International Series in
Operations Research & Management Science

Percy H. Brill

Level Crossing Methods in Stochastic Models

Second Edition

Springer

International Series in Operations Research & Management Science

Volume 250

More information about this series at http://www.springer.com/series/6161

Percy H. Brill

Level Crossing Methods in Stochastic Models

Second Edition

Springer

Percy H. Brill
Departments of Management Science
and Mathematics and Statistics
University of Windsor
Windsor, ON
Canada

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-50330-1 ISBN 978-3-319-50332-5 (eBook)
DOI 10.1007/978-3-319-50332-5

Library of Congress Control Number: 2016961662

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Out of great complexity comes great simplicity.

Adapted From Winston Churchill

Life is like riding a bicycle. To keep your balance, you must keep moving.

Albert Einstein

To the memory of my parents

Preface

The Second Edition adds 17 new figures bringing the total number to 124; many of the original figures have been revised. It also includes a new section with a novel derivation of the Beneš series for M/G/1 queues. It describes new results on the service time for three M/G/1 queueing models with bounded workload. It analyzes new applications of queues where zero-wait customers get exceptional service, including several examples on M/G/1 queues, and a new section on G/M/1 queues. There are two other important new sections: on the level crossing derivation of the finite time-*t* probability distributions of excess, age, and total life in renewal theory; and on a level crossing analysis of a risk model in insurance. These two new sections make it convenient to split the original Chap. 10 into two chapters. The new Chap. 10 is on renewal theory. The new first section of Chap. 11 is on a risk model. More explicit use is made of the renewal reward theorem throughout. Many technical and editorial changes have been made to facilitate readability. The Second Edition has many more specific citations and references to enhance efficiency of perusing the book: 145 references (102 in the First Edition); 125 works listed in the Partial Bibliography (56 in the First Edition). The Index has been expanded to expedite locating information in the book.

Windsor, ON, Canada | Percy H. Brill
October 2016

From Preface of First Edition

From 1972 to 1974, I was working on a Ph.D. thesis entitled *Multiple Server Queues with Service Time Depending on Waiting Time*. The method of analysis was the embedded Markov chain technique, described in the papers [109] and [103]. My analysis involved lengthy, tedious derivations of systems of integral equations for the probability density function (PDF) of the waiting time. After pondering for many months whether there might be a faster, easier way to derive the integral equations, I finally discovered the basic theorems for such a method in August 1974. The theorems establish a connection between sample-path level crossing rates of the virtual wait process and the PDF of the waiting time. This connection was not found anywhere else in the literature at the time. I immediately developed a comprehensive new methodology for deriving the integral equations based on these theorems, and called it *system point theory.* (Subsequently, it was called *system point method*, or *system point level crossing method:* abbreviated *SPLC* or simply *LC*.) I rewrote the entire Ph.D. thesis from November 1974 to March 1975, using LC to reach solutions. The new thesis was called *System Point Theory in Exponential Queues*. On June 12, 1975, I presented an invited talk on the new methodology at the Fifth Conference on Stochastic Processes and their Applications at the University of Maryland. Many queueing theorists were present. Ever since, LC has become an increasingly used technique for analyzing a large class of stochastic models. LC can be used to derive integro-differential equations for transient distributions, or integral equations for steady-state distributions.

This monograph elucidates LC for obtaining probability distributions of state variables in a variety of stochastic models. Most of the analyses are for steady-state distributions. However, some results for transient distributions are also given. The book is intended for research- and applications-oriented workers in operations research, management science, engineering, probability and statistics, actuarial science, mathematics, and the natural sciences.

To date, many researchers have applied LC. Applications have appeared in refereed journals, conference proceedings, technical reports, master's and Ph.D. theses, and in chapters and sections of books, worldwide.

One reason for this great interest and consequent proliferation of publications is that LC is very intuitive. Furthermore, it leads to exact analytical solutions. An LC analysis starts with a typical sample path of a stochastic process. A sample path (sample function, realization, and tracing) can be thought of dynamically. That is, the path evolves in the state space over time, governed by the probability laws of the model.

The LC method focuses on *time rates* at which a sample path exits and enters certain measurable state-space sets. Level-crossing theorems equate these transition rates to simple algebraic expressions of the PDF and/or CDF (cumulative distribution function) of the state variable. In a steady-state analysis, the algebraic expressions often appear in separate terms of Volterra integral equations of the second kind with parameter. Thus, "physical" sample-path transition rates are in one-to-one correspondence with terms of the integral equations. The integral equations themselves are constructed by applying rate conservation laws, e.g., rate balance. The upshot is that we can write down the integral equations "by inspection," upon observing the sample-path structure of a model!

The integral equations are solved simultaneously with a normalizing condition, which specifies that all probabilities sum to 1. The system of equations is solved for the PDF and/or CDF of the state variable. We may use analytical, numerical, algorithmic, simulation, or approximation techniques to solve the system of equations. We can derive operating characteristics of the model using the solution and/or LC concepts.

It is axiomatic that one can reach solutions for mathematical models by applying alternative techniques. My own experience, and that of many other researchers, has demonstrated that LC often leads quickly and easily to solutions. It provides useful intuition about the model dynamics. This is due to the perspective taken: geometric sample-path structure; rate conservation laws; connection to concepts of natural science such as physics. LC may free the analyst from lengthy derivations of a system of model equations. Thus, it facilitates focusing on model dynamics and on operating characteristics. An LC analysis quite often suggests new creative approaches for studying a model.

I hope that readers will find the monograph interesting, and useful for research. The concepts, techniques, examples, applications, and theoretical results in this book may suggest potentially new theory and new applications.

Windsor, ON, Canada Percy H. Brill

Acknowledgements

I am grateful to Dr. Mei Ling Huang, Department of Mathematics and Statistics, Brock University; Dr. Myron Hlynka, Department of Mathematics and Statistics, University of Windsor; Dr. Stanley L. Sclove, Department of Information and Decision Sciences, University of Illinois at Chicago, for many technical discussions, comments and suggestions, which have helped to improve the presentation.

I thank Dr. Fred S. Hiilier, Stanford University, for encouraging me to write the Second Edition. Nicholas Philipson, Neil Levine, and Matthew Amboy, of Springer, New York, have been helpful and encouraging toward completing the book. I thank faculty and staff in the dean's office of the Odette School of Business, and staff in the Research Grants accounting office, of the University of Windsor, for facilitating the research process. The long-term research support via a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) has provided me with the practical means necessary to work on the Second Edition.

From Acknowledgements of First Edition

Dr. Julian Keilson was the external examiner for my Ph.D. thesis. He independently initiated a formal written invitation from the conference chair Dr. R. Syski, for me to make the first international conference presentation on the level crossing methodology at the Fifth Conference on Stochastic Processes and their Applications in June, 1975.

Contents

Chapter 1
Origin of Level Crossing Method

1.1 Prologue

This chapter presents a condensed version of the original development of the *level crossing method (**LC**)* for deriving probability distributions of state variables in stochastic models. I developed LC concomitantly with the more general *system point method (SP method)*. Thus LC is actually an essential component of the system point method. A more precise nomenclature for the overall technique is the *system point level crossing method (**SPLC**)*. In this monograph, for simplicity we usually use the abbreviation LC to refer to the overall procedure.

The LC technique was developed during the period January 1974–August 1974, while I was working on my Ph.D. thesis of a different topic, namely *Multiple Server Queues with Service Time Depending on Waiting Time*. The work, since May 1972, involved analyzing the steady-state distribution of customer wait in an M/M/c queue with service time depending on wait before service. This had been my original Ph.D. thesis topic, suggested by my supervisor M.J.M. Posner. The goal had been to generalize to multiple server M/M/c queues, the (then) forthcoming paper by M.J.M. Posner [117] on single server M/M/1 queues, *using the method of embedded Markov chains*, a purely algebraic technique [103]. That analysis formulates Lindley recursions for successive customer waits and their probability distributions [109]. The approach utilizes inequalities, conditional probabilities, and the law of total probability. It also involves multiple integration, transformation of variable, differentiation, and limit operations.

The embedded Markov-chain analysis can be tedious and time consuming, especially for complex models. I worked for several thousand hours (about

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_1

50 h per week) developing, simplifying and solving "fifty-page" integral equations on computer paper (the old kind 10"×17") over a two year period. Much experience and many observations had shown that the analyses of different model variants ultimately converge to a common stage. Each analysis culminates with its own system of Volterra integral equations of the second kind with parameter, for the steady-state pdf (probability density function) of the customer wait. At this point, all of my analyses were purely algebraic.

While I pondered the complexity and tediousness of various embedded Markov-chain analyses, the question gradually surfaced as to whether there may exist an alternative, more intuitive technique for deriving the integral equation(s) for the pdf. After considerable analysis, finally in August 1974, I discovered the basic LC theorems and the related methodology.

For queues, the LC method *starts* by constructing a *typical* sample path (sample function, realization, trajectory, tracing, orbit) of the virtual wait process (see Sect. 2.2). Then we apply LC theorems. These theorems utilize sample-path structure to write an integral equation, or system of integral equations, for the steady-state pdf, ***by inspection!*** The LC approach can save an enormous amount of time when analyzing complex stochastic models. LC provides a common systematic procedure for studying a wide variety of stochastic models. It focuses attention on sample paths. Therefore it often leads to new insights into the model dynamics and its subtleties. In complex models, construction of a sample path may itself be a challenge. However, the benefit of this construction is that it often leads to a deeper understanding of the model.

In order to construct the integral equation(s), the LC method employs a one-to-one correspondence between: (1) the set of algebraic terms in the integral equation(s) for the pdf, and (2) a set of mutually exclusive and exhaustive sample-path transitions relative to state-space levels or state-space sets (see Sects. 2.4.3 and 2.4.4).

My original thesis using embedded Markov chains and Lindley recursions was published in two working papers [48] and [49]. Immediately after my discovery of LC, I completely rewrote my original Ph.D. thesis using SPLC, from November 1974 to March 1975. Two additional chapters were added as well, outlining the System-Point/Level-Crossing theory (Chap. 2 in [11]), and examples from the literature (Chap. 8 in [11]). The results using LC corroborated the original results in [48] and [49], and also pointed to a whole new array of possible applications. The new thesis was called *System Point Theory in Exponential Queues* [11] . This led to the subsequent publications [50–53].

Two years later in 1976, J.W. Cohen [61] discussed the same level crossing ideas, as related to regenerative processes (e.g., [134]), and followed that work with [62]. The resulting regenerative-process connection to LC is useful for obtaining certain steady-state (limiting) results in a variety of stochastic models.

The following abridged version of my development of LC in 1974 deals with the single server queue. (This preserves the main ideas, which originally evolved from analyzing complex M/M/c queues with service time depending on waiting time.) For background, we first derive an integral equation based on the *classical* algebraic Lindley-recursion/embedded-Markov-chain method for GI/G/1 and M/G/1 queues. This was the method used to analyze my original Ph.D. thesis topic. (Due to multiple servers, that derivation started with a more general Lindley recursion, employed in working papers [48] and [49]. It ended with a *system* of Volterra integral equations for the steady-state pdf of wait, and its solution.) Then we outline the original development of the LC method, and use it to derive the same integral equation—*by inspection*.

1.2 Lindley Recursion for GI/G/1 Wait

Let W_n, S_n, T_{n+1} denote respectively the waiting time of customer n before service, the service time of customer n, and the time interval $\tau_{n+1} - \tau_n$ between the arrival instants (epochs) τ_n, τ_{n+1} of customers n and $n + 1$ at the system, $n = 1, 2, \ldots$. The well-known Lindley recursion for the waiting time is

$$W_{n+1} = \max\{W_n + S_n - T_{n+1}, 0\}, \; n = 1, 2, \ldots. \tag{1.1}$$

Referring to Fig. 1.1, we have the following inequalities. For fixed $x \geq 0$,

$$\left.\begin{array}{l} 0 \leq W_{n+1} \leq x \\ \Longleftrightarrow W_n + S_n - T_{n+1} \leq x \\ \Longleftrightarrow y + S_n - z \leq x \\ \Longleftrightarrow S_n \leq x + z - y, \end{array}\right\} \tag{1.2}$$

given $W_n = y$ and $T_{n+1} = z$. (Symbol "$\Longleftrightarrow$" is equivalent to "if and only if" or "iff".)

Let $P(A)$ denote the probability of an event A.

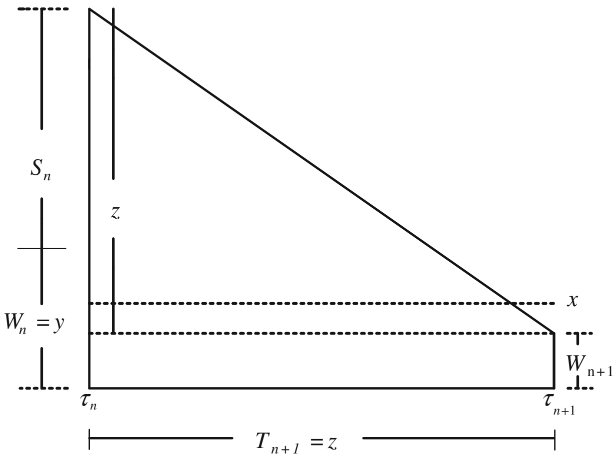


Fig. 1.1 Lindley recursion for $W_n \to W_{n+1}$ geometrically

Definition 1.1 For $n = 1, 2, ...$

$$\left.\begin{array}{l} F_n(x) = P(W_n \leq x), x \geq 0, \\ f_n(x) = \frac{d}{dx}F_n(x), x > 0, \text{ where the derivative exists,} \\ P_n(0) = F_n(0), \\ B(y) = P(S_n \leq y),\ y \geq 0, n = 1, 2, ..., \\ \overline{B}(y) = 1 - B(y), y \geq 0. \end{array}\right\} \tag{1.3}$$

$F_n(\cdot)$ is the cdf of W_n; $f_n(\cdot)$ is the pdf on the positive part of W_n; $F_n(\infty) = P_n(0) + \int_{x=0}^{\infty} f_n(x)dx = 1, n = 1, 2, ...$ Assume that the parameters of the queue are such that the steady state cdf $F(\cdot)$ and mixed pdf $\{P_0, f(x)\}_{x>0}$ of the wait exist, and $\lim_{n\to\infty} F_n(x) = F(x), x \geq 0$, $\lim_{n\to\infty} P_n(0) = P_0$, $\lim_{n\to\infty} f_n(x) = f(x), x > 0$. We define $f(\cdot)$ to be right continuous. Thus $f(x^+) = f(x), x > 0$. For consistency, we extend the domain of $f(\cdot)$ to include $x = 0$, and define $f(0^+) = f(0)$. Note that $f(0)$ adds zero probability to P_0.

1.3 Integral Equation for M/G/1 PDF of Wait via Lindley Recursion

Assume that the arrival process is Poisson at rate λ, and that the random variables $\cup_{n\in\mathbb{N}^+}\{S_n, T_{n+1}\}$ are mutually independent (where $\mathbb{N}^+ = \{1, 2, ...\}$). For this model assume S_n, W_n are independent of each other, $n = 1, 2, ...$ The classical approach applies inequalities (1.2) to derive an integral equation, which expresses $F_{n+1}(\cdot)$ in terms of $P_n(0)$ and $f_n(\cdot)$. The notation $P(A|B)$ denotes the conditional probability of event A given that event B occurs. Conditioning on T_{n+1} and then on W_n, gives for $x \geq 0$,

$$\begin{aligned} &F_{n+1}(x) \\ &= \int_{z=0}^{\infty} P(W_n + S_n - z \leq x | T_{n+1} = z)\lambda e^{-\lambda z} dz \\ &= \int_{z=0}^{\infty}\int_{y=0^-}^{x+z} P(S_n \leq x + z - y | W_n = y, T_{n+1} = z) f_n(y)\lambda e^{-\lambda z} dydz. \end{aligned}$$

where 0^- emphasizes that the probability of the atom $\{0\}$ is included in the integral. (See Sect. 2.4.9 for a definition of atom.) Substituting from (1.3),we obtain for $x \geq 0$,

$$\begin{aligned} F_{n+1}(x) &= \int_{z=0}^{\infty}\int_{y=0^-}^{x+z} B(x + z - y) f_n(y)\lambda e^{-\lambda z} dydz \\ &= P_n(0)\int_{z=0}^{\infty} B(x + z)\lambda e^{-\lambda z} dz \\ &+ \int_{z=0}^{\infty}\int_{y=0}^{x+z} B(x + z - y) f_n(y)\lambda e^{-\lambda z} dydz. \end{aligned} \tag{1.4}$$

The transformation $w = x + z$ in (1.4) gives, for $x \geq 0$,

$$\begin{aligned} F_{n+1}(x) &= P_n(0)\int_{w=x}^{\infty} B(w)\lambda e^{-\lambda(w-x)} dw \\ &+ \int_{w=x}^{\infty}\int_{y=0}^{w} B(w - y) f_n(y)\lambda e^{-\lambda(w-x)} dydw. \end{aligned} \tag{1.5}$$

For $x > 0$, take $\frac{d}{dx}$ on both sides of (1.5) wherever it exists. Then

$$\begin{aligned} f_{n+1}(x) &= \lambda F_{n+1}(x) - \lambda P_n(0)B(x) \\ &- \lambda \int_{y=0}^{x} B(x - y) f_n(y) dy, \; x > 0. \end{aligned} \tag{1.6}$$

By definition,

$$F_{n+1}(x) = P_{n+1}(0) + \int_{y=0}^{x} f_{n+1}(y) dy, \; x \geq 0.$$

Substituting into (1.6) yields

$$f_{n+1}(x) = \lambda\left(P_{n+1}(0) + \int_{y=0}^{\infty} f_{n+1}(y)dy\right) - \lambda P_n(0)B(x)$$
$$- \lambda\int_{y=0}^{x} B(x-y)f_n(y)dy,\, x > 0,$$

which simplifies to

$$f_{n+1}(x) = \lambda\left[P_{n+1}(0) - P_n(0)B(x)\right] + \lambda\int_{y=0}^{x}(f_{n+1}(y) - B(x-y)f_n(y))dy,\, x > 0. \tag{1.7}$$

In (1.7), letting $n \to \infty$ gives the desired integral equation for the steady state pdf, namely,

$$f(x) = \lambda P_0\overline{B}(x) + \lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy,\, x > 0. \tag{1.8}$$

The normalizing condition that all probabilities sum to 1, is

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{1.9}$$

Equations (1.8) and (1.9) are then solved simultaneously to obtain the steady-state pdf of wait $\{P_0, f(x)\}_{x>0}$. Steady-state operating characteristics can be computed from $\{P_0, f(x)\}_{x>0}$: the cdf $F(\cdot)$; the Laplace-Stieltjes transform $\int_{y=0^-}^{\infty} e^{-sy}dF(y), s > 0$; the expected values of the waiting time, system time and number in the system, by applying Little's theorem $\boldsymbol{L = \lambda \cdot W}$ [110]; quantiles of $F(\cdot)$; the probability mass function (pmf) of the number in the system, by conditioning on the wait and applying the PASTA principle [145]; etc.

When analyzing more general stochastic models, e.g., state-dependent models, we obtain variations and generalizations of integral equation (1.8). Examples are: single and multiple server queues with service time or arrival rate depending on current workload; inventories where demand rate or demand size depends on current inventory level (stock on hand); general storage systems where input size depends on current content; risk reserve systems in Insurance where claim size depends on current risk reserve; systems in the physical and natural sciences with state-dependent parameters.

The algebraic steps in (1.1)–(1.8) illustrate the *classical* approach. In complex state-dependent models, the classical approach begins with more general Lindley recursions than (1.1). Then, significantly more algebra is typically required to derive an integral equation, or system of integral equations, for the steady state pdf of the state variable, e.g., [48] and [49].

It is important to note that the classical method based on Lindley recursions is very useful both theoretically and computationally, for studying the waiting time in queues, and state variables in many stochastic models.

The following question gradually evolved while continuing to derive integral equations for the pdf in complex state-dependent M/M/c models using the classical method [48] and [49]. Does there exist an alternative way to derive integral equation (1.8), and analogous integral equations in complex state-dependent models, which: (a) bypasses starting from (1.1); (b) reduces the amount of accompanying algebra? The goal was to derive equations like (1.8) in a manner similar to the well-known, intuitively appealing *rate into state = rate out of state* balance equations for the state probabilities in discrete-state, continuous-time Markov chains, e.g., [125]. Persevering with this idea, while continuing to apply the classical method, ultimately led to the SPLC methodology. The developmental process is outlined in Sects. 1.4–1.7.

1.4 Observations and Questions

The following elementary observations and simple questions considered together, lead to a very powerful approach for analyzing stochastic models.

1. For each $x \geq 0$, the cdf $F(x) \in [0, 1]$. Thus $F(x)$ is a dimensionless quantity. It is a real number without associated units.
2. For each $x > 0$, the pdf $f(x)$ $\left(= \frac{dF(x)}{dx}\right)$, has dimension $1/\,[Time]$. This follows because Δx has the same dimension as x, namely $[Time]$ because $f(x)$ is the pdf of waiting time, in the defining formula $f(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x}$.
3. In integral equation (1.8), the dimension of both left and right hand sides is $\left[\frac{1}{Time}\right]$. Note that the parameter λ has dimension $\left[\frac{1}{Time}\right]$.
4. A number having dimension $[1/Time]$ is the measure of a *rate,* a notion from Physics.
5. Each side of integral equation (1.8) is the measure of some unknown (in 1974) *rate.*
6. In integral equation (1.8), the left hand side $f(x)$ and the right hand side $\lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y) dy$, may represent two different rates, which have the same value.

7. **Question:** What *geometric* or *physical rate*, if any, does $f(x)$ measure?
8. **Question:** What *geometric* or *physical rate*, if any, does $\lambda P_0\overline{B}(x) + \lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy$ measure?

Remark 1.1 The classical approach, starting from Lindley recursions , is a completely algebraic technique. There was no inkling whatsoever in 1974, of the geometric picture that was about to emerge, as described in Sect. 1.5.

1.5 Further Properties of Integral Equation for PDF of Waiting Time in M/G/1

To answer Questions 7 and 8 of Sect. 1.4, we study (1.8) further. Let $x \downarrow 0$ on both sides of (1.8). This yields

$$f(0^+) = \lambda P_0. \tag{1.10}$$

Observation: *For the M/G/1 queue in steady state (equilibrium), consider two discrete states that the system may present from the viewpoint of an arriving customer:* $\{0\}$*: no wait;* $\{1\}$*: wait. Over time the system alternates between presenting states* $\{0\}$ *and* $\{1\}$ *to the arrival stream. An arrival waits: (a) zero time iff the server is idle at the arrival instant; (b) a positive time iff the server is busy at the arrival instant. Thus we may equivalently redefine the states from the viewpoint of the system (or server) as:* $\{0\}$ *: idle;* $\{1\}$*: busy.*

The rate at which busy periods start is λP_0, due to Poisson arrivals, and the *rate out of state* $\{0\} = \lambda P_0$, as in continuous-time, discrete-state Markov chains. By conservation of rates out of and into $\{0\}$, the rate at which busy periods end must also be λP_0. Furthermore, a connection is made to integral equation (1.8) via the relation (1.10), $f(0^+) = \lambda P_0$.

Figure 1.2 depicts the motion between the two states $\{0\}, \{1\}$. The sojourn times of visits to $\{0\}$ are i.i.d. (independently and identically distributed) random variables distributed as an idle period. An idle period is exponentially distributed with mean $1/\lambda$. The sojourn times of visits to $\{1\}$ are i.i.d. random variables distributed as a busy period. A sample path corresponds to that of a two-state alternating renewal process. It is a special case of a Markov renewal process or semi-Markov process with 2×2 Markov transition matrix $\|P_{ij}\|$ where $P_{01} = P_{10} = 1$ (see pp. 457–460 in [125]). Let $\{A(t)\}_{t\geq 0}$ denote this two-state process, where $A(t) = 0$ if $t \in$ idle period and $A(t) = 1$ if $t \in$ busy period. A sample path consists of alternating horizontal, right-continuous line segments (Fig. 1.2).

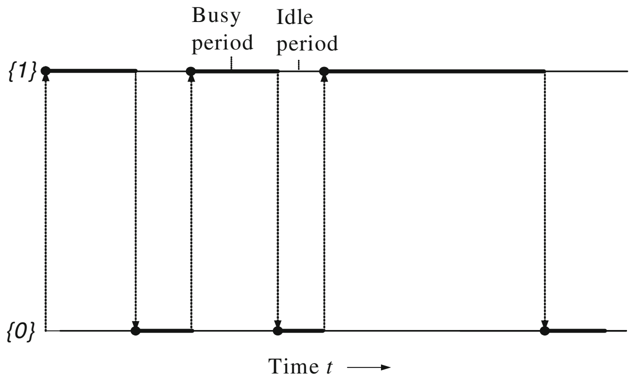


Fig. 1.2 Sample path of alternating renewal process $\{A(t)\}_{t\geq 0}$

1.5.1 Connection with Virtual Wait Process

Reflecting on the structure of the alternating renewal process $\{A(t)\}_{t\geq 0}$, led to the recognition of a close correspondence with the well-known ***virtual wait*** process (thanks to [140] which the author had become aware of in 1964). The virtual wait represents how long a customer that arrives at time t must wait to start service (same as the workload at time t in standard M/G/1). For the standard M/G/1 queue, the virtual wait $\{W(t)\}_{t\geq 0}$ is a continuous-time, continuous-state process with state space $[0, \infty)$. Sample paths of $\{W(t)\}_{t\geq 0}$ are real-valued, non-negative, right-continuous functions on $[0, \infty)$. Characteristically,

$$\frac{dW(t)}{dt} = \begin{cases} -1 \text{ if } W(t) > 0, \\ \ 0 \text{ if } W(t) = 0 \end{cases}$$

(Fig. 1.3). Jumps occur at Poisson rate λ. Jump sizes are distributed as the service time. Table 1.1 shows the correspondence between $\{A(t)\}_{t\geq 0}$ and $\{W(t)\}_{t\geq 0}$.

Observation: *Sample paths of* $\{W(t)\}_{t\geq 0}$ *are strictly positive during busy periods and equal to zero during idle periods. Sample paths of* $\{A(t)\}_{t\geq 0}$ *have the same property, if we make the correspondence as in Table 1.1.*

Interestingly, for the process $\{A(t)\}_{t\geq 0}$ state $\{1\}$ can be viewed as a "black box" containing all possible busy periods. Whenever the sample path enters $\{1\}$, a random busy period is generated.

Observation: *For the M/G/1 queue, it is well known that the cdf and pdf of* $W(t)$ *as* $t \to \infty$ *are respectively equal to the cdf and pdf of* W_n *as* $n \to \infty$, *provided the limits exist (*e.g., [140] *).*

The above discussion leads to the following observation.

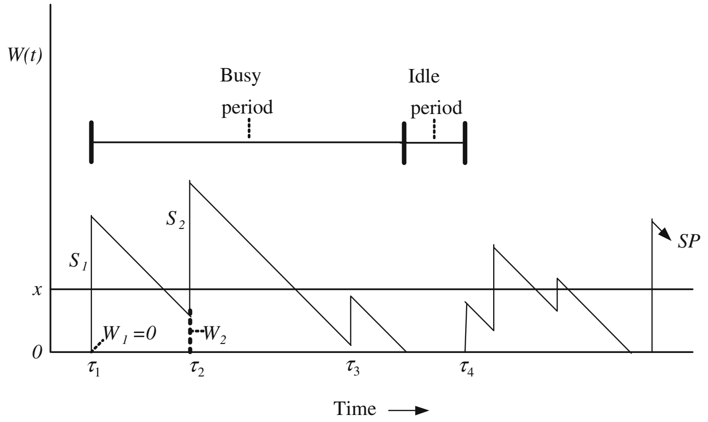


Fig. 1.3 Sample path of virtual wait $\{W(t)\}_{t \geq 0}$ in M/G/1 showing: actual waits $\{W_n\}_{n=1,2,\ldots}$; busy and idle periods; system point SP; fixed level x

Table 1.1 Correspondence Between {A(t)} and {W(t)}

Time $t \geq 0$	$A(t)$	$W(t)$
$t \in$ idle period	0	0
$t \in$ busy period	1	$\in (0, \infty)$

Observation: *$f(0^+)$ = rate at which a typical sample path of $\{W(t)\}_{t \geq 0}$ hits level 0 from above at a 45° angle (Fig. 1.3). Hits of level 0 from above occur at the ends of busy periods.*

Insight: *Shift attention to sample paths of the virtual wait $\{W(t)\}_{t \geq 0}$! Focus on the geometry of a typical sample path of $\{W(t)\}_{t \geq 0}$!*

The last observation provides an alternative interpretation of Eq. (1.10). In complex systems, this observation may lead to extra conditions to help solve for unknown constants of integration arising in the solution of a system of integral (or differential) equations. More importantly, the foregoing considerations suggest the key question and conjecture given in Sect. 1.5.2.

1.5.2 Looking Upward from Level Zero

Key Question: *At what rate does a typical sample path of $\{W(t)\}_{t \geq 0}$ hit any state-space level $x \geq 0$, from above?*

To answer the key question, imagine, temporarily, that the M/G/1 model under consideration were really an M/M/1 model with service rate μ. The jump sizes of the virtual wait process (Fig. 1.3) would then be *exponentially* distributed with mean $1/\mu$. Fix level $x > 0$ in the state space. Consider a jump that starts at some level $y < x$ and ends above x. By the memoryless property of the exponential distribution, the excess jump above x would have the same distribution as the total service time. That is, $P(S_n > x-y+z|S_n > x-y) = e^{-\mu z}$, $n = 1, 2, ...$, independent of y and x . This implies that each sojourn time of a sample path above *every* $x \geq 0$, would be statistically identical to a busy period, *independent* of x! Thus, the picture during sojourns above level x would be a probabilistic replica of Fig. 1.3 during busy periods above level 0. However, the sojourns at or below level x, would be of different durations depending on x (see Sect. 3.4.16). This leads to the key conjecture. Recall that $f(0) = f(0^+)$.

Key Conjecture: *For each* $x \geq 0, f(x)$ *is the rate at which a sample path of* $\{W(t)\}$ *hits level* x *from above.*

The key conjecture generalizes the last observation in Sect. 1.5.1. The conjecture is readily confirmed mathematically for M/M/1, M/G/1 and GI/G/1 queues. Furthermore, in many *general*, state-dependent stochastic models, analogous results connect sample-path hits of a state-space level, and the pdf of the state variable at that level. The notions of sample-path *smooth hits* of a level and *jumps across* a level, naturally suggest the concept of ***level crossings***: in particular, ***downcrossings*** and ***upcrossings***.

Remark 1.2 Various areas of real analysis and stochastic processes utilize level crossing concepts. In stochastic processes most work deals with level crossings of processes having continuous sample paths. Prior to 1974, level crossings had not been directly connected with, or used to obtain integral equations to solve for probability distributions of state random variables. The level crossing method is particularly useful in continuous-time continuous-state stochastic models where sample paths have discontinuous jumps. It is also applicable to processes with strictly continuous sample paths, as in a dam with alternating influx and efflux (see Sect. 11.8).

In this monograph, we shall regularly use the terms: *level crossing, downcrossing, upcrossing*. In the present context it is sufficient to use their intuitive meaning, as in Fig. 1.4. Roughly speaking, for the virtual wait of a standard M/G/1 queue, a downcrossing of a level at instant t_0 is a smooth or left-continuous hit of that level from above at t_0. An upcrossing at instant t_0 is made by a jump, which starts below, and ends above the level, at t_0. These concepts are discussed more precisely in Chap. 2.

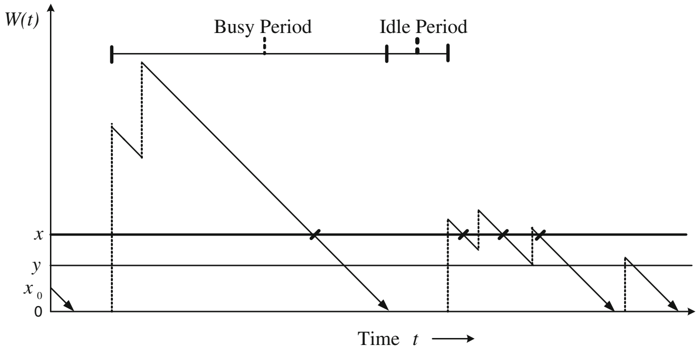


Fig. 1.4 Sample path of $\{W(t)\}_{t\geq 0}$ indicating downcrossings of level x and hits of level 0

1.5.3 Integral Equation in Light of the Sample Path

Consider the left side of (1.8). For each $x > 0, f(x)$ is equal to the sample-path *downcrossing rate* of level x. That is, $f(x)$ corresponds to the rate of a particular type of sample-path transition across level x. This correspondence has an intuitive appeal, which we now explore further.

Question: *Does the right side of Eq. (1.8), $\lambda P_0\overline{B}(x)+\lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy$, correspond to the rate of a particular type of sample-path transition across level x?*

The last question prompts consideration of the idea *conservation law*, or *principle of set balance (rate balance across a boundary separating two disjoint state-space sets)*. Referring to $W(t), t \geq 0$, (Fig. 1.4), let $x_0 = W(0)$, and fix $x > 0$. The state space is $\boldsymbol{S} = [0, \infty) = [0, x] \cup (x, \infty)$ (union of two disjoint sets). The long-run sample-path *exit* and *entrance* rates of state-space set (x, ∞) are equal, independent of the initial state x_0. Exits and entrances of (x, ∞) alternate in time, and correspond to sample-path downcrossings and upcrossings of level x, respectively. Set balance (rate balance across level x) suggests interpreting $\lambda P_0\overline{B}(x) + \lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy$ as the sample-path *upcrossing* rate of level x. We now show that this interpretation is correct.

For the process $\{W(t)\}_{t\geq 0}$ the following property holds for a sample-path jump starting at level $y < x$ (Fig. 1.4):

$$\begin{aligned}P(\text{end of jump} > x \mid \text{start of jump} = y < x)\\ &= P(\text{service time} > x - y)\\ &= \overline{B}(x-y). \qquad (1.11)\end{aligned}$$

If a jump upcrosses x, it starts either at level 0 or at a level $y \in (0, x)$. Setting $y = 0$ in (1.11) shows that the rate of upcrossings of x, starting at level 0, is $\lambda P_0 \overline{B}(x)$. The rate of jumps starting in a small interval $(y, y + dy)$ is $\lambda f(y)dy$. From (1.11), the rate of upcrossings of x, starting in $(0, x)$ is $\lambda \int_{y=0}^{x} \overline{B}(x - y) f(y)dy$. Thus, there is a one-to-one correspondence between the set of three algebraic terms of (1.8) and the set of three mutually exclusive and exhaustive sample-path crossing rates of level x (see Fig. 1.6).

1.6 Basic Level Crossing Theorem for M/G/1

The foregoing notions lead to the basic level crossing theorem for the steady-state pdf of wait in the standard M/G/1 queue, namely Theorem 1.1 below. Assume $\lambda E(S) < 1$, where λ is the arrival rate and $E(S)$ is the expected value of the service time. Consider a sample path of the virtual wait process.

1.6.1 Downcrossing and Upcrossing Rates

For fixed $x > 0$ and fixed $t > 0$, let $\mathcal{D}_t(x)$, $\mathcal{U}_t(x)$ denote the *number* of down- and upcrossings of level x during $(0, t)$, respectively. The *average* rates of down- and upcrossings during $(0, t)$ are $\frac{\mathcal{D}_t(x)}{t}$ and $\frac{\mathcal{U}_t(x)}{t}$, respectively. Let $E(X)$ denote the expected value of a generic random variable X. The average rates of the expected number of down- and upcrossings during $(0, t)$ are $\frac{E(\mathcal{D}_t(x))}{t}$ and $\frac{E(\mathcal{U}_t(x))}{t}$, respectively. Note that the singleton discrete state $\{0\}$ is an atom having steady-state probability $P_0 > 0$. (See Sect. 2.4.9 for a definition of atom.) Let $\mathcal{O}_t(\{0\})$ denote the number of exits out of, and $\mathcal{I}_t(\{0\})$ the number of entrances into, the discrete state $\{0\}$ during $(0, t)$. Here, an intuitive notion of *exit* and *entrance* suffices. Define $\mathcal{D}_t(0) = \mathcal{I}_t(\{0\})$ and $\lim_{t\to\infty} \frac{\mathcal{D}_t(0)}{t} = \lim_{t\to\infty} \frac{\mathcal{I}_t(\{0\})}{t}$. These notions are specified further in Chap. 2.

Theorem 1.1 (P.H. Brill, 1974) *For the virtual wait process* $\{W(t)\}_{t\geq 0}$ *in the stable M/G/1 queue* ($\rho = \lambda E(S) < 1$)

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x), x \geq 0, \tag{1.12}$$

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \underset{a.s.}{=} f(x), x \geq 0, \tag{1.13}$$

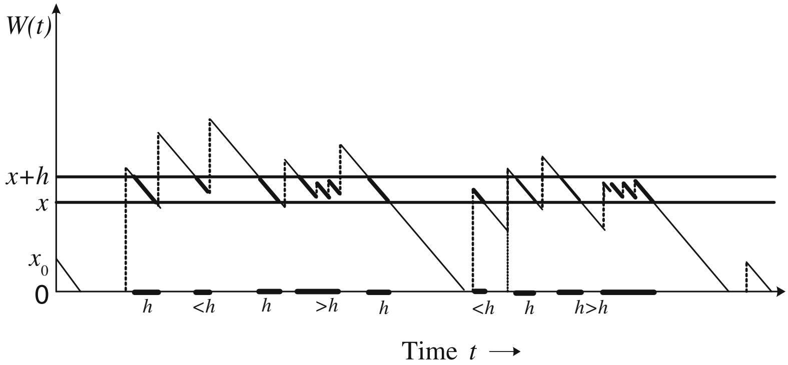


Fig. 1.5 Sample path of virtual wait in M/G/1 queue. Shows levels x and $x+h$ and various sojourn times in interval $(x, x+h)$, used in proof of Theorem 1.1

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y)dy, x > 0, \tag{1.14}$$

$$\lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} \underset{a.s.}{=} \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y)dy, x > 0, \tag{1.15}$$

where '$\underset{a.s.}{=}$' means 'equal almost surely' or 'with probability 1'.

Proof (Note: A different proof is given in Corollary 3.6 of Theorem 3.4 in Sect. 3.2.7 for the *transient* pdf of $\{W(t)\}_{t\ge 0}$. Also see [50], [11], [52])

Here we demonstrate some of the simple intuition underlying the SPLC methodology. Consider a sample path of the virtual wait on $(0, t)$, i.e., $\{W(s)\}_{0<s<t}$ and fix levels $x > 0$ and $x+h$, where $h > 0$ is small (Fig. 1.5).

Sojourns in $(x, x+h)$ after downcrossing of level $x+h$

The contribution to the *expected sojourn time* in $(x, x+h)$ due to sojourn times $= h$ is

$$h \cdot e^{-\lambda h} = h \cdot [1 - \lambda h + o(h)] = h + o(h)$$

due to the memoryless property of exponential interarrivals. The sample path spends a shorter or longer time than h in $(x, x+h)$ with probability *less than* $[\lambda h + o(h)]$ because in either case a jump must occur before the sample path exits $(x, x+h)$. That jump ends either above or below $x+h$. Thus the contribution to the *expected* sojourn time in $(x, x+h)$ is *less than* $h \cdot [\lambda h + o(h)] = o(h)$.

Sojourns in $(x, x+h)$ after upcrossings of x that end in $(x, x+h)$

The probability that a jump upcrosses level x and ends in $(x, x + h)$ is $b(x - y)h$ for some $y \in [0, x)$. We assume $b(\cdot)$ is bounded. The contribution to the expected value of the subsequent sojourn in $(x, x + h)$ is: (a) less than $h \cdot b(x - y)h = o(h)$ if there is no arrival before the sample path falls to level x, or (b) less than $A \cdot b(x-y)h\,[\lambda h + o(h)] = o(h)$ if the sojourn time in $(x, x + h)$ is extended due to an arrival, where $0 < A < t$. (We use the fact that

$$\left(\frac{A \cdot b(x-y)h\,[\lambda h + o(h)]}{t}\right) < b(x-y)h\,[\lambda h + o(h)] = o(h)$$

below to get the left side of (1.16).)

Thus, the contributions to the expected sojourn times in $(x, x + h)$, that are $\neq h$ is $o(h)$. Hence during the interval $(0, t)$, $t > 0$, the expected *total time* spent in $(x, x+h)$ is $E(\mathcal{D}_{t-h}(x+h)) \cdot [h + o(h)]$. The *limiting expected proportion* of time that the sample path spends in $(x, x + h)$ is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_{t-h}(x+h)) \cdot [h + o(h)]}{t} = F(x+h) - F(x), \tag{1.16}$$

by the definition of $F(x)$, $x > 0$. Dividing both sides of (1.16) by h and letting $h \downarrow 0$ gives

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x^+))}{t} = f(x),$$

since $E(\mathcal{D}_{t^-}(x^+)) = E(\mathcal{D}_t(x^+))$. At downcrossing instants the sample path is continuous from the left, so that $E(\mathcal{D}_t(x^+)) = E(\mathcal{D}_t(x))$. Hence

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x),\ \ x > 0.$$

This proves (1.12). The counting process $\{\mathcal{D}_t(x)\}_{t\geq 0}$ is a renewal process due to Poisson arrivals. Therefore $\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} \underset{a.s.}{=} \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t}$ [125], and (1.13) follows.

An intuitive proof of (1.14) and (1.15) follows from the discussion in Sect. 1.5.3. ■

Corollary 1.1 *For the M/G/1 queue in equilibrium*

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t(0))}{t} = f(0^+) = f(0) = \lambda P_0, \tag{1.17}$$

$$\lim_{t\to\infty}\frac{\mathcal{D}_t(0)}{t} \underset{(a.s.)}{=} f(0^+) = f(0) = \lambda P_0. \tag{1.18}$$

Proof Let $x \downarrow 0$ in (1.12)–(1.15) and apply (1.10) ■

Note that (1.17) and (1.18) equate the sample-path: (1) downcrossing rate of level 0 (= *entrance rate into* discrete state $\{0\}$); (2) *exit rate from* $\{0\}$; (3) the pdf $f(0)$ at level 0. An important notion is that sample-path rates into and out of a *discrete* state, are equal to a particular value of the pdf of a *continuous* random variable! This relation connects $\{0\}$, which is a boundary of $[0, \infty)$, to the state-space interval of continuous states $(0, \infty)$.

Formula (1.19) below, gives the principle of *set balance* for a state-space set (x, ∞), $x > 0$, in terms of rate balance across level x.

Principle of Rate Balance for Level x

This is the same as *set balance* for (x, ∞), i.e.,

$$\left.\begin{aligned} \lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} &= \lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t}, x > 0, \ (a.s), \\ \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t} &= \lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}, x > 0. \end{aligned}\right\} \tag{1.19}$$

Formula (1.19) means that for each x, the (long-run) SP down- and upcrossing rates of level x are equal, independent of the initial state $W(0) = x_0$ at $t = 0$. Rate balance for levels (set balance for sets having the level as a boundary) is discussed more fully in Chap. 2, Sect. 2.4.7.

1.7 Integral Equation for M/G/1 Waiting Time Using Level Crossing Method

We now derive (1.8) using LC, by applying Theorem 1.1 and rate balance (1.19). *Start with a typical sample path of* $\{W(t)\}_{t\geq 0}$. Fix level $x > 0$. Apply the one-to-one correspondence that exists between the set of mutually exclusive and exhaustive sample-path crossing rates of level x, and the set of algebraic expressions which contain $\{P_0, f(x)\}_{x>0}$. Write integral equation (1.8) as a *rate-balance equation* using (1.19), *by inspection of the sample path* (Fig. 1.6)! Note that starting from level 0, the upcrossing rate of level $x > 0$ is

$$\lim_{t\to\infty}\frac{E(\mathcal{O}_t(\{0\}))}{t}\cdot \overline{B}(x) = \lambda P_0 \overline{B}(x).$$

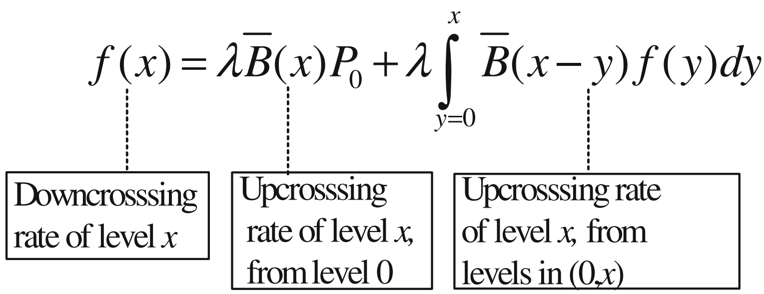


Fig. 1.6 One-to-one correspondence between virtual-wait sample-path rates of crossing level x and terms of integral equation (1.8) for $f(x)$

Summary of Steps in LC Derivation of Integral Equation (1.8)

1. *Construct a sample path of* $\{W(t)\}_{t\geq 0}$ *(Fig.* 1.4*).*
2. *Substitute from (*1.12*) and (*1.14*) term by term into (*1.19*).*
3. *Write integral equation (*1.8*) (Fig.* 1.6*).*

This completes an abbreviated outline of the original development in 1974, of the system-point level-crossing method for analyzing stochastic models. Note that the SPLC method was developed first for multiple-server M/M/c queues and then for M/G/1 queues immediately after. For the M/M/c case, the method of *sheets* (or *pages*), was developed simultaneously, since it is a vital component of the SPLC method (see Sect. 4.5.7, and Refs. [11] and [52]).

Chapter 2
Sample Path and System Point

2.1 Introduction

When applying the system point level crossing method (abbreviated SPLC, or briefly LC) to obtain probability distributions of state variables in stochastic models, *intuitive* notions of sample-path *transitions* often suffice (see Fig. 1.6). In some models, however, *more rigorous* notions of such transitions are useful for applying SPLC. This chapter provides definitions and examples which enhance intuitive background about sample paths and SP motion in state spaces which are subsets of $\mathbb{R}$ (the set of real numbers). Pertinent sample-path transitions include *exits* and *entrances* of state-space sets; *boundary crossings*, *downcrossings*, *upcrossings*, *tangents, hits*, *egresses* and *pass-bys* of state-space levels. We discuss the useful principles of *rate balance* across state-space levels, and of *set balance* between disjoint state-space sets. These transitions and principles are relatively easy to discern from a "typical" sample path of the state variable of the model. They relate directly to the probability distribution of the state variable (as in Theorem 1.1). Thus they help us develop model equations for the probability distribution by inspection of the sample path (as in Fig. 1.6). Section 2.5 summarizes geometrically 35 types of sample-path and SP transitions with respect to state-space levels.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_2

2.2 State Space and Sample Paths

2.2.1 Sample Paths

For each stochastic model considered, we will tacitly assume the existence of a basic probability space $(\boldsymbol{\Omega}, \mathcal{F}, \boldsymbol{P})$, where $\boldsymbol{\Omega}$ is the set of all possible outcomes of the associated random experiment, $\mathcal{F}$ is a σ-field of events, and $\boldsymbol{P}$ is a probability measure on $\mathcal{F}$. The first step of the LC method is to construct a "typical" sample path of the state variable of interest over Time, from the sequences of random variables and the rules defining the model (e.g., Fig. 1.4). Examples of such sequences occur in: queues—inter-arrival and service times; inventories—inter-demand times and demand sizes; dams—inter-input times and input sizes; actuarial models—inter-claim times and claim amounts; pharmacokinetics—inter-dose times and dose amounts.

A "typical" sample path is one which is "reasonable" or "not rare". Examples are sample paths of: the virtual wait in an M/G/1 queue where the averages of the alternating busy and idle periods converge to their respective theoretical values (Fig. 2.1); the net inventory of an $\langle s, S \rangle$ inventory system with product decay where the average of the replenishment cycles converge to the theoretical value (Fig. 2.2).

We assume that: the state space $\boldsymbol{S}$ consists of continuous and/or discrete states (atoms); *the number of atoms is finite in finite state-space intervals*. For example, the state space of the virtual wait process in M/G/1 queues has exactly one atom, at level 0 (Fig. 2.1).

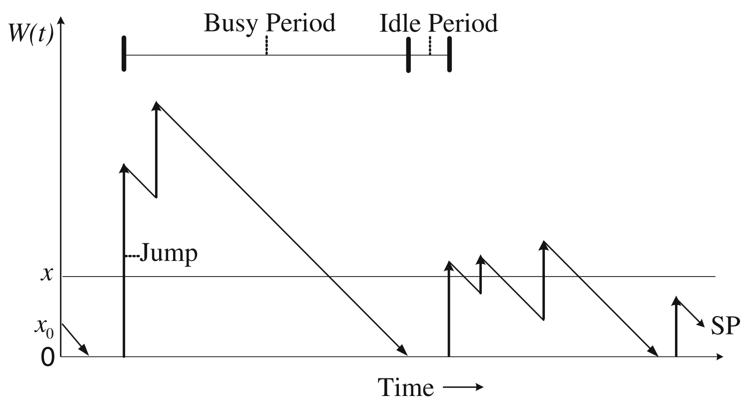


Fig. 2.1 Sample path of virtual wait in $M/G/1$ queue. Emphasizes SP jumps and hits of level 0 from above

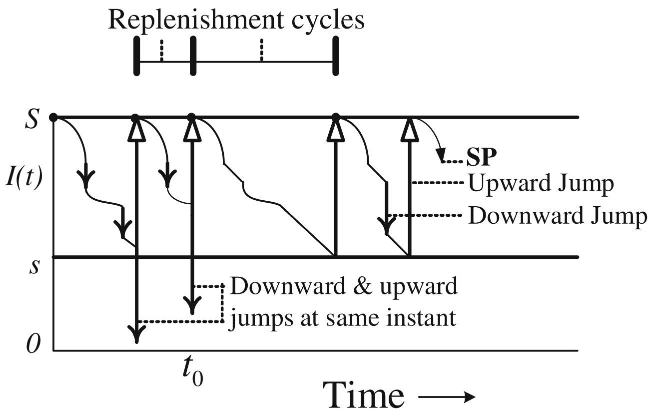


Fig. 2.2 Sample path of net inventory in $\langle s, S\rangle$ model with product decay, no lead time

Let $\boldsymbol{T}$ denote the continuous parameter set of the model. Usually, $\boldsymbol{T} = \{t \mid t \in [0, \infty)\}$, the *time axis*.

We employ the following "working" definition of a sample path. It is sufficiently general for a large class of stochastic models in OR (Operations Research), and applies to the models analyzed in this monograph.

Definition 2.1 A **sample path** is a bounded real-valued or vector-valued, right-continuous function $X(t)$, $t \in \boldsymbol{T}$, with domain $\boldsymbol{T}$ and range a subset of the state space $\boldsymbol{S}$. Left limits exist for all $t > 0$. All sample-path discontinuities are jumps. During arbitrary finite time intervals, the number of jumps and number of relative extrema (excluding "trivial" extrema during sojourns in discrete states) are finite with finite expectations.

Sample paths are also called sample functions, realizations, trajectories, tracings, orbits. (Reference [69] contains a comprehensive treatment of sample functions.) A sample path is a possible outcome $\omega \in \boldsymbol{\Omega}$ of the background random experiment associated with a model; each ω corresponds uniquely with a function $\omega_t : T \to \mathbb{R}$. For fixed t, $X(t)$ is a random variable with domain $\boldsymbol{\Omega}$ and *range* a subset of $\boldsymbol{S}$. If $\boldsymbol{S} \subseteq \mathbb{R}$ then $X(t)$ has cdf $P(X(t) \leq x) = P(\{\omega\,|\omega_t \leq x|\}, x \in \boldsymbol{S}$, and pdf $\frac{d}{dx}P(X(t) \leq x)$, where the derivative exists. If t_0 is not an instant of jump, then $X(\cdot)$ is continuous at t_0. If t_0 *is* an instant of "jump" then generally $X(t_0^-) \neq X(t_0)$; however possibly $X(t_0^-) = X(t_0)$ if there is a double jump at t_0 (see Example 2.3 and Fig. 2.6, on $\langle s, S\rangle$ inventory with no decay). (**Note**: The symbol $\langle s, S\rangle$ denotes an inventory system with a reorder point s and order-up-to level S.) For many models in this monograph, sample paths are piecewise continuous

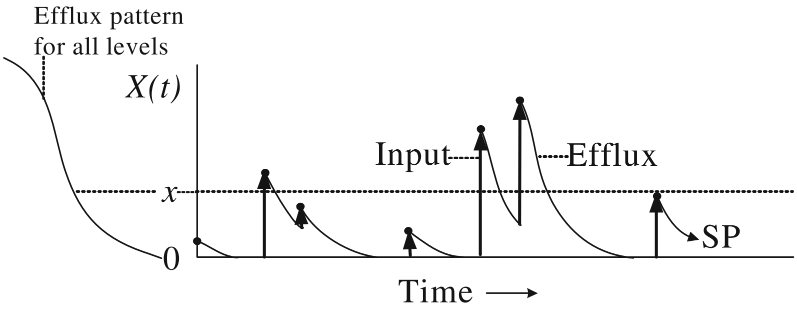


Fig. 2.3 Sample path of content in dam with general release pattern (efflux). Emphasizes sample-path jumps, right continuity, and slopes for all positive levels

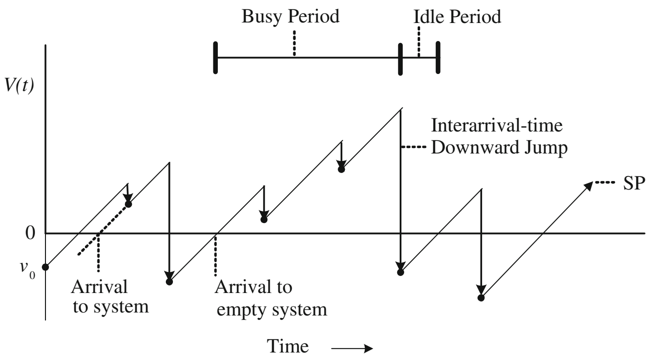


Fig. 2.4 Sample path of extended age process in G/M/1 queue. $V(t)$ = time customer in service at t has been in system, if $V(t) > 0$. $-V(t)$ = remaining time at t until next arrival, if $V(t) < 0$ (see Sect. 5.1). Emphasizes downward jumps and right continuity of sample path

and differentiable between jumps; the slope at a fixed state-space level x is independent of t (see Figs. 2.2, 2.3, 2.4 and 2.5).

2.2.2 Sample-Path Properties and Jumps

Proposition 2.1 The total number of sample-path jumps and/or relative extrema for a model with time domain $\boldsymbol{T} = [0, \infty)$ is a countable set (a.s.) (almost surely, with probability 1 with respect to $(\boldsymbol{\Omega}, \mathcal{F}, \boldsymbol{P})$).

Proof The time domain $\boldsymbol{T} = \bigcup_{n=1}^{\infty} [n-1, n)$ is a countable union of disjoint finite intervals. Each interval contains at most a finite number of sample-path

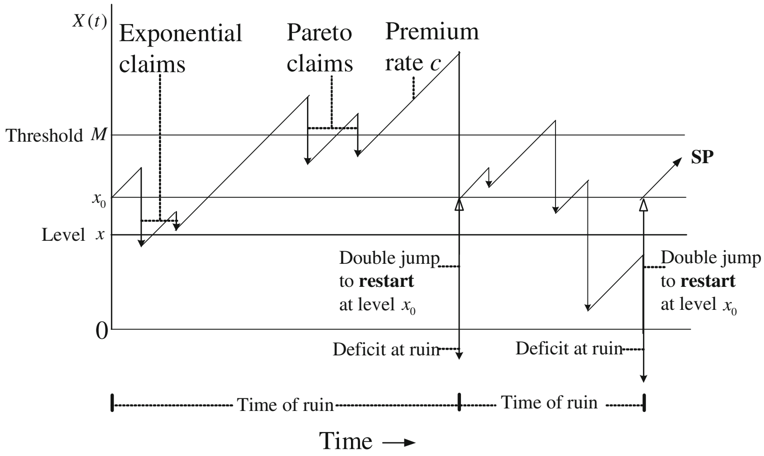


Fig. 2.5 Sample path of surplus (risk reserve) in ruin-like model in Insurance. Shows first 2 regenerative cycles (see [54]). In this example claim size distribution depends on surplus at claim instant: exponential below M; Pareto above M

jumps and/or relative extrema (a.s.), by Definition 2.1. A countable union of countable sets is countable (e.g., [6, 56]). ■

For continuous time models, in practice it is possible to observe a jump of a state variable $X(\cdot)$ at any instant $t \in T$. For some models it is possible that two "jumps", e.g., downward and upward, occur at the same instant, which can affect the physical behavior of the system (see Examples 2.1, 2.2 and Remark 2.1). We discuss such multiple jumps in Sect. 2.3.

Example 2.1 Consider a typical sample path of the stock on hand (net inventory) $\{I(t)\}_{t\geq 0}$ in a **continuous review** $\langle s, S\rangle$ **inventory** with a single product, random demand stream, random demand sizes, no lead time, and continuous product decay (Fig. 2.2). The "wide-sense" state space is $(-\infty, S] \subseteq \mathbb{R}$ (see Sect. 2.3.1). The reorder point is s, and the order-up-to level is S, $0 \leq s < S$. Arrivals of demands generate downward jumps. The OR analyst **prescribes** an **order to replenish the stock up to S, corresponding to sample-path** upward jumps, in response to the following signals: (a) a demand causes a downward jump that ends **at or below** s, (b) the stock on hand **decays continuously from above** into level s. All upward jumps start below or at level s, and end at level S. At each instant when signal (a) is detected, **both** downward and upward jumps occur, resulting in a **net** upward jump of the sample path.

In Fig. 2.2 the sample path consists of piecewise deterministic, continuous curved segments with negative slope. The relative extrema (peaks and troughs) are contained within the state space interval $[s, S]$. The jumps are not part of the sample path per se. Nevertheless, the jumps are observable, and they determine the structure of the sample path over Time. Downward jumps that signal instants to place an order, occur at the same instants as the corresponding prescribed upward jumps, which replenish the stock to level S. There are no discrete states in $\boldsymbol{S}$.

Remark 2.1 We briefly discuss examples of sample paths in some models of applied probability. For: (1) the **virtual wait process** in a G/G/1 queue, upward jumps occur at arrival instants (Fig. 2.1); (2) the **content in a dam** with instantaneous inputs, upward jumps occurring at input instants, and general pattern of efflux (Fig. 2.3); (3) the **extended age process** in a G/M/1 queue (the time that a customer in service has been in the system, or, negative of remaining time until the next arrival to an empty system) (see [19]), downward jumps equal in distribution to the inter-arrival times, occur at departure instants (i.e., completions of service) (Fig. 2.4); (4) the **risk reserve process** (**surplus**) in actuarial science (Insurance), downward jumps occur at claim instants, and upward jumps may occur at ruin instants (epochs) when using a recent LC analysis (see Fig. 2.5; [54]); Sect. 11.1 below); (5) the **concentration of a drug** in a one-compartment pharmacokinetic model with multiple bolus dosing, upward jumps occur at instants of dosing (Fig. 11.2).

2.3 System Point Motion and Jumps

Empirically, sample paths may be viewed as evolving in Time. We assume that sample paths evolve in the same direction in Time: $-\infty \to +\infty$, or *left* $\to$ *right* in diagrams, unless otherwise specified.

We call the *leading point* of an evolving sample path the *System Point* (*abbreviated* SP) (Figs. 2.2, 2.3, 2.4, 2.5 and 2.6). The author coined the term *system point* in this context because the leading *point* $(t, X(t))$ of a sample path at instant t contains relevant information about the *system*, due to the history up to t. In Markov processes, the information conveyed by $(t, X(t))$ is sufficient to statistically predict the future evolution after t, independent of the history up to t. (Interestingly, the SP can also be considered as the *trailing* point of the future sample path!)

Assume the state space is $\mathbf{S} \subseteq \mathbb{R}$. If $t_0 > 0$ is a point of continuity of the sample path $X(t), t \geq 0$ then the SP moves in the direction defined by $\frac{d}{dt}X(t)|_{t=t_0}$ where the derivative exists (see Sect. 6.2 and Fig. 6.1). If the derivative does not exist, then the SP changes direction to the slope of the right derivative $\frac{d}{dt}X(t)|_{t\downarrow t_0}$. If t_0 is an instant of jump, then the SP moves

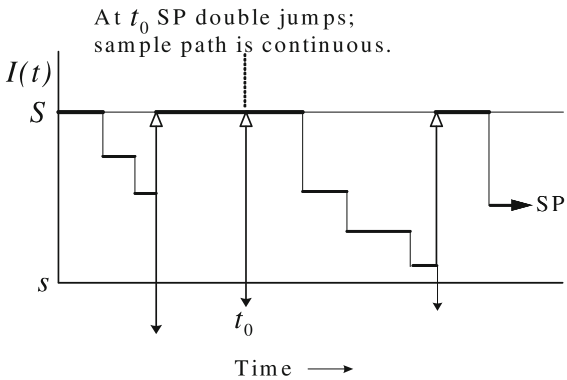


Fig. 2.6 Sample path, and SP (system point) motion in $\langle s, S\rangle$ inventory with no product decay. At t_0 sample path is continuous. At other jump instants, sample path is discontinuous

in the direction *up*, *down* or *both (i.e., double jumps), not in the direction of Time* (Figs. 2.2, 2.5, 2.6). Technically, the SP jumps are not part of the sample path, which is by definition a mathematical function on $\boldsymbol{T} = [0, \infty)$.

Given the time, placement, size and direction of a jump, the immediately following continuous sample-path segment is determined assuming its slope at all levels is known, e.g., is independent of time (see Fig. 2.3). The end-point levels of two continuous segments which are contiguous and separated by a jump, determine the *net* sample-path jump (see Fig. 2.2). While tracing the continuous segments of a sample path the SP is imagined to have a "finite velocity" in Time. When the SP jumps, it has "infinite speed". The completely-evolved sample path is an inert graph on $[0, \infty)$. On the other hand, the SP is like the moving tip of a stylus that is imagined to plot the graph (see Sect. 4.5.2).

We quantify the foregoing description of jumps further. Consider a typical sample path $X(t), t \geq 0$. Let t_0 be an instant of jump (possibly a double jump). Let u_{t_0} and d_{t_0} denote respectively the sizes of the upward and downward jumps at t_0, where $u_{t_0} \geq 0$, $d_{t_0} \geq 0$ and $u_{t_0}^2 + d_{t_0}^2 > 0$. At least one of u_{t_0}, d_{t_0} is positive. The resultant position of the SP at t_0 due to the jump(s) is the sample-path value

$$X(t_0) = X(t_0^-) + u_{t_0} - d_{t_0} = \lim_{t \downarrow t_0} X(t).$$

The *net* sample-path jump is $X(t_0) - X(t_0^-)$, which may be positive, negative or zero (Figs. 2.2, 2.5, 2.6). If $X(t_0) - X(t_0^-) = 0$ the sample path is continuous at t_0 although the SP makes jumps at t_0. In this case, real changes occur in the associated physical system at t_0 (e.g., an order is placed and received), but the sample path is continuous (Fig. 2.6).

Example 2.2 Consider the stock on hand in a continuous review $\langle s, S\rangle$ **inventory** with a single product, $0 \le s < S < \infty$. Assume a random demand stream, random demand sizes, no lead time, and **continuous product decay**. Downward jumps occur at demand instants (Fig. 2.2). Let t_0 be a demand instant when $I(t_0^-) = y, s < y < S$, and let the demand be $d_{t_0} > y - s$. The would-be resulting stock on hand at t_0, due to the demand, is $y - d_{t_0} < s$ so that $y - d_{t_0} \notin (s, S]$. The unsatisfied demand (deficit) at t_0 is $s - y + d_{t_0}$. The downward jump that ends below s is a signal at t_0 to place an order and replenish the stock up to level S immediately (no lead time). There is a **prescribed** upward jump of stock at t_0 equal to $u_{t_0} = S - y + d_{t_0}$. This satisfies the deficit and restores the stock up to S, i.e., $I(t_0) = S$. The SP makes **both** downward and upward jumps at t_0 (**double jumps**) which **result in a single net sample-path upward jump** of size $S - y = u_{t_0} - d_{t_0}$. The SP upward jump is a **prescribed** or **policy** jump. In summary, **at t_0 the SP makes two jumps in opposite directions**; **the sample path has one net upward jump.**

Example 2.3 In Example 2.2 with **no product decay** (see Fig. 2.6), at instant t_0 the SP makes two jumps of equal size in opposite directions: one downward (demand) and one upward (replenishment). The sample path makes a **net jump of size 0**. That is, $I(t_0^-) = S$, a demand of size $d_{t_0} > S - s$ occurs, impelling the SP below level s. The order-up-to level S policy prescribes an immediate upward jump at t_0 of size $u_{t_0} = d_{t_0}$, ending at level S. Thus $I(t_0) = I(t_0^-) = S$, which implies the sample path is **continuous** at t_0 by right continuity.

Remark 2.2 Examples 2.2 and 2.3 show that at least two SP jumps can occur at an instant, not in Time (orthogonal to the time axis). SP **multiple jumps** are compatible with a common assumption about the occurrence of multiple events in continuous time stochastic models. That is, **multiple probabilistic events** cannot occur *at the same instant in Time. The latter assumption technically applies to sample paths and to the sequences of random variables defining the model. It usually prohibits* to occur more than one: arrival; service completion in a queue; demand of an inventory; input to a dam; insurance claim; etc., at a particular instant in Time. The LC method is based on the count or rate of SP transitions across levels or state-space boundaries, or between state-space sets. The transitions may be due to SP jumps, or sample-path smooth descents or ascents to a level. In the $\langle s, S\rangle$ inventory, LC counts jumps due to chance events like demands, and jumps due to prescribed responses like replenishments, when computing rates of crossing state-space levels.

Remark 2.3 Consider Example 2.1. An observer of the **sample path** who is aware of the $\langle s, S\rangle$ policy, and observes $X(t_0) = S$ just after a jump at t_0, cannot determine whether the SP has made a double jump, or a single upward jump at t_0 (Fig. 2.2). That is, the jump resulting in $X(t_0) = S$ could have been caused by a signal of either type (a) **left-continuous descent** to level s, or type (b) **demand** that causes the SP to jump below s (see [43]).

Remark 2.4 In Example 2.3 (Fig. 2.6), assume an observer of the sample path knows the policy is $\langle s, S \rangle$ and that $X(t_0) = S$. The observer cannot distinguish t_0 as being an instant of SP activity (placing an order and replenishing to S) or an instant of SP inactivity, since the SP motion is "invisible" at t_0. Knowledge of the sample path is sufficient to derive **probability distributions** of the net inventory. However, knowledge of the SP motion over Time **including** SP motion at instants of jump, implies knowledge of the sample path structure as well as of the ongoing actual activity of the real-world inventory.

Remark 2.5 In a **real-world** $\langle s, S \rangle$ **inventory** the signal to place an order **precedes** the replenishment. The signal is the **cause** of the replenishment under the $\langle s, S \rangle$ policy. There is a time order of the signal and the replenishment, even if the separation is only a nanosecond or picosecond. In the **mathematical model**, both signal and replenishment occur at the same instant.

To summarize, the *SP* moves in Time during sample-path *continuous* segments, and moves in the state space orthogonal to Time at instants of jump. (It is a coincidence that '*sample path*' and '*system point*' have the same initials.)

2.3.1 State Space in the Wide Sense

Examples 2.1–2.3 pose a conceptual question. The state space is usually considered to be the interval $(s, S]$, since all *states* describing net inventory are subsets of $(s, S]$. However, observations of the jumps are required in order to construct the sample path. Jumps may end or start in the interval $(-\infty, S]$ (some outside $(s, S]$). Hence it is crucial to be able to observe SP motion in $(-\infty, S] = (-\infty, s] \cup (s, S]$. In these examples we call $(-\infty, S]$ the ***state space in the wide sense***.

In this monograph the term *state space* will mean ***state space in the wide sense***, unless otherwise specified. The state space in the wide sense contains the range of all possible SP jumps.

2.4 State Space a Subset of $\mathbb{R}$

In the models discussed so far, the state space is an interval subset of $\mathbb{R}$. Most models in this monograph fit this category. We now discuss such models more formally, to develop intuitive background about the SPLC methodology. (Models with more general state spaces are discussed in Chaps. 4 and 7.)

Consider a stochastic model having state-space interval $\boldsymbol{S} \subseteq \mathbb{R}$. Set $\boldsymbol{S}$ is often an *infinite* interval. In Table 2.1 **S** is in the 'wide sense' (see Sect. 2.3.1). To illustrate, in Example 2.2, using the state space in the wide sense has no effect on the values of the cdf and pdf of stock on hand; all probability is supported on $(s, S]$.

2.4.1 Picture of a Subset of S Over Time

Let $X(t)$, $t \geq 0$, denote a sample path in the two-dimensional Cartesian product space $\boldsymbol{T} \times \boldsymbol{S} = [0, \infty) \times \boldsymbol{S} = \{(t, x)|t \in [0, \infty), x \in \boldsymbol{S}\}$. Let $\boldsymbol{A} \subset \boldsymbol{S}$ be a proper *interval* subset of **S**. Then $\boldsymbol{T} \times \boldsymbol{A} \subset [0, \infty) \times \boldsymbol{S}$. For set $\boldsymbol{A}$, let $\boldsymbol{A}^c$ denote the complement in $\boldsymbol{S}$, $\partial \boldsymbol{A}$ the boundary, $\boldsymbol{A}^o$ the interior set, and $\boldsymbol{A}^e$ the exterior set (= interior of $\boldsymbol{A}^c$) (see, e.g., [136], P. 5ff). Set $\boldsymbol{A}^c$ may be the union of two disjoint intervals. The two-dimensional Cartesian product sets $\boldsymbol{T} \times \boldsymbol{A}$, $\boldsymbol{T} \times \boldsymbol{A}^c$, $\boldsymbol{T} \times \partial \boldsymbol{A}$, $\boldsymbol{T} \times \boldsymbol{A}^o$, $\boldsymbol{T} \times \boldsymbol{A}^e$, are proper subsets of $\boldsymbol{T} \times \boldsymbol{S}$ (Fig. 2.7). Sets $\boldsymbol{A}^o$, $\delta \boldsymbol{A}$, $\boldsymbol{A}^e$ are mutually disjoint, as are their respective Cartesian products with $\boldsymbol{T}$.

Table 2.1 Examples of models with state space a subset of $\mathbb{R}$ and corresponding figures

Stochastic model	State space **S**	See figures
Virtual wait in M/G/1 queue	$[0, \infty)$	Figure 2.1
Extended age process in G/M/1 queue	$(-\infty, +\infty)$	Figure 2.4
Stock on hand in $\langle s, S \rangle$ inventory	$(-\infty, S]$, $S > 0$	Figures 2.2, 2.6
Content in dam	$[0, \infty)$	Figure 2.3
Surplus in risk model	$(-\infty, +\infty)$	Figure 2.5

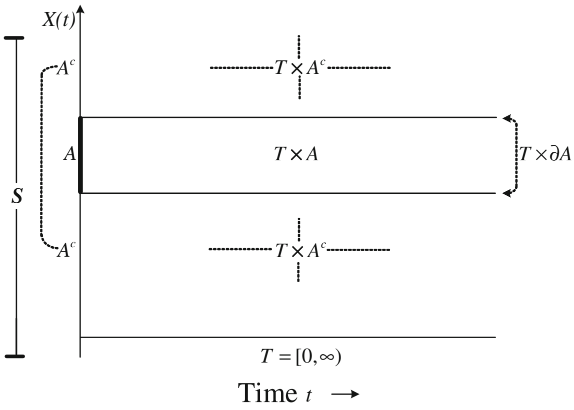


Fig. 2.7 Sets $\boldsymbol{A}$, $\boldsymbol{A}^c$ and $\mathbf{T} \times \boldsymbol{A}$, $\boldsymbol{T} \times \boldsymbol{A}^{\mathbf{c}}$, $\boldsymbol{T} \times \partial \boldsymbol{A}$ when interval $\boldsymbol{A} \subset \boldsymbol{S} \subseteq \mathbb{R}$

2.4.2 Levels in S

A level-x **contour** in $\boldsymbol{T} \times \boldsymbol{S}$ is defined as a straight line $\boldsymbol{T} \times \{x\}$, $x \in \boldsymbol{S}$. We call this line ***level*** x for brevity. Level x is parallel to the t axis at a distance $|x|$ from the line $\boldsymbol{T} \times \{0\}$ (t axis). When we discuss transitions of a sample path (or motion of the SP) with respect to level x, we mean with respect to the level-x contour in $\boldsymbol{T} \times \boldsymbol{S}$. We also use the terminology ***level*** x ***in the state space***, or ***level*** $x \in \boldsymbol{S}$, since these expressions convey the idea intuitively. (Technically level $x \in \boldsymbol{S}$ is the projection of the level-x contour in $\boldsymbol{T} \times \boldsymbol{S}$ onto $\boldsymbol{S}$.)

We consider arbitrary levels $x \in \boldsymbol{S}$, *because of the basic level crossing theorem for M/G/1* (Theorem 1.1, Sect. 1.6). That theorem connects the probability distribution of the state variable (virtual wait) at an arbitrary value x, with sample-path and SP down- and upcrossing rates across level x (e.g., Fig. 1.6). Similarly, we can gain empirical background about an arbitrary stochastic model by observing the motion of the SP and the structure of a sample path in $\boldsymbol{T} \times \boldsymbol{S}$.

For fixed $x \in \boldsymbol{S}$, we may observe rates of SP or sample-path down- and upcrossings, and of tangents (see Definition 2.2). Applying level crossing theorems like Theorem 1.1, greatly facilitates the derivation of integral equations or algebraic equations for the pdf and/or cdf of the state variable, which are valid *for each* $x \in \boldsymbol{S}$ (as in Fig. 1.6). We can solve such equations by analytical, numerical, or simulation techniques.

2.4.3 Sample Path Transitions

Consider an interval $\boldsymbol{A} \subseteq \boldsymbol{S}$ (Fig. 2.7). We first define the following transitions: *sample-path exits, entrances, tangents, boundary crossings and level crossings* with respect to $\boldsymbol{T} \times \boldsymbol{A}$, using elementary topological concepts of real analysis (see, e.g., [6, 56, 127], or [136]). Let $X(t), t \geq 0$ denote a sample path. Assume $t_0 > 0$ is an instant of either sample-path continuity or jump. Let $X(t_0^-) = \lim_{t \uparrow t_0} X(t)$ (left limit at t_0 exists, Definition 2.1).

Definition 2.2
Sample-path Exit:
$X(\cdot)$ **exits** $\boldsymbol{A}$ at instant t_0 if $\exists\, \varepsilon > 0 \ni X(t) \in \boldsymbol{T} \times \boldsymbol{A}$ for $t \in (t_0 - \varepsilon, t_0)$ and $X(t) \in \boldsymbol{T} \times \boldsymbol{A}^c$ for $t \in (t_0, t_0 + \varepsilon)$.
Sample-path Entrance:
$X(\cdot)$ **enters** $\boldsymbol{A}$ at instant t_0 if $X(\cdot)$ exits $\boldsymbol{A}^c$ at t_0.
Sample-path Interior Tangent:
$X(\cdot)$ is **interior tangent** to $\boldsymbol{A}$ at instant t_0 if $\exists\, \varepsilon > 0 \ni X(t) \in \boldsymbol{T} \times \boldsymbol{A}^o$, $t \in (t_0 - \varepsilon, t_0 + \varepsilon) \setminus \{t_0\}$ and either $X(t_0^-) \in \boldsymbol{T} \times \partial \boldsymbol{A}$, or $X(t_0) \in \boldsymbol{T} \times \partial \boldsymbol{A}$.
Sample-path Exterior Tangent:
$X(\cdot)$ is **exterior tangent** to $\boldsymbol{A}$ at instant t_0 if $X(\cdot)$ is interior tangent to $\boldsymbol{A}^c$ at instant t_0.
Sample-path Boundary Crossing:
$X(\cdot)$ crosses boundary $\partial \boldsymbol{A}$ at instant t_0 if $X(\cdot)$ exits $\boldsymbol{A}^o$ and enters $\boldsymbol{A}^e$ (denoted $A^o \to A^e$), or $X(\cdot)$ exits $\boldsymbol{A}^e$ and enters $\boldsymbol{A}^o$ (denoted $A^e \to A^o$) at t_0.

In Definition 2.3 below fix $x \in \boldsymbol{S}$ and let $\boldsymbol{A} = (x, \infty) \cap \boldsymbol{S}$. Then

$$\boldsymbol{A}^o = (x, \infty) \cap \boldsymbol{S} = \boldsymbol{A}, \qquad \boldsymbol{A}^e = (-\infty, x) \cap \boldsymbol{S}, \qquad \partial \boldsymbol{A} = \{x\} \cap \boldsymbol{S}.$$

Definition 2.3
Sample-path Downcrossing:
$X(\cdot)$ **downcrosses** level x at instant t_0 if $X(\cdot)$ crosses the boundary $\boldsymbol{T} \times \{x\}$ (denoted $\boldsymbol{T} \times \boldsymbol{A}^o \to \boldsymbol{T} \times \boldsymbol{A}^e$) at t_0. Equivalently, $X(\cdot)$ exits $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ and enters $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$ at t_0.
Sample-path upcrossing:
$X(\cdot)$ **upcrosses** level x at instant t_0 if $X(\cdot)$ crosses the boundary $T \times \{x\}$ ($\boldsymbol{T} \times A^e \to \boldsymbol{T} \times A^o$) at t_0. Equivalently, $X(\cdot)$ exits $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$ and enters $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ at t_0.

Definitions 2.2 and 2.3 apply at an instant of either sample-path continuity or sample-path jump. At instants of *continuity* of $X(t), t \geq 0$, system point (SP) transitions are defined identically as for sample-path transitions in Definitions 2.2 and 2.3. However, at instants of *jump* (sample-path discontinuity), SP transitions are defined differently, since the SP moves orthogonally to the direction of Time; either upward or downward in $T \times \boldsymbol{S}$.

2.4.4 System Point (SP) Transitions

We now define SP transitions with respect to $\boldsymbol{T} \times \boldsymbol{A}$ *at an instant of jump*, say t_0. Assume that at t_0 the SP makes a *single* jump either of size d_{t_0} downward or size u_{t_0} upward. Let

$$\theta = \begin{cases} 1 \text{ if direction of the jump is } \textit{downward,} \\ 0 \text{ if direction of the jump is } \textit{upward.} \end{cases}$$

Definition 2.4
SP Exit at Instant of Jump:
The SP **exits** $\boldsymbol{A}$ at t_0 if $X(t_0^-) \in \boldsymbol{T} \times \boldsymbol{A}$ and

$$X(t_0) = X(t_0^-) - \theta d_{t_0} + (1-\theta)u_{t_0} \in \boldsymbol{T} \times \boldsymbol{A}^c.$$

SP Entrance at Instant of Jump:
The SP **enters** $\boldsymbol{A}$ at t_0 if the SP exits $\boldsymbol{T} \times \boldsymbol{A}^c$ at t_0.
SP Boundary Crossing:
The SP makes a **boundary crossing** of $\partial \boldsymbol{A}$ at t_0 if $X(t_0^-) \in \boldsymbol{T} \times \boldsymbol{A}^o$ and

$$X(t_0) = X(t_0^-) - \theta d_{t_0} + (1-\theta)u_0 \in \boldsymbol{T} \times \boldsymbol{A}^e(\boldsymbol{A}^o \to \boldsymbol{A}^e),$$

or **if** $X(t_0^-) \in \boldsymbol{T} \times A^e$ and

$$X(t_0) = X(t_0^-) - \theta d_{t_0} + (1-\theta)u_{t_0} \in \boldsymbol{T} \times \boldsymbol{A}^o(\boldsymbol{A}^e \to \boldsymbol{A}^o).$$

For Definition 2.5 below we fix $x \in \boldsymbol{S}$. Then $T \times \{x\}$ is a boundary of both $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ and $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$.

Definition 2.5
SP Downcrossing:

The SP **downcrosses** level x at t_0 if the SP crosses boundary $\boldsymbol{T} \times \{x\}$ from $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ to $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$ at t_0 $((x, \infty) \to (-\infty, x))$.
SP Upcrossing:
The SP **upcrosses** level x at t_0 if the SP crosses boundary $\boldsymbol{T} \times \{x\}$ from $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$ to $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ at t_0 $((-\infty, x) \to (x, \infty))$.

To motivate Definition 2.6 below, consider Example 2.1 (see Fig. 2.2). Assume a demand for the product is placed at t_0^- causing the SP to jump downward to level $z < s$. An order is placed, the SP immediately rebounds with a *prescribed* upward jump to level S, to replenish the product. Thus the SP *touches* level z from above and immediately rebounds with an upward jump; but the SP has not entered state $\{z\}$ at t_0. We say that the SP makes a *pass-by* of level z (see [59]). State $\{z\}$ is a boundary of the intervals (z, S) and $(-\infty, z)$. Definition 2.6 of pass-by assumes that a *double* jump occurs at t_0 because the SP must make *two* jumps in opposite directions.

Definition 2.6
SP Pass-by of a Boundary at Instant of Jump:
The SP makes a **pass-by** of a boundary $\partial \boldsymbol{A}$ at t_0 if

$$X(t_0^-) \in \boldsymbol{T} \times (\boldsymbol{A}^o \cup \boldsymbol{A}^e), \qquad X(t_0) \in \boldsymbol{T} \times (\boldsymbol{A}^o \cup \boldsymbol{A}^e)$$
$$\text{and } X(t_0^-) - \theta d_{t_0} + (1-\theta) u_{t_0} = z \in \boldsymbol{T} \times \partial \boldsymbol{A},$$

where $\theta = 1$ if the downward jump occurs first, touching level z; $\theta = 0$ if the upward jump occurs first, touching level z. Thus, $\theta = 1$ implies $X(t_0) = z + u_{t_0}$, and $\theta = 0$ implies $X(t_0) = z - d_{t_0}$.

2.4.5 *Continuous and Jump Crossings*

Definition 2.7
Left-continuous crossing:
An SP down- or upcrossing of level x at instant t_0 is called **left-continuous** if $X(t_0^-) = x$.
Continuous crossing:
A down- or upcrossing of level x at instant t_0 is called a *continuous crossing* if $X(t_0^-) = x = X(t_0)$.

Thus a continuous crossing is a left-continuous crossing, but a left-continuous crossing is not necessarily a continuous crossing (see Figs. 2.12, 2.13).

Definition 2.8
Left-continuous jump downcrossing:
A downcrossing of level x at instant t_0 is called a **left-continuous jump downcrossing** if $X(t_0^-) = x$ and $X(t_0) < x$.
Left-continuous jump upcrossing:
An upcrossing of level x at instant t_0 is called a **left-continuous jump upcrossing** if $X(t_0^-) = x$ and $X(t_0) > x$.

Notation 2.1
$\mathcal{D}_t(x), \mathcal{U}_t(x)$: number of downcrossings and number of upcrossings respectively of level x during time interval $(0, t)$.

$\mathcal{D}_t^c(x), \mathcal{U}_t^c(x)$: number of *left-continuous* downcrossings and number of left-continuous upcrossings of level x respectively, during time interval $(0, t)$.

$\mathcal{D}_t^j(x), \mathcal{U}_t^j(x)$: number of *jump* downcrossings and number of jump upcrossings of level x respectively during time interval $(0, t)$.

Figures 2.12, 2.13, 2.14, 2.15 and 2.16 picture various types of sample-path and SP transitions. Note that

$$\begin{aligned} \mathcal{D}_t(x) &= \mathcal{D}_t^c(x) + \mathcal{D}_t^j(x), \\ \mathcal{U}_t(x) &= \mathcal{U}_t^c(x) + \mathcal{U}_t^j(x). \end{aligned} \tag{2.1}$$

Remark 2.6 $\mathcal{D}_t^c(x), \mathcal{U}_t^c(x)$ count **all** left-continuous down- and upcrossings respectively, including those continuous from the right, and those not continuous from the right.

2.4.6 Number of Transitions in a Finite Time Interval

Consider state-space interval $\boldsymbol{A} \subset \boldsymbol{S}$.

Notation 2.2
$\mathcal{O}_t(\boldsymbol{A}), \mathcal{I}_t(\boldsymbol{A})$:
number of SP exits, and number of SP entrances of $\boldsymbol{T} \times \boldsymbol{A}$ during $(0, t)$, respectively.
$\mathcal{T}_t^o(\boldsymbol{A}), \mathcal{T}_t^e(\boldsymbol{A})$:
number of sample-path interior tangents, and number of sample-path exterior tangents, of set $\boldsymbol{A}$ during $(0, t)$, respectively.

In Proposition 2.2 below, the term sample-path *relative extrema* includes: maximum, minimum, supremum and infimum (see Definition 2.1).

Proposition 2.2 Fix $t > 0$ in T and level $x \in \boldsymbol{S}$. The random variables

$$\mathcal{O}_t(\boldsymbol{A}),\ \mathcal{I}_t(\boldsymbol{A}),\ \mathcal{T}_t^o(\boldsymbol{A}),\ \mathcal{T}_t^e(\boldsymbol{A}),\ \mathcal{D}_t(x),\ \mathcal{U}_t(x)$$

and their corresponding expected values are finite.

Proof
(1) **Exits and Entrances**: At most one sample-path exit or entrance of $(0, t) \times \boldsymbol{A}$ can occur between two successive sample-path relative extrema during $(0, t)$ (see Fig. 2.8).

(2) **Tangents**: Interior or exterior tangents can occur only at instants of relative extrema during $(0, t)$.
(3) **Down- and upcrossings**: At most one SP down- or upcrossing of level x can occur between successive relative extrema during $(0, t)$.

By Definition 2.1, a sample path has at most a finite number of relative extrema during $(0, t)$. Thus the random variables in the hypothesis are discrete and finite. Their expected values are finite because the expected value of the number of extrema in $(0, t)$ is finite. ■

Corollary 2.1
$\lim_{t\to\infty}\left(\mathcal{O}_t(\boldsymbol{A}) + \mathcal{I}_t(\boldsymbol{A}) + \mathcal{T}_t^o(\boldsymbol{A}) + \mathcal{T}_t^e(\boldsymbol{A}) + \mathcal{D}_t(x) + \mathcal{U}_t(x)\right)$ is a countable set.

Proof The time axis $\boldsymbol{T} = [0, \infty) = \lim_{t\to\infty}[0, t) = \cup_{n=0}^{\infty}[n, n+1)$. In Proposition 2.2 the number of exits from set $\boldsymbol{A}$ during time interval $[n, n+1) = \mathcal{O}_{n+1}(\boldsymbol{A}) - \mathcal{O}_n(\boldsymbol{A})$, which is finite. Similarly, the values of the other random variables in the hypothesis are finite during $[n, n+1)$. Therefore the sum of random variables in the hypothesis is finite in each time interval $[n, n+1)$. Countability follows since T is a countable union of finite numbers (e.g., [6]). ■

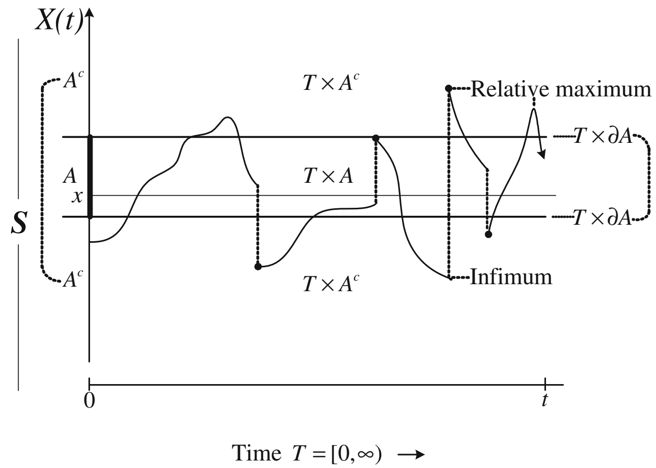


Fig. 2.8 $X(t), t \geq 0$, during an arbitrary finite time interval $(0, t)$, showing relative extrema, transitions with respect to set $\mathbf{T} \times \mathbf{A}$, and transitions with respect to a fixed level x

2.4.7 Principle of Set Balance

Consider a proper subset $\boldsymbol{A} \subset \boldsymbol{S}$.

Proposition 2.3 Instants of sample-path and/or SP exits and entrances of $\boldsymbol{T} \times \boldsymbol{A}$ alternate in time.

Proof The proposition follows from Definitions 2.1, 2.2, 2.3, 2.4 and Proposition 2.2. ∎

From Proposition 2.3 for fixed $t > 0$

$$\mathcal{O}_t(\boldsymbol{A}) - \mathcal{I}_t(\boldsymbol{A}) = \begin{cases} -1 \\ 0 \\ +1 \end{cases} \tag{2.2}$$

depending on whether $X(0)$, $X(t)$ are in $\boldsymbol{A}$ or $\boldsymbol{A}^c$. Dividing both sides of (2.2) by t and letting $t \to \infty$, gives the *principle of set balance* for exits and entrances of set A, assuming the limits exist, as follows.

Principle of Set Balance
For every set $\boldsymbol{A} \subset \boldsymbol{S}$,

$$\left.\begin{array}{c} \lim_{t\to\infty} \frac{\mathcal{O}_t(\boldsymbol{A})}{t} \underset{(a.s.)}{=} \lim_{t\to\infty} \frac{\mathcal{I}_t(\boldsymbol{A})}{t}, \\ \lim_{t\to\infty} \frac{E(\mathcal{O}_t(\boldsymbol{A}))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{I}_t(\boldsymbol{A}))}{t}. \end{array}\right\} \tag{2.3}$$

When emphasizing entrance and exit rates of sets, we usually refer to (2.3). Set balance expressed by (2.2) and (2.3) is applied in a number of places in this monograph, e.g., (3.5) of Theorem 3.1, Sect. 4.5, in reference [11], and numerous other publications.

2.4.8 Rate Balance for Down- and Upcrossings

By Definition 2.3 and Proposition 2.3, instants of down- and upcrossing alternate in time. Thus for each $t > 0$.

$$\mathcal{D}_t(x) - \mathcal{U}_t(x) = \begin{cases} -1 \\ 0 \\ +1 \end{cases} \tag{2.4}$$

depending on whether the values of $X(0)$, $X(t)$ are in (x, ∞) or in $(-\infty, x)$. Dividing (2.4) by t and letting $t \to \infty$, gives the *principle of rate balance* for down- and upcrossings across level $x \in \mathbf{S}$, assuming the limits exist,

$$\left.\begin{array}{c} \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t}, \\ \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t}. \end{array}\right\} \tag{2.5}$$

When referring to level crossings, we usually refer to (2.5) as *rate balance* across level x. Occasionally we call (2.5) *set balance* if we emphasize that crossings are *exits* or *entrances* of the sets $\mathbf{T} \times (x, \infty) \cap \mathbf{S}$ and $\mathbf{T} \times (-\infty, x) \cap \mathbf{S}$, as in (2.3).

Remark 2.7 When applying LC, the choice of state-space intervals and boundaries, is flexible and somewhat arbitrary. This facilitates potential creativity in obtaining solutions. Thoughtful choices may yield straightforward, simple derivations of systems of integral equations for the pdf and cdf of state variables in complex models. Examples given in the following chapters indicate the potentially wide applicability of LC.

2.4.9 Continuous and Discrete States

Atom: Definition In this monograph, a singleton discrete point $\{x\}$ in the state space is called an **atom** if it has a positive probability (see, e.g., p. 137ff in [74]).

A singleton state $\{x\} \subset \mathbf{S}$ may be either *continuous* or *discrete* with respect to the distribution of the state random variable (Table 2.2).

Table 2.2 Examples of atoms (discrete states) in various models; and corresponding figures

Stochastic model	State space	Atoms	Figures
Virtual wait, M/G/1	$[0, \infty)$	$x = 0$	Figure 2.1
Extended age, G/M/1	$(-\infty, +\infty)$	None	Figure 2.4
$\langle s, S\rangle$ inventory, decay	$(-\infty, +S]$	None	Figure 2.2
$\langle s, S\rangle$ inventory, no decay	$(-\infty, +S]$	$x = S$	Figure 2.6
Content, dam	$[0, \infty)$	Possibly $x = 0$	Figure 2.3
Risk model, no barrier	$(-\infty, +\infty)$	None	Figure 2.5
Birth-death, standard	$0, \ldots, N$	$0, \ldots, N$	Figure 2.10
Birth-death, extended	$[0, N]$	$0, \ldots, N$	Figure 2.10
Elevator-like model	$[0, N]$	$0, \ldots, N$	Figure 2.11

Any other states are continuous

Continuous State in State Space
A continuous state $\{x\}$ is characterized by having probability 0. That is, $P(X(t) = x) = 0$, $t \geq 0$, and $\lim_{t\to\infty} P(X(t) = x) = 0$. The long-run proportion of time that $X(\cdot)$ spends in $\mathbf{T} \times \{x\}$ is 0.

The LC method gains much power to analyze stochastic models from the one-to-one correspondence between: (a) sample-path left-limit down- and upcrossing rates of an arbitrary level $x \in \boldsymbol{S}$ (and other types of transitions related to level x), and (b) the transient and/or limiting pdfs of the state variable, and integral transforms of them, at level x (see Fig. 1.6).

Discrete States (Atoms)
A discrete state or **atom** is a singleton $\{x\}$ characterized by having *positive probability*. That is, $P(X(t) = x) > 0$ for some $t \geq 0$ and $\lim_{t\to\infty} P(X(t) = x) > 0$, when the limit exists. The long-run proportion of time that $X(\cdot)$ spends in $\mathbf{T} \times \{x\}$ is positive.

Proposition 2.4 The number of sample-path sojourns inside of a discrete state $\{x\} \subset \mathbf{S}$ is finite in finite time intervals, and is a countable set in $\mathbf{T} = [0, \infty)$.

Proof Sojourns in $\{x\}$ start at instants of sample-path *entrance* into $\{x\}$ and end at instants of *exit* from $\{x\}$. Countability follows from Proposition 2.2 and Corollary 2.1. If $X(\cdot) = x$ at the start and/or end of a finite time interval, the result is the same. ■

Set Balance for Discrete States
Substituting $\{x\} = \boldsymbol{A}$ in (2.2) and (2.3) yields

$$\mathcal{O}_t(\{x\}) - \mathcal{I}_t(\{x\}) = \begin{cases} +1 \\ 0 \\ -1 \end{cases}, \ t > 0, \text{ and}$$

$$\left.\begin{array}{l} \lim_{t\to\infty} \frac{\mathcal{O}_t(\{x\})}{t} = \lim_{t\to\infty} \frac{\mathcal{I}_t(\{x\})}{t} \text{ (a.s.)}, \\ \lim_{t\to\infty} \frac{E(\mathcal{O}_t(\{x\}))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{I}_t(\{x\}))}{t}. \end{array}\right\} \tag{2.6}$$

Equations (2.6) are equivalent to the well-known balance equations used for the rates into and out of discrete states, in continuous-time discrete-state Markov chains—CTMCs (e.g., [125]). The balance equations for CTMCs originally suggested to the author in 1973, the possibility of extending the "rate balance" technique for *discrete* states to analyze *continuous states* in continuous-time continuous-state Markov processes (or to mixed-state Markov processes). This was another motivation leading the author to the discovery of SPLC.

2.4.10 Hits and Egresses of Levels

Hits
Sample-path *hits* of a level describe the sample path in time ***left neighborhoods before*** "touching" the level. Hits describe the SP ***approach*** to the level from above or below. Intuitively, hits can be thought of as landings, touch downs, dives to, impacts with, descents to, ascents to, etc.

Egresses
Sample-path *egresses from* a level describe the sample path in time ***right neighborhoods after*** touching the level. Egresses describe SP ***departures*** *from the level above or below*. Egresses can be thought of as takeoffs, leaps from, rebounds from, jumps or dives away from, descents from, ascents from, etc.

Sample-path hit:

$X(\cdot)$ **hits** level x at instant t_0 if $X(t_0^-) = x$ (left limit) or if $X(t_0) = x$ and $\exists\varepsilon > 0 \ni X(t) \neq x, t \in (t_0 - \varepsilon, t_0)$.

Sample-path hit from above:

A sample-path hit of level x at t_0 is **from above** if $X(t) > x, t \in (t_0 - \varepsilon, t_0)$.

Sample-path hit from below: A sample-path hit of level x at t_0 is **from below** if $X(t) < x, t \in (t_0 - \varepsilon, t_0)$.

Sample-path egress:

A sample path makes an egress from level x at t_0 if $X(t_0^-) = x$ or if $X(t_0) = x$ and $\exists\varepsilon > 0 \ni X(t) \neq x, t \in (t_0, t_0 + \varepsilon)$.

Sample-path egress above:

A sample-path egress from level x at t_0 is **above** if $X(t) > x, t \in (t_0, t_0 + \varepsilon)$.

Sample-path egress below:

A sample-path egress from level x at t_0 is **below** if $X(t) < x, t \in (t_0, t_0 + \varepsilon)$.

Level as boundary. A level is a boundary of a set in $\boldsymbol{T} \times \boldsymbol{S}$. For example, level $x \in \boldsymbol{S}$ is a boundary of the sets:

$$\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S};\ \ \boldsymbol{T} \times [x, \infty) \cap \boldsymbol{S};\ \ \boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S};\ \ \boldsymbol{T} \times (-\infty, x] \cap \boldsymbol{S},$$

and an infinite number of other subsets of $\boldsymbol{S}$. The choice by an analyst of a set whose boundary is $\boldsymbol{T} \times \{x\}$ may simplify derivations of integral equations for the pdf and/or cdf at level x of the state variable. When applying "level crossing" theorems, we may require knowledge of the rate of sample-path hits of a level from above or below. On the other hand, we may require knowledge of the rate of sample-path egresses above or below (see Fig. 2.9).

Hits and egresses may be due to different types of transitions, such as sample-path exits, entrances, level crossings, or tangents.

A hit of level x from above at instant t_0 may be due to having $X(t_0^-) = x$; e.g., a left-limit downcrossing of x or left-limit tangent from above (interior tangent of $\boldsymbol{T} \times (x, \infty)$). A hit of level x from below may be due to a left-limit upcrossing of x or tangent from below (exterior tangent of $\boldsymbol{T} \times (x, \infty)$).

An egress from level x above at t_0 may be due to a continuous upcrossing of x or interior tangent of $\boldsymbol{T} \times (x, \infty)$ having $X(t_0) = x$. An egress from level x below at t_0 may be due to a continuous downcrossing of x or exterior tangent of $\boldsymbol{T} \times (x, \infty)$ having $X(t_0) = x = \lim_{t \downarrow t_0} X(t)$.

The rate at which a sample-path hits level x from above is not necessarily equal to the rate of egress from x below (see Fig. 2.9). When such transition rates on opposite sides of a boundary are unequal, LC theorems often facilitate the derivation of analytical properties of the pdf and cdf of the state variable, such as the position, size, and direction of any discontinuities. Different sample-path transition rates on opposite sides of a boundary occur in a variety of stochastic models (see Example 2.4).

Example 2.4 Consider a typical sample path of the **virtual wait** $W(t), t \geq 0$, for the **M/D/1 queue**. The state space is $\boldsymbol{S} = [0, \infty)$. Arrivals occur at a Poisson rate λ; every customer gets the same service time $D > 0$ (Fig. 2.9).

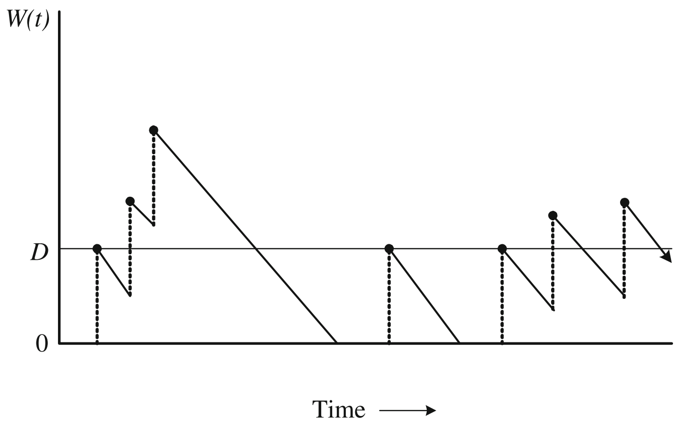


Fig. 2.9 Sample path of $\{W(t)\}_{t \geq 0}$ in M/D/1 queue. Service time $\equiv D$. Rate of egresses from level D below = Rate of hits of level D from above + arrival rate to empty system

All SP jumps are upward of size D. Consider level D, i.e., the line $\boldsymbol{T} \times \{D\}$. The SP hit rate of $\boldsymbol{T} \times \{D\}$ from above is due exclusively to continuous left-limit downcrossings of level D. This rate is **less** than the rate of egresses from level D below. The latter rate is due to two sources: (a) continuous downcrossings of level D **and** (b) exterior, right-continuous (same as right-limit) tangents of the set (D, ∞) (tangents of D from below). The tangents touch level D at the ends of SP jumps that start at level 0, at arrival instants when the system is empty. We show in Example 2.5, and also in Sect. 3.10 that the singleton state $\{D\}$ is a **continuous** state (not an atom). Assuming the traffic intensity $\lambda D < 1$, the limiting pdf of wait $\{P_0, f(x)\}_{x>0}$ exists. Level crossing theorems can be used to prove that there is a discontinuity $f(D^-) - f(D) = \lambda P_0$, where $P_0 = \lim_{t\to\infty} P(W(t) = 0)$ (Example 2.5).

2.4.11 Principle of Rate Balance for Hits and Egresses

Superscripts will have the following roles:

"a": from above, or to above (depending on transition type);

"b": from below, or to below (depending on transition type);

"c": left-limit (= left-continuous) (e.g., $X(t_0^-) = x$; same as continuous if $X(t_0^-) = X(t_0) = x$).

"j": jump transition.

The meaning of the superscripts will be clear from the context. Superscript "c" plays a dual role, which suffices because given a level x and an instant of transition t_0, SPLC is concerned, e.g., with **state-space** intervals like $(x - \varepsilon, x)$, $(x, x + \varepsilon)$, $\varepsilon > 0$, and with **Time** open neighborhoods like $(t_0 - \varepsilon, t_0)$, $(t_0, t_0 + \varepsilon)$, $\varepsilon > 0$.

$\mathcal{H}_t^a(x)$, $\mathcal{H}_t^{ac}(x)$: number of sample-path hits and left-limit hits of level x, from above during $(0, t)$, respectively.

$\mathcal{H}_t^b(x)$, $\mathcal{H}_t^{bc}(x)$: number of sample-path hits and left-limit hits of level x, from below during $(0, t)$, respectively.

$\mathcal{T}_t^a(x)$, $\mathcal{T}_t^{ac}(x)$: number of tangents and left-limit tangents of x, from above during $(0, t)$ (interior tangents of $\boldsymbol{T} \times (x, \infty) \cap \mathbf{S}$), respectively.

$\mathcal{T}_t^b(x)$, $\mathcal{T}_t^{bc}(x)$: number of tangents and left-limit tangents of x, from below during $(0, t)$ (exterior tangents of $\boldsymbol{T} \times (x, \infty) \cap \mathbf{S}$), respectively.

$\mathcal{E}_t^a(x)$, $\mathcal{E}_t^b(x)$: number of egresses from level x, to above and to below during $(0, t)$, respectively.

For example,

$$\begin{aligned}
&\mathcal{H}_t^a(x) = \mathcal{D}_t(x) - \mathcal{D}_t^j(x) + \mathcal{T}_t^a(x),\\
&\mathcal{H}_t^{ac}(x) = \mathcal{D}_t^c(x) + \mathcal{T}_t^{ac}(x),\\
&\mathcal{H}_t^b(x) = \mathcal{U}_t(x) - \mathcal{U}_t^j(x) + \mathcal{T}_t^b(x),\\
&\mathcal{H}_t^{bc}(x) = \mathcal{U}_t^c(x) + \mathcal{T}_t^{bc}(x),\\
&\mathcal{T}_t^b(x) = \mathcal{T}_t^e(\boldsymbol{T} \times (x,\infty) \cap \mathbf{S}),\\
&\mathcal{E}_t^b(x) = \mathcal{D}_t^c(x) + \mathcal{T}_t^{bc}(x).
\end{aligned} \tag{2.7}$$

In (2.7) dividing all terms by t and letting $t \to \infty$ gives rate equations. Each rate corresponds to some transition rate of the sample path in the corresponding model of interest. SPLC theorems like (1.1) give these rates in terms the limiting pdfs or cdfs of the state variable of the model, analogous to $\{P_0, f, (x)\}_{x>0}$ in Fig. 1.6.

Example 2.5 For the **M/D/1 queue** (Example 2.4, Fig. 2.9), $\mathcal{T}_t^a(x) = 0$ and $\mathcal{D}_t^j(x) = 0$ for all $x \in \mathbf{S}$ $(a.s.)$, i.e., there are no tangents from above and no downward jumps. Hence $\mathcal{H}_t^{ac}(x) = \mathcal{D}_t^c(x), x \geq 0$. For level D

$$\mathcal{H}_t^{ac}(D) = \mathcal{D}_t^c(D). \tag{2.8}$$

Since all hits of level D from above are left-limit (left-continuous) downcrossings,

$$\mathcal{E}_t^b(D) = \mathcal{H}_t^{ac}(D) + \mathcal{T}_t^b(D) = \mathcal{D}_t^c(D) + \mathcal{T}_t^b(D), \tag{2.9}$$

upon substitution from (2.8). In (2.9) dividing by t and letting $t \to \infty$ yields

$$\lim_{t\to\infty} \frac{\mathcal{D}_t^c(D)}{t} = \lim_{t\to\infty} \frac{\mathcal{E}_t^b(D)}{t} - \lim_{t\to\infty} \frac{\mathcal{T}_t^b(D)}{t}. \tag{2.10}$$

From Theorem 1.1

$$\lim_{t\to\infty} \frac{\mathcal{D}_t^c(D)}{t} = f(\mathcal{D}), \quad \lim_{t\to\infty} \frac{\mathcal{E}_t^b(D)}{t} = f(\mathcal{D}^-), \quad \lim_{t\to\infty} \frac{\mathcal{T}_t^b(D)}{t} = \lambda P_0,$$

where $\{P_0, f(x)\}_{x>0}$ is the limiting pdf of virtual wait. (All tangents of level D are due to jumps from level 0.) Substitution into (2.10) yields

$$f(D^+) = f(D) = f(D^-) - \lambda P_0, \text{ or } f(D^-) - f(D^+) = \lambda P_0. \tag{2.11}$$

Equation (2.11) expresses the analytical property that the limiting pdf has a jump discontinuity **downward** at $x = D$ of size λP_0 (see Sect. 3.10.1). The limiting pdf has no other discontinuities for $x > 0$. In addition, every downcrossing and tangent from below of level D, has no motion in the direction of Time in the line $\boldsymbol{T} \times \{D\}$. The total number of such transitions of level D in $\boldsymbol{T} = [0, \infty)$ is countable (Proposition 2.2, Corollary 2.1). Therefore the long-run **proportion** of time spent at level D is 0. So $\{D\}$ is a continuous state.

2.4.12 Hits and Egresses for Discrete States (Atoms)

A hit of a *discrete* state (atom) $\{x\} \in \boldsymbol{S} \subseteq \mathbb{R}$ may be an SP *entrance* into $\{x\}$, a *left-limit* down- or upcrossing of level x, a tangent of level x, etc. In the model of interest there must be, with positive probability, at least one way to enter and sojourn for a positive time in $\{x\}$ (e.g., there may be different rules regarding entering $\{x\}$ by jumps that start in disjoint state-space intervals below level x). This ensures that the long-run proportion of time spent in $\{x\}$ is positive, making $\{x\}$ an atom.

An egress out of a discrete state $\{x\}$ may be an SP exit from $\{x\}$, a *right-continuous* down- or upcrossing of level x, a *right-continuous* tangent of level x, etc. (see Figs. 2.14, 2.15).

Example 2.6 Consider a sample path of the **virtual wait** $\{W(t)\}_{t \geq 0}$ for the standard M/G/1 queue (Fig. 2.1). (The M/D/1 queue is a special case of M/G/1.) Let the arrival rate be λ and the service time S. Assume the traffic intensity $\lambda E(S) < 1$, so that the limiting distribution of wait exists. Let $\{P_0, f(x)\}_{x>0}$ be the limiting mixed pdf of wait. State $\{0\}$ is the **only discrete state** (atom) in the state space $\boldsymbol{S} = [0, \infty)$. P_0 is the long-run proportion of time that the sample path is in $\{0\}$, i.e., $\lim_{t \to \infty} P(W(t) = 0) = P_0 > 0$.

All hits of level 0 are due to sample-path **left-continuous** entrances into $\{0\}$ from $(0, \infty)$; at a hit instant say t_0, $W(t_0^-) = W(t_0) = 0$. The hit rate of level 0 from above is the entrance rate of state $\{0\}$, namely

$$\lim_{t \to \infty} \frac{\mathcal{H}_t^a(0)}{t} = \lim_{t \to \infty} \frac{\mathcal{I}_t(\{0\})}{t} = \lim_{t \to \infty} \frac{\mathcal{D}_t^c(0^+)}{t} = \lim_{x \downarrow 0} f(x) = f(0^+) \equiv f(0),$$

by Theorem 1.1.

The SP egress rate from level 0 above is the exit rate from state $\{0\}$. This is the rate at which customers arrive when the system is empty, namely λP_0. Set balance between the sets $\{0\}$ and $(0, \infty)$, equates entrance and exit rates of the atom $\{0\}$ (formula (2.6)). It yields the equation $f(0) = \lambda P_0$, which relates (curiously) the continuous part $f(x)$, $x > 0$, of $\{P_0, f(x)\}_{x>0}$ to the positive probability P_0 of the atom $\{0\}$. Thus, the SP entrance rate into and exit rate out of the discrete state {0} is $f(0)$ ($\equiv f(0^+)$). This type of relationship appears in different forms in various models, and is useful for computing limiting distributions of state variables. The PASTA principle (Poisson arrivals see time averages [145]) ensures that the "arrival-point P_0" is the same as the "time-average P_0" in $\{P_0, f(x)\}_{x>0}$.

At an instant of egress from level 0, the SP jumps upward by a realized value of the r.v. S, say s. This jump upcrosses every state-space level in interval $(0, s)$. The end point of the jump is tangent to level s from below. If S is a continuous r.v., the probability of hitting level s from below again due to a jump occurring at any other instant, is 0.

Example 2.7 Consider a standard **birth-death process** having states 0, ..., N (Fig. 2.10, [125]). Let the Poisson rate of jumps from n to $n+1$ be λ_n, and from n to $n-1$ be μ_n, $n = 1, ..., N$. The conventional state space is the set of discrete states $\boldsymbol{S} = \{0, 1, ..., N\}$ having limiting probabilities $P_0, ..., P_N$ respectively. Let $\boldsymbol{S}$ be **extended to the state space in the wide sense**, i.e., the closed interval $[0, N]$. This extension does not change the probability distribution associated with the model. All probability is still concentrated on the discrete states 0, ..., N. The SP moves at jump instants in $\boldsymbol{S}$ orthogonal to $\boldsymbol{T} \times \boldsymbol{S}$ (not in Time), through state-space intervals $(n, n+1)_{n=0,...,N-1}$, implying $P(U_{n=0}^{N-1}(n, n+1)) = 0$.

We derive the values of P_0,..., P_N using SPLC. Fix level x, $n < x < n+1$, $n \in \{0, ..., N-1\}$. The down- and upcrossing rates of x are respectively

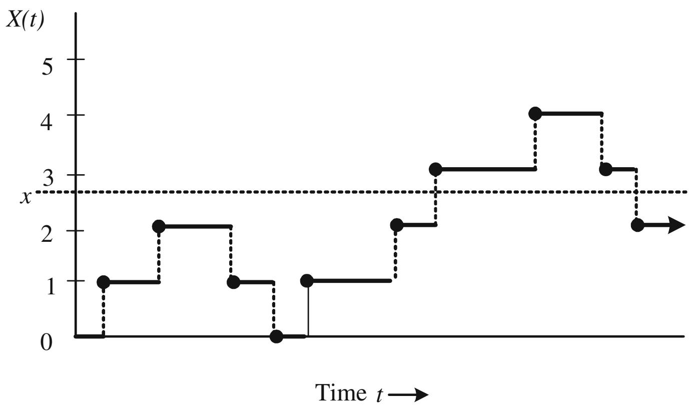


Fig. 2.10 Sample path of birth-death process with $N = 5$ discrete states (atoms). State space (in the wide sense) is the interval $[0, 4]$

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{D}_t^j(x)}{t} = \mu_{n+1} P_{n+1} \text{ and}$$
$$\lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t^j(x)}{t} = \lambda_n P_n,$$

respectively. By rate balance (2.5)

$$\lim_{t\to\infty} \frac{\mathcal{D}_t^j(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t^j(x)}{t},$$
$$\mu_{n+1} P_{n+1} = \lambda_n P_n, n = 0, ..., N-1,$$
$$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n, n = 0, ..., N-1,$$

Thus we obtain the well-known formula

$$P_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} P_0, n = 1, ..., N.$$

Substituting into the normalizing condition $P_0 + \cdots + P_N = 1$ yields P_0. and $P_1, ..., P_N$ (e.g., [125]). The above derivation appears to be identical to the conventional "**rate in = rate out**" argument for discrete states ([64, 125]). However, the extension of the state space to the wide-sense state space, includes continuous states. This allows us to use SPLC directly. The SPLC approach displays a subtle difference, which is prescient regarding solving more complex discrete-state continuous-time models (see Example 2.8).

Example 2.8 Consider an "**elevator-like" model** (Fig. 2.11). An elevator may stop at $N + 1$ floors, 0, ..., N. Assume the elevator travels at constant

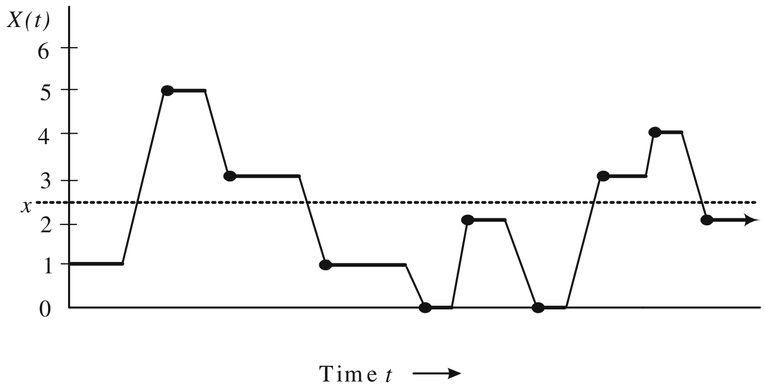


Fig. 2.11 Sample path of elevator-like model, $N = 6$ floors. Discrete states (atoms) are 0, ..., 5. Continuous states are open intervals $(n, n + 1)$, $n = 0, ..., 4$

speeds k and h meters per minute when moving respectively upward and downward between floors. We ignore the start-up acceleration and slow-down deceleration phases, for exposition. To fix ideas, assume the motion is in a semi-Markov environment (see Sect. 11.4 and, e.g., [125]). Assume that from the instant the elevator stops at floor i, its sojourn time at floor i has mean μ_i minutes until the next motion starts to a different floor. Its next stop will be at floor j with probability $P_{ij}, i \neq j \in \{0, ..., N\}$. The $(N+1) \times (N+1)$ matrix $\|P_{ij}\|$ is a Markov matrix. Assume the stationary probabilities of $\|P_{ij}\|$ are $\pi_i, i = 0, ..., N$. Let the limiting probability that the elevator is at floor i be $P_i, i = 0, ..., N$. Let the partial pdf of the elevator's position when it is moving upward and downward between floors i and $i+1$ be respectively $f_{i1}(x), f_{i2}(x), x \in (i, i+1), i = 0,..., N-1$. Let

$$f_i(x) = f_{i1}(x) + f_{i2}(x), x \in (i, i+1), i = 0, ..., N-1,$$

so that $f_i(x), x \in (i, i+1)$ is the limiting pdf of the elevator's position between floors i and $i+1$ regardless of its direction of motion. The state space is $\boldsymbol{S} = [0, N]$. The discrete states (atoms) are $0, ..., N$, representing the floors. The continuous states are points in the open intervals between floors, $(i, i+1), i = 0, ..., N-1$. The total probability is concentrated on both the discrete and continuous states. Hence the total pdf of the elevator's position will be "mixed". The normalizing condition is

$$\sum_{i=0}^{N} P_i + \sum_{i=0}^{N-1} \int_{x=i}^{i+1} f_i(x)dx = 1.$$

The goal is to determine for $i = 0, ..., N-1$: P_i, $f_{ij}(x), j = 1, 2$ and $f_i(x), x \in (i, i+1)$. To solve for these quantities we can apply the **method of pages** (also called **method of sheets**) originated and applied by the author in [11], and explained and applied in Sects. 4.11, 11.8, references [39, 42, 44, 52], and elsewhere. The relationship between $f_i(x)$ and the slope of the sample path (elevator speed) at level x is given in Theorem 6.4 below, reference [31]; and Sect. 11.4.

2.5 Transition Types Geometrically

Figures 2.12, 2.13, 2.14, 2.15 and 2.16 summarize ***geometrically*** 35 different types of sample-path and SP transitions with respect to a level $x \in \boldsymbol{S} \subseteq \mathbb{R}$ that can occur in various models.

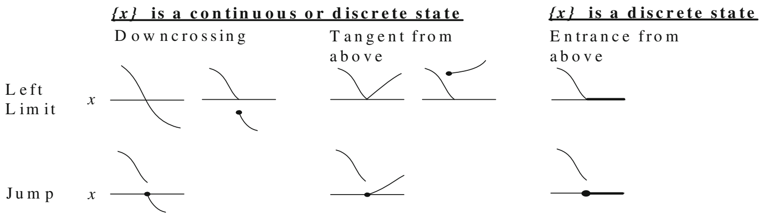


Fig. 2.12 Sample-path hits of level $x \in S$ *from above*

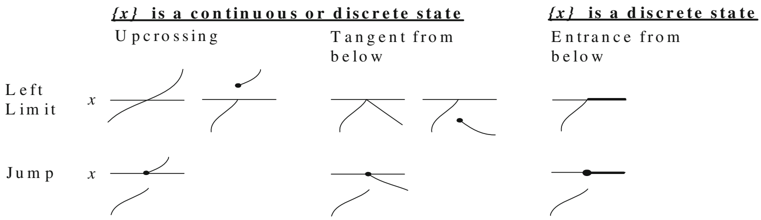


Fig. 2.13 Sample-path hits of level $x \in S$ *from below*

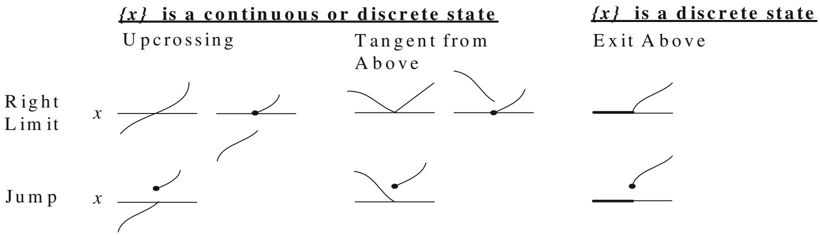


Fig. 2.14 Sample-path egresses from level $x \in S$ *above*

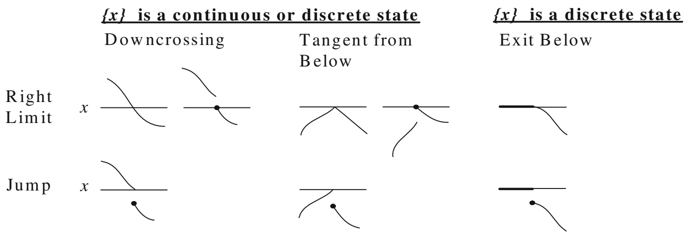


Fig. 2.15 Sample-path egresses from level $x \in S$ *below*

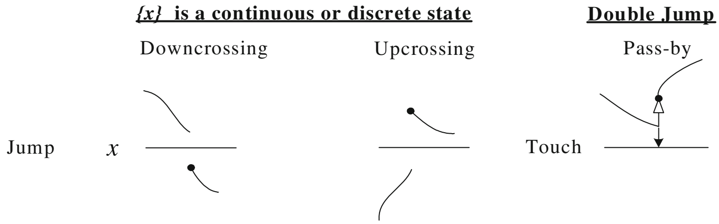


Fig. 2.16 Jump downcrossing and upcrossing—no hit of, or egress from level x. Touch of level x—pass-by due to double jump

Figures 2.12, 2.13, 2.14 and 2.15 illustrate four categories of transitions: SP hits from above and below; egresses above and below. In these figures, the instant of contact with level x is assumed to be $t_0 > 0$. In Figs. 2.12, 2.13, "**left limit**" means $X(t_0^-) = x$; "**jump**" means $X(t_0^-) \neq x$ and $X(t_0) = x$. In Figs. 2.14, 2.15, "**right limit**" means $X(t_0) = x$; "**jump**" means $X(t_0^-) = x$ and $X(t_0) \neq x$. Figure 2.16 illustrates level crossings that are not hits of, or egresses from level x; it also depicts a "touch" of a randomly determined level x from a "**pass-by**" due to a double jump.

Example 2.9 We illustrate the use of Figs. 2.12, 2.13, 2.14, 2.15 and 2.16. In Fig. 2.12 consider the 2 sub-diagrams in the position (**Left Limit, Tangent from Above**). The SP makes a hit from above of level x which is a **left-limit tangent from above**. These 2 sub-diagrams apply when $\{x\}$ is a continuous or a discrete state (atom).

Chapter 3
M/G/1 Queues and Variants

3.1 Introduction

This chapter considers the virtual wait process (workload) and related concepts in the M/G/1 queue, and variants of the model. It first develops relationships between sample-path level crossings and the *time dependent* (transient) distribution of wait. These relationships provide sample-path quantities obtainable via simulation or computation, which can estimate the analytical transient pdf of wait. They also lead in Sect. 3.3 to an alternative proof of the basic LC theorem for the *steady-state* pdf of wait (Theorem 1.1 in Sect. 1.6), by taking limits as time tends to infinity. The relationships are also of inherent interest for general time-dependent methods of analysis.

Next, in Sect. 3.3.1, alternative *forms* of Eq. (1.8) are derived by using a different, but very useful LC interpretation of sample-path jumps. These equation forms facilitate the analysis of certain variants of M/G/1 queues such as M/Discrete/1 where the service time has a general discrete distribution (Sect. 3.11.3).

The remainder of the chapter gives LC analyses of M/M/1 and M/G/1 models in the steady state, which illustrate the effectiveness of LC in practice.

3.2 Transient Distribution of Wait

Consider an M/G/1 queue with Poisson arrival rate λ, positive service times with cdf $B(x)$, $x \geq 0$, and pdf $\frac{d}{dx}B(x) = b(x)$, where the derivative exists. Let $\overline{B}(x) \equiv 1 - B(x)$. Consider a sample path of the virtual wait $\{W(t)\}_{t\geq 0}$, and fix level $x > 0$ in the state space $\mathbf{S} = [0, \infty)$ (Figs. 2.1 and 3.1). Let $\mathcal{D}_t(x)$

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_3

and $\mathcal{U}_t(x)$ denote the number of down- and upcrossings of level $x \geq 0$ during $(0, t)$, respectively. Both $\{\mathcal{D}_t(x)\}_{t\geq 0}$ and $\{\mathcal{U}_t(x)\}_{t\geq 0}$ are counting processes (e.g., p. 312 in [125]). The existence of $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$ and $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$ is important for the transient analysis, so we consider this property in Sects. 3.2.1 and 3.2.2.

3.2.1 Derivative $\partial E(\mathcal{D}_t(x))/\partial t$, $x \geq 0$

For economy of notation, we define $\mathcal{D}_t(0) \equiv \mathcal{D}_t(0^+) = \mathcal{H}_t^{a,c}(0)$ (number of left-limit continuous hits of 0 from above during $(0, t)) = \mathcal{I}_t(0)$ (number of SP entrances into $\{0\}$ during $(0, t)$) (see Sect. 2.4.11). (Here all downcrossings are continuous downcrossings).

Proposition 3.1 The partial derivative $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$, $x \geq 0$, exists and is positive for $t > 0$.

Proof The memoryless property of the exponential distribution implies that for each $x \geq 0$ $\{\mathcal{D}_t(x)\}_{t\geq 0}$ is a *delayed renewal* process (e.g., p. 466 in [125]; p. 197 in [99]), i.e., after each downcrossing of level x the future is a probabilistic replica of the whole process starting at time d_0. The delay d_0 depends on the initial wait $W(0) = x_0$. If $x_0 = x$, $d_0 = 0$. If $x_0 \neq x$, d_0 is the time from $t = 0$ to the first downcrossing of x. Starting from time d_0, let the level-x inter-downcrossing times be $d_1, d_2, \ldots$, where $d_k \underset{dis}{\equiv} d_1, k = 2, 3, \ldots$ (Fig. 3.1). Let $H_{d_0}(\cdot)$, $h_{d_0}(\cdot)$ denote the cdf and pdf of d_0, respectively. We prove the result when $d_0 > 0$; if $d_0 = 0$, the proof is similar.

The following well-known relationship (e.g., p. 423 in [125]; p. 167 in [99]) holds for $n = 1, 2, \ldots$, and $t > 0$:

$$\mathcal{D}_t(x) \geq n \iff d_0 + d_1 + \cdots + d_{n-1} \leq t.$$

$$\text{Thus } P(\mathcal{D}_t(x) \geq n) = P(d_0 + d_1 + \cdots + d_{n-1} \leq t).$$

Summing on both sides over $n = 1, 2, \ldots$, gives, by mutual independence of $\{d_i\}_{i=0,1,2,\ldots}$,

$$E(\mathcal{D}_t(x)) = \sum_{n=1}^{\infty} F_{d_0+d_1+\cdots+d_{n-1}}(t) = \sum_{n=1}^{\infty} \int_{s=0}^{t} F_{d_1}^{n-1}(t-s)h_{d_0}(s)ds,$$

where $F_{d_0+d_1+\cdots+d_{n-1}}(t)$, $t > 0$, is the cdf of $d_0 + d_1 + \cdots + d_{n-1}$ and $F_{d_1}^{n-1}(\cdot)$ is the $(n-1)$-fold self convolution of $F_{d_1}(\cdot)$. Since $\{\mathcal{D}_t(x)\}_{t\geq 0}$ is a delayed renewal process, $E(\mathcal{D}_t(x))$ is the renewal function. Thus $E(\mathcal{D}_t(x))$

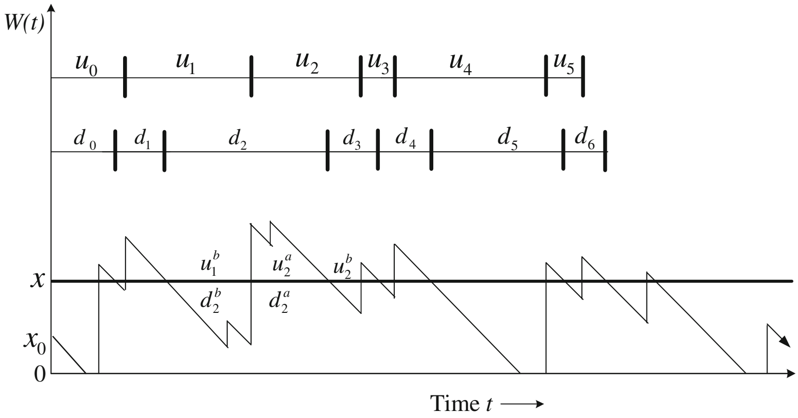


Fig. 3.1 Sample path of virtual wait in M/G/1 showing inter down- and upcrossing times for level x, $\{d_n\}_{n=1,2,...}$, $\{u_n\}_{n=1,2,...}$, and their components, e.g., d_2^b, d_2^a, u_2^a, u_2^b, etc.

is finite for all t, and the series $\sum_{n=1}^{\infty} F_{d_0+d_1+\cdots+d_{n-1}}(t)$ converges uniformly (e.g., p. 182 in [99]).

Since $F_{d_1}^{n-1}(0) = 0$, we obtain the derivative of the $(n-1)$-th summand as

$$\begin{aligned}
&\tfrac{\partial}{\partial t} \textstyle\int_{s=0}^{t} F_{d_1}^{n-1}(t-s)h_{d_0}(s)ds \\
&= \textstyle\int_{s=0}^{t} \tfrac{\partial}{\partial t} F_{d_1}^{n-1}(t-s)h_{d_0}(s)ds + F_{d_1}^{n-1}(0)h_{d_0}(t) \\
&= \textstyle\int_{s=0}^{t} f_{d_1}^{n-1}(t-s)h_{d_0}(s)ds,
\end{aligned}$$

where $f_{d_1}^{n-1}(\cdot)$ is the pdf of the $(n-1)$-fold convolution of d_1. (Due to Poisson arrivals, d_i is a continuous random variable implying $d_0 + d_1 + \cdots + d_{n-1}$ is continuous.) If we assume the parameters of the M/G/1 queue are such that the series of derivatives $\left\{\int_{s=0}^{t} f_{d_1}^{n-1}(t-s)h_{d_0}(s)ds\right\}_{n=1,2,...}$ also converges uniformly, then we can interchange the order of differentiation and summation (e.g., p. 317 in [6]), giving

$$\begin{aligned}
\frac{\partial}{\partial t} E(\mathcal{D}_t(x)) &= \sum_{n=1}^{\infty} \frac{\partial}{\partial t} \int_{s=0}^{t} F_{d_1}^{n-1}(t-s)h_{d_0}(s)ds \\
&= \sum_{n=1}^{\infty} \int_{s=0}^{t} f_{d_1}^{n-1}(t-s)h_{d_0}(s)ds\ .
\end{aligned}$$

Moreover, $\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) > 0$ since both $f_{d_1}^{n-1}(t-s) > 0$ and $h_{d_0}(s) > 0$. (Alternatively, $\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) > 0$ follows since $E(\mathcal{D}_t(x))$ is a non-decreasing function of t.) ■

Let $d_i = d_i^a + d_i^b$, where d_i^a is the time interval spent above x and d_i^b is the immediately preceding time interval spent below x by $W(\cdot)$ during d_i, $i = 1, 2, \ldots$.

3.2.2 *Derivative $\partial E(\mathcal{U}_t(x))/\partial t$, $x \geq 0$*

Consider a sample path of $\{W(t)\}_{t\geq 0}$. The process $\{\mathcal{U}_t(x)\}_{t\geq 0}$ is a "delayed" process. In general, however, $\{\mathcal{U}_t(x)\}_{t\geq 0}$ is not a renewal process. The delay u_0, is the time from $t = 0$ to the first (jump) *upcrossing* of x after time d_0. The subsequent level-x *inter-upcrossing* times starting at u_0 are denoted by $u_1, u_2, \ldots$ (Fig. 3.1). Let $u_i = u_i^a + u_i^b$ where u_i^a is the time interval spent above x and u_i^b is the immediately following time interval spent below x by $W(\cdot)$ during u_i. In general the random variables $\{u_i\}_{i=1,2,\ldots}$ are *not i.i.d.* (see Remark 3.1).

Remark 3.1 Let $\gamma_{x|y}$ denote the excess of a jump over level x given that the jump starts at level $y < x$ and initiates the interval u_i. Then $P(\gamma_{x|y} > z) = \frac{\overline{B}(x-y+z)}{\overline{B}(x-y)}$. Thus u_i^a depends on $x - y$, and $u_i^a \underset{dis}{=} \mathcal{B}_{\gamma_{x|y}}$, where $\mathcal{B}_{\gamma_{x|y}}$ denotes an M/G/1 busy period that starts with $W(0) \underset{dis}{=} \gamma_{x|y}$. (**Note**: The symbol "$\underset{dis}{=}$" means "is equal in distribution to", henceforth.) If $i \neq j$ then $u_i^a \underset{dis}{\neq} u_j^a$ *a.s.* (almost surely, i.e., with probability 1) because the start-of-jump position y is a continuous random variable. Now $u_i^b \underset{dis}{\equiv} u_1^b$, $i = 2, 3, \ldots$, because at the start of u_i^b $\{W(t)\}_{t\geq 0}$ downcrosses x and the future evolution is a probabilistic replica of $\{W(t)\}_{t\geq 0}$ from the start time of d_1, due to Poisson arrivals (that is, due to the memoryless property of the exponential interarrival times).

Proposition 3.2 The partial derivative $\partial E(\mathcal{U}_t(x))/\partial t$, $x \geq 0$, exists and is positive for $t > 0$.

Proof The delay time u_0 is a continuous random variable (r.v.). The process $\{\mathcal{U}_t(x)\}_{t\geq 0}$ is a counting process, but is not a renewal process (Fig. 3.1). Let $H_{u_0}(\cdot)$, $h_{u_0}(\cdot)$ denote the cdf and pdf of u_0, respectively.

The following relationship, usually applied for a renewal process, also holds for a counting process even though the inter-occurrence times are not independent. Thus

$$\mathcal{U}_t(x) \geq n \iff u_0 + u_1 + \cdots + u_{n-1} \leq t, n = 1, 2, \ldots$$

$$P(\mathcal{U}_t(x) \geq n) = P(u_0 + u_1 + \cdots + u_{n-1} \leq t).$$

Summing on both sides over $n = 1, 2, \ldots$ gives

$$\begin{aligned} E(\mathcal{U}_t(x)) &= \sum_{n=1}^{\infty} F_{u_0+u_1+\cdots+u_{n-1}}(t) \\ &= \sum_{n=1}^{\infty} \int_{s=0}^{t} F_{u_1+\cdots+u_{n-1}}(t-s) h_{u_0}(s) ds \end{aligned}$$

where $F_{u_1+\cdots+u_{n-1}}(t)$ is the cdf of $u_1 + \cdots + u_{n-1}$. Since u_i is continuous for each $i = 1, 2, \ldots$, the sum $u_0 + u_1 + \cdots + u_{n-1}$ is a continuous r.v. Taking $\frac{\partial}{\partial t}$ on both sides (differentiating under the integral) gives

$$\begin{aligned} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) &= \sum_{n=1}^{\infty} \Big(\int_{s=0}^{t} \frac{\partial}{\partial t} F_{u_1+\cdots+u_{n-1}}(t-s) h_{u_0}(s) ds \\ &\qquad + F_{u_1+\cdots+u_{n-1}}(0) h_{u_0}(t) \Big) \\ &= \sum_{n=1}^{\infty} \int_{s=0}^{t} f_{u_1+\cdots+u_{n-1}}(t-s) h_{u_0}(s) ds, \end{aligned}$$

where $f_{u_1+\cdots+u_{n-1}}(\cdot)$ is the pdf of $u_1 + \cdots + u_{n-1}$, since $F_{u_1+\cdots+u_{n-1}}(0) = 0$. The rest of the proof is similar to that in Proposition 3.1. Positiveness follows since $E(\mathcal{U}_t(x))$ is an increasing function of t. ■

The derivatives $\frac{\partial}{\partial t} E(\mathcal{U}_t(x))$ and $\frac{\partial}{\partial t} E(\mathcal{D}_t(x))$ are fundamentally related to the transient cdf of $W(t)$ (Sects. 3.2.3–3.2.8).

Remark 3.2 Assume the service time S is exponentially distributed with mean $1/\mu$ as in M/M/1 (Sect. 3.5). Then for any sample path

$$P(\gamma_{x|y} > z) = \frac{\overline{B}(x-y+z)}{\overline{B}(x-y)} = \frac{e^{-(x-y+z)}}{e^{-(x-y)}} = e^{-\mu z},$$

which is independent of x, y, and $x - y$, so that $u_i^a \underset{dis}{\equiv} \text{Exp}_\mu$ (:= exponential r.v. with mean $1/\mu$). In that case $u_i^a \underset{dis}{\equiv} \mathcal{B}_{M/M/1}, i - 1, 2, \ldots$ where $\mathcal{B}_{M/M/1}$:= busy period of an M/M/1 queue with $S \underset{dis}{=} \text{Exp}_\mu$ (Sect. 3.5.6). Then $\{u_i\}_{i=1,2,\ldots}$ **is a renewal process**.

3.2.3 Level Crossings and Transient CDF of Wait

Denote the transient time-t cdf, pdf and probability of a zero wait respectively as

$$\begin{aligned} &F_t(x) = P(W(t) \leq x), x \geq 0, t \geq 0, \\ &f_t(x) = \tfrac{\partial}{\partial x} F_t(x), x > 0, t \geq 0, \text{ wherever } \tfrac{\partial}{\partial x} F_t(x) \text{ exists}, \\ &P_0(t) = F_t(0), t \geq 0. \end{aligned} \tag{3.1}$$

Define the joint cdf of $(W(t_1), W(t_2))$ as

$$F_{t_1,t_2}(x_1, x_2) = P(W(t_1) \leq x_1, W(t_2) \leq x_2), t_1 \neq t_2 \geq 0, x_1, x_2 \geq 0. \tag{3.2}$$

The marginal cdfs are

$$F_{t_1}(x_1) = F_{t_1,t_2}(x_1, \infty), \;\; F_{t_2}(x_1) = F_{t_1,t_2}(\infty, x_2), x_1, x_2 \geq 0 \tag{3.3}$$

Note that $\mathcal{D}_t(x) - \mathcal{U}_t(x) \in \{0, +1, \; -1\}$ for every $x \geq 0$, $t \geq 0$, since down- and upcrossings of a fixed state-space level alternate in time (formulas (2.4) and (2.2)). The simple but useful Theorem 3.1 below connects $E(\mathcal{U}_t(x))$, $E(\mathcal{D}_t(x))$ and the transient cdf $F_t(x)$, by using (3.3) with $t_1 = 0$, $t_2 = t > 0$, and $x_1 = x_2 = x$ (Fig. 3.2).

In M/G/1, $\mathcal{D}_t(x) = \mathcal{D}_t^c(x)$ since all downcrossings are *left-continuous* (see Sect. 2.4.5). Also $\mathcal{U}_t(x) = \mathcal{U}_t^j(x)$, since all upcrossings are *jump* upcrossings.

Theorem 3.1 (P.H. Brill 1983) In the M/G/1 queue, for fixed $x \geq 0$, $t \geq 0$,

$$E(\mathcal{D}_t(x)) = E(\mathcal{U}_t(x)) + F_t(x) - F_0(x). \tag{3.4}$$

Proof Equation (3.4) holds when $t = 0$ because $\mathcal{D}_0(x) = \mathcal{U}_0(x) = 0$, $x \geq 0$. For $t > 0$, we compare possible sample path values of $\{W(s)\}_{0 \leq s \leq t}$ at $s = 0$ and $s = t$ with respect to level x, and relate the possible values to $F_0(x)$, $F_t(x)$ and $F_{0,t}(x, x)$ (Fig. 3.2). This procedure leads to the following values and probabilities for $\mathcal{D}_t(x) - \mathcal{U}_t(x)$:

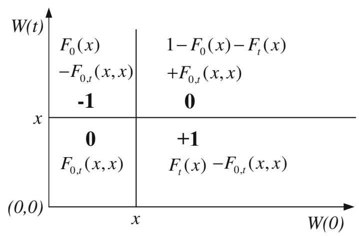


Fig. 3.2 Values of $\mathcal{D}_t(x) - \mathcal{U}_t(x)$ are $+1, 0, -1$, with probabilities shown in the finite and infinite sub-squares and two infinite rectangles of the $(W(0), W(t))$ plane

$\mathcal{D}_t(x) - \mathcal{U}_t(x)$	Probability
0	$1 - F_t(x) - F_0(x) + 2F_{0,t}(x,x)$
$+1$	$F_t(x) - F_{0,t}(x,x)$
-1	$F_0(x) - F_{0,t}(x,x)$

(3.5)

From (3.5) we obtain for fixed $x \geq 0$,

$$E(\mathcal{D}_t(x)) - E(\mathcal{U}_t(x)) = F_t(x) - F_0(x), t \geq 0, \tag{3.6}$$

equivalent to (3.4). ■

In (3.5) the term $\mathcal{D}_t(x) - \mathcal{U}_t(x) = 0$ contributes 0 to $E\left(D_t(x) - \mathcal{U}_t(x)\right)$; it is included for completeness. In further similar computations of expected value, terms with value 0 may be omitted.

Equation (3.6) leads to Theorem 3.2 below, which is fundamental for relating the transient probability distribution of wait and sample-path properties (see Remark 3.3 below).

3.2.4 Relating the Transient CDF and Level Crossings

Theorem 3.2 (P.H. Brill 1983) In the M/G/1 queue

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) = \frac{\partial}{\partial t}F_t(x) + \frac{\partial}{\partial t}E(\mathcal{U}_t(x)), t > 0, x \geq 0. \tag{3.7}$$

Proof Differentiating (3.4) with respect to t gives formula (3.7). (Existence of $\frac{\partial}{\partial t}(\cdot)$ of each term in (3.4) is considered in Sects. 3.2.1 and 3.2.2.) ■

Remark 3.3 Theorem 3.2 is a special case of the more general Theorem B (Theorem 4.1 in Sect. 4.2.1), which connects the sample-path marginal entrance rate and marginal exit rate of an arbitrary measurable set $\boldsymbol{A} \subset \boldsymbol{S}$ ($\boldsymbol{S}$:= state space) to $P_t(\boldsymbol{A})$ (:= probability of $\boldsymbol{A}$ at time t). In Theorem 3.2 $\boldsymbol{A} = [0, x]$.

3.2.5 Downcrossings and Transient PDF of Wait

Theorem 3.3 below shows that the sample-path quantity $\partial E(\mathcal{D}_t(x))/\partial t$ equals the analytic pdf $f_t(x), x \geq 0$, where $f_t(0) :\equiv f_t(0^+)$. We now briefly outline an important consequence, realizable by a computer program.

Simulate a finite number of independent sample paths of $\{W(s)\}_{s\geq 0}$ on $[0, t+h]$ (h small) to estimate $E(\mathcal{D}_t(x))$ and $E(\mathcal{D}_{t+h}(x))$, respectively. Then use $(E(\mathcal{D}_{t+h}(x)) - E(\mathcal{D}_t(x)))\,/\,h$ to estimate $\partial E(\mathcal{D}_t(x))/\partial t \approx f_t(x)$, $x \geq 0$. Adjust the values of t, x and h as needed to fit the particular model being considered.

Another important consequence is Corollary 3.2 below, which leads to an alternative proof of the crucial '*downcrossing*' part of the basic LC theorem for the *steady-state* pdf of wait (i.e., $\lim_{t\to\infty} E(\mathcal{D}_t(x))/t = f(x)$, in Theorem 1.1) in Chap. 1.

Theorem 3.3 *In the M/G/1 queue, for each* $t > 0$,

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = f_t(x), x > 0, \tag{3.8}$$

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = f_t(0). \tag{3.9}$$

Proof For the virtual wait $\{W(t)\}_{t\geq 0}$, fix state-space level $x > 0$. Consider the state-space triangular set $\triangle_{t,x,h} := \{(t, x+h), (t, x), (t+h, x)\}, t > 0$, where h is "small" (see Fig. 3.3). Examination of some possible sample paths $W(s)$ with respect to $\triangle_{t,x,h}$ leads to the possible values of $\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x)$ and their corresponding probabilities given in the table in (3.10); a brief explanation follows immediately after Theorem 3.3.

$\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x)$	Probability
$+1$	$F_t(x+h) - F_t(x) + o(h)$
-1	0, since $\mathcal{D}_t(x)$ increases with t
≥ 2	$o(h)$

(3.10)

Computing $E\,(D_{t+h}(x) - D_t(x))$ using (3.10), and dividing by h yields

$$\frac{E(\mathcal{D}_{t+h}(x)) - E(\mathcal{D}_t(x))}{h} = \frac{F_t(x+h) - F_t(x)}{h} + \frac{o(h)}{h}.$$

Letting $h \downarrow 0$ gives (3.8); then letting $x \downarrow 0$ yields (3.9). ∎

Explanation of the Probabilities in (3.10)

Let the Poisson arrival rate be λ. Assume the pdf $b(x)$ of the common service time is bounded on $(0, \infty)$. In this paragraph we denote the event $\{\mathcal{D}_t(x+h) - \mathcal{D}_t(x) = 1\}$ by $\{$*Diff1*$\}$ for brevity. Consider $P($*Diff1*$)$ for the three types of paths that enter and exit set $\triangle_{t,x,h}$ in Fig. 3.3. Employing the memoryless property of the exponential distribution, we get

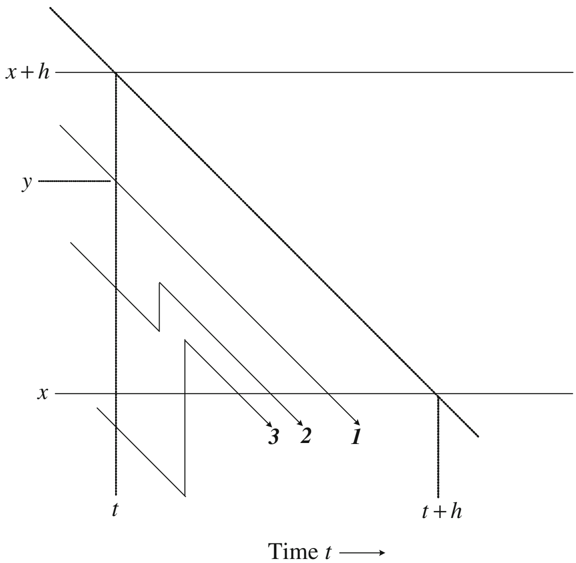


Fig. 3.3 Sample-path examples in triangle $\Delta_{t,x,h}$ resulting in $\mathcal{D}_t(x+h) - \mathcal{D}_t(x) = 1$. Probabilities are: P(path type 1) $= 1 - \lambda(y - x) + o(y-x)$; P(path type 2) $\leq o(h)$; P(path type 3) $\leq o(h)$

$$P(\textit{Diff1}|\text{path 1}) = e^{-\lambda(y-x)} \to 1 \text{ as } h \downarrow 0 \text{ since } (y-x) < h,$$

$$P(\textit{Diff1}|\text{path is type 2}) = [\lambda h + o(h)]\,[b(\cdot)h + o(h)] = o(h),$$

$$P(\textit{Diff1}|\text{path is type 3}) < [\lambda h + o(h)]\,[b(\cdot)h + o(h)] = o(h),$$

where $\lambda h + o(h) = P$(an arrival occurs in $(0, h)$), and $b(\cdot)h + o(h) = P$(a service-time jump ends in an interval of size $< h$). Similar consideration of other possible paths implies that $P(\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x) = n) = o\,(h)$ $(n = 2, 3, \ldots)$. Then $P(\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x) \geq n) = o\,(h)$ results because a countable sum of $o(h)$s = $o(h)$.

Alternative Proof of Formula (3.8) for Perspective

We can write

$$\begin{aligned}
&E(D_{t+h}(x) - D_t(x)) = 1 \cdot P(D_{t+h}(x) - D_t(x) = 1) + o(h) \\
&= \int_{y=x}^{x+h} e^{-\lambda(y-x)} f_t(y)dy + o(h) \\
&= \int_{y=x}^{x+h} [1 - \lambda(y-x) + o(y-x)]\, f_t(y)dy + o(h) \\
&= F_t(x+h) - F_t(x) - \int_{y=x}^{x+h} [\lambda(y-x) - o(y-x)]\, f_t(y)dy + o(h) \\
&= F_t(x+h) - F_t((x) - h\left[\lambda(y^* - x) - o(y^* - x)\right] f_t(y^*)dy + o(h), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x < y^* < x+h,
\end{aligned} \tag{3.11}$$

by the mean value theorem for integrals (e.g., p. 237 in [137]). Dividing both sides by h gives

$$\frac{E(D_{t+h}(x)-D_t(x))}{h}$$
$$= \frac{F_t(x+h)-F_t((x)}{h} - \lambda(y^* - x) + o(y^* - x)\cdot f_t(y^*) + \frac{o(h)}{h}.$$

Letting $h \downarrow 0$ leads to (3.8), because $h \downarrow 0$ implies: $(y^* - x) \downarrow 0$; $o(y^* - x) \downarrow 0$; $f_t(y^*) \to f_t(x)$; $o(h)/h \to 0$.

Corollary 3.1 For fixed $t > 0$,

$$E(\mathcal{D}_t(x)) = \int_{s=0}^{t} f_s(x)ds, x > 0, t > 0. \tag{3.12}$$

$$E(\mathcal{D}_t(0)) = \int_{s=0}^{t} f_s(0)ds, t > 0. \tag{3.13}$$

Proof Solving (3.8) for $E(\mathcal{D}_t(x))$ and (3.9) for $E(\mathcal{D}_t(0))$ by integrating with respect to t, and applying the initial condition $E(\mathcal{D}_0(x)) \equiv 0, x \geq 0$, gives (3.12) and (3.13), respectively. ■

3.2.6 Alternative Proof of $\lim_{t\to\infty} E(\mathcal{D}_t(x))/t = f(x)$

Starting from the transient analysis, Corollaries 3.2 and 3.3 below provide an alternative proof of the *downcrossing-rate* part of Theorem 1.1 in Sect. 1.6, Chap. 1, i.e., Eqs. (1.12) and (1.13). Let $\{P_0, f(x)\}_{x>0}$ denote the limiting (*steady-state*) mixed pdf of $\{W(t)\}_{t\geq 0}$. We assume $\lambda E(S) < 1$ (condition for existence of steady state).

Corollary 3.2

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x), x > 0 \tag{3.14}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(0))}{t} = f(0^+) \equiv f(0). \tag{3.15}$$

Proof Let $t \to \infty$ in (3.8) and (3.9) giving respectively

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = \lim_{t\to\infty} f_t(x) = f(x), x > 0,$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = \lim_{t\to\infty} f_t(0) = f(0).$$

In (3.12) and (3.13) divide both sides by $t > 0$, and let $t \to \infty$. Then

$$\lim_{t\to\infty} E(\mathcal{D}_t(x))/t = \lim_{t\to\infty} \frac{1}{t}\int_{s=0}^{t} f_s(x)ds = f(x), x > 0,$$
$$\lim_{t\to\infty} E(\mathcal{D}_t(0))/t = \lim_{t\to\infty} \frac{1}{t}\int_{s=0}^{t} f_s(0)ds = f(0).$$

Then (3.14) and (3.15) follow. ■

Corollary 3.3

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = f(x), x \geq 0\,(a.s.)\,. \tag{3.16}$$

Proof Since $\{\mathcal{D}_t(x)\}_{y\geq 0}$ is a renewal process due to Poisson arrivals, by the elementary renewal theorem,

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t}, x \geq 0\,(a.s.)\,.$$

Thus (3.16) follows from (3.14) and (3.15). ■

Corollary 3.4 gives an alternative perspective of set and rate balance (see Sect. 2.4.7) in Chap. 2.

Corollary 3.4

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t}, x \geq 0, \tag{3.17}$$
$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t}, x \geq 0\ \ (a.s.)\,. \tag{3.18}$$

Proof $\mathcal{D}_t(x) - \mathcal{U}_t(x) \in \{0, +1, -1\}$, $t \geq 0$, $x \geq 0$, for all possible sample paths of $\{W(t)\}_{t\geq 0}$. Hence $-1 \leq \mathcal{D}_t(x) - \mathcal{U}_t(x) \leq +1$, and $-1 \leq E(\mathcal{D}_t(x)) - E(\mathcal{U}_t(x)) \leq +1$. Dividing by $t > 0$ and letting $t \to \infty$ gives (3.17) and (3.18). ■

Remark 3.4 Formulas (3.17) and (3.18) also state the principle of set balance for sets $[0, x)$ and $[x, \infty)$, $x \geq 0$. That is, the equation *sample-path exit rate from set* $[0, x)$ = *sample-path entrance rate into* $[0, x)$ holds. The same principle applies to set $[x, \infty)$. Moreover, **SP** motion contains the sample path as a subset; i.e., **SP** motion includes the "not-in-Time" state-space jumps (see Sect. 2.3 in Chap. 2). Hence the same principle applies to **SP** exits and entrances.

3.2.7 Upcrossings and Transient PDF of Wait

Theorem 3.4 below connects the sample-path quantity $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$ to the analytical transient mixed pdf $\{P_0(t),\ f_t(y)\}_{0<y<x}, t>0$.

Theorem 3.4 In the M/G/1 queue with arrival rate λ and service time cdf $B(\cdot)$

$$\frac{\partial}{\partial t}E(\mathcal{U}_t(x)) = \lambda\overline{B}(x)P_0(t) + \lambda\int_{y=0}^{x}\overline{B}(x-y)f_t(y)dy \tag{3.19}$$

$$\frac{\partial}{\partial t}E(\mathcal{U}_t(0)) = \lambda P_0(t). \tag{3.20}$$

Proof We define $\mathcal{U}diff := \mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)$ here for brevity. Let $x>0, t>0$, and small $h>0$ be given. We examine possible SP upcrossings of level x in the state-space infinite rectangle $\sqcup := \{(t, t+h)\times(0,\infty)\}$ (Fig. 3.4). We consider possible SP entrances into $\sqcup$ at time t. Entrances that occur: above $x+h$ imply $\mathcal{U}diff = 0$; within $(x, x+h)$ imply $\mathcal{U}diff = 0$ ***or*** $P(\mathcal{U}diff = 1) = [\lambda h + o(h)]\,\overline{B}(x-y)f_t(y)dy$ for some $y \in (x-h, x)$; within $(0,x)$ imply $\mathcal{U}diff = 0$ ***or*** $P(\mathcal{U}diff = 1) = [\lambda h + o(h)]\,\overline{B}(x-y)f_t(y)dy$ for some $y \in (0, x-h)$; at level 0 imply $\mathcal{U}diff = 0$ ***or*** $P(\mathcal{U}diff = 1) = [\lambda h + o(h)]\,\overline{B}(x)P_t(0)$. For any $n \geq 2$, $P(\mathcal{U}diff = n) = o(h)$ because at least

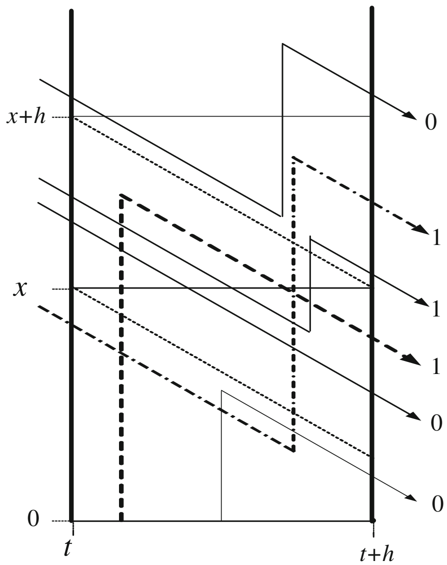


Fig. 3.4 Examples of sample-path (and SP) entrances and exits of set shaped like ⊔. Numbers at the ends of path segments are values of *Udiff*

n arrivals would be required during $(t, t+h)$ (equivalently during $(0, h)$ due to the memoryless property). Finally we get

$\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)$	Probability
$+1$	$[\lambda h + o(h)]\, P_0(t)\overline{B}(x) + [\lambda h + o(h)] \int_0^x \overline{B}(x-y) f_t(y)dy + o(h)$
≥ 2	$o(h)$.

(3.21)

In (3.21), the values $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) \leq 0$ are omitted since $\{\mathcal{U}\ \mathit{diff} = 0\}$ does not affect $E(\mathcal{U}\mathit{diff})$, and negative values are not possible because $\mathcal{U}_t(x)$ is a counting process (non-decreasing). Utilizing (3.21) gives

$$E(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)) = [\lambda h + o(h)]\, P_0(t)\overline{B}(x) + [\lambda h + o(h)] \int_{y=0}^x \overline{B}(x-y) f_t(y)dy + o(h).$$

Dividing both sides by h and taking limits as $h \downarrow 0$ gives (3.19). Letting $x \downarrow 0$ in (3.19) gives (3.20) since $\mathcal{U}_t(0) \equiv \mathcal{U}_t(0^+)$, and $\overline{B}(0) = 1$, since $\overline{B}(x)$ is right continuous. ■

Corollary 3.5 For fixed $t > 0$,

$$E(\mathcal{U}_t(x)) = \lambda \int_{s=0}^t \overline{B}(x) P_0(s)ds + \lambda \int_{s=0}^t \int_{y=0}^x \overline{B}(x-y) f_s(y)dyds,\ x > 0, \tag{3.22}$$

$$E(\mathcal{U}_t(0)) = \lambda \int_{s=0}^t P_0(s)ds. \tag{3.23}$$

Proof Integrate over time from 0 to t in (3.19) and (3.20). The constants of integration are 0 because $E(\mathcal{U}_0(x)) = 0,\ x \geq 0$. ■

Corollary 3.6 If the steady state exists, then

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = \lambda\overline{B}(x)P_0 + \lambda \int_0^x \overline{B}(x-y) f(y)dy, \tag{3.24}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(0)) = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t} = \lambda P_0. \tag{3.25}$$

Proof Note that

$$\lim_{t\to\infty} F_t(x) = F(x), \quad \lim_{t\to\infty} f_t(x) = f(x), \quad \lim_{t\to\infty} P_0(t) = P_0.$$

In (3.24) and (3.25), the results for

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) \text{ and } \lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(0))$$

follow from (3.19) and (3.20) respectively. The results for

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t}$$

follow from (3.22) and (3.23). ■

Corollary 3.6 completes the alternative transient-analysis derivation of Theorem 1.1, which seems to provide a more general perspective than the equilibrium-analysis approach of Sect. 1.6.

3.2.8 Integro-differential Equation for PDF of Wait

We apply LC to derive the ***Takács integro-differential equation*** for the transient probability distribution of wait, by utilizing Theorems 3.2, 3.3 and 3.4 above. (See Remarks 3.5, 3.6 and 3.7 below.)

Theorem 3.5 For an M/G/1 queue with arrival rate λ and service time cdf $B(\cdot)$, the transient distribution of the virtual wait satisfies the following equations for each $t > 0$:

$$f_t(x) = \frac{\partial}{\partial t} F_t(x) + \lambda \overline{B}(x) P_0(t) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy, \; x > 0, \tag{3.26}$$

$$f_t(0) = \frac{\partial}{\partial t} P_0(t) + \lambda P_0(t), \tag{3.27}$$

$$P_0(t) + \int_{y=0}^{\infty} f_t(y) dy = 1. \tag{3.28}$$

Proof The theorem follows by applying (3.7), substituting from (3.8), (3.9), (3.19), (3.20), and using (3.1). Equation (3.28) is the normalizing condition. ■

Remark 3.5 Equation (3.26) was derived by Takács in [139] by a different technique. Also, see formula (17), p. 87 in [140].

Remark 3.6 Minor extensions of the proofs in this section yield relationships and integro-differential equations for the transient pdf of wait in the important cases where the arrival rate and probability distribution of the service time are also time-dependent. In the formulas (3.26) and (3.27) we can replace λ by $\lambda(t)$ so that the arrival process is non-homogeneous Poisson; and $B(y)$ by $B_t(y)$ so that the service time is time-dependent (see Sect. 3.2.9 below). Equation (3.26) is called in the literature the Takács integro-differential equation (see [139]; formula (5.172), p. 227 in [104]).

Remark 3.7 The LC proofs of (3.26) and (3.27) have important ramifications. The relationship of both sides of (3.26) and (3.27) to $E(\mathcal{D}_t(x))$, $E(\mathcal{U}_t(x)), x \geq 0$, leads to techniques for **LC estimation of the transient distributions** by simulation of multiple independent sample paths (see Remark 9.2 in Sect. 9.2). **LC estimation (computation, approximation) for steady-state distributions is discussed in** Chap. 9. LC estimation is a form of non-parametric distribution (or density) estimation.

Example 3.1 below illustrates how transient sample-path quantities can be used to solve transient integro-differential equations numerically for analytical transient pdfs or transient probabilities.

Example 3.1 Assume $W(0) = x_0$ (≥ 0) so that $P_0(x_0) = 1$. Note that $P_0(s) = 0, 0 \leq s \leq x_0$. What is a point estimate of $P_t(0)$ for a finite time $t > x_0$? From Eqs. (3.27) and (3.9) we have the differential equation

$$\frac{\partial}{\partial t} P_0(t) + \lambda P_0(t) = \frac{\partial}{\partial t} E(\mathcal{D}_t(0)), t > x_0. \tag{3.29}$$

Using integrating factor $e^{\lambda t}$ in (3.29) and solving for $P_0(t)$ we get

$$\frac{d}{dt}\left(e^{\lambda t} P_0(t)\right) = e^{\lambda t} \frac{d}{dt} E(\mathcal{D}_t(0)),$$
$$e^{\lambda t} P_0(t) = \int_{s=x_0}^{t} e^{\lambda s}\left[\frac{d}{ds} E(\mathcal{D}_s(0))\right] ds + A$$

where A is a constant. Letting $t \downarrow x_0$ implies $A = e^{\lambda x_0} P_0(x_0) = e^{\lambda x_0}$, and

$$P_0(t) = e^{-\lambda t}\left(\int_{s=x_0}^{t} e^{\lambda s}\left[\frac{d}{ds} E(\mathcal{D}_s(0))\right] ds\right) + e^{-\lambda(t-x_0)}, t > x_0. \tag{3.30}$$

We estimate the function $\frac{d}{ds} E(\mathcal{D}_s(0)), x_0 \leq s \leq t$ in (3.30) as follows. Select a partition on $[x_0, t]$ having small norm h such that $t - x_0 = \nu h, \nu \in \mathbb{N}^+$ (set of positive integers). E.g., if $t - x_0$ is rational or irrational select $h = 0.001\,(t - x_0)$ or $0.0002\,(t - x_0)$, etc. Simulate N independent sample paths of $W(s), s \in [0, t + h]$, where N is large. Let $\mathcal{D}_{i,j}(0)$:= number of downcrossings of level 0 (left continuous hits from above) during time intervalls $[x_0 + (j - 1)h, x_0 + jh]$, $j = 1, \ldots, \nu + 1$ for the ith sample path, $i = 1, \ldots, N$. Let $\overline{\mathcal{D}}_j(0) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}_{i,j}(0)$, $j = 1, \ldots, \nu + 1$. An estimate of $\frac{d}{ds}E(\mathcal{D}s(0))$ is the step function

$$\frac{\overline{\mathcal{D}}_j(0)}{h},\ x_0 + (j-1)\,h < s < x_0 + jh,\ j = 1, \ldots, \nu + 1.$$

Substituting $\frac{\overline{\mathcal{D}}_j(0)}{h}$ into (3.30), we get the point estimate of $P_0(t)$ as

$$\begin{aligned}\widehat{P}_0(t) &= \frac{e^{-\lambda t}}{h}\sum_{j=1}^{\nu+1}\overline{\mathcal{D}}_j(0)\int_{s=x_0+(j-1)h}^{x_0+jh} e^{\lambda s}ds + e^{-\lambda(t-x_0)} \\ &= \frac{e^{-\lambda t}}{\lambda}\sum_{j=1}^{\nu+1}\left(\overline{\mathcal{D}}_j(0)\cdot\frac{e^{\lambda(x_0+jh)} - e^{\lambda(x_0+(j-1)h)}}{h}\right) + e^{-\lambda(t-x_0)}. \quad (3.31)\end{aligned}$$

FORTRAN-programmed computations were carried out in the Masters project [120] to estimate $P_0(t)$ when $x_0 = 0$, using a special case of the method outlined in this example. The computations generally agreed with the known analytical value of $P_0(t)$, $t > 0$, computed from the analytic formula given in [140].

Remark 3.8 Concepts in Example 3.1 relate to renewal theory since downcrossings of level 0 occur at the ends of **busy cycles**, which are i.i.d. random variables forming a renewal process (see formula (5.1) p. 189 in [99]). This will be discussed further in Chap. 10.

3.2.9 PDF When Arrivals and Service Are Time Dependent

We very briefly revisit the *transient pdf* of wait in the M/G/1 queue in Theorem 3.5 above in Sect. 3.2.8. We can prove by a slight generalization of the proofs in Sect. 3.2, that the theory holds for models where the arrival rate λ and cdf of service time $B(x)$, $x > 0$, depend on time t. Denoting them by $\lambda(t)$ and $B_t(x)$, $x > 0$, respectively, we obtain

$$\begin{aligned}f_t(x) &= \tfrac{\partial}{\partial t}F_t(x) + \lambda(t)\,\overline{B}_t(x)P_0(t) \\ &\qquad + \lambda(t)\int_{y=0}^{x}\overline{B}_t(x-y)f_t(y)dy,\ x > 0, \qquad (3.32) \\ f_t(0) &= \tfrac{\partial}{\partial t}P_0(t) + \lambda(t)\,P_0(t).\end{aligned}$$

The solution of the differential equation for $P_0(t)$ in (3.32) is

$$P_0(t) = e^{-m(t)} \int_{s=0}^{t} e^{m(s)} f_s(0) ds + P_0(0) e^{-m(t)}, \tag{3.33}$$

where $m(t) = \int_{s=0}^{t} \lambda(s)\, ds$ and $P_0(0) = \begin{cases} 1 \text{ if } W(0) = 0, \\ 0 \text{ otherwise.} \end{cases}$

3.2.10 Steady-State PDF of Wait from Transient PDF

Equation (1.8) for the steady state distribution of wait, is now proved directly from the foregoing LC connections between sample paths and the transient distribution of wait. The next theorem gives two such proofs.

Theorem 3.6 For an M/G/1 queue with arrival rate λ and service time S having cdf $B(\cdot)$, where $\lambda E(S) < 1$, the steady state pdf of the virtual wait $\{P_0, f(x)\}_{x>0}$, is given by

$$f(x) = \lambda \overline{B}(x) P_0 + \lambda \int_0^x \overline{B}(x-y) f(y) dy, \; x > 0, \tag{3.34}$$

$$f(0) = \lambda P_0, \tag{3.35}$$

$$P_0 + \int_0^\infty f(y) dy = 1. \tag{3.36}$$

Proof Since $\lambda E(S) < 1$, the transient probability distribution converges to the steady state probability distribution, i.e., $\lim_{t\to\infty} F_t(x) = F(x)$, $\lim_{t\to\infty} f_t(x) = f(x)$, $\lim_{t\to\infty} P_0(t) = P_0$. Moreover

$$\lim_{t\to\infty} \frac{\partial}{\partial t} F_t(x) = 0, \; x \geq 0, \; \lim_{t\to\infty} \frac{\partial}{\partial t} P_0(t) = 0.$$

Then (3.34) and (3.35) follow from Theorem 3.5 by letting $t \to \infty$.

Alternatively, (3.34) and (3.35) follow from the principle of rate balance expressed in (3.17), (3.18), and substituting from (3.14), (3.15), (3.24), and (3.25). ∎

Remark 3.9 For the M/G/1 queue with $\lambda E(S) < 1$, it is well known that

$$\lim_{t\to\infty} P(W(t) \leq x) = \lim_{n\to\infty} P(W_n \leq x), \; x \geq 0,$$

where W_n is the waiting time of the nth customer (arrival-point wait) (see [140]). Hence Eqs. (3.34)–(3.36) hold for the steady-state distributions of both the arrival-point wait and the virtual wait.

Remark 3.10 Using LC to derive (3.34)–(3.36) is useful because each algebraic term corresponds to a unique down- or upcrossing rate of level $x \geq 0$. This one-to-one correspondence enables the derivation of exact analytical integral equations for steady-state distributions of state variables in many complex stochastic models, intuitively and straightforwardly, using the sample path as a template. The idea is to construct a pertinent typical sample path of the stochastic model; then write the integral equation(s) by inspection using LC theorems and the principle of rate and/or set balance. The solution of the equation(s) is found with the aid of initial conditions (e.g., $f(0) = \lambda P_0$, $f'(0) = -\lambda P_0 b(0) + \lambda^2 P_0$). This procedure can save time and help the analyst focus on the model dynamics.

3.3 Steady-State Distribution of Wait

We begin with Example 3.2 below, which illustrates the derivation of the steady-state pdf of wait in an M/G/1 queue, where $G := \text{Erl}_{k,\mu}$ is the sum of k i.i.d. Exp_μ r.v.s. ($\text{Erl}_{k,\mu}$ denotes an Erlang r.v.; Exp_μ denotes an exponential r.v. with mean $1/\mu$.) In the M/$\text{Erl}_{k.\mu}$/1 queue $E(S) = k \cdot E(\text{Exp}_\mu) = k/\mu$.

Example 3.2 Consider the M/$\text{Erl}_{k.\mu}$/1 queue with arrival rate λ. Let $\mathcal{S}_k(x)$:= event {sum of k i.i.d. Exp_μs $\leq x$}, and $\mathcal{G}_k(x)$:= event {number of Poi_μ events in $(0, x)$ is $\geq k$}, where Poi_μ denotes a Poisson process with rate μ (see pp. 312–316 in [125]). Since $\mathcal{S}_k(x) \iff \mathcal{G}_k(x)$, we have $P(\mathcal{S}_k(x)) = P(\mathcal{G}_k(x))$, and cdf $B(x) = P(S \leq x) = P(\mathcal{S}_k(x)) = P(\mathcal{G}_k(x))$. Therefore

$$B(x) = P(\mathcal{G}_k(x)) = \sum_{i=k}^{\infty} \frac{e^{-\mu x}(\mu x)^i}{i!}, x > 0. \tag{3.37}$$

(See Sect. 2.3.2 for Exp_μ; Chap. 5 for Poi_μ, in [125].) Taking $\frac{d}{dx}$ in (3.37) readily shows that $b(x)$ $(:= \frac{d}{dx}B(x))$ is given by

$$b(x) = e^{-\mu x}\frac{(\mu x)^{k-1}\mu}{(k-1)!}, x > 0. \tag{3.38}$$

(Intuitively, (3.38) is equivalent to '$b(x)dx = P((k-1)$ Poi_μ events in $(0, x)$ and the kth event occurs at time $x)\, dx$'.) Since $\sum_{i=0}^{\infty} e^{-\mu x}(\mu x)^i/i! = 1$,

$$\overline{B}(x) = 1 - B(x) = e^{-\mu x}\left(\sum_{i=0}^{k-1} \frac{(\mu x)^i}{i!}\right), x \geq 0. \tag{3.39}$$

The condition for existence of the steady state is $\lambda E(S) < 1$ or $\lambda < \mu/k$.

Substituting $e^{-\mu x}\left(\sum_{i=0}^{k-1}\frac{(\mu x)^i}{i!}\right)$ for $\overline{B}(x)$ in (3.34), we obtain

$$\begin{aligned} f(x) &= \lambda P_0 e^{-\mu x}\left(\sum_{i=0}^{k-1}\frac{(\mu x)^i}{i!}\right) \\ &\quad +\lambda\int_{y=0}^{x} e^{-\mu(x-y)}\left(\sum_{i=0}^{k-1}\frac{(\mu(x-y))^i}{i!}\right) f(y)dy,\ \ x>0. \end{aligned} \tag{3.40}$$

where $P_0 = 1-\lambda E(S) = 1-(\lambda k)/\mu$. The normalizing condition is

$$P_0+\int_{x=0}^{\infty} f(x)dx = 1.$$

Case $k=2$: We illustrate the solution when $k = 2$, which corresponds to the M/Erl$_{2,\mu}$/1 queue. From (3.40) we have

$$f(x) = \lambda P_0 e^{-\mu x}(1+\mu x)+\lambda\int_{y=0}^{x} e^{-\mu(x-y)}(1+\mu(x-y))f(y)dy, x>0. \tag{3.41}$$

Differentiating (3.41) with respect to x twice results in the second order differential equation

$$f''(x)+(2\mu-\lambda)f'(x)+(\mu^2-2\lambda\mu)f(x) = 0, x>0,$$

with solution

$$f(x) = a_1 e^{r_1 x}+a_2 e^{r_2 x}, x>0, \tag{3.42}$$

where a_1, a_2 are constants to be determined, and

$$r_1 = -\mu+\frac{\lambda}{2}-\frac{1}{2}\sqrt{\lambda^2+4\mu\lambda},\ \ r_2 = -\mu+\frac{\lambda}{2}+\frac{1}{2}\sqrt{\lambda^2+4\mu\lambda},$$

are the solutions of the characteristic function

$$z^2+(2\mu-\lambda)z+(\mu^2-2\lambda\mu) = 0.$$

Both $r_1<0$, $r_2<0$. The constants a_1, a_2 and P_0 can be determined from the initial condition $f(0) = \lambda P_0$, and the normalizing condition $P_0+\int_{y=0}^{\infty} f(y)dy = 1$, giving

$$a_1 = \frac{r_1 r_2}{r_1 - r_2}(1 - P_0 + \frac{\lambda P_0}{r_2}),$$
$$a_2 = \lambda P_0 - a_1,$$
$$P_0 = 1 - \frac{2\lambda}{\mu}.$$

3.3.1 Alternative LC Equations for PDF of Wait

We now give two different forms of the basic integral equation (1.8) for the limiting pdf of wait in the M/G/1 queue (see Fig. 1.6). The alternative forms are useful due to their applicable LC interpretation. We can write (1.8) as

$$\begin{aligned} f(x) &= \lambda(1 - B(x))P_0 + \lambda \int_{y=0}^{x} (1 - B(x-y)) f(y) dy \\ &= \lambda \left(P_0 + \int_{y=0}^{x} f(y)dy \right) - \lambda \left(B(x)P_0 + \int_{y=0}^{x} B(x-y) f(y) dy \right) \end{aligned}$$

which gives *two* alternative forms of the LC equation:

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} B(x-y) dF(y), \; x \geq 0; \tag{3.43}$$

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} F(x-y) dB(y), \; x \geq 0, \tag{3.44}$$

noting that $F(x) = P_0 + \int_{y=0}^{x} f(y)dy$, and $F(\infty) = 1$. Formulas (3.43) and (3.44) have intuitive LC interpretations which help us write them immediately. Consider a sample path of the virtual wait (Fig. 1.4) and observe a one-to-one correspondence between the set of algebraic terms in the equations and a set of mutually exclusive and exhaustive sample-path crossings of level x, different from those depicted in Fig. 1.6.

In (3.43) and (3.44) the left side is the SP downcrossing rate of level x, as usual (see formula (3.14)). However, on the right side, $\lambda F(x)$ is the rate of *all* SP jumps that start in the state-space interval $[0, x]$. The second term subtracts off the rate of such jumps *that end below level* x (do not upcross x). Therefore the right side is precisely the total rate at which SP jumps upcross level x. Rate balance, (3.17) or (3.18), gives these equations directly. Note that (3.43) yields (3.44) by using the transformation $z = x - y$, $dz = -dy$, and integrating by parts.

Equations (3.43) and (3.44) are useful when analyzing variants of M/D/1 and M/Discrete/1 queues (Sects. 3.10 and 3.11); they help us derive the steady-state cdf $F(x)$ directly since $f(x) = F'(x)$. They are also useful in theoretical applications, such as in TAM (transform approximation method) [87, 129, 130]. The LC *intuitive* interpretations of (3.43) and (3.44) also suggest how to use LC to develop integral equations for the pdf of state variables in more general models.

Example 3.3 Consider the M/U$_{(0,c)}$/1 queue with arrival rate λ, where the service time $S \underset{dis}{=} \mathrm{U}_{(0,c)}$, a uniformly distributed r.v. on $(0, c)$, $c > 0$, i.e.,

$$B(x) = \begin{cases} 0, x < 0, \\ \frac{x}{c}, 0 \leq x < c, \\ 1, x \geq c. \end{cases} \tag{3.45}$$

We assume: $\{W(t)\}_{t \geq 0}$ is unbounded, i.e., $0 \leq W(t) < \infty$; the steady state pdf $\{P_0, f(x)\}_{x>0}$ and cdf $F(x), x \geq 0$, exist. A necessary and sufficient condition for the steady state to exist is $\lambda E(S) < 1 \iff (\lambda c/2) < 1$. Then busy periods are finite $(a.s.)$, and $P_0 = 1 - \lambda E(S) = 1 - \lambda \frac{c}{2}$.

Solution Approach in Example 3.3

We first solve (3.52) for $f(x), 0 < x < c$; then we indicate the iteration on successive state-space intervals $[c, 2c), [2c, 3c), \ldots$. (In Sect. 3.10 we obtain a complete solution for the M/D/1 queue, using a similar technique.)

Substituting from (3.45) into (3.43) and using $F(x - c) = 0, 0 < x < c$, gives (3.46).

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} \frac{(x-y)}{c} dF(y), 0 < x < c, \tag{3.46}$$

$$f(x) = \lambda F(x) - \lambda F(x - c) - \lambda \int_{y=x-c}^{x} \frac{(x-y)}{c} dF(y), x \geq c. \tag{3.47}$$

The LC explanation of (3.46) is the same as for 3.43. In (3.47) on the right side, $\lambda F(x)$ is the total rate of jumps that start below x. The term $-\lambda F(x - c)$ subtracts off the rate of jumps that start at any $y \in [0, x - c)$, and thus cannot upcross level x. The term $-\lambda \int_{y=x-c}^{x} \frac{(x-y)}{c} dF(y)$ subtracts off the rate of jumps that start in $(x - c, x)$ but are too small to upcross x.

Differentiating (3.46) twice with respect to x results in the second order linear homogeneous differential equation

$$f''(x) - \lambda f'(x) + \frac{\lambda}{c} f(x) = 0, x \in (0, c). \tag{3.48}$$

with characteristic (also called "auxiliary") equation

$$r^2 - \lambda r + \frac{\lambda}{c} = 0,$$

having solution

$$r = \frac{1}{2}\left(\lambda \pm \sqrt{\lambda^2 - 4\frac{\lambda}{c}}\right). \tag{3.49}$$

This gives

$$f(x) = a_1.e^{\frac{\lambda}{2}x}\cos\left(\frac{1}{2}\sqrt{\frac{4\lambda}{c} - \lambda^2}\cdot x\right) + a_2 \cdot e^{\frac{\lambda}{2}x}\sin\left(\frac{1}{2}\sqrt{\frac{4\lambda}{c} - \lambda^2}\cdot x\right), x \in (0, c), \tag{3.50}$$

where a_1, a_2 are constants to be determined. The *cos* and *sin* functions occur because we assumed that $\lambda < 2/c < 4/c$, so that the discriminant $\sqrt{\lambda^2 - 4\frac{\lambda}{c}}$ in (3.49) is a complex number (see Sect. 3.5, pp. 106–114 in [10]). In (3.46), applying the initial conditions $f(0) = \lambda P_0$, $f'(0) = \lambda^2 P_0 - \frac{\lambda P_0}{c}$ with $P_0 = 1 - \frac{\lambda c}{2}$, gives a_1, a_2 in (3.50) as

$$a_1 = \lambda(1 - \frac{\lambda c}{2}), \quad a_2 = \frac{(1 - \frac{\lambda c}{2})\lambda(\lambda - \frac{1}{c})}{\sqrt{\frac{4\lambda}{c} - \lambda^2}}.$$

We can iterate to solve for $f(x)$, $x \in [c, 2c)$, $x \in [2c, 3c)$, etc., by using (3.50). For $x \in [c, 2c)$, we have

$$f(x) = \lambda F(x) - \lambda \int_{y=c}^{x} \frac{(x-y)}{c} dF(y) - \lambda \int_{y=x-c}^{c} \frac{(x-y)}{c} f(y)dy - \lambda F(x-c), c \le x < 2c. \tag{3.51}$$

We solve (3.51) by substituting $f(y)$ from (3.50) on the interval $(x - c, c)$ into the second integral in (3.51), Then use discontinuity at $x = c$, i.e., $f(c^+) - f(c^-) = -\lambda P_0$ (letting $x \downarrow c$ in (3.51), $x \uparrow c$ in (3.50), and subtracting). The computation of $f(x)$, $c < x < 2c$ by stepping upward from state-space interval $(0, c)$ to interval $[c, 2c)$ is iterated on intervals $[ic, (i + 1)c)$, $i \ge 2$. (A similar discontinuity in the pdf $f(x)$ occurs at $x = D$ in the M/D/1 queue considered below in Sect. 3.12.)

Example 3.4 Now we assume a workload-bounded M/U$_{(0,c)}$/1 queue, i.e., $\{W(t)\}_{t\geq 0}$ is bounded at level $\boldsymbol{K} > 0$. To demonstrate the solution technique we let $\boldsymbol{K} := c$, and assume all service times that cause the virtual wait to exceed level c are truncated at level c. (See variant 1 in Sect. 3.16 and Fig. 3.33.)

The steady-state cdf $F(x)$ exists for all $\lambda > 0$ (see Sect. 2.1 in [25]). Substituting from (3.45) into (3.43) and using $F(x-c) = 0$, $0 < x < c$, gives

$$\begin{aligned} F'(x) &= \lambda F(x) - \lambda \int_{y=0}^{x} \tfrac{(x-y)}{c} dF(y) \\ &= \lambda F(x) - \lambda \int_{y=0}^{x} \tfrac{(x-y)}{c} f(y)dy - \lambda P_0 \tfrac{x}{c}, 0 < x < c, \end{aligned} \tag{3.52}$$

for the steady-state cdf $F(x)$.

Solution Approach for Example 3.4
Taking d/dx in (3.52) leads to the second order differential equation

$$F''(x) - \lambda F'(x) + \frac{\lambda}{c} F(x) = 0, 0 \leq x \leq c. \tag{3.53}$$

Assuming $\lambda > 4/c$, the solution of (3.53) is

$$F(x) = a_1 \cdot e^{r_1 x} + a_2 \cdot e^{r_2 x}, 0 \leq x \leq c, \tag{3.54}$$

where

$$r_1 = \frac{\frac{1}{2}\left(\lambda c + \sqrt{c^2\lambda^2 - 4c\lambda}\right)}{c},\ r_2 = \frac{\frac{1}{2}\left(\lambda c - \sqrt{c^2\lambda^2 - 4c\lambda}\right)}{c},$$

and a_1 and a_2 are constants to be determined. Using the initial conditions

$$\begin{aligned} F(0) &= a_1 + a_2 = P_0, \\ F'(0^+) &= a_1 \cdot r_1 + a_2 \cdot r_2 = \lambda P_0, \end{aligned}$$

results in

$$a_1 = \frac{r_2 - \lambda}{r_1 - r_2} P_0, \quad a_2 = \frac{r_1 - \lambda}{r_1 - r_2} P_0. \tag{3.55}$$

From (3.54) we get

$$F(x) = P_0 \left(\frac{r_2 - \lambda}{r_1 - r_2} \cdot e^{r_1 x} + \frac{r_1 - \lambda}{r_1 - r_2} \cdot e^{r_2 x} \right), 0 \leq x \leq c; \tag{3.56}$$

by using the boundary condition $F(c) = 1$,

$$P_0 = \left[\frac{r_2 - \lambda}{r_1 - r_2} \cdot e^{r_1 c} + \frac{r_1 - \lambda}{r_1 - r_2} \cdot e^{r_2 c} \right]^{-1}.$$

Generalization When Workload Bound is Greater Than c
Suppose the workload bound $\boldsymbol{k}$ is such that $c < \boldsymbol{k} < 2c$. Define $F(x) := F_0(x) \cdot \boldsymbol{I}_{[0,c]}(x) + F_1(x) \cdot \boldsymbol{I}_{[c,\boldsymbol{k}]}(x)$, where $\boldsymbol{I}_{\boldsymbol{A}}(x) = 1$ if $x \in A$, and 0 if $x \notin A$. The corresponding pdfs are $f_i(x) = dF_i(x)/dx$, $i = 0, 1$. Thus $F_0(x) = a_0 \cdot e^{r_1 x} + b_0 \cdot e^{r_2 x}$, $0 \le x \le c$ as in (3.54). (Here a_0 and b_0 will have different values than a_1, a_2 in (3.55) because now $F_1(\boldsymbol{k}) = 1$.) An integral equation for $F_1(x)$, $c \le x \le \boldsymbol{k}$, is given in terms of $F_0(\cdot)$ and $f_0(\cdot)$ as

$$\begin{aligned} F_1'(x) =& \lambda F_1(x) - \lambda \int_{y=c}^{x} \frac{(x-y)}{c} f_1(y)dy - \lambda F_0(x-c) \\ & - \lambda \int_{y=x-c}^{c} \frac{(x-y)}{c} f_0(y)dy, c < x \le k. \end{aligned} \tag{3.57}$$

If the bound $\boldsymbol{k} \in (jc, (j+1)c]$, $j \in \mathbb{N}^+$, we can iterate to solve for $F_{j+1}(x)$, $x \in [jc, k)$, $j = 1, 2, \ldots$, similarly as in (3.57). In the solution, we can use $F_{j+1}(jc^-) = F_j(jc^+)$ by continuity of the cdf at jc, $j = 1, 2, \ldots, \lfloor k/c \rfloor$ to facilitate solving for the constants. A related solution technique is applied for the M/D/1 queue in Sect. 3.12. When numerics are substituted for the parameters λ and c, the solution procedure can be programmed on a computer.

3.3.2 Relating System and Waiting Times Using LC

Let σ denote the *system time* in the M/G/1 queue. Denote the pdf and cdf of σ as $f_\sigma(x)$ and $F_\sigma(x)$, $x > 0$, respectively (see, e.g., Sect. 3.5.2). Then $\sigma = W_q + S$, where W_q is the wait before service and S is the common service time. The pdf and cdf of W_q are $f(x)$, $x > 0$, and $F(x)$, $x \ge 0$, respectively. We use LC interpretations of sample-path quantities to develop an analytical equation relating $f(x)$, $F(x)$ and $F_\sigma(x)$. This is an example where using LC interpretations of sample-path quantities can lead directly to analytical results, or to estimation methods for analytical quantities in particular models (see LC estimation in Chap. 9).

Peaks, Troughs and Downcrossings
A sample path of $\{W(s)\}_{s \ge 0}$ (Fig. 3.5) has a sequence of peaks (relative maxima) and troughs (relative 'minima', which are infima, i.e., greatest lower bounds, due to sample-path *right continuity*). A trough at level 0 is considered

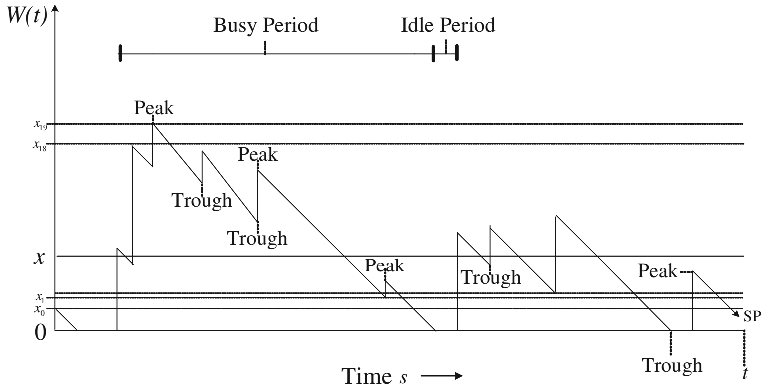


Fig. 3.5 Sample path of virtual wait $\{W(s)\}_{0 \le s \le t}$ showing peaks, troughs, level x, and subset of the associated partition $\{0 = x_0 < x_1 < \cdots < x_{18} < x_{19}\}$ which depends on the realized sample path over $[0, t]$

to be an interval which starts at an instant the SP hits level 0 from above, and ends at the next instant the SP leaps (jumps upward) from level zero.

Fix time $t > 0$ and level $x \ge 0$. Let $\mathcal{P}_t^+(x)$, $\mathcal{T}_t^+(x)$ denote respectively the number of peaks and troughs strictly above x during $[0, t)$. When $s = t$ the point $(t, W(t))$ is a trough since $\frac{d}{ds}W(s) = -1 \cdot \boldsymbol{I}_{(0,\infty)}(W(s)) + 0 \cdot \boldsymbol{I}_{\{0\}}(W(s))$. Then $\mathcal{D}_t(x)$ (number of SP downcrossings of x during $(0, t)$), is a step function with respect to x, with constant integer values on subintervals of the partition $\{0 = x_0 < x_1 < \cdots x_{n-1} < x_n\}$, where x_j is the ordinate of a peak or trough ($j = 1, \ldots, n-1$) and x_n is the highest peak during $[0, t]$. Such a fixed partition exists for each realized sample path (Fig. 3.5). An LC interpretation leads to

$$\mathcal{D}_t(x) = \mathcal{P}_t^+(x) - \mathcal{T}_t^+(x), x > 0. \tag{3.58}$$

The values of $\mathcal{D}_t(x)$ in adjacent subintervals, (x_{j-1}, x_j) and (x_j, x_{j+1}), j = 1, 2, …, differ by ±1, or 0 if S is a continuous random variable. If S has discontinuities, as in M/D/1 in which $S \equiv D$, then the values of $\mathcal{D}_t(x)$ in the two subintervals abutting on D will generally differ by more than 1; in this case, a difference >1 is the result of a discontinuity in the pdf of wait at $x = D$. Formula (3.58) can be useful when simulating sample paths for estimating state-space pdfs.

Equation Relating $f(x)$, $F(x)$ and $F_\sigma(x)$
Let $N_A(t)$ denote the number of arrivals during $(0,t)$. Assume $N_A(t) > 0$. Dividing (3.58) by $t > 0$, we obtain

$$\begin{aligned}\frac{\mathcal{D}_t(x)}{t} &= \frac{P_t^+(x)}{t} - \frac{T_t^+(x)}{t}\\ &= \frac{N_A(t)}{t}\cdot\frac{P_t^+(x)}{N_A(t)} - \frac{N_A(t)}{t}\cdot\frac{T_t^+(x)}{N_A(t)},\ t>0.\end{aligned} \tag{3.59}$$

Note that $P_t^+(x)$ represents the number of *system times* greater than x in $(0,t)$. Also $\mathcal{T}_t^+(x)$ represents the number of *waiting times* greater than x in $(0,t)$. Taking limits of the terms on the right side of (3.59) as $t \to \infty$ yields

$$\lim_{t\to\infty}\frac{N_A(t)}{t} \underset{a.s}{=} \lambda,\quad \lim_{t\to\infty}\frac{P_t^+(x)}{N_A(t)} \underset{a.s}{=} 1 - F_\sigma(x),\quad \lim_{t\to\infty}\frac{T_t^+(x)}{N_A(t)} \underset{a.s}{=} 1 - F(x),$$

which provides two more alternative forms of the M/G/1 integral equation for the pdf of wait, namely

$$f(x) = \lambda(1 - F_\sigma(x)) - \lambda(1 - F(x)), \tag{3.60}$$

and

$$f(x) = \lambda F(x) - \lambda F_\sigma(x). \tag{3.61}$$

LC Interpretations of (3.60) and (3.61)
On the right side of (3.60) the first term is the rate of *all* jumps that *end above* level x (system time $> x$). The second term subtracts off the rate of those jumps that *start above* level x (wait $> x$). Thus, the right side is the rate of SP jumps that upcross x.

The LC interpretation of (3.61) is that the first term on the right side is the rate of all jumps that *start* in $[0,x]$ (wait $\le x$). The second term subtracts off the rate of those jumps that *end* at levels in $[0,x]$ (system time $\le x$). Thus the right side is the rate of SP jumps that upcross x. Equation (3.61) is equivalent to (3.43) since, by independence of S and W_q

$$\begin{aligned}F_\sigma(x) = P(S + W_q \le x) &= \int_{y=0}^{x} P(S \le x - y | W_q = y) dF(y)\\ &= \int_{y=0}^{x} B(x-y) dF(y).\end{aligned}$$

Remark 3.11 Equation (3.59) combines sample-path peaks and troughs and the key part of the basic LC theorem $\lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} \underset{a.s.}{=} f(x)$, for a concrete derivation of integral equation (1.8) (same as (3.34)) in Fig. 1.6, based on LC interpretations of SP motion.

3.4 Waiting Time Properties in Steady State

We derive several familiar properties of the steady-state distribution of the waiting time before service starting from the basic LC integral equation (3.34). We let W_q := *wait before start of service*.

3.4.1 Probability of Zero Wait

In (3.34) integrate both sides with respect to x over $(0, \infty)$. This yields

$$1 - P_0 = \lambda P_0 \int_{x=0}^{\infty} \overline{B}(x)dx + \lambda \int_{x=0}^{\infty} \int_{y=0}^{x} \overline{B}(x-y) f(y) dy dx;$$

interchanging the order of integration in the double integral leads to

$$\begin{aligned} 1 - P_0 &= \lambda P_0 E(S) + \lambda E(S)(1 - P_0), \\ P_0 &= 1 - \lambda E(S) = 1 - \rho. \end{aligned} \tag{3.62}$$

3.4.2 Pollaczek-Khinchine (P-K) Formula

In (3.34) multiply both sides by x and integrate with respect to x over $(0, \infty)$. We obtain

$$\int_{x=0}^{\infty} x f(x) dx = \lambda P_0 \int_{x=0}^{\infty} x \overline{B}(x) dx + \lambda \int_{x=0}^{\infty} \int_{y=0}^{x} x \overline{B}(x-y) f(y) dy dx.$$

In the double integral, interchange the order of integration, write $x = x - y + y$, and simplify, giving

$$E(W_q) = \lambda P_0 \frac{E(S^2)}{2} + \lambda(1 - P_0)\frac{E(S^2)}{2} + \lambda E(W_q) E(S),$$

from which we obtain the well-known Pollaczek-Khinchine (P-K) formula

$$E(W_q) = \frac{\lambda E(S^2)}{2(1 - \lambda E(S))} = \frac{\lambda E(S^2)}{2\,(1-\rho)} = \frac{\lambda(var(S) + (E(S))^2)}{2P_0}, \tag{3.63}$$

where $var(S) := E(S^2) - (E(S))^2$. (See pp. 220–225 in [84] for a discussion and variations of the P-K formula.)

3.4.3 Expected Number in Queue and in System

Let N_q denote the number of customers waiting in the queue before service; let $L_q = E(N_q)$. From Little's formula $\boldsymbol{L = \lambda W}$ (see [110]), and formula (3.63), we get

$$L_q = \lambda E(W_q) = \frac{\lambda^2 E(S^2)}{2(1-\lambda E(S))} = \frac{\lambda^2 E(S^2)}{2(1-\rho)} = \frac{\lambda^2 E(S^2)}{2P_0}. \tag{3.64}$$

The expected number in the system is

$$L = L_q + L_s$$

where L_s denotes the expected number in service, given by

$$L_s = 1 \cdot (1 - P_0) + 0 \cdot P_0 = \lambda E(S) = \rho.$$

Thus

$$L = \frac{\lambda^2 E(S^2)}{2(1-\lambda E(S))} + \lambda E(S) = \frac{\lambda^2 E(S^2)}{2P_0} + \rho. \tag{3.65}$$

3.4.4 Laplace-Stieltjes Transform (LST) of a PDF

Before deriving the LST of $f(x)$, i.e., the pdf of W_q, we very briefly define the LST and related Laplace transform LT of a function. (See pp. 455–460 in [84] for a concise, clear introduction to the LST and LT.) The LST applies when the function has atoms or is continuous. The LT applies when the function is continuous. (Sect. 11.9 in Chap. 11 below presents an LC technique for estimating the LST and LT.)

LST
The Laplace-Stieltjes transform of $f(x)$ is

$$F^*(s) := E(e^{sW_q}) = \int_{x=0}^{\infty} e^{-sx} dF(x) = P_0 + \int_{x=0}^{\infty} e^{-sx} f(x)dx, s > 0. \tag{3.66}$$

The LST of $b(x)$, i.e., the pdf of the service time, is

$$B^*(s) := \int_{x=0}^{\infty} e^{-sx} dB(x) = \int_{x=0}^{\infty} e^{-sx} b(x)dx.$$

LT

The Laplace transform (LT) of $B(x)$, i.e., the cdf of the service time, is

$$\tilde{B}(s) := \int_{x=0}^{\infty} e^{-sx} B(x)dx.$$

Integrating $\tilde{B}(s)$ by parts shows that $B^*(s) = s\tilde{B}(s)$, $s > 0$.

In (3.34), the basic Volterra integral equation for $f(x)$, $x > 0$, we multiply both sides by e^{-sx} and integrate with respect to x over $(0, \infty)$, giving

$$\begin{aligned}\widetilde{f}(s) &= F^*(s) - P_0 = \textstyle\int_{x=0}^{\infty} e^{-sx} f(x)dx \\ &= \lambda P_0 \textstyle\int_{x=0}^{\infty} e^{-sx}\overline{B}(x)dx + \lambda \textstyle\int_{x=0}^{\infty} e^{-sx} \textstyle\int_{y=0}^{x} \overline{B}(x-y)f(y)dydx.\end{aligned} \tag{3.67}$$

In the double integral, express e^{-sx}as $e^{-sy} \cdot e^{-s(x-y)}$ and interchange the order of integration, giving

$$\begin{aligned}\widetilde{f}(s) = \lambda P_0 \textstyle\int_{x=0}^{\infty} e^{-sx}\overline{B}(x)dx \\ +\lambda \textstyle\int_{y=0}^{\infty} e^{-sy} f(y) \textstyle\int_{x=y}^{\infty} e^{-s(x-y)}\overline{B}(x-y)dxdy\end{aligned} \tag{3.68}$$

Simplifying yields the well-known formula

$$\begin{aligned}\widetilde{f}(s) &= \frac{sP_0}{s - \lambda(1 - B^*(s))} \\ &= \frac{s(1 - \lambda E(S))}{s - \lambda(1 - B^*(s))} = \frac{1-\rho}{1 - \rho\left(\frac{1-B^*(s)}{sE(S)}\right)}, s > 0,\end{aligned} \tag{3.69}$$

(see p. 237 in [84]). Substituting $\rho := \lambda E(S)$ and expanding $\widetilde{f}(s)$ as a geometric series gives

$$\widetilde{f}(s) = (1-\rho)\sum_{k=0}^{\infty} \rho^k \left(\frac{1 - B^*(s)}{sE(S)}\right)^k. \tag{3.70}$$

3.4.5 Series for PDF of W_q by Inverting $\widetilde{f}(s)$

Let γ_S denote the limiting excess service time having pdf $g(x)$, $x > 0$. Generally $g(x) = \overline{B}(x)/E(S)$, $x \in (0, \infty) \cap$ (*domain* of S). (See, e.g., p. 193 in [99]; p. 453 in [125]; p. 317 in [143]; and others.) (In Chap. 10 below we use

LC to derive an analytical expression for $g(x)$, which is denoted as $f_\gamma(x)$ therein.) Then

$$\widetilde{g}(s) = \frac{1}{E(S)} \int_{x=0}^{\infty} e^{-sx}(1 - B(x))dx = \frac{1}{E(S)} \left(\frac{1}{s} - \int_{x=0}^{\infty} e^{-sx} B(x)dx \right)$$
$$= \frac{1}{E(S)} \left(\frac{1}{s} - \frac{B^*(s)}{s} \right) = \frac{1 - B^*(s)}{sE(S)},$$

which is raised to the power k in the series (3.70). Moreover, $(\widetilde{g}(s))^k$ is the LT of the kth self convolution of $g(x)$, which we denote by $g_{(k)}(x)$, with $g_{(0)}(x) \equiv 1$. Since the LT uniquely defines a function and conversely, we can write (3.70) as the series

$$f(x) = (1 - \rho) \sum_{k=0}^{\infty} \rho^k g_{(k)}(x), x > 0, \tag{3.71}$$

which is known as the Beneš series (see [8]). Due to its importance in queueing theory we give several additional references for (3.71): pp. 200–201 in [104]; p. 236 in [84]; Example 7.24, p. 453 in [125]; pp. 169–170 in [99] ; Theorem 18, p. 37 in [118]; and also see [32, 65]; Sect. 10.1.3 in Chap. 11 below. Section 3.17 shows that (3.71) is a special case of a more general series having a term by term level-crossing interpretation.

Probabilistic Interpretation of LT

Remark 3.12 Equations (3.67) and (3.69) can be interpreted as the probability that the waiting time in queue is less than an independent 'catastrophe' random variable, say $Y \underset{dis}{=} \text{Exp}_s$. That is, the wait in queue finishes before the catastrophe occurs with probability $F^*(s)$. This probabilistic interpretation is useful for deriving Laplace transforms of random variables associated with stochastic models (see, e.g., Sect. 7.2, p. 267ff in [104]; Sect. 3 in [41]; and also see [92, 126]; many major papers supervised by M. Hlynka, University of Windsor).

3.4.6 Another Look at System Time

Here we use the notation of Sect. 3.3.2. For an arbitrary arrival, $\sigma > x$ iff the arrival waits in queue $y \leq x$ and its service time exceeds $x - y$, or, the arrival waits in queue $> x$. Thus

$$
\begin{aligned}
1 - F_\sigma(x) &= P(\sigma > x) \\
&= P_0 \overline{B}(x) + \int_{y=0}^{x} \overline{B}(x-y) f(y) dy + 1 - F(x) \\
&= \frac{f(x)}{\lambda} + 1 - F(x),
\end{aligned} \tag{3.72}
$$

implying

$$
f(x) = \lambda F(x) - \lambda F_\sigma(x),
$$

which is the same as (3.61). If $f(x)$ is known, then $F(x)$ can be computed. Then $F_\sigma(x)$ and $F'_\sigma(x) \equiv f_\sigma(x)$ can be obtained.

3.4.7 Connecting PDFs of System and Waiting Times

We now give a new LC-derived equation connecting $f_\sigma(x)$ directly with $f(x)$. Consider a sample path of the virtual wait and fix level $x > 0$. We view the SP jumps at arrival instants from the *ends* of the jumps (rather than from the starts of the jumps). The level of the end of a jump represents the system time of the corresponding arrival.

The downcrossing rate of level x is given by

$$
\lambda \int_{y=x}^{\infty} e^{-\lambda(y-x)} f_\sigma(y) dy,
$$

since $\lambda f_\sigma(y) dy$ is the rate of SP jumps that *end* within a "dy" neighborhood about level $y > x$, and $e^{-\lambda(y-x)}$ is the probability that the next customer arrives more than $y - x$ later. Thus the time interval of duration $y - x$ is devoid of new arrivals and associated SP jumps. The SP descends with slope -1 to level x, making a left-continuous downcrossing of x. (In this scenario, the jumps that end 'at' y may start either below x or in state-space interval (x, y). The end level y is the system time of the associated arrival.)

By Theorem 1.1, another expression for the SP downcrossing rate of x is $f(x)$ (also equal to upcrossing rate of x). Hence we have the equation

$$
f(x) = \lambda \int_{y=x}^{\infty} e^{-\lambda(y-x)} f_\sigma(y) dy. \tag{3.73}
$$

Multiplying both sides of (3.73) by $e^{-\lambda x}$ and differentiating with respect to x yields

$$
f_\sigma(x) = f(x) - \frac{f'(x)}{\lambda}, x > 0, \tag{3.74}
$$

wherever $f'(x)$ exists. Thus, if $f(x)$ is known, $f_\sigma(x)$ can be found directly using (3.74).

Example 3.5 In $M_\lambda/M_\mu/1$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$, $x > 0$ (see (3.112) and (3.117) in Sect. 3.5.2). Substituting $f(x)$ into (3.74) yields

$$\begin{aligned} f_\sigma(x) &= (\mu-\lambda)\, e^{-(\mu-\lambda)x}, x > 0, \\ F_\sigma(x) &= \textstyle\int_{y=0}^{x} f_\sigma(y)dy = 1 - e^{-(\mu-\lambda)x}, x \geq 0, \end{aligned} \tag{3.75}$$

Example 3.6 In $M/Erl_{2,\mu}/1$, the continuous part of the pdf of wait is

$$f(x) = a_1 e^{r_1 x} + a_2 e^{r_2 x}, x > 0;$$

thus

$$f_\sigma(x) = a_1 e^{r_1 x} + a_2 e^{r_2 x} - \frac{a_1 r_1 e^{r_1 x} + a_2 r_2 e^{r_2 x}}{\lambda}, x > 0,$$

where $a_i, r_i, i = 1, 2$ are given in Example 3.2, Sect. 3.3.

3.4.8 Number in System Probability Distribution

We obtain the steady-state probability distribution of the number in the system in two ways: by conditioning on W_q, or conditioning on σ. Let $P_n, n = 0, 1, \ldots$, denote the probability of n customers in the system at an arbitrary time point ($P_n :=$ proportion of time n are in the system). Let a_n, $d_n, n = 0, 1, \ldots$, denote the steady-state probability of n in the system just before an arrival, and just after a departure, respectively ($a_n :=$ proportion of arrivals that "see" n; $d_n :=$ proportion of departures that leave n).

For the M/G/1 queue it is well known that $P_n = a_n$ due to Poisson arrivals, and generally $a_n = d_n$ (e.g., pp. 501–502 in [125]; see also in [145]).

Conditioning on W_q, we obtain

$$\begin{aligned} P_n &= \int_{y=0}^{\infty} P(n-1 \text{ arrivals during } y | W_q = y) f(y)dy \\ &= \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!} f(y)dy, n = 1, 2, \ldots . \end{aligned} \tag{3.76}$$

Equation (3.76) is consistent with $P_0 + \int_{y=0}^{\infty} f(y)dy = 1$ since the proportion of time the system presents a positive wait to a potential arrival is

$$\sum_{n=1}^{\infty} P_n = \int_{y=0}^{\infty} e^{-\lambda y} \sum_{n=1}^{\infty} \frac{(\lambda y)^{n-1}}{(n-1)!} \cdot f(y)dy$$
$$= \int_{y=0}^{\infty} e^{-\lambda y} e^{\lambda y} f(y)dy = \int_{y=0}^{\infty} f(y)dy = 1 - P_0.$$

Alternatively, conditioning on σ,

$$P_n = \int_{y=0}^{\infty} P(n \text{ arrivals during } y|\sigma = y) f_\sigma(y)dy$$
$$= \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\lambda y)^n}{n!} f_\sigma(y)dy, n = 0, 1, \ldots, \tag{3.77}$$

which is also consistent with $P_0 + \int_{y=0}^{\infty} f(y)dy = 1$ since

$$\sum_{n=0}^{\infty} P_n = \int_{y=0}^{\infty} e^{-\lambda y} \left(\sum_{n=0}^{\infty} \frac{(\lambda y)^n}{n!} \right) \cdot f_\sigma(y)dy = \int_{y=0}^{\infty} f_\sigma(y)dy = 1.$$

If $f(\cdot)$, $f_\sigma(\cdot)$ are known for a particular M/G/1 model, either Eq. (3.76) or (3.77) can be applied to yield $\{P_n\}_{n=0,1,\ldots}$. Note that both a_n and d_n are also given by (3.76) or (3.77).

Interestingly

$$P_0 = \int_{y=0}^{\infty} e^{-\lambda y} f_\sigma(y)dy = \widetilde{f}_\sigma(\lambda), \tag{3.78}$$

the Laplace transform of $f_\sigma(\cdot)$. Using the probabilistic interpretation of the LT, formula (3.78) says that $P_0 = P(\sigma < Y)$ where Y is a an independent exponentially distributed “catastrophe” variable having rate λ (see Remark 3.12 in Sect. 3.4.4).

3.4.9 Renewal Reward Theorem: Statement

We state here the renewal reward theorem for easy reference, due to intermittent use in the sequel. The theorem applies generally to regenerative processes, although we state it here with respect to busy cycles in the standard M/G/1 queue. This brief section is based on the references in the Proof section immediately after Eq. (3.79) below.

Theorem Let R_n denote the amount of ‘reward’ earned during the busy cycle $\mathcal{C}_n$, where $\{R_n\}_{n=1,2,\ldots}$ are i.i.d. random variables. Assume $E(|R_1|) < \infty$, and let $R(t)$ denote the *total* reward earned during the time interval $(0, t)$, $t > 0$. Then $\{R(t)\}_{t\geq 0}$ is called the *renewal reward process*. The key result is

$$\frac{E(R_1)}{E(C)} = \lim_{t\to\infty} \frac{R(t)}{t} \text{ with probability 1.} \tag{3.79}$$

Proof Proofs of (3.79), and related material, are given in the following references: p. 41ff in [143]; p. 439ff in [125]; Proposition 3.4.1, p. 192 in [122]. ■

3.4.10 Expected Busy Period in M/G/1

Let $\mathcal{B}$ denote a busy period, $\mathcal{I}$ an idle period, and $\mathcal{C}$ a busy cycle. Then $\mathcal{C} = \mathcal{B} + \mathcal{I}$. The sequence $\{\mathcal{C}_n\}_{n=1,2,...}$, where $\mathcal{C}_n \underset{dis}{=} \mathcal{C}$, forms a renewal process. Consider a sample path of the virtual wait $\{W(t)\}_{t\geq 0}$. $\{W(t)\}_{t\geq 0}$ is a regenerative process with respect to $\{\mathcal{C}_n\}_{n=1,2,...}$. (For discussions on regenerative processes see, e.g., p. 447ff in [125]; p. 215ff in [122]; also see [132, 134], and others.)

Expected Busy Period

We now look at several ways to derive $E(\mathcal{B})$, for perspective.

{1} An expression for the (long-run) expected proportion of time that the sample path is in the state-space interval $(0, \infty)$ is $1 - P_0 = \rho := \lambda E(S)$. A different expression for the same proportion of time is

$$\lim_{t\to\infty} \frac{\mathcal{U}_t(0) E(\mathcal{B})}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t(0)}{t} E(\mathcal{B}) = \lambda P_0 E(\mathcal{B}),$$

since each exit of level 0 above (upcrossing of 0) initiates an independent busy period; moreover $\lim_{t\to\infty} \mathcal{U}_t(0)/t = \lambda P_0$. Equating these two different expressions gives

$$\lambda P_0 E(\mathcal{B}) = \lambda E(S),$$

$$E(\mathcal{B}) = \frac{E(S)}{P_0}. \tag{3.80}$$

{2} From the elementary renewal theorem (see, e.g., Proposition 7.1, p. 428 and Theorem 7.1, p. 432 in [125]), and LC theory,

$$E(\mathcal{C}) = \frac{1}{\text{downcrossing rate of level 0}} = \frac{1}{f(0)} = \frac{1}{\lambda P_0}. \tag{3.81}$$

In the *renewal reward theorem* let $R_n = \mathcal{B}_n$, where $\mathcal{B}_n$ is the busy period embedded in $\mathcal{C}_n$, $\mathcal{B}_n \underset{dis}{=} \mathcal{B}$, $n = 1, 2, \ldots$. Then $E(R_1) = E(\mathcal{B})$. Equation (3.79) gives

$$\frac{E(\mathcal{B})}{E(\mathcal{C})} = \frac{E(\mathcal{B})}{\frac{1}{\lambda P_0}} = \lim_{t\to\infty} \frac{\text{amount of time server is busy during } (0,t)}{t}$$
$$= \text{proportion of time workload is in } (0,\infty) \; = \; \rho = \lambda E(S).$$
$$E(\mathcal{B}) = \frac{\lambda E(S)}{\lambda P_0} = \frac{E(S)}{P_0},$$

which agrees with (3.80).
{3} Since $\mathcal{C} = \mathcal{B} + \mathcal{I}$,

$$E(\mathcal{B}) = E(\mathcal{C}) - E(\mathcal{I}) = \frac{1}{\lambda P_0} - \frac{1}{\lambda} = \frac{1-P_0}{\lambda P_0} = \frac{E(S)}{P_0}.$$

{4} Intuitively $E(\mathcal{B})$ is the $(1-P_0)$-th proportion of $E(\mathcal{C})$, i.e.,

$$E(\mathcal{B}) = (1-P_0)\cdot E(\mathcal{C}) = \frac{1-P_0}{\lambda P_0} = \frac{\lambda E(S)}{\lambda P_0} = \frac{E(S)}{P_0};$$

this is really a version of the renewal-reward-theorem method.

The appearance of P_0 in the denominator of (3.80) follows from the renewal reward theorem, or from $f(0) = \lambda P_0$ in Theorem 1.1, Corollary 1.1. The expression

$$E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0} \tag{3.82}$$

appears to be more fundamental than the expression $E(\mathcal{B}) = \frac{E(S)}{1-\lambda E(S)}$, since in some well-known variants of the standard M/G/1 queue, $P_0 \neq 1 - \lambda E(S)$ (e.g., if the workload has a positive barrier (see [25]; also Sects. 3.9 and 3.13 below).
{5} Busy periods and idle periods form an alternating renewal process. Hence

$$P_0 = \frac{E(\mathcal{I})}{E(\mathcal{B}) + E(\mathcal{I})} = \frac{\frac{1}{\lambda}}{E(\mathcal{B}) + \frac{1}{\lambda}} = 1 - \lambda E(S);$$

the last equality implies (3.82). This derivation also assumes the renewal reward theorem, so is similar to derivation {2}. However, it does not directly "explain" the appearance of P_0 in the denominator; derivation {2} does provide the explanation.

Remark 3.13 Formula (3.82) shows immediately that

$$E(\mathcal{B}) < \infty \text{ iff } 0 < P_0 \leq 1,$$

and equivalently

$$E(\mathcal{B}) = \infty \text{ iff } P_0 = 0.$$

The **stability condition** for the standard M/G/1 queue is $P_0 > 0$ (same as $\lambda E(S) < 1$). The queue is stable iff state $\{0\}$ is positive recurrent iff $\mathcal{B}$ is finite $(a.s.)$

Remark 3.14 Formula $E(\mathcal{B}) = \frac{1-P_0}{f(0)}$ is even more fundamental than $E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0}$, since in some M/G/1 variants $f(0) \neq \lambda P_0$. For example $f(0) = \lambda P_0 B(K)$ in a workload-barrier M/G/1 queue with finite barrier $K > 0$, where a customer balks if its service time would cause the workload to overshoot the barrier (variant 2 of Sect. 3.16.3); in that case $E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0 B(K)}$.

3.4.11 Equation for $f(x)$ via Renewal Reward Theorem

Consider $\{W(t)\}_{t\geq 0}$. Let $\{P_0, f(x)\}_{x>0}$ be the limiting pdf of wait in M/G/1. We have $f(x) = \lim_{t\to\infty} \mathcal{D}_t(x)/t$ by Theorem 1.1. We now apply the *renewal reward theorem* to derive the *right hand side* of Eq. (1.8), as a check on the upcrossing-rate interpretation in Theorem 1.1, and because the renewal reward theorem is useful for analyzing many complex models as well (see references following Eq. (3.79)). Let $\mathcal{C}$:= an M/G/1 *busy cycle*, and $A_{\mathcal{C}}$:= *number of arrivals during C* (same as number of SP jumps of the embedded busy period $\mathcal{B}$). Denote the customers served in $\mathcal{B}$ as $\{\mathrm{C}_i\}_{i=1,\ldots,A_{\mathcal{C}}}$. Let

$$\mathcal{U}_i(x) = \begin{cases} 1 \text{ if customer-}i\text{'s service jump upcrosses level } x, \ i = 1, \ldots, A_{\mathcal{C}}, \\ 0 \text{ otherwise.} \end{cases}$$

Assume we do not know the order of arrival of the C_i's. Conditioning on the starting levels of the SP jumps, we have

$$P(\mathcal{U}_i(x) = 1) = P(S > x | W_i = 0) P(W_i = 0)$$
$$+ \int_{y=0}^{x} P(S > x - y | W_i = y) d\dot{y}, i = 1, \ldots, A_{\mathcal{C}}.$$

where the events $\{W_i = 0\}$ and $\{W_i = y\}_{y>0}$ are mutually exclusive and exhaustive. Thus

$$P(\mathcal{U}_i(x) = 1) = \overline{B}(x)P_0 + \int_{y=0}^{x} \overline{B}(x-y)f(y)dy, i = 1, \ldots, A_C;$$

$$E(\mathcal{U}_i(x) = \overline{B}(x)P_0 + \int_{y=0}^{x} \overline{B}(x-y)f(y)dy, i = 1, \ldots, A_C.$$

Since $E(A_\mathcal{C}) = 1/P_0$ (see 3.4.14 below), The number of upcrossings of x during $A_\mathcal{C}$ is,

$$\begin{aligned}
\mathcal{U}_\mathcal{C}(x) &= \sum_{i=1}^{A_\mathcal{C}} \mathcal{U}_i(x), x > 0, \\
E(\mathcal{U}_\mathcal{C}(x)) &= E(A_\mathcal{C})E(\mathcal{U}_i(x) \\
&= \frac{1}{P_0}\left(\overline{B}(x)P_0 + \int_{y=0}^{x} \overline{B}(x-y)f(y)dy\right) \\
&= \overline{B}(x) + \frac{1}{P_0}\int_{y=0}^{x} \overline{B}(x-y)f(y)dy.
\end{aligned}$$

Finally the renewal reward theorem implies

$$\begin{aligned}
\lim \frac{\mathcal{U}_t(x)}{t} = \frac{E(\mathcal{U}_\mathcal{C}(x))}{E(\mathcal{C})} &= \frac{\overline{B}(x) + \frac{1}{P_0}\int_{y=0}^{x} \overline{B}(x-y)f(y)dy}{1/(\lambda P_0)} \\
&= \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy;
\end{aligned}$$

rate balance across level x, viz., $\lim_{t\to\infty} \mathcal{D}_t(x)/t = \lim \mathcal{U}_t(x)/t$, yields Eq. (1.8).

3.4.12 Busy Period Structure in Standard M/G/1

The M/G/1 busy period $\mathcal{B}$ can be partitioned into a set of sub-busy periods, different from a classical partition (see pp. 206–211 and p. 220ff in [104]; also [140]). Direct observation of a sample path of $\{W(t)\}_{t\geq 0}$ in Fig. 3.6, leads to a partition of $\mathcal{B}$ which preserves the scale with respect to the time axis '$t \to$' and the ordinates $W(t)$ throughout $\mathcal{B}$. Suppose a customer arrives at t_A^- and $W(t_A^-) = y \geq 0$; the SP then has coordinates (t_A^-, y). The SP immediately jumps an amount $\underset{dis}{=} S$, ending at $(t_A, y + S)$. Let

$$t_y = \min\{t > t_A | W(t) = y\}.$$

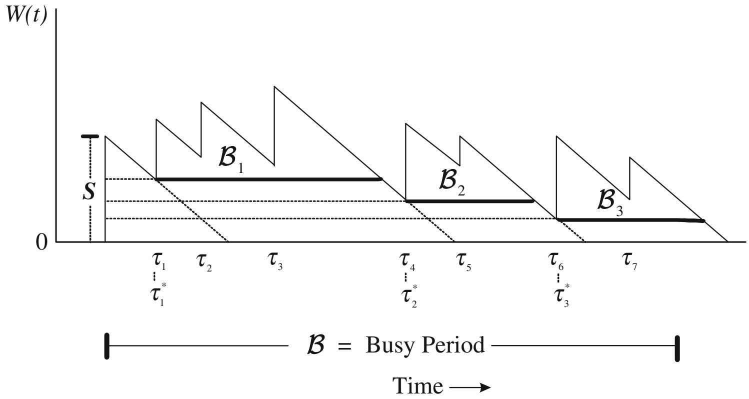


Fig. 3.6 Busy period $\mathcal{B} \underset{dis}{=} S + \sum_{i=1}^{N_S} \mathcal{B}_i$. $\mathcal{B}_i \underset{dis}{=} \mathcal{B}$, $i = 1, \ldots, N_S$. N_S = number of "tagged" (pseudo) arrivals in $\mathcal{B}$. Here $N_S = 3$. $N_S \underset{dis}{=}$ number of arrivals during S. Tagged arrival times are $\tau_1^* = \tau_1$, $\tau_2^* = \tau_4$, $\tau_3^* = \tau_6$. Tagged arrivals 1, 4, 6 during $\mathcal{B}$ initiate sub-busy periods $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$

The interval (t_A, t_y) having length $t_y - t_A$ is a busy period $\mathcal{B}$ if $y = 0$; (t_A, t_y) is a sub-busy period $\underset{dis}{=} \mathcal{B}$ if $y > 0$. The time interval $t_y - t_A$ is independent of y, since the initial SP jump at t_A is $\underset{dis}{=} S$. We utilize this partition of $\mathcal{B}$ to study its structure. (The foregoing definition of busy period is equivalent to the usual definition made for $y = 0$ only, e.g., [140]; see also p.10 and p. 102 in [84].)

Consider $\mathcal{B}$ within which $n \geq 1$ customers arrive. Denote their arrival times within $\mathcal{B}$ by $\tau_1 < \tau_2 < \cdots < \tau_n$, implying that τ_1 occurs within the initial service time S. Then $W(\tau_i^-) > 0, i = 1, 2, \ldots$. Define $\tau_1^* = \tau_1$ and $\tau_j^* = \min\{t > \tau_{j-1}^* | 0 < W(t) < W(\tau_{j-1}^*)\}$, $j = 2, \ldots, n$. Due to the memoryless property of the inter-arrival times and since $\frac{d}{dt}W(t) = -1$ $(W(t) > 0)$, the ordinates $\{W(\tau_j^{*-})\}_{j=1,\ldots,n}$ are distributed the same as n customer arrival times *during the first service time* S of $\mathcal{B}$. We call the customers that arrive at $\{\tau_j^*\}_{j=1,\ldots,n}$ "*tagged*" or "*pseudo*" *arrivals with respect to the initial* S *of* $\mathcal{B}$ (see Fig. 3.6).

Let N_S denote the number of *tagged arrivals* **during** $\mathcal{B}$. Then N_S is distributed as *the number of arrivals to the system during the* **service time** S. Tagged arrivals initiate their own sub-busy periods starting at $\{(\tau_n^{-*}, W(\tau_n^{-*}))\}_{n=1,\ldots,N_S}$ similar to $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ depicted in Fig. 3.6 (where $\tau_1^* = \tau_1$, $\tau_2^* = \tau_4$, $\tau_3^* = \tau_6$). The tagged arrivals during $\mathcal{B}$ are customers 1, 4 and 6, which initiate $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$, respectively. Note that $(\tau_n^{-*}, W(\tau_n^{-*}))_{n=1,\ldots,N_S}$ are strict descending ladder points [74] within $\mathcal{B}$. Then

$$\mathcal{B} = S + \sum_{i=1}^{N_S} \mathcal{B}_i, \tag{3.83}$$

where $\{\mathcal{B}_i\}_{i=1,2,\ldots}$ are i.i.d. sub-busy periods $\underset{dis}{=} \mathcal{B}$, and independent of N_S. Equation (3.83) is known, and is usually derived by different, but equivalent, reasoning (see Example 5.27, pp. 347–349 in Ross [125]). From (3.83), we obtain

$$E(\mathcal{B}) = E(S) + E(N_S)E(\mathcal{B}) = E(S) + \lambda E(S)E(\mathcal{B}) \tag{3.84}$$

which gives $E(\mathcal{B})$ as in (3.80).

Also, we can obtain (3.80) by recursively substituting for $\mathcal{B}_i$ in (3.83). This gives an infinite series of terms

$$\mathcal{B} \underset{dis}{=} S + \sum_{i=1}^{N_S} S_i + \sum_{i=1}^{N_S}\sum_{j=1}^{N_S} S_{i,j} + \sum_{i=1}^{N_S}\sum_{j=1}^{N_S}\sum_{k=1}^{N_S} S_{i,j,k} + \cdots$$

where $S_i, S_{i,j}, S_{i,j,k}, \ldots$, are $\underset{dis}{=} S$. Assuming $0 < \lambda E(S) < 1, \{P_0, f(x)\}_{x>0}$ exists and $\mathcal{B} < \infty$ $(a.s.)$. Then

$$\begin{aligned} E(\mathcal{B}) &= E(S) + \geq(E(S))^2 + \geq^2(E(S))^3 + \cdots \\ &= E(S) \cdot (1 + \geq E(S) + (\geq E(S))^2 + \cdots) \\ &= \frac{E(S)}{1 - \lambda E(S)}. \end{aligned}$$

If $\lambda E(S) \geq 1$ it is possible for the busy period to be infinite. Then its mean and variance do not exist.

We compute the known formula for the variance of $\mathcal{B}$ assuming it exists from (3.83) and the definition $Var(\mathcal{B}) = E(\mathcal{B}^2) - (E(\mathcal{B}))^2$,for completeness; we intend to use the result for $E(\mathcal{B}^2)$ when discussing M/G/1 priority queues in Sect. 3.14 (see p. 349 in [125]).

To compute $E(\mathcal{B}^2)$, we first obtain a formula for $\mathcal{B}^2$ from (3.83) as

$$\mathcal{B}^2 = S^2 + 2S\sum_{i=1}^{N_S} \mathcal{B}_i + \left(\sum_{i=1}^{N_S} \mathcal{B}_i\right)^2.$$

Conditioning on $S = s$, gives the conditional expected value

$$E(\mathcal{B}^2|S = s) = s^2 + 2sE\left(\sum_{i=1}^{N_s} \mathcal{B}_i\right) + E\left(\left(\sum_{i=1}^{N_s} \mathcal{B}_i\right)^2\right).$$

In the second term on the right $\sum_{i=1}^{N_s} \mathcal{B}_i$ is a compound Poisson process with rate λ (see p. 346 in [125]). Thus

$$E\left(\sum_{i=1}^{N_s} \mathcal{B}_i\right) = \lambda s E(\mathcal{B}).$$

The third term on the right is

$$\begin{aligned} E\left(\left(\sum_{i=1}^{N_s} \mathcal{B}_i\right)^2\right) &= E\left(\sum_{i=1}^{N_s} \mathcal{B}_i^2 + \sum_{i \neq j=1}^{N_s} \mathcal{B}_i \mathcal{B}_j\right) \\ &= \lambda s E(\mathcal{B}^2) + E(N_s(N_s - 1)\mathcal{B}_i \mathcal{B}_j) \\ &= \lambda s E(\mathcal{B}^2) + E(N_s(N_s - 1))(E(\mathcal{B}))^2 \\ &= \lambda s E(\mathcal{B}^2) + (\lambda s)^2 (E(\mathcal{B}))^2. \end{aligned}$$

since

$$E(N_s(N_s - 1)) = \sum_{n=2}^{\infty} \frac{n(n-1)e^{-\lambda s}(\lambda s)^n}{n!} = (\lambda s)^2.$$

Thus

$$E(\mathcal{B}^2|S = s) = s^2 + 2\lambda s^2 E(\mathcal{B}) + \lambda s E(\mathcal{B}^2) + (\lambda s)^2 (E(\mathcal{B}))^2.$$

Unconditioning with respect to the service time distribution, substituting from (3.80) and simplifying yields

$$E(\mathcal{B}^2) = \frac{E(S^2)(1 + \lambda E(\mathcal{B}))^2}{1 - \lambda E(S)} = \frac{E(S^2)}{(1 - \lambda E(S))^3} = \frac{E(S^2)}{(1 - \rho)^3}, \quad (3.85)$$

where $\rho := \lambda E(S)$.

Since $Var(\mathcal{B}) = E(\mathcal{B}^2) - (E(B))^2$, from (3.80) and (3.85)

$$Var(\mathcal{B}) = \frac{Var(S) + \lambda(E(S))^3}{(1 - \lambda E(S))^3} = \frac{Var(S) + \lambda \rho^3}{P_0^3}. \quad (3.86)$$

3.4.13 Probability Distribution of the Busy Period

Starting from formula (3.83) above, we can proceed as on pp. 211–226 in [104] to derive $F_{\mathcal{B}}(y),\ y > 0 :=$ the cdf of $\mathcal{B}$. Formula (5.169) on p. 226 in [104] gives an explicit expression for $F_{\mathcal{B}}(y),\ y > 0$ as

$$F_{\mathcal{B}}(y) = \int_{s=0}^{y} \sum_{n=1}^{\infty} e^{-\lambda s} \frac{(\lambda s)^n}{n!} b_{(n)}(s) ds,\ y > 0, \tag{3.87}$$

where $b_{(n)}(s) :=$ the n-fold self convolution of $b(s)$. The paragraph following (5.169) therein observes that the "study of the busy period has really been the study of a transient phenomenon", which makes it more complicated than the analysis of a phenomenon in steady state.

3.4.14 Expected Number Served in Busy Period

Let $N_{\mathcal{B}} :=$ the number of customers served in a busy period. Let $A_C :=$ number of arrivals in a busy cycle. Then $N_{\mathcal{B}} = A_C$. Let $A(t)$ denote the number of arrivals to the system during time interval $(0, t)$. We get $E(N_{\mathcal{B}})$ by applying the renewal reward theorem; thus

$$\frac{E(N_{\mathcal{B}})}{E(\mathcal{C})} = \frac{E(A_{\mathcal{C}})}{E(\mathcal{C})} = \lim_{t\to\infty} \frac{A(t)}{t} = \lambda,$$
$$E(N_{\mathcal{B}}) = \lambda E(\mathcal{C}) = \lambda \frac{1}{\lambda P_0} = \frac{1}{P_0}. \tag{3.88}$$

(See Exercise 17, p. 233 in [64].)

Another View for $E(N_{\mathcal{B}})$ using $N_{\mathcal{B}}$ as a Stopping Time
Let S_i, T_i denote the ith service and inter-arrival times during $\mathcal{B}$, respectively, $i = 1, 2, \ldots$. Then $N_{\mathcal{B}} = \min\{n | \sum_{i=1}^{n} (S_i - T_i) \le 0\}$ is a *stopping time* for the sequence $\{(S_i - T_i)\}_{n=1,2,\ldots}$ (see, e.g., Exercise 13, p. 486, and pp. 678–679 in Ross [125]). Since $T_i \underset{dis}{\equiv} \text{Exp}_\lambda$, the excess inter-arrival time at the end of $\mathcal{B}$ is also distributed as Exp_λ due to the memoryless property. Hence $\sum_{i=1}^{N_{\mathcal{B}}} (S_i - T_i)$ ends a distance *below* 0, which is $\underset{dis}{=} \text{Exp}_\lambda$, implying

$$E\left(\sum_{i=1}^{N_{\mathcal{B}}} (S_i - T_i)\right) = -\frac{1}{\lambda}.$$

Applying Wald's equation (aka Wald's identity; see, e.g., p. 47ff in [122]) gives

$$E(N_{\mathcal{B}})\left(E(S)-\frac{1}{\lambda}\right)=-\frac{1}{\lambda}, \tag{3.89}$$

$$E(N_{\mathcal{B}})=\frac{1}{1-\lambda E(S)}=\frac{1}{P_0}. \tag{3.90}$$

We may also write $N_{\mathcal{B}}=\min\{n|\sum_{i=1}^{n}S_i\le\sum_{i=1}^{n}T_i\}$. In this form it is seen that $N_{\mathcal{B}}$ is a stopping time for both sequences $\{S_i\}_{i=1,2,\ldots}$ and $\{T_i\}_{i=1,2,\ldots}$. That is, we observe the r.v.s in the order $S_1, T_1, S_2, T_2, \ldots$ and stop at n in both sequences when the stopping criterion $(\sum_{i=1}^{n}S_i\le\sum_{i=1}^{n}T_i)$ is first satisfied. Thus the event $\{N_{\mathcal{B}}=n\}$ is independent of $S_{n+1}, T_{n+1}, \ldots$. Moreover, since $\mathcal{B}=\sum_{i=1}^{N_{\mathcal{B}}}S_i$ where $S_i \underset{dis}{\equiv} S$, from (3.80) we have

$$E(\mathcal{B})=E(N_{\mathcal{B}})E(S)=\frac{E(S)}{1-\lambda E(S)},$$

which yields (3.90). (Interestingly, $E(S)\times E(N_{\mathcal{B}})$ is an intuitive way of thinking about $E(\mathcal{B})$.)

Note that $\mathcal{C}=\sum_{i=1}^{N_{\mathcal{B}}}T_i$ (one interarrival time precedes each arrival in a *busy cycle*). From $E(\mathcal{C})=1/(\lambda P_0)$ we have

$$E(\mathcal{C})=\frac{1}{\lambda P_0}=E(N_{\mathcal{B}})E(T)=(E(N_{\mathcal{B}}))\frac{1}{\lambda}, \tag{3.91}$$

which also gives (3.90).

We may write

$$N_{\mathcal{B}}=1+\sum_{i=1}^{N_S}N_{\mathcal{B}_i}$$

where $N_{\mathcal{B}_i}\underset{dis}{\equiv}N_{\mathcal{B}}$, and $N_S\equiv$ number of arrivals in the first service time of $\mathcal{B}$ (see Fig. 3.6; one sub-busy period for each arrival during the first service time). Then

$$E(N_{\mathcal{B}})=1+E(N_S)E(N_{\mathcal{B}})=1+\lambda E(S)E(N_{\mathcal{B}}),$$

again leading to (3.90).

In (3.90) if $P_0\lessapprox 1$ (close to 1) corresponding to a very low traffic intensity ρ, then $E(N_{\mathcal{B}})\gtrapprox 1$ (close to 1) meaning most customers in service are alone in the system.

The role of LC in this section, is that the downcrossing rate of level 0 (SP hit rate of 0 from above) is $f(0)$, which implies $E(\mathcal{C}) = \frac{1}{f(0)} = \frac{1}{\lambda P_0}$. Also, applying the stopping-time definition of a busy cycle just preceding (3.89), leads to (3.90).

3.4.15 Inter-Downcrossing Time of a State-Space Level

Consider a sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 3.7). Let d_x denote the time between two successive downcrossings of level $x \geq 0$. Starting at the instant of the first downcrossing of state-space level x, d_x is an interval of a renewal process $\{\mathcal{D}_t(x)\}_{t\geq 0}$ due to exponential inter-arrival times. The renewal rate is $\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x)$ (see Corollary 3.2 in Sect. 3.2.5 above; Theorem 7.1, p. 432 in Ross [125]). Thus,

$$E(d_x) = \frac{1}{f(x)}, x \geq 0 \tag{3.92}$$

where $f(x)$ is the solution of (3.34) and (3.36).

Since $d_0 = \mathcal{C} :=$ busy cycle, $d_0 = \mathcal{B} + \mathcal{I}$ ($\mathcal{B} :=$ busy period; $\mathcal{I} :=$ idle period). Letting $x \downarrow 0$ in (3.92) gives

$$E(d_0) = \frac{1}{f(0)} = E(\mathcal{B}) + E(\mathcal{I}).$$

Thus, using method {3} in Sect. 3.4.10 we get $E(\mathcal{B})$ in (3.80).

3.4.16 Sojourn Below a Level of $\{W(t)\}_{t\geq 0}$

Let b_x denote a sojourn time below, or at, level $x \geq 0$ (Fig. 3.7). Assuming the queue is stable ($\rho < 1$), the proportion of time a sample path spends at or below x, is $\lim_{t\to\infty} E(\mathcal{D}_t(x))/t \cdot E(b_x) = f(x)E(b_x)$, *and is also equal to* the limiting cdf $F(x)$. Hence

$$E(b_x) = \frac{F(x)}{f(x)} \tag{3.93}$$

(see Remark 3.15 below). Letting $x \downarrow 0$, reduces (3.93) to the expected idle period

$$E(b_0) = \frac{F(0)}{f(0)} = \frac{P_0}{\lambda P_0} = \frac{1}{\lambda}.$$

Also, from (3.93)

$$\frac{d}{dx} \ln F(x) = \frac{1}{E(b_x)},$$

which leads to expressions for the cdf $F(x)$ and pdf $f(x)$ $(= F'(x))$ of wait in terms of $E(b_y)$, $0 < y < x$,

$$F(x) = P_0 e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x \geq 0, \tag{3.94}$$

$$f(x) = \frac{P_0}{E(b_x)} e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x > 0\,. \tag{3.95}$$

3.4.17 *Sojourn Above a Level of* $\{W(t)\}_{t \geq 0}$

Let a_x denote a sojourn time above level $x \geq 0$ (Fig. 3.7). Then $a_0 = \mathcal{B}$. By Theorem 1.1 in Sect. 1.6 the down- and upcrossing rates of level x are both equal to $f(x)$, $x \geq 0$. The proportion of time that a sample path spends above x is $\lim_{t\to\infty} (\mathcal{U}_t(x) \cdot E(a_x))/t = \lim_{t\to\infty} (\mathcal{U}_t(x)/t) \cdot E(a_x) = f(x)E(a_x)$, *and is also equal to* $1 - F(x)$. Therefore

$$E(a_x) = \frac{1 - F(x)}{f(x)}, x \geq 0. \tag{3.96}$$

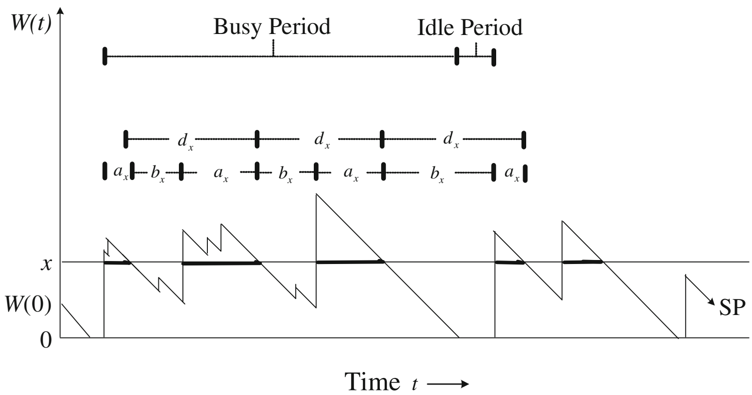


Fig. 3.7 Sample path of $\{W(t)\}_{t\geq 0}$ in M/G/1. Shows inter-downcrossing time d_x, sojourns a_x and b_x, busy and idle periods

Letting $x \downarrow 0$, in (3.96) gives $E(a_0) = (1 - P_0)\,/\,(\lambda P_0) = E(\mathcal{B})$.

Remark 3.15 Formula (3.96) can also be proved using the renewal reward theorem (Sect. 3.4.9), since $\{\mathcal{D}_t(x)\}_{t \geq 0}$ is a renewal process (starting from the first downcrossing of level x) since interarrival times are $\underset{dis}{=}$ Exp_λ having the memoryless property. Thus

$$\frac{E(a_x)}{E(d_x)} = \lim_{t\to\infty} \frac{\text{time spent above } x \text{during}(0, t)}{t} \underset{a.s.}{=} 1 - F(x),$$
$$E(a_x) = E(d_x) \cdot (1 - F(x)) = \frac{1 - F(x)}{f(x)}.$$

We can derive formula (3.93) for $E(b_x)$ similarly.

Proposition 3.3 below shows that *if* $E(a_x) \equiv E(\mathcal{B})$, $x \geq 0$, *then* the absolutely continuous part of $\{P_0, f(x)\}_{x>0}$ has an exponential form. Assume $\rho := \lambda E(S) < 1$.

Proposition 3.3 If $E(a_x) = E(\mathcal{B})$ for all $x \geq 0$, then the steady-state cdf of wait is $F(x) = 1 - \rho e^{-\frac{x}{E(\mathcal{B})}}$ and $\{P_0, f(x)\}_{x>0}$ is given by

$$P_0 = 1 - \rho, \quad f(x) = \lambda P_0 e^{-\frac{x}{E(\mathcal{B})}}, x > 0.$$

Proof If $E(a_x) \equiv E(\mathcal{B})$, $x \geq 0$, then from (3.96)

$$\frac{f(x)}{1 - F(x)} \equiv \frac{1}{E(\mathcal{B})}, x > 0, \tag{3.97}$$
$$\frac{d}{dx} \ln(1 - F(x)) \equiv -\frac{1}{E(\mathcal{B})}, x > 0.$$

Formula (3.97) is the *hazard rate (failure rate)* of the pdf of wait at x. (See Sect. 3.4.18 below.) Integration with respect to x yields

$$1 - F(x) = Ae^{-\frac{x}{E(\mathcal{B})}}, x > 0,$$

where A is a constant. Letting $x \downarrow 0$ gives

$$\begin{aligned} A &= 1 - F(0) = 1 - P_0 = \rho; \\ F(x) &= 1 - \rho e^{-\frac{x}{E(\mathcal{B})}}, x \geq 0. \end{aligned} \tag{3.98}$$

Differentiation of $F(x)$ in (3.98) with respect to $x > 0$ gives

$$f(x) = \frac{1}{E(\mathcal{B})}\rho e^{-\frac{x}{E(\mathcal{B})}} = \frac{1}{\frac{E(S)}{P_0}}\lambda E(S) \cdot e^{-\frac{x}{E(\mathcal{B})}} = \lambda P_0 e^{-\frac{x}{E(\mathcal{B})}}, x > 0. \tag{3.99}$$

■

Remark 3.16 The standard $M_\lambda/M_\mu/1$ queue satisfies the hypothesis of Proposition 3.3 because $S \underset{dis}{=} \text{Exp}_\mu$. All jumps that upcross level x have excess above $x \underset{dis}{=} \text{Exp}_\mu$ by the memoryless property, implying $a_x \underset{dis}{=} \mathcal{B}$, $x \geq 0$ (see Sect. 3.5.6).

3.4.18 Hazard Rate of PDF of Waiting Time

The term **hazard rate**, also called **failure rate**, is usually defined for positive continuous random variables in renewal theory, and failure time of components in reliability models (see, e.g., pp. 1–7 in [66]). In this monograph, we apply the 'hazard rate' to the pdf at x of waiting time (and other state variables, e.g., pdf at x of content of a dam in Sect. 6.2.12 in Chap. 6, etc.). In M/G/1 we may think of sojourn a_x as a 'lifetime' spent above level x. Thus, $\phi(x)E(a_x) = 1$, where $\phi(x) := f(x)/(1 - F(x))$, the hazard rate at x. Then $E(a_x)$ (E(lifetime above x)) varies inversely with $\phi(x)$. This idea fits the notion of failure rate in reliability models. Let $X :=$ *lifetime of a component* (also called *failure time*). The failure rate at lifetime x is the conditional pdf of lifetime given the lifetime exceeds x. Following pp. 1–4 in [66],

$$\begin{aligned}
\phi(x) &= \lim_{\Delta x \downarrow 0} \frac{P(x < X \leq x + \Delta x | X > x)}{\Delta x} \\
&= \lim_{\Delta x \downarrow 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x P(X > x)} \\
&= \lim_{\Delta x \downarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \frac{1}{(1 - F(x)} \\
&= \frac{f(x)}{(1 - F(x))}.
\end{aligned}$$

For the pdf of waiting time, the dimension of $\phi(x)$ is the same as that of $f(x)$, viz., $1/[Time]$. In other stochastic models the dimension of $\phi(x)$ is the same as that of the pdf of the state variable.

3.4.19 Sojourn Above a Level and Distribution of Wait

Proposition 3.4 below relates $E(a_y)$, $y \in (0, x)$, to $F(x)$ and $f(x)$, $x > 0$. In general $E(a_y)$ varies with $y > 0$. (However, in M/M/1 $E(a_y) \equiv E(\mathcal{B})$, $y > 0$.)

Proposition 3.4 For the M/G/1 queue in equilibrium ($\rho < 1$),

$$F(x) = 1 - \rho \cdot e^{-\int_{y=0}^{x} \frac{1}{E(a_y)} dy}, x \geq 0. \tag{3.100}$$

$$f(x) = \frac{\rho}{E(a_x)} \cdot e^{-\int_{y=0}^{x} \frac{1}{E(a_y)} dy}, x > 0. \tag{3.101}$$

Proof Consider a sample path of $\{W(t)\}_{t \geq 0}$. The pdf $f(x)$ is the SP upcrossing (and downcrossing) rate of level x. Hence the long-run proportion of time $\{W(t)\}_{t \geq 0}$ spends above level x is

$$f(x)E(a_x) = 1 - F(x).$$

Thus

$$\frac{f(x)}{1 - F(x)} = \frac{1}{E(a_x)}, x > 0. \tag{3.102}$$

The term $f(x)/(1 - F(x))$ is the hazard rate of the waiting time at level x (see Sect. 3.4.18 above). From (3.102)

$$\frac{d}{dx} \ln(1 - F(x)) = -\frac{1}{E(a_x)}, x > 0.$$

Integrating with respect to x gives

$$1 - F(x) = Ae^{-\int_{y=0}^{x} \frac{1}{E(a_y)} dy},$$

where A is a constant. Letting $x \downarrow 0$, gives

$$A = 1 - F(0^+) = 1 - F(0) = 1 - P_0 = \rho.$$

Hence we obtain (3.100); (3.101) follows by taking $dF(x)/dx$ in (3.100). ■

Equivalence of Formulas for $F(x)$ in Terms of $E(b_x)$ and $E(a_x)$
We now check that the right sides of (3.100) and (3.94) are both equal to $F(x)$, $x > 0$, and therefore to each other. Thus

$$P_0 e^{\int_{y=0}^{x} \frac{f(y)dy}{F(y)}} = P_0 e^{\int_{y=0}^{x} d\ln F(y)} = P_0 e^{\ln\left(\frac{F(x)}{F(0)}\right)}$$
$$= P_0 \frac{F(x)}{F(0)} = F(x), x \geq 0,$$

and

$$1 - \rho e^{-\int_{y=0}^{x} \frac{f(y)dy}{1-F(y)}} = 1 - \rho e^{\int_{y=0}^{x} d\ln(1-F(y))}$$
$$= 1 - \rho e^{\ln\left(\frac{1-F(x)}{1-F(0)}\right)} = 1 - \rho\left(\frac{1-F(x)}{1-F(0)}\right)$$
$$= 1 - \rho\left(\frac{1-F(x)}{\rho}\right) = F(x), x \geq 0,$$

proving the equivalence of the two formulas.

3.4.20 Computing $F(x)$ via $E(a_x)$

Suppose we do not have an explicit formula for $F(x)$ in a particular M/G/1 model. We can compute $E(a_x)$ (reciprocal of hazard rate) either analytically or using simulation, and apply formula (3.100) to obtain an analytical formula for $F(x)$, or an estimate of $F(x)$.

Analytical We can get analytic expressions for $E(a_x)$ in some models. We know $E(a_x) = \mathcal{B}$ in M/M/1. In general, however, $E(a_x)$ may be difficult to compute analytically. Example 3.7 below computes $E(a_x)$ analytically in an $\text{M}_\lambda/\text{Erl}_{2,\mu}/1$ queue. ($\text{Erl}_{k,\mu}$:= Erlang random variable; see Gamma distribution in Table 2.2, p. 66 in [125].)

Example 3.7 In $\text{M}/\text{Erl}_{2,\mu}/1$ with arrival rate λ, $E(S) = 2/\mu$ and $\rho = \lambda \cdot \frac{2}{\mu} < 1$, consider a sample path of $\{W(t)\}_{t\geq 0}$ (see also Example 3.2 in Sect. 3.3). $S \underset{dis}{=} \text{Erl}_{2,\mu}$ is the sum of two i.i.d. Exp_μ random variables; we call these phase 1 and phase 2 respectively. Either $a_x \underset{dis}{=} \mathcal{B}$ for the **standard** $\text{M}_\lambda/\text{Erl}_{2.\mu}/1$ queue, or $a_x \underset{dis}{=} \mathcal{B}$ for the $\text{M}_\lambda/\text{Erl}_{2,\mu}/1$ queue **where zero-wait customers have** $S \underset{dis}{=} \text{Exp}_\mu$ (i.e., special (exceptional) service for 'zero-wait' arrivals), depending on the initial service-time phase that covers x. That is, a_x's initial SP upcrossing of x covers x either during phase 1 or during phase 2 of the $\text{Erl}_{2,\mu}$ service time. If phase 1 covers x, then the excess jump above $x \underset{dis}{=} \text{Erl}_{2,\mu}$, due to the memoryless property of Exp_μ. If phase 2 covers x, then the excess jump above $x \underset{dis}{=} \text{Exp}_\mu$. If phase 1 covers x, applying (3.82) we get

$E(\mathcal{B}) = \frac{2}{\mu - 2\lambda}$. If phase 2 covers x then the initial $S \underset{dis}{=} Exp_\mu$; this results in an $M_\lambda/Erl_{2,\mu}/1$ with $E(a_x) = E(\mathcal{B}) = \frac{1}{\mu-2\lambda}$, because $a_x \underset{dis}{=} \mathcal{B}$ in which zero-wait customers receive "special" service Exp_μ different from the rest of the service times which are $Erl_{2,\mu}$s (see Sect. 3.6.1 below). Thus,

$$E(a_x) = p_1(x)\left(\frac{2}{\mu - 2\lambda}\right) + p_2(x)\left(\frac{1}{\mu - 2\lambda}\right),$$

where $p_i(x) = P$ (phase i of an SP jump covers $x|$SP upcrosses x), $i = 1, 2$. From (3.100)

$$F(x) = 1 - \rho \exp\left(-\int_{y=0}^{x} \frac{1}{p_1(y)\left(\frac{2}{\mu-2\lambda}\right) + p_2(y)\left(\frac{1}{\mu-2\lambda}\right)} dy\right). \tag{3.103}$$

In Example (3.2), Eq. (3.41) for M_λ /$Erl_{2,\mu}$/1 yields

$$\begin{aligned} p_1(x) &= \frac{\lambda\left(P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y) dy\right)}{f(x)} \\ &= \frac{P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y) dy}{P_0 e^{-\mu x}(1+\mu x) + \int_{y=0}^{x} e^{-\mu(x-y)}(1+\mu(x-y)) f(y) dy, x>0}, \qquad (3.104) \\ p_2(x) &= \frac{P_0 e^{-\mu x}\mu x + \int_{y=0}^{x} e^{-\mu(x-y)}\mu(x-y) f(y) dy}{P_0 e^{-\mu x}(1+\mu x) + \int_{y=0}^{x} e^{-\mu(x-y)}(1+\mu(x-y)) f(y) dy, x>0}, \; x > 0. \end{aligned}$$

where $\{P_0, f(y)\}_{y \ge 0}$ is specified in (3.42).

Example 3.8 In Example 3.7, $S \underset{dis}{=} Erl_{2,\mu}$ and $E(B) = E(S)/P_0 = \frac{2/\mu}{1-\lambda(2/\mu)} = \frac{2}{\mu-2\lambda}$. Then $E(a_x) = p_1(x)\left(\frac{2}{\mu-2\lambda}\right) + p_2(x)\left(\frac{1}{\mu-2\lambda}\right)$ where $p_1(x) + p_2(x) = 1$ and $p_1(x) > 0$, $p_2(x) > 0$. Thus $E(a_x) < \frac{2}{\mu-2\lambda} = E(B)$.

Alternatively, we could *estimate* $p_1(x)$, $p_2(x)$, $x > 0$, from a simulated sample path of $\{W(t)\}_{t \ge 0}$. Then substitute the estimated values into (3.103) to estimate $F(x)$, $x > 0$. This **hybrid technique** combines estimated values from simulation and analytical results.

Simulation to Estimate $E(a_y)$

To estimate $E(a_y)$, $y \in [0, x]$, simulate a single sample path of $\{W(t)\}$, $0 \le t \le T_{sim}$, where T_{sim} is "large". We utilize $\{W(t)\}_{0 \le t \le T_{sim}}$ to *estimate* $E(a_{y_j})$ where y_j is a level of a state-space partition of $[0, x]$: $0 = y_0 < y_1 < \cdots < y_N$, and choose the subintervals to be "small", e.g., $y_{j+1} - y_j \equiv h > 0$, $j = 0, \ldots, N-1$ (depending on the required accuracy). Take $N = \lfloor x/h \rfloor$ where

$\lfloor \alpha \rfloor$ denotes the greatest integer $\leq \alpha$. Suppose in the simulated sample path there are M_j sojourns above level y_j during $[0, T_{sim}]$; let their observed *values* be $a_{y_j,1}, \ldots, a_{y_j,M_j}$. Assume T_{sim} is sufficiently large so that each M_j is "large" enough for the required accuracy. Then estimate $E(a_{y_j})$ using

$$\widehat{E}(a_{y_j}) = \frac{1}{M_j} \sum_{i=0}^{M_j} a_{y_j,i}, \; j = 0, 1, \ldots, N.$$

We can estimate the value of $\int_{y=0}^{x} \frac{1}{E(a_y)} dy$ in (3.100) by

$$\int_{y=0}^{x} \frac{1}{E(a_y)} dy \underset{est}{=} h \sum_{j=0}^{N} \frac{1}{\widehat{E}(a_{y_j})}.$$

(We consider LC estimation in Chap. 9.)

Intuitive Meaning of the Hazard Rate
Denote the hazard rate of wait at x by $\phi(x)$. From (3.102), a plausible estimate of $\phi(x)$ is

$$\widehat{\phi}(x) = \frac{1}{\widehat{E}(a_x)}. \tag{3.105}$$

By definition

$$\phi(x)dx = P(W_q \in (x, x+dx) | W_q > x) = \frac{P(x < W_q < x + dx)}{P(W_q > x)},$$

where W_q is the teady-state queue wait (see, e.g., p. 299 in [125]). Formula (3.102) suggests an intuitive meaning based on $\phi(x) = 1/E(a_x)$, i.e., $\phi(x)$ is large iff $E(a_x)$ is small, and $\phi(x)$ is small iff $E(a_x)$ is large. This suggests studying connections between hazard rates of state random variables, and their sample-path expected sojourn times with respect to state-space levels in related stochastic models. (See Sect. 3.4.18 for pertinent comments.)

3.4.21 Events During an Inter-downcrossing Time

Consider $\{W(t)\}_{t \geq 0}$. We derive formulas for E(number of SP downcrossings of an arbitrary level $x \geq 0$) during d_y, $y \geq 0$, and E(number of customer arrivals) during d_y, $y \geq 0$; see Sect. 3.4.15 and Fig. 3.1. (See also Sect. 3.5 below regarding the M/M/1 queue.)

Consider a sample path of $\{W(t)\}_{t\geq 0}$ for an M/G/1 queue with $\rho < 1$. Denote the steady-state pdf of wait by $\{P_0, f(x)\}_{x\geq 0}$. Fix level $y \geq 0$. Let $\mathcal{D}_{d_y}(x)$ denote the number of SP downcrossings of an **arbitrary** level $x \geq 0$ during a sample-path inter-downcrossing d_y.

Proposition 3.5

$$E(\mathcal{D}_{d_y}(x)) = \frac{f(x)}{f(y)}. \tag{3.106}$$

Proof Since $\{\mathcal{D}_t(y)\}_{t\geq 0}$ is a renewal process starting at the first downcrossing of y,

$$\frac{E(\mathcal{D}_{d_y}(x))}{E(d_y)} = \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = f(x),$$

by Theorem 1.1 and the renewal reward theorem. Also, $E(d_y) = 1/f(y)$ by the elementary renewal theorem. Equation (3.106) follows. ■

Observe that for arbitrary x, $E(\mathcal{D}_{d_y}(x))/E(d_y)$ is invariant for all $y \geq 0$. For example, if $y = 0$ then $d_0 = \mathcal{C}$ (busy cycle), and $E(\mathcal{D}_{d_0}(x))/E(d_0) = E(\mathcal{D}_{d_0}(x))/(1/(\lambda P_0)) = f(x)$.

If $y = 0$ and $x \downarrow 0$ then $E(\mathcal{D}_{d_0}(0)) = E(\mathcal{D}_{\mathcal{C}}(0)) = 1$. Thus

$$E(\mathcal{D}_{d_0}(0))/(1/(\lambda P_0)) = 1/(1/(\lambda P_0)) = \lambda P_0 = f(0),$$

which is compatible with Theorem 1.1.

Let A_{d_y} := number of customer arrivals during d_y. Let $A(t)$:= number of customer arrivals during $(0, t)$.

Proposition 3.6

$$E(A_{d_y}) = \frac{\lambda}{f(y)}, \; y \geq 0. \tag{3.107}$$

Proof

$$\frac{E(A_{d_y})}{E(d_y)} = \frac{E(A_{d_y})}{1/f(y)} = \lim_{t\to\infty} \frac{A(t)}{t} = \lambda,$$

by Theorem 1.1 and the renewal reward theorem, resulting in (3.107). ■

Letting $y \downarrow 0$ in 3.106 gives

$$E(A_{d_0}) = E(A_{\mathcal{C}}) = E(N_B) = \frac{\lambda}{f(0)} = \frac{\lambda}{\lambda P_0} = \frac{1}{P_0},$$

which is an additional proof of Eq. (3.88).

3.4.22 Boundedness of PDF in Steady State

Why is it potentially useful to know that in the limiting pdf of wait $\{P_0, f(x)\}_{x>0}$, $f(x)$ is bounded by a finite quantity? Suppose we want to estimate $f(x)$ in an analytically intractable M/G/1 model by means of simulation of a sample path. It would be helpful to know this fact when writing a computer program for the simulation.

In the standard M/G/1 queue let the arrival rate be λ, let S have cdf $B(y)$, $y > 0$, and $\rho < 1$. Assume $B(\cdot)$ is absolutely continuous.

Proposition 3.7

$$f(x) \leq \lambda, x > 0. \tag{3.108}$$

Proof (1) In the integral equation for $\{P_0, f(x)\}_{x>0}$ (1.8), repeated here for convenience,

$$f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y) dy, x > 0.$$

$\overline{B}(x) < 1$, $x > 0$ (for any cdf $H(\cdot)$, $0 \leq H(x) \leq 1$, where $H(x)$ is right-continuous and monotone increasing). Thus

$$f(x) < \lambda P_0 + \lambda \int_{y=0}^{x} f(y) dy = \lambda \left(P_0 + \int_{y=0}^{x} f(y) dy \right) = \lambda F(x) \leq \lambda, x > 0.$$

(2) On the right side of the alternative form of the LC integral equation (3.43) (repeated here)

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} B(x-y) f(y) dy, x > 0.$$

the subtracted term is > 0. Thus

$$f(x) < \lambda F(x) \leq \lambda.$$

(3) Consider a sample path of $\{W(t)\}_{t\geq 0}$. Let $\mathcal{D}_t(x)$ and $A(t)$ denote the number of SP downcrossings of level x, and number of arrivals to the system during $(0,t)$, respectively. Examination of the sample path yields $\mathcal{D}_t(x) < A(t), x \geq 0, t > 0$, $(a.s.)$. Hence

$$f(x) = \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \leq \lim_{t\to\infty} \frac{A(t)}{t} = \lambda,$$

since $\{A(t)\}$ is a Poisson process with rate λ. ■

Example 3.9 In $M_\lambda/M_\mu/1$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0$, $P_0 = 1-\rho > 0$ (Sect. 3.5.1). Both $0 < P_0 < 1$ and $0 < e^{-(\mu-\lambda)x} < 1$, $x > 0$. Therefore $f(x) < \lambda P_0, x > 0$.

Inequality (3.108) also holds in: the workload-bounded M/G/1 queue (Sect. 3.16); the M/D/1 queue (Sect. 3.12); and others.

3.5 M/M/1 Queue

We now derive some steady-state results for the standard M/M/1 queue with FCFS (first come first served) discipline. Some well-known results are included to develop facility with LC and reinforce intuitive background. Let λ := arrival rate, service time $S \underset{dis}{=} \text{Exp}_\mu$, $\overline{B}(x) = e^{-\mu x}, x \geq 0$, $B(x) = 1 - e^{-\mu x}, x \dot{\geq} 0$, $\rho := \lambda E(S) = \lambda/\mu < 1$.

3.5.1 Waiting Time PDF and CDF

Consider a sample path of $\{W(t)\}_{t\geq 0}$ (e.g., Fig. 3.5). From the basic LC integral equation (3.34), or Fig. 1.6 in Sect. 1.7, we get

$$f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy,\ x > 0, \tag{3.109}$$

where $\{P_0, f(x)\}_{x>0}$ is the steady-state pdf of wait.

Differentiating both sides of (3.109) with respect to x, yields the ordinary differential equation

$$f'(x) + (\mu - \lambda) f(x) = 0, x > 0, \tag{3.110}$$

with solution

$$f(x) = Ae^{-(\mu-\lambda)x}, x > 0; \tag{3.111}$$

the constant A is determined by letting $x \downarrow 0$ in both (3.109) and (3.111). Thus $A = f(0^+) = \lambda P_0$, giving

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0, \tag{3.112}$$

where, for the standard M/G/1 (see e.g., Eq. (3.62))

$$P_0 = 1 - \rho = 1 - \frac{\lambda}{\mu}. \tag{3.113}$$

We may also compute P_0 by substituting (3.112) into the normalizing condition,

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1, \tag{3.114}$$

which yields (3.113) directly.

From (3.112) the cdf of wait is

$$F(x) = P_0 + \int_{y=0}^{x} \lambda(1-\rho)e^{-(\mu-\lambda)y}dy = 1 - \rho e^{-(\mu-\lambda)x}, x > 0. \tag{3.115}$$

3.5.2 System Time PDF and CDF

Let σ denote the system time, $f_\sigma(x)$ its pdf, $F_\sigma(x)$ its cdf, $x > 0$ (see Sect. 3.3.2). Since $\sigma = W_q + S$, we obtain

$$\begin{aligned} P(\sigma > x) &= P(S > x|W_{q=0})P_0 + \textstyle\int_{y=0}^{x} P(S > x-y|W_q = y)f(y)dy \\ &\qquad + P(W_q > x) \\ &= P_0 e^{-\mu x} + \lambda P_0 \textstyle\int_{y=0}^{x} e^{-(\mu-\lambda)y} e^{-\mu(x-y)}dy + \int_{y=x}^{\infty} \lambda P_0 e^{-(\mu-\lambda)y}dy \\ &= \frac{P_0}{1-\frac{\lambda}{\mu}} e^{-(\mu-\lambda)x} = e^{-(\mu-\lambda)x}, x > 0. \end{aligned} \tag{3.116}$$

We can also obtain (3.116) using Eq. (3.61) (or equivalently Eq. (3.72)). Thus $\sigma \underset{dis}{=} \text{Exp}_{\mu-\lambda}$, i.e.,

$$\begin{aligned} f_\sigma(x) &= (\mu-\lambda)\, e^{(\mu-\lambda)x}, x > 0 \\ F_\sigma(x) &= 1 - e^{(\mu-\lambda)x}, x \geq 0. \end{aligned} \tag{3.117}$$

Additionally, we can obtain $f_\sigma(x)$ directly in terms of $f(x)$ using (3.74), thus getting (3.117) similarly as in Example 3.5 in Sect. 3.4.7.

3.5.3 Number in System Probability Distribution

Let N denote the number of units in the M/M/1 system at an arbitrary time point in the steady state. Let $P(N = n) = P_n, n = 0, 1, \ldots$ (see Sect. 3.4.8). We obtain the distribution of N by conditioning on W_q, or on σ.

Conditioning on W_q, substitute $f(x)$ in (3.112) into (3.76), getting

$$\begin{aligned} P_n &= \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!} \lambda P_0 e^{-(\mu-\lambda)y} dy \\ &= P_0 \left(\frac{\lambda}{\mu}\right)^n \int_{y=0}^{\infty} e^{-\mu y} \frac{(\mu y)^{n-1}}{(n-1)!} \mu dy = P_0 \rho^n, n = 0, 1, \ldots , \end{aligned}$$

since the integrand $e^{-\mu y}(\mu y)^{n-1}\mu/(n-1)!$ is the pdf of $\text{Erl}_{n,\mu}$ (see formula (3.38) in Example 3.2, Sect. 3.3).

The normalizing condition $\sum_{n=0}^{\infty} P_n = 1$ yields $P_0 \sum_{n=0}^{\infty} \rho^n = 1$, whence $P_0 = 1 - \rho$, giving the well-known geometric distribution

$$P_n = P_0 (1 - P_0)^n = (1 - \rho)\rho^n, n = 0, 1, \ldots . \tag{3.118}$$

Conditioning on σ, substitute $f_\sigma(x)$ from (3.117) into (3.77) in Sect. 3.4.8, getting

$$\begin{aligned} P_n &= \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\lambda y)^n}{n!} (\mu - \lambda) e^{-(\mu-\lambda)y} dy \\ &= \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\mu y)^n}{n!} \mu dy \\ &= \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n = (1 - \rho)\rho^n, n = 0, 1, \ldots, \end{aligned}$$

because the integrand $e^{-\lambda y}(\mu y)^n \mu/n!$ is the pdf of $\text{Erl}_{n+1,\mu}$; P_n so derived is consistent with (3.118).

E(**number in system**) The right tail probability is $P(N \geq n) = \rho^n, n = 0, \ldots$. Thus

$$E(N) = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} \rho^n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}, \tag{3.119}$$

which agrees with the M/G/1 result $E(N) = L = \frac{\lambda^2 E(S^2)}{2P_0} + \rho$ in (3.65), since $E(S^2) = 2/\mu^2$ when $S \underset{dis}{=} Exp_\mu$.

Remark 3.17 A classical way to derive P_n, $n = 0, 1, \ldots$, in M/M/1 is via birth and death processes (e.g., pp. 49–55 in [84]; Sect. 6.3, p. 374 and Example 6.14, p. 395 in [125]; and others). Using the birth-death derived values of P_n, the pdf $\{P_0, f(x)\}_{x>0}$ of wait is then derived by conditioning on N. Here, we reason in the opposite direction: first derive the pdf $\{P_0, f(x)\}_{x>0}$ or $f_\sigma(x)$, $x > 0$, then condition on W_q or on σ to derive the values of P_n, $n = 1, 2, \ldots$. Similar remarks apply to other exponential models, like multiple server M/M/c queues (Chap. 4).

3.5.4 Expected Busy Period

The $M_\lambda/M_\mu/1$ queue is an $M_\lambda/G/1$ queue having $E(S) = \frac{1}{\mu}$. Substituting $\frac{1}{\mu}$ into (3.80) in Sect. 3.4.10 gives the well-known result

$$E(\mathcal{B}) = \frac{E(S)}{P_0} = \frac{1}{\mu\left(1 - \frac{\lambda}{\mu}\right)} = \frac{1}{\mu - \lambda}. \tag{3.120}$$

3.5.5 CDF and PDF of Busy Period in M/M/1

Applying formula (3.87) in Sect. 3.4.13 we obtain, since the n-fold convolution of $b(y)$ $(:= \mu e^{-\mu y})$ is $b_{(n)}(y) \underset{dis}{=} \text{Erl}_{n,\mu}(y)$. The cdf of $\mathcal{B}$ is

$$\begin{aligned}
F_{\mathcal{B}}(x) &= \int_{y=0}^{x} \sum_{n=1}^{\infty} \frac{e^{-\lambda y}\,(\lambda y)^{n-1}}{n!} \frac{e^{-\mu y}(\mu y)^{n-1}\mu}{(n-1)!} dy \\
&= \int_{y=0}^{x} e^{-(\lambda+\mu)y} \frac{\mu}{\sqrt{\lambda\mu}y} \sum_{n=1}^{\infty} \frac{\left(\sqrt{\lambda\mu}y\right)^{2n-1}}{n!\,(n-1)!} dy \\
&= \int_{y=0}^{x} e^{-(\lambda+\mu)y} \sqrt{\frac{\mu}{\lambda}} \frac{1}{y} \sum_{n=1}^{\infty} \frac{\left(\frac{2\sqrt{\lambda\mu}y}{2}\right)^{2n-1}}{n!\,(n-1)!} dy, \; x > 0,
\end{aligned}$$

which yields $F_{\mathcal{B}}(x)$ and pdf $f_{\mathcal{B}}(x)$ of $\mathcal{B}$ as

$$F_{\mathcal{B}}(x) = \int_{y=0}^{x} \frac{\sqrt{\mu/\lambda}e^{-(\lambda+\mu)y} I_1(2\sqrt{\lambda\mu}y)}{y} dy, \tag{3.121}$$

$$f_{\mathcal{B}}(x) = \frac{\sqrt{\mu/\lambda}e^{-(\lambda+\mu)x} I_1(2\sqrt{\lambda\mu}x)}{x}, x > 0, \tag{3.122}$$

where $I_1(z) :=$ *modified Bessel function of the first kind of order* 1 given by

$$I_1(z) = \sum_{n=1}^{\infty} \frac{(z/2)^{2n-1)}}{(n-1)!n!}$$

(see, e.g., pp. 101–102 in Gross et al. [84]).

3.5.6 Geometric Derivation of CDF and PDF of Wait

Consider a sample path of $\{W(t)\}_{t\geq 0}$ in M/M/1. Let $\mathcal{B}$ denote a busy period. Given that the SP upcrosses level x, the sojourn above x is $a_x \underset{dis}{=} \dot{\mathcal{B}}$, *independent* of $x \geq 0$, due to the memoryless property of Exp_μ (Fig. 3.8). (See Proposition 3.4 in Sect. 3.4.19; also paragraph following "Key Question" in Sect. 1.5.2.)

Substituting $E(\mathcal{B})$ for $E(a_x)$ in formulas (3.100) and (3.101), and applying (3.120) yields

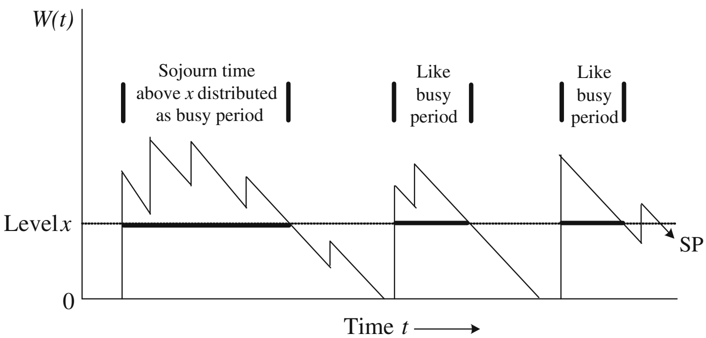


Fig. 3.8 Sample path of $\{W(t)\}_{t\geq 0}$ in $\text{M}_\lambda/\text{M}_\mu/1$ queue showing $a_x \underset{dis}{=} \mathcal{B}$. SP excess jumps above x are $\underset{dis}{=} \text{Exp}_\mu$

$$F(x) = 1 - \rho e^{-\frac{x}{E(\mathcal{B})}} = 1 - \rho e^{-(\mu-\lambda)x}, x \geq 0, \tag{3.123}$$
$$f(x) = \lambda(1-\rho)e^{-(\mu-\lambda)x} = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0. \tag{3.124}$$

The M/M/1 model satisfies Proposition 3.3 in Sect. 3.4.17.

3.5.7 Inter-crossing Time of Level x

We now consider d_x, b_x, a_x, defined in Sects. 3.4.15, 3.4.16 and 3.4.17, respectively. We look at the time between SP successive upcrossings (inter-*upcrossing* time), and E(number of SP crossings of a level) during a busy cycle or during sojourns above or below an arbitrary level.

Inter-downcrossing Time of Level x
We have

$$d_x = b_x + a_x,\ E(d_x) = E(b_x) + E(a_x).$$

In M/M/1 the inter-arrival and service times are $\underset{dis}{=}$ Exp_λ and Exp_μ, respectively. For fixed $x \geq 0$, successive triplets $\{d_{x,n}, b_{x,n}, a_{x,n}\}_{n=1,2,...}$ form a sequence of i.i.d. random variables ($d_{x,n} \underset{dis}{=} d_x$, $b_{x,n} \underset{dis}{=} b_x$, $a_{x,n} \underset{dis}{=} a_x$). Thus $\left\{d_{x,n}\right\}_{n=1,2,...}$ forms a renewal process and $\{b_{x_n}, a_{x,n}\}_{n-1,2,...}$ forms an alternating renewal process. As in Sects. 3.4.15, 3.4.16 and 3.4.17,

$$E(d_x) = \frac{1}{f(x)},\ E(b_x) = \frac{F(x)}{f(x)},\ E(a_x) = \frac{1-F(x)}{f(x)}. \tag{3.125}$$

Since $a_x \underset{dis}{\equiv} \mathcal{B}$

$$E(a_x) = \frac{1}{\mu-\lambda}, x \geq 0, \tag{3.126}$$
$$E(d_x) = \frac{F(x)}{f(x)} + \frac{1}{\mu-\lambda}, x \geq 0. \tag{3.127}$$

Letting $x = 0$ in (3.127) gives $E(d_0) = E(\mathcal{C})$ where $\mathcal{C}$:= busy cycle = d_0. Thus

$$E(\mathcal{C}) = \frac{F(0)}{f(0)} + \frac{1}{\mu-\lambda} = \frac{P_0}{\lambda P_0} + \frac{1}{\mu-\lambda} = \frac{1}{\lambda(1-\rho)} = \frac{1}{\lambda P_0}, \tag{3.128}$$

which agrees with formula (3.81). We obtain $E(d_x)$ by substituting $f(x)$ from (3.124) into (3.127). Thus

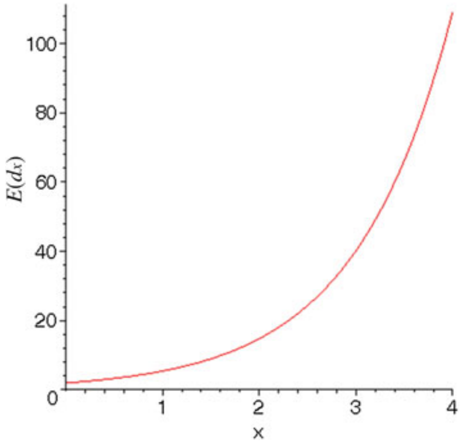


Fig. 3.9 Expected inter-downcrossing (or inter-upcrossing) time of level x, $E(d_x)$ (or $E(u_x)$) in M/M/1: $\lambda = 1.0$, $\mu = 2.0$, $\rho = 0.5$

$$E(d_x) = \frac{e^{(\mu-\lambda)x}}{\lambda(1-\rho)}, x \geq 0, \tag{3.129}$$

which increases exponentially with x (Fig. 3.9).

Inter-upcrossing Time of a Level
Denote the inter-upcrossing time of level x by u_x. Inspection of sample paths of $\{W(t)\}_{t\geq 0}$ indicates that $u_x \underset{dis}{=} d_x$ due to the memoryless property of both the inter-arrival and service times in M/M/1. Hence the plot of $E(u_x)$ versus x is identical to that of $E(d_x)$ versus x in Fig. 3.9.

3.5.8 Number of Crossings of a Level in a Busy Cycle

Denote the number of downcrossings of level $x \geq 0$ during $d_0(= \mathcal{C})$ by $\mathcal{D}_{d_0}(x)(= \mathcal{D}_{\mathcal{C}}(x))$. Since $\mathcal{D}_t(x)$ is the number of downcrossings of x during time interval $(0, t)$, from the renewal reward theorem

$$\frac{E(\mathcal{D}_{d_0}(x))}{E(d_0)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x \geq 0.$$

Hence,

$$\begin{aligned} E(\mathcal{D}_{d_0}(x)) &= \lambda P_0 e^{-(\mu-\lambda)x} \cdot E(d_0) = \lambda P_0 e^{-(\mu-\lambda)x} \cdot \tfrac{1}{\lambda P_0} \\ &= e^{-(\mu-\lambda)x}, x \geq 0. \end{aligned} \tag{3.130}$$

Since $\lambda < \mu$, $E(\mathcal{D}_{d_0}(x)) \leq 1$. From (3.130), $E(\mathcal{D}_{d_0}(x))$ decreases exponentially as x increases.

Let $\mathcal{U}_{d_0}(x)$:= number of upcrossings of level x during d_0. Since $\mathcal{D}_{d_0}(x) = \mathcal{U}_{d_0}(x)$, $x \geq 0$, formula (3.130) gives

$$E(\mathcal{D}_{d_0}(0)) = E(\mathcal{U}_{d_0}(0)) = \lim_{x \downarrow 0} e^{-(\mu-\lambda)x} = 1. \tag{3.131}$$

Equation (3.131) is intuitive, since during $\mathcal{C}$ the SP hits level 0 from above exactly once, and egresses from level 0 above (upcrosses 0) exactly once. The SP hit occurs at the end of the embedded $\mathcal{B}$. The SP egress occurs at the start of the embedded $\mathcal{B}$.

3.5.9 Downcrossings at Different Levels

From formula (3.106) in Sect. 3.4.21 for the M/G/1 queue, E(number of SP downcrossings of x) during an inter-downcrossing time d_y is given by

$$E(\mathcal{D}_{d_y}(x)) = \frac{f(x)}{f(y)}, x \geq 0, \tag{3.132}$$

which implies in M/M/1

$$E(\mathcal{D}_{d(y)}(x)) = e^{-(\mu-\lambda)(x-y)},\ x \geq 0,\ y \geq 0, \tag{3.133}$$

since $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$, $x \geq 0$. From (3.133)

$$E(\mathcal{D}_{d_y}(x)) \begin{cases} < 1 \text{ if } x > y, \\ = 1 \text{ if } x = y, \\ > 1 \text{ if } x < y. \end{cases} \tag{3.134}$$

In (3.134) $E(\mathcal{D}_{d_x}(x)) = e^{-(\mu-\lambda)(x-x)} = 1$, $x \geq 0$, in agreement with intuition, upon examining a sample path of $\{W(t)\}_{t \geq 0}$.

Proposition 3.8 For arbitrary state-space levels $x, y, y_1, y_2, \ldots, y_n$

$$E(\mathcal{D}_{d_y}(x)) = E(\mathcal{D}_{d_y}(y_1)) \cdot E(\mathcal{D}_{d_{y_1}}(y_2)) \cdots E(\mathcal{D}_{d_{y_{n-1}}}(y_n)) \cdot E(\mathcal{D}_{d_{y_n}}(x)) \tag{3.135}$$

Proof From (3.132) we obtain

$$E(\mathcal{D}_{d_y}(x)) = \frac{f(x)}{f(y)} = \frac{f(y_1)}{f(y)} \cdot \frac{f(y_2)}{f(y_1)} \cdots \frac{f(y_n)}{f(y_{n-1})} \cdot \frac{f(x)}{f(y_n)}, n = 1, 2, \ldots$$

which is equivalent to (3.135). ■

Remark 3.18 The results in (3.132) and (3.135) hold for the standard M/G/1 queue, since the proofs depend only on having a Poisson arrival process. In order to apply these formulas to a specific M/G/1 model, we must have a formula for $f(x)$. The pdf $f(x)$ is known analytically in many M/G/1 models, e.g., M/D/1, M/Erl$_{k,\mu}$/1 and variants; otherwise $f(x)$ can be approximated or estimated by numerical or simulation methods.

3.5.10 Number Served in a Busy Period

Equation (3.88) in Sect. 3.4.14 yields

$$E(N_{\mathcal{B}}) = \frac{1}{P_0} = \frac{1}{1-\rho}. \tag{3.136}$$

It follows that the number served in a k-busy period, starting with k customers in the system at time 0, is equal to k/P_0 (see Exercise 17, p. 233 in Cooper [64]).

Remark 3.19 Sect. 5.1.15 in Chap. 5 considers the number of system times above or below a state-space level x during a sojourn a_y, $y \geq 0$, and related quantities. The M/M/1 results are presented in Sect. 5.1.15 because they follow as special cases of related results for G/M/1 queues given in Sects. 5.1.13 and 5.1.14.

3.5.11 Relationship Between M/M/1 and M/M/1/1

The M/M/1/1 queue is an M/M/1 variant restricted to having *at most one* customer in the system at all $t \geq 0$. The second /1 in the notation M/M/1/1

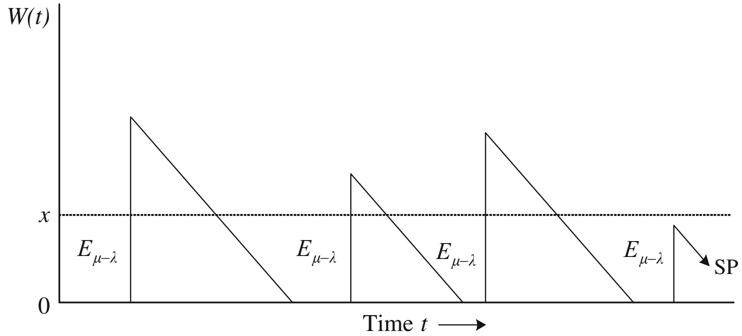


Fig. 3.10 Sample path of workload for $M_\lambda/M_{\mu-\lambda}/1/1$ queue with arrival rate λ and service rate $\mu-\lambda$. Blocked customers are cleared

refers to the *queue discipline*, namely: customers arriving when there is a customer in service, are blocked and cleared; customers arriving when the system is empty start service immediately. We compare the virtual wait process (same as workload) $\{W(t)\}_{t\geq 0}$ for M/M/1 (Fig. 3.8) and the *workload* process $\{W(t)\}_{t\geq 0}$ for M/M/1/1 (Fig. 3.10). (In M/M/1/1 all customers that get served wait 0.) The LC approach immediately connects the two models in steady-state. The cdf (3.123) and pdf (3.124) of *wait* (workload) in $M_\lambda/M_\mu/1$, are respectively *identical to* the steady-state cdf and pdf of *workload* in $M_\lambda/M_{\mu-\lambda}/1/1$ (arrival rate λ, ***service rate*** $\boldsymbol{\mu-\lambda}$).

This exact similarity of cdfs and pdfs is evident from a sample path of the workload $\{W(t)\}_{t\geq 0}$ in $M_\lambda/M_{\mu-\lambda}/1/1$ (Fig. 3.10). Fix level $x>0$. The SP downcrossing rate of x is $f(x)$, as in Theorem 1.1. The SP upcrossing rate of x is $\lambda P_0 P(S>x)=\lambda P_0 e^{-(\mu-\lambda)x}$, since *all* SP jumps start at level 0, and are distributed as $\text{Exp}_{\mu-\lambda}$, where $\mu>\lambda$. In *both* M/M/1 and M/M/1/1, $E(\mathcal{B})=\frac{1}{\mu-\lambda}$ and $P_0=1-\lambda/\mu$. In $M_\lambda/M_{\mu-\lambda}/1/1$, the busy period $\mathcal{B}$ and the blocking time are identical, and are $\underset{dis}{=}(\mu-\lambda)e^{-(\mu-\lambda)x}$, $x>0$. Also, the system times are both $\underset{dis}{=}(\mu-\lambda)e^{-(\mu-\lambda)x}$, $x>0$. Although the expected busy periods are identical, their busy-period probability distributions are quite different—evident from formulas (3.121) and 3.122 involving Bessel functions. These probability distributions depend on the (different) jump structures of the $\{W(t)\}_{t\geq 0}$s. The $M_\lambda/M_{\mu-\lambda}/1/1$ workload has the same distribution as the wait (workload) in $M_\lambda/M_{\mu}/1$, namely

$$P_0=1-\frac{\lambda}{\mu},\ f(x)=\lambda P_0 e^{-(\mu-\lambda)x}, x>0.$$

A key point is that the pdf of workload $\{P_0, f(x)\}_{x>0}$ in $\text{M}_\lambda/\text{M}_{\mu-\lambda}/1/1$ is derived *immediately by inspection*, since *all* SP jumps start at level 0.

The foregoing relationship suggests re-examining integral equation (3.109). We substitute the $\text{M}_\lambda/\text{M}_{\mu-\lambda}/1/1$ solution into the integral in (3.109), i.e., $f(y) = \lambda P_0 e^{-(\mu-\lambda)y}$, and simplify. The immediate result is the solution for the $\text{M}_\lambda/\text{M}_\mu/1$ model $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$, $x > 0$, obtained while bypassing differential equation (3.110). This solution for $\text{M}_\lambda/\text{M}_{\mu-\lambda}/1/1$ "solves" integral equation (3.109) for $\text{M}_\lambda/\text{M}_\mu/1$.

This solution procedure suggests exploring conditions that facilitate solving for the pdf of state variables "by inspection" in more general models than M/M/1. The idea is to identify a "companion" or "isomorphic" model having a much simpler sample-path jump structure.

3.6 M/G/1: Service Time Depending on Wait

Consider an M/G/1 queue with arrival rate λ and service time depending on the wait before service, $S(W_q)$. Let $P(S(W_q) \leq x | W_q = y) = B_y(x)$, $x \geq 0$, $y \geq 0$, having pdf $b_y(x) = \frac{\partial}{\partial x} B_y(x)$, $x > 0$, $y \geq 0$, wherever the derivative exists. Let W_q have steady-state cdf $F(x)$, $x \geq 0$ and pdf $\{P_0, f(x)\}_{x>0}$ (assuming $\frac{d}{dx} F(x) = f(x)$ exists). We define $f(0) \equiv f(0^+)$ for convenience (does not add probability to P_0). A sample path of $\{W(t)\}_{t\geq 0}$ resembles that for the standard M/G/1 queue, except that the SP jump size (service time) generated by each arrival depends on the SP level at the start of the jump (actual wait).

Consider a fixed state-space level $x \geq 0$ in a sample path of $\{W(t)\}_{t\geq 0}$. The downcrossing rate of x is $f(x)$, by Theorem 1.1. The total *upcrossing* rate of x is

$$\lambda P_0 \overline{B}_0(x) + \lambda \int_{y=0}^{x} \overline{B}_y(x-y) f(y) dy; \; x > 0. \tag{3.137}$$

In (3.137) the term $\lambda P_0 \overline{B}_0(x)$ is the upcrossing rate of x by SP jumps at arrival instants when the system is empty. The term $\lambda \int_{y=0}^{x} \overline{B}_y(x-y) f(y) dy$ is the upcrossing rate of x by SP jumps at arrival instants when $\{W(t)\}_{t\geq 0}$ is at state-space levels $y \in (0, x)$. Rate balance across level x yields the integral equation for $f(x)$,

$$f(x) = \lambda P_0 \overline{B}_0(x) + \lambda \int_{y=0}^{x} \overline{B}_y(x-y) f(y) dy, \; x \geq 0. \tag{3.138}$$

As in the *standard* M/G/1 queue, letting $x \downarrow 0$ gives $f(0) = \lambda P_0 \overline{B}_0(0) = \lambda P_0$.

Integrating (3.138) on both sides with respect to x over $(0, \infty)$ gives

$$1 - P_0 = \rho_0 P_0 + \int_{y=0}^{\infty} \rho_y f(y)dy,$$

$$P_0 = \frac{1 - \int_{y=0}^{\infty} \rho_y f(y)dy}{1 + \rho_0}, \tag{3.139}$$

where $\rho_y \equiv \lambda E(S_y)$, $y \geq 0$. (Eq. (3.139) is an implicit formula for P_0 since, from (3.138), $f(y)$ in the integral contains P_0. See Eq. (3.144) below for an explicit value for P_0 in the case where zero-wait customers receive special service.)

One way to deal with Eq. (3.138) is to partition the state space using $\{x_i\}_{i=0,\ldots,M+1}$, where integer $M \geq 0$, and

$$0 \equiv x_0 < x_1 < x_2 < \cdots < x_M < x_{M+1} \equiv \infty, \tag{3.140}$$

as in the paper by Posner [117]. Denote the service time of a zero-wait customer as S_0, and of a y-waiting customer, $y \in (x_{i-1}, x_i)$, as S_i. Assume the service-time is S_0 for all arrivals who wait zero, and S_i for all arrivals who wait $y \in (x_{i-1}, x_i)$. Thus the cdf of service time is

$$\begin{aligned} &B_0(x),\ x > 0 \text{ for zero-wait arrivals,} \\ &B_j(x),\ x > 0 \text{ for all arrivals who wait } y \in (x_{i-1}, x_i),\ i = 1, \ldots, M+1. \end{aligned} \tag{3.141}$$

Integral equation (3.138) can then be written

$$\begin{aligned} f(x) = \lambda P_0 \overline{B}_0(x) + \lambda \textstyle\sum_{i=1}^{j-1} \int_{y=x_{i-1}}^{x_i} \overline{B}_i(x-y) f(y)dy \\ + \lambda \textstyle\int_{y=x_{j-1}}^{x} \overline{B}_j(x-y) f(y)dy,\ x \in (x_{j-1}, x_j),\ j = 1, \ldots, M+1. \end{aligned} \tag{3.142}$$

where $\sum_{i=1}^{0} \equiv 0$. In (3.142), for any fixed $x > 0$, the right side is the upcrossing rate of level x. Thus, we have constructed integral equation (3.142) in a fast, easy, intuitive, straightforward manner using LC.

Queues with service time depending on wait appear in the literature in, e.g., ([57, 58]). The single-server model was treated in the literature using Laplace transforms in [108], and by the embedded Markov chain technique using a Lindley recursion in [117], who obtained an explicit solution for $\{P_0, f(x)\}_{x>0}$.

Remark 3.20 Deriving (3.142) using the embedded Markov chain technique is "relatively" tedious and purely algebraic (see Sect. 1.3 in Chap. 1). The single-server model was generalized to multiple servers using the embedded Markov chain technique in [48, 49] (the *original* topic and methodology of the author's Ph.D. thesis). After my discovery of LC in 1974, the model solution was completely revised using LC in the Ph.D. thesis [11], which greatly simplified the derivation of the integral equations. An analysis of an M/M/2 model with service time depending on wait is given in [53]; a revised version appears in Sect. 4.11 below.

3.6.1 M/G/1: Zero-Wait Arrivals Get Special Service

A particular case of M/G/1 with service time depending on wait, which has many useful applications, is a model where the initial customer of each busy period receives special service; we set $M = 0, x_0 = 0, x_1 = \infty$ in the state-space partition (3.140). (e.g., see, [144]; also Example 3.7 in Sect. 3.4.19; the last division of this Section; Example 3.11 in Sect. 3.8.5 below).

The integral equation (3.142) reduces to

$$f(x) = \lambda P_0 \overline{B}_0(x) + \lambda \int_{y=0}^{\infty} \overline{B}_1(x-y) f(y) dy, x > 0. \tag{3.143}$$

Integrating (3.143) with respect to x over $(0, \infty)$, using $\int_{x=0}^{\infty} f(x)dx = 1 - P_0$, gives

$$P_0 = \frac{1 - \lambda E(S_1)}{1 - \lambda E(S_1) + \lambda E(S_0)} = \frac{1 - \rho_1}{1 - \rho_1 + \rho_0}. \tag{3.144}$$

A necessary condition for stability is $\rho_1 < 1$ (guarantees $P_0 > 0$ and $\{0\}$ is a positive recurrent state). (If $\rho_1 > 1$ then $1 - \rho_1 < 0$. We would then need $1 - \rho_1 + \rho_0 < 0$ to ensure that $P_0 > 0$, causing $|1 - \rho_1 + \rho_0| < |1 - \rho_1|$. But that would imply $P_0 > 1$ in (3.144), a contradiction. If $\rho_1 = 1$, then $P_0 = 0$, which would imply the queue is unstable.)

Multiplying both sides of (3.143) by x, and integrating for $x \in (0, \infty)$ gives a Pollaczek-Khinchine (P-K)-like result for the expected wait before service

$$E(W_q) = \frac{\lambda(E(S_0^2) + E(S_1^2))}{2(1 - \lambda E(S_1))}. \tag{3.145}$$

Expected Busy Period When $M = 0$ in Partition (3.140)

In this case there are two types of arrivals. Customers that don't have to wait (wait time = 0) have service time S_0. Customers that wait a positive time have service time S_1. We determine $E(\mathcal{B})$ where $\mathcal{B}$:= busy period when initial service is $\underset{dis}{=} S_0$ and all other services are $\underset{dis}{=} S_1$.

Method 1
The busy period is

$$\mathcal{B} = S_0 + \sum_{i=1}^{A_{S_0}} \mathcal{B}_{1,i}, \tag{3.146}$$

where $A_{S_0} \underset{dis}{=}$ number of arrivals, including pseudo arrivals during S_0, the initial service time of $\mathcal{B}$ (see Sect. 3.4.12 and Fig. 3.6 therein); the sub-busy periods $\{\mathcal{B}_{1,i}\}_{j=1,2,...}$ are i.i.d. r.v.s distributed as a busy period $\mathcal{B}_1$ in a standard M_λ/G/1 queue with service time S_1 (see Fig. 3.11). The $\mathcal{B}_{1,i}$s are independent of A_{S_0}. Taking the expected value in (3.146) gives

$$\begin{aligned} E(\mathcal{B}) &= E(S_0) + \lambda E(S_0)E(\mathcal{B}_1) = E(S_0) + \lambda E(S_0)\frac{E(S_1)}{1-\lambda E(S_1)} \\ &= \frac{E(S_0)}{1-\rho_1} = \frac{E(S_0)}{P_{0,1}}, \end{aligned} \tag{3.147}$$

where $P_{0,1} = P(\text{wait} = 0)$ in the standard M/G/1 with common service time S_1.

Method 2
Applying the LC-based result for the expected busy period in M/G/1 (3.82), and using P_0 in (3.144) we get (3.147) as follows:

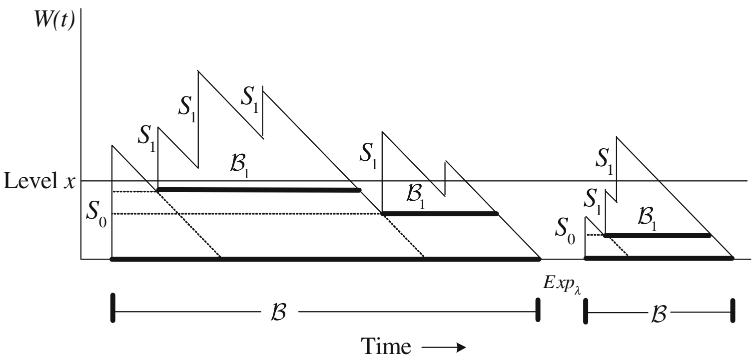


Fig. 3.11 $\mathcal{B}$'s are busy periods in M_λ/G/1 with zero-waits receiving service time $\underset{dis}{=} S_0$. $\mathcal{B}_1$'s are busy periods of M_λ/G/1 with all service times $\underset{dis}{=} S_1$, generated by pseudo arrivals during S_0

$$E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0} = \frac{1-\frac{1-\lambda E(S_1)}{1-\lambda E(S_1)+\lambda E(S_0)}}{\lambda\frac{1-\lambda E(S_1)}{1-\lambda E(S_1)+\lambda E(S_0)}} = \frac{E(S_0)}{1-\lambda E(S_1)} = \frac{E(S_0)}{1-\rho_1} = \frac{E(S_0)}{P_{0,1}}.$$

We can now derive the expression for P_0 directly using the expression for $E(\mathcal{B})$. Thus

$$P_0 = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+E(\mathcal{B})} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+\frac{E(S_0)}{1-\rho_1}} = \frac{1-\rho_1}{1-\rho_1+\rho_0}.$$

3.6.2 M/M/1: Zero-Wait Arrivals Get Special Service

We now derive the pdf $\{P_0, f(x)\}_{x>0}$ when service times are exponentially distributed with $B_0(x) = 1 - e^{-\mu_0 x}$, $B_1(x) = 1 - e^{-\mu_1 x}$. Substituting $e^{-\mu_0 x}$ for $\overline{B}_0(x)$ and $e^{-\mu_1 x}$ for $\overline{B}_1(x-y)$ in (3.143) and applying differential operator $\langle D+\mu_0\rangle \langle D+\mu_1\rangle$ (equivalent to differentiating twice with respect to x, followed by some algebra) yields a second order differential equation

$$\langle D+\mu_1-\lambda\rangle \langle D+\mu_0\rangle f(x) = 0,$$

with solution

$$f(x) = ae^{-(\mu_1-\lambda)x} + be^{-\mu_0 x}, \ x > 0, \tag{3.148}$$

provided $\mu_0 \neq \mu_1 - \lambda$ (if $\mu_0 = \mu_1 - \lambda$, $f(x)$ in the differential equation has a different solution; see, e.g., pp. 106–113 in [10]). Constants a, b are obtained from two independent initial conditions:

$$f(0) = \lambda P_0 \text{ and } f'(0) = -\mu_0\lambda P_0 + \lambda f(0),$$

giving

$$a = \frac{-\lambda^2 P_0}{\mu_1-\mu_0-\lambda},\ b = \frac{\lambda(\mu_1-\mu_0)P_0}{\mu_1-\mu_0-\lambda},\ P_0 = \frac{1-\rho_1}{1-\rho_1+\rho_2}, \tag{3.149}$$

where $\rho_i = \lambda/\mu_i, i = 1, 2$. (See Example 3.12 in Sect. 3.17.3 for an alternative solution technique to derive $f(x)$, $x > 0$.)

Expected Busy Period When Service Times Are Exponential
From Eq. (3.147),

$$E(\mathcal{B}) = \frac{\frac{1}{\mu_0}}{1-\frac{\lambda}{\mu_1}} = \frac{\mu_1}{\mu_0(\mu_1-\lambda)}.$$

Mild check on $E(\mathcal{B})$: If $\mu_0 = \mu_1 = \mu$ then $E(\mathcal{B}) = 1/(\mu - \lambda)$, as in the standard $M_\lambda/M_\mu/1$ queue.

Sojourn Above Level x When Service Times Are Exponential
Let γ_x denote the excess above x given $\{W(t)\}_{t\geq 0}$ upcrosses level x. Then $\gamma_x \underset{dis}{=} \text{Exp}_{\mu_0}$ or $\gamma_x \underset{dis}{=} \text{Exp}_{\mu_1}$, where $S_0 \underset{dis}{=} \text{Exp}_{\mu_0}$ and $S_1 \underset{dis}{=} \text{Exp}_{\mu_1}$. Then

$$E(a_x) = p_1(x)\frac{1/\mu_0}{1-\frac{\lambda}{\mu_1}} + p_2(x)\frac{1/\mu_1}{1-\frac{\lambda}{\mu_1}},$$

where $p_i(x) := P(\text{an upcrossing is due to the service time jump of a type-}i\text{ arrival})$, $i = 1, 2$, and $p_1(x) + p_2(x) = 1$. If $1/\mu_0 < 1/\mu_1$ then $E(a_x) < E(\mathcal{B})$. If $1/\mu_0 > 1/\mu_1$ then $E(a_x) > E(\mathcal{B})$. Moreover (see derivation of (3.170) in Sect. 3.8.6 below)

$$p_1(x) = \frac{P_0 e^{-\mu_0 x}}{f(x)/\lambda}, \qquad p_2(x) = \frac{\int_{y=0}^{x} B(x-y) f(y) dy}{f(x)/\lambda},$$

where $\{P_0, f(x)\}_{x>0}$ is given in (3.148) and (3.149).

3.7 Expected Sojourn Above Level *x* in M/G/1

We derive $E(a_x)$ in M/G/1 with general service time S, where $a_x :=$ *sojourn by* $\{W(t)\}_{t\geq 0}$ *above a fixed level* $x \geq 0$. The derivation utilizes a connection with M/G/1 where *zero-wait customers receive special service* (Sect. 3.6.1). Consider a sample path of $\{W(t)\}_{t\geq 0}$. A sojourn a_x is initiated by the excess of an upcrossing of x. We derive a formula for $E(a_x)$ when S is a positive continuous random variable having pdf $b(y)$, $y > 0$, cdf $B(y)$, $y > 0$, and $\overline{B}(y) = 1 - B(y)$, $y \geq 0$. Let $\{P_0, f(x)\}_{x>0}$ be the limiting mixed pdf of $\{W(t)\}_{t\geq 0}$ as $t \to \infty$. From Theorem 1.1 in Chap. 1, $\{P_0, f(x)\}_{x>0}$ is determined by the equations

$$\begin{gathered} f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y) dy, \; x > 0, \\ P_0 + \int_{x=0}^{\infty} f(x) dx = 1. \end{gathered} \tag{3.150}$$

Let $\gamma_x :=$ *excess over* x, which initiates an a_x whenever $\{W(t)\}_{t\geq 0}$ upcrosses level x. The a_xs are i.i.d. random variables since they occur within regenerative cycles delimited by successive level-x downcrossings (one a_x per cycle) (see Fig. 3.12).

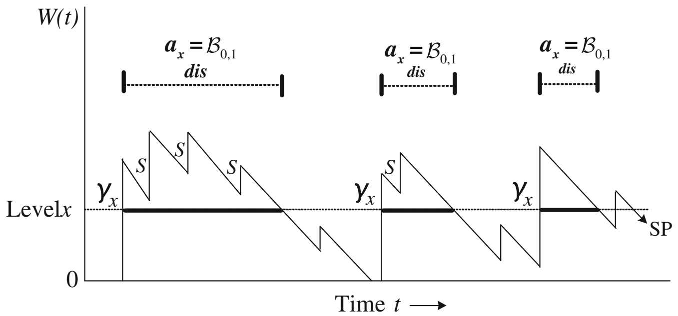


Fig. 3.12 γ_x := excess over level x. a_x := sojourn above level x. $\mathcal{B}_{0,1}$:= busy period when zero-wait service $\underset{dis}{=} \gamma_x$ and other services $\underset{dis}{=} S$

The *first* jump size of each a_x is $\underset{dis}{=} \gamma_x$. However, *during* a_x, all jump sizes are $\underset{dis}{=} S$. Thus, $a_x \underset{dis}{=}$ *busy period of an MG/1 queue where the first service is exceptional (special), denoted as* $\mathcal{B}_{0,1}$ in Fig. 3.12 (see formula (3.147) in Sect. 3.6.1).

Let $G_{\gamma_x}(z)$, $z > 0$, denote the cdf of γ_x.

Theorem 3.7 For fixed $x \geq 0$,

$$E(a_x) = \frac{E(\gamma_x)}{P_0} = \frac{\int_{z=0}^{\infty} \left(1 - G_{\gamma_x}(z)\right) dz}{P_0}$$
$$= \frac{\int_{z=0}^{\infty} \left[\lambda \int_{y=0}^{x} \overline{B}(x+z-y) dF(y)\right] dz}{f(x) P_0}. \quad (3.151)$$

Proof We employ the equation

$$\left(1 - G_{\gamma_x}(z)\right) f(x) = \lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y) f(y) dy, x \geq 0, z > 0, \quad (3.152)$$

where the LHS = $P(_{\gamma_x} > z | SP$ upcrosses level $x) \times$ (rate at which SP upcrosses level x), mindful that $f(x)$ is both the up- and downctossing rate of x (see Theorem 1.1). Thus the LHS is the upcrossing rate of level $x + z$ by the excess over x, of jumps starting below x. The RHS is a *different expression for* the upcrossing rate of level $x + z$ by jumps staring below x; all upcrosses of $x + z$ occur during the excess over x. Therefore Eq. (3.152) follows, implying

$$1 - G_{\gamma_x}(z) = \frac{\lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y) f(y) dy}{f(x)},$$

so that

$$E(\gamma_x) = \int_{z=0}^{\infty} \big(1 - G_{\gamma_x}(z)\big)\, dz = \frac{\int_{z=0}^{\infty} \Big[\lambda \int_{y=0}^{x} \overline{B}(x+z-y) dF(y)\Big] dz}{f(x)}.$$

Formula (3.151) follows since $E(a_x) = E(\mathcal{B}_{0,1}) = \frac{E(\gamma_x)}{P_0}$ by formula (3.147) in Sect. 3.6.1. ■

We can solve the equations in (3.150) for $\{P_0, f(x)\}_{x \geq 0}$ analytically, numerically or by simulation; or obtain an approximate solution. Theorem 3.7 then enables us to calculate $E(a_x)$, $x \geq 0$.

3.8 M/G/1 with Multiple Poisson Inputs

Customers arrive at a single-server system in N independent Poisson streams at rates λ_i, $i = 1, \ldots, N$, $\sum_{i=1}^{N} \lambda_i = \lambda$. Let the corresponding service times be S_i having cdf $B_i(x)$, $\overline{B}_i(x) = 1 - B_i(x)$, $x \geq 0$, and pdf $b_i(x) = \frac{d}{dx} B_i(x)$, $x > 0$, wherever the derivative exists. The service discipline is FCFS. The service time of an arbitrary arrival is S_i with probability λ_i/λ. Denote the steady-state pdf and cdf of the wait before service, W_q, by $\{P_0, f(x)\}_{x>0}$ and $F(x)$, $x \geq 0$, respectively.

Due to independent Poisson arrivals, we may view the system as an M/G/1 queue with arrival rate λ and service time

$$S = \begin{cases} S_1 \text{ with probability } \frac{\lambda_1}{\lambda}, \\ \cdots \\ S_N \text{ with probability } \frac{\lambda_N}{\lambda}. \end{cases} \tag{3.153}$$

S has a *mixture* probability distribution with mixture components S_i and mixture weights λ_i/λ (> 0) such that $\sum_{i=1}^{N} (\lambda_i/\lambda) = 1$. Hence the nth moment $S^n = S_i^n$ with probability λ_i/λ, $i = 1, \ldots, N$; $n = 1, 2, \ldots$. Thus

$$E(S) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i),\ E(S^2) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i^2). \tag{3.154}$$

Employing $\rho_i = \lambda_i E(S_i)$, $i = 1, \ldots, N$,

$$P_0 = 1 - \lambda E(S) = 1 - \sum_{i=1}^{N} \rho_i. \tag{3.155}$$

Stability
The system is stable iff every typical sample path of $\{W(t)\}_{t\geq 0}$ returns to state $\{0\}$ iff $P_0 > 0$, i.e.,

$$\sum_{i=1}^{N} \rho_i < 1. \tag{3.156}$$

3.8.1 Integral Equation for PDF of Wait

Sample paths of $\{W(t)\}_{t\geq 0}$ resemble those of the standard M/G/1 queue, except that the jump size due to an arrival is $\underset{dis}{=} S_i$ with probability λ_i/λ, having cdf $B_i(\cdot)$, $i = 1, \ldots, N$. Thus jumps $\underset{dis}{=} S_i$ occur at Poisson rate λ_i. By Theorem 1.1, for a fixed state-space level $x > 0$, the SP downcrossing rate is $f(x)$. The SP upcrossing rate for type i arrivals is

$$\lambda_i P_0 \overline{B}_i(x) + \lambda_i \int_{y=0}^{x} \overline{B}_i(x-y) f(y) dy, i = 1, \ldots, N.$$

Balancing the *total* SP down- and upcrossing rates of level x for all customer types, yields the integral equation for $f(x)$,

$$f(x) = \sum_{i=1}^{N} \lambda_i \left[P_0 \overline{B}_i(x) + \int_{y=0}^{x} \overline{B}_i(x-y) f(y) \right] dy,$$

or

$$f(x) = \lambda P_0 \left(\sum_{i=1}^{N} \frac{\lambda_i}{\lambda} \overline{B}_i(x) \right) + \lambda \int_{y=0}^{x} \left(\sum_{i=1}^{N} \frac{\lambda_i}{\lambda} \overline{B}_i(x-y) \right) f(y) dy. \tag{3.157}$$

Integral equation (3.157) is in the form of the analogous integral equation (3.34) for the pdf of wait in a *standard* M/G/1 queue with $\lambda = \sum_{i=1}^{N} \lambda_i$, and $\overline{B}(x) = \sum_{i=1}^{N} (\lambda_i/\lambda) \overline{B}_i(x)$, $\dot{x} > 0$.

3.8.2 Expected Wait Before Service

Since $E(S^2) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i^2)$, the Pollaczek-Khinchine (P-K) formula (3.63) gives the expected wait before service as

$$E(W_q) = \frac{\lambda E(S^2)}{2(1 - \lambda E(S))} = \frac{\lambda \sum_{i=1}^{N} (\lambda_i/\lambda)\, E(S_i^2)}{2(1 - \sum_{i=1}^{N} \rho_i)} = \frac{\sum_{i=1}^{N} \lambda_i E(S_i^2)}{2P_0}. \tag{3.158}$$

Alternatively, we can obtain $E(W_q)$ in (3.158) directly from (3.157) upon multiplying both sides by x, then integrating both sides with respect to $x \in (0, \infty)$, changing the order of integration in the double integral, and doing some algebra.

3.8.3 Expected Number in Queue

Let L_q = expected number of units in the queue before service in the steady state. Then by $\boldsymbol{L} = \boldsymbol{\lambda W}$ (Little [110]) and (3.158)

$$L_q = \lambda E(W_q) = \frac{\lambda \sum_{i=1}^{N} \lambda_i E(S_i^2)}{2P_0}. \tag{3.159}$$

Denote the steady-state expected number of type i units in the queue by $L_{q,i}$. Let the wait of an arbitrary *type i customer* be $W_{q,i}$, and the wait of an *arbitrary customer* be W_q. Then $W_{q,i} \underset{dis}{=} W_q$, because the waiting time of any arrival depends only on the current workload at the arrival instant. Thus $E(W_{q,i}) = E(W_q)$, $i = 1, \ldots, N$, and by $\boldsymbol{L} = \boldsymbol{\lambda W}$,

$$L_{q,i} = \lambda_i E(W_{q,i}) = \lambda_i E(W_q) = \frac{\lambda_i \sum_{i=1}^{N} \lambda_i E(S_i^2)}{2P_0}, i = 1, \ldots, N. \tag{3.160}$$

3.8.4 Expected Busy Period

Applying (3.82) and (3.155), the expected busy period is given by

$$E(\mathcal{B}) = \frac{1 - P_0}{f(0)} = \frac{1 - P_0}{\lambda P_0} = \frac{\sum_{i=1}^{N} \rho_i}{\lambda \left(1 - \sum_{i=1}^{N} \rho_i\right)}. \tag{3.161}$$

As a mild check on formula (3.161), let $S = S_i$ with probability $1/N$, $i = 1, \ldots, N$. Then $\lambda_i/\lambda \equiv 1/N$ so that $\sum_{i=1}^{N} (\lambda_i/\lambda) = \sum_{i=1}^{N} (1/N) = 1$, $\rho_i \equiv \lambda_i E(S_i) = (\lambda/N)\, E(S_i)$ and $\sum_{i=1}^{N} \rho_i = (\lambda/N) \sum_{i=1}^{N} E(S_i)$. The multiple Poisson input model reduces to a standard M/G/1 queue with arrival rate λ and $E(S) = \frac{1}{N} \sum_{i=1}^{N} E(S_i)$. From (3.161)

$$E(\mathcal{B}) = \frac{\frac{\lambda}{N} \sum_{i=1}^{N} E(S_i)}{\lambda \left(1 - \sum_{i=1}^{N} \rho_i\right)} = \frac{\frac{1}{N} \sum_{i=1}^{N} E(S_i)}{1 - \sum_{i=1}^{N} \rho_i} = \frac{E(S)}{P_0},$$

which is the formula for $E(\mathcal{B})$ for the standard M/G/1 queue.

3.8.5 M/M/1 with Multiple Poisson Inputs

To outline a solution technique for integral equation (3.157), we assume the service times are $\underset{dis}{=} \text{Exp}_{\mu_i}$ with $\overline{B}_i(x) = e^{-\mu_i x}$, $i = 1, 2, \ldots, N$. Then (3.157) becomes

$$f(x) = \sum_{i=1}^{N} \lambda_i \left[P_0 e^{-\mu_i x} + \int_{y=0}^{x} e^{-\mu_i (x-y)} f(y) dy \right], \; x > 0. \qquad (3.162)$$

We can apply the differential operator $\langle D + \mu_1 \rangle \ldots \langle D + \mu_N \rangle$ to Eq. (3.162), to derive and solve analytically an Nth order differential equation with constant coefficients for $f(x)$, using initial conditions to obtain the constants of integration.

The differential operator $\langle D + \mu_i \rangle$ is commutative with respect to exponential functions of the form $e^{\alpha x + \beta}$, where α and β are constants, i.e., for any permutation $(i_1, i_2, \ldots, i_N)$ of the numbers $(1, 2, \ldots, N)$

$$\begin{aligned} \langle (D + \mu_1) \cdots (D + \mu_N) \rangle e^{\alpha\alpha x + \beta} &= \langle D + \mu_1 \rangle \cdots \langle D + \mu_N \rangle \, e^{\alpha x + \beta} \\ &= \left\langle D + \mu_{i_1} \right\rangle \cdots \left\langle D + \mu_{i_N} \right\rangle e^{\alpha x + \beta} \\ &= \langle (D + \mu_{i_1}) \cdots (D + \mu_{i_N}) \rangle e^{\alpha x + \beta}. \end{aligned}$$

The commutativity property simplifies the transformation of an integral equation into a differential equation, when the kernel of any integral in the equation has an exponential form like $e^{-\mu_i (x-y)}$ in (3.162).

Expected Wait and Expected Number in Queue

If $S_i \underset{dis}{=} \text{Exp}_{\mu_i}$ then $E(S_i^2) = 2/\mu_i^2$, which substituted into (3.158) and (3.159) respectively, yield

$$E(W_q) = \frac{\sum_{i=1}^{N} \frac{\lambda_i}{\mu_i^2}}{1 - \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i}}, \tag{3.163}$$

$$L_q = \frac{\lambda \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i^2}}{1 - \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i}}. \tag{3.164}$$

Similarly, substituting into (3.160), gives

$$L_{q,i} = \frac{\lambda_i \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i^2}}{1 - \sum_{i=1}^{N} \frac{\lambda i}{\mu_i}}, \, i = 1, \ldots, N. \tag{3.165}$$

Two Customer Types

To illustrate the solution, we consider two distinct customer types, and derive $\{P_0, f(x)\}$. Set $N = 2$ in (3.162). Applying differential operator $\langle D + \mu_1 \rangle \langle D + \mu_2 \rangle$ to both sides, gives a second order differential equation

$$\left\langle D^2 + (\mu_1 + \mu_2 - \lambda)D + (\mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1) \right\rangle f(x) = 0$$

having solution

$$f(x) = ae^{R_1 x} + be^{R_2 x}, \tag{3.166}$$

where R_i, $i = 1, 2$ are the roots for z in the characteristic equation

$$z^2 + (\mu_1 + \mu_2 - \lambda)z + \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1 = 0.$$

Both roots are negative since the product $R_1 R_2 = \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1 > 0$ (equivalent to $1 - \rho_1 - \rho_2 > 0$, the stability condition), and $R_1 + R_2 = -(\mu_1 + \mu_2 - \lambda) < 0$. Constants a and b are determined by applying two independent initial conditions for $f(0) = a + b$ and $f'(0) = R_1 a + R_2 b$, obtained from (3.166) and also from (3.162), resulting in two equations for a, b:

$$\begin{aligned} &a + b = \lambda P_0, \\ &R_1 a + R_2 b = -(\mu_1\lambda_1 + \mu_2\lambda_2)P_0 + \lambda f(0) = -(\mu_1\lambda_1 + \mu_2\lambda_2 - \lambda^2)P_0. \end{aligned} \tag{3.167}$$

Thus $f(x)$ is given by (3.166) where $[a, b]$ is the solution of the two equations in (3.167):

$$\begin{aligned} a &= -P_0(R_2\lambda - \lambda^2 + \lambda_1\mu_1 + \lambda_2\mu_2)/(R_1 - R_2), \\ b &= P_0(R_1\lambda - \lambda^2 + \lambda_1\mu_1 + \lambda_2\mu_2)/(R1 - R2), \end{aligned} \tag{3.168}$$

and

$$\left.\begin{aligned} &P_0 = 1 - \tfrac{\lambda_1}{\mu_1} - \tfrac{\lambda_2}{\mu_2}, \\ &R_1 = \tfrac{-\boldsymbol{B}}{2} + \tfrac{\sqrt{\boldsymbol{B}^2 - 4\boldsymbol{AC}}}{2},\ R_2 = \tfrac{-\boldsymbol{B}}{2} - \tfrac{\sqrt{\boldsymbol{B}^2 - 4\boldsymbol{AC}}}{2}, \\ &\boldsymbol{A} = 1,\ \boldsymbol{B} = \mu_1 + \mu_2 - \lambda,\ \boldsymbol{C} = \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1. \end{aligned}\right\} \tag{3.169}$$

Example 3.10 Consider a numerical example with $N = 2$, $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\mu_1 = 3$, $\mu_2 = 2$. Then $P_0 = 0.4167$, $R_1 = -1.0$, $R_2 = -2.5$, $a = 0.555555$, $b = 0.069444$, and

$$f(x) = 0.555555\, e^{-1.0x} + 0.069444\, e^{-2.5x}, x > 0.$$

A computational check shows that $F(\infty) = 1$, and $f(0) = \lambda P_0$, i.e.,

$$\begin{aligned} F(\infty) &= P_0 + \int_{x=0}^{\infty} f(x)dx \\ &= 0.4167 + \int_{x=0}^{\infty} (0.555555\, e^{-1.0x} + 0.069444\, e^{-2.5x})dx = 1, \\ f(0) &= a + b = \lambda P_0 = 0.625. \end{aligned}$$

3.8.6 Expected Sojourn Above Level x − E(a_x)

Let $p_i(x) := P$ (arrival type $i | \{W(t)\}_{t\geq 0}$ upcrosses x), $i = 1, \ldots N$. Using Bayes' formula, and Eq. (3.162),

$$\begin{aligned} p_i(x) &= \frac{P(\{W(t)\}_{t\geq 0} \text{ upcrosses } x | \text{ arrival type } i) \cdot P(\text{arrival type } i)}{P(\{W(t)\}_{t\geq 0} \text{ upcrosses } x)} \\ &= \frac{\left(P_0\overline{B}_i(x) + \int_{y=0}^{x} \overline{B}_i(x-y)f(y)dy\right) \cdot (\lambda_i/\lambda)}{\sum_{i=1}^{N}\left[P_0\,(\lambda_i/\lambda)\,\overline{B}_i(x) + \int_{y=0}^{x} (\lambda_i/\lambda)\,\overline{B}_i(x-y)f(y)dy\right]} \\ &= \frac{P_0\,(\lambda_i/\lambda)\,\overline{B}_i(x) + \int_{y=0}^{x} (\lambda_i/\lambda)\,\overline{B}_i(x-y)f(y)dy}{f(x)/\lambda}. \end{aligned} \tag{3.170}$$

Let $\gamma_{S_i}(x)$:= excess above level x due to a type-i upcrossing of x. Sojourn $a_x \underset{dis}{=} \mathcal{B}_{(i)}$ with probability $p_i(x)$, i.e., a_x has a mixture distribution with components $\mathcal{B}_{(i)}$ and mixture probabilities $p_i(x)$. which is a busy period where zero-waits get service $S_0 \underset{dis}{=} \gamma_{S_i}(x)$, and positive-waits get service time S_1 (see formula (3.153)). Applying (3.147) in Sect. 3.6.1 and formula (3.161) gives

$$E(a_x) = \frac{E(\gamma_S(x))}{1 - \sum_{i=1}^{N} \rho_i}. \tag{3.171}$$

$E(a_x)$ in M/M/1 with Two Types of Poisson Inputs

Consider two types of input with $S_i \underset{dis}{=} \text{Exp}_{\mu_i}$, $i = 1, 2$. By the memoryless property $\gamma_{S_i}(x) \underset{dis}{=} \text{Exp}_{\mu_i}$, $i = 1, 2$, for all $x \geq 0$. If the upcrossing of x that initiates a_x is due to a type-i arrival then $a_x \underset{dis}{=} \mathcal{B}_i$, an M/G/1 busy period with *initial* service $S_0 = \text{Exp}_{\mu_i}$, and other service times a mixture of Exp_{μ_1} and Exp_{μ_2}. Thus

$$E(a_x|\text{excess} = \gamma_{S_i}(x)) = \frac{1}{\mu_i \left(1 - \sum_{i=1}^{2} \rho_i\right)}, i = 1, 2,$$

and

$$E(a_x) = \sum_{i=1}^{2} p_i(x) \frac{1}{\mu_i \left(1 - \sum_{i=1}^{2} \rho_i\right)}, \tag{3.172}$$

From (3.170) the probability of a type-i upcrossing of level x is

$$p_i(x) = \frac{P_0 (\lambda_i/\lambda) e^{-\mu_i x} + \int_{y=0}^{x} (\lambda_i/\lambda) e^{-\mu_i (x-y)} f(y) dy}{f(x)/\lambda}, i = 1, 2.$$

Example 3.11 Using the input values and $f(x)$, $x > 0$, in Example 3.10, Sect. 3.8.5, we compute

$$p_1(x) = \frac{0.27777\, e^{-3.x}(0.5\, e^{0.5x} + e^{2.x})}{0.55555\, e^{-x} + 0.06944\, e^{-2.5x}}, \qquad p_2(x) = 1 - p_1(x), x \geq 0.$$

Applying formula (3.172) and using $p_2(x) = 1 - p_1(x)$ gives

$$E(a_x) = p_1(x) \frac{1}{1.25} + p_2(x) \frac{1}{0.83333}.$$

Fig. 3.13 $E(a_x)$ versus x in Example 3.11. $E(a_0) = E(\mathcal{B})$

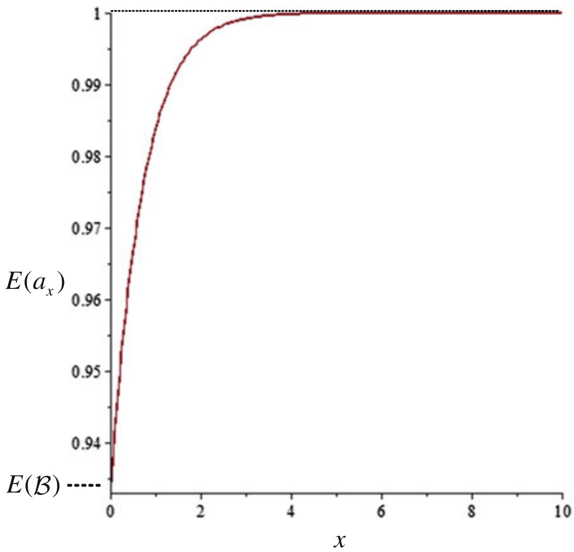
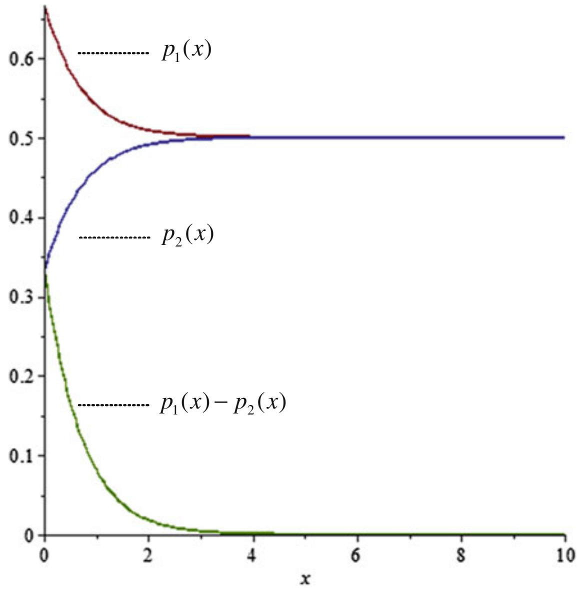


Fig. 3.14 Changes in $p_i(x)$, $i = 1, 2$ and $p_1(x) - p_2(x)$, $x > 0$, in Example 3.11



From (3.161) the expected value of the M/G/1 busy period with multiple Poisson inputs is $E(\mathcal{B}) = 0.9333$. We see from a plot that $E(a_x) > E(\mathcal{B}), x > 0, E(a_0) = E(\mathcal{B})$ and $\lim_{x\to\infty} E(a_x) = 1$ (Fig. 3.13). Moreover $\lim_{x\to\infty} p_i(x) = 0.5, i = 1, 2$ (Fig. 3.14). These observations are readily verified analytically. The growth of $E(a_x)$ as x increases is due to the fact that $E(\gamma_x) = 1/\mu_i$ independent of x, implying that the evolution of $E(a_x)$ on $(0, \infty)$ is determined by the $p_i(x)$s which do change as x increases. (In examples where $E(\gamma_x)$ depends on x the properties of $E(a_x)$ would be different.)

3.9 M/G/1: Wait-Number Dependent Service

Arrivals occur at Poisson rate λ. The queue discipline is FCFS. The service time is denoted as $S(N_q)$ where N_q := number of customers left waiting in the queue *just after* a start of service. Thus $N_q \in \{0, 1, \ldots\}$. For exposition, we assume there are two different types of service. Let

$$S(N_q) = \begin{cases} S_0, & \text{if } N_q = 0, \\ S, & \text{if } N_q = 1, 2, \ldots. \end{cases}$$

Let $P(S_0 \le x) = B_0(x)$, $\overline{B}_0(x) = 1 - B_0(x)$, $P(S \le x) = B(x)$, $\overline{B}(x) = 1 - B(x)$. Denote the steady-state wait before service as W_q having cdf $P(W_q \le x) = F(x)$ and mixed pdf $\{P_0, f(x)\}_{x>0}$ wherever $\frac{d}{dx}F(x)$ $(=f(x))$ exists.

We represent this M/G/1 queue by M/G(N_q)/1. The analysis utilizes the construction of a sample path of the ***virtual wait*** $\{W(t)\}_{t\ge 0}$ by applying the definition of virtual wait *literally*. The virtual wait $W(t)$ at instant t, is defined as the time that a potential (would-be) arrival at t must wait before starting service. The virtual wait is a continuous-state continuous-time process. Its value at any instant t is conditional on an arrival occurring at instant t.

Remark 3.21 In order to validate the LC method immediately after its discovery in 1974, the author applied LC to derive $\{P_0, f(x)\}_{x>0}$ in M/G(N_q)/1 (and in several other queueing models in the literature; and in multiple-server state-dependent queues in his original Ph.D. thesis topic, where solutions had been derived using Lindley recursions and embedded Markov chains [48, 49]). The author included an LC analysis of M/G(N_q)/1 in his Ph.D.

thesis (pp. 206–213 in [11]). The results agreed with the classically-based analysis of M/G(N_q)/1 in C.M. Harris's 1966 Ph.D. thesis [85] and in C.M. Harris's 1967 journal article [86]. (See also Example 5 in [51].)

3.9.1 Sample Path of $\{W(t)\}_{t\geq 0}$

Consider Fig. 3.15. The first customer C_1 arrives, initiates a busy period and receives a service time S_0, since zero customers are left behind in the queue when C_1 starts service. Customer C_2 arrives at t^- during C_1's service time and is allotted a "virtual" service time S, although C_2's actual service time is not known until later at C_2's start-of-service instant. The reason is that the virtual wait may be considered to be the answer to the following question asked a non-countably infinite number of times, i.e., at every instant $t \geq 0$: "***How long would a new arrival at instant*** t ***have to wait before its start-of-service instant?***" The answer to this question forces us to allot service time S to C_2 at its arrival instant. That is, a would-be new arrival *immediately after*

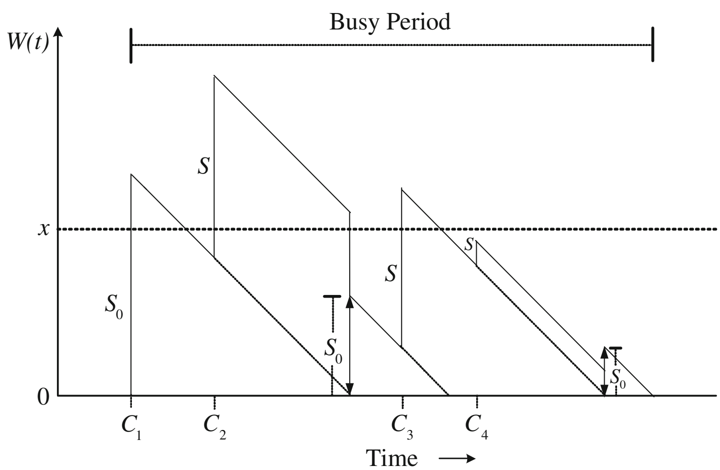


Fig. 3.15 Sample path of $\{W(t)\}_{t\geq 0}$ in M/G(N_q)/1 during a busy period. Shows jumps of size S_0 from level 0 and size S from positive levels. Illustrates a double jumps in the virtual wait $\{W(t)\}_{t\geq 0}$

C_2's arrival, would force C_2 to start service with at least one customer left waiting in the queue. In other words, if C_2 arrives at t^-, the virtual wait at t is the time that a would-be new arrival would have to wait before service.

Suppose, as depicted in Fig. 3.15, *zero* customers arrive during C_2's wait. Then at C_2's start-of-service instant, C_2 must receive an actual service time S_0. This cancels S assigned at C_2's arrival epoch, and substitutes an actual service time S_0. The SP jumps both downward to level 0, and upwards by an amount S_0, at the start-of-service instant of C_2. All SP upward jumps from level 0 are $\underset{dis}{=} S_0$, and all SP upward jumps from positive levels are $\underset{dis}{=} S$.

At instants like the start-of-service instant of C_2 the SP makes a *double* jump (for other examples of double jumps see Examples 2.2 and 2.3 in Sect. 2.3, and Figs. 2.3, 2.5 and 2.6 in Chap. 2).

Next we discuss and derive the steady-state distribution of the **virtual wait** (in contrast to workload).

3.9.2 Integral Equation for PDF of Virtual Wait

Consider a sample path of $\{W(t)\}_{t\geq 0}$ and fix level $x > 0$ in the state space (Fig. 3.15). The SP downcrossing rate of x has *two* components:

1. $f(x)$ by Theorem 1.1,
2. $\lambda\overline{B}(x)\tilde{f}(\lambda)$ due to SP *downward* jumps similar to that at the start-of-service instant of C_2. Here $\tilde{f}(s) := \int_{y=0}^{\infty} e^{-sy} f(y)dy, s > 0$, is the Laplace transform of $f(x)$, and $\tilde{f}(\lambda) = \tilde{f}(s)|_{s=\lambda}$. ($\tilde{f}(s)$ is also denoted by $\mathcal{L}_f(s)$ or other symbols; Sect. 3.4.4 briefly discusses the Laplace transform.)

In component 2, S must be greater than x in order for a downcrossing of x to occur at instants such as the *start of service* of C_2 in Fig. 3.15. The rate of such downcrossings is

$$\begin{aligned}
&\lambda P(S > x, \text{ and zero customers arrive in a waiting time}) \\
&\quad = \lambda P(S > x) P(\text{zero customers arrive in a waiting time}) \\
&\quad = \lambda P(S > x) \int_{y=0}^{\infty} e^{-\lambda y} f(y)dy = \lambda\overline{B}(x)\tilde{f}(\lambda),
\end{aligned}$$

by independence of S and the arrival stream. The total downcrossing rate of x is

$$f(x) + \lambda\overline{B}(x)\tilde{f}(\lambda), x > 0. \tag{3.173}$$

The SP upcrossing rate of x has *three* components:

1. $\lambda\overline{B}_0(x)P_0$, due to arrivals when the system is empty,
2. $\lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy$, due to arrivals when the virtual wait is $y\in(0,x)$,
3. $\lambda\overline{B}_0(x)\tilde{f}(\lambda)$, due to arrivals that must wait a positive time and have zero customers arrive behind them during their wait in queue. The total upcrossing rate is

$$\lambda\overline{B}_0(x)P_0+\lambda\int_{y=0}^{x}\overline{B}(x-y)f(y)dy+\lambda\overline{B}_0(x)\tilde{f}(\lambda). \tag{3.174}$$

Rate balance across level x equates (3.173) and (3.174), leading to the integral equation for $f(\cdot)$,

$$\begin{aligned} f(x)= \ &\lambda\overline{B}_0(x)P_0+\lambda\textstyle\int_{y=0}^{x}\overline{B}(x-y)f(y)dy \\ &+\lambda\left(\overline{B}_0(x)-\overline{B}(x)\right)\cdot\tilde{f}(\lambda),\ \ x>0. \end{aligned} \tag{3.175}$$

3.9.3 Exponential Service

Assume $\overline{B}_0(x)=e^{-\mu_0 x}$, $\overline{B}(x)=e^{-\mu x}$, $x>0$, and let $\rho_0=\frac{\lambda}{\mu_0}$, $\rho=\frac{\lambda}{\mu}$. Then (3.175) reduces to

$$\begin{aligned} f(x)= \ &\lambda e^{-\mu_0 x}P_0+\lambda\textstyle\int_{y=0}^{x}e^{-\mu(x-y)}f(y)dy \\ &+\lambda\left(e^{-\mu_0 x}-e^{-\mu x}\right)\cdot\tilde{f}(\lambda),\ x>0. \end{aligned} \tag{3.176}$$

Applying differential operator $\langle D+\mu_0\rangle\langle D+\mu\rangle$ to both sides of (3.176) yields the second order differential equation with constant coefficients

$$\langle D^2+(\mu_0+\mu-\lambda)D+\mu_0(\mu-\lambda)\rangle f(x)=0, \tag{3.177}$$

with general solution

$$f(x)=ae^{-(\mu-\lambda)x}+be^{-\mu_0 x},\ x>0, \tag{3.178}$$

assuming $\mu_0\neq\mu-\lambda\neq 0$. From the first term of formula (3.178), a necessary condition for stability is $\lambda<\mu$, since necessarily $f(\infty)=0$.

Using the initial condition $f(0) = \lambda P_0$, substituting $f(y)$ from (3.178) into (3.176), and equating coefficients of common exponents, we obtain the parameters in (3.178) as

$$P_0 = \frac{1-\rho}{1-\rho+\rho_0+\rho_0^2-\rho_0\rho}, \tag{3.179}$$

and

$$a = \frac{-\lambda\rho_0^2 P_0}{\rho_0-\rho-\rho_0\rho}, \; b = \frac{\lambda(1+\rho_0)(\rho_0-\rho)P_0}{\rho_0-\rho-\rho_0\rho}. \tag{3.180}$$

Expected Busy Period

The rate at which the SP makes left-continuous hits of level 0 from above is $f(0) = \lambda P_0$ (Fig. 3.15). Hence the expected busy period is, from (3.82),

$$E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0} = \frac{\rho_0+\rho_0^2-\rho_0\rho}{\lambda(1-\rho)}. \tag{3.181}$$

As a mild check on $E(\mathcal{B})$, set $\rho_0 = \rho = \frac{\lambda}{\mu}$. Then the model reduces to a standard M/M/1 queue. Formula (3.181) reduces to $E(\mathcal{B}) = \frac{1}{\mu-\lambda}$, corresponding to $E(\mathcal{B})$ for the standard M/M/1 queue.

Distribution of Number in System

Applying formula (3.76) and using (3.178) and (3.180) we obtain the steady-state probability of n customers left in the system at *departure* instants,

$$\begin{aligned} d_n &= \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} f(x)dx \\ &= \frac{\rho_0 \cdot \left(\rho_0^{n-1} - \rho\rho_0^{n-2} - \rho^n(1+\rho_0)^{n-1}\right)}{(\rho_0-\rho-\rho_0)(1+\rho_0)^{n-1}} P_0, \; n = 1, 2, \ldots, \end{aligned} \tag{3.182}$$

where P_0 ($=d_0$) is given in (3.179). The values in (3.182) agree with the values of d_n obtained in the earlier works [85, 86].

3.9.4 Workload

In the standard M/G/1, $\{W(t)\}_{t\geq 0}$ is the same as the workload at instant t. In M/G(N_q)/1, the workload is not known at the instant just after an arrival, because the added service time is either S_0 or S depending on *future* arrivals during its wait before service. We can determine the probabilities of these two service times, which allows us to proceed with the analysis.

Consider the **workload process** which we designate $\{W_{wk}(t)\}_{t\geq 0}$. Then $W_{wk}(t)$:= amount of remaining work in the system at time t. Denote the steady-state pdf of $\{W_{wk}(t)\}_{t\to\infty}$ by $\left\{P_{0,wk}, g(x)\right\}_{x>0}$.

In order to construct a sample path, we ask the question immediately after an arrival when the actual workload is y: "**What is the workload just after the arrival**?". The answer logically causes the SP to make a jump of size S with probability $(1-e^{-\lambda y})$ (P(at least 1 arrival in time y)), or size S_0 with probability $e^{-\lambda y}$ (P(no arrivals in time y)). This leads to the upcrossing rate of level x as the right side of (3.183) below. The downcrossing rate of x is $g(x)$. Rate balance across level x gives

$$g(x) = \lambda \overline{B}_0(x)P_{0,wk} + \lambda \int_{y=0}^{x} \overline{B}(x-y)(1-e^{-\lambda y})g(y)dy$$
$$+\lambda \int_{y=0}^{x} \overline{B}_0(x-y)e^{-\lambda y}g(y)dy. \tag{3.183}$$

If service time $S_0 \underset{dis}{=} \text{Exp}_{\mu_0}$, $S \underset{dis}{=} \text{Exp}_{\mu}$ then $\overline{B}_0(z) = e^{-\mu_0 z}$, $\overline{B}(z) = e^{-\mu z}$, $z > 0$, in (3.183). Applying $\langle D+\mu\rangle \langle D+\mu_0\rangle$ to the resulting integral equation yields a second order differential equation with a variable coefficient for $g(x)$

$$\left\langle D^2 + (\mu_0+\mu-\lambda)D + \mu_0(\mu-\lambda) - (\mu-\mu_0)\lambda e^{-\lambda x}\right\rangle g(x) = 0.$$

The solution is given by

$$g(x) = e^{\frac{1}{2}(-\mu-\mu_0+\lambda)x}\left(a \text{ BesselJ}\left(-\frac{|-\lambda-\mu+\mu_0|}{\lambda}, \frac{2\sqrt{\mu_0-\mu}\, e^{-\frac{1}{2}\lambda x}}{\sqrt{\lambda}}\right)\right.$$
$$\left. + b \text{ BesselY}\left(-\frac{|-\lambda-\mu+\mu_0|}{\lambda}, \frac{2\sqrt{\mu_0-\mu}\, e^{-\frac{1}{2}\lambda x}}{\sqrt{\lambda}}\right)\right), \tag{3.184}$$

where a, b are constants to be determined using the initial conditions $g(0) = \lambda P_{0,wk}$, $g'(0) = -(\mu_0-\lambda)\lambda P_{0,wk}$; and BesselJ := first kind ($\nu = 1$), BesselY := second kind ($\nu = 2$), which satisfy Bessel's equation

$$xy'' + xy' + (-\nu^2 + x^2)y = 0, \nu = 1, 2,$$

(see Bessel functions in Maple 17 software). We solve for $P_{0,wk}$ using the normalizing condition $P_{0,wk} + \int_{x=0}^{\infty} g(x)dx = 1$. Due to the Bessel functions in (3.184), it is difficult to get analytic solutions for a, b and $P_{0,wk}$. However, one can obtain numerical solutions when the input parameters λ, μ, and μ_0 have numerical values.

3.10 M/D/1 Queue

The M/D/1 queue is a classical model in queueing theory, first analyzed by A.K. Erlang in 1909 [72].

Here we use LC to derive the cdf of the wait before service, $F(x), x \geq 0$, the mixed pdf $\{P_0, f(x)\}_{x>0}$, where $f(x) = dF(x)/dx, x > 0$, wherever the derivative exists. We also obtain the probability distribution of the number of customers in the system P_n, $n = 0, 1, 2, \ldots$, and related quantities.

The arrival stream is Poisson at rate λ. The service time for each customer is deterministic $S = D > 0$. The traffic intensity is $\rho = \lambda E(S) = \lambda D < 1$, implying stability. Consider the virtual wait $\{W(t)\}_{t\geq 0}$, (Fig. 3.16) and the actual waiting times $\{W_n\}_{n=1,2,\ldots}$. Denote $P(W_n \leq x)$ by $H_n(x), x \geq 0$ and $\lim_{n\to\infty} H_n(x) = H(x), x \geq 0$. Due to Poisson arrivals (e.g., [140])

$$F(x) \equiv \lim_{t\to\infty} P(W(t) \leq x) = \lim_{n\to\infty} P(W_n \leq x) = H(x), x \geq 0.$$

The $\{W(t)\}_{t\geq 0} \leftrightarrow \{W_n\}_{n=0,1,\ldots}$ connection ensures that a study of the virtual wait yields considerable information about both processes.

We define $f(x), x > 0$, to be right continuous, and for notational convenience $f(0) = f(0^+)$ which adds zero probability to $F(0)$. The probability

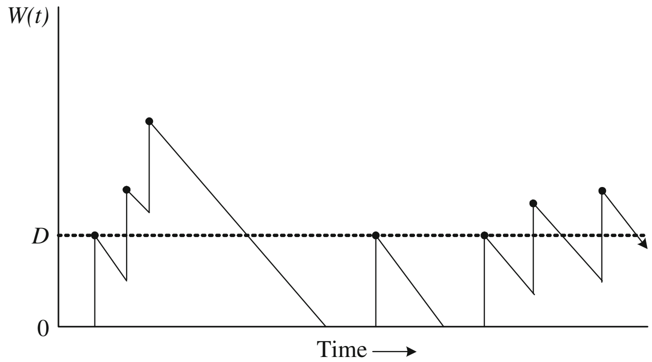


Fig. 3.16 Sample path of $\{W(t)\}_{t\geq 0}$ in M/D/1 queue. Black circles at peaks indicate right continuity

of a zero wait is $P_0 = F(0) = 1 - \rho = 1 - \lambda D$. The mixed pdf $\{P_0, f(x)\}_{x>0}$ is related to $F(x)$ by

$$F(x) = P_0 + \int_{y=0}^{x} f(y)dy, x \geq 0, \qquad F(\infty) = P_0 + \int_{y=0}^{\infty} f(x)dx = 1. \tag{3.185}$$

3.10.1 Properties of PDF and CDF of Wait

We use LC to derive three properties of $\{P_0, f(x)\}_{x>0}$ and a property of $F(x), x \geq 0$.

Proposition 3.9 For the M/D/1 queue: (1) $\{P_0, f(x)\}_{x>0}$ has exactly one atom, which is at $x = 0$; (2) $f(x)$ has a downward jump discontinuity of size λP_0 at $x = D$; (3) $f(x)$ is continuous for all $x > 0, x \neq D$.

Proof Consider sample paths of $\{W(t)\}_{t\geq 0}$ in Figs. 3.16 and 3.17.

1. State $\{0\}$ is an atom since a sample path spends a positive proportion of time in $\{0\}$ $(a.s.)$, namely $P_0 = 1 - \lambda D > 0$ (from (3.62) in Sect. 3.4). The state space $\boldsymbol{S} = [0, \infty)$ has no other atoms, since the proportion of time the SP spends in each state $x > 0$, is 0.
2. Consider state-space levels D and $D - \varepsilon$, $0 < \varepsilon < D$. Fix time $t > 0$. $\mathcal{T}_t^b(D)$ is the number of tangents of level D from below during $(0, t)$

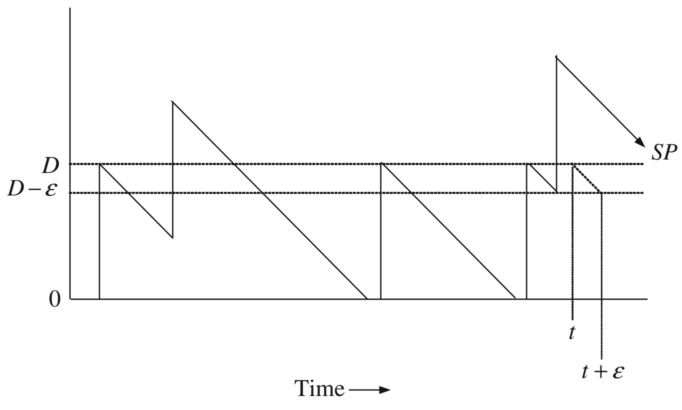


Fig. 3.17 Sample path of $\{W(t)\}_{t\geq 0}$ in M/D/1 showing levels D and $D - \varepsilon$ and instants $t, t + \varepsilon$. See Proposition 3.7, Proof, Part (2)

(see Fig. 2.13 (row 2, column 2) in Sect. 2.5; and Examples 2.4 and 2.5 in Sect. 2.4.10). We have

$$\mathcal{D}_{t+\varepsilon}(D-\varepsilon) = \sum_{j=1}^{\mathcal{D}_t(D)+\mathcal{T}_t^b(D)} I_j(D,\varepsilon); \tag{3.186}$$

where $I_j(\chi,\varepsilon) = 1$ if the jth downcrossing *or* tangent of level χ from below, is followed by a downcrossing of level $\chi - \varepsilon$ exactly ε time units later, and $I_j(\chi,\varepsilon) = 0$ otherwise. Due to the memoryless property $P(I_j(\chi,\varepsilon) = 1) = e^{-\lambda\varepsilon}, \chi > 0$. Set $\chi = D$; $I_j(D,\varepsilon)$ is independent of $\mathcal{D}_t(D) + \mathcal{T}_t^b(D)$, and $E(I_j(D,\varepsilon)) = e^{-\lambda\varepsilon}$, $j = 1, 2, \ldots$. Taking expected values on both sides of (3.186) gives

$$E(\mathcal{D}_{t+\varepsilon}(D-\varepsilon)) = E(\mathcal{D}_t(D) + \mathcal{T}_t^b(D))e^{-\lambda\varepsilon}. \tag{3.187}$$

By Corollary 3.2 of Theorem 3.3 in Sect. 3.2.5

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(D))}{t} = f(D) \, and \, \lim_{t\to\infty} \frac{E(\mathcal{D}_t(D-\varepsilon))}{t} = f(D-\varepsilon).$$

Also, $\lim_{t\to\infty} \frac{E\left(\mathcal{T}_t^b(D)\right)}{t} = \lambda P_0$, due to the one-to-one correspondence between zero-wait arrivals and tangents of level D from below. Dividing both sides of (3.187) by t, writing $\frac{1}{t} = \frac{1}{t+\varepsilon} \cdot \frac{t+\varepsilon}{t}$ on the left side, and letting $t \to \infty$ gives

$$f(D-\varepsilon) = (f(D) + \lambda P_0)e^{-\lambda\varepsilon}.$$

Letting $\varepsilon \downarrow 0$ yields, since $f(D) = f(D^+)$,

$$f(\mathcal{D}^-) - f(D^+) = \lambda \mathcal{P}_0. \tag{3.188}$$

3. **Case** $x > D$. With probability 1, sample paths are not tangent to level x due to continuous inter-arrival times ($\underset{dis}{=}\text{Exp}_\lambda$). Let ε be $< (x - D)$ and small. Then

$$\mathcal{D}_{t+\varepsilon}(x-\varepsilon) = \sum_{j=1}^{\mathcal{D}_t(x)} I_j(x,\varepsilon) + \sum_{j=1}^{A_t(\varepsilon)} \nu_j(\varepsilon), \tag{3.189}$$

where $\nu_j(\varepsilon) = 1$ if an arrival occurs when $W(t) = \xi \in (x - \varepsilon - D, x - D)$ causing a jump ending at $\xi + D \in (x - \varepsilon, x)$. Note that $P(\nu_j(\varepsilon) = 1)$ $= \int_{x-\varepsilon-D}^{x-D} f(y)dy = \varepsilon f(\xi^*)$, $\xi^* \in (x - D - \varepsilon, x - D)$. But $f(\xi^*) <$

λ, (see Proposition 3.7 in Sect. 3.4.22). So $P(\nu_j(\varepsilon) = 1) < \varepsilon\lambda$. Thus $E(\nu_j(\varepsilon)) < \varepsilon\lambda$, which tends to 0 as $\varepsilon \downarrow 0$.
Taking expected values in (3.189) and dividing both sides by t, gives

$$E\left(\mathcal{D}_{t+\varepsilon}(x-\varepsilon)\right) = E\left(\mathcal{D}_t(x)\right)\cdot e^{-\lambda\varepsilon} + E(A_t)\cdot E(\nu_j(\varepsilon))$$
$$\lim_{t\to\infty}\frac{E\left(\mathcal{D}_{t+\varepsilon}(x-\varepsilon)\right)}{t} = \lim_{t\to\infty}\frac{E\left(\mathcal{D}_t(x)\right)e^{-\lambda\varepsilon}}{t} + \lim_{t\to\infty}\frac{\lambda t E(\nu_j(\varepsilon))}{t}$$
$$f(x-\varepsilon) = f(x)e^{-\lambda\varepsilon} + \lambda E(\nu_j(\varepsilon)).$$

Letting $\varepsilon \downarrow 0$ gives $f(x^-) = f(x)$.

3. **Case** $0 < x < D$. If $0 < x < D$ then, similar to Eq. (3.186) with D replaced by x, and omitting $\mathcal{T}_t^b(x)$, we have

$$\mathcal{D}_{t+\varepsilon}(x-\varepsilon) = \sum_{j=1}^{\mathcal{D}_t(x)} I_j(D,\varepsilon).$$

Taking expected values on both sires, dividing by t, letting $t \to \infty$, then letting $\varepsilon \downarrow 0$, gives $f(x^-) = f(x)$. ■

Proposition 3.10 (1) $F(x)$, $x \geq 0$, has a jump discontinuity at $x = 0$ of size P_0; (2) $F(x)$ is continuous for all $x > 0$.

Proof (1) $F(x)$ has a discontinuity at $x = 0$, since 0 is an atom having probability $F(0) = P_0$. (2) Fix $x > 0$ in the state space. Then x is not an atom (Proposition 3.9 Part (1)); therefore $P(\{x\}) = 0$. That is, x is not a point of increase in probability. Thus x is a point of continuity of $F(\cdot)$. ■

3.10.2 Integral Equation for PDF of Wait

Applying the alternative form of the basic LC integral equation (3.44) with $B(x-y) = 0$ if $x - y < D$ and $B(x-y) = 1$ if $x - y \geq D$, we immediately write an equation for $f(x)$ in terms of $F(\cdot)$, which is a differential equation for the cdf $F(x)$ since $f(x) = F'(x)$,

$$f(x) = \lambda F(x) - \lambda F(x-D),\ x > 0. \tag{3.190}$$

To explain (3.190) in terms of LC, consider a sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 3.16). In (3.190) the left side $f(x)$ is the SP downcrossing rate of level x. SP jumps occur at rate λ, all upward of size D. On the right side of (3.190), the first term $\lambda F(x)$ is the rate of SP jumps that start in state set $[0, x]$. The second term, $-\lambda F(x-D)$ subtracts off the rate of jumps that start in $[0,x]$ and end below x, because jumps starting below $x-D$ cannot upcross x. Thus the right side is the upcrossing rate of x. Rate balance across level x then yields (3.190).

Remark 3.22 The properties in Proposition 3.9, and Eq. (3.190) are readily inferred intuitively upon considering a sample path (Fig. 3.16), and applying LC interpretations of transition rates. Such intuitive insights often lead to formal proofs as in Proposition 3.9.

3.10.3 Analytic Solution for CDF and PDF of Wait

CDF of Wait We give the solution of (3.190), for completeness. For $x \in [0,D)$, $F(x-D) \equiv 0$; thus $f(x) = \lambda F(x)$, or

$$F^{'}(x) - \lambda F(x) = 0,$$

having solution

$$F(x) = A_0 e^{\lambda x}, x \in [0,D)$$

where A_0 is a constant. Letting $x \downarrow 0$, gives the constant $A_0 = P_0 = 1-\rho$. Thus

$$F(x) = P_0 e^{\lambda x}, x \in [0,D).$$

For $x \in [D,2D)$, (3.190) is equivalent to

$$F^{'}(x) - \lambda F(x) = -\lambda P_0 e^{\lambda(x-D)}, x \in [D,2D).$$

Multiplying both sides by the integrating factor $e^{-\lambda(x-D)}$ and then integrating both sides over $[D,x)$ yields the solution up to a constant

$$F(x) = -P_0\lambda(x-D)e^{\lambda(x-D)} + A_1 e^{\lambda(x-D)}, x \in [D,2D).$$

The constant A_1 is determined from the *continuity* of $F(x), x > 0$ (Proposition 3.10). Thus $F(D^-) = F(D)$, or $A_1 = P_0 e^{\lambda D}$ resulting in the solution

$$F(x) = P_0\left(-\lambda(x-D)e^{\lambda(x-D)} + e^{\lambda x}\right), x \in [D,2D).$$

Mathematical induction on (3.190) yields the classical formula for the cdf of wait originally derived in [72],

$$F(x) = P_0 \sum_{i=0}^{m} (-\lambda)^i \frac{(x - iD)^i}{i!} e^{\lambda(x-iD)}, \; x \in [m, (m+1)D), \; m = 0, 1, 2, \ldots \tag{3.191}$$

An alternative form of (3.191) is (e.g., p. 385 in [84]),

$$F(x) = P_0 \sum_{i=0}^{\lfloor x/D \rfloor} (-\lambda)^i \frac{(x - iD)^i}{i!} e^{\lambda(x-iD)}, \; x \geq 0, \tag{3.192}$$

where $\lfloor \alpha \rfloor :=$ greatest integer $\leq \alpha$.

PDF of Wait The pdf $f(x)$ may be obtained by differentiating $F(x)$ with respect to x. More simply, we obtain $f(x)$ by substituting (3.191) into (3.190) giving

$$f(x) = \lambda P_0 e^{\lambda x}, 0 < x < D$$

and for $x \in [mD, (m+1)D)$, $m = 0, 1, 2, \ldots$,

$$\begin{aligned} & f(x) \\ &= \lambda P_0 \sum_{i=0}^{m} (-\lambda)^i \frac{(x-iD)^i}{i!} e^{\lambda(x-iD)} - \sum_{i=0}^{m-1} (-\lambda)^i \frac{(x-(i+1)D)^i}{i!} e^{\lambda(x-(i+1)D)} \\ &= \lambda P_0 (-\lambda)^m \frac{(x-mD)^m}{m!} e^{\lambda(x-mD)} \\ &\quad + \sum_{i=0}^{m-1} \frac{(-\lambda)^i}{i!} [(x - iD)^i e^{\lambda(x-iD)} - (x - (i+1)D)^i e^{\lambda(x-(i+1)D)}]. \end{aligned} \tag{3.193}$$

The pdf $f(x)$ in (3.193) has a discontinuity at $x = D$ (Proposition 3.9 Part (2)). That is $f(D^-) = \lambda P_0 e^{\lambda D}$, and $f(D^-) - f(D) = \lambda P_0$, illustrating that $f(x)$ has a downward jump of size λP_0 at $x = D$. Moreover $f(x)$ is continuous for all other $x > 0$ (see Fig. 3.18). In Fig. 3.18 there is a concave wave in $f(x)$ for $x \in [D, 2D)$, the waviness dampens to the right of $x = 2D$. The cdf $F(x)$,

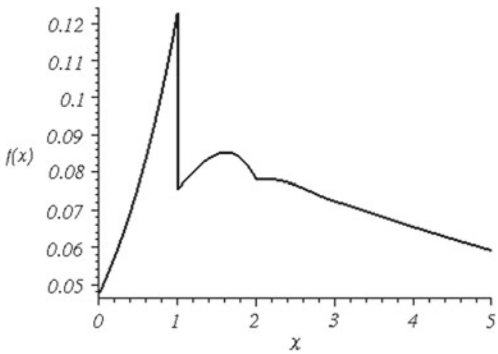


Fig. 3.18 PDF $f(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$, $\rho = 0.95$ (high traffic). Shows discontinuity and downward jump of size λP_0 at $x = D$; and extreme waviness in right neighborhood $[D, 2D)$

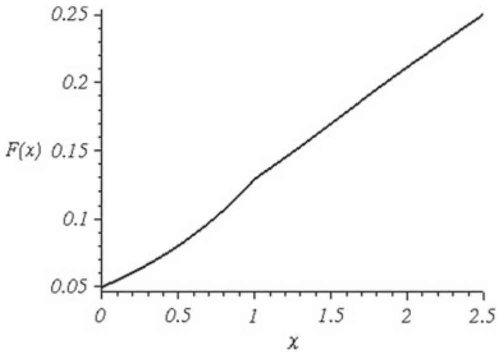


Fig. 3.19 CDF $F(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$. Shows continuity of $F(x), x > 0$; and decrease in slope of $F(x)$ at $x = D$

for the same example, is given in formula (3.191) and plotted in Fig. 3.19, where the continuity of $F(x)$, $x > 0$, and discontinuity of $\frac{d}{dx}F(x)|_{x=D}$ are evident.

Remark 3.23 An LC examination of a typical sample path of $\{W(t)\}_{t \geq 0}$ suggests an isomorphism: $\{$sample-path properties$\} \leftrightarrow \{$analytical properties of $f(x)$ and $F(x)\}$.

3.10.4 Probability Distribution of Number in System

Let N be the number of customers in the system at an arbitrary time point and let W_q (≥ 0) be the wait before service, in the steady-state. Let $P_n := P(N = n)$. Consider a_n, d_n, the probabilities that the number of customers in the system is n just before an arrival, and just after a departure, respectively. Due to Poisson arrivals, $a_n = P_n = d_n$, $n = 0, 1, 2, \ldots$. Arrivals "see" n customers in the system iff $W_q \geq 0$ and $W_q \in ((n-1)D, nD]$, $n = 0, 1, 2, \ldots$. Thus

$$a_n = F(nD) - F((n-1)D) = P_n = d_n, n = 0, 1, 2, \ldots.$$

From (3.191)

$$\begin{aligned}
P_0 &= F(0) - F(-D) = F(0) = P_0 \\
P_1 &= F(D) - F(0) = P_0 e^{\lambda D} - P_0 = P_0(e^{\lambda D} - 1) \\
P_2 &= F(2D) - F(D) = P_0 e^{\lambda D}(-\lambda D + e^{\lambda D} - 1) \\
&\ldots \\
P_n &= F(nD) - F((n-1)\,D), n = 0, 1, 2, \ldots
\end{aligned} \tag{3.194}$$

The cdf of N is

$$P(N \leq n) = \sum_{i=0}^{n} P_i = F(nD), \;\; n = 0, 1, 2, \ldots, \tag{3.195}$$

where $F(nD)$ is computed using (3.191) or (3.192).

3.11 M/Discrete/1 Queue Aka M/D$_n$/1

We look at the M/D$_n$/1 queue, which is an M/G/1 queue with multiple Poisson inputs where the service times are discrete quantities $\{D_n\}_{n=1,2,\ldots}$ (also called an M/Discrete/1 queue). We study the wait before service W_q, and derive analytical properties of its cdf $F(x)$, $x \geq 0$, and pdf $\{P_0, f(x)\}_{x>0}$, where $f(x) = \frac{d}{dx}F(x)$, $x > 0$, wherever the derivative exists. Consider a typical sample path of the virtual wait $\{W(t)\}_{t\geq 0}$ (Fig. 3.20).

Customers arrive in a Poisson stream at rate λ at a single server. For each arrival,

$$P(S = D_i) = p_i, \quad \sum_{i=1}^{N} p_i = 1,$$

where $D_i > 0$, $i = 1, \ldots, N$, and N is a positive integer. Then $E(S) = \sum_{i=1}^{N} p_i D_i$. Without loss of generality, reorder the D_is if necessary, such that

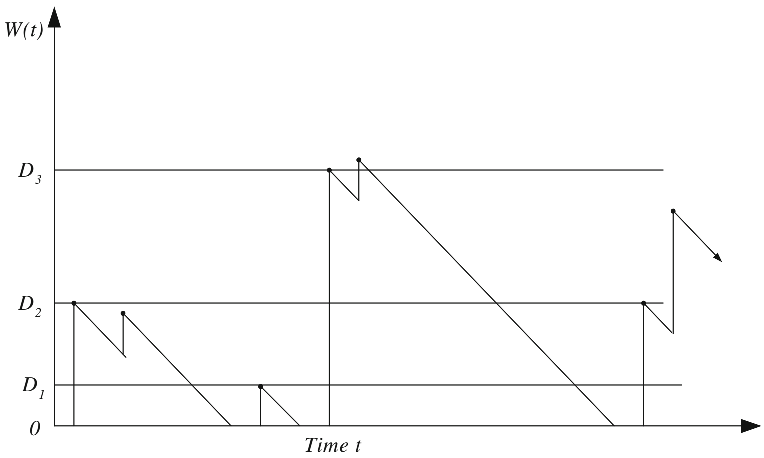


Fig. 3.20 Sample path of $\{W(t)\}_{t\geq 0}$ in $M/\{D_n\}/1$ queue with $N = 3$ service levels

$$0 \equiv D_0 < D_1 < \cdots < D_N < D_{N+1} \equiv \infty.$$

Customers that receive a service time D_i arrive at rate λp_i. The traffic intensity is $\rho = \lambda E(S) < 1$ (stability). Due to Poisson arrivals (e.g., [140]),

$$\lim_{t\to\infty} P(W(t) \le x) = \lim_{n\to\infty} P(W_n \le x),$$

where $\{W_n\}_{n=1,2,\ldots}$ is the process of actual (arrival-point) waits.

We define $f(x)$, $x \ge 0$, as right continuous. The probability of a zero wait is

$$P_0 \equiv F(0) = 1 - \rho = 1 - \lambda \sum_{i=1}^{N} D_i p_i.$$

The cdf and pdf are related by

$$F(x) = P_0 + \int_{y=0}^{x} f(y)dy, \; F(\infty) = P_0 + \int_{y=0}^{\infty} f(x)dx = 1. \quad (3.196)$$

Remark 3.24 The arrival stream may be viewed in two distinct ways:

1. A homogeneous class of customers arrives at rate λ. For each arrival the service time S has a mixture probability distribution with components D_i and mixture probabilities (weights) p_i, $\sum_{i=1}^{N} p_i = 1$.
2. N classes of customers arrive in independent Poisson processes at rates $\lambda_i \equiv \lambda p_i$, $\sum_{i=1}^{N} p_i = 1$, and receive independent service times D_i, $i = 1, \ldots, N$, respectively. This way shows that M/D$_n$/1 is an M/G/1 queue with multiple Poisson inputs.

These two viewpoints yield the same steady-state distribution of wait, as reflected in the two equivalent forms for the traffic intensity $\rho = \lambda \left(\sum_{i=1}^{N} p_i D_i\right) = \sum_{i=1}^{N} \lambda_i D_i$, where $\lambda_i = \lambda p_i$ (see Sect. 5.3.4, p. 319 in [125]).

Remark 3.25 A similar analysis of the M/D$_n$/1 queue applies if $N = \infty$.

3.11.1 Properties of PDF and CDF of Wait

The steady-state distribution of wait has analytical properties given in Proposition 3.11.

Proposition 3.11 In the M/D$_n$ /1 queue, $\{P_0; f(x), x > 0\}$: (1) has exactly one atom which is at $x = 0$ (state $\{0\}$ is an atom); (2) has exactly $\boldsymbol{N}$ down-

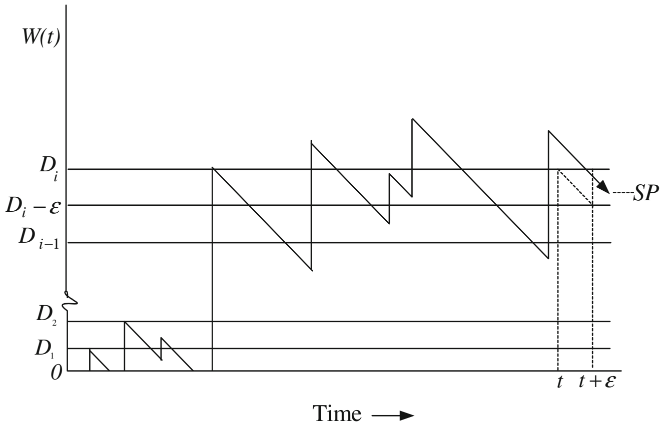


Fig. 3.21 Sample path in M/{D$_n$}/1 showing levels D_i, $D_i - \varepsilon$ and instants $t, t + \varepsilon$. See Proposition 3.11, Proof, Part (2)

ward jump discontinuities of sizes $\lambda p_i P_0$ at $x = D_i$, $i = 1, \ldots, N$; (3) is continuous for all $x > 0$, $x \neq D_i$, $i = 1, \ldots, N$.

Proof Check a typical sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 3.20).

1. State {0} is an atom since a sample path spends a positive proportion of time in {0} (a.s.), namely $P_0 = 1 - \lambda \sum_{i=1}^{N} p_i D_i$. Each sojourn time in $\{0\} \underset{dis}{=} \text{Exp}_\lambda$. There are no other atoms in the state space, since the proportion of time that a sample path spends in each state $x > 0$, is 0.
2. The proof is similar to the proof of Part (2) in Proposition 3.9, Sect. 3.10, upon replacing D, $D - \varepsilon$ by D_i, $D_i - \varepsilon$; λ by λp_i; and where $\varepsilon \in (0, D_i - D_{i-1})$, $i = 1, \ldots, N$; (as in Fig. 3.21). Using similar reasoning as in Proposition 3.9 we obtain

$$f(D_i - \varepsilon) = (f(\mathcal{D}_i) + \lambda p_i P_0)e^{-\lambda\varepsilon}, i = 1, \ldots, N$$

where $\lambda p_i P_0$ is the rate at which the SP makes a tangent to level D_i from below, which is the same as the arrival rate of type-i customers when the system is empty (rate of SP jumps of size D_i from level 0). Letting $\varepsilon \downarrow 0$ results in

$$f(\mathcal{D}_i^-) - f(\mathcal{D}_i) = \lambda p_i P_0, i = 1, \ldots, N.$$

verifying downward jumps at D_i of size $\lambda p_i P_0$, $i = 1, \ldots, N$.

3. The proof is similar to the proof of Part (3) in Proposition 3.9. We thus obtain for $x > 0$, $x \notin \{D_i\}_{i=1,\ldots,N}$

$$f(x-\varepsilon) = f(x) \cdot e^{-\lambda\varepsilon}.$$

Letting $\varepsilon \downarrow 0$ yields $f(x^-) = f(x)$ so that x is a point of continuity. ■

Remark 3.26 From Part (2) of Proposition 3.11, the sum of the downward jumps at points of discontinuity of the pdf $f(x)$ is $\lambda P_0 \sum_{i=1}^{N} p_i = \lambda P_0$. This formula is the same as the size of the single downward jump in the pdf of wait in the M/D/1 model, independent of N.

Proposition 3.12 In the M/$\{D_n\}$/1 queue the steady-state cdf of wait $F(x)$, $x \geq 0$, has a single jump discontinuity at $x = 0$ of size P_0, and is continuous for all $x > 0$.

Proof $F(\cdot)$ has a jump discontinuity at level 0, since $\{0\}$ is an atom having probability $P_0 = F(0)$ (Proposition 3.11, Part (2)). Fix $x > 0$ in the state space. Then x is not an atom (Proposition 3.11, Part (3)). Hence x has probability 0. Thus x is a point of continuity of $F(\cdot)$. ■

3.11.2 Expected Busy Period

From (3.80) the expected busy period is

$$E(\mathcal{B}) = \frac{E(S)}{1-\lambda E(S)} = \frac{1-P_0}{\lambda P_0} = \frac{\sum_{i=1}^{N} D_i p_i}{1-\lambda \sum_{i=1}^{N} p_i D_i}.$$

Another way to compute P_0 is, letting $\mathcal{I}$ denote an idle period,

$$P_0 = \frac{E(\mathcal{I})}{E(\mathcal{I}) + E(\mathcal{B})} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + \frac{\sum_{i=1}^{N} p_i D_i}{1-\lambda\sum_{i=1}^{N} p_i D_i}} = 1 - \lambda \sum_{i=1}^{N} p_i D_i.$$

3.11.3 Integral Equation for PDF of Wait

The alternative form of the LC integral equation for M/G/1 (3.44) leads immediately to an "integral" equation for the pdf $f(x)$ (*differential equation* for cdf $F(x)$),

$$f(x) = \lambda F(x) - \lambda \sum_{i=1}^{N} p_i F(x - D_i)$$
$$= \lambda F(x) - \sum_{i=1}^{N} \lambda_i F(x - D_i), x > 0. \tag{3.197}$$

To explain (3.197) consider a virtual-wait sample path (Fig. 3.20). In (3.197), the left side $f(x)$ is the downcrossing rate of level x. SP jumps occur at rate $\lambda = \sum_{1=1}^{N} \lambda_i$; having size D_i with probability $p_i = \lambda_i/\lambda$. On the right side, the first term $\lambda F(x)$ is the rate at which SP jumps start in state-space set $[0, x]$. The second term, $-\lambda \sum_{i=1}^{N} F(x - D_i)p_i$, subtracts off the rate of those jumps which start in state set $[0, x]$ and end *below* level x. SP jumps of size D_i that start below $x - D_i$, cannot upcross level x. Thus the right side is the sample-path upcrossing rate of x. Rate balance across level x gives (3.197).

3.11.4 Solution for CDF of Wait

Differential equation (3.197) for $F(x)$ is solvable. However the form of $F(x)$ differs in the state-space intervals

$$[0, D_1),\ [D_1, 2D_1),$$
$$\ldots, [j_{11}D_1, D_2),\ [D_2, (j_{11}+1)D_1),\ [(j_{11}+1)D_1, (j_{11}+2)D_1),$$

etc., where $j_{11} = \left\lfloor \frac{D_2}{D_1} \right\rfloor$ (greatest integer $\le \frac{D_2}{D_1}$). At D_3 in the state space, we need to consider $j_{12} = \left\lfloor \frac{D_3}{D_1} \right\rfloor$ and $j_{22} = \left\lfloor \frac{D_3}{D_2} \right\rfloor$, etc. This makes the solution procedure complex. We must keep track of the positions in the state space of the break points where the functional form changes, by considering the relative sizes of $D_1, D_2, \ldots, D_N$. Section 3.11.5 discusses another approach to solve for $F(x), x \ge 0$.

3.11.5 Alternative Approach for CDF of Wait

We can obtain a solution for $F(x)$, $x \ge 0$, using a "specialized" M/D$_n$/1 queue. Assume, without loss of *computational accuracy*, that all D_is are rational numbers. (Rationals can approximate irrational numbers arbitrarily closely.). Let

$$D_1 = k_1 D,\ D_2 = k_2 D, \ldots, D_N = k_N D,$$

where $D = \gcd\{D_1, \ldots, D_N\}$ (gcd denotes greatest common divisor); and $0 < k_1 < k_2 < \cdots < k_N$ are positive integers.

Consider an M/D$_n$/1 queue where $D_i = iD$, $i = 1, \ldots, N$. We call this model an M/$\{iD\}$/1 queue. It is somewhat easier to obtain an analytical solution for the cdf and pdf of wait in M/$\{iD\}$/1 than in M/D$_n$/1. Once a solution for M/$\{iD\}$/1 is obtained, then adjust the *arrival rates* for customers that get service times $k_i D$ $(=D_i)$ so that they correspond to those of the original M/D$_n$/1 queue. The arrival rates for intermediate service time values $\{iD | iD \neq D_i, i = 1, \ldots, N\}$ are set to 0 in that solution. The resulting cdf for M/$\{iD\}$/1 is equal to $F(x)$, $x \geq 0$, for the original M/D$_n$ /1 model (i.e., the solution of (3.197)).

Thus M/$\{iD\}$/1 (where $D = \gcd\{D_1, \ldots, D_N\}$) may be considered as equivalent M/D$_n$/1. Also, it is more amenable analytically and computationally. We next examine the M/$\{iD\}$/1 queue.

3.12 M/{*iD*}/1 Queue

We analyze the M/$\{iD\}$/1 queue, mindful of its close relationship to M/D$_n$ /1 (Sect. 3.11.5).

In M/$\{iD\}$/1 there are N types of arrivals at Poisson rates $\lambda_i, i = 1, \ldots, N$, where N is a positive integer. Customers of type i receive a service time $S = iD$, where $D > 0$ is fixed. Equivalently, customers arrive at Poisson rate λ and get $S = iD$ with probability $p_i, \sum_{i=1}^N p_i = 1$. Thus $\lambda p_i \equiv \lambda_i$. The expected service time is $E(S) = \sum_{i=1}^N iDp_i$. Assume $\lambda E(S) < 1$ (stability). Let P_0 denote the steady-state probability that the system is empty. Then

$$P_0 = 1 - \lambda E(S) = 1 - \lambda \sum_{i=1}^{N} iDp_i = 1 - \sum_{i=1}^{N} iD\lambda_i. \tag{3.198}$$

The M/D/1 queue is a special case of M/$\{iD\}$/1 with $N = 1$. The M/$\{iD\}$/1 queue is a special case of M/$\{$D$_n\}$/1, with $D_n = k_n D$, $n = 1, \ldots, N$, k_n is an integer in the set $\{1, \ldots, N\}$, and $D = \gcd\{D_1, \ldots, D_N\}$ (gcd := greatest common divisor). Paradoxically, M/$\{iD\}$/1 may also be considered as a *generalization* of M/D$_n$ /1! (Sect. 3.11.5).

3.12.1 Integral Equation for CDF of Wait

Let W_q denote the wait before service in the steady state, having cdf $F(x) \equiv P(W_q \leq x), x \geq 0$ and pdf $f(x) = \frac{d}{dx}F(x), x > 0$, wherever the derivative exists. We apply the 'alternative LC' Eq. (3.43) (see also Eq. (3.190) for the M/D/1 queue) relating $f(x)$ and $F(x)$ of wait to obtain

$$f(x) = \lambda F(x) - \lambda \sum_{i=1}^{N} F(x - iD)p_i = \lambda F(x) - \sum_{i=1}^{N} \lambda_i F(x - iD), x > 0. \tag{3.199}$$

Consider the virtual wait process $\{W(t)\}_{t \geq 0}$ (similar to Fig. 3.20). To explain (3.199) the left side is the sample path downcrossing rate of x. On the right side, the term $\lambda F(x)$ is the rate of jumps that start at levels in $[0, x]$. The term $-\sum_{i=1}^{N} \lambda_i F(x - iD)$ subtracts off the rate of jumps that start at levels in $[0, x]$ and *end below* x. For example, $\lambda_i F(x - iD)$ is the rate of type-i jumps of size iD that do not upcross x, since they start below $x - iD$. Hence, the right side is the upcrossing rate of x. Equation (3.199) results by rate balance across level x.

3.12.2 Recursion for CDF of Wait

We now outline a procedure to solve (3.199) recursively for $F(x)$, $x \in [mD, (m+1)D)$, $m = 0, 1, 2, \ldots$. Let

$$F(x) \equiv F_m(x), \;\; f(x) \equiv f_m(x), \;\; x \in [mD, (m+1)D), m = 0, 1, 2, \ldots$$

and $F_{-k}(x) \equiv 0$ if k is a positive integer (see Figs. 3.22 and 3.23). Then write (3.199) as

$$f_m(x) = \lambda F_m(x) - \sum_{i=1}^{N} \lambda_i F_{m-i}(x - iD),$$
$$x \in [mD, (m+1)D), \;\; m = 0, 1, 2, \ldots \tag{3.200}$$

First, let us consider the state-space interval $[0, D)$. The cdf $F(x - D) = 0$ if $x - D < 0$. For $x \in [0, D)$, Eq. (3.200) reduces to

$$f_0(x) = \lambda F_0(x), x \in [0, D),$$

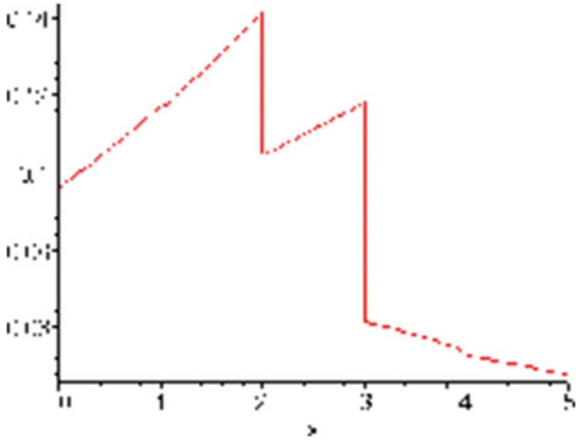


Fig. 3.22 PDF of wait in $M/\{iD\}/1$ queue: $D = 1.0$, four arrival types ($N = 4$), $\lambda = 0.2$, $p_1 = p_4 = 0.01$, $p_2 = 0.39$, $p_3 = .59$. Downward jumps at $x = 1, 2, 3, 4$

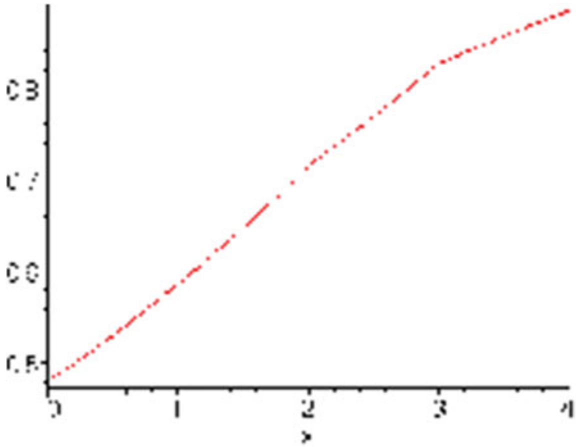


Fig. 3.23 CDF of wait in $M/\{iD\}/1$ queue. $D = 1.0$, $N = 4$, $\lambda = 0.2$, $p_1 = p_4 = 0.01$, $p_2 = 0.39$, $p_3 = 0.59$. The slope decreases abruptly at $x = 1, 2, 3, 4$

or differential equation

$$F_0'(x) = \lambda F_0(x), x \in (0, D),$$

with solution, using the initial condition $F(0) = P_0$,

$$F_0(x) = P_0 e^{\lambda x}, x \in [0, D).$$

Next, on interval $[D, 2D)$, Eq. (3.200) reduces to

$$f_1(x) = \lambda F_1(x) - F_0(x - D)\lambda_1, x \in [D, 2D),$$
$$\text{or } F_1'(x) = \lambda F_1(x) - P_0 e^{\lambda(x-D)}\lambda_1, x \in [D, 2D).$$

The last equation is a differential equation in $F_1(x)$, which is readily solved up to a constant, by using continuity $F_0(D^-) = F_1(D)$, resulting in

$$F_1(x) = P_0\left(e^{\lambda x} + \lambda_1(D-x)e^{-\lambda(D-x)}\right), x \in [D, 2D).$$

Imagine extending the domain of $F_0(x)$ to $[0, \infty)$. The last equation can then be written as

$$F_1(x) = F_0(x) + P_0\lambda_1(D-x)e^{-\lambda(D-x)}, x \in [D, 2D).$$

Similarly we obtain recursively

$$F_2(x), x \in [2D, 3D), \quad F_3(x), x \in [3D, 4D), \quad F_4(x), x \in [4D, 5D),$$

where we extend the domains of $F_m(x)$ to $[m, \infty)$, $m = 0, 1, \ldots$. The recursive formulas in (3.201) below summarize the values of $F(x)$ on state-space interval $[0, 5D)$ by specifying the corresponding functions on intervals $[0, D)$, ..., $[4D, 5D)$, respectively.

$$\begin{aligned}
F_0(x) =& \ P_0 e^{\lambda x}, \\
F_1(x) =& \ F_0(x) + P_0\lambda_1(D-x)e^{-\lambda(D-x)}, \\
F_2(x) =& \ F_1(x) + P_0\left(\lambda_2(2D-x) + \tfrac{\lambda_1^2(2D-x)^2}{2!}\right)e^{-\lambda(2D-x)}, \\
F_3(x) =& \ F_2(x) + P_0\Big(\lambda_3(3D-x) + \lambda_2\lambda_1(3D-x)^2 \\
& + \tfrac{\lambda_1^3(3D-x)^3}{3!}\Big)e^{-\lambda(3D-x)}, \\
F_4(x) =& \ F_3(x) + P_0\Big(\lambda_4(4D-x) + \lambda_3\lambda_1(4D-x)^2 + \tfrac{\lambda_2^2(4D-x)^2}{2!} \\
& + \tfrac{\lambda_2\lambda_1^2(4D-x)^3}{3!} + \tfrac{\lambda_1^4(4D-x)^4}{4!}\Big)e^{-\lambda(4D-x)}.
\end{aligned} \tag{3.201}$$

The recursion (3.201) can be continued indefinitely. The general solution appeared in an article in 2005 by J.F. Shortle and P.H. Brill (see [128]), and is stated below in Sect. 3.12.3.

3.12.3 Solution for CDF and PDF of Wait

Using mathematical induction, it can be shown that an analytical solution of the *indefinitely extended* recursion (3.201) for the cdf of W_q is

$$F_m(x) = P_0 \sum_{i=0}^{m} e^{-\lambda(iD-x)} \sum_{\mathcal{L}\in\mathcal{P}(i)} \frac{(iD-x)^{|\mathcal{L}|}}{H(\mathcal{L})} \prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j, \qquad x \in [mD, (m+1)D),\ m = 0, 1, \ldots, \tag{3.202}$$

where: $\mathcal{P}(i)$ is the set of partitions of integer i; $\mathcal{L}$ is a partition in $\mathcal{P}(i)$; $r_1 > r_2 > \cdots > r_d$ are the distinct integers in $\mathcal{L}$ with multiplicities $n_1, \ldots, n_d$, respectively; $H(\mathcal{L}) \equiv n_1!\,n_2! \cdots n_d!$; $|\mathcal{L}| = n_1 + n_2 + \cdots + n_d$; $\prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j \equiv \lambda_{r_1}^{n_1} \lambda_{r_2}^{n_2} \cdots \lambda_{r_d}^{n_d}$. Also, if $i = 0$, then

$$\sum_{\mathcal{L}\in\mathcal{P}(0)} \frac{(iD-x)^{|\mathcal{L}|}}{H(\mathcal{L})} \prod_{j\in\mathcal{L}} \lambda_j \equiv 1.$$

The pdf of wait is $f_m(x) = F_m'(x)$. Differentiating (3.202) with respect to x, gives for $x \in (mD, (m+1)D)$, $m = 0, 1, 2, \ldots$,

$$f_m(x) = P_0 \sum_{i=0}^{m} e^{-\lambda(iD-x)} \sum_{\mathcal{L}\in\mathcal{P}(i)} (\lambda(iD-x) - |\mathcal{L}|) \frac{(iD-x)^{|\mathcal{L}|-1}}{H(\mathcal{L})} \prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j. \tag{3.203}$$

As a mild check on the cdf of W_q in M/$\{iD\}$/1 given in (3.202), we obtain from it the cdf of W_q in M/D/1 (formula (3.191)), namely

$$F_m(x) = P_0 \sum_{i=0}^{m} e^{-\lambda(iD-x)} \frac{(iD-x)^i}{i!} \lambda^i = P_0 \sum_{i=0}^{m} (-\lambda)^i \frac{(x-iD)^i}{i!} e^{-\lambda(iD-x)}, \qquad x \in [mD, (m+1)D), m = 0, 1, \ldots.$$

To explain, the latter M/D/1 formula it results since: (1) $\lambda_1 = \lambda$ and $\lambda_i = 0$, $i > 1$; (2) for each i, the only partition in $\mathcal{P}(i)$ that contributes positive terms is that of i 1s, $\{1, \ldots, 1\}$; (3) each i yields one such partition with $n_1 = i$, $H(\mathcal{L}) = i!$, and $\prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j = \lambda^i$.

Remark 3.27 In [128], formula (3.202) was derived by inversion of the Laplace transform of wait (see Eq. (3.69). The inversion procedure is at least as involved as the foregoing LC derivation. Moreover, it also requires the induction step. The advantages of the LC approach are: (1) the analysis prior to the induction step is intuitive and completely in the time domain; (2) the effect on the solution, due to the discontinuities in $f(x)$, and the continuity of

$F(x)$, is clear using LC; (3) since LC emphasizes sample paths, it enhances intuitive understanding of the model dynamics, and suggests new avenues for research.

3.13 M/G/1: Wait Related Reneging/Balking

We analyze an M/G/1 queue in which arrivals either: (1) join the system and stay for full service, or (2) balk from joining the system or renege from the waiting line, depending on their estimated (approximate) required arrival-point wait and on their staying resolve (e.g., patience). We assume that the arrivals to the system occur according to a Poisson process at rate λ, from a homogeneous source.

Let $\{W(t)\}_{t\geq 0}$ denote the virtual wait process, and τ_n the arrival time of customer C_n, $n = 1, 2, \ldots$ (Fig. 3.24). Let the service time be

$$S_n = \begin{cases} S \text{ if } C_n \text{ obtains full service,} \\ 0 \text{ if } C_n \text{ balks/reneges before starting service} \end{cases}, n = 1, 2, \ldots,$$

where S has cdf $B(x)$, $x > 0$, and $\overline{B}(x) = 1 - B(x)$, $x \geq 0$, independent of n. The arrival-point waiting time $W(\tau_n^-)$ (:=W_n) is the *required wait* before service of C_n, $n = 1, 2, \ldots$. We assume that a system manager informs C_n

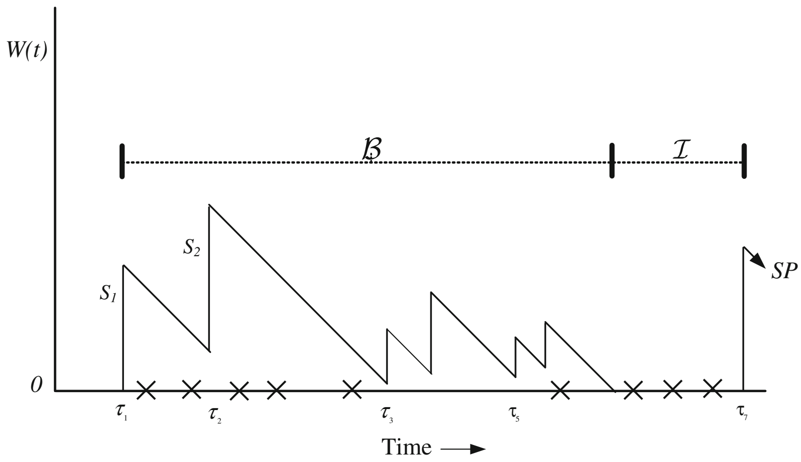


Fig. 3.24 M/G/1 with wait-dependent reneging: busy period $\mathcal{B}$, idle period $\mathcal{I}$; stayers arrive at τ_n, $n = 1, 2, \ldots$; balkers arrive at $\times$

at time τ_n^-, the *estimated* (*approximate)* waiting time W_n. Some arriving customers will balk immediately upon arrival. Others will wait hoping that the approximate wait is higher than the true wait, or that their patience will endure the true wait, or joining has high personal priority. For example, a bus terminal continuously displays electronically the (approximate) wait until the next bus departure; a doctor's office informs an arriving patient about the (approximate) required wait to see the doctor; a telephone answering service informs the caller about the (approximate) wait for the next available agent; etc. If the customers are mechanical devices needing service, the manager may accept or reject entrance to the system, according to the (approximate) required wait before service.

Define, for $n = 1, 2, \ldots$,

$$\theta_n = \begin{cases} 1 \text{ if } C_n \text{ stays and receives full service,} \\ 0 \text{ if } C_n \text{ balks from joining and is cleared.} \end{cases} \tag{3.204}$$

Since the customer source is homogeneous, we define the common random variable $\theta \equiv \theta_n$. Thus θ is a Bernoulli random variable taking the value 1 (***stay***), or 0 (***balk***).

Our aim here is to determine the steady-state mixed pdf of wait denoted by $\{P_0, f(x)\}_{x \geq 0}$, where $f(x)$:= pdf of customers who join and wait for service; and related quantities.

3.13.1 The Staying Function $\overline{R}(y)$, $y \geq 0$

For each $y \geq 0$, we define the common *conditional* probabilities

$$\begin{aligned} \overline{R}(y) &:= P(\theta = 1 | W_n = y),\ y \geq 0, \\ R(y) &:= P(\theta = 0 | W_n = y),\ y \geq 0, \end{aligned} \tag{3.205}$$

independent of $n = 1, 2, \ldots$. From (3.204)

$$\overline{R}(y) + R(y) = 1,\ y \geq 0; \tag{3.206}$$

$P(C_n \text{ stays} | W_n = y) = \overline{R}(y)$; $P(\mathrm{C}_n \text{ balks} | W_n = y) = R(y)$.

3.13.2 Sample Path of $\{W(t)\}_{t\geq 0}$

The r.v., $W(t)$, is the required wait until service of a would-be time-t arrival. Consider a sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 3.24) and an arrival at τ_n^-. If $W_n = y$ then the SP jump size $\underset{dis}{=}$ S having cdf $B(\cdot)$ with probability $\overline{R}(y)$, and jump size = 0 with probability $R(y)$. A would-be arrival at time τ_n just after a balker/reneger arrives (and is cleared), also would have a required wait y until service. This implies $W(\tau_n) = W(\tau_n^-) = y$ if $y > 0$. The sample path would be continuous with slope -1 at τ_n (such t_ns are denoted by $\times$ in Fig. 3.24).

Integral Equation for $\{P_0, f(x)\}_{x\geq 0}$
An integral equation for $\{P_0, f(x)\}_{x\geq 0}$ is (see, e.g., Fig. 1.6 in Sect. 1.7, and Eq. (3.34) in Sect. 3.2.10)

$$f(x) = \lambda\overline{R}(0)P_0\overline{B}(x) + \lambda\int_0^x \overline{B}(x-y)\overline{R}(y)f(y)dy, x > 0, \quad (3.207)$$

with normalizing condition $P_0 + \int_0^\infty f(x)dx = 1$. The left and right sides of Eq. (3.207) are equal to the sample path down- and upcrossing rates of level x, respectively. The upward jump sizes are related to $\overline{B}(\cdot)$ on the right side. Jumps occur at rates that customers stay for service. These rates are state-dependent, viz., $\lambda\overline{R}(y)$, $y \geq 0$. The pdf on the left side is a time-average pdf, the pdf on the right side is the arrival-point pdf at arrival instants; their equality is addressed below. We first briefly discuss the system dynamics with respect to the state-dependence.

E**(Idle Period) and State Dependence**
Consider an idle period $\mathcal{I}$ (Fig. 3.24). When $y = 0$ arrivals enter the system at Poisson rate $\lambda\overline{R}(0)$, implying $E(\mathcal{I}) = 1/\left(\lambda\overline{R}(0)\right)$. Viewed alternatively, customers arrive at Poisson rate λ; at each arrival instant P(the customer stays for service) = $\overline{R}(0)$, and P(the customer balks) = $R(0)$. When $y = 0$, the number of balks until the next stay is distributed as a geometric random variable where a start of service is a *success* and a balk is a *failure*. Thus, E(number of arrivals until a start of service) = $1/\overline{R}(0)$ (see, e.g., p. 37 in [125]). The expected time between arrivals is $1/\lambda$. By independence of the interarrival times and random variable θ, $E(\mathcal{I}) = (1/\lambda)\cdot\left(1/\overline{R}(0)\right) = 1/\left(\lambda\overline{R}(0)\right)$, which agrees with taking $\lambda\overline{R}(0)$ to be the Poisson rate of *stayers*. A similar argument holds for any fixed arrival-point wait $y > 0$.

Equality of Time-Average and Arrival-Point PDFs
In Sect. 8.4.2, Chap. 8, we use the *embedded* LC method to show that the time-average pdf $\{P_0, f(x)\}_{x>0}$ is identical to the limiting arrival-point pdf

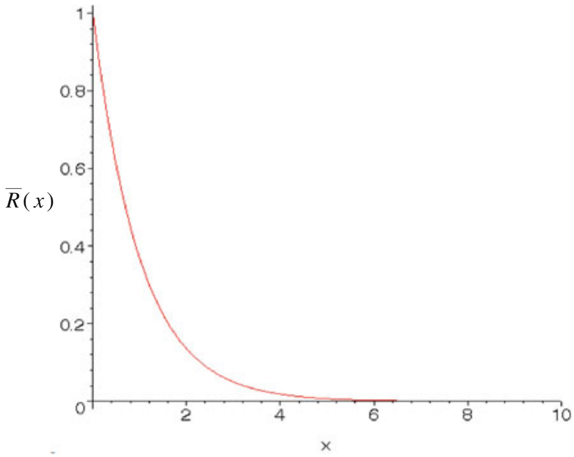


Fig. 3.25 Staying function $\overline{R}(x) = e^{-rx}$. ($r = 1$ in the diagram); limit $L = 0$. $R(x) = 1 - \overline{R}(x) = 1 - e^{-rx\cdot}$, $x \geq 0$, the cdf of an Exp_r random variable

as $n \to \infty$, denoted by $\left\{P_{\iota,0},\ f_\iota(x)\right\}_{x>0,\ldots}$, because both pdfs satisfy integral equation (3.207) and the same normalizing condition.

Form of the Staying Function $\overline{R}(\cdot)$

We assume $\overline{R}(y)$, $y \geq 0$, is a monotone, piecewise continuous, non-increasing function (decreasing in the wide sense), with range a subset of [0, 1]. (See Figs. 3.25, 3.26 and 3.27.) Then $\lim_{y\to\infty} \overline{R}(y) := L \in [0, 1]$ exists. If $\overline{R}(y) \equiv 1$, $y \geq 0$, then $L = 1$, and there would be no reneging or balking; each arrival would receive full service. The model would be a standard M/G/1 queue.

Remark 3.28 In a more general model, $\overline{R}(y)$ may be an arbitrary function such that $\overline{R}(y) \in [0, 1]$, $y \geq 0$, is not necessarily monotone. In that case, the

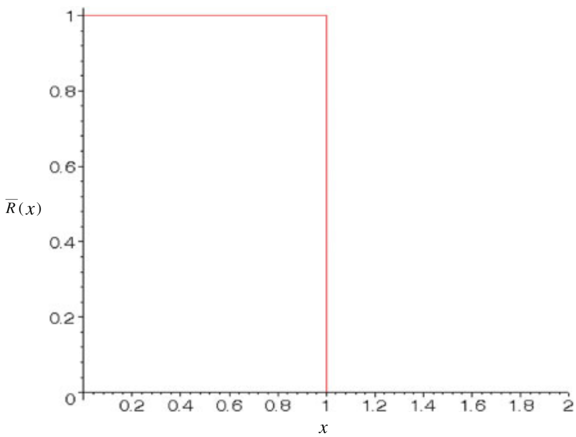


Fig. 3.26 Staying function $\overline{R}(x) = 1$, $0 < x < 1$, $\overline{R}(x) = 0$, $x \geq 1$. Limit $L = 0$. $R(x) = 1 - \overline{R}(x)$, is the cdf of a deterministic r.v.

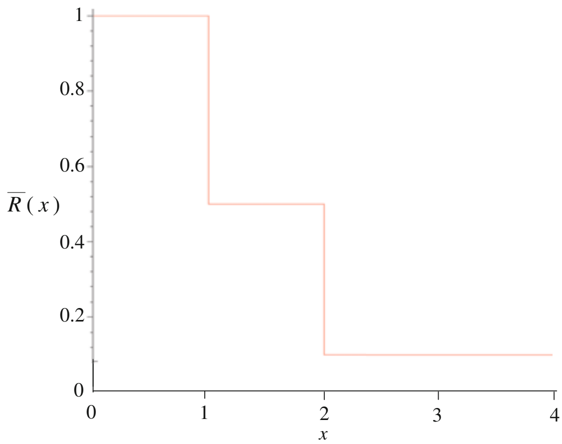


Fig. 3.27 $\overline{R}(x) = 1$, $x < 1$, $\overline{R}(x) = 0.5$, $1 \le x < 2$, $\overline{R}(x) = 0.1$, $x \ge 2$. $R(x)$ is not a cdf. Limit $L > 0$

presented analysis applies as well. However, the stability condition would be slightly modified (see Theorem 3.8 and Remark 3.31 below).

Interestingly, the renege/balk M/G/1 queue where $\overline{R}(x) = 1 \cdot \boldsymbol{I}_{[0,1)}(x) + 0 \cdot \boldsymbol{I}_{[1,\infty)}(x)$ is essentially the same as M/G/1 with a threshold at level 1 denoted as Variant 3 (with $K = 1$) in Sect. 3.16.6 below.

Proportion of Customers that Get Full Service
Denote by q_S the proportion of arrivals that are stayers. Then $q_S := P$(an arbitrary arrival gets full service), namely

$$q_S = \overline{R}(0)P_0 + \int_{y=0}^{\infty} \overline{R}(y) f(y) dy. \tag{3.208}$$

The proportion of customers that balk upon knowing their actual or approximate required wait before service is

$$q_B = 1 - q_S = R(0)P_0 + \int_{y=0}^{\infty} R(y) f(y) dy.$$

3.13.3 M/M/1: Wait Dependent Reneging/Balking

We now study the particular case where the service times of stayers are $\underset{dis}{=} \text{Exp}_\mu$, with $\overline{B}(x) = e^{-\mu x}$, $x \ge 0$. Then (3.207) becomes

$$f(x) = \lambda P_0 \overline{R}(0) e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy. \tag{3.209}$$

Applying differential operator $\langle D + \mu \rangle$ to both sides of (3.209) yields the first order differential equation

$$\begin{aligned} \langle D + \mu \rangle\, f(x) &= \lambda \overline{R}(x) f(x), \\ f'(x) + (\mu - \lambda \overline{R}(x)) f(x) &= 0, \\ \frac{f'(x)}{f(x)} = \frac{d \ln f(x)}{dx} &= -(\mu - \lambda \overline{R}(x)). \end{aligned}$$

Integration on both sides of the last equation with respect to x, followed by exponentiation gives

$$f(x) = A e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)}, x > 0, \tag{3.210}$$

where A is a constant. Letting $x \downarrow 0$ in (3.209) and (3.210) implies

$$f(0) = A = \lambda P_0 \overline{R}(0).$$

From LC, $f(0)$ is the SP *entrance* rate into $\boldsymbol{T} \times \{0\}$ (i.e., into level 0) from above. The term $\lambda P_0 \overline{R}(0)$ is the SP *exit* rate of level 0 above (i.e., into state-space interval $(0, \infty)$). The resulting pdf of wait is

$$f(x) = \lambda P_0 \overline{R}(0) e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)}, x > 0. \tag{3.211}$$

The normalizing condition $P_0 + \int_{x=0}^{\infty} f(x) dx = 1$ leads to

$$P_0 = \frac{1}{1 + \lambda \overline{R}(0) \int_{x=0}^{\infty} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)} dx}. \tag{3.212}$$

3.13.4 M/M/1: Reneging/Balking-Stability Condition

In the $M_\lambda/M_\mu/1$ queue, Theorem 3.8 below gives a necessary and sufficient condition relating λ and μ, such that $\{P_0, f(x)\}_{x>0}$ exists (stability).

Theorem 3.8 In $M_\lambda/M_\mu/1$ with wait-time dependent reneging/balking assume the staying function $\overline{R}(x)$, $x \geq 0$, is monotone non-increasing and piecewise continuous. Let $L = \lim_{x\to\infty} \overline{R}(x)$. A necessary and sufficient condition for stability is

$$\begin{aligned} &0 < \lambda < \tfrac{\mu}{L} \quad \text{if } 0 < L \le 1, \\ &0 < \lambda < \infty \quad \text{if } L = 0. \end{aligned} \tag{3.213}$$

Proof (Adapted from [90]) By the hypothesis $1 \ge \overline{R}(a) \ge \overline{R}(b) \ge 0$ whenever $a < b$; hence $\lim_{x\to\infty} \overline{R}(x) := L \in (0, 1]$ exists (see, e.g., Problem *8, p. 119, in Chap. 8, in [137]). Stability holds iff the discrete state $\{0\}$ is positive recurrent iff $0 < P_0 \le 1$. Let

$$I := \int_{x=0}^{\infty} e^{-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy} dx,$$

in the denominator of (3.212). Stability holds iff

$$I < \infty. \tag{3.214}$$

We now show that the condition (3.214) is equivalent to the condition (3.213) above. We have $L \le \overline{R}(x), x \ge 0$, because L is the greatest lower bound (i.e., *glb*, *infimum*) of the range of $\overline{R}(\cdot)$. Hence

$$\begin{aligned} \lambda L x \quad &= \quad \lambda \int_{y=0}^{x} L dx \le \lambda \int_{y=0}^{x} \overline{R}(y)dy \\ &\Longleftrightarrow \quad e^{-\mu x + \lambda L x} \le e^{-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy} \\ &\Longleftrightarrow \quad \int_{x=0}^{\infty} e^{-(\mu - \lambda L)x} dx \le I. \end{aligned} \tag{3.215}$$

For a given $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that $\overline{R}(x) < \varepsilon + L$ for $x > M_\varepsilon$. Thus

$$\begin{aligned} \lambda \int_{y=0}^{x} \overline{R}(y)dy &< \lambda \int_{y=0}^{M_\varepsilon} \overline{R}(y)dy + \lambda \int_{y=M_\varepsilon}^{x} (\varepsilon + L)\, dy \\ &= C_1 + \lambda\,(\varepsilon + L)\, x, \;\; x > M_\varepsilon \end{aligned}$$

$$\begin{aligned} &\Longrightarrow \; e^{-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy} < C_2 e^{-\mu x + \lambda(\varepsilon + L)x}, \;\; x > M_\varepsilon \\ &\Longrightarrow \; \int_{x=M_\varepsilon}^{\infty} e^{-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy} dx < C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu + \lambda L + \lambda\varepsilon)x} dx \\ &\Longrightarrow \; I < C_3 + C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu + \lambda L + \lambda\varepsilon)x} dx, \end{aligned} \tag{3.216}$$

where C_1, C_2, C_3 are positive constants. Combining inequalities (3.215) and (3.216) gives

$$\int_{x=0}^{\infty} e^{-(\mu-\lambda L)x} dx \leq I < C_3 + C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu+\lambda L+\lambda\varepsilon)x} dx. \qquad (3.217)$$

In (3.217), if $I < \infty$ then

$$\int_{x=0}^{\infty} e^{-(\mu-\lambda L)x} dx < \infty \iff \mu - \lambda L > 0. \qquad (3.218)$$

If $\mu - \lambda L > 0$ then choose ε so that $-\mu + \lambda L + \lambda\varepsilon < 0$, i.e., $\varepsilon < \frac{\mu-\lambda L}{\lambda}$. Then

$$\int_{x=M_\varepsilon}^{\infty} e^{(-\mu+\lambda L+\lambda\varepsilon)x} dx < \infty \implies I < \infty. \qquad (3.219)$$

The stability condition (3.213) is equivalent to (3.218) and (3.219). ■

Remark 3.29 To shed additional perspective on the stability condition (3.213), consider the exponent in the integrand of

$$I \equiv \int_{x=0}^{\infty} e^{-\mu x+\lambda \int_{y=0}^{x} \overline{R}(y)dy} dx.$$

The function μx is linear with slope $\mu > 0$. The function of x, $\int_{y=0}^{x} \overline{R}(y)dy$, $x > 0$, is positive and increasing with slope $\frac{d}{dx}\int_{y=0}^{x} \overline{R}(y)dy = \overline{R}(x)$, $x > 0$. If $\overline{R}(x)$, $x > 0$, is *strictly* decreasing and differentiable, then $\int_{y=0}^{x} \overline{R}(y)dy$ is concave since $\frac{d^2}{dx^2}\int_{y=0}^{x} \overline{R}(y)dy = \frac{d}{dx}\overline{R}(x) < 0$, $x > 0$. Additionally, $\lim_{x\to\infty} \frac{d}{dx}\int_{y=0}^{x} \overline{R}(y)dy = \lim_{x\to\infty} \overline{R}(x) = L$. We compare the graphs of μx and $\lambda \int_{y=0}^{x} \overline{R}(y)dy$, $x > 0$ in Fig. 3.28.

If $L > 0$ then there exists $M \geq 0$ such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for all $x \geq M$ iff $\mu - \lambda L > 0$ iff $\lambda < \mu/L$. If $L = 0$, there exists $M \geq 0$ such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for all $x \geq M$ iff $\mu \geq \lambda \cdot 0$. In that case λ can assume any positive value, i.e., $\lambda \in (0, \infty)$.

Remark 3.30 If $\overline{R}(x)$ is piecewise continuous, we can obtain similar perspective as in Remark 3.29.

Another Look at Theorem 3.8

We provide an alternative verification of the stability condition, in order to clarify the intuition behind the result. Consider an *optimization problem*

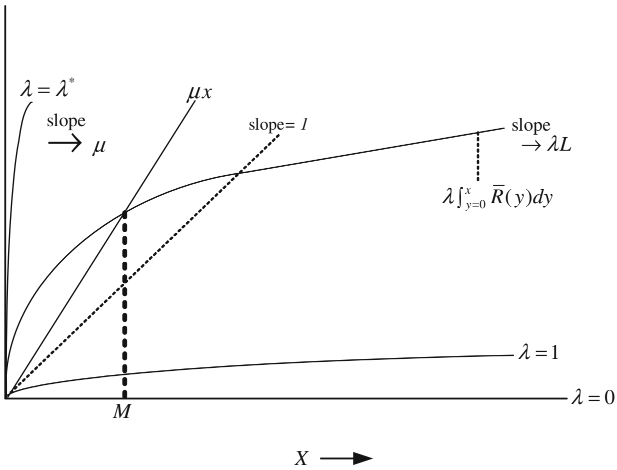


Fig. 3.28 Functions μx and $\lambda \int_{y=0}^{x} \overline{R}(y)dy$, indicating M such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for $x \geq M$. Indicates range $0 < \lambda < \lambda^*$ such that stability holds. The system is stable for λ if $\lambda \int_{y=0}^{x} \overline{R}(y)dy$ intersects and remains below μx thereafter

where λ is the decision variable. We shall derive a range $0 < \lambda < \lambda^*$ for which there exists $M \geq 0$ such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for all $x \geq M$, thereby making the system stable (see Fig. 3.28). The value λ^* is the solution of the following optimization problem **P**. (Note that $\mu > 0$, $L \geq 0$.)

Problem **P**	
Maximize	λ
such that	$\mu - \lambda L \geq 0$
subject to	$\lambda > 0.$

The solution of problem **P** is readily seen to be

$$\lambda^* = \begin{cases} \frac{\mu}{L} \text{ if } L > 0, \\ \infty \text{ if } L = 0, \end{cases}$$

which is the same result as in Theorem 3.8.

Remark 3.31 The stability condition given in Theorem 3.8 was originally proved in [16] together with a theorem in which $\overline{R}(y)$, $y \geq 0$ may be other than monotone non-increasing. That proof is based on the fact that

$$\int_{x=0}^{\infty} e^{-\mu x+\lambda \int_{y=0}^{x} \overline{R}(y)dy} dx = \int_{x=0}^{\infty} e^{-\mu x} \cdot e^{\lambda \int_{y=0}^{x} \overline{R}(y)dy} dx$$

is the Laplace transform of $e^{\lambda \int_{y=0}^{x} \overline{R}(y)dy}$ evaluated at the parameter μ. A sufficient condition for the Laplace transform to be finite is that $e^{\lambda \int_{y=0}^{x} \overline{R}(y)dy}$ is of exponential order. Let $\overline{L} = \limsup_{x\to\infty} \overline{R}(x)$. A *sufficient* condition for stability is

$$\lambda < \frac{\mu}{\overline{L}} \text{ if } \overline{L} > 0,$$
$$\lambda < \infty \text{ if } \overline{L} = 0.$$

3.13.5 M/M/1: Reneging/Balking-Exponential $\overline{R}(\cdot)$

We illustrate the M/G/1 model by taking G(·) := Exp$_\mu$. Let $\overline{B}(x) = e^{-\mu x}$, $x \geq 0$, and $\overline{R}(y) = e^{-ry}$, $y > 0$, $r > 0$. Thus $\overline{R}(y)$ is monotone decreasing and $L = \lim_{y\to\infty} \overline{R}(y) = 0$ in the notation of Sect. 3.13.3. Also, $\overline{R}(0) = 1$, so that all zero-wait customers join the system.

Equation (3.209) becomes

$$f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} e^{-ry} f(y)dy. \tag{3.220}$$

Substituting e^{-ry} for $\overline{R}(y)$ in (3.211) gives the pdf of wait $\{P_0, f(x)\}_{x>0}$ as

$$f(x) = \lambda P_0 e^{-\mu x+\frac{\lambda}{r}(1-e^{-rx})} = \lambda e^{\lambda/r} P_0 e^{-\mu x-\frac{\lambda}{r}e^{-rx}}, x > 0. \tag{3.221}$$

Substituting (3.221) into (3.212) yields

$$P_0 = \frac{1}{1+\lambda e^{\lambda/r} \int_{x=0}^{\infty} e^{-\mu x-\frac{\lambda}{r}e^{-rx}} dx} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+e^{\lambda/r} \int_{x=0}^{\infty} e^{-\mu x-\frac{\lambda}{r}e^{-rx}} dx}. \tag{3.222}$$

In the denominator of P_0 the term $\int_{x=0}^{\infty} e^{-\mu x-\frac{\lambda}{r}e^{-rx}} dx < 1/\mu < \infty$ for *every* trio of positive numbers $\{\lambda, \mu, r\}$, since the integrand $e^{-\mu x-\frac{\lambda}{r}e^{-rx}} < e^{-\mu x}$, $x \geq 0$. Thus $P_0 > 0$ for all positive $\{\lambda, \mu, r\}$. In particular $P_0 > 0$ for *every* arrival rate $\lambda > 0$. This adds credence to Theorem 3.8 above when $\lim_{x\to\infty} \overline{R}(x) = L = 0$.

Expected Busy Period $E(\mathcal{B})$

In the standard M/G/1 queue, $E(\mathcal{B}) = E(S)/(1 - \lambda E(S))$. However, in M/G/1 with balking $P_0 \neq 1 - \lambda E(S)$. Hence, we use the more fundamental formula for $E(\mathcal{B})$ in terms of P_0. From (3.82) and (3.222),

$$E(\mathcal{B}) = \frac{1 - P_0}{f(0)} = \frac{1 - P_0}{\lambda P_0}$$

$$= e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx = \int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1 - e^{-rx})} dx. \tag{3.223}$$

We can infer formula (3.223) immediately since P_0 in (3.222) has the form

$$P_0 = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + E(\mathcal{B})} = \frac{E(\mathcal{I})}{E(\mathcal{I}) + E(\mathcal{B})},$$

and $E(\mathcal{I}) = 1/\lambda$, because all zero-wait customers stay for service if $\overline{R}(y) = e^{-ry},\ y \geq 0$.

3.13.6 M/M/1: Reneging/Balking and Standard M/M/1

Assume $\lambda < \mu$ (stability condition for standard M/M/1). In (3.223), $(1 - e^{-rx}) < rx, x > 0$ and $(1 - e^{-r \cdot 0}) = 0$. Letting subscript 'b' represent M/M/1 with reneging/balking, and subscript 's' the standard M/M/1, we have, since $(1 - e^{-rx})/r < x, x > 0$,

$$E(\mathcal{B}_{\mathrm{b}}) = \int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1 - e^{-rx})} dx < \int_{x=0}^{\infty} e^{-(\mu - \lambda)x} dx = \frac{1}{\mu - \lambda} = E(\mathcal{B}_{\mathrm{s}}).$$

In (3.222), we again apply the inequality

$$\int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1 - e^{-rx})} dx < \frac{1}{\mu - \lambda},$$

which gives

$$P_{\mathrm{b},0} > \frac{1}{1 + \lambda \cdot \frac{1}{\mu - \lambda}} = 1 - \frac{\lambda}{\mu} = P_{\mathrm{s},0}.$$

The comparisons for $E(\mathcal{B})$ and P_0 are intuitive. In the reneging/balking model, the arrival rate of customers that increase workload is $\lambda \overline{R}(y),\ y \geq 0$. In the standard model, it is $\lambda > \lambda \overline{R}(y),\ y > 0$.

3.13.7 M/M/1: Reneging/Balking-Number in System

Let P_n, a_n, d_n denote the steady-state probabilities of n stayers in the system at an arbitrary time point, just before an arrival and just after a departure, respectively. Then $a_n = d_n = P_n$, $n = 0, 1, 2, \ldots$, (see Sect. 8.2.2, p. 500 in [125]); and P_0 is given in (3.222). Furthermore, since $\overline{R}(y) = e^{-ry}$, $y \geq 0$,

$$\begin{aligned} d_n &= \int_{x=0}^{\infty} \left(e^{-\lambda \int_0^x \overline{R}(y)dy}\right) \frac{\left(\lambda \int_0^x \overline{R}(y)dy\right)^{n-1}}{(n-1)!} f(x)dx \\ &= \lambda P_0 \int_{x=0}^{\infty} \frac{(\frac{\lambda}{r}(1-e^{-rx}))^{n-1}}{(n-1)!} e^{-\mu x} dx, n = 1, 2, \ldots. \end{aligned} \tag{3.224}$$

(see Eq. (3.76) in Sect. 3.4.8).

In formula (3.224), $\lambda\overline{R}(y)$ $(=\lambda e^{-ry})$ is the arrival rate of stayers when the required wait is y.

Remark 3.32 We outline a derivation of (3.224) using an approximation of $\overline{R}(x)$ by a step function. Let $[0,\Omega)$ be a large waiting-time interval in the state space. Partition $[0, \Omega)$ into m subintervals $\Delta_i = [x_i, x_{i+1})$, $i = 0, \ldots, m-1$, where $x_0 = 0$, $x_m = \Omega$. We then approximate $\overline{R}(y)$ by $\overline{R}(y) \equiv \overline{R}(x_i)$, $y \in \Delta_i$. Thus the arrival rate of stayers is a constant $\lambda\overline{R}(x_i)$ if the required wait $y \in [x_i, x_{i+1})$ at an arrival instant. The probability that $n-1$ stayers arrive during an individual required wait $y \in \Delta_i$ is approximately

$$\frac{e^{-\lambda\overline{R}(x_i)x_i'}(\lambda\overline{R}(x_i)x_i')^{n-1}}{(n-1)!}$$

where $x_i' \in \Delta_i$. The probability that $n-1$ stayers arrive during $(0, \Omega)$ is approximately the Riemann sum

$$\sum_{i=0}^{m-1} \frac{e^{-\lambda\overline{R}(x_i)x_i'}(\lambda\overline{R}(x_i)x_i')^{n-1}}{(n-1)!} f(x_i'')\Delta_i$$

where $x_i'' \in \Delta_i$. Let $m \to \infty$ and $\Delta_i \downarrow 0$, $i = 0, \ldots, m-1$. Then $x_i, x_i', x_i'' \to x$ and

$$\begin{aligned} &\lim_{\substack{m\to\infty \\ \Delta_i \downarrow 0}} \sum_{i=0}^{m-1} e^{-\lambda\overline{R}(x_i)x_i'} \frac{(\lambda\overline{R}(x_i)x_i')^{n-1}}{(n-1)!} f(x_i'')\Delta_i \\ &= \int_{x=0}^{\Omega} e^{-\lambda\overline{R}(x)x} \frac{(\lambda\overline{R}(x)x)^{n-1}}{(n-1)!} f(x)dx. \end{aligned}$$

Letting $\Omega \to \infty$ implies (3.224), where $f(x)$ is given by (3.221).

3.13.8 Proportion of Customers Served

In M/M/1 with wait-time dependent reneging/balking and $\overline{R}(y) = e^{ry}$, $y \geq 0$, from (3.208), (3.221) and (3.222), the proportion of customers that get complete service is

$$q_S = P_0 + \int_{x=0}^{\infty} e^{-rx} f(x)dx = \frac{1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx} - rx} dx}{1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx}. \tag{3.225}$$

The proportion of customers that renege from the waiting line is $1 - q_S$.

In the expressions for P_0, $E(\mathcal{B})$, and q_S the integrals do not have closed forms. They can be evaluated using series expansion or numerical methods, for given values of λ, μ, and r.

3.14 M/G/1 with Priorities

Assume N types of customers arrive at a single-server system at independent Poisson rates $\lambda_i, i = 1, \ldots, N$. We denote the type-$i$ service time as S_i having cdf $B_i(x), x > 0, \overline{B}_i(x) = 1 - B_i(x), x \geq 0$, and pdf $b_i(x), x > 0$. We assume type 1 ($i = 1$) has the highest priority, type 2 the next highest, …, and type N ($i = N$) the lowest priority. The service discipline is FCFS within priority classes. The priority discipline is non-preemptive, i.e., any customer that starts service is allowed to complete it without interruption. The customer at the head of the highest-priority line, among all waiting customers, will start service immediately after the next service completion.

Denote the steady-state pdf and cdf of wait before service of a type i customer, by $\{P_0, f_i(x)\}_{x>0}$, and $F_i(x)$, $x \geq 0$ respectively. The probability P_0 of a zero wait, is independent of customer type.

3.14.1 Two Priority Classes

For exposition we consider two priority classes, so that $N = 2$. We will confirm the well-known stability condition, $\lambda_1 E(S_1) + \lambda_2 E(S_2) < 1$, using an

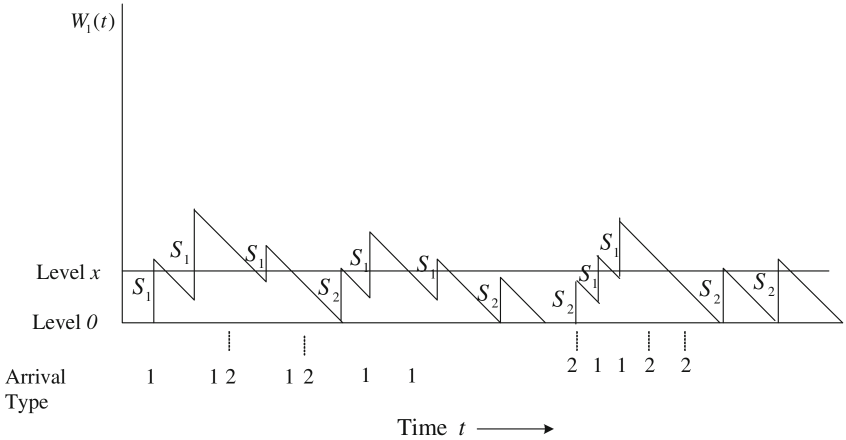


Fig. 3.29 Sample path of virtual wait for high priority type-1 arrivals. Low priority type-2 arrivals that must wait, start service at the end of a $\mathcal{B}_1$ or a $\mathcal{B}_{21}$ (Fig. 3.30) busy period. All type 2 jumps start at level 0

LC approach. Let $\{W_1(t)\}_{t\geq 0}$ be the virtual wait process for *type*-1 *customers*; a sample path is shown in Fig. 3.29. Fix level $x > 0$ in the state space.

3.14.2 Integral Equation for $\{P_0, f_i(x)\}_{x>0}$

From the sample path, we construct the integral equation

$$\begin{aligned} f_1(x) = {} & \lambda_1 \overline{B}_1(x) P_0 + \lambda_2 \overline{B}_2(x) P_0 + \lambda_1 \int_{y=0}^{x} \overline{B_1}(x-y) f_1(y) dy \\ & + \lambda_2 (1 - P_0) \overline{B}_2(x). \end{aligned} \tag{3.226}$$

To explain (3.226), the left side $f_1(x)$ is the SP downcrossing rate of x (as in Theorem 1.1 in Chap. 1). On the right side, the terms $\lambda_1 \overline{B}_1(x) P_0$ and $\lambda_2 \overline{B}_2(x) P_0$ are respectively the SP upcrossing rates of x due to type-1 and type-2 arrivals, when the system is empty. The term $\lambda_1 \int_{y=0}^{x} \overline{B_1}(x - y) f_1(y) dy$ is the upcrossing rate of x due to type-1 arrivals that wait a positive time $y \in (0, x)$. The term $\lambda_2(1 - P_0)\overline{B}_2(x)$ is the upcrossing rate of x due to type-2 arrivals that wait positive times before they start service. The first-in-line of such type 2s must wait *until the end* of a type 1 busy period to start service. Any other such type 2s wait longer before they start service. Those type 2s can start service only when the type-1 virtual wait hits level 0. The corresponding SP jumps of size S_2 start at level 0. The *long-run rate* at which such type 2s start service is $\lambda_2(1 - P_0)$ since all type 2s must eventually get served in a finite time, due to stability.

3.14.3 Stability Condition

Integrate both sides of (3.226) with respect to x on $(0, \infty)$. Since $\int_{x=0}^{\infty} f_1(x)dx = 1 - P_0$, and $\int_{x=0}^{\infty} \overline{B}_i(x)dx = E(s_i)$ some algebra yields

$$P_0 = 1 - \lambda_1 E(S_1) - \lambda_2 E(S_2) = 1 - \rho_1 - \rho_2, \tag{3.227}$$

where $\rho_i = \lambda_i E(S_i)$, $i = 1, 2$. For stability, we must have $0 < P_0 < 1$, or

$$0 < \rho_1 + \rho_2 < 1, \tag{3.228}$$

which implies both $\rho_1 < 1$ and $\rho_2 < 1$.

3.14.4 Expected Wait of High Priority Customers

We confirm the known formula for the expected wait of type-1 customers using (3.226). Denote the wait in queue before service of an arbitrary type-1 arrival by $W_{q,1}$. Multiplying both sides of (3.226) by x and integrating on $(0, \infty)$ with respect to x, the left side becomes $\int_0^{\infty} x f_1(x)dx = E(W_{q1})$; the right side results in the equation

$$E(W_{q1}) = \left(\lambda_1 \frac{E(S_1^2)}{2} + \lambda_2 \frac{E(S_2^2)}{2}\right) P_0 + \lambda_1 E(S_1) E(W_{q1})$$
$$+\lambda_1(1 - P_0)\frac{E(S_1^2)}{2} + \lambda_2(1 - P_0)\frac{E(S_2^2)}{2}.$$

Simplifying yields the familiar result (e.g., p. 545 in [125])

$$E(W_{q1}) = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1 - \rho_1)}. \tag{3.229}$$

3.14.5 Equation for PDF of Wait of Type-2 Customers

Let $\{W_2(t)\}_{t\geq 0}$ be the virtual wait process of type-2 customers. Let $W_{q,2}$ be the steady-state wait. Denote the pdf of $W_{q,2}$ by $\{P_0, f_2(x)\}_{x>0}$, for which we now develop an integral equation.

Preliminaries

Let $\mathcal{B}_1$ denote a an M/G/1 type-1 *busy period*, consisting of type-1s only, having cdf $\boldsymbol{B}_1(x), x > 0$ and $\overline{\boldsymbol{B}}_1(x) = 1 - \boldsymbol{B}_1(x), x \geq 0$. We let $\mathcal{B}_{2,1}$ denote a busy period in which the first service is type 2, and all subsequent services are type 1 (Fig. 3.30). Let random variable $N_{S_{2,1}}$ denote the number of strict descending ladder points that occur in a sample path of a $\mathcal{B}_{2,1}$ busy period. Then $N_{S_{2,1}}$ has the same distribution as the number of type-1 customers that arrive during a type-2 *service time* S_2. Thus we have

$$\mathcal{B}_{2,1} \underset{dis}{=} S_2 + \sum_{i=1}^{N_{S_{2,1}}} \mathcal{B}_{1,i}, \tag{3.230}$$

where the $\mathcal{B}_{1,i}$s are i.i.d. random variables distributed as an M/G/1 type-1 busy period $\mathcal{B}_1$ independent of $N_{S_{2,1}}$. Equation (3.230) follows due to the memoryless property of the type-1 inter-arrival times ($\underset{dis}{=}\text{Exp}_{\lambda_1}$). (A related discussion of busy period structure is given above in Sect. 3.4.12.)

We illustrate the meaning of $N_{S_{2,1}}$ in Fig. 3.30, with $N_{S_{21}} = 3$. There are three type-1 sub-busy periods in $B_{2,1}$. There are four vertical gaps, each distributed as an inter-arrival time, separating and bordering on these three sub-busy periods. The basic observation is that the sum of the four gaps is equal to S_2.

From (3.80)

$$E(\mathcal{B}_1) = \frac{E(S_1)}{1 - \lambda_1 E(S_1)}. \tag{3.231}$$

Taking expected values in (3.230) we obtain

$$\begin{aligned} E(\mathcal{B}_{2,1}) &= E(S_2) + \lambda_1 E(S_2) E(\mathcal{B}_1) \\ &= E(S_2) + \lambda_1 E(S_2) \frac{E(S_1)}{1 - \lambda_1 E(S_1)} \\ &= \frac{E(S_2)}{1 - \lambda_1 E(S_1)} = \frac{E(S_2)}{1 - \rho_1}. \end{aligned} \tag{3.232}$$

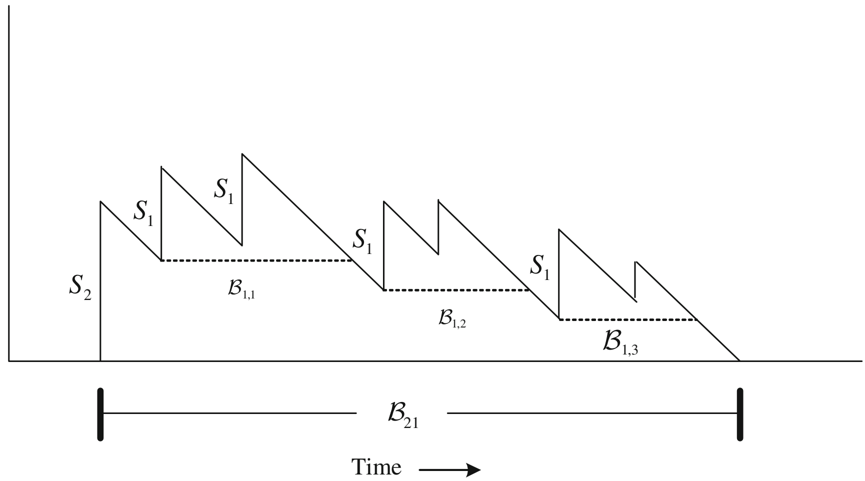


Fig. 3.30 Busy period $\mathcal{B}_{2,1}$. Initial jump is a type 2 service S_2. Each subsequent jump is a type 1 service S_1. $\mathcal{B}_{1,j},\ j = 1, 2, \ldots,$ are M/G/1 type 1 busy periods

Remark 3.33 $E(\mathcal{B}_{2,1})$ is the same as the expected busy period in an M/G/1 queue in which zero-waiting customers receive exceptional service. Thus we can obtain (3.232) immediately as a special case of (3.147).

Let $\boldsymbol{B}_{2,1}(x)$ denote the cdf of $\mathcal{B}_{2,1}$, and $\overline{\boldsymbol{B}}_{2,1}(x) = 1 - \boldsymbol{B}_{2,1}(x)$, $x \geq 0$. Consider a sample path of the virtual wait of type-2 customers $\{W_2(t)\}_{t\geq 0}$ (Fig. 3.31). The sample path illustrates that *type*-2 customers may view the model as a queue with server vacations (see Sect. 3.15). When a type 1 arrives to an empty system, the server vacation is $\mathcal{B}_1$. When a type 2 arrives, the server vacation consists of $N_{S_{21}}\mathcal{B}_1$s. By (3.230), type-2 generated SP jumps are $\underset{dis}{=} \mathcal{B}_{2,1}$.

Integral Equation for $f_2(x)$
We now construct the integral equation

$$f_2(x) = \lambda_1 \overline{\boldsymbol{B}}_1(x)P_0 + \lambda_2 \overline{\boldsymbol{B}}_{2,1}(x)P_0 + \lambda_2 \int_{y=0}^{x} \overline{\boldsymbol{B}}_{2,1}(x-y) f_2(y)dy. \tag{3.233}$$

In (3.233) the left side $f_2(x)$ is the sample-path downcrossing rate of level x (Theorem 1.1 in Chap. 1). On the right side the term $\lambda_1 \overline{\boldsymbol{B}}_1(x)P_0$ is the SP upcrossing rate of x due to type-1 arrivals when the system is empty. A potentially arriving type-2 customer, immediately after the initial type-1 starts service, would wait a type-1 busy period before starting service. The term $\lambda_2 \overline{\boldsymbol{B}}_{2,1}(x)P_0$ is the SP upcrossing rate of x due to type-2 arrivals when the

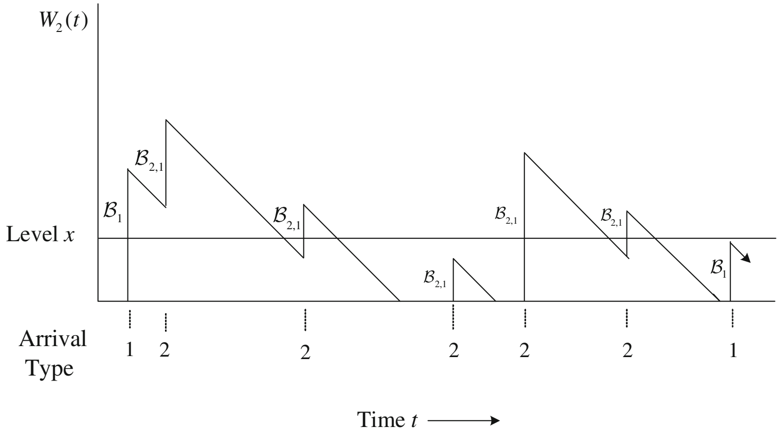


Fig. 3.31 Sample path of virtual wait for *low priority, type 2 arrivals*. High priority type 1's that arrive when the system is empty generate jumps distributed as $\mathcal{B}_1$ busy periods. All type 2 arrivals generate jumps distributed as $\mathcal{B}_{2,1}$ busy periods (see Fig. 3.30). All type 1's that must wait, are counted in the $\mathcal{B}_{2,1}$ jumps

system is empty. A potentially arriving type-2 customer, immediately after the type 2 starts service, would wait a busy period $\mathcal{B}_{2,1}$ before starting service. It is possible that $\mathcal{B}_{2,1}$ consists of the initial type-2 service only. Possibly no type 1s arrive during the initial service time. Generally, $\mathcal{B}_{2,1}$ includes an additional run of $N_{S_{2,1}}$ $\mathcal{B}_1$s (Fig. 3.30). The term $\lambda_2 \int_{y=0}^{x} \overline{\boldsymbol{B}}_{2,1}(x-y) f_2(y) dy$ is the upcrossing rate of x due to type-2 arrivals that must wait a positive time $y \in (0, x)$. A would-be type-2 customer that arrives immediately after such a type-2 arrival, would face an additional wait equal to $\mathcal{B}_{2,1}$ before starting service.

The three terms on the right of (3.233) account for all arrivals to the system. The type 2s are counted in the last two terms; they include all type 2s that wait ≥ 0. The type 1s are counted in all three terms. The type 1s that wait zero are counted in the first term. The type 1s that wait a positive time are counted in all three terms.

Both Types Have the Same P_0

We test for consistency of integral equations (3.233) and (3.226), by checking whether they give the same value of P_0. It is required to show that (3.227) results from (3.233). We integrate both sides of (3.233) with respect to x on $(0, \infty)$. Simplification gives

$$\begin{aligned} 1 - P_0 &= \lambda_1 E(\mathcal{B}_1) P_0 + \lambda_2 E(\mathcal{B}_{21}) P_0 + \lambda_2 E(\mathcal{B}_{21})(1 - P_0) \\ &= \lambda_1 E(\mathcal{B}_1) P_0 + \lambda_2 E(\mathcal{B}_{21}). \end{aligned}$$

Substituting for $E(\mathcal{B}_1)$, $E(\mathcal{B}_{21})$ from (3.231), (3.232) respectively we obtain

$$1 - P_0 = \lambda_1 \frac{E(S_1)}{1 - \lambda_1 E(S_1)} P_0 + \lambda_2 \frac{E(S_2)}{1 - \lambda_1 E(S_1)},$$

or

$$P_0 = 1 - \lambda_1 E(S_1) - \lambda_2 E(S_2) = 1 - \rho_1 - \rho_2,$$

which is identical to (3.227) QED.

3.14.6 Expected Wait of Type-2 Customers

We obtain the expected wait $E(W_{q,2})$ by multiplying integral equation (3.233) by x on both sides and integrating with respect to x on $(0, \infty)$. Some algebra gives

$$\begin{aligned} E(W_{q2}) = {} & \lambda_1 \tfrac{E(\mathcal{B}_1^2)}{2} P_0 + \lambda_2 \tfrac{E(\mathcal{B}_{21}^2)}{2} P_0 \\ & + \lambda_2 \tfrac{E(\mathcal{B}_{21}^2)}{2} (1 - P_0) + \lambda_2 E(\mathcal{B}_{21}) E(W_{q2}) \end{aligned}$$

or

$$E(E(W_{q2})) = \frac{\lambda_1 E(\mathcal{B}_1^2) P_0 + \lambda_2 E(\mathcal{B}_{21}^2)}{2(1 - \lambda_2 E(\mathcal{B}_{21}))}.$$

Substituting from (3.85), (3.227) and (3.232) gives

$$E(W_{q,2}) = \frac{\left(\lambda_1 \frac{E(S_1^2)}{(1-\rho_1)^3} (1 - \rho_1 - \rho_2) + \lambda_2 E(\mathcal{B}_{21}^2)\right) \cdot (1 - \rho_1)}{2\,(1 - \rho_1 - \rho_2)}. \quad (3.234)$$

The term $\lambda_2 E(\mathcal{B}_{2,1}^2)$ in the numerator of (3.234) is

$$\begin{aligned} \lambda_2 E(\mathcal{B}_{2,1}^2) = {} & \lambda_2 E\left(\left(S_2 + \sum_{i=1}^{N_{S_{2,1}}} \mathcal{B}_{1,i}\right)^2\right) \\ = {} & \lambda_2 E(S_2^2) + 2\lambda_2 E\left(S_2 \sum_{i=1}^{N_{S_{21}}} \mathcal{B}_{1,i}\right) + \lambda_2 E\left(\left(\sum_{i=1}^{N_{S_{21}}} \mathcal{B}_{1,i}\right)^2\right). \end{aligned}$$

We condition on $N_{S_{2,1}} = n$, $S_2 = s$ in the last two terms. Then $N_{S_{2,1}}$ is a Poisson random variable with parameter $\lambda_1 s$. We then carry out some algebra, and “uncondition”. This procedure yields

$$\lambda_2 E(\mathcal{B}_{2,1}^2) = \lambda_2 E(S_2^2) + 2\lambda_2 E(S_2^2)\frac{\rho_1}{1-\rho_1} + \lambda_2(\lambda_1 E(S_2)E(\mathcal{B}_1^2) + \lambda_1^2 (E(\mathcal{B}_1))^2 E(S_2^2)).$$

Substituting from (3.85) into the last equation gives

$$\lambda_2 E(\mathcal{B}_{2,1}^2) = \lambda_2 E(S_2^2) + 2\lambda_2 E(S_2^2)\tfrac{\rho_1}{1-\rho_1} + \rho_2\lambda_1 \tfrac{E(S_1^2)}{(1-\rho_1)^3} + \lambda_2 \tfrac{\rho_1^2}{(1-\rho_1)^2} E(S_2^2). \tag{3.235}$$

Substituting the expression in (3.235) for $\lambda_2 E(\mathcal{B}_{2,1}^2)$ in the numerator of (3.234) gives

$$\text{coefficient of } E(S_1^2) = \frac{\lambda_1}{(1-\rho_1)},$$
$$\text{coefficient of } E(S_2^2) = \frac{\lambda_2}{(1-\rho_1)}.$$

Hence

$$E(W_{q2}) = \frac{\frac{\lambda_1}{(1-\rho_1)} E(S_1^2) + \frac{\lambda_2}{(1-\rho_1)} E(S_2^2)}{2\,(1-\rho_1-\rho_2)} = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1-\rho_1)\,(1-\rho_1-\rho_2)}, \tag{3.236}$$

which agrees with the result in the literature (e.g., p. 545 in [125]).

Remark 3.34 We have used LC to derive $E(W_{q,1})$ from the integral equation for $f_1(x)/$, and $E(W_{q,2})$ from the integral equation for $f_2(x)$. The importance of this approach is that we essentially have an analytic solution for the pdfs and cdfs of wait of both priority classes. The LC analysis is in the time domain without use of transforms. Integral equations (3.226), (3.233) can be solved analytically in some cases; or else numerically. The LC analysis highlights conceptual properties of the priority queue that are in common with queues having: (1) service time depending on wait, (2) multiple Poisson inputs, (3) server vacations. In addition, the exercise of constructing the sample paths of wait for the different priority classes, leads to an intuitive understanding of the model dynamics.

3.14.7 Exponential Service

We now solve for $\{P_0,\ f_1(x)\}_{x>0}$ in an M/M/1 queue with two priority types. Here $S_i \underset{dis}{=} Exp_{\mu_i}, i = 1, 2$. Substituting $\overline{B}_i(x) = e^{-\mu_i x}$ into (3.226) gives an integral equation for $f_1(x)$,

$$\begin{aligned} f_1(x) = \ & \lambda_1 e^{-\mu_1 x} P_0 + \lambda_2 e^{-\mu_2 x} P_0 + \lambda_1 \int_{y=0}^{x} e^{-\mu_1(x-y)} f_1(y) dy \\ & + \lambda_2 (1 - P_0) e^{-\mu_2 x}. \end{aligned} \tag{3.237}$$

We apply differential operator $\langle D + \mu_1 \rangle \langle D + \mu_2 \rangle$ to both sides of (3.237), obtaining the second order differential equation

$$\langle D + \mu_2 \rangle \langle D + \mu_1 - \lambda \rangle f_1(x) = 0,$$

with solution

$$f_1(x) = a e^{-(\mu_1 - \lambda_1)x} + b e^{-\mu_2 x}, x \geq 0, \tag{3.238}$$

where constants a, b are to be determined.

Letting $x \downarrow 0$ in (3.237) and (3.238) yields

$$a + b = \lambda_1 P_0 + \lambda_2. \tag{3.239}$$

Taking $\frac{d}{dx}$ on both sides of (3.237) and letting $x \downarrow 0$ gives

$$f_1'(0) = -\lambda_1 \mu_1 P_0 + \lambda_1^2 P_0 + \lambda_1 \lambda_2 - \lambda_2 \mu_2. \tag{3.240}$$

Taking $\frac{d}{dx}$ in (3.238), letting $x \downarrow 0$, and equating to (3.240) gives

$$-(\mu_1 - \lambda_1) a - \mu_2 b = -\lambda_1 \mu_1 P_0 + \lambda_1^2 P_0 + \lambda_1 \lambda_2 - \lambda_2 \mu_2. \tag{3.241}$$

We use (3.238) and the condition $P_0 + \int_{x=0}^{\infty} f_1(x) dx = 1$ to obtain

$$P_0 + \frac{a}{\mu_1 - \lambda_1} + \frac{b}{\mu_2} = 1. \tag{3.242}$$

We now solve the system of three Eqs. (3.239), (3.241), (3.242) for P_0, a, b to obtain

$$P_0 = \frac{(\mu_2\mu_1 - \mu_2\lambda_1 - \mu_1\lambda_2)}{\mu_2\mu_1}, \tag{3.243}$$

$$a = \frac{\lambda_1(\mu_2\mu_1^2 + 2\mu_2\mu_1\lambda_1 + \mu_2^2\mu_1 - \mu_2\lambda_1^2 - \mu_2^2\lambda_1 + \mu_1^2\lambda_2 - \mu_1\lambda_2\lambda_1)}{(-\mu_1 + \lambda_1 + \mu_2)\mu_2\mu_1}, \tag{3.244}$$

$$b = \frac{\lambda_2(\mu_2 - \mu_1)}{(-\mu_1 + \lambda_1 + \mu_2)}. \tag{3.245}$$

Check on the Values of P_0, a, b

We conduct a mild check (indicated by ✓) on the values of P_0, a, b. Set λ_2 = 0. The model reverts to a standard $M_{\lambda_1}/M_{\mu_1}/1$ queue in which $f(x)$ and P_0 are given in (3.112) and (3.113), respectively.

Substituting $\lambda_2 = 0$ in (3.243), (3.244) and (3.245) yields: $P_0 = 1 - \lambda_1/\mu_1$; $a = \lambda_1\,(1 - \lambda_1/\mu_1)$; $b = 0$. ✓

3.15 M/G/1 with Server Vacations

There are many M/G/1 server-vacation models. During a server vacation the server is not available to serve customers. For example, vacations may start after each service completion, or when the server becomes idle, or both. (See, e.g., Problems 9.2 and 9.6, pp. 420–422 in [143], and see also [39] in which consecutive vacations are connected by a Markov chain.)

Here we apply LC to a basic M/G/1 server-vacation model. Let the arrival rate be λ and service time be S having cdf $B(x), x > 0$. Assume that after each service completion the server goes on vacation for a time U having cdf $V(x), x > 0$. During U the server may be doing required work after each service. For example, a doctor updates a record after seeing each patient, a bank teller does required paper work after serving each customer, an auto service manager fills out forms after receiving a car for service. Consider the virtual wait process $\{W(t)\}_{t\geq 0}$ (Fig. 3.32).

Denote the complementary cdf of $S + U$ by $\overline{B * V}(x)$. An integral equation for the steady-state pdf of wait $\{P_0, f(x)\}_{x\geq 0}$ is

$$f(x) = \lambda P_0 \overline{B * V}(x) + \lambda \int_{y=0}^{x} \overline{B * V}(x - y) f(y)dy, x \geq 0. \tag{3.246}$$

In (3.246) the left side $f(x)$ is the SP downcrossing rate of level x. On the right side $\lambda P_0 \overline{B * V}(x)$ is the SP upcrossing rate of level x, starting from

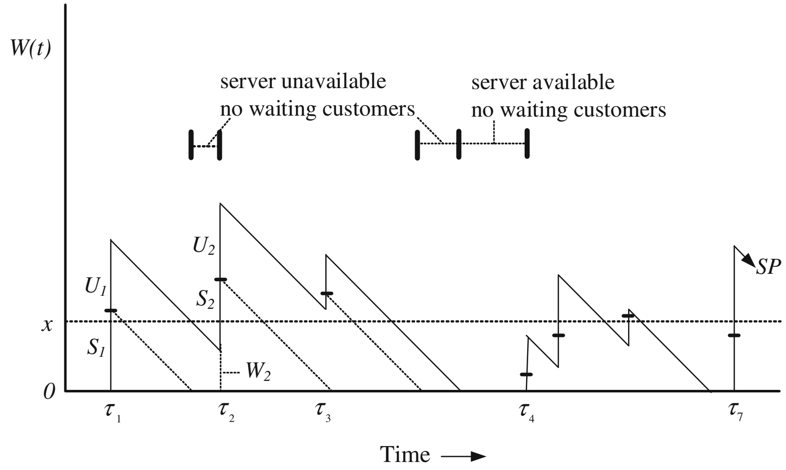


Fig. 3.32 Sample path of $\{W(t)\}_{t\geq 0}$ in M/G/1 queue with a server vacation after each service completion

level 0. The term $\lambda \int_{y=0}^{x} \overline{B * V}(x-y)f(y)dy$ is the SP upcrossing rate of level x, starting from levels in the state-space interval $(0, x)$.

Comparing (3.246) and (3.34) indicates that the server-vacation and standard M/G/1 models are equivalent with regard to the integral equation for the pdf of queue wait; only the "service time" cdfs differ.

3.15.1 Probability of Zero Wait

Since the queue behaves like $M_\lambda/G/1$ with common service time $S+U$ with respect to the customer wait until service, then

$$P_0 = 1 - \lambda E(S+U) \qquad (3.247)$$

provided $\lambda E(S+U) < 1$.

3.15.2 Expected Busy and Idle Period

Define the idle period $\mathcal{I}$ as the time interval when the server is available to start service and no customers are waiting. Then $E(I) = 1/\lambda$. Let $\mathcal{B}_s$:= time that the server is busy serving customers, $\mathcal{B}_u$:= time that server is "on vacation", during a "busy period" $\mathcal{B}$, where $\mathcal{B} = \mathcal{B}_s + \mathcal{B}_u$. Then $\mathcal{B}$ is distributed as a regular busy period in a standard M$_\lambda$/G/1 queue with service time $S + U$. Applying (3.247)

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = \frac{\lambda E(S+U)}{\lambda(1 - \lambda E(S+U))}. \tag{3.248}$$

Given the server is "busy", the pairs $\{S_i, U_i\}, i = 1, 2, \ldots$, form an alternating renewal process (Fig. 3.32). During a "busy" period, the proportion of time the server is busy serving customers $= \frac{E(S)}{E(S)+E(U)}$; "on vacation" = $\frac{E(U)}{E(S)+E(U)}$. Thus

$$E(\mathcal{B}_s) = \frac{E(S)}{E(S) + E(U)} \cdot E(\mathcal{B}),\ E(\mathcal{B}_u) = \frac{E(U)}{E(S) + E(U)} \cdot E(\mathcal{B});$$

from (3.248)

$$E(\mathcal{B}_s) = \frac{E(S)}{1 - \lambda E(S+U)},\ E(\mathcal{B}_u) = \frac{E(U)}{1 - \lambda E(S+U)}.$$

3.15.3 Number in System

Let d_n denote the probability of n customers in the system *just after the server returns from vacation*. Then (see Eq. (3.76) in Sect. 3.4.8)

$$d_n = \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} f(x) dx.$$

Let a_n denote the probability that an arrival "sees" n customers in the system. Then $a_n = d_n = P_n$ due to Poisson arrivals, P_n is the long-run proportion of time there are n customers in the system.

3.15.4 M/M/1 with Server Vacations $\underset{dis}{=}$ Exp$_\nu$

Let $\overline{V}(x) = e^{-\nu x}$, $\overline{B}(x) = e^{-\mu x}$, $x \geq 0$. Assume $\nu \neq \mu > 0$. Then

$$\overline{B * V}(x) = P(S + V > x) = \frac{\mu e^{-\nu x} - \nu e^{-\mu x}}{\mu - \nu}, x \geq 0,$$

and (3.246) reduces to

$$f(x) = \lambda P_0 \frac{\mu e^{-\nu x} - \nu e^{-\mu x}}{\mu - \nu} + \lambda \frac{1}{\mu - \nu} \int_{y=0}^{x} \left(\mu e^{-\nu(x-y)} - \nu e^{-\mu(x-y)}\right) f(y)dy, \ x \geq 0. \tag{3.249}$$

In (3.249), applying differential operator $\langle D+\nu\rangle\langle D+\mu\rangle$ to both sides results in the differential equation

$$f''(x) + (\nu + \mu - \lambda) f'(x) + (\nu\mu - \lambda\mu - \lambda\nu) f(x) = 0,$$

with solution

$$f(x) = c_1 e^{R_1 x} + c_2 e^{R_2 x}, \ x \geq 0,$$

where roots R_1, R_2 are the (negative) roots of the characteristic equation

$$z^2 + (\nu + \mu - \lambda)z + (\nu\mu - \lambda\mu - \lambda\nu) = 0.$$

Applying the initial conditions $f(0) = \lambda P_0$, $f'(0) = \lambda^2 P_0$, and the normalizing condition $P_0 + \int_{y=0}^{\infty} f(x)dx = 1$ yields

$$c_1 = \lambda P_0 \frac{\lambda - R_2}{R_1 - R_2}, \qquad c_2 = -\lambda P_0 \frac{-R_1 + \lambda}{R_1 - R_2}, \qquad P_0 = \frac{c_1 R_2 + c_2 R_1 + R_1 R_2}{R_1 R_2}.$$

Busy Period

The expected values of $\mathcal{B}$, $\mathcal{B}_s$, $\mathcal{B}_u$ are

$$E(\mathcal{B}) = \frac{\frac{1}{\mu} + \frac{1}{\nu}}{1 - \lambda\left(\frac{1}{\mu} + \frac{1}{\nu}\right)}, \qquad E(\mathcal{B}_s) = \frac{\frac{1}{\mu}}{\frac{1}{\mu} + \frac{1}{\nu}} E(\mathcal{B}), \qquad E(\mathcal{B}_u) = \frac{\frac{1}{\nu}}{\frac{1}{\mu} + \frac{1}{\nu}} E(\mathcal{B}).$$

Number in System

The probability that the server finds n in the system just after a vacation is for $n = 1, 2, \ldots,$

$$\begin{aligned} d_n &= \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} \left(c_1 e^{R_1 x} + c_2 e^{R_2 x}\right) dx \\ &= \frac{1}{\lambda}\left(\left(\frac{\lambda}{\lambda - R_1}\right)^n c_1 + \left(\frac{\lambda}{\lambda - R_2}\right)^n c_2\right), \end{aligned}$$

where $R_i, c_i, i = 1, 2$ are given above. The probability that an arrival "sees" n customers in the system is $a_n = d_n = P_n$.

3.16 M/G/1 with Bounded Workload

We look at three M/G/1 variants with a finite barrier $K > 0$ on the virtual wait (workload) process. These and related models (e.g., risk models with a dividend barrier in actuarial science) have been discussed widely in the literature (e.g., [78]; Example 5.5.2, p. 213 and Exercise 9.9, p. 423 in [143], and also in [25]; M/M/c queues with bounded wait in Example 1, p. 44 in [52], and also in [54, 79, 100]; and others). They are also useful in the proof of Proposition 9.1 in Sect. 9.4, Chap. 9 on level crossing estimation. As $K \to \infty$, variants 1–3 tend to a standard M/G/1 queue with infinite waiting buffer, under mild conditions. We illustrate this property with M/M/1 in Sects. 3.16.2, 3.16.4, and 3.16.6. In all three variants, we denote the arrival rate by λ; the requested full service time for each arrival, by S having cdf $B(x)$, $x > 0$, $\overline{B}(x) = 1 - B(x)$, $x \geq 0$; and the virtual wait (workload) process as $\{W_K(t)\}_{t \geq 0}$.

3.16.1 Variant 1

All customers join the system, and all waiting times (before start of service) are in $[0, K)$. Each arrival gets either full service S, or truncated service if S causes $\{W_K(t)\}_{t \geq 0}$ to exceed K, i.e., customers in service must *renege if and when* their total system time reaches K. We define the service time S_K due the level-K barrier, in terms of S as follows. If a customer must wait $y \geq 0$ then $S_K = \min(S, K - y)$. Thus for all customers, *wait* + *service time* $\leq K$. Consider a sample path of $\{W_K(t)\}_{t \geq 0}$ (Fig. 3.33). Let the mixed pdf

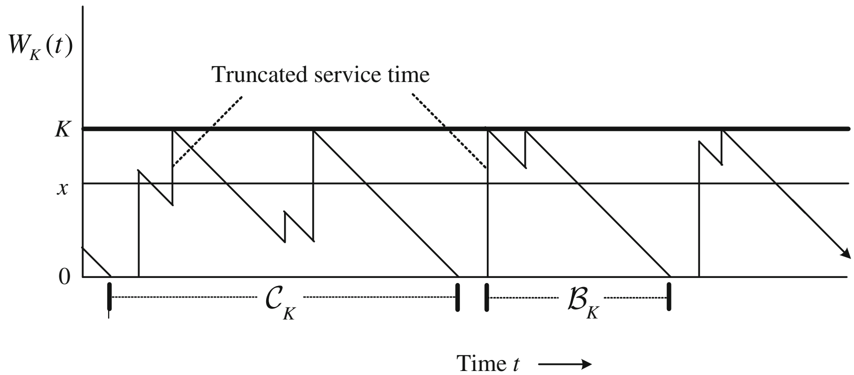


Fig. 3.33 Variant 1. Sample path of $\{W_K(t)\}_{t \geq 0}$ in M/G/1 with bounded workload. $\mathcal{C}_K$:= busy cycle, $\mathcal{B}_K$:= busy period

of wait be $\left\{P_{K,0}, f_K(x)\right\}_{x>0}$. Rate balance across level x gives immediately Eq. (3.250) for $f_K(x)$, where the left and right sides are the SP down- and upcrossing rates, respectively:

$$f_K(x) = \lambda P_{K,0}\overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f_K(y)dy, 0 < x < K, \quad (3.250)$$

$$P_{K,0} + \int_{y=0}^{K} f_{K,0}(x)dx = 1. \quad (3.251)$$

If $K \in (0, \infty)$ then $\left\{P_{K,0}, f_K(x)\right\}_{x>0}$ exists for all values of $\lambda > 0$ (see Sect. 2.1 in [25]). Also, [25] gives the pdf of S_K and shows the important result that $E(S_K) = \left(1 - P_{K,0}\right)/\lambda$, equivalently $P_{K,0} = 1 - \lambda E(S_K)$. (Interestingly, this is similar to $P_0 = 1 - \lambda E(S)$ in the standard *no-barrier* M/G/1 queue in steady state.) If there exists $M > 0$ such that $P_{K,0} > 0$ for all $K > M$, and we assume $\lambda E(S) < 1$ then $\left\{P_{K,0}, f_K(x)\right\}_{x>0} \to \{P_0, f(x)\}_{x>0}$ in the standard no-barrier M/G/1, since Eqs. (3.250), (3.251) would converge to Eqs. (3.34)–(3.36).

3.16.2 Variant 1: M/M/1 Model

In the $M_\lambda/M_\mu/1$ model $\overline{B}(x) = e^{-\mu x}$; the solution of (3.250) and (3.251) is

$$\left.\begin{aligned} f_K(x) &= \lambda P_{K,0} e^{-(\mu-\lambda)x}, 0 < x < K, \\ P_{K,0} &= \frac{\mu - \lambda}{\mu + e^{-(\mu-\lambda)K}}. \end{aligned}\right\} \quad (3.252)$$

If we assume $\lambda < \mu$ so that the no-barrier M/M/1 is stable, and let $K \to \infty$, then $P_{K,0} \to 1 - \lambda/\mu$ and the domain $(0, K)$ of $f_K(\cdot)$, tends to $(0, \infty)$. This results in the solution for the standard no-barrier $M_\lambda/M_\mu/1$ queue (see formulas (3.112) and (3.113)).

3.16.3 Variant 2

Upon arrival customers *balk and are cleared* if their system times would exceed K. We assume that the workload $W_K(t^-)$ and the service time S of a would-be time-t arrival are known to a "system manager" by some means. A time-t arrival joins the system only if $W_K(t^-) + S < K$. We define the service time S_K due the level-K barrier in terms of S as follows. If a customer must wait $y \geq 0$ then

$$S_K = \begin{cases} S \text{ if } y + S \le K, \\ 0 \text{ if } y + S > K. \end{cases}$$

Customers that are allowed to join receive full service S, and depart upon completing service. Consider a sample path of $\{W_K(t)\}_{t\ge0}$ (Fig. 3.34). We obtain via LC the integral equation for $f_K(x)$:

$$\begin{aligned} f_K(x) &= \lambda P_{K,0}\left(\overline{B}(x) - \overline{B}(K)\right) \\ &\quad + \lambda \int_{y=0}^{x} \left(\overline{B}(x-y) - \overline{B}(K-y)\right) f_K(y)dy, \; 0 < x < K, \end{aligned} \tag{3.253}$$

with normalizing condition $P_{K,0} + \int_{y=0}^{K} f_K(x)dx = 1$. In (3.253), the term $\overline{B}(x) - \overline{B}(K) = P(x < S < K)$ and the term $\overline{B}(x-y) - \overline{B}(K-y)$ = $P(x - y < S < K - y)$. Using the technique in [25] for Variant 1, we can also find in Variant 2, the pdf of S_K and show that $E(S_K) = (1 - P_{K,0})/\lambda$.

3.16.4 Variant 2: M/M/1 Model

In the $M_\lambda/M_\mu/1$ queue with $\overline{B}(x) = e^{-\mu x}$, we obtain immediately the solution of (3.253) for $\left\{P_{K,0},\, f_K(x)\right\}_{x\in(0,K)}$ as a special case of the M/M/c queue with bounded system time. (In Example 1, p. 44 in [52], we set *number of servers* = 1.) We get

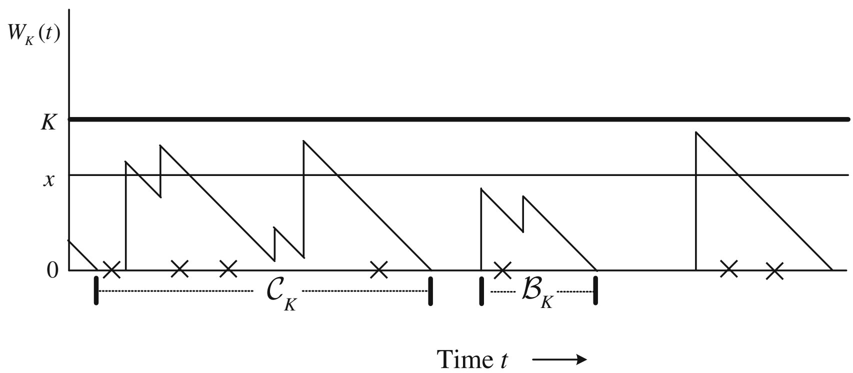


Fig. 3.34 Variant 2. Sample path of $\{W_K(t)\}_{t\ge0}$ in M/G/1 with bounded workload. '$\times$' indicates arrivals who balk because $wait + S > K$. $\mathcal{C}_K$:= busy cycle, $\mathcal{B}_K$:= busy period

$$\left.\begin{aligned} f_K(x) &= \lambda e^{\rho\beta} P_{K,0} e^{\mu(\rho-1)x}(1-\beta e^{\mu x})e^{-\mu\beta e^{\mu x}}, 0 < x < K, \\ P_{K,0} &= \frac{1}{1+\lambda e^{\rho\beta}\int_{x=0}^{K} e^{\mu(\rho-1)x}(1-\beta e^{\mu x})e^{-\mu\beta e^{\mu x}}dx}, \end{aligned}\right\} \quad (3.254)$$

where $\rho = \lambda/\mu$, $\beta = e^{-\mu K}$. The solution in (3.254) checks with the single-server Markovian result obtained in [78], and is more complex than the solution (3.252) for variant 1.

If $K \to \infty$ then $\beta \downarrow 0$. Additionally, if $\lambda < \mu$ then (3.254) becomes

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0, \qquad P_0 = 1 - \frac{\lambda}{\mu},$$

as in the standard no-barrier M/M/1 queue.

3.16.5 Variant 3

All arrivals that "see" a *wait* $< K$ join the system and receive full service S. Some of these service tines will cause jumps that upcross level K (Fig. 3.35). (In variant 3 we call level K a *threshold* rather than a barrier, because sample-path conditions switch at level K.) Arrivals that "see" a wait $> K$, are blocked from joining, and are cleared. (Effectively, they balk upon arrival. With respect to the arrival-point waiting time Variant 3 is identical to M/G/1 with reneging/balking and having a *staying function*

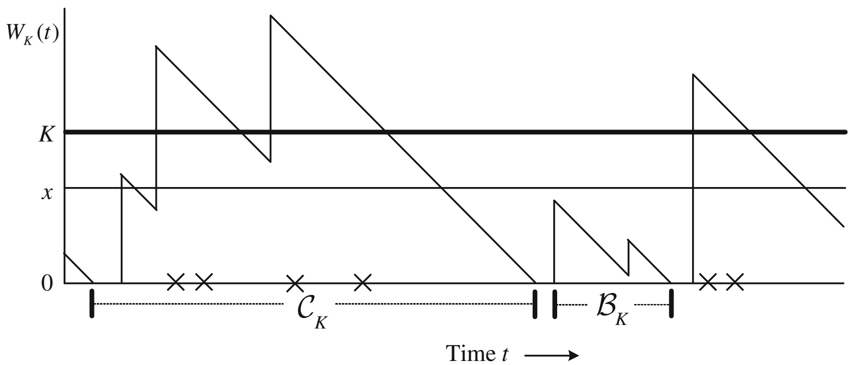


Fig. 3.35 Variant 3. Sample path of $\{W_K(t)\}_{t\geq 0}$ in M/G/1 with threshold at level K. '$\times$' indicates arrivals who balk because $W_K(\cdot) > K$ upon their arrival. $\mathcal{C}_K$:= busy cycle, $\mathcal{B}_K$:= busy period

$\overline{R}(y) = 1 \cdot \boldsymbol{I}_{[0,k)}\ (W_\iota) + 0.\boldsymbol{I}_{[k,\infty)}\ (W_\iota)$, where W_ι := arrival-point wait ... see Fig. 3.35 and Sect. 3.13, which analyzes the renege/balk M/G/1 queue. We define the service time S_K due to the level-K threshold, in terms of S as follows. If a customer must wait y then

$$S_K = \begin{cases} S \text{ if } y \in [0, K)\,, \\ 0 \text{ if } y \in [K, \infty)\,, \end{cases}$$

which may be written as $S_K = S \cdot \boldsymbol{I}_{[0,k)}\ (W_\iota) + 0.\boldsymbol{I}_{[k,\infty)}\ (W_\iota)$, where $\boldsymbol{I}_{\boldsymbol{A}}(\cdot)$ is the characteristic function of set $\boldsymbol{A}$.

We denote the mixed pdf of wait as $\left\{P_{K,0},\ f_{K,i}(x)\right\}_{i=0,1}$ where the domain of $f_{K,0}(x)$ is $(0, K)$ and the domain of $f_{K,1}(x)$ is $[K, \infty)$. Using LC we can write integral equations for $f_{K,i}(x)$, $i = 1, 2$, by inspection of Fig. 3.35, as follows.

$$\left.\begin{aligned} f_{K,0}(x) = \lambda P_{K,0}\overline{B}(x) + \lambda \textstyle\int_{y=0}^{x} \overline{B}(x-y) f_{K,0}(y)dy,\ x \in (0, K)\,, \\ f_{K,1}(x) = \lambda P_{K,0}\overline{B}(x) + \lambda \textstyle\int_{y=0}^{K} \overline{B}(x-y) f_{K,0}(y)dy,\ x \in [K, \infty)\,, \\ P_{K,0} + \textstyle\int_{y=0}^{K} f_{K,0}(x)dx + \int_{x=K}^{\infty} f_{K,1}(x)dx = 1. \end{aligned}\right\} \tag{3.255}$$

We infer from Fig. 3.35 and Theorem 1.1, the continuity condition at K

$$f_{K,1}(K^+) = f_{K,0}(K^-), \tag{3.256}$$

noting $\lim_{t\to\infty} D_t(x)/t = \lim_{t\to\infty} D_t(x^-)/t$, and there are no SP tangents at level K. (Contrast this property with that at level D in M/D/1 where there is a *discontinuity*; see Proposition 3.9 Part (2) **in** Sect. 3.10.1.)

Expected Sojourn Above Level K

Let γ_K := *excess of a jump* over level K, a_K := *sojourn* above level K. Then $a_K = \gamma_K$, and $E(a_K) = E(\gamma_K)$. Let $F_{\gamma_K}(z) := P(\gamma_K \le z)$, $z > 0$. Two different expressions for $\lim_{t\to\infty} \mathcal{U}_t(K+z)/t$ are

$$\left(1 - F_{\gamma_K}(z)\right) f_{K,0}(K^-)$$

$$\text{and} \qquad \lambda P_{K,0}\overline{B}(K+z) + \lambda \int_{y=0}^{K} \overline{B}(K+z-y) f_{K,0}(y)dy.$$

In the first expression $f_{K,0}(K^-)$ is the upcrossing rate (also the downcrossing rate) of level K, and $1 - F_{\gamma_K}(z)$ is the upcrossing rate of level $K+z$ given the SP upcrosses level K. The second term is the upcrossing rate of level $K+z$ due to upward jumps that start in $[0, K)$. Thus

$$
\begin{aligned}
1 - F_{\gamma_K}(z) = & \ \frac{\lambda P_{K,0}\overline{B}(K+z) + \lambda \int_{y=0}^{K} \overline{B}(K+z-y) f_{K,0}(y)dy}{f_K, 0(K^-)} \\
= & \ \frac{\lambda P_{K,0}\overline{B}(K+z) + \lambda \int_{y=0}^{K} \overline{B}(K+z-y) f_{K,0}(y)dy}{\lambda P_{K,0}\overline{B}(K) + \lambda \int_{y=0}^{K} \overline{B}(K-y) f_{K,0}(y)dy}
\end{aligned}
$$

and

$$
\begin{aligned}
E(a_K) &= E(\gamma_K) = \int_{z=0}^{\infty} \left(1 - F_{\gamma_K}(z)\right) dz \\
&= \int_{z=0}^{\infty} \left[\frac{\lambda P_{K,0}\overline{B}(K+z)+\lambda \int_{y=0}^{K} \overline{B}(K+z-y) f_{K,0}(y)dy}{\lambda P_{K,0}\overline{B}(K)+\lambda \int_{y=0}^{K} \overline{B}(K-y) f_{K,0}(y)dy} \right] dz.
\end{aligned} \tag{3.257}
$$

Using the technique in [25] for Variant 1, we can also find in Variant 3, the pdf of S_K and show that $E(S_K) = (1 - P_{K,0})/\lambda$.

3.16.6 Variant 3: M/M/1 Model

Setting $\overline{B}(x) = e^{-\mu x}$ in (3.255), and solving by converting to differential equations, gives

$$
\left.
\begin{aligned}
f_{K,0}(x) &= \lambda P_{k,0} e^{-(\mu-\lambda)x}, \ x \in (0, K), \\
f_{K,1}(x) &= \lambda P_{k,0} e^{-(\mu x - \lambda K)}, \ x \in [K, \infty), \\
P_{K,0} &= \frac{1}{1+\frac{\lambda}{\mu-\lambda}(1-e^{-(\mu-\lambda)K})+\frac{\lambda}{\mu}e^{-(\mu-\lambda)K}}.
\end{aligned}
\right\} \tag{3.258}
$$

In (3.258) if $x > (\lambda K)/\mu$ then $\mu x - \lambda K > 0$ and $\int_{x=K}^{\infty} f_{K,1}(x)dx$ is finite. If additionally $\lambda < \mu$ then as $K \to \infty$ the denominator of $P_{K,0} \to \frac{1}{1+\lambda/(\mu-\lambda)}$ = $1 - \lambda/\mu$, which is P_0 in the no-threshold M/M/1 queue. Also $f_{K,0}(x) \to \lambda P_0 e^{-(\mu-\lambda)x}$, $x \in (0, \infty)$ which is $f(x)$, $x > 0$ in the no-threshold M/M/1 queue.

From (3.257), $E(a_K) = \int_{z=0}^{\infty} e^{-\mu z} dz = 1/\mu$.

3.17 Generalized Beneš Series for PDF of Wait

In this Section we use LC to generalize the Beneš series for the pdf of wait in M/G/1 (see formula (3.71) in Sect. 3.4.5). We use LC, the busy-period structure (Fig. 3.6 in Sect. 3.4.12), the multiplicative structure (Fig. 3.36), and the ***renewal reward theorem*** (see references following Eq. (3.79)) to develop a *series* for the pdf of wait (W_q). Combining LC and the renewal reward theorem facilitates creating more general series for the pdf of W_q in MG/1 variants as well. We illustrate the more generalized series in an M/G/1 model where zero-wait arrivals receive exceptional service (see Sect. 3.6.1).

3.17.1 Model Description

The arrival rate is λ. Zero-wait arrivals (initiators of busy periods) receive service time S_0. Positive-wait arrivals receive service time S_1 ($\neq_{dis} S_0$). We denote: the cdf of S_i as $B_i(x)$, $x > 0$, $\overline{B}_i(x) = 1 - B_i(x)$, $x \geq 0$, $i = 0, 1$; the steady-state pdf of wait as $\{P_0, f(x)\}_{x>0}$; the limiting excess of S_i as γ_{S_i};; *the pdf* of γ_{S_i} by $g_i(x)$, $x \in (0, \infty) \cap$ (*domain of* S_i). It is well known that $g_i(x) = (1/E(S_i))\,\overline{B}_i(x)$, $i = 0, 1$ (see Example 7.24, p. 453 in [125] ; formula (6.2), p. 193 in [99]). Also $\rho_i := \lambda E(S_i)$, $i = 0, 1$.

Examining a busy period of the virtual wait process $\{W(t)\}_{t\geq 0}$ (Fig. 3.36) and applying LC rate balance across level x (>0), yields Eq. (3.143) of Sect. 3.6.1 (repeated here for handy reference)

$$f(x) = \lambda P_0 \overline{B}_0(x) + \lambda \int_{y=0}^{x} \overline{B}_1(x-y) f(y) dy, x > 0. \tag{3.259}$$

Integrating both sides of (3.259) with respect to $x \in (0, \infty)$ and simplifying leads to formula (3.144) for P_0, whose form implies $P_0 \in (0, 1)$ iff $\rho_1 < 1$.

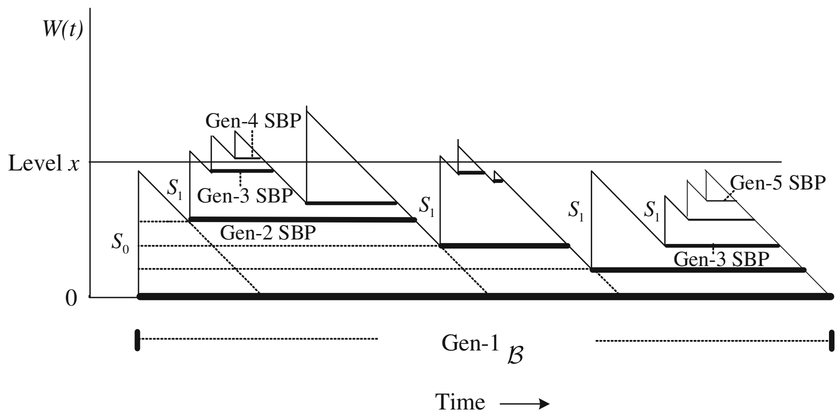


Fig. 3.36 Multiplicative structure of $\mathcal{B}$ of $\{W(t)\}_{t\geq 0}$ for Beneš series analysis. Each arrival generates the initial jump of a $\mathcal{B}$ or sub-busy period (SBP). Initial jumps of all busy/sub-busy periods account for all arrivals (Gen := generation)

3.17.2 Applying the Renewal Reward Theorem

Consider the gen-1 (abbreviation for generation-1) busy period $\mathcal{B}$ in Fig. 3.36. Fix level $x > 0$. Let: $E(\mathcal{U}_{\text{gen-}k}(x)) := E$(number of upcrossings of x by gen-k initiated jumps in $\mathcal{B}$); $E(\mathcal{U}_{\text{gen-}k,t}(x)) := E$(number of upcrossings of x by gen-k initiated jumps *during* $(0,t)$). Then

$$E(\mathcal{U}_{\text{gen-1}}(x)) = 1 \times \overline{B}_0(x);\; E(\mathcal{C}) = 1/(\lambda P_0)$$

(for $E(\mathcal{C})$ see formula (3.81) in Sect. 3.4.10). All jumps during $\mathcal{C}$ (busy cycle) occur during its embedded $\mathcal{B}$. By the *renewal reward theorem*, the *long-run* upcrossing rate of x due to gen-1 busy period initiated jumps is

$$\begin{aligned}\frac{\lim_{t\to\infty}\mathcal{U}_{\text{gen-1},t}(x)}{t} &= \frac{E(\mathcal{U}_{\text{gen-1}}(x))}{E(\mathcal{C})}\\ = \frac{\overline{B}_0(x)}{1/(\lambda P_0)} &= \lambda P_0\overline{B}_0(x) = P_0\rho_0\overline{g}_0(x).\end{aligned} \tag{3.260}$$

A similar analysis of gen-2 sub-busy period initiated upcrossings of x gives

$$E(\mathcal{U}_{\text{gen-2}}(x)) = \lambda E(S_0)\int_{y=0}^{x}\overline{B}_1(x-y)g_0(y)dy,$$

since E(*number of gen-2 jumps during C – same as number in* $\mathcal{B}$) = $\lambda E(S_0)$. Thus

$$\begin{aligned}E(\mathcal{U}_{\text{gen-2}}(x)) &= (1/\mu_1)\lambda E(S_0)\int_{y=0}^{x}\mu_1\overline{B}_1(x-y)g_0(y)dy\\ &= (1/\mu_1)\rho_0\left(g_{1(1)} * g_0\right)(x).\end{aligned}$$

By the *renewal reward theorem*

$$\begin{aligned}\frac{\lim_{t\to\infty}\mathcal{U}_{\text{gen-2},t}(x)}{t} &= \frac{E(\mathcal{U}_{\text{gen-2}}(x))}{E(\mathcal{C})}\\ = \frac{(1/\mu_1)\rho_0\left(g_{1(1)}*g_0\right)(x)}{1/(\lambda P_0)} &= P_0\rho_0\rho_1\left(g_{1(1)} * g_0\right)(x),\end{aligned} \tag{3.261}$$

where $\left(g_{1(1)} * g_0\right)(x) = \int_{y=0}^{x} g_{1(1)}(x-y)g_0(y)dy$.

Similarly, for gen-3 sub-busy period initiated upcrossings of x,

$$E(\mathcal{U}_{\text{gen-3}}(x)) = (\lambda E(S_1))(\lambda E(S_0)) \int_{y=0}^{x} \overline{B}_1(x-y)\left(g_{1(1)} * g_0\right)(y)dy$$
$$= (1/\mu_1)(\lambda E(S_1))(\lambda E(S_0)) \int_{y=0}^{x} \mu_1 \overline{B}_1(x-y)\left(g_{1(1)} * g_0\right)(y)dy$$
$$= (1/\mu_1)\rho_0\rho_1\left(g_{1(2)} * g_0\right)(x);$$

the factor $\lambda E(S_1)$ occurs because each gen-2 sub-busy period initiated jump is $\underset{dis}{=} S_1$—the initial service time of a gen-3 sub-busy period. By the renewal reward theorem

$$\frac{\lim_{t\to\infty} \mathcal{U}_{\text{gen-3},t}(x)}{t} = \frac{E(\mathcal{U}_{\text{gen-3}}(x))}{E(\mathcal{C})}$$
$$= \frac{(1/\mu_1)\rho_0\rho_1\left(g_{1(2)} * g_0\right)(x)}{1/(\lambda P_0)} = P_0\rho_0\rho_1^2\left(g_{1(2)} * g_0\right)(x), x > 0.$$

Similar reasoning for gen-k sub-busy period initiated upcrossings of x, yields

$$\begin{aligned} \frac{\lim_{t\to\infty} \mathcal{U}_{\text{gen-}k,t}(x)}{t} &= \frac{E(\mathcal{U}_{\text{gen-}k}(x))}{E(\mathcal{C})} \\ &= P_0\rho_0\rho_1^{k-1}\left(g_{1(k-1)} * g_0\right)(x), k = 1, 2, \ldots, \end{aligned} \tag{3.262}$$

where $g_{1(k-1)}(\cdot)$ is the $(k-1)$-*fold self-convolution of* $g_1(\cdot)$, and $g_{1(0)} \equiv 1$.

The principle of rate balance across level x gives

$$\begin{aligned} \frac{\lim_{t\to\infty} \mathcal{D}_t(x)}{t} &= \sum_{k=1}^{\infty} \frac{\lim_{t\to\infty} \mathcal{U}_{\text{gen-}k,t}(x)}{t}, x > 0, \\ f(x) &= P_0\rho_0 \sum_{k=1}^{\infty} \rho_1^{k-1}\left(g_{1(k-1)} * g_0\right)(x), x > 0, \end{aligned} \tag{3.263}$$

upon applying formula (3.262). In (3.263) the right side is the total upcrossing rate of level x; term 1 is the upcrossing rate of x due to gen-1 busy period initiated jumps, and term k is the upcrossing rate of x due to gen-$(k-1)$ sub-busy period initiated jumps, $k = 2, 3, \ldots$.

3.17.3 LC Equation for $\{P_0, f(x)\}_{x\geq 0}$ via a Series

In (3.263) term k is the SP upcrossing rate of level x due to the gen-k busy/sub-busy period initiated jumps, where $\left(g_{1(0)} * g_0\right)(x) \equiv g_0(x)$. From Fig. 3.36

every arrival is the initiator of some gen-1 busy period or some gen-k sub-busy period, $k = 2, 3, \ldots$ Hence, the initial jumps of all the gen-k busy/sub-busy periods, $k = 1, 2, \ldots$, *account for all arrivals to the system*. In (3.263) the left and right sides are the SP down- and upcrossing rates of level x, respectively. Hence, (3.263) is an *alternative* way of viewing the LC balance equation for $f(x)$. Due to the geometric factors ρ_1^{k-1}, $k = 1, 2, \ldots$, $(\rho_1 < 1)$, the series converges geometrically fast, to $f(x)$. Formula (3.263) is a series solution of the standard Volterra integral equation for the pdf given by (3.259). Moreover, because (3.263) is the sum of gen-k initiated upcrossing rates of level x, (3.263) is an alternative LC equation for $\{P_0, f(x)\}_{x>0}$. (In fact, the right side of (3.259) is the series expansion of the integral in (3.259)). Interestingly, we now have a geometric/physical interpretation of each term via LC. By computing or approximating the convolutions $\left(g_{1(k-1)} * g_0\right)(x)$, $k = 1, 2, \ldots$, we can quickly estimate $f(x)$ by summing the first N appropriate terms of (3.263).

In the standard M/G/1 queue, $g_0(x) \equiv g_{1(k-1)(x)}$ and the series (3.263) simplifies to the well-known Beneš series (3.71) (see [8]; formula (5.111), p. 201 in [104]).

Example 3.12 In M/M/1 where zero-wait arrivals get exceptional service $g_i(y) = \mu_i e^{-\mu_i y}$, and $E(S_i)$ =$1/\mu_i$, $\rho_i = \lambda/\mu_i$, $i = 0, 1$. Then

$$g_0(y) = e^{-\mu_0 y}\mu_0 \equiv g_{1(0)}, \qquad g_{1(k-1)}(y) = \frac{e^{-\mu_1 y}\,(\mu_1 y)^{k-2}\,\mu_1}{(k-2)!}, k = 2, 3, \ldots,$$

so that

$$\left(g_{1(k-1)} * g_0\right)(x) = \begin{cases} e^{-\mu_0 x}\mu_0, \ k = 1, \\ \int_{y=0}^{x} \frac{e^{-\mu_1(x-y)}(\mu_1(x-y))^{k-2}\mu_1}{(k-2)!} \cdot e^{-\mu_0 y}\mu_0 dy, \ k = 2, 3, \ldots \end{cases} \tag{3.264}$$

where $k-2 := (k-1)-1$ (see formula (3.39) for the pdf of $\text{Erl}_{k,\mu}$). Substituting from (3.264) into (3.263) gives the first term of the series as

$$P_0\rho_0 e^{-\mu_0 x}\mu_0 = P_0\lambda e^{-\mu_0 x}.$$

The sum of the subsequent terms of the series is

$$P_0\rho_0 \sum_{k=2}^{\infty} \rho_1^{k-1} \int_{y=0}^{x} \frac{e^{-\mu_1(x-y)}(\mu_1(x-y))^{k-2}\mu_1}{(k-2)!} \cdot e^{-\mu_0(y)}\mu_0 dy$$
$$= P_0\rho_0 e^{-\mu_1 x} \int_{y=0}^{x} \lambda\mu_0 e^{\mu_1 y} \cdot \sum_{k=2}^{\infty} \frac{(\lambda(x-y))^{k-2}}{(k-2)!} \cdot e^{-\mu_0 y} dy$$
$$= P_0\rho_0 e^{-\mu_1 x} \int_{y=0}^{x} \lambda\mu_0 e^{\mu_1 y} \cdot e^{\lambda(x-y)} \cdot e^{-\mu_0 y} dy$$
$$= P_0\rho_0\lambda\mu_0 e^{-(\mu_1-\lambda)x} \int_{y=0}^{x} e^{(\mu_1-\lambda-\mu_0)y} d\dot{y}$$
$$= P_0\lambda^2 \frac{e^{-\mu_0 x} - e^{-(\mu_1-\lambda)x}}{\mu_1-\lambda-\mu_0}.$$

Summing all the terms gives

$$f(x) = P_0\lambda e^{-\mu_0 x} + P_0\lambda^2 \frac{e^{-\mu_0 x} - e^{-(\mu_1-\lambda)x}}{\mu_1-\lambda-\mu_0}$$
$$= P_0\left(\frac{-\lambda^2}{\mu_1-\lambda-\mu_0} e^{-(\mu_1-\lambda)x} + \frac{\lambda(\mu_1-\mu_0)}{\mu_1-\lambda-\mu_0} e^{-\mu_0 x}\right),$$
$$P_0 = \frac{\mu_1-\lambda}{\mu_1-\lambda-\mu_0},$$

which is identical to formulas (3.148) and (3.149) in Sect. 3.6.1, which were obtained by converting an integral equation to a differential equation, solving the latter, and then using initial conditions to obtain the constants of integration.

The foregoing example illustrates important properties of the level crossing method.

1. We can partition the sample-path jumps of $\{W(t)\}_{t\geq 0}$ into subsets, such as jumps that initiate generation-k sub-busy periods, in order to obtain new views of the queueing kinetics directly from the structure of the sample path. In this Section the partition into gen-k jumps results in a generalization of the Beneš series for M/G/1.
2. Once the convolutions in the series are specified, it is straightforward in many cases to derive the pdf $f(x)$, $x > 0$. Comparing the above example with the solution method for $f(x)$, $x > 0$ in Sect. 3.6.1 shows that the LC-derived generalized Beneš series approach is more straightforward, and computes the coefficients of $e^{-(\mu_1-\lambda)x}$ and $e^{-\mu_0 x}$ directly without resorting to differential equations and using initial conditions.

3.17.4 Brief Discussion

We have indicated how to apply LC to derive transient and steady-state properties of the waiting time in several M/G/1 and M/M/1 queues, emphasizing steady-state results. Many of the LC-derived properties have been obtained in the literature by different methods, but some properties and results given in Chap. 3 are new.

A vast array of additional models and variants have been analyzed using LC, since 1976. For example, M/G/1 queues with Markov-generated server vacations [39] generalizes the standard M/G/1 server-vacation model. The vacation time following a service completion depends on the length of the immediately preceding *vacation*, via a Markov chain. Such dependency arises in many situations. A teller in a bank may do paper work following each service. After the next service completion, the paper work required may depend on the quality and quantity of the paperwork completed during the preceding vacation. Similar remarks apply to workers who write a report after completing a service, e.g., medical practitioners after seeing a patient; dentists after treating a patient; repairmen after completing a job; salesmen after completing a sale; and so forth.

Variants of the M/G$^{(a,b)}$/1 queue with bulk service were analyzed using LC in [20, 93]. The model utilizes a two-dimensional state $\{W(t), M(t)\}_{t\geq 0}$ where $W(t)$ is the virtual wait. Random variable $M(t)$ is discrete; it represents at time t, the number of 'customers in the waiting line confirmed' to form a service group, where $M(t) \in \{a, \ldots, b\}$ and b is the maximum service-group size. $M(t)$ is called the *system configuration*, which is explained for M/M/c queues in Sects. 4.4 and 4.5 in Chap. 4. The idea of *system configuration* was introduced by the author in [11] (see also Sect. 2 in [52]). System configurations are very useful in many stochastic models, by giving the LC method much flexibility for modelling various situations (see, e.g., the effective system configuration due to L. Green in [38]). A system configuration introduces sufficient detailed information, to make a model Markov ian. Creating a useful system configuration requires thinking through the system dynamics carefully.

Chapter 4
M/M/c Queue

4.1 Introduction

In Sect. 4.2 below we prove a useful general result (which we call Theorem B) about SP transitions. This theorem facilitates the analysis of *transient* distributions of state variables, and will be applied variously in the sequel.

Sections 4.3–4.5 explain the sample-path structure and dynamics of the generalized M/M/c queue. In the generalized model the SP can make a *transition* between disjoint state-space *sets* (called *pages* or *sheets*). Geometrically, sheets are analogous to a package of sheets of paper, cards in a deck, or pages of a book. They form a discrete number of disjoint subsets of the state space, not connected by a continuous segment of the sample path. (We can also model complex *single-server* queues using the ***method of sheets*** (aka ***method of pages***) (see, e.g., [39, 53, 93]). The *method of sheets* provides LC with great flexibility to analyze different types of stochastic models: e.g., queues; dams (see Sect. 11.8); inventories; production-inventories; actuarial risk models, replacement models; models in the natural sciences, etc.

Sections 4.6.1–4.6.9 develop equations for transient and steady-state pdfs of wait in the generalized M/M/c model. Sections 4.7–4.12 provide steady-state analyses of M/M/c variants using LC. In particular, Sect. 4.8 derives known results for the standard M/M/c queue as a special case of the generalized model. The remaining Sections of the Chapter study variants of M/M/c queues. All Sections provide empirical background for potentially novel applications of LC.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_4

4.2 Theorem B for Transient Analysis

We state and prove Theorem B. This straightforward theorem facilitates the transient analyses of a variety of stochastic models.

4.2.1 Theorem B

We first give a fundamental generalization of Theorems 3.1 and 3.2 of Chap. 3, which is useful for LC derivations of integro-differential equations for transient distributions in general.

Let $\{X(t)\}_{t\geq 0}$ denote a sample path of a general stochastic process with state space $\boldsymbol{S}$. Let $\boldsymbol{A}$, $\boldsymbol{B}$ be arbitrary measurable subsets of $\boldsymbol{S}$. Denote the transient probability $P(X(t) \in \boldsymbol{A})$ at instant t by $P_t(\boldsymbol{A}), t \geq 0$. Let $P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B})$ be the joint probability $P(X(t_1) \in \boldsymbol{A},\ X(t_2) \in \boldsymbol{B})$ at instants $t_1, t_2 \geq 0$. Let $\mathcal{I}_t(\boldsymbol{A})$ be the number of SP *entrances* and $\mathcal{O}_t(\boldsymbol{A})$ the number of SP *exits* of $\boldsymbol{A}$, during $(0, t)$ (see Fig. 2.7). Assume the derivatives

$$\frac{\partial}{\partial t}P_t(\boldsymbol{A}),\ \frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})),\ \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A}))$$

exist for all $t > 0$. Both

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) > 0 \text{ and } \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) > 0$$

hold wherever the derivatives exist, since $\mathcal{I}_t(\boldsymbol{A})$ and $\mathcal{O}_t(\boldsymbol{A})$ are counting processes which increase (wide sense, i.e., not strictly; they may be step functions) as t increases. The following useful result holds.

Theorem 4.1 *Theorem B* (P.H. Brill, 1983)

$$E(\mathcal{I}_t(\boldsymbol{A})) = E(\mathcal{O}_t(\boldsymbol{A})) + P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}) \tag{4.1}$$

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}). \tag{4.2}$$

Proof We give two proofs in order to develop intuition about the result.

Proof 1: This proof is similar to that of Theorems 3.1 and 3.2 in Sect. 3.2.3 above. We make the correspondence:

$$\boldsymbol{A} \leftrightarrow (-\infty, x], \quad \mathcal{I}_t(\boldsymbol{A}) \leftrightarrow \mathcal{D}_t(x), \quad \mathcal{O}_t(\boldsymbol{A}) \leftrightarrow \mathcal{U}_t(x),$$
$$P_t(\boldsymbol{A}) \leftrightarrow F_t(x), t \geq 0, \quad P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{A}) \leftrightarrow F_{t_1,t_2}(x, x), t_1, t_2 \geq 0.$$

SP down- and upcrossings of level x are entrances and exits of sets (Definitions 2.2–2.5). Note that

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = +1 &\iff X(0) \in \boldsymbol{A}^c,\ X(t) \in \boldsymbol{A},\\
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = -1 &\iff X(0) \in \boldsymbol{A},\ X(t) \in \boldsymbol{A}^c,\\
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = 0 &\iff X(0) \in \boldsymbol{A},\ X(t) \in \boldsymbol{A}\\
&\qquad \text{or } X(0) \in \boldsymbol{A}^c,\ X(t) \in \boldsymbol{A}^c.
\end{aligned}$$

We thus obtain the following values and corresponding probabilities:

$\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A})$	Probability
$+1$	$P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) = P_t(\boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$
-1	$P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c) = P_0(\boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$
0	$1 - P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}) + 2P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$

Taking the expected value $E(\mathcal{I}_t(A) - \mathcal{O}_t(A))$ and then the derivative $\frac{\partial}{\partial t} E(\mathcal{I}_t(A) - \mathcal{O}_t(A))$ yields (4.1) and (4.2).

Proof 2: Fix $t \geq 0$. The probability of the sure event $\boldsymbol{S}$ is

$$P_t(\boldsymbol{S}) = P_t(\boldsymbol{A} \cup \boldsymbol{A}^c) = P_t(\boldsymbol{A}) + P_t(\boldsymbol{A}^c) = 1.$$

Consider $P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{S})$. Events $\{X(t_1) \in \boldsymbol{A}\}$ and $\{X(t_2) \in \boldsymbol{S}\}$ are independent for every $0 \leq t_1 \neq t_2$. Knowledge that $\{X(t_1) \in \boldsymbol{A}\}$ has occurred, does not effect the probability of event $\{X(t_2) \in \boldsymbol{S}\}$, which is $P_{t_2}(\mathbf{S}) = 1$, and vice versa. Similarly, the events $\{X(t_1) \in \boldsymbol{S}\}$ and $\{X(t_2) \in \boldsymbol{B}\}$ are independent. Note that $\boldsymbol{S} = \boldsymbol{A} \cup \boldsymbol{A}^c = \boldsymbol{B} \cup \boldsymbol{B}^c$. Hence

$$\left.\begin{aligned}
P_{t_1}(\boldsymbol{A}) &= P_{t_1,t_2}(\boldsymbol{A}, \mathbf{S}) = P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B} \cup \boldsymbol{B}^c),\\
P_{t_2}(\boldsymbol{B}) &= P_{t_1,t_2}(\boldsymbol{S}, \boldsymbol{B}) = P_{t_1,t_2}(\boldsymbol{A} \cup \boldsymbol{A}^c, \boldsymbol{B}),
\end{aligned}\right\}$$

or

$$\left.\begin{aligned}
P_{t_1}(\boldsymbol{A}) &= P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}) + P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}^c),\\
P_{t_2}(\boldsymbol{B}) &= P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}) + P_{t_1,t_2}(\boldsymbol{A}^c, \boldsymbol{B}).
\end{aligned}\right\} \tag{4.3}$$

The possible values of $\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$ and corresponding joint probabilities at time points $t_1 = 0$ and $t_2 = t > 0$ are:

$\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$	Probability
0	$P_{0,t}(\boldsymbol{A}, \boldsymbol{A}) + P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}^c)$
$+1$	$P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A})$
-1	$P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c)$

(4.4)

Taking the expected value of $\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$ in (4.4) yields

$$\begin{aligned}E\left(\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})\right) &= P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c)\\ &= P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) + P_{0,t}(\boldsymbol{A}, \boldsymbol{A})\\ &\quad - \left(P_{0,t}(\boldsymbol{A}, \boldsymbol{A}) + P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c)\right)\\ &= P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}),\end{aligned}$$

which gives (4.1). Taking $\partial/\partial t$ in (4.1) yields (4.2). ■

Remark 4.1 Theorem B also applies to multi-dimensional processes with state space $\boldsymbol{S} \subseteq \mathbb{R}^n$, $n = 2, \ldots$, whose states are described by more than one continuous random variable. (**Note**: The symbol $\mathbb{R}$ denotes the set of real numbers.) We analyze two multi-dimensional inventory models in steady state, in Chap. 7.

4.3 Generalized M/M/c Queue

Customers arrive at an M/M/c queue in a Poisson stream at rate λ. There is one waiting line and c servers. Arrivals start service from the first available server, in order of arrival. We assume that for each arrival, the service time is exponentially distributed with rate selected from a nonempty set $\boldsymbol{\mu} = \{\mu_0, \ldots, \mu_J\}$ of $J + 1$ positive constants, depending on a *server-assignment policy* specified for the model; this allows service rates to be state dependent. Thus the standard M/M/c queue is a special case (see Sect. 4.8 below; p. 66ff in [84]).

For the generalized M/M/c queue we use a ‘partition/synthesis’ technique. We partition the state space into zero-wait and positive-wait states, and analyze the partitioned states to obtain ‘partial’ pdfs of the waiting time using LC. Then we synthesize those results to obtain the ‘total’ pdf of the waiting time, and related quantities of interest.

We next discuss: virtual wait; server workload; system configuration; the system point process (SP process); and give examples. (References for this section are [11] and [52], and others cited below.)

4.3.1 Virtual Wait and Server Workload

Notation 4.2 In the remainder of Sect. 4.3 we use two symbols for customers arriving to the system, depending on the context. (1) C (t) denotes a **would-**

be (potential*)* time-t arrival, $t \geq 0$. (2) $C_{a,t}$ denotes an **actual** time-t arrival, that arrives at t^-.

Let $C(t)$ be a would-be (potential) time-t arrival to the system, $t \geq 0$. Let $R_i(t)$ denote the (remaining) *workload* (in time units) at instant $t \geq 0$, of server i, $i = 1, \ldots, c$ (server numbering is arbitrary). Let $\{W(t)\}_{t \geq 0}$ be the *virtual wait process*. The random variable $W(t)$ is the would-be wait required by $C(t)$ measured from time t until the start of service of $C(t)$. Thus $W(t) = \min_{i=1,\ldots,c} \{R_i(t)\}, t \geq 0$. We assume: sample paths of $\{W(t)\}_{t \geq 0}$ and of $\{R_i(t)\}_{t \geq 0}$ are right continuous and have left limits; the model parameters are such that the steady state exists (condition relaxed for the transient analysis in Sects. 4.6.1–4.6.8 below).

Remark 4.2 In M/M/c ($c \geq 2$), virtual wait $\neq$ system workload. The system workload at time t is $\sum_{i=1}^{c} R_i(t)$.

Since $C_{a,t}$ is an *actual* time-t arrival to the system, that arrives at t^-, $C_{a,t}$'s wait is $W(t^-)$ before starting service, from some server i_t^*. If $R_j(t^-) = 0$ for some server j, then $W(t^-) = 0$, and i_t^* is one of the idle servers at t^-. For zero-wait arrivals, server i_t^* is selected from the idle servers according to the model's server-assignment policy (e.g., randomly, or by server number, etc.). If $R_i(t^-) > 0, i = 1, \ldots, c$, then $W(t^-) > 0$, and $C_{a,t}$ will start service from server i_t^* at instant $t + W(t^-)$ if i_t^* has the minimum workload among the c occupied servers at t^- (further explained in Sect. 4.4 below).

4.3.2 *Sample Paths of Workload and Virtual Wait*

In some models, sample paths of $R_i(t), i = 1, \ldots, c$, are useful for the overall analysis. We now outline how to construct a sample path of each $R_i(t), t \geq 0$, $i = 1, \ldots, c$. (Refer to Fig. 4.1, which depicts sample paths for a *special case* of generalized M/M/c with $c = 2$.) Without loss of generality, assume the system is empty at $t = 0$. Then $R_i(t) = 0$, $i = 1, \ldots, c$, from $t = 0$ until the first arrival instant (C_1). A new arrival (C_2) starts service from a server i^* and $R_{i^*}(\cdot)$ jumps upward to the ordinate $\underset{dis}{=} \text{Exp}_{\mu_{i^*}}$, where $\mu_{i^*} \in \boldsymbol{\mu}$. $R_{i^*}(\cdot)$, then decreases steadily with slope $= -1$ as service progresses.

Eventually all c servers become occupied simultaneously (just after C_2 arrives). Let $t_1 := \min\{t | \text{all } c \text{ servers are occupied}\}$. If the next customer C_τ arrives at time $\tau > t_1$ before any further service completion, then C_τ is the sole customer waiting to start service at time τ (C_3). C_τ's server will be i_τ^* if $R_{i_\tau^*}(\tau^-) = \min_{i=1,\ldots,c} \left\{R_i(\tau^-)\right\} := W(\tau^-)$ (virtual wait). The workload

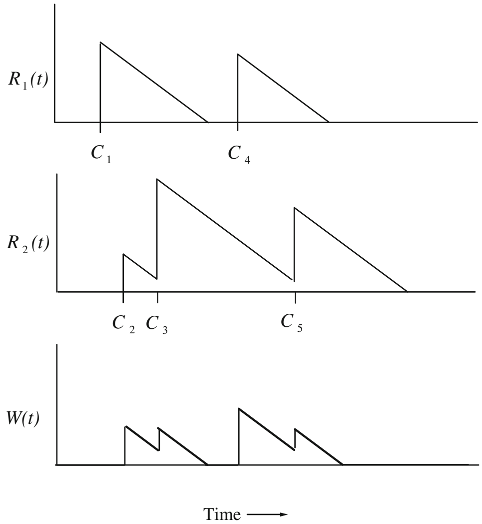


Fig. 4.1 Sample paths of $R_j(t)$, $j = 1, 2$, and $W(t)$, $t \geq 0$ in M/M/2. C_n, $n = 1, 2, \ldots$, denote customers at successive arrival instants

$R_{i_\tau^*}(\tau^-)$ jumps upward by C_τ's service time $s_{i_\tau^*} \underset{dis}{=} \text{Exp}_{\mu_{i_\tau^*}}$, where $\mu_{i_\tau^*} \in \boldsymbol{\mu}$. Thus $R_{i_\tau^*}(\tau) = R_{i_\tau^*}(\tau^-) + s_{i_\tau^*} = W(\tau^-) + s_{i_\tau^*}$. For *all other servers*, $R_i(\tau) = R_i(\tau^-)$, $i \neq i_\tau^*$. Subsequently $W(\tau) = \min_{i=1,\ldots,c}\{R_i(\tau)\}$.

The next arrival that "sees" *at least one idle server* (C_4), will cause the $\{W(t)\}_{t\geq 0}$ to evolve similarly. The next arrival that finds *all servers busy* will be assigned to that server which has minimum workload (C_5) and so forth. If arrivals find several customers waiting in line, the dynamics are similar to the case '*all servers busy'* (described in Sect. 4.4 below).

4.3.3 Distinguishable Servers

When tracking *server workloads*, we regard the servers as distinguishable (Fig. 4.1). However, we are often interested in the statistical properties of the entire system, rather than the processing of each individual customer, or the action of a particular server. Here we analyze the system by constructing a sample path of $\{W(t)\}_{t\geq 0}$ generated according to the model's prescribed probability laws for the service and interarrival times, and operational policies.

Suppose we can keep track of the c server workloads in continuous time. Then we could assign a '*ticket*' to each arrival, which points to its up-coming server, identified because it has the minimum workload at the arrival instant.

This procedure distributes *theoretical* waiting lines to the c servers, although there is only one physical waiting line in the waiting room.

4.3.4 Indistinguishable Servers

In many M/M/c models, it is not necessary to construct sample paths of the server workloads $\{R_i(t)\}_{t\geq 0}, i = 1, \ldots, c$, in order to construct a sample path of $\{W(t)\}_{t\geq 0}$. It suffices to regard the servers as *indistinguishable*. Then it is not necessary to track individual server workloads. To analyze important statistical properties of the model, it is sufficient to track ***directly*** the virtual wait $W(t) := \min_{i=1,\ldots,c}\{R_i(t)\}$. Thus we utilize what we call the ***system configuration*** (Sect. 4.4).

4.4 System Configuration

In generalized M/M/c, assume a 'system manager' knows the up-coming target server i_t^* to be occupied at instant $t + W(t^-)$ by a would-be time-t arrival, denoted by C(t) (i.e., the manager knows the server having minimum workload at time t). Let $\boldsymbol{M}(t) :=$ *system configuration* at time t. The process $\{\boldsymbol{M}(t)\}_{t\geq 0}$ tracks the service rates of the $c-1$ servers *other than* i_T^*. We assume that the model specifies $J+1$ possible exponential service rates: $\boldsymbol{\mu}$ = $\{\mu_0, \mu_1, \ldots, \mu_J\}$. Each arrival is assigned a service rate selected from the set $\boldsymbol{\mu}$. Recall that if t is not an arrival instant, sample-path right continuity implies $W(t^-) = W(t)$.

Definition 4.1 The system configuration $\boldsymbol{M}(t)$ is a $J+1$ vector of **server occupancy numbers** $m_j \geq 0$, namely $\boldsymbol{M}(t) = (m_0, \ldots, m_J)$, where $m_j :=$ **number of servers having service rate** $\mu_j \in \mu$, *among the* $c-1$ *servers other than* i_t^* *at* $t + W(t^-)$.

Definition 4.2 The set of all possible configurations is denoted by $\boldsymbol{M}$ = $\{\boldsymbol{m}|\boldsymbol{m} = (m_0, \ldots, m_J)\}$.

For each configuration $\boldsymbol{m} \in \boldsymbol{M}$, $0 \leq \sum_{j=0}^{J} m_j \leq c-1$; C($t$) would start service at instant $t + W(t^-)$ and would be assigned a service rate $\mu_t(W(t^-), \boldsymbol{m}) \in \boldsymbol{\mu}$, which nay be a function of three variables: t, $W(t^-)$, and $\boldsymbol{m}$. That is

$$(t, W(t^-), \boldsymbol{m}) \to \mu_j \in \boldsymbol{\mu}, \text{ for some } j = 0, \ldots, J.$$

Remark 4.3 In various models, the service rate $\mu_t(\cdot, \cdot)$ may also depend on other variables as well. It may be selected randomly from the set $\boldsymbol{\mu}$. Additionally, the number of possible service rates in $\boldsymbol{\mu}$ may be countable.

4.4.1 Inter Start-of-service Depart Time $\mathcal{S}_t$

In generalized M/M/c a *key random variable* is the time-t *'look-ahead'* ***inter start-of-service depart time,*** denoted by $\mathcal{S}_t$. For example, let the state $(W(t^-), M(t^-))$ be $(x, \boldsymbol{m})$ when customer $\mathrm{C}_{a,t}$ arrives. Then $\mathrm{C}_{a,t}$'s required wait before starting service is $x \geq 0$ and the configuration is $\boldsymbol{m}$. If $0 \leq \sum_{i=0}^{J} m_j \leq c-1$ then $x = 0$ and $\mathrm{C}_{a,t}$ starts service immediately by one of the idle servers. If $\sum_{i=0}^{J} m_j = c-1$ there are two possibilities for x. If $x = 0$ then $\mathrm{C}_{a,t}$ starts service immediately by the ***single idle server***. If $x > 0$ then $\mathrm{C}_{a,t}$ waits time x before starting service by the first available server thereafter. Just after $\mathrm{C}_{a,t}$ starts service at time $t + x$ all c servers will be occupied.

Definition 4.3 The ***inter start-of-service depart time*** $\mathcal{S}_t$ is the time measured from $t + x$ (start of $\mathrm{C}_{a,t}$'s service time) until the first departure from the system after $t + x$. In other words $\mathcal{S}_t :=$ ***time from the start of service of*** $\mathrm{C}_{a,t}$ ***until the first departure from the system thereafter***.

Importantly,

$$\mathcal{S}_t \underset{dis}{=} \min\left\{\mathrm{Exp}_{m_0\mu_0}, \ldots, \mathrm{Exp}_{m_J\mu_J}, \mathrm{Exp}_{\mu_t(W(t^-),\boldsymbol{m})}\right\},$$

which is the *minimum* of $\sum_{i=0}^{J} m_j + 1\ (= c)$ independent exponential r.v.s. Among these, m_j servers have rate μ_j, $j = 0, \ldots, J$, and one server has rate $\mu_t(x, \boldsymbol{m})$ (assigned to $\mathrm{C}_{a,t}$). Thus $\mathcal{S}_t \underset{dis}{=} \mathrm{Exp}_{\nu_t}$ where $\nu_t = \sum_{j=0}^{J} m_j\mu_j + \mu_t(x, \boldsymbol{m})$.

An important aspect of the forgoing definition of system configuration and use of S_t when $W(t^-) > 0$, is that once all c servers are occupied, the probability distribution of S_t is independent of future arrivals to the system. That is, the set of active servers functions like a separate sub-system until the first departure thereafter, mindful of the memoryless property of the exponential service times. Although the concept '*system configuration*' may appear 'different', it is straightforward to apply when developing model equations for the pdf of the waiting time in complex M/M/c queues (see, e.g., pp. 80–97 in [11], and also in [38, 52, 53]).

4.4.2 Number of Configurations

Let $(W(t), M(t)) = (x, \boldsymbol{m})$ (see Definition 4.1 above). Looking ahead to $t + W(t)$, assume $\boldsymbol{m} = (m_0, \ldots, m_J)$ is such that

$$\sum_{j=0}^{J} m_j = k, 0 \leq k \leq c - 1.$$

The servers are considered to be *indistinguishable* (as in subsequent models in this monograph, unless otherwise noted). We track only the *number of servers occupied* with service rate $\mu_j \in \boldsymbol{\mu}$, $j = 0, \ldots, k$.

The number of possible configurations such that exactly k servers are occupied, is the number of non-negative integer solutions of the equation

$$m_0 + \cdots + m_J = k.$$

It is the same as the number of ways of distributing k *indistinguishable* balls in $J + 1$ *distinguishable* cells, namely $\binom{J+k}{J} = \binom{J+k}{k}$ (see Lemma, p. 36, Chap. II in [73]). Thus, the total number of possible configurations is

$$\sum_{k=0}^{c-1} \binom{J+k}{J} = \binom{J+c}{J+1} = \binom{J+c}{c-1}. \tag{4.5}$$

The first equality in (4.5) is readily proved by induction.

Example 4.1 Consider an M/M/c queue with $c = 3$ and $J = 2$, so that $\boldsymbol{\mu} = \{\mu_0, \mu_1, \mu_2\}$. If a potential arrival C(t) finds the system **empty**, then $(W(t), M(t)) = (0, (0, 0, 0))$; thus $m_i = 0$, $i = 0, 1, 2$. C(t) would wait zero and "see" zero servers occupied, (0, 0, 0). The number of solutions of $m_0 + m_1 + m_2 = 0$ is $\binom{J+0}{J} = \binom{2}{2} = 1$. C($t$) would wait $W(t^-) = 0$ and start service from one the three unoccupied servers, per the server-assignment policy.

If C(t) would find **one** customer in the system (one occupied server), then $W(t^-) = 0$ and the configuration that C(t) would "see" is one of **three** possible vectors

$$M(t^-) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}.$$

The number of solutions of $m_0 + m_1 + m_2 = 1$ is $\binom{J+1}{J} = \binom{3}{2} = 3$. C($t$) would start service from one of the two unoccupied servers, per the server-assignment policy.

If C(t) would find **two** customers in the system, then $W(t^-) = 0$ and the configuration that C(t) would "see" is one of **six** possible vectors

$$M(t^-) \in \{(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)\}.$$

The number of solutions of $m_0 + m_1 + m_2 = 2$ is $\binom{J+2}{J} = \binom{4}{2} = 6$. C($t$) would start service from the **one unoccupied** server.

If C(t) would find **three or more** customers in the system, then all **three** servers would be occupied at t^-. The look-ahead configuration that C(t) would "see" **just before start of service** at $t + W(t^-)$ is also one of **six** possible vectors

$$M(t) \in \{(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)\}.$$

The six possible configurations are the *same* as when C(t) sees **two occupied** servers. This is because a configuration tracks the service-rate occupancies of those servers **other than** C(t)'s eventual server. Customer C(t) would wait a **positive time** and start service at $t + W(t^-)$ from some server i_t^*. We "look ahead" to the **start of service instant** $t + W(t^-)$ **and assign rate** $\mu_t(W(t^-), M(t^-))$ *to* i_t^*. The random variable $M(t)$ tracks the service-rate occupancies of the two servers other than i_t^* at $t + W(t^-)$. (The look-ahead idea is not new. For example, it is tacitly assumed for the virtual wait in the standard M/G/1 queue, where we increase the virtual wait by a service time at an *arrival* instant, although the service is not started until the end of the waiting time. The generalized M/M/c generalizes the M/G/1 look-ahead idea to the start-of-service time.)

At instant t, the state $(W(t), M(t))$ conveys sufficient information to determine the probabilities of the m_j $(j = 0,\ldots,J)$ occupied servers that will have the minimum service time among all the occupied servers at time $t + W(t^-)$. These probabilities depend on the Markovian property. We shall illustrate this more fully in Example 4.2 below.

The total number of possible configurations is

$$\sum_{k=0}^{c-1} \binom{J+k}{J} = \sum_{k=0}^{2} \binom{J+k}{J} = \binom{J+c}{J+1} = \binom{5}{3} = 10.$$

4.4.3 Border States

In Example 4.1 the zero-wait state $\{(0, \boldsymbol{m})\}$ is a ***border*** state if

$$\boldsymbol{m} \in \{(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)\}.$$

Definition 4.4 We call the state $(W(t), M(t))$ a **border state**if $W(t) = 0$ and $M(t))$ is such that $\sum_{j-0}^{J} m_j = c - 1$. A border state is a discrete zero-wait state in a **boundary** separating other discrete zero-wait states and a set of continuous positive-wait states.

In the above definition, the *other* zero-wait states are *non-border* states such that $0 \leq \sum_{j-0}^{J} m_j < c - 1$. Border states communicate with continuous positive-wait states in one step: at arrival instants (zero-wait → positive-wait); or at departure instants (positive-wait → zero-wait). When the SP moves on a path *from* a *non-border* zero-wait state *to* a continuous positive-wait state the path must pass through a border state at an arrival instant. In the opposite direction, *from* a positive-wait state to a non-border zero-wait state, the path must pass through a border state at a departure instant.

We denote the set of border states by $\boldsymbol{S}_b$, and the set of border configurations by $\boldsymbol{M}_b$. Thus

$$
\begin{aligned}
\boldsymbol{S}_b &= \left\{(0, \boldsymbol{m}) \mid \textstyle\sum_{j=0}^{J} m_j = c - 1\right\}, \\
\boldsymbol{M}_b &= \{\boldsymbol{m} \mid (0, \boldsymbol{m}) \in \boldsymbol{S}_b\} = \left\{\boldsymbol{m} \mid \textstyle\sum_{j=0}^{J} m_j = c - 1\right\}.
\end{aligned}
\tag{4.6}
$$

4.4.4 The Next Configuration

Consider an actual arrival $C_{a,t}$ at instant t. $C_{a,t}$ "sees" configuration $M(t^-)$. Just *after* the arrival the configuration is $M(t)$. Either $M(t) = M(t^-)$ or $M(t) \neq M(t^-)$. We illustrate by example how to compute the probability mass function of $M(t)$.

Example 4.2 Consider Example 4.1 for M/M/c with $c = 3$, $J = 2$. Suppose $C_{a,t}$ arrives when the wait is $W(t^-)$ and the configuration is $(m_0, m_1, m_2) = (1, 1, 0)$. The state is

$$(W(t^-), M(t^-)) = (0, (1, 1, 0)).$$

Suppose that $C_{a,t}$ is assigned service rate μ_0, i.e., $\mu_t(W(t^-), (1, 1, 0)) = \mu_0$.

At instant $t + 0$, **just after** $C_{a,t}$ **starts service**, there will be **two** servers with rate μ_0 since $m_0 = 1$. There will be *one* server with rate μ_1, since $m_1 = 1$. The inter-start-of-service-depart time $\mathcal{S}_t \underset{dis}{=} \text{Exp}_{2\mu_0+\mu_1}$.

We now compute the probability distribution of the **next configuration** at instant $t + W(t)$. Thus,

$$
\begin{aligned}
P(M(t) = (2, 0, 0)) &= P(\text{rate-}\mu_1 \text{ server finishes first}) \\
&= \frac{\mu_1}{2\mu_0 + \mu_1},
\end{aligned}
$$

$$
\begin{aligned}
P(M(t) = (1, 1, 0)) &= P(\text{rate-}\mu_0 \text{ server finishes first}) \\
&= \frac{2\mu_0}{2\mu_0 + \mu_1}.
\end{aligned}
$$

Importantly

$$P(M(t) = (2, 0, 0)) + P(M(t) = (1, 1, 0)) = 1.$$

The only two possible configurations for $M(t)$ are $(2, 0, 0)$ and $(1, 1, 0)$, independent of whether $W(t^-) = 0$ or $W(t^-) > 0$ (illustrated below in Example). No other configuration is possible for $M(t)$ once the arrival at t^- has been assigned rate μ_0. Knowledge of their probabilities is sufficient to analyze the sample path to write down model equations using LC.

Remark 4.4 The service mechanism can be generalized considerably. We can expand the domain of $\mu_t(w, \boldsymbol{m})$ to include: type or priority class of $\mathrm{C}_{a,t}$; type of customer replaced by $\mathrm{C}_{a,t}$ in server i_t^*; type of any customer followed by $\mathrm{C}_{a,t}$ into service; identity of server i_t^* (e.g., server number or unique property); number of customers in the system or waiting for service at the arrival or start of service instant of $\mathrm{C}_{a,t}$; various types of bounds on the virtual wait; reneging indices; blocked and cleared customers, etc. (see, e.g., [38] for an effective definition of $M(t)$ due to L. Green; [42]; also see [39, 44, 53]; and others).

Other generalizations may incorporate: a non-homogeneous Poisson arrival process with intensity λ_t, or a Poisson arrival rate $\lambda(W(t), M(t))$ which is a function of the current state $(W(t), M(t))$; or various Markov arrival processes.

4.5 System Point Process

We now discuss the ***system point process*** and the geometry of its state space (see Fig. 4.2).

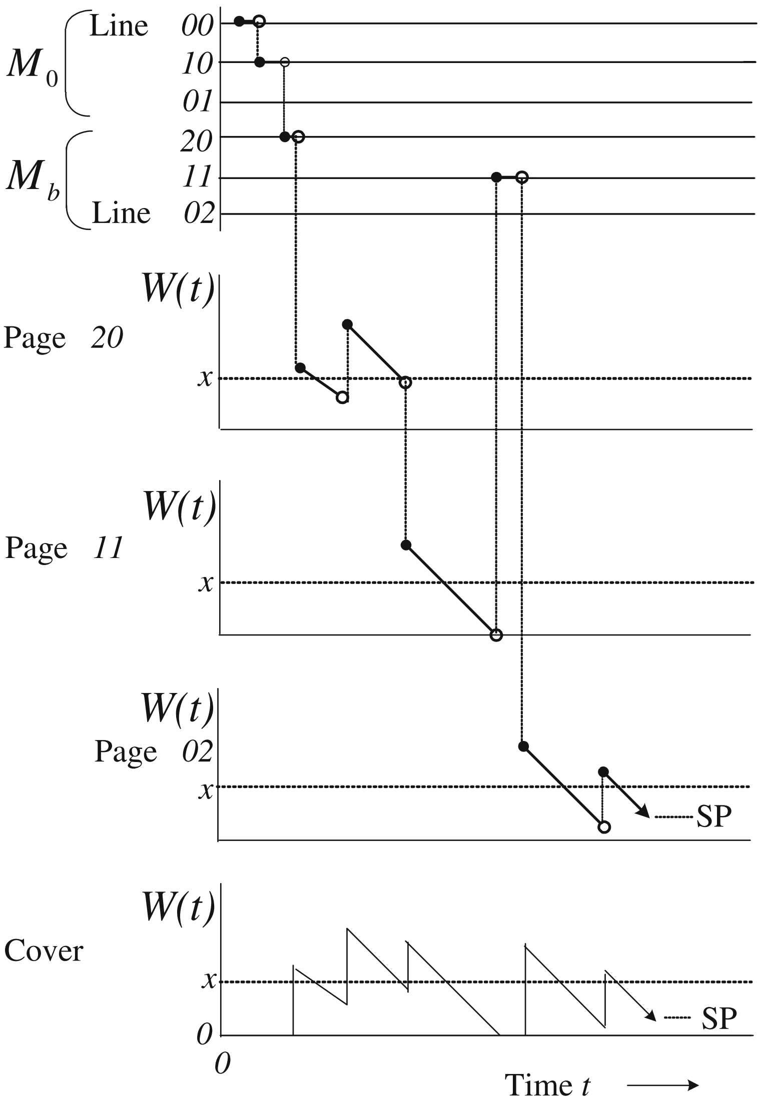


Fig. 4.2 Sample path of SP process $\{W(t), M(t)\}_{t \geq 0}$ for Example 4.4 with random assignment of service rates independent of state at arrival instants ($c = 3$, $J = 1$). The space $\boldsymbol{T} \times \boldsymbol{S}$ has 6 lines for zero-wait states, and 3 pages (sheets) for positive-wait states (pages 20, 11, 02). The cover is the projection of the sample path from all lines and pages onto one non-negative planar quadrant

We call $\{W(t), M(t)\}_{t \geq 0}$ the *system point (SP) process*. Its nomenclature derives from the fact that the SP traces out a sample path as the system evolves over time. The SP process for M/M/c queues is a generalization (with exponential service times) of the virtual wait process for M/G/1 queues. State variable $W(t) :=$ *virtual wait*; state variable $M(t) :=$ *system configuration at time t*. Random variable $M(t)$ is discrete. The SP process is a Markov process (Sect. 4.5.7 below).

We partition the *state space* $\boldsymbol{S}$ into three disjoint state-space sets $\boldsymbol{S}_0$, $\boldsymbol{S}_b$, $\boldsymbol{S}_1$. $\boldsymbol{S}_0$ contains the zero-wait states that are non-boundary states. $\boldsymbol{S}_b$ contains the zero-wait states that are boundary states. $\boldsymbol{S}_1$ contains the positive-wait states. The states in $\boldsymbol{S}_0 \cup \boldsymbol{S}_b$ are atoms (see Sect. 2.4.9 for 'atom'). The states in $\boldsymbol{S}_1$ are points in a continuum, e.g., $(x, \boldsymbol{m})$, $x > 0$, and $\boldsymbol{m}= (m_{0,\ldots,} m_J)$. Specifically,

$$\begin{aligned} \boldsymbol{S}_0 &= \{(0, \boldsymbol{m})|0 \le \textstyle\sum_{j=0}^{J} m_j \le c-2\}, \\ \boldsymbol{S}_b &= \{(0, \boldsymbol{m})| \textstyle\sum_{j=0}^{J} m_j = c-1\}, \\ \boldsymbol{S}_1 &= \{(x, \boldsymbol{m})|x > 0, \textstyle\sum_{j=0}^{J} m_j = c-1\}. \end{aligned} \tag{4.7}$$

Note that $\boldsymbol{S} = \boldsymbol{S}_0 \cup \boldsymbol{S}_b \cup \boldsymbol{S}_1$, and $\boldsymbol{S}_0 \cap \boldsymbol{S}_b = \boldsymbol{S}_0 \cap \boldsymbol{S}_1 = \boldsymbol{S}_b \cap \boldsymbol{S}_1 = \phi$, the empty set. The corresponding sets of system configurations are

$$\begin{aligned} \boldsymbol{M}_0 &= \{\boldsymbol{m}|(0, \boldsymbol{m}) \in \boldsymbol{S}_0) = \{\boldsymbol{m}|0 \le \textstyle\sum_{j=0}^{J} m_j \le c-2\}; \\ \boldsymbol{M}_b &= \{\boldsymbol{m}|(0, \boldsymbol{m}) \in \boldsymbol{S}_b) = \{\boldsymbol{m}| \textstyle\sum_{j=0}^{J} m_j = c-1\}; \\ \boldsymbol{M}_1 &= \{m|(x, \boldsymbol{m}) \in \boldsymbol{S}_1\} = \{\boldsymbol{m}|x > 0, \textstyle\sum_{j=0}^{J} m_j = c-1\}; \end{aligned} \tag{4.8}$$

thus $\boldsymbol{M}_b = \boldsymbol{M}_1$ (see Sect. 4.4 above).

$\boldsymbol{W(t^-) = 0}$ An arrival C(t) would "see" $W(t^-) = 0$ if and only if the state at time t^- is in $\boldsymbol{S}_0 \cup \boldsymbol{S}_b$. C($t$) would then wait zero and start service from some server, say i_t^*, at time t. Geometrically, we associate a distinct horizontal line $\boldsymbol{T} \times (0, \boldsymbol{m})$ with each state $(0, \boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b$ where $\boldsymbol{T}$ is the time axis $[0, \infty)$. We call the line $\boldsymbol{T} \times (0, \boldsymbol{m})$ "*line* $\boldsymbol{m}$" (e.g., Fig. 4.2).

$\boldsymbol{W(t^-) > 0}$ C(t) would "see" $W(t^-) > 0$ if and only if the state is in $\boldsymbol{S}_1$. C(t) would wait time $W(t^-)$ and start service from some server, say i_t^*, at time $t + W(t^-)$. Geometrically, we associate the quadrant of the plane $\boldsymbol{T} \times (0, \infty)$ with each set of continuous states $(x, \boldsymbol{m}) \in \boldsymbol{S}_1$, $x > 0$. We call the positive quadrant $\boldsymbol{T} \times ((0, \infty), \boldsymbol{m})$ ***sheet*** (or ***page)*** $\boldsymbol{m}$ (e.g., Fig. 4.2).

Sample Path Diagram The LC analyst draws (or visualizes) a plot of $W(t)$ versus t on page $\boldsymbol{m}$ while the system is in the state corresponding to configuration $\boldsymbol{m}$. In a diagram, we may place the zero-wait *border* lines (corresponding to the states in $\boldsymbol{S}_b$) $\boldsymbol{T} \times (0, \boldsymbol{m})$, $(0, \boldsymbol{m}) \in \boldsymbol{S}_b$, alongside the zero-wait *non-border* lines for states $(0, \boldsymbol{m}) \in \boldsymbol{S}_0$; or else at level 0 of the corresponding sheets for the positive-wait states $\boldsymbol{S}_1$ having the same configurations $\boldsymbol{m}$. There is a one-to-one correspondence between sheets and states in $\boldsymbol{S}_b$ (Fig. 4.2).

4.5.1 Sample Path of SP Process

A sample path of $\{W(t), M(t)\}_{t\geq 0}$ is a piecewise right-continuous function of t having left limits. It has a finite number of jumps during finite time intervals (see Sect. 2.2 and Definition 2.1 in Chap. 2). We plot a sample path within a Cartesian product space $\boldsymbol{T} \times \boldsymbol{S} = \boldsymbol{T} \times (\boldsymbol{S_0} \cup \boldsymbol{S_b} \cup \boldsymbol{S_1})$. The direction of time is from left to right. It is useful to envisage each Cartesian product $\boldsymbol{T} \times (0, \boldsymbol{m}), (0, \boldsymbol{m}) \in \boldsymbol{S_0} \cup \boldsymbol{S_b}$ as a ***line***; and each quadrant $\boldsymbol{T} \times ((0, \infty), \boldsymbol{m})$, $\boldsymbol{m} \subseteq \boldsymbol{M}_1$, as a sheet (or page in a ***book***).

Description of a Sample Path

Assume that the system starts empty. The SP moves among the zero-wait lines, jumping from line to line at arrival and departure instants, eventually reaching a zero-wait *border* line (often placed at level zero of some sheet $\boldsymbol{m} \in \boldsymbol{M}_b$). Eventually the SP jumps from line $\boldsymbol{m} \in \boldsymbol{M}_b$, to a positive level on some sheet '$\boldsymbol{k}$', at an arrival instant. It then moves steadily with slope -1 on sheet $\boldsymbol{k}$. It is possible that either $\boldsymbol{m} = \boldsymbol{k}$, or $\boldsymbol{m} \neq \boldsymbol{k}$, depending on the probabilities governing the motion of the SP. (See Fig. 4.2 and Example 4.4 in Sect. 4.5.5 for a detailed example.) Other clarifying examples are in the author's Ph.D. thesis (Fig. 4.3, p. 79, Chap. 4 in [11]; and in [52]).

At an arrival instant while the SP is on sheet $\boldsymbol{k}$, the SP may jump to another sheet, say $\boldsymbol{m}'$, and move steadily with slope -1 on sheet $\boldsymbol{m}'$ for a positive time. Otherwise the SP may jump, and stay on the same sheet $\boldsymbol{k}$. On each sheet it moves downward with slope -1. If the SP hits level 0 from above on page $\boldsymbol{k}$ before the next arrival, it starts moving immediately on the border line $\boldsymbol{k}$ (no customers waiting, $c - 1$ servers occupied).

If the SP is in a state in $\boldsymbol{S}_b \cup \boldsymbol{S}_1$ having configuration $\boldsymbol{m}$ at some arrival instant, it makes a jump ending either on page $\boldsymbol{m}$ or on some page $\boldsymbol{k} \neq \boldsymbol{m}$. If $\boldsymbol{k} \neq \boldsymbol{m}$, the SP is said to make an $\boldsymbol{m} \to \boldsymbol{k}$ *transition*. This may be an upward jump from a border line $\boldsymbol{m}$, or from sheet $\boldsymbol{m}$ to sheet $\boldsymbol{k}$, at an arrival instant. Generally, $\boldsymbol{m} \to \boldsymbol{k}$ transitions do not give rise to 'typical' level crossings as in M/G/1 models that have exactly one 'page'. However, $\boldsymbol{m} \to \boldsymbol{m}$ transitions from a border line $\boldsymbol{m}$ or from a point on sheet $\boldsymbol{m}$ to a higher point on sheet $\boldsymbol{m}$, are similar to SP jumps as discussed for models with a single sheet (Sect. 2.3 in Chap. 2).

Remark 4.5 In some model variants, an $\boldsymbol{m} \to \boldsymbol{k}$ transition may be a **parallel jump**. That is, the SP makes a jump from a level y on page $\boldsymbol{m}$ to level y on page $\boldsymbol{k} \neq \boldsymbol{m}$, at an arrival instant. For example, in an M/G/1 queue, we may utilize a modified configuration $M(t) = n$, where n is the number of customers waiting for service, and the virtual wait is unchanged at some arrival instants. Such parallel jumps occur in M/G/1 or M/M/c queues with bulk service (see, e.g., [93]).

4.5.2 A Metaphor for Sample Path and SP Motion

The SP motion over the state space is like the motion of the tip of a pen writing out a *single-page* history of the system over time. The writing takes place in a book of transparent pages all the same size. The cover is also transparent. The pen moves from left to right, and never overlays what has been written already. After writing flat or sloped lines on a page for a random amount of time, the pen jumps to a different page, and continues writing. The pen jumps in this manner at random time points from page to page. The next page is selected by a random mechanism depending on where it is presently. The *entire* history up to an instant in time can be seen only by holding all the pages one behind the other, like pages in a book, and viewing the projected history on the ***cover***. The projected history on the cover is invariant with respect to shuffling or mixing the pages, which change their relative positions. An analyst that views an arbitrary page in isolation, sees only local segments of the history specific to that page (see Fig. 3.2, p. 49 and Fig. 4.3, p. 79 in [11]).

The global history is like the *total* sample path of the SP process over the state space $\boldsymbol{S_0} \cup \boldsymbol{S_b} \cup \boldsymbol{S_1}$. The local histories on various pages are like sample-path segments due to sojourns on the 'lines' and 'sheets' of the state space. On the cover, SP motion on all the lines occurs at level 0. That is, when all the lines are projected onto the cover, they are placed at level zero—to form a single "line 0".

We may think of the overall method as having several steps.

1. **Partition** the **Time-State space** into mutually exclusive and exhaustive lines and sheets.
2. **Analyze** the sample-path segments on the lines and sheets using LC methods.
3. **Project** the sample-path segments from the lines and sheets onto the 'cover' of the 'book'. Analyze the projected path on the cover using LC.
4. **Combine** all the LC results with a normalizing condition. Construct the model equations (usually Volterra integral equations of the second kind with parameter for the pdf's of interest) and derive probability distributions of the model.

The LC method utilizes statistical properties of the local path segments on the lines and sheets. It also uses statistical properties of the projected path on the cover. It employs the one-step communication properties among the lines and sheets (at successive arrivals and/or departures) to construct a sample path. Basic LC theorems apply to each page $\boldsymbol{m} \in \boldsymbol{M}$. Jumps out of, and into lines and sheets, follow rate-conservation laws.

Equations for PDF of Wait

We use sample-path structure, and transition rates into and out of state-space sets, to construct (by inspection of the sample path) integro-differential and differential equations in a *transient* analysis. Similarly, we construct integral equations and algebraic equations in a *steady-state* analysis. These are equations for the joint pdf and/or cdf of *wait and system configuration*. We can also derive equations for the marginal (total) pdf and cdf of wait, or for the probabilities of the system configurations.

Remark 4.6 The author originally had the idea for partitioning the state space, visualizing the positive-wait states over time in separate quadrants, and having a 'system point' moving on the quadrants over time, from an analogy with Riemann sheets and diagrams of winding numbers in complex variable theory (see, e.g., Sect. 3.4, p. 137ff in [96]). My Ph.D. thesis used the term 'sheets'. The term 'pages' was introduced soon after. I also thought of using the term 'cards', analogous to boxes of computer cards for data and programs, which were widely in use in 1974. Then, the state space could be pictured like a box or deck of rectangular cards. Such ('IBM') cards had been ubiquitous until personal computers became common in the 1980s.

4.5.3 Notation: Probabilities and Distributions

Transient Probabilities and Distributions
We denote: the zero-wait probabilities by

$$P_t(0, \boldsymbol{m}) = P(W(t) = 0, M(t) = \boldsymbol{m}), (0, \boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b;$$

the mixed joint cdf of $(W(t), M(t))$ by

$$\begin{aligned} F_t(x, \boldsymbol{m}) &= P(W(t) \leq x, M(t) = \boldsymbol{m}) \\ &= P_t(0, \boldsymbol{m}) + P(0 < W(t) \leq x, M(t) = \boldsymbol{m}) \\ &= P_t(0, \boldsymbol{m}) + \int_{y=0}^{x} f_t(y, \boldsymbol{m}) dy, \ x \geq 0, t \geq 0, \\ &\qquad (0, \boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b, (x, \boldsymbol{m}) \in \boldsymbol{S}_1, \end{aligned}$$

where $P(0 < W(t) \leq x, \boldsymbol{M}(t) = \boldsymbol{m}) = P(\phi) = 0$ if $x = 0$; and the mixed joint pdf of $(W(t), M(t))$ is

$$f_t(x, m) = \frac{\partial}{\partial x} F_t(x, \boldsymbol{m}), x > 0, t \geq 0, (x, \boldsymbol{m}) \in \boldsymbol{S}_1,$$

wherever $\frac{\partial}{\partial x} F_t(x, \boldsymbol{m})$ exists.

We *assume*:

1. $F_t(x, \boldsymbol{m})$ and $f_t(x, \boldsymbol{m})$ are right continuous in x for every $t \geq 0, \boldsymbol{m} \in \boldsymbol{M}_1$.
2. $\frac{\partial}{\partial t} F_t(x, \boldsymbol{m})$ and $\frac{\partial}{\partial t} f_t(x, \boldsymbol{m})$, $t > 0$, $x \geq 0$, exist and are finite for every $\boldsymbol{m} \in \boldsymbol{M}_1$.

Let $P_0(t) := P(W(t) = 0)$ be the marginal probability of a zero wait at t. Then

$$\begin{aligned} P_0(t) &= \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b} P_t(0, \boldsymbol{m}) \\ &= \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_0} P_t(0, \boldsymbol{m}) + \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_b} P_t(0, \boldsymbol{m}), t \geq 0. \end{aligned}$$

The transient marginal cdf of wait $P(W(t) \leq x)$ is

$$\begin{aligned}
F_t(x) &= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0} P_t(0,\boldsymbol{m}) + \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_b} F_t(x,\boldsymbol{m}) \\
&= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0\cup \boldsymbol{S}_b} P_t(0,\boldsymbol{m}) + P(0 < W(t) \leq x) \\
&= P_0(t) + P(0 < W(t) \leq x) \\
&= P_0(t) + \int_{y=0}^{x} f_t(y)dy, \;\; x \geq 0, t \geq 0.
\end{aligned}$$

Note that $P_t(0,\boldsymbol{m}) = F_t(0,\boldsymbol{m})$ for $(0,\boldsymbol{m}) \in \boldsymbol{S}_b$.

(Recall the definitions of $\boldsymbol{M}_b$ and $\boldsymbol{S}_b$ in (4.6), and $\boldsymbol{M}_b = \boldsymbol{M}_1$, which is the set of system configurations for positive-wait states.)

The transient marginal pdf of $W(t)$ is

$$f_t(x) = \frac{\partial}{\partial x} F_t(x) = \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} f_t(x,\boldsymbol{m}), x > 0, t \geq 0.$$

A potential (would-be) arrival C(t) would find the system configuration to be $\boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$ with probability $P_t(0,\boldsymbol{m})$. C(t) would find the configuration to be $\boldsymbol{m} \in \boldsymbol{M}_1$ with probability $F_t(\infty,\boldsymbol{m})$. The normalizing condition for fixed $t \geq 0$, is

$$\begin{aligned}
F_t(\infty) &= \sum_{\boldsymbol{m}\in \boldsymbol{M}_0} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} F_t(\infty,\boldsymbol{m}) \\
&= \sum_{\boldsymbol{m}\in \boldsymbol{M}_0\cup \boldsymbol{M}_b} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} \int_{y=0}^{\infty} f_t(y,\boldsymbol{m})dy \\
&= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0\cup \boldsymbol{S}_b} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} \int_{y=0}^{\infty} f_t(y,\boldsymbol{m})dy = 1.
\end{aligned}$$

Steady-State Probabilities and Distributions

We denote the steady-state zero-wait probabilities, pdfs and cdfs of wait by dropping the subscript t in the immediately foregoing notation for the transient quantities.

4.5.4 Configuration Just After an Arrival

Example 4.3 below demonstrates the probability of a system configuration just after an arrival. Assume that an actual customer $C_{a,t}$ arrives and finds the state to be $(W(t^-), \boldsymbol{M}(t^-)) = (x, \boldsymbol{m})$. The service rate assigned to $C_{a,t}$ is $\mu_t(x, \boldsymbol{m}) \in \boldsymbol{\mu}$. Recall that sample paths are right continuous and have left limits.

Example 4.3 Consider Example 4.2 in Sect. 4.4.4, where $c = 3$, $J = 2$. Let each arrival receive a service rate selected **with equal probability** from the set $\boldsymbol{\mu} := \{\mu_0, \mu_1, \mu_2\}$. Then

$$P(C_{a,t} \text{ starts service at } t + W(t^-) \text{ with service rate } \mu_i) = \frac{1}{3}, i = 0, 1, 2,$$

independent of t and $W(t^-)$. Assume $(W(t^-), M(t^-)) = (x, (2, 0, 0))$, $x > 0$, *just before* $C_{a,t}$ arrives. Then $C_{a,t}$ will wait a positive time x. Looking ahead to time $t + W(t^-)$, two other occupied servers will have service rates μ_0 ($m_0 = 2$, $m_1 = m_2 = 0$) when $C_{a,t}$ starts service at $t + W(t^-)$, in the just-vacated server. **Question**: What is the configuration $M(t)$ *just after* $C_{a,t}$ arrives? It can be either $(2, 0, 0)$, $(1, 1, 0)$, or $(1, 0, 1)$. The probabilities for $M(t)$ are:

$$\begin{aligned}
P(M(t) = (2,0,0)) &= P(\mu_t(x,(2,0,0)) = \mu_0) \cdot 1 \\
&\quad + P(\mu_t(x,(2,0,0)) = \mu_1) \cdot \tfrac{\mu_1}{2\mu_0+\mu_1} \\
&\quad + P(\mu_t(x,(2,0,0)) = \mu_2) \tfrac{\mu_2}{2\mu_0+\mu_2} \\
&= \tfrac{1}{3}\left(1 + \tfrac{\mu_1}{2\mu_0+\mu_1} + \tfrac{\mu_2}{2\mu_0+\mu_2}\right). \\
P(M(t) = (1,1,0)) &= P(\mu_t(x,(2,0,0)) = \mu_1) \cdot \tfrac{2\mu_0}{2\mu_0+\mu_1} \\
&= \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0+\mu_1}. \\
P(M(t) = (1,0,1)) &= P(\mu_t(x,(2,0,0)) = \mu_2) \cdot \tfrac{2\mu_0}{2\mu_0+\mu_2} \\
&= \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0+\mu_2}.
\end{aligned}$$

Thus

$$\begin{aligned}
&P(M(t) = (2,0,0)) + P(M(t) = (1,1,0)) + P(M(t) = (1,0,1)) \\
&= \tfrac{1}{3}\left(1 + \tfrac{\mu_1}{2\mu_0+\mu_1} + \tfrac{\mu_2}{2\mu_0+\mu_2}\right) + \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0+\mu_1} + \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0+\mu_2} = 1.
\end{aligned}$$

The resulting virtual wait at time t is

$$W(t) = W(t^-) + \mathcal{S}_t = x + \mathcal{S}_t,$$

where $\mathcal{S}_t$ is the inter start-of-service-depart time, distributed as a mixture

$$\mathcal{S}_t \underset{dis}{=} \begin{cases} \text{Exp}_{3\mu_0}, \\ \text{Exp}_{2\mu_0+\mu_1}, & \text{with probability } 1/3 \text{ each.} \\ \text{Exp}_{2\mu_0+\mu_2}, \end{cases}$$

(see Sect. 4.4.1). The sample path will have a jump whose size is distributed as $\mathcal{S}_t$ at instant t (see Fig. 4.2).

4.5.5 Sample Path of SP Process Revisited

We first describe a typical sample path of the virtual wait in Example 4.4 wherein $J = 1$ and $\boldsymbol{\mu} = \{\mu_0, \mu_1\}$, to facilitate exposition. If $J > 1$, sample-path construction would be similar, but with more lines and pages (sheets) in the product space $\boldsymbol{T} \times \boldsymbol{S}$ (see Fig. 4.2). Next, Example 4.5 discusses the general nature of a typical sample path with reference to Example 4.4 and then we outline the mechanics of a *specific* sample path in Example 4.4, based on the M/M/3 queue in Example 4.3 above.

Example 4.4 Consider M/M/c with $c = 3$, $J = 1$. (Here we take J= 1 for exposition.) A typical sample path of the virtual wait is given in Fig. 4.2).

Arrivals are assigned an exponential service rate from $\boldsymbol{\mu} = \{\mu_0, \mu_1\}$ with equal probability 1/2. (In general the probabilities can be, e.g., p_0, $p_1 = 1 - p_0$.) The total number of possible configurations is $\binom{J+c}{J+1} = \binom{4}{2} = 6$. The full set of configurations is

$$\boldsymbol{M} = \{00, 10, 01, 11, 20, 02\}.$$

We write (2, 0) as 20 when $m_0 = 2, m_1 = 0$, indicating that 2 servers are occupied with rate μ_0; and similar notation for the other system configurations.

Example 4.5 General Nature of Sample Path with reference to Example 4.4. The state space consists of: (1) six discrete points for the zero-wait states $(0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$. Thus $\boldsymbol{M}_0 = \{00, 10, 01\}$ and $\boldsymbol{M}_b = \boldsymbol{M}_1 = \{11, 20, 02\}$; (2) three intervals $((0, \infty), \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_1$. The three border states are $(0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_b$.

Arrival Waits Zero. Assume an arrival "sees" state $(0, m)$, $m \in \boldsymbol{M}_0$. The SP moves horizontally at time-rate 1 on a line $\boldsymbol{T} \times (0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_0$. If the next arrival occurs before a departure, the SP jumps to a line $\boldsymbol{T} \times (0, \boldsymbol{m}')$, $\boldsymbol{m}' \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$, where

$$m_0' + m_1' = m_0 + m_1 + 1,$$

because there is one more occupied server. If a departure occurs before an arrival, the SP jumps to a line $\boldsymbol{T} \times (0, \boldsymbol{m}'')$, $\boldsymbol{m}'' \in \boldsymbol{M}_0$, where

$$m_0'' + m_1'' = m_0 + m_1 - 1,$$

because there is one less occupied server. If $\boldsymbol{m} = (0, 0)$, the state can change only due to an arrival.

If an arrival finds the system to be in state $(0, m)$, $m \in \boldsymbol{M}_b$ the SP jumps to a sheet $\boldsymbol{T} \times ((0, \infty), \boldsymbol{k})$, $\boldsymbol{k} \in \boldsymbol{M}_1$. Configuration $\boldsymbol{k}$ is determined by the service rate assigned to the new arrival, and which server finishes first after the new arrival starts service. Denote the service time of an arrival $\mathrm{C}_{a,t}$ by s_t. Then $s_t \underset{dis}{=} \mathrm{Exp}_{\mu_t}$ where

$$P(\mu_t = \mu_i) = \frac{1}{2}, i = 0, 1;$$

(see Fig. 4.2).

To fix ideas, let the SP be on the border line $\boldsymbol{T} \times (0, 20)$ at arrival instant t. Thus $\left\langle W(t^-), M(t^-)\right\rangle = \langle 0, 20\rangle$. $\mathrm{C}_{a,t}$ starts service upon arrival in the one idle server and is assigned either rate μ_0 or μ_1 with probability 1/2 each. Let $\mathcal{S}_t$ denote the time from the start-of-service of $\mathrm{C}_{a,t}$ until the first departure from the system thereafter. (Since all 3 servers are busy $\mathcal{S}_t$ is independent of any future arrivals that join the waiting line.)

Case 1: Let us assume the service time s_t has been assigned rate μ_0. Then $\mathcal{S}_t \underset{dis}{=} \mathrm{Exp}_{3\mu_0}$ because $\mathcal{S}_t = \min\left\{3 \text{ i.i.d. } \mathrm{Exp}_{\mu_0}\mathrm{s}\right\}$. The SP jumps upward an amount $\mathcal{S}_t$. The virtual wait at time t is

$$W(t) = W(t^-) + \mathcal{S}_t = 0 + \mathcal{S}_t = \mathcal{S}_t.$$

At instant $t + W(t^-) + \mathcal{S}_t$ $(= t + \mathcal{S}_t)$, one of the three occupied servers completes service. The service rate of each of the resulting two occupied servers at $t + \mathcal{S}_t$ must be μ_0. By the **look-ahead process**, the configuration at t is $M(t) := M(t^-) = 20$. In this scenario the configuration remains the same as when the test customer arrived. Geometrically, at instant t, the look-ahead process impels the SP to jump from *line* 20 to *page* 20, at a height $\underset{dis}{=}$

$\text{Exp}_{3\mu_0}$ (see Fig. 4.2). In Case 1, the SP jumps from line 20 to page 20 at ordinate $\underset{dis}{=}$ $\text{Exp}_{2\mu_0+\mu_1}$, resulting in $W(t) \underset{dis}{=} \text{Exp}_{2\mu_0+\mu_1}$ and $M(t) = 20$.

Case 2. Let us assume that s_t has been assigned rate μ_1. Then $\mathcal{S}_t \underset{dis}{=} \text{Exp}_{2\mu_0+\mu_1}$. At $t + \mathcal{S}_t$ one of the three servers completes service. The service rates of the other two still-occupied servers at $t + \mathcal{S}_t$ are either: (a) both μ_0 with probability $\frac{\mu_1}{2\mu_0+\mu_1}$ (the rate-μ_1 server finishes first), or (b) μ_0 and μ_1 with probability $\frac{2\mu_0}{2\mu_0+\mu_1}$ (a rate-μ_0 server finishes first).

In Case 2(a) at instant t, the SP jumps from line 20 to page 20 at an ordinate $\underset{dis}{=}$ $\text{Exp}_{2\mu_0+\mu_1}$. Thus $W(t) \underset{dis}{=} \text{Exp}_{2\mu_0+\mu_1}$ and $M(t) = 20$. In Case 2(b) at instant t, the SP jumps from line 20 to page 11 at ordinate $\underset{dis}{=}$ $\text{Exp}_{2\mu_0+\mu_1}$, resulting in $W(t) \underset{dis}{=} \text{Exp}_{2\mu_0+\mu_1}$ and $M(t) = 11$.

Arrival Waits a Positive Time. Assume $C_{a,t}$ arrives when the state is $(x, 20)$, $x > 0$ (SP is at ordinate x on page 20). If the service-rate assignment policy assigns $s_t \underset{dis}{=} \text{Exp}_{\mu_0}$, the SP jumps upward an amount $\text{Exp}_{3\mu_0}$, and moves with slope -1 steadily on page 20. If the service-rate assignment policy assigns $s_t \underset{dis}{=} \text{Exp}_{\mu_1}$, the SP can end up on either page 20 or page 11 at t. The SP jumps upward to $W(t) \underset{dis}{=} W(t^-) + \text{Exp}_{2\mu_0+\mu_1}$ and moves with slope -1 steadily on page 20, with probability $\frac{\mu_1}{2\mu_0+\mu_1}$. The SP jumps upward to ordinate $W(t) \underset{dis}{=} W(t^-)+\text{Exp}_{2\mu_0+\mu_1}$ on page 11, with probability $\frac{2\mu_0}{2\mu_0+\mu_1}$.

If the SP descends to the bottom of page 20 and hits level 0 from above in a continuous manner before a new arrival occurs, it immediately enters border line 20, and continues its motion along line 20.

4.5.6 A Specific Sample Path

We expound further on a possible realization of the SP motion as it traces out the sample path, with reference to Fig. 4.2. Assume that initially the system is empty. The SP moves on line 00. Arrival 1 (C_1) sees an empty system. The server-assignment policy assigns C_1 service rate μ_0. The SP jumps to, and moves on, line 10. C_2 arrives before C_1 completes service and is also assigned rate μ_0. At C_2's arrival the SP jumps to line 20. C_3 arrives while both C_1 and

C_2 are in service. C_3 receives rate μ_1. The SP jumps to an ordinate $\text{Exp}_{2\mu_0+\mu_1}$, and if the rate-$\mu_1$ customer finishes first among the three customers in service, the resulting configuration is again 20. The probability of this event is $\frac{\mu_1}{2\mu_0+\mu_1}$, due to the memoryless property of exponential variates. This explains why at C_3's arrival instant the SP jumps to page 20.

Just before C_4 arrives the SP is descending at slope -1 on page 20. C_4 is assigned service rate μ_0. The SP jumps upward an amount $\text{Exp}_{3\mu_0}$. It remains on page 20. That is, whichever server finishes first, the two remaining active service rates will be μ_0, resulting in configuration 20. C_5 arrives when the SP is on page 20. C_5 is assigned rate μ_1. Suppose a server with rate μ_0 finishes first. The probability of this event is $\frac{2\mu_0}{2\mu_0+\mu_1}$. The SP jumps upward by $\text{Exp}_{2\mu_0+\mu_1}$. It simultaneously makes a $20 \to 11$ transition from page 20 to page 11, since the two remaining occupied servers have rates μ_0 and μ_1 when the first service ends. The configuration changes immediately from 20 to 11.

No new arrivals occur prior to the completion of the first rate-μ_0 customer. The SP descends on page 11 with slope -1 and hits level 0 from above, exactly when the first rate-μ_0 customer finishes service. The system now presents a zero wait to a potential arrival. When the SP hits level 0, it enters border line 11 (in Fig. 4.2 it jumps to line 11). C_6 arrives, and starts service immediately. C_6 is assigned rate μ_1. The SP jumps to page 02, with probability $\frac{\mu_0}{\mu_0+2\mu_1}$ (μ_0-rate service finishes first), the next configuration is 02.

The system continues to evolve. The SP continues to trace a sample path on the lines and pages according to the probability laws of the model. The sample path gives us a precise picture of the evolving system over time. Construction of the sample path goes hand in hand with understanding the model dynamics, and writing the model equations by inspection.

Remark 4.7 In Sect. 4.8 below we develop the **steady-state** theory. We will then return to Example 4.3, and formulate the balance equations for the zero-wait probabilities $P(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M} \equiv \boldsymbol{M}_0 \cup \boldsymbol{M}_b$; integral equations for the 'partial' pdfs of wait $f(x, m), x > 0, \boldsymbol{m} \in \boldsymbol{M}_1$, and for the total pdf $\{P_0, f(x), x > 0\}$.

4.5.7 SP Process Is Markovian

We outline a proof that the SP process is a Markov process. For $t \geq 0$, let $(x, \boldsymbol{m})_t :=$ *event* $\{(W(t), M(t)) = (x, \boldsymbol{m})\}$. It is required to show that for $x, y \geq 0, \boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}$,

$$P((y,\boldsymbol{k})_{t+h}|(x,\boldsymbol{m})_t,(W(u),M(u)_{0\le u<t}) \\ = P((y,\boldsymbol{k})_{t+h}|(x,\boldsymbol{m})_t),\ t\ge 0,\ h>0. \tag{4.9}$$

Formula (4.9) states that the probability of event $(y,\boldsymbol{k})_{t+h}$ given that event $(x,\boldsymbol{m})_t$ occurred, is independent of the history $(W(u),M(u))_{0\le u<t}$. We sketch the proof in two steps: (1) **zero-wait states**; (2) **positive-wait states**.

For a Poisson (or non-homogeneous Poisson) process, the probability of more than one event occurring in $(t,t+h)$ is $o(h)$ (e.g., Definition 5.3, p. 314, and Definition 5.4, p. 339 in [125]).

Zero-Wait Non-border States

Assume state $(0,\boldsymbol{m})_t \in \left\{(0,\boldsymbol{m})|0\le \sum_{j=0}^{J} m_j \le c-2\right\}$ ($\boldsymbol{m}\in \boldsymbol{M}_0$, SP$\in \boldsymbol{S}_0$ at time t).

No Departure or Arrival in $(t,t+h)$ The state remains $(0,\boldsymbol{m})$ in $(t,t+h)$ iff no arrival or departure occurs during $(t,t+h)$, or an event with probability $o(h)$ occurs. Thus

$$P((x,\boldsymbol{m})_{t+h}|(x,\boldsymbol{m})_t) = 1-\left(\lambda+\sum_{j=0}^{J} m_j\mu_j\right)h+o(h),$$

which is independent of $(W(u),M(u))_{0\le u<t}$.

Arrival in $(t,t+h)$ Possibly there is an arrival during $(t,t+h)$. The next configuration will have the form

$$\boldsymbol{m}_{L^+} := (m_0,\ldots,m_L+1,\ldots,m_J),$$

for some $L\in\{0,\ldots,J\}$. Then

$$P((0,\boldsymbol{m}_{L^+})_{t+h}|(0,\boldsymbol{m})_t) \\ = (\lambda h+o(h))\cdot P(\mu_t((0,\boldsymbol{m}))=\mu_L) \\ = \lambda h P(\mu_t((0,\boldsymbol{m}))=\mu_L)+o(h),\ L\in\{0,\ldots,J\}\cdot \tag{4.10}$$

Formula (4.10) is the probability that there is an arrival during $(t,t+h)$ assigned service rate μ_L, which is independent of the history given by $(W(u),M(u))_{0\le u<t}$. Note that

$$\sum_{L=0}^{J} P(\mu_t((0,\boldsymbol{m})=\mu_L)=1.$$

Departure in $(t,t+h)$ Possibly there is a departure during $(t,t+h)$. Let configuration

$$\boldsymbol{m}_{L^-} := (m_0, \ldots, \theta_L \cdot (m_L - 1), \ldots, m_J), L \in \{0, \ldots, J\},$$

where

$$\theta_L = \begin{cases} 1 \text{ if } m_L \geq 1, \\ 0 \text{ if } m_L = 0. \end{cases}$$

Assume $\boldsymbol{m} \neq (0, \ldots, 0)$. Then

$$P((0, \boldsymbol{m}_{L-})_{t+h} | (0, \boldsymbol{m})_t) = (m_L \cdot \mu_L)h + o(h), \tag{4.11}$$

which is the probability of a rate-μ_L departure during $(t, t+h)$ (*rate-μ_L service finishes first)*. Expression (4.11) is independent of the history $(W(u), M(u))$,$0 \leq u < t$. Note that $\left(\sum_{L=0}^{J} m_L \mu_L\right) h + o(h)$ is the probability of a departure during $(t, t+h)$.

Zero-Wait Border States

Consider zero-wait border states $\left\{(0, \boldsymbol{m})_t | \sum_{j=0}^{J} m_j = c - 1\right\}$ ($\boldsymbol{m} \in \boldsymbol{M}_b$, $(0, \boldsymbol{m}) \in \boldsymbol{S}_b$).

No Arrival in $(t, t+h)$ If no arrival or departure occurs, or only a departure occurs, during $(t, t+h)$, the Markov property follows similarly as for the zero-wait non-border states given above.

Arrival in $(t, t+h)$ Possibly there is an arrival during $(t, t+h)$. In *this* case, the SP jumps to a *positive level on a sheet (page)*. Let configuration

$$\begin{aligned} \boldsymbol{k} &:= (m_0, \ldots, m_L + 1, \ldots, m_R - 1, \ldots, m_J) \\ &= (k_0, \ldots, k_J), \end{aligned}$$

for some $L, R \in \{0, \ldots, J\}$. Thus $\sum_{j=0}^{J} k_j = \sum_{j=0}^{J} m_j = c - 1$. Let

$$\nu_L = \sum_{j=0}^{J} m_j \mu_j + \mu_L.$$

The probability that the SP jumps to sheet $\boldsymbol{k}$ during $(t, t+h)$ and is in state-space interval $((y, y+dy), \boldsymbol{k})_{y>0}$ at $t+h$, is

$$\begin{aligned} &P((W(t+h), M(t+h)) \in ((y, y+dy), \boldsymbol{k}) | (0, \boldsymbol{m})_t) \\ &= (\lambda h + o(h)) \cdot P(\mu_t(0, \boldsymbol{m}) = \mu_L) \cdot \frac{m_R \mu_R}{\nu_L} \cdot \nu_L \cdot e^{-\nu_L \cdot y} dy \\ &= \lambda h \cdot P(\mu_t(0, \boldsymbol{m}) = \mu_L) \cdot m_R \mu_R \cdot e^{-\nu_L y} dy + o(h), L \in \{0, \ldots, J\}, \end{aligned}$$

which is independent of the history $(W(u), M(u))_{0 \le u < t}$. The right side is the probability that there is an arrival in $(t, t+h)$, which is assigned service rate μ_L, and a rate-μ_R *service* finishes first among the occupied servers, at a time in the state space interval $(y, y+dy)$.

Positive-Wait States
Arrival In $(t, t+h)$ Given $(x, \boldsymbol{m})_t$, $x > 0$, where $\sum_{j=0}^{J} m_j = c - 1$, there may be an arrival during $(t, t+h)$. Let

$$\boldsymbol{k} = (m_0, \ldots, m_L + 1, \ldots, m_R - 1, \ldots, m_J).$$

Reasoning as for zero-wait border states, we obtain

$$\begin{aligned} &P((W(t+h), M(t+h) \in ((x+y, x+y+dy), \boldsymbol{k})|(x, \boldsymbol{m})_t)) \\ &= \lambda h \cdot P(\mu_t(0, \boldsymbol{m}) = \mu_L) \cdot m_R \mu_R \cdot e^{-\nu_L (y-x)} dy + o(h), \end{aligned}$$

which is independent of $(W(u), M(u))_{0 \le u < t}$.

Virtual Wait in $(0, h)$
Consider the case where all servers are occupied, no customers are waiting and $W(t) \in (0, h)$, where h is "small". Assume a server completes service before a new arrival occurs. Given $(x, \boldsymbol{m})_t$, $0 < x < h$, $\sum_{j=0}^{J} m_j = c - 1$, we obtain

$$P((0, \boldsymbol{m})_{t+h} | (x, \boldsymbol{m})_t) = 1 - \lambda x + o(x).$$

The SP hits level 0 from above in a continuous manner at $t + x$. It immediately enters border line $\boldsymbol{m}$ corresponding to the border state $(0, \boldsymbol{m})$, and continues its motion in the direction of Time. This is independent of the past history prior to t.

The above cases cover all possible situations. Formula (4.9) follows in each case, implying that the SP process has the Markov property.

4.5.8 Departures from Positive-Wait States

We examine the departure rates during a sojourn on a sheet (page). (Table 4.1 describes the symbols in Fig. 4.3.)

Suppose the SP is at a positive level on page $\boldsymbol{m} \in \boldsymbol{M}_1$ ($\sum_{j=0}^{J} m_j = c - 1$ and all c servers are occupied, including the last arrival). The occupancy number of service rate μ_j among the $c - 1$ servers, not occupied by the last arrival, is m_j, $j \in \{0, \ldots, J\}$.

The single remaining server, which *is occupied by the last arrival*, may have an arbitrary service rate $\mu^* \in \boldsymbol{\mu}$. Assume μ^* does not match a positive

Table 4.1 Description of Symbols in Fig. 4.3

Symbol	Description
τ_n	Arrival instant
C_{τ_n}	Customer that arrives at τ_n
σ_n	Start of service instant of C_{τ_n}
$\mathcal{S}_{\tau_n}$	$\sigma_{n+1} - \sigma_n$ = inter start-of-service depart time

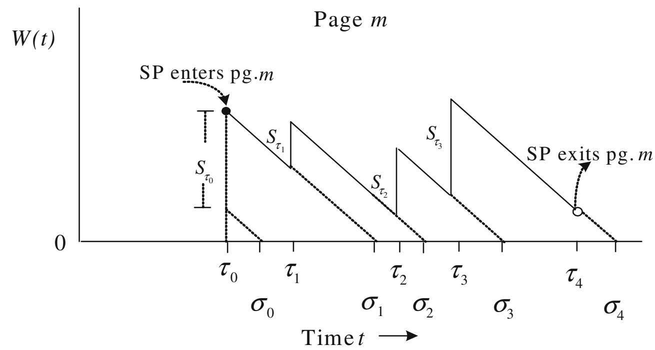


Fig. 4.3 SP sojourn on page m. Departure rate may differ on intervals (τ_0, σ_0), (σ_0, σ_1), (σ_1, σ_2), (σ_2, σ_3), (σ_3, τ_4). At instants σ_0, σ_1, σ_2, arrivals C_{τ_0}, C_{τ_1}, C_{τ_2} start service. Just after departure instants $\sigma_0 + \mathcal{S}_{\tau_0}$, $\sigma_1 + \mathcal{S}_{\tau_1}$, $\sigma_2 + \mathcal{S}_{\tau_2}$, the remaining $c - 1$ servers will have server occupancies $\boldsymbol{m} = (m_0, \ldots, m_J)$

component in configuration $\boldsymbol{m}$. In order for the SP to *remain* on page $\boldsymbol{m}$ just after that arrival, the rate-μ^* server must complete service first among the c occupied servers (see Fig. 4.3).

While the SP is on page $\boldsymbol{m}$, the system exponential departure rate will, in general, differ during inter-departure intervals. These possibly different exponential departure rates have no effect on the *Markov property* of the SP process. The configurations are determined *at arrival instants* (i.e., earlier when service rates are assigned) (Fig. 4.3).

4.6 Transient Analysis of Generalized M/M/c

Sections 4.6.1–4.6.6 develop LC relations and definitions leading to the formulation of integro-differential equations for the transient time-t pdf of the virtual wait, in Sect. 4.6.8 below. This development is based on the author's working papers [21, 23]. The transient analysis complements the results in [52], which focuses on the generalized M/M/c queue in steady-state. Section 4.6.7 derives the steady-state integral equations by letting $t \to \infty$ in the transient equations. Section 4.7 serves as a brief tutorial on writing steady-state model equations using LC and sample paths.

4.6.1 Transient PDF of Wait and Downcrossings

We next determine relationships between the transient pdf of wait and sample-path transitions. Let $\mathcal{D}_t(x, \boldsymbol{m})$:= number of sample-path downcrossings of level x on page $\boldsymbol{m} \in \boldsymbol{M}_1$ during $[0, t]$. Let

$$\mathcal{D}_t(x) = \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} \mathcal{D}_t(x, \boldsymbol{m})$$

denote the total number of downcrossings of level x on all pages during $[0, t]$. Theorem 4.3 connects the instantaneous rate of change of the expected number of downcrossings of level x in $[0, t]$, to the time-t transient pdf of wait at level x.

Theorem 4.3 For each configuration $\boldsymbol{m} \in \boldsymbol{M}_1$,

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(x, \boldsymbol{m})) = f_t(x, \boldsymbol{m}), x > 0, t > 0, \tag{4.12}$$

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(0, \boldsymbol{m})) = f_t(0^+, \boldsymbol{m}) \ (= f_t(0, \boldsymbol{m})), t > 0, \tag{4.13}$$

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = f_t(x), x > 0, t > 0, \tag{4.14}$$

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = f_t(0^+) \ (= f_t(0)), t > 0. \tag{4.15}$$

Proof Fix state-space level $x > 0$. Consider instants t and $t + h$, where $t > 0$, and $h > 0$ is small. To prove (4.12) and (4.13) for page $\boldsymbol{m}$, we develop a table similar to (3.10) in Chap. 3 for the M/G/1 queue, and proceed as in the proof of (3.8) and (3.9). Formulas (4.14) and (4.15) follow from the definitions of $\mathcal{D}_t(x)$ and the total pdf $f_t(x), x > 0$. ■

Corollary 4.1

$$E(\mathcal{D}_t(x,\boldsymbol{m})) = \int_{s=0}^{t} f_s(x,\boldsymbol{m})ds, \tag{4.16}$$

$$E(\mathcal{D}_t(0,\boldsymbol{m})) = \int_{s=0}^{t} f_s(0^+,\boldsymbol{m})ds, \tag{4.17}$$

$$E(\mathcal{D}_t(x)) = \int_{s=0}^{t} f_s(x)ds, \tag{4.18}$$

$$E(\mathcal{D}_t(0)) = \int_{s=0}^{t} f_s(0^+)ds. \tag{4.19}$$

Proof Integrating both sides of (4.12), (4.13), (4.14) and (4.15) with respect to s over the interval $[0, t]$ and applying the initial conditions

$$E(\mathcal{D}_0(x,\boldsymbol{m})) = E(\mathcal{D}_0(x)) = 0, x \geq 0,$$

yield (4.16), (4.17), (4.18) and (4.19), respectively. ■

4.6.2 Steady-State PDF of Wait and Downcrossings

Corollary 4.2 below connects the SP limiting downcrossing rate as $t \to \infty$ and the steady-state pdf of wait, at a state-space level. It is analogous to Corollary 3.2 for M/G/1. It also demonstrates the equality of the limit of the instantaneous rate of change of the expected number of downcrossings in $[0, t]$, and the limit of the average downcrossing rate over $[0, t]$.

Let $\boldsymbol{S}_{\boldsymbol{m}} = ([0,\infty),\boldsymbol{m})\,, \boldsymbol{m} \in \boldsymbol{M}_1$. The results below apply to each page $\boldsymbol{T} \times \boldsymbol{S}_{\boldsymbol{m}}, \boldsymbol{m} \in \boldsymbol{M}_1$ as well as to the "book" $\boldsymbol{T} \times \left(\cup_{\boldsymbol{m}\in\boldsymbol{M}_1} \boldsymbol{S}_{\boldsymbol{m}}\right)$.

Corollary 4.2 *Assume the following limits exist*

$$\lim_{t\to\infty} f_t(x,\boldsymbol{m}) \equiv f(x,\boldsymbol{m}), x \in \boldsymbol{S}_{\boldsymbol{m}}, \boldsymbol{m} \in \boldsymbol{M}_1.$$

Then

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x,\boldsymbol{m})) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x,\boldsymbol{m}))}{t} = f(x,\boldsymbol{m}), x > 0, \tag{4.20}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0,\boldsymbol{m})) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(0,\boldsymbol{m}))}{t} = f(0^+,\boldsymbol{m}) \equiv f(0,\boldsymbol{m}), \tag{4.21}$$

$$\lim_{t\to\infty}\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) = \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t} = f(x), x > 0, \tag{4.22}$$

$$\lim_{t\to\infty}\frac{\partial}{\partial t}E(\mathcal{D}_t(0)) = \lim_{t\to\infty}\frac{E(\mathcal{D}_t(0))}{t} = f(0^+) \equiv f(0). \tag{4.23}$$

Proof In (4.20), (4.21), (4.22) and (4.23), the equalities of the left-most terms to the pdfs on the right, follow by letting $t \to \infty$ in (4.12), (4.13), (4.14) and (4.15), respectively. The equalities of the middle terms to the pdfs on the right, follow by dividing both sides of (4.16), (4.17), (4.18) and (4.19) by $t > 0$ and letting $t \to \infty$. ∎

4.6.3 SP $\boldsymbol{m} \to \boldsymbol{k}$ Transitions

Before discussing the relationship between the transient pdf of wait and SP upcrossings, we define SP $\boldsymbol{m} \to \boldsymbol{k}$ transitions. We say that the SP makes an $\boldsymbol{m} \to \boldsymbol{k}$ transition at instant t_0 if it *exits* state-space set $\boldsymbol{S_m}$ and *enters* state-space set $\boldsymbol{S_k}$ at t_0. That is, the SP exits $([0,\infty), \boldsymbol{m})$ and enters $([0,\infty), \boldsymbol{k})$ at t_0. If $\boldsymbol{m} = \boldsymbol{k}$, then an $\boldsymbol{m} \to \boldsymbol{k}$ transition maintains the SP on page $\boldsymbol{m}$ at t_0. Similar remarks apply to zero-wait lines $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_0$, or line $\boldsymbol{m} \in \boldsymbol{M}_b$ and $\boldsymbol{S_k}$ (see Sects. 2.4.3, 2.4.4 for definitions of entrance and exit).

$\boldsymbol{m} \to \boldsymbol{k}$ Upcrossing of a Level Consider $\boldsymbol{S_m}, \boldsymbol{S_k}$. Fix level $x > 0$. An $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing of level x occurs at instant t_0 if the SP exits set $([0,x), \boldsymbol{m})$ and enters set $((x,\infty), \boldsymbol{k})$ at t_0. That is, the SP makes both an $\boldsymbol{m} \to \boldsymbol{k}$ transition and an upcrossing of level x at t_0. Thus the SP moves instantaneously (not in Time) from page $\boldsymbol{m}$ to page $\boldsymbol{k}$ and from a level below x to a level above x. Viewed from the "cover" of the "book", the upcrossing of level x resembles an "ordinary" upcrossing of x by a sample path of the virtual wait in the M/G/1 queue (see Fig. 4.2). Similar definitions apply to line $\boldsymbol{m}$ and $\boldsymbol{S_k}$ (page $\boldsymbol{k}$).

$\boldsymbol{m} \to \boldsymbol{k}$ Parallel Transition In some variants of the M/M/c queue, the SP may make "parallel" transitions. The SP makes an $\boldsymbol{m} \to \boldsymbol{k}$ *parallel* transition at t_0 if it exits $\boldsymbol{S_m}$ from a level y and enters $\boldsymbol{S_k}$ at the *same level* y, at t_0. SP parallel transitions can also occur in variants of *single-server* queues (e.g., queues with bulk service [20, 93]) and in other stochastic models. The concepts of system configuration, pages (sheets), cover, $\boldsymbol{m} \to \boldsymbol{k}$ transitions, etc., are useful in analyzing many other stochastic models.

4.6.4 SP m → K Upcrossings Viewed from "Cover"

Let

$$\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}), \boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$$

denote the number of SP $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings of level x during $[0, t]$. Denote the *total* number of upcrossings of level x during $[0, t]$ (as viewed from the "cover" of the "book") by

$$\mathcal{U}_t(x) = \sum_{\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1} \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}). \tag{4.24}$$

In (4.24) $\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$ will be positive only if $\boldsymbol{m}$, $\boldsymbol{k}$ are such that page $\boldsymbol{k}$ is accessible from page $\boldsymbol{m}$ in one step at an arrival instant (considering lines $\boldsymbol{m}$ and $\boldsymbol{k}$ as zero-levels of pages $\boldsymbol{m}$, $\boldsymbol{k}$ respectively). For an $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing of level x to occur, the "target" page $\boldsymbol{k}$ can be either page $\boldsymbol{m}$ itself ($\boldsymbol{k} = \boldsymbol{m}$) or a different page ($\boldsymbol{k} \neq \boldsymbol{m}$).

4.6.5 Number of Types of m → K Upcrossings

A *type* of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing is an ordered pair $(\boldsymbol{m}, \boldsymbol{k})$. The *total* number of possible types of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings depends on how many pages communicate in one step at arrival instants. An upper bound on the total number of possible $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings is

$$\begin{aligned} \textit{number of ordered pairs } (\boldsymbol{m}, \boldsymbol{k}) &= (\textit{number of configurations in } \boldsymbol{M}_1)^2 \\ &= \binom{J + c - 1}{c - 1}^2 = \binom{J + c - 1}{J}^2. \end{aligned}$$

This maximum number $\binom{J+c-1}{c-1}^2$ is realized only if *all* $\binom{J+c-1}{c-1}$ pages communicate in one step. In that case, there are $\binom{J+c-1}{c-1}$ ways to select the "source" page $\boldsymbol{m}$ and $\binom{J+c-1}{c-1}$ ways to select the "target" page $\boldsymbol{k}$ (with replacement).

Example 4.6 Consider an M/M/c queue with $c = 3$ and $J = 1$, as in Example 4.4 (see Fig. 4.2). The set of configurations corresponding to pages is $\boldsymbol{M}_1 = \{20, 11, 02\}$. Here $\binom{J+c-1}{c-1} = \binom{3}{2} = 3$. An upper bound on the number of types of $\boldsymbol{m} \to \boldsymbol{k}$ transitions (ordered pairs $(\boldsymbol{m}, \boldsymbol{k})$) is $3^2 = 9$. This maximum can be realized only if all configurations in $\boldsymbol{M}_1$ communicate with each other in one step. This will depend on the probabilities governing the evolution of the states over time. In the present example, configurations 20 and 02 do not communicate in one step (at an arrival instant). There are *seven* possible **types** of one-step transitions, namely,

$$\{20 \to 20, 20 \to 11, 11 \to 20, 11 \to 11, 11 \to 02, 02 \to 11, 02 \to 02\}.$$

Transition types $20 \to 02$ and $02 \to 20$ are not possible.

The Probability $p_t(z, \boldsymbol{m} \to \boldsymbol{k})$ We denote the probability that page $\boldsymbol{k}$ is accessible in one step from level z on page $\boldsymbol{m}$ at an arrival instant t, by $p_t(z, \boldsymbol{m} \to \boldsymbol{k})$. Thus for each $\boldsymbol{m} \in \boldsymbol{M}_1$

$$\sum_{\boldsymbol{k} \in \boldsymbol{M}_1} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) = 1.$$

Usually, for fixed z, there is some $\boldsymbol{k}$ for which $p_t(z, \boldsymbol{m} \to \boldsymbol{k}) = 0$. Then page $\boldsymbol{k}$ is not accessible from level z on page $\boldsymbol{m}$ in one step. If such inaccessibility applies for all $(z, \boldsymbol{m})$, $z \geq 0$, then page $\boldsymbol{k}$ is not accessible from page $\boldsymbol{m}$ in one step. This is the case in Example 4.6: for $\boldsymbol{m} = 20$ and $\boldsymbol{k} = 02$,

$$p_t(z, 20 \to 02) = p_t(z, 02 \to 20) = 0, z \geq 0;$$

so, pages $\boldsymbol{m}$ and $\boldsymbol{k}$ do not communicate in one step.

4.6.6 Transient PDF of Wait and Upcrossings

If a time-t arrival $C_{a,t}$ finds the state to be $(z, \boldsymbol{m})$, then $C_{a,t}$ is assigned a service rate $\mu_t(z, \boldsymbol{m}) \in \boldsymbol{\mu}$. We assume that $\mu_t(z, \boldsymbol{m})$ is a right continuous with respect to both z and t. In Theorem 4.4 we use the fact that $\boldsymbol{M}_1 = \boldsymbol{M}_b = \left\{\boldsymbol{m} | \sum_{j=0}^{J} m_j = c - 1\right\}$ (defined in Sect. 4.4.3).

Theorem 4.4 For $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$, the instantaneous rate of change of the expected number of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings in $[0, t]$ is given by

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) = \lambda \int_{z=0}^{x} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(z,\boldsymbol{m})(x-z)} dF_t(z, \boldsymbol{m}), x \geq 0, t \geq 0, \quad (4.25)$$

where

$$\nu_t(z, \boldsymbol{m}) = \sum_{j=0}^{J} m_j \mu_j + \mu_t(z, \boldsymbol{m}).$$

Proof Fix level $x > 0$ on page $\boldsymbol{m}$, and time $t > 0$. Examination of a sample path on page $\boldsymbol{m}$ over the time interval $(t, t+h)$, $h > 0$, leads to the non-zero values of $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$, and corresponding probabilities in (4.26) below. We omit $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}) = 0$, which contributes 0 to $E(\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}))$. We omit negative values, because $\{\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})\}_{t \geq 0}$ is a counting process implying $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}) \geq 0$.

$\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k})$ $-\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$	Probability
$+1$	$\lambda h P_0(t) p_t(0, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(0,\boldsymbol{m})x}$ $+\lambda h \int_h^x p_t(z, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(z,\boldsymbol{m})(x-z)} f_t(z) dz + o(h)$
≥ 2	$o(h)$.

(4.26)

In (4.26), taking the expected value of $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$, dividing by $h > 0$ and letting $h \downarrow 0$, yields

$$\begin{aligned}\frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) = \lambda \cdot P_0(t) \cdot p_t(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_t(0,\boldsymbol{m})x} \\ + \lambda \int_{z=0}^{x} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_t(z,\boldsymbol{m})(x-z) \cdot} f_t(z) dz \\ = \lambda \int_{z=0}^{x} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_t(z,\boldsymbol{m})(x-z)} \cdot dF_t(z, \boldsymbol{m}),\end{aligned} \tag{4.27}$$

which is the same as (4.25). ■

Corollary 4.3 *For* $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$,

$$\begin{aligned}&E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) \\ &= \lambda \int_{s=0}^{t} \int_{z=0}^{x} p_s(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_s(z,\boldsymbol{m})(x-z)} \cdot dF_s(z, \boldsymbol{m}) ds, \ x \geq 0, t \geq 0.\end{aligned} \tag{4.28}$$

Proof In (4.25) change the variable from t to s on both sides, integrate with respect to s over the interval $[0, t]$, and apply the initial condition $E(\mathcal{U}_0(x, \boldsymbol{m}, \boldsymbol{k})) = 0$. This yields (4.28). ■

Corollary 4.4 *Consider the "cover". For $x \geq 0, t \geq 0$,*

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(x)) = \lambda \sum_{\boldsymbol{m},\boldsymbol{k} \in \boldsymbol{M}_1} \int_{z=0}^{x} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_t(z,\boldsymbol{m})(x-z)} \cdot dF_t(z, \boldsymbol{m}) \tag{4.29}$$

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(0)) = \lambda \sum_{\boldsymbol{m},\boldsymbol{k} \in \boldsymbol{M}_1} p_t(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot F_t(0, \boldsymbol{m}). \tag{4.30}$$

Proof We define $\mathcal{U}_t(x), x \geq 0$ in (4.24). Equations (4.29) and (4.30) follow by setting $x > 0$, and $x = 0$, respectively, in (4.27), and applying (4.24). (The sample path viewed from the cover is the projection of the sample-path segments from all pages onto a single sheet.) ■

Corollary 4.5 *For $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$ and $x \geq 0, t \geq 0$,*

$$E(\mathcal{U}_t(x)) = \lambda \sum_{\boldsymbol{m},\boldsymbol{k}} \int_{s=0}^{t} \int_{z=0}^{x} p_s(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu_s(z,\boldsymbol{m})(x-z)} \cdot dF_s(z, \boldsymbol{m}) ds,$$

$$E(\mathcal{U}_t(0)) = \lambda \sum_{\boldsymbol{m},\boldsymbol{k}} \int_{s=0}^{t} p_s(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot F_s(0, \boldsymbol{m}) ds.$$

Proof In (4.29) and (4.30) change t to s and integrate with respect to s on $[0, t]$. Then apply the initial condition $\mathcal{U}_0(x) = 0, x \geq 0$. ■

4.6.7 Steady-State PDF of Wait and Upcrossings

Corollary 4.6 below proves

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}))}{t},$$

by relating both limits to the steady-state pdf of wait. Let

$$p(z, \boldsymbol{m} \to \boldsymbol{k}),\ \nu(z, \boldsymbol{m}),\ F(z, \boldsymbol{m}),\ \text{and } f(z, \boldsymbol{m})$$

be the limiting values of

$$p_t(z, \boldsymbol{m} \to \boldsymbol{k}),\ \nu_t(z, \boldsymbol{m}),\ F_t(z, \boldsymbol{m}),\ f_t(z, \boldsymbol{m}),$$

respectively, as $t \to \infty$ (for definition of: $p_t(z, \boldsymbol{m} \to \boldsymbol{k})$ see Sect. 4.6.5; $\nu_t(z, \boldsymbol{m})$ see formula 4.25).

Corollary 4.6 *For* $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$ *and* $x \geq 0$*,*

$$\begin{aligned}
\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) &= \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}))}{t} \\
&= \lambda \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z, \boldsymbol{m}) \\
&= \lambda p(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(0,\boldsymbol{m})x} P_0 \\
&\quad + \lambda \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot f(z, \boldsymbol{m}) dz. \qquad (4.31)
\end{aligned}$$

Proof The equality

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) = \lambda \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z, \boldsymbol{m}),$$

follows by letting $t \to \infty$ on both sides of (4.25). The equality

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}))}{t} = \lambda \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z, \boldsymbol{m})$$

is obtained upon dividing both sides of (4.28) by $t > 0$, letting $t \to \infty$, and using L'Hôpital's rule (e.g., Theorem 9, p. 179 in [137]; and many Calculus texts). Equation (4.31) then follows. ∎

The next corollary relates the limits

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t},$$

for the expected *total number* of upcrossings in $[0, t]$, to the steady-state *total* probability distribution of wait.

Corollary 4.7 *For* $x \geq 0$*,*

$$\begin{aligned}
&\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} \\
&= \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z, \boldsymbol{m}) \\
&= \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1} p(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(0,\boldsymbol{m})x} P_{0,\boldsymbol{m}} \\
&\quad + \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot f(z, \boldsymbol{m}) dz. \qquad (4.32)
\end{aligned}$$

Proof The result (4.32) follows from (4.31) and the definition of $\mathcal{U}_t(x)$ in (4.24). ■

4.6.8 Equations for Transient PDF of Wait

We derive the transient model equations for the generalized M/M/c model. These equations comprise: (1) $\binom{J+c-1}{c-1}$ integro-differential equations for the partial pdfs $f_t(x, \boldsymbol{m}), x > 0, \boldsymbol{m} \in \boldsymbol{M}_1$; (2) $\binom{J+c-1}{c-1}$ differential equations for the zero-wait probabilities $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1 (= \boldsymbol{M}_b)$; (3) $\binom{J+c-1}{c-2}$ differential equations for the zero-wait probabilities $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0$; (4) one equation for the normalizing condition. Also $\boldsymbol{M}_0 = \left\{\boldsymbol{m}|0 \le \sum_{i=0}^{J} m_j \le c-2\right\}$ (see definition in formula 4.8).

We also derive the model equations for the *total* transient mixed pdf of wait $\{P_0(t), f_t(x)\}_{x>0,t\ge 0}$ (cover of book—see definition in formula 4.5.3).

Formula (4.1) and especially (4.2) of Theorem B (Sect. 4.2.1) play important roles in these derivations. In Theorem B we take the set $\boldsymbol{A}$ to be an interval in the state space having one of its boundaries equal to x.

Equations for Partial Transient PDFs of Wait

Before stating Theorem 4.5, we introduce/review some definitions.

Definition 4.5 page $\boldsymbol{i} := T \times ((0, \infty), \boldsymbol{i})$ where system configuration $\boldsymbol{i} \in \boldsymbol{M}_1$ (=$\boldsymbol{M}_b$)—technically page $\boldsymbol{i}$ excludes line $\boldsymbol{i} := T \times (0, (0, \boldsymbol{i}))$, which may be separately depicted, or appended to the bottom of page $\boldsymbol{i}$, in geometric figures (see, e.g., Fig. 4.2); $J_t^{(0,x)}(\boldsymbol{k}, \boldsymbol{m})$:= *number of* $k \to m$ *jumps that* ***start*** *in state set* $((0, x), \boldsymbol{k})$ *during* $[0, t]$ and end in $((0, \infty), \boldsymbol{m})$; $J_t^0(\boldsymbol{k}, \boldsymbol{m})$:= *number of* $k \to m$ *jumps that* ***start*** *in state set* $(0, \boldsymbol{k})$ *during* $[0, t]$ and end in $((0, \infty), \boldsymbol{m})$; $\mathcal{U}_t(x, \boldsymbol{k}, \boldsymbol{m})$:= *number of SP* $\boldsymbol{k} \to \boldsymbol{m}$ *transitions that start in* ([*0,x*) , k) and ***jump-upcross*** *level x during* $[0, t]$ (start in $((0, x), \boldsymbol{k})$ or in $(0, \boldsymbol{k})$ during $[0, t]$ and end in $((x, \infty), \boldsymbol{m})$).

Theorem 4.5 (1) The integro-differential equations for $f_t(x, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$, are

$$\begin{aligned} f_t(x, \boldsymbol{m}) &+ \lambda \sum_{\boldsymbol{k} \ne \boldsymbol{m}} \int_{z=0}^{x} p_t(z, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu_t(z,\boldsymbol{m})(x-z)}) f_t(z, \boldsymbol{k}) dz \\ &+ \lambda \sum_{\boldsymbol{k} \ne \boldsymbol{m}} p_t(0, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu_t(0,\boldsymbol{k})(x-z)}) P_t(0, \boldsymbol{k}) \\ &= \frac{\partial}{\partial t} F_t(x, \boldsymbol{m}) - \frac{\partial}{\partial t} P_t(0, \boldsymbol{m}) + f_t(0, \boldsymbol{m}) \end{aligned} \quad (4.33)$$

$$+\lambda \int_{z=0}^{x} p_t(z, \boldsymbol{m}, \boldsymbol{m}) e^{-\nu_t(z,\boldsymbol{m})(x-z))} f_t(z, \boldsymbol{m}) dz$$

$$+\lambda \sum_{\boldsymbol{k} \neq \boldsymbol{m}} \int_{z=0}^{x} p_t(z, \boldsymbol{m}, \boldsymbol{k}) f_t(z, \boldsymbol{m}) dz, x \geq 0, t \geq 0,$$

where configuration $\boldsymbol{k} \in \boldsymbol{M}_1$.
(2) The differential equation for $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$, is

$$f_t(0, \boldsymbol{m}) + \lambda \sum_{\boldsymbol{k}} p_t(0, \boldsymbol{k}, \boldsymbol{m}) P_t(0, \boldsymbol{k})$$

$$= \frac{\partial}{\partial t} P_t(0, \boldsymbol{m}) + \left(\lambda + \sum_{j=0}^{J} m_j \mu_j \right) P_t(0, \boldsymbol{m}) \tag{4.34}$$

where $\boldsymbol{k}$ is such that $\sum_{j=0}^{J} k_j = c - 2$.
(3) The differential equations for $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0$, are

$$\lambda \sum_{\boldsymbol{r} \neq \boldsymbol{m}} p_t(0, \boldsymbol{r}, \boldsymbol{m}) P_t(0, \boldsymbol{r}) + \sum_{\boldsymbol{s} \neq \boldsymbol{m}} s_j \mu_j p_t(0, \boldsymbol{s}, \boldsymbol{m}) P_t(0, \boldsymbol{s})$$

$$= \frac{\partial}{\partial t} P_t(0, \boldsymbol{m}) + \left(\lambda + \sum_{j=0}^{J} m_j \mu_j \right) P_t(0, \boldsymbol{m}), \tag{4.35}$$

where state $(0, \boldsymbol{m})$ is accessible in one step from state $(0, \boldsymbol{r})$ at an arrival instant, and in one step from $(0, \boldsymbol{s})$ at a departure instant. That is,

$$\sum_{j=0}^{J} m_j = \sum_{j=0}^{J} r_j + 1 = \sum_{j=0}^{J} s_j - 1.$$

(4) The normalizing condition is

$$\sum_{\boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_1} P_t(0, \boldsymbol{m}) + \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} \int_{x=0}^{\infty} f_t(x, \boldsymbol{m}) dx = 1. \tag{4.36}$$

Proof **(1)** We derive (4.33) by applying Theorem B (Sect. 4.2.1).
Choose $\boldsymbol{A}$. In (4.1) and (4.2), choose $\boldsymbol{A} := ((0, x), \boldsymbol{m})$ (i.e., $\boldsymbol{A}$ is open interval $(0, x)$ on page $\boldsymbol{m}$). The measure of set $\boldsymbol{A}$ at time t is

$$P_t(\boldsymbol{A}) = F_t(x, \boldsymbol{m}) - F_t(0, \boldsymbol{m}) = F_t(x, \boldsymbol{m}) - P_t(0, \boldsymbol{m}).$$

Entrance rate into $\boldsymbol{A}$. The SP can *enter* $\boldsymbol{A}$ by: (i) downcrossing level x on page $\boldsymbol{m}$; (ii) making a $\boldsymbol{k} \to \boldsymbol{m}$ ($\boldsymbol{k} \neq \boldsymbol{m}$) upward jump starting in $((0, x), \boldsymbol{k})$, *that* ***ends in*** $((0, x), \boldsymbol{m})$; (iii) making a jump that *starts from state (0,k)* ($\boldsymbol{k} \in \boldsymbol{M}_1$) (sometimes located *at level* 0 on page $\boldsymbol{k}$ in figures), and *ends* ***in*** $((0, x), \boldsymbol{m})$.

The number of SP entrances into set $\boldsymbol{A}$ during $[0, t]$ is

$$\mathcal{I}_t(\boldsymbol{A}) = \mathcal{D}_t(x, \boldsymbol{m}) + \sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} J_t^{(0,x)}(\boldsymbol{k}, \boldsymbol{m}) + \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} J_t^0(\boldsymbol{k}, \boldsymbol{m}) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t(x, \boldsymbol{k}, \boldsymbol{m}). \tag{4.37}$$

In (4.37) the algebraic sum

$$\sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} J^{(0,t)}(\boldsymbol{k}, \boldsymbol{m}) + \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} J_t^0(\boldsymbol{k}, \boldsymbol{m}) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t(x, \boldsymbol{k}, \boldsymbol{m}) \tag{4.38}$$

= (number of SP jumps that start in $([0, x), \boldsymbol{k})$ on any pages or zero-wait lines $k \in \boldsymbol{M}_1$, and end in $((0, \infty), \boldsymbol{m})$ – (number of such jumps that end in $((x, \infty), \boldsymbol{m})$ on page m during $[0, t]$)). Thus, (4.38) is the number of SP entrances into $((0, x), \boldsymbol{m})$ during $[0, t]$, due to jumps that start below x on pages or lines outside of $\boldsymbol{T} \times ((0, x), \boldsymbol{m})$ and end in $((0, x), \boldsymbol{m})$. Therefore $\mathcal{I}_t(\boldsymbol{A})$ is the *total* number of SP entrances into $((0, x), \boldsymbol{m})$ from all sources in one step during $[0, t]$.

Taking expected values and then $\frac{\partial}{\partial t}$ in (4.37) yields

$$\frac{\partial}{\partial t} E\left(\mathcal{I}_t(A)\right) = \frac{\partial}{\partial t} E\left(\mathcal{D}_t(x, \boldsymbol{m})\right) + \sum_{\boldsymbol{k}\neq\boldsymbol{m}} \frac{\partial}{\partial t} E\left(J_t^{(0,x)}(\boldsymbol{k}, \boldsymbol{m})\right) + \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \frac{\partial}{\partial t} E\left(J_t^0(\boldsymbol{k}, \boldsymbol{m})\right) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \frac{\partial}{\partial t} E\left(\mathcal{U}_t(x, \boldsymbol{k}, \boldsymbol{m})\right). \tag{4.39}$$

Exit Rate of set $\boldsymbol{A}$. The SP can *exit* set $\boldsymbol{A}$ by: (i) hitting level 0 on page $\boldsymbol{m}$ from above in a continuous fashion, (i.e., exiting $((0, x), \boldsymbol{m})$ and simultaneously entering state $(0, \boldsymbol{m})$); (ii) starting in $((0, x), \boldsymbol{m})$ at an arrival instant and making an $\boldsymbol{m} \to \boldsymbol{k}$ (including $\boldsymbol{m} \to \boldsymbol{m}$) *upcrossing* of level x, ending in $((x, \infty), \boldsymbol{k})$, $\boldsymbol{k} \in \boldsymbol{M}_1$; (iii) starting in $((0, x), \boldsymbol{m})$ at an arrival instant instant and making an $\boldsymbol{m} \to \boldsymbol{k}$ ($\boldsymbol{k} \neq \boldsymbol{m}$) jump-transition that ends *below* x on any page $\boldsymbol{k} \neq \boldsymbol{m}$, i.e., in $((0, x), \boldsymbol{k})$, $\boldsymbol{k} \in \boldsymbol{M}_1$, $\boldsymbol{k} \neq \boldsymbol{m}$

The total number of exits from set $\boldsymbol{A}$ during $[0, t]$ is

$$\mathcal{O}_t(\boldsymbol{A}) = \mathcal{D}_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}) + \sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} J_t^{(0,x)}(\boldsymbol{m},\boldsymbol{k}) - \sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}). \tag{4.40}$$

Explanation of (4.40). On the right side, $\mathcal{D}_t(0,m)$ is the number of exits from $\boldsymbol{A}$ during $[0,t]$ by downcrossing level 0 on page $\boldsymbol{m}$ (entering $(0,\boldsymbol{m})$). The term $\sum_{\boldsymbol{k}\in\boldsymbol{M}_1}\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$ is the number of SP jump exits from $([0,x),\boldsymbol{m})$ during $[0,t]$ that *upcross* level x on any page $\boldsymbol{k}$ (including $\boldsymbol{k}=\boldsymbol{m}$). $\sum_{\boldsymbol{k}\neq m\in\boldsymbol{M}_1} J_t^{(0,x)}(\boldsymbol{m},\boldsymbol{k})$ is the number jump-exits that start in $((0,x),\boldsymbol{m})$ and end in $((0,\infty),\boldsymbol{k})$ for any $\boldsymbol{k}\neq\boldsymbol{m}$. Term $-\sum_{\boldsymbol{k}\neq m\in\boldsymbol{M}_1}\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$ cancels the extra number of jump-exits from $([0,x),\boldsymbol{m})$ during $[0,t]$ that *upcross* level x on any page $\boldsymbol{k}\neq\boldsymbol{m}$, i.e., ending in $((x,\infty),\boldsymbol{k})$.

Taking expected values and then $\frac{\partial}{\partial t}$ in (4.40) results in

$$\begin{aligned}&\frac{\partial}{\partial t}E\left(\mathcal{O}_t(\boldsymbol{A})\right)\\&=\frac{\partial}{\partial t}E\left(\mathcal{D}_t(0,m)\right)+\sum_{\boldsymbol{k}\in\boldsymbol{M}_1}\frac{\partial}{\partial t}E\left(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})\right)\\&\quad+\sum_{\boldsymbol{k}\neq m}\frac{\partial}{\partial t}E\left(\mathcal{U}_t^{(0,x)}(x,\boldsymbol{m},\boldsymbol{k})\right)-\sum_{\boldsymbol{k}\neq m}\frac{\partial}{\partial t}E\left(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})\right).\end{aligned} \tag{4.41}$$

Integro-differential Equation: We substitute in (4.41) from (4.12), (4.13), (4.25). This yields the integro-differential equation (4.33).

(2) We derive (4.34) by letting set $\boldsymbol{A}=(0,\boldsymbol{m})$ in Theorem B, and substituting formulas from Section 4.6.1 relating downcrossings and the transient distribution of wait, as in the proof of (1).

(3) We derive (4.35) in a similar manner as in (2).

(4) The final equation is the normalizing condition

$$\sum_{\boldsymbol{m}\in\boldsymbol{M}_0\cup\boldsymbol{M}_1} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in\boldsymbol{M}_1}\int_{x=0}^{\infty} f_t(x,\boldsymbol{m})dx = 1.$$

■

Remark 4.8 In practice we can derive an equivalent set of model equations by letting set $\boldsymbol{A}=((x,\infty),\boldsymbol{m})$, $x>0$, in Theorem B (instead of substituting $((0,x],\boldsymbol{m})$). This choice of $\boldsymbol{A}$ may simplify the derivation of the model equations for $f_t(x,\boldsymbol{m})$. We would then consider SP jumps that start below and end above level x. This would yield terms of the form $e^{-\nu_t(z,\boldsymbol{m})(x-z)}$

rather than $(1 - e^{-\nu_t(z,\boldsymbol{m})(x-z)})$ in the integrands. In real-world applications, writing the integro-differential equations is much simpler than it may seem at this point. Some practice on a few simple models will quickly establish the method. It is very intuitive.

Remark 4.9 We can generalize the model upon replacing λ by λ_t, depending on t. The arrival stream would then be a non-homogeneous Poisson process. This generalization holds because the developments in the foregoing sections involving λ are essentially the same if λ_t is substituted for λ.

Model Equations for Total Transient PDF

In the following theorem, we utilize the previously defined equivalent notation $F_t(0,\boldsymbol{m}) \equiv P_t(0,\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1,\ F_t(0) \equiv P_0(t),\ f_t(0) \equiv f_t(0^+)$.

Theorem 4.6 For the total pdf of wait $\{P_0(t),\ f_t(x)\}_{x>0}$, as viewed from the 'cover', the following integro-differential and differential equations hold :

$$\begin{aligned} f_t(x) =& \frac{\partial}{\partial t}F_t(x) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu_t(z,\boldsymbol{m})(x-z)} dF_t(z,\boldsymbol{m}) \\ =& \frac{\partial}{\partial t}F_t(x) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P_t(0,\boldsymbol{m}) e^{-\nu_t(z,\boldsymbol{m})x} \\ &+ \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu_t(z,\boldsymbol{m})(x-z)} f_t(z,\boldsymbol{m}) dz,\ x > 0,\ t \ge 0, \end{aligned} \quad (4.42)$$

$$f_t(0) = \frac{\partial}{\partial t}P_0(t) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P_t(0,\boldsymbol{m}),\ t \ge 0. \quad (4.43)$$

Proof In Theorem B (Sect. 4.2.1), consider the set

$$\boldsymbol{A} = \left(\cup_{\boldsymbol{m}\in\boldsymbol{M}_0\cup\boldsymbol{M}_1}(0,\boldsymbol{m})\right) \cup \left(\cup_{\boldsymbol{m}\in\boldsymbol{M}_1}((0,x]\,,\boldsymbol{m}),\ x > 0\right).$$

Set $\boldsymbol{A}$ includes all $\binom{J+c}{c-1}$ zero-wait states $\left\{(0,\boldsymbol{m}) | 0 \le \sum_{j=0}^{J} m_j \le c-1\right\}$, as well as all positive-wait states $\left\{(y,\boldsymbol{m}) | \sum_{j=0}^{J} m_j = c-1,\ y \in (0,x]\right\}$.

Every SP *entrance* into $\boldsymbol{A}$ must occur from above at level x. Therefore all entrances are due to (continuous) SP downcrossings of level x. Every *exit* out of $\boldsymbol{A}$ must be due to a jump starting below level x on a page and ending at a level above level x on some page. Therefore all SP exits from set $\boldsymbol{A}$ are due to upcrossings of level x.

Thus

$$\mathcal{I}_t(A) = \mathcal{D}_t(x),\ \mathcal{O}_t(A) = \mathcal{U}_t(x),$$
$$E\left(\mathcal{I}_t(A)\right) = E\left(\mathcal{D}_t(x)\right),\ E\left(\mathcal{O}_t(A)\right) = E\left(\mathcal{U}_t(x)\right),$$
$$\tfrac{\partial}{\partial t}E\left(\mathcal{I}_t(A)\right) = \tfrac{\partial}{\partial t}E\left(\mathcal{D}_t(x)\right),\ \tfrac{\partial}{\partial t}E\left(\mathcal{O}_t(A)\right) = \tfrac{\partial}{\partial t}E\left(\mathcal{U}_t(x)\right).$$

We then substitute these expressions into formulas (4.14), (4.15), (4.29) and (4.30). This substitution yields the integro-differential equation (4.42) and the differential equation (4.43). ■

The normalizing condition

$$P_0(t) + \int_{x=0}^{\infty} f_t(x)dx = 1,$$

is used along with (4.42), (4.43) to solve for the unknown time-t zero-wait probabilities and positive-wait pdfs.

When it is not feasible to obtain an analytical solution, we can use numerical, simulation or approximation techniques to solve for the transient zero-wait probabilities and positive-wait pdfs.

4.6.9 Equations for Steady-State PDF of Wait

We obtain the model equations for the steady-state pdf of wait by letting $t \to \infty$ in (4.34)–(4.36). All quantities subscripted by t have limits as $t \to \infty$. We denote the limits utilizing the same notation, omitting subscript t. If stability holds, then

$$\lim_{t\to\infty} \frac{\partial}{\partial t} F_t(x, \boldsymbol{m}) = \lim_{t\to\infty} \frac{\partial}{\partial t} F_t(0, \boldsymbol{m}) = 0.$$

This corresponds to the cdf $F(x, \boldsymbol{m})$ being independent of t.

Theorem 4.7 The integral equation for the steady-state pdf $f(x, m)$, $\boldsymbol{m} \in \boldsymbol{M}_1$, is

$$\begin{aligned}
&f(x, \boldsymbol{m}) + \lambda \textstyle\sum_{\boldsymbol{k}\neq \boldsymbol{m}\in \boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z, \boldsymbol{k})dz \\
&\quad + \lambda \textstyle\sum_{\boldsymbol{k}\in \boldsymbol{M}_1} p(0, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu(0,\boldsymbol{k})(x-z)}) P(0, \boldsymbol{k}) \\
&= f(0, \boldsymbol{m}) \\
&\quad + \lambda \textstyle\int_{z=0}^{x} p(z, \boldsymbol{m}, \boldsymbol{m}) e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z, \boldsymbol{m})dz \\
&\quad + \lambda \textstyle\sum_{\boldsymbol{k}\neq \boldsymbol{m}\in \boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{m}, \boldsymbol{k}) f(z, \boldsymbol{m})dz,\ x \geq 0.
\end{aligned} \tag{4.44}$$

Proof We obtain (4.44) by letting $t \to \infty$ in (4.33). ■

Theorem 4.8 The model equation for the total steady-state pdf is

$$f(x) = \lambda \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} P(0, \boldsymbol{m}) e^{-\nu(z,\boldsymbol{m})x} + \lambda \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu(z,\boldsymbol{m})(x-z)} f(z, \boldsymbol{m}) dz, \; x > 0. \tag{4.45}$$

Proof Let $t \to \infty$ in (4.42). ■

Remark 4.10 In practice, it is often more efficient to derive balance equations for SP exit/entrance rates with respect to the state-space sets $((x, \infty), \boldsymbol{m})_{\boldsymbol{m} \in \boldsymbol{M}_1}$, $x > 0$, rather than with respect to the state-space sets $((0, x), \boldsymbol{m})_{\boldsymbol{m} \in \boldsymbol{M}_1}$, $x > 0$. The derived equations will be equivalent, no matter which state-space sets are employed for rate balance.

Interpretation of Equations in Theorem 4.7 for Sheets

We now interpret (4.44) in terms of rate balance across levels and between pages. This interpretation gives LC power for deriving steady-state model equations by inspecting a typical sample path, in a vast array of complex stochastic models.

In (4.44) the *left* side is the SP *entrance rate* into $((0, x), \boldsymbol{m})$. The term $f(x, \boldsymbol{m})$ is the SP downcrossing rate of level x on page $\boldsymbol{m}$. The term

$$\lambda \sum_{\boldsymbol{k} \neq \boldsymbol{m} \in \boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z, \boldsymbol{k}) dz$$

is the rate at which the SP enters composite state $((0, x), \boldsymbol{m})$ due to jumps at arrival instants that originate in $((0, x), \boldsymbol{k})$ on pages $\boldsymbol{k} \neq \boldsymbol{m}$. The term

$$\lambda \sum_{\boldsymbol{k} \in \boldsymbol{M}_1} p(0, \boldsymbol{k}, \boldsymbol{m})(1 - e^{-\nu(0,\boldsymbol{k})(x-z)}) P(0, \boldsymbol{k})$$

is the rate at which the SP enters composite state $((0, x), \boldsymbol{m})$ due to jumps that originate at level $(0, \boldsymbol{k})$ on any zero-wait line $\boldsymbol{k} \in \boldsymbol{M}_1$). These three terms exhaust the possible paths by which the SP can enter $((0, x), \boldsymbol{m})$.

The *right* side of (4.44) is the SP *exit rate* of $((0, x), \boldsymbol{m})$. The term $f(0, \boldsymbol{m})$ is the rate at which the SP exits $((0, x), \boldsymbol{m})$ and simultaneously enters the zero-wait boundary state $(0, \boldsymbol{m})$, due to downcrossings of level 0. The term

$$\lambda \int_{z=0}^{x} p(z, \boldsymbol{m}, \boldsymbol{m}) e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z, \boldsymbol{m}) dz$$

is the rate at which the SP exits $((0, x), \boldsymbol{m})$ and simultaneously enters $([x, \infty), \boldsymbol{m})$ due to jumps at arrival instants. The term

$$\lambda \sum_{\boldsymbol{k} \neq \boldsymbol{m}} \int_{z=0}^{x} p(z, \boldsymbol{m}, \boldsymbol{k}) f(z, \boldsymbol{m}) dz$$

is the rate at which the SP exits $((0, x), \boldsymbol{m})$ and simultaneously enters any page $\boldsymbol{k} \neq \boldsymbol{m}$. These three terms exhaust the possible paths by which the SP can exit $((0, x), \boldsymbol{m})$.

Thus equation (4.44) is a rate-balance equation of the form:

$$\boldsymbol{Rate\ \ into}\ \ ((0, x), \boldsymbol{m}) = \boldsymbol{Rate\ \ out\ \ of}\ \ ((0, x), \boldsymbol{m}),$$

which is a well-known principle for stochastic processes with discrete states, e.g., birth-death processes.

Interpretation of Equation for Total PDF

We now provide an LC interpretation of (4.45). We may view the LC analysis of the sheets as a *dissection* of the states of the model (into a partition). The total equation is like a *synthesis* , i.e., reconstruction of the parts into a single whole. This idea helps to derive model equations in complex models directly from sample-path considerations. It utilizes LC ideas for the sheets, lines and the 'cover'.

In (4.45) the *left* term $f(x)$ is the total downcrossing rate of level x, on all pages. On the right side, the term $\lambda \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} P(0, \boldsymbol{m}) e^{-\nu(z, \boldsymbol{m})x}$ is the total rate at which the SP upcrosses level x at arrival instants, due to jumps starting at level 0. The term

$$\lambda \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu(z, \boldsymbol{m})(x-z)} f(z, \boldsymbol{m}) dz$$

is the total rate at which the SP upcrosses level x at arrival instants, starting from levels in $(0, x)$ on all pages $\boldsymbol{m} \in \boldsymbol{M}_1$. We form Eq. (4.45) by rate balance, with respect to level x

$$\boldsymbol{Downcrossing\ \ rate} = \boldsymbol{Upcrossing\ \ rate}.$$

The normalizing condition is

$$P_0 + \int_{x=0}^{\infty} f(x) dx = 1,$$

which too has an LC interpretation. That is, multiply both sides by λ. This yields

$$\lambda P_0 + \lambda \int_{x=0}^{\infty} f(x)dx = \lambda.$$

On the left side, λP_0 is the rate at which the SP makes jumps at arrival instants out of zero-wait states. The term $\lambda \int_{x=0}^{\infty} f(x)dx$ is the rate at which the SP makes jumps at arrival instants from positive-wait states. The right side λ is the total rate at which the SP makes jumps at arrival instants. The left and right sides are equal.

4.6.10 Discussion of Rate Balance in Complex Models

The rate-balance interpretation provides the analyst with a powerful technique for constructing model equations for steady-state distributions in very complex models. The method is straightforward, intuitive, and relatively easy.

1. Select a state-space interval with boundary x.
2. Express the SP entrance and exit rates of the interval algebraically in terms of the unknown probability of the interval and/or unknown pdf at x.
3. Apply rate balance to construct an integral equation (or other type of balance equation) for the probability and/or pdf at x.
4. Repeat (1)–(3) for every sub-partition of the state space as required, to form a complete system of Volterra integral equations of the second kind (as above), plus other relevant equations, depending on the model.
5. Write the normalizing condition.
6. Solve the entire system of equations simultaneously for the probabilities and pdfs of the model. This can be done analytically, numerically, by approximation, or by LC estimation (see Chap. 9).

Remark 4.11 The author realized in 1974 that the steady-state model equations discussed here, are really **rate-balance equations**. Originally, these steady-state equations had been derived by starting with Lindley recursions, analogous to those described for M/G/1 in Sect. 1.2 of Chap. 1. The derivation for M/M/c queues started, however, with more complex Lindley recursions.

4.7 Example of Steady-State Equations

This Section serves as a brief tutorial on writing steady-state model equations using LC and sample paths. We derive model equations for the steady-state pdf of wait in the *specific* M/M/c queue with $c = 3$ and $J = 1$, discussed in Example 4.4 in Sect. 4.5.5, with a sample path in Fig. 4.2. There are two possible service rates: $\boldsymbol{\mu} = \{\mu_0, \mu_1\}$. We make a slight generalization for the *service-rate* assignment policy. For each arrival, the rates $\{\mu_0, \mu_1\}$ are assigned with probabilities $\{\alpha_0, \alpha_1\}$, $\alpha_0 + \alpha_1 = 1$ (instead of 1/2 each). Our present example reduces to Example 4.4 if $\alpha_0 = \alpha_1 = 1/2$.

We use α_0, α_1 to make it easier to follow the intuitive derivation of the model equations, since α_0, α_1 appear explicitly in the equations.

The set of possible configurations is $\boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{(m_0, m_1)\}$, where m_j denotes the number of servers occupied by customers with service rate μ_j, $j = 0, 1$. From the definition of system configuration (Sect. 4.4),

$$0 \leq \sum_{j=0}^{1} m_j \leq c - 1 = 2.$$

We abbreviate (m_0, m_1) as $m_0 m_1$. There are *six* possible configurations (same as in Example 4.4):

$$\boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{00, 10, 01, 20, 11, 02\}, \tag{4.46}$$

where

$$\boldsymbol{M}_0 = \{00, 10, 01\}, \qquad \boldsymbol{M}_1 = \{20, 11, 02\}.$$

When an arrival finds more than one server idle, it immediately occupies one of them in accordance with a *server-assignment* rule, and starts service at rate μ_i with probability α_i, $i = 0, 1$.

First we will derive the equations for the zero-wait states (atoms). These are represented in the virtual-wait diagram by the six lines $\boldsymbol{T} \times (0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_1$ (Fig. 4.2).

Next we will derive the integral equations for the pdfs of the positive-wait states (continuous states). These states are represented by pages $\boldsymbol{T} \times ((0, \infty), m)$, $\boldsymbol{m} \in \boldsymbol{M}_1$ (Fig. 4.2). Fix level $x > 0$. For the equation corresponding to $\boldsymbol{m} \in \boldsymbol{M}_1$, the left side is the SP exit rate (*out of)* state-space interval $((x, \infty), \boldsymbol{m})$, and the right side is the SP entrance rate (*into)* $((x, \infty), \boldsymbol{m})$. (We use interval (x, ∞) instead of $(0, x)$, since (x, ∞) results in simpler (equivalent) equations.) Since $\boldsymbol{M}_1 = \{20, 11, 02\}$, there are three pages, three pdfs, and three corresponding integral equations.

Remark 4.12 To summarize, the zero-wait states are $(0, m), \boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_1$. The positive-wait states we use for the derivation, are composite states $((x, \infty), \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$. We could use alternative state-space intervals having a fixed level-x boundary , such as $((0, x), \boldsymbol{m})$ or $((x, a), \boldsymbol{m})$, where constant $a > x$, or $((x, bx), \boldsymbol{m})$, $b > 1$, etc. For different interval selections we would derive a different, but equivalent set of model equations. A creative choice of state-space interval may simplify the derivation and final form of the equations. It may lead to new identities or insights about the model. It may also suggest easier ways to obtain solutions of the equations.

The configurations for the zero-wait states are given in $\boldsymbol{M}_0 \cup \boldsymbol{M}_1$ and for the pages in $\boldsymbol{M}_1$, in (4.46) above; (see also Fig. 4.2).

We now derive the model equations; a detailed explanation follows immediately after.

4.7.1 Equations for Zero-Wait States

Notation 4.9 State $(0, m_0 m_1)$ means $m_0 + m_1$ servers are occupied: m_i serve at rate μ_i, $i = 0, 1$. $P_{m_0 m_1} := P(\text{system is in state } (0, m_0 m_1))$.

Using the principle for discrete states ***rate out = rate in***, we obtain the equations for the zero-wait states, as in (4.47). A detailed explanation follows below.

State	Rate out		Rate in
(0, 00)	λP_{00}	$=$	$\mu_0 P_{10} + \mu_1 P_{01}$
(0, 10)	$(\lambda + \mu_0) P_{10}$	$=$	$\lambda\alpha_0 P_{00} + 2\mu_0 P_{20} + \mu_1 P_{11}$
(0, 01)	$(\lambda + \mu_1) P_{01}$	$=$	$\lambda\alpha_1 P_{00} + 2\mu_1 P_{02} + \mu_0 P_{11}$
(0, 20)	$(\lambda + 2\mu_0) P_{20}$	$=$	$\lambda\alpha_0 P_{10} + f_{20}(0^+)$
(0, 11)	$(\lambda + \mu_0 + \mu_1) P_{11}$	$=$	$\lambda\alpha_1 P_{10} + \lambda\alpha_0 P_{01} + f_{11}(0^+)$
(0, 02)	$(\lambda + 2\mu_1) P_{02}$	$=$	$\lambda\alpha_1 P_{01} + f_{02}(0^+)$

(4.47)

Explanation for Discrete States $(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$

In (4.47) the first three equations are derived as in a "bubble" diagram for discrete-state continuous-time Markov chains, using *rate out = rate in.* The last three equations are derived similarly, except for the terms $f_{20}(0^+)$, $f_{11}(0^+)$, $f_{02}(0^+)$. These are the exit rates from $((0, \infty), 20)$, $((0, \infty), 11)$, and $((0, \infty), 02)$ into discrete states $(0, 20)$, $(0, 11)$, $(0, 02)$ respectively. At instants of these exits, the SP simultaneously enters the corresponding line $\boldsymbol{T} \times (0, 20)$, $\boldsymbol{T} \times (0, 11)$, or $\boldsymbol{T} \times (0, 02)$. It continues its motion.

4.7.2 Equations for States $((x, \infty), m)$, $m \in M_1$

We now derive the model equations for pages $\boldsymbol{m} \in \boldsymbol{M}_1$. Detailed explanations follow immediately after Eq. (4.51) below.

Rate balance of rates out (left side) and in (right side) for composite state $((x, \infty), 20)$, $x > 0$, result in the equation

$$
\begin{aligned}
&f_{20}(x) + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{20}(y)dy \\
&= \lambda \left(\alpha_0 e^{-3\mu_0 x} + \alpha_1 \frac{\mu_1}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} \right) P_{20} \\
&\quad + \lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} P_{11} + \lambda\alpha_0 \int_{y=0}^{x} e^{-3\mu_0(x-y)} f_{20}(y)dy \\
&\quad + \lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y)dy \\
&\quad + \lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{11}(y)dy. \qquad (4.48)
\end{aligned}
$$

Rate balance for composite state $((x, \infty), 11)$, $x > 0$, gives the equation

$$
\begin{aligned}
&f_{11}(x) + \lambda\alpha_1 \frac{\mu_0}{\mu_0 + 2\mu_1} \int_{y=x}^{\infty} f_{11}(y)dy + \lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{11}(y)dy \\
&= \lambda \left(\alpha_1 \frac{2\mu_1}{\mu_0 + 2\mu_1} e^{-(\mu_0+2\mu_1)x} + \alpha_0 \frac{2\mu_0}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} \right) P_{11} \\
&\quad + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} P_{20} + \lambda\alpha_0 \frac{2\mu_1}{\mu_0 + 2\mu_1} e^{-(\mu_0+2\mu_1)x} P_{02} \\
&\quad + \lambda\alpha_1 \frac{2\mu_1}{\mu_0 + 2\mu_1} \int_{y=0}^{x} e^{-(\mu_0+2\mu_1)(x-y)} f_{11}(y)dy \\
&\quad + \lambda\alpha_0 \frac{2\mu_0}{2\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y)dy \\
&\quad + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{20}(y)dy \\
&\quad + \lambda\alpha_0 \frac{2\mu_1}{\mu_0 + 2\mu_1} \int_{y=0}^{x} e^{-(\mu_0+2\mu_1)(x-y)} f_{02}(y)dy \\
&\quad + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{20}(y)dy + \lambda\alpha_0 \frac{2\mu_1}{\mu_0 + 2\mu_1} \int_{y=x}^{\infty} f_{02}(y)dy. \qquad (4.49)
\end{aligned}
$$

Rate balance for composite state $((x,\infty),02)$, $x>0$, gives the equation

$$
\begin{aligned}
& f_{02}(x)+\lambda\alpha_0\frac{2\mu_1}{\mu_0+2\mu_1}\int_{y=x}^{\infty}f_{02}(y)dy \\
&= \lambda\left(\alpha_1 e^{-3\mu_1 x}+\alpha_0\frac{\mu_0}{\mu_0+2\mu_1}e^{-(\mu_0+2\mu_1)x}\right)P_{02} \\
&\quad +\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}e^{-(\mu_0+2\mu_1)x}P_{11}+\lambda\alpha_1\int_{y=0}^{x}e^{-3\mu_1(x-y)}f_{02}(y)dy \\
&\quad +\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}\int_{y=0}^{x}e^{-(\mu_0+2\mu_1)(x-y)}f_{11}(y)dy \\
&\quad +\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}\int_{y=x}^{\infty}f_{11}(y)dy. \qquad (4.50)
\end{aligned}
$$

The normalizing condition is

$$
\begin{aligned}
& P_{00}+P_{10}+P_{01}+P_{20}+P_{11}+P_{02} \\
& +\int_{x=0}^{\infty}\left[f_{20}(x)+f_{11}(x)+f_{02}(x)\right]dx=1. \qquad (4.51)
\end{aligned}
$$

Explanation of Equations for States $((x,\infty),\boldsymbol{m}),\ \boldsymbol{m}\in\boldsymbol{M}_1$

Left Side of Equation (4.48) In (4.48), the left side represents the SP *exit* rate out of $((x,\infty),20)$. There are two routes by which the SP can exit this composite state: (1) downcrossing level x on page 20; (2) jumping to page 11 pursuant to an arrival that is assigned rate μ_1. The term $f_{20}(x)$ is the downcrossing rate of level x on page 20.

The term

$$
\lambda\alpha_1\frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=x}^{\infty}f_{20}(y)dy
$$

is the rate at which the SP jumps to page 11 at arrival instants. In this expression, $\lambda f_{20}(y)dy$ is the rate at which arrivals find the SP in interval $(y,y+dy)$ on page 20. The term α_1 is the probability that an arrival gets assigned rate μ_1, resulting in two servers having rate μ_0 and one server having rate μ_1 just after the arrival starts service. The term $\frac{2\mu_0}{2\mu_0+\mu_1}$ is the probability that a rate-μ_0 customer finishes first, causing the SP to jump to page 11. The SP cannot jump to page 02 if an arrival finds the configuration to be 20. The sum of the two terms on the left of side of (4.48) is the *exit* rate of the SP out of (x,∞) on page 20.

Right Side of Equation (4.48) The right side of (4.48) is the SP *entrance* rate into $((x, \infty), 20)$. The fist term

$$\lambda \left(\alpha_0 e^{-3\mu_0 x} + \alpha_1 \frac{\mu_1}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} \right) P_{20}$$

is the entrance rate into $((x, \infty), 20)$ due to arrivals that find the state to be $(0, 20)$. In it, the product λP_{20} is the rate at which arrivals find the state to be $(0, 20)$. The arrival does not wait, and immediately starts service from the one available server. The term $\alpha_0 e^{-3\mu_0 x}$ is the product of two probabilities: α_0, that the arrival is assigned rate μ_0; $e^{-3\mu_0 x}$, that the minimum of three independent service times, each having rate μ_0, exceeds x.

The term

$$\alpha_1 \frac{\mu_1}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x}$$

is the product of three probabilities: α_1, that the arrival is assigned rate μ_1; $\mu_1 / (2\mu_0 + \mu_1)$, that the minimum of three service times, two having rate μ_0 and one having rate μ_1, is the rate μ_1; $e^{-(2\mu_0+\mu_1)x}$, that the minimum of the three service times exceeds x. Both terms result in the SP landing above x on page 02. The entire term is the rate at which the SP moves from level 0 on page 20 to interval (x, ∞) on page 20.

The term

$$\lambda \alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} e^{-(2\mu_0+\mu_1)x} P_{11}$$

is the rate at which arrivals find the state to be $(0, 11)$ (rate λP_{11}), are assigned service rate μ_0 (probability α_0), the minimum service time is a rate-μ_1 service (probability $\mu_1 / (2\mu_0 + \mu_1)$), and the minimum exceeds x (probability $e^{-(2\mu_0+\mu_1)x}$). This is the rate at which the SP moves from discrete level 0 on page 11 to (x, ∞) on page 20.

The term

$$\lambda \alpha_0 \int_{y=0}^{x} e^{-3\mu_0(x-y)} f_{20}(y) dy$$

is the rate at which arrivals find the state to be $(y, 20)$, $y \in (0, x)$, are assigned service rate μ_0 (probability α_0), and the minimum of three service times each having rate μ_0 exceeds $x - y$ (probability $e^{-3\mu_0(x-y)}$) integrated over all $y \in (0, x)$. This is the rate at which the SP moves from $(0, x)$ on page 20 to (x, ∞) on page 20 (makes $20 \rightarrow 20$ upcrossings of x).

The term

$$\lambda \alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y) dy$$

is the rate at which arrivals find the state to be in $((y, y+dy), 11)$, $y \in (0, x)$ (factor $\lambda f_{11}(y)dy$), are assigned service rare μ_0, the rate-μ_1 service ends first, and the minimum of three exponential r.v.s (two having rate μ_0 and one rate μ_1) exceeds $x - y$, integrated over all $y \in (0, x)$. This is the rate at which the SP moves from $(0, x)$ on page 11 to (x, ∞) on page 20 (makes $11 \to 20$ upcrossing of x).

The term

$$\lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{11}(y)dy$$

is the rate at which arrivals find the state to be in $((y, y+dy), 11)$, $y > x$, are assigned service rate μ_0, the rate-μ_1 service finishes first, and the minimum of three exponential service times (two having rate μ_0 and one having rate μ_1) has any value in (x, ∞). This is the rate at which the SP moves from (x, ∞) on page 11 to (x, ∞) on page 20 (makes $11 \to 20$ transition, from and to, points above x).

Integral Equations (4.49) and (4.50)
We derive integral equations (4.49) and (4.50) for the pdfs $f_{11}(x)$ and $f_{02}(x)$ (pages 11 and 02), in a similar manner as for $f_{20}(x)$ above.

Normalizing Condition
The normalizing condition (4.51) ensures that the sum of all zero-wait and positive-wait probabilities is 1.

4.8 Standard M/M/c: Steady-State Analysis

We analyze the standard M/M/c queue as a special case of the generalized M/M/c queue developed in Sects. 4.3–4.7. It is instructive to derive known results for M/M/c using LC. *The standard M/M/c queue is one of the first models the author analyzed in 1974, to validate the LC method* (see pp. 37–39 in [11]).

We assume the number of servers is $c \geq 2$, each customer receives the same exponential service rate μ, and $\lambda < c\mu$. Using the notation of Sect. 4.3, we have here $J = 0$, $\boldsymbol{\mu} = \{\mu_0\} := \{\mu\}$. A system configuration has one component m_0, which can take values in $\{0, 1, \ldots, c-1\}$. The virtual wait process is denoted as $\{W(t)\}_{t \geq 0}$.

In this model, a system configuration is a scalar $m_0 :=$ *number of customers in the other servers just after an arrival starts service*. Thus $m_0 \in \{0, 1, \ldots, c-1\}$. Equivalently m_0 is the number of *other* occupied servers at a start of service instant. The set of all configurations, $\boldsymbol{M} = \boldsymbol{M}_0 \cup \boldsymbol{M}_1$, has size $\binom{J+c}{c-1} = \binom{0+c}{c-1} = \binom{c}{1} = c$ (see Sect. 4.4.2). That is,

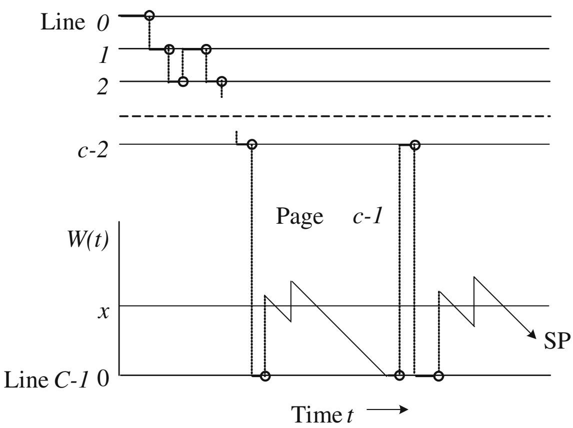


Fig. 4.4 Sample path of $\{W(t)\}_{t\geq 0}$ in *standard* M/M/c queue. There are c lines and *one* page. Line $c-1$ is at level 0 of page $c-1$

$$\boldsymbol{M}_0 = \{0, 1, \ldots, c-2\}, \ \boldsymbol{M}_1 = \{c-1\}.$$

(Recall that $\boldsymbol{M}_1 = \boldsymbol{M}_b$, the set of 'border' configurations.)

A sample path of $\{W(t)\}_{t\geq 0}$ has c *lines* for the zero-wait states $(0, j)$, $j = 0, \ldots, c-1$, and *one page* (*sheet*) for the composite state $((0, \infty), c-1)$ (Fig. 4.4). Line $c-1$, the *border* line corresponding to state $(0, c-1)$, is usually placed as the bottom line of page $c-1$, but is arbitrarily located among the other 0-wait states in Fig. 4.4. This does not change the analysis because rate balance across level $x > 0$ (*downcrossing rate* = *upcrossing rate*) is equivalent to rate balance between sets $((x, \infty), c-1)$ and $([0, x], c-1)$ i.e., *rate from* $((x, \infty), c-1)$ *into* $([0, x], c-1)$ = *rate from* $([0, x], c-1)$ *into* $((x, \infty), c-1)$.

Denote the zero-wait probabilities as P_n, $n = 0, \ldots, c-1$, the pdf of wait as $f(x)$, $x > 0$, and the steady-state cdf of wait by $F(x)$, $x \geq 0$. Then

$$F(x) = \sum_{n=0}^{c-1} P_n + \int_0^x f(x)dx, x \geq 0,$$

$$F(0) = \sum_{n=0}^{c-1} P_n.$$

4.8.1 Equations for Steady-State PDF of Wait

We derive model equations for the steady-state pdf of wait, and give further explanations in Sect. 4.8.2 below.

Zero-Wait States
For the zero-wait states (atoms) the model equations are (using *rate out* = *rate in*)

$$\begin{aligned} \lambda P_0 &= \mu P_1 \\ (\lambda + \mu)P_1 &= \lambda P_0 + 2\mu P_2 \\ (\lambda + 2\mu)P_2 &= \lambda P_1 + 3\mu P_3 \\ &\cdots \\ (\lambda + (c-2)\mu)P_{c-2} &= \lambda P_{c-3} + (c-1)\mu P_{c-1} \\ (\lambda + (c-1)\mu)P_{c-1} &= \lambda P_{c-2} + f(0^+). \end{aligned} \tag{4.52}$$

The term $f(0^+)$ in the last equation in (4.52) connects the *pdf of a continuous random variable* (*waiting time*) with the *probabilities of atoms* (states $(0, c-1)$ and $(0, c-2)$). This observation (and other examples) led the author to coin the term "*border state*" (i.e., state $(0, c-1)$ in the present context).

Positive-Wait States
For the composite state $((0,\infty), c-1)$ (the single page) the model equation is

$$f(x) = \lambda P_{c-1} e^{-c\mu x} + \lambda \int_{y=0}^{x} e^{-c\mu(x-y)} f(y)dy, \, x > 0. \tag{4.53}$$

Composite state $((0,\infty), c-1)$ is accessible in one step at an arrival instant, only from the border state $(0, c-1)$. The normalizing condition is

$$F(\infty) = \sum_{n=0}^{c-1} P_n + \int_{y=0}^{\infty} f(x)dx = 1. \tag{4.54}$$

4.8.2 Explanation of Equations (4.52) and (4.53)

Linear Equations (4.52)
Equation (4.52) are rate-balance equations, which equate SP rates out of, and into, the discrete zero-wait states $(0, n)$, $n = 0, \ldots, c-1$. The term $f(0^+)$ $(:= f(0))$ is the SP downcrossing rate of level 0, i.e., the SP *entrance rate into state* $(0, c-1)$ from above. (In sample-paths, line $c-1$ may be equally placed at level 0 of page $(c-1)$. If it is placed separately as in Fig. 4.4, we can still imagine it to be *at level 0* of the page with respect to SP motion.)

Integral Equation (4.53)
To derive the positive-wait integral equation (4.53) consider composite state $((x,\infty), c-1)$ on the (single) page (Fig. 4.4). We equate the SP *exit rate*

(i.e., downcrossing rate of level x) to the *entrance rate* (i.e., 'upcrossing' rate of level x starting from line $c-1$ thought of as being at the bottom of the page). The downcrossing rate of level x is $f(x)$ (see Corollary 4.2 in Sect. 4.6.2).

The SP entrance rate into $((x,\infty), c-1)$ is from two sources:
(1) Entrances are generated by jumps due to arrivals when the state is the border state $(0, c-1)$, starting from level 0 of the page and ending above level x on the page. Since there is only one page, the only access to $((x,\infty), c-1)$ in one step from a zero-wait state is from state $(0, (c-1))$, i.e., line $c-1$ in the sample path. The SP entrance rate from this source is $\lambda P_{c-1} \cdot P(\mathcal{S} > x)$, where P_{c-1} is the limiting probability of state $(0, c-1)$, and $\mathcal{S}$ is the *inter start-of-service depart time*. (See Sect. 4.4.1 for a discussion of *inter start-of-service depart time*.) Random variable $S \underset{dis}{=} \text{Exp}_{c\mu}$, since there would be c customers with rate μ in service just after such an arrival starts service, and $S :=$ *minimum of* c *i.i.d.* $\textit{Exp}_\mu$ *random variables*. Thus, $P(\mathcal{S} > x) = e^{-c\mu x}$. This gives the term $\lambda P_{c-1} e^{-c\mu x}$ in (4.53).
(2) Entrances into $((x,\infty), c-1)$ are generated by jumps due to arrivals when the state is a continuous state $(y, c-1)$, $y \in (0, x)$. Such jumps start at level y and end above level x. Just after such an arrival begins service (y after its arrival), all c servers will be occupied and $S \underset{dis}{=} \text{Exp}_{c\mu}$, independent of any new arrivals to the system. The SP will enter $((x,\infty), c-1)$ with probability $e^{-c\mu(x-y)}$. This leads to Eq. (4.53).

4.8.3 Solution of Equations

We first solve (4.53). Differentiating both sides with respect to x and solving the resulting first-order differential equation, gives

$$f(x) = Ae^{-(c\mu-\lambda)x}, x > 0,$$

where A is a constant. Letting $x \downarrow 0$, we get the initial condition

$$f(0) = A = \lambda P_{c-1} \tag{4.55}$$

since $f(0)$ $(:= f(0^+))$ is the SP downcrossing rate of level 0, and λP_{c-1} is the "upcrossing" rate of level 0 (rate of egress from $(0, c-1)$ above). (Equivalently, $f(0)$ is the exit rate out of $((0,\infty), c-1)$ and λP_{c-1} is the entrance rate into $((0,\infty), c-1)$.) Thus $A = \lambda P_{c-1}$ and

$$f(x) = \lambda P_{c-1} e^{-(c\mu-\lambda)x}, x > 0. \tag{4.56}$$

Note that the condition (4.55) is itself a rate-balance equation for the rates out of, and into, $((0, \infty), c-1)$.

Next, from (4.52) and (4.56) we obtain

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} P_0, n = 0, \ldots, c-1,$$
$$P_{c-1} = \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0. \tag{4.57}$$

Substituting (4.57) into (4.56) gives

$$f(x) = \lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0 \cdot e^{-(c\mu-\lambda)x}, x > 0.$$

The normalizing condition (4.54) is

$$\left(\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}\right) P_0 + \lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = 1.$$

This gives the well-known value

$$P_0 = \frac{1}{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \left(\frac{\lambda}{\mu}\right)^c \frac{c\mu}{c!(c\mu-\lambda)}}. \tag{4.58}$$

The cdf of wait is

$$\begin{aligned} F(x) &= P_0 + \int_{y=0}^{x} \lambda P_{c-1} e^{-(c\mu-\lambda)y} dy \\ &= P_0 \left(1 + \lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!(c\mu-\lambda)} \left(1 - e^{-(c\mu-\lambda)x}\right)\right), x \geq 0. \end{aligned} \tag{4.59}$$

Boundedness of PDF of Wait

From (4.56) $f(x) < \lambda$, $x > 0$, since $P_{c-1} < 1$ and $e^{-(c\mu-\lambda)x} < 1$ $(c\mu - \lambda > 0)$.

4.8.4 CDF and PDF of Wait Geometrically

It is insightful and intuitive to derive the steady-state cdf and pdf of wait geometrically, directly from sample path properties. This derivation bypasses model equation (4.53). A similar geometric derivation for the cdf of wait in the M/M/1 queue is given in Sect. 3.5.6.

Consider level $x > 0$ on the single page (Fig. 4.4). Rate balance across level x applies the principle

$$\textit{Upcrossing rate of } x = \textit{Downcrossing rate of } x = f(x).$$

Equivalently, in symbols

$$\lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t} = \lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} = f(x)\ (a.s.),$$

or

$$\lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t} = \lim_{t\to\infty}\frac{E\left(\mathcal{D}_t(x)\right)}{t} = f(x).$$

The sojourn time above level $x > 0$ on the page, initiated by each upcrossing of x, is := *busy period of a standard* $M_\lambda/M_{c\mu}/1$ queue with arrival rate λ and *service rate* $c\mu$ because, when the SP is on the page, all c servers are occupied and each is serving at rate μ. Thus the inter start-of-service depart time (see Definition 4.3 in Sect. 4.4.1) $\mathcal{S} \underset{dis}{=}$ *size* of *each jump ending on the page* $\underset{dis}{=} \text{Exp}_{c\mu}$. Moreover, by the memoryless property, excess jumps above level x are $\underset{dis}{=} \text{Exp}_{c\mu}$.

Let a_x denote an SP sojourn time above x. Then $a_x \underset{dis}{=}$ *busy period in* $M_\lambda/M_{c\mu}/1$ $(\lambda < c\mu)$. Thus

$$E(a_x) = \frac{1}{c\mu - \lambda}, \tag{4.60}$$

independent of x, since the expected value of the busy period in $M_\lambda/M_{c\mu}/1$ is $1/(c\mu - \lambda)$.

Let d_x := *inter-downcrossing time at level* $x \geq 0$. Since level-x downcrossings are regenerative points, similarly as in Sect. 3.4.15 we have

$$E(d_x) = 1/f(x). \tag{4.61}$$

The renewal reward theorem (Sect. 3.79), now yields

$$\frac{E(a_x)}{E(d_x)} = \lim_{t\to\infty} \frac{\text{time } W(\cdot) \in (x,\infty) \text{ during } (0,t)}{t} = 1 - F(x),$$

$$\frac{1/(c\mu - \lambda)}{1/f(x)} = 1 - F(x),$$

or

$$\frac{f(x)}{1-F(x)} = c\mu - \lambda, x > 0, \tag{4.62}$$

equivalent to the differential equation

$$\frac{\frac{d}{dx}(1-F(x))}{1-F(x)} = -(c\mu - \lambda),$$

$$\frac{d}{dx}\ln(1-F(x)) = -(c\mu - \lambda),$$

with solution

$$1 - F(x) = A \cdot e^{-(c\mu-\lambda)x},$$

where A is a constant, evaluated by letting $x \downarrow 0$, and yielding the cdf of wait

$$F(x) = 1 - (1 - F(0))e^{-(c\mu-\lambda)x}, x \geq 0, \tag{4.63}$$

where $F(0) = P(\text{zero wait})$. Taking $dF(x)/dx$, $x > 0$, in (4.63) gives the pdf of wait

$$f(x) = (1 - F(0))(c\mu - \lambda)e^{-(c\mu-\lambda)x}, x > 0. \tag{4.64}$$

We next employ the equations in (4.57) to get

$$F(0) = \sum_{n=0}^{c-1} P_n = P_0 \sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}. \tag{4.65}$$

Note that $f(0) = \lambda P_{c-1}$., i.e., the SP entrance rate into state $(0, c-1)$ from above (downcrossing rate of level 0) is equal to the SP exit rate from state $(0, c-1)$ at arrival instants. Letting $x \downarrow 0$ In (4.64) yields

$$f(0) = (1 - F(0))(c\mu - \lambda) = \lambda P_{c-1}. \tag{4.66}$$

From (4.66) and (4.57)

$$F(0) = 1 - \frac{\lambda}{c\mu - \lambda} P_{c-1} = 1 - \frac{\lambda}{c\mu - \lambda}\left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0. \tag{4.67}$$

Substituting the value of $F(0)$ from (4.65) into (4.67) and solving for P_0 gives (4.58). The upshot is two different ways to determined P_0; and two different, equivalent formulas for $f(x), x > 0$: (4.56) and (4.64).

Remark 4.13 Another way to obtain the second equality in (4.66) is to note that the SP expected sojourn time above 0 is

$$E(a_0) = E(\text{busy period of } M_\lambda/M_{c\mu}/1) = \frac{1}{c\mu - \lambda}.$$

The proportion of time the SP spends above level 0 is therefore

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t} \cdot \frac{1}{c\mu - \lambda} = \lambda P_{c-1} \cdot \frac{1}{c\mu - \lambda} = 1 - F(0).$$

Busy Period in M/M/c

Note that a_0 is equal to a busy period in M/M/c, denoted by $\mathcal{B}_{c-1,c}$, defined as the time measured from an arrival instant when the state is $(0, c-1)$ until the first departure instant thereafter that leaves the system in state $(0, c-1)$ again. (The arrival increases the number in the system to c. The departure decreases the number to $c-1$.) Since $a_x \equiv a_0, x \geq 0$,

$$E\left(\mathcal{B}_{c-1,c}\right) = E\left(a_0\right) = E\left(a_x\right) = \frac{1}{c\mu - \lambda}, x \geq 0. \tag{4.68}$$

We also call $\mathcal{B}_{c-1,c}$ a $[c-1, c]$ busy period.

4.8.5 PMF of Number in the System

We use the foregoing pdf of wait (4.56) to derive $P_n, n = c, c+1, \ldots$. This approach is the reverse order of the usual derivation, which first derives the pmf (probability mass function) of the number-in-system using a birth-death analysis. It then obtains the pdf of wait by conditioning on the number in the system when there is an arrival. The method we apply here utilizes partly birth-death analysis and partly LC. It provides a different perspective on the M/M/c model.

Due to Poisson arrivals, $P_n = a_n = d_n$, where a_n, d_n are the steady-state probabilities of n units in the system just before an arrival, and just after a departure, respectively (in this Section). Reasoning as for M/M/1 (see Sect. 3.5.3), we get

$$P_n = d_n = P(n - c \text{ arrivals during a waiting time}), n = c, c+1, \ldots .$$

Substituting from (4.56) and (4.57)

$$
\begin{aligned}
P_n &= \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-c}}{(n-c)!} f(x)dx \\
&= \left(\frac{\lambda}{\mu}\right)^{n-c+1} \frac{1}{c^{n-c+1}} P_{c-1} \int_{x=0}^{\infty} c\mu e^{-c\mu x} \frac{(c\mu x)^{n-c}}{(n-c)!} dx \\
&= \left(\frac{\lambda}{\mu}\right)^{n} \frac{1}{c^{n-c} c!} P_0, \; n = c, c+1, \dots .
\end{aligned}
$$

In summary, we obtain the well-known formulas (e.g., p. 67 in [84])

$$
\left.
\begin{aligned}
P_0 &= \frac{1}{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \left(\frac{\lambda}{\mu}\right)^c \frac{c\mu}{c!(c\mu-\lambda)}} \\
P_n &= \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} P_0, \; n = 0, \dots, c-1, \\
P_n &= \left(\frac{\lambda}{\mu}\right)^n \frac{1}{c^{n-c}} \frac{1}{c!} P_0, \; n = c, c+1, \dots .
\end{aligned}
\right\} \tag{4.69}
$$

The probability that all servers are occupied is

$$
\begin{aligned}
\sum_{n=c}^{\infty} P_n = P(\text{wait} > 0) &= \int_{x=0}^{\infty} f(x)dx \\
&= \lambda P_{c-1} \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = \frac{\lambda}{c\mu - \lambda} P_{c-1} \\
&= \frac{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{c!}}{c\mu - \lambda} P_0.
\end{aligned} \tag{4.70}
$$

The probability that there is at least one idle server is

$$
\sum_{n=0}^{c-1} P_n = P(\text{wait} = 0) = 1 - \frac{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{c!}}{c\mu - \lambda} P_0. \tag{4.71}
$$

4.8.6 Inter-downcrossing and Sojourn Times

Consider d_x, a_x, b_x $(x \geq 0)$, respectively: time between successive SP downcrossings of level x; sojourn time above x initiated by an upcrossing of x; sojourn time at or below x initiated by a downcrossing of x. Formula (4.61) shows

$$E(d_x) = \frac{1}{f(x)} = \frac{e^{(c\mu-\lambda)x}}{\lambda P_{c-1}}, x \geq 0; \tag{4.72}$$

formula (4.60) shows that, independent of x,

$$E(a_x) = \frac{1}{c\mu - \lambda}.$$

Note that

$$\lim_{t\to\infty} \frac{(\textit{time that the SP is above } x \textit{ during } (0,t))}{t} = 1 - F(x);$$

by the renewal reward theorem (Sect. 3.4.9),

$$\frac{E(a_x)}{E(d_x)} = \frac{E(a_x)}{1/f(x)} = 1 - F(x),$$
$$E(a_x) = \frac{1-F(x)}{f(x)} = \frac{1}{c\mu-\lambda}., x > 0,$$

The last equality above corroborates formula (4.62) when solving for $f(x)$ geometrically in Sect. 4.8.4. Also, we can validate (4.62) in Sect. 4.8.4 using

$$\frac{1-F(x)}{f(x)} = \frac{\int_{y=x}^{\infty} f(y)dy}{\lambda P_{c-1}e^{-(c\mu-\lambda)x}}$$
$$= \frac{\int_{y=x}^{\infty} \lambda P_{c-1}e^{-(c\mu-\lambda)y}dy}{\lambda P_{c-1}e^{-(c\mu-\lambda)x}} = \frac{1}{c\mu-\lambda}.$$

Note that $F(x) = \lim_{t\to\infty}(\textit{time the SP spends at or below } x \textit{ during } (0,t))/t$. Each instant that the SP downcrosses $x \geq 0$ is a regenerative point, due to the memoryless property of the interarrival times. From the renewal reward theorem (i.e., the theory of regenerative processes, e.g., [134]) $E(b_x)/E(d_x) = F(x)$, implying

$$E(b_x) = \frac{F(x)}{f(x)} = \frac{1-(1-F(0))e^{-(c\mu-\lambda)x}}{\lambda P_{c-1}e^{-(c\mu-\lambda)x}}$$
$$= \frac{e^{(c\mu-\lambda)x}}{\lambda P_{c-1}} - \frac{(1-F(0))}{\lambda P_{c-1}}$$
$$= \frac{1}{\lambda\left(\frac{\lambda}{\mu}\right)^{c-1}\frac{1}{(c-1)!}P_0}\left(e^{(c\mu-\lambda)x} - (1 - P_0\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^n\frac{1}{n!})\right)$$

$$= \frac{e^{(c\mu-\lambda)x} - 1}{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0} + \frac{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!}}. \tag{4.73}$$

Remark 4.14 From (4.60), when the SP upcrosses x, it next downcrosses x after a time a_x where $E(a_x)$ is independent of x. By contrast, (4.73) implies when the SP downcrosses level x, it next upcrosses x after a time b_x where $E(b_x)$ grows exponentially with increasing x.

The foregoing results for

$$d_x,\ a_x,\ b_x, E(d_x),\ E(a_x), E(b_x)$$

generalize analogous results for M/M/1 (Sect. 3.5.7).

4.9 M/M/c/c and Standard M/M/c Queues

The M/M/c/c queue is a special case of M/M/c/k, in which an upper limit k is placed on the number of customers allowed in the system at any time (see, e.g., Sect. 2.5, p. 76ff in [84]). Here, we develop a relationship between M/M/c/c and the standard $M_\lambda/M_\mu/c$ queue. By a judicious choice of parameters for M/M/c/c, the pdf of the virtual wait in the two models have identical forms. However, the jump structure of the sample path of M/M/c/c is much simpler than that of the corresponding M/M/c model, for positive values of the virtual wait. This jump structure makes it much easier to derive the pdf of the virtual wait in M/M/c/c. The point of this exercise is to obtain the pdf of wait in the parameter-modified M/M/c/c queue, which can be derived in one line, without having to solve an integral equation (as in M/M/c), and the derived pdf is the same as in M/M/c. This relationship suggests a broader prospect. For a given complex model, can we identify a related model having the same solution form, that can be solved more easily?

The M/M/c/c queue is usually analyzed using a birth-death analysis. Here, we employ an LC approach. Consider an M/M/c/c queue where the service time for each customer that enters the system has exponential rate $\mu - \frac{\lambda}{c} > 0$. (We choose $\lambda < c\mu$ because our related model is a standard $M_\lambda/M_\mu/c$ queue in equilibrium.)

In M/M/c/c all *actual* waits are 0 – there is no waiting line. In a queue where blocking is possible, we shall define the virtual wait as the time that a potential arrival *would* wait to start service, if it were not blocked and cleared.

Thus the virtual wait is not 0 for every arrival. In M/M/c/c, customers that arrive when the *virtual* wait is positive, are blocked and cleared from the system. In both models, the virtual wait is positive if and only if all c servers are occupied.

For M/M/c/c, consider the 'system point' process $\{W(t), M(t)\}_{t\geq 0}$, where $W(t)$ is the virtual wait and $M(t) \in \{0, \ldots, c-1\}$ is the system configuration at time t. $M(t)$ is the number of occupied servers at instant t^-, if there is an idle server at t^-. We denote the c *discrete states* by $\{(0, 0), \ldots, (0, c-1)\}$. Thus $M(t) = n$ if n *other* servers are occupied when a customer joins the system and starts service, $n = 0, \ldots, c-1$. Denote the steady-state probability of $(0, n)$ as $P_n, n = 0, \ldots, c-1$. Denote the *positive* virtual-wait states as $\{(x, c-1), x \in (0, \infty)\}$.

4.9.1 Sample Path of $\{W(t), M(t)\}_{t\geq 0}$

Consider a sample path of $\{W(t), M(t)\}_{t\geq 0}$ (Fig. 4.5). Without loss of generality, assume the system starts empty. The SP is on line 0 at $t = 0$. As the system evolves, the SP moves among the lines until $c-1$ of the servers are occupied, just as in a standard $M_\lambda/M_{\mu-\frac{\lambda}{c}}/c$ model. In Fig. 4.5 we situate line $c-1$ at level 0 of the page; this layout makes it easier to depict SP exchanges between line $c-1$ and the virtual-wait positive states.

Suppose a customer arrives when $c-1$ servers are occupied. The arrival joins the system and starts service in the one free server. All c servers are busy just after the arrival starts service. The configuration is $c-1$, since $c-1$ *other* servers are occupied just after the arrival instant. Each of the c servers has service time $\underset{dis}{=} \text{Exp}_{\mu-\frac{\lambda}{c}}$ once the arrival starts service, due to the memoryless property of exponential service times. The SP jumps to ordinate $y \in (0, \infty)$ on the page, where $y \underset{dis}{=} \text{Exp}_{c\mu-\lambda}$ which is distributed as the minimum of c i.i.d. exponential r.v.s each distributed with rate $\mu - \frac{\lambda}{c}$.

The SP descends at rate 1 (slope $= -1$), until it makes a continuous hit of level 0 from above. New arrivals are blocked and cleared, and have no effect on the sample path during this descent. Once the SP hits level 0, it continues its motion among the states $(0, 0), \ldots, (0, c-1)$, until it makes another jump out of state $(0, c-1)$ onto the page.

All upward jumps that end on the page start at level 0. Hence the jump structure for M/M/c/c is much simpler than that of the standard M/M/c queue, in which jumps that end on the page may start at any point in $[0, \infty)$.

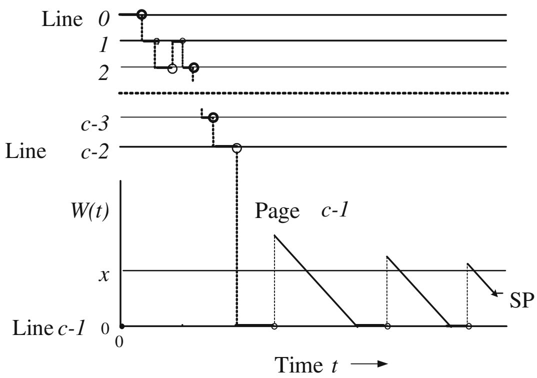


Fig. 4.5 Sample path of $\{W(t)\}_{t\geq 0}$ in $M/M/c/c$ queue. All jumps ending in $(0,\infty)$ begin at level 0 (state $(0, c-1)$) with size $\underset{dis}{=}$ $\text{Exp}_{c\mu-\lambda}$

4.9.2 PDF of Virtual Wait

Denote the pdf of the virtual wait as $f_{c-1}(x) \equiv f(x)$, $x > 0$. To derive the pdf of $f(x)$, fix level $x > 0$. The SP downcrossing rate of x is $f(x)$. Since all SP jumps ending on the page start from state $(0, c-1)$ at arrival instants, and all jumps sizes are $\underset{dis}{=}$ $\text{Exp}_{c\mu-\lambda}$, the upcrossing rate of x is $\lambda P_{c-1}e^{-(c\mu-\lambda)x}$. Balancing SP rates out of and into set $((x,\infty), c-1)$ yields

$$f(x) = \lambda P_{c-1}e^{-(c\mu-\lambda)x}, x > 0. \tag{4.74}$$

Remark 4.15 Formula (4.74) has precisely the same **form** as the steady-state pdf of wait in the standard M_λ /M_μ/c queue given by (4.56), except that P_{c-1} has a **different value**. For the M_λ/$\text{M}_{\mu-\lambda/c}$/c/c queue, formula (4.74) is derived "instantly" from observing a sample path of the virtual wait. There is no need to solve an integral equation, as in M/M/c. In M/M/c/c, the pdf formula for $f(x)$ is inherently a model equation. This is the main relationship between the two models we discuss here. The result for M_λ/$\text{M}_{\mu-\lambda/c}$/c/c allows us to write the **form** of the pdf of wait in M_λ/M_μ /c immediately.

4.9.3 Non-blocking States

The rate-balance equations for the non-blocking states $(0,0), \ldots, (0, c-1)$ are the same as in (4.52) for M_λ/M_μ/c, with $\mu - \frac{\lambda}{c}$ substituted for μ. Thus in M/M/c/c

$$P_n = \left(\frac{\lambda}{\mu - \frac{\lambda}{c}}\right)^n \frac{1}{n!} P_0, n = 0, \ldots, c-1,$$

so that

$$P_{c-1} = \left(\frac{\lambda}{\mu - \frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} P_0.$$

The normalizing condition is

$$\left(\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!}\right) P_0 + \int_{x=0}^{\infty} f(x)dx = 1.$$

Applying (4.74) gives

$$\left(\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu - \frac{\lambda}{c}}\right)^n \frac{1}{n!}\right) P_0$$
$$+ \lambda\left(\frac{\lambda}{\mu - \frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = 1,$$

$$P_0 = \frac{1}{\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!} + \lambda\left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} \frac{1}{c\mu-\lambda}}$$
$$= \frac{1}{\sum_{n=0}^{c}\left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!}}. \checkmark$$

4.9.4 Blocking Time T_B

Let T_B denote the time from the instant the system gets blocked (all c servers occupied) until the first instant that it becomes unblocked thereafter (at which $c-1$ servers are occupied). We call T_B the blocking time.

The pdf of the virtual wait in $M_\lambda/M_{\mu-\lambda/c}$/c/c is the same as the pdf of $\mathcal{S}$ (inter start-of-service depart time) when an arrival "sees" state $(0, c-1)$. Also, $\mathcal{S} \underset{dis}{=} T_B$.

Then $E(T_B) = \text{Exp}_{c\mu-\lambda} = 1/(c\mu - \lambda)$. Let P_c denote the *proportion* of time the system is blocked. Then

$$P_c = \int_{x=0}^{\infty} f(x)dx = \lambda P_{c-1} \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx$$

$$= \lambda \left(\frac{\lambda}{\mu - \frac{\lambda}{c}} \right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu - \lambda)x} dx$$

$$= \left(\frac{\lambda}{\mu - \frac{\lambda}{c}} \right)^{c} \frac{1}{c!} P_0$$

$$= \frac{\left(\frac{\lambda}{\mu - \frac{\lambda}{c}} \right)^{c} \frac{1}{c!}}{\sum_{n=0}^{c} \left(\frac{\lambda}{\mu - \frac{\lambda}{c}} \right)^{n} \frac{1}{n!}}.$$

P_c is the probability that a right-truncated Poisson variate (truncated at c), has value c. It is the classical *Erlang-B loss formula* for the $\text{M}_\lambda/\text{M}_{\mu - \frac{\lambda}{\mu}}$/c/c queue (see, e.g., p. 82, Sect. 2.6 in [84]).

Note that the blocking time is a $[c-1, c]$ busy period, denoted by $\mathcal{B}_{c-1,c}$, so that $T_B \underset{dis}{=} \mathcal{B}_{c-1,c}$. From Remark 4.8.4, $E\left(\mathcal{B}_{c-1,c}\right) = \frac{1}{c\mu - \lambda}$.

Remark 4.16 Suppose that in the M/M/c/c model the servers were numbered $1, \ldots, c$. Let the service rates assigned to arrivals depend on which server is occupied, say rates $\nu_i, i = 1, \ldots, c$. Assume $\sum_{i=1}^{c} \nu_i = c\mu - \lambda > 0$, where μ, λ are the parameters of a stable $\text{M}_\lambda/\text{M}_\mu$/c queue. Then the distribution of T_B would be the same as in (4.74). So this specialized M/M/c/c model can also be used as a "companion" model to obtain the pdf of wait in the $\text{M}_\lambda/\text{M}_\mu$/c queue.

4.9.5 Discussion

We can derive formula (4.74) for $f(x)$ geometrically as in Sect. 4.8.4. Let $F(x)$, $x \geq 0$, be the cdf of the virtual wait. We get

$$\frac{d}{dx} \ln(1 - F(x) = \frac{-1}{E(B_{c-1,c})} = -(c\mu - \lambda),$$

$$F(x) = 1 - (1 - F(0))e^{-(c\mu - \lambda)x}, x \geq 0,$$

$$f(x) = (c\mu - \lambda)(1 - F(0))e^{-(c\mu - \lambda)x}. \tag{4.75}$$

Comparing (4.74) and (4.75) shows that

$$\lambda P_{c-1} = (c\mu - \lambda)(1 - F(0)) = (c\mu - \lambda) P_c, \tag{4.76}$$

where P_c is the probability of c units in the system.

In M/M/c/c an arrival enters the system iff the virtual wait is 0. Thus $F(0) = P$(an arrival enters the system). Hence $(1 - F(0)) = P$(an arrival is blocked and cleared) $= P_c$. Equation (4.76) is precisely the balance equation that would appear in a birth-death analysis of the system.

4.10 M/M/c in Which Zero-Wait Customers Get Special Service

Consider an M/M/c ($c \geq 2$) queue with arrival rate λ, in which zero-wait customers get service rate μ_0, and positive-wait customers get service rate μ_1 ($\neq \mu_0$). Thus, the assigned service rate is *state-dependent*. We derive below the steady-state pdf of wait, distribution of the number-in-system, and related model characteristics.

Denote the state of the system as $\{W(t), \boldsymbol{M}(t)\}_{t\geq 0}$, where $W(t) \geq 0$ is the virtual wait and $M(t)$ is the system configuration. Thus

$$\boldsymbol{M}(t) = (m_0, m_1), 0 \leq m_0 + m_1 \leq c - 1,$$

where m_j is the number of occupied servers operating at rate μ_j, $j = 0, 1$. In the notation of Sect. 4.4, integer $J = 1$. The number of zero-wait states is the total number of non-negative integer solutions for m_0, m_1 in the c equations

$$m_0 + m_1 = k, k = 0, \ldots, c - 1,$$

which is, since $J = 1$,

$$\sum_{k=0}^{c-1} \binom{J+k}{J} = \binom{J+c}{J+1} = \binom{c+1}{2}$$
$$= \frac{c(c+1)}{2} = 1 + 2 + \cdots + c.$$

From Sect. 4.4.2, $\boldsymbol{M}_0 = \left\{(0, \boldsymbol{m}) | 0 \leq \sum_{j=0}^{J} m_j \leq c - 2\right\}$, which contains $\frac{(c-1)c}{2}$ configurations. Set $\boldsymbol{M}_b = \left\{\boldsymbol{m} | \sum_{j=0}^{J} m_j = c - 1\right\}$ comprises the discrete *boundary* states, and contains $\binom{J+c-1}{J} = \binom{c}{1} = c$ configurations. (Note that $\boldsymbol{M}_b = \boldsymbol{M}_1$.)

Zero-wait Probabilities Let $P_{m_0 m_1}$ denote the steady-state probability that an arrival "sees" m_j rate-μ_j customers in service, $j = 0, 1$, and waits zero before starting service. $P_{m_0 m_1}$is the steady-state probability of state $(0, (m_0, m_1))$.

There are c positive-wait *pages (sheets)*, one for each configuration in $\boldsymbol{M}_b$, where

$$\boldsymbol{M}_b = \{(c-1, 0), (c-2, 1), \ldots, (1, c-2), (0, c-1)\}.$$

Positive-wait PDFs Let $f_{\boldsymbol{m}}(x)$, $x > 0$, denote the steady-state pdf of the virtual wait when the occupancies of the *other* $c-1$ servers will be $\boldsymbol{m} \in \boldsymbol{M}_b$ at start of service ('look-ahead' property of virtual wait).

4.10.1 Equations for Probabilities of Zero-Wait States

The $\frac{c(c+1)}{2}$ zero-wait states, having configurations in $\boldsymbol{M}_0 \cup \boldsymbol{M}_b$, viz.,

$$(0, (m_0, m_1)), 0 \leq m_0 + m_1 \leq c - 1,$$

give rise to $\frac{c(c+1)}{2}$ linear equations for their probabilities, using the principle *rate out = rate in*, as in (4.77)–(4.79) below.

First consider states $(0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M}_0$. For $m_0 = m_1 = 0$, (empty system) there is one equation:

$$\lambda P_{00} = \mu_0 P_{10} + \mu_1 P_{01}. \tag{4.77}$$

For states $(0, (m_0, m_1))$, $1 \leq m_0 + m_1 \leq c - 2$, there are $\frac{(c-1)c}{2} - 1$ equations, each of the form

$$\begin{aligned}(\lambda + m_0\mu_0 + m_1\mu_1) P_{m_0 m_1} = \ & \lambda P_{(m_0-1)m_1} \\ & + (m_0 + 1)\mu_0 P_{(m_0+1)m_1} \\ & + (m_1 + 1)\mu_1 P_{m_0(m_1+1)}.\end{aligned} \tag{4.78}$$

For states $(0, (m_0, m_1)) \in \boldsymbol{M}_b$, there are c equations, each of the form

$$(\lambda + m_0\mu_0 + m_1\mu_1) P_{m_0 m_1} = \lambda P_{(m_0-1)m_1} + f_{m_0 m_1}(0). \tag{4.79}$$

In (4.79) the term $f_{m_0 m_1}(0)$ $(= f_{m_0 m_1}(0^+))$ is the rate at which the SP enters border state $(0, (m_0, m_1))$ due to left continuous hits of level 0 from above on page $m_0 m_1$.

4.10.2 Equations for PDF of Positive-Wait States

There are c Volterra integral equations for the positive-wait states. Consider composite state $((x, \infty), \boldsymbol{m})$, $x > 0$, on page $\boldsymbol{m} \in \boldsymbol{M}_b$. For positive-wait states $(y, m_0 m_1)$, $y > 0$, $m_0 + m_1 = c - 1$. We first specify the SP exit and entrance rates of the pertinent composite states in the state space. Then we will write the equations.

Rate Out of $((x, \infty), m_0 m_1)$

Because (m_0, m_1) is a configuration, $m_0 + m_1 = c - 1$. The SP rate out of $((x, \infty), m_0 m_1)$ is

$$f_{m_0 m_1}(x) + \lambda \frac{m_0 \mu_0}{m_0 \mu_0 + (m_1 + 1)\mu_1} \int_{y=x}^{\infty} f_{m_0 m_1}(y) dy. \tag{4.80}$$

Explanation of Terms in (4.80)

The first term $f_{m_0 m_1}(x)$ is the SP downcrossing rate of level x on page $m_0 m_1$. The second term

$$\lambda \frac{m_0 \mu_0}{m_0 \mu_0 + (m_1 + 1)\mu_1} \int_{y=x}^{\infty} f_{m_0 m_1}(y) dy$$

is the rate of arrivals when the state is $(y, m_0 m_1)$, $y > x$ (being assigned service rate μ_1 thereby adding one rate-μ_1 occupied server *upon start of service*); and a rate-μ_0 service completes first thereafter. At the arrival instant the SP jumps to level $y+ \text{Exp}_{m_0 \mu_0 + (m_1+1)\mu_1}$ on page $(m_0 - 1, m_1 + 1)$ (i.e., page $(m_0 - 1, c - m_0)$). If $m_0 = 0$, the SP would be on page $(0, c - 1)$. The only exit from the page would be via a downcrossing of level 0. All arrivals would be assigned service rate μ_1 and cause the SP to jump upward but remain on page $(0, c - 1)$; the second term in (4.80) would equal 0 if $m_0 = 0$.

Rate into $((x, \infty), m_0 m_1)$

The SP rate into $((x, \infty), m_0 m_1)$ is

$$\begin{aligned}
&\lambda \frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1 \mu_1} e^{-((m_0+1)\mu_0 + m_1 \mu_1)x} P_{m_0 m_1} \\
&+ \lambda \frac{(m_1 + 1)\mu_1}{m_0 \mu_0 + (m_1 + 1)\mu_1} e^{-(m_0 \mu_0 + (m_1+1)\mu_1)x} P_{m_0 - 1, m_1 + 1} \\
&+ \lambda \frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1 \mu_1} \int_{y=x}^{\infty} f_{m_0+1, m_1-1}(y) dy \\
&+ \lambda \frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1 \mu_1} \int_{y=0}^{x} e^{-((m_0+1)\mu_0 + m_1 \mu_1)(x-y)} f_{m_0+1, m_1-1}(y) dy
\end{aligned}$$

$$+\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy. \tag{4.81}$$

where we have inserted a comma in subscripts like m_0-1, m_1+1, for clarity.

Explanation of Terms in (4.81)

The term

$$\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1}$$

is the rate at which the SP jumps at arrival instants from level 0 on page m_0m_1 into $((x,\infty), m_0m_1)$. At arrival instants customers are assigned service rate μ_0 (wait $=0$), resulting in (m_0+1) rate-μ_0 and m_1 rate-μ_1 customers in service. A rate-μ_0 service finishes first with probability

$$\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1},$$

in which case the SP jumps to page m_0m_1. SP jumps from level 0 over level x have probability $e^{-((m_0+1)\mu_0+m_1\mu_1)x}$ since $\mathcal{S} \underset{dis}{=} \text{Exp}_{(m_0+1)\mu_0+m_1\mu_1}$.

The term

$$\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}e^{-(m_0\mu_0+(m_1+1)\mu_1)x}P_{m_0-1,m_1+1}$$

is the rate at which the SP jumps at arrival instants, from level 0 on page (m_0-1, m_1+1) into $((x,\infty), m_0m_1)$. The arriving customer is assigned service rate μ_0 (wait $=0$), resulting in m_0 rate-μ_0 and (m_1+1) rate-μ_1 customers in service. If a rate-μ_1 service finishes first thereafter, the SP jumps to page m_0m_1; the probability is

$$\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}.$$

SP jumps from level 0 upcross level x with probability $e^{-(m_0\mu_0+(m_1+1)\mu_1)x}$ since the inter-start-of-service depart time $\mathcal{S} \underset{dis}{=} \text{Exp}_{m_0\mu_0+(m_1+1)\mu_1}$.

The term

$$\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=x}^{\infty} f_{m_0+1,m_1-1}(y)dy$$

is the rate at which the SP jumps at arrival instants, out of (x, ∞) on page $(m_0 + 1, m_1 - 1)$ into $((x, \infty), m_0m_1)$. The arriving customer is assigned service rate μ_1 (wait > 0) resulting in $(m_0 + 1)$ rate-μ_0 and m_1 rate-μ_1 customers in service just after the start of service of the arrival. If a rate-μ_0 service finishes first, the SP jumps to page m_0m_1; this has probability

$$\frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1\mu_1}.$$

A jump $\mathcal{S}$ of any size will cause such a jump to enter $((x, \infty), m_0m_1)$ since the start of the jump is already above level x.

The term

$$\lambda \frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1\mu_1} \int_{y=0}^{x} e^{-((m_0+1)\mu_0+m_1\mu()x-y)} f_{m_0+1,m_1-1}(y)dy$$

is the rate at which the SP jumps upward at arrivals, out of

$$((0, x),(m_0 + 1, m_1 - 1)) \text{ into } ((x, \infty), m_0m_1)\,.$$

That is, the SP makes a $(m_0 + 1, m_1 - 1) \to (m_0m_1)$ upcrossing of level x. An arrival is assigned service rate μ_1 (wait > 0). Just after the arrival starts service there are $m_0 + 1$ rate-μ_0 and m_1 rate-μ_1 customers in service. The probability that a rate-μ_0 service finishes first is

$$\frac{(m_0 + 1)\mu_0}{(m_0 + 1)\mu_0 + m_1\mu_1},$$

causing the SP to jump to page m_0m_1. Starting at level $y < x$ the SP will upcross level x if $\mathcal{S} > x - y$; since $\mathcal{S} \underset{dis}{=} \text{Exp}_{(m_0+1)\mu_0+m_1\mu_1}$, this event has probability $e^{-((m_0+1)\mu_0+m_1\mu_1)(x-y)}$.

The term

$$\lambda \frac{(m_1 + 1)\mu_1}{m_0\mu_0 + (m_1 + 1)\mu_1} \int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy$$

is the rate at which the SP jumps at arrival instants from $((0, x), m_0m_1)$ upward into $((x, \infty), m_0m_1)$, i.e., it upcrosses level x on page m_0m_1. Arrivals are assigned service rate μ_1 (wait > 0). Just after the arrival starts service there are m_0 rate-μ_0 and $(m_1 + 1)$ rate-μ_1 customers in service. A rate-μ_1 service ends first with probability

$$\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}.$$

causing the SP to jump to page m_0m_1. If the SP starts at level y it will upcross level x provided $\mathcal{S} > x - y$; since $\mathcal{S} \underset{dis}{=} \text{Exp}_{m_0\mu_0+(m_1+1)\mu_1}$, this event has probability $e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}$.

Writing Equations for Positive-Wait States

The model equation for the positive-wait states on page m_0m_1 is written by using the principle of rate balance with respect to set $((x,\infty), m_0m_1)$, *exit rate* = *entrance rate*. Equating exit rate (4.80) and entrance rate (4.81) gives

$$\begin{aligned}
&f_{m_0m_1}(x) + \lambda\frac{m_0\mu_0}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=x}^{\infty} f_{m_0m_1}(y)dy\\
&= \lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1}\\
&\quad+\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}e^{-(m_0\mu_0+(m_1+1)\mu_1)x}P_{m_0-1,m_1+1}\\
&\quad+\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=x}^{\infty} f_{m_0+1,m_1-1}(y)dy\\
&\quad+\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=0}^{x} e^{-((m_0+1)\mu_0+m_1\mu_1)(x-y)} f_{m_0+1,m_1-1}(y)dy\\
&\quad+\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy.
\end{aligned} \tag{4.82}$$

Equation for "Cover"

The total probability of a zero wait is

$$P_0 = \sum_{\boldsymbol{m}\in \boldsymbol{M}_0\cup \boldsymbol{M}_b} P_{\boldsymbol{m}} = \sum_{0\le m_0+m_1\le c-1} P_{m_0m_1}. \tag{4.83}$$

The total pdf of wait is

$$f(x) = \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} f_{\boldsymbol{m}}(x) = \sum_{m_0+m_1=c-1} f_{m_0m_1}(x),\ x>0. \tag{4.84}$$

Let $x > 0$ be fixed. The total SP downcrossing rate of x is $f(x)$. The total SP upcrossing rate of x due to jumps starting from level 0 at arrival instants, is

$$\lambda \sum_{m_0+m_1=c-1} e^{-((m_0+1)\mu_0+m_1\mu_1)x} P_{m_0m_1}.$$

The total SP upcrossing rate of x due to jumps starting from levels $y \in (0, x)$ at arrival instants, is

$$\lambda \sum_{m_0+m_1=c-1} \int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy.$$

Rate balance across level x gives the model equation for the cover,

$$\begin{aligned} f(x) = \lambda &\sum_{m_0+m_1=c-1} e^{-((m_0+1)\mu_0+m_1\mu_1)x} P_{m_0m_1} \\ &+ \lambda \sum_{m_0+m_1=c-1} \int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy. \end{aligned} \tag{4.85}$$

Normalizing Condition
The normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ can be expressed as

$$\sum_{0\leq m_0+m_1\leq c-1} P_{m_0m_1} + \sum_{m_0+m_1=c-1} \int_{x=0}^{\infty} f_{m_0m_1}(x)dx = 1. \tag{4.86}$$

4.10.3 Solution of Model Equations

In Sect. 4.11 below, we formulate and solve the foregoing M/M/2 model with zero-wait customers receiving exceptional service, whose solution illustrates relevant SPLC ideas and related insights. A more general solution procedure of a two-server M/M/2 queue where service time depends on waiting time in *a general manner* is detailed in Chap. 4 of [11].

4.11 M/M/2: Zero-Waits Get Special Service

M/M/2/(μ_0, μ_1), $(0, (0, \infty))$

To fix ideas and clarify the system dynamics of M/M/c with special service for zero-wait customers, we formulate the model with $c = 2$ servers. We discuss the solution for the zero-wait probabilities and the positive-wait pdfs. We denote the model by M/M/2/(μ_0, μ_1) , (0, (0, ∞)). This notation indicates that 0-wait arrivals get service rate μ_0 and (0, ∞)-wait arrivals get service rate μ_1; diagrammatically, $\mu_0 \leftrightarrow$ 0-wait, $\mu_1 \leftrightarrow$ (0, ∞)-wait.

There are only three zero-wait states in $\boldsymbol{M}_0 \cup \boldsymbol{M}_b$ (compare Sect. 4.10),

$$\{(0, m_0 m_1)\} = \{(0, 00), (0, 10), (0, 01)\}.$$

Denote the steady-state probabilities of the zero-wait states by P_{00}, P_{10}, P_{01} respectively.

For example, state (0, 10) indicates that an arrival would wait 0 and would "see" a rate-μ_0 customer being served by the *other* server. The arrival would be assigned rate μ_0 since it waits 0. There would then be two rate-μ_0 customers in service. The inter start-of-service depart time $\mathcal{S}$ would be $\underset{dis}{=}$ $\text{Exp}_{2\mu_0}$.

There are only two zero-wait states such that $m_0 + m_1 = 1$ (both border states). Denote the pdfs of the positive-wait states $(x, 10)$, $(x, 01)$, by $f_{10}(x)$, $f_{01}(x)$, $x > 0$, respectively. A would-be arrival that finds the state $(x, 10)$, $x > 0$, for example, would wait x before service, and be assigned service rate μ_1 (wait > 0). Just after its start of service, it would have a rate-μ_0 customer as neighbor in the other server. Inter start-of-service depart time $\mathcal{S}$ $\underset{dis}{=}$ $\text{Exp}_{\mu_0+\mu_1}$. The rate-μ_1 customer would finish service first with probability $\frac{\mu_1}{\mu_0+\mu_1}$, leaving the rate-$\mu_0$ customer in service ($m_0 m_1 = 10$). The rate-μ_0 customer would finish service first with probability $\frac{\mu_0}{\mu_0+\mu_1}$, leaving the rate-$\mu_1$ customer in service ($m_0 m_1 = 01$).

If an arrival "sees" state $(x, 01)$, $x > 0$, $\mathcal{S}$ would be $\underset{dis}{=}$ $\text{Exp}_{2\mu_1}$. The first customer to complete service would have rate μ_1 with certainty. The customer remaining in service just after that service completion would have service rate μ_1 ($m_0 m_1 = 01$).

A sample-path diagram of the virtual wait process $\{W(t)\}_{t \geq 0}$, has three lines and two pages (Fig. 4.6).

The total (marginal) probability of a zero wait is

$$P_0 = P_{00} + P_{10} + P_{01}.$$

The total pdf of wait is

$$f(x) = f_{10}(x) + f_{01}(x), x > 0.$$

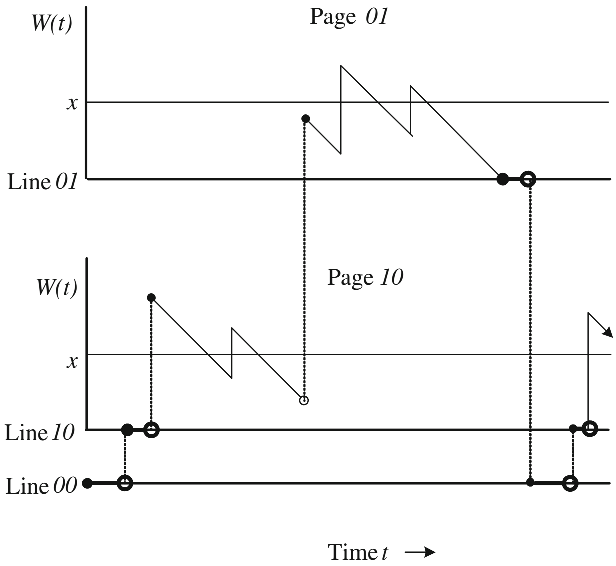


Fig. 4.6 Sample path of virtual wait in M/M/2/(μ_0, μ_1), $(0, (0, \infty))$. Lines for states $(0, 10)$, $(0, 01)$ are at level 0 of corresponding pages. Line for state $(0, 00)$ is isolated. The SP can enter state $(0, 01)$ only by downcrossing level 0 on page 01 (See Fig. 4.10.)

4.11.1 Model Equations

Zero-Wait States

Applying *SP exit rate* = *SP entrance rate* for the zero-wait states $(0, 00)$, $(0, 10)$, $(0, 01)$ gives, respectively,

$$\begin{aligned} \lambda P_{00} &= \mu_0 P_{10} + \mu_1 P_{01}, \\ (\lambda + \mu_0) P_{10} &= \lambda P_{00} + f_{10}(0), \\ (\lambda + \mu_1) P_{01} &= f_{01}(0). \end{aligned} \tag{4.87}$$

In (4.87), the terms $f_{10}(0)$, $f_{01}(0)$ (same as $f_{10}(0^+)$, $f_{01}(0^+)$) are the rates at which the SP hits level 0 from above on pages 10 and 01 respectively. Immediately following such hits, the SP moves on lines 10 and 01 respectively.

Positive-Wait States

Applying *SP exit rate* = *SP entrance rate* for $((x, \infty), 10)$ (on page 10) yields the integral equation

$$
\begin{aligned}
& f_{10}(x) + \lambda \int_{y=x}^{\infty} \frac{\mu_0}{\mu_0+\mu_1} f_{10}(y)dy \\
& = \lambda P_{10} e^{-2\mu_0 x} + \lambda \frac{\mu_1}{\mu_0+\mu_1} P_{01} e^{-(\mu_0+\mu_1)x} \\
& \quad + \lambda \frac{\mu_1}{\mu_0+\mu_1} \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} f_{10}(y)dy, x > 0.
\end{aligned}
\tag{4.88}
$$

When formulating equation (4.88), note that the SP cannot jump directly from a positive-wait state on page 01 into set $((x,\infty),10)$. An arrival that "sees" state $(y,01)$, $y>0$, will be assigned rate μ_1 and start service after a wait y; its neighbor in the other server will also have service rate μ_1 (because $m_0 m_1 = 01$). The random variable $\mathcal{S}$ will be distributed as $\text{Exp}_{2\mu_1}$, and the remaining customer in service just after the first departure thereafter, will have rate μ_1. At the arrival instant, the SP will start a jump at level y on page 01, which ends at level $y+\mathcal{S} \underset{dis}{=} y + \text{Exp}_{2\mu_1}$, also on page 01. The configuration remains 01 just after the arrival. The only exit route from page 01 is via a downcrossing of level 0 (continuous hit of 0 from above—see Fig. 4.6).

Now we balance the SP exit and entrance rates for $((x,\infty),01)$ (page 01), giving integral equation

$$
\begin{aligned}
f_{01}(x) = {} & \lambda \frac{\mu_0}{\mu_0+\mu_1} P_{01} e^{-(\mu_0+\mu_1)x} \\
& + \lambda \int_{y=0}^{x} e^{-2\mu_1(x-y)} f_{01}(y)dy \\
& + \lambda \frac{\mu_0}{\mu_0+\mu_1} \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} f_{10}(y)dy \\
& + \lambda \frac{\mu_0}{\mu_0+\mu_1} \int_{y=x}^{\infty} f_{10}(y)dy.
\end{aligned}
\tag{4.89}
$$

When formulating (4.89), note that the SP can exit $((x,\infty),01)$ only by downcrossing level x. Also, the SP cannot enter $((x,\infty),01)$ from state $(0,10)$ at arrivals, since all jumps that start from line 10 (corresponding to state $(0,10)$) must end on page 10, at an ordinate $\underset{dis}{=} \text{Exp}_{2\mu_0}$.

The equation for the total pdf is

$$
f(x) = f_{10}(x) + f_{01}(x),
$$

as viewed from the "cover", the result of projecting sample-path segments on pages 10 and 01 onto a single sheet. An integral equation for $f(x)$ is obtained by balancing the SP *total* down- and upcrossing rates of level $x > 0$. This is equivalent to equating the exit and entrance rates for the state-space set

$$((x, \infty), 10) \cup ((x, \infty), 01).$$

The resulting equation is

$$\begin{aligned} f(x) &= \lambda P_{10} e^{-2\mu_0 x} + \lambda P_{01} e^{-(\mu_0+\mu_1)x} \\ &\quad + \lambda \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} f_{10}(y)dy \\ &\quad + \lambda \int_{y=0}^{x} e^{-2\mu_1(x-y)} f_{01}(y)dy, \; x > 0. \end{aligned} \tag{4.90}$$

Equation (4.90) can also be derived by summing the corresponding sides of (4.88) and (4.89). However, it is intuitive and instructive to interpret equation (4.90) as total SP rate-balance across level $x > 0$.

The normalizing condition is

$$P_{00} + P_{10} + P_{01} + \int_{x=0}^{\infty} f_{10}(x)dx + \int_{x=0}^{\infty} f_{01}(x)dx = 1,$$

or

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{4.91}$$

4.11.2 Solution of Equations

Equation (4.88) is an integral equation in $f_{10}(x)$, which is not confounded by the presence of $f_{01}(x)$; so we utilize it to obtain the functional form of $f_{10}(x)$. Applying differential operator $\langle D\rangle\langle D + \mu_0 + \mu_1\rangle$ to both sides of (4.88) leads to the second order differential equation

$$\begin{aligned} &f_{10}''(x) + (\mu_0 + \mu_1 - \lambda) f_{10}'(x) - \lambda\mu_0 f_{10}(x) \\ &= 2\lambda\mu_0(\mu_0 - \mu_1) P_{10} e^{-2\mu_0 x}, \; x > 0. \end{aligned} \tag{4.92}$$

with solution

$$f_{10}(x) = C_{10}e^{ax} + C_{10}^1 e^{bx} + \lambda K_{10} P_{10} e^{-2\mu_0 x},\ x > 0,$$

where a and b are the roots of the auxiliary quadratic equation, namely

$$a = \frac{1}{2}\left(\lambda - \mu_0 - \mu_1 - \sqrt{\lambda^2 + 2\lambda\mu_0 - 2\lambda\mu_1 + \mu_0^2 + 2\mu_0\mu_1 + \mu_1^2}\right) < 0,$$
$$b = \frac{1}{2}\left(\lambda - \mu_0 - \mu_1 + \sqrt{\lambda^2 + 2\lambda\mu_0 - 2\lambda\mu_1 + \mu_0^2 + 2\mu_0\mu_1 + \mu_1^2}\right) > 0,$$
$$K_{10} = \frac{2(\mu_0 - \mu_1)}{\lambda + 2\mu_0 - 2\mu_1},$$

and C_{10}, C_{10}^1 are constants of integration. A necessary condition for system stability is $\lim_{x\to\infty} f_{10}(x) = 0$, which implies $C_{10}^1 = 0$ (since $b > 0$). Thus the functional form of $f_{10}(x)$ is

$$f_{10}(x) = C_{10}e^{ax} + \lambda K_{10} P_{10} e^{-2\mu_0 x}, x > 0, \tag{4.93}$$

where C_{10} is a constant to be determined.

The term K_{10} will be undefined if $\lambda + 2\mu_0 - 2\mu_1 = 0$. If $\lambda + 2\mu_0 - 2\mu_1 \neq 0$ and $\mu_0 - \mu_1 \neq 0$, then K_{10} may be positive or negative. If $\mu_0 - \mu_1 = 0$ the model reduces to a standard M/M/c queue with $c = 2$ (Sect. 4.8); the computed distribution of wait should then match that of a standard M/M/2 queue. (We will utilize this property later as a mild check on the correctness of the present solution.)

We obtain the functional form of $f_{01}(x)$ by substituting the expression for $f_{10}(x)$ (4.93) into (4.90). Since

$$f_{01}(x) = f(x) - f_{10}(x),$$

this substitution gives the integral equation

$$\begin{aligned} f_{01}(x) = {} & \lambda(1 - K_{10})P_{10}e^{-2\mu_0 x} + \lambda P_{01} e^{-(\mu_0+\mu_1)x} - C_{10}e^{ax} \\ & + \lambda \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} (C_{10}e^{ay} + \lambda K_{10} P_{10} e^{-2\mu_0 y}) dy \\ & + \lambda \int_{y=0}^{x} e^{-2\mu_1(x-y)} f_{01}(y) dy. \end{aligned} \tag{4.94}$$

The first integral term in (4.94) is

$$\lambda\int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)}(C_{10}e^{ay}+\lambda K_{10}P_{10}e^{-2\mu_0 y})dy$$
$$=\frac{\lambda C_{10}}{\mu_0+\mu_1+a}e^{ax}-\frac{\lambda^2 K_{10}P_{10}e^{-2\mu_0 x}}{\mu_0-\mu_1}$$
$$-\left(\frac{\lambda C_{10}}{\mu_0+\mu_1+a}-\frac{\lambda^2 K_{10}P_{10}}{\mu_0-\mu_1}\right)e_1^{-(\mu_0+\mu_1)x}.$$

Thus (4.94) is equivalent to the integral equation

$$f_{01}(x)=H_{01}C_{10}e^{ax}+\lambda B_{01}P_{10}e^{-2\mu_0 x}+D_{01}e^{-(\mu_0+\mu_1)x}+\lambda\int_{y=0}^{x}e^{-2\mu_1(x-y)}f_{01}(y)dy, \tag{4.95}$$

where

$$H_{01}=\frac{\lambda}{\mu_0+\mu_1+a}-1,$$
$$B_{01}=1-K_{10}-\frac{\lambda K_{10}}{\mu_0-\mu_1},$$
$$D_{01}=\lambda P_{01}-\frac{\lambda C_{10}}{\mu_0+\mu_1+a}+\frac{\lambda^2 K_{10}P_{10}}{\mu_0-\mu_1}.$$

Applying the differential operator $\langle D+2\mu_1\rangle$ to both sides of (4.95) yields the differential equation for $f_{01}(x)$,

$$f_{01}'(x)+(2\mu_1-\lambda)f_{01}(x)=(2\mu_1+a)H_{01}C_{10}e^{ax}+2\lambda(\mu_1-\mu_0)B_{01}P_{10}e^{-2\mu_0 x}+(\mu_1-\mu_0)D_{01}e^{-(\mu_1+\mu_0)x}. \tag{4.96}$$

whose solution is

$$f_{01}(x)=\frac{2\lambda(\mu_1-\mu_0)}{2\mu_1-\lambda-2\mu_0}B_{01}P_{10}e^{-2\mu_0 x}+\frac{\mu_1-\mu_0}{\mu_1-\lambda-\mu_0}D_{01}e^{-(\mu_1+\mu_0)x}$$

$$+ \frac{2\mu_1 + a}{2\mu_1 - \lambda + a} H_{01} C_{10} e^{ax}$$
$$+ C_{01} e^{-(2\mu_1 - \lambda)x}, \tag{4.97}$$

where C_{01} is a constant of integration to be determined (see Sect. 4.11.4).

4.11.3 Stability Condition

Consider the functional forms of $f_{10}(x)$ and $f_{01}(x)$ in (4.93) and (4.97). In the exponents, all the coefficients of x are negative except possibly the coefficient $-(2\mu_1 - \lambda)$ in $e^{-(2\mu_1-\lambda)x}$ of (4.97). A necessary condition for stability is that

$$f_{10}(\infty) = f_{01}(\infty) = f(\infty) = 0;$$

implying $-(2\mu_1 - \lambda) < 0$, equivalent to $\lambda < 2\mu_1$. That is, the arrival rate must be less than the system departure rate when both servers are occupied by positive-wait customers, regardless how large x is. Thus, for stability, if the waiting time is large and customers are arriving, then the mean inter-arrival time should exceed the mean inter-departure time. This ensures that the waiting time will return to zero in a finite time.

4.11.4 Determination of Constants

A complete solution for the distribution of wait requires the values of five unknown constants

$$P_{00},\ P_{10},\ P_{01},\ C_{10},\ C_{01},$$

which we obtain from five independent equations.

In (4.93) letting $x \downarrow 0$ to obtain $f_{10}(0)$, and referring to (4.87) gives

$$C_{10} + \lambda K_{10} P_{10} = (\lambda + \mu_0) P_{10} - \lambda P_{00}. \tag{4.98}$$

In (4.97) letting $x \downarrow 0$ to obtain $f_{01}(0)$ gives

$$f_{01}(0) = \frac{2\lambda(\mu_1 - \mu_0)}{2(\mu_1 - \mu_0) - \lambda} B_{01} P_{10}$$
$$+ \frac{\mu_1 - \mu_0}{\mu_1 - \mu_0 - \lambda} D_{01}$$
$$+ \frac{2\mu_1 + a}{2\mu_1 + a - \lambda} H_{01} C_{10} + C_{01}. \tag{4.99}$$

Substituting $f_{01}(0)$ from (4.99) into (4.87) gives

$$C_{01} = (\lambda + \mu_1)P_{01} - \frac{2\lambda(\mu_1 - \mu_0)}{2(\mu_1 - \mu_0) - \lambda} B_{01} P_{10} - \frac{\mu_1 - \mu_0}{\mu_1 - \mu_0 - \lambda} D_{01} - \frac{2\mu_1 + a}{2\mu_1 + a - \lambda} H_{01} C_{10}. \tag{4.100}$$

We get another independent equation by substituting the functional form

$$f_{10}(x) = C_{10}e^{ax} + \lambda K_{10} P_{10} e^{-2\mu_0 x}$$

into the integral equation (4.88) and equating the coefficients of corresponding exponential terms on both sides after evaluating the integral (different exponentials are linearly independent—see, e.g., Sect. 3.3, pp. 99ff and p. 205 in [10]). The coefficient of $e^{-(\mu_0+\mu_1)x}$ on the right side of (4.88) must be 0. This yields the linear equation

$$\frac{\lambda\mu_1}{\mu_0 + \mu_1} P_{01} - \frac{1}{\mu_0 + \mu_1 + a} - \frac{\lambda K_{10}}{\mu_1 - \mu_0} P_{10} = 0. \tag{4.101}$$

The normalizing condition is

$$1 = P_{00} + P_{10} + P_{01} + \frac{C_{10}}{(-a)} + \frac{\lambda K_{10} P_{10}}{2\mu_0} + \frac{\lambda(\mu_1 - \mu_0)}{\mu_0(2(\mu_1 - \mu_0) - \lambda)} B_{01} P_{10} + \frac{\mu_1 - \mu_0}{(\mu_1 + \mu_0)(\mu_1 - \mu_0 - \lambda)} D_{01} + \frac{2\mu_1 + a}{(-a)(2\mu_1 + a - \lambda)} H_{01} C_{10} + \frac{1}{2\mu_1 - \lambda} C_{01}. \tag{4.102}$$

We now have a set of five equations to solve for the five constants: from (4.87)

$$\lambda P_{00} = \mu_0 P_{10} + \mu_1 P_{01},$$

and (4.98), (4.100), (4.101), (4.102).

Remark 4.17 In the derivation of the functional forms of $f_{10}(x)$, $f_{01}(x)$ the expressions

$$\mu_1 - \mu_0, \quad 2\mu_1 - \lambda - 2\mu_0, \quad \mu_1 - \lambda - \mu_0, \quad 2\mu_1 - \lambda + a$$

appear in various denominators. If any of these four expressions were equal to 0, the functional forms would have to be modified. The five equations used to solve for the constants in the present model would have to be modified accordingly. In this monograph we emphasize the system-point level-crossing approach to derive model equations, and various techniques to solve them. However, there are many techniques to solve systems of integral equations, requiring additional study, outside the scope of the present volume. We give **numerical** solutions of the equations in several examples below.

Remark 4.18 It would be interesting to explain the appearance of the immediately above expressions in the denominators. Does the system reduce to a particular queueing model when a denominator is equal to 0? For example, when $\mu_1 - \mu_0 = 0$, the M/M/2/(μ_0, μ_1) , (0, (0, ∞)) system reduces to a standard M/M/2 model. In M/M/2/(μ_0, μ_1) , (0, (0, ∞)) the only criterion necessary for stability is $\lambda < 2\mu_1$. What do these exceptional denominators mean with regard to physical models?

Another question is how to select a set of linearly independent equations to solve for the constants. Once a set of equations is derived, it can be checked for independence using matrix methods. But this amounts to trial and error. Is there a way to derive five independent equations directly? Taking derivatives may be the answer to this question.

Example 4.7 We first give a mild numerical check on the five equations by letting $\mu_1 - \mu_0 = 0$. In this case M/M/2/(μ_0, μ_1) , (0, (0, ∞)) reduces to a standard M/M/2 queue. We arbitrarily take

$$\lambda = 1,\ \ \mu_0 = 1.5,\ \ \mu_1 = 1.5.$$

Then $a = -2.581139$. The solution for the constants is

$$C_{10} = 0.0,\ \ P_{01} = .133333,\ \ P_{10} = .20,\ \ C_{01} = .333333,\ \ P_{00} = .50.$$

We compare this solution with that of the standard M/M/2 queue with $\lambda = 1$, $\mu = 1.5$. In M/M/2, the probability of an empty system is $P_0 = 0.5$. The probability of 1 customer in the system *is indeed* $P_1 = 0.33333$. The values match P_{00} and $P_{10} + P_{01}$ in M/M/2/(μ_0, μ_1) , (0, (0, ∞)) model, as expected.

Also, in M/M/2/(μ_0, μ_1) , (0, (0, ∞)), we see from (4.97) that

$$\begin{aligned} f_{01}(x) &= C_{01}e^{-(2\mu_1-\lambda)x} \\ &= \lambda P_1 e^{-(2\mu_1-\lambda)x} \\ &= 1 \cdot (0.33333)e^{-2x}, x > 0, \end{aligned}$$

since $\mu_1 - \mu_0 = 0$ and $C_{10} = 0$.

Example 4.8 Let $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$. These values preclude that any of the four above-mentioned denominators is 0. We get $a = -2.715136$. We solve the equations and obtain

$$P_{00} = .417715,\ P_{10} = 0.339103,\ P_{01} = 0.0202270,$$
$$C_{01} = 0.022818,\ C_{10} = -0.322655.$$

The functions $f_{10}(x), x > 0$, and $f_{01}(x), x > 0$, are linear combinations of exponentials,

$$\begin{aligned} f_{10}(x) &= -0.322655e^{-2.715136x} + 0.617056e^{-2.2x}, \\ f_{01}(x) &= 0.505784e^{-2.2x} + 0.067831e^{-3.31x} \\ &\quad - 0.531504e^{-2.715136x} + 0.022818e^{-3.42x}. \end{aligned}$$

We substitute the values of $P_{00}, P_{10}, P_{01}, f_{10}(x), f_{01}(x)$ into the normalizer (4.91), and obtain 1; it checks.

The partial pdfs of wait $f_{10}(x)$, $f_{01}(x)$ and total pdf of wait $f(x)$ are depicted in Figs. 4.7, 4.8, and 4.9 respectively.

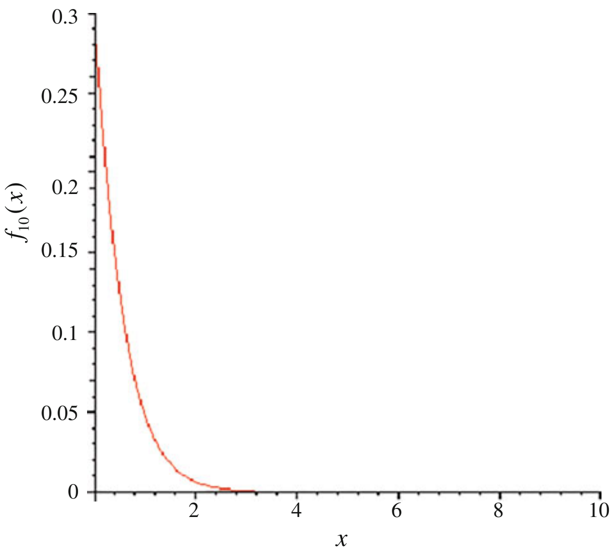


Fig. 4.7 Partial pdf of wait $f_{10}(x)$ in M/M/2/(μ_0, μ_1), $(0, (0, \infty))$. $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$

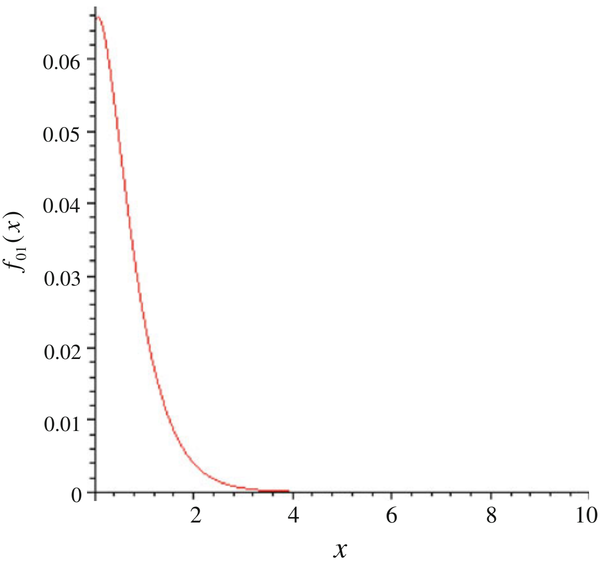
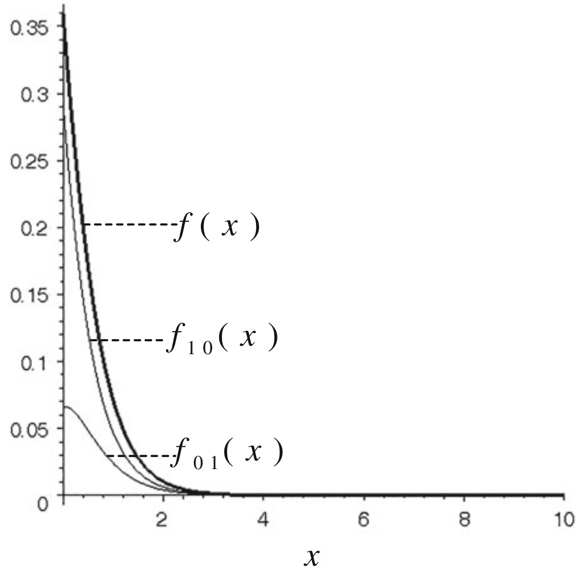


Fig. 4.8 Partial pdf of wait $f_{01}(x)$ in M/M/2/(μ_0, μ_1), (0, (0, ∞)). $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$

Fig. 4.9 Total pdf of wait $f(x) = f_{10}(x) + f_{01}(x)$ in M/M/2/(μ_0, μ_1), (0, (0, ∞)). $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$



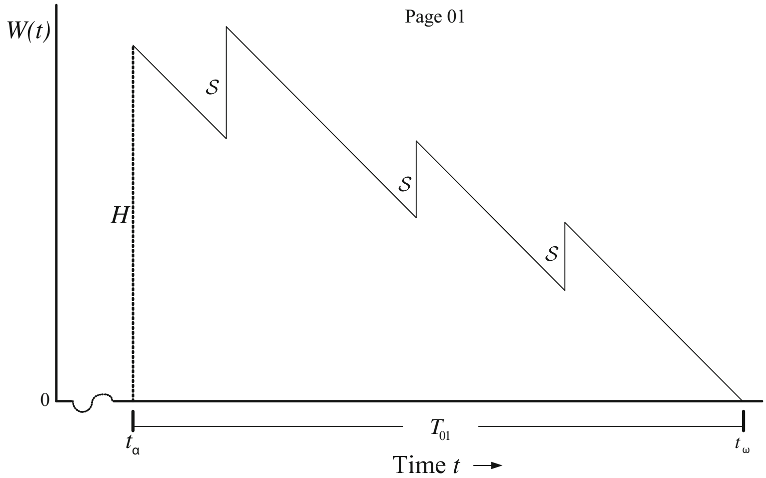


Fig. 4.10 T_{01} := sojourn on page 01. t_α := start of T_{01}, t_ω := end of T_{01}. $\mathcal{S} \underset{dis}{=} \text{Exp}_{2\mu_1}$. (See Fig. 4.6.)

4.11.5 Expected Sojourn Time on a Page

Consider page 01. The SP can enter page 01 from discrete state $(0, 01)$ or from page 10, due to a jump at an arrival (Fig. 4.6). It cannot enter directly from state (0, 10) at an arrival instant, since zero-wait arrivals are assigned rate μ_0 resulting in both servers being occupied with rate-μ_0 customers; so any SP jump to a positive level must end on page 10.

In a sojourn on page 01, the first inter start-of-service depart time will be $\underset{dis}{=} \text{Exp}_{\mu_0+\mu_1}$; any other inter start-of-service depart times that follow while on page 01 will be $\underset{dis}{=} \text{Exp}_{2\mu_1}$. While the SP is on page 01, each departure will leave a rate-μ_1 customer in the neighboring occupied server. Given that the SP enters page 01, its source state was $(0, 01)$ with probability (using Bayes' rule)

$$q = \frac{P_{01}}{P_{01} + \int_{y=0}^{\infty} f_{10}(y)dy}.$$

Its source was composite state $((0, \infty), 10)$ with probability

$$1 - q = \frac{\int_{y=0}^{\infty} f_{10}(y)dy}{P_{01} + \int_{y=0}^{\infty} f_{10}(y)dy}.$$

Let H denote the height above level 0 (ordinate) at which the SP enters page 01 (see Fig. 4.10). A sojourn on page 01 starts at level H, where

$$E(H|\text{source is } (0,01)) = \frac{1}{\mu_0+\mu_1},$$

and

$$E(H|\text{source is level } y \text{ on page } 10) = y + \frac{1}{\mu_0+\mu_1}, y>0,$$

since the size of a jump from either source onto page 01 is $\underset{dis}{=} \text{Exp}_{\mu_0+\mu_1}$. Thus

$$E(H) = \frac{1}{\mu_0+\mu_1}\cdot q + \left(\int_{y=0}^{\infty}\left(y+\frac{1}{\mu_0+\mu_1}\right) f_{10}(y)dy\right)\cdot(1-q)\,.$$

From (4.93) $f_{10}(y)$ is given by

$$f_{10}(y) = C_{10}e^{ay} + \lambda K_{10}P_{10}e^{-2\mu_0 y}, y>0,$$

and thus

$$\begin{aligned}
E(H) &= \frac{1}{\mu_0+\mu_1}\cdot q \\
&\quad + \left(\int_{y=0}^{\infty}\left(y+\frac{1}{\mu_0+\mu_1}\right)\left(C_{10}e^{ay}+\lambda K_{10}P_{10}e^{-2\mu_0 y}\right)dy\right)\cdot(1-q) \\
&= \frac{1}{\mu_0+\mu_1}\cdot q + \Bigg(\frac{1}{4}\Big(4C_{10}\mu_0^2\mu_1 + 4C_{10}\mu_0^3 \\
&\quad + 3\lambda K_{10}P_{10}a^2\mu_0 + \lambda K_{10}P_{10}a^2\mu_1 \\
&\quad - 4C_{10}a\mu_0^2\Big)\Big/\left(a^2\mu_0^2(\mu_0+\mu_1)\right)\Bigg)\cdot(1-q).
\end{aligned} \tag{4.103}$$

Let T_{01} denote a sojourn time on page 01, i.e., the time from SP entrance until the first exit from page 01 thereafter. The only possible exit is due to a downcrossing of level 0 (Fig. 4.6). Thus

$$T_{01} = H + \sum_{i=1}^{N_H}\mathcal{B}_i$$

where N_H is the number of arrivals during time H and $\mathcal{B}_i$ represents a busy period of an *M/M/1 queue* with service rate $2\mu_1$, since both servers are busy with rate-μ_1 customers. (See Sect. 3.4.12 and Fig. 3.6.) The expected busy period is obtained from (3.120) with $2\mu_1$ substituted for μ. Thus

$$E(\mathcal{B}_i) = \frac{1}{2\mu_1-\lambda}, i=1,\ldots,N_H.$$

The r.v.s N_H and $\mathcal{B}_i, i = 1, \ldots, N_H$ are independent, since the $\mathcal{B}_i$s are i.i.d. each distributed as an $\mathrm{M}_\lambda/\mathrm{M}_{2\mu_1}/1$ busy period. The expected sojourn time on page 01 is

$$E(T_{01}) = E(H) + E\left(\sum_{i=1}^{N_H} \mathcal{B}_i\right) = E(H) + E(N_H)E(\mathcal{B}_i)$$
$$= E(H) + \lambda E(H)\frac{1}{2\mu_1 - \lambda} = \frac{E(H)}{1 - \lambda/(2\mu_1)}, \tag{4.104}$$

where $E(H)$ is given in formula (4.103). It is noteworthy that T_{01} is distributed as the *busy period* in an $\mathrm{M}_\lambda/\mathrm{M}_{2\mu_1}/1$ queue in which zero-wait arrivals obtain special service $\underset{dis}{=} H$, and positive-wait arrivals get service rate $2\mu_1$. This structure of T_{01} illustrates an interesting application, and the versatility, of the M/G/1 queue where zero-wait arrivals get exceptional service (see Sect. 3.6.1).

Example 4.9 In Example 4.8 with $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$, we obtain

$$q = .111216, \quad 1 - q = .888784, \quad E(H) = 0.151416.$$

The expected sojourn time on page 01 is $E(T_{01}) = 0.195689$.

Remark 4.19 Various questions arise regarding Example 4.9. What is the *proportion* of time that the SP spends circulating on page 01, page 10, or in the zero-wait states? Can this question be answered for a general M/M/c/(μ_0, μ_1) , (0, (0, ∞)) queue with $c > 2$? If yes, then it would be straightforward to determine P_{00}. This would facilitate solving for all the zero-wait probabilities and the partial pdfs of wait.

4.12 M/M$_i$/c with Reneging

Consider an M/M/c queue, with $c \geq 2$ *distinguishable* servers having fixed exponential service rates μ_i, $i = 1, \ldots, c$. Thus, the queue has ***heterogeneous servers***. This model is denoted by M/M$_i$/c. Using the notation for the generalized M/M/c model (Sects. 4.3, 4.4 and 4.5), let $\{W(t), M(t)\}_{t\geq 0}$ denote the *system point process*, where $W(t) :=$ *virtual wait* at time t and $M(t) :=$ *system configuration* at time t (see Sect. 4.5). The set of possible exponential service rates is $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_c\}$. A new arrival receives one of those service rates, depending on which server it engages. We assume the

μ_is are distinct. When some or all of the μ_is are equal, the analysis is similar with slight modification.

Assume zero-wait arrivals start service immediately (no balking). In general, the zero-wait server-assignment policy is arbitrary. When formulating equations for the zero-wait probabilities in a specific model, however, we must specify a zero-wait server-assignment policy (see Sect. 4.12.7 below).

4.12.1 Staying Function

Let $\{\tau_n\}_{n=1,2,...}$ be the arrival times of customers C$_n$, $n = 1, 2, \ldots$, respectively. Then $W(\tau_n^-) \equiv W_n :=$ *required wait* before start of service of C$_n$.

Define

$$\theta_n = \begin{cases} 1 \text{ if } C_n \text{ stays for a full service} \\ 0 \text{ if } C_n \text{ reneges while waiting for service} \end{cases}, n = 1, 2, \ldots.$$

With respect to the steady-state statistical properties of the waiting time, this model is equivalent to one in which customers balk from joining the system at arrival instants, depending on their required wait before service, i.e., on their arrival-point W_ns. (See a sample path of $\{W(t)\}_{t\geq 0}$ in Fig. 3.24 for a similar M/G/1 model with reneging.)

We define the *staying function* $\overline{R}(\cdot)$ similarly as in Sect. 3.13.1. For each $y \geq 0$, define the *conditional* probabilities

$$\overline{R}(y) \equiv P(\theta_n = 1|W_n = y), \quad R(y) \equiv P(\theta_n = 0|W_n = y),$$

independent of $n = 1, 2, \ldots$. Note that $\overline{R}(0) = 1$, and $\overline{R}(y) + R(y) = 1$, $y \geq 0$.

For each $y \geq 0$, given $W_n = y$, θ_n has a Bernoulli distribution (e.g., p. 26 in [125]). The staying function $\overline{R}(y)$ is the conditional probability of an arrival staying for a full service, given $W_n = y$. Its complement $R(y)$ is the probability of an arrival reneging while in the waiting line, given $W_n = y$.

Using the foregoing definition, $1 - \overline{R}(y)$, $y \geq 0$, is not necessarily a cdf.

We assume: $\overline{R}(0) = 1$; $\overline{R}(y)$, $y \geq 0$, is monotone decreasing in the wide sense (i.e., not strictly monotone—it may be non-increasing); $\overline{R}(y)$, $y > 0$, is *bounded from below* by 0. $\overline{R}(y)$ may be continuous or piecewise continuous; it may be a step function.

Due to boundedness from below and monotonicity, $\lim_{y\to\infty} \overline{R}(y)$ exists. Let

$$\lim_{y\to\infty} \overline{R}(y) = L, 0 \le L \le 1.$$

If $\overline{R}(y) \equiv 1$, $y \ge 0$, the model reverts to a standard M/M$_i$/c queue with no reneging; in that case L= 1 (see Sect. 3.13 and Theorem 3.8.)

4.12.2 System Configuration

The set of possible system configurations is

$$\boldsymbol{M} = \boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{\boldsymbol{m}|(m_1, m_2, \ldots, m_c)|0 \le \sum_{i=1}^{c} m_i \le c-1\},$$

where $m_i = \begin{cases} 1 \text{ if server } i \text{ is occupied} \\ 0 \text{ if server } i \text{ is idle} \end{cases}$, *just after* a start of service in some server, since the configuration represents the service rates of those servers *other than* the one just occupied.

There are $\binom{c}{j}$ configurations in which exactly j servers are occupied (i.e., $\sum_{i=1}^{c} m_i = j$). The total number of configurations in $\boldsymbol{M}$ is

$$\sum_{j=0}^{c-1} \binom{c}{j} = 2^c - 1.$$

The number of configurations in $\boldsymbol{M}_0 := \{\boldsymbol{m}|0 \le \sum_{i=1}^{c} m_i \le c-2\}$, is $2^c - 1 - c$. The number of configurations in $\boldsymbol{M}_1 := \{\boldsymbol{m}| \sum_{i=1}^{c} m_i = c-1\}$ (border configurations), is c. (Recall that $\boldsymbol{M}_1 = \boldsymbol{M}_b$.)

4.12.3 State of System and Sample Path

State of System
Denote the state of the system as $\{W(t), M(t)\}_{t\ge0}$, where $W(t) \ge 0 :=$ *virtual wait*, and $M(t) \in \boldsymbol{M} :=$ *system configuration*, at instant t.

Sample Path
Consider a sample path of $\{(W(t), M(t))\}_{t\ge0}$. A sample-path diagram has $2^c - 1$ lines corresponding to the zero-wait states $(0, m)$, $m \in \boldsymbol{M}$ (i.e., $W(t) = 0$); and c sheets (pages) corresponding to the positive-wait states (y, m), $y > 0$ (i.e., $W(t) > 0$). (See Fig. 4.11 for the special case $c = 2$.)

Assume the system starts empty at $t = 0$. Initially, arriving customers wait 0, complete service and depart. Eventually customers in service accumulate until $c - 1$ servers are occupied. Concurrently the SP moves among the $2^c - 1 - c$ lines for the non-border zero-wait states. It resides on each such line for an exponentially distributed time, making transitions from line to line. Various states unfold until the SP ends up on one of the c *border* lines.

All zero-wait arrivals stay for full service (no balking). Assume that a new arrival C_τ finds $c - 1$ servers occupied (SP on a border line). Then C_τ waits 0, and starts service in the single idle server. At τ^- the configuration is some $\boldsymbol{m} \in \boldsymbol{M}_b$. At instant τ all c servers are occupied. The SP jumps at instant τ to one of the c sheets, depending on which service will finish first. The probability that server k will finish first is μ_k/μ where $\mu := \mu_1 + \cdots + \mu_c$. The SP will be at a height $\underset{dis}{=} \text{Exp}_\mu$, since the inter start-of-service depart time $\mathcal{S}$ is the *minimum* of c independent exponentially distributed r.v.s with rates $\mu_1, \ldots, \mu_c$, due to the memoryless property.

Let $\boldsymbol{m}_{\bar{i}}$ denote a *border* configuration such that the rate-μ_i server (i.e., server i) is *idle* (see Remark 4.20). In configuration $\boldsymbol{m}_{\bar{i}}$, $m_j = 1$, if $j \neq i$, and $m_i = 0$, i.e.,

$$m_1 + \cdots + m_{i-1} + 0 + m_{i+1} + \cdots + m_c = c - 1.$$

At time τ the SP will end up at a positive height on page $\boldsymbol{m}_{\bar{k}}$ with probability μ_k/μ, $k = 1, \ldots, c$.

Remark 4.20 We use the notation $\bar{\imath}$ to shorten the representation of $\boldsymbol{m}$ if c is large. If c is small, e.g. $c = 3$, we can use notation like 100, 010, 001, 110, 101, 011. If $c = 2$, we can use 01, 10—see Sect. 4.12.8.

4.12.4 Zero-Wait Probabilities

Let P_n, $n = 0, \ldots, c - 1$ denote the steady-state probability of n customers in the system at an arbitrary point in time. Let $P_{n,\boldsymbol{m}}$ denote the probability that there are n customers in the system and the configuration is $\boldsymbol{m} \in \boldsymbol{M}$. There are $\binom{c}{n}$ configurations such that $\sum_{i=1}^{c} m_i = n$. Let

$$\boldsymbol{M}_n = \{\boldsymbol{m} | \sum_{i=1}^{c} m_i = n\}.$$

Thus

$$P_n = \sum_{\boldsymbol{m} \in \boldsymbol{M}_n} P_{n,\boldsymbol{m}}, n = 0, \ldots, c-1.$$

Due to Poisson arrivals P_n is the probability that an arrival waits 0 and "sees" n other customers in service just before it starts service (using PASTA, e.g., [145]).

Remark 4.21 For the **zero-wait** states, a configuration specifies the service rates in the servers at an arbitrary time point. Due to Poisson arrivals, this is the same as the service rates **just before** an arrival. It is also the same as the service rates in the **other** servers **just after** an arrival starts service in an available server.

The probability of a zero wait is denoted by $F(0)$, where

$$F(0) = \sum_{n=0}^{c-1} P_n = \sum_{n=0}^{c-1} \sum_{\boldsymbol{m} \in \boldsymbol{M}_n} P_{n,\boldsymbol{m}}. \tag{4.105}$$

4.12.5 Positive-Wait PDF and CDF

For the **positive-wait** states, a configuration defines the service rates in the **other** servers **just after** start of service.

Let $f_{\boldsymbol{m}}(x), x > 0$, denote the *partial* pdf of wait for page $\boldsymbol{m} \in \boldsymbol{M}_b$. Denote the marginal pdf for the *cover* as

$$f(x) = \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} f_{\boldsymbol{m}}(x), x > 0.$$

The *total* pdf of wait is $\{\boldsymbol{P}_0, f(x), x > 0\}$. The cdf of wait is $F(x) = F(0) + \int_{y=0}^{x} f(y)dy, x \geq 0$, where $F(0)$ is defined in (4.105). The normalizing condition is

$$\lim_{x \to \infty} F(x) = F(0) + \int_{y=0}^{\infty} f(y)dy = 1. \tag{4.106}$$

4.12.6 Equations for Positive-Wait PDFs

A key assumption of this model is that each positive-wait arrival reneges from the waiting line with probability $R(y)$, and stays for complete service with probability $\overline{R}(y)$ (=$1 - R(y)$), where $y \geq 0$ is the required wait before service.

Equation for Total PDF $f(x)$ We first derive an integral equation for $f(x)$, the total pdf of wait of stayers (who wait and receive a full service), namely,

$$f(x) = \lambda P_{c-1}e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)}\overline{R}(y)f(y)dy, x > 0, \tag{4.107}$$

directly using the sample path, as follows (see Fig. 4.11).

Explanation of Equation (4.107) On the left side, $f(x)$ is the *total SP downcrossing rate* of level x over all c sheets, *projected onto the "cover"*. On the right side, since all zero-wait arrivals stay for full service ($\overline{R}(0) = 1$), the term $\lambda P_{c-1}e^{-\mu x}$ is the total SP upcrossing rate of level x due to jumps starting at level 0 (i.e., line 0) of any of the c sheets (from border states $\{(0, \boldsymbol{m}_{\bar{i}})\}$, $i = 1, \ldots, c$), at arrival instants. These jumps have size $\mathcal{S} \underset{dis}{=} \text{Exp}_{\mu}$ ($\underset{dis}{=} \min_{i=1,\ldots,c}\left\{\text{Exp}_{\mu_i}\right\}$). The term $\lambda \int_{y=0}^{x} e^{-\mu(x-y)}\overline{R}(y)f(y)dy$ is the rate at which the SP upcrosses level x due to jumps starting at levels $y \in (0, x)$, on any page, at arrival instants of stayers. The right side is, therefore, the total SP upcrossing rate of level x. Rate balance across x yields (4.107).

Comparing (4.107) with Eq. (3.211) implies that the solution of (4.107) is

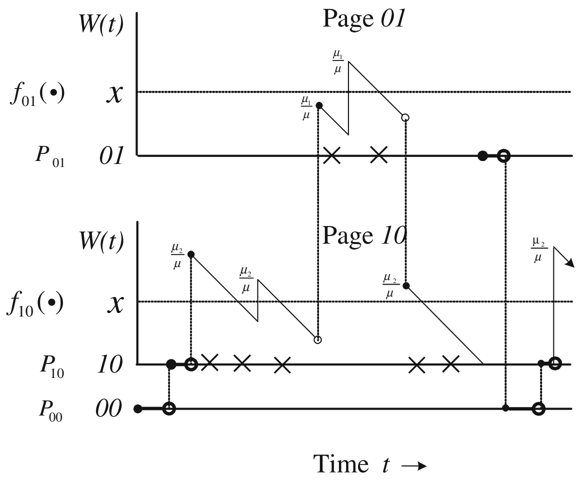
$$f(x) = \lambda P_{c-1}e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)}, x > 0, \tag{4.108}$$

where $\mu = \sum_{i=1}^{c} \mu_i$ and $P_{c-1} = \sum_{i=1}^{c} P_{c-1,\boldsymbol{m}_{\bar{i}}}$.

Equations for Partial PDFs $f_{\bar{i}}(x), x > 0, i = 1, \ldots, c$ We now obtain integral equations for the pdfs $f_{\bar{i}}(x)$, $x > 0$, on the c sheets (see Fig. 4.11); they are

$$\begin{aligned}
&f_{\bar{i}}(x) + \lambda(1 - \frac{\mu_i}{\mu}) \int_{y=x}^{\infty} \overline{R}(y) f_{\bar{i}}(y)dy \\
&= \lambda\frac{\mu_i}{\mu} P_{c-1}e^{-\mu x} + \lambda\frac{\mu_i}{\mu} \int_{y=0}^{x} e^{-\mu(x-y)}\overline{R}(y)f(y)dy \\
&\qquad + \lambda\frac{\mu_i}{\mu} \int_{y=x}^{\infty} \overline{R}(y) \left(f(y) - f_{\bar{i}}(y)\right) dy, i = 1, \ldots, c.
\end{aligned} \tag{4.109}$$

Fig. 4.11 Sample path of $\{W(t), M(t)\}_{t\geq 0}$ in M/M$_i$/2, where zero-wait arrivals join, and positive-wait arrivals may renege from the waiting line. Times marked × indicate arrivals that renege (do not contribute to the limiting pdf of wait)



Explanation of Equation (4.109) On the left side, $f_{\bar{i}}(x)$ is the SP *exit* rate from composite state $((x,\infty),\bar{i})$ due to SP downcrossings of level x; term $\lambda(1-\frac{\mu_i}{\mu})\int_{y=x}^{\infty}\overline{R}(y)f_{\bar{i}}(y)dy$ is the SP rate of jumps *out of* $((x,\infty),\bar{i})$ into the composite states $((x,\infty),\bar{j})$, $j\neq i$, on other sheets. On the right side, the first two terms are SP *entrance* rates into $((x,\infty),\bar{i})$ due to jumps starting at level-0 border states, and jumps starting at levels $y\in(0,x)$ on any sheet, respectively (recall $f(y)=\sum_{i=1}^{c}f_{\bar{i}}(y)$). The third term is the SP *entrance* rate into $((x,\infty),\bar{i})$ due to jumps starting in $\cup_{j\neq i}((x,\infty),\bar{j})$. Rate balance of SP exits and entrances of $((x,\infty),\bar{i})$ yields (4.109).

Solution of Equation (4.109) We obtain the solution of (4.109) in terms of the solution for $f(x)$, which is given in formula (4.108), using the following Proposition.

Proposition 4.1 *The partial pdf is given by*

$$f_{\bar{i}}(x)=\frac{\mu_i}{\mu}f(x), x>0, i=1,\ldots,c. \tag{4.110}$$

Proof Substitute $\frac{\mu_i}{\mu}f(x)$ for $f_{\bar{i}}(x)$ in Eq. (4.109), and cancel like terms. The proposition is true if and only if the following is an identity:

$$\frac{\mu_i}{\mu} f(x) + \lambda \int_{y=x}^{\infty} \frac{\mu_i}{\mu} \overline{R}(y) f(y))dy$$
$$= \lambda \frac{\mu_i}{\mu} P_{c-1} e^{-\mu x} + \lambda \frac{\mu_i}{\mu} \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy$$
$$+ \lambda \frac{\mu_i}{\mu} \int_{y=x}^{\infty} \overline{R}(y) f(y) dy, x > 0, \tag{4.111}$$

if and only if

$$f(x) = \lambda P_{c-1} e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y) \overline{R}(y)} f(y) dy, x > 0, \tag{4.112}$$

is an identity. Equation (4.112) is identical to Eq. (4.107). Hence the Proposition is true. ■

Exponential Staying Function
Consider an exponential staying function, $\overline{R}(x) := e^{-rx}, r > 0, x \geq 0$. (Note that $0 < e^{-rx} \leq 1$, and is strictly decreasing on $(0, \infty)$, satisfying the definition of staying function.) The total pdf $f(x)$ is now obtained by substituting e^{-ry} for $\overline{R}(y)$ in (4.108), which substituted into (4.110), gives

$$f_{\bar{i}}(x) = \lambda \frac{\mu_i}{\mu} e^{\frac{\lambda}{r}} P_{c-1} e^{-\mu x - \frac{\lambda}{r} e^{-rx}}, x > 0, i = 1, \ldots, c. \tag{4.113}$$

We shall solve an M/M$_i$/2 model using $\overline{R}(x) := e^{-rx}, r > 0, x \geq 0$, in Sect. 4.12.8 below.

4.12.7 Equations for Zero-Wait Probabilities

Assume that the zero-wait server assignment policy is: arrivals that find k available servers, $1 \leq k \leq c$, get served by a particular available server with probability $1/k$. (Other policies are also viable, e.g., the arrival gets served by the lowest-numbered available server, or by the fastest-available service rate, etc.) Using the principle *SP exit rate = SP entrance rate* for the zero-wait states, we obtain the equations (notation explained below)

$$
\begin{aligned}
(\lambda+\mu-\mu_i)P_{c-1,\bar{i}} &= f_{\bar{i}}(0)+\tfrac{\lambda}{2}\textstyle\sum_{j\in \boldsymbol{J}_i}^{c} P_{c-2,\overline{ij}},\ i=1,\ldots,c,\\
(\lambda+\mu-\mu_i-\mu_j)P_{c-2,\overline{ij}} &= \mu_j P_{c-1,\bar{i}}+\mu_i P_{c-1,\bar{j}}\\
&\quad +\tfrac{\lambda}{3}\textstyle\sum_{k\in J_{ij}}^{c} P_{c-3,\overline{ijk}},\ j=1,\ldots,c,\\
&\cdots\\
(\lambda+\mu_i)P_{1,i} &= \textstyle\sum_{k\neq i=1}^{c}\mu_k P_{2,ik}+\dfrac{\lambda}{c}P_{00},\ i=1,\ldots,c,\\
\lambda P_{00} &= \textstyle\sum_{i=1}^{c}\mu_i P_{1,i}.
\end{aligned}
\tag{4.114}
$$

Notation in equations (4.114) In the first c equations, the index j of the sum takes values in $\boldsymbol{J}_i := \{j|j=0,\ldots,c,\ j\neq i\}$, and the subscript $\overline{ij}$ means both servers$i$ and j are idle. In the second set of $\binom{c}{2}$ equations, the index k of the sum takes values in $\boldsymbol{J}_{ij}=\{k|k=0,\ldots,c, k\neq i, k\neq j\}$, and the subscript $\overline{ijk}$ means all three servers i, j and k are idle. The row of dots "$\cdots$" indicates similar rate balance equations for $P_{c-3,\overline{ijk}}$, ..., $P_{2,\cdot}$. In the second last equation, for $P_{1,i}$, on the right side $P_{2,ik}$ denotes the probability of two units in the system, in servers i and k having service rates μ_i and μ_k respectively.

We solve Eq. (4.114) explicitly in Sect. 4.12.8 below for M/M$_i$/2, in order to convey some characteristics of the solution.

4.12.8 Solution for M/M$_i$/2 with Reneging

Notation When there is a small number of servers we can use an alternative, perhaps more familiar notation. If $c = 2$, there are two sheets corresponding to configurations $\bar{1}$ and $\bar{2}$, which we now replace by 01 and 10 respectively. Thus, configuration 01 means server 1 is available and server 2 is occupied; configuration 10 means server 2 is available and server 1 is occupied.

Applying formula (4.113), the partial pdfs of wait are now denoted by

$$
\begin{aligned}
f_{10}(x) &= \lambda\tfrac{\mu_2}{\mu}e^{\frac{\lambda}{r}}P_1 e^{-\mu x-\frac{\lambda}{r}e^{-rx}},\ x>0,\\
f_{01}(x) &= \lambda\tfrac{\mu_1}{\mu}e^{\frac{\lambda}{r}}P_1 e^{-\mu x-\frac{\lambda}{r}e^{-rx}},\ x>0.
\end{aligned}
\tag{4.115}
$$

The marginal ("total") pdf of wait is

$$
f(x)=f_{10}(x)+f_{01}(x)=\lambda e^{\frac{\lambda}{r}}P_1 e^{-\mu x-\frac{\lambda}{r}e^{-rx}},\ x>0. \tag{4.116}
$$

The zero-wait probabilities are $P_{1,\bar{i}}$, $i = 1, 2$, and P_{00}; using the alternative notation we have

$$\begin{aligned}
P_1 &= P_{1,\bar{2}} + P_{1,\bar{1}} = P_{10} + P_{01}, \\
P_0 &= P_{00} + P_{1,\bar{2}} + P_{1,\bar{1}} = P_0 + P_{10} + P_{01} \\
&= P_{00} + P_1.
\end{aligned}$$

The rate-balance equations for the zero-wait probabilities are

$$\begin{aligned}
(\lambda + \mu_1) P_{10} &= \tfrac{\lambda}{2} P_{00} + f_{10}(0), \\
(\lambda + \mu_2) P_{01} &= \tfrac{\lambda}{2} P_{00} + f_{01}(0), \\
\lambda P_{00} &= \mu_1 P_{10} + \mu_2 P_{01}.
\end{aligned} \tag{4.117}$$

Substituting for $f_{10}(0)$, $f_{01}(0)$ from (4.115), we rewrite the equations in (4.117) as

$$\begin{aligned}
(\lambda + \mu_1) P_{10} &= \tfrac{\lambda}{2} P_{00} + \lambda \tfrac{\mu_2}{\mu} P_1, \\
(\lambda + \mu_2) P_{01} &= \tfrac{\lambda}{2} P_{00} + \lambda \tfrac{\mu_1}{\mu} P_1, \\
\lambda P_{00} &= \mu_1 P_{10} + \mu_2 P_{01}.
\end{aligned} \tag{4.118}$$

The solution of (4.118) in terms of P_{00} is

$$\begin{aligned}
P_{01} &= \tfrac{\lambda}{2\mu_2} P_{00}, \\
P_{10} &= \tfrac{\lambda}{2\mu_1} P_{00}, \\
P_1 &= \tfrac{\lambda(\mu_1+\mu_2)}{2\mu_1\mu_2} P_{00} = \tfrac{\lambda\mu}{2\mu_1\mu_2} P_{00}.
\end{aligned} \tag{4.119}$$

The normalizing condition

$$P_{00} + P_1 + \int_{x=0}^{\infty} f(x)dx = 1,$$

yields

$$P_{00} = \left(1 + \frac{\lambda(\mu_1 + \mu_2)}{2\mu_1\mu_2} + \frac{\lambda(\mu_1 + \mu_2)}{2\mu_1\mu_2} \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx \right)^{-1}. \tag{4.120}$$

The analytic solution comprises the results in (4.120), (4.119), (4.116) and (4.115).

Example 4.10 We present a numerical example for the M/M$_i$/2 queue with reneging allowed from the waiting line (see Fig. 4.12). Let

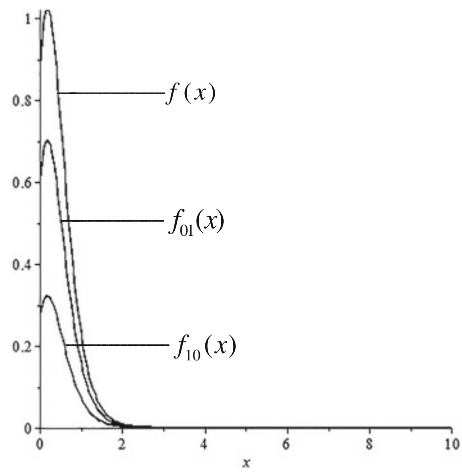


Fig. 4.12 Plot of $f(x)$, $f_{01}(x)$ $(= f_{\overline{1}}(x))$, $f_{10}(x)$ $(= f_{\overline{2}}(x))$, in Example 4.10

$$\lambda = 5.2,\ \mu_1 = 2.4,\ \mu_2 = 1.1,\ \mu = \mu_1 + \mu_2 = 3.5,\ r = 2.1.$$

Then

$$\int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx = 0.074741,$$

and

$$\begin{aligned}
P_{00} &= 0.049059, \quad P_{01} = 0.115958, \quad P_{10} = 0.053147,\\
P_1 &= 0.169105,\\
F(0) &= P_{00} + P_1 = 0.218164,\\
F(\infty) &= F(0) + \lambda e^{\frac{\lambda}{r}} P_1 \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx\\
&= 0.218164 + 0.781836 = 1.0,\\
f(x) &= \lambda e^{\frac{\lambda}{r}} P_1 e^{-\mu x - \frac{\lambda}{r} e^{-rx}} = 10.461 \cdot e^{-3.5x - 2.476 e^{-2.1x}},\\
f_{01}(x) &= \frac{\mu_1}{\mu} f(x) = 7.173 \cdot e^{-3.5x - 2.476 e^{-2.1x}},\\
f_{10}(x) &= \frac{\mu_2}{\mu} f(x) = 3.288 \cdot e^{-3.5x - 2.476 e^{-2.1x}}.
\end{aligned}$$

Remark 4.22 In M/M$_i$/c with reneging from the waiting line allowed, we can generalize the staying function $\overline{R}(x)$, $x \geq 0$. For example, $\overline{R}(x)$ may depend on the server that would be occupied by an arrival, i.e., on the system configuration at the arrival instant. We may then use the notation $\overline{R}_{\overline{i}}(x)$. Thus $\overline{R}_{\overline{i}}(x)$ may depend on, not only customer required wait before service, but also on customer attraction or aversion to the "target" server. A natural question arises. Can this model be modified to study attraction or aversion in natural processes such as: electrically charged particles approaching an electrically charged or magnetized environment; asteroids approaching a planet; particles adhering or falling away from a surface; laser pulses affecting cells containing certain chemicals in biological or medical applications; etc.?

4.12.9 Stability Condition

Consider the M$_\lambda$/M$_i$/c ($c \geq 2$) queue with heterogeneous servers having rates $\mu_1, \ldots, \mu_c$ in which reneging depending on required wait is allowed before service begins. Let the staying function $\overline{R}(x), x \geq 0$, be monotone decreasing (includes non-increasing), let $\overline{R}(0) = 1$ (no balking upon arrival), and assume $0 \leq \overline{R}(x) \leq 1$, $x \geq 0$. Let $L = \lim_{x\to\infty} \overline{R}(x)$, which exists by monotonicity and boundedness. The ideas in Theorem 3.8 also apply in the M/M$_i$/c environment, as follows.

Theorem 4.10 In M$_\lambda$/M$_i$/c ($c \geq 2$) with reneging from the waiting line allowed, as described immediately above, a necessary and sufficient condition for stability is

$$\lambda < \frac{\mu}{L} \text{ if } 0 < L \leq 1,$$
$$\lambda < \infty \text{ if } L = 0,$$

where $\mu = \sum_{i=1}^{c} \mu_i$.

Proof The proof is similar to that of Theorem 3.8. The alternative proof given there, Remark 3.31 and Fig. 3.28 also apply for the present M/M$_i$/c queue with reneging, upon substituting $\mu = \sum_{i=1}^{c} \mu_i$. ■

4.13 Discussion

We can use LC to analyze a vast array of additional M/M/c models. We mention only a few examples.

LC has been applied to M/M/c queues in which customers receive simultaneous service from a random number of servers. The original source for such queueing models is the Ph.D. thesis of L. Green [81, 82]. An LC analysis, motivated by the work of L. Green, is given in [38].

LC has been applied to M/M/c with bounded system time (wait + service). An arrival balks upon arrival if its system time would exceed an upper bound K, e.g., a system manager informs an arrival of the current expected system time (see Example 1, p. 44 in [52]). This generalizes variant 2 of the M/G/1 model discussed above in Sect. 3.6. It is straightforward to apply LC to analyze an M/M/c model analogous to variant 1 in Sect. 3.16. In that model customers renege from *service* if their *age* in the system *(elapsed system time)* reaches K. Similar remarks apply to M/M/c where the actual waits are bounded by K (as in variant 3 in Sect. 3.16). In that case the workload can exceed K. We can develop an expression for the tail of the steady-state pdf of workload, from its integral equation.

LC can be used to analyze a variety of M/M/c queues with server vacations; priorities; and many others.

Chapter 5
G/M/1 and G/M/c Queues

This chapter applies a system-point level-crossing approach (SPLC, abbreviated LC) to derive the steady-state pdf of the virtual wait and the actual wait (arrival-point wait) in single-server G/M/1, and multiple-server G/M/c queues. Sections 5.1 and 5.2 treat G/M/1 and Sect. 5.3 treats G/M/c ($c \geq 2$). We assume arrivals occur according to a renewal process and service times are exponentially distributed.

We will not derive transient distributions in this chapter. However, for G/M/c ($c \geq 2$), we could use LC to derive the transient distribution of *extended age*, which is related to the virtual wait (Sect. 5.1.1), by applying techniques similar to those utilized in Sects. 3.2, 4.3, 6.2.6, 11.7. Those analyses provide background for deriving transient distributions using LC in G/M/c queues, as well as in a great variety of stochastic models. (The extended age process is defined and utilized in [19].)

5.1 G/M/1 Queue

In the G/M/1 queue in steady state, we assume arrivals occur according to a renewal process (e.g., pp. 167–175 in [99]). For the common inter-arrival time denote the cdf, complementary cdf, and pdf, respectively, by $A(x)$, $x > 0$, $\overline{A}(x) = 1 - A(x)$, $x \geq 0$ and $a(x) = dA(x)/dx$ wherever the derivative exists. Assume the service time of each customer is $\underset{dis}{=} \text{Exp}_{\mu}$. We derive the steady-state pdf and cdf of the *virtual* wait, the steady-state pdf and cdf of the *actual* (arrival-point) wait just before arrival instants, the expected busy and idle periods, and related results.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_5

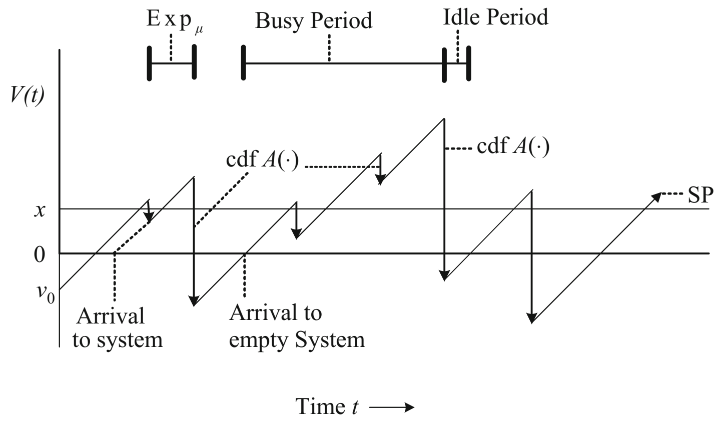


Fig. 5.1 Sample path of extended age process $\{V(t)\}_{t\geq 0}$ in G/M/1 queue. Interarrival times have cdf $A(\cdot)$ (cdf of downward jump sizes). Service times are $\underset{dis}{=} \text{Exp}_{\mu}$. $dV(t)/dt = +1$ between jumps

5.1.1 Virtual Wait and Extended Age Processes

Let $\{W(t)\}_{t\geq 0}$ denote the virtual wait process having state space $\boldsymbol{S} = [0, \infty)$ (e.g., similar to Fig. 3.5).

We consider the *extended age* process $\{V(t)\}_{t\geq 0}$ having state space $\boldsymbol{S} = (-\infty, \infty)$, defined as follows. For $t > 0$,

$$V(t) = \begin{cases} \textit{age of customer in service at time } t, \text{ if } V(t) \geq 0, \\ -\textit{time from } t \textit{ until the next arrival instant}, \text{ if } V(t) < 0. \end{cases} \tag{5.1}$$

In (5.1) 'age' means '*time spent in the system*' measured from the arrival instant. A sample path of $\{V(t)\}_{t\geq 0}$ is depicted in Fig. 5.1. All SP jumps start from positive levels (i.e., $V(\cdot) > 0$) at customer *departure* instants, and are *downward* in direction. By contrast, jumps of $\{W(t)\}_{t\geq 0}$ occur at *arrival* instants, and are *upward* in direction.

5.1.2 Duality Between Extended Age and Virtual Wait

Consider a sample path of $\{V(t)\}_{t\geq 0}$ (Fig. 5.2). Assume $V(t) \geq 0$. There is a one-to-one correspondence between the *peaks* (relative maxima) of $\{V(t)\}_{t\geq 0}$ and the sample-path *peaks* of $\{W(t)\}_{t\geq 0}$, as well as a one-to-one correspondence between their respective *troughs* (relative minima or infima). Within busy periods, corresponding peaks have equal ordinates and occur in the same time order in both processes; similarly for corresponding troughs (Fig. 5.2).

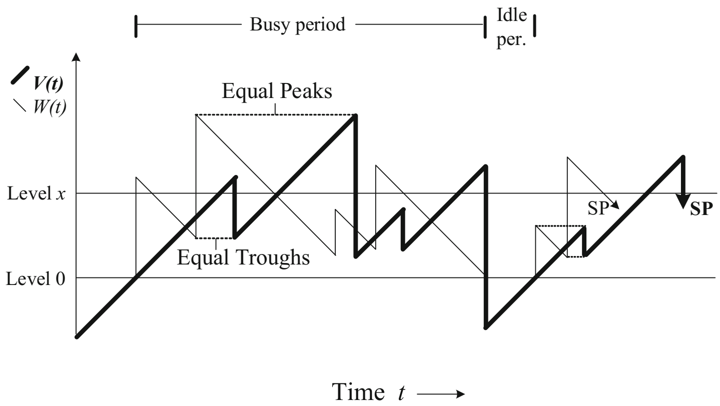


Fig. 5.2 Sample path of extended age process "↗" compared with sample path of virtual wait process "↘" for G/M/1 queue. Illustrates duality properties. Corresponding peaks and corresponding troughs have equal ordinates and the same time order. Busy periods, idle periods, and busy cycles are equal

Properties of Busy Period in G/M/1

The process $\{V(t)\}_{t\geq 0}$ has slope $+1$ between SP downward jumps. However, $\{W(t)\}_{t\geq 0}$ has slope -1 between SP upward jumps, within a busy period; the slope is 0 within an idle period. The length of busy periods is identical in both processes. These properties guarantee that the long-run proportion of time that the SP spends within any state-space interval, is the same in both processes (see Proposition 5.1 below).

The sojourn time of $\{V(t)\}_{t\geq 0}$ below level 0 is identical to an idle period in the $\{W(t)\}_{t\geq 0}$ process (see Remark 5.2). The length of busy cycles are identical in both processes (Fig. 5.2).

The stability condition is $\dfrac{1}{E(\text{inter-arrival time})\cdot\mu} < 1$ (e.g., pp. 261–262 in [84]). Intuitively, the expected number of arrivals in a service time is < 1. (See Proposition 5.3 in Sect. 5.1.8 below.)

Probability Distributions

Denote the steady-state cdf of the extended age by

$$F(x) = \lim_{t\to\infty} P(V(t) \leq x), -\infty < x < \infty,$$

having pdf

$$\begin{aligned} f(x) &= \tfrac{dF(x)}{dx}, x \geq 0; \\ h(x) &= \tfrac{dF(x)}{dx}, x < 0, \end{aligned} \tag{5.2}$$

wherever the derivatives exist. The probability of an empty system is

$$P_0 = F(0) = \int_{y=-\infty}^{0} h(y)dy. \tag{5.3}$$

Then

$$\begin{aligned} &F(x) = P_0 + \textstyle\int_{y=0}^{x} f(y)dy, x \geq 0, \\ &F(x) = \textstyle\int_{y=-\infty}^{x} h(y)dy, x \leq 0, \\ &F(0) = P_0, \\ &F(\infty) = P_0 + \textstyle\int_{y=0}^{\infty} f(y)dy = 1. \end{aligned} \tag{5.4}$$

Proposition 5.1 The **steady-state** cdf of the $\{V(t)\}_{t\geq 0}$ and of $\{W(t)\}_{t\geq 0}$ as $t \to \infty$, are identical. That is,

$$F(x) = \lim_{t\to\infty} P(V(t) \leq x) = \lim_{t\to\infty} P(W(t) \leq x), x \geq 0.$$

Proof There is a one-to-one correspondence between sample paths of $\{V(t)\}_{t\geq 0}$ and $\{W(t)\}_{t\geq 0}$ because of the duality properties discussed above (see Fig. 5.2). The *proportion* of time spent in each state-space interval is the same in corresponding sample paths for every $\omega \in \Omega$, where Ω is the sample space of the 'underlying experiment', and ω is a possible outcome (see more details in [19]).

For $\{V(t)\}_{t\geq 0}$ a sojourn time below level 0 is the same as an idle period in $\{W(t)\}_{t\geq 0}$. Thus $F(0) = P_0 = \lim_{t\to\infty} P_0(t)$ is the same for both processes (where $P_0(t)$ is the probability of a zero wait at time t). ■

The main reason for employing $\{V(t)\}_{t\geq 0}$ to obtain the limiting pdf of $\{W(t)\}_{t\geq 0}$ in G/M$_\mu$/1, is that we can apply LC, since SP downward jumps occur at end-of-service instants at the *Poisson rate* μ. Thus, in $\{V(t)\}_{t\geq 0}$ the PASTA principle —Poisson arrivals see time averages—applies ([145]). But PASTA does not apply in the process $\{W(t)\}_{t\geq 0}$ in general, since the interarrival times are not exponentially distributed.

Remark 5.1 We emphasize that the *transient probability distributions* of $V(t)$ and $W(t)$ are **not** equal in general. Proposition 5.1 holds for *steady-state* distributions only.

Remark 5.2 We may also define an **extended 'virtual wait' process** $\{W(t)\}_{t\geq 0}$ with state space $(-\infty, +\infty)$. If $W(t) > 0$, then $W(t)$ is the usual virtual wait. If $W(t) < 0$, then $-W(t)$ is the time **since the last departure of the immediately previous busy period**. For the extended **virtual wait**, the slope is -1 between upward jumps. Sojourn times below level 0 are equal to idle periods. If arrivals are Poisson, an integral equation for the pdf of $\{W(t)\}_{t\geq 0}$ when $W(t) < 0$ can be obtained by applying LC. All results for the usual virtual wait can be derived using the extended virtual wait. If arrivals are Poisson at rate λ, the expected sojourn time below level 0 is $1/\lambda$ = E(idle period).

5.1.3 *Equation for Steady-State PDF of Age*

By Proposition 5.1 the steady-state pdfs of $\{V(t)\}_{t\geq 0}$ and $\{W(t)\}_{t\geq 0}$ are the same. Thus, for G/M/1 we will obtain the steady-state pdf of $\{W(t)\}_{t\geq 0}$ by deriving the steady-state pdf of $\{V(t)\}_{t\geq 0}$.

Consider a sample path of $\{V(t)\}$ (Fig. 5.1). Fix level $x > 0$ in the state space. The SP *upcrossing* rate of x is

$$\lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t} = f(x), \tag{5.5}$$

The SP *downcrossing* rate of level x is

$$\lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t} = \mu\int_{y=x}^{\infty}\overline{A}(y-x)f(y)dy, \tag{5.6}$$

Formulas (5.5) and (5.6) are proved similarly as for the down- and upcrossing crossing rates in M/G/1 (e.g., Theorem 1.1 in Chap. 1).

LC Interpretation of (5.6)

The SP rate of downward jumps starting from level $y > 0$ is the rate at which *service times end* when customers have been in the system for a time y, namely $\mu f(y)dy$. If $y > x$,

$$\begin{aligned}&P(\text{downward jump size} > y-x)\\&= P(\text{inter-arrival time} > y-x) = \overline{A}(y-x).\end{aligned}$$

Summing over all $y > x$ gives the right-most term of (5.6).

The principle of rate balance across level x, i.e.,

$$\lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t} = \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t},$$

gives the integral equation for $f(x)$,

$$f(x) = \mu\int_{y=x}^{\infty}\overline{A}(y-x)f(y)dy. \tag{5.7}$$

5.1.4 Alternative Form of Equation for PDF of Age

An alternative form of integral equation (5.7) is

$$f(x) = \mu(1 - F(x)) - \mu \int_{y=x}^{\infty} A(y - x)f(y)dy, x > 0. \tag{5.8}$$

LC Interpretation of (5.8)

The left side is the SP upcrossing rate of level x. On the right side, the first term is the rate of service completions which generate SP downward jumps that *start above* level x. The second term subtracts off the rate of service completions which generate SP downward jumps that *start above* level x *and end above* level x. Thus, the right side is the SP downcrossing rate of level x. Rate balance across level x gives the equation.

Alternative equation (5.8) for $f(x)$ in G/M/1 obtained using LC, is similar to the alternative equations (3.43) and (3.44) in Sect. 3.3.1 for the M/G/1 queue, and alternative equation (6.22) for the M/G/r(·) dam in Chap. 6. All these equations are reached directly using LC interpretations of SP motion in state spaces, thereby enhancing background intuition.

5.1.5 Exponential Form of PDF of $\{V(t)\}_{t\geq 0}$

We demonstrate *geometrically* using LC, that $f(x), x > 0$, the limiting pdf of $\{V(t)\}_{t\geq 0}$ (therefore of $\{W(t)\}_{t\geq 0}$) as $t \to \infty$, has an *exponential form* over the state-space interval $(0, \infty)$, and an atom at 0. This LC derivation differs from previous derivations (e.g., pp. 261–263 in [84]; pp. 251–253 in [104]; pp. 400–401 in [143]).

Consider a sample path of $\{V(t)\}_{t\geq 0}$. Due to the memoryless property of the exponential service-time distribution, an SP *sojourn time above* an arbitrary level $x \geq 0$ is $\underset{dis}{=} \mathcal{B}$ *independent of* x, where $\mathcal{B}$ denotes a busy period (Figs. 5.1 and 5.2). Let u_x denote the *inter-upcrossing* time of level x (see Sect. 5.1.12 below). Combining formula (5.5), the fact that $1 - F(x)$ is the long-run proportion of time spent by $\{V(t)\}_{t\geq 0}$ above level x, and the elementary renewal theorem, implies $E(u_x) = 1/f(x)$. The sequence of u_xs forms a renewal process. By the renewal reward theorem

$$\frac{E(\mathcal{B})}{E(u_x)} = \frac{E(\mathcal{B})}{1/f(x)} = 1 - F(x) \,,$$

$$\frac{f(x)}{1 - F(x)} = \frac{1}{E(\mathcal{B})}, \tag{5.9}$$

which is equivalent to the differential equation

$$\frac{\frac{d}{dx}(1 - F(x))}{1 - F(x)} = -\frac{1}{E(\mathcal{B})},$$
$$\frac{d \ln(1 - F(x))}{dx} = -\frac{1}{E(\mathcal{B})},$$

with solution

$$F(x) = 1 - (1 - P_0)e^{-\frac{1}{E(\mathcal{B})}x}, x \geq 0,$$
$$f(x) = \frac{1 - P_0}{E(\mathcal{B})}e^{-\frac{1}{E(\mathcal{B})}x}, x > 0, \tag{5.10}$$

where $F(0) = P_0$, and $f(0) = (1 - P_0)\,/E(\mathcal{B})$.

From (5.10), $f(x)$ has the exponential form

$$f(x) = Ke^{-\gamma x}, x \geq 0 \tag{5.11}$$

where

$$K = \frac{1 - P_0}{E(\mathcal{B})}, \qquad \gamma = \frac{1}{E(\mathcal{B})}. \tag{5.12}$$

From (5.12),

$$P_0 = 1 - KE(\mathcal{B}) = 1 - \frac{K}{\gamma}. \tag{5.13}$$

Substituting from (5.11) into (5.7) and cancelling K gives an equation for γ,

$$e^{-\gamma x} = \mu \int_{y=x}^{\infty} \overline{A}(y - x)e^{-\gamma y} dy.$$

Letting $z = y - x$ yields

$$\int_{z=0}^{\infty} \overline{A}(z)e^{-\gamma z} dz = \frac{1}{\mu}. \tag{5.14}$$

Equation (5.14) for γ is *fundamental* for G/M/1; its left side is the Laplace transform of $\overline{A}(z)$ evaluated with parameter γ. It is also an equation for $E(\mathcal{B})$ since $\gamma = \frac{1}{E(\mathcal{B})}$.

Let $A^*(\gamma)$ denote the LST (Laplace-Stieltjes transform) of $A(\cdot)$. (See Sect. 3.4.4 above for definitions of LST and LT). Integrating (5.14) by parts gives

$$A^*(\gamma) = 1 - \frac{\gamma}{\mu}. \tag{5.15}$$

Thus γ is the solution of Eq. (5.14), or equivalently equation (5.15). Some forms of $\overline{A}(\cdot)$ allow for an analytical solution for γ. Generally, however, γ is computed by numerical methods (e.g., Newton's or other successive approximation methods—see p. 18 in [1]; or using computational software such as Maple).

Value of P_0

Consider a sample path of $\{V(t)\}_{t\geq 0}$ on the state-space interval $(-\infty, 0)$, and fix level $x \in (-\infty, 0)$. The SP *upcrossing* rate of level x ($x < 0$) is equal to $h(x)$ (proved as for the downcrossing rate in M/G/1). The SP *downcrossing* rate of level x is

$$\mu \int_{y=0}^{\infty} \overline{A}(y-x)f(y)dy = \mu \int_{y=0}^{\infty} \overline{A}(y-x)Ke^{-\gamma y}dy,$$

since all downward jumps originate at end-of-service instants when the SP is in state-space set $(0, \infty)$. Rate balance across level x gives

$$h(x) = \mu \int_{y=0}^{\infty} \overline{A}(y-x)Ke^{-\gamma y}dy, x < 0. \tag{5.16}$$

Invoking (5.16) and (5.3) leads to

$$P_0 = \int_{x=-\infty}^{0} h(x)dx = K \int_{x=-\infty}^{0} \mu \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx.$$

Making the transformation $u = -x$ gives

$$P_0 = K \int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu.$$

Thus

$$P_0 = \frac{K}{C_\gamma}, \text{ or } K = P_0 C_\gamma \tag{5.17}$$

where

$$C_\gamma := \left(\int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu \right)^{-1}, \tag{5.18}$$

and $C_\gamma > 0$.

We evaluate P_0 from the normalizing condition and (5.17). Thus

$$P_0 + K \int_{y=0}^{\infty} e^{-\gamma x} dx = 1,$$

$$P_0 + C_\gamma P_0 \int_{y=0}^{\infty} e^{-\gamma x} dx = 1.$$

These equations yield

$$P_0 = 1 - \frac{K}{\gamma}. \tag{5.19}$$

$$= \frac{\gamma}{\gamma + C_\gamma}. \tag{5.20}$$

From (5.17)

$$K = \frac{\gamma \cdot C_\gamma}{\gamma + C_\gamma}, \tag{5.21}$$

which implies $K < \gamma$.

Due to exponentially distributed service times, instants of SP egress from level 0 above, are regenerative points of $\{V(t)\}_{t \geq 0}$ initiating busy cycles (see Fig. 5.1; Sect. 2.4.10 for definitions of SP egresses). Thus, steady-state properties over busy cycles recapitulate limiting properties over the time axis as $t \to \infty$.

Expected Idle Period

Let C represent a *busy cycle* and $\mathcal{I}$ an *idle period*. Then

$$\mathcal{C} = \mathcal{B} + \mathcal{I}.$$

By the renewal reward theorem, and from (5.20),

$$P_0 = \frac{E(\mathcal{I})}{E(\mathcal{C})} = \frac{E(\mathcal{I})}{E(\mathcal{B}) + E(\mathcal{I})} = \frac{\gamma}{\gamma + C_\gamma} = \frac{\frac{1}{C_\gamma}}{\frac{1}{\gamma} + \frac{1}{C_\gamma}}. \tag{5.22}$$

From (5.12) $E(\mathcal{B}) = \frac{1}{\gamma}$. Comparing the middle and last ratios in (5.22), and using (5.18), we obtain

$$E(\mathcal{I}) = \frac{1}{C_\gamma} = \int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u) e^{-\gamma y} dy du. \tag{5.23}$$

5.1.6 PDF of Idle Period

Consider a sample path of $\{V(t)\}_{t\geq 0}$ (Fig. 5.1). Let $\mathcal{I}$ denote an idle period, and $\mathcal{C}$ a busy cycle; then $\mathcal{C} = \mathcal{B} + \mathcal{I}$. Note that $\mathcal{I}$:= *excess below level 0 of the downward SP jumps ending the busy periods*. Denote the cdf and pdf of $\mathcal{I}$ by $F_{\mathcal{I}}(z)$, and $f_{\mathcal{I}}(z), z > 0$, respectively. The *downcrossing rate* of level $-z$ ($z > 0$) can be expressed in two different ways. First, by $\lim_{t\to\infty} \mathcal{D}_t(-z)/t = (1 - F_{\mathcal{I}}(z)) \cdot f(0) = P(\mathcal{I} > z|SP\ downcrosses\ level\ 0) \times \lim_{t\to\infty} \mathcal{D}_t(0)/t$ (since $f(0)$ is both the up- and downcrossing rate of level 0). Second, by $\lim_{t\to\infty} \mathcal{D}_t(-z)/t = \mu \int_{y=0}^{\infty} \overline{A}(y+z)f(y)dy$, since all downward jumps start above level 0. Equating the two formulas gives

$$(1 - F_{\mathcal{I}}(z)) \cdot f(0) = \mu \int_{y=0}^{\infty} \overline{A}(y+z)f(y)dy, z > 0,$$

$$(1 - F_{\mathcal{I}}(z)) \cdot K = \mu \int_{y=0}^{\infty} \overline{A}(y+z)Ke^{-\gamma y}dy, z > 0,$$

$$F_{\mathcal{I}}(z) = 1 - \mu \int_{y=0}^{\infty} \overline{A}(y+z)e^{-\gamma y}dy, z > 0. \quad (5.24)$$

Directly using $F_{\mathcal{I}}(z), z > 0$, in (5.24), we obtain

$$E(\mathcal{I}) = \int_{z=0}^{\infty} (1 - F_{\mathcal{I}}(z))\, dz = \mu \int_{z=0}^{\infty} \int_{y=0}^{\infty} \overline{A}(y+z)e^{-\gamma y}dydz, \quad (5.25)$$

which agrees with formula (5.23) above.

Remark 5.3 As a mild check on the above results, if G/M$_\mu$/1 were an M$_\lambda$/M$_\mu$/1 queue then, in (5.11) $E(\mathcal{B}) = 1/(\mu - \lambda)$, $\gamma = \mu - \lambda$, $C_\gamma = \lambda$, implying

$$K = \frac{\gamma \cdot C_\gamma}{\gamma + C_\gamma} = \lambda\left(1 - \frac{\lambda}{\mu}\right) = \lambda P_0,$$

giving $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0$ ✓. (See formula (3.112) for M/M/1.)

5.1.7 PDF of Actual Wait

For G/M/1, generally the limiting pdf of the actual wait (arrival-point wait) is not equal to the limiting pdf of the virtual wait. In particular, these pdfs are equal only when the arrival stream is Poisson (M/M/1). We can utilize results for $\{V(t)\}_{t\geq 0}$ in Sects. 5.1.1–5.1.5 to determine the pdf of the actual wait.

The subscript 'ι' (Greek iota) is used to signify actual wait. Let the limiting cdf of actual wait be $F_\iota(x) := P(\text{actual wait} \leq x)$, $x \geq 0$, with pdf $dF_\iota(x)/dx = f_\iota(x)$, $x > 0$, and $P_{0,\iota} := P(\text{actual wait} = 0)$. Recall that γ $(= 1/E(\mathcal{B}))$ is the solution of (5.14) or (5.15).

Proposition 5.2 The steady-state cdf and mixed pdf of **actual wait** are

$$F_\iota(x) = 1 - \left(1 - \frac{\gamma}{\mu}\right) e^{-\gamma x}, \tag{5.26}$$

$$f_\iota(x) = K_\iota e^{-\gamma x}, x > 0, \tag{5.27}$$

$$P_{0,\iota} = 1 - \frac{K_\iota}{\gamma}, \tag{5.28}$$

where

$$K_\iota = \gamma \left(1 - \frac{\gamma}{\mu}\right). \tag{5.29}$$

Proof The long-run *proportion* of actual waits that are $> x$ is

$$1 - F_\iota(x) = \frac{\mu(1 - F(x)) - f(x)}{\mu\,(1 - F(0))}, x \geq 0, \tag{5.30}$$

where $F(x)$, $f(x)$ denote the limiting cdf and pdf respectively of the age process $\{V(t)\}_{t\geq 0}$ (same as for *virtual wait*).

Explanation of (5.30). In the numerator, $\mu(1 - F(x))$ is the departure rate of customers that have been in the system $> x$ time units (*wait* + *service* $> x$), since departures occur at sample-path peaks. The departures generate SP *downward* jumps that end at the next actual wait > 0 if the end point is > 0 (at a trough > 0), or at the start of an idle period implying that the next actual wait is = 0 if the end point is < 0 (at a trough < 0) (Fig. 5.1). The term $-f(x)$ subtracts off the rate at which SP jumps *start above* x and *end below* x, since $f(x)$ is equal to the level-x *downcrossing* rate of level x (equal also to the *upcrossing* rate of x—see derivation of Eq. (5.7)). Thus the numerator is the rate at which '*next*' *actual waits before service* (troughs) are $> x$. The denominator $\mu\,(1 - F(0))$ in (5.30) is the *total* departure rate, which is also the long-run rate of downward jumps, and is the same as the long-run arrival rate.

From (5.10) and (5.11) we obtain (5.26) and (5.27).

Probability $P_{0,\iota}$ and **Value of** K_ι. Setting $x = 0$ in (5.30) yields, since $P_{0,\iota} = F_\iota(0)$,

$$
\begin{aligned}
1 - P_{0,\iota} &= 1 - \frac{f(0)}{\mu\left(1 - F(0)\right)} = 1 - \frac{K}{\mu\left(1 - \left(1 - \frac{K}{\gamma}\right)\right)} \\
P_{0,\iota} &= \frac{\gamma}{\mu}. \qquad (5.31)
\end{aligned}
$$

We ascertain K_ι using the normalizing condition for $\left\{P_{0,\iota}, f_\iota(x)\right\}_{x>0}$,

$$
\begin{aligned}
P_{0,\iota} + \int_{y=0}^{\infty} K_\iota e^{-\gamma y} dy &= 1, \\
\frac{\gamma}{\mu} + \frac{K_\iota}{\gamma} &= 1,
\end{aligned}
$$

yielding (5.29) and (5.28). ■

Proposition 5.2 demonstrates that the *form* of the pdf of actual wait $f_\iota(x)$, is the same as the *form* of the pdf of the virtual wait $f(x)$, $x > 0$; the term $e^{-\gamma x}$ is common to both pdfs. The form of $P_{0,\iota}$ is similar to the form of P_0 (formula 5.13); however, $K_\iota \neq K$, except when the arrival stream is Poisson.

Remark 5.4 Formula (5.28) for $P_{0,\iota}$ matches the result derived later in formula (8.23), Sect. 8.3.3 *in Chap.* 8, via **embedded** *LC*. The embedded LC result **is** indeed the value of $P_{0,\iota}$, since it is the steady-state pdf of the **actual wait** W_n as $n \to \infty$. This match validates the standard 'continuous' LC approach utilized here. In many models, it is easier to apply standard LC than *embedded LC*. Embedded LC is useful per se: for checking results obtained by other means; analyzing new models; combining with continuous LC to obtain new results; and so forth.

Remark 5.5 $P_{0,\iota}$ and $f_\iota(x)$ in (5.28) and (5.27) agree with formulas obtained by different techniques (e.g., pp. 250–254 in [83]; pp. 259–263 in [84]). The constant γ above is equal to $\mu(1 - r_0)$, where r_0 is the solution for z in the equation '$z = A^*(\mu(1 - z))$, $z \in (0, 1)$', in those references.

5.1.8 Stability Condition for G/M/1

Stability occurs iff the solution of Eq. (5.14) for γ is positive and finite, i.e., iff the steady-state pdfs $f(x) = Ke^{-\gamma x}$ (virtual wait) and $f_\iota(x) = K_\iota e^{-\gamma x}$ (arrival-point wait) exist. These pdfs exist provided γ is positive and finite, in which case K and K_ι are also positive and finite by (5.12) and (5.29) respectively.

Denote the expected inter-arrival time by $1/\boldsymbol{a}$; the expected service time is $1/\mu$.

Proposition 5.3 The G/M/1 queue is stable if and only if $\boldsymbol{a} < \mu$.

Proof The queue is stable iff $0 < E(\mathcal{B}) < \infty$ iff $0 < \gamma < \infty$, since $\mathcal{B} = 1/\gamma$. If $0 < \gamma < \infty$ then $0 < e^{-\gamma y} < 1$ for all $y > 0$. If in addition γ satisfies Eq. (5.14), then

$$\frac{1}{\mu} = \int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy < \int_{y=0}^{\infty} \overline{A}(y)dy = \frac{1}{\boldsymbol{a}} \implies \boldsymbol{a} < \mu.$$

Hence $\boldsymbol{a} < \mu$ is a *necessary* condition for stability.

Conversely, if $\boldsymbol{a} < \mu$ then

$$\frac{1}{\mu} < \frac{1}{\boldsymbol{a}} = \int_{y=0}^{\infty} \overline{A}(y)dy.$$

We construct a function of γ, $\phi(\gamma) := \int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy$, $0 < \gamma < \infty$, with $\phi(\gamma) > 0$, $\lim_{\gamma\downarrow 0} \phi(\gamma) = \frac{1}{\boldsymbol{a}}$, $\lim_{\gamma\to\infty} \phi(\gamma) = 0$, $\phi'(\gamma) = -\gamma\phi(\gamma) < 0$, $\phi''(\gamma) = \gamma^2\phi(\gamma) > 0$. Thus $\phi(\gamma)$ is continuous, convex and strictly monotone decreasing on $(0, \infty)$. Consequently $\phi(\gamma)$ assumes each value in its range $\left(0, \frac{1}{\boldsymbol{a}}\right)$. For a given μ such that $\frac{1}{\mu} \in (0, \frac{1}{\boldsymbol{a}})$, there is a unique value $\gamma \in (0, \infty)$ such that $\phi(\gamma) = \frac{1}{\mu}$. Hence for each such μ there exists exactly one *finite* root $\gamma > 0$ of Eq. (5.14) such that $\frac{1}{\mu} < \frac{1}{\boldsymbol{a}}$, which implies $\boldsymbol{a} < \mu$ is a *sufficient* condition for stability. ■

In conclusion, $\boldsymbol{a} < \mu$ is a necessary and sufficient condition for stability.

5.1.9 Steady-State Distribution of System Time

Let $W_{q,\iota}$, S and σ denote respectively the steady-state actual wait before service, the service time, and the system time of a customer; then $\sigma = W_{q,\iota} + S$. The cdf of $W_{q,\iota}$ is $P(W_{q,\iota} \le x) = F_\iota(x)$, $x \ge 0$, having pdf $f_\iota(x)$, $x > 0$. Also $P(W_{q,\iota} = 0) = F_\iota(0) = P_{0,\iota}$ (see Proposition 5.2 in Sect. 5.1.8). Let $F_\sigma(x)$, $x \ge 0$, and $f_\sigma(x) = \frac{d}{dx}F_\sigma(x)$, $x > 0$, wherever the derivative exists, denote the steady-state cdf and pdf of σ, respectively. For the standard G/M/1 queue, S and $W_{q,\iota}$ are independent.

From Proposition 5.2, the cdf of σ is the convolution

$$\begin{aligned} F_\sigma(x) &= P_{0,\iota}P(S \le x) + \textstyle\int_{y=0}^{x} P(S \le x-y) f_\iota(y)dy \\ &= \tfrac{\gamma}{\mu}(1-e^{-\mu x}) + \textstyle\int_{y=0}^{x} \left(1-e^{-\mu(x-y)}\right)\gamma\left(1-\tfrac{\gamma}{\mu}\right)e^{-\gamma y}dy. \end{aligned} \tag{5.32}$$

The last integral in (5.32) is equal to

$$\frac{1}{\mu}\left[\mu e^{(\mu+\gamma)x} - \gamma e^{(\mu+\gamma)x} + \gamma e^{\gamma x} - \mu e^{\mu x}\right]e^{-(\mu+\gamma)x}. \tag{5.33}$$

Summing (5.33) and $\frac{\gamma}{\mu}(1-e^{-\mu x})$ simplifies (5.32) to

$$F_\sigma(x) = 1 - e^{-\gamma x}, x \ge 0, \tag{5.34}$$

with pdf

$$f_\sigma(x) = \gamma e^{-\gamma x}, x > 0. \tag{5.35}$$

Remark 5.6 The expressions for $F_\sigma(x)$ and $f_\sigma(x)$ in (5.34) and (5.35) for G/M/1 are analogous to those for the standard $M_\lambda/M_\mu/1$ queue given in (3.117), with $\gamma=\mu-\lambda$. In the term $e^{-\gamma x}$ the coefficient of x is $-\gamma=-1/E(\mathcal{B})$ in both G/M/1 and M/M/1 ($\mathcal{B}$:= busy period).

5.1.10 Arrival-Point PMF of Number in System

This section derives the steady-state *arrival-point pmf* (probability mass function) of the number of units in the system. Specifically, let N_ι denote the number in the system *just before an arrival instant*, and let $P_{n,\iota} := P(N_\iota = n)$, $n = 0, 1, \ldots$. From formulas (5.29) and (5.28),

$$P_{0,\iota} = \frac{\gamma}{\mu}. \tag{5.36}$$

Let d_n be the steady-state probability of n remaining in the system *just after a departure instant*. Then $P_{n,\iota} = d_n$, $n = 0, 1, \ldots$ (see pp. 500–502 in [125]). Let $A^{(n)}(y)$ be the cdf of the n-fold convolution of the interarrival time evaluated at $y > 0$.

Proposition 5.4 In the standard G/M/1 queue, for $n = 1, 2, \ldots$,

$$\begin{aligned} P_{n,\iota} = d_n &= \int_{y=0}^{\infty}\left[A^{(n)}(y) - A^{(n+1)}(y)\right]f_\sigma(y)dy \\ &= \gamma\int_{y=0}^{\infty}\left[A^{(n)}(y) - A^{(n+1)}(y)\right]e^{-\gamma y}dy. \end{aligned} \tag{5.37}$$

Proof Let $N_A(t)$ be the number of arrivals in $(0, t)$ and let $\mathcal{S}_n$ be the time of the n^{th} arrival. A basic renewal equivalence relation is (see, p. 423, Sect. 7.2 in [125]; pp. 167–168 in [99])

$$N_A(t) \geq n \iff \mathcal{S}_n \leq t.$$

Thus

$$\begin{aligned} P(N_A(t) = n) &= P(N_A(t) \geq n) - P(N_A(t) \geq n + 1) \\ &= P(\mathcal{S}_n \leq t) - P(\mathcal{S}_{n+1} \leq t) \\ &= A^{(n)}(t) - A^{(n+1)}(t), t > 0 \end{aligned}$$

Also, $d_n = P(n$ arrivals *during a system time* $\sigma)$. By independence of the arrival stream and σ,

$$\begin{aligned} d_n &= \int_{y=0}^{\infty} P(N_A(y) = n | \sigma = y) f_\sigma(y) dy \\ &= \int_{y=0}^{\infty} P(N_A(y) = n) f_\sigma(y) dy \\ &= \int_{y=0}^{\infty} \left[A^{(n)}(y) - A^{(n+1)}(y) \right] f_\sigma(y) dy. \end{aligned}$$

Also, $P_{n,\iota} = d_n$. Substituting from formula (5.35) gives (5.37). ■

Compact Expression for PMF

Proposition 5.4 leads to a compact expression for $P_{n,\iota}$, $n = 0, 1, \ldots$. Integration by parts gives

$$\int_{y=0}^{\infty} A^{(n)}(y) e^{-\gamma y} dy = \frac{1}{\gamma} \int_{y=0}^{\infty} a^{(n)}(y) e^{-\gamma y} dy = \frac{A^{n*}(\gamma)}{\gamma},$$

where $a^{(n)}(y)$ is the pdf of the n-fold convolution of inter-arrival times. Thus (5.37) becomes

$$P_{n,\iota} = A^{n*}(\gamma) - A^{(n+1)*}(\gamma), n = 1, 2, \ldots . \tag{5.38}$$

From Laplace-transform theory and (5.15)

$$A^{n*}(\gamma) = \left(A^*(\gamma)\right)^n = \left(1 - \frac{\gamma}{\mu}\right)^n, n = 1, 2, \ldots .$$

Substituting into (5.38) yields

$$\begin{aligned}P_{n,\iota} &= \left(1-\frac{\gamma}{\mu}\right)^n - \left(1-\frac{\gamma}{\mu}\right)^{n+1}\\ &= \frac{\gamma}{\mu}\left(1-\frac{\gamma}{\mu}\right)^n\\ &= P_{0,\iota}\left(1-P_{0,\iota}\right)^n, n = 0, 1, ...,\end{aligned} \tag{5.39}$$

using (5.36). Notably, formula (5.39) is analogous to the result for P_n in terms of P_0 in M/M/1, given in (3.118) in Chap. 3.

As a caveat to Proposition 5.4, the probabilities of n in the system *at an arbitrary time point* are *not* equal to $P_{n,\iota}, n = 0, 1, 2, \ldots$. Equality holds only if the arrival stream is Poisson (see Proposition 8.2, p. 502 in [125], and also [145]).

5.1.11 G/M/1 with Poisson Arrivals

To enhance intuition, we specialize the foregoing G/M$_\mu$/1 results to M$_\lambda$/M$_\mu$/1, where arrivals are Poisson with rate λ.

Virtual Wait

From Eq. (5.14) for γ, with $\overline{A}(x) := e^{-\lambda x}$, $x \geq 0$, the solution is then $\gamma = \mu - \lambda = 1/E(\mathcal{B})$, where $\mathcal{B}$ is the busy period in M/M/1. From formula (5.18) we get $C_\gamma = \lambda$. The formulas in Sect. 5.1.5 yield

$$\begin{gathered}P_0 = \frac{\gamma}{\gamma+C_\gamma} = \frac{\mu-\lambda}{\mu-\lambda+\lambda} = 1-\frac{\lambda}{\mu},\\ K = \frac{\gamma\cdot C_\gamma}{\gamma+C_\gamma} = \lambda(1-\frac{\lambda}{\mu}) = \lambda P_0,\\ f(x) = Ke^{-\gamma x} = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0,\\ E(\mathcal{B}) = \frac{1}{\gamma} = \frac{1}{\mu-\lambda},\\ E(\mathcal{I}) = \frac{1}{C_\gamma} = \frac{1}{\lambda},\end{gathered} \tag{5.40}$$

which check with the steady-state virtual-wait quantities for M/M/1.

For $x < 0$, the pdf of extended age is

$$h(x) = \mu\int_{y=0}^{\infty} e^{-\lambda(y-x)}Ke^{-(\mu-\lambda)y}dy = Ke^{\lambda x}, x < 0,$$

whence $P_0 = \int_{x=-\infty}^{0} h(x)dx = 1-\frac{\lambda}{\mu}$, which agrees with (5.40).

Actual Wait

For the *actual wait* in G/M/1 with Poisson arrivals, $\gamma = \mu - \lambda$, $P_{0,\iota} = \gamma/\mu = 1 - \lambda/\mu$, and $K_\iota = \gamma(1 - \gamma/\mu) = \lambda(\frac{\mu-\lambda}{\mu}) = \lambda P_{0,\iota}$, giving $f_\iota(x) = \lambda P_{0,\iota} e^{-(\mu-\lambda)x}$, $x > 0$. These actual-wait results agree with P_0 and $f(x)$, $x > 0$ for the virtual wait, as they must in M/M/1 (see (3.112)), where the Poisson arrival stream implies

$$P_{0,\iota} = P_0 = 1 - \frac{\lambda}{\mu},\ f_\iota(x) = f(x), x > 0,$$
$$\text{and } P_{n,\iota} = P_n = \left(\frac{\lambda}{\mu}\right)^n P_0,$$

in agreement with PASTA.

5.1.12 Sojourn Time Above or Below a Level

We next determine the expected value of the sojourn time above or below a state-space level (see Fig. 5.1).

Inter-upcrossing Time u_x of Level x

Let u_x denote the *inter-upcrossing* time of level x (time between two successive upcrossings). We consider the cases $x \geq 0$ and $x < 0$ separately, since the SP motion above and below level 0 are of a different nature (Fig. 5.1).

Level $x \geq 0$ For $x \geq 0$, upcrossings of x are regenerative points, since both service times and remaining service times after upcrossing x are $\underset{dis}{=} \text{Exp}_\mu$. By the elementary renewal theorem and LC,

$$E(u_x) = \frac{1}{\lim_{t\to\infty} \mathcal{U}_t(x)/t} = \frac{1}{f(x)} = \frac{e^{\gamma x}}{K}, x \geq 0, \tag{5.41}$$

where γ and K are given in (5.14) and (5.21), respectively. (To compute K, we may use C_γ given in (5.18).)

Level $x < 0$ For $x < 0$, $-x$ is the time until the next arrival instant, at which a sample path of $\{V(t)\}_{t\geq 0}$ hits level 0 from below. Upcrossings of x are regenerative points, since the time $-x$ is followed by a service time $\underset{dis}{=} \text{Exp}_\mu$. By the elementary renewal theorem, LC, and (5.16),

$$E(u_x) = \frac{1}{\lim_{t\to\infty} \mathcal{U}_t(x)/t} = \frac{1}{h(x)} = \frac{1}{\mu K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy}, x < 0. \tag{5.42}$$

Sojourn Time a_x Above Level x

Let a_x := *sojourn time of* $\{V(t)\}_{t\ge 0}$ *above level* x.

Level $x \ge 0$ For $x \ge 0$, $a_x \underset{dis}{=} \mathcal{B}$ ($\mathcal{B}$:= busy period) *independent* of x, due to exponential service times. By (5.12)

$$E(a_x) = E(\mathcal{B}) = \frac{1}{\gamma}, x \ge 0. \tag{5.43}$$

Level $x < 0$ Since $u_x = a_x + b_x$,

$$\begin{aligned} E(a_x) &= E(u_x) - E(b_x) = \tfrac{1}{h(x)} - E(b_x) \\ &= \frac{1}{\mu K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy} \\ &\quad - \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} \overline{A}(y-x+z)Ke^{-\gamma y}dy}{\int_{y=0}^{\infty} \overline{A}(y-x)Ke^{-\gamma y}dy} \right] dz, x < 0, \end{aligned} \tag{5.44}$$

where b_x := *sojourn time below level* x, and $E(b_x)$, $x < 0$, is derived in Proposition 5.5 below in this section.

Sojourn Time b_x Below Level x

Level $x \ge 0$ Because $u_x = a_x + b_x$,

$$E(b_x) = E(u_x) - E(a_x) = \frac{e^{\gamma x}}{K} - \frac{1}{\gamma}, x \ge 0.$$

Level $x < 0$ For $x < 0$, b_x is given by the following proposition.

Proposition 5.5 The expected sojourn time of $\{V(t)\}_{t\ge 0}$ below level x is

$$E(b_x) = \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} \overline{A}(y-x+z)e^{-\gamma y}dy}{\int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy} \right] dz, x < 0. \tag{5.45}$$

Proof We present two proofs, which complement each other.

Proof 1. Consider an SP downward jump that ends below level $x < 0$ (all jumps start above level 0). Let r_x := *excess of a jump ending below* x. Since a sample path of $\{V(t)\}_{t\ge 0}$ increases steadily at rate $+1$ when $V(t) < 0$,

$b_x \underset{dis}{=} r_x$, and $E(b_x) = E(r_x)$. Using a 'probabilistic' approach, condition on the event 'a jump downcrosses level x', yielding

$$\begin{aligned}
&P(r_x > z) \\
&= P(\text{a jump downcrosses level } x - z | \text{the jump downcrosses level } x) \\
&= \frac{P(\text{a jump downcrosses level } x - z \textit{ and } \text{downcrosses level } x)}{P(\text{the jump downcrosses level } x)} \\
&= \frac{P(\text{a jump downcrosses level } x - z)}{P(\text{the jump downcrosses level } x)} \\
&= \frac{\int_{y=0}^{\infty} \overline{A}(y - x + z)\cancel{K}e^{-\gamma y} dy}{\int_{y=0}^{\infty} \overline{A}(y - x)\cancel{K}e^{-\gamma y} dy}, x < 0, z > 0.
\end{aligned}$$

Thus

$$\begin{aligned}
E(b_x) = E(r_x) &= \int_{z=0}^{\infty} P(r_x > z) dz \\
&= \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} \overline{A}(y - x + z)e^{-\gamma y} dy}{\int_{y=0}^{\infty} \overline{A}(y - x)e^{-\gamma y} dy} \right] dz, x < 0.
\end{aligned}$$

Proof 2. Alternatively, denote the cdf of r_x by $F_{r_x}(z), z > 0$. Using a 'rate' approach, we have

$$\left(1 - F_{r_x}(z)\right) \cdot h(x) = \mu \int_{y=0}^{\infty} \overline{A}(y - x + z)Ke^{-\gamma y} dy,$$

where the left and right sides are two *different* expressions for the *SP downcrossing rate of level x-z*. (A similar argument is used in the derivation of $E(a_x)$ in Variant 3 of M/G/1 with bounded virtual wait, given in formula (3.257).) Since $h(x) = \mu \int_{y=0}^{\infty} \overline{A}(y - x)Ke^{-\gamma y} dy$, and $b_x = r_x$,

$$\begin{aligned}
1 - F_{r_x}(z) &= \frac{\mu \int_{y=0}^{\infty} \overline{A}(y - x + z)\cancel{K}e^{-\gamma y} dy}{\mu \int_{y=0}^{\infty} \overline{A}(y - x)\cancel{K}e^{-\gamma y} dy}, \\
E(b_x) &= \int_{z=0}^{\infty} \left(1 - F_{r_x}(z)\right) dz \\
&= \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} \overline{A}(y - x + z)e^{-\gamma y} dy}{\int_{y=0}^{\infty} \overline{A}(y - x)e^{-\gamma y} dy} \right] dz, x < 0, \qquad (5.46)
\end{aligned}$$

in agreement with Proof 1. ■

Example 5.1 Assume G/M/1 is an $M_\lambda/M_\mu/1$ queue. From (5.45)

$$E(b_x) = E(r_x) = \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} e^{-\lambda(y-x+z)} e^{-\gamma y} dy}{\int_{y=0}^{\infty} e^{-\lambda(y-x)} e^{-\gamma y} dy} \right] dz$$
$$= \int_{z=0}^{\infty} e^{-\lambda z} dz = \frac{1}{\lambda}, x < 0,$$

using the memoryless property of the interarrival tines ($\underset{dis}{=} \text{Exp}_\lambda$).

5.1.13 *Events During a Sojourn a_x Above a Level*

System Time σ

System time σ := *wait* + *service time* = $W_{q,\iota} + S$. The σs are realized at end-of-service instants (at departures from the system—i.e., sample-path peaks). The $W_{q,\iota}$s are realized at start of service instants—i.e., sample-path troughs. (See Figs. 5.1 and 5.2.)

Number $N_{a_x}^{\sigma}$ of Realized System Times in a_x

Level $x \geq 0$ Let $N_{a_x}^{\sigma}$:= *number of service completions (departures) during* a_x, $x \geq 0$, *including*the departure that that ends a_x (i.e., the length of a run of system times $> x$, or number of sample-path peaks $> x$ in a_x). Let S_i and T_i, $i = 1, 2,...,$ denote the service times and immediately following inter-arrival times, respectively, counting from the instant a sample path of $\{V(t)\}_{t\geq 0}$ upcrosses level x (start of a_x). If $x = 0$, then S_1 is a full service time. If $x > 0$, then S_1 is the remaining service time measured from level x, and $S_1 \underset{dis}{=} \text{Exp}_\mu$, due to the memoryless property. Then (see Fig. 5.1)

$$N_{a_x}^{\sigma} = \min\left\{ n \middle| \sum_{i=1}^{n} (S_i - T_i) \leq 0 \right\}, x \geq 0. \tag{5.47}$$

$N_{a_x}^{\sigma}$ is a stopping time (e.g., pp. 678–679 in [125]) for both sequences $\{S_i - T_i\}_{i=1,2....}$ and $\{S_i\}_{i=1,2,...}$. Since $a_x = \sum_{i=1}^{N_{a_x}^{\sigma}} S_i$ and $a_x \underset{dis}{=} \mathcal{B}$ for all $x \geq 0$, by Wald's equation (e.g., p. 47ff in [122]; p. 679 in [125])

$$E(a_x) = E(N_{a_x}^{\sigma})E(S_i),$$
$$E(N_{a_x}^{\sigma}) = \frac{E(a_x)}{E(S_i)} = \frac{E(\mathcal{B})}{E(S)}, \tag{5.48}$$

independent of x. Substituting from (5.43) into (5.48), gives

$$E(N_{a_x}^{\sigma}) = \frac{\frac{1}{\gamma}}{\frac{1}{\mu}} = \frac{\mu}{\gamma}, x \geq 0. \tag{5.49}$$

From (5.49) $E(N_{a_x}^{\sigma}) > 1$ since $\mu > \gamma$ (see Remark 5.7 below). This agrees with intuition, because there must be at least one departure (sample-path peak) counted in a_x (the last departure in a_x).

Number Served in a Busy Period $\mathcal{B}$

Let $N_{\mathcal{B}}^{\sigma}$:= *number of system-time realizations in* $\mathcal{B}$ (:= number of sample-path peaks in $\mathcal{B}$, i.e., number of customers served in $\mathcal{B}$). (This section is related to Sect. 3.4.14 in Chap. 3 for M/G/1.) Since $\mathcal{B} = a_0$, $N_{\mathcal{B}}^{\sigma} = N_{a_0}^{\sigma}$, and from (5.49), $E(N_{\mathcal{B}}^{\sigma}) = \mu/\gamma$. From (5.31), we obtain a notable formula for G/M/1,

$$E\left(N_{\mathcal{B}}^{\sigma}\right) = \frac{\mu}{\gamma} = \frac{1}{P_{0,\iota}}, \tag{5.50}$$

which is similar to formula (3.90) for M/G/1 queues.

Number of Realized Waiting Times (Service Starts) $N_{a_x}^{w}$ in a_x

We assume $x \geq 0$, since service starts can occur only when $V(t) \geq 0$. Let $N_{a_x}^{w}$ denote the number of customers that *start* service during a_x, $x \geq 0$. Then $N_{a_x}^{w}$ is the number of customers that *wait* $> x$ time units (strictly greater than x) *before starting service* during a_x. Figure 5.1 shows that $N_{a_x}^{w} = N_{a_x}^{\sigma} - 1, x \geq 0$, since the count of service starts *with waits* $> x$ during a_x is one less than the count of service completions in a_x (including the a_x-ending completion). Hence

$$E(N_{a_x}^{w}) = E(N_{a_x}^{\sigma}) - 1 = \frac{\mu}{\gamma} - 1 > 0, x \geq 0. \tag{5.51}$$

Remark 5.7 In (5.51) the inequality $\frac{\mu}{\gamma} - 1 > 0$ holds because of (5.14), i.e., $\int_{y=0}^{\infty} \overline{A}(y) e^{-\gamma y} dy = \frac{1}{\mu}$; $\overline{A}(0) = 1$; $\overline{A}(y) = 1 - A(y)$ is non-increasing with $\lim_{y \to \infty} \overline{A}(y) = 0$. Thus there exists finite $M_{\varepsilon} > 0$ such that $\overline{A}(y) < \varepsilon < 1$ (strictly) for $y > M_{\varepsilon}$. Hence

$$\begin{aligned}
\frac{1}{\mu} &< \int_{y=0}^{M_{\varepsilon}} 1 \cdot e^{-\gamma y} dy + \int_{y=M_{\varepsilon}}^{\infty} \varepsilon \cdot e^{-\gamma y} dy \\
&= \frac{1}{\gamma}\left[1 - e^{-\gamma M_{\varepsilon}}(1-\varepsilon)\right] < \frac{1}{\gamma} \quad \Longrightarrow \frac{\mu}{\gamma} > 1.
\end{aligned}$$

5.1.14 Events Above Level x During a Busy Period $\mathcal{B}$

A busy period $\mathcal{B}$ may contain several sojourns above level x, e.g., in Fig. 5.1 the first three busy periods contain 2, 2, and 1 sojourns above level x, respectively.

Number of SP Sojourns $N_{a_x}^{\text{soj}}(\mathcal{B})$ above level x during $\mathcal{B}$

Let $N_{a_x}^{\text{soj}}(\mathcal{B})$:= *number of* ***sojourns*** *above level* $x \geq 0$ during $\mathcal{B}$ ($N_{a_x}^{\text{soj}}(\mathcal{B}) =$ 0, 1, 2,...). We first obtain $E(N_{a_x}^{\text{soj}}(\text{B}))$.

Level $x \geq 0$

Let $\mathcal{C}$ denote a busy *cycle*. Let $\mathcal{U}_{\mathcal{C}}(x)$ denote the number of SP *upcrossings* of level x during $\mathcal{C}$, *all of which occur during the embedded* $\mathcal{B}$. There is a one-to-one correspondence between the starts of sojourns a_x and SP upcrossings of x, implying $N_{a_x}^{\text{soj}}(\mathcal{B}) = \mathcal{U}_{\mathcal{C}}(x)$ (see Fig. 5.1). By the renewal reward theorem,

$$\frac{E(N_{a_x}^{\text{soj}}(\mathcal{B}))}{E(\mathcal{C})} = \frac{E(\mathcal{U}_{\mathcal{C}}(x))}{E(\mathcal{C})} = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = f(x) = Ke^{-\gamma x}, x \geq 0. \tag{5.52}$$

Since $E(\mathcal{C}) = 1/f(0) = 1/K$, from (5.52)

$$E\left(N_{a_x}^{\text{soj}}(\mathcal{B})\right) = \frac{1}{\not K} \cdot \not K e^{-\gamma x} = e^{-\gamma x}, x \geq 0. \tag{5.53}$$

To satisfy intuition, setting $x = 0$ in (5.53) implies $E\left(N_{a_0}^{\text{soj}}(\mathcal{B})\right) = 1$, which “says” that $\mathcal{B}$ consists of exactly one sojourn a_0 above level 0.

Number $N_{a_x}^{\sigma}(\mathcal{B})$ of System Times $> x$ in a Busy Period $\mathcal{B}$

Let $N_{a_x}^{\sigma}(\mathcal{B})$:= number of system times $> x$ during $\mathcal{B}$, and $N_{a_x}^{\sigma}(\mathcal{C})$:= number of system times $> x$ during $\mathcal{C}$ respectively (see Fig. 5.1). $N_{a_x}^{\sigma}(\mathcal{B}) = N_{a_x}^{\sigma}(\mathcal{C})$ since all departures in $\mathcal{C}$ occur during, or at the end of, the embedded $\mathcal{B}$. All departures having $\sigma > x$ occur *within sojourns* a_x in $\mathcal{B}$. So,

$$N_{a_x}^{\sigma}(\mathcal{B}) = N_{a_x}^{\sigma}(\mathcal{C}) = \sum_{i=1}^{N_{a_x}^{\text{soj}}(\mathcal{C})} N_{a_x,i}^{\sigma}, \tag{5.54}$$

where $N_{a_x,i}^{\sigma}$ is the number of system times $> x$ in the i^{th} *sojourn* above x during $\mathcal{C}$. The $N_{a_x,i}^{\sigma}$s are i.i.d. r.v.s with $E\left(N_{a_x,i}^{\sigma}\right) = \mu/\gamma$, by (5.49), independent of the number of sojourns $N_{a_x}^{\mathrm{soj}}(\mathcal{C})$ (because excess service above $x \underset{dis}{=} \mathrm{Exp}_\mu$). Taking expected values in (5.54) and using (5.49) and (5.53), gives

$$E\left(N_{a_x}^{\sigma}(\mathcal{B})\right) = E\left(N_{a_x}^{\sigma}(\mathcal{C})\right)$$
$$= E\left(N_{a_x,i}^{\sigma}\right) \cdot E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{C})\right) = \frac{\mu}{\gamma} e^{-\gamma x}, x \geq 0. \tag{5.55}$$

Number $N_{a_x}^{w}(\mathcal{B})$ of Waiting Times $> x$ in a Busy Period $\mathcal{B}$

We obtain the expected number of *waiting times* $> x$ in $\mathcal{B}$, similarly as for the derivation of (5.51) (see Remark 5.7 in Sect. 5.1.13). Thus

$$E\left(N_{a_x}^{w}(\mathcal{B})\right) = \left(\frac{\mu}{\gamma} - 1\right) e^{-\gamma x}, x \geq 0. \tag{5.56}$$

Setting $x = 0$ in (5.56) gives $E\left(N_{a_0}^{w}(\mathcal{B})\right) = \frac{\mu}{\gamma} - 1$, which is also the expected number of customers in $\mathcal{B}$ that wait > 0 (same as (5.51) since $a_x \underset{dis}{=} \mathcal{B}$). Only the first customer in $\mathcal{B}$ waits 0.

Proportion of Customers that Wait $> x$

N.B. Here we use 'proportion' to mean the ratio $E\left(N_{a_x}^{w}(\mathcal{B})\right)/E\left(N_{\mathcal{B}}^{\sigma}\right)$, not $N_{a_x}^{w}(\mathcal{B})/N_{\mathcal{B}}^{\sigma}$ or $E\left(N_{a_x}^{w}(\mathcal{B})/N_{\mathcal{B}}^{\sigma}\right)$.

We assume level $x \geq 0$, since all waits in line are ≥ 0, and connect the 'proportion' of customers that wait $> x$ in $\mathcal{B}$ with other parameters of the model. For example, the *proportion* of customers that wait $> x$ in $\mathcal{B}$ is, using (5.31),

$$\frac{E\left(N_{a_x}^{w}(\mathcal{B})\right)}{E\left(N_{\mathcal{B}}^{\sigma}\right)} = \frac{\left(\frac{\mu}{\gamma} - 1\right) e^{-\gamma x}}{\frac{\mu}{\gamma}} = \left(1 - \frac{\gamma}{\mu}\right) e^{-\gamma x} = \left(1 - P_{0,\iota}\right) e^{-\gamma x}, x \geq 0, \tag{5.57}$$

where $N_{\mathcal{B}}^{\sigma} :=$ *number of service completions in* $\mathcal{B}$ (= number of arrivals in $\mathcal{C}$). When $x = 0$, (5.57) reduces to $E\left(N_{a_x}^{w}(\mathcal{B})\right)/E\left(N_{\mathcal{B}}^{\sigma}\right) = 1 - P_{0,\iota}$, which is intuitive, because $\mathcal{C}$ is a probabilistic microcosm of the evolution of the system over the entire time axis and the long-run proportion of customers that wait > 0 is $1 - P_{0,\iota}$.

Level $x < 0$

Number Served in a Sojourn a_x Above Level x

Fix a level $x < 0$. After upcrossing x, a sample path of $\{V(t)\}_{t \geq 0}$ ascends steadily at rate $+1$ to level 0. Hence the *number* of service completions during a_x is

$$N_{a_x}^{\sigma} = \min\left\{n| \sum_{i=1}^{n} \left(r_{0,i} + S_i - T_i\right) > -x\right\}, \tag{5.58}$$

where $r_{0,i}$:= *excess below level* 0 *due to an SP jump that downcrosses level* 0 (see Fig. 5.3). The values of $r_{0,i},\ i = 1, ..., n$, are given by

$r_{0,1} := -x$,
$r_{0,i} = 0$ if the jump does not downcross level 0,
$0 < r_{0,i} < -x$ iff a_x continues with another service,
$r_{0,i} > -x$ iff a_x ends with the i^{th} jump (which leaps below x).

The cdf of $r_{0,i}$ and $E(r_{0,i})$ are given above (see formula (5.46) as $x \uparrow 0$, in Proposition 5.5, in Sect. 5.1.12). Thus

$$E(r_{0,i}) = E(b_0) = \int_{z=0}^{\infty} \left[\frac{\int_{y=0}^{\infty} \overline{A}(y+z)e^{-\gamma y}dy}{\int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy}\right] dz.$$

$N_{a_x}^{\sigma}$ is a stopping time for $\left\{r_{0,i} + S_i - T_i\right\}_{i=1,2,...}$ and for $\left\{r_{0,i} + S_i\right\}_{i=1,2,...}$. The *sojourn time* above level x is $a_x = \sum_{i=1}^{N_{a_x}} \left(r_{0,i} + S_i\right)$, implying

$$E(a_x) = E(N_{a_x}^{\sigma}) \cdot \left[E(r_{0,i}) + E(S_i)\right].$$

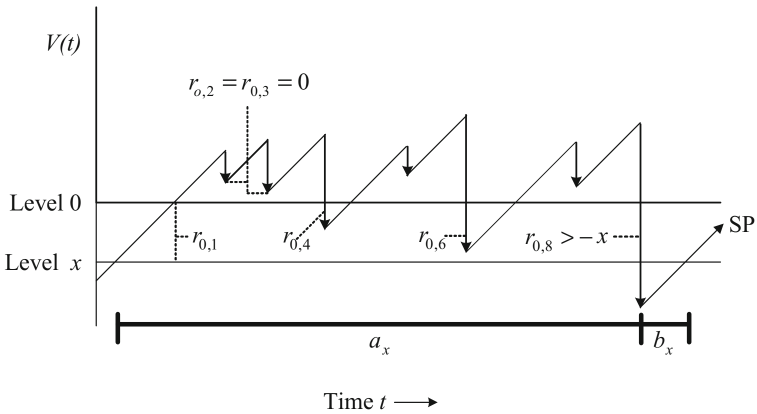


Fig. 5.3 a_x and b_x in G/M/1 queue for $x < 0$. Illustrates jump overshoots: $r_{0,1} := -x$; $r_{0,i}$= 0 if jump ends above 0; $0 < r_{0,i} < -x$ iff a_x continues; if $r_{0,i} > -x$ then a_x ends and b_x begins

Thus

$$E(N_{a_x}^{\sigma}) = \frac{E(a_x)}{E(r_{0,i}) + E(S_i)} = \frac{E(a_x)}{E(r_i) + 1/\mu}, \tag{5.59}$$

where $E(r_{0,i})$ is given in (5.44), and $E(r_i)$ is given in (5.45).

5.1.15 Revisit of M/M/1

Consider the M/M/1 queue in the light of the G/M/1 results in Sects. 5.1.13–5.1.14. Let 'G' in G/M/1 be Poisson at rate λ. Then $\gamma = \mu - \lambda$. From Eq. (5.56) for G/M/1,

$$\begin{aligned} E\left(N_{a_x}^{w}(\mathcal{B})\right) &= \left(\frac{\mu}{\gamma} - 1\right) e^{-\gamma x} = \left(\frac{\mu}{\mu - \lambda} - 1\right) e^{-\gamma x} \\ &= \left(\frac{1}{1 - \frac{\lambda}{\mu}} - 1\right) e^{-\gamma x} = \left(\frac{1}{P_0} - 1\right) e^{-\gamma x}, x \geq 0. \end{aligned} \tag{5.60}$$

When $x = 0$, (5.60) reduces to $E\left(N_{a_0}^{w}(\mathcal{B})\right) = 1/P_0 - 1$.

In M/M/1 (and M/G/1), $E\left(N_{\mathcal{B}}^{\sigma}\right) := E($*number of customers* ***served in*** $\mathcal{B}) = 1/P_0$ (formula (3.90)). The customer initiating $\mathcal{B}$ waits 0; all others served in $\mathcal{B}$ wait > 0. This explains intuitively why $E\left(N_{a_0}^{w}(\mathcal{B})\right) = E\left(N_{\mathcal{B}}^{\sigma}\right) - 1$.

In M/M/1 (and M/G/1) the *proportion* of customers that wait > 0 in a busy period is

$$\frac{E\left(N_{a_0}^{w}(\mathcal{B})\right)}{E\left(N_{\mathcal{B}}^{\sigma}\right)} = \frac{\frac{1}{P_0} - 1}{\frac{1}{P_0}} = 1 - P_0 = \frac{\lambda}{\mu} = \rho,$$

which agrees with the result for G/M/1 given in (5.57).

Related Results for $\text{M}_\lambda/\text{M}_\mu/1$

Similarly as in the analyses above for G/M/1, we obtain the following results for $\text{M}_\lambda/\text{M}_\mu/1$ (see Fig. 3.7). The expected number of system times completed in a sojourn a_x is

$$E\left(N_{a_x}^{\sigma}\right) = E\left(N_{\mathcal{B}}^{\sigma}\right) = \frac{\mu}{\mu - \lambda} = \frac{1}{P_0}, x \geq 0, \tag{5.61}$$

since $N_{a_x}^{\sigma} \underset{dis}{=} N_{\mathcal{B}}^{\sigma}$ in both M/M/1 and G/M/1, $x \geq 0$. Also, $E\left(N_{a_x}^{\sigma}\right) > 1$ since $\mu > \mu - \lambda$ (i.e., for stability $0 < \lambda < \mu$).

$E($number that wait $> x$ during $a_x)$ is

$$E(N_{a_x}^{w}) = \frac{\mu}{\mu - \lambda} - 1 = \frac{1}{P_0} - 1 = \frac{\lambda}{\mu - \lambda} > 0. \tag{5.62}$$

E(number of sojourns above level x in $\mathcal{B}$) is, using (5.53),

$$E(N_{a_x}^{\text{soj}}(\mathcal{B})) = e^{-\gamma x} = e^{-(\mu-\lambda)x}, x \geq 0; \tag{5.63}$$

if $x = 0$ then

$$E\left(N_{a_x}^{\text{soj}}(\mathcal{B})\right) = e^0 = E\left(N_{a_0}^{\text{soj}}(\mathcal{B})\right) = 1.$$

$\mathcal{B}$ consists of exactly one sojourn above level 0, and $\mathcal{B}$ may contain a random number of sojourns above an arbitrary level $x > 0$.

The number of system times (service completions—sample-path peaks) above level x in $\mathcal{B}$ is

$$N_{a_x}^{\sigma}(\mathcal{B}) = N_{a_x}^{\sigma}(\mathcal{C}) = \sum_{i=1}^{N_{a_x}^{\text{soj}}(\mathcal{C})} N_{a_x i}^{\sigma},$$

where $N_{a_x i}^{\sigma} = N_{a_x}^{\sigma}$ independent of $N_{a_x}^{\text{soj}}(\mathcal{C})$. By (5.61) and (5.63),

$$\begin{aligned} E\left(N_{a_x}^{\sigma}(\mathcal{B})\right) &= E\left(N_{a_x}^{\sigma}\right) = E\left(N_{\mathcal{B}}^{\sigma}\right) \cdot E\left(N_{a_x}^{\text{soj}}(\mathcal{C})\right) \\ &= \frac{\mu}{\mu - \lambda} \cdot e^{-(\mu - \lambda)x}, \geq 0. \end{aligned} \tag{5.64}$$

E(number of waiting times $> x$ in $\mathcal{B}$) is

$$\begin{aligned} E(N_{a_x}^{w}(\mathcal{B})) &= \left(\frac{\mu}{\mu - \lambda} - 1\right) \cdot e^{-(\mu - \lambda)x} \\ &= \frac{\lambda}{\mu - \lambda} \cdot e^{-(\mu - \lambda)x}, x \geq 0; \end{aligned} \tag{5.65}$$

if $x = 0$, then $E(N_{a_x}^{w}(\mathcal{B})) = \frac{\lambda}{\mu-\lambda}$ = expected number that wait > 0 in $\mathcal{B}$.

The '*proportion*' of customers that wait > 0 in $\mathcal{B}$ is

$$\frac{E\left(N_{a_0}^{w}(\mathcal{B})\right)}{E\left(N_{\mathcal{B}}^{\sigma}\right)} = \frac{\frac{\lambda}{\mu-\lambda}}{\frac{\mu}{\mu-\lambda}} = \frac{\lambda}{\mu} = 1 - P_0,$$

where $N_{\mathcal{B}}^{\sigma} :=$ *number served in B*. The intuitive explanation of the last formula is that the long-run proportion of customers that wait > 0 is $1 - P_0$ ($\mathcal{C}$ is a probabilistic replica of the entire time line. All arrivals take place in the embedded $\mathcal{B}$).

Let $N^{\sigma}_{b_x}(\mathcal{B})$:= *number of* σs $\le x$ *in* $\mathcal{B}$ (occur in a sojourn b_x below level x).

Proposition 5.6 In M/M/1 the expected number of system times $\le x$ in a busy period $\mathcal{B}$ is

$$\begin{aligned} E(N^{\sigma}_{b_x}(\mathcal{B})) &= \frac{\mu}{\mu-\lambda} - \frac{\mu}{\mu-\lambda}e^{-(\mu-\lambda)x} \\ &= \frac{\mu}{\mu-\lambda}(1-e^{-(\mu-\lambda)x}),\ x \ge 0. \end{aligned} \tag{5.66}$$

Proof $N^{\sigma}_{b_x}(\mathcal{B}) + N^{\sigma}_{a_x}(\mathcal{B}) = N^{\sigma}_{\mathcal{B}}(\mathcal{B})$. From (5.61)

$$E(N^{\sigma}_{b_x}(\mathcal{B})) + E(N^{\sigma}_{a_x}(\mathcal{B})) = E\left(N^{\sigma}_{\mathcal{B}}\right) = \frac{\mu}{\mu-\lambda}.$$

Then (5.66) follows from (5.61) and (5.64). ■

Proposition 5.7 For M/M/1 the expected number of waiting times $\le x$ in a busy period $\mathcal{B}$ is

$$E(N^{w}_{b_x}(\mathcal{B})) = \frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu-\lambda}e^{-(\mu-\lambda)x},\ x \ge 0. \tag{5.67}$$

Proof In $\mathcal{B}$, the number of customers with *waiting* times $\le x$ plus the number with waiting times $> x$, is equal to the number served in $\mathcal{B}$, namely $N^{\sigma}_{\mathcal{B}}$. By (5.61),

$$E(N^{w}_{b_x}(\mathcal{B})) + E(N^{w}_{a_x}(\mathcal{B})) = E\left(N^{\sigma}_{\mathcal{B}}\right) = \frac{\mu}{\mu-\lambda}.$$

Thus, (5.67) follows from (5.62) and (5.65). ■

Remark 5.8 In $M_{\lambda}/M_{\mu}/1$:

If $x = 0$ then $E(N^{\sigma}_{b_x}(\mathcal{B})) = 0$.

If $x \to \infty$ then $E(N^{\sigma}_{b_x}(\mathcal{B})) \to \mu/(\mu-\lambda) = 1/P_0$.

If $x = 0$ then $E(N^{w}_{b_x}(\mathcal{B})) = 1$ (initiator of $\mathcal{B}$ waits 0).

If $x \to \infty$ then $E(N^{w}_{b_x}(B)) \to \mu/(\mu-\lambda) = 1/P_0$.

5.1.16 Boundedness of PDF of Wait f(x), x > 0

In G/M/1 let $S = \text{Exp}_{\mu}$; let the inter-arrival time have cdf $A(y)$, $y > 0$, with $A(0^{+}) = 0$. Assume the steady-state pdf of wait $f(x)$, $x > 0$, exists.

The pdf of the *virtual* wait, $f(x) = Ke^{-\gamma x}$, $x > 0$, is decreasing and convex on $(0, \infty)$. Thus

$$f(x) < K = \gamma C_\gamma / \left(\gamma + C_\gamma\right) < \gamma < \mu,\ x > 0,$$

where K is given by (5.21), and C_γ by (5.18).

The pdf of the *actual* wait, $f_\iota(x) = K_\iota e^{-\gamma x}$, $x > 0$, has similar properties. Thus

$$f_\iota(x) < K_\iota = \gamma\left(1 - \frac{\gamma}{\mu}\right) < \gamma < \mu,\ x > 0,$$

where K_ι is given in (5.29).

Proposition 5.8 below derives boundedness of $f(x)$ from "first principles" without drawing on the result $f(x) = Ke^{-\gamma x}$, $x > 0$. We include it for ideas that may be useful to obtain bounds on the pdf of wait in variants of G/M/1 (or random variables in other models), from basic LC considerations.

Proposition 5.8 In G/M/1 with the foregoing assumptions

$$f(x) < \mu\,(1 - F(x)) < \mu, x > 0. \tag{5.68}$$

Proof We present two proofs for perspective.

Proof 1. An alternative form of the LC integral equation for G/M/1 (formula (5.8)) is

$$f(x) = \mu(1 - F(x)) - \mu\int_{y=x}^{\infty} A(y-x)f(y)dy, x > 0. \tag{5.69}$$

Since $A(0^+) = 0$, there exists $\varepsilon > 0$ such that $A(z) < 1$ for $z \in (0, \varepsilon)$. Thus

$$\begin{aligned}\int_{y=x}^{\infty} A(y-x)f(y)dy &< \int_{y=x}^{x+\varepsilon} 1.f(y)dy + \int_{y=x+\varepsilon}^{\infty} A(y-x)f(y)dy\\ &\le \int_{y=x}^{x+\varepsilon} 1.f(y)dy + \int_{y=x+\varepsilon}^{\infty} 1.f(y)dy\\ &= \int_{y=x}^{\infty} f(y)dy = 1 - F(x).\end{aligned}$$

The subtracted term in (5.69) is such that

$$0 < \mu\int_{y=x}^{\infty} A(y-x)f(y)dy < \mu\,(1 - F(x))\,, x > 0,$$

so that (5.68) follows.

Proof 2. Consider a sample path of $\{V(t)\}$ (see formula (5.1) and Fig. 5.1). From formula (5.9) the hazard rate is $f(x)/(1-F(x)) = 1/E(\mathcal{B})$. Also, $E(\mathcal{B}) > 1/\mu$ since $E(\textit{number served in } \mathcal{B}) := E(N_B^\sigma) = E(\mathcal{B})/E(S) = \mu E(\mathcal{B}) > 1$ (formula (5.48)). (The inequality holds since $\mathcal{B}$ consists of *at least one* service time.) Thus $1/E(\mathcal{B}) < \mu$ and

$$f(x) = (1-F(x))\frac{1}{E(\mathcal{B})} < \mu(1-F(x)) < \mu. \tag{5.70}$$

■

Example 5.2 **M$_\lambda$/M$_\mu$/1** is a special case of G/M//1 in which $\rho = \lambda/\mu < 1$ for stability, with $F(x) = 1 - \rho e^{-(\mu-\lambda)x}$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$, and $E(\mathcal{B}) = 1/(\mu-\lambda)$. Using (5.70),

$$f(x) = (1-F(x))(\mu-\lambda) = \mu\left(1-\frac{\lambda}{\mu}\right)(1-F(x)) = \mu P_0(1-F(x)),$$

which satisfies (5.68).

5.2 G/M/1: Zero-Waits Receive Special Service

Due to ubiquitous applications of queues where zero-wait customers receive exceptional service, we now derive the pertinent limiting pdf $\{P_0, f(x)\}_{x>0}$ in G/M/1 by using the method of pages (see, e.g., Fig. 4.2 in Sect. 4.5.5). (Sect. 3.6.1 above, derives $\{P_0, f(x)\}_{x>0}$ in M/G/1 where zero-waits receive special service.)

Let $\{V(t), M(t)\}_{t\geq 0}$ denote the extended age process, where $V(t) :=$ *system time of a customer in service at t*; $M(t) = 0$ if a zero-wait customer (initiator of a busy period) is in service or the system is empty, at time t; $M(t) = 1$ if a positive-wait customer is in service at time t. Let $S_0 :=$ *service time of zero-waits* and $S_1 :=$ *service time of positive-waits*, where $S_i \underset{dis}{=} \text{Exp}_{\mu_i}$, $i = 0, 1$. Assume the arrival stream is a renewal process with inter-arrivals having cdf $A(\cdot)$, $\overline{A}(\cdot) = 1 - A(\cdot)$, and mean $1/\boldsymbol{a}$. (In Sect. 5.2.2 below we assume the arrival process is Poisson with rate λ, and get the same limiting pdf of wait as for M/M/1 in Sect. 3.6.2.) Let $f_i(x)$, $x > 0$, $i = 0, 1$, denote the limiting pdf of $V(t)$ as $t \to \infty$ when a zero-wait and a nonzero-wait is in service, respectively. Let $h(x)$, $x < 0$, denote the pdf of ($-$ *remaining time until the next arrival of a* ***zero-wait customer***). The domains (and ranges) of f_0, h and f_1 are respectively $\{(0,\infty), 0\}$, $\{(-\infty, 0), 0\}$, and $\{(0,\infty), 1\}$, in which the right-most symbols 0, 1 denote pages. Let $P_0 := P(\text{an arrival waits } 0)$.

5.2.1 Integral Equations for $f_0(x)$, $f_1(x)$ and $h(x)$

Figure 5.4 illustrates a sample path of $\{V(t), M(t)\}_{t\geq 0}$, which aids in deriving the integral equations.

$\boldsymbol{f_0(x)}$ Consider state set $\{(x,\infty),0\}_{x\geq 0}$. The rate ***in*** is $\lim_{t\to\infty}\mathcal{U}_t(x)/t = f_0(x)$; the rate ***out*** is $\mu_0\int_x^\infty f_0(y)dy$. Since the ends of services of zero-waits initiate downward jumps ($\underset{dis}{=}$ Exp$_\lambda$) ending in $\{(-\infty,0)\,,0\}$ or $\{(0,\infty)\,,1\}$, rate balance gives

$$f_0(x) = \mu_0\int_x^\infty f_0(y)d\dot{y}, x > 0. \tag{5.71}$$

Converting to a differential equation and solving, gives the solution

$$f_0(x) = \alpha e^{-\mu_0 x}, x > 0, \tag{5.72}$$

where α is a constant to be determined.

$\boldsymbol{f_1(x)}$ Consider state set $\{(x,\infty),1\}_{x\geq 0}$. The balance equation is

$$f_1(x) + \mu_0\int_x^\infty A(y-x)f_0(y)dy = \mu_1\int_x^\infty \overline{A}(y-x)f_1(y)dy,$$

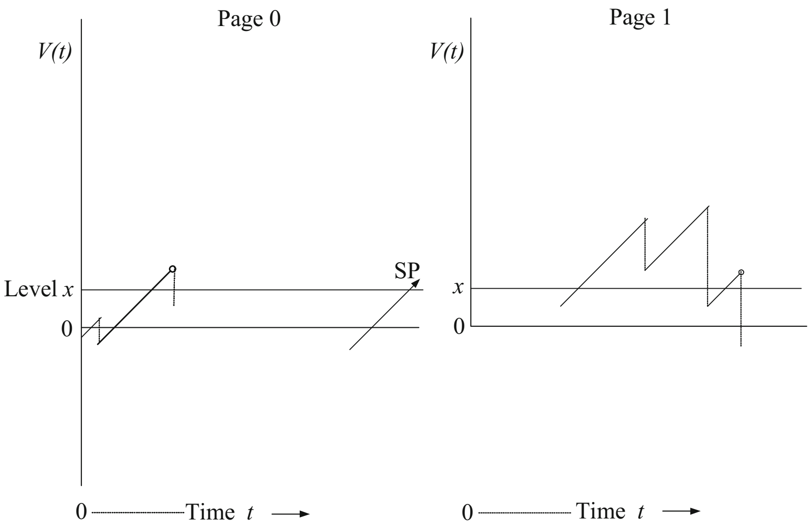


Fig. 5.4 Sample path of $\{V(t), M(t)\}_{t\geq 0}$ in G/M/1 where zero-waits get special service. Pages have same time axis. Page 0 overlays page 1

where the left side is the entrance rate due to upcrossing level x on page 1 or leaving set $\{(x, \infty), 0\}$, and the right side is the exit rate due to jumps ending below level x in $\{(0, x), 1\}$ or in $\{(-\infty, 0), 0\}$. Transposing terms and substituting from (5.72) yields

$$f_1(x) - \mu_1 \int_x^\infty \overline{A}(y-x) f_1(y) dy = -\mu_0 \alpha \int_x^\infty A(y-x) e^{-\mu_0 y} dy, x > 0, \tag{5.73}$$

***h*(*x*)** Consider state set $\{(-\infty, x), 0\}_{-\infty < x \leq 0}$. The balance equation is

$$h(x) = \mu_0 \int_0^\infty \overline{A}(y-x) \alpha e^{-\mu_0 y} dy + \mu_1 \int_0^\infty \overline{A}(y-x) f_1(y) dy, x < 0, \tag{5.74}$$

where the left side is the *exit* rate ($\lim_{t \to \infty} \mathcal{U}_t(x)/t$) and the right side is the *entrance* rate (the first term from $\{(0, \infty), 0\}$, the second term from $\{(0, \infty), 1\}$).

Also

$$P_0 = \int_{-\infty}^0 h(x) dx. \tag{5.75}$$

The normalizing condition is

$$P_0 + \int_{x=0}^\infty f_0(x) dx + \int_{x=0}^\infty f_1(x) dx = 1 \tag{5.76}$$

We shall not solve equations (5.72)–(5.76) explicitly here for a general $A(\cdot)$. Section 5.2.2 gives a full solution when the arrival rate is Poisson at rate λ, which serves as an example of the method of pages for single-server queues, and a check on the model in this Section.

5.2.2 *M/M/1 as Special Case of G/M/1*

We now derive the pdf of wait in the corresponding M/M/1, as a special case of G/M/1 where zero-waits get exceptional service, by letting $G \underset{dis}{=} \text{Exp}_\lambda$. Assume $A(y) = 1 - e^{-\lambda y}$, $y > 0$, $\overline{A}(y) = e^{-\lambda y}$, $y \geq 0$. Observe that the upcrossing rate of level 0 is equal to $h(0) = f_0(0) =$ *arrival rate of busy-period initiators (zero-waits)*. Therefore, $f_0(0) = \lambda P_0 = \alpha$. Substituting in Eq. (5.73) yields

$$f_1(x) - \mu_1 \int_x^\infty e^{-\lambda(y-x)} f_1(y) dy = -\mu_0 \lambda P_0 \int_x^\infty \left[1 - e^{-\lambda(y-x)}\right] e^{-\mu_0 y} dy,$$

or

$$f_1(x) - \mu_1 \int_x^\infty e^{-\lambda(y-x)} f_1(y)dy = -\frac{\lambda^2 P_0}{\lambda + \mu_0} e^{-\mu_0 x}, x > 0,$$

Applying $\langle D - \lambda \rangle$ to both sides results in the differential equation

$$f_1'(x) + (\mu_1 - \lambda) f_1(x) = \lambda^2 P_0 e^{-\mu_0 x}.$$

Using the condition that $f_1(0) = 0$ (because the upcrossing rate of level 0 on page 1, is 0), the solution of the differential equation is

$$f_1(x) = \frac{\lambda^2 P_0}{\mu_1 - \lambda - \mu_0} e^{-\mu_0 x} - \frac{\lambda^2 P_0}{\mu_1 - \lambda - \mu_0} e^{-(\mu_1 - \lambda)x},$$

assuming $\mu_0 \neq \mu_1 - \lambda$.

The total pdf of wait is

$$\begin{aligned} f(x) &= f_0(x) + f_1(x) \\ &= \frac{\lambda(\mu_1 - \mu_0) P_0}{\mu_1 - \lambda - \mu_0} e^{-\mu_0 x} - \frac{\lambda^2 P_0}{\mu_1 - \lambda - \mu_0} e^{-(\mu_1 - \lambda)x}. \end{aligned}$$

Equation (5.76) gives

$$P_0 \left[1 + \frac{\lambda}{\mu_1 - \lambda - \mu_0} \left(\frac{\mu_1 - \mu_0}{\mu_0} - \frac{\lambda}{\mu_1 - \lambda} \right) \right] = 1,$$

which leads to

$$P_0 = \frac{1 - \rho_1}{1 - \rho_1 + \rho_0},$$

where $\rho_i = \lambda / \mu_i$, the same as in Sect. 3.6.2 for M/M/1, and in the open literature.

5.3 Multiple-Server G/M/c Queue

The G/M/c queue (c = 2, 3, ...) generalizes G/M/1 of Sect. 5.1 to multiple parallel servers. The same symbols as for G/M/1 in Sect. 5.1 denote the inter-arrival time cdf $A(\cdot)$, pdf $a(\cdot)$, complementary cdf $\overline{A}(\cdot)$, and mean $1/\mathbf{a}$. Each customer has service time $S \underset{dis}{=} \mathrm{Exp}_\mu$, independent of any other customer's service time.

This section uses LC to derive the steady-state pdfs of the *virtual wait* and *actual wait (arrival-point wait)* in the *standard* G/M/c queue. We obtain exact formulas for the pdfs in G/M/2, and check them with the pdfs in M/M/2, to mildly validate the LC approach. Also, we derive related quantities in G/M/c.

5.3.1 Extended Age Process for G/M/c

To analyze G/M/c ($c \geq 2$), we employ the stochastic process

$$\{V(t), M(t)\}_{t \geq 0}, -\infty < V(t) < \infty, M(t) \in \boldsymbol{M}.$$

The random variable $V(t)$ is the 'extended age' at time t. When $c \geq 2$, $V(t)$ is a generalization of $V(t)$ defined for G/M/1 in Sect. 5.1.1. (Age at time t is defined immediately after the two following remarks.)

$M(t)$ is the 'system configuration'. For the *standard* G/M/c, $M(t)$ is defined more simply than for the generalized M/M/c queue, given in Sects. 4.4 and 4.5). Here, let $M(t) :=$ *number of customers* ***in service*** *at time* t. Thus $M(t) \in \boldsymbol{M} = \{0, 1, 2, ..., c\}$. If $M(t) = c$ then *at least* c customers are *in the system* at time t.

The state space of $\{V(t), M(t)\}_{t \geq 0}$ is $\boldsymbol{S} = \mathbb{R} \times \boldsymbol{M}$ where $\mathbb{R} := (-\infty, +\infty)$. $\{V(t), M(t)\}_{t \geq 0}$ is a type of '*system point process*' (see Sect. 4.5). The state is two-dimensional; $V(t)$ is continuous, $M(t)$ is discrete.

Remark 5.9 The definition of system configuration $M(t)$ is **flexible,** i.e., it may depend on the current task. (See [45, 46] for concepts of flexibility and adaptivity.) We utilize an appropriate 'configuration' that expedites the analysis. We could define $M(t)$ for G/M/c analogously as for the generalized M/M/c in Sect. 4.4. However, our simpler definition of $M(t)$ here is sufficient to study the **standard** G/M/c model. If the objective were to analyze a 'generalized' G/M/c queue, we might utilize $M(t)$ along the lines of Sect. 4.4. This flexibility applies well to state-dependent models with, e.g., service time depending on wait; service time depending on the types of other customers in service at start of service times; service rate selected at random from a set of possible service rates; etc.

Remark 5.10 For G/M/c, the definition of $M(t) \in \{0, 1, ..., c-1, c\}$ is a variant of the definition of $M(t)$ given in Sect. 4.4 for M/M/c. For G/M/c, $M(t)$ is the number of occupied servers "seen" **at an arrival instant** t. Thus $M(t)$ includes a 'sheet c' which denotes "all servers are occupied" **at the arrival instant t**. On the other hand, in M/M/c, Chap. 4, sheet $c-1$' corresponds to "$c-1$ servers are occupied **just before a start of service in the remaining 'target' unoccupied server"**. In Sect. 4.4 $M(t)$ **"looks ahead" to the instant that an arrival starts service at** $t + \mathcal{S}$. In G/M/c, $M(t)$ focuses on the arrival

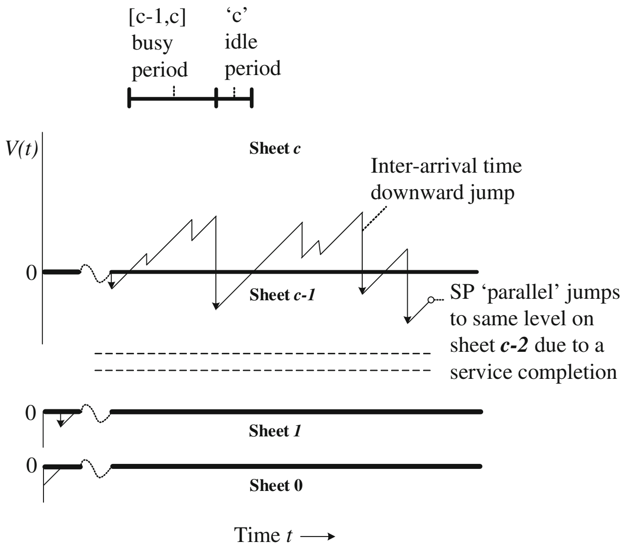


Fig. 5.5 Sample path of $\{V(t), M(t)\}_{t\geq 0}$ for G/M/c queue. There are $c+1$ sheets. Range of sheet c is $[0, \infty)$. Range of each sheet 0,..., $c-1$ is $(-\infty, 0)$. Sheet $c-1$ abuts on the bottom of sheet c for geometric convenience. Time between jumps originating on sheet $c \underset{dis}{=} \text{Exp}_{c\mu}$

time t (does not look ahead). In G/M/c sheet $c-1$ corresponds to "$c-1$ servers are occupied at an arrival instant t" (Fig. 5.5).

Extended Age and inter start-of-service depart Time in Standard G/M/c

If $M(t) = c$ then $M(t)$:= *"age" (time already spent in the system) of the* ***last*** *customer to start service at time* $\leq t$, so that $V(t) > 0$ (Fig. 5.5).

Let $\mathcal{S}$ denote the time from the instant a customer starts service until the first departure from the system thereafter, i.e., $\mathcal{S}$ is the ***inter start-of-service depart time***. Then $\mathcal{S} = \min\{S_1, ..., S_c\}$ where the S_is are mutually independent and each $S_i \underset{dis}{=} \text{Exp}_{\mu}$. One S_i is a full service time; $c-1$ S_is are *remaining* service times. Hence $\mathcal{S} \underset{dis}{=} \text{Exp}_{c\mu}$.

Relationship Between $V(t)$ and $M(t)$

When $M(t) \in \{0, 1, ..., c-1\}$, the random variable '$-V(t)$' denotes the ***remaining inter-arrival time*** required until the next arrival joins the system (Fig. 5.5). Thus

$$\text{if} \begin{cases} M(t) = c \text{ then } V(t) \geq 0; \\ M(t) \in \{0, 1, 2, ..., c-1\} \text{ then } V(t) < 0. \end{cases}$$

5.3.2 Steady-State PDF of Virtual Wait

Let $\boldsymbol{T} = \{t \in [0, \infty)\}$ be the time axis. A sample path of $\{V(t), M(t)\}_{t \geq 0}$ is given in Fig. 5.5. The *rate* at which the SP moves in $\boldsymbol{T} \times \boldsymbol{S}$ between downward jumps when $M(t) = c$, is $dV(t)/dt = +1$, $0 < V(t) < \infty$, $t > 0$.

The steady-state pdf of $V(t)$ as $t \to \infty$, is the same as that of the *virtual wait* $W(t)$ as $t \to \infty$ (proved similarly as in Proposition 5.1 for G/M/1).

Denote the steady-state cdf of the virtual wait by $F(x)$, $x \geq 0$, having pdf $f(x) = dF(x)/dx$, $x > 0$, wherever the derivative exists. $F(0)$ is the *proportion of time* with fewer than c customers in service, i.e., the probability that the system presents a zero wait to a potential arrival. Let P_i be the proportion of time that an arrival "sees" i customers in service, $i = 0, \ldots, c-1$. The P_is are *zero-wait probabilities*, and $F(0) = \sum_{i=0}^{c-1} P_i$.

Integral Equation for PDF of Wait

Consider a sample path of $\{V(t), M(t)\}_{t \geq 0}$ (Fig. 5.5). $\boldsymbol{T} \times \boldsymbol{S}$ is partitioned into $c+1$ sheets (or 'pages'), which are planar subsets of $\boldsymbol{T} \times \boldsymbol{S}$. Sheets $0, \ldots, c-1$ can be thought of as being one behind the other like pages in a book, *below* level 0 (the time axis). Only sheet c is *above* level 0. Sheet c is pictured as being directly above, and contiguous to, sheet $c-1$.

Case $M(t) = c$ When $M(t) = c$ the SP is on sheet c, moving upward with slope $+1$. Fix level $x > 0$. The SP *upcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} \underset{a.s}{=} \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = f(x)$$

(proved similarly as in Theorem 1.1 for the downcrossing rate in M/G/1; see also Sect. 2.1 in [19]). All servers are occupied just before a departure, the inter start-of-service depart time is $\mathcal{S} \underset{dis}{=} \text{Exp}_{c\mu}$, implying the departure rate is $c\mu$. The SP *downcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \underset{a.s}{=} \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = c\mu \int_{y=x}^{\infty} \overline{A}(y-x) f(y) dy. \tag{5.77}$$

In (5.77), $c\mu$ is the customer departure rate. Each departure generates a downward jump with size $\underset{dis}{=} A(\cdot)$ (*inter-arrival* time cdf). $\overline{A}(y-x)$ is $P(\text{jump} > y-x | \text{jump starts at level } y) = P(\text{jump downcrosses level } x | \text{jump starts at level } y)$.

Rate balance across level x, i.e., $\lim_{t\to\infty} E(\mathcal{U}_t(x))/t = \lim_{t\to\infty} E(\mathcal{D}_t(x))/t$, gives a basic LC integral equation for G/M/c

$$f(x) = c\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy, x > 0. \tag{5.78}$$

Equation (5.78) for G/M/c, where the downward jump rate is $c\mu$, is a straightforward generalization of the LC integral equation for G/M/1, where the downward jump rate is μ (formula 5.7).

Alternative Form of Integral Equation

An alternative form of (5.78) is

$$f(x) = c\mu\,(1 - F(x)) - c\mu \int_{y=x}^{\infty} A(y-x)f(y)dy, x > 0. \tag{5.79}$$

Explanation of (5.79) On the left side, $f(x)$ is the *upcrossing rate* of level x. On the right side, $c\mu\,(1 - F(x))$ is the rate at which all downward jumps start in state-space set $\{((x, \infty), c)\}$. The subtracted term including the integral, subtracts the rate at which downward jumps start in $\{((x, \infty), c)\}$ *and end in* $\{((x, \infty), c)\}$; such jumps *do not downcross level* x. Thus the right side of (5.79) is the *downcrossing rate* of level x.

5.3.3 Form of PDF of Wait in G/M/c Geometrically

Let $\mathcal{B}_{c-1,c}$ denote a $[c-1, c]$ busy period , the time from the instant the number of customers in service increases from $c-1$ to c until the first instant thereafter when the number of customers in service decreases back to $c-1$ (Fig. 5.5). During $\mathcal{B}_{c-1,c}$ the number of customers in the system is $\geq c$. $\mathcal{B}_{c-1,c}$ is a sojourn time on sheet c , which starts with an SP upcrossing of level 0 (via top of sheet $c-1$), and ends with a downcrossing of level 0 that terminates on sheet $c-1$. Let a_x denote a sojourn time above level $x \geq 0$ starting with an upcrossing of level x (on sheet c). Then $\mathcal{B}_{c-1,c}= a_0$, and $E\left(\mathcal{B}_{c-1,c}\right) = E\,(a_0)$.

The memoryless property of $\mathcal{S}$ $(\underset{dis}{=} \text{Exp}_{c\mu})$ implies the excesses $\mathcal{S}$ above level $x \underset{dis}{=} \text{Exp}_{c\mu}$ so that $E(a_x) = E(\mathcal{B}_{c-1,c})$, $x \geq 0$, *independent of* x. By LC and the elementary renewal theorem, $E(u_x) = 1/f(x)$, where $u_x :=$ *inter-upcrossing time of level* x. The sequence of u_xs forms a renewal process because each upcrossing of level $x > 0$ has an excess over level $x \underset{dis}{=} \text{Exp}_{c\mu}$, implying the probabilistic future evolution is independent of the past. By the renewal reward theorem

$$\frac{E(a_x)}{E(u_x)} = \frac{E(a_x)}{1/f(x)} = 1 - F(x)$$

$$E(a_x) = E(\mathcal{B}_{c-1,c}) = \frac{1 - F(x)}{f(x)}$$

$$\frac{f(x)}{1 - F(x)} = \frac{1}{E(\mathcal{B}_{c-1,c})}, x \geq 0$$

Similarly as for G/M/1 in Sect. 5.1.5,

$$\frac{d}{dx} \ln (1 - F(x)) = -\frac{1}{E(\mathcal{B}_{c-1,c})} ,$$

a differential equation whose solution for the cdf of wait is

$$F(x) = 1 - (1 - F(0)) \cdot e^{-\frac{1}{E(\mathcal{B}_{c-1,c})}x}, x \geq 0. \tag{5.80}$$

Taking $dF(x)/dx$ in (5.80) gives the pdf of wait as

$$f(x) = \frac{1 - F(0)}{E(\mathcal{B}_{c-1,c})} \cdot e^{-\frac{1}{E(\mathcal{B}_{c-1,c})}x}, x \geq 0. \tag{5.81}$$

Hence

$$f(x) = Ke^{-\gamma x}, x > 0, \tag{5.82}$$

where

$$K = \frac{1 - F(0)}{E(\mathcal{B}_{c-1,c})}, \qquad \gamma = \frac{1}{E(\mathcal{B}_{c-1,c})}. \tag{5.83}$$

From (5.83)

$$E(\mathcal{B}_{c-1,c}) = \frac{1}{\gamma}. \tag{5.84}$$

Substituting $f(x)$ from (5.82) into (5.78) gives a transcendental equation for γ,

$$\int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy = \frac{1}{c\mu}. \tag{5.85}$$

The Laplace-Stieltjes transform of the inter-arrival distribution evaluated at γ, is $A^*(\gamma) = \int_{y=0}^{\infty} a(y)e^{-\gamma y}dy$. Integration by parts in (5.85) gives an alternative equation for γ,

$$A^*(\gamma) = 1 - \frac{\gamma}{c\mu}. \tag{5.86}$$

To specify the mixed pdf of wait $\{F(0), f(x)\}_{x>0}$, it is required to solve for $F(0)$ in (5.81) or equivalently for K in (5.82). From (5.83) we obtain

$$F(0) = 1 - \frac{K}{\gamma}. \tag{5.87}$$

From the form of $f(x)$, $x > 0$, as $Ke^{-\gamma x,}$ we can also obtain (5.87) from the normalizing condition

$$\begin{aligned} F(0) + \int_{x=0}^{\infty} f(x)dx &= 1, \\ F(0) + \int_{x=0}^{\infty} Ke^{-\gamma x}dx &= 1, \\ F(0) + \frac{K}{\gamma} &= 1. \end{aligned}$$

Remark 5.11 Another way to obtain (5.87) is directly from the sample path of $\{V(t)\}_{t\geq 0}$ and SP motion. We include this derivation because it highlights the close relationship between probabilities of the model and SP motion. $F(0)$ is the **proportion** of time that the system presents a zero wait. The expected time between successive SP upcrossings of level 0 due to arrivals that "see" $c-1$ customers in service, is $1/f(0)$ (starts of $\mathcal{B}_{c-1,c}$s). Since $f(x) = Ke^{-\gamma x}$,

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t} = f(0) = K.$$

After the SP moves onto sheet c, it eventually leaves sheet c when a departure propels a downward jump onto sheet $c-1$. The SP then sojourns among some or all sheets 0,..., $c-1$. During this SP sojourn, any arrival to the system would wait zero. The sojourn below level 0 continues until the SP next upcrosses level 0 from sheet $c-1$ onto sheet c. Thus upcrossings of level 0 from sheet $c-1$ are regenerative points. From the theory of regenerative processes (renewal reward theorem)

$$\begin{aligned} F(0) &= \frac{E(\text{sojourn time among sheets } 0, ..., c-1)}{E(\text{time between SP entrances to sheet } c)} \\ &= \frac{\frac{1}{K} - E(\mathcal{B}_{c-1,c})}{\frac{1}{K}} = \frac{\frac{1}{K} - \frac{1}{\gamma}}{\frac{1}{K}} = 1 - \frac{K}{\gamma}. \end{aligned} \tag{5.88}$$

Explanation of (5.88). In the numerator $1/K = 1/f(0) = E($time between SP upcrossings from sheet $c-1$ onto sheet $c)$. These upcrossings are regenerative points. The term '$-E(\mathcal{B}_{c-1,c})$' subtracts off the expected embedded busy period $E\left(B_{c-1,c}\right) = 1/\gamma$, thus leaving E(sojourn time among sheets $0, ..., c-1$).

Value of K

We must solve for the value of K in order to specify $F(0)$ and $f(x)$, $x > 0$, in terms of the model parameters. This requires a further analysis of SP motion on sheets $0, ..., c-1$.

Remark 5.12 Applying the normalizing condition

$$F(0) + \int_{x=0}^{\infty} f(x)dx = 1,$$

and using (5.81), does not give $F(0)$ in terms of the model parameters, since it yields the tautology 1 = 1. Section 5.3.4 below develops integral equations for the steady-state partial pdfs of $V(t)$ **on sheets** $0, ..., c-1$, which leads to an independent expression for $F(0)$; the normalizing condition then gives $F(0)$. We shall not solve for $F(0)$ explicitly in the general G/M/c queue here. However, we will solve for $F(0)$ explicitly in G/M/2 in Sect. 5.4, to indicate the solution procedure.

5.3.4 Partial PDFs of Extended Age for Sheets **0** *to* $c - 1$

Let $g_i(x)$, $x < 0$, denote the steady-state pdf of $\{V(t), M(t) = i\}_{t\to\infty}$, $i = 0, ..., c-1$. In Fig. 5.5 the partial pdfs $\{g_i(x)\}_{x<0}$ are the partial pdfs when $M(t) = i \in 0,..., c-1$. We derive integral equations for $g_i(x)$, $i = 0,..., c-1$, by applying rate balance to SP *exits* and *entrances* of state-space intervals $\{(-\infty, x), i\}_{x<0}$ on sheets $i = 0, ..., c-1$.

The probability $F(0)$ is the proportion of time that potential arrivals wait 0 for service. Thus

$$F(0) = \sum_{i=0}^{c-1} \int_{x=-\infty}^{0} g_i(x)dx = \sum_{i=0}^{c-1} P_i \tag{5.89}$$

where $P_i := \int_{x=-\infty}^{0} g_i(x)dx$ is the steady-state probability of i customers in service.

Integral Equation for PDF $g_{c-1}(x)$**: Sheet** $c-1$

First consider interval $\{(-\infty, x), c-1\}$ for fixed $x < 0$, on sheet $c-1$.

Exit Rate The SP *exit* rate from $\{(-\infty, x), c-1\}$ is

$$g_{c-1}(x) + (c-1)\,\mu \int_{y=-\infty}^{x} g_{c-1}(y)dy. \tag{5.90}$$

Explanation of (5.90) The first term is the SP (continuous) *upcrossing* rate of level x. The second term is the rate at which customers *depart* the system when $c-1$ servers are occupied *and* the *remaining time* until the *next* arrival to the system is $-y$, summed over all $y \in (-\infty, x)$. Departures occur at rate $(c-1)\,\mu$ since there are $c-1$ customers in service; and service times are independent of the remaining time until the next arrival. Such customer departures generate SP ***parallel*** jumps from sheet $c-1$ to sheet $c-2$ starting and ending ***at the same level***, because just after these departures there would be $c-2$ units in service, and the remaining inter-arrival time would still be the same as it was just before each departure.

Entrance Rate The SP *entrance* rate into $\{(-\infty, x), c-1\}$ is

$$c\mu \int_{y=0}^{\infty} \overline{A}(y-x)f(y)dy + g_{c-2}(0) \cdot \overline{A}(-x). \tag{5.91}$$

Explanation of (5.91) The first term is the rate at which the SP jumps downward from level $y > 0$ on sheet c into interval $\{(-\infty, x), c-1\}$, due to customer departures that leave $c-1$ units in service. An inter-arrival time which is $> y-x$ causes the SP to downcross level x on sheet $c-1$ (probability is $\overline{A}(y-x)$). In the second term, $g_{c-2}(0)$ is the SP hit rate of level 0 from below on page $c-2$ (upcrossing rate of level 0), i.e., the *arrival* rate to the system when $c-2$ servers are occupied. Such arrivals *increase* the number of occupied servers to $c-1$. $\overline{A}(-x)$ is the probability that the *immediately following inter-arrival time* exceeds $-x$, thereby propelling the SP below level x and into $\{(-\infty, x), c-1\}$.

Equating (5.90) and (5.91) gives the integral equation for $g_{c-1}(x)$,

$$\begin{aligned} &g_{c-1}(x) + (c-1)\,\mu \int_{y=-\infty}^{x} g_{c-1}(y)dy \\ &= c\mu \int_{y=0}^{\infty} \overline{A}(y-x)f(y)dy + g_{c-2}(0)\overline{A}(-x), x < 0. \end{aligned} \tag{5.92}$$

Integral Equations for PDF: Sheets $1, \ldots, c-2$

Consider the state-space interval $\{(-\infty, x), i\}, x < 0$ on sheet i where $i \in \{1, \ldots, c-2\}$ (Fig. 5.5). Reasoning as in the derivation of (5.92) for sheet $c-1$, we obtain integral equations

$$\begin{aligned} g_i(x) &+ i\mu \int_{y=-\infty}^{x} g_i(y)dy \\ &= (i+1)\mu \int_{y=-\infty}^{x} g_{i+1}(y)dy + g_{i-1}(0)\overline{A}(-x), \\ &\qquad\qquad i = 1, ..., c-2, x < 0. \end{aligned} \tag{5.93}$$

Explanation of (5.93) The left side is the SP ***exit rate*** from interval $\{(-\infty, x), i\}$, composed of upcrossings of level x *plus* the customer departure rate when there are i in service. The right side is the SP ***entrance rate*** into $\{(-\infty, x), i\}$, composed of $(i+1)\mu \int_{y=-\infty}^{x} g_{i+1}(y)dy$, the rate of ***parallel*** jumps starting in $\{(-\infty, x), i+1\}$ and ending at the same level in $\{(-\infty, x), i\}$ *plus* the rate of customer arrivals (upcrossings of level 0) when there are $i-1$ customers in service followed by inter-arrivals that exceed $-x$.

Integral Equation for PDF: Sheet 0

Consider state-space interval $\{(-\infty, x), 0\}, x < 0$.

Exit Rate The SP can exit $\{(-\infty, x), 0\}, x < 0$ only by means of a (left) continuous hit of level x from below (upcrossing of level x). The system is empty, so no customer departures can occur when $M(t)$= 0. Therefore the exit rate of $\{(-\infty, x), 0\}$ is $g_0(x)$.

Entrance Rate The SP can enter $\{(-\infty, x), 0\}$ only by a parallel jump from $\{(-\infty, x), 1\}$ on sheet 1. There must be *one* customer in service, which departs before any new arrivals to the system occur, and the remaining inter-arrival time, say y, is $> -x$, so that $y \in \{(-\infty, x), 0\}$. The rate of such parallel jumps $\mu \int_{y=-\infty}^{x} g_1(y)dy$.

Rate balance of exit and entrance rates of $\{(-\infty, x), 0\}$ gives an integral equation for sheet 0 ($M(t) = 0$),

$$g_0(x) = \mu \int_{y=-\infty}^{x} g_1(y)dy. \tag{5.94}$$

Form of $F(0)$

The probability of a potential wait of zero is given in (5.89). Here we shall not detail a procedure to compute $F(0)$ for the virtual wait in G/M/c for general values of c. Nevertheless, in Sect. 5.4.1 below, we provide a detailed derivation of $F(0)$ in G/M/2.

5.3.5 Stability Condition for G/M/c

The stability condition for G/M/c follows directly from (5.82) and (5.85). The system is stable iff the steady-state pdf in (5.82) exists iff there exists a positive finite solution γ for Eq. (5.85). Using an analysis similar to that given in Proposition 5.3 for G/M/1, we obtain a necessary and sufficient condition for stability in G/M/c, namely $\boldsymbol{a} < c\mu$.

5.3.6 Form of PDF of Actual Wait $W_{q,\iota}$

Let $W_{q,\iota}$ denote the actual wait in line before service (arrival-point wait), in steady state. Let $F_\iota(x) := P(W_{q,\iota} \leq x), x \geq 0$, and $f_\iota(x) = dF_\iota(x)/dx), x > 0$, be the pdf of $W_{q,\iota}$.

Proposition 5.9 In the G/M/c queue, the form of the pdf of $W_{q,\iota}$ is

$$f_\iota(x) = K_\iota e^{-\gamma x}, x > 0, \tag{5.95}$$

where $K_\iota > 0$.

Proof The proportion of arrivals that wait $W_{q,\iota} > x$ is

$$1 - F_\iota(x) = \frac{c\mu(1 - F(x)) - f(x)}{c\mu\,(1 - F(0)) + \sum_{i=1}^{c-2} g_i(0)}, x > 0. \tag{5.96}$$

In formula (5.96) $F(x)$ and $f(x)$ are respectively the cdf and pdf of the *virtual wait*; $F(0) := P(\textit{virtual wait} = 0)$; $g_i(0), i = 1, ..., c - 1$, are respectively the arrival rates to the system when i customers are in service. (See Proposition 5.2 for G/M/1 in Sect. 5.1.7).)

Explanation of (5.96). $c\mu(1 - F(x))$ is the rate of *downward* jumps that start at peaks $> x$ (i.e., in $\{(x, \infty), c\}$ on sheet c), and end at an ordinate equal to the next *actual wait* > 0 if the end point is > 0 (in $\{(0, \infty), c\}$, i.e., at a trough > 0); or at a trough in $\{(-\infty, 0), c - 1\}$ implying that the next actual wait is = 0, if the end point is < 0 (Fig. 5.5). The term $-f(x)$ subtracts off the rate of such downward jumps that *end below* x. (Note that $f(x) = \lim \mathcal{U}_t(x)/t = \lim \mathcal{D}_t(x)/t, x \in \{(0, \infty), c\}$.) Thus the numerator is the rate at which ‘ *next*’ $W_{q,\iota}$s (troughs on page c) are $> x$. In the denominator, $c\mu\,(1 - F(0))$ is the rate of all *downward* jumps that start on sheet c; $\sum_{i=1}^{c-2} g_i(0)$ is the rate of all *downward* jumps that start at level 0 on sheets $1, ..., c - 2$. Thus, the denominator is the total rate of all downward jumps, which is precisely the total rate at which '*next*' customers start service. Thus the right side of (5.96) is the *proportion* of downward jumps that start *above* level $x > 0$ and end *above* level x, on sheet c. This is the same as the proportion of *all* customers whose actual (arrival point) wait is $> x$.

From Eqs. (5.80) and (5.82), $1 - F(x) = \kappa e^{-\gamma x}$ where $\kappa > 0$, and $f(x) = Ke^{-\gamma x}$. Also, $c\mu\,(1 - F(0)) + \sum_{i=1}^{c-2} g_i(0)$ is > 0. Substituting these values into the right side of (5.96) and taking d/dx on both sides of (5.96), yields (5.95) for some constant $K_\iota > 0$. ■

5.3.7 *Probability That Actual Wait is Zero:* $F_\iota(0)$

The total rate at which *zero-wait* customers arrive is equal to the total rate at which the SP hits level 0 from below, namely $\sum_{i=0}^{c-1} g_i(0)$ (see definition of $g_i(\cdot)$, $i = 0, ..., c-1$ in Sect. 5.3.4). So, $g_i(0)$ is the rate at which customers arrive at the system (remaining inter-arrival time = 0), when there are i customers in service, $i = 0, ..., c-1$.

Let N_t, $N_t^{=0}$, $N_t^{>0}$ denote respectively the total number of arrivals, the number of arrivals that wait 0, and the number of arrivals that wait > 0, during $(0, t)$.

Consider a sample path of $\{V(t), M(t)\}_{t\geq 0}$ (Fig. 5.5). Let $\mathcal{U}_t^i(x)$ denote the number of SP upcrossings of level x on sheet i during $(0, t)$, $i = 0, ..., c-1$. Then

$$\lim_{t\to\infty} \frac{\mathcal{U}_t^i(x)}{t} \underset{a.s.}{=} \lim_{t\to\infty} \frac{E\left(\mathcal{U}_t^i(x)\right)}{t} = g_i(x), x \leq 0, i = 0, ..., c-1.$$

Note that $N_t^{=0} = \sum_{i=0}^{c-1} \mathcal{U}_t^i(0)$.

The *proportion* of arrivals that wait 0 is

$$\begin{aligned}
F_\iota(0) &= \lim_{t\to\infty} \frac{N_t^{=0}}{N_t} = \lim_{t\to\infty} \frac{N_t^{=0}}{N_t^{=0} + N_t^{>0}} \\
&= \frac{\lim_{t\to\infty} \frac{N_t^{=0}}{t}}{\lim_{t\to\infty} \frac{N_t^{=0}}{t} + \lim_{t\to\infty} \frac{N_t^{>0}}{t}} \\
&= \frac{\lim_{t\to\infty} \frac{\sum_{i=0}^{c-1} \mathcal{U}_t^i(0)}{t}}{\lim_{t\to\infty} \frac{\sum_{i=0}^{c-1} \mathcal{U}_t^i(0)}{t} + \lim_{t\to\infty} \frac{N_t^{>0}}{t}} \\
&= \frac{\sum_{i=0}^{c-1} g_i(0)}{\sum_{i=0}^{c-1} g_i(0) + \lim_{t\to\infty} \frac{N_t^{>0}}{t}}. \qquad (5.97)
\end{aligned}$$

In the denominator of (5.97)

$$\lim_{t\to\infty}\frac{N_t^{>0}}{t} \underset{a.s.}{=} \lim_{t\to\infty}\frac{E\left(N_t^{>0}\right)}{t} = c\mu\int_{y=0}^{\infty}A(y)f(y)dy$$
$$= c\mu\int_{y=0}^{\infty}A(y)Ke^{-\gamma y}dy = c\mu\int_{y=0}^{\infty}\left(1-\overline{A}(y)\right)Ke^{-\gamma y}dy$$
$$= \frac{c\mu}{\gamma}K - K, \tag{5.98}$$

upon utilizing (5.82) and (5.85).

Explanation of (5.98). $c\mu\int_{y=0}^{\infty}A(y)f(y)dy$ is the rate at which customers depart after being in the system for a time y, and the immediately *next* inter-arrival time is $< y$, summed over all $y > 0$; this is the rate at which *next* customers wait > 0.

Substituting from (5.98) into (5.97) gives

$$F_\iota(0) = \frac{\sum_{i=0}^{C-1} g_i(0)}{\sum_{i=0}^{C-1} g_i(0) + \frac{c\mu}{\gamma}K - K}. \tag{5.99}$$

In (5.92) let $x \uparrow 0$. The SP exit rate from sheet $c-1$ across level 0 is equal to the SP entrance rate of interval $\{(0,\infty), c\}$ (sheet c). Thus

$$\lim_{t\to\infty}\frac{E\left(\mathcal{U}_t^{c-1}(0)\right)}{t} = g_{c-1}(0) = f(0) = K.$$

Here we do not carry out the procedure to compute $F_\iota(0)$ for general values of c (Eq. (5.99)). In Sect. 5.4.2 below we derive $F_\iota(0)$ explicitly for G/M/2, to indicate the computational procedure.

5.4 G/M/2: PDF of Virtual and of Actual Wait

We now derive the steady-state pdf of the virtual wait and of the actual wait for G/M/2. Consider the process $\{V(t), M(t)\}_{t\geq 0}$. When $c = 2$, $M(t) \in \boldsymbol{M} = \{0, 1, 2\}$. Graphically, there are three corresponding sheets in $\boldsymbol{T} \times \boldsymbol{S}$ labeled 0, 1, 2 (see Fig. 5.6). The analyses below suggest the type of solution approach that may be used for $c \geq 2$. (The results for $c = 2$ are applied in [87].)

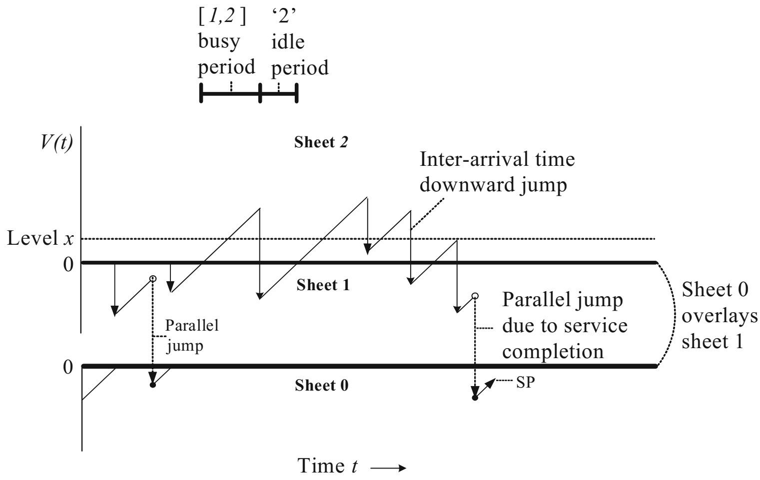


Fig. 5.6 G/M/2: sample path of $\{V(t), M(t)\}_{t \geq 0}$. Sheets 0 and 1 are quarter planes that overlay each other like pages in a book with common edges along the $V(0)$-vertical axis and the level-0 line

5.4.1 PDF of Virtual Wait

In G/M/2 the pdf of the virtual wait has the same form as in the general G/M/c model,

$$f(x) = Ke^{-\gamma x}, x > 0.$$

For $c = 2$ the integral equations for sheets 1 and 0 are respectively

$$g_1(x) + \mu \int_{y=-\infty}^{x} g_1(y)dy = 2\mu K \int_{y=0}^{\infty} \overline{A}(y - x)e^{-\gamma y}dy + g_0(0)\overline{A}(-x), x < 0, \tag{5.100}$$

$$g_0(x) = \mu \int_{y=-\infty}^{x} g_1(y)dy, x < 0, \tag{5.101}$$

as in equations (5.92) and (5.94).

Also $g_1(0) = K$. The proportion of time that the system has less than 2 occupied servers is

$$F(0) = \int_{x=-\infty}^{0} (g_1(x) + g_0(x))dx = 1 - \frac{K}{\gamma}, \tag{5.102}$$

as in (5.88).

Adding corresponding sides of (5.100) and (5.101) and integrating with respect to $x \in (-\infty, 0)$, gives

$$F(0) \equiv \int_{x=-\infty}^{0} (g_1(x) + g_0(x))dx$$
$$= 2\mu K \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + g_0(0)\frac{1}{\boldsymbol{a}}, \quad (5.103)$$

where $\frac{1}{\boldsymbol{a}} = \int_{u=0}^{\infty} \overline{A}(u)du$ is the mean inter-arrival time.

Taking d/dx in (5.101) gives the relation

$$g_1(x) = \frac{g_0'(x)}{\mu}. \quad (5.104)$$

Substituting (5.104) and (5.101) into (5.100) gives a differential equation for $g_0(x)$

$$g_0'(x) + \mu g_0(x) = 2\mu^2 K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy + +\mu g_0(0)\overline{A}(-x), x < 0, \quad (5.105)$$

whose solution is

$$g_0(x) = 2\mu^2 K e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z} \int_{y=0}^{\infty} \overline{A}(y-z)e^{-\gamma y}dydz$$
$$+ \mu g_0(0)e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z}\overline{A}(-z)dz, x < 0. \quad (5.106)$$

We obtain (5.106) because the constant of integration when solving (5.105) is 0, upon using $\lim_{x\downarrow-\infty} g_0(x) = 0$ and $\lim_{x\downarrow-\infty} \int_{z=-\infty}^{x} (\cdots)\, dz = 0$.

Since $\lim_{x\uparrow 0} e^{-\mu x} = e^0 = 1$, in (5.106) letting $x \uparrow 0$ gives an equation for $g_0(0)$ in terms of K (after making the transformation $u = -z$),

$$g_0(0) = 2\mu^2 K \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu$$
$$+\mu g_0(0) \int_{u=0}^{\infty} e^{-\mu u}\overline{A}(u)du,$$

or

$$g_0(0) = \left(\frac{2\mu^2 \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu}{1 - \mu \int_{u=0}^{\infty} e^{-\mu u}\overline{A}(u)du,} \right) K := H_0 \cdot K. \quad (5.107)$$

Equation (5.107) defines the constant H_0, which is independent of K.

We now obtain an equation for K. From (5.102) and (5.100),

$$F(0) = 1 - \frac{K}{\gamma}$$
$$= 2\mu K \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + H_0 K \frac{1}{\boldsymbol{a}}. \quad (5.108)$$

Solving (5.108) for K gives

$$K = \frac{1}{\frac{1}{\gamma} + 2\mu \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + H_0 \cdot \frac{1}{\boldsymbol{a}}}. \quad (5.109)$$

where H_0 is defined in (5.107).

Thus

$$F(0) = 1 - \frac{K}{\gamma}$$
$$= 1 - \frac{1}{1 + 2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}$$
$$= \frac{2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}{1 + 2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}$$
$$= \frac{2\mu\gamma \int_{u=0}^{\infty} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}{1 + 2\mu\gamma \int_{u=0}^{\infty} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}, \quad (5.110)$$

upon making the transformation $u = -x$.

The pdf of the *virtual wait* is $\{F(0), f(x)\}_{x>0}$, where $f(x) = Ke^{-\gamma x}$, $x > 0$ and K is specified in (5.109). The probability of a zero wait $F(0)$, is given by (5.110).

5.4.2 PDF of Actual Wait

Equation (5.95) becomes

$$f_\iota(x) = K_\iota e^{-\gamma x}, x > 0, \quad (5.111)$$

where

$$K_\iota = \frac{1 - F_\iota(0)}{E(\mathcal{B}_{1,2})}, \qquad \gamma = \frac{1}{E(\mathcal{B}_{1,2})}, \qquad F_\iota(0) = P_{0,\iota} + P_{1,\iota}.$$

From (5.99) the *proportion of arrivals* that wait 0 is

$$F_\iota(0) = \frac{\sum_{i=0}^{1} g_i(0)}{\sum_{i=0}^{1} g_i(0) + \frac{2\mu}{\gamma}K - K}. \tag{5.112}$$

Taking d/dx on both sides of (5.100) gives an ordinary differential equation for $g_1(x)$ with solution

$$\begin{aligned} e^{\mu x} g_1(x) = 2\mu \int_{z=-\infty}^{x} e^{\mu z} \int_{y=0}^{\infty} a(y-x) K e^{-\gamma y} dy dz \\ + g_0(0) \int_{z=-\infty}^{x} e^{\mu z} a(-z) dz + H_1, x < 0, \end{aligned} \tag{5.113}$$

where H_1 is a constant. Necessarily $\lim_{x\downarrow -\infty} g_1(x) = 0$, and utilizing $\lim_{x\downarrow -\infty} e^{\mu x} g_1(x) = 0$ and $\lim_{x\downarrow -\infty} \int_{z=-\infty}^{x} (\cdots) dz = 0$, leads to $H_1 = 0$.

Additionally, $\lim_{x\uparrow 0} e^{\mu x} g_1(x) = g_1(0) = f(0) = K$. Letting $x \uparrow 0$ in (5.113) yields

$$g_0(0) = K \cdot B_0, \tag{5.114}$$

where

$$B_0 = \frac{1 - 2\mu \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} a(y+u) e^{-\gamma y} dy du}{\int_{u=0}^{\infty} e^{-\mu u} a(u) du}, \tag{5.115}$$

using the transformation $u = -z$.

Thus, with B_0 given in (5.115)

$$g_1(0) + g_0(0) = K + KB_0.$$

From (5.112)

$$\begin{aligned} F_\iota(0) &= \frac{\sum_{i=0}^{1} g_i(0)}{\sum_{i=0}^{1} g_i(0) + \frac{c\mu}{\gamma}K - K} \\ &= \frac{K + KB_0}{K + KB_0 + \frac{2\mu}{\gamma}K - K} = \frac{1 + B_0}{B_0 + \frac{2\mu}{\gamma}}, \end{aligned} \tag{5.116}$$

which is independent of K.

We then calculate K_ι from the normalizing condition

$$F_\iota(0) + \int_{x=0}^{\infty} f_\iota(x)dx = 1,$$
$$F_\iota(0) + \int_{x=0}^{\infty} K_\iota e^{-\gamma x}dx = 1.$$

Applying (5.116) gives

$$\frac{1+B_0}{B_0 + \frac{2\mu}{\gamma}} + \frac{K_\iota}{\gamma} = 1,$$

which yields

$$K_\iota = \gamma\left(\frac{2\mu-\gamma}{2\mu+\gamma B_0}\right) = \gamma\left(1 - F_\iota(0)\right). \tag{5.117}$$

From (5.95) and (5.117)

$$f_\iota(x) = K_\iota e^{-\gamma x} = \gamma\left(1 - F_\iota(0)\right)e^{-\gamma x}, x > 0,$$

and

$$F_\iota(0) = 1 - \frac{K_\iota}{\gamma} = 1 - \left(\frac{2\mu-\gamma}{2\mu+\gamma B_0}\right) = \frac{\gamma\left(1+B_0\right)}{2\mu+\gamma B_0}. \tag{5.118}$$

5.4.3 Reduction of G/M/2 PDF to M/M/2 PDF

To enhance intuition, we check that the G/M/2 pdf for the actual wait, given above, reduces to the M/M/c pdf given in (4.56), (4.57) and (4.58) when c = 2. In M/M/2 let P_0, P_1 be the steady-state probabilities of 0 units and 1 unit in the system, respectively. For M/M/2 the pdfs of the virtual wait and actual wait are the same, due to Poisson arrivals. We now show that for G/M/2 with Poisson arrivals (e.g., G $\underset{dis}{=}$ Exp_λ), $F_\iota(0) = P_0 + P_1$.

In $\text{M}_\lambda/\text{M}_\mu/2$, the standard formulas for P_0, P_1 and $f(x)$ are

$$\begin{aligned} P_0 &= \frac{1}{1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu(2\mu-\lambda)}} \\ P_1 &= \frac{\lambda}{\mu}P_0 \\ f(x) &= \lambda P_1 e^{-(2\mu-\lambda)x}, x > 0. \end{aligned} \tag{5.119}$$

In M/M/2, (5.119) gives

$$P_0 + P_1 = \frac{(2\mu-\lambda)(\lambda+\mu)}{\lambda\mu + 2\mu^2}. \tag{5.120}$$

To obtain these values from G/M/2 when $G \underset{dis}{=} Exp_{\lambda}$, we first specialize the G/M/2 formula for B_0 in (5.115) to M/M/2, by letting $a(z) = \lambda e^{-\lambda z}$, $z > 0$, and set $\gamma = 2\mu - \lambda$. This substitution yields $B_0 = \mu/\lambda$. Combining with (5.118) gives

$$F_{\iota}(0) = \frac{(2\mu - \lambda)(\lambda + \mu)}{\lambda\mu + 2\mu^2} \tag{5.121}$$

in agreement with (5.120).

The pdf is

$$\begin{aligned} f_{\iota}(x) &= K_{\iota} e^{-\gamma x} = \gamma (1 - F_{\iota}(0)) e^{-(2\mu - \lambda)x} \\ &= \lambda P_1 e^{-(2\mu-\lambda)x}, x > 0, \end{aligned} \tag{5.122}$$

since $\gamma = 2\mu - \lambda$, and from (5.119),

$$\begin{aligned} \gamma (1 - F_{\iota}(0)) &= (2\mu - \lambda) \frac{(2\mu - \lambda)(\lambda + \mu)}{\lambda\mu + 2\mu^2} \\ &= \lambda \frac{\lambda (2\mu - \lambda)}{\lambda\mu + 2\mu^2} = \lambda P_1. \end{aligned}$$

Hence the G/M/2 pdf $\{F_{\iota}(0); f_{\iota}(x), x > 0\}$ in (5.121) and (5.122) when $G \underset{dis}{=} Exp_{\lambda}$, agrees with the $M_{\lambda}/M_{\mu}/2$ pdf.

5.4.4 Moments of Actual Wait for G/M/2

All statistical moments (about 0) of $W_{q,\iota}$ can be found using

$$E(W_{q,\iota}^n) = \int_{y=0}^{\infty} y^n K_{\iota} e^{-\gamma y} dy = K_{\iota} \frac{n!}{\gamma^{n+1}}, n = 0, 1, 2, ...,$$

where K_{ι} is given in (5.117). In particular the mean and variance of the actual wait are

$$E(W_{q,\iota}) = \frac{K_{\iota}}{\gamma}, \quad Var(W_{q,\iota}) = \frac{K_{\iota}(2\gamma - K_{\iota})}{\gamma^4}.$$

The Laplace-Stieltjes transform of the actual wait is

$$F_{\iota}(0) e^{-s0} + \int_{y=0}^{\infty} e^{-sy} K_{\iota} e^{-\gamma y} dy = F_{\iota}(0) + \frac{K_{\iota}}{s + \gamma}, s > 0.$$

5.5 Discussion

5.5.1 Heavy-Tailed Inter-arrivals

For the LC analysis of G/M/c the inter-arrival times may have a ***heavy-tailed distribution*** (see [131]), e.g., a Pareto distribution with

$$A(x) = 1 - \frac{1}{(1+x)^{\alpha}}, \quad \overline{A}(x) = \frac{1}{(1+x)^{\alpha}}, \quad a(x) = \frac{\alpha}{(1+x)^{\alpha+1}}, x \geq 0,$$

where α is the shape parameter. All moments exist up to $\lceil \alpha - 1 \rceil$, where $\lceil u \rceil$ denotes the smallest integer $\geq u$. The LC solution technique used in this chapter applies because the solution for γ depends only on the complementary cdf $\overline{A}(\cdot)$, the probability of the *tail of the distribution*, and not on whether the distribution mean and variance exist.

Similar remarks apply to inter-arrival times which have a folded Cauchy, or inverse-log distribution, etc. Additional LC results for heavy-tailed inter-arrival times are given in [87].

5.5.2 Model Variants

The LC solution technique in this chapter is useful for analyzing state dependent models. For example, inter-arrival times and/or service rates of arrivals may depend on the number of customers in service, or on the system time of the last departure from the system. LC can be used to analyze other generalizations, e.g., bounded workload, or service rate depending on waiting time. In generalized models, we could derive integral equations for the pdf of wait in a manner similar to that for the standard G/M/c or G/M/1 queue (see, e.g., [19]).

Chapter 6
Dams and Inventories

6.1 Introduction

In this chapter we analyze several models of dams and inventories with state space $\boldsymbol{S} \subseteq \mathbb{R}$, using LC. When the content in a dam, or stock on hand in an inventory, is positive-valued, it can decline at varying instantaneous rates in accordance with a general release rule specified in the model. Thus the efflux differs from the virtual wait or workload in M/G/1 queues, which *decreases* at rate 1 when positive, or the extended age in G/M/c queues, which *increases* at rate 1.

Section 6.2 describes a dam with general release rule, denoted by M/G/$r(\cdot)$ (or 'M/G/1 dam'). The function $r(x)$, $x \geq 0$, denotes the *efflux rate* when the content is at level x, having dimension [(*Content unit)/Time*]. We discuss sample-path and SP transitions in the time-state space, and derive *integro-differential* equations for the *transient* (time-dependent) distribution of the content. The subscript "t" is used to indicate transience. Integral equations for the *steady-state* (limiting) distribution of content are then obtained by taking limits as $t \to \infty$.

Sections 6.3–6.9 apply SPLC to analyze several models of dams and inventories in steady state.

6.2 M/G/r($\cdot$) Dam

6.2.1 Model Description

Consider a dam with state space $\boldsymbol{S} = [0, \infty)$. Denote the content at instant t by $W(t), t \geq 0$. Assume inputs occur at a Poisson rate λ. Denote the instants of input by τ_n, $n = 1, 2, \ldots,$ where $0 \equiv \tau_0 < \tau_1 < \tau_2 < \cdots$. Denote the input size at τ_n by S_n. We assume $\{S_n\}_{n=1,2,\ldots}$ are i.i.d. positive r.v.s independent of n, with $S_n \underset{dis}{\equiv} S$. Let $B(x) = P(S \leq x)$, and $\overline{B}(x) = 1 - B(x)$.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_6

In some state-dependent model variants, the input size may depend on the content $W(\tau_n^-)$ just before input instant τ_n (denoted by $S(W(\tau_n^-))$, or on a Markovian environment (e.g., denoted by $S_{(i)}$ where i is a state of a continuous-time Markov chain describing the environment). Other input-time dependencies are possible.

If S depends on the current content only, the conditional cdf of $S(W(\tau_n^-))$ given $W(\tau_n^-) = y$, is denoted as

$$B_y(x) = P(S(W(\tau_n^-)) \leq x | W(\tau_n^-) = y), y \geq 0, n = 1, 2, \ldots .$$

The *efflux rate* of content out of the dam, is denoted by $r(W(t))$, defined in Sect. 6.2.2 below. Generally, the efflux rate depends on the current content (see Sect. 5 in [77]).

In M/G/r(·), we assume that the entire input amount goes into the dam instantaneously at an input instant. Under this assumption the model applies to some real-world situations, e.g., systems involving torrential rainfalls, repeated shocks, bolus injections of a prescription medication in pharmacokinetics, instillment of certain eye drops, consumer response to a particular product when exposed to repeated non-uniform advertising in marketing-science models (e.g., [40, 47]), etc.

6.2.2 General Efflux Rate

Let $r(W(t))$ denote the instantaneous efflux rate at which the content decreases (flows out of the dam) at instant t, when the content is $W(t)$. Assume $r(W(t))$ is finite and

$$\left.\begin{array}{l} r(x) > 0 \text{ if } x > 0, \\ r(x) = 0 \text{ if } x = 0. \end{array}\right\} \tag{6.1}$$

The rate of decline of $W(t)$ between input instants is (see Sect. 5 in [77])

$$\frac{dW(t)}{dt} = -r(W(t)), \tau_n \leq t < \tau_{n+1}, n = 0, 1, 2, \ldots, \tag{6.2}$$

independent of n. The variable $r(W(t))$ has 'physical' dimension (unit) $\frac{[content\ unit]}{[Time]}$, e.g., $\frac{[Volume]}{[Time]} = [L^3T^{-1}]$, where $L := Length$ and $T := Time$.

This section assumes that $r(x)$, $x \in \boldsymbol{S}$ is a time-homogeneous piecewise right-continuous function, except at level 0. Usually, $r(0) \neq r(0^+) = \lim_{x\downarrow 0} r(x)$. However, equality of $r(0)$ and $r(0^+)$ is possible in some models.

Example 6.1 Suppose $r(x) = (x + 1)^2, x > 0, r(0) = 0$. Then $r(0^+) = 1 \neq r(0)$. On the other hand, suppose $r(x) = x^2, x > 0$, and $r(0) = 0$. Then $r(0^+) = r(0)$.

In some model variants, $r(x), x \geq 0$, may have different functional forms on separate state space intervals. In such cases, consider a state-space partition $\{x_j\}_{j=0,1,\ldots,n+1}$ where $0 \equiv x_0 < x_1 < x_2 < \cdots < x_n < x_{n+1} \equiv \infty$. Let $\boldsymbol{I}_1 := (x_0, x_1)$, and

$\boldsymbol{I}_j := \left[x_{j-1}, x_j\right), j = 2, \ldots, n+1$. Define $\left\{r_j(x)\right\}_{j=0,1\ldots,n+1}$ by

$$r(x) = \begin{cases} r_0(0) = 0 \\ r_1(x), x \in (0, x_1) \equiv \boldsymbol{I}_1 \\ r_2(x), x \in [x_1, x_2) \equiv \boldsymbol{I}_2 \\ \quad \cdots\cdots \\ r_n(x), x \in [x_{n-1}, x_n) \equiv \boldsymbol{I}_n \\ r_{n+1}(x), x \in [x_n, \infty) \equiv \boldsymbol{I}_{n+1}, \end{cases} \tag{6.3}$$

where $r_j(x), x \in \boldsymbol{I}_j$ is positive and continuous, $j = 1, 2, \ldots, n+1$. (See, e.g., Sects. 3.1 and 3.2 in [19] for examples using state-space partitions.)

Remark 6.1 In some model generalizations $r(W(t))$ may also depend on t. We would then append a subscript t, i.e., denote the efflux rate as $r_t(W(t))$.

6.2.3 Sample Paths

We use the symbol '$W(t)$' to denote the content of the dam, and also to denote the ordinate of a sample path of the content at instant t (unless specified otherwise), for economy of notation, and because the usage will be clear from the context.

A sample path of $\{W(t)\}_{t\geq 0}$ is a piecewise *deterministic* function plotted in the *time-state* plane $\boldsymbol{T} \times \boldsymbol{S}$, where $\boldsymbol{T} := \{t | t \geq 0\}$ (Fig. 6.1).

6.2.4 Time for $\{W(t)\}_{t\geq 0}$ to Decrease to a Level

In Eq. (6.2), separating variables gives the differential equation

$$\frac{dW(t)}{r(W(t))} = -dt, \tau_n \leq t < \tau_{n+1}, n = 0, 1, 2, \ldots;$$

integrating both sides gives

$$\int_{W(t_x)}^{W(t_y)} \frac{1}{r(W(t))} dW(t) = -\int_{t_x}^{t_y} dt = t_x - t_y.$$

(See Fig. 6.2). The *time* required for a sample path of $\{W(t)\}_{t\geq 0}$ to descend from level $W(t_y) = y$ at instant t_y to a lower level $W(t_x) = x \geq 0$ at instant t_x, if no inputs to the dam intervene, i.e., if

$$W(\tau_n) > y > x \geq W(\tau_{n+1}^-) \geq 0, \text{ for some fixed } n,$$

is

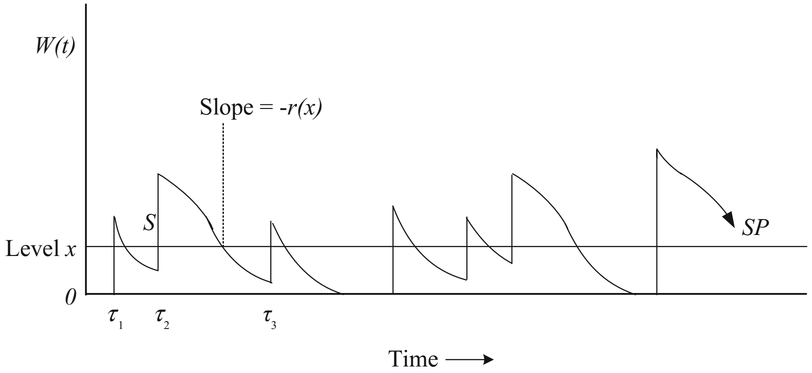


Fig. 6.1 Sample path of $\{W(t)\}_{t\geq 0}$ in $M/G/r(\cdot)$ dam

$$t_x - t_y = \int_{z=x}^{y} \frac{1}{r(z)} dz. \tag{6.4}$$

upon substituting $W(t) = z$.

Formula (6.4) is useful when analyzing models of dams and inventories in continuous time (as in this chapter), or when analyzing a dam via the *embedded LC method* (see Sect. 8.2 in Chap. 8).

6.2.5 Condition for $\{W(t)\}_{t\geq 0}$ to Return to Level 0

Formula (6.4) implies that a necessary and sufficient condition for a return by $\{W(t)\}_{t\geq 0}$ to level 0, is

$$\lim_{x\downarrow 0} \int_{z=x}^{y} \frac{1}{r(z)} dz < \infty \text{ for every finite } y > 0, \tag{6.5}$$

(see pp. 116–117 in [77]).

6.2.6 Transient Probability Distribution of Content

Transient CDF and PDF

Denote the transient cdf of $W(t)$ by $F_t(x)$, $x \geq 0$, and let $F_t(0) := P_0(t)$. Let $f_t(x) := \partial F_t(x)/\partial x$, $x > 0$, wherever the derivative exists. We denote the transient pdf of $W(t)$ by $\{P_0(t), f_t(x)\}_{t\geq 0}$. Assume $F_t(x)$, $f_t(x)$ are right continuous in x. We use $f_t(0^+)$ and $f_t(0)$ interchangeably for notational convenience since $f_t(0)$ adds zero probability to $P_0(t)$. The function $f_t(x)$ may have jump discontinuities depending on the distribution of the input r.v.s. (See, e.g., Sects. 3.10 and 3.11 regarding the pdf of wait in M/D/1 and M/Discrete/1 queues.)

For each $t \geq 0$,

$$F_t(x) = P_0(t) + \int_{y=0}^{x} f_t(y)dy,$$

and the normalizing condition is

$$F_t(\infty) = P_0(t) + \int_{y=0}^{\infty} f_t(y)dy = 1.$$

Steady-State Probability Distribution

We mention the steady-state cdf and pdf now because we will derive them in Sect. 6.2.11, immediately after the discussion of the transient cdf and pdf below. The steady-state cdf and pdf of content are denoted as $F(x), x \geq 0$, and $\{P_0, f(x)\}_{x>0}$ respectively, and are obtained by letting $t \to \infty$, i.e.,

$$F(x) = \lim_{t\to\infty} F_t(x), x \geq 0, f(x) = \lim_{t\to\infty} f_t(x), x > 0, P_0 = \lim_{t\to\infty} P_0(t).$$

Remark 6.2 P_0 exists if and only if a sample path of $\{W(t)\}_{t\geq 0}$ returns to level 0 with probability 1. However, some forms of $r(W(t))$ make returns to level 0 impossible (see pp. 116–117 in [77], and Sect. 6.2.5).

6.2.7 Sample-Path and SP Downcrossings

Consider a sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 6.1). Fix level $x \in \boldsymbol{S}$. Let $\mathcal{D}_t(x)$ denote the number of SP downcrossings of level x during $(0, t)$. The SP traces the sample path during piecewise continuous segments between input instants. At sample-path discontinuities, the SP makes an upward jump, *not in Time* (see Sects. 2.4.3 and 2.4.4). Let $\mathcal{D}_t^c(x)$ and $\mathcal{D}_t^j(x)$ denote respectively the number of SP *left-continuous* downcrossings and SP *jump* downcrossings of level x during $(0, t)$. Then

$$\mathcal{D}_t(x) = \mathcal{D}_t^c(x) + \mathcal{D}_t^j(x), x \geq 0, t \geq 0.$$

In the basic M/G/r(·) dam of this section, $\mathcal{D}_t^j(x) \equiv 0$, $t \geq 0$. In variations of the basic model, however, SP downward jumps can indeed occur. Both SP left-continuous downcrossings and SP jump downcrossings also occur in a vast number of *inventory* and *production-inventory* models. Thus, we shall distinguish $\mathcal{D}_t(x)$ from $\mathcal{D}_t^c(x)$ in Theorem 6.1 in Sect. 6.2.8. Note that $\mathcal{D}_t^c(0)$ may equal 0 in certain cases of $r(W(t))$ (see Sect. 6.2.5).

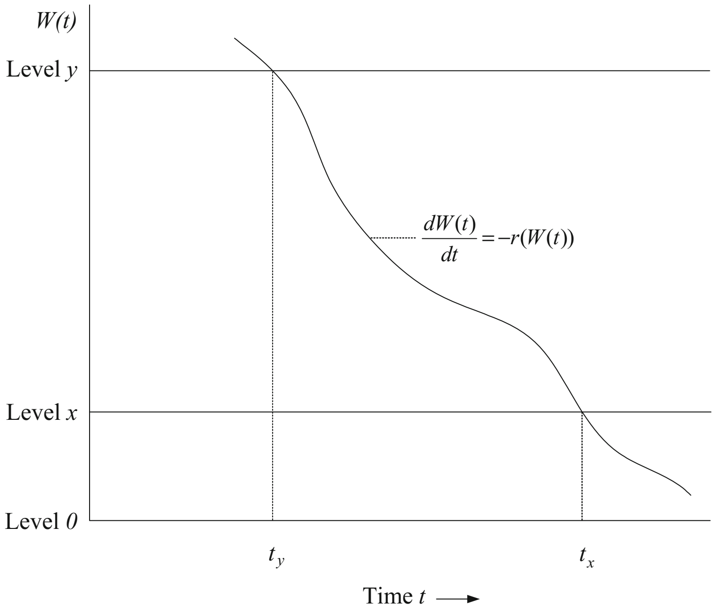


Fig. 6.2 M/G/r(·) dam: time to descend from level y to level $x > 0$ is $t_x - t_y = \int_{z=x}^{y} \frac{1}{r(z)} dz$

6.2.8 Level Crossings and Transient PDF of Content

In a sample path of $\{W(t)\}_{t \geq 0}$, fix level $x \in \boldsymbol{S}$ (Fig. 6.1). Let $\mathcal{U}_t(x) :=$ *number of SP upcrossings of level x during* $(0, t)$. It can be shown, along the lines of Sects. 3.2.1 and 3.2.2, that $\frac{\partial}{\partial t} E(\mathcal{D}_t^c(x))$, $\frac{\partial}{\partial t} E(\mathcal{U}_t(x))$ exist and are positive.

Theorems 6.1 and 6.2 were originally proved using LC in [23].

Downcrossings

Theorem 6.1 For the M/G/r(·) dam

$$\frac{\partial}{\partial t} E(\mathcal{D}_t^c(x)) = r(x) f_t(x), x > 0, \tag{6.6}$$

$$\frac{\partial}{\partial t} E(\mathcal{D}_t^c(0)) = r(0^+) f_t(0). \tag{6.7}$$

Proof Consider a sample path of $\{W(t)\}_{t \geq 0}$, and fix state-space level $x \in \boldsymbol{I}_j$ for some $j \in \{1, \ldots, n+1\}$ in (6.3). Fix instant t. Consider $t + h$, $(h > 0)$ and define $\delta > 0$ by

$$\int_{z=x}^{x+\delta} \frac{1}{r(z)} dz = h. \tag{6.8}$$

Assume h is sufficiently small so that level $x + \delta \in \boldsymbol{I}_j$; h is the time for the content to decrease from level $x + \delta$ to level x if there are no inputs during $(t, t + h)$ (see formula (6.4)). Applying the (first) mean value theorem for integrals with continuous integrand (see, e.g., Problems 27–28, p. 237 in [137]) to Eq. (6.8) yields

$$h = \frac{1}{r(z^*)}\delta \iff \delta = r(z^*)h \tag{6.9}$$

for some z^* such that $x < z^* < x + \delta$.

The event $\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x) = 1$ occurs iff $W(t) \in (x, x + \delta)$ and there is no input in a time subinterval $(t, t + \xi) \subseteq (t, t + h)$, or an event with probability $o(h)$ occurs. From (6.9)

$$\begin{aligned} P(\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x) = 1) &= f_t(x) \cdot \delta \cdot (1 - \lambda h) + o(h) \\ &= f_t(x) \cdot r(z^*) \cdot h \cdot (1 - \lambda h) + o(h). \end{aligned}$$

The value $\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x) = 0$ has no affect on $E(\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x))$. Due to the Poisson input stream, $P(\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x) \geq 2) = o(h)$. Hence the expected value

$$\begin{aligned} E(\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x)) &= 1 \cdot P(\mathcal{D}^c_{t+h}(x) - \mathcal{D}^c_t(x) = 1) + o(h), \\ E(\mathcal{D}^c_{t+h}(x)) - E(\mathcal{D}^c_t(x)) &= f_t(x) \cdot r(z^*) \cdot h \cdot (1 - \lambda h) + o(h). \end{aligned} \tag{6.10}$$

Dividing both sides of (6.10) by h and letting $h \downarrow 0$ gives (6.6) since $z^* \downarrow x$ and $r(z^*) \downarrow r(x^+) = r(x)$, $x > 0$, as $h \downarrow 0$. Then letting $x \downarrow 0$ in (6.6) gives (6.7). ■

Corollary 6.1 For each $t \geq 0$,

$$\begin{aligned} E(\mathcal{D}^c_t(x)) &= r(x) \int_{s=0}^{t} f_s(x) ds, \; x > 0, \\ E(\mathcal{D}^c_t(0)) &= r(0^+) \int_{s=0}^{t} f_s(0) ds. \end{aligned}$$

Proof In (6.6) and (6.7) set $t = s$, integrate with respect to $s \in [0, t]$, and apply the initial condition $E(\mathcal{D}^c_0(x)) = 0$, $x \geq 0$. ■

Corollary 6.2 The steady-state pdf of $\{W(t)\}_{t \geq 0}$ as $t \to \infty$ is given in terms of downcrossing rates by

$$\lim_{t \to \infty} \frac{E(\mathcal{D}^c_t(x))}{t} = r(x) f(x), \; x > 0, \tag{6.11}$$

$$\lim_{t \to \infty} \frac{E(\mathcal{D}^c_t(0))}{t} = r(0^+) f(0). \tag{6.12}$$

Proof In Corollary 6.1, since $\lim_{s\to\infty} f_s(x) = f(x)$, for every $\varepsilon > 0$ there exists t_ε such that $|f_s(x) - f(x)| < \varepsilon$ for $s > t_\varepsilon$, implying

$$\int_{s=0}^{t} f_s(x)ds < C_\varepsilon + \int_{t_\varepsilon}^{t} (f(x) + \varepsilon)\, ds = C_\varepsilon + (t - t_\varepsilon)\,(f(x) + \varepsilon)\,,$$
$$\int_{s=0}^{t} f_s(x)ds > C_\varepsilon + \int_{t_\varepsilon}^{t} (f(x) + \varepsilon)\, ds = C_\varepsilon + (t - t_\varepsilon)\,(f(x) - \varepsilon)\,,$$

where the constant $C_\varepsilon := \int_{s=0}^{t_\varepsilon} f_s(x)ds$ and $t > t_\varepsilon$. Combining both inequalities yields

$$C_\varepsilon + (t - t_\varepsilon)\,(f(x) - \varepsilon) < \int_{s=0}^{t} f_s(x)ds < C_\varepsilon + (t - t_\varepsilon)\,(f(x) + \varepsilon)\,.$$

Dividing throughout by t gives

$$\frac{C_\varepsilon}{t} + \left(1 - \frac{t_\varepsilon}{t}\right)(f(x) - \varepsilon) < \frac{1}{t}\int_{s=0}^{t} f_s(x)ds < \frac{C_\varepsilon}{t} + \left(1 - \frac{t_\varepsilon}{t}\right)(f(x) + \varepsilon)\,.$$

Letting $t \to \infty$ gives

$$\begin{aligned} f(x) - \varepsilon \quad &< \quad \lim_{t\to\infty} \tfrac{1}{t} \textstyle\int_{s=0}^{t} f_s(x)ds < f(x) + \varepsilon \\ &\implies \quad \lim_{t\to\infty} \tfrac{1}{t} \textstyle\int_{s=0}^{t} f_s(x)ds = f(x) \end{aligned}$$

since $\varepsilon > 0$ is arbitrarily small, thus yielding (6.11); then setting $x = 0$ gives (6.12). ■

Upcrossings

Theorem 6.2 For the M/G/r(·) dam

$$\begin{aligned} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) &= \lambda \int_{z=0}^{x} \overline{B}(x - z)dF_t(z) \\ &= \lambda P_0(t)\overline{B}(x) + \lambda \int_{z=0}^{x} \overline{B}(x - z)f_t(z)dz, x > 0, \end{aligned} \tag{6.13}$$

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(0)) = \lambda P_0(t). \tag{6.14}$$

Proof Fix instants t and $t + h, t \geq 0, h > 0$ (h small). Fix level $x > 0$. Then $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) = 1$ iff $W(s) = z < x$ at an instant $s \in (t, t+h)$ at which there is an input of size $S > x - z$, or an event having probability $o(h)$ occurs. The value $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) = 0$ does not contribute to $E(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x))$. Also $P(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) \geq 2) = o(h)$.

$$\begin{aligned} E(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)) &= E(\mathcal{U}_{t+h}(x)) - E(\mathcal{U}_t(x)) \\ &= \lambda \int_{z=0}^{x} \int_{s=0}^{h} \overline{B}(x-z)dsdF_{t+s}(z) + o(h) \\ &= \lambda h \int_{z=0}^{x} \overline{B}(x-z)dF_{t+s^*}(z) + o(h) \end{aligned} \tag{6.15}$$

where $0 < s^* < h$. Dividing both sides of (6.15) by h and letting $h \downarrow 0$ gives (6.13) since $s^* \downarrow 0$ as $h \downarrow 0$, and $F_t(\cdot)$ is right-continuous in t. Then letting $x \downarrow 0$ in (6.13) gives (6.14).

Note: If $r(0^+) = 0$ then $P_0(t) = 0$, $t > 0$ (see Sect. 6.2.5 and Remark 6.2) in Sect. 6.2.6. ■

Corollary 6.3

$$\begin{aligned} E(\mathcal{U}_t(x)) &= \lambda \int_{s=0}^{t} \int_{z=0}^{x} \overline{B}(x-z)dF_s(z)ds \\ &= \lambda \int_{s=0}^{t} P_0(s)\overline{B}(x)ds + \lambda \int_{s=0}^{t} \left[\int_{z=0}^{x} \overline{B}(x-z)f_s(z)dz \right] ds, x > 0, \\ E(\mathcal{U}_t(0)) &= \lambda \int_{s=0}^{t} P_0(s)\overline{B}(x)ds. \end{aligned}$$

Proof Set $t = s$ in (6.13) and (6.14), integrate with respect to $s \in [0, t]$ and apply the initial condition $E(\mathcal{U}_0(x)) = 0, x \geq 0$. ■

Corollary 6.4

$$\begin{aligned} \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} &= \lambda \int_{z=0}^{x} \overline{B}(x-z)dF(z)dz \\ &= \lambda P_0 \overline{B}(x) + \lambda \int_{z=0}^{x} \overline{B}(x-z)f(z)dz, x > 0, \\ \lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t} &= \lambda P_0 \overline{B}(x). \end{aligned}$$

Proof In Corollary 6.3, interchange the order of integration. Divide both sides by t and let $t \to \infty$. The result follows since $\lim_{s\to\infty} f_s(z) = f(z)$ implying $\lim_{t\to\infty} (1/t) \int_{s=0}^{t} f_s(z)ds = f(z)$. (See the proof of Corollary 6.2 above in this Section. Also see the **Note** immediately after Theorem 6.2 above, regarding the condition ensuring $P_0 > 0$). ■

6.2.9 *Equation for Transient Distribution of Content*

The following theorem has been proved using classical methods by various authors (see, e.g., Eq. (5.4) in [77]). Here we prove it using LC (based on [23]).

Theorem 6.3 In the M/G/r(·) dam, the transient pdf of content, $f_t(x), x > 0$, satisfies the **integro-differential equation**

$$\begin{aligned} r(x)f_t(x) &= \tfrac{\partial}{\partial t}F_t(x) + \lambda \int_{z=0}^{x} \overline{B}(x-z)dF_t(z) \\ &= \tfrac{\partial}{\partial t}F_t(x) + \lambda\overline{B}(x)P_0(t) \\ &\quad +\lambda \int_{z=0}^{x} \overline{B}(x-z)f_t(z)dz, x > 0, \end{aligned} \tag{6.16}$$

and $P_0(t)$ satisfies the **differential equation**

$$\frac{d}{dt}P_0(t) + \lambda P_0(t) = r(0^+)f_t(0). \tag{6.17}$$

Proof In Theorem 4.1 (i.e., Theorem B in Sect. 4.2), substitute set $[0, x] = \boldsymbol{A}$, $\mathcal{D}_t^c(x) = \mathcal{I}_t(x), \mathcal{U}_t(x) = \mathcal{O}_t(x)$. This gives

$$\frac{\partial}{\partial t}E(\mathcal{D}_t^c(x)) = \frac{\partial}{\partial t}F_t(x) + \frac{\partial}{\partial t}E(\mathcal{U}_t(x)) \tag{6.18}$$

Substituting from (6.6) and (6.13) into (6.18) gives (6.16). Equation (6.17) then follows by letting $x \downarrow 0$ in (6.16), noting that $F_t(0) = P_0(t)$.

See the **Note** at the end of Theorem 6.2 in Sect. 6.2.8. ■

Remark 6.3 The dimension of $r(x)$ is $\left[\frac{content\ unit}{Time}\right]$. The dimension of $f_t(x)$ is $\left[\frac{1}{content\ unit}\right]$. The dimension of the left sides of (6.16) and of (6.17), is

$$\left[r(x)f_t(x)\right] = \left[\frac{content\ unitt}{Time}\right] \cdot \left[\frac{1}{content\ unit}\right] = \frac{1}{[Time]}, x \geq 0,$$

which matches the dimensional unit of the right side.

6.2.10 Estimate of Transient Probability $P_0(t)$

We briefly outline an '***LC estimation***' procedure for the *transient* probability $P_0(t)$, $t \geq 0$, *assuming* $P_0(t)$ *exists for all* $t > 0$, which occurs provided returns to level 0 are regenerative points (i.e., $r(0^+) > 0$). (See Sect. 6.2.5 and Remark 6.2 in Sect. 6.2.6.) We also call this procedure LCE, or LC computation. LCE to compute a pdf $f_t(x)$, $x > 0$, would be similar. We do not detail LCE for *transient pdfs* elsewhere in this monograph. See Remark 9.2 in Sect. 9.2 in Chap. 9. We detail LCE for *limiting distributions* in Chap. 9.

To solve differential equation (6.17) multiply by the integrating factor $e^{\lambda t}$, and integrate with respect to t, yielding

$$P_0(t) = \left[\int_{s=0}^{t} e^{\lambda s}\frac{\partial}{\partial s}E(\mathcal{D}_s^c(0))ds + P_0(0)\right]e^{-\lambda t}, \tag{6.19}$$

where

$$P_0(0) = \begin{cases} 1 \text{ if } W(0) = 0, \\ 0 \text{ if } W(0) \neq 0. \end{cases}$$

Formula (6.19) connects $P_0(t)$ and $\partial E(\mathcal{D}_s^c(0))/ds$, $0 < s < t$, which appears as a factor in the integrand. This connection leads to an *estimation method* for $P_0(t)$, by estimating the integral in (6.19).

The idea is to first simulate N *independent* sample paths of $\{W(t)\}_{t\geq 0}$ denoted as $\{W_n(s)\}_{s\geq 0, n=1,\ldots,N}$ on the same time interval $[0, t_M + r]$, where t_M is the maximum finite time of interest, r is an "extra" finite time which ensures that t_M is not the right end point of the simulated time interval. N is a large positive integer. A reasonable value of N would be in the range [400, 1,000]. Due to the high speed of today's computers, N may be considerably larger than 1,000. Let $h = t_M/m$ be small, where m is a positive integer. We can use, e.g., $h = 0.001$ or 0.0001, or any small value $h < r$. The accuracy of the estimation of $P_0(t)$, $t \in [0, T_M]$, improves with larger values of N and smaller values of h.

We then compute the number of SP *left-continuous downcrossings* (hits of level 0) denoted by $\mathcal{D}_{ih,n}^c(0)$, $i = 0, \ldots, m$, for each sample path, $\{W_n(s)\}_{n=1,\ldots,N}$. For fixed i and n, the $\mathcal{D}_{ih,n}^c(0)$s are independent since the N sample paths are independent. We compute point estimates of the true downcrossing rates $\mathcal{D}_{ih,n}^c(0)$ and $\mathcal{D}_{(i+1)h,n}^c(0)$ at times ih and $(i+1)\,h$ respectively by averaging over the N sample paths. Then we compute estimates of $E(\mathcal{D}_{ih}^c(0))$ and $E(\mathcal{D}_{(i+1)h}^c(0))$ using

$$\widehat{E}(\mathcal{D}_{ih}^c(0)) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{D}_{ih,n}^c(0), \widehat{E}(\mathcal{D}_{(i+1)h}^c(0)) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{D}_{(i+1)h,n}^c(0).$$

An estimate of the *derivative* $\partial E(\mathcal{D}_{ih}^c(0))/dt$ is then given by the difference quotient

$$\widehat{\frac{\partial}{\partial t}}E(\mathcal{D}_{ih}^c(0)) = \frac{\widehat{E}(\mathcal{D}_{(i+1)h}(0)) - \widehat{E}(\mathcal{D}_{ih}^c(0))}{h}, i = 0, \ldots, m.$$

Finally, we approximate the integral $\int_{s=0}^{kh} e^{\lambda s}\frac{\partial}{\partial s}E(\mathcal{D}_s^c(0))ds$ as a finite Riemann sum

$$h\sum_{i=0}^{k} e^{\lambda ih}\widehat{\frac{\partial}{\partial t}}E(\mathcal{D}_{ih}^c(0)), k = 1, \ldots, m$$

A point estimate of $P_0(kh)$ is

$$\widehat{P}_0(kh) = \left[h \sum_{i=0}^{k} e^{\lambda ih} \widehat{\frac{\partial}{\partial t}} E(\mathcal{D}_{ih}^c(0)) + P_0(0) \right] e^{-\lambda t}, k = 1, \ldots, m, \tag{6.20}$$

where $mh = t_M$. This technique results in estimates of $P_0(h)$, $P_0(2h)$, …, $P_0(mh)$. Thus, we estimate $P_0(t)$, $t = 0, h, 2h, \ldots, t_M$. Smoothing techniques can be applied to estimate intermediate values. Then we can plot $\widehat{P}_0(t)$, $0 < t < t_M$. (We can also develop interval estimates for $P_0(kh)$, $k = 1, \ldots, m$.)

Generalizations and variations of this technique can be used to estimate transient distributions of state variables in many stochastic models having a continuous time parameter.

The foregoing example of **LCE** relates to Chap. 9, which describes LCE for steady-state distributions. LCE has also been discussed in [17] and [24]. (Also, see Remark 9.2 in Sect. 9.2 in Chap. 9.)

Remark 6.4 Future computer speeds will undoubtedly increase. Thus the computational method described above will achieve better and better accuracy. It will be possible to increase N and decrease h, while completing the computations in a much shorter amount of real time.

Remark 6.5 In the M/G/r(·) dam, possibly $P_0(t) = 0$ for all $t \geq \tau_1$ (instant of first input). For example, if $r(x) = kx$, $x > 0$, $k > 0$, the decline of the sample-path has a negative exponential form between inputs. In theory the content will never reach level zero after the first input at τ_1. If the inter-input time is very long, the content eventually declines below any preassigned level $\varepsilon > 0$ however small, but never reaches level 0. In that case we may use downcrossings of an arbitrary level $\varepsilon > 0$ as regeneration points of a regenerative process. $\{W(t)\}_{t \geq 0}$ will then move along level ε until the next arrival. We may then use '$P_t(\varepsilon)$' like '$P_0(t)$'. Alternatively, we just estimate $f_t(x)$, $x > 0$ (see Remark 6.2 in Sect. 6.2.6).

6.2.11 *Equation for Steady-State PDF of Content*

Assume the system is stable *and* $\{W(t)\}_{t \geq 0}$ returns to level 0 with probability 1. Then

$$F(x) = \lim_{t \to \infty} F_t(x), \qquad f(x) = \lim_{t \to \infty} f_t(x), \qquad P_0 = F(0) = \lim_{t \to \infty} P_0(t)$$

all exist, and $\lim_{t \to \infty} \frac{\partial}{\partial t} F_t(x) = 0$. In Eq. (6.16), taking limits of all terms as $t \to \infty$ yields

$$\begin{aligned} r(x)f(x) &= \lambda \textstyle\int_{y=0}^{x} \overline{B}(x-y)dF(y), x > 0, \\ r(x)f(x) &= \lambda P_0 \overline{B}(x) + \lambda \textstyle\int_{y=0}^{x} \overline{B}(x-y)f(y)dy, x > 0, \\ r(0^+)f(0) &= \lambda P_0. \end{aligned} \tag{6.21}$$

Alternative Forms of Equation for Steady-State PDF

Two alternative forms of the integral equation in (6.21) are

$$r(x)f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} B(x-y)f(y)dy, x > 0; \tag{6.22}$$

$$r(x)f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} F(x-y)b(y)dy, x > 0, \tag{6.23}$$

where $b(y) = dB(y)/dy$.

Explanation of (6.22) and (6.23). In each equation the left side is $\lim \mathcal{D}_t(x)/t$, the SP *downcrossing* rate of level x. On the right side, the first term $\lambda F(x)$ is the rate of inputs when the content is $\leq x$; these inputs generate upward jumps that *start* in state-space interval $[0, x]$. The second term subtracts off the rate of such jumps that do not upcross level x. Hence the right side is $\lim \mathcal{U}_t(x)/t$, the *upcrossing* rate of level x. Applying rate balance $\lim \mathcal{D}_t(x)/t = \lim \mathcal{U}_t(x)/t$, gives the alternative equations.

Equations (6.22) and (6.23) are analogous to Eqs. (3.43) and (3.44) in Sect. 3.3.1 for the M/G/1 queue.

Stability

A condition for stability of the M/G/r(·) dam is

$$\lambda E(S) < \lim_{x \to \infty} r(x). \tag{6.24}$$

Formula (6.24) asserts the rate at which the content increases is less than the efflux rate when the content is at high levels. So the content is prevented from increasing to indefinitely high amounts. Condition (6.24) guarantees the return of $\{W(t)\}_{t\geq 0}$ *to every level* $x > 0$ in a finite time (see pp. 116–117 in [77]; Theorem 2 in [134].).

A condition that guarantees the content *will return to level* 0, therefore implying $P_0 > 0$, is Eq. (6.5) in Sect. 6.2.5 above.

Example 6.2 The M/G/1 queue is a special case of the M/G/r(·) dam with $r(x) \equiv 1$, $x > 0$, and $r(0) = 0$. Stability holds iff $\lambda E(S) < \lim_{x\to\infty} r(x) = 1$, the well-known stability condition for M/G/1 queues; if stability holds $\{W(t)\}_{t\geq 0}$ returns to level 0 (a.s.) since for all **finite** $x > 0$

$$\lim_{u\downarrow 0} \int_{y=u}^{x} \frac{1}{r(y)} dy = \lim_{u\downarrow 0} \int_{y=u}^{x} 1 \cdot dy = \lim_{u\downarrow 0}(x-u) = x < \infty$$

if no arrivals intervene.

Example 6.3 In the M/G/r(·) dam with $\lambda > 0$, $E(S) < \infty$, and $r(x) = kx$, $k > 0$,

$$\lim_{u\downarrow 0} \int_{y=u}^{x} \frac{1}{ky} dy = \frac{1}{k} \lim_{u\downarrow 0} \left(\ln \left(\frac{x}{u} \right) \right) = \infty,$$

for every finite $x > 0$. Hence the content **does not return to level** 0**,** implying $P_0 = 0$. On the other hand, this dam is **stable** for every $k > 0$ because

$$\lambda E(S) < \lim_{x\to\infty} r(x) = \lim_{x\to\infty} kx = \infty.$$

6.2.12 Sojourn Times Related to State-Space Level x

Consider a sample path of $\{W(t)\}_{t\geq 0}$. Fix level $x > 0$. Due to Poisson arrivals and the level-dependent slope of the efflux, $\{\mathcal{D}_t(x)\}_{t\geq 0}$ (same as $\left\{\mathcal{D}_t^c(x)\right\}_{t\geq 0}$) is a renewal counting process. The times between successive downcrossings of level x (renewals) are i.i.d. r.v.s. The instants of SP downcrossings of level x are regenerative points with respect to the process $\{W(t)\}_{t\geq 0}$, where $\{W(t)\}_{t\geq 0}$ restarts independent of the past.

Let d_x := *time between successive downcrossings of level* x. Let a_x, b_x denote sojourn times above and below level x, respectively. A sojourn a_x begins with an upcrossing of x and ends with the first downcrossing of x thereafter. A sojourn b_x begins with a downcrossing of x and ends with the first upcrossing of x thereafter. Thus $d_x = b_x + a_x$.

Inter-downcrossing Time d_x
For the process $\{\mathcal{D}_t(x)\}_{t\geq 0}$, using (6.11) and the elementary renewal theorem, the renewal rate is

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t^c(x))}{t} = \lim_{t\to\infty}\frac{\mathcal{D}_t^c(x)}{t} \underset{a.s.}{=} r(x)f(x) = \frac{1}{E(d_x)}, x > 0.$$

Hence

$$E(d_x) = \frac{1}{r(x)f(x)}, x > 0. \tag{6.25}$$

Sojourn a_x **Above Level** x
From the renewal reward theorem

$$\frac{E(a_x)}{E(d_x)} = \frac{\lim_{t\to\infty}(time\ \ SP\ \ is\ \ above\ \ level\ \ x\ \ during\ (0,t))}{t} = 1 - F(x),$$

$$E(a_x) = (1 - F(x))\cdot E(d_x) = \frac{1 - F(x)}{r(x)f(x)}, x > 0. \tag{6.26}$$

From (6.26)

$$\frac{f(x)}{1 - F(x)} = \frac{1}{r(x)E(a_x)}, x > 0, \tag{6.27}$$

$$\frac{d}{dx}\ln(1 - F(x)) = \frac{-1}{r(x)E(a_x)}, x > 0. \tag{6.28}$$

Integrating on both sides of (6.28) with respect to x and computing the constant of integration by letting $x \downarrow 0$, gives

$$\begin{aligned} F(x) &= 1 - (1 - P_0)e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x \geq 0,\\ f(x) &= \tfrac{1-P_0}{r(x)E(a_x)}e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x > 0. \end{aligned} \tag{6.29}$$

(Possibly $0 < P_0 < 1$ or $P_0 = 0$.) The normalizing condition $F(\infty) = 1$, is

$$1 - (1 - P_0)e^{-\int_{y=0}^{\infty} \frac{1}{r(y)E(a_y)}dy} = 1,$$

which implies that $e^{-\int_{y=0}^{\infty} \frac{1}{r(y)E(a_y)}dy} = 0$ if $0 < P_0 < 1$.

Hazard Rate of PDF of Content at x

The left side of (6.27) is the *hazard rate* of the steady-state pdf of content at x. An inverse relation holds between it and the product $r(x)E(a_x)$. The hazard rate has the same dimension as $f(x)$, i.e., 1/[*content unit*]. (See Sect. 3.4.18 for definition and discussion of hazard rate.)

Sojourn b_x **Below Level** x

By the renewal reward theorem, $E(b_x)/E(d_x) = F(x)$. Thus

$$E(b_x) = F(x) \cdot E(d_x) = \frac{F(x)}{r(x)f(x)}, x > 0, \tag{6.30}$$

implying

$$\frac{f(x)}{F(x)} = \frac{d}{dx} \ln F(x) = \frac{1}{r(x)E(b_x)}, x > 0,$$

and

$$\begin{aligned} F(x) &= P_0 e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy}, x \geq 0, \\ f(x) &= \frac{P_0}{r(x)E(b_x)} e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy}, x > 0, \end{aligned} \tag{6.31}$$

using $F(0^+) = P_0$.

Interestingly, formulas (6.29) and (6.31) give two different expressions for $F(x)$ and $f(x)$, in terms of $E(a_x)$ and $E(b_x)$, respectively.

If $r(x) \equiv 1, x > 0$, the right side of the second equation in (6.31) reduces to the pdf of wait in the M/G/1 queue, i.e., since $E(b_x) = F(x)/f(x), x \geq 0$,

$$\begin{aligned} \frac{P_0}{1 \cdot E(b_x)} e^{\int_{y=0}^{x} \frac{1}{1 \cdot E(b_y)}dy} &= \frac{P_0 \cdot f(x)}{F(x)} e^{\int_{y=0}^{x} \frac{f(y)}{F(y)}dy} \\ = \frac{P_0 \cdot f(x)}{F(x)} e^{(\ln F(x) - \ln F(0))} &= \frac{P_0 \cdot f(x)}{F(x)} F(x) P_0^{-1} = f(x). \end{aligned}$$

as in formula (3.95).

As a mild check on (6.31), we compute $f(x)$ for the M/M/1 queue in which

$$E(b_x) = \frac{F(x)}{f(x)} = \frac{1 - (1 - (1 - \frac{\lambda}{\mu}))e^{-(\mu - \lambda)x}}{\lambda(1 - \frac{\lambda}{\mu})e^{-(\mu - \lambda)x}}, x \geq 0,$$

and $F(0) = P_0 = 1 - \frac{\lambda}{\mu}$. Substituting these values directly for $E(b_x)$ and P_0 in (6.31) leads to $f(x) = \lambda(1 - \frac{\lambda}{\mu})e^{-(\mu - \lambda)x}$, $x \geq 0$, the steady-state pdf of wait in M/M/1 (formula (3.112) in Sect. 3.5).

6.2.13 CDF and PDF of Excess of Jump over Level x

Let $\gamma_x :=$ excess of an input upcrossing of x (jump starts below x). Let $G_x(z), z > 0$, $g_x(z) = \partial G_x(z)/\partial z$, $z > 0$, denote the cdf and pdf of γ_x, respectively. We determine these quantities by means of an argument analogous to that in the proof of Theorem (3.7) in Sect. 3.7. In steady state,

$$\lim_{t\to\infty} \mathcal{U}_t(x)/t = \lim_{t\to\infty} \mathcal{D}_t(x)/t = r(x)f(x).$$

The rate at which the SP *upcrosses level* $x+z$ *whenever an input amount jump-upcrosses level* x is

$$r(x)f(\dot{x})\,[1 - G_x(z)]\,.$$

A different expression for the *upcrossing rate of level* $x+z$, whenever the input amount jump-upcrosses level x is

$$\lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y)f(y)dy,$$

which is the rate of jumps that start below level x, having excesses over x that upcross level $x+z$.

Therefore

$$r(x)f(x)\,[1 - G_x(z)] = \lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y)f(y)dy,$$

$$\begin{aligned}1 - G_x(z) &= \frac{\lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y)f(y)dy}{r(x)f(x)} \\ &= \frac{\lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y)f(y)dy}{\lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy},\end{aligned}$$

and

$$G_x(z) = 1 - \frac{\lambda P_0 \overline{B}(x+z) + \lambda \int_{y=0}^{x} \overline{B}(x+z-y)f(y)dy}{\lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy}, \tag{6.32}$$

$$g_x(z) = \frac{\lambda P_0 \overline{b}(x+z) + \lambda \int_{y=0}^{x} \overline{b}(x+z-y)f(y)dy + \lambda \overline{B}(z)f(x)}{\lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy}. \tag{6.33}$$

6.2.14 Expected Nonempty Period

Let $\mathcal{B}_D$:= *nonempty period of the dam*. Then $\mathcal{B}_D = a_0$. Generally, the structure of $\mathcal{B}_D$ differs from that of the busy period $\mathcal{B}$ in the M/G/1 queue given in (3.83), because in M/G/r(·) the efflux rate $r(x)$ varies as x varies. This variation causes the sub-nonempty periods to depend on the beginning ordinate of their initial inputs. For example, in M/G/r(·), a_0 is infinite if $P_0 = 0$, corresponding to the case $r(x) = kx, x > 0, k > 0$, since sample paths decay exponentially between inputs and never decay completely to level 0 (see Example 6.3, Sect. 6.2.11).

Constant Efflux Rate $k > 0$

In the particular case where there is some constant $k > 0$ such that $r(x) \equiv k, x > 0$, the structure of $\mathcal{B}$ given by (3.83) and Fig. 3.6, Sect. 3.4.12, is preserved for $\mathcal{B}_D$, except that the slope of the sample path between inputs is $-k$. Then $0 < P_0 < 1$. Let S := *input size*. In particular, S is the size of the first input of a nonempty period. Let N_S := *number of inputs during the time required for the first* S *to deplete*, i.e., during a time $\int_{y=0}^{S} \frac{1}{r(y)} dy = \int_{y=0}^{S} \frac{1}{k} dy = S/k$ time units. Then

$$\mathcal{B}_D = \frac{S}{k} + \sum_{i=1}^{N_S} \mathcal{B}_{D,i}, \tag{6.34}$$

where $\mathcal{B}_{D,i}, i = 1, \ldots, N_S$ are sub-nonempty periods $\underset{dis}{=} \mathcal{B}_D$, independent of N_S. Taking expected values on both sides of (6.34) gives

$$E(\mathcal{B}_D) = \frac{E(S)}{k} + E(N_S)E(\mathcal{B}_D) = \frac{E(S)}{k} + \lambda \frac{E(S)}{k} E(\mathcal{B}_D), \tag{6.35}$$

since $E(N_S) = \lambda\,(E(S)/k)$. Equation (6.35) gives

$$E(\mathcal{B}_D) = E(a_0) = \frac{E(S)}{k\left(1 - \frac{\lambda}{k}E(S)\right)}. \tag{6.36}$$

Alternative Derivation of $E(\mathcal{B}_D)$

We can obtain P_0 directly from formula (6.21) when $r(x) \equiv k, x > 0$, by dividing by k and integrating both sides with respect to $x \in (0, \infty)$. Since $1 - P_0 = \int_{x=0}^{\infty} f(x)dx$, we get

$$P_0 = 1 - \frac{\lambda}{k} E(S). \tag{6.37}$$

We now use P_0 in (6.37) and the renewal reward theorem. Since $E($*nonempty cycle)* := $E(d_0) = 1/\left(r(0^+)f(0)\right) = 1/\,(\lambda P_0)$, we get

$$\frac{E(\mathcal{B}_D)}{E(d_0)} = 1 - P_0$$
$$E(\mathcal{B}_D) = \frac{1 - P_0}{r(0^+)f(0)} = \frac{1 - P_0}{\lambda P_0}.$$

Substituting for P_0 from (6.37) gives

$$E(\mathcal{B}_D) = E(a_0) = \frac{E(S)}{k\left(1 - \frac{\lambda}{k}E(S)\right)}. \tag{6.38}$$

Formula (6.38), derived by LC and the renewal reward theorem, illustrates the usefulness of the formula

$$E(a_0) = \frac{1 - P_0}{\lambda P_0}, \tag{6.39}$$

which also applies to $E(\mathcal{B})$ in M/G/1 queues, as well as to the nonempty period in M/G/r(·) dams where $0 < P_0 < 1$.

6.3 M/M/r(·) Dam

Assume inputs are of size $S \underset{dis}{=} \text{Exp}_\mu$ occurring at a Poisson rate λ. Assume the dam is stable, i.e., $\lambda E(S) < \lim_{x\to\infty} r(x)$ (see formula (6.24)), so the steady-state distribution of content exists.

6.3.1 *Equation for Steady-State PDF of Content*

Substitute $\overline{B}(x - y) = e^{-\mu(x-y)}$, $0 \le y < x$, in Eq. (6.21), resulting in the integral equation for the steady-state pdf of content $f(x)$,

$$r(x)f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy, x > 0, \tag{6.40}$$

$$f(x) = \frac{\lambda}{r(x)} \left(P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy \right), x > 0. \tag{6.41}$$

6.3.2 *Solution of Equation (6.40) for PDF of Content*

Assume $P_0 > 0$. (Recall $P_0 > 0$ iff $\{W(t)\}_{t\ge 0}$ returns to 0, i.e., (6.5) holds.) Applying differential operator $\langle D + \mu \rangle$ to both sides of (6.40), leads to the differential equation for $f(x)$,

$$\frac{f'(x)}{f(x)} = -\frac{r'(x) + \mu r(x) - \lambda}{r(x)}, x > 0,$$
$$\frac{d}{dx}\ln\left(r(x)f(x)\right) = -\mu + \frac{\lambda}{r(x)}, x > 0, \tag{6.42}$$

by transposing $r'(x)/r(x)$ (=$d \ln r(x)/dx$) to the left side and using well-known properties of derivatives and logarithms. The solution of (6.42) is

$$f(x) = \frac{\lambda P_0}{r(x)} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \frac{dy}{r(y)}\right)}, x > 0, \tag{6.43}$$

upon applying the initial condition $r(0^+)f(0) = \lambda P_0$.

Substituting $f(x)$ from (6.43) into the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ gives

$$P_0 = \frac{1}{1 + \lambda \int_{x=0}^{\infty} \frac{1}{r(x)} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \frac{1}{r(y)} dy\right)} dx}. \tag{6.44}$$

As a mild check, let $r(x) = k > 0$. From (6.44)

$$P_0 = 1/\left(1 + \lambda/\left(k\left(\mu - \lambda/k\right)\right)\right) = \left(k\mu - \lambda\right)/\left(k\mu\right) = 1 - \lambda/\left(k\mu\right),$$

which agrees with (6.37), since $E(S) = 1/\mu$. In the M/M/1 queue, $k = 1$, so $r(x) \equiv 1$, $x > 0$. Substituting $r(x) \equiv 1$ in (6.43) and (6.44) gives (3.112) and (3.113) respectively, agreeing with the analogous results for M/M/1.

6.3.3 Sojourn Times and State-Space Levels

Assume $P_0 > 0$. From (6.25) and (6.26) with $x = 0$, we get E(*nonempty* ***cycle***) and E(*nonempty* ***period***) as

$$E(d_0) = \frac{1}{r(0^+)f(0)} = \frac{1}{\lambda P_0},$$
$$\text{and } E(a_0) = E(\mathcal{B}_D) = (1 - P_0)E(d_0) = \frac{1 - P_0}{\lambda P_0},$$

respectively, with P_0 given in (6.44).

In M/M/r(·), all upward jumps are $\underset{dis}{=}\text{Exp}_\mu$. By the memoryless property, the excess of a jump over any level x is also $\underset{dis}{=}\text{Exp}_\mu$. But, generally a_x depends on x. This differs from the M/M/1 queue or M/M/r(·) dam with $r(x) = k > 0$, $x > 0$, where a_x is *independent of* x, and $E(a_x) \equiv E(\mathcal{B})$, and $E(a_x) \equiv E(\mathcal{B}_D)$; respectively. The structure of $\mathcal{B}$ and $\mathcal{B}_D$ guarantees this independence (see formula (3.83)). However, generally In M/M/r(·), $r(x)$ varies with x; so a_x depends on the values of $r(y)$, $y > x$, $x \geq 0$. Nevertheless, we can still determine $E(a_x)$, $E(b_x)$ and $E(d_x)$ as long as we can solve for $\{P_0, f((x)\}_{x>0}$ as in Sects. 6.3.1–6.3.2.

Constant Efflux Rate

When $r(x) \equiv k$, $k > 0$, $x > 0$, the structure of $\mathcal{B}_D$ is similar to that of $\mathcal{B}$ in M/G/1. Thus, from (6.37) and (6.38),

$$P_0 = 1 - \frac{\lambda}{k\mu}$$

$$E(a_x) = E(B_D) = \frac{\frac{1}{\mu}}{k\left(1 - \frac{\lambda}{k\mu}\right)} = \frac{1}{k\mu - \lambda}, x \geq 0.$$

6.4 M/M/r(·) Dam with *r*(*x*) = *kx*

When the efflux rate ***varies directly with content***, $r(x) = kx$, $x > 0$, for some fixed $k > 0$, and $P_0 = 0$ (i.e., the efflux rate is proportional to content). (See Example 6.3 in Sect. 6.2.11).The sample path of $\{W(t)\}_{t\geq 0}$ has a negative exponential shape between input instants, because $r(W(t)) = kW(t) = -dW(t)/dt$, implying that $dW(t)/W(t) = -kdt$, with solution $W(t) = W(\tau_n)e^{-k(t-\tau_n)}$, $\tau_n \leq t < \tau_{n+1}$, $n = 0, 1, 2, \ldots$(see Formula (6.2) in Sect. 6.2.2).

6.4.1 PDF of Content and Its Laplace Transform

Upon substituting $r(x) = kx$ in (6.41) with $P_0 = 0$, we solve for $f(x)$ using Laplace transforms (see Sect. 3.4.4). The Laplace transform of $f(x)$ is

$$\widetilde{f}(s) \equiv \int_{x=0}^{\infty} e^{-sx} f(x)dx, s > 0.$$

In (6.41), multiplying both sides by e^{-sx}, and integrating on $x \in (0, \infty)$ yields

$$\widetilde{f}(s) = \lambda \int_{x=0}^{\infty} e^{-sx} \frac{1}{kx} \int_{y=0}^{x} e^{-\mu(x-y)} f(y) dy dx. \tag{6.45}$$

Taking d/ds on both sides of (6.45) and interchanging the order of integration gives

$$\frac{d}{ds}\widetilde{f}(s) = -\frac{\lambda}{k} \int_{y=0}^{\infty} e^{-sy} f(y) \int_{x=y}^{\infty} e^{-(s+\mu)(x-y)} dx dy.$$

The inner integral is $1/(\mu + s)$, implying the right side is $-(\lambda/k)\widetilde{f}(s)/(\mu + s)$, yielding differential equation

$$\frac{d}{ds}\widetilde{f}(s) + \frac{\lambda}{k}\left(\frac{1}{\mu + s}\right)\widetilde{f}(s) = 0. \tag{6.46}$$

Separation of variables in (6.46), and integration with respect to s gives

$$\widetilde{f}(s) = A(\mu + s)^{-\frac{\lambda}{k}},$$

for some constant A. The identity $\widetilde{f}(s) \equiv \int_{x=0}^{\infty} e^{-sx} f(x) dx$ implies $\widetilde{f}(0^{+}) = \int_{x=0}^{\infty} f(x) dx = 1$ (normalizing condition since $P_0 = 0$). Thus

$$\widetilde{f}(0^{+}) = A\mu^{-\frac{\lambda}{k}} = 1 \text{ and } A = \mu^{\frac{\lambda}{k}}.$$

Hence

$$\widetilde{f}(s) = \left(\frac{\mu}{\mu + s}\right)^{\frac{\lambda}{k}} = \left(\frac{1}{1 + \frac{s}{\mu}}\right)^{\frac{\lambda}{k}} = \left(1 + \frac{s}{\mu}\right)^{-\frac{\lambda}{k}}, s > 0. \tag{6.47}$$

In (6.47) $\widetilde{f}(s)$ is the Laplace transform of a Gamma pdf (e.g., p. 128, Sect. 3.3.1, in [84]); p. 166ff in [97]; p. 109 in [75]), namely

$$f(x) = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right) \mu^{-\frac{\lambda}{k}}} x^{\left(\frac{\lambda}{k}-1\right)} e^{-\mu x} = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} (\mu x)^{\frac{\lambda}{k}-1} e^{-\mu x} \mu, x > 0. \tag{6.48}$$

In (6.48), letting $u = \mu x$ gives

$$\int_{x=0}^{\infty} (\mu x)^{\frac{\lambda}{k}-1} e^{-\mu x} \mu dx = \int_{0}^{\infty} u^{\left(\frac{\lambda}{k}-1\right)} e^{-u} du = \Gamma\left(\frac{\lambda}{k}\right),$$

implying $\int_{x=0}^{\infty} f(x)dx = 1$. The uniqueness of $f(x)$ in (6.48) is guaranteed due to a one-to-one correspondence between $\widetilde{f}(s)$ and its inverse, up to a set of measure 0 (see pp. 13–14 in [95]).

The statistical moments of $f(x)$ about 0, are

$$E(W^n) = (-1)^n \left. \frac{d^n}{ds^n} \widetilde{f}(s) \right|_{s=0}, n = 1, 2, \ldots .$$

The first and second moments are

$$E(W) = \frac{\lambda}{k\mu}, E(W^2) = \frac{\lambda}{k\mu^2}\left(\frac{\lambda}{k} + 1\right).$$

The variance is

$$Var(W) = E(W^2) - (E(W))^2 = \frac{\lambda}{k\mu^2}.$$

Remark 6.6 In the literature the formula for a standard Gamma pdf is often

$$g(x) = \frac{1}{b\Gamma(c)}\left(\frac{x}{b}\right)^{c-1} e^{-\frac{x}{b}}, x > 0,$$

where $b > 0$, $c > 1$ (see p. 109 in [75]), having Laplace transform

$$\widetilde{g}(s) = (1 + bs)^{-c}, s > -\frac{1}{b}.$$

Since $b > 0$, it is sufficient to take $s > 0$. (The significance of $s > 0$ is discussed on pp. 13–14ff in [95].) Setting $b = \frac{1}{\mu}$, $c = \frac{\lambda}{k}$ gives $\widetilde{g}(s) = \widetilde{f}(s)$ in (6.47).

6.4.2 CDF of Content

The steady-state cdf $F(x)$ and pdf $f(x)$ of the content are

$$F(x) = \int_{y=0}^{x} f(y)dy = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} \int_{y=0}^{x} (\mu y)^{\left(\frac{\lambda}{k}-1\right)} e^{-\mu y} \mu\, dy = \frac{\Gamma(\frac{\lambda}{k}, x)}{\Gamma\left(\frac{\lambda}{k}\right)}, x > 0, \quad (6.49)$$

where $\Gamma(\frac{\lambda}{k}, x) = \int_{y=0}^{x} u^{\left(\frac{\lambda}{k}-1\right)} e^{-u} du$ is the *incomplete* Gamma function (e.g., p. 15 in [138]). Thus $F(\infty) = \Gamma(\frac{\lambda}{k}, \infty)/\,\Gamma\left(\frac{\lambda}{k}\right) = 1$. Generally, $F(x)$ in (6.49) cannot be expressed in closed form for finite $x > 0$, but can be evaluated numerically.

6.4.3 Sojourns with Respect to a Level x

We examine next the inter-downcrossing time d_x, and sojourns a_x and b_x. Consider a sample path of $\{W(t)\}_{t\geq 0}$. Referring to Eqs. (6.25), (6.26), and (6.30) above, we get

$$E(d_x) = \frac{1}{r(x)f(x)} = \frac{1}{kxf(x)} = \frac{\Gamma\left(\frac{\lambda}{k}\right)}{kx(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}\mu}, x > 0; \tag{6.50}$$

$$E(a_x) = (1 - F(x))E(d_x) = \frac{\frac{1}{\Gamma(\frac{\lambda}{k})}\int_{y=x}^{\infty}(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}\mu dy}{kx\frac{1}{\Gamma(\frac{\lambda}{k})}(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}\mu}$$

$$= \frac{\int_{y=x}^{\infty}\mu(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}dy}{kx(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}\mu}, x > 0; \tag{6.51}$$

$$E(b_x) = F(x)E(d_x) = \frac{\int_{y=0}^{x}(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}\mu dy}{kx(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}\mu}, x > 0. \tag{6.52}$$

Naturally, $E(a_x)+E(b_x) = E(d_x)$. Quantities $E(d_x)$, $E(a_x)$, $E(b_x)$ can be evaluated numerically and plotted over a range of x in the state space, for any valid triplet of model parameters $\{\lambda, k, \mu\}$ (see Figs. 6.3, 6.4, 6.5, 6.6, and 6.7).

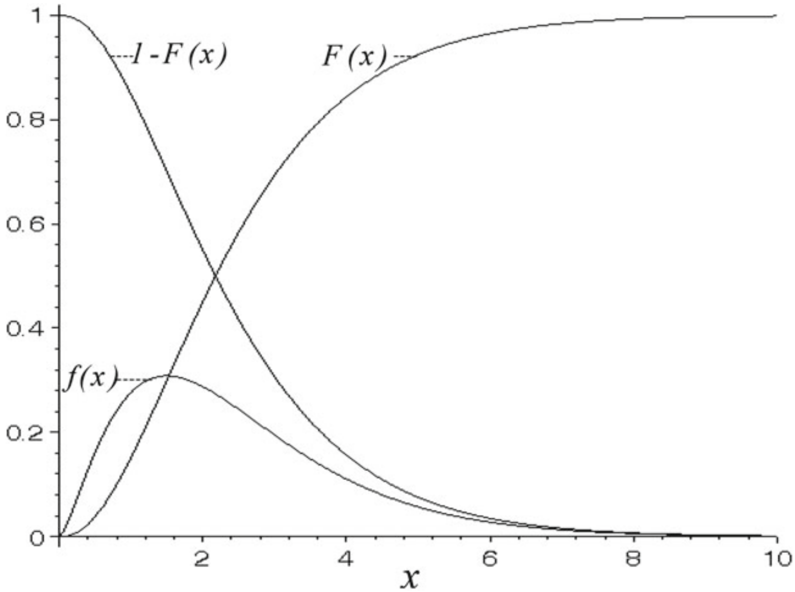


Fig. 6.3 Steady-state pdf $f(x)$, cdf $F(x)$, and complementary cdf $1 - F(x)$, in M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$

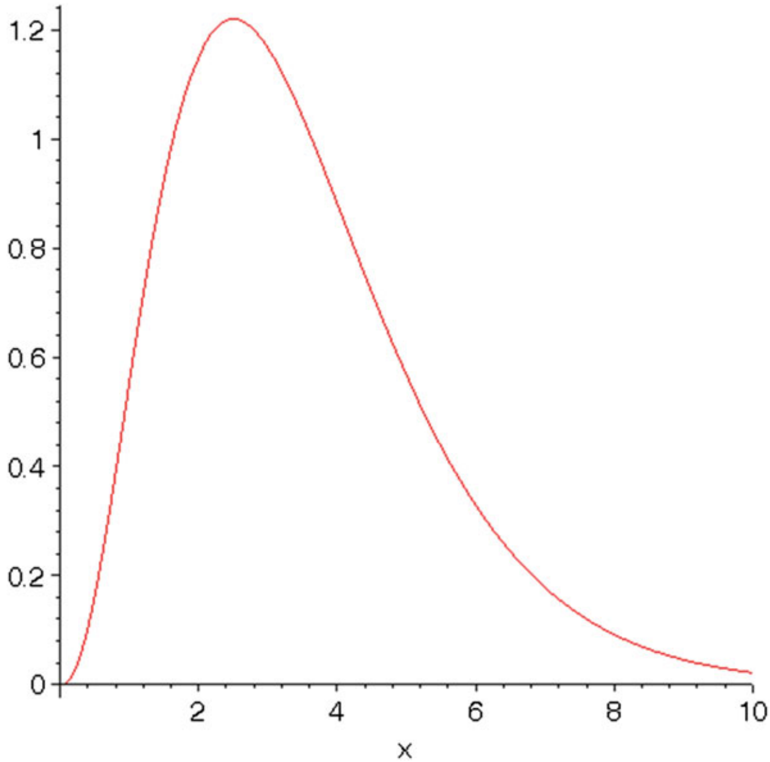


Fig. 6.4 $\mathrm{Lim}_{t\to\infty}\mathcal{D}_t(x)/t$ versus x in M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. $\mathrm{Lim}_{t\to\infty}\mathcal{D}_t(x)/t = \mathrm{Lim}_{t\to\infty}\mathcal{U}_t(x)/t$

Example 6.4 Consider an M/M/r(·) dam with $r(x) = kx$, $x > 0$. (See Figs. 6.3, 6.4, 6.5, 6.6, 6.7 and 6.8.) Set $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. The steady-state pdf of content is

$$f(x) = 0.752253x^{1.5}e^{-x}, x > 0.$$

The cdf of content is, for $x > 0$,

$$F(x) = -0.188063\left(4.0x^{3/2} + 6.0x^{1/2} - 5.317362 \cdot erf(x^{1/2}) \cdot e^{x}\right) \cdot e^{-x},$$

where $erf(x) := \left(2/\sqrt{\pi}\right)\int_0^x e^{-t^2}dt$, the **error function** (see p. 262 in [1]). Because $\mu = 1$ in this example,

$$E(W) = \frac{\lambda}{k\mu} = Var(W) = \frac{\lambda}{k\mu^2} = 2.5.$$

The hazard rate of $f(x)$ is plotted for values of x in Fig. 6.8.

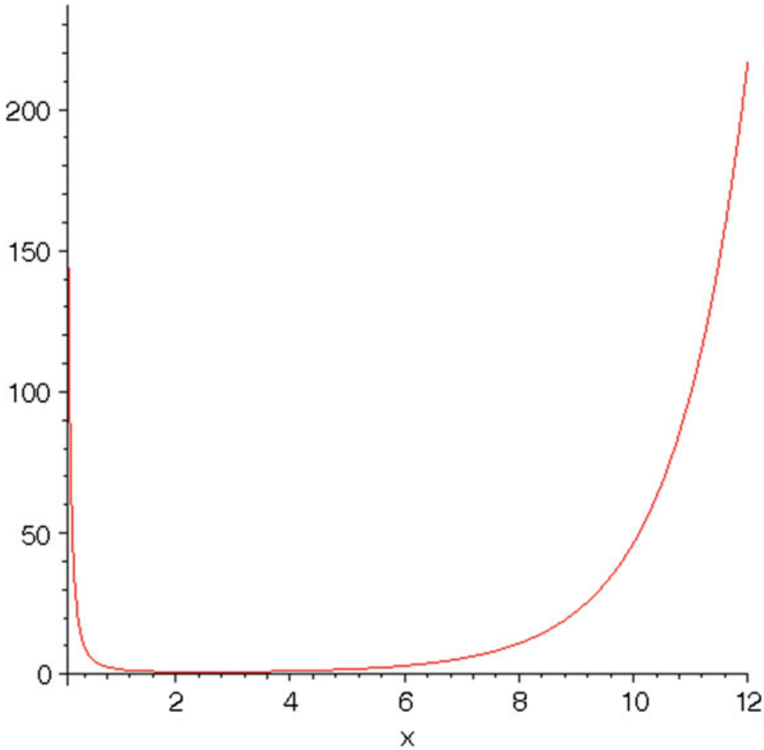


Fig. 6.5 $E(d_x)$ versus x, in M/M/r($\cdot$) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. The "bathtub" shape of $E(d_x)$ is intuitive

6.5 M/M/$r(\cdot)$ with Special Zero-Content Inputs

Assume $P_0 > 0$, and inputs when the dam is empty, have a special size S_0 having cdf $B_0(x), x > 0$, pdf $b_0(x), x > 0$, and $\overline{B_0}(x) = 1 - B_0(x), x \geq 0$. Rate balance across level x, and the law of total probability (normalizing condition), imply

$$\begin{gathered} r(x)f(x) = \lambda P_0 \overline{B_0}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy, x > 0, \\ P_0 + \int_{y=0}^{\infty} f(x)dx = 1. \end{gathered} \tag{6.53}$$

We can solve (6.53) for $\{P_0, f(x)\}_{x>0}$ (analytically or numerically); then obtain $F(x)$, and $E(\mathcal{C}_D)$ (=$E(d_0)$) = $1/\left(r(0^+)f(0)\right)$ using equality $r(0^+)f(0) = \lambda P_0$. Applying the renewal reward theorem, we obtain $E(a_x) = (1 - F(x))/(\lambda P_0)$ and $E(b_x) = F(x)/(\lambda P_0), x \geq 0$. Thus $E(\mathcal{B}_D) = (1 - P_0)\,/\,(\lambda P_0)$. In particular, the first 'input' of

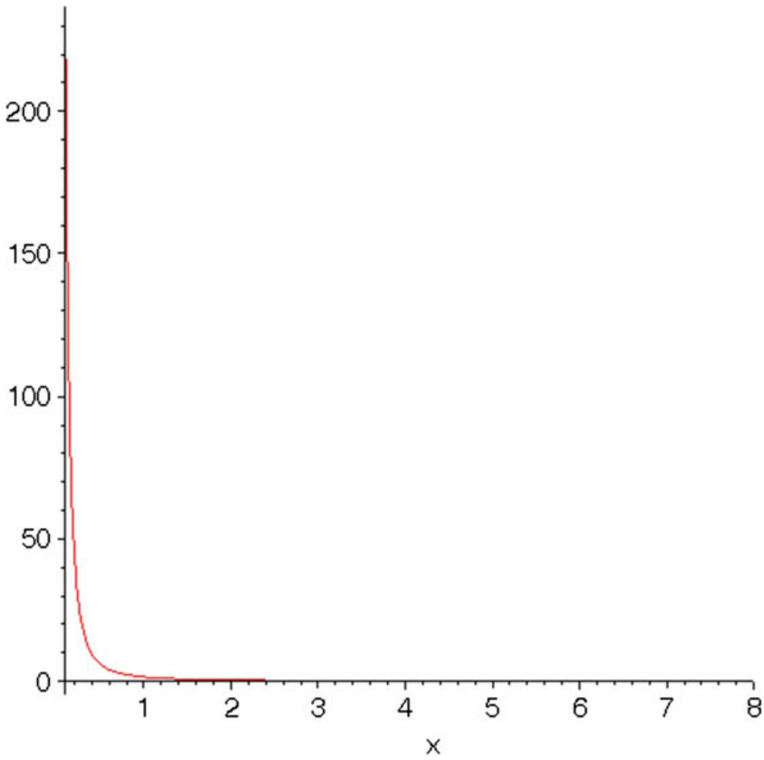


Fig. 6.6 $E(a_x)$ versus x in M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$

a_x is $\underset{dis}{=}\gamma_x$, the excess of a jump over x, distributed differently from all other jumps during a_x. In that case the cdf of γ_x, denoted $G_x(\cdot)$, is given by formula (6.32) in Sect. 6.2.13.

An interesting inference about the structure of $\mathcal{B}_D$ (including a_x and b_x) follows because $\{\mathcal{D}_t(x)\}_{t\geq 0}$ is a renewal process. *Although the structure of* $\mathcal{B}_D$ *generally differs from that of* $\mathcal{B}$ *in M/G/1, or* $\mathcal{B}_D$ *in M/G/r(·)* ($r(x) = k > 0, x > 0$)*, we can always derive* $E(a_x)$ and $E(b_x)$ *once* $F(x)$*,* $x \geq 0$ *and* P_0 are known. Thus, the LC-connected derivation of $E(a_x)$ or $E(b_x)$, is more general than the derivation based directly on structure.

6.6 Generalization of M/G/r(·) Dam

We discuss a generalization of the M/G/r(·) dam considered in Sects. 6.2–6.3.The generalized model allows for SP *downward jumps due to exogenous* events; *state-dependent prescribed jumps* just after the SP *hits or jump-crosses designated state-*

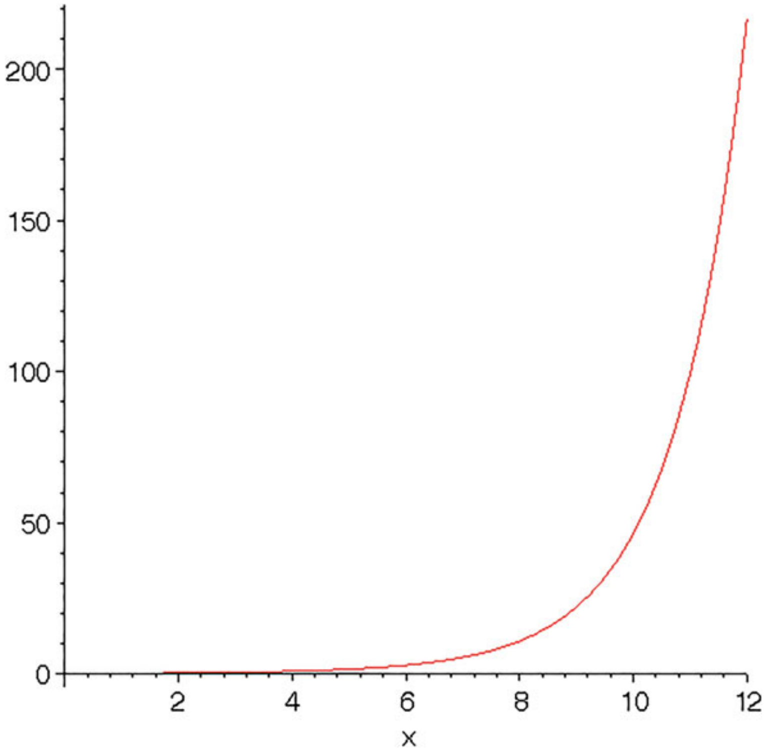


Fig. 6.7 $E(b_x)$ versus x in M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$

space levels, e.g., thresholds or barriers; *specialized state-dependent jumps if an exogenous event occurs when the SP is in a designated state-space interval*; *etc.*

For example, in Marketing Science a target population of repeated advertisements for a product may develop a "rebound" effect against purchasing the product due to "overselling" (i.e., over-advertising). Let $\{W(t)\}_{t\geq 0}$ represent the *consumer response* process for the product, where high measures are favorable, and low measures are unfavorable. The SP may take a sudden jump downward if a new advertisement occurs while the SP is above a 'tolerance' threshold. A sample path of $\{W(t)\}_{t\geq 0}$ would increase in a roughly "saw-tooth", possibly non-linear, pattern, but make exceptional downward jumps from levels above the threshold. (See, e.g., [40].)

A related model applies in multiple dosing of a medication in pharmacokinetics. Suppose the control of a patient's illness depends on lowering to a therapeutic range, systolic blood pressure denoted by $\{BP(t)\}_{t\geq 0}$. The goal of the dosing regime is to maintain $BP(t)$ within a specified finite range, say (L, H) measured in millimeters of mercury (mmHg). This implies the concentration in the blood of the medication, denoted by $\{BC(t)\}_{t\geq 0}$, should be in a corresponding therapeutic range, say (α, β) measured in milligrams per liter (mg/L). If $BC(t)$ upcrosses threshold β, then $BP(t)$

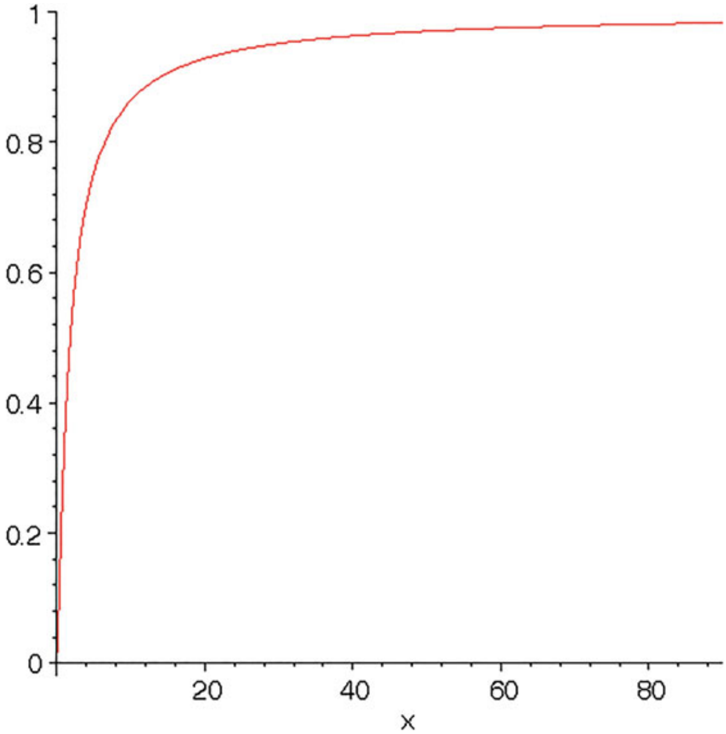


Fig. 6.8 ***Hazard rate*** $\frac{f(x)}{1-F(x)}$ for steady-state distribution of content in M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Note the inverse relation with $E(a_x)$

will drop below threshold L. If $BC(t)$ downcrosses threshold α, then $BP(t)$ will upcross threshold H. A sample path of $\{BP(t)\}_{t\geq 0}$ would move in a roughly "saw-tooth" pattern, inversely emulating the pattern of $\{BC(t)\}_{t\geq 0}$. Similar remarks apply to illnesses depending on blood-thinning medications to protect against strokes (see Sect. 11.6; also [47]).

6.6.1 Model and Steady-State Distribution of Content

Let $\{W(t)\}_{t\geq 0}$ denote the content of a dam with "wide-sense" state space $\boldsymbol{S} \subseteq \mathbb{R}$, which may contain sets having probability 0 (see Sect. 2.3.1). For example, in the standard $\langle s, S\rangle$ inventory, the usual state space is interval $(s, S]$, which supports the probability distribution of inventory. The wide-sense state space is $(-\infty, S]$ because some demands propel the sample path of $\{W(t)\}_{t\geq 0}$ below the reorder point s. Prescribed replenishments then cause the SP to jump immediately up to level S (double jump), so $\{W(t)\}_{t\geq 0}$ spends *zero time* below level s (see Example 2.2 and

Fig. 2.2 in Sect. 2.2.2). The proportion of time $\{W(t)\}_{t\geq 0}$ spends below level s is zero, so the probability of $(-\infty, s]$ is zero.

A particular model may permit jumps due to exogenous events or by *prescription*. Assume that the SP makes *upward* and *downward* jumps at *exogenous* Poisson rates λ_u, λ_d respectively, which are independent of each other and of the current state of the system. Let the corresponding upward and downward jump magnitudes have cdfs $B_u(\cdot)$, $B_d(\cdot)$, and complementary cdfs $\overline{B_u}(\cdot)$, $\overline{B_d}(\cdot)$, respectively. Let $F(\cdot), f(\cdot)$ denote, respectively, the steady-state cdf and pdf of $W(t)$ as $t \to \infty$. Our immediate aim is to derive an integral equation for $f(x)$.

Let the downward jumps occur at instants $0 \equiv \tau_{d,1} < \tau_{d,2} < \cdots$, and upward jumps at instants $0 \equiv \tau_{u,1} < \tau_{u,2} < \cdots$, respectively. Possibly, the SP makes both an upward and downward jump at the same instant (see Sect. 2.3). Without loss of generality, we assume the initial state is $W(0) > 0$. Let $\{\tau_n\}_{n=1,\ldots} = \{\tau_{d,i}\}_{i=1,2,\ldots} \cup \{\tau_{u,i}\}_{i=1,2,\ldots}$.Thus $\{\tau_n\}_{n=1,2,\ldots}$ is a refinement of $\{\tau_{d,i}\}_{i=1,2,\ldots}$ and $\{\tau_{u,i}\}_{i=1,2,\ldots}$. The SP jumps occur at instants $0 < \tau_1 < \tau_2 < \cdots$.

Efflux Rate

The efflux rate $r(W(t)) := dW(t)/dt$ is specified by Eqs. (6.2) and (6.3), above.

Sample Path

A typical sample path of $\{W(t)\}_{t\geq 0}$ is a piecewise continuous function in the time-state plane, which decreases continuously between jumps (see Definition 2.1 in Sect. 2.2.1).

6.6.2 SP Downcrossings

Consider the following types of downcrossings of level x, and their number during $(0, t)$.

$\mathcal{D}_t^c(x) :=$ *number of left-continuous downcrossings of level* x.

$\mathcal{D}_{t,d}^j(x) :=$ number of *jump* downcrossings of level x *at exogenous Poisson rate* λ_d.

$\mathcal{D}_{t,p}^j(x) :=$ *number of state-dependent policy (i.e., prescribed) jump* downcrossings of x, e.g., following hits of a threshold or barrier above x.

$\mathcal{D}_t^j(x) :=$ *total* number of SP downward jumps of x.

Then $\mathcal{D}_t^j(x) = \mathcal{D}_{t,d}^j(x) + \mathcal{D}_{t,p}^j(x)$.

Theorem 6.4

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t^c(x))}{t} = \lim_{t\to\infty} \frac{\mathcal{D}_t^c(x)}{t} \underset{a.s.}{=} r(x)f(x), x \in \boldsymbol{S}, \tag{6.54}$$

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_{t,d}^j(x))}{t} = \lim_{t\to\infty} \frac{\mathcal{D}_{t,d}^j(x)}{t} \underset{a.s.}{=} \lambda_d \int_{y=x}^{\infty} \overline{B_d}(y-x)f(y)dy, x \in \boldsymbol{S}. \tag{6.55}$$

Proof Formula (6.54) follows similarly as in Theorem 6.1 and Corollary 6.2 in Sect. 6.2.8. Formula (6.55) follows as in Theorem 6.2 in Sect. 6.2.8. ■

6.6.3 SP Upcrossings

Consider the following types of upcrossings of level x and their number during $(0, t)$.
$\mathcal{U}^j_{t,u}(x) :=$ *number of jump upcrossings of level x due to the exogenous Poisson rate* λ_u.
$\mathcal{U}^j_{t,p}(x) :=$ *number of prescribed or policy state-dependent jump upcrossings of level* x.
$\mathcal{U}^j_t(x) :=$ *total number of SP jump upcrossings.*

Then $\mathcal{U}^j_t(x) = \mathcal{U}^j_{t,u}(x) + \mathcal{U}^j_{t,p}(x)$. In this model, every upcrossing is a ***jump upcrossing***.

Theorem 6.5

$$\lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,u}(x))}{t} = \lim_{t\to\infty} \frac{\mathcal{U}^j_{t,u}(x)}{t} \underset{a.s.}{=} \lambda_u \int_{y=-\infty}^{x} \overline{B_u}(x-y)f(y)dy, x \in \boldsymbol{S}, \tag{6.56}$$

Proof Similar to proof of Theorem 6.2 in Sect. 6.2.8. ■

Remark 6.7 All three terms in Theorem 6.5 represent the long-run rate of SP upward jumps **due to Poisson rate** λ_u, from state-space set $(-\infty, x]$ into (x, ∞).

6.6.4 Integral Equation for PDF of Content

Applying rate balance across level x, *total downcrossing rate* of x = *total upcrossing rate of* x. Thus

$$\lim_{t\to\infty} \frac{E(\mathcal{D}^c_t(x))}{t} + \lim_{t\to\infty} \frac{E(\mathcal{D}^j_t(x))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,u}(x))}{t} + \lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,p}(x))}{t}. \tag{6.57}$$

Substituting from Theorems 6.4 and 6.5 gives

$$\begin{aligned} &r(x)f(x) + \lambda_d \textstyle\int_{y=x}^{\infty} \overline{B_d}(y-x)f(y)dy + \lim_{t\to\infty} \frac{E(\mathcal{D}^j_{t,p}(x))}{t} \\ &= \lambda_u \textstyle\int_{y=-\infty}^{x} \overline{B_u}(x-y)f(y)dy + \lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,p}(x))}{t}, x \in \boldsymbol{S}. \end{aligned} \tag{6.58}$$

In models where Eq. (6.58) applies, the terms

$$\lim_{t\to\infty} \frac{E(\mathcal{D}^j_{t,p}(x))}{t} \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,p}(x))}{t}$$

are usually expressed in terms of $f(x)$, or as constants. For example, in a standard $\langle s, S\rangle$ inventory model,

$$\lambda_u = 0, \lim_{t\to\infty} \frac{E(\mathcal{D}^j_{t,p}(x))}{t} = 0,$$

$$\text{and } \lim_{t\to\infty} \frac{E(\mathcal{U}^j_{t,p}(x))}{t} = r(s)f(s) + \lambda_d \int_{y=s}^{S} \overline{B}(y-s)f(y)dy,$$

where λ_d is the demand rate (see Sect. 6.8, in which $\lambda_d \equiv \lambda$).

Remark 6.8 Integral Eq. (6.58) can serve as a **template** for variants of the M/G/r(·) dam. We do not solve the equation here. In any related variant, Eq. (6.58) will have a particular form, depending on the model parameters. It can then be solved for $f(x)$ (see Sect. 6.8 below).

$\langle S, S\rangle$ **Inventory**
The $\langle s, S\rangle$ continuous review inventory system is a special case of this model. If there is no lead time and no backlogging, then $r(x) > 0$ for all $x \in (s, S]$. If there is a lead time and backlogging is allowed, then the regular state space and wide-sense state space are both equal to the interval $(-\infty, S]$; also $r(x) = 0$ for $x < s$ (see, e.g., [4]).

In the $\langle s, S\rangle$ model, **prescribed (i.e., policy)** jump upcrossings occur, due to replenishments up to level S whenever the inventory jumps to or below level s, or makes a left-continuous hit of level s from above.

6.7 r(·)/G/M Dam

Consider a dam with a continuous *influx* when the content is positive. The influx is interrupted by "demands" for content (i.e., instantaneous outputs), which occur in an independent Poisson process. The demand sizes are i.i.d. positive random variables, having a common general distribution. If a demand exceeds the current content, the dam becomes empty. Empty periods are exponentially distributed with a common mean, independent of other factors. We may regard an empty period as "setup time" needed before starting a new influx cycle. We call a dam having these properties as an '**r(·)/G/M**' dam, to emphasize the continuous influx rate $r(\cdot)$. The r(·)/G/M dam generalizes the "extended age" process for a G/M/1 queue (Sect. 5.1.1).

The r(·)/G/M dam can be regarded as a template for a variety of production-inventory models where the production rate depends on the current stock level. There are many related variations. We can include: a fixed upper bound on content; thresholds indicating changes in influx rate; several fixed levels at which production may pause for a time; lost sales; backlogging, etc. The r(·)/G/M model is related to the surplus (risk reserve) in a risk model in actuarial science, where the influx is the rate of increase of surplus due to premium payments, and the outputs correspond to claim amounts (see Fig. 2.5 in Sect. 2.2.2, Table 2.1 in Sect. 2.4; Sect. 11.1).

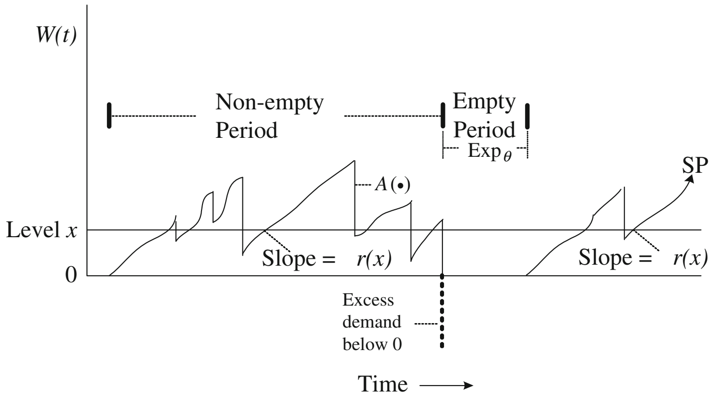


Fig. 6.9 Sample path of standard r(·)/G/M dam

6.7.1 Model Specification and Notation

Let $\{W(t)\}_{t\geq 0}$ denote the content of the dam at time $t \geq 0$. The influx goes on continuously at a positive rate $dW(t)/dt = r(W(t))$, when $W(t) > 0$. Demands for content occur at a Poisson rate μ, and are filled immediately (e.g., a sudden demand for water from a reservoir, oil from a storage tank; a rush order for a product; etc.). The demand sizes are positive with common cdf $A(\cdot)$ and complementary cdf $\overline{A}(\cdot)$. If a demand at t_0^- exceeds the current content, the resulting "content" would be negative. The corresponding end point of the SP downward jump would be below level 0 (Fig. 6.9). Only part of the demand is filled. Various policies can be used regarding the excess demand (e.g., backlogging). To focus on the LC analysis, we shall assume 'no backlogging'. Then the content at t_0 would be $W(t_0) = 0$. It *remains at level* 0 for a time = Exp$_\theta$ independent of the unfilled excess demand below 0. During an empty period $dW(t)/dt = 0$. At the end of an empty period, the content begins to rise from level 0 at rate $r(0^+)$, and continues to rise in a roughly 'saw-tooth pattern' until some future demand takes the content 'below level 0' (see *pass by* in Fig. 2.16). The content alternates between nonempty and empty periods (Fig. 6.9).

If the dam is stable then the content will return to level 0 (state $\{0\}$ is positive recurrent). Denote the *transient* pdf and cdf of content, $t \geq 0$, by $\{P_0(t), f_t(x)\}_{x>0}$ and $F_t(x), x \geq 0$, respectively. Then $P_0(t) = F_t(0)$. Denote the *steady-state* pdf and cdf of content by $\{P_0, f(x)\}_{x,>0}$ and $F(x), x \geq 0$, respectively.

6.7.2 Equation for Transient PDF of Content

Consider a sample path of $\{W(t)\}_{t\geq 0}$. Let $\mathcal{U}_t(x)$, $\mathcal{D}_t(x)$ denote the number of up- and downcrossings of x during $(0, t)$, respectively. In r(·)/G/M sample paths rise steadily at rate $r(x)$ depending on x. Even so we can readily modify the proofs in Theorem 3.4 in Sect. 3.2.7 and Theorem 3.3, in Sect. 3.2.5, getting

$$\begin{aligned}
&\tfrac{\partial}{\partial t}E(\mathcal{U}_t(x)) = r(x)f_t(x), x > 0, t > 0,\\
&\tfrac{\partial}{\partial t}E(\mathcal{U}_t(0)) = r(0^+)f_t(0) = \theta P_0(t), t > 0,\\
&\tfrac{\partial}{\partial t}E(\mathcal{D}_t(x)) = \mu \textstyle\int_{y=x}^{\infty} \overline{A}(y-x)f_t(y)dy, x \geq 0, t > 0.
\end{aligned} \tag{6.59}$$

Consider set $\boldsymbol{A} = [0, x]$, $x \geq 0$, in the state space. Theorem B (see Theorem 4.2 in Sect. 4.2.1 and Theorem 3.2 in Sect. 3.2.4) gives

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}), \tag{6.60}$$

where $\mathcal{I}_t(\boldsymbol{A})$, $\mathcal{O}_t(\boldsymbol{A})$ are the number of SP entrances and exits of $\boldsymbol{A}$ during $(0, t)$, respectively. We assume that a 'pass by' of level 0 due to a downcrossing of level 0 results immediately in an entrance of $\{0\}$. Thus $\mathcal{I}_t(\boldsymbol{A}) = \mathcal{D}_t(x)$, $\mathcal{O}_t(\boldsymbol{A}) = \mathcal{U}_t(x)$, $P_t(\boldsymbol{A}) = F_t(x)$. Substitution from (6.59) into (6.60) results in an *integro-differential* equation for $f_t(x)$ and a *differential* equation for $P_0(t)$, namely

$$\begin{aligned}
\mu \int_{y=x}^{\infty} \overline{A}(y-x)f_t(y)dy &= r(x)f_t(x) + \frac{\partial}{\partial t}F_t(x), x > 0,\\
&= r(0^+)f_t(0) + \frac{\partial}{\partial t}P_0(t) = \theta P_0(t) + \frac{\partial}{\partial t}P_0(t).
\end{aligned} \tag{6.61}$$

The normalizing condition is

$$P_0(t) + \int_{x=0}^{\infty} f_t(x)dx = 1, \text{ for each } t \geq 0.$$

Remark 6.9 In Eq. (6.61), the terms $\partial F_t(x)/\partial t$, $\partial P_0(t)/\partial t$ appear on the opposite side from the integrals, in contrast to Eqs. (6.16) and (6.17). The reason is that the sample path of content **increases** in r(·)/G/M, whereas it **decreases** in the M/G/r(·) dam discussed in Sect. 6.2.9.

6.7.3 *Equation for Steady-State PDF of Content*

If the dam is stable

$$\lim_{t\to\infty} f_t(x) = f(x), \quad \lim_{t\to\infty} \frac{\partial}{\partial t}F_t(x) = 0, x \geq 0, \quad \lim_{t\to\infty} P_0(t) = P_0.$$

We get an integral equation for $f(x)$ by letting $t \to \infty$ in (6.61), implying also a rate-balance equation for state $\{0\}$, viz.,

$$r(x)f(x) = \mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy,\, x > 0,$$
$$r(0^+)f(0) = \mu \int_{y=0}^{\infty} \overline{A}(y)f(y)dy = \theta P_0. \tag{6.62}$$

The normalizing condition is

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{6.63}$$

We can also derive (6.62) directly by considering a sample path of $\{W(t)\}_{t\geq 0}$ (Fig. 6.9). Fix level $x > 0$. The upcrossing rate of level x is $r(x)f(x)$. The downcrossing rate of x is $\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy$. Rate balance across level x gives the first equation in (6.62); the second equation follows by balancing the SP entrance and exit rates of state $\{0\}$.

Remark 6.10 In r(·)/G/M, Eq. (6.62) for the steady-state pdf of content generalizes that for the pdf of "extended" age in the G/M/1 queue, i.e., $r(x)f(x)$ replaces $f(x)$ on the left side of Eq. (5.7) in Sect. 5.1.3.

6.7.4 Sojourn Times Above and Below a Level

Let $a_x :=$ *sojourn time above level* x, $b_x :=$ *sojourn time at or below level* x. Due to Poisson arrivals, the upcrossing instants of level x form a renewal process with $d_x :=$ *interarrival time*, between successive upcrossings, and $E(d_x) = 1/(r(x)f(x)), x \geq 0$.

***E*(a_x)**
By the renewal reward theorem

$$\frac{E(a_x)}{E(d_x)} = 1 - F(x).$$

Also,

$$\frac{E(\mathcal{B}_D)}{E(d_0)} = \frac{E(a_0)}{E(d_0)} = 1 - F(0).$$

Thus,

$$E(a_x) = \frac{1-F(x)}{r(x)f(x)}, x > 0, \tag{6.64}$$
$$E(\mathcal{B}_D) = E(a_0) == \frac{1-P_0}{r(0^+)f(0)} = \frac{1-P_0}{\theta P_0}. \tag{6.65}$$

where $f(x), f(0)$ and P_0 are the solutions of (6.62), and (6.63); and *empty period* $\underset{dis}{=}$ Exp_θ.

Relating $f(x)$ and $E(a_x)$
From (6.64), the hazard rate of $f(x)$ is

$$\frac{f(x)}{1-F(x)} = \frac{1}{r(x)E(a_x)}, x > 0,$$

$$\text{and } \frac{d}{dx}\ln(1-F(x)) = \frac{-1}{r(x)E(a_x)},$$

$$1-F(x) = Ce^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, \tag{6.66}$$

where C is a constant, evaluated by letting $x \downarrow 0$ in (6.66), resulting in $C = 1 - P_0$. Thus,

$$F(x) = 1 - (1-P_0)e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x \geq 0. \tag{6.67}$$

Taking d/dx in (6.67) gives the pdf

$$f(x) = \frac{1-P_0}{r(x)E(a_x)}e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x > 0. \tag{6.68}$$

From (6.65)

$$f(x) = \frac{E(a_0)\theta P_0}{r(x)E(a_x)}e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x > 0. \tag{6.69}$$

The normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$, gives

$$P_0 = \frac{1}{1 + E(a_0)\theta\int_{x=0}^{\infty}\left(\frac{e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}}{r(x)E(a_x)}\right)dx}; \tag{6.70}$$

from (6.65), another expression for P_0 is

$$P_0 = \frac{1}{1+\theta E(a_0).} \tag{6.71}$$

Formula (6.71) implies that the integral in the denominator of (6.70), $\int_{x=0}^{\infty}\left(\frac{e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}}{r(x)E(a_x)}\right)dx = 1$. This equality is verified by letting $u = \int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy$, whence $du = 1/(r(x)E(a_x))\,dx$, resulting in

$$\int_{u=0}^{\infty} e^{-u}du = \left[-e^{-u}\right]_{u=0}^{\infty} = -0 + e^0 = 1.$$

$E(b_x)$

Similarly, we obtain (using (6.62)),

$$E(b_x) = \frac{F(x)}{r(x)f(x)} = \frac{F(x)}{\mu\int_{y=x}^{\infty}\overline{A}(y-x)f(y)dy}, x \geq 0, \tag{6.72}$$

$$E(b_0) = \frac{P_0}{r(0^+)f(0)} = \frac{P_0}{\mu\int_{y=0}^{\infty}\overline{A}(y-x)f(y)dy} = \frac{P_0}{\theta P_0} = \frac{1}{\theta}. \tag{6.73}$$

Hence,

$$\begin{aligned}
&\frac{f(x)}{F(x)} = \frac{1}{r(x)E(b_x)},\\
&\frac{d}{dx}\ln F(x) = \frac{1}{r(x)E(b_x)},\\
&F(x) = C_1 e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy},\\
&F(x) = P_0 e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy}, x \geq 0,\\
&f(x) = \frac{P_0}{r(x)E(b_x)} e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy}, x > 0.
\end{aligned} \tag{6.74}$$

In (6.74) $F(0) = P_0$ and $F(\infty) = 1$.

Example 6.5 As a mild check on (6.74) let $r(x) \equiv 1, x > 0, \overline{A}(x) = e^{-\lambda x}, x \geq 0$, inter-demand time $\underset{dis}{=}\text{Exp}_\lambda$ (i.e., $\theta = \lambda$), corresponding to an M/M/1 queue (i.e., age in G/M/1 specialized to M/M/1).

In M/M/1 the relevant quantities are: $F(x) = 1 - \frac{\lambda}{\mu}e^{-(\mu-\lambda)x}$, $x \geq 0$; $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$, $x > 0$; $P_0 = 1 - \lambda/\mu$. Then

$$E(b_x) = \frac{F(x)}{f(x)} = \frac{1 - \frac{\lambda}{\mu}e^{-(\mu-\lambda)x}}{\lambda P_0 e^{-(\mu-\lambda)x}}.$$

In (6.74), using $F(0) = P_0$,

$$\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy = \int_{y=0}^{x}\frac{1}{E(b_y)}dy = \int_{y=0}^{x}\frac{f(y)}{F(y)}dy = \ln\left(\frac{F(x)}{F(0)}\right),$$

$$F(x) = P_0 e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy} = P_0 e^{\ln\left(\frac{F(x)}{F(0)}\right)} = P_0\frac{F(x)}{F(0)}$$

$$= 1 - \frac{\lambda}{\mu}e^{-(\mu-\lambda)x}, x \geq 0.$$

6.7.5 k/G/M Dam

Let the influx rate be $r(x) = k, x > 0, k > 0$, and assume the *output sizes* are $\underset{dis}{=}\text{Exp}_\lambda$. Thus $\overline{A}(x) = e^{-\lambda x}$, $x > 0$, $\lambda > 0$ (see Fig. 6.10). Since the inter-output times when $W(t) > 0$ are $\underset{dis}{=}$ Exp_μ, the sojourn time $a_x \underset{dis}{=} a_0$ (nonempty period), and $E\left(a_x\right) = (1 - P_0) / (\theta P_0)$, $x > 0$. From (6.68)

$$f(x) = \frac{E(a_0)\theta P_0}{r(x)E(a_x)} e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy} = \frac{\theta P_0}{k} e^{-\frac{\theta P_0}{k(1-P_0)}x}, x > 0. \tag{6.75}$$

From (6.70)

$$P_0 = \frac{1}{1 + \frac{\theta}{k}\int_{x=0}^{\infty} e^{-\int_{y=0}^{x}\frac{dy}{kE(a_0)}}dx} = \frac{\mu - \lambda k}{\mu - \lambda k + \theta} = \frac{1}{1 + \theta/(\mu - \lambda k)}, \tag{6.76}$$

where the term $(\mu - \lambda k)/(\mu - \lambda k + \theta)$ in (6.76) is derived in formula (6.79) in Sect. 6.7.7 below.

6.7.6 *E*(Nonempty Period)

Assume a nonempty (aka non-empty) period a_0 starts at time τ_0. Let $\tau_1 < \tau_2 < \cdots$, be the times of successive *outputs within* a_0, that occur after τ_0. Let $\tau_1^* = \tau_1$ and

$$\tau_{n+1}^* = \min\{\tau_i > \tau_n^* | 0 < W(\tau_i) < W(\tau_n^*)\}, n = 1, 2, \ldots.$$

The ordinates $\{W(\tau_n^*)\}_{n=1,2,\ldots}$ are *strictly descending ladder points* (Fig. 6.10). (See Example (d), p. 280 in [73]; Fig. 1, p. 192, and pp. 390–394 in [74])). Let I^* be the initial influx amount, i.e., up to the first output (decrement) at τ_1 ($I^* = W(\tau_1^-)$). Let N_{I^*} denote the number of descending ladder points during a_0. Since output sizes are $\underset{dis}{=}\text{Exp}_\lambda$, the memoryless property implies these ladder points are distributed as Poisson "arrivals" in a length I^*, where $E(I^*) = E(\tau_1 - \tau_0) \cdot k = k/\mu$. If the output at τ_1 should empty the dam, then $a_0 = \tau_1 - \tau_0 = I^*/k$. In general,

$$a_0 = \frac{I^*}{k} + \sum_{i=1}^{N_{I^*}} a_{0,i}, \tag{6.77}$$

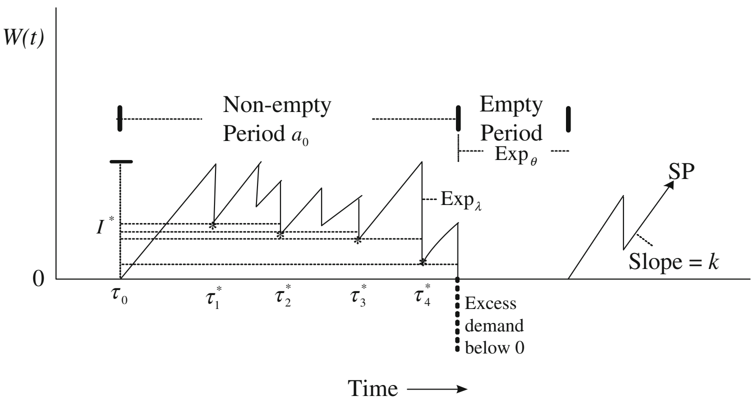


Fig. 6.10 Sample path for r(·)/G/M dam with $r(x) \equiv k$, $\overline{A}(x) = e^{-\lambda x}$. Shows $I^* = k(\tau_1 - \tau_0)$, and descending ladder points at $\tau_1^*, \ldots, \tau_4^*$. The indicated ladder point ordinates are equivalent to four Poisson arrivals (rate λ) within length I^*

where the $a_{0,i}$s are i.i.d. r.v.s $\underset{dis}{=} a_0$, independent of N_{I^*} (see Fig. 6.10 and Sect. 3.4.12).

From (6.77)

$$
\begin{aligned}
E(a_0) &= \frac{E(I^*)}{k} + E(N_{I^*}) \cdot E(a_0) \\
&= \frac{1}{\mu} + \lambda E(I^*) \cdot E(a_0) \\
&= \frac{1}{\mu} + \lambda \frac{k}{\mu} \cdot E(a_0), \\
E(a_0) &= \frac{\frac{1}{\mu}}{1 - \frac{\lambda k}{\mu}} = \frac{1}{\mu - \lambda k}.
\end{aligned} \tag{6.78}
$$

6.7.7 Probability of Emptiness and PDF of Content

We compute $\{P_0, f(x)\}_{x>0}$. Since $E(\textit{empty period}) = E(b_0) = 1/\theta$, from (6.78)

$$
P_o = \frac{E(b_0)}{E(b_0) + E(a_0)} = \frac{\frac{1}{\theta}}{\frac{1}{\theta} + \frac{1}{\mu - \lambda k}} = \frac{\mu - \lambda k}{\mu - \lambda k + \theta}. \tag{6.79}
$$

Substituting for P_0 into (6.75) gives

$$
f(x) = \frac{\theta(\mu - \lambda k)}{k(\mu - \lambda k + \theta)} e^{-\frac{(\mu - \lambda k)}{k} x}, x > 0. \tag{6.80}
$$

6.8 $\langle S, S \rangle$ Inventory with Product Decay

Consider a continuous review $\langle s, S \rangle$ inventory system with reorder point $s \geq 0$, and order-up-to level $S > s$. Assume that demands for stock occur at a Poisson rate λ. The demand quantities, $D_i, i = 1, 2, \ldots$, are i.i.d. random variables with common cdf $B(x)$, and $\overline{B}(x) = 1 - B(x), x \geq 0$. Denote the stock on hand process by $\{I(t)\}_{t \geq 0}$. Assume the stock decays at rate $dI(t)/dt = -r(I(t)) < 0$ when the stock is at level $I(t) \in (s, S]$. The ordering policy is as follows. If the stock either decays

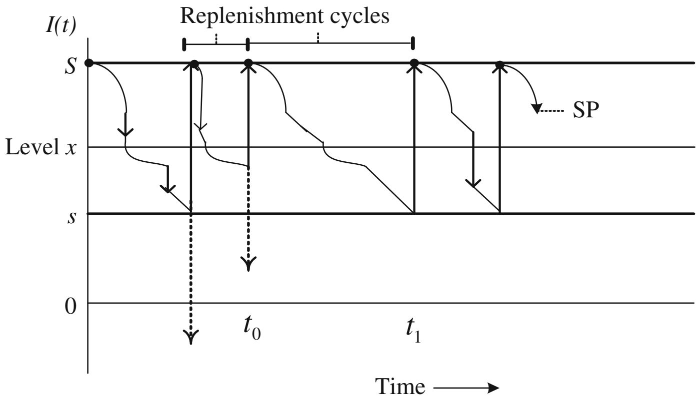


Fig. 6.11 Sample path $\{I(t)\}_{t\geq 0}$ in $\langle s, S\rangle$ inventory with general deacy

continuously to level s, or jumps downward to, or below level s due to a demand, then an order is placed and received immediately, replenishing the stock up to level S. All SP *upward jumps* end at level S. The *regular* state space is $(s, S]$ since all probability is concentrated on $(s, S]$. The *wide-sense* state space is $(-\infty, S]$, which accounts for SP downward jumps ending below s, and immediately jumping up to end at S (double jumps). The latter upward SP jumps correspond to replenishments (see Figs. 6.11 and 2.2 and Example 2.2). In order to focus on the LC analysis, we assume there is no lead time or backlogging. These extensions, as well as others, can be incorporated into the analysis (e.g., [4, 31]).

Let $f(x)$, $s < x \leq S$, and $F(x)$, $x \leq S$, denote, respectively, the steady-state pdf and cdf of $I(t)$ as $t \to \infty$. Assume each order size $\underset{dis}{=}$ Exp$_\mu$.

6.8.1 PDF of Inventory with Constant Decay Rate k

To focus on the LC solution technique, we let the rate of decay be $r(x) := k > 0$, $x \in (s, S]$, and assume demand sizes are $\underset{dis}{=}$Exp$_\mu$. We derive an integral equation for $f(x), x \in (s, S]$ in the following section. (The decay rate is generalized to $r(x) = kx$ in [36].)

This $\langle s, S\rangle$ model is a special case of the generalized M/G/r(·) dam in Sect. 6.6, with demand rate $\lambda_d \equiv \lambda$, the rate of sample-path downward jumps. The cdf of the demand sizes is $B_d(x) \equiv B(x) = 1 - e^{-\mu x}$, $x > 0$. The upward jump rate due to *exogenous* factors is $\lambda_u \equiv 0$. In this standard $\langle s, S\rangle$ model all upward jumps are due to replenishments (*prescribed, i.e., policy* jumps). (The decay rate is generalized to $r(x) = kx$ in [36].)

6.8.2 Equation and Solution for PDF of Inventory

We develop an integral equation for $f(x)$, $x \in (s, S]$. Consider a sample path of $\{I(t)\}_{t\geq 0}$ (similar to Fig. 6.11 with slope $= -k$). Fix level $x \in (s, S)$. The rate at which the SP *decays* into level $x \in (s, S)$ from above (due to left-continuous strict downcrossings of x) is

$$\lim_{t\to\infty} \frac{E\left(\mathcal{D}_t^c(x)\right)}{t} = r(x)f(x) = kf(x).$$

(We use the terms "rate" and "expected rate" synonymously when they are equal $a.s.$). The SP decay rate into level s is

$$\lim_{t\to\infty} \frac{E\left(\mathcal{D}_t^c(s)\right)}{t} = r(s^+)f(s^+) \equiv r(s)f(s) = kf(s).$$

The rate at which the SP *jump-downcrosses* level $x \in [s, S)$ due to demands is

$$\lim_{t\to\infty} \frac{E\left(\mathcal{D}_t^j(x)\right)}{t} = \lambda \int_{y=x}^{S} \overline{B}(y - xf(y)dy = \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y)dy$$

(jumps start at $y \in (x, S)$ and are greater than $y - x$). The *total* SP *downcrossing* rate of level $x \in (s, S]$ is

$$\lim_{t\to\infty} \frac{E\left(D_t^c(x)\right)}{t} + \lim_{t\to\infty} \frac{E\left(\mathcal{D}_t^j(x)\right)}{t} = r(x)f(x) + \lambda \int_{y=x}^{S} \overline{B}(y - xf(y)dy$$
$$= kf(x) + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y)dy, x \in (s, S].$$

The total "downcrossing" rate of the reorder point s is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(s))}{t} = r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y - s)f(y)dy$$
$$= kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy,$$

where we have counted a left-continuous hit of level s from above as a downcrossing of s.

The SP total downcrossing rate of level s is equal to the SP egress rate out of the order-up-to level S *below* (related to right limit tangent below in Fig. 2.15 in Chap. 2). This equality is due to the ordering policy, which replenishes stock on hand to level S with each left continuous hit, or jump downcrossing, of level s. (There is a one-to-one correspondence between SP *egresses from S below*, and *downcrossings of level s*.) Rate balance into and out of state $\{S\}$ results in the equality

$$\begin{aligned} r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y-s)f(y)dy &= r(S)f(S), \\ kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)}f(y)dy &= kf(S), \end{aligned} \tag{6.81}$$

where $r(S) := r(S^-)$ and $f(S) := f(S^-)$.

A crucial simplifying feature of this model is: the total SP *upcrossing* rate of *every* level $x \in (s, S]$ *is equal to* the total *downcrossing* rate of level s (replenishment rate). Applying Eq. (6.81), and rate balance across level x, yields an integral equation for $f(x)$

$$\begin{aligned} r(x)f(x) + \lambda \int_{y=x}^{S} \overline{B}(y-x)f(y)dy &= r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y-s)f(y)dy \\ &= r(S)f(S), x \in (s, S], \end{aligned} \tag{6.82}$$

or, since $r(x) \equiv k$,

$$\begin{aligned} kf(x) + \lambda \int_{y=x}^{S} e^{-\mu(y-x)}f(y)dy &= kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)}f(y)dy \\ &= kf(S), x \in (s, S]. \end{aligned} \tag{6.83}$$

The state space has no atoms, i.e., there is no state in which the SP spends a positive time. The probability distribution of stock on hand is concentrated on $(s, S]$. The normalizing condition is

$$\int_{x=s}^{S} f(x)dx = 1. \tag{6.84}$$

6.8.3 Solution of Integral Equation (6.83)

Taking d/dx in (6.83) gives

$$kf'(x) + \mu\left[kf(S) - kf(x)\right] - \lambda f(x) = 0. \tag{6.85}$$

In (6.85), using the second equality in Eq. (6.83) we replace the term $\mu\lambda \int_{y=x}^{S} e^{-\mu(y-x)}f(y)dy$ by $\mu\left[kf(S) - kf(x)\right]$. The term $-\lambda f(x)$ results from taking the derivative under the integral sign. Simplifying (6.85) gives the differential equation

$$f'(x) - \left(\mu + \frac{\lambda}{k}\right) f(x) = -\mu f(S). \tag{6.86}$$

Multiplying both sides of (6.86) by the integrating factor $e^{-(\mu+\frac{\lambda}{k})x}$, integrating with respect to x, and simplifying gives

$$f(x) = \frac{\mu f(S)}{\mu + \frac{\lambda}{k}} + Ce^{-(\mu+\frac{\lambda}{k})x}, \tag{6.87}$$

where C is a constant. Setting $x = S$ in (6.87) yields

$$f(S) = \frac{\mu + \frac{\lambda}{k}}{\frac{\lambda}{k}} \cdot Ce^{-(\mu+\frac{\lambda}{k})S};$$

which substituted back into (6.87) gives

$$f(x) = \frac{\mu k}{\lambda} \cdot Ce^{-(\mu+\frac{\lambda}{k})S} + Ce^{-(\mu+\frac{\lambda}{k})x},\ s < x \leq S. \tag{6.88}$$

Applying the normalizing condition (6.84) leads to

$$\frac{\mu k}{\lambda} \cdot Ce^{-(\mu+\frac{\lambda}{k})S}(S-s) + C\frac{e^{-\left(\mu+\frac{\lambda}{k}\right)S} - e^{-\left(\mu+\frac{\lambda}{k}\right)s}}{\mu+\frac{\lambda}{k}} = 1,$$
$$\text{or } \frac{1}{C\frac{\mu k}{\lambda}e^{-\left(\mu+\frac{\lambda}{k}\right)S}} = (S-s) + \frac{\lambda}{\mu k}\frac{1-e^{-\left(\mu+\frac{\lambda}{k}\right)(S-s)}}{\mu+\frac{\lambda}{k}}.$$

Factoring (6.88) results in

$$f(x) = C\frac{\mu k}{\lambda}e^{-(\mu+\frac{\lambda}{k})S}\left(1 + \frac{\lambda}{\mu k}e^{-(\mu+\frac{\lambda}{k})(S-x)}\right),\ s < x \leq S.$$

Letting $A := C\frac{\mu k}{\lambda}e^{-(\mu+\frac{\lambda}{k})S}$ gives

$$f(x) = A\left(1 + \frac{\lambda}{\mu k}e^{-(\mu+\frac{\lambda}{k})(S-x)}\right),\ s < x \leq S, \tag{6.89}$$

wherein (applying (6.84)),

$$A = \left[(S-s) + \frac{\lambda}{\mu k\left(\mu+\frac{\lambda}{k}\right)}\left(1 - e^{-(\mu+\frac{\lambda}{k})(S-s)}\right)\right]^{-1}.$$

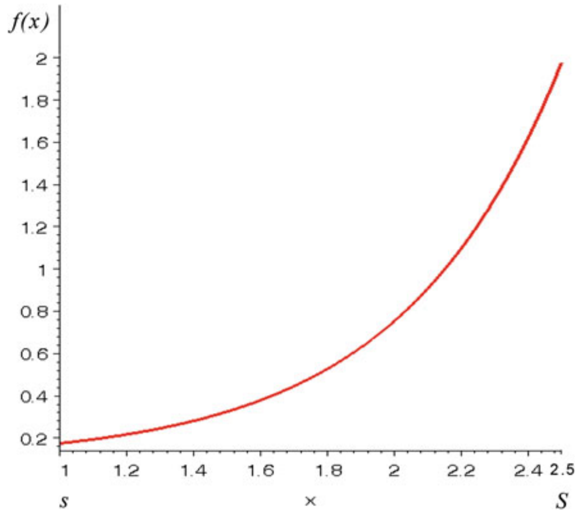


Fig. 6.12 PDF $f(x)$ in $\langle s, S\rangle$ inventory with decay rate k. $\lambda = 2$, $\mu = 0.10$, $k = 1$, $S = 2.5$, $s = 1$. Note that $f(x) = 0 \begin{cases} x < 1, \\ x > 2.5 \end{cases}$

In (6.89) $f(x)$ is convex and increasing on (s, S) (i.e., $f'(x) > 0, f''(x) > 0$). This property agrees with intuition which suggests that the stock resides a large proportion of time at high levels closer to S and a smaller proportion of time near s, for every $k > 0$. This accumulation of inventory near S is a consequence of the re-order policy, which instantaneously replenishes the stock up to level S at each replenishment instant since there is no lead time. (See the numerical example in Sect. 6.8.8 and Figs. 6.12 and 6.13.)

6.8.4 Sojourns Above and Below Level x

Let $a_x :=$ *sojourn time above* x, and $b_x :=$ *sojourn time at or below* x, $x \in (s, S]$, in $\{I(t)\}_{t\geq 0}$. Every instant $t \geq 0$ such that $I(t) = S$ is a regenerative point of $\{I(t)\}_{t\geq 0}$. The regenerative property holds whether replenishments up to S are due to SP smooth decays into level s from above, or due to SP jumps that end at or below level s as a result of a demand. For example, consider Fig. 6.11. At time points like t_1 the SP makes a left-continuous hit of level s from above, and jumps upward to level S. The Poisson arrival process for demands ensures that the excess arrival time until the next demand is $\underset{dis}{=}\text{Exp}_\lambda$ (memoryless property).

Time Between Successive Hits of Level S
The times between successive instants when $I(t) = S$, form a renewal process. Let $d_S :=$ *inter-level-S time*. From (6.83), the total SP downcrossing rate of level s is identical to the

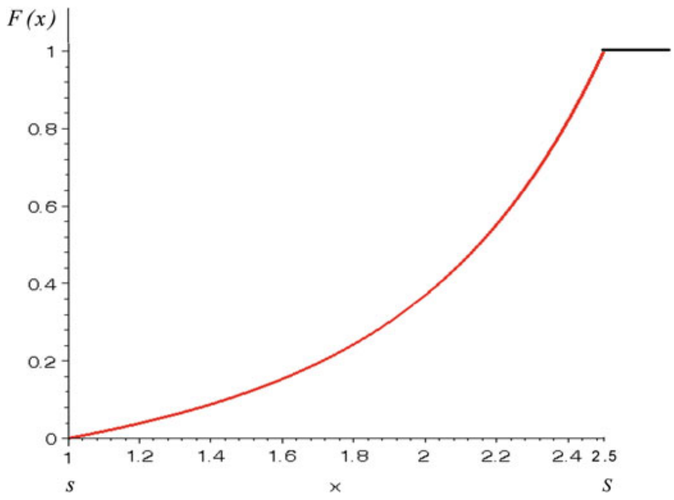


Fig. 6.13 CDF $F(x)$ in $\langle s, S\rangle$ inventory with decay rate k. $\lambda = 2$, $\mu = 0.10$, $k = 1$, $S = 2.5$, $s = 1$

$$\text{replenishment rate } = kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy = kf(S).$$

The value of $f(S)$ is obtained from (6.89). Thus, with A as in (6.89),

$$E(d_S) = \frac{1}{kf(S)} = \frac{\mu}{A(k\mu + \lambda)}, \tag{6.90}$$

***E*(a_x)**
The *proportion* of time the SP spends above level x is $1 - F(x)$, which is equal to $\frac{E(a_x)}{E(d_S)} = E(a_x)kf(S)$ (by the renewal reward theorem). Thus

$$E(a_x) = \frac{1 - F(x)}{kf(S)} = \frac{\mu(1 - F(x))}{A(k\mu + \lambda)}, \tag{6.91}$$

where $F(x) = \int_{y=s}^{x} f(y)dy$ is obtained from (6.89).

***E*(b_x)**
Similarly,

$$E(b_x) = \frac{F(x)}{kf(S)} = \frac{\mu F(x)}{A(k\mu + \lambda)}. \tag{6.92}$$

We can also obtain $E(b_x)$ using (6.91) and

$$E(a_x) + E(b_x) = E(d_S) = \frac{1}{kf(S)}.$$

A check on $E(a_x)$, $E(b_x)$ when $x = s$, using $F(s) = 0$, $1 - F(s) = 1$, and (6.90), is

$$E(a_s) = \frac{1 - F(s)}{kf(S)} = \frac{1}{kf(S)} = \frac{\mu}{A(k\mu + \lambda)} = E(d_S),$$

the expected replenishment cycle, as intuitively expected. Also $E(b_s) = F(s)/kf(S)$ = 0, which agrees with the SP spending zero time below level s. (State-space jumps, including jumps that end below s, occur *not in Time* because they are perpendicular to the time axis. See Remark 2.2 in Sect. 2.3).

6.8.5 Replenishments Due to Two Types of Signal

The replenishment rate (*total ordering*) is the SP total *downcrossing* rate of level s, which is the right side of (6.83), namely $kf(S) = A(k\mu + \lambda)/\mu$ (A as in (6.89)).

Replenishments are made up to level S whenever one of two types of signals occurs. A **type-$\boldsymbol{c}$** signal denotes an SP *left continuous decay into level s from above* (time point t_1 in Fig. 6.11). A **type-$\boldsymbol{j}$** signal denotes an SP *downward jump that ends at or below s, due to a demand* (time point t_0 in Fig. 6.11). (We use the term 'type-$\boldsymbol{k}$ order' to mean an order is due to a type-$\boldsymbol{k}$ signal, $\boldsymbol{k} = \boldsymbol{c}, \boldsymbol{j}$.) An **order cycle** (same as **replenishment** cycle) is the time between two successive instants when an order is *received*, namely, d_S. Due to Poisson arrivals of demands, the sequence $\left\{d_{S,i}\right\}_{i=1,2,...}$ with $d_{S,i} \underset{dis}{=} d_S$, is a renewal process. An order initiating d_S is either type-$\boldsymbol{c}$ or type-$\boldsymbol{j}$. Let $P_{\boldsymbol{c}} := P(\text{an order is type-}\boldsymbol{c})$, $P_{\boldsymbol{j}} := P(\text{an order is type-}\boldsymbol{j})$. Then $P_c + P_j = 1$.

We now determine $P_{\boldsymbol{c}}$ and $P_{\boldsymbol{j}}$. The counting process $\{\mathcal{D}_t^c(s)\}_{t\geq 0}$ is a renewal process due to Poisson arrivals of demand. Since there is exactly 1 type-$\boldsymbol{c}$ or exactly 1 type-$\boldsymbol{j}$ order in an ordering cycle,

$$E(\text{number of type-}\boldsymbol{c}\text{ orders in } d_S) = 1 \cdot P_{\boldsymbol{c}} + 0 \cdot P_{\boldsymbol{j}} = P_{\boldsymbol{c}},$$
$$E(\text{number of type-}\boldsymbol{j}\text{ orders in } d_S) = 0 \cdot P_{\boldsymbol{c}} + 1 \cdot P_{\boldsymbol{j}} = P_{\boldsymbol{j}}.$$

By the renewal reward theorem,

$$\frac{E(\text{number of type-}\boldsymbol{c}\text{ orders in } d_S)}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^c(s))}{t} = r(s)f(s)$$
$$\Longrightarrow \frac{P_{\boldsymbol{c}}}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^c(s))}{t} = r(s)f(s).$$

Since $E(d_S) = 1/\,(r(S)f(S))$,

$$P_{\boldsymbol{c}} = r(s)f(s) \cdot E(d_S) = \frac{r(s)f(s)}{r(S)f(S)}. \tag{6.93}$$

Intuitively, in (6.93) the numerator is the rate of type-$\boldsymbol{c}$ orders; the denominator is the total rate of orders.

Also, $\{D_t^j(s)\}_{,\geq 0}$ is a renewal process (due to Poisson arrivals of demands). Therefore

$$\frac{E(\text{number of type-}\boldsymbol{j}\text{ orders in } d_S)}{E(d_S)} = \frac{1 \cdot P_{\boldsymbol{j}} + 0 \cdot P_{\boldsymbol{c}}}{E(d_S)}$$
$$= \frac{P_{\boldsymbol{j}}}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^j(s))}{t},$$

and (see Sect. 6.6),

$$P_{\boldsymbol{j}} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^j(s))}{t} E(d_S) = \frac{\lambda \int_{y=s}^{S} \overline{B}(y-s) f(y) dy}{r(S) f(S)}. \tag{6.94}$$

Intuitively, in (6.94) the numerator is the rate of type-$\boldsymbol{j}$ orders; the denominator is the total rate of orders.

If $r(x) \equiv k$, $x \in (s, S]$, and the demand sizes are $\underset{dis}{=} \text{Exp}_\mu$, then

$$P_{\boldsymbol{c}} = \frac{kf(s)}{kf(S)} = \frac{f(s)}{f(S)}, P_{\boldsymbol{j}} = \frac{\lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y) dy}{kf(S)}.$$

implying, with A as in (6.89),

$$P_{\boldsymbol{c}} = \frac{k\mu + \lambda e^{-(\frac{k}{\lambda}+\mu)(S-s)}}{k\mu + \lambda}, \tag{6.95}$$

$$P_{\boldsymbol{j}} = \frac{\mu\lambda A \int_{y=s}^{S} e^{-\mu(y-s)} \left(1 + \frac{\lambda e^{-\left(\frac{\lambda}{k}+\mu\right)(S-y)}}{k\mu}\right) dy}{k\mu + \lambda}. \tag{6.96}$$

6.8.6 Expected Order Size

Denote the replenishment order by R. If an order is type-$\boldsymbol{c}$ then $R = S - s$. If an order is type-$\boldsymbol{j}$ then $R = S - s + r_s$ where $r_s :=$ *excess demand below s*. If the demand sizes are $\underset{dis}{=} \text{Exp}_\mu$ then $r_s \underset{dis}{=} \text{Exp}_\mu$ (by memoryless property). Since $P_{\boldsymbol{c}} + P_{\boldsymbol{j}} = 1$, the expected order size is

$$E(R) = (S-s)P_{\boldsymbol{c}} + \left(S - s + \frac{1}{\mu}\right) P_{\boldsymbol{j}} = S - s + P_{\boldsymbol{j}} \frac{1}{\mu}, \tag{6.97}$$

where $P_{\boldsymbol{j}}$ is given in (6.96).

6.8.7 Cost Rate

Since there is no backlogging or lead-time costs in the $\langle s, S\rangle$ model considered here, the cost function accounts only for costs of setup of placing orders, and for holding inventory. Let $\mathcal{C}$:= total average *cost rate*, $\mathcal{C}_s$:= setup cost per type-***c*** order, $\mathcal{C}_j$, := setup cost per type-***j*** order. Let $\mathcal{C}_H$ be the holding cost per unit per unit time. Then

$$\begin{aligned} C &= \mathcal{C}_s \cdot (\text{type-}\boldsymbol{c}\text{ ordering rate}) \\ &\quad + \mathcal{C}_j \cdot (\text{type-}\boldsymbol{j}\text{ ordering rate}) + \mathcal{C}_H \int_{x=s}^{S} xf(x)dx \\ &= \mathcal{C}_s kf(s) + \mathcal{C}_j \int_{x=s}^{S} e^{-(x-s)\mu} f(x)dx + \mathcal{C}_H \int_{x=s}^{S} xf(x)dx, \end{aligned} \tag{6.98}$$

where $f(x)$ is given by (6.89).

6.8.8 Numerical Example

In $\langle s, S\rangle$ with $r(x) \equiv k$ and all demand sizes $\underset{dis}{=} \text{Exp}_\mu$, assume $\lambda = 2$, $\mu = 0.10$, $k = 1, S = 2.5, s = 1$. Calculations give $A = 0.094200$ in (6.89). The pdf of inventory is

$$f(x) = 0.094200 + 1.88400e^{(-5.250+2.10x)},\ 1 < x \le 2.5,$$
$$f(1) = 0.1749, \qquad f(2.5) = 1.9782.$$

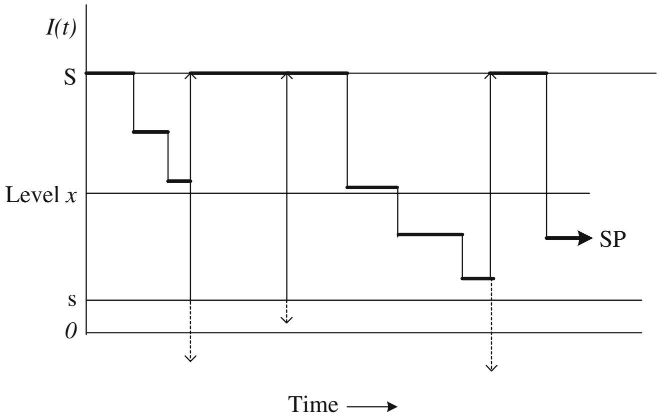


Fig. 6.14 Sample path of $\{I(t)\}_{t\ge 0}$ in $\langle s, S\rangle$ inventory with *no* decay. The SP stays at a level for a time $\underset{dis}{=} \text{Exp}_\lambda$

The cdf of inventory is

$$F(x) = -0.132645 + 0.094200x + 0.897144e^{(-5.25+2.1x)}, 1 < x \leq 2.5,$$
$$F(1) = 0, \qquad F(2.5) = 1.0.$$

Functions $f(x)$ and $F(x)$ are plotted in Figs. 6.12 and 6.13, which demonstrate convexity and the probability massed towards level S.

6.9 ⟨S, S⟩ Inventory with No Product Decay

Consider an $\langle s, S\rangle$ model as in Sect. 6.8, having demand sizes $\underset{dis}{=} \text{Exp}_\mu$ and ***no decay of inventory***. Thus $r(x) \equiv 0$. Once the stock on hand enters a level in $(s, S]$, it remains at that level for a time $\underset{dis}{=} \text{Exp}_\lambda$, until the next demand instant (see Figs. 2.6 and 6.14 above). The state space has an *atom* at level S (positive probability). Every state $\{x\} \in \{y|y \in (s, S)\}$ is continuous (not an atom), because the probability of entering and remaining in such $\{x\}$ for a positive time is 0, due to continuous demand sizes.

Let $\Pi_S = P(\{I(t)\}_{t\geq 0}$ is at level $S)$ in the steady state. Equating the SP down- and upcrossing rates of level $x \in (s, S)$ yields an integral equation for $f(x)$,

$$\lambda\Pi_S e^{-\mu(S-x)} + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y)dy$$
$$= \lambda\Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy = \lambda\Pi_S, \tag{6.99}$$

$$\lambda\Pi_S e^{-\mu(S-x)} + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y)dy = \lambda\Pi_S,\ s < x < S. \tag{6.100}$$

The first equality in Eq. (6.99) indicates that the upcrossing rate of every $x \in (s, S)$ is equal to the downcrossing rate of level s.

Equation (6.100) employs the second equality in (6.99), which expresses the fact *SP rate into* $\{S\}$ = *SP rate out of* $\{S\}$. The normalizing condition is

$$\Pi_S + \int_{x=s}^{S} f(x)dx = 1. \tag{6.101}$$

6.9.1 PDF of Inventory

Some algebra using (6.100) and (6.101), shows that $\{\Pi_S, f(x)\}_{x\in(s,S)}$ is given by

$$\Pi_S = \frac{1}{1+\mu(S-s)}, f(x) = \frac{\mu}{1+\mu(S-s)}, x \in (s, S). \tag{6.102}$$

That is, $f(x)$ is uniformly distributed on state-space interval (s, S), and there is an atom at S. In (6.102), if $\mu < 1$ then $\Pi_S > f(x), x \in (s, S)$; if $\mu > 1$ then $\Pi_S < f(x)$, $x \in (s, S)$. If $\mu = 1$ then $\Pi_S = f(x), x \in (s, S)$.

6.9.2 Sojourn Times Above and Below a Level

The rate of replenishments up to S is the total SP downcrossing rate of level s, namely

$$\lambda \Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy = \lambda \Pi_S,$$

since all replenishments are due to type-$\boldsymbol{j}$ orders that include the excess s. Let $d_S :=$ *time between two successive replenishments up to level S* (same as an ordering cycle). Then

$$E(d_S) = \frac{1}{\lambda \Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy} = \frac{1}{\lambda \Pi_S}. \tag{6.103}$$

Fix level $x \in (s, S]$. From the theory of regenerative processes (specifically the renewal reward theorem)

$$\frac{E(a_x)}{E(d_S)} = 1 - F(x), \qquad \frac{E(b_x)}{E(d_S)} = F(x), x \in [s, S],$$

$$E(a_x) = (1 - F(x)) E(d_S) = \frac{1 - F(x)}{\lambda \Pi_S}, \tag{6.104}$$

$$E(b_x) = F(x) E(d_S) = \frac{F(x)}{\lambda \Pi_S}, \tag{6.105}$$

where

$$F(x) = \int_{y=s}^{x} f(y)dy = \frac{\mu(x-s)}{1+\mu(S-s)}, s < x \le S.$$

The law of total probability gives

$$F(S) = \frac{\mu(S-s)}{1+\mu(S-s)} + \Pi_S = \frac{\mu(S-s)}{1+\mu(S-s)} + \frac{1}{1+\mu(S-s)} = 1.$$

6.9.3 Ordering Characteristics

Ordering Rate

The total ordering rate is the right hand side of (6.100), i.e., $\lambda\Pi_S = \lambda/\left(1+\mu(S-s)\right)$.

Expected Order Size

All orders are type-$\boldsymbol{j}$, signalled by demands ending at or below s. Thus $P_c = 0$ and $P_j = 1$ (see Sect. 6.8.5 and 6.8.6). Hence the expected order size is

$$E(R) = S - s + \frac{1}{\mu}. \tag{6.106}$$

Expected Number of Demands in an Ordering Cycle

Let N_{d_S} := number of demands in an ordering cycle d_S. By the renewal reward theorem,

$$\frac{E(N_{d_S})}{E(d_S)} = \lambda,$$

$$E(N_{d_S}) = \lambda E(d_S) = \frac{\lambda}{\lambda\Pi_S} = \frac{1}{\Pi_S} = 1 + \mu(S-s). \tag{6.107}$$

An alternative derivation of (6.107) using stopping times, is of interest. First,

$$N_{d_S} = \min\left\{n \Big| \sum_{i=1}^{n} D_i > S - s\right\}, \tag{6.108}$$

implying N_{d_S} is a stopping time for the sequence $\{D_i\}_{i=1,2,\ldots}$. Also

$$d_S = \sum_{i=1}^{N_{d_S}} T_i,$$

where $T_i \underset{dis}{\equiv} T \underset{dis}{=} \text{Exp}_\lambda$. N_{d_S} is also a stopping time for the sequence $\{T_i\}_{i=1.2\ldots.}$ because there is a 1–1 correspondence between $\{T_i\}_{i=1.2\ldots.}$ and $\{D_i\}_{i=1,2,\ldots}$. Thus

$$E(d_S) = E(N_{d_S}) \cdot E(T),$$

and

$$E(N_{d_S}) = \frac{E(d_S)}{E(T)} = \frac{\frac{1}{\lambda\Pi_S}}{\frac{1}{\lambda}} = \frac{1}{\Pi_S} = 1 + \mu(S-s), \tag{6.109}$$

corroborating (6.107).

6.9.4 *Cost Rate*

Let $\mathcal{C}_O, \mathcal{C}_H$ denote the setup cost per order and holding cost per order per unit time, respectively. The total average cost rate is

$$C = \mathcal{C}_O \cdot (\text{ordering rate}) + \mathcal{C}_H \int_{x=s}^{S^+} xf(x)dx.$$

The ordering rate is $\lambda\Pi_S = \lambda/ (1 + \mu(S - s))$. The average stock on hand is

$$\begin{aligned}\int_{x=s}^{S^+} xf(x)dx &= S\Pi_S + \int_{x=s}^{S} \frac{\mu x}{1+\mu(S-s)}dx \\ &= \frac{S}{1+\mu(S-s)} + \frac{\mu(S^2-s^2)}{2(1+\mu(S-s))} \\ &= \frac{2S+\mu(S^2-s^2)}{2(1+\mu(S-s))}.\end{aligned}$$

Thus

$$C = \frac{\lambda}{1+\mu(S-s)} \cdot C_O + \frac{2S+\mu(S^2-s^2)}{2(1+\mu(S-s))} \cdot C_H. \tag{6.110}$$

Remark 6.11 In the standard $\langle s, S\rangle$ with **no decay**, suppose the inter-demand times form a **renewal process** (not necessarily a Poisson process). Then the results will be the same as in this section, except for the **arrival-point mixed pdf** denoted $\left\{\Pi_{S,\iota}, f_\iota(x)\right\}_{x\in(s,S)}$, because PASTA will not apply. The integral equation for the pdf $f_\iota(x)$ would be the same as (6.100), where λ represents the renewal rate of the demand process. The arrival rate λ cancels out of the equation. Thus the formulas for $\Pi_{S,\iota}$ and $f_\iota(x)$, $x \in (s, S)$ in (6.102) are independent of λ. The underlying reason for this property is that all orders are type-***j*** at the ends (and starts) of inter-demand times. When the SP jumps up to level S, the time until the next demand is a **full inter-arrival time**. Hence $\left\{d_{S,i}\right\}$ is a renewal process, where $d_{S,i} \underset{dis}{\equiv} d_S$.

Remark 6.12 For exposition, we have applied LC to only two basic $\langle s, S\rangle$ inventory systems. We emphasize that LC equally applies to a vast array of other inventory systems as well, e.g., $\langle r, nQ\rangle$, variations of *EOQ* models, models with lead time and backlogging, production-inventory models of various complexity, models with a variety of state-dependent control policies (e.g., [3, 4]).

Chapter 7
Multi-dimensional Models

7.1 Introduction

In many stochastic models the system state is defined by n real-valued random variables at time $t \geq 0$, where $n \in \{2, 3, \ldots\}$. Thus, the state space $\boldsymbol{S}$ is a subset of n-dimensional Euclidean space, denoted as $\mathbb{R}^{\boldsymbol{n}}$. (Note that n-dimensional may be abbreviated n-D.) To determine the statistical behavior of the system at a transient time t, or as $t \to \infty$, we must derive the time-t, or limiting (as $t \to \infty$), joint pdf or cdf of the state r.v.s. Here, we use LC to analyze a particular inventory system whose state has two continuous real-valued, non-negative random variables; thus, $\boldsymbol{S} \subseteq \mathbb{R}^{\boldsymbol{2}}$.

7.2 Models with State Space *S* a Subset of $\mathbb{R}^2$

Suppose $\boldsymbol{S} \subseteq \mathbb{R}^{\boldsymbol{2}}$, where $\mathbb{R}^{\boldsymbol{2}} = \{(x, y)\,|x \in \mathbb{R}, y \in \mathbb{R}\}$, and $\mathbb{R}$ is the set of real numbers. We assume the time axis is $\boldsymbol{T} = [0, \infty)$.

Notation 7.1 We denote an n**-dimensional** stochastic model as type-$\boldsymbol{n_{c,d}}$ if c state variables are **mixed** (i.e., having **continuous** components and possibly **some atoms**), and d state variables are **discrete (atoms)**, where $c + d = n$.

Example 7.1 Consider the **system point process** $\{W(t), M(t)\}_{t \geq 0}$ in the general M/M/c queue (Sects. 4.3–4.5). Random variable $W(t)$ is the waiting time at time t, which is **continuous** on $(0, \infty)$ with an **atom** at 0 (i.e., **mixed**). Random variable $M(t)$ is the system configuration, which is **discrete**, with values in a set $\boldsymbol{M}$. Assume $M(t)$:= number of **other servers** occupied at

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_7

the start-of-service instant of a time-t arrival. Then, the system state would consist of two random variables ($t \geq 0$). In the standard M/M/c, the system point process is two-dimensional, described as type-$\mathbf{2_{1,1}}$ (1 continuous, 1 discrete variable). The state space is $\boldsymbol{S} \subseteq \mathbb{R} \times \boldsymbol{M}$ ($\times$:= cross product). Similar remarks apply if $M(t)$ is a **vector** of d discrete r.v.s.; then the process would be type-$(\mathbf{1} + \boldsymbol{d})_{1,d}$.

Remark 7.1 If the state of a model with two continuous r.v.s and system configuration having a single random variable $M(t) \in \mathbf{M}$, the state space is $\boldsymbol{S} \subseteq \mathbb{R}^2 \times \mathbf{M}$. The model is type-$\mathbf{3}_{2,1}$. Analogous descriptions apply to models with states consisting of c continuous r.v.s, c = 3, 4,..., and system configuration having a single random variable. The model would be type-$(\boldsymbol{c} + \mathbf{1})_{c,1}$, the state space would be $\boldsymbol{S} \subseteq \mathbb{R}^c \times \mathbf{M}$.

In the other chapters of this monograph, LC is used to analyze type-$\mathbf{1}_{1,0}$, type-$\mathbf{1}_{0,1}$, or type-$\mathbf{2}_{1,1}$ queues, inventories, dams, renewal processes, counter models, etc. LC techniques can be applied to analyze type-$\mathbf{2}_{2,0}$ models with state space $\boldsymbol{S} \subseteq \mathbb{R}^2$, or type-$\mathbf{3}_{2,1}$ models with $\boldsymbol{S} \subseteq \mathbb{R}^2 \times \mathbf{M}$, etc. Similar techniques also apply to type-$(\boldsymbol{n} + \boldsymbol{d})_{n,d}$ models with $\boldsymbol{S} \subseteq \mathbb{R}^n$ ($d = 0$) or $\boldsymbol{S} \subseteq \mathbb{R}^n \times \mathbf{M}$, $n = 3, 4 \ldots$.

Here we focus on two variants of a type-$\mathbf{2}_{2,0}$ model with $\boldsymbol{S} \subseteq \mathbb{R}^2$. The idea is to fix a point $(x, y) \in \boldsymbol{S}$, and select a region $\mathcal{R}_{x,y} \subseteq \boldsymbol{S}$ having a boundary $\partial\mathcal{R}_{x,y}$ which is expressible as a function $\phi(x, y)$. That is, $\partial\mathcal{R}_{x,y}$ may be a *level set* $\phi(x, y) = constant$. Alternatively, $\partial\mathcal{R}_{x,y}$ may be defined as the union of sets (paths in $\mathbb{R}^2$) which describe a "curve" in $\boldsymbol{S}$; we call such a piece-wise connected path a two-dimensional level. If such a level is specified (as in Sect. 7.2.2 below), we obtain SP rates across it, ***expressed in terms of the joint pdf*** of the *continuous* (or mixed) state variables, using familiar LC techniques applied previously in this monograph. Specifically, we use the principle of rate balance across the level to formulate integral equations for the steady-state *joint pdf* of the state variables. The integral equations are solved using analytic, simulation or numerical methods.

We will illustrate the LC technique by modelling a two-product inventory system in which the products share the same limited total storage space in a warehouse (Sect. 7.3 below). Before discussing the two-product inventory model, we briefly review some properties of a rectangle (same as interval) in $\mathbb{R}^2$.

7.2.1 Rectangle in $\mathbb{R}^2$

Let $[x_1, y_1]$, $[x_2, y_2]$, $x_i < y_i$, $i = 1, 2$, denote two finite closed *intervals* in $\mathbb{R}$. A closed *rectangle* (*aka interval)* in $\mathbb{R}^2$ is the cross product

$$[x_1, y_1] \times [x_2, y_2] = \{(\alpha, \beta) \,|\alpha \in [x_1, y_1], \beta \in [x_2, y_2]\}\,.$$

Similar definitions apply to open intervals, partially open intervals, etc. For an arbitrary finite interval in $\mathbb{R}$, there are two choices for *each* end point. The end point either belongs to the 1-dimensional interval associated with the edge, or it does not. Since a rectangle in $\mathbb{R}^2$ has 4 edges and 4 vertices, there are $2^4 = 16$ possible combinations of open and closed edges (Fig. 7.1).

We can determine the count of various types of edges by considering whether vertices are "filled in". (see Fig. 7.1.) Column 1 portrays 0 vertices filled-in (1 way)—all 4 edges are open. In column 2, 1 vertex is filled-in (4 ways), 2 edges are half-open and 2 edges are open. In column 3, 2 vertices are filled-in ($\binom{4}{2} = 6$ ways), giving 2 distinct cases: case 1—the filled-in vertices are adjacent, there are, 1 closed edge, 2 half-open edges, 1 open edge (4 ways); case 2—the filled-in vertices are diagonal, all 4 edges are half-open (2 ways). In column 4, 3 vertices are filled in (4 ways), 2 edges are closed, 2 edges are half open. In column 5, all 4 vertices are filled in (1 way), all 4 edges are closed.

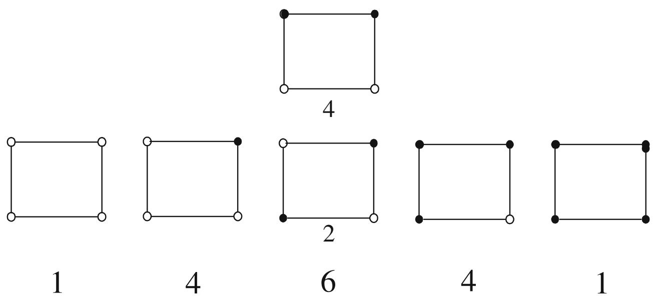


Fig. 7.1 Rectangles in 2-dimensional space. Shows the number of combinations of open, half-open, and closed edges for each number of filled-in vertices, $0, \ldots, 4$; e.g., column 3 shows 2 vertices filled-in: adjacent (4 ways), diagonally (2 ways). The total number of combinations is 16

7.2.2 Two-Dimensional Levels

In many models of Operations Research and related fields, there are two non-negative state variables with mixed pdfs, and the states are *restricted to the non-negative quadrant of* $\mathbb{R}^2$, i.e.,

$$\boldsymbol{S} = \{(x, y) \,|x \ge 0, y \ge 0\}.$$

The LC analysis uses an arbitrary 2-D (two-dimensional) rectangle $\mathcal{R}_{x,y} \subseteq \boldsymbol{S}$ with a fixed corner $(x, y) \in \boldsymbol{S}$ $(x, y > 0)$, as a "test" region with respect to SP entrance and exit rates, where

$$\mathcal{R}_{a,b}^{x,y} = (a, x) \times (b, y), 0 < a < x, 0 < b < y.$$

Rectangle $\mathcal{R}_{a,b}^{x,y}$ is an open set. (Since the state r.v.s have continuous components, the analysis *generally* leads to identical results for joint pdfs, whether choosing an open or closed test rectangle. Care must be taken in some models if the joint pdf takes different forms on subsets of $\boldsymbol{S}$ that are half open, e.g., if one of the variables is the waiting time in an M/D/1 queue (see formulas (3.192) and (3.193) in Sect. 3.10.3).

A Level as a Boundary of the Test 2-D Rectangle

Denote the '*north-east*' boundary of $\mathcal{R}_{a,b}^{x,y}$ as $\rceil_{(a,b)}^{(x,y)}$, so that

$$\rceil_{(a,b)}^{(x,y)} \equiv \{(x, \beta) \,|b < \beta < y\} \cup \{(\alpha, y) \,|a < \alpha < x\}.$$

The boundary $\rceil_{(a,b)}^{(x,y)}$ is the union of the *east* and *north* edges of $\mathcal{R}_{x,y}$ shaped like the letter 'L' rotated 180° to look like '$\rceil$' (see Fig. 7.2, in which $a = b = 0$). The set $\boldsymbol{T} \times \rceil_{(a,b)}^{(x,y)}$ is 'called' a 'level-$\rceil_{(a,b)}^{(x,y)}$ *contour* in $\boldsymbol{T} \times \boldsymbol{S}$', equivalently a '2 *-D level* in $\boldsymbol{T} \times \boldsymbol{S}$', or simply *a* '*level in* $\boldsymbol{S}$'.

The Time Axis and a 2-D Level

The time axis is $\boldsymbol{T} = \{t \in [0, \infty)\}$. The level $\boldsymbol{T} \times \rceil_{(a,b)}^{(x,y)}$ contour in $\boldsymbol{T} \times \boldsymbol{S}$ is the union of two perpendicular planar strips having width $x - a$, $y - b$, extending to infinity along the $\boldsymbol{T}$-axis, in the 3-D (3-dimensional) space $\boldsymbol{T} \times \boldsymbol{S}$. Pictorially, $\boldsymbol{T} \times \rceil_{(a,b)}^{(x,y)}$ contour is a surface resembling an "edge guard" (or edge protector in carpentry), extending to infinity in the direction of $\boldsymbol{T}$. Figure 7.2 plots a $\boldsymbol{T} \times \rceil_{(a,b)}^{(x,y)}$ contour in the 3-D *time-state* space when $a = b = 0$.

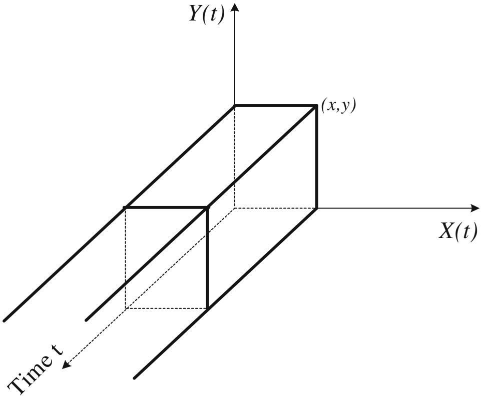


Fig. 7.2 Assume $\{X(t), Y(t)\}, t \geq 0$ is a $2_{(2,0)}$-dimensional process with $(X(t), Y(t)) \in \boldsymbol{R}^2$ restricted to the non-negative quadrant. The level set $\rceil_{(0,0)}^{(x,y)} \equiv \{(x,\beta)\,|0 < \beta < y\} \cup \{(\alpha, y)\,|0 < \alpha < x\}$ is shown over time, and also projected on the $(X(t), Y(t))$ plane. Cross product $\boldsymbol{T} \times \rceil_{(0,0)}^{(x,y)}$ is a surface in 3-dimensional Euclidian space

The Role of Levels in the LC Analysis

Assume system stability holds. Denote the limiting *joint* pdf of the continuous r.v.s by $f(x, y), x \geq 0, y \geq 0$. We compute SP transition rates across level $\rceil_{(a,b)}^{(x,y)}$ in terms of values $\{f(\alpha, \beta)|\text{a crossing of } \rceil_{(a,b)}^{(x,y)} \text{ is possible starting from } (\alpha, \beta)\}$. These transition rates are determined using the probability laws, and any prescribed laws, governing the model dynamics. Rate balance across level $\rceil_{(a,b)}^{(x,y)}$ yields one or more integral equations for $f(x, y)$.

7.3 Two Products Sharing Limited Storage

The model we analyze here is an inventory system with two products that share a common limited storage facility having total capacity Q. Assume that product 1 is governed by an $\langle s, S\rangle$-like ordering policy, and product 2 by an EOQ-like ordering policy. Let $\{I_1(t), I_2(t)\}_{t\geq 0}$ denote the stock on hand at time t of products 1 and 2 respectively. We assume that $I_1(t), I_2(t)$ are continuous r.v.s, therefore type-$2_{2,0}$. Suppose that the parameters are such that the joint process $\{I_1(t), I_2(t)\}_{t\geq 0}$ is stable.

Denote the steady-state joint cdf of $I_1(t), I_2(t)$ as $t \to \infty$ by

$$F(x, y) = \lim_{t\to\infty} P(I_1(t) \leq x, I_2(t) \leq y), (x, y) \in \boldsymbol{S},$$

having joint pdf $f(x, y) = \partial^2 F(x, y)/\partial x \partial y$, wherever the underlying partial derivatives exist.

We will analyze two elementary versions of this model, using LC. We assume no lead times, no backlogging, and no product decay, to further focus on the LC technique. There are many plausible variations on the state space $\boldsymbol{S}$ and on the ordering policies. For elucidation, we analyze the model when the state space $\boldsymbol{S}$ is *relatively* simple, and the ordering policies are well known.

The units of the two products are assumed to be the same as the units of Q. For example, suppose Q is measured in cubic meters (m^3). If the products' units are not the same, we convert them to m^3. Product 1 may consist of 2-m (length) $\times$ 5-cm (outer diameter) plastic pipes which can be cut into continuously variable lengths. Product 2 may be 0.5-in. thick 8-ft $\times$ 4-ft ($8'$ by $4'$) plywood sheets, which can be cut into continuously variable rectangles in the $8'$ by $4'$ sheet. We would convert all volumes to m^3.

An example where the product units are the same as that of Q, is where the products are two different agricultural grains sharing a single storage space, such that they can be retrieved separately to satisfy demands. We will not address the accompanying "packing" problems here. We treat the inventory problem only, and assume that a model with continuous state variables is appropriate.

7.4 Two Products Sharing Storage: Model 1

7.4.1 Policies for the Products

Product 1

Product 1 follows a modified $\langle s, S\rangle$ policy with *no decay* (see Sect. 6.9 above). Assume there is a Poisson demand rate $\lambda > 0$ per unit time and the demand sizes are $\underset{dis}{=} \text{Exp}_\mu$. Its stock on hand remains constant until the next demand for it occurs.

If a demand for product 1 at instant t_0^- causes the stock-on-hand of product 1 to jump downward below a fixed level s, $0 \le s < Q$, an order is placed for product 1, and is filled immediately. The amount received satisfies any shortage caused by the demand and replenishes the stock up to the ***available space*** $Q - I_2(t_0^-)$ at instant t_0.

Product 2

Product 2 follows a modified EOQ policy (see, e.g., p. 210ff in [114]). There is a *constant* demand rate equal to $k > 0$ units per unit time independent of the amount of stock on hand of product 2.

If the stock on hand of product 2 hits level 0 from above at t_0^- an order is placed and received immediately. The order replenishes the stock of product 2 up to the available space $Q - I_1(t_0^-)$ at t_0.

7.4.2 State Space S

The *regular* state space $\boldsymbol{S}_r$ is a finite right-angled triangle with vertices at $(s, 0)$, $(Q, 0)$ and $(s, Q - s)$ (Fig. 7.3). We assume $s \geq 0$. The state space in the *wide sense* is

$$\boldsymbol{S} = \boldsymbol{S}_r \cup \{(\alpha, \beta)\,|\alpha < s, 0 \leq \beta \leq Q - s|\},$$

which appends an infinite rectangular region to the left of $\boldsymbol{S}_r$. We use $\boldsymbol{S}$ since the SP jumps into the infinite rectangular area

$$\{(\alpha, \beta)\,|\alpha < s, 0 \leq \beta \leq Q - s|\}$$

when an order for product 1 occurs (see Sect. 2.3.1 regarding a wide-sense state space).

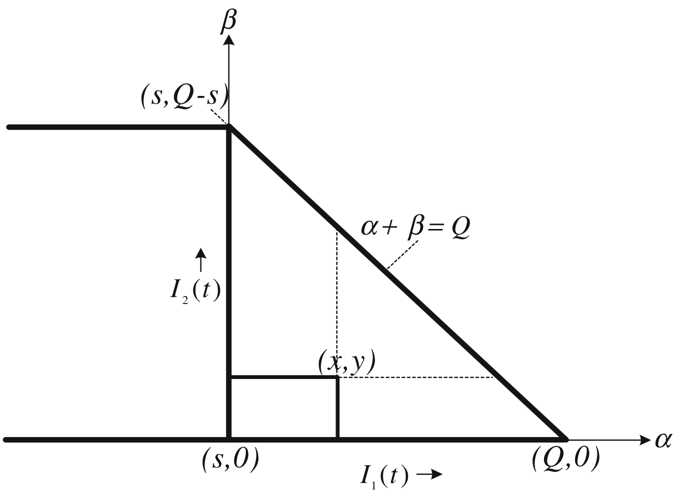


Fig. 7.3 State space $\boldsymbol{S}$ in inventory model with two products sharing limited storage of capacity Q. Shows level $\rceil_{(s,0)}^{(x,y)}$. Indicates two trapezoidal regions of $\boldsymbol{S}$ from which SP can traverse $\rceil_{(s,0)}^{(x,y)}$ due to product demands. $\boldsymbol{S}$ includes infinite rectangle $\{(\alpha, \beta)\,|\alpha < s, 0 \leq \beta \leq Q - s|\}$ appended to $\boldsymbol{S}_r$ on the left

Remark 7.2 There will always be a positive amount of each product on hand, except possibly for an initial finite time period. For, suppose at $t = 0$ the state is $(I_1(0), I_2(0)) = (Q, 0)$ (all space is used for product 1, and no product 2 is present). The state will remain $(Q, 0)$ until a demand for product 1 of some size $D_1 < Q - s$, occurs for the first time. A storage space of size $Q - D_1$ will become available at that instant.

The number of product-1 demands until such a D_1 occurs is geometrically distributed. That is, product 1 may have successive demands of size $> Q - s$ before a demand of size D_1 occurs. The probability of "failure" $= e^{-\mu(Q-s)}$; the probability of "success" $= (1 - e^{-\mu(Q-s)})$. The expected number of demands until obtaining $D_1 < Q - s$ is $1/(1 - e^{-\mu(Q-s)})$. The time between demands has mean $1/\lambda$. Thus, the expected **time** until a size D_1 demand occurs is $1/\left(\lambda(1 - e^{-\mu(Q-s)})\right)$.

We assume that if a product-1 demand of size D_1 occurs, then an order for product 2 of size D_1 is placed and received immediately to fill the space. The resulting state becomes $(Q - D_1, D_1)$. From that instant on, the ordering policies preclude the system ever returning to the state $(Q, 0)$, due to the continuity of the demand sizes of product 1. Thus state $(Q, 0)$ is not recurrent. Therefore $(Q, 0)$ is not an atom (has probability 0).

7.4.3 Sample Path

Figure 7.4 shows a sample path of $\{I_1(t), I_2(t)\}_{t\geq 0}$, containing *vertical* line segments corresponding to the decay of Product-2 with slope k. The vertical segments are traced in planes parallel to the $(t, I_2(t))$-plane, on time intervals $\underset{dis}{=} \text{Exp}_\lambda$ ($\underset{dis}{=}$ *interarrival time between product-1 demands*). The slope-k segments appear as *vertical lines* in the cross-section depicted, which is perpendicular to the time axis in Fig. 7.4. At the lower end points of these line segments, the SP jumps *leftward* due to product-1 demands, to planes closer to the $(t, I_2(t))$-plane, unless it jumps past the reorder plane $\alpha = s$. The leftward jumps occur in planes parallel to the $(t, I_1(t))$ plane with sizes $\underset{dis}{=} \text{Exp}_\mu$. If a leftward jump crosses the plane $\alpha = s$, the SP jumps immediately *rightward* (double jumps), to the order-up-to plane $\alpha + \beta = Q$, corresponding to a product-1 replenishment up to the available space.

If the SP makes a continuous slope-k hit from above of level 0, it jumps immediately *upward* (double jumps) parallel to the $(t, I_2(t))$ plane, the plane

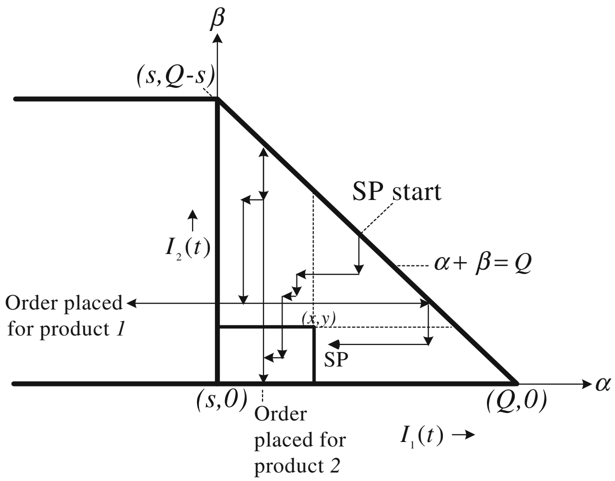


Fig. 7.4 Sample path projected onto $(I_1(t), I_2(t))$ plane in Model 1 of inventory with two products sharing total space Q. The vertical line segments are projections of a line of slope k relative to the $(t, I_2(t))$ plane. The horizontal line segments are projections of horizontal line segments relative to the $(t, I_1(t))$ plane

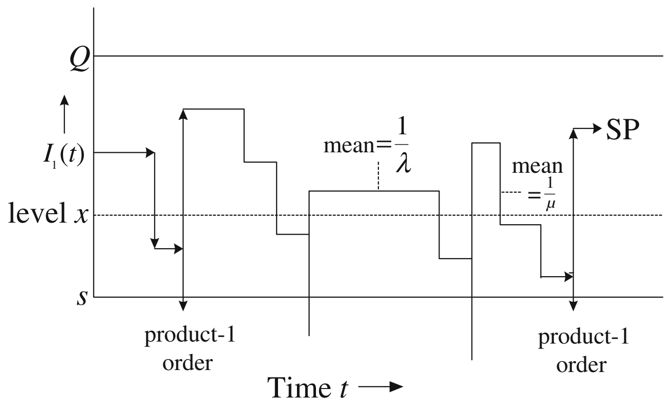


Fig. 7.5 Possible sample path in Model 1 of two-product inventory with limited storage, projected onto the $(t, I_1(t))$ plane. Shows perspective of SP motion for product 1

$\alpha + \beta = Q$, corresponding to a product-2 replenishment up to the available space.

Perspectives of possible sample paths projected onto the $(t, I_1(t))$ -plane and $(t, I_2(t))$-plane are given in Figs. 7.5 and 7.6, respectively.

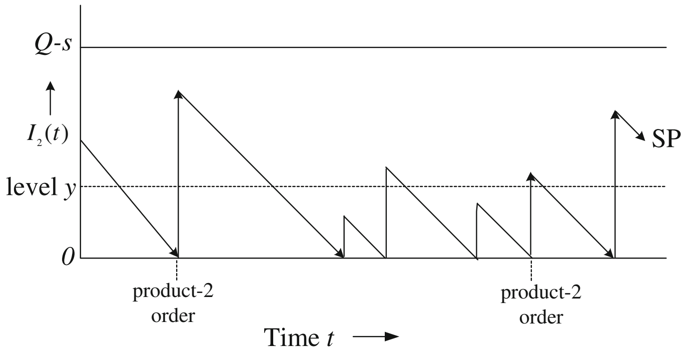


Fig. 7.6 Possible sample path in Model 1 of two-product inventory with limited storage, projected onto the $(t, I_2(t))$ plane. Shows perspective of SP motion for product 2. $I_2(t) \le Q - s$ because product 1 ordering policy implies $I_1(t) \ge s$

7.4.4 Integral Equation for Steady-State Joint PDF

Fix $(x, y) \in \boldsymbol{S}, x > s, y > 0$. Consider the 2-D level $\rceil_{(s,0)}^{(x,y)}$ (Fig. 7.4). The rate at which the SP crosses $\rceil_{(s,0)}^{(x,y)}$ leftward (from the right) is

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)} f(\alpha, \beta) d\alpha d\beta,$$

since all SP leftward crossings are due to jumps corresponding to demands for product 1, having demand rate λ. A jump starting at $\alpha > x$ causes the SP to cross the vertical edge of level $\rceil_{(s,0)}^{(x,y)}$ with probability $e^{-\mu(\alpha-x)}$. Jumps that cross level $\rceil_{(s,0)}^{(x,y)}$ leftward must originate in the trapezoidal region $\{((\alpha, \beta)) \,|x < \alpha < Q - \beta, 0 < \beta < y\}$.

The rate at which the SP downcrosses level $\rceil_{(s,0)}^{(x,y)}$ is

$$k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha,$$

since the demand for product 2 is constant at rate k and the SP downcrossing rate of a point (α, y) is $kf(\alpha, y)$ (see Corollary 6.2). That the SP downcrossing rate is $kf(\alpha, y)$ at (α, y), can be proved by a slight modification of Theorem 1.1 in Chap. 1, such that the SP declines at an arbitrary slope $k > 0$.

Thus, the total SP crossing rate of level $\rceil_{(s,0)}^{(x,y)}$ leftward and downward (from above) is

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)} f(\alpha, \beta) d\alpha d\beta + k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha. \tag{7.1}$$

Similarly the SP crossing rate of level $\rceil_{(s,0)}^{(x,y)}$ rightward (*from the left*) and upward (*from below*) is

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=s}^{Q-\beta} e^{-\mu(\alpha-s)} f(\alpha, \beta) d\alpha d\beta + k \int_{\alpha=s}^{x} f(\alpha, 0) d\alpha, \tag{7.2}$$

by applying the ordering policies for product 1 and product 2. The first term in (7.2) is due to horizontal SP jumps across plane-$(s, 0)$ from the right due to product-1 demands, each of which double-jumps (rebounds) to the right, ending at the plane $\{(\alpha, \beta)\,|\alpha + \beta = Q\}$. The second term in (7.2) is due to SP continuous hits of level 0 from above signalling an instantaneous double jump vertically upward (rebound) to the plane $\{(\alpha, \beta)\,|\alpha + \beta = Q\}$.

Rate balance across level $\rceil_{(s,0)}^{(x,y)}$ equates (7.1) and (7.2), giving the integral equation for $f(x, y), 0 < x + y < Q$.

$$\begin{aligned} &\lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)} f(\alpha, \beta) d\alpha d\beta + k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha \\ &= \lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=s}^{Q-\beta} e^{-\mu(\alpha-s)} f(\alpha, \beta) d\alpha d\beta + k \int_{\alpha=s}^{x} f(\alpha, 0) d\alpha. \end{aligned} \tag{7.3}$$

Form of Solution of (7.3)

Taking $\partial/\partial y$ once and $\partial/\partial x$ twice on both sides of (7.3) leads to a second order PDE (partial differential equation) for $f(x, y)$

$$\begin{aligned} \tfrac{\partial^2}{\partial x \partial y} f(x, y) - \mu \tfrac{\partial}{\partial y} f(x, y) - \tfrac{\lambda}{k} \tfrac{\partial}{\partial x} f(x, y) = 0, \\ s < x < Q - y, 0 < y < Q - s. \end{aligned} \tag{7.4}$$

Applying *separation of variables* for PDEs to (7.4), let $f(x, y) = g(x)h(y)$ (e.g., pp. 422–428 in [10]). Then (7.4) reduces to

$$\frac{g'(x)}{g(x)} = \frac{\mu h'(y)}{h'(y) - \frac{\lambda}{k} h(y)} := \sigma,$$

where the derivatives are taken with respect to the corresponding variables and σ is a constant to be determined. Thus

$$\frac{d \ln g(x)}{dx} = \sigma, \ \frac{d \ln h(y)}{dy} = \frac{\lambda \sigma}{k(\sigma - \mu)},$$

with solutions

$$g(x) = Ae^{\sigma x}, \qquad h(y) = Be^{\frac{\lambda\sigma}{k(\sigma-\mu)}y}, \tag{7.5}$$

where A, B are constants. We next evaluate the constant σ.

Value of Constant σ

We utilize a *boundary* condition to evaluate σ. Consider a point $(x, Q-x)$ on the north-east boundary $\{(\alpha, \beta) \,|\alpha + \beta = Q\}$ of $\boldsymbol{S}$.

In steady state, the SP total rate *into* $(x, Q-x)$ from the left and from below, is

$$\lambda \int_{\alpha=s}^{x} e^{-\mu(\alpha-s)} f(\alpha, Q-x)d\alpha + kf(x,0),$$

where the first term is due to product-1 demands that signal product-1 replenishment orders up to the available space, when product 2 is at level $Q-x$; and the second term is the rate of product 2 demands that signal product 2 replenishment orders up to the available space when product 1 is at level x.

The SP rate *out of* $(x, Q-x)$ leftward and downward, is

$$\lambda f(x, Q-x) + kf(x, Q-x),$$

where the first term is due to product-1 demands when the state is $(x, Q-x)$ and the second term is the SP rate out of $(x, Q-x)$ due to the constant demand rate k for product 2.

Equating the SP rates into and out of line $(x, Q-x)$ gives

$$\begin{aligned} \lambda \int_{\alpha=s}^{x} e^{-\mu(\alpha-s)} f(\alpha, Q-x)d\alpha + kf(x,0) \\ = \lambda f(x, Q-x) + kf(x, Q-x). \end{aligned} \tag{7.6}$$

From (7.5), substituting for $g(\cdot)$, $h(\cdot)$ in (7.6), simplifying and letting $x \downarrow s$ leads to

$$\sigma = \frac{k\ln(1+\frac{\lambda}{k})}{k\ln(1+\frac{\lambda}{k}) + \lambda(Q-s)}, \tag{7.7}$$

provided $\sigma \neq \mu$. (The value $\sigma \neq \mu$; otherwise $h(y)$ would be infinite for all y.)

7.4.5 Solution for Joint PDF of $\{I_1(t), I_2(t)\}_{t\to\infty}$

From (7.5), the steady-state joint pdf of inventory is

$$f(x, y) = Ae^{\sigma x} \cdot Be^{\frac{\lambda\sigma}{k(\sigma-\mu)}y} = ce^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y},$$

where constant $c := AB$, and the constant σ is given in (7.7). The value of c is obtained from the normalizing condition

$$\int_{y=0}^{Q-s}\int_{x=s}^{Q-y} f(x, y)dxdy = \int_{y=0}^{Q-s}\int_{x=s}^{Q-y} ce^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y} = 1,$$
$$\text{implying} \quad c = \frac{1}{\int_{y=0}^{Q-s}\int_{x=s}^{Q-y} e^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y}}. \tag{7.8}$$

Example 7.2 Consider Model 1 with arbitrary parameter values

$$Q = 5, s = 1, \mu = 1, \lambda = 1.5, k = 2.5.$$

Then

$$f(x, y) = ce^{\sigma x + \frac{0.6\sigma}{\sigma-1}y}$$

and from (7.7) $\sigma = 0.1638$. From (7.8) $c = 0.0971$. Thus (Fig. 7.7)

$$f(x, y) = 0.0971e^{0.1638x - 0.1175y}, 1 < x < 5 - y, 0 < y < 4.$$

The marginal pdf of product 1 is $f_1(x) = \int_{y=0}^{Q-x} f(x, y)dy$ or

$$f_1(x) = 0.8266(e^{0.16376x} - e^{0.2813x - 0.5875)})$$

(Figure 7.8). The marginal pdf of product 2 is $f_2(y) = \int_{x=s}^{Q-y} f(x, y)dy$ or

$$f_2(y) = -0.5931(e^{0.1638 - 0.1175y} + e^{0.8188 - 0.2813y})$$

(Figure 7.9).

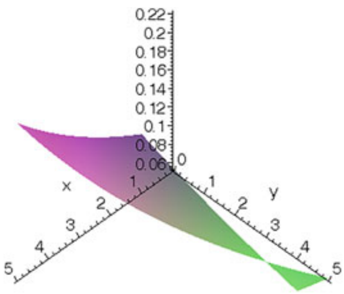


Fig. 7.7 Joint pdf $f(x.y), s < x < Q - y, 0 < y < Q - s$ in model 1 of two product inventory with limited storage: example with $Q = 5, s = 1, \mu = 1, \lambda = 1.5, k = 2.5$

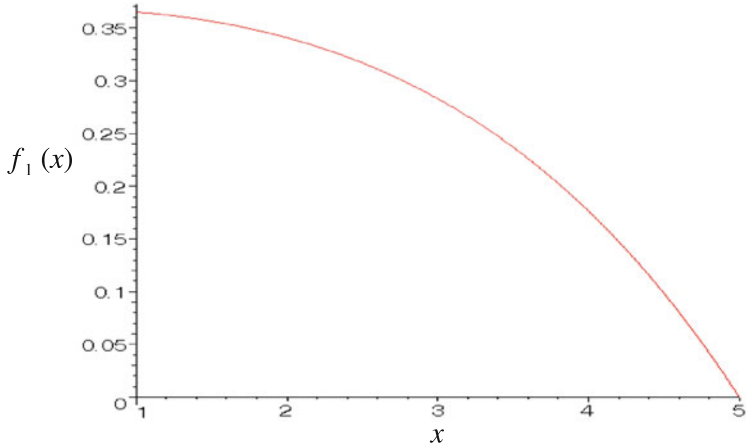


Fig. 7.8 Marginal pdf $f_1(x)$, $s < x < Q$ for product 1 in Model 1 of inventory with two products sharing limited storage: example with $Q = 5$, $s = 1$, $\mu = 1$, $\lambda = 1.5$, $k = 2.5$

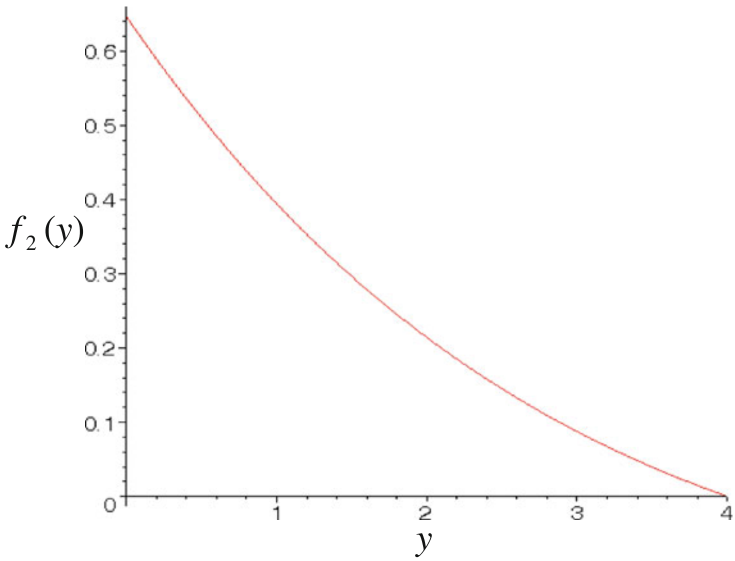


Fig. 7.9 Marginal pdf $f_2(y)$, $0 < y < Q - s$ for product 2 in Model 1 of inventory with two products sharing limited storage: example with $Q = 5$, $s = 1$, $\mu = 1$, $\lambda = 1.5$, $k = 2.5$

Let $\{I_1, I_2\} := \lim_{t\to\infty} \{I_1(t), I_2(t)\}_{t\geq 0}$ denote the limiting state variables for products 1 and 2. The expected values and variances of I_1 and I_2 are respectively

$$E(I_1) = 2.5392, \qquad E(I_2) = 1.1617,$$

$$Var(I_1) = 0.9791, \qquad Var(I_2) = 0.7851.$$

The covariance is $Cov(I_1, I_2) = -0.4457$. The correlation coefficient between I_1 and I_2 is

$$\rho_{I_1,I_2} = \frac{Cov(I_1, I_2)}{\sqrt{Var(I_1)}\sqrt{Var(I_2)}} = -0.5084.$$

Intuitively, we expect ρ_{I_1,I_2} to be negative. That is, if there is a high stock on hand of product i, then there is generally a low stock on hand of product $3-i$, $i = 1, 2$, and vice versa, since $I_1 + I_2 = Q$.

7.5 Two Products Sharing Storage: Model 2

We study a variant (Model 2) of the two-product inventory model in which the products share storage space (see Sect. 7.4.1). This variant has a type-$2_{2,0}$ state with an *atom*. Model 2 places finite limits on the current amounts of each product in storage, so that the total amount in storage at any instant is $\leq Q$. The state space $\boldsymbol{S}$ is such that the derived joint pdf of stock on hand stimulates one's intuition about space-sharing inventory systems, and serves also as a mild check on the (following) method of analysis.

7.5.1 Model 2 Description

Suppose the finite amount of *each* product on hand is restricted to being $\leq$ capacity C $(:= \frac{Q}{2})$. The total amount of stock on hand never exceeds $2C$ $(= Q)$. The "regular" state space is the $(C - s) \times C$ rectangle

$$\boldsymbol{S}_r = \{(\alpha, \beta)\,|s < \alpha \leq C, 0 < \beta \leq C\},$$

in the positive quadrant in Fig. 7.10. The "wide-sense" state space (see Sect. 2.3.1) is

$$\boldsymbol{S} = \boldsymbol{S}_r \cup \{(\alpha, \beta)\,| -\infty < \alpha < s, 0 < \beta < C\},$$

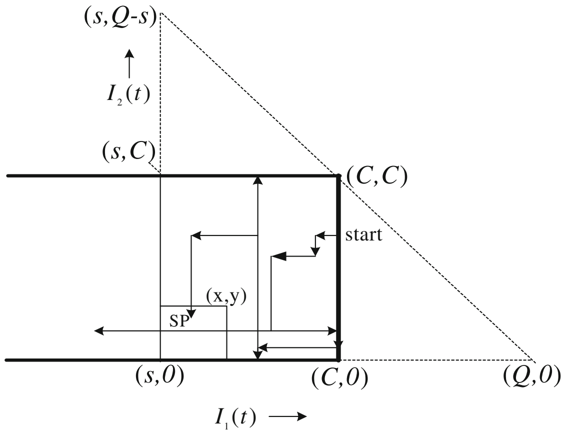


Fig. 7.10 State space $\boldsymbol{S}$ for Model 2, with atom along right edge $\{(\alpha,\beta)\,|\alpha=C, 0<\beta<C\}$. Shows possible sample path projected onto $(I_1(t), I_2(t))$ plane. Also shows level $\rceil_{(s,0)}^{(x,y)}$

the set union of $\boldsymbol{S}_r$ and the infinite rectangle depicted to the left of $\boldsymbol{S}_r$.The ordering policies for products 1 and 2 are the same as in Model 1 (see Sect. 7.4.1). In Model 2, product 1 can have C units in storage for a positive time period $\underset{dis}{=}$ Exp$_\lambda$, which ends with a demand having size $\underset{dis}{=}$ Exp$_\mu$. Product 2 can have C units in storage only for an instant (at a product-2 replenishment instant), since its demand rate is continuously k (see Figs. 7.6 and 7.10). In Model 2 the boundary $\{(C,\beta)\,|0<\beta<C\}$ is an ***atom*** with positive probability, due to the $\langle s,S\rangle$ policy governing product-1 (see Figs. 7.5 and 7.10).

Denote the *steady-state* joint pdf of $\{I_1(t), I_2(t)\}_{t\geq 0}$ by

$$f(x,y), \{(x,y)\,|s<x<C, 0<y\leq C\},$$

and denote the pdf along edge $\{(C,\beta)\,|0<\beta<C\}$ by $\Pi_C(y), 0<y\leq C$. The normalizing condition is

$$\int_{y=0}^{C}\int_{x=s}^{C} f(x,y)dxdy + \int_{y=0}^{C}\Pi_C(y)dy = 1. \tag{7.9}$$

7.5.2 Integral Equation for Joint PDF of Inventory

Fix $(x,y)\in \boldsymbol{S}$, $s<x<C$, $0<y<C$ (Fig. 7.10). Reasoning as in Sect. 7.4.4 in Model 1, the SP leftward rate (from the right) and downward rate (from above) across the 2-D level $\rceil_{(s,0)}^{(x,y)}$ is

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=x}^{C} e^{-\mu(\alpha-x)} f(\alpha,\beta) d\alpha d\beta$$
$$+ \lambda \int_{\beta=0}^{y} e^{-\mu(C-x)} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha.$$

The SP total rate across level $\rceil_{(s,0)}^{(x,y)}$ rightward (from the left) and upward (from below), is

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta) d\alpha d\beta$$
$$+ \lambda \int_{\beta=0}^{y} e^{-\mu(C-s)} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, 0) d\alpha.$$

Rate balance of exits and entries of S_r across 2-D level $\rceil_{(s,0)}^{(x,y)}$ yields the integral equation

$$\begin{aligned} &\lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=x}^{C} e^{-\mu(\alpha-x)} f(\alpha,\beta) d\alpha d\beta \\ &\qquad + \lambda \textstyle\int_{\beta=0}^{y} e^{-\mu(C-x)} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha \\ &= \lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta) d\alpha d\beta \\ &\qquad + \lambda \textstyle\int_{\beta=0}^{y} e^{-\mu(C-s)} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, 0) d\alpha. \end{aligned} \tag{7.10}$$

In (7.10) the sum of the first two right-side terms satisfies

$$\begin{aligned} &\lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta) d\alpha d\beta + \lambda \int_{\beta=0}^{y} e^{-\mu(C-s)} \Pi_C(\beta) d\beta \\ &\quad = \lambda \textstyle\int_{\beta=0}^{y} \Pi_C(\beta) d\beta, \end{aligned} \tag{7.11}$$

since the SP total leftward crossing rate of the $\langle s, S\rangle$ re-order point boundary $\{(s,\beta)|0 < \beta < y\}$, is equal to the SP leftward jump rate out of atomic boundary $\{(C,\beta)\,|0 < \beta < y\}$ due to product-1 demands. Thus (7.10) simplifies to

$$\begin{aligned} &\lambda \textstyle\int_{\beta=0}^{y} \int_{\alpha=x}^{C} e^{-\mu(\alpha-x)} f(\alpha,\beta) d\alpha d\beta \\ &\quad + \lambda \textstyle\int_{\beta=0}^{y} e^{-\mu(C-x)} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, y) d\alpha \\ &= \lambda \textstyle\int_{\beta=0}^{y} \Pi_C(\beta) d\beta + k \int_{\alpha=s}^{x} f(\alpha, 0) d\alpha, \; s < x < C, 0 < y \le C. \end{aligned} \tag{7.12}$$

7.5.3 Solution of Integral Equation

Taking $\partial/\partial y$ once and $\partial/\partial x$ twice in (7.12) leads to the second order PDE

$$\frac{\partial^2}{\partial x \partial y} f(x, y) - \mu \frac{\partial}{\partial y} f(x, y) - \frac{\lambda}{k} \frac{\partial}{\partial x} f(x, y) = 0, \quad s < x < C, 0 < y \leq C.$$

Applying the separation of variables technique for PDEs (as in Sect. 7.4.4) yields the solution

$$f(x, y) = Ae^{\sigma x} \cdot Be^{\frac{\lambda\sigma}{k(\sigma-\mu)} y} = ABe^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)} y}, \tag{7.13}$$

where the constant σ is to be determined.

Taking $\partial/\partial y$ in (7.11) gives

$$\Pi_C(y)(1 - e^{-\mu(C-s)}) = \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha, y) d\alpha, 0 < y \leq C. \tag{7.14}$$

Solution for Constant σ

Consider the rectangular edge $\{(\alpha, C), s < \alpha < C\} \in \boldsymbol{S}$. The SP downcrossing of the edge is $k \int_{\alpha=s}^{C} f(\alpha, C) d\alpha$ due to the constant demand rate k of product 2. A demand for product 1 will not move the SP out of this edge. The SP upcrossing rate of the edge is $k \int_{\alpha=s}^{C} f(\alpha, 0) d\alpha$ due to replenishments of product 2 when it becomes depleted to 0. Rate balance across this edge gives

$$k \int_{\alpha=s}^{C} f(\alpha, C) d\alpha = k \int_{\alpha=s}^{C} f(\alpha, 0) d\alpha. \tag{7.15}$$

Substitute from (7.13) into (7.15) and cancel k and AB from both sides.

If $\sigma = 0$ then $\int_{\alpha=s}^{C} e^{\sigma\alpha} da = C - s > 0$. If $\sigma \neq 0$ then $\int_{\alpha=s}^{C} e^{\sigma\alpha} da = \frac{e^{\sigma C} - e^{\sigma s}}{\sigma} \neq 0$. Thus we may also cancel $\int_{\alpha=s}^{C} e^{\sigma\alpha} da$ from both sides. This leads to the equation for σ

$$e^{\frac{\lambda\sigma}{k(\sigma-\mu)} C} = 1. \tag{7.16}$$

Solving (7.16) for σ gives the value

$$\sigma = 0. \tag{7.17}$$

Solution for AB and Other Constants

From (7.13) and (7.17)

$$f(x, y) = AB, s < x < C, 0 < y \le C.$$

Substituting into (7.14) gives

$$\Pi_C(y)(1 - e^{-\mu(C-s)}) = \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} ABd\alpha,$$

or

$$\Pi_C(y) = \frac{AB}{\mu}, 0 < y \le C.$$

The normalizing condition (7.9) gives

$$AB\left(\int_{y=0}^{C}\int_{x=s}^{C} dxdy + \frac{1}{\mu}\int_{y=0}^{C} dy\right) = 1.$$

Hence

$$AB = \frac{\mu}{C(1 + \mu(C - s))}. \tag{7.18}$$

From (7.18)

$$f(x, y) = \frac{\mu}{C(1 + \mu(C - s))}, s < x < C, 0 < y \le C, \tag{7.19}$$

and

$$\Pi_C(y) = \frac{1}{C(1 + \mu(C - s))}, 0 < y \le C. \tag{7.20}$$

Let $\Pi_C = \int_{y=0}^{C} \Pi_C(y)dy.$

From (7.20)

$$\Pi_C = \frac{1}{(1 + \mu(C - s))}. \tag{7.21}$$

7.5.4 Marginal PDFs of Stock on Hand

From (7.19) the marginal pdf for product 1 in the interval $s < x < C$ is

$$\begin{aligned} f_1(x) &= \int_{y=0}^{C} f(x, y)dy \\ &= \frac{\mu}{(1 + \mu(C - s))}, s < x < C. \end{aligned} \tag{7.22}$$

The complete mixed marginal pdf for product 1 is

$$\{f_1(x); \Pi_C\} = \left\{ \frac{\mu}{(1+\mu(C-s))}, s < x < C; \frac{1}{(1+\mu(C-s))} \right\},$$

where $\int_{x=s}^{C} f_1(x)dx + \Pi_C = 1$.

From (7.19) and (7.20) the marginal pdf for product 2 is

$$\begin{aligned} f_2(y) &= \int_{x=s}^{C} f(x, y)dx + \Pi_C(y) \\ &= \frac{\mu(C-s)}{C(1+\mu(C-s))} + \frac{1}{C(1+\mu(C-s))} \\ &= \frac{1+\mu(C-s)}{C(1+\mu(C-s))} = \frac{1}{C}, 0 < y \le C. \end{aligned} \tag{7.23}$$

Formula (7.22) is identical to (6.102) with the order-up-to level S replaced by C. Intuitively, this result holds because the ordering policy for product 1 is $\langle s, S \rangle$ with no decay, and the state space $\boldsymbol{S}$ is rectangular.

Similarly, (7.23) is uniform on $(0, C]$, which is a well-known result for the stationary distribution in a standard EOQ model.

The motion of the SP in $\boldsymbol{T} \times \boldsymbol{S}$ is affected by orders of both product types. Nevertheless the stock on hand of products 1 and 2 in steady state are statistically independent, corroborated by the relationships between the joint pdf and marginal pdfs, namely

$$\begin{aligned} &f(x, y) = f_1(x) \cdot f_2(y), s < x < C, 0 < y \le C, \\ &\Pi_C(y) = \Pi_C \cdot f_2 C), 0 < y \le C. \end{aligned} \tag{7.24}$$

Remark 7.3 Model 2 serves as a mild check on the LC method for analyzing type-$2_{2,0}$ models. Intuitively we expect statistical independence of the stock on hand of the two products. Indeed, the marginal pdfs turn out as expected for such independence. The stock on hand of each product is independent of the stock on hand of the companion product.

7.5.5 Summary

In this chapter we have used LC to analyze two variants of a model in which two products share the same total storage space. There are many different ordering policies, different types of constraints, and modified state spaces possible for such variants. The model variants would have unique corresponding steady-state joint and marginal pdfs of stock on hand for the products.

We can analyze a vast array of type-$2_{2,0}$ models by applying LC. These include various types of inventory, production-inventory, queuing-network, and natural-science models. A similar remark applies to type-$n_{n,0}$ models, with $n = 3, 4, \ldots$ We can also extend the analysis to type-$n_{c,d}$ models where $c + d = n$ and both $c > 0$, $d > 0$.

Chapter 8
Embedded Level Crossing Method

Much of this chapter is based on [15]. Section 8.4, however, was written by the author for the first edition of this monograph. The ELC (embedded level crossing) method, along with the continuous LC method used in Chaps. 1–7 and later in the monograph, often get results faster than with Lindley recursions (see Sect. 1.2).

8.1 Dams and Queues

Consider a system modelled by $\{W(t)\}_{t\geq 0}$, a continuous-parameter process with state space $\boldsymbol{S} = [0, \infty)$. (The state space can be extended to $\boldsymbol{S} \subseteq \mathbb{R}^{\boldsymbol{n}}$ in more general models.) Let $\{\tau_n\}_{n=1,2,...}$ be an infinite set of embedded *time points* such that

$$0 \leq \tau_1 < \tau_2 < \cdots < \tau_n < \tau_{n+1} < \cdots .$$

Let $\{W_n\}_{n=1,2,...}$ be the embedded discrete-parameter process, where $W(\tau_n^-) \equiv W_n$ and $W(\tau_n) \equiv W_n + S_n$, $n = 1, 2,...$. Assume $W(t)$ is monotone in the interval $[\tau_n, \tau_{n+1})$, and let

$$\frac{dW(t)}{dt} = -r(W(t)), t \in [\tau_n, \tau_{n+1}), n = 1, 2, ...,$$

where $r(x) \geq 0$. Denote the cdf of S_n, $n = 1, 2,...$, by $B(x)$, $x \geq 0$, with $B(0) = 0$, and pdf $b(x) = dB(x)/dx$, $x > 0$, wherever the derivative exists. Denote

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_8

the cdf of W_n by $F_n(x), x \geq 0$, with pdf $dF_n(x)/dx = f_n(x)$, wherever it exists.

8.1.1 Embedded Downcrossings and Upcrossings

Definition 8.1 An **embedded downcrossing** of state-space level x occurs during the closed interval $[\tau_n, \tau_{n+1}]$ if $W_n > x \geq W_{n+1}$.

An **embedded upcrossing** of level x occurs during $[\tau_n, \tau_{n+1}]$ if $W_n \leq x < W_{n+1}$.

Fix level $x \in \boldsymbol{S}$. Definition 8.1 classifies the set of intervals

$$\{[\tau_n, \tau_{n+1}]\}, n = 1, 2, ...$$

into three mutually exclusive and exhaustive subsets with respect to level x (see Fig. 8.1):

1. intervals that contain an embedded downcrossing,
2. intervals that contain an embedded upcrossing,
3. intervals that contain no embedded level crossing.

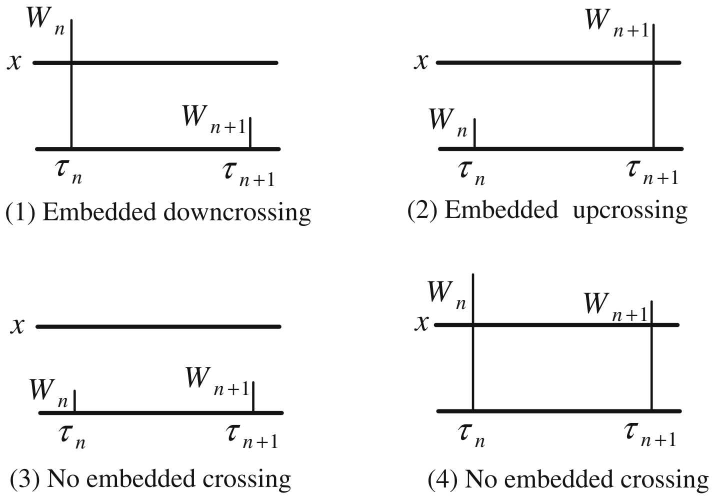


Fig. 8.1 Embedded level crossings and non-crossings during the time interval $[\tau_n, \tau_{n+1}]$

8.1.2 Rate Balance Across a State-Space Level

Consider the time interval $[0, \tau_n]$, $n \geq 2$ and a fixed level $x \in \boldsymbol{S}$. Let $\mathcal{D}_n(x)$, $\mathcal{U}_n(x)$ denote respectively the number of embedded down- and upcrossings of level x during $[0, \tau_n]$. Assume that a typical sample path has an infinite number of embedded time points $\tau_n, n = 1, 2, ...$ with probability 1. The principle of rate balance across level x is

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathcal{D}_n(x)}{n} &= \lim_{n\to\infty} \frac{\mathcal{U}_n(x)}{n} \ (a.s.), \\
\lim_{n\to\infty} \frac{E(\mathcal{D}_n(x))}{n} &= \lim_{n\to\infty} \frac{E(\mathcal{U}_n(x))}{n}.
\end{aligned} \tag{8.1}$$

8.1.3 Method of Analysis

If $\{W(t)\}_{t\geq 0}$ is stable, the steady-state probability distribution of $W(t)$ as $t \to \infty$ and of W_n as $n \to \infty$, exist. Let $f(x) = \lim_{n\to\infty} f_n(x)$, $F(x) = \lim_{n\to\infty} F_n(x)$, $x \in \boldsymbol{S}$. In the following sections, we shall derive an integral equation for $f(x)$ and $F(x)$ by using only:

1. the concept of embedded level crossings,
2. the principle of rate balance,
3. properties of the model,
4. knowledge of the efflux function $r(x), x \geq 0$.

8.2 GI/G/r(·) Dam

Assume inputs to the dam occur in a renewal process with inter-input times having common cdf $A(\cdot)$. The model description is the same as for the M/G/r(·) dam in Sect. 6.2.1, except for the general renewal input stream here.

The embedded process $\{W_n\}_{n=1,2,...}$ is a Markov chain, since

$$W_{n+1} = \max\{W_n + S_n - \Delta_n, 0\}$$

where S_n is the input amount at instant τ_n and Δ_n is the change in content during the time interval $[\tau_n, \tau_{n+1})$.

Define $\mathcal{G}(x)$ as the anti-derivative of $\frac{1}{r(x)}$ for $r(x) > 0$. Then $\mathcal{G}(x)$ is a continuous increasing function of x, since $d\mathcal{G}(x)/dx = 1/r(x)$ is > 0. The time for the content to decline from state-space level v to level $u > 0$, is (see Sect. 6.2.4)

$$\int_u^v \frac{1}{r(x)}dx = \mathcal{G}(v) - \mathcal{G}(u).$$

A necessary and sufficient condition for the content of the dam to return to level 0 is: for every $v > 0$,

$$\lim_{u\downarrow 0}\int_{x=u}^v \frac{1}{r(x)}dx = \lim_{u\downarrow 0}[\mathcal{G}(v) - \mathcal{G}(u)] = \mathcal{G}(v) - \lim_{u\downarrow 0}\mathcal{G}(u) < \infty \qquad (8.2)$$

(see Sect. 6.2.5). For example, in a pharmacokinetic model (Sect. 11.6) with "first order" kinetics, $r(x) = kx, x > 0$. In theory the drug concentration never returns to level 0. In practice, the drug may be entirely removed from the body after some finite time.

8.2.1 Embedded Downcrossing Rate

Proposition 8.1 The probability of an embedded downcrossing of level x occurring in $[\tau_n, \tau_{n+1}]$ is

$$\begin{aligned} d_n(x) &= \int_{y=0}^{\infty}\int_{\alpha=x}^{\gamma(x,y)} B(\gamma(x,y)-\alpha)dF_n(\alpha)dA(y) \\ &= \int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF_n(\alpha), n = 1, 2, \dots , \end{aligned} \qquad (8.3)$$

where $\gamma(x, y) = \mathcal{G}^{-1}(\mathcal{G}(x) + y)$, and $\eta(\alpha, x) = \mathcal{G}(\alpha) - \mathcal{G}(x)$.

Proof An embedded downcrossing occurs in $[\tau_n, \tau_{n+1}] \iff W_n > x$ and the time for $W(t)$ to descend from level $W_n + S_n$ to level x is $\leq (\tau_{n+1} - \tau_n) \iff$

$$\int_{z=x}^{W_n+S_n} \frac{1}{r(z)}dz = \mathcal{G}(W_n + S_n) - \mathcal{G}(x) \leq \tau_{n+1} - \tau_n. \qquad (8.4)$$

Conditioning on $\tau_n - \tau_{n+1} = y$, (8.4) is equivalent to

$$\begin{aligned} &\mathcal{G}(W_n + S_n) - \mathcal{G}(x) \leq y, \\ &\mathcal{G}(W_n + S_n) \leq \mathcal{G}(x) + y. \end{aligned} \qquad (8.5)$$

Note that $\mathcal{G}(\cdot)$ and its inverse $\mathcal{G}^{-1}(\cdot)$ are both continuous and increasing functions. Taking the inverse $\mathcal{G}^{-1}$ on both sides of (8.5) gives

$$S_n \leq \mathcal{G}^{-1}(\mathcal{G}(x) + y) - W_n = \gamma(x, y) - W_n.$$

Conditioning on $W_n = \alpha$, gives

$$P(\text{embedded downcrossing in } [\tau_n, \tau_{n+1}] | \tau_n - \tau_{n+1} = y)$$
$$= \int_{\alpha=x}^{\gamma(x,y)} B(\gamma(x, y) - \alpha) dF_n(\alpha).$$

We obtain the unconditional probability of an embedded downcrossing of x during $[\tau_n, \tau_{n+1}]$ by integrating with respect to the inter-input time y having distribution $A(y)$. This yields $d_n(x)$ given in (8.3). □

Let

$$\delta_n(x) = \begin{cases} 1 \text{ if there is an embedded downcrossing of } x \text{ in } [\tau_n, \tau_{n+1}], \\ 0 \text{ if there is no embedded downcrossing of } x \text{ in } [\tau_n, \tau_{n+1}]. \end{cases}$$

Then $E(\delta_n(x)) = d_n(x)$. The number of embedded downcrossings of level x in $[0, \tau_{n+1}]$ is

$$\mathcal{D}_n(x) = \sum_{i=1}^{n} \delta_i(x).$$

Thus

$$E(\mathcal{D}_n(x)) = \sum_{i=1}^{n} d_i(x).$$

The long-run expected embedded downcrossing *rate* of level x is

$$\lim_{n\to\infty} \frac{E(D_n(x))}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(x).$$

From (8.3), since $\lim_{n\to\infty} F_n(x) \equiv F(x)$, then $\lim_{n\to\infty} d_n(x) = d(x)$, where

$$d(x) = \int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x, y) - \alpha) dA(y) dF(\alpha).$$

Also,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(x) = \lim_{n\to\infty} d_n(x) = d(x)$$

implies the expected embedded level downcrossing rate of level x is

$$\lim_{n\to\infty}\frac{E(\mathcal{D}_n(x))}{n}=\int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty}B(\gamma(x,y)-\alpha)dA(y)dF(\alpha). \qquad (8.6)$$

8.2.2 Embedded Upcrossing Rate

Proposition 8.2 The probability of an embedded upcrossing of level x occurring in $[\tau_n, \tau_{n+1}]$ is

$$\begin{aligned} u_n(x) &= \int_{y=0}^{\infty}\int_{\alpha=0}^{x}\overline{B}(\gamma(x,y)-\alpha)dF_n(\alpha)dA(y) \\ &= \int_{\alpha=0}^{x}\int_{y=0}^{\infty}\overline{B}(\gamma(x,y)-\alpha)dA(y)dF_n(\alpha), n=1,2,\dots . \end{aligned} \qquad (8.7)$$

Proof An embedded upcrossing of level x occurs in $[\tau_n, \tau_{n+1}] \iff W_n \leq x$, $W_n+S_n > x$, and the time for $W(t)$ to descend from level W_n+S_n to level x exceeds $\tau_{n+1}-\tau_n$

$$\begin{aligned} &\iff \textstyle\int_{z=x}^{W_n+S_n}\frac{1}{r(z)}dz=\mathcal{G}(W_n+S_n)-\mathcal{G}(x)>\tau_{n+1}-\tau_n \\ &\iff S_n>\mathcal{G}^{-1}(\mathcal{G}(x)+y)-W_n=\gamma(x,y)-W_n, \end{aligned}$$

where we have conditioned on $\tau_n-\tau_{n+1}=y$. Therefore

$$\begin{aligned} &P(\text{embedded upcrossing in } [\tau_n,\tau_{n+1}]|\tau_n-\tau_{n+1}=y) \\ &\quad=\int_{\alpha=0}^{x}\overline{B}(\gamma(x,y)-\alpha)dF_n(\alpha), \end{aligned}$$

where $\overline{B}(z)=1-B(z)$, $z\geq 0$. Therefore, the unconditional probability of an embedded upcrossing of x in $[\tau_n,\tau_{n+1}]$ is given by (8.7). □

As in the derivation of (8.4), it follows that the long-run expected embedded upcrossing rate of level x is

$$\lim_{n\to\infty}\frac{E(\mathcal{U}_n(x))}{n}=\int_{\alpha=0}^{x}\int_{y=0}^{\infty}\overline{B}(\gamma(x,y)-\alpha)dA(y)dF(\alpha). \qquad (8.8)$$

8.2.3 Integral Equation for Steady-State PDF of Content

Applying (8.1), rate balance across level x, to formulas (8.6) and (8.8) gives an integral equation for $f(x)$ and $F(x)$, namely,

$$\begin{aligned}&\int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha)\\&\quad-\int_{\alpha=0}^{x}\int_{y=0}^{\infty}\overline{B}(\gamma(x,y)-\alpha)dA(y)dF(\alpha)=0,\ x\geq 0.\end{aligned}\tag{8.9}$$

CDF Form of Integral Equation

In the second term of (8.9) write $\overline{B}(\cdot)=1-B(\cdot)$ and apply $F(x)=\int_{\alpha=0}^{x} dF(\alpha)$. This yields a ***cdf form*** with $F(x)$ on the left side explicitly,

$$\begin{aligned}F(x)=&\int_{\alpha=0}^{x}\int_{y=0}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha)\\&+\int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha),\ x\geq 0.\end{aligned}\tag{8.10}$$

PDF Form of Integral Equation

Differentiation of (8.10) with respect to $x>0$, gives a **pdf form** with $f(x)$ explicitly on the left side,

$$\begin{aligned}f(x)=&\int_{\alpha=0}^{x}\int_{y=0}^{\infty}\varrho(x,y)\cdot b(\gamma(x,y)-\alpha)dA(y)dF(\alpha)\\&+\int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty}\varrho(x,y)\cdot b(\gamma(x,y)-\alpha)dA(y)dF(\alpha),\ x>0,\end{aligned}\tag{8.11}$$

where $\varrho(x,y)=\partial\gamma(x,y)/\partial x=r(\gamma(x,y))/r(x)$.

Probability of Zero Content

Letting $x\downarrow 0$ in (8.10) gives

$$F(0)=\frac{\int_{\alpha=0^+}^{\infty}\int_{y=\eta(\alpha,0)}^{\infty} B(\gamma(0,y)-\alpha)dA(y)dF(\alpha)}{\int_{y=0}^{\infty}\overline{B}(\gamma(0,y))dA(y)}.\tag{8.12}$$

The normalizing condition is

$$F(0) + \int_{\alpha=0}^{\infty} f(\alpha) d\alpha = 1 \tag{8.13}$$

If condition (8.2) does not hold, then $F(0) = 0$ (recall that $f(0) \equiv f(0^+)$).

Solution Method
The solution method in the following sections will be to obtain the functional form of $f(x)$ and $F(x)$ using (8.10) or (8.11), and applying the boundary conditions (8.12) and (8.13) to specify $f(x)$, $F(x)$, $x \geq 0$.

8.2.4 M/G/r(·) Dam

In this model, $A(y) = 1 - e^{-\lambda y}$, $y \geq 0$. Note that

$$\frac{\partial\,(\gamma(x, y))}{\partial y} = \frac{\partial(\mathcal{G}^{-1}(\mathcal{G}(x) + y))}{\partial y} = r(\gamma(x, y)) = r(\mathcal{G}^{-1}(\mathcal{G}(x) + y)).$$

Integrating (8.11) *by parts*, using the parts

$$\frac{\lambda e^{-\lambda y}}{r(y)} \quad \text{and} \quad r(\gamma(x, y)) \cdot b(\gamma(x, y) - \alpha) dy,$$

simplifying and substituting from (8.10), results in

$$r(x) f(x) = \lambda \int_{\alpha=0}^{x} \overline{B}(x - \alpha) dF(\alpha), x > 0. \tag{8.14}$$

Equation (8.14) is identical to the integral equation (6.21) for the steady-state pdf of content in the M/G/r(·) dam (derived using "continuous" LC).

Remark 8.1 In Eq. (8.14) $f(x) = \lim_{n\to\infty} f_n(x)$ since (8.14) has been derived using **embedded** LC. In Chap. 6, Eq. (6.21), $f(x) = \lim_{t\to\infty} f_t(x)$ is the **time-average** steady-state pdf of content. The fact that $\lim_{n\to\infty} f_n(x)$ and $\lim_{t\to\infty} f_t(x)$ satisfy the same integral equation, demonstrates that the content of an M/G/r(·) dam satisfies the PASTA principle that Poisson arrivals "see" time averages (see [145]). Here we have derived PASTA for the M/G/r(·) dam by using continuous and embedded LC concepts only.

8.3 GI/G/1 Queue

The GI/G/1 queue is closely related to the Gi/G/r(·) dam (see Table 8.1). For the virtual wait of the GI/G/1 queue $r(x) = \begin{cases} 1, x > 0, \\ 0, x = 0. \end{cases}$

The anti-derivative of $1/r(x), x > 0$, is

$$\mathcal{G}(x) = \int \frac{1}{r(x)} dx = \int 1 \cdot dx = x.$$

Thus,

$$\begin{aligned}
\gamma(x, y) &= \mathcal{G}^{-1}(\mathcal{G}(x) + y)) = \mathcal{G}^{-1}(x + y)) = x + y \\
\eta(\alpha, x) &= \mathcal{G}(\alpha) - \mathcal{G}(x) = \alpha - x, \\
\varrho(x, y) &= \frac{r(\gamma(x, y))}{r(x)} = \frac{r(x + y)}{1} = \frac{1}{1} = 1.
\end{aligned}$$

In the GI/G/1 *queue*, Eqs. (8.10), (8.11) and (8.13) reduce respectively to

$$\begin{aligned}
F(x) = &\int_{\alpha=0}^{x} \int_{y=0}^{\infty} B(x + y - \alpha) dA(y) dF(\alpha) \\
&+ \int_{\alpha=x}^{\infty} \int_{y=\alpha-x}^{\infty} B(x + y - \alpha) dA(y) dF(\alpha), x \geq 0,
\end{aligned} \tag{8.15}$$

$$\begin{aligned}
f(x) = &\int_{\alpha=0}^{x} \int_{y=0}^{\infty} b(x + y - \alpha) dA(y) dF(\alpha) \\
&+ \int_{\alpha=x}^{\infty} \int_{y=\alpha-x}^{\infty} b(x + y - \alpha) dA(y) dF(\alpha), x > 0,
\end{aligned} \tag{8.16}$$

$$F(0) = \frac{\int_{\alpha=0^+}^{\infty} \int_{y=\alpha}^{\infty} B(y - \alpha) dA(y) dF(\alpha)}{\int_{y=0}^{\infty} \overline{B}(y)) dA(y)}. \tag{8.17}$$

Table 8.1 GI/G/r(.) dam versus GI/G/1queue

GI/G/r(·) Dam	Gi/G/1 Queue
Input instant τ_n^-	Customer arrival instant τ_n^-
Input amount at τ_n^-	Service time (jump size) S_n
Content at τ_n^-	Customer wait W_n in queue at τ_n^-
Content at instant τ_n	Virtual wait $W(\tau_n) = W_n + S_n$
Content at time $t \geq 0$	Virtual wait $W(t)$ at time $t \geq 0$
$r(x) > 0, x > 0; r(0) = 0$	$r(x) = 1, x > 0; r(0) = 0$
Distribution of content	Distribution of waiting time

The normalizing condition is

$$F(0) + \int_{\alpha=0}^{\infty} f(\alpha)d\alpha = 1. \tag{8.18}$$

8.3.1 Applications of Embedded LC

Some single-server queueing models can be analyzed using embedded LC, by applying Eqs. (8.15)–(8.18). Other models are analyzed by deriving integral equations for the pdf of the state variables from first principles using embedded LC. The next four sections illustrate some applications.

8.3.2 M/G/1 Queue

The M/G/1 queue is a special case of the M/G/r(·) dam, with $r(x) = 1, x > 0$, and $A(y) = 1 - e^{-\lambda y}$, $y \geq 0$. Substituting directly into Eq. (8.14) or into (8.16) followed by some algebra yields

$$\begin{aligned} f(x) &= \lambda \int_{\alpha=0}^{x} \overline{B}(x-\alpha)dF(\alpha) \\ &= \lambda P_0 \overline{B}(x) + \lambda \int_{\alpha=0}^{x} \overline{B}(x-\alpha)f(\alpha)d\alpha, x > 0, \end{aligned} \tag{8.19}$$

which is identical to Eqs. (3.34) in Sect. 3.2.10. Remark 8.1 above applies also to this queueing model.

8.3.3 GI/M/1 Queue

The GI/M/1 queue is a special case of the GI/G/1 queue with

$$B(x) = 1 - e^{-\mu x}, x \geq 0, \;\; b(x) = \mu e^{-\mu x} = \mu - \mu B(x), x > 0.$$

Substituting $b(x) = \mu - \mu B(x)$ into (8.16), simplifying and combining with (8.15), gives the integral equation

$$f(x) = \mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy, \;\; x > 0, \tag{8.20}$$

Table 8.2 Interchanged roles of terms in integral equations for M/G/1 and G/M/1

Equation (8.19) for M/G/1	**Equation (8.20) for G/M/1**
λ	μ
x is upper bound of integral	x is lower bound of integral
$\overline{B}(x-y)$	$\overline{A}(y-x)$
P_0 appears explicitly	P_0 does not appear explicitly

which is identical to Eq. (5.7) in Sect. 5.1.3.

Duality of M/G/1 and GI/M/1 Queues

Upon comparing integral equations (8.19) and (8.20) it is evident that they are duals, in the sense that the roles of certain terms are interchanged (see Table 8.2). The significance of this "duality" is that we analyze the M/G/1 queue via LC using the virtual wait process. On the other hand, we are led to analyzing the G/M/1 queue via LC using the extended "age" process (see Sect. 5.1.1 and [15]).

Remark 8.1 applies also to GI/M/1, provided we analyze the extended age process, for which departures from the system occur in a Poisson process at rate μ conditional on the server being occupied. This implies that in (8.20), $f(x)$ on the left side (equal to time-average pdf of virtual wait) is the same function as $f(y)$ in the integrand on the right side (pdf of system time at departure instants).

Solution for Steady-State PDF of Wait in GI/M/1

The pdf of wait has the form $f(x) = Ke^{-\gamma x}$, $x > 0$ (see formula (5.11) in Sect. 5.1.5). Substituting $Ke^{-\gamma x}$ into (8.20) yields the equation for γ

$$\int_{z=0}^{\infty} \overline{A}(z)e^{-\gamma z}dz = \frac{1}{\mu},$$

or

$$\frac{1}{\gamma} - \frac{1}{\gamma}A^*(\gamma) = \frac{1}{\mu}, \tag{8.21}$$

where $A^*(\cdot)$ is the Laplace-Stieltjes transform of $A(\cdot)$ defined by

$$A^*(s) = \int_{y=0}^{\infty} e^{-sy}a(y)dy, s \geq 0,$$

and $a(y) = dA(y)/dy$, assuming the inter-arrival times are continuous r.v.s. We obtain an expression for $P_0 = F(0)$ upon substituting $B(y) = 1 - e^{-\mu y}$ and $f(\alpha) = Ke^{-\gamma\alpha}$ in (8.17), i.e.,

$$F(0) = [A^*(\mu)]^{-1} \left[\frac{\gamma - \mu + \mu A^*(\gamma) - \gamma A^*(\mu)}{\gamma(\gamma - \mu)} \right] \cdot K. \tag{8.22}$$

From (8.21)

$$\mu - \mu A^*(\gamma) = \gamma,$$

which substituted into (8.22) leads directly to

$$F(0) = \frac{K}{\mu - \gamma}. \tag{8.23}$$

The normalizing condition (8.18) gives

$$\frac{K}{\mu - \gamma} + \frac{K}{\gamma} = 1.$$

Then (8.23) implies

$$F(0) = \frac{\gamma}{\mu}. \tag{8.24}$$

Formula (8.24) is important because $F(0) = P_{0,\iota}$ in (5.31) which was derived using "continuous" or "time-average" LC. (This provides further evidence of the overall logical validity of the LC methodology.)

Check with M/M/1 Queue

It is instructive to check the result for the M/M/1 queue. Consider M/M/1 with arrival rate λ and service rate μ. Then $A^*(s) = \frac{\lambda}{\lambda+s}$. From (8.21) $\gamma = \mu - \lambda$, which substituted into (8.22), gives $F(0) = P_0 = \frac{K}{\lambda}$. Applying the normalizing condition $F(0) + \int_{y=0}^{\infty} f(y)dy = 1$, gives

$$\frac{K}{\lambda} + K \int_{y=0}^{\infty} e^{-(\mu-\lambda)y} dy = 1,$$

$$K = \lambda(1 - \frac{\lambda}{\mu}).$$

Thus

$$P_0 = \frac{K}{\lambda} = 1 - \frac{\lambda}{\mu}, \checkmark$$

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0, \checkmark$$

which checks with the M/M/1 solution given in (3.112) and (3.113) in Sect. 3.5.1.

8.3.4 $Erl_{k,\lambda}$/M/1 Queue

Assume the common pdf of the inter-arrival times is $a(\cdot) :=$ pdf of $\mathrm{Erl}_{k,\lambda}$. For integers $k = 1, 2, ..., a(y) = e^{-\lambda y} \frac{(\lambda y)^{k-1}}{(k-1)!} \lambda,\ y > 0$. Let $A(\cdot)$ denote the cdf corresponding to $a(\cdot)$ (see Example 3.2 in Sect. 3.3). The LST of $A(\cdot)$ is $A^*(\gamma) = \left(\frac{\lambda}{\lambda+\gamma}\right)^k$, which substituted into Eq. (8.21) gives an equation for γ,

$$\frac{1}{\gamma} - \frac{1}{\gamma}\left(\frac{\lambda}{\lambda + \gamma}\right)^k = \frac{1}{\mu}, k = 1, 2, \ldots . \tag{8.25}$$

We seek a unique positive solution of (8.25) for γ. Assume that $\lambda, \mu > 0$ and $\lambda < k\mu$ (stability condition for G/M/1 is $\boldsymbol{a} < \mu$, where $\boldsymbol{a} = k/\lambda =$ arrival rate). Then Eq. (8.25) has exactly one *real* positive root for γ (see [15]). If k is odd, all other roots are *complex*. If k is even, one other root is negative real and all other roots are complex. Thus the solution for γ is unique. Denote it by γ_k.

To solve for $K \equiv \eta_k$ we first substitute γ_k into (8.22) and use (8.25) to obtain

$$F(0) = \frac{\eta_k}{\mu - \gamma_k}.$$

(We use η_k instead of K_k in this section only, for notational contrast.) Then apply the normalizing condition (8.18) to obtain

$$\eta_k = \frac{\gamma_k(\mu - \gamma_k)}{\mu} = \gamma_k\left(1 - \frac{\gamma_k}{\mu}\right).$$

The steady-state pdf of wait is then given by

$$P_0 = \frac{\eta_k}{\mu - \gamma_k} = \frac{\gamma_k}{\mu},$$
$$f(x) = \eta_k e^{-\gamma_k x} = \gamma_k\left(1 - \frac{\gamma_k}{\mu}\right) e^{-\gamma_k x}, x > 0.$$

Remark 8.2 The solution of Eq. (8.25) can be readily obtained numerically for any specified values of λ, μ, k.

8.3.5 D/M/1 Queue

Assume the common inter-arrival time is $D > 0$. Then $A^*(s) = e^{-sD}$, $s > 0$. Let the steady-state pdf of wait be $f(x) = Ke^{-\gamma x}$, $x > 0$. Substituting $A^*(\gamma) = e^{-\gamma D}$ into (8.21) gives the equation

$$\mu e^{-\gamma D} + \gamma - \mu = 0$$

for γ, whose solution we call γ_D. From (8.22)

$$F(0) = \frac{K}{\mu - \gamma_D}.$$

Let $K_D := K$. Substituting into (8.18) gives

$$\frac{K_D}{\mu - \gamma_D} + \frac{K_D}{\gamma_D} = 1,$$
$$K_D = \gamma_D \left(1 - \frac{\gamma_D}{\mu}\right).$$

The steady-state pdf of wait is

$$P_0 = \frac{K_D}{\mu - \gamma_D},$$
$$f(x) = K_D e^{-\gamma_D \cdot x}, x > 0.$$

8.4 M/G/1: Wait Related Reneging/Balking

We revisit the M/G1 queue with balking/reneging in Sect. 3.13, in which customers can balk from joining the system upon arrival, or renege from the waiting line, depending on the required wait and staying resolve. Here, we apply the embedded LC method to analyze the system. Assume the *staying function* is $\overline{R}(y) = P(\text{arrival stays for service}|\text{required wait} = y)$. We show that embedded LC will verify that the pdf $f(x)$, $x > 0$, on the left and right sides of Eq. (3.207) are the same functions. This is important because on the left side $f(x) = \lim_{t\to\infty} f_t(x)$ (*a time-average pdf*). On the right side $f(y) = \lim_{n\to\infty} f_{\iota,n}(x) := f_\iota(y)$ (*an arrival-point pdf*), and $P_0 = P_{0,\iota}$ (arrival point probability of a zero wait). We now use embedded LC to derive an integral equation for $f_\iota(x)$, $x > 0$, and show that it is identical to Eq. (3.207).

8.4.1 Embedded Level Crossing Probabilities

The limiting probability of an SP *embedded upcrossing* of level x is

$$u = \int_{y=0^-}^{x} \int_{z=0}^{\infty} \overline{B}(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy, \tag{8.26}$$

where the lower limit $y = 0^-$ means that the term $\overline{B}(x+z)P_{0,\iota}$ for the atom $\{0\}$, is included in the evaluation of u. The right side of (8.26) holds because an embedded upcrossing of x occurs iff $0 \leq W_n = y < x$, the arrival at τ_n stays for service (probability $\overline{R}(y)$), and given that the time to the next arrival is z, the service time exceeds $x - y + z$.

The limiting probability of an SP *embedded downcrossing* of level x consists of two terms,

$$\begin{aligned} d = &\int_{y=x}^{\infty} \int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy \\ &+ \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} R(y)f_\iota(y)\lambda e^{-\lambda z}dzdy. \end{aligned} \tag{8.27}$$

The first term on the right of (8.27) is similar to (8.26), except that an SP jump starts at a level $y > x$ and the service time must be less than $x - y + z$ for an embedded downcrossing to occur. The second term is due to arrivals that *do not stay for service* (balk at joining the system or renege from the waiting line); arrivals renege with probability $R(y) = 1 - \overline{R}(y)$. We can assume that an SP "jump" is of size 0 (probability $R(y)$) when a reneger arrives; equivalently there is *no SP jump* when a balker arrives. In this case the SP makes an embedded downcrossing of level x provided the next inter-arrival time is $z > y - x$. The second term in (8.27) simplifies to $\int_{y=x}^{\infty} R(y)f_\iota(y)e^{-\lambda(y-x)}dy$.

Since $\overline{B}(x) \equiv 1 - B(x)$, $x \geq 0$, Eq. (8.26) can be written as

$$u = \int_{y=0^-}^{x} \overline{R}(y)f_\iota(y)dy - \int_{y=0^-}^{x} \int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy \tag{8.28}$$

8.4.2 Steady-State PDF of Wait of Stayers

Applying *embedded* rate balance across level x, we set $u = d$. This yields, from Eqs. (8.27) and (8.28), the integral equation

$$\int_{y=0^-}^{x} \overline{R}(y)f(y)dy = \int_{y=0^-}^{x}\int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy$$
$$+ \int_{y=x}^{\infty}\int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy$$
$$+ \int_{y=x}^{\infty}\int_{z=y-x}^{\infty} R(y)f(y)\lambda e^{-\lambda z}dzdy. \quad (8.29)$$

We take d/dx on both sides of (8.29), which involves differentiation under the integral sign. Some algebra, including cancellation of terms and using $R(y) + \overline{R}(y) = 1$, gives

$$f_\iota(x) = \int_{y=0^-}^{x}\int_{z=0}^{\infty} b(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy$$
$$+ \int_{y=x}^{\infty}\int_{z=y-x}^{\infty} b(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy$$
$$+ \lambda\int_{y=x}^{\infty}\int_{z=y-x}^{\infty} R(y)f_\iota(y)\lambda e^{-\lambda z}dzdy. \quad (8.30)$$

Integrating each of the inner integrals

$$\int_{z=0}^{\infty} b(x-y+z)\lambda e^{-\lambda z}dz \;\text{ and }\; \int_{z=y-x}^{\infty} b(x-y+z)\lambda e^{-\lambda z}dz$$

in (8.30) by parts, using the parts $\lambda e^{-\lambda z}$ and $b(x-y+z)$, leads to the integral equation (assuming $B(0) = 0$)

$$\begin{aligned} f_\iota(x) = &-\lambda \textstyle\int_{y=0^-}^{x} \overline{R}(y)f_\iota(y)B(x-y)dy \\ &+\lambda \textstyle\int_{y=0^-}^{x}\int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy \\ &+\lambda \textstyle\int_{y=x}^{\infty}\int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f_\iota(y)\lambda e^{-\lambda z}dzdy \\ &+\lambda \textstyle\int_{y=x}^{\infty}\int_{z=y-x}^{\infty} R(y)f_\iota(y)\lambda e^{-\lambda z}dzdy. \end{aligned} \quad (8.31)$$

From (8.29) the sum of the last three terms on the right of (8.31) is

$$\lambda\int_{y=0^-}^{x} \overline{R}(y)f(y)dy.$$

Hence

$$f_\iota(x) = \lambda \int_{y=0^-}^{x} \overline{R}(y) f_\iota(y) dy - \lambda \int_{y=0^-}^{x} \overline{R}(y) f_\iota(y) B(x-y) dy,$$

$$f_\iota(x) = \lambda \int_{y=0^-}^{x} \overline{B}(x-y) \overline{R}(y) f_\iota(y) dy. \tag{8.32}$$

Equation (8.32) is *identical to* (3.207). Hence, in (3.207), the *time-average pdf* of stayers (left side) is equal to the *arrival-point pdf* of stayers (which occurs in the integral on right side). The derivation of (3.207) using "continuous-time" LC is far simpler than the derivation of (8.32). Nevertheless, the embedded LC method is very useful in this case, and elsewhere. It helps to confirm that "continuous" LC works in the wait-time dependent reneging/balking model. The embedded LC method can often be applied to determine whether the time-average and arrival-point pdfs are equal. The embedded LC method is inherently very intuitive, and has additional applications as well.

Chapter 9
Level Crossing Estimation

9.1 Introduction

This chapter describes a basic level crossing estimation method (LCE) for *steady-state (i.e.,limiting) probability distributions* in queues, storage processes and related stochastic models. LCE is also called: level crossing computation, system point estimation (or computation). LCE is related to non-parametric density estimation methods (e.g., [133]). In standard density estimation the data is assumed to be a random sample from an unknown pdf. The data is used to construct histograms, naive density estimators, kernel-density estimators, etc., for the unknown pdf, utilizing associated smoothing techniques. (LCE for *transient distributions* is discussed briefly in Example 3.1 in Sect. 3.2.8 in Chap. 3.)

In LCE, for estimating steady-state distributions, we obtain the data from a *single simulated sample path* of the stochastic process of interest over a long simulated time t. We compute estimates of the pdf of the state variable from the appropriate sample-path level-crossing time averages, or related averages. The estimators used in LCE can be combined with smoothing techniques to improve the estimates (e.g., [93, 94, 98]).

9.1.1 Main Steps of Level Crossing Estimation

The basic LCE procedure described here for steady-state distributions, has three main steps:

1. Simulate a *single* sample path of the process over a *long simulated* time period, say $[0, t]$, where t is "large".

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_9

2. Using the simulated sample path, compute *point estimates* of the pdf and cdf of the state variable, in terms of level-crossing time averages calculated on a state-space partition with "small" fixed norm Δ (defined in Definition 9.1 in Sect. 9.4.1 below); or on a *natural* state-space partition obtained from the peaks and troughs of the simulated sample path (see Fig. 9.1). Also, compute point estimators of moments and expected values of measurable functions of the state variable.
3. Obtain *confidence intervals* for the pdf, cdf, moments and expected values of measurable functions.

Remark 9.1 Step 2 above may also include a sensitivity analysis of the estimates. We may vary the simulated total time t, and/or the state-space partition norm size Δ (see Definition 9.1 in Sect. 9.4.1 below), to ensure that estimates remain within preassigned tolerances.

In addition to the three main steps above, we also characterize the steady-state pdf and cdf according to continuity, boundedness, convexity, differentiability, etc., by utilizing the sample-path properties on $[0, t]$. For example, in M_λ/G/1 and in G/M_μ/1 queues, the steady-state pdfs of wait are bounded by λ and μ respectively (see Proposition 3.7 in Sect. 3.4.22, and Proposition 5.8 in Sect. 5.1.16).

The author has carried out numerous LCE computational experiments using the procedure described here, as well as other LCE procedures using BASIC computer software (e.g., [17, 26, 27, 44]). These experiments detected accurately, such pdf properties as: discontinuities, intervals of convexity or concavity, etc., in benchmark models, where the pdf properties are known analytically. For example, the M/Discrete/1 queue may serve as a benchmark. Proposition 3.11 in Sect. 3.11.1 specifies continuity/discontinuity properties of the pdf of wait. We can also apply LCE to estimate the pdf of wait in *variants* of M/Discrete/1 queues with state dependencies, etc., in which analytical results are either tedious to obtain, or are not otherwise available.

9.2 Theoretical Basis for LC Estimation

LCE is based on level crossing theorems. Consider M/G/1. Theorem 1.1 in Chap. 1 implies that virtual-wait sample-path level-crossing time averages converge to the steady-state pdf of wait ($a.s.$) as time $t \to \infty$ (Sect. 9.2.2). This implies that time averages computed from a simulated sample path over a long simulated time t, should approximate the pdf accurately for all state-space values up to the maximum state-space level attained during $[0, t]$, say χ_t. Thus, the state-space interval $[0, \chi_t]$ will contain an increasing measure of the total probability as t increases (Sect. 9.2.5). The measure will grow to 1 as $t \to \infty$.

Remark 9.2 The LCE method described here is one of several LC estimation methods developed by the author. For example, a version of LCE for estimating **transient distributions of state variables** is based on Theorem 3.2 in Sect. 3.2.4 and Theorem 3.3 in Sect. 3.2.5, and earlier related theorems (e.g., see Remark 3.7 in Sect. 3.2.8 above). Talks on this technique were presented by the author at several conferences, e.g., P.H. Brill (1982), System Point Monte Carlo Simulation of Time Dependent Probability Distributions of Waiting Times in Queues, TIMS/ORSA National Meeting, Chicago, April.

9.2.1 Boundedness of Steady-State PDF

A bound on the steady-state pdf of the virtual wait $f(x)$ in M_λ /G/1 queues is given in Proposition 3.7, Sect. 3.4.22, and on the steady-state *arrival-point* pdf of wait $f_\iota(x)$ in G/M_μ/1 queues in Proposition 5.8, Sect. 5.1.16. In M_λ/G/1, $f(x) < \lambda, x > 0$. In G/M_μ/1, $f_\iota(x) < \mu, x > 0$. Recall that $\mathcal{D}_t(x)$ and $\mathcal{U}_t(x)$ are the numbers of SP down- and upcrossings of level x during $(0, t]$, respectively. Boundedness implies that for a *typical* sample path in M_λ/G/1,

$$f(x) = \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} < \lambda, x \geq 0.$$

In G/M_μ/1,

$$f_\iota(x) = \lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} < \mu, x > 0.$$

Similarly, we can develop bounds on $f(x)$ for other models, e.g., M/M/c, G/M/c, etc. In M_λ/G/r(·) dams, boundedness follows from integral equation (6.21) in Sect. 6.2.11 for the steady-state pdf of content $f(x)$. Therein, if the efflux rate satisfies $r(x) > m > 0, x > 0$, then $f(x) < \lambda/m, x > 0$, because $\int_{y=0}^{x} \overline{B}(x-y)dF(y) < 1, x > 0$.

9.2.2 Role of Level Crossing Theorems in LCE

Consider the virtual wait $\{W(t)\}_{t\geq 0}$ in the M_λ/G/1 queue. (See sample paths in Fig. 3.5, Sect. 3.3.2, and in Fig. 9.1 below.) Let $F(x), x \geq 0$, $f(x), x > 0$, be the steady-state cdf and pdf of wait respectively. Let $P_0 := F(0)$. Theorem 1.1 in Chap. 1 asserts

$$\lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} = f(x), x \geq 0, \qquad \lim_{t\to\infty}\frac{\mathcal{D}_t(0)}{t} = f(0) = \lambda P_0 \ (a.s.)$$

(recall that $f(0) := f(0^+)$, for notational convenience). Given $\varepsilon > 0$, for each $x > 0$ there exists $t_{x,\varepsilon}$ such that

$$t > t_{x,\varepsilon} \implies \left|\frac{\mathcal{D}_t(x)}{t} - f(x)\right| < \varepsilon f(x) \ (a.s.), \tag{9.1}$$

since $f(x)$ is bounded, i.e., $0 < f(x) < \lambda < \infty$, $x > 0$ (see Sect. 9.2.1). Also, since $\lim_{t\to\infty}\mathcal{D}_t(0)/t = \lambda P_0$, there exists $t_{0,\varepsilon}$ such that

$$t > t_{0,\varepsilon} \implies \left|\frac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| < \varepsilon P_0. \tag{9.2}$$

Choose an arbitrary "small" δ, $0 < \delta << 1$. Let $W_q :=$ *steady-state queue wait* having pdf $f(x)$. Define $z_\delta > 0$ by $P(W_q > z_\delta) = \delta$. Then δ is the probability of the right tail of $f(x)$, namely (z_δ, ∞). Thus

$$1 - F(z_\delta) = \int_{y=z_\delta}^{\infty} f(y)dy = \delta. \tag{9.3}$$

Suppose we could determine (finite) $t_\delta^* = \max_x\{t_{x,\varepsilon} | x \in [0, z_\delta)\}$, where $t_{x,\varepsilon}$, $x > 0$ is defined in (9.1) and $t_{0,\varepsilon}$ is defined in (9.2). Then $t > t_\delta^*$ implies

$$\begin{aligned} \left|\tfrac{\mathcal{D}_t(x)}{t} - f(x)\right| &< \varepsilon f(x) \text{ for all } x \in (0, z_\delta)(a.s.), \\ \text{and } \left|\tfrac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| &< \varepsilon P_0 \ (a.s.). \end{aligned} \tag{9.4}$$

Using the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$, gives

$$P_0 + \int_{x=0}^{z_\delta} f(x)dx = 1 - \int_{x=z_\delta}^{\infty} f(x)dx = 1 - \delta > 0. \tag{9.5}$$

Summing over all $x \in [0, \infty)$ in (9.4) and using (9.5), shows that $t > t_\delta^*$ implies

$$\begin{aligned} \left|\tfrac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| + \int_{x=0}^{z_\delta}\left|\tfrac{\mathcal{D}_t(x)}{t} - f(x)\right|dx &< \varepsilon P_0 + \varepsilon\int_{x=0}^{z_\delta} f(x)dx \\ &= \varepsilon(1-\delta) < \varepsilon(a.s.). \end{aligned} \tag{9.6}$$

Let $\left\{\widehat{P_0}, \widehat{f}(x)\right\}_{x>0}$ denote the estimate of $\{P_0, f(x)\}_{x>0}$. We assume that a sample path over a fixed simulated time interval $[0, t]$ is used to compute

$\left\{\widehat{P_0}, \widehat{f}(x)\right\}_{x>0}$. (We omit subscript "$t$" in the symbols $\widehat{P_0}$ and $\widehat{f}(x)$, in order to distinguish $\widehat{P_0}$, $\widehat{f}(x)$ from estimators "$\widehat{P_{0,t}}$, $\widehat{f_t}(x)$" for the *transient* pdf of wait, which we use outside this monograph.)

Assume we use the "natural" estimator based on the peaks and troughs of the sample path, viz., $\widehat{P_0} = \frac{\mathcal{D}_t(0)}{\lambda t}$, $\widehat{f}(x) = \frac{\mathcal{D}_t(x)}{t}$, $t > t^*_\delta$. Then (9.6) implies that the *total absolute error* of $\left\{\widehat{P_0}, \widehat{f}(x)\right\}_{x>0}$ in estimating $\{P_0, f(x)\}_{x\in(0,z_\delta)}$ is less than ε.

Since the simulated sample path over the finite time period $(0, t^*_\delta)$ is bounded above, we assume the estimate $\widehat{f}(x) = 0, x > z_\delta$. Then $t > t^*_\delta$ implies the total absolute error in $\widehat{f}(x)$, $x > z_\delta$, is equal to δ, i.e.,

$$t > t^*_\delta \implies \int_{x=z_\delta}^{\infty} \left|\widehat{f}(x) - f(x)\right| dx = \int_{x=z_\delta}^{\infty} f(x)dx = \delta, (a.s.). \quad (9.7)$$

Suppose we could simulate a sample path over a sufficiently large time interval $(0, t)$, $t > t^*_\delta$. Statements (9.6) and (9.7) imply the total absolute error would be

$$\begin{aligned} &\left|\widehat{P_0} - P_0\right| + \int_{x=0}^{\infty} \left|\widehat{f}(x) - f(x)\right| dx \\ &= \left|\frac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| + \int_{x=0}^{\infty} \left|\frac{\mathcal{D}_t(x)}{t} - f(x)\right| dx < \varepsilon + \delta, (a.s.). \end{aligned} \quad (9.8)$$

In practice we can choose ε and δ arbitrarily small. Then we can simulate a sample path over a long simulated time $t > t^*_\delta$ and ensure that the total absolute error of $\left\{\widehat{P_0}, \widehat{f}(x)\right\}_{x>0}$ in estimating $\{P_0, f(x)\}_{x>0}$ is arbitrarily small. This procedure would amount to *computation* of the entire true pdf $\{P_0, f(x)\}_{x>0}$ within a preassigned tolerance. The total error on $[0, z_\delta)$ is $< \varepsilon$. The total error on (z_δ, ∞) is $\le \delta$.

9.2.3 Natural Partition of State Space

We illustrate a *natural partition* of the state space with an example.

Example 9.1 Consider a **sample path of the virtual wait** $\{W(t)\}_{t\ge0}$ **in an M/G/1 queue** (Fig. 9.1). The state space is $\boldsymbol{S} = [0, \infty)$. For **fixed** $x \in \boldsymbol{S}$, $\{\mathcal{D}_t(x)\}_{t\ge0}$ is a counting process and a renewal process (due to Poison arrivals). For **fixed** $t > 0$, $\mathcal{D}_t(x)$ is a step function on $\boldsymbol{S}$. The step-function jumps occur at the peaks $\{W_n + S_n\}_{n=1,2,...}$ and troughs $\{W_n\}_{n=1,2,...}$, where W_n is an actual wait in queue, and S_n is a service time. In Fig. 9.1 level $W(0)$ is a peak and level $W(t)$ is a trough. We **merge the peaks and troughs** to form a natural state-space partition

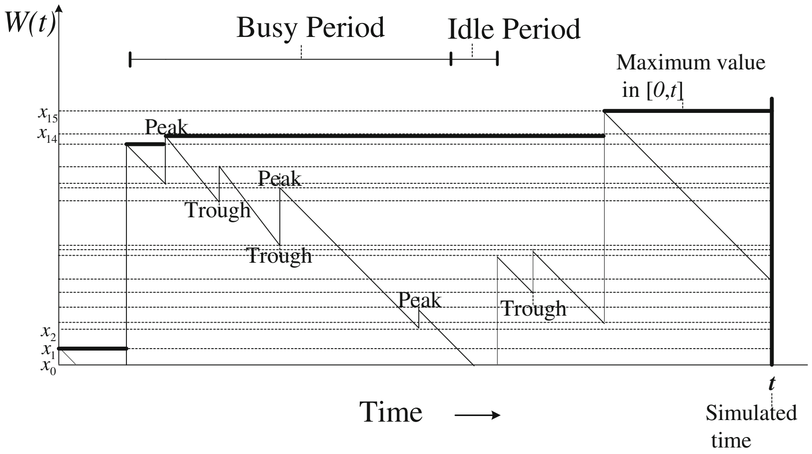


Fig. 9.1 Sample path of virtual wait $\{W(t)\}_{t\geq 0}$ in M/G/1. Shows peaks $\{W_n + S_n\}_{n=1,2,...}$; troughs $\{W_n\}_{n=1,2,...}$; state-space partition $0 = x_0 < x_1 < x_2 \cdots < x_{15}$ in time interval $(0, t)$; maximum sample-path x_{15} value attained in $[0, t]$

$$\{x_i\}_{i=0,1,2,...} = W(0) \cup \{W_n\}_{n=1,2,...} \cup \{W_n + S_n\}_{n=1,2,...} \cup W(t),$$

arranged **in ascending order of magnitude** in $\boldsymbol{S}$, namely,

$$0 = x_0 < x_1 < \cdots < x_{M(t)} < \infty,$$

where $M(t)$ is the number of subintervals of partition $\{x_i\}_{i=0,1,2,....}$. The first partition point x_0 corresponds to all troughs of $W(0) \cup \{W_n\}_{n=1,2,...} \cup W(t)$ such that the ordinate is 0. The second partition point is

$$x_1 = \min_n \left\{W(0) \cup \{W_n\} \cup \{W_n + S_n\} \cup W(t) \diagdown \ \{W_n = 0\}\right\}$$

That is, $\min_n \{\cdot\}$ excludes the troughs corresponding to x_0 $(= 0)$. The jth partition point x_j is obtained recursively, excluding those troughs and/or peaks corresponding to $\left\{x_0, x_1, ..., x_{j-1}\right\}$. $M(t)$ satisfies the inequality $M(t) \leq 2N_a(t)$, where $N_a(t)$ is the number of customer arrivals during $[0, t]$. In Fig. 9.1, $N_a(t) = 8$, $M(t) = 15$.

9.2.4 Step Function for Downcrossings on a Partition

Similarly as in Example 9.1, we fix the interval $[0, t]$ prior to the simulation of the sample path. Observing Fig. 9.1, shows that the number of downcrossings in $[0, t]$ is the same for every x in a state-space subinterval of the partition $[x_i, x_{i+1}), i = 0, 1, ..., M(t) - 1$. Let $\mathcal{D}_t(x) :=$ *number of down-crossings during* $[0, t]$ *of level* $x \in [0, M(t)]$. Then

$$\mathcal{D}_t(x) = d_i, \text{ for all } x \in [x_i, x_{i+1}), i = 0, 1, ..., M(t) - 1,$$

where $d_i \geq 0$ is a constant. Thus

$$\frac{\mathcal{D}_t(x)}{t} = \frac{d_i}{t}, x \in [x_i, x_{i+1}), i = 0, 1, ..., M(t)$$

is a step function of $x \in \left[0, x_{M(t)}\right]$.

Suppose we can determine t^*_δ as in (9.4). From (9.8)

$$t > t^*_\delta \implies \left|\frac{d_0}{\lambda t} - P_0\right| + \sum_{i=0}^{M(t)} \int_{x=x_i}^{x_{i+1}} \left|\frac{d_i}{t} - f(x)\right| dx < \varepsilon + \delta, (a.s.). \tag{9.9}$$

In Fig. 9.1,

$$d_0 = 2, \quad d_1 = 1, \quad d_2 = 2, \quad d_3 = 3, ..., d_{14} = 1, \quad d_{15} = 0.$$

The recursion (9.10) below may simplify computation of $\{d_i\}_{0=M(t)-1}$ in a computer program.

$$d_{i+1} = \begin{cases} d_i + 1 & \text{if } x_{i+1} \text{ is a trough,} \\ d_i - 1 & \text{if } x_{i+1} \text{ is a peak, } i = 0, ..., M(t) - 1, \end{cases} \tag{9.10}$$

Since $\mathcal{D}_t(x) = 0$, for $x_{M(t)} \leq x < \infty$, we may write $d_{M(t)+1} = 0$.

The sub-interval lengths of the partition are

$$x_{i+1} - x_i, i = 0, ..., M(t) - 1.$$

These lengths vary in a natural way (variable bin sizes), because they depend on the probability laws governing the simulation of the sample path.

9.2.5 Ladder Points and LCE Estimates

For $\{W(t)\}_{t\geq 0}$, let $\chi_t :=$ *maximum sample-path level in* S *attained during* $[0, t]$. For fixed $t > 0$, $\chi_t = x_{M(t)}$, the highest finite point of the partition $\{x_i\}_{i=0,...,M(t)}$. As t increases $\{\chi_t\}_{t\geq 0}$ is a non-decreasing step function with the inter-jump time at level x distributed as b_x, the sojourn time of $\{W(t)\}_{t\geq 0}$ below level x (see Fig. 9.1). A sample path of $\{\chi_t\}_{t\geq 0}$ is a non-decreasing right-continuous step function with upward jumps $\underset{dis}{=}$ *excess over level x* of the service-time jump that ends sojourn b_x. Let $\tau_{l,n}, n = 1, 2,...,$ denote the embedded arrival instants when ladder points occur (subscript 'l' stands for 'ladder point'). Since $\chi_{\tau,n-1} = \chi_{\tau_{l,n}^-}$, $d\chi_t/dt = 0$, $\tau_{l,n-1} < t < \tau_{l,n}$, $n = 0, 1, 2,...,$ where $\tau_{l,0} \equiv 0$. The increase in $\{\chi_t\}_{t\geq 0}$ at ladder-point instant $\tau_{l,n}$ equals $\chi_{\tau_{l,n}} - \chi_{\tau_{l,n}^-} =$ *excess service time above level* $\chi_{\tau_{ln}^-}$.

Random variables $\chi_{\tau_{l,n}}, n = 1, 2, ...,$ are ordinates of the *strict ascending ladder points* $\left\{\left(\tau_{l,n}, \chi_{\tau_{l,n}}\right)\right\}_{n=1,2,...}$ of the virtual wait process $\{W(s)\}_{s\geq 0}$. The points $\left(\tau_{l,n}, \chi_{\tau_{l,n}}\right) \in \boldsymbol{T} \times \boldsymbol{S}, n = 1, 2, ...,$ are analogous to the strict ascending ladder points for a random walk (see Fig. 1, p. 192, and pp. 390–394, in [74]). The LCE estimate of the pdf of wait $f(x), x \geq \chi_t$, is $\widehat{f}(x) = 0$, since $\chi_t = x_{M(t)}$. The number of strict ascending ladder points $\left\{\left(\tau_{l,n}, \chi_{\tau_{l,n}}\right)\right\}_{n=1,2,...}$ in time interval $[0, t]$ form a counting process (as t increases). If the sample-path jump sizes are $\underset{dis}{=}$ Exp_μ, all excess jumps over a level are $\underset{dis}{=}$ Exp_μ by the memoryless property. Then the nth ascending ladder point is $\underset{dis}{=}$ $\text{Erl}(n, \mu)$ (see Example 3.2 in Sect. 3.3 for the pdf of $\text{Erl}(n, \mu)$). (We mention ladder points here because of their importance in the overall LCE method, and their connection with random walk. We shall not discuss them further in this introduction to LCE.)

9.3 Computer Program for LCE

LCE can be used to obtain extremely accurate estimates of *steady-state* or *transient* pdfs, of state variables. This section focuses on estimating the *steady-state* pdf' of $\{W(t)\}_{t\geq 0}$. For a fixed $t > 0$, we simulate a sample path of the virtual wait over a simulated time interval $[0, t]$. We count $\mathcal{D}_t(x), x > 0$:= *number of SP downcrossings of every state-space level* $x \in S$ *during* $[0, t]$. This is easier than it may seem at first glance, due to the step-function structure of $\mathcal{D}_t(x), x > 0$, for any fixed $t > 0$.

9.3.1 Designs for a Computer Program

The associated computer program for the LCE estimates of $\{P_0, f(x)\}_{x>0}$ can utilize different logical designs. We discuss two feasible program designs for the LCE computer program, which specify the state-space partition to be used.

State-Space Partition with a Variable Subinterval Size
One design is based directly on the discussion in Sect. 9.2, using the *natural* partition $\{x_i\}_{i=0,..,M(t)}$ having *variable* sub-interval sizes $\Delta_i = x_{i+1} - x_i$. The Δ_is occur naturally as part of the simulated sample path (see Fig. 9.1).

Embedded processes $\{W_n\}_{n=1,2,...}$ (troughs) and $\{W_n + S_n\}_{n=1,2,...}$ (peaks) are Markov processes. Thus, in a sample path the union $\{W_n\} \cup \{W_n + S_n\}, n = 1, 2,...$, of troughs and peaks, is everywhere dense in $\boldsymbol{S} = [0, \infty)$ as $t \to \infty$ $(a.s.)$. That is, the entire state space will be 'almost' completely covered *eventually* by the ordinates of the peaks and troughs.

An advantage of this design is that it takes every sample-path peak and trough during $[0, t]$ into consideration. In theory, any computed estimator using the natural partition will utilize all the information available in the sample path.

A possible disadvantage of this design is from a programming point of view. The points in $\{x_i\}_{i=0,..,M(t)}$ become more dense as the sample path is generated over time. The Δ_is in the regions of higher probability, will become extremely small as simulated time t increases. The partition $\{x_i\}_{i=0,..,M(t)}$ will contain on the order of $2N_a(t)$ distinct points, where $N_a(t)$ is the number of 'arrivals' in $(0, t)$ (each arrival produces 1 trough and 1 peak of the sample path). If t is large, $N_a(t)$ will be large. Many Δ_is will become less than a practical resolution size required for the estimation of the pdf of wait.

State-Space Partition with a Fixed Subinterval Length
A second design starts $\{x_i\}_{i=0,..,M(t)}$ with $x_0 = 0$ and a *fixed* partition norm length Δ. Thus $x_i = x_{i-1} + \Delta$, $i = 1, ...$ The computer program updates the count of SP downcrossings of each state-space level x_i, $i = 0, ..., M(t)$, as the sample path evolves over the time interval $[0, t]$. We compute the maximum peak χ_s during $[0, s]$ as we generate the sample path over time $s \in [0, t]$. The state-space partition $\{x_i\}_{i=0,..,M(t)}$ covers the state-space interval $[0, \chi_t]$. Generally the time intervals between successive ladder points of $\{W(s)\}_{0<s<t}$ increase, because the sojourn times b_x below level x, increase as x increases. That is, generally $\tau_{l,n+1} - \tau_{l,n} > \tau_{l,n} - \tau_{l,n-1}$, after some integer $n \in \{1, 2, ...\}$. Estimates of $\{P_0, f(x)\}_{x \geq 0}$ computed using a fixed-Δ partition, very closely approximate estimates using a natural partition with variable Δ_is, for most practical purposes. Moreover, the fixed-Δ design is easy to program.

9.4 LCE for the M/G/1 Queue

This section describes LCE for the steady-state pdf of wait and related quantities for M/G/1 queues. A numerical example using this method is given in the next section. Let $\{W(t)\}_{t\geq 0}$ denote the virtual wait. Without loss of generality assume $W(0) = 0$. The state space is $\boldsymbol{S} = [0, \infty)$. Let the arrival rate be λ. Let $S_n, n = 1, 2,...$ denote the service times, which may be state dependent. Assume the parameters are such that the queue is stable, e.g., $\lambda E(S) < 1$. Assume $W(t) \underset{dis}{\rightarrow} W$ as $t \to \infty$ (weak convergence). Denote the cdf and pdf of W by $F(x)$, $x \geq 0$, and $\{P_0, f(x)\}_{x>0}$ respectively. Here $P_0 = F(0) > 0$ and $f(x) = dF(x)/dx$ wherever the derivative exists. Denote the nth moment of W by $m_n = \int_{x=0}^{\infty} x^n f(x)dx, n = 1, 2,....$ Let $\psi(W)$ denote an arbitrary measurable function of W.

We use a computer program based on the fixed-norm size design (fixed Δ) of Sect. 9.3.1 to compute the estimates. Definition 9.1 below in 9.4.2 incorporates minor modifications of the "basic" estimators, that retain theoretical consistency, and are suitable in practice.

9.4.1 Quantities Computed from a Sample Path

Fix finite time $t > 0$. Consider a simulated sample path $\{W(s)\}_{0\leq s\leq t}$. The SP is the leading point of a sample path when thought of as evolving over time (Sect. 2.3 in Chap. 2).

Definition 9.1

$\mathcal{D}_t(x)$ number of SP downcrossings of level x, $x \geq 0$ during $[0, t]$,
χ_t $\max\{W(s)|0 \leq s \leq t\}$,
Δ norm of preassigned uniform partition on $\boldsymbol{S}$,
ν $\max\{n|n\Delta \leq \chi_t, n = 0, 1, 2, ...\}$,
x_j $x_j = j\Delta,\ j = 0, ..., \nu + 1;\ x_{\nu+2} \equiv \infty$,
$\{x_j\}$ preassigned uniform partition on $[0, (\nu + 1)\,\Delta]$ with norm Δ,
$\boldsymbol{J}_j$ interval $\boldsymbol{J}_j = [x_j, x_{j+1}),\ j = 0, 1, ..., \nu$,
d_j $\mathcal{D}_t(x_j),\ j = 0, ..., \nu + 1$,
A_t $A_t = \frac{1}{t}\left(\frac{d_0}{\lambda} + \Delta\sum_{j=0}^{\nu} d_j\right) = \frac{1}{t}\left(\frac{\mathcal{D}_t(0)}{\lambda} + \Delta\sum_{j=0}^{\nu}\mathcal{D}_t(x_j)\right)$.

Remark 9.3 Definition 9.1 retains the argument "t" for $\mathcal{D}_t(x)$, χ_t and A_t. Both ν and d_j also depend on t. We omit subscript t: for ν to simplify notation since ν often appears as a subscript or index; for d_j for computer-programming purposes. The quantities Δ, x_j and $\boldsymbol{J}_j$ are defined in the state space, and are generally independent of t. (We may vary t and Δ jointly in various runs of the program for a **sensitivity analysis** in order to study accuracy of the estimates.) The term A_t is such that $A_t > 0, t > \tau_1$ ($\tau_1 =$ first customer arrival instant).

Remark 9.4 Note the inequality $x_\nu = \nu \cdot \Delta \leq \chi_t < (\nu+1)\Delta = x_{\nu+1}$. Also, for every $x \geq x_{\nu+1}$, $\mathcal{D}_t(x) = 0$, implying $d_{\nu+1} \equiv 0$.

Proposition 9.1

$$\lim_{\substack{t\to\infty \\ \Delta\downarrow 0}} A_t = 1 \ (a.s.). \tag{9.11}$$

Proof We *sketch* a proof of (9.11) in three steps.
(1) $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ (normalizing condition).
(2) For the first term of A_t we have

$$\lim_{t\to\infty} \frac{d_0}{t\lambda} = \lim_{t\to\infty} \frac{\mathcal{D}_t(0)}{t\lambda} = \frac{f(0)}{\lambda} = \frac{\lambda P_0}{\lambda} = P_0 \ (a.s.). \tag{9.12}$$

(3) For the second term of A_t, first we assume the virtual wait $W(t) \leq K$, $t > 0$, for some upper bound $K > 0$. (For example, Variants 1 and 2 of M/G/1 with bounded $W(t) \leq K$ are discussed in Sect. 3.16 in Chap. 3.) Then $\chi_t \leq K$ for all $t > 0$. Also $\nu \leq \left[\frac{K}{\Delta}\right]$ where $[z]$ denotes the greatest integer $\leq z$, $z \in \mathbb{R}$. Thus ν is finite and positive for all values of t. Moreover,

$$\begin{aligned} \lim_{\substack{t\to\infty \\ \Delta\downarrow 0}} \left(\Delta \textstyle\sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right) &= \lim_{\Delta\downarrow 0}\left(\lim_{t\to\infty}\left(\textstyle\sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\Delta\right)\right) \\ &= \lim_{\Delta\downarrow 0}\left(\textstyle\sum_{j=0}^{\nu}\left(\lim_{t\to\infty} \frac{\mathcal{D}_t(x_j)}{t}\right)\Delta\right) \\ &= \lim_{\Delta\downarrow 0}\left(\textstyle\sum_{j=0}^{\upsilon} f(x_j)\Delta\right) \\ &= \textstyle\int_{x=0}^{K} f(x)dx (a.s.), \end{aligned} \tag{9.13}$$

since $\lim_{t\to\infty} \mathcal{D}_t(x_j)/t = f(x_j)$ $(a.s.)$ by Theorem 1.1 in Chap. 1. In the last equality of (9.13), the expression $\sum_{j=0}^{\upsilon} f(x_j)\Delta$ is a Riemann sum. It converges to the definite integral $\int_{x=0}^{K} f(x)dx$ as $\Delta \downarrow 0$, since $K - \Delta < x_\nu \leq K$.

The result (9.13) holds for every $K > 0$. If $K \to \infty$, then

$$\lim_{\substack{t\to\infty \\ \Delta\downarrow 0}} \left(\Delta \sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right) = \int_{x=0}^{\infty} f(x)dx \ (a.s.). \tag{9.14}$$

Thus, Eq. (9.11) follows from (9.12), (9.13) and the normalizing condition. ■

9.4.2 Point Estimators

For fixed $t > 0$ let

$$\widehat{f}(x), x > 0, \quad \widehat{F}(x), x \geq 0, \quad \widehat{P}_0, \quad \widehat{m}_n, n = 1, 2, ..., \quad \widehat{E}(\psi(W)),$$

denote *point estimators* of the corresponding quantities under the circumflexes, specified in Definition 9.2 below. Assume a "small" norm Δ is given (Δ = "bin size").

Definition 9.2 For each fixed $t > 0$, the **point estimators** are (see Definition 9.1 above):

1. $\widehat{f}(x) \equiv \frac{d_j}{tA_t} = \frac{\mathcal{D}_t(x_j)}{tA_t}, x \in \boldsymbol{J}_j, j = 0, ..., \nu,$
2. $\widehat{P}_0 = \frac{d_0}{\lambda t A_t} = \frac{\mathcal{D}_t(0)}{\lambda t A_t},$
3. $\widehat{F}(x) = \widehat{P}_0 + \Delta \sum_{i=0}^{j-1} \widehat{f}(x_i) + (x - x_j)\widehat{f}(x_j), x \in \boldsymbol{J}_j, j = 0, ..., \nu,$
4. $\widehat{m}_n = \Delta \sum_{i=0}^{\nu} x_i^n \widehat{f}(x_i),$
5. $\widehat{E}(\psi(W)) = \psi(0)\widehat{P}_0 + \Delta \sum_{i=0}^{\nu} \psi(x_i)\widehat{f}(x_i).$

Estimator of the Laplace-Stieltjes Transform of Wait
In Definition 9.2, set $\psi(W) = e^{-sW}$, $s > 0$. Then $E(\psi(W))$ is the Laplace-Stieltjes transform (LST) of W, namely

$$E(e^{-sW}) = \int_{x=0}^{\infty} e^{-sx} dF(x)dx.$$

The estimator of $E(e^{-sW})$ is

$$\widehat{E}(e^{-sW}) = \widehat{P}_0 + \Delta \sum_{i=0}^{\nu} e^{-sx_i} \widehat{f}(x_i), s > 0.$$

We can compute $\widehat{E}(e^{-sW})$, $s = 0, h, 2h,...,$ where h is a small positive constant, and plot $\widehat{E}(e^{-sW})$ versus s. Then we can substitute $\widehat{E}(e^{-sW})$ for the LST in formulas where it appears.

The value of Δ may be adjusted after a computer run, to increase accuracy or investigate an estimator's convergence rate with respect to Δ.

Remark 9.5 In Definition 9.2 the quantities under the symbol "$\widehat{\ }$" omit the argument t, to distinguish them from *estimators of transient distributions.* (The latter estimators are not included in this monograph, but are discussed briefly in Remark 9.2 in Sect. 9.2 above and related cited remarks therein.) The quantities also omit the argument Δ for notational simplicity.

Remark 9.6 For fixed $t > 0$, $\widehat{f}(x)$ is a step function of $x \in \cup_{j=0}^{\nu+1} \boldsymbol{J}_j$ having constant values on the intervals $\{\boldsymbol{J}_j\}_{j=0,...,\nu+1}$. The **term A_t is a normalizing constant** which guarantees that $\widehat{F}(x) = 1, x \geq x_{\nu+1}$, for any $t > 0$. Also, $\widehat{f}(x) = 0$, $x \in J_{\nu+1}$.

Consistency of Estimators

Definition 9.3 An estimator $\widehat{\varphi}_t$ of the quantity ϕ is said to be *consistent* if $\lim_{t\to\infty} P\,(\widehat{\varphi}_t = \phi) = 1$; $\widehat{\varphi}_t$ is *strongly consistent* if $P\,(\lim_{t\to\infty} \widehat{\varphi}_t = \phi) = 1$ – equivalently $\lim_{t\to\infty} \widehat{\varphi}_t = \phi\ (a.s.)$. (See, e.g., p. 322ff in [107]; p. 132ff in [9], for discussions on consistency of estimators.)

Strong consistency implies $\lim_{t\to\infty} \widehat{\varphi}_t = \phi$ for every simulated (typical) sample path of $\{W(s)\}_{s\in(0,t)}$, if t is "large". This is because every possible sample path is a "point" in the background sample space of the process $\{W(s)\}_{s\in(0,\infty)}$. Thus any simulated sample path will give the same accurate estimates. The estimators

$$\widehat{f}(x), x > 0, \quad \widehat{F}(x), x \geq 0, \widehat{P}_0, \quad \widehat{m}_n, n = 1, 2, ..., \quad \widehat{E}\,(\psi(W))$$

in Definition 9.2 are strongly consistent, which is proved in the following proposition by using level crossing theorems discussed in Sect. 9.2.2.

Proposition 9.2

1. (a). For each x_j, $\widehat{f}(x_j)$ is strongly consistent.
   (b). For each fixed $x \neq x_j$ $\lim_{\Delta\downarrow 0} \widehat{f}(x)$ is strongly consistent.
2. (a). For each fixed $t > 0$, $0 \leq \widehat{P}_0 \leq 1$.
   (b). $\widehat{P}_0$ is strongly consistent.
3. (a). For each fixed $t > 0$, $0 \leq \widehat{F}(x) \leq 1$, $x \geq 0$, and $\widehat{F}(\infty) = 1$.
   (b). For each fixed $x \geq 0$, $\lim_{\Delta\downarrow 0} \widehat{F}(x)$ is strongly consistent.
4. $\lim_{\Delta\downarrow 0} \widehat{m}_n$ is strongly consistent, $n = 1, 2, ...$
5. $\lim_{\Delta\downarrow 0} \widehat{E}(\psi(W))$ is strongly consistent.

Proof 1(a). By formula (9.11), since $\lim_{t\to\infty} A_t = 1$,

$$\lim_{t\to\infty} \widehat{f}(x_j) = \lim_{t\to\infty} \frac{d_j}{tA_t} = \lim_{t\to\infty} \frac{D_t(x_j)}{tA_t} \underset{a.s.}{=} \frac{f(x_j)}{\lim_{t\to\infty} A_t} = f(x_j).$$

1(b). Fix $t > 0$. Fix $x \in \boldsymbol{S}$. Let $\delta > 0$ be given. We can make the fixed norm size Δ arbitrarily small. There exists $\Delta > 0$ and x_j in the fixed norm partition such that $0 < x - x_j < \Delta$. Also we have $x - x_j < \Delta \implies \left|f(x) - f(x_j)\right| < \delta$, since $f(\cdot)$ is defined to be right continuous. Note that $\widehat{f}(x) \equiv \widehat{f}(x_j)$. Now let $t > t_{x_j\varepsilon}$, such that $t > t_{x_j\varepsilon} \implies \left|\widehat{f}(x_j) - f(x_j)\right| < \varepsilon$. (Such $t_{x_j\varepsilon}$ exists by 1(a).) Hence for Δ sufficiently small and $t > t_{x_j\varepsilon}$,

$$\begin{aligned}\left|f(x) - \widehat{f}(x)\right| &= \left|f(x) - \widehat{f}(x_j)\right| = \left|f(x) - f(x_j) + f(x_j) - \widehat{f}(x_j)\right| \\ &\leq \left|f(x) - f(x_j)\right| + \left|f(x_j) - \widehat{f}(x_j)\right| < \delta + \varepsilon.\end{aligned}$$

As $t \to \infty$, $\left|f(x_j) - \widehat{f}(x_j)\right| \downarrow 0$. Thus $\left|f(x) - \widehat{f}(x)\right| < \delta$, implying that $\lim_{t\to\infty}\left(\lim_{\Delta\downarrow 0} \widehat{f}(x)\right) = f(x)$ $(a.s.)$.
2(a). For fixed $t > 0$, $\mathcal{D}_t(0) \geq 0$. Hence

$$0 \leq \frac{D_t(0)}{\lambda t A_t} = \widehat{P_0} = \frac{\frac{D_t(0)}{\lambda t}}{\left(\frac{\mathcal{D}_t(0)}{\lambda t} + \Delta \sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right)} \leq 1.$$

2(b). For a stable queue, state $\{0\}$ is positive recurrent. Hence

$$\lim_{t\to\infty} \widehat{P_0} = \lim_{t\to\infty} \frac{D_t(0)}{\lambda t A_t} = \frac{f(0)}{\lambda \lim_{t\to\infty} A_t} = \frac{\lambda P_0}{\lambda \cdot 1} = P_0 \ (a.s.).$$

3(a). This follows because the denominators of $\widehat{P_0}$ and $\widehat{f}(x_j)$, $j = 1, ..., \upsilon$ contain the normalizing factor $A_t = \widehat{P_0} + \Delta \sum_{j=0}^{\nu} \widehat{f}(x_j)$, which exceeds or equals the value of the total numerator.

3(b). This follows because $\lim_{t\to\infty} \widehat{P_0} = P_0$. Also,

$$\begin{aligned}&\lim_{t\to\infty}\left(\lim_{\Delta\downarrow 0}\left(\Delta \sum_{i=0}^{j-1} \widehat{f}(x_i) + (x - x_j)\widehat{f}(x_j)\right)\right) \\ &= \lim_{\Delta\downarrow 0}\left(\lim_{t\to\infty}\left(\Delta \sum_{i=0}^{j-1} \widehat{f}(x_i) + (x - x_j)\widehat{f}(x_j)\right)\right)\end{aligned}$$

$$= \lim_{\Delta \downarrow 0} \left(\Delta \sum_{i=0}^{j-1} f(x_i) + (x - x_j) f(x_j) \right),$$

since for fixed Δ, the values of the partition points $\{x_j\}$ are fixed (thus interchange of limits permitted). Hence

$$\lim_{t \to \infty} \left(\lim_{\Delta \downarrow 0} \widehat{F}(x) \right) = P_0 + \int_{y=0}^{x} f(x) dx = F(x) \ (a.s.).$$

4, 5. These follow using similar reasoning as in the proof of 3(b). ■

Remark 9.7 In the estimation procedure of this section, we must make two important preset choices: (1) the value of simulated time t; (2) the value of Δ. Since t is finite and $\Delta > 0$, the estimators in Proposition 9.2 are **approximately** consistent. We consider the partition norm Δ to be sufficiently "small" if the following holds. We repeat the estimation procedure with a smaller Δ, say $\frac{\Delta}{10}$ or $\frac{\Delta}{100}$, etc., and several repetitions leave the estimates within a preassigned tolerance.

Similarly, we consider t to be sufficiently "large" if repeating the procedure with a larger t, say $10t$ or $100t$, etc., leaves the estimates within a preassigned tolerance (compare with the Cauchy condition for convergence of series—see, e.g., p. 390 in [137]). The joint choice of (t, Δ) poses an interesting exercise. Experimentation may be informative. A discussion is given in [24]. Computational experimentation has shown that the estimation procedure is robust over a wide range of (t, Δ) values. With the advent of fast computer processors, fast random access memories, fast storage drives, etc., a sensitivity analysis can be carried out very efficiently. Computer speeds will increase in the future. Sensitivity analyses of the estimates with respect to (t, Δ) will become ever more efficient.

9.4.3 Statistical Properties and Confidence Limits

For an arbitrary sample path $W(s), 0 \leq s \leq t$, define the following quantities.

d_x	time between successive SP downcrossings of level x,
$Var(d_x)$	variance of d_x,
$\sqrt{Var(d_x)}$	standard deviation of d_x,
b_x	time SP is in state-space interval $[0, x]$ during d_x = sojourn time at or below level x,

$\mathcal{A}((W(\cdot))^n)$ area under the sample path of $(W(s))^n$ during a busy cycle of $\{W(s)\}_{0\leq s\leq t}$,

λP_0 long-run rate at which arrivals initiate busy periods.

Asymptotic Normality of Estimators The following proposition describes the asymptotic normality of the estimators. Let $N(0, 1)$ denote a standard normal random variate with mean 0 and variance 1. Let $Var(Z)$ denote the variance of a generic random variable Z.

Proposition 9.3 1. For every x_j, $j = 0, ..., \nu$

$$\frac{\widehat{f}(x_j) - f(x_j)}{Var(d_{x_j})\,((tA_t)^{-1}\,(f(x_j))^3)^{\frac{1}{2}}} \to N(0, 1) \text{ as } t \to \infty.$$

2.

$$\frac{\widehat{P}_0 - P_0}{Var(d_0)\,((tA_t)^{-1}\,\lambda\,(P_0)^3)^{\frac{1}{2}}} \to N(0, 1) \text{ as } t \to \infty.$$

3. If Δ is small then for every $x \geq 0$ approximately

$$\frac{\widehat{F}(x) - F(x)}{\big((tA_t)^{-1} Var(b_x - b_0) f(x)\big)^{\frac{1}{2}}} \to N(0, 1) \text{ as } t \to \infty.$$

4. If Δ is small then approximately

$$\frac{\widehat{m}_n - m_n}{(t^{-1} Var(\mathcal{A}((W(\cdot))^n))\lambda P_0)^{\frac{1}{2}}} \to N(0, 1) \text{ as } t \to \infty.$$

Proof The proofs of statements 1–4 follow from the asymptotic normality of the number of renewals in the time interval $(0, t)$ as $t \to \infty$, for renewal processes (also known as the central limit theorem for renewal processes). (See pp. 36–42, especially p. 40, in [66]; p. 438 in [125].). This proposition is also discussed in Sect. 6 of [24]. ■

Confidence Intervals for Estimators

Assume t is large and define $z_{\frac{\alpha}{2}}$ by $P(N(0, 1) > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. The following $100(1 - \alpha)\%$ confidence limits apply.

1. $f(x_j)$: $\quad \widehat{f}(x_j) \pm z_{\frac{\alpha}{2}} \cdot \widehat{Var}(d_{x_j}) \cdot \left((tA_t)^{-1} \widehat{f}(x_j)^3\right)^{\frac{1}{2}}$,
2. P_0: $\quad \widehat{P}_0 \pm z_{\frac{\alpha}{2}} \cdot \widehat{Var}(d_0) \cdot \left((tA_t)^{-1} \lambda \left(\widehat{P}_0\right)^3\right)^{\frac{1}{2}}$,
3. $F(x)$: $\quad \widehat{F}(x) \pm z_{\frac{\alpha}{2}} \cdot \left((tA_t)^{-1} \widehat{Var}(b_x - b_0) \widehat{f}(x_j)\right)^{\frac{1}{2}}$,
4. m_n: $\quad \widehat{m}_n \pm z_{\frac{\alpha}{2}} \cdot \left(t^{-1} \widehat{Var}(\mathcal{A}((W(\cdot))^n)) \lambda \widehat{P}_0\right)^{\frac{1}{2}}$.

Proof The proofs are based on Proposition 9.3. ∎

9.5 LCE Example: M/M/1 with Reneging

Consider an $M_\lambda/M_\mu/1$ queue in which customers can renege from the waiting line at any time. Otherwise customers wait and stay for complete service. (See Sects. 3.13, 3.13.5 and Eqs. (3.211), (3.212) in Chap. 3.) We compute LCE estimates of the steady-state pdf, cdf and mean wait of *stayers* and compare them with the analytical solutions for the same quantities.

Let us assume customers who wait less than 1 time unit stay ("reach" the server) and get complete service; and customers required to wait ≥ 1 time unit to reach the server, renege from the waiting line before reaching the server, or balk at their arrival time (being informed by the system manager the approximate required wait until service starts). In the notation of Sect. 3.13 the *staying function* $\overline{R}(x)$, $x \geq 0$, has the same form as in Fig. 3.26, i.e.,

$$\overline{R}(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & x \geq 1. \end{cases} \tag{9.15}$$

The arrival rate λ and service rate μ may be arbitrary positive numbers since the queue is stable for all values of λ, μ (see Theorem 3.8 in Sect. 4.12.9). We arbitrarily set $\lambda = 1$, $\mu = 5$.

9.5.1 Analytical Solution

We obtain the analytical solution for the pdf of the wait of stayers $\{P_0, f(x)\}_{x>0}$ from the model equations

$$f(x) = \begin{cases} \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy, 0 < x < 1, \\ \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{1} e^{-\mu(x-y)} f(y)dy, x \geq 1. \end{cases} \tag{9.16}$$

We briefly explain the second equation in (9.16). When $x \geq 1$ the pdf $f(x)$ is positive due to those arrivals who stay for service (required wait < 1), whose service time jumps cause $\{W(t)\}_{t \geq 0}$ to upcross level 1. The solution of (9.16) is

$$f(x) = \begin{cases} \lambda P_0 e^{-(\mu-\lambda))x}, \ 0 < x < 1, \\ \lambda P_0 e^{\lambda} e^{-\mu x}, \ 1 \leq x < \infty. \end{cases} \tag{9.17}$$

We substitute (9.17) into the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$, yielding

$$P_0 = \frac{1}{1 + \frac{\lambda}{\mu-\lambda}(1 - e^{-(\mu-\lambda)}) + \frac{\lambda}{\mu} e^{-(\mu-\lambda)}}. \tag{9.18}$$

Interestingly, formulas (9.16), (9.17) and (9.18) are identical to those in display (3.255) with $\overline{B}(x) = e^{-\mu x}$ and $K = 1$, and in (3.258) with $k = 1$, in Sect. 3.16.6, in Chap. 3. Thus Variant 3 of M/G/1 with a threshold on the workload at level 1 is mathematically connected to the renege/balk M/G/1 queue due to the staying function $\overline{R}(y) = 1 \cdot \boldsymbol{I}_{[0.k)}(y) + 0 \cdot \boldsymbol{I}_{[k,\infty)}(y)$, where $\boldsymbol{I}_A(\cdot)$ is the characteristic function of set $\boldsymbol{A}$. The two differently conceived queueing models express the same idea! (Fig. 3.35 in Chap. 3 shows a valid sample path of $\{W(t)\}_{t \geq 0}$, in both models.)

Substituting $\lambda = 1$, $\mu = 5$ in (9.18) and (9.17) results in (see Fig. 9.3)

$$P_0 = 0.8006, \tag{9.19}$$

$$f(x) = \begin{cases} 0.8006 \cdot e^{-4.0x}, 0 < x < 1, \\ 2.1763 \cdot e^{-5.0x}, 1 \leq x < \infty. \end{cases} \tag{9.20}$$

From (9.20) the derivative is

$$f'(x) = \begin{cases} -3.2024 \cdot e^{-4.0x}, 0 < x < 1, \\ -10.8815 \cdot e^{-5.0x}, 1 \leq x < \infty. \end{cases}$$

The pdf $f(x)$ is continuous at $x = 1$. The derivative $f'(x)$ is discontinuous at $x = 1$. Thus $f'(1^-) = -0.058654$, $f'(1) = -0.073319$. The pdf is bounded above by the arrival rate λ, i.e.,

$$\max_{x \geq 0} f(x) = f(0) = 0.8006 < 1 = \lambda.$$

9.5.2 LCE Estimates of PDF and CDF of Wait of Stayers

We present the LCE estimates of $f(x)$, $F(x)$ and P_0 in Table 9.1, using $t = 3000$, $\Delta = 0.1$ in the computer program. See also Figs. 9.2 and 9.3.

9.5.3 LCE Estimates of Mean of Wait of Stayers and P_0

From (9.20), $E(W_q) \equiv \int_{x=0}^{\infty} xf(x)dx \equiv m_1 = 0.049$, where W_q denotes the required wait of stayers before service. Simulation of 10 independent sample paths using $t = 3000$, $\Delta = 0.1$, resulted in the sample-average point estimate

Table 9.1 Comparison of LCE estimates with analytic values in renege/balk M/M/1 queue

LC Estimation using $t = 3000$, $\Delta = 0.1$				
Estimated values			Analytical Values	
$\widehat{P}_0 = 0.7995$			$P_0 = .800587$	
x	$\widehat{f}(x)$	$\widehat{F}(x)$	$f(x)$	$F(x)$
0.1	.7995	.7995	.8006	.8006
0.2	.5265	.8652	.5366	.8666
0.3	.2447	.9395	.2411	.9403
0.4	.1602	.9591	.1616	.9603
0.5	.1142	.9734	1083	.9736
0.6	.0729	.9828	.0726	.9826
0.7	.0484	.9809	.0487	.9886
0.8	.0317	.9929	.0326	.9926
0.9	.0208	.9955	.0219	.9953
1.0	.0147	.9973	.0147	.9971
1.1	.0092	.9984	.0089	.9982
1.2	.0058	.9992	.0054	.9989
1.3	.0031	.9996	.0033	.9993
1.4	.0010	.9998	.0020	.9996
1.5	.0007	.9999	.0012	.9998
1.6	.0003	1.000	.0007	.9999
1.7	.0000	1.000	.0004	.9999

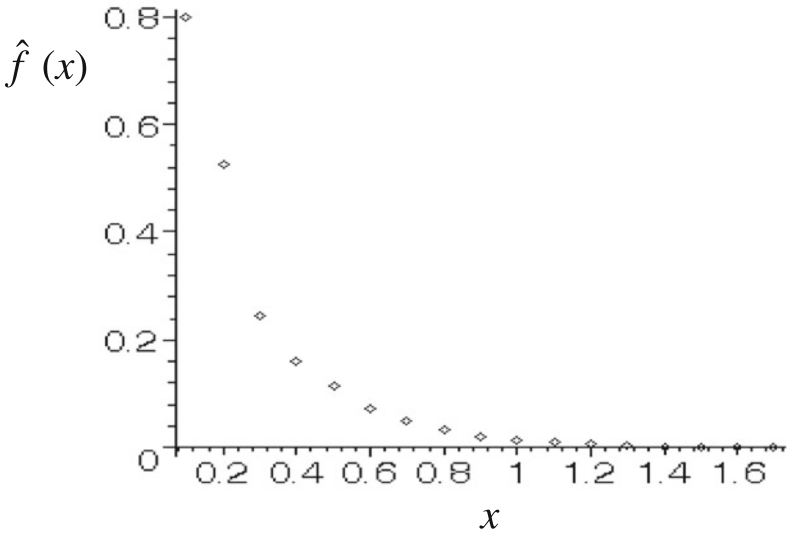


Fig. 9.2 Point estimate $\widehat{f}(x)$ of $f(x)$ based on Table 9.1, in $M_{\lambda}/M_{\mu}/1$ queue with reneging or balking at service: $\lambda = 1.0$, $\mu = 5.0$. Compare with Fig. 9.3

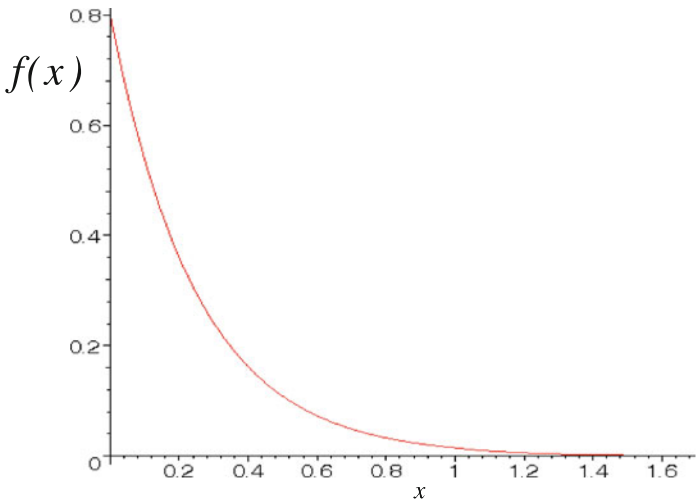


Fig. 9.3 Analytical solution for $f(x)$ in $M_{\lambda}/M_{\mu}/1$ queue with reneging or balking at service: $\lambda = 1$, $\mu = 5$. See formulas (9.15), (9.17), (9.20). $f(x)$ is continuous at $x = 1$. $f'(x)$ is discontinuous at $x = 1$

$\overline{\widehat{m}}_1 = 0.0489$. A 95% confidence interval for m_1 is obtained using $t_{9,0.025} \cdot s_{\widehat{m}_1}$ where $t_{9,0.025}$ is the upper 2.5% point of the Student "t" distribution with 9 degrees of freedom (Student "t" because 10 is a small sample size) and $s_{\widehat{m}_1}$ is the sample standard deviation of $\widehat{m}_1$. The value of $t_{9,0.025} \cdot s_{\widehat{m}_1}$ turned out to be 0.0013. Thus a 95% confidence interval is $m_1 = \overline{\widehat{m}}_1 \pm t_{9,0.025} \cdot s_{\widehat{m}_1}$ or $m_1 = 0.0489 \pm 0.0013$, which covers the true mean wait. Similarly a 95% confidence interval for P_0 is $P_0 = \overline{\widehat{P}}_0 \pm t_{9,0.025} \cdot s_{\widehat{P_0}}$ or $P_0 = 0.7996 \pm 0.0025$, which covers the true value of P_0.

9.5.4 Discussion of Numerical Example

The probability that an arbitrary arrival stays and receives full service is

$$\begin{aligned} q_S &= P_0 + \int_{x=0}^{\infty} \overline{R}(x) f(x) dx = P_0 + \int_{x=0}^{1} f(x) dx \\ &= 0.8006 + \int_{x=0}^{1} 0.8006 e^{-4.0x} dx = 0.9971. \end{aligned}$$

For the particular choice λ =1, μ = 5, and $\overline{R}(\cdot) = \boldsymbol{I} \cdot I_{[0,1)} + 0 \cdot \boldsymbol{I}_{[1,\infty)}$, in this example, nearly all customers stay, i.e., wait and get full service. Only $(1 - q_S) \cdot 100\% = 0.29\%$ either renege or balk at their arrival instant. The reason is that the service rate μ is very fast relative to the arrival rate λ. The vast majority of arrivals (99.71%) are required to wait less than one time unit, and therefore stay for a full service.

The expected busy period is

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = 0.24906.$$

The expected idle period is $E(\mathcal{I}) = \frac{1}{\lambda} = 1$. The proportion of time the server is idle is $E(\mathcal{I})/(E(\mathcal{I}) + E(\mathcal{B})) = 0.8006 = P_0$. Different values of λ, μ would of course, give quite different results.

9.6 Brief Discussion of LCE

LCE is useful for confirming theoretical results derived by various methods of analysis. LCE can be used to investigate the pdf of a state variable in a new model where the model equations are difficult to formulate, or if formulated, are analytically intractable. It is an alternative approach for estimating pdfs, cdfs, moments, and expected values of functions of state variables (e.g., Laplace transforms) in stochastic models.

LCE for steady-state distributions has several advantages. It uses a single simulated sample path of the underlying process of interest. It requires the analyst to be sufficiently familiar with the model dynamics to construct a sample path using a computer program. It may help to uncover and explain subtleties about the pdf and cdf of the state variable, which enhance intuition and understanding about the model. It may help to discover unexpected properties about the pdf of the state variable, or connections with apparently unrelated models.

LCE can be incorporated into a *hybrid* technique combining partially-known analytical solutions and statistical estimation. For example, in a single-server queue, the theoretical values of P_0 (probability of a zero wait) and $E(\mathcal{B})$ (expected busy period) may be known in terms of the model parameters. On the other hand, equations for the pdf of wait $f(x), x > 0$, may be analytically intractable. It may be possible to utilize the theoretical values of P_0 and $E(\mathcal{B})$ in the LCE computer program, to estimate $f(x), x > 0$.

LCE methods similar to that described here for M/G/1, have been applied to M/G/r(·) dams including cases where G is deterministic or discrete [27]; and to more complex models such as M/G^{a,b}/1 bulk-service queues [44]. The LCE technique is applicable in a vast array of other stochastic models as well. The basis of all such applications stems from the '$\lim_{t\to\infty} \mathcal{D}_t(x)/t = f(x)$' part of Theorem 1.1 in Chap. 1, and its extensions in multi-server queues and dams, etc.

We may classify the LCE method as an estimation method, or a *computational* method. With sensible values of the simulated time t and state-space partition norm size Δ, the technique gives almost-analytical values for the distribution of the state variable and related values, in many benchmark computational experiments already carried out.

Chapter 10
Renewal Theory Using LC

In this chapter, Sect. 10.1 gives an LC analysis of a replacement model, which is structured using two interconnected renewal processes. We derive efficiently, via sample paths and LC, the *limiting* pdfs of the excess life , the age, and the total life, of both renewal processes. Section 10.2 gives an LC analysis of a classical renewal problem with a barrier. Section 10.3 uses LC to derive the *finite time-t probability distributions* of the excess, age and total life, of a renewal process.

10.1 Replacement Model via Renewal Theory

We first describe a replacement model, which is a variant of the GI/G/r(·) dam. (see Sect. 6.2 for a related M/G/r(·) dam.) Sects. 10.1.3 and 10.1.4 derive the ***steady-state (limiting)*** pdfs of the excess, age and total life of two connected, renewal processes in the model.

10.1.1 The Model

Let $\{X(t)\}_{t\geq 0}$ denote a continuous-time continuous-state stochastic process having upward jumps of i.i.d. sizes $X_n > 0$, all starting at level 0, at times τ_n^-, where $0 = \tau_0 < \tau_1 \cdots < \tau_n < \cdots$, such that $X(\tau_n) = X_n, n = 0, 1, 2, \ldots$ Let the state space be $\boldsymbol{S} = [0, \infty)$. Figure 10.1 shows a sample path of $\{X(t)\}_{t\geq 0}$ (we use $X(t)$ to denote both the state variable and a sample path, for

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_10

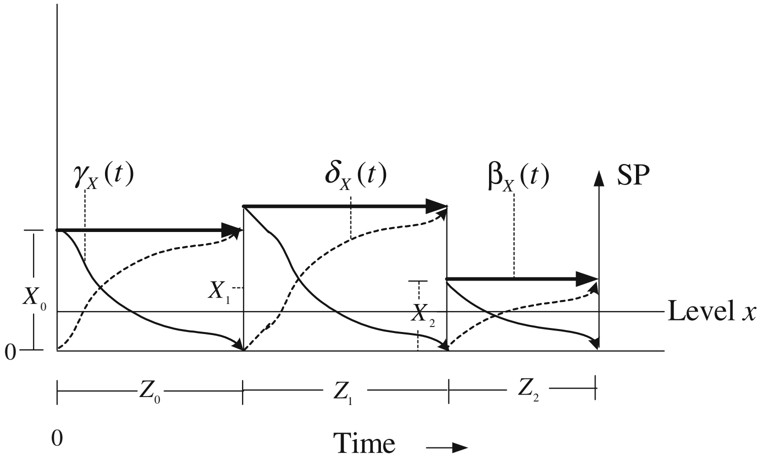


Fig. 10.1 Sample path of excess life $\gamma_X(t)$, age $\delta_X(t)$, total life $\beta_X(t)$. Also shows a state-space level x

economy of notation). Assume $dX(t)/dt = -r(X(t)), t \in [\tau_n, \tau_{n+1}), n = 0, 1, \ldots$, where $r(x) > 0, x > 0$. Thus $X(t)$ is a piecewise, decreasing deterministic function between jumps. Assume that for all $v > 0$,

$$\lim_{u\downarrow 0}\int_{y=u}^{v}\frac{1}{r(y)}dy < \infty, \tag{10.1}$$

which guarantees that a sample path $X(t)$, $t \geq 0$, starting from any level $v > 0$, returns to level 0 in a finite time. The process $\{X(t)\}_{t\geq 0}$ is a variant of the GI/G/r(·) dam subject to inputs $\{X_n\}_{n=0,1,\ldots}$ occurring at all instants τ_n when the dam becomes empty, and at no other time points. This mechanism can be thought of as that of a *replacement model*. New inputs replace the immediately preceding used-up inputs. Thus $\{X(t)\}_{t\geq 0}$ is never at level 0 for a positive duration, and $\tau_n, n = 1, 2, \ldots$, are replacement times.

Denote the inter-replacement times by $\{Z_n\}_{n=0,1,\ldots}$. The random variables Z_n and X_n are related by

$$Z_n = \int_{y=0}^{X_n}\frac{1}{r(y)}dy, n = 0, 1, \ldots. \tag{10.2}$$

From (10.2), Z_n is the *time required* for $\{X(t)\}_{t\geq 0}$ to descend from level X_n to level 0. The Z_n, $n = 0, 1,$ are i.i.d. because X_n, $n = 0, 1,$ are i.i.d.

10.1.2 Renewal Processes $\{Z_n\}_{n=0,1,...}$ and $\{X_n\}_{n=0,1,...}$

$\{Z_n\}_{n=0,1,...}$ is in one-to-one correspondence with $\{X_n\}_{n=0,1,...}$, and with the piecewise deterministic continuous efflux rate $r(X(t))$, $\tau_n < \tau_{n+1}$, $n = 0, 1, \ldots$.

Let $X_n \underset{dis}{=} X$ and $Z_n \underset{dis}{=} Z$, $n = 0, 1, \ldots$

Example 10.1 Consider a newly-installed battery at τ_0 with initial electrical charge $X_0 \underset{dis}{\equiv} X$, to power a device. Assume that the charge declines at a rate that depends on the present charge. That is, $dX(t)/dt = -r(X(t)) < 0$, $t \in [\tau_0, \tau_1)$. Suppose the battery's charge dissipates non-uniformly and descends to 0 after a time $\tau_1 = Z_0 \underset{dis}{\equiv} Z$. The battery is immediately replaced by a new fully-charged one. This procedure is repeated as batteries wear out, at times τ_n, $n = 2, 3, \ldots$. Thus $Z_n \underset{dis}{\equiv} Z$, and $X_n \underset{dis}{\equiv} X$, $n = 0, 1, 2, \ldots$, and

$$Z = \int_{y=0}^{X} \frac{1}{r(y)} dy, \tag{10.3}$$

is the inter-replacement time (see formula (6.4) in Sect. 6.2.4). The dimension of Z is $[Time]$. The dimension of X is $[Coulombs]$. The function $r\,(X(t))$ has dimension $[Coulomb][Time]^{-1}$.

10.1.3 The Renewal Process $\{X_n\}_{n=0,1,...}$

Let $\gamma_X(t) :=$ *excess life* of *content at instant* $t \geq 0$. The process $\{\gamma_X(t)\}_{t\geq 0}$ has the same sample path as $\{X(t)\}_{t\geq 0}$, since we assume that all input jumps start at level 0. Then $d(\gamma_X(t))/dt = -r(\gamma_X(t))$. Let $\delta_X(t) :=$ *age of the content at instant* $t \geq 0$, i.e., amount of content used up at instant t, from the last renewed amount prior to t. Then $d\,(\delta_X(t))\,/dt = +r(\delta_X(t))$. Let $\beta_X(t) :=$ *total life* (*span*) *of the latest renewed amount of content at* t (Fig. 10.1). (In Example 10.1, $\gamma_X(t))$, $\delta_X(t)$, $\beta_X(t)$ are respectively, the remaining charge, the charge used up, and the total charge, of the battery in use at time t.)

In the sample paths of the processes $\{\gamma(t)\}_{t\geq 0}$, $\{\delta(t)\}_{t\geq 0}$, $\{\beta(t)\}_{t\geq 0}$ all upward jumps start at level 0 and are $\underset{dis}{=} X$. All downward jumps start at a random level X and end at level 0.

Limiting Distributions in $\{X_n\}_{n=0,1,...}$ Using LC
We derive the *limiting* pdfs $f_{\gamma_X}(x)$, $f_{\delta_X}(x)$, $f_{\beta_X}(x)$, $x > 0$, of r.v.s $\gamma_X(t)$, $\delta_X(t)$, $\beta_X(t)$, as $t \to \infty$, assuming $E(X) < \infty$, which is the condition for their existence. Consider sample paths of $\{\gamma_X(t)\}$, $\{\delta_X(t)\}$, $\{\beta_X(t)\}$, $t \geq 0$ (Fig. 10.1).

Let $F_X(x)$, $f_X(x)$, μ_X be the cdf, pdf and expected value respectively of r.v. X. Let $\overline{F}_X(x) = 1 - F_X(x)$, $x \geq 0$.

Limiting PDF of Excess Life in $\{X_n\}_{n=0,1,...}$
Consider a sample path of $\{\gamma(t)\}_{t\geq 0}$. The long-run SP expected *downcrossing* rate of level $x > 0$, is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = r(x) f_{\gamma_X}(x). \tag{10.4}$$

(as in Corollary 6.2 in Sect. 6.2.8).

The long-run SP expected *upcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = \frac{1}{E(Z)} \cdot \overline{F}_X(x), \tag{10.5}$$

since the expected time between upward jumps starting from level 0 is $E(Z)(= E(\tau_{n+1} - \tau_n), n = 0, 1, \ldots)$, and $\overline{F}_X(x) = P(\textit{SP jump starting at level } 0 \textit{ is} > x)$. In (10.3), substituting from (10.2), and conditioning on $X = x$ gives

$$\begin{aligned} E(Z) &= \int_{x=0}^{\infty} \left(\int_{y=0}^{x} \frac{1}{r(y)} dy \right) f_X(x)dx = \int_{y=0}^{\infty} \int_{x=y}^{\infty} \frac{1}{r(y)} f_X(x)dxdy \\ &= \int_{y=0}^{\infty} \frac{1}{r(y)} \left(\int_{x=y}^{\infty} f_X(x)dx \right) dy = \int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy. \end{aligned} \tag{10.6}$$

Equating (10.4) and (10.5) by the principle of rate balance across level x, and using (10.6), yields the equation

$$r(x) f_{\gamma_X}(x) = \frac{\overline{F}_X(x)}{E(Z)} = \frac{\overline{F}_X(x)}{\int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}, \tag{10.7}$$

$$f_{\gamma_X}(x) = \frac{\overline{F}_X(x)}{r(x) \int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}. \tag{10.8}$$

The dimension of $f_{\gamma_X}(x)$ is $[content]^{-1}$ (= $[Coulomb]^{-1}$ in Example 10.1)).

Limiting PDF of Excess Life in $\{X_n\}_{n=0,1,...}$ when $r(x) \equiv 1$
If the efflux rate $r(x) \equiv 1$, formula (10.8) reduces to

$$f_{\gamma_X}(x) = \frac{\overline{F}_X(x)}{\int_{y=0}^{\infty} \overline{F_X}(y)dy} = \frac{\overline{F}_X(x)}{\mu_X}, \tag{10.9}$$

since $\int_{y=0}^{\infty} \overline{F}_X(y)dy = E(X) = \mu_X$. ($\gamma_X$ represents the limiting excess life of *content*, having pdf $f_{\gamma_X}(x)$.) Formula (10.9) is exactly the same as the well-known limiting pdf of the excess life in a "standard" renewal process. However, here the dimension of $f_{\gamma_X}(x)$ is $[content]^{-1}$ instead of $[Time]^{-1}$.

Limiting PDF of Age in $\{X_n\}_{n=0,1,...}$
For the age process $\{\delta_X(t)\}_{t\geq 0}$, the long-run SP expected *upcrossing rate of level* $x > 0$ is

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = +r(x) f_{\delta_X}(x), \tag{10.10}$$

(see Corollary 6.2 in Sect. 6.2.8). The long-run SP *expected downcrossing rate of level* x is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \frac{1}{E(Z)} \int_{y=x}^{\infty} f_X(y)dy = \frac{\overline{F}_X(x)}{E(Z)}, \tag{10.11}$$

because (1) downward jumps occur at rate $1/E(Z)$, (2) in order for the SP to downcross level x by a jump at some τ_n^-, the *upward* jump at τ_{n-1} starting from level 0 must have been such that $X_{n-1} > x$. Moreover, X_{n-1} at τ_{n-1} is also equal to the downward jump size at τ_n^- (see Fig. 10.1).

Equating (10.10) and (10.15) (rate balance across level x), gives

$$r(x) f_{\delta_X}(x) = \frac{\overline{F}_X(x)}{E(Z)} = \frac{\overline{F}_X(x)}{\int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy};$$

$$f_{\delta_X}(x) = \frac{\overline{F}_X(x)}{r(x) \int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}. \tag{10.12}$$

Comparing (10.8) with (10.12) shows that $f_{\delta_X}(x) = f_{\gamma_X}(x)$. The dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Age in $\{X_n\}_{n=0,1,...}$ when $r(x) \equiv 1$
If $r(x) \equiv 1$, we obtain similarly as in (10.9), the limiting pdf

$$f_{\delta_X}(x) = \frac{\overline{F}_X(x)}{\mu_X}. \tag{10.13}$$

The dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$. It is well known that for an ordinary renewal process, the limiting distributions of the excess life and age are identical. In the variant of a GI/G/r(·) dam possessing the renewal structure outlined here, these distributions are also identical with respect to the content, even when the efflux rate has a general form $r(x)$, $x > 0$. That is, formulas (10.8) and (10.12) are identical.

Limiting PDF of Total Life in $\{X_n\}_{n=0,1,...}$
For the process $\{\beta_X(t)\}_{t\geq 0}$, the long-run SP expected *downcrossing rate of level* $x > 0$, is

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t}=\int_{y=x}^{\infty}\left(\frac{1}{\int_{u=0}^{y}\frac{1}{r(u)}du}\right)f_{\beta_X}(y)dy, \tag{10.14}$$

where we have conditioned on $\beta_X(t) = y > x$. In (10.14), $1/\int_{u=0}^{y}\frac{1}{r(u)}du$, *the reciprocal of the expected sojourn time of* $\{\beta_X(t)\}_{t\geq 0}$ *at level* $y > x$, is equal to the *downward jump rate* across level x starting from level y (Fig. 10.1). At the end of a level-y ($y > x$) sojourn time, the SP jumps downward to level 0, and *downcrosses every state-space level* in $(0, y)$, including level x.

The SP long-run (expected) *upcrossing* rate of level x is

$$\lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}=\frac{\overline{F}_X(x)}{E(Z)}=\frac{1}{E(Z)}\int_{y=x}^{\infty}f_X(y)dy, \tag{10.15}$$

since the expected time between SP upward jumps *out of level* 0 is $E(Z)$, and the probability that such an SP jump exceeds level x is $\overline{F}_X(x)$. The SP makes a *double jump* in opposite directions at each renewal instant of the sequence $\{Z_n\}_{n=0,1,...}$. One jump is downward ending at level 0; the "opposite jump" is upward starting at level 0.

Equating (10.14) and (10.15) (rate balance across level x), results in the integral equation for $f_{\beta_X}(\cdot)$,

$$\int_{y=x}^{\infty}\frac{1}{\left(\int_{u=0}^{y}\frac{1}{r(u)}du\right)}f_{\beta_X}(y)dy=\frac{\overline{F_X}(x)}{E(Z)}. \tag{10.16}$$

In (10.16), taking d/dx on both sides yields

$$-\frac{1}{\left(\int_{u=0}^{x}\frac{1}{r(u)}du\right)}f_{\beta_X}(x)=-\frac{f_X(x)}{E(Z)}.$$

Hence

$$f_{\beta_X}(x) = \frac{\left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(x)}{E(Z)} = \frac{\left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(x)}{\int_{y=0}^{\infty} \frac{\overline{F}(y)}{r(y)} dy}. \tag{10.17}$$

The dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Total Life in $\{X_n\}_{n=0,1,...}$ when $r(x) \equiv 1$
If $r(x) \equiv 1$, then $Z_n = X_n$ and $E(Z_n) = E(X_n) = \mu_X$ in *magnitude*. (However, their dimensions differ; since $[X_n] = [contentunit]$ and $[Z_n] = [Time]$.) Formula (10.17) reduces to the well-known limiting pdf of total life (span) for an *ordinary* renewal process,

$$f_{\beta_X}(x) = \frac{x f_X(x)}{E(Z)} = \frac{x f_X(x)}{\mu_X}, \tag{10.18}$$

except that the dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$ instead of $[Time]^{-1}$. That is, in the variant of the GI/G/r(·) dam described here, the "life" is measured in 'content' units.

Remark 10.1 The foregoing variant of GI/G/r(·) exhibits the phenomenon of SP multiple jumps at the same (renewal) instant. Recall that SP jumps in the state space **do not occur in Time**. (See Examples 2.2, 2.3 in Sect. 2.3, regarding SP multiple jumps.)

Example 10.2 Suppose $r(x) = kx$, $x > 0$, where $k > 0$ is a constant. Then the inequality (10.1) does not hold because sample paths decay as a negative exponential function (see Sect. 6.4). The SP returns to every level $x > 0$, however small. Let us choose a small level $\varepsilon > 0$ to indicate that it is time for a new replenishment of content. Thus, whenever the content hits level ε from above, it increases by an amount $\underset{dis}{=} X$. We consider two cases.

Case 1: All replenishments start at level 0.
Whenever the content decays to level ε, a new replenishment starts at level 0. (We assume that the added amount is greater than ε; otherwise it is discarded.) Many systems are of this type. For example, heat and smoke alarms make a beep and/or show a red light, when the charge in the alarm's battery decays to a certain level. This signals that the battery needs replacing. In Example 10.1, this corresponds to replacing a battery with a new one when the preceding charge decreases to ε Coulombs.

Case 2: All replenishments start at level $\varepsilon > 0$
For each positive $v > \varepsilon$, the time to decay to level ε is

$$\int_{y=\varepsilon}^{v} \frac{1}{kx} dx = \frac{1}{k} \ln \frac{v}{\varepsilon} < \infty,$$

meaning the content returns to level ε in a finite time from any level $\varepsilon + X$. In this case, $\{X_n\}_{n=0,1,...}$ is a delayed renewal process, where the first inter-renewal time is $X_0 \underset{dis}{=} X - \varepsilon$, because all future inputs (replenishments) will start at level ε. The inter-renewal amounts $X_1, X_2, \ldots$ are $\underset{dis}{=} X$. In the renewal process $\{X_n\}_{n=0,1,...}$, the support of X_0 equals the support of X minus ε, because the remaining inputs start at level ε. This model structure is similar to that in the example in Sect. 10.3.8. We will not continue the analysis here; the limiting and time-t distributions of the quantities of interest can be derived from the reasoning given below in Sect. 10.3.8. This model would apply to cases where it is important to maintain the concentration of a solute in a solution above a certain level (say ε). Examples include: pharmacokinetics (see Sect. 11.6); consumer response to nonuniform advertisements (see [40]); and many others.

10.1.4 The Renewal Process $\{Z_n\}_{n=0,1,...}$

Limiting Excess Life, Age, Total Life in $\{Z_n\}_{n=0,1,...}$
Let $\gamma_Z(t)$, $\delta_Z(t)$, $\beta_Z(t)$ denote the excess life, age, and total life respectively, at a fixed time $t > 0$.

Define $\mathcal{G}(x) := \int_{y=0}^{x} \frac{1}{r(y)} dy :=$ *time for* $\{X(t)\}_{t \geq 0}$ *to decay from level* $x > 0$ *to level* 0. Then $\mathcal{G}(x)$ is an increasing differentiable function of x (since $d\mathcal{G}(x)/dx = 1/r(x)$), which implies the inverse $\mathcal{G}^{-1}(x)$ of $\mathcal{G}(x)$ exists, and

$$\frac{d}{dx}\mathcal{G}^{-1}(x) = \frac{1}{\frac{d}{dx}\mathcal{G}(x)} = \frac{1}{\frac{1}{r(x)}} = r(x), x > 0$$

(see, e.g., pp. 206–207 in [137], and other Calculus texts). Thus $\mathcal{G}^{-1}(x)$ is also an increasing (differentiable) function of x. Moreover, $\mathcal{G}^{-1}(x)$ is that level in the Time-state space $\boldsymbol{T} \times \lfloor 0, \infty)$ from which a descent to level 0 takes time period x, as $\mathcal{G}(\mathcal{G}^{-1}(x)) = x$. We now derive the pdfs of $\gamma_Z, \delta_Z, \beta_Z$ from the results for the pdfs of $\gamma_X, \delta_X, \beta_X$, respectively, given in Sect. 10.1.3.

Limiting PDF of Excess Life γ_Z in $\{Z_n\}_{n=0,1,\ldots}$
The relation between Z_n and $X(t)$ implies

$$\gamma_Z \le x \text{ iff } \gamma_X \le \mathcal{G}^{-1}(x)).$$

(see Fig. 10.1). Hence

$$F_{\gamma_Z}(x) = F_{\gamma_X}(\mathcal{G}^{-1}(x)). \tag{10.19}$$

Taking d/dx on both sides of (10.19) and substituting from (10.8) gives

$$\begin{aligned} f_{\gamma_Z}(x) &= f_{\gamma X}(\mathcal{G}^{-1}(x)) \cdot \frac{d}{dx}\mathcal{G}^{-1}(x) = f_{\gamma X}(\mathcal{G}^{-1}(x)) \cdot r(x) \\ &= \frac{r(x) \cdot \overline{F}_X(\mathcal{G}^{-1}(x))}{r(\mathcal{G}^{-1}(x)) \int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}. \end{aligned} \tag{10.20}$$

The dimension of $f_{\gamma_Z}(x)$ is $[Time]^{-1}$.
If $r(y) \equiv 1$, then $\mathcal{G}(x) = \mathcal{G}^{-1}(x) = x$ and

$$f_{\gamma_Z}(x) = \overline{F}_X(x) / \int_{y=0}^{\infty} \overline{F}_X(y) dy = \overline{F}_X(x)/\mu_X = f_{\gamma X}(x).$$

The dimension of $f_{\gamma_Z}(x)$ is $[Time]^{-1}$, whereas the dimension of $f_{\gamma X}(x)$ is $[content]^{-1}$.

Limiting PDF of Age δ_Z in $\{Z_n\}_{n=0,1,\ldots}$
In a similar manner as for the excess life, the age satisfies

$$\delta_Z \le x \text{ iff } \delta_X \le \mathcal{G}^{-1}(x).$$

Thus, $F_{\delta_Z}(x) = F_{\delta_X}(\mathcal{G}^{-1}(x))$. Taking $\frac{d}{dx}$ then yields

$$f_{\delta_Z}(x) = \frac{r(x)\overline{F}_X(\mathcal{G}^{-1}(x))}{r(\mathcal{G}^{-1}(x)) \int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}. \tag{10.21}$$

Thus $f_{\delta_Z}(x) = f_{\gamma_Z}(x)$, $x > 0$. The dimension of $f_{\delta_Z}(x)$ is $[Time]^{-1}$.
If $r(y) \equiv 1$ then $\mathcal{G}(x) = \mathcal{G}^{-1}(x) = x$, and

$$f_{\delta_Z}(x) = \overline{F}_X(x) / \int_{y=0}^{\infty} \overline{F}_X(y) dy = f_{\delta_X}(x).$$

The dimension of $f_{\delta_Z}(x)$ is $[Time]^{-1}$, whereas the dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Total Life β_Z in $\{Z_n\}_{n=0,1,...}$
Since $\beta_Z \le x$ iff $\beta_X \le \mathcal{G}^{-1}(x)$ then we obtain similarly as for $f_{\delta_Z}(x)$ and $f_{\gamma_X}(x)$ above,

$$f_{\beta_Z}(x) = f_{\beta_X}(\mathcal{G}^{-1}(x)) \cdot \frac{d}{dx}\mathcal{G}^{-1}(x) = f_{\beta_X}(\mathcal{G}^{-1}(x)) \cdot r(x).$$

From (10.17) we get

$$f_{\beta_Z}(x) = \frac{r(x) \cdot \left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(\mathcal{G}^{-1}(x))}{\int_{y=0}^{\infty} \frac{\overline{F}_X(y)}{r(y)} dy}. \tag{10.22}$$

The dimension of $f_{\beta_Z}(x)$ is $[Time]^{-1}$ whereas the dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$. If $r(x) \equiv 1$ then

$$f_{\beta_Z}(x) = \frac{x f_X(x)}{\int_{y=0}^{\infty} \overline{F}_X(y)dy} = \frac{x f_X(x)}{\mu_X},$$

having dimension $[Time]^{-1}$.

10.1.5 Limiting PDFs in Ordinary Renewal Process

We now give the steady-state pdfs of excess, age and total life for the ordinary (i.e., standard) renewal process as a *special case* of those for the replacement model above. In the *ordinary* renewal process, we have $X_n = Z_n, n = 0, 1, 2, ...,$ since $r(X(t)) \equiv 1$ (see Fig. 10.2). The dimensions of X_n and Z_n are the same, usually $[Time]$. The pdfs of the excess, age and total life, i.e., $f_{\gamma_Z}(x)$, $f_{\delta_Z}(x)$, $f_{\beta_Z}(x)$, $x > 0$, are the same as formulas (10.9), (10.13), (10.18) respectively, and all have dimension $[Time]^{-1}$.

Direct Derivation of Limiting PDFs $f_{\gamma_Z}(x)$, $f_{\delta_Z}(x)$, $f_{\beta_Z}(x)$
We can derive these limiting pdfs very simply and directly in the ordinary renewal process. For example, to get $f_{\gamma_Z}(x)$, $X > 0$, we examine the sample path of $\gamma_Z(t)$, $t \ge 0$, in Fig. 10.2. The downcrossing rate of level x is $f_{\gamma_Z}(x)$; the upcrossing rate of level x is $\overline{F}_Z(x)/E(Z)$. Rate balance gives $f_{\gamma_Z}(x) = \overline{F}_Z(x)/E(Z)$. Similarly, examining the sample path of $\delta_Z(t), t \ge 0$, gives $f_{\delta_Z}(x) = \overline{F}_Z(x)/E(Z)$, $x > 0$. To derive $f_{\beta_Z}(x)$, $x > 0$,

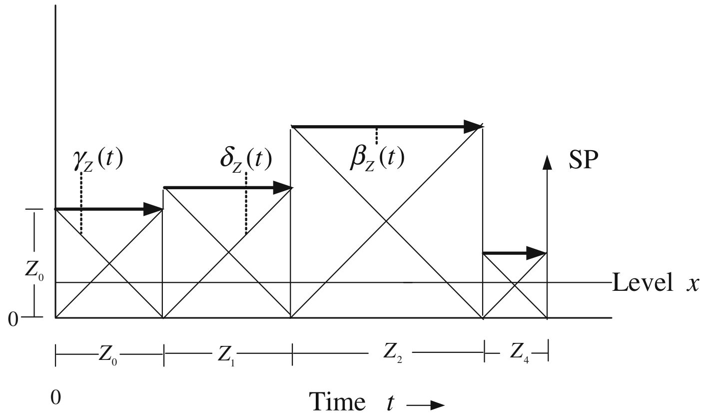


Fig. 10.2 Sample paths of excess $\gamma_Z(t)$, age $\delta_Z(t)$ and total life $\beta_Z(t), t \geq 0$, in the ordinary (standard) renewal process

examine the sample path of $\beta_Z(t), t \geq 0$. The downcrossing rate of level x is $\int_{y=x}^{\infty} (1/y) f_{\beta_Z}(y)dy$ and the upcrossing rate of level x is $\overline{F}_Z(x)/E(Z)$. Rate balance gives $\int_{y=x}^{\infty} (1/y) f_{\beta_Z}(y)dy = \overline{F}_Z(x)/E(Z)$. Taking d/dx of both sides yields $f_{\beta_Z}(x) = x f_Z(x)/E(Z)$.

Remark 10.2 The LC derivations of the limiting pdfs of excess life, age and total life, at time t as $t \to \infty$, are **relatively easy** in the replacement model, and are **much simpler** for the ordinary (standard) renewal process.

Remark 10.3 All the derivations in Sect. 10.1 are based directly on the author's unpublished notes of June 18–July 26, 1992 [28]. These notes were motivated by a talk on the ordinary renewal process by van Harn and Steutel (see Partial Bibliography below) at the 21st Conference on Stochastic Processes and their Applications at York University, Toronto, June 15–19, 1992. Their presented work differs completely from LC conceptually. Results for the **ordinary renewal process** using LC were published independently in Katayama (2002) (see Partial Bibliography).

10.2 A Renewal Problem with Barrier

Consider a renewal process $\{Z_n\}_{n=1,2,...}$, where $Z_n \underset{dis}{=} U_{(0,1)} :=$ *uniform random variable on* (0, 1) (Fig. 10.3). Let N_K denote the number of renewals required to *first exceed* a barrier $K > 0$. In this section we derive the expected

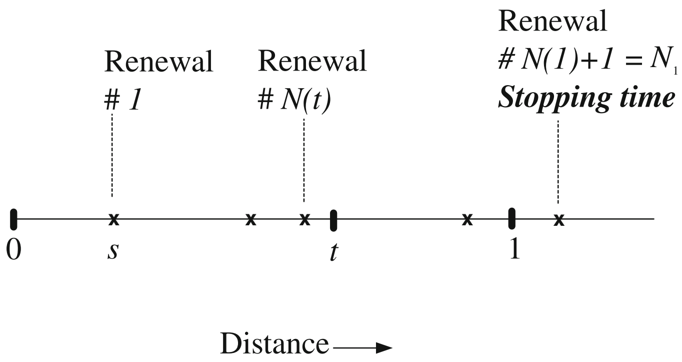


Fig. 10.3 Renewal process $\{Z_n\}_{n=1,2...}$ showing renewals. $N(t)$ is the number of renewals within $(0, t)$. $N_1 = N(1) + 1$ is number of renewals required to first exceed barrier $K = 1$. N_1 is a stopping time for the sequence $\{Z_n\}_{n=1,2,...}$ where $Z_n \underset{dis}{=} U_{(0,1)}$

value $E(N_K)$, $K = 1, 2, 3, \ldots$, and related quantities. It is well known that $E(N_1) = e$, the base of natural logarithms (see Problem 5, p. 485 in [125]). Usually, it is shown that $E(N_1) = e$ by a standard renewal argument, i.e., conditioning on the first renewal distance s (Fig. 10.3), deriving a renewal equation, and solving it. However, the general formula for $E(N_K)$, $K = 2, 3,$ ... is not well known.

Here we derive $E(N_1)$ by an alternative method, which also applies to derive $E(N_K)$, $K = 2, 3, \ldots$. The idea is to embed statistically independent replicas of the one-dimensional renewal process $\{Z_n\}_{n=1,2,....}$ into the cycles of a regenerative process such that the time axis of the embedded processes is perpendicular to the time axis of the regenerative process. Thus, the one-dimensional process $\{Z_n\}_{n=1,2,....}$ becomes transformed into an infinite sequence of statistically independent copies of $\{Z_n\}_{n=1,2,...}$, in a two-dimensional construct having two different perpendicular time axes. One time axis is for the regenerative process; the other is for $\{Z_n\}_{n=1,2,....}$. The type of construction in this alternative method, facilitates finding the expected number of renewals required to exceed a barrier or threshold, in other (seemingly unrelated) stochastic models as well.

10.2.1 Method for $E(N_K)$ Using a Regenerative Process

We construct a continuous-time continuous-state positive recurrent regenerative process

$$\{X(t)\}_{t\geq 0}, \qquad X(0) = 0,$$

which embeds statistically independent reproductions of $\{Z_n\}_{n=1,2,\ldots}$ in all cycles of $\{X(t)\}_{t\geq 0}$ (Fig. 10.4). A sample path of $\{X(t)\}_{t\geq 0}$ is a non-decreasing step function, which makes SP *upward* jumps of size $\underset{dis}{=} U_{(0,1)}$ at an ***arbitrary Poisson rate*** λ. (*We select* $\lambda = 1$, for convenience.) The upward jumps are denoted by

$$b_n := Z_{n:} \underset{dis}{\equiv} U_{(0,1)}, n = 1, 2, \ldots .$$

(We replace symbol Z_n by b_n for generality beyond the threshold $K = 1$, and because of applicability to other models. See [33]).

Define random variable N_K by

$$N_K = \min\{n| \sum_{i=1}^{n} b_i > K\}, K = 1, 2, \ldots ; \tag{10.23}$$

thus N_K is a *stopping time for the sequence* $\{b_n\}_{n=1,2,\ldots}$.

Let random variable $a \underset{dis}{=} \text{Exp}_\lambda = \text{Exp}_1$, implying $E(a) = 1$.

Define random variable c by

$$c = \sum_{i=1}^{N_K} a_i, \qquad \text{where each } a_i \underset{dis}{=} a, \tag{10.24}$$

and the a_is are i.i.d. r.v.s.

Let $\{c_n\}_{n=1,2,\ldots}$ be a renewal process where $c_n \underset{dis}{\equiv} c$; the c_ns are i.i.d. Then $\{c_n\}_{n=1,2,\ldots}$ are "compound" cycles of a regenerative process with sub-components $\{a_i\}_{n=1,2,\ldots}$. Since there is a one-to-one correspondence between a_n and b_n, $n = 1, 2, \ldots$, the random variable N_K is also a stopping time for the sequence $\{a_i\}_{n=1,2,\ldots}$. Taking the expected value in (10.24) yields

$$E(c) = E(N_K)E(\alpha) = E(N_K), \tag{10.25}$$

by Wald's equation (e.g., see Exercises 13–24, p. 486–489 in [125]).

Just after each instant when a sample path of $\{X(t)\}_{t\geq 0}$ upcrosses level K, *the SP jumps downward* (rebounds) *immediately to level* 0, *and the process* $\{X(t)\}_{t\geq 0}$ *restarts*. Our construction guarantees that the limiting distribution of $X(t)$ exists as $t \to \infty$ (see [132]). Random variable N_K equals the number of SP jumps required for $\{X(t)\}_{t\geq 0}$ to first exceed level K. A simple, but key observation, is that N_K is equal to the number of sub-intervals with lengths $\underset{dis}{=}$ a, comprising a cycle c. The state space of $\{X(t)\}_{t\geq 0}$ is $\boldsymbol{S} = [0, K+1)$, because the excess of the jumps that exceed level K is less than 1 (due to jump sizes $\underset{dis}{=} U_{(0,1)}$).

Relation to $\langle s, S\rangle$ Inventory with No Decay
Other stochastic models have a related structure. For example, the $\langle s, S\rangle$ inventory *with no decay* in Sect. 6.9 is the "flip" (like $\updownarrow$) of the $\{X(t)\}_{t\geq 0}$ process, where $K := S - s$, and the jump sizes are distributed as Exp_μ. In the $\langle s, S\rangle$ model $E(N_K)$ $(= E(N_{S-s}))$ is the expected number of demands in an ordering cycle.

10.2.2 Derivation of $E(N_1)$

Let the limiting mixed pdf of $\{X(t)\}_{t\geq 0}$ as $t \to \infty$, be $\{\pi_1, f_0(x)\}_{0<x<1}$. Consider a sample path of $\{X(t)\}_{t\geq 0}$; fix level $x \in (0,1)$ (Fig. 10.4). SP upcrossings of level x are due to jumps $\underset{dis}{=} U_{(0,1)}$ starting at level 0, or starting at some level $y \in (0, x)$. Thus, the SP upcrossing rate of level x is

$$1 \cdot \pi_1 \cdot P(b > x) + 1 \cdot \int_{y=0}^{x} P(b > x - y) \cdot f_0(y)dy$$

where $b \underset{dis}{\equiv} U_{(0,1)}$, and upward jumps occur at rate $1/E(a) = \lambda = 1$. This leads to the equation

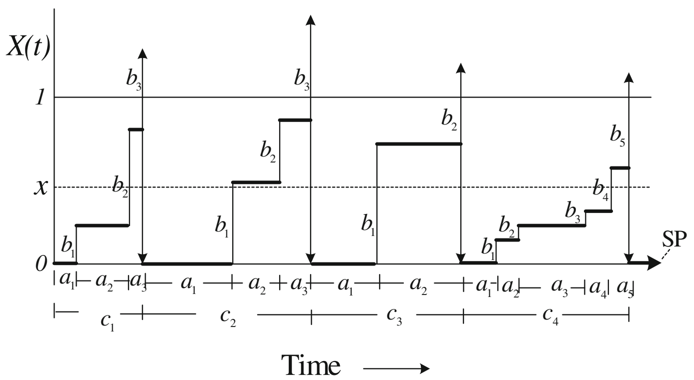


Fig. 10.4 Sample path of $\{X(t)\}_{t\geq 0}$, in renewal problem to determine $E(N_1)$ when renewal times are $\underset{dis}{=} U_{(0,1)}$

$$1 \cdot \pi_{0.1} \cdot P(b > x) + 1 \cdot \int_{y=0}^{x} P(b > x - y) \cdot f_0(y)dy = \pi_1, x \in (0, 1)\,, \tag{10.26}$$

explained as follows. On the right side of (10.26), π_1 is the downcrossing rate of level x because the rate of SP downward jumps is the same as the rate of SP *entrances into state* $\{0\}$ *from above* (also called here 'downcrossings' of level 0). From the principle of *set balance*, this entrance rate is equal to the exit rate of {0}, namely $\lambda\pi_1 = 1 \cdot \pi_1 = \pi_1$.

The SP downcrossing rate of every level $x \in [0, 1)$ is equal to the total *upcrossing rate of level* 1. The SP rebounds into level 0 immediately after it upcrosses level 1. (The SP makes a *double jump*. Compare with the $\langle s, S\rangle$ inventory with no decay in Example 2.3 in Sect. 2.3; see also Sect. 2.3.) In the inventory model, whenever the stock on hand jumps below the reorder point s, it is replenished immediately up to level S.

Letting $x = 1$ in equation (10.26) gives

$$\pi_1 \cdot P(b > 1) + \int_{y=0}^{1} P(b > 1 - y) \cdot f_0(y)dy = \pi_1. \tag{10.27}$$

Since $b \underset{dis}{=} U_{(0,1)}$, we substitute $P(b > x) = 1 - x$, $0 < x < 1$, into (10.26), resulting in

$$\pi_{0,0}(1 - x) + \int_{y=0}^{x} (1 - x + y) f(y)dy = \pi_{0,0}, 0 < x < 1. \tag{10.28}$$

Taking d/dx twice in (10.28), and solving the resulting ordinary differential equation gives

$$f(x) = \pi_1 e^x, 0 < x < 1, \tag{10.29}$$

which we substitute into the normalizing condition $\pi_0 + \int_{x=0}^{1} f(x)dx = 1$, giving

$$\pi_1 = \frac{1}{e}. \tag{10.30}$$

The renewal rate of $\{c_n\}_{n=1,2,...}$ is $1/E(c)$ = SP entrance (or exit) rate of $\{0\} = \pi_1$. Thus $E(c) = 1/\pi_1$. From (10.25) and (10.30),

$$E(N_1) = E(c) \cdot E(a) = \frac{1}{\pi_1} \cdot 1 = e = 2.71828. \tag{10.31}$$

We have derived $E(N_1)$ in detail using the compound-cycle regenerative process structure, to fix ideas. The following values of $E(N_K)$, $K = 2, 3, \ldots$, in this Section are not well known.

10.2.3 Derivation of $E(N_2)$

Let $\pi_2 := \lim_{t\to\infty} P(X(t) = 0)$. Let the steady-state PDF of $\{X(t)\}_{t\geq 0}$ be

$$f(x) = f_0(x)\boldsymbol{I}_{(0,1]}(x) + f_1(x)\boldsymbol{I}_{[1,2)}(x), x \in (0, 2),$$

where $\boldsymbol{I}_{\boldsymbol{A}}(x) = 1$ if $x \in \boldsymbol{A}$, and $\boldsymbol{I}_{\boldsymbol{A}}(x) = 0$ if $x \notin \boldsymbol{A}$ (the characteristic function of set $\boldsymbol{A}$).

Consider a sample path of $\{X(t)\}_{t\geq 0}$ (Fig. 10.5), where the state space is $\boldsymbol{S} = [0, 3)$. Balancing SP up- and downcrossing rates of $x \in (0, 1)$, as in the case $K = 1$, gives

$$\pi_2(1 - x) + \int_{y=0}^{x} (1 - x + y) f_0(y)dy = \pi_2, x \in (0, 1). \tag{10.32}$$

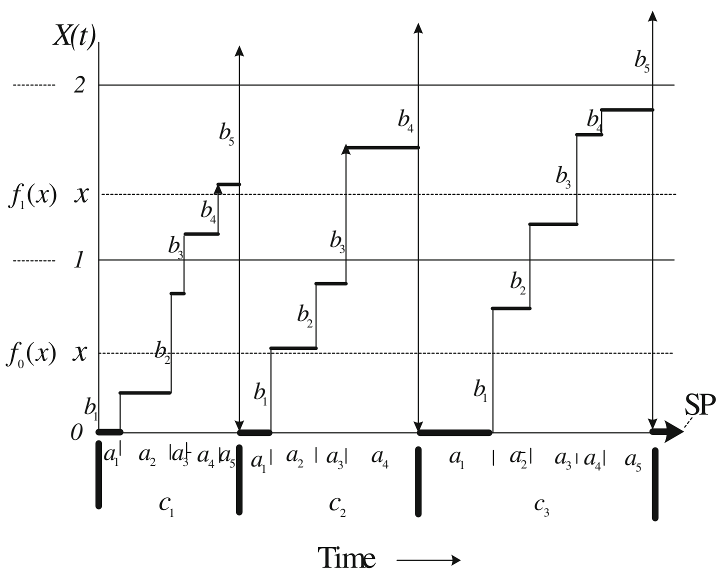


Fig. 10.5 Sample path of $\{X(t)\}_{t\geq 0}$ for renewal problem, with state space $\boldsymbol{S} = [0, 3)$. Facilitates solution for $E(N_2)$

Fix $x \in [1, 2)$ (see Fig. 10.5). Balancing SP up- and downcrossing rates of x, gives

$$\int_{y=x-1}^{1} (1-x+y) f_0(y)dy + \int_{y=1}^{x} (1-x+y) f_1(y)dy = \pi_2, x \in [1,2)\,. \tag{10.33}$$

In (10.33), the first integral is the upcrossing rate of level x due to jumps starting in $(0, 1)$. The lower limit is $y = x - 1$ because an SP jump can upcross level $x \in (1, 2)$ only if it starts in interval $(x - 1, 1)$, which is a subset of $[0, 1)$. The second integral is the upcrossing rate of level x due to jumps starting in $[1, x)$.

Taking d/dx in (10.33) gives

$$-\int_{y=x-1}^{1} f_0(y)dy - 0 - \int_{y=1}^{x} f_1(y)dy + f_1(x) = 0, x \in [1,2). \tag{10.34}$$

Substituting $\pi_2 e^y$ from (10.29), for $f_0(y)$ in (10.34) with π_1 replaced by π_2, and letting $x \downarrow 1$ yields

$$f_1(1^+) = \pi_2\,(e-1) = \pi_2 e - \pi_2 = f_0(1^-) - \pi_2.$$

which shows that $f(x)$ has a discontinuity at $x = 1$ of size π_2. (In $f_0(x)$, π_2 replaces π_1 because at this stage we are solving for $E(N_2)$). Taking d/dx in (10.34) gives

$$\begin{aligned} f_1'(x) - f_1(x) &= -f_0(x-1) = -\pi_2 e^{(x-1)}, \\ \frac{d}{dx}(e^{-x} f_1(x)) &= -\pi_2 e^{-1}, \\ f_1(x) &= -\pi_2 e^{-1} x e^x + C e^x, x \in [1,2), \end{aligned} \tag{10.35}$$

where C is a constant, evaluated by letting $x \downarrow 1$ in (10.35), resulting in

$$f_1(1^+) = -\pi_2 + Ce = \pi_2 e - \pi_2,$$
$$\text{and} \quad C = \pi_2.$$

Substituting $C = \pi_2$ into (10.35) gives

$$f_1(x) = \pi_2 \left(-e^{-1}x + 1\right) e^x, x \in [1,2).$$

Thus we obtain

$$f_0(x) = \pi_2 e^x, x \in (0, 1),$$
$$f_1(x) = \pi_2(1 - e^{-1}x)e^x, x \in [1, 2. \tag{10.36}$$

From (10.36) we check the discontinuity at $x = 1$,

$$f_1(1^+) = \pi_2 e - \pi_2 = f_0(1^-) - \pi_2,$$

Fig. 10.6 shows that the discontinuity stays in place at $x = 1$ for $K = 2, 3, \ldots$, but with the sizes π_K, $K = 2, 3, \ldots$ respectively.

The normalizing condition is

$$\pi_2 + \int_{x=0}^{1} f_0(x)dx + \int_{x=1}^{2} f(x)dx = 1. \tag{10.37}$$

Substituting from (10.36) into (10.37) gives

$$\pi_2 = \frac{1}{-e + e^{2.}}. \tag{10.38}$$

From (10.25),

$$E(N_2) = E(c)E(a) = \frac{1}{\pi_2} \cdot 1 = -e + e^2 = 4.67077. \tag{10.39}$$

10.2.4 Derivation of $E(N_3)$

We now explore further the pattern of $E(N_K)$, $K = 1, 2, \ldots$. For deriving $E(N_3)$, the state space is $\boldsymbol{S} = [0, 4)$. Let $\pi_3 := \lim_{t\to\infty} P(X(t) = 0)$. Let the steady state pdf of $\{X(t)\}_{t\geq 0}$ be

$$f(x) = f_0(x)\boldsymbol{I}_{(0,1]}(x) + f_1(x)\boldsymbol{I}_{[1,2)}(x) + f_2(x)\boldsymbol{I}_{[2,3)}(x), x \in (0, 3),$$

(plotted as $f(x)/\pi_3$ in Fig. 10.6 since π_3 is a factor of each $f_j(x)$, $j = 0, 1, 2$—see (10.43) below). We now balance SP up- and downcrossing rates across arbitrary levels $x \in (0, 1)$; $x \in [1, 2)$; $x \in [2, 3)$, which gives respectively, integral equations

$$\pi_3(1 - x) + \int_{y=0}^{x} (1 - x + y) f_0(y)dy = \pi_3, \tag{10.40}$$

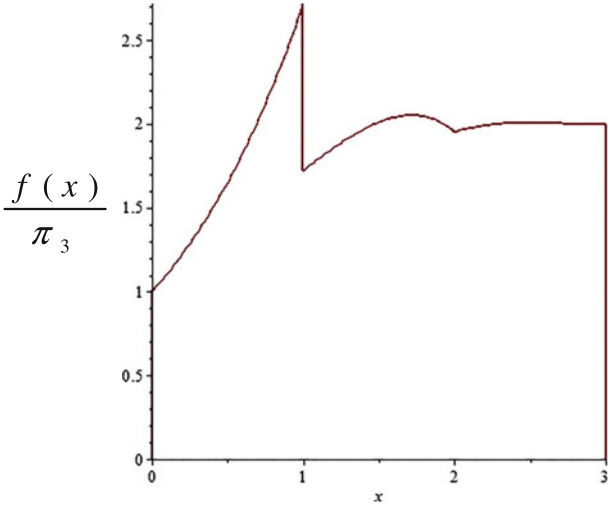


Fig. 10.6 Plot of $f(x)/\pi_3, x \in (0,3). E(N_3) = 1/\pi_3 = 1 + \int_0^3 f(x)dx = 1 + \sum_{j=0}^{2} \int_j^{j+1} f_j(x)dx$. The only discontinuity of $f(x)$ is at $x = 1$, of size $= \pi_3$; $f(x)$ is continuous for all $x \in (1,3)$

$$\int_{y=x-1}^{1} (1-x+y) f_0(y)dy + \int_{y=1}^{x} (1-x+y) f_1(y)dy = \pi_3, \quad (10.41)$$

$$\int_{y=x-1}^{2} (1-x+y) f_1(y)dy + \int_{y=2}^{x} (1-x+y) f_2(y)dy = \pi_3. \quad (10.42)$$

Solving integral equations (10.40), (10.41) in a similar manner as for $K = 1, 2$ above, gives

$$\begin{aligned} f_0(x) =& \pi_3 e^x, x \in (0,1), \\ f_1(x) =& \pi_3 (1 - e^{-1}x)e^x, x \in [0,1, \\ f_2(x) =& \frac{1}{2}\pi_3(-2xe^{-2} + e^{-2}x^2 - 2xe^{-1} + 2)e^x, x \in [2,3). \end{aligned} \quad (10.43)$$

The normalizing condition is

$$\pi_3 + \int_{x=0}^{1} f_0(x)dx + \int_{x=1}^{2} f_1(x)dx + \int_{x=2}^{3} f_2(x)dx = 1, \quad (10.44)$$

yielding

$$\pi_3 = \frac{1}{\frac{1}{2}e - 2e^2 + e^3}.$$

Substituting from (10.43) into (10.44) gives

$$E(N_3) = \frac{1}{\pi_3} = \frac{1}{2}e - 2e^2 + e^3 = 6.66656563. \tag{10.45}$$

The form of $f(x)$ is reminiscent of the pdf of wait in the M/D/1 queue in Fig. 3.18 in Sect. 3.10.2. In both models, the pdf has a discontinuity at exactly one point: at $x = 1$ in the present renewal problem, and at $x = D$ in the M/D/1 queue.

10.2.5 Derivation of $E(N_K)f$ for General K

Repeating the foregoing procedure for several more values of K with the aid of mathematical software (e.g., Maple) gives, for $x \in [3, 4)$

$$f_3(x) = \pi_4 \Big(1 - e^{-1}x - \frac{3}{2}e^{-3}x - e^{-2}x + \frac{1}{2}e^{-2}x^2 - \frac{1}{6}e^{-3}x^3 + e^{-3}x^2e^x\Big),$$

and for $x \in [4, 5)$

$$f_4(x) = \pi_5 \Big(1 + \frac{1}{24}e^{-4}x^4 - \frac{1}{2}e^{-4}x^3 + 2e^{-4}x^2 - e^{-2}x - \frac{3}{2}e^{-3}x - \frac{8}{3}e^{-4}x - \frac{1}{6}e^{-3}x^3 + e^{-3}x^2 + \frac{1}{2}e^{-2}x^2 + e^{-1}x\Big)e^x.$$

Applying the normalizing conditions for $K = 4$ and 5 respectively then results in

$$E(N_4) = \frac{1}{\pi_4} = -\frac{1}{6}e + 2e^2 - 3e^3 + e^4,$$
$$E(N_5) = \frac{1}{\pi_5} = \frac{1}{24}e - \frac{4}{3}e^2 + \frac{9}{2}e^3 - 4e^4 + e^5.$$

The author hypothesized that $E(N_K)$ is the sum of powers of $e, e^2, \ldots, e^k$, with coefficients given in the series

$$\frac{1}{\pi_K} = E(N_K) = \sum_{j=1}^{K} \frac{(-j)^{K-i}}{(K-j)!} e^j, \; K = 1, 2, \ldots. \tag{10.46}$$

This can be verified by mathematical induction, carried out by first deriving the formulas for $f_j(x)$, $j = 0, \ldots, K$, in the same way as for those in (10.43). Then we obtain (10.46) in a similar manner as for the derivation of (10.45).

10.2.6 Asymptotic Formula for $E(N_K)$ as $K \to \infty$

We now show that in (10.46),

$$E(N_K) \approx 2K + 2/3,$$

i.e.,

$$\lim_{K\to\infty} \frac{E(N_K)}{2K + \frac{2}{3}} = 1. \tag{10.47}$$

For example, using (10.47) with the "large" number $K = 20$, we immediately have the approximation $E(N_{20}) \approx 2(20) + 2/3 = 40.6667$. The analytical value up to the same number of decimals using (10.46) is also 40.6667, whose accuracy depends on the number of digits carried, and on the computational algorithm used.

Remarkably, from the analytical values of $E(N_2)$ and $E(N_3)$ given in (10.39) and (10.45), the approximation (10.47) is very accurate for $K = 2, 3, \ldots$. Even for $K = 1$, the "asymptotic" approximation $2K + \frac{2}{3} = 2.6666$, which is within 1.90% of $e = 2.71828$.

Derivation of Asymptotic Formula (10.47)
Let γ denote the excess life at an arbitrary point $x \in \boldsymbol{S}$, as $x \to \infty$. Then $f_\gamma(y) = \frac{1}{\mu}(1 - B(y))$, $y > 0$, where $B(y)$ is the common cdf of the inter-renewal time, having mean μ (formula (10.9) above; see also Example 7.24, p. 453 in [125]. Here, the inter-renewal times are $\underset{dis}{=} U_{(0,1)}$. Thus $B(y) = y$, $0 < y < 1$, $\mu = 1/2$, and

$$E(\gamma) = \frac{1}{\mu} \int_{y=0}^{\infty} y \cdot f_\gamma(y) dy = 2 \int_{y=0}^{1} y\,(1 - y)\, dy = \frac{1}{3}. \tag{10.48}$$

Let γ_K denote the excess life at K; if K is large then $E(\gamma_K) \approx \frac{1}{3}$. If K is finite then

$$K + \gamma_K = \sum_{j=1}^{N_K} Z_j, \tag{10.49}$$

where the Z_js are i.i.d. $\underset{dis}{\equiv} U_{(0,1)}$, and N_K is a stopping time for $\{Z_j\}_{j=1,2,\ldots}$. Taking expected values in (10.49) yields

$$K + E(\gamma_K) = E(N_K) \cdot \frac{1}{2},$$

$$\text{implying} \qquad E(N_K) = 2K + 2 \cdot E(\gamma_K).$$

If K is large, $E(\gamma_K) \approx E(\gamma)$; substituting from (10.48) gives

$$E(N_K) \approx 2K + \frac{2}{3},$$

which is equivalent to formula (10.47). Also, if $\alpha > 0$ is a "large" real number then $E(N_\alpha) \approx 2\alpha + \frac{2}{3}$, where N_α is the number of renewals required to first exceed α.

10.2.7 Number of Renewals Within an Arbitrary Interval

Let $N(\alpha, \beta)$ denote the number of renewals *within* an interval $(\alpha, \beta) \subseteq (0, K)$, during a single cycle c_n, $n \in \{1, 2, \ldots\}$, of $\{X(t)\}_{t\geq 0}$. Without loss of generality, $X(0) = 0$, and we stop after N_K subintervals of $\{a_n\}_{n=1,2,\ldots}$. Then

$$N(0, K) = N_K - 1, \quad \text{and} \quad E(N(0, K)) = E(N_K) - 1.$$

Thus the values of $E(N_1)$; $E(N_2)$; $E(N_3)$ lead to the expected number of renewals within intervals (0, 1); (0, 2), (0, 3), (1, 2); (2, 3), as follows:

$$\begin{aligned}
E(N(0,1)) &= E(N_1) - 1 = e - 1 \approx 1.7183,\\
E(N(0,2)) &= E(N_2) - 1 = -e + e^2 - 1 \approx 3.6708,\\
E(N(0,3)) &= E(N_3) - 1 = \frac{1}{2}e - 2e^2 + e^3 - 1 \approx 5.6666,\\
E(N(1,2)) &= E(N(0,2)) - E(N(0,1)) = E(N_2) - E(N_1) \approx 1.9525,\\
E(N(2,3)) &= E(N(0,3)) - E(N(0,2)) = E(N_3) - E(N_2) \approx 1.9958.
\end{aligned} \tag{10.50}$$

For large K,

$$E(N(K, K+1) = E(0, K+1) - E(0, K) = E(N_{K+1}) - E(N_K) \approx 2.0.$$

In (10.50), the values of $E(N(1, 2))$, $E(N(2, 3))$ are already within 2.38% and 1.40% of the limiting value 2.0, respectively.

If $0 < \alpha < \beta < 1$, where α and β are arbitrary real numbers then $E(N_\alpha) = e^\alpha$, and $E(N_\beta) = e^\beta$, obtained similarly as in the solution for $E(N_1)$. Hence, $E(N(0, \alpha)) = e^\alpha - 1$, $E(N(0, \beta)) = e^\beta - 1$, implying the expected number of renewals within (α, β) is

$$E(N(\alpha, \beta)) = E(N_\beta) - E(N_\alpha) = e^\beta - e^\alpha, 0 < \alpha < \beta < 1. \quad (10.51)$$

In particular,

$$E\left(N\left(\frac{2}{3}, 1\right)\right) = e - e^{\frac{2}{3}} \approx 0.77055,$$
$$E\left(N\left(\frac{1}{3}, \frac{2}{3}\right)\right) = e^{\frac{2}{3}} - e^{\frac{1}{3}} \approx 0.55212,$$
$$E\left(N\left(0, \frac{1}{3}\right)\right) = e^{\frac{1}{3}} - e^0 \approx 0.39561.$$

Thus, approximately 44.84% of the renewals occur *in the top third*, 32.13% *in the middle third* and 23.02% *in the bottom third*, of interval (0, 1), indicating renewal instants tend to accumulate *in the top portion* of (0, 1). For a possible intuitive explanation of this phenomenon, fix the length of a “sliding interval” $\boldsymbol{I}_h$ using $|\boldsymbol{I}_h| = h$, $0 < h < 1$. As we slide $\boldsymbol{I}_h$ steadily from position $(0, h)$ to position $(1 - h, 1)$, the expected number of renewals in $\boldsymbol{I}_h$ increases steadily.

We can extend the analysis to determine the expected number of renewals within an arbitrary interval (α, β), $0 \leq \alpha < \beta < \infty$.

10.2.8 Discussion

We can apply the compound-cycle regenerative process model of this section, to an arbitrary renewal process $\{b_n\}_{n=1,2,\ldots}$, where b_n is a non-lattice positive r.v. The analysis can also be extended to models where $-\infty < b_n < \infty$, so that $\{b_n\}_{n=1,2,\ldots}$ is not a renewal process, but the cycles $\{c_n\}_{n=1,2,\ldots}$ and subintervals $\{a_n\}_{n=1,2,\ldots}$ are inter-renewal times of renewal processes.

Possible applications are to problems where it is required to determine the expected number of events until a stopping criterion is satisfied. Examples are the number of: customers served in a busy period of a queue; demands in an ordering cycle of an inventory system; inputs until overflow of a dam;

shocks until failure of a machine part; claims until ruin in an actuarial model; doses of a prescription drug until an overdose occurs; advertisements until a favorable consumer response occurs for a product.

10.3 The Time-t PDFs of a Renewal Process

We now apply LC to obtain the pdfs of the *excess life*, *age* and *total life at an arbitrary finite time* $t > 0$, based on concepts in [34]. The time-t probability distributions have been analyzed classically in [66, 99, 123, 135], and in many other studies.

Consider an ***ordinary renewal process*** $\{Z_n\}_{n=1,2,...}$ with continuous inter-arrival times $Z_n \underset{dis}{=} Z$ having cdf $B(\cdot)$, pdf $b(\cdot)$, ccdf (the complementary cdf) $\bar{B}(\cdot) = 1 - B(\cdot)$, and support $(0, \boldsymbol{U})$, $\boldsymbol{U} > 0$. We transform the one-dimensional process $\{Z_n\}_{n=1,2,...}$ into a regenerative process $\{X(s)\}_{s\geq 0}$ having state space $[0, \infty)$ with upward jumps $\underset{dis}{=} Z$ occurring at an arbitrary Poisson rate, and with regenerative epochs at instants of level-t exceedance (see the first subsection of Sect. 10.2; Sect. 10.2.1; Figs. 10.3 and 10.4). The limiting pdf of $\{X(s)\}_{s\geq 0}$ as $s \to \infty$, exists (see, e.g., [132]), and is concentrated on the state-space interval $[0, t)$. Knowledge of *this limiting pdf* leads directly to the finite time-t distributions of $\{Z_n\}_{n=1,2,...}$.

10.3.1 Structure of Regenerative Process $\{X(s)\}_{s\geq 0}$

The process $\{X(s)\}_{s\geq 0}$ is built up from i.i.d. replicas of $\{Z_n\}_{n=1,2,...}$ and an *independent* Poisson process of arbitrary rate λ ($\lambda := 1$ for simplicity). Let $X(0) = 0$. Sample paths of $\{X(s)\}_{s\geq 0}$ make upward jumps at Poisson rate 1, with inter-jump times $a_i \underset{dis}{=} a$ and $E(a) \equiv 1$. The jumps originate in state-space subset $[0, t)$ and are in one-to-one correspondence with the horizontal intervals $\{a_i\}_{i=1,2,...}$(see Fig. 10.7). When a jump upcrosses level t the SP immediately jumps *downward to level* 0 (double jumps—see Example 2.2 in Sect. 2.3 and subsequent material therein). At that instant $\{X(s)\}_{s\geq 0}$ restarts (Fig. 10.7). (Fig. 10.8 depicts the regenerative process when $Z_n \underset{dis}{\equiv} U_{(0,1)}$, and $t \in (1, 2)$.)

Let $\left\{\pi^{(t)}, f^{(t)}(x)\right\}_{0<x<t}$ be the limiting mixed pdf of $\{X(s)\}_{s\geq 0}$ as $s \to \infty$, where $\pi^{(t)} = \lim_{s\to\infty} P(X(s) = 0)$. Then $\left\{\pi^{(t)}, f^{(t)}(x)\right\}_{0<x<t}$ is a time-average pdf since upward jumps occur at a Poisson rate, implying the arrival-

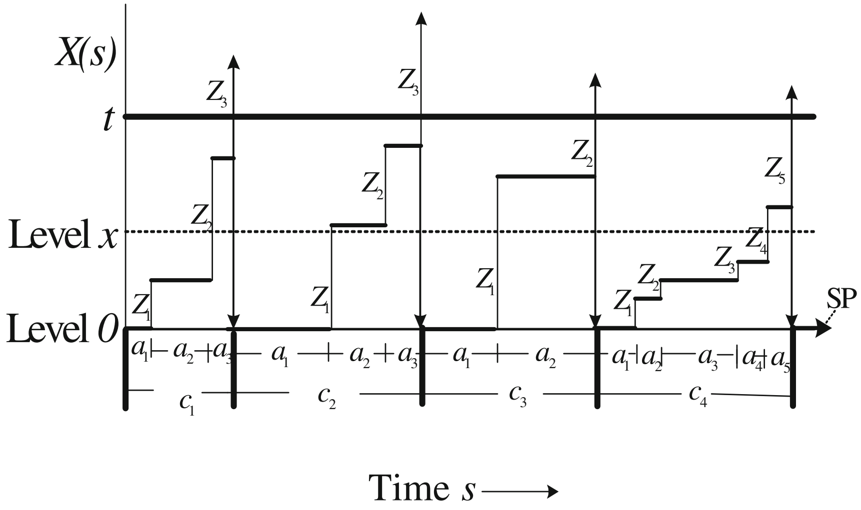


Fig. 10.7 Sample path of regenerative process $\{X(s)\}_{s\geq 0}$. Indicates embedded renewal process $\{Z_n\}_{n=1,2,,...}$, the fixed time t, cycles $c_1, c_2, \ldots$, interarrival times between upward jumps $a_i \underset{dis}{=} \text{Exp}_1$, a fixed state-space level x, SP motion

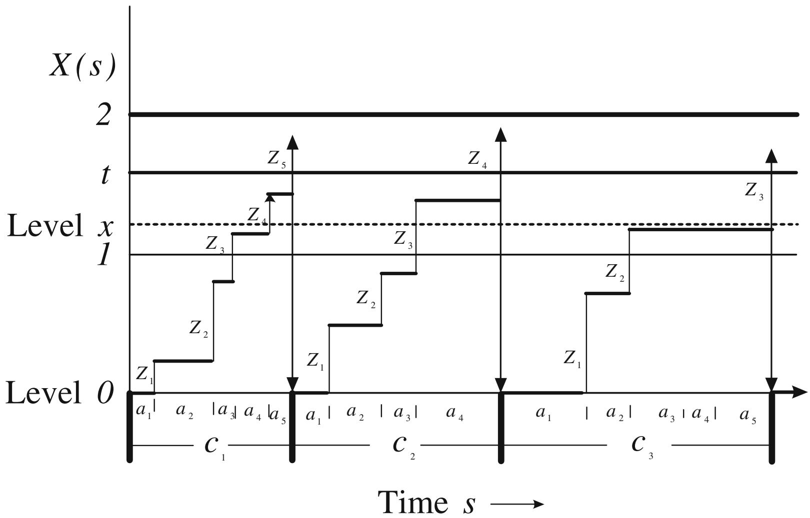


Fig. 10.8 Sample path of $\{X(s)\}_{s\geq 0}$, where $Z_n \underset{dis}{\equiv} U_{(0,1)}$ and $t \in (1,2)$. In each cycle having at least j upward jumps $Z_j \underset{dis}{=} U_{(0,1)}$, but the Z_j's in different cycles have different sizes. (See Fig. 10.4.)

point pdf is the same as the time-average pdf (e.g., [145]). A rate-balance equation for $\left\{\pi^{(t)}, f^{(t)}(x)\right\}_{0<x<t}$ is

$$\lim_{s\to\infty}\frac{\mathcal{U}_s^j(x)}{s} = \lim_{s\to\infty}\frac{\mathcal{D}_s^j(x)}{s}, x \in [0,t). \tag{10.52}$$

Substituting the formulas for the rates in (10.52) gives

$$\pi^{(t)}\bar{B}(x) + \int_0^x \overline{B}(x-y)f^{(t)}(y)dy = \pi^{(t)}, x \in [0,t), \tag{10.53}$$

which is to be solved with the normalizing condition

$$\pi^{(t)} + \int_0^t f^{(t)}(y)dy = 1. \tag{10.54}$$

10.3.2 Solution of Equation for $\left\{\pi^{(t)}, f^{(t)}(x)\right\}_{0<x<t}$

The study [34] shows that the solution of (10.53) and (10.54) is

$$f^{(t)}(x) = \frac{M'(x)}{M(t)+1}, 0 < x < t, \tag{10.55}$$

$$\pi^{(t)} = \frac{1}{M(t)+1}, \tag{10.56}$$

where

$$M(x) = \sum_{n=1}^{\infty} B^{*n}(x); \qquad M'(x) = \sum_{n=1}^{\infty} b^{*n}(x), x \in (0,t], \tag{10.57}$$

and $B^{*n}(x)$, $b^{*n}(x)$ are the n-fold convolutions of $B(\cdot)$ and $b(\cdot)$, respectively. $M(x)$ is the *renewal function* for $\{Z_n\}_{n=1,2,\ldots}$; $M(x) = E($*number of renewals up to time* $x)$—see pp. 167–169 in [99]. (Step 1 in Sect. 10.3.8 has a more detailed derivation of $\left\{\pi^{(t)}, f^{(t)}(x)\right\}_{0<x<t}$ in a particular *modified* renewal process.) From (10.56)

$$\frac{1}{\pi^{(t)}} = M(t) + 1 = E(number\ of\ renewals\ required\ to\ exceed\ x). \tag{10.58}$$

Formula (10.58) connects $M(t)$ to $1/\pi_1$ in formula (10.31) when $K = 1$ in Sect. 10.2.2 (and also to $1/\pi_K$ when $K = 2, 3, \ldots$, in Sects. 10.2.3–10.2.5).

10.3.3 Time-t Probability Distributions of $\{Z_n\}_{n=1,2,...}$

We look separately at the two cases: $t \leq \boldsymbol{U}$, and $t > \boldsymbol{U}$, where $(0, \boldsymbol{U})$ is the support of the inter-renewal pdf $b(\cdot)$. (See Sect. 4, pp. 191–195, in [34]).

10.3.4 PDF of Excess Life γ_t

Case $t \leq \boldsymbol{U}$
If $t \leq \boldsymbol{U}$ then $\boldsymbol{U}$ is finite or infinite. Integral equation (10.59) below for the ccdf $\overline{F}_{\gamma_t}(x)$ equates two different upcrossing rates of level $t + x$ (see Sect. 3.16.5 for similar reasoning in an M/G/1 queue with bounded virtual wait). For $0 < x < \boldsymbol{U}$,

$$\pi^{(t)}\overline{F}_{\gamma_t}(x) = \pi^{(t)}\overline{B}(t+x) + \int_{y=(t+x-U,0)^+}^{t} \overline{B}(t+x-y)f^{(t)}(y)dy, \tag{10.59}$$

$$\overline{F}_{\gamma_t}(x) = \overline{B}(t+x) + \int_{y=(t+x-U,0)^+}^{t} \overline{B}(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy. \tag{10.60}$$

where $(t + x - \boldsymbol{U}, 0)^+ := \max(t + x - \boldsymbol{U}, 0)$.

Taking d/dx in (10.60) gives, since $\overline{B}(t+x-(t+x-U)) = \overline{B}(U) = 0$,

$$f_{\gamma_t}(x) = b(t+x) + \int_{y=(t+x-U,0)^+}^{t} b(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy, 0 < x < U. \tag{10.61}$$

From (10.61) we obtain

$$f_{\gamma_t}(x) = \begin{cases} b(t+x) + \int_{y=0}^{t} b(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy, 0 < x < \boldsymbol{U} - t, \\ \int_{y=t+x-U}^{t} b(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy, \boldsymbol{U} - t < x < \boldsymbol{U}. \end{cases} \tag{10.62}$$

If $\boldsymbol{U} = \infty$, only the first formula in (10.62) applies; if $\boldsymbol{U} < \infty$, both formulas in (10.62) apply.

Remark 10.4 If $t < \boldsymbol{U} < \infty$ then $f_{\gamma_t}(x)$ has a jump discontinuity at $x = \boldsymbol{U} - t$ of magnitude $f_{\gamma_t}\left((\boldsymbol{U} - t)^+\right) - f_{\gamma_t}\left((\boldsymbol{U} - t)^-\right) = -b(\boldsymbol{U}^-)$. This follows by letting $x \downarrow (\boldsymbol{U} - t)$ in the second formula of (10.62), and $x \uparrow (\boldsymbol{U} - t)$ in the first formula of (10.62), and subtracting.

Case $t > \boldsymbol{U}$
If $t > \boldsymbol{U}$, then $\boldsymbol{U}$ must be finite; assume $t \in [NU, (N+1)U)$ for some integer $N \geq 1$. Upward jumps starting below $t + x - \boldsymbol{U}$, cannot upcross level $t + x$. Thus, an equation analogous to (10.60) is

$$\overline{F}_{\gamma_t}(x) = \int_{y=t+x-\boldsymbol{U}}^{t} \overline{B}(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy, \ 0 < x < U. \tag{10.63}$$

Taking d/dx in (10.63) and noting that $\overline{B}(U) = 0$, gives

$$f_{\gamma_t}(x) = \int_{y=t+x-\boldsymbol{U}}^{t} b(t+x-y)\frac{f^{(t)}(y)}{\pi^{(t)}}dy, \ 0 < x < U. \tag{10.64}$$

From (10.64), with

$$f_n^{(t)}(y) := f^{(t)}(y) \cdot \boldsymbol{I}_{(n\boldsymbol{U},(n+1))\boldsymbol{U}}(y), \, y > 0, n = 0, \ldots, N,$$

we get

$$f_{\gamma_t}(x) = \begin{cases} \int_{y=t+x-\boldsymbol{U}}^{NU} b(t+x-y)\frac{f_{N-1}^{(t)}(y)}{\pi^{(t)}}dy \\ \qquad + \int_{y=N\boldsymbol{U}}^{t} b(t+x-y)\frac{f_N^{(t)}(y)}{\pi^{(t)}}dy, 0 < x < (N+1)U - t, \\ \int_{y=t+x-\boldsymbol{U}}^{t} b(t+x-y)\frac{f_N^{(t)}(y)}{\pi^{(t)}}dy, (N+1)U - t < x < U. \end{cases}$$

10.3.5 PDF of $\{X(s)\}_{s\geq 0}$ Just Before a Jump Over t

Let $X_{JB}^{(t)}$:= ordinate of $\{X(s)\}_{s\geq 0}$ just before the jump that first exceeds level t. Denote its mixed pdf by $\left\{\pi_{JB}^{(t)}, \ f_{JB}^{(t)}(x)\right\}_{0<x<t}$ where $\pi_{JB}^{(t)} = P(X_{JB}^{(t)} = 0)$. ($X_{JB}^{(t)}$ is an important random variable in various stochastic models, such as actuarial ruin models – see, e.g., [79] and [54].) We now state the results for $\left\{\pi_{JB}^{(t)}, \ f_{JB}^{(t)}(x)\right\}_{0<x<t}$. (For detailed derivations see [34].)

Case $t \leq \boldsymbol{U}$
Considering a sample path of $\{X(s)\}_{s\geq 0}$ (Fig. 10.7), and applying Bayes' rule, leads to

$$f_{JB}^{(t)}(x) = \overline{B}(t-x)\frac{f^{(t)}(x)}{\pi^{(t)}}, 0 < x < t\ . \tag{10.65}$$

$$\pi_{JB}^{(t)} = \overline{B}(t). \tag{10.66}$$

Case $t > U$

$$f_{JB}^{(t)}(x) = \begin{cases} \overline{B}(t-x)\dfrac{f_{N-1}^{(t)}(x)}{\pi^{(t)}}, t-U \leq x < NU, \\ \overline{B}(t-x)\dfrac{f_{N}^{(t)}(x)}{\pi^{(t)}}, NU \leq x < t. \end{cases} \tag{10.67}$$

Random variable $X_{JB}^{(t)}$ is related to the age δ_t.

10.3.6 PDF of Age δ_t

We get the mixed pdf $\left\{\pi_{\delta_t}, f_{\delta_t}(x)\right\}_{0<x<t}$ using (10.65)–(10.67). Since $\delta_t = t - X_{JB}^{(t)}$, $\pi_{\delta_t} = P(\delta_t = t) = P(X_{JB}^{(t)} = 0) = \pi_{JB}^{(t)}$, and $f_{\delta_t}(x) = f_{JB}^{(t)}(t-x)$, $0 < x < t$.

Case $t \leq U$
Using (10.65) and (10.66) yields

$$\pi_{\delta_t} = \overline{B}(t), \qquad f_{\delta_t}(x) = \overline{B}(x)\frac{f^{(t)}(t-x)}{\pi^{(t)}}, 0 < x < t. \tag{10.68}$$

Case $t > U$
Probability $\pi_{\delta_t} = 0$ since $t > U$. Using (10.67) and applying $f_{\delta_t}(x) = f_{JB}^{(t)}(t-x)$, yields

$$f_{\delta_t}(x) = \begin{cases} \dfrac{\overline{B}(x) f_{N}^{(t)}(t-x)}{\pi^{(t)}}, 0 < x < t - NU, \\ \dfrac{\overline{B}(x) f_{N-1}^{(t)}(t-x)}{\pi^{(t)}}, t - NU < x < U. \end{cases} \tag{10.69}$$

10.3.7 PDF of Total life β_t

The total life is $\beta_t := \gamma_t + \delta_t$. Hence $P(\beta_t = x)dx = P(\gamma_t = x - \delta_t)dx$, $x > \delta_t$.

Case $t \le \boldsymbol{U}$

$$f_{\beta_t}(x|\delta_t = y)dx = P(Z = x|Z > y)dx = \frac{P(Z = x)dx}{P(Z > y)} = \frac{b(x)dx}{\overline{B}(y)}.$$

Unconditioning $f_{\beta_t}(x|\delta_t = y)$ with respect to $f_{\delta_t}(y)$, and substituting for $f_{\delta_t}(y)$ from (10.68), gives

$$\begin{aligned} f_{\beta_t}(x) &= \int_{y=0}^{x} \frac{b(x)}{\overline{B}(y)} f_{\delta_t}(y)dy = \int_{y=0}^{x} \frac{b(x)}{\overline{B}(y)} \overline{B}(y) \frac{f^{(t)}(t-y)}{\pi^{(t)}} dy \\ &= b(x) \int_{y=0}^{x} \frac{f^{(t)}(t-y)}{\pi^{(t)}} dy, \; 0 < x < t. \end{aligned} \tag{10.70}$$

Similar reasoning yields

$$\begin{aligned} f_{\beta_t}(x) &= \frac{b(x)}{\overline{B}(t)} \pi_{\delta_t} + b(x) \int_{y=0}^{t} \frac{f^{(t)}(t-y)}{\pi^{(t)}} dy \\ &= b(x) \left(1 + \int_{y=0}^{t} \frac{f^{(t)}(t-y)}{\pi^{(t)}} dy \right), \; t < x < \boldsymbol{U}. \end{aligned} \tag{10.71}$$

Formulas (10.70) and (10.71) imply $f_{\beta_t}(x)$ has a jump discontinuity at $x = t$ of magnitude

$$f_{\beta_t}(t^+) - f_{\beta_t}(t^-) = b(t). \tag{10.72}$$

If $\boldsymbol{U} = \infty$ then $\lim_{t\to\infty} b(t) = 0$ in (10.72).

Check on $\lim_{t\to\infty} f_{\beta_t}(x)$
Formula (10.70) gives the known result $\lim_{t\to\infty} f_{\beta_t}(x) = xb(x)/E(Z), x > 0$, since (10.55) and (10.56) imply $f^{(t)}(t-y)/\pi^{(t)} = M'(t-y)$, and (10.70) gives

$$f_{\beta_t}(x) = b(x) \int_{y=0}^{x} M'(t-y)dy = b(x)\left(M(t) - M(t-x)\right), 0 < x < t.$$

Applying Blackwell's theorem (p. 191 in [99]) implies

$$\lim_{t\to\infty} (M(t) - M(t-x)) = \frac{x}{E(Z)}, x > 0 \implies \lim_{t\to\infty} f_{\beta_t}(x) = \frac{xb(x)}{E(Z)}, x > 0.$$

Case $t > \boldsymbol{U}$
Expressing $f_{\beta_t}(x)$ in terms of $f_{\beta_t}(x|\delta_t = y)$ and substituting for $f_{\delta_t}(y)$ from (10.69) yields

$$f_{\beta_t}(x) = b(x) \int_{y=0}^{x} \frac{f_N^{(t)}(t-y)}{\pi^{(t)}} dy, 0 < x < t - N\boldsymbol{U}. \tag{10.73}$$

Reasoning as for (10.73) and using also (10.69), yields

$$f_{\beta_t}(x) = b(x) \left(\int_{y=0}^{t-N\boldsymbol{U}} \frac{f_N^{(t)}(t-y)}{\pi^{(t)}} dy + \int_{y=t-N\boldsymbol{U}}^{x} \frac{f_{N-1}^{(t)}(t-y)}{\pi^{(t)}} dy \right), \quad t - N\boldsymbol{U} < x < \boldsymbol{U}. \tag{10.74}$$

10.3.8 Example—A Modified Renewal Process

Consider a ***modified renewal process*** $\{Z_n\}_{n=1,2,...}$ where $Z_1 \underset{dis}{=} U_{(0,1)}$ and $Z_n \underset{dis}{=} \text{Exp}_\mu, n = 2, 3, \ldots$ This is also called a ***delayed renewal process*** (see pp. 197–199 in [99]; pp. 27–29 in [66]). Thus $\overline{B}_0(x) := P(Z_1 > x) = 1 - x, x \in (0, 1)$, and $\overline{B}_1(x) := P(Z_n > x) = e^{-\mu x}, x \in (0, \infty)$. We now derive $f_{\gamma_t}(\cdot)$, $f_{\delta_t}(\cdot)$ and $f_{\beta_t}(\cdot)$, for the case $t \in (0, 1)$. The support of Z_1 is $(0, \boldsymbol{U}_0)$, where $\boldsymbol{U}_0 = 1$, and that of $Z_n, n = 2, 3, \ldots$, is $(0, \boldsymbol{U}_1)$, where $\boldsymbol{U}_1 = \infty$. Therefore, this example deals with the case $0 < t\boldsymbol{U}_0 < \boldsymbol{U}_1$, and differs from the three examples in Sect. 5, pp. 195–200 in [34].

Step 1. Derive pdf $\left\{\pi_0^{(t)}, f^{(t)}(x)\right\}_{x\in(0,t)}$ for process $\{X(s)\}_{s\geq 0}$.
Since $t \in (0, 1)$, equating up- and downcrossing rates of level $x \in (0, t)$ gives

$$\pi_0^{(t)} (1-x) + \int_0^x e^{-\mu(x-y)} f^{(t)}(y) dy = \pi_0^{(t)}, x \in (0, t) \, .$$

Taking d/dx results in

$$-\pi_0^{(t)} - \mu \left(\pi_0^{(t)} - \pi_0^{(t)} (1-x)\right) + f^{(t)}(x) = 0,$$
$$f^{(t)}(x) = \pi_0^{(t)} (\mu x + 1), x \in (0, t) \, . \tag{10.75}$$

The law of total probability (normalizing condition) (10.54) gives

$$\pi_0^{(t)} = \frac{2}{\mu t^2 + 2t + 2}. \tag{10.76}$$

Step 2. Derive $f_{\gamma_t}(z)$, $z > 0$, **using** (10.75) **and** (10.76)
Equating *two different expressions* for the upcrossing rate of level $t + z$ gives

$$\begin{aligned}\pi_0^{(t)}\overline{F}_{\gamma_t}(z) &= \pi_0^{(t)}\overline{B}_0(t+z) + \textstyle\int_0^t \overline{B}_1(t+z-y)\pi_0^{(t)}(\mu y+1)\,dy \\ &= \pi_0^{(t)}(1-t-z) + \textstyle\int_0^t e^{-\mu(t+z-y)}\pi_0^{(t)}(\mu y+1)\,dy,\; z \in (0, 1-t)\,,\end{aligned}$$

and

$$\begin{aligned}\pi_0^{(t)}\overline{F}_{\gamma_t}(z) &= \textstyle\int_0^t \overline{B}_1(t+z-y)\pi_0^{(t)}(\mu y+1)\,dy \\ &= \textstyle\int_0^t e^{-\mu(t+z-y)}\pi_0^{(t)}(\mu y+1)\,dy,\; z \in (1-t, \infty)\,.\end{aligned}$$

Thus

$$\overline{F}_{\gamma_t}(z) = 1 - t - z + \int_0^t e^{-\mu(t+z-y)}(\mu y+1)\,dy,\; z \in (0, 1-t)\,,$$
$$\overline{F}_{\gamma_t}(z) = \int_0^t e^{-\mu(t+z-y)}(\mu y+1)\,dy,\; z \in (1-t, \infty)\,;$$

taking d/dz in both equations gives

$$f_{\gamma_t}(z) = 1 + \mu\int_0^t e^{-\mu(t+z-y)}(\mu y+1)\,dy,\; z \in (0, 1-t)\,,$$
$$f_{\gamma_t}(z) = \mu\int_0^t e^{-\mu(t+z-y)}(\mu y+1)\,dy,\; z \in (1-t, \infty)\,,$$

implying, respectively,

$$f_{\gamma_t}(z) = 1 + \mu t e^{-\mu z},\; z \in (0, 1-t)\,, \tag{10.77}$$
$$f_{\gamma_t}(z) = \mu t e^{-\mu z},\; z \in (1-t, \infty)\,, \tag{10.78}$$

which satisfies the normalizing condition

$$\int_0^{t-1} f_{\gamma_t}(z)dz + \int_{t-1}^{\infty} f_{\gamma_t}(z)dz = 1.$$

The pdf $f_{\gamma_t}(z)$ has a discontinuity at $z = 1 - t$ of size

$$f_{\gamma_t}((1-t)^+) - f_{\gamma_t}((1-t)^-) = -1 = -b_0\left(\boldsymbol{U}_0^-\right).$$

(See Remark 10.4 in Sect. 10.3.4 for a similar discontinuity in the ordinary renewal process.)

Step 3. Derive $\left\{\pi_{\delta_t}, f_{\delta_t}(x)\right\}_{x\in(0,t)}$, $t \in (0,1)$ **using** (10.75) **and** (10.76)
Substituting $\left\{\pi_0^{(t)}, f^{(t)}(x)\right\}_{x\in(0,t)}$ from Step 1 above into (10.68) gives

$$\pi_{\delta_t} = \overline{B}_0(t) = 1 - t, \tag{10.79}$$

$$f_{\delta_t}(x) = \overline{B}_1(x)\frac{f^{(t)}(t-x)}{\pi^{(t)}} = e^{-\mu x}\left(\mu\left(t-x\right)+1\right), x \in (0,t). \tag{10.80}$$

The normalizing condition $\pi_{\delta_t} + \int_0^t f_{\delta_t}(x)dx = 1$ is readily checked.

Step 4. Derive $f_{\beta_t}(x)$, $x > 0$, $t \in (0,1)$ **using** (10.75) **and** (10.76)
Using similar reasoning as for (10.70) in Sect. 10.3.7, we obtain

$$\begin{aligned} f_{\beta_t}(x) &= b_1(x)\int_{y=0}^{x}\frac{1}{\overline{B}(y)}\overline{B}(y)\frac{f^{(t)}t-y)}{\pi^{(t)}}dy \\ &= \mu e^{-\mu x}\int_{y=0}^{x}\left(\mu\left(t-y\right)+1\right)dy \\ &= \mu e^{-\mu x}\left[\left(\mu t+1\right)x - \mu\frac{x^2}{2}\right], x \in (0,t). \end{aligned} \tag{10.81}$$

Since $Z_1 \underset{dis}{=} U_{(0,1)}$ has support in $(0,1)$ (and substituting $\overline{B}_0(t) = \pi_{\delta_t}$ from (10.79)), we get

$$\begin{aligned} f_{\beta_t}(x) &= \frac{b_0(x)}{\overline{B}_0(t)}\pi_{\delta_t} + b_1(x)\int_{y=0}^{t}\frac{1}{\overline{B}(y)}\overline{B}(y)\frac{f^{(t)}t-y)}{\pi^{(t)}}dy \\ &= b_0(x) + b_1(x)\int_{y=0}^{t}\frac{1}{\overline{B}(y)}\overline{B}(y)\frac{f^{(t)}t-y)}{\pi^{(t)}}dy, x \in (t,1), \end{aligned}$$

which differs from formula (10.71) for the ordinary renewal process. Thus

$$\begin{aligned} f_{\beta_t}(x) &= b_0(x) + e^{-\mu x}\left[\left(\mu t+1\right)t - \mu\frac{t^2}{2}\right] \\ &= 1 + e^{-\mu x}\left[\left(\mu t+1\right)t - \mu\frac{t^2}{2}\right], x \in (t,1). \end{aligned} \tag{10.82}$$

Subtracting (10.81) at t^- from (10.82) at t^+ gives $f_{\beta_t}(t^+) - f_{\beta_t}(t^+) = 1 = b_0(\boldsymbol{U}_0^-)$. The magnitude of this jump discontinuity is similar to that in formula (10.72) for the ordinary renewal process, but in this particular delayed renewal process, only the support of Z_1 contributes to the size of the discontinuity.

Since jumps starting from level 0 cannot upcross level $x = 1$, we obtain

$$
\begin{aligned}
f_{\beta_t}(x) &= b_1(x)\int_{y=0}^{t}\frac{1}{\overline{\overline{B}}(y)}\overline{B}(y)\frac{f^{(t)}t-y)}{\pi^{(t)}}dy \\
&= e^{-\mu x}\left[(\mu t+1)\,t-\mu\frac{t^2}{2}\right] \\
&= e^{-\mu x}\left[(\mu t+1)\,t-\mu\frac{t^2}{2}\right], x\in(1,\infty).
\end{aligned} \tag{10.83}
$$

The formula for $f_{\beta_t}(x)$, $x > 0$, satisfies the normalizing condition, since

$$
\int_0^t f_{\beta_t}(x)dx = t - \frac{1}{2}e^{-\mu t}\mu t^2 - e^{-\mu t}t,
$$

$$
\int_t^1 f_{\beta_t}(x)dx + \int_1^{\infty} f_{\beta_t}(x)dx = 1 - t + \frac{1}{2}e^{-\mu t}\mu t^2 + e^{-\mu t}t,
$$

implying $\int_0^{\infty} f_{\beta_t}(x)dx = 1$.

Chapter 11
ADDITIONAL APPLICATIONS of LC

This chapter applies SPLC to a variety of stochastic models in order to illustrate the scope, applicability and flexibility of the methodology, and to motivate additional new applications. Section 11.1 analyzes a variant of the classical Cramér-Lundberg (C-L) risk process, based on Model 1, pp. 289–302 in [54]. Section 11.2 gives a general technique for transient distributions in a stochastic process. Section 11.10 discusses the application of LC to simple harmonic motion. The intervening sections analyze other potentially motivational models.

11.1 Risk Model: Barrier and Reinvestment

Let $\{X(t)\}_{t\geq 0}$ denote the risk reserve (also called the *surplus*) at time $t \geq 0$ in a standard C-L (Cramér-Lundberg) risk model with initial reserve $x_0 > 0$ (see, e.g., pp. 22–28 in [71]). Insurance claims occur in a Poisson process with rate λ. Claim sizes are i.i.d. r.v.s $\{Z_i\}_{i=1,2,\ldots}$ with common cdf $B(x)$, $x \geq 0$. Then

$$X(t) = x_0 + c\,t - \sum_{i=1}^{\mathbf{N}(t)} Z_i,\ t \geq 0, \tag{11.1}$$

where $c > 0$ is the premium rate, and $\{\mathbf{N}(t)\}_{t\geq 0}$ is a Poisson process having rate λ. Denote the time until ruin by

$$\tau = \inf\{t > 0 \mid X(t) \text{ downcrosses level } 0\}. \tag{11.2}$$

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5_11

11.1.1 Variant of the Cramér-Lundberg Model

Here, we consider a variant of the standard C-L model (characterized by formula (11.1)), with the following properties. There is a constant barrier at level $M > x_0$, and a *reinvestment strategy*. Whenever $\{X(t)\}_{t\geq 0}$ reaches the barrier M, a portion of net gain $M - x_0$ is transferred out immediately and moved into alternative investments (e.g., a conservative investment portfolio). Specifically, if $X(t^-) = M$ then $X(t) = a$, where $a \in (x_0, M)$. Every $t_M \in \{t | X(t^-) = M\}$ is a regenerative point of $\{X(t)\}_{t\geq 0}$. At such t_Ms the capital reserve jumps downward by the reinvestment amount $M - a$ to level a (see Fig. 11.1). Let k be the number of instants t_M occurring in $(0, t)$. The following formulas characterize this variant.

$$\begin{aligned} &X(t) = x_0 + c\,t - \textstyle\sum_{i=1}^{\mathbf{N}(t)} Z_i - k(M - a),\ X(t) < M, t \in (0, \tau) \\ &X(t) = a \text{ if } X(t^-) = M, t \in (0, \tau)\,, \end{aligned} \tag{11.3}$$

where τ is the ruin time of $\{X(t)\}_{t\geq 0}$.

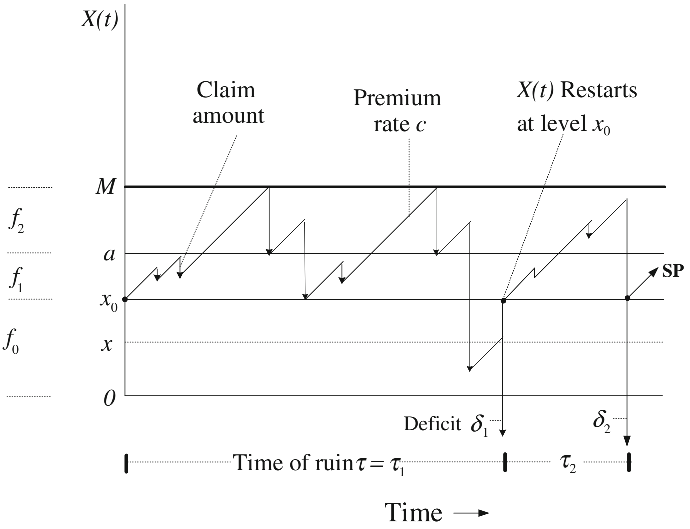


Fig. 11.1 Sample path of regenerative risk process $\{X(t)\}_{t\geq 0}$; two "ruin" cycles; barrier at M; initial reserve x_0; reinvestment indicator level a; deficits at ruin δ_1, δ_2

11.1.2 Extending $\{X(t)\}$ *from* $[0, \tau)$ *to* $[0, \infty)$

We create a regenerative process whose cycles are i.i.d. replicas of the process $\{X(t)\}_{t\in[0,\tau)}$, as follows. At the ruin time τ, we restore the capital reserve to the initial value x_0, and *restart the process* $\{X(t)\}_{t\in[0,\tau)}$. We repeat this restoration at each successive "ruin" instant. This procedure forms a regenerative process, denoted by $\{X(t)\}_{t\geq 0}$. Figure 11.1 shows the first two cycles of $\{X(t)\}_{t\geq 0}$. The regenerative process has the advantage of containing an infinite number of i.i.d. ruin cycles, and enables us to analyze the ruin model using its properties. The time points $\tau_1, \tau_1 + \tau_2, \tau_1 + \tau_2 + \tau_3$, ..., are regenerative points at which the SP makes a double jump—one downward below level 0 causing a deficit ending a ruin cycle, and one upward to level x_0, starting the next cycle. Since $\{X(t)\}_{t\geq 0}$ is a regenerative process, its limiting pdf exists as $t \to \infty$, which we denote by $f(x)$, $x \in (0, M)$.

Expected Time of Ruin $E(\tau)$

Consider the renewal process $\{\tau_n\}_{n=1,2,...}$, where each $\tau_n \underset{dis}{=} \tau$. Let $N_\tau(t)$:= number of renewals (i.e., "ruins") in $(0, t)$. Let $\mathcal{D}_t(x)$:= number of SP downcrossings of level x in $(0, t)$ (Fig. 11.1). From the elementary renewal theorem and LC theory.

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \lambda \int_x^M \overline{B}(y-x)f(y)dy, 0 \leq x \leq M,$$
$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(0))}{t} = \lim_{t\to\infty} \frac{N_\tau(t)}{t} \underset{a.s.}{=} \lim_{t\to\infty} \frac{E(N_\tau(t))}{t} = \frac{1}{E(\tau)}, \quad (11.4)$$

where $\overline{B}(\cdot) = 1 - B(\cdot)$. In particular

$$\frac{1}{E(\tau)} = \lambda \int_0^M \overline{B}(y)f(y)dy,$$
$$E(\tau) = \frac{1}{\lambda \int_0^M \overline{B}(y)f(y)dy}. \quad (11.5)$$

Let $\mathcal{U}_t(x)$:= number of SP upcrossings of level x in $(0, t)$. All upcrossings of any level $x \in (0, M)$ occur when the sample path is continuous, except at the "ruin" instants, when upcrossings of level x are jumps over level $x \in (0, x_0)$. From LC theory and (11.5),

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = cf(x), x_0 \le x < M,$$

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = cf(x) + \lambda \int_0^M \overline{B}(y) dF(y)$$

$$= cf(x) + \frac{1}{E(\tau)}, 0 < x < x_0.$$

11.1.3 CDF and PDF of the Deficit at Ruin

A random variable of interest in risk theory is the deficit at ruin, denoted by $\boldsymbol{\delta}$, having cdf denoted by $F_{\boldsymbol{\delta}}(y)$, and pdf $f_{\boldsymbol{\delta}}(y)$, $y > 0$ (Fig. 11.1). R.v. $\boldsymbol{\delta}$ is equal to the excess below level 0 of a downward jump due to a claim.

Theorem 11.1 The cdf and pdf of the deficit at ruin $\boldsymbol{\delta}$ are given by

$$F_{\boldsymbol{\delta}}(x) = 1 - \lambda E(\tau) \int_0^M \overline{B}(y + x) f(y) dy, x > 0, \tag{11.6}$$

$$f_{\boldsymbol{\delta}}(x) = \lambda E(\tau) \int_{[0,M]} b(y + x) dF(y), x > 0, \tag{11.7}$$

where $E(\tau)$ is given by (11.5) and $b(y) = dB(y)/dy$, $y > 0$, is the pdf of the claim size.

Proof Fix level $x < 0$. By (11.5), $1/E(\tau)$ is the downcrossing *rate* of level 0. The term $1 - F_{\boldsymbol{\delta}}(|x|)$ is the probability that a downward jump that crosses level 0, also crosses level x, where $|\cdot|$ stands for absolute value. Thus $(1/E(\tau))\,(1 - F_{\boldsymbol{\delta}}(|x|))$ is the downcrossing rate of level x . Since downcrossings of level 0 are caused by claims only, we have the following equation for $F_{\boldsymbol{\delta}}(|x|)$

$$\frac{1}{E(\tau)}(1 - F_{\boldsymbol{\delta}}(|x|)) = \lambda \int_0^M \overline{B}(y - x) f(y) dy, x < 0, \tag{11.8}$$

which is equivalent to (11.6) (since $x < 0$). Taking d/dx on both sides of (11.6) gives (11.7). □

11.1.4 Analysis of the Risk Model

Let

$$f(x) := f_0(x)\boldsymbol{I}_{[0,x_0)}(x) + f_1(x)\boldsymbol{I}_{[x_0,a)}(x) + f_2(x)\boldsymbol{I}_{[a,M)}(x),$$

where $\boldsymbol{I}_{\boldsymbol{A}}(x) = 1$ if $x \in \boldsymbol{A}$ and $\boldsymbol{I}_{\boldsymbol{A}}(x) = 0$ if $x \notin \boldsymbol{A}$.

Define $f_2(M) := f_2(M^-)$ for convenience; this does not affect any probability measures. By LC and rate balance across level x (see explanations after (11.13) below), we obtain

$$cf_2(x) = cf_2(M^-) + \lambda \int_x^M \overline{B}(y-x)f_2(y)\,dy,\ a \le x < M, \tag{11.9}$$

$$cf_1(x) = \lambda \int_x^a \overline{B}(y-x)f_1(y)\,dy + \lambda \int_a^M \overline{B}(y-x)f_2(y)\,dy,\ x_0 \le x < a, \tag{11.10}$$

$$\begin{aligned} cf_0(x) + \frac{1}{E(\tau)} = \lambda \int_x^{x_0} \overline{B}(y-x)f_0(y)\,dy + \lambda \int_{x_0}^a \overline{B}(y-x)f_1(y)\,dy \\ +\lambda \int_a^M \overline{B}(y-x)f_2(y)\,dy,\ 0 < x < x_0, \end{aligned} \tag{11.11}$$

where

$$\frac{1}{E(\tau)} = \lambda \int_0^{x_0} \overline{B}(y)f_0(y)\,dy + \lambda \int_{x_0}^a \overline{B}(y)f_1(y)\,dy + \lambda \int_a^M \overline{B}(y)f_1(y)\,dy. \tag{11.12}$$

The normalizing condition is

$$\int_0^{x_0} f_0(y)\,dy + \int_{x_0}^a f_1(y)\,dy + \int_a^M f_2(y)\,dy = 1. \tag{11.13}$$

Explanation of (11.9)–(11.11).
In (11.9), the left side is the upcrossing rate of level x ($a < x < M$). On the right side, the first term is the downcrossing rate of level x, due to hits of level M from below causing jumps downward to level a. The second term is the downcrossing rate of level x due to claims occurring when the capital reserve is in the interval (x, M). Equation (11.10) is explained similarly.

In (11.11), the left side is the upcrossing rate of level x ($0 < x < x_0$). The first term is the upcrossing rate at continuous points of the sample path. The second term is the total downcrossing rate of level 0 , by (11.12), resulting in immediate upcrossings of level x (a double jump). The right side of (11.11) is the downcrossing rate of x caused by claims when the SP is above x.

Rate balance across level x yields equations (11.9), (11.10) and (11.11).

Remark 11.1 Consider the hit and egress rates of level x_0. We obtain the following rate balance equation

$$cf_0(x_0^-) + \frac{1}{E(\tau)} = cf_1(x_0). \tag{11.14}$$

In (11.14), the left side is the total hit rate of x_0; the right side is the egress rate out of x_0 above (Fig. 11.1). The term $cf_0(x_0^-)$ is the hit rate of level x_0 from below at continuous sample-path points. Since downward jumps ending below level 0 cause ruin and immediate jumps up to level x_0, $1/E(\tau)$ is also the hit rate of level x_0 from below. Finally, on the right side, $cf_1(x_0)$ is the egress rate out of level x_0 above.

From (11.14), $f(x)$ has a jump discontinuity at x_0, given by

$$f_1(x_0) - f_0(x_0^-) = \frac{1}{cE(\tau)}. \tag{11.15}$$

Similar reasoning for level a gives

$$f_2(a) - f_1(a^-) = f_2(M) \tag{11.16}$$

(The derivations of (11.15) and (11.16) are examples of how LC can lead to analytical properties of the pdf in an intuitive manner. These formulas also serve as a check on integral equations (11.9)–(11.11).)

11.1.5 Solution of Model with Exponential Claim Sizes

We now solve (11.9)–(11.11) and (11.13) for $f_0(x), f_1(x)$ and $f_2(x)$ when the claim sizes are $\underset{dis}{=}$ Exp$_\mu$, to illustrate the solution technique. Thus, $e^{-\mu(\cdot)}$ is substituted for $\overline{B}(\cdot)$ in (11.9)–(11.11).

General Case $c \neq \frac{\lambda}{\mu}$

Taking d/dx on both sides of (11.9)–(11.11) and solving the resulting differential equations for $f_2(\cdot), f_1(\cdot)$ and $f_0(\cdot)$ yields

$$\begin{cases} f_0(x) = A'_{10} + B'_{10} e^{(\mu - \frac{\lambda}{c})x}, 0 \le x < x_0, \\ f_1(x) = A'_{11} e^{(\mu - \frac{\lambda}{c})x}, x_0 \le x < a, \\ f_2(x) = f_2(a) \frac{\mu - \left(\frac{\lambda}{c}\right) e^{-(\mu - \frac{\lambda}{c})(M-x)}}{\mu - \left(\frac{\lambda}{c}\right) e^{-(\mu - \frac{\lambda}{c})(M-a)}}, a \le x < M. \end{cases} \tag{11.17}$$

where $A'_{10}, B'_{10}, A'_{11}$ are constants.

Substituting $f_2(\cdot), f_1(\cdot)$ and $f_0(\cdot)$ into (11.9)–(11.11) and (11.13) yields

$$\begin{cases} A'_{10} = \left(C'_{11}\right)^{-1} \lambda \mu e^{x_0 \mu} \left(e^{\frac{\lambda}{c} M + a\mu} - e^{\frac{\lambda}{c} a + M\mu}\right), \\ B'_{10} = -A'_{10}, \\ A'_{11} = \left(C'_{11}\right)^{-1} \left(\frac{\lambda}{c}\right) \left(e^{\frac{\lambda}{c} M + a\mu} - e^{\frac{\lambda}{c} a + M\mu}\right) \left(\lambda e^{\frac{\lambda}{c} x_0} - c\mu e^{x_0 \mu}\right) \\ f_2(a) = \left(C'_{11}\right)^{-1} c e^{a(\mu - \frac{\lambda}{c})} \left(\frac{\lambda}{c} e^{\frac{\lambda}{c} x_0} - \mu e^{x_0 \mu}\right) \left(\frac{\lambda}{c} e^{\frac{\lambda}{c} M + a\mu} - \mu e^{\frac{\lambda}{c} a + M\mu}\right), \end{cases} \tag{11.18}$$

where

$$C'_{11} = (a - M)\, c\mu e^{(M+a)\mu} \left(\frac{\lambda}{c} e^{\frac{\lambda}{c} x_0} - \mu e^{x_0 \mu}\right) + \lambda (1 + x_0 \mu) e^{x_0 \mu} \left(e^{\frac{\lambda}{c} M + a\mu} - e^{\frac{\lambda}{c} a + M\mu}\right).$$

The expected time of ruin is, from (11.12),

$$E(\tau) = C'_{11} \lambda^{-1} (\lambda - c\mu)^{-1} e^{-x_0 \mu} \left(e^{\frac{\lambda}{c} M + a\mu} - e^{\frac{\lambda}{c} a + M\mu}\right)^{-1}. \tag{11.19}$$

Expected Number of Claims Before Ruin

Let N be the number of claims before ruin. From $E(\tau)$ given in (11.19) we obtain

$$E(N) = \lambda E(\tau) = C'_{11} (\lambda - c\mu)^{-1} e^{-x_0 \mu} \left(e^{\frac{\lambda}{c} M + a\mu} - e^{\frac{\lambda}{c} a + M\mu}\right)^{-1}. \tag{11.20}$$

Special Case $c = \frac{\lambda}{\mu}$

By a similar analysis, we can show that when $c = \lambda/\mu$

$$\begin{cases} f_0(x) = B''_{10} x, & 0 \le x < x_0, \\ f_1(x) = A''_{11}, & x_0 \le x < a, \\ f_2(x) = A''_{12} + B''_{12} x, & a \le x < M, \end{cases} \tag{11.21}$$

where

$$B''_{10} = 2\mu^2 \left(C''_{11}\right)^{-1}, \quad A''_{11} = 2\mu\,(1 + x_0\mu)\, C''^{-1}_{11},$$
$$B''_{12} = \left(\frac{1}{a - M}\right) A''_{11}, \quad A''_{12} = -\left(M + \frac{1}{\mu}\right) B''_{12} \tag{11.22}$$

and $C''_{11} = 2 - x_0^2\mu^2 + \mu\,(a + M)\,(1 + x_0\mu)$.

The expected time of ruin is

$$E(\tau) = \left(\frac{1}{2\lambda}\right) C''_{11}. \tag{11.23}$$

Expected Capital Transferred Out Before Ruin

Let N_M be the number of hits of level M and let CAP be the capital transferred out, before ruin. We give a method to derive $E(CAP)$, since each illustrates useful intuitive notions.

E(CAP) Using the renewal reward theorem we get

$$\frac{E(N_M)}{E(\tau)} = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(M))}{t} = cf_2(M)$$
$$\Longrightarrow E(N_M) = cE(\tau)f_2(M). \tag{11.24}$$

When $c \neq \lambda/\mu$, substituting $f_2(M)$ and $E(\tau)$ from (11.17) and (11.19) into (11.24) yields

$$E(CAP) = (M - a)\,E(N_M) = (M - a)\,\left(\frac{c}{\lambda}\right)\left(\frac{\left(\frac{\lambda}{c}\right) e^{\left(\frac{\lambda}{c}-\mu\right)x_0} - \mu}{e^{\left(\frac{\lambda}{c}-\mu\right)M} - e^{\left(\frac{\lambda}{c}-\mu\right)a}}\right). \tag{11.25}$$

When $c = \lambda/\mu$, $E(\tau)$ is given by (11.23), and $f_2(M)$ is obtained from (11.21). Then it can be shown that

$$E(CAP) = x_0 + \frac{1}{\mu}. \tag{11.26}$$

Note that when $c = \lambda/\mu$, the expected capital transferred out for alternative investments before ruin is independent of the barrier M.

For additional analytical details, numerical examples, and applications of LC to other risk models, see [54].

11.2 A Technique for Transient Distributions

This section outlines a technique for deriving transient distributions of continuous-parameter processes with a continuous or discrete state space, denoted as $\{X(t)\}_{t\geq 0}$. The technique is based on the general version of Theorem B (Theorem 4.1 in Sect. 4.2.1). We repeat here formulas (4.1) and (4.2) of Theorem B for easy reference, i.e.,

$$E(\mathcal{I}_t(\boldsymbol{A})) = E(\mathcal{O}_t(\boldsymbol{A})) + P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}),\ \ t \geq 0, \tag{11.27}$$

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}), t > 0, \tag{11.28}$$

where $\mathcal{I}_t(\boldsymbol{A})$ is the number of SP entrances, and $\mathcal{O}_t(\boldsymbol{A})$ is the number of SP exits, of state-space set $\boldsymbol{A}$ during $[0, t]$. Let the parameter set be $\boldsymbol{T} = [0, \infty)$.

Remark 11.2 If the *limiting* distribution of $\{X(t)\}_{t\geq 0}$ exists, it is obtained by taking the limit of the derived transient distribution as $t \to \infty$.

11.2.1 State-Space Set with Variable Boundary

State Space $\boldsymbol{S} \subseteq \mathbb{R}$

In formulas (11.27) and (11.28) assume set $\boldsymbol{A}$ depends on a continuous variable x and define $\boldsymbol{A} := \boldsymbol{A}_x$, $x \in \boldsymbol{S}$, where x is a state-space level, e.g., $\boldsymbol{T} \times \{x\}$ (a line in the $\boldsymbol{T}$-$\boldsymbol{S}$ coordinate system parallel to the time axis). For fixed x, replace formulas (11.27) and (11.28) by

$$E(\mathcal{I}_t(\boldsymbol{A}_x)) = E(\mathcal{O}_t(\boldsymbol{A}_x)) + P_t(\boldsymbol{A}_x) - P_0(\boldsymbol{A}_x), \tag{11.29}$$

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A}_x)) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}_x). \tag{11.30}$$

Assume the following mixed partial derivatives exist and are equal, i.e.,

$$\frac{\partial^2}{\partial x \partial t}E(\mathcal{O}_t(\boldsymbol{A}_x)) = \frac{\partial^2}{\partial t \partial x}E(\mathcal{O}_t(\boldsymbol{A}_x)),$$
$$\frac{\partial^2}{\partial x \partial t}P_t(\boldsymbol{A}_x) = \frac{\partial^2}{\partial t \partial x}P_t(\boldsymbol{A}_x).$$

Taking $\partial/\partial x$ in (11.30) we obtain

$$\frac{\partial^2}{\partial x \partial t}E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial^2}{\partial t \partial x}E(\mathcal{O}_t(\boldsymbol{A}_x)) + \frac{\partial^2}{\partial t \partial x}P_t(\boldsymbol{A}_x). \tag{11.31}$$

State Space $\boldsymbol{S} \subseteq \mathbb{R}^n$

Let $\{\boldsymbol{X}(t)\}_{t\geq 0}$ denote a continuous-time process with n-dimensional state space $\boldsymbol{S} \subseteq \mathbb{R}^n$. The state space may be continuous or discrete. Let vector $\boldsymbol{x} = (x_1, ..., x_n)$, and let state-space set $\boldsymbol{A_x} = \cap_{i=1}^{n}(-\infty, x_i] \subseteq \boldsymbol{S}$. Then $P_t(\boldsymbol{A_x}) = F_t(\boldsymbol{x}) = F_t(x_1, ..., x_n)$ is the joint cdf of the n state variables at time $t \geq 0$.

From the general formula (11.29) the joint cdf is given by

$$F_t(\boldsymbol{x}) = E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x)) + F_0(\boldsymbol{x})$$

where $F_0(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{X}(0) \in \boldsymbol{A}_x, \\ 0 & \text{if } \boldsymbol{X}(0) \notin \boldsymbol{A}_x \end{cases}$.

Provided the derivatives exist, we obtain

$$\begin{aligned} \frac{\partial F_t(\boldsymbol{x})}{\partial x_i} &= \frac{\partial}{\partial x_i}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right], i = 1, ..., n, \\ \frac{\partial^n F_t(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n} &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right], \\ \frac{\partial F_t(\boldsymbol{x})}{\partial t} &= \frac{\partial}{\partial t}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right]. \end{aligned}$$

If $\frac{\partial E(\mathcal{I}_t(\boldsymbol{A}_x))}{\partial t}$ and $\frac{\partial E(\mathcal{O}_t(\boldsymbol{A}_x))}{\partial t}$ can be expressed as functions of $F_t(\boldsymbol{x})$ or $f_t(\boldsymbol{x})$, then we may be able to derive an integro-differential equation for $F_t(\boldsymbol{x})$ or $f_t(\boldsymbol{x})$.

If $n = 1$ the state space is one-dimensional, and $\boldsymbol{A}_x = (-\infty, x]$. Thus

$$f_t(x) = \frac{\partial}{\partial x}\left[E(\mathcal{I}_t((-\infty, x])) - E(\mathcal{O}_t((-\infty, x]))\right]$$

where $f_t(x) :=$ *transient pdf of* $\boldsymbol{X}(t)$.

LC Computation

The expressions in this Section can aid in estimating or computing the transient cdf and pdf of a continuous-parameter n-dimensional process using LCE (level crossing estimation or computation) for transient distributions. We will not expound on this transient LCE technique further in this monograph. LCE for transient distributions is discussed briefly in Remark 3.7 and Example 3.1 in Sect. 3.2.8, and briefly mentioned in Remark 9.2 in Sect. 9.2.

11.3 Discrete-Parameter Processes

Let $\{X_n\}_{n=0,1,2,...}$ denote a discrete-parameter process taking values in a state space $\boldsymbol{S}$, which may be discrete or continuous. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be (measurable) subsets of $\boldsymbol{S}$. Let $P_n(\boldsymbol{A}) = P(X_n \in \boldsymbol{A})$, and $P_{m,n}(\boldsymbol{B}, \boldsymbol{C}) = P(X_m \in \boldsymbol{B}, \boldsymbol{X}_n \in \boldsymbol{C})$.

Definition 11.1 (a) The SP **exits** set $\boldsymbol{A}$ at time n if $X_n \in \boldsymbol{A}$ and $X_{n+1} \notin \boldsymbol{A}$. (b) The SP **enters** set $\boldsymbol{A}$ at time n if $X_{n-1} \notin \boldsymbol{A}$ and $X_n \in \boldsymbol{A}$. (c) $\mathcal{I}_n(\boldsymbol{A}))$:= **number of** SP entrances into $\boldsymbol{A}$ during $[0, n]$. (d) $\mathcal{O}_n(\boldsymbol{A})$:= **number of** SP exits out of $\boldsymbol{A}$ during $[0, n]$.

We now state a theorem for discrete-time processes which is analogous to Theorem B (see formulas (11.29) and (11.30)).

Theorem 11.2 Let $\{X_n\}_{n=0,1,2,...}$ be a discrete-time process with state space $\boldsymbol{S}$. Let $\boldsymbol{A} \subseteq \boldsymbol{S}$.

$$E(\mathcal{I}_n(\boldsymbol{A})) = E(\mathcal{O}_n(\boldsymbol{A})) + P_n(\boldsymbol{A}) - P_0(\boldsymbol{A}), \tag{11.32}$$

where $P_0(\boldsymbol{A}) = \begin{cases} 1 \text{ if } X_0 \in A, \\ 0 \text{ if } X_0 \notin A \end{cases}.$

Proof The proof is similar to that of Theorem 4.1 in Sect. 4.2.1 of Chap. 4, upon replacing t by n. □

11.3.1 Application to Markov Chains

Let $\{X_n\}_{n=0,1,...}$ be a Markov chain with the discrete state space $\boldsymbol{S}$. For example, let $\boldsymbol{S} = \{0, \pm 1, \pm 2, ...\}$. Let the set $\boldsymbol{A} := j \in \boldsymbol{S}$. Then

$$E(\mathcal{I}_n(j)) = \sum_{i \neq j} \sum_{m=0}^{n-1} P_i^m P_{ij}, \text{ and } E(\mathcal{O}_n(j)) = \sum_{i \neq j} \sum_{m=0}^{n} P_j^m P_{ji},$$

where P_{ij} is the one-step transition probability from i to j and $P_j^m \equiv P_m(\boldsymbol{A}) = P_m(j)$. Substituting into (11.32) gives

$$P_j^n = \sum_{i \neq j} \sum_{m=0}^{n-1} P_i^m P_{ij} - \sum_{i \neq j} \sum_{m=0}^{n} P_j^m P_{ji} + P_j^0. \tag{11.33}$$

Assume the following limiting probabilities exist:

$$\lim_{n\to\infty} P^n_{i,j} = \lim_{n\to\infty} P^n_{jj} = \lim_{n\to\infty} P^n_j \equiv \pi_j,$$

where $P^n_{i,j}$ is the n-step transition probability from i to j. That is, the Markov chain is positive recurrent and aperiodic, and $\sum_{j\in \boldsymbol{S}} \pi_j = 1$. Dividing both sides of (11.33) by n and letting $n \to \infty$ yields

$$\begin{aligned}
\lim_{n\to\infty} \tfrac{P^n_j}{n} = \textstyle\sum_{i\neq j} \left(\lim_{n\to\infty} \tfrac{1}{n} \sum_{m=0}^{n-1} P^m_i\right) P_{i,j} \\
- \textstyle\sum_{i\neq j} \left(\lim_{n\to\infty} \tfrac{1}{n} \sum_{m=0}^{n} P^m_j\right) P_{j,i} + \lim_{n\to\infty} \tfrac{P^0_j}{n}, \\
0 = \textstyle\sum_{i\neq j} \pi_i P_{i,j} - \sum_{i\neq j} \pi_j P_{j,i} + 0, \\
\textstyle\sum_{i\neq j} \pi_j P_{j,i} = \sum_{i\neq j} \pi_i P_{i,j}, \\
\pi_j (1 - P_{j,j}) = \textstyle\sum_{i\neq j} \pi_i P_{i,j}, \\
\pi_j = \textstyle\sum_{i\in \boldsymbol{S}} \pi_i P_{i,j}, j \in \boldsymbol{S}.
\end{aligned}$$

Thus we have derived the classical equations for the limiting probabilities $\{\pi_j\}_{j\in \boldsymbol{S}}$ by using an LC method, namely

$$\begin{aligned}
\pi_j = \textstyle\sum_{i\in \boldsymbol{S}} \pi_i P_{i,j}, j \in \boldsymbol{S}, \\
\textstyle\sum_{j\in \boldsymbol{S}} \pi_j = 1.
\end{aligned} \tag{11.34}$$

Remark 11.3 We have applied the discrete-time analog of Theorem B to a standard Markov chain in order to demonstrate its applicability to discrete-time discrete-state models. Theorem B emphasizes the **system point aspect** of the SPLC method. SPLC utilizes SP entrance/exit rates of state-space sets. (SP level crossings are special cases of SP entrances and exits.)

11.4 Semi-Markov Process

Consider a semi-Markov process (SMP) $\{X(t)\}_{t\geq 0}$, with discrete state space $\boldsymbol{S}$ (also called a *Markov renewal process*). (See, e.g., pp. 207–208 in [99]; pp. 457–460 in [125].) Let the sojourn time in state $j \in \boldsymbol{S}$ have a *general* distribution with mean $\mu_j > 0$. The type of distribution of the sojourn time may differ from state to state; only the *means* are utilized in this analysis. At the end of a sojourn in state i, say at instant τ^-, assume $P\left(X(t) = j | X(t^-) = i\right) = P_{i,j}, j \neq i, j \in \boldsymbol{S}$. The matrix $\|P_{i,j}\|$ is a Markov matrix. Assume the Markov chain with transition matrix $\|P_{i,j}\|$ is positive recurrent and aperiodic so that the limiting probabilities $\pi_j, j \in \boldsymbol{S}$ exist.

Let $P_j(t) := P(X(t) = j)$, $t \geq 0$ and $P_j := \lim_{t\to\infty} P_j(t)$, $j \in \boldsymbol{S}$. We now derive the probabilities $P_j, j \in \boldsymbol{S}$, using LC.

Consider a sample path of $\{X(t)\}_{t\geq 0}$. Let $T_t(i)$ denote the total time spent by the SP in state i during $(0, t)$, and

$$\boldsymbol{I}_i(X(s)) = \begin{cases} 1 & \text{if } X(s) = i, s \in [0, t], \\ 0 & \text{if } X(s) \neq i, s \in [0, t] \end{cases}.$$

Then $E\left(\boldsymbol{I}_i(X(s))\right) = P_i(s)$, and $T_t(i) = \int_{s=0}^{t} \boldsymbol{I}_i(X(s)ds$, implying

$$E(T_t(i)) = \int_{s=0}^{t} E\left(\boldsymbol{I}_i(X(s)\right) ds = \int_{s=0}^{t} P_i(s)ds. \tag{11.35}$$

The expected number of SP exits from state i during $(0, t)$ is $E(T_t(i))/\mu_i$ since the mean of each sojourn time in i is μ_i. The expected number of SP $i \to j$ transitions during $(0, t)$ is $E(T_t(i))/\mu_i P_{i,j}$. The expected *total number of SP transitions* into (*entrances* into) state j during $(0, t)$ is

$$E(\mathcal{I}_t(j)) = \sum_{i\neq j} \frac{E(T_t(i))}{\mu_i} P_{i,j}. \tag{11.36}$$

Similarly, the expected number of SP *exits* out of j during $(0, t)$ is

$$E(\mathcal{O}_t(j)) = \frac{E(T_t(j))}{\mu_j}. \tag{11.37}$$

Substituting from (11.36) and (11.37) into Theorem B, formula (11.27), gives

$$\sum_{i\neq j} \frac{E(T_t(i))}{\mu_i} P_{i,j} = \frac{E(T_t(j))}{\mu_j} + P_j(t) - P_j(0). \tag{11.38}$$

(We assume the interchange of summation and the limit operation is valid. This applies if, e.g., $\boldsymbol{S}$ has a finite number of states.)

From (11.35), the long-run proportion of time the SP is in state $i \in \boldsymbol{S}$ is

$$\lim_{t\to\infty} \frac{E(T_t(i))}{t} = P_i, i \in \boldsymbol{S}.$$

Since $0 \leq P_j(t) \leq 1$, $t \geq 0$,

$$\lim_{t\to\infty} \frac{P_j(t)}{t} = \lim_{t\to\infty} \frac{P_j(0)}{t} = 0,$$

Dividing both sides of (11.38) by $t > 0$ and letting $t \to \infty$ gives

$$\sum_{i \neq j} \frac{P_i}{\mu_i} P_{ij} = \frac{P_j}{\mu_j}, j \in \boldsymbol{S} \tag{11.39}$$

Suppose $\sum_{j \in S} \frac{1}{\mu_j} P_j = K > 0$; then $\sum_{j \in S} \left(\frac{1}{K\mu_j} P_j \right) = 1$. Dividing both sides of (11.39) by K and transposing terms gives the system of equations for $P_i, i \in \boldsymbol{S}$,

$$\begin{aligned} &\tfrac{1}{K\mu_j} P_j = \textstyle\sum_{i \neq j} \left(\frac{1}{K\mu_j} P_i \right) P_{ij}, j \in \boldsymbol{S} \\ &\textstyle\sum_{j \in S} \left(\frac{1}{K\mu_j} P_j \right) = 1. \end{aligned} \tag{11.40}$$

The system of equations (11.40) for $\left\{ 1/\left(K\mu_j \right) \cdot P_j \right\}_{j \in \boldsymbol{S}}$ is identical to the system of equations (11.34) for $\left\{ \pi_j \right\}_{j \in \boldsymbol{S}}$ in Markov chains. Thus

$$\frac{1}{K\mu_j} P_j = \pi_j, j \in \boldsymbol{S},$$

$$P_j = (\pi_j \mu_j) K, j \in \boldsymbol{S}, \tag{11.41}$$

and K is obtained from the normalizing condition

$$\sum_{j \in S} P_j = K \sum_{j \in S} \pi_j \mu_j = 1,$$

namely

$$K = \frac{1}{\sum_{j \in S} \pi_j \mu_j}, \tag{11.42}$$

which substituted into (11.41) gives the well-known formula

$$P_j = \frac{\pi_j \mu_j}{\sum_{j \in \boldsymbol{S}} \pi_j \mu_j},\ j \in \boldsymbol{S}. \tag{11.43}$$

The key steps in this LC derivation of (11.43) are: (1) obtain expressions for the expected SP *entrance and exit rates* of each state; (2) apply formula (11.27) of Theorem B; (3) divide by t and take $\lim_{t \to \infty}$; (4) evaluate the constant K by recognizing the role of the **linear Markov-chain equations** (11.34) for $\left\{ \pi_j \right\}_{j \in \boldsymbol{S}}$.

11.5 Non-homogeneous Pure Birth Processes

Consider the pure birth process $\{X(t)\}_{t\geq 0}$, where $X(t)$ denotes the population at time $t > 0$. Let the initial population be $X(0) = i$, a non-negative integer. Let the sequence of positive functions $\lambda_k(t)$, $k = i, i+1, ...,$ $(i = 0, 1, ...)$, denote the birth rate at time t given that the population at t is k, with the property

$$P(X(t+h) - X(t) = 1 | X(t) = k) = \lambda_k(t)h + o(h),$$
$$P(X(t+h) - X(t) = 0 | X(t) = k) = 1 - \lambda_k(t)h + o(h),$$

where $h > 0$. Define $P_n(t) := P(X(t) = n)$.

We now derive an expression for $P_n(t), t > 0$, $n = i, i+1, ...$, using Theorem B, i.e., formulas (11.27) and (11.28).

The expected number of SP *entrances into state i* during $(0, t)$ is $E(\mathcal{I}_t(i))$ $= 0$, since $X(0) = i$, and $X(t), t > 0$, never visits state i once it increases from i to $i+1$. On the other hand the expected number of SP *exits out of state i* during $(0, t)$ is $E(\mathcal{O}_t(i)) = \int_{s=0}^{t} \lambda_s(i) P_i(s) ds$, since an SP $i \to i+1$ transition can occur at any instant $s \in (0, t)$. Note that $P_i(0) = 1$. Substituting $E(\mathcal{I}_t(i))$, $E(\mathcal{O}_t(i))$ and $P_i(0)$ into (11.27), we obtain

$$0 = \int_{s=0}^{t} \lambda_i(s) P_i(s) ds + P_i(t) - 1. \tag{11.44}$$

Differentiating (11.44) with respect to t gives

$$\frac{d}{dt} P_i(t) + \lambda_i(t) P_i(t) = 0$$

having solution, since $P_i(0) = 1$,

$$P_i(t) = e^{-m_i(t)}, t \geq 0, \tag{11.45}$$

where

$$m_i(t) = \int_{s=0}^{t} \lambda_i(s) ds.$$

For an arbitrary state $j > i$,

$$E(\mathcal{I}_t(j)) = \int_{s=0}^{t} \lambda_{j-1}(s) P_{j-1}(s) ds, \tag{11.46}$$

$$E(\mathcal{O}_t(j)) = \int_{s=0}^{t} \lambda_j(s) P_j(s) ds. \tag{11.47}$$

Substituting from (11.46) and (11.47) into (11.27) gives

$$\int_{s=0}^{t} \lambda_{j-1}(s)P_{j-1}(s)ds = \int_{s=0}^{t} \lambda_j(s)P_j(s)ds + P_j(t) - 0. \tag{11.48}$$

Taking d/dt on both sides of (11.48) yields

$$\frac{d}{dt}P_j(t) + \lambda_j(t)P_j(t) = \lambda_{j-1}(t)P_{j-1}(t),$$

with solution

$$P_j(t) = e^{-m_j(t)} \int_{s=0}^{t} e^{m_j(s)} \lambda_{j-1}(s)P_{j-1}(s)ds, \, t \geq 0. \tag{11.49}$$

Formula (11.49) provides a recursive solution expressing $P_j(t)$ in terms of $P_{j-1}(t)$, $j = i + 1$,...., and $P_i(0) = 1$.

11.5.1 Non-homogeneous Poisson Process

The non-homogeneous Poisson process is a special case of the pure birth process (see, e.g., pp. 339–345 in [125]). Assume $X(0) = 0$, $\lambda_j(t) \equiv \lambda(t)$ independent of the state j, so that $m(t) = \int_{s=0}^{t} \lambda(s)ds$. Setting $i = 0$ in (11.45) gives $P_0(t) = e^{-m(t)}$. From (11.49) we obtain (by induction) the well-known formula

$$P_n(t) = e^{-m(t)} \frac{(m(t))^n}{n!}, \, n = 0, 1, 2, \ldots . \tag{11.50}$$

Formula (11.50) is the pmf (probability mass function) of a Poisson distribution with mean $m(t)$. Then $P_n(t)$, $n = 0, 1,...,$ for the standard Poisson process are obtained from (11.50) by setting $\lambda(t) \equiv \lambda$, so that $m(t) \equiv \lambda t$.

11.5.2 Yule Process

The Yule process is a special case of the pure birth process, where the birth rate (growth rate) is *directly proportional to the current population size*, but independent of time t. Assume $X(0) = 1$ and $\lambda_i(t) = i\lambda$, $t \geq 0$, $i = 1, 2,...$. Then $P_1(t) = e^{-\lambda t}$ (= probability of no births during $(0, t)$). Using (11.49) and mathematical induction, we obtain the well-known geometric distribution for the Yule process

$$P_n(t) = (1 - e^{-\lambda t})^{n-1} e^{-\lambda t}, n = 1, 2,.... \tag{11.51}$$

Let $P_{k,(i)}(t) := P(i$ *independent* Yule processes *with the same parameter* λ yield a total of $k \geq i$ individuals at time $t > 0)$. Assume each process starts in state 1 at time 0. Since $P_n(t)$ in (11.51) is a geometric distribution, we obtain the convolution of i i.i.d. geometric distributions as the negative binomial distribution

$$P_{k,(i)}(t) = \binom{k-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{k-i}, k = i, i+1,.... \tag{11.52}$$

(see, e.g., p. 164ff in [73]). Formulas (11.51) and (11.52) can be derived in several different ways (e.g., pp. 122–123 in [99]; pp. 383–384 in [125]). We now outline a direct proof of (11.52) using LC.

We derive similarly as for (11.49),

$$P_{i,(k)}(t) = (k+1)\lambda e^{-k\lambda t} \int_{s=0}^{t} e^{k\lambda s} P_{i,(k-1)}(s) ds + C_k e^{-k\lambda t}, k \geq i, \tag{11.53}$$

where $C_k = \begin{cases} 1 \text{ if } k = i, \\ 0 \text{ if } k > i \end{cases}$. Since $P(\text{no births in } (0, t)) = P(\text{Exp}_{i\lambda} > t)$ we have

$$P_{i,(i)}(t) = e^{-i\lambda t}. \tag{11.54}$$

Thus (11.52) holds for $k = i$. From (11.54) and (11.53) with $k = i + 1$, we obtain

$$P_{i,(i+1)}(t) = ie^{-i\lambda t}\left(1 - e^{-\lambda t}\right) = \binom{i+1-1}{i-1} e^{-i\lambda t}\left(1 - e^{-\lambda t}\right). \tag{11.55}$$

Therefore (11.52) holds for $k = i + 1$.

Assume (11.52) holds for an arbitrary integer $k > i$. We then show using (11.53) that (11.52) holds for $k + 1$. Hence it holds for all $k = i, i + 1,...$, by the principle of mathematical induction.

11.6 Pharmacokinetic Model

This Section outlines an LC approach to multiple dosing in pharmacokinetics, by briefly discussing a simplified one-compartment model. We assume bolus dosing, i.e., a full dose of a drug is absorbed into the blood stream immediately at each dosing instant. Also, inter-dose times are $\underset{dis}{=} \text{Exp}_{\lambda}$. Thus doses occur in a Poisson process at rate λ. This assumption may be valid outside of a controlled environment. Statistical tests have shown that many patients take certain medications over time in a Poisson process [47].

We assume first-order kinetics. That is, the concentration of the drug in the blood stream decays at a rate which is proportional to the concentration. This is equivalent to a plot of the concentration over time having a negative exponential shape between doses (similar to Fig. 11.2 below).

11.6.1 Model Description

Let $\{W(t)\}_{t \geq 0}$, denote the drug concentration at time t. Let the dosing times be $\{\tau_n\}$, $\tau_n < \tau_{n+1}$, $n = 0, 1, 2, \ldots$ The rate of concentration decay due to drug elimination is

$$\frac{dW(t)}{dt} = -kW(t), \tau_n \leq t < \tau_{n+1}, n = 0, 1, 2, \ldots, \tag{11.56}$$

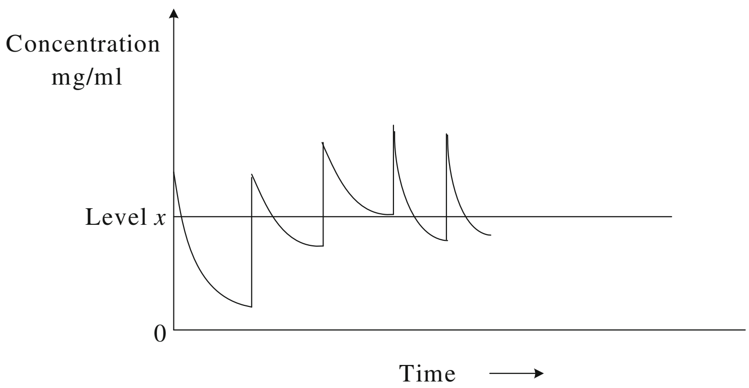


Fig. 11.2 Sample path of drug concentration $\{W(t)\}_{t \geq 0}$ in one-compartment model with bolus dosing and first-order kinetics

where $k > 0$. The dimension of the concentration $W(t)$ is $[W(t)] = \left[\frac{Mass}{Volume}\right]$; of decay rate is $\left[\frac{dW(t)}{dt}\right] = \left[\frac{Mass}{Volume}\right] \cdot \left[Time^{-1}\right]$; of the constant k is $[k] = [Time]^{-1}$.

Let $\{P_0, f(x)\}_{x>0}$, denote the *steady-state pdf* of concentration. Then $P_0 = 0$ due to the negative exponential shape of the decay graph between doses (see Sects. 6.2.4, 6.2.5 and 6.4 in Chap. 6). In theory, the concentration of the drug never vanishes. In practice, it goes to 0 or is negligible. (We are not discussing the treatment effects of multiple dosing; only the concentration dynamics.) Table 11.1 below indicates the close relationship between the M/G/r($\cdot$) Dam and the Pharmacokinetic model.

11.6.2 Dose Amounts Exponentially Distributed

We first analyze a model which assumes exponentially distributed dose sizes. This assumption may be valid if the amount of each dose absorbed is affected randomly by the dosing environment (e.g., acidity, presence of enzymes, interaction with other medications, etc.). Another instance could occur when eye drops are instilled by a patient, say *approximately* every 6 h. Often, the sizes of the individual drops may vary considerably, due to using a hand-squeezed medicine dropper. The location on the cornea of the instillation may vary from dose to dose, possibly affecting absorption. This could create random increases in concentration with the successive doses during a dosing regime. Similar remarks apply to fast-acting sprays, such as nitrolingual pump sprays, or to nasal sprays (Also, for certain medications it may be feasible to study the effect of using random dose sizes as an exponential random variable inherently in a prescription, in order to test whether it will decrease the long-run variance of concentration.).

Table 11.1 M/G/r($\cdot$) Dam versus Pharmakokinetic model

M/G/r($\cdot$) Dam	Pharmacokinetic Model
Input instant	Bolus dose instant
Input amount (jump size)	Dose amount (jump size)
Content $W(t), t \geq 0$	Concentration $W(t), t \geq 0$
Sample-path slope $-r(x), x > 0$	Sample-path slope $-r(x), x > 0$
CDF/PDF of content	CDF/PDF of concentration
Mean content	Average drug concentration
Variance of content	Variance of concentration

Let us assume the bolus dose amounts are randomly distributed as Exp_μ. Since $P_0 = 0$ the LC balance equation for the pdf of concentration is

$$kxf(x) = \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy. \tag{11.57}$$

Equation (11.57) has the solution

$$f(x) = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} (\mu x)^{(\frac{\lambda}{k}-1)} e^{-\mu x} \mu, x > 0. \tag{11.58}$$

where $\Gamma(\cdot)$ is the Gamma function (see Sect. 6.4, formulas (6.48) and (6.49)). Let W denote the steady-state concentration of $W(t)$ as $t \to \infty$, having the pdf in (11.58). The first and second moments of W are

$$E(W) = \frac{\lambda}{k\mu}, \quad E(W^2) = \frac{\lambda}{k\mu^2}\left(\frac{\lambda}{k} + 1\right).$$

The variance of W is

$$Var(X) = E(W^2) - (E(W))^2 = \frac{\lambda}{k\mu^2}.$$

We can find the probability that the W is between two threshold limits, say $\alpha < \beta$, using

$$P(\alpha < W < \beta) = \int_{x=\alpha}^{\beta} \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} \mu(\mu x)^{(\frac{\lambda}{k}-1)} e^{-\mu x} dx. \tag{11.59}$$

The information in (11.59) can be useful when multiple dosing continues for a long time, e.g., when administering the blood thinner coumadin (warfarin). If the concentration is $<\alpha$ coumadin is not effective for the intended treatment. If the concentration is $>\beta$ the blood becomes too thin. The interval (α, β) is thought of as the therapeutic range (see Fig. 11.2).

The type of analysis outlined briefly here can be extended to various pharmacokinetic models of varying complexity.

11.6.3 Dose Amounts Deterministic

We next suppose the dose amounts are all of size $D > 0$ units (e.g., milligrams). This model is equivalent to an M/D/r($\cdot$) dam (see, e.g., Sect. 6.2).

Equation for Stationary CDF and PDF of Concentration
Consider a sample path of $\{W(t)\}_{t\geq 0}$ (similar to Fig. 11.2 with all jump amounts equal to D). Fix level $x > 0$. The SP downcrossing rate of level $x > 0$ is $kxf(x)$. The SP upcrossing rate of x is equal to $\lambda F(x) - \lambda F(x-D)$ where $F(x) = \int_{y=0}^{x} f(y)dy$. (see Sect. 3.10). The principle of rate balance across level x gives an equation for the pdf $f(x)$ and cdf $F(x)$ of W, namely

$$kxf(x) = \lambda F(x) - \lambda F(x-D), x > 0. \tag{11.60}$$

Due to jumps of size D we derive $f(x)$ and $F(x)$ on successive non-overlapping intervals of length D. Denote the pdf and cdf of W as $f_0(x)$, $F_0(x)$, $x \in (0, D]$, and as $f_n(x)$, $F_n(x)$, $x \in [nD, (n+1)D)$, $n = 1, 2,...$

In differential equation (11.60) for $F(\cdot)$ (since $f(x) = F'(x)$), observe that $F(x-D) = 0$ for $x \in (0, D)$. Thus

$$kxf_0(x) = \lambda F_0(x), x \in (0, D),$$

implying

$$\frac{F_0'(x)}{F_0(x)} = \frac{d}{dx}\ln F_0(x) = \frac{\lambda}{kx}, x \in (0, D)$$

with solution

$$F_0(x) = Ax^{\frac{\lambda}{k}}, x \in (0, D), \tag{11.61}$$

where A is a constant to be determined. For $x \in [D, 2D)$, substituting from (11.61) into equation (11.60) gives

$$\begin{aligned}
kxf_1(x) &= \lambda F_1(x) - \lambda F_0(x-D), \\
f_1(x) &= \frac{\lambda}{kx}F_1(x) - \frac{\lambda}{kx}A(x-D)^{\frac{\lambda}{\kappa}}, \\
F_1'(x) - \frac{\lambda}{kx}F_1(x) &= -\frac{\lambda}{kx}A(x-D)^{\frac{\lambda}{\kappa}}, x \in [D, 2D).
\end{aligned} \tag{11.62}$$

To solve the differential equation (11.62) multiply both sides by the integrating factor $e^{-\int_D^x \frac{\lambda}{ku}du}$ and apply continuity of the CDF at level D, i.e., $F_1(D^+) = F(D^-)$, to obtain

$$\begin{aligned}
F_1'(x)e^{-\int_D^x \frac{\lambda}{ku}du} - \frac{\lambda}{kx}F_1(x)e^{-\int_D^x \frac{\lambda}{ku}du} &= -\frac{\lambda}{kx}A(x-D)^{\frac{\lambda}{\kappa}}e^{-\int_D^x \frac{\lambda}{ku}du}, \\
\frac{d}{dx}F_1(x)e^{-\int_D^x \frac{\lambda}{ku}du} &= -\frac{\lambda}{kx}A(x-D)^{\frac{\lambda}{\kappa}}e^{-\int_D^x \frac{\lambda}{ku}du}, \\
F_1(x) = -e^{+\frac{\lambda}{k}\ln(x/D)}\int_D^x \frac{\lambda}{ky}A(y-D)^{\frac{\lambda}{\kappa}}e^{-\frac{\lambda}{k}\ln(y/D)}dy &+ C_1 e^{+\frac{\lambda}{k}\ln(x/D)},
\end{aligned}$$

where C_1 is a constant. Setting $x = D$ in the last equality gives $F_1(D) = C_1 = AD^{\frac{\lambda}{k}} = F_0(D^-)$, from (11.61) since ln (1) = 0 , which leads to

$$F_1(x) = Ax^{\frac{\lambda}{k}}\left[-\frac{\lambda D}{k}\int_D^x \left(\frac{y}{D}-1\right)^{\frac{\lambda}{\kappa}}\left(\frac{y}{D}\right)^{-\left(\frac{\lambda}{k}+1\right)}dy+1\right], x \in [D, 2D), \tag{11.63}$$

where A is defined in (11.61). Setting $x = D$ in (11.63), checks with $F_1(D^+) = AD^{\frac{\lambda}{k}} = F_0(D^-)$.

The solution for $F_n(x)$, $x \in [nD, (n+1)D)$, $n = 2, 3, ...$, can be obtained by a similar recursive procedure and mathematical induction. We now give the induction step. Suppose we know $F_n(x)$, $x \in [nD, (n+1)D)$. For $x \in [(n+1)D, (n+2)D)$, letting the integrating factor be $\mathbf{\Phi}(x) := e^{-\frac{\lambda}{k}\int_{(n+1)D}^x \frac{1}{u}du}$, we get

$$\begin{aligned}
F'_{n+1}(x)\mathbf{\Phi}(x) - \frac{\lambda}{kx}F_{n+1}(x)\mathbf{\Phi}(x) &= -\frac{\lambda}{kx}F_n(x-D)\mathbf{\Phi}(x),\\
\frac{d}{dx}F_{n+1}(x)\mathbf{\Phi}(x) &= -\frac{\lambda}{kx}F_n(x-D)\mathbf{\Phi}(x),\\
F_{n+1}(x)\left(\frac{x}{(n+1)D}\right)^{-\frac{\lambda}{k}} &= -\frac{\lambda}{k}\int_{(n+1)D}^x \frac{1}{y}F_n(y-D)\left(\frac{y}{(n+1)D}\right)^{-\frac{\lambda}{k}}\\
&\quad +C_{n+1},\\
F_{n+1}(x) = -\frac{\lambda}{k}\left(\frac{x}{(n+1)D}\right)^{+\frac{\lambda}{k}} &\int_{(n+1)D}^x \frac{1}{y}\left(\frac{y}{(n+1)D}\right)^{-\frac{\lambda}{k}}F_n(y-D)dy\\
&\quad +C_{n+1}\left(\frac{x}{(n+1)D}\right)^{+\frac{\lambda}{k}}.
\end{aligned}$$

Letting $x = (n+1)D$ gives $C_{n+1} = F_{n+1}((n+1)D) = F_n((n+1)D^-)$ which is known on the assumption we know $F_n(x)$, $x \in [nD, (n+1)D)$. Moreover, C_{n+1} will be in terms of the factor A. Thus, in principle, we can derive $F_n(x)$, $n = 0, 1, 2, \ldots$, in terms of A. The constant A can then be determined (or closely approximated) using the normalizing condition $F(\infty) = 1$. Once $F(x)$ is obtained, we can determine $f(x)$ by substituting into (11.60) (as in Sect. 3.10).

11.6.4 Using LCE to Compute $f(x)$

As an alternative to the foregoing analytical solution, we can solve for $f(x)$ using LCE (LC Estimation) via a typical simulated sample path of $\{W(s)\}_{s\geq 0}$ over a long time $t > 0$, to estimate the pdf $f(x)$ for $x \in (0, x_{M(t)})$ where $x_{M(t)}$ is the maximum state-space partition level (see Sects. 9.3.1 and 9.6).

Remark 11.4 We mention that it is possible to apply Theorem B to compute the **time-dependent pdf and cdf of concentration** (see formulas (11.27)–(11.30) in Sect. 11.2). Knowledge of transient distributions may be useful in multiple dosing regimes where it is important to estimate the concentration after a short dosing period, or to manage dosing quantities.

Remark 11.5 Some related stochastic models have characteristics in common with the pharmacokinetic model. One group of such models involves consumer response (CR) to non-uniform advertisements, which have been analyzed along similar lines using LC (see, e.g., [40]).

11.7 Counter Models

For a description of a classical example of a counter model see pp. 128–131 in [99]. In this section we analyze the transient total output of type-1 and type-2 counters, using LC.

11.7.1 Type-2 Counter

We first analyze a type-2 counter. Electrical pulses arrive in a Poisson process at rate λ. Each *arriving* pulse is followed immediately by a fixed *locked* period of length $D > 0$, during which new arrivals cannot be detected by the counter. However, if a new arrival occurs at a time t while the counter is locked, then the locked period is extended to time $t + D$. Thus the locked periods "telescope" to form a '*total locked period*' denoted by L, implying $L \geq D$. Arrivals can be detected only when the counter is unlocked or *free*. Let us assume that the counter is free at time $t = 0$ (see Fig. 11.3).

Let the amplitudes of the pulses be $\underset{dis}{\equiv} X$, having cdf $B(y)$, $y > 0$. Let $\eta_i(t)$, $t \geq \tau_i$, denote the output at time t due to the *detected pulse* X_i occurring at τ_i. Assume that $\eta_i(t)$ dissipates at rate

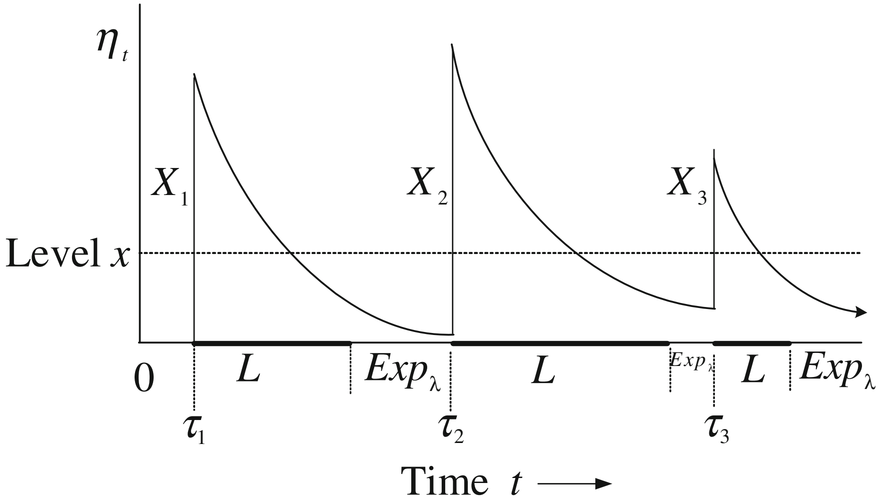


Fig. 11.3 Sample path of total output η_t for type-2 counter model. Total locked periods are each $\underset{dis}{=} L \geq D$. Arrivals during locked periods are not detected, but extend it by D. Pulses arrive at Poisson rate λ

$$\frac{d\eta_i(t)}{dt} = -k \cdot \eta_i(t), t > \tau_i, \tag{11.64}$$

where the constant $k > 0$ is the same for all $i = 1, 2, \ldots$.

Let η_t denote the *total output* at time t, due to all *registered* (detected) pulses that arrive during $(0, t)$ (see Fig. 11.3). If the number of detected pulses in $(0, t)$ is n, then

$$\eta_t = \sum_{i=1}^{n} \eta_i(t), \tau_n \leq t < t_{n+1}, n = 1, 2, \ldots, \tag{11.65}$$

From (11.65)

$$\frac{d}{dt}\eta_t = -k \sum_{i=1}^{n} \eta_i(t) = -k\eta_t, \tau_n \leq t < t_{n+1}, n = 1, 2, \ldots. \tag{11.66}$$

Denote the cdf and pdf of η_t respectively by $F_t(x)$ and $f_t(x)$ $(= \frac{d}{dt}F_t(x), x > 0,$ wherever the derivative exists).

11.7.2 Sample Path of Total Output η_t

A sample path of the process $\{\eta_t\}_{t\geq 0}$ consists of segments that decay exponentially with decay constant k, between the τ_is, which are instants when arrivals are detected (Fig. 11.3). That is,

$$\eta_t = \sum_{i=1}^{n} X_i e^{-k(t-\tau_i)}, \tau_n \leq t < t_{n+1}, n = 1, 2, \ldots \tag{11.67}$$

Note that a sample path cannot descend to level 0 due to exponential decay.

Probability that the Counter Is Free at Time t
Let $p(t) := P(\textit{counter is free to detect a new arriving pulse at time } t \geq 0)$. Then

$$p(t) = \begin{cases} e^{-\lambda t}, 0 < t < D, \\ e^{-\lambda D}, t \geq D. \end{cases} \tag{11.68}$$

The reason for (11.68) is that for $0 < t < D$, the counter is free at t iff there is no arrival in $(0, t)$, which has probability $e^{-\lambda t}$. For $t \geq D$, the counter is free at time t iff there has not been an arrival during the interval $(t - D, t)$. The probability of this event is $e^{-\lambda D}$, by the memoryless property of Exp_λ (see, e.g., pp. 179–181 in [99]).

11.7.3 Integro-differential Equation for PDF of Output

Consider level $x > 0$ in the state space, and state-space set $\boldsymbol{A}_x := (0, x]$. Similarly as in the theorems on down- and upcrossings of level x in Sect. 6.2.8, and also in Theorem 6.3 in Sect. 6.2.9, we infer that for SP *entrances* into set $\boldsymbol{A}_x$ (all entrances are downcrossings of level x)

$$\frac{\partial}{\partial t} E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = kxf_t(x), t > 0. \tag{11.69}$$

For SP *exits* out of $\boldsymbol{A}_x$ (all exits are upcrossings of level x), using (11.68),

$$\begin{aligned} &\tfrac{\partial}{\partial t} E(\mathcal{U}_t(x)) = \tfrac{\partial}{\partial t} E(\mathcal{O}_t(A_x)) \\ &= \begin{cases} \lambda e^{-\lambda t} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy, x > 0, 0 < t < D, \\ \lambda e^{-\lambda D} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy, x > 0, t \geq D. \end{cases} \end{aligned} \tag{11.70}$$

Substituting (11.69) and (11.70) into Theorem B (i.e., Theorem 4.1 in Sect. 4.2.1) , and noting that $\frac{\partial}{\partial t} F_t(x) = -\frac{\partial}{\partial t}(1 - F_t(x))$, we get the integro-differential equations for the pdf $f_t(x)$,

$$kxf_t(x) = \lambda e^{-\lambda t} \cdot \int_{y=0}^{x} \overline{B}(x-y)f_t(y)dy - \frac{\partial}{\partial t}(1 - F_t(x)),$$
$$x > 0, 0 < t < D, \qquad (11.71)$$

$$kxf_t(x) = \lambda e^{-\lambda D} \cdot \int_{y=0}^{x} \overline{B}(x-y)f_t(y)dy - \frac{\partial}{\partial t}(1 - F_t(x)),$$
$$x > 0, t \geq D, \qquad (11.72)$$

where the arrival rate is λ, and a time-t arrival is registered at time t iff the counter is unlocked (free).

11.7.4 Expected Value of Total Output

We obtain the expected value of η_t by integrating both sides of (11.71) and (11.72) with respect to $x \in (0, \infty)$. In (11.71) and (11.72), we assume that $\frac{\partial}{\partial t}F_t(x)$ is continuous with respect to $t > 0$, which is required to apply theorems on interchanging the operations $\int_{x=0}^{\infty}$ and $\frac{\partial}{\partial t}$ (see, e.g., the dominated convergence theorem, Fubini's theorem, and related comments on p. 274 in [116]; p. 111 in [74]; p. 269 in [127]; p.273 in [6]).

Upon integrating (11.71) with respect to x, for t on $(0, D)$ we obtain

$$kE(\eta_t) = \lambda e^{-\lambda t}E(X) - \frac{\partial}{\partial t}E(\eta_t),$$
$$\frac{\partial}{\partial t}e^{kt}E(\eta_t) = \lambda e^{(k-\lambda)t}E(X),$$
$$E(\eta_t) = \frac{\lambda e^{-\lambda t}E(X)}{k-\lambda} + Ae^{-kt}, 0 < t < D, (A \text{ constant}),$$
$$E(\eta_t) = \frac{\lambda E(X)}{k-\lambda}\left(e^{-\lambda t} - e^{-kt}\right), 0 < t < D, \qquad (11.73)$$

since $E(\eta_0) = 0$ by assumption.

Integrating (11.72) with respect to x, for t on (D, ∞) we obtain

$$\begin{aligned}
kE(\eta_t) &= \lambda e^{-\lambda D} E(X) - \frac{\partial}{\partial t} E(\eta_t), \\
\frac{\partial}{\partial t} e^{kt} E(\eta_t) &= \lambda e^{-\lambda D} E(X) e^{kt}, \\
E(\eta_t) &= \frac{\lambda e^{-\lambda D} E(X)}{k} + A e^{-kt}, t \geq D, \qquad (11.74)
\end{aligned}$$

where the constant A is given by

$$A = \lambda E(X) \left(\frac{e^{-(\lambda - k)D} - 1}{k - \lambda} - \frac{e^{-(\lambda - k)D}}{k} \right).$$

To evaluate A, we have used continuity of η_t at $t = D$, i.e., $\eta_{D^-} = \eta_D$ (see Fig. 11.3), which implies continuity of $E(\eta_t)$ at $t = D$ $(a.s.)$. Thus, from (11.73),

$$E(\eta_D) = \frac{\lambda E(X)}{k - \lambda} \left(e^{-\lambda D} - e^{-kD} \right).$$

If $t \to \infty$, then (11.74) reduces to

$$\lim_{t \to \infty} E(\eta_t) = \frac{\lambda e^{-\lambda D} E(X)}{k}.$$

If $D = 0$, then $A = -\frac{\lambda E(X)}{k}$. We then obtain $E(\eta_t) = \frac{\lambda E(X)}{k} \left(1 - e^{-kt}\right)$ and $\lim_{t \to \infty} E(\eta_t) = \frac{\lambda E(X)}{k}$, as on p. 131 in [99].

11.7.5 Type-1 Counter

A type-1 counter differs from the type-2 counter analyzed in Sect. 11.7.1, only in the locking mechanism (see, e.g., pp. 177–179 in [99]). In a type-1 counter, only *registered* (detected) arrivals when the counter is free, generate locked periods. Arrivals when the counter is locked, do not effect the locked period. Thus every locked period has length $D > 0$. Aside from the locking mechanism, we generally use the same notation and assumptions for type-1 and type-2 counters. Thus equations (11.64)–(11.67) hold for type-1 counters.

11.7.6 Sample Path of Total Output

A sample path of the total-output process $\{\eta_t\}_{t\geq 0}$ consists of segments that decay exponentially with decay constant k, between successive detection times τ_n, n =1, 2,... (Fig. 11.4).

Probability that the Counter Is Free at Time t
The probability that the counter is free to register a newly arriving pulse at time t is given by the following recursion ([91]).

$$
\begin{aligned}
p_1(t) &= e^{-\lambda t}, 0 < t < D,\\
p_2(t) &= e^{-\lambda(t-D)} p_1(D) + \frac{(\lambda(t-D))\, e^{-\lambda(t-D)}}{1!}, D \leq t < 2D,\\
&\cdots\\
p_n(t) &= \sum_{j=1}^{n-1} \frac{(\lambda(t-(n-1)D))^{j-1} \cdot e^{-\lambda(t-(n-1)D)}}{(j-1)!} p_{n-j}((n-j)D)\\
&\quad + \frac{(\lambda(t-(n-1)D))^{n-1}\, e^{-\lambda(t-(n-1)D)}}{(n-1)!},\\
&\qquad (n-1)\, D \leq t < nD, n = 1, 2, \ldots, \qquad (11.75)
\end{aligned}
$$

where $\sum_{j=1}^{0} \equiv 0$.

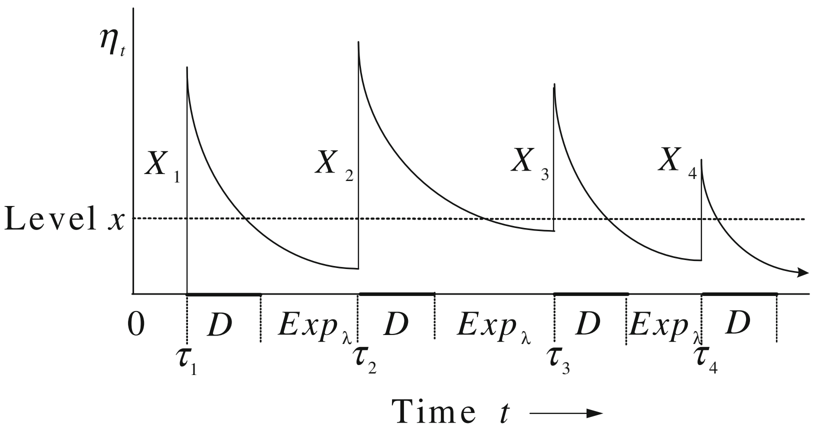


Fig. 11.4 Sample path of total output η_t for type-1 counter model. Locked periods are each $= D$. Undetected arrivals do not effect a locked period. Pulses arrive at Poisson rate λ

Remark 11.6 The successive time intervals (**free**, **locked**) with mean lengths $1/\lambda$ and D, respectively, form an alternating renewal process (see time axis in Fig. 11.4). Let $p(t) = P$(the counter is free at time t), $t \geq 0$. Then $\lim_{t\to\infty} p(t)$ = $(1/\lambda)$ / $(1/\lambda + D)$ (a known result for alternating renewal processes–e.g., pp. 84–86 in [66]). Hence we have proved using probability arguments that

$$\lim_{n\to\infty} p_n(nD) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + D},$$

where $\{p_n(nD)\}_{n=1,2,...}$ is the series obtained by substituting $t = nD$ in (11.75). Another way of stating the limiting time-t result is: for every $\alpha \in [0, 1]$, the same limit holds for any convex combination of the time points $(n-1)D$ and nD, i.e.,

$$\lim_{n\to\infty} p_n(\alpha(n-1)D + (1-\alpha)nD) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + D}.$$

11.7.7 Integro-differential Equation for PDF of Output

Consider level $x > 0$ in the state space; and state-space set $\boldsymbol{A}_x = (0, x]$. We can show as in Sect. 6.2.9, that for SP entrances into set $\boldsymbol{A}_x$,

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial}{\partial t}E(\mathcal{D}_t(x)) = kxf_t(x), \, t > 0. \tag{11.76}$$

For SP exits out of $\boldsymbol{A}_x$,

$$\frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A}_x)) = \frac{\partial}{\partial t}E(\mathcal{U}_t(x)) = \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y)f_t(y)dy,$$
$$(n-1)\,D \leq t < nD, n = 1, 2, ... \tag{11.77}$$

In (11.77), the factor $p_n(t)$ occurs because an arrival is registered if it arrives when the counter is free.

Substituting (11.76) and (11.77) into Theorem B (Theorem 4.1 in Sect. 4.2.1) , we get an integro-differential equation for the pdf $f_t(x)$,

$$kxf_t(x) = \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y)f_t(y)dy + \frac{\partial}{\partial t}F_t(x), x > 0,$$
$$kxf_t(x) = \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y)f_t(y)dy - \frac{\partial}{\partial t}(1 - F_t(x)), x > 0,$$
$$(n-1)\,D \le t < nD, n = 1, 2, \dots . \tag{11.78}$$

11.7.8 Expected Value of Total Output

We obtain the expected value of η_t by dividing both sides by k and then integrating both sides of (11.78) for $x \in (0, \infty)$, yielding

$$E(\eta_t) = \frac{\lambda E(X)}{k-\lambda}\left(e^{-\lambda t} - e^{-kt}\right), 0 < t < D \tag{11.79}$$

in the same manner as for (11.73). Similarly, we can obtain $E(\eta_t)$, $nD \le t < (n+1)\,D$, $n = 1, 2, \dots$. (We do not carry out the latter computation here.)

Remark 11.7 If the locked period has value $D = 0$, then $p_n(t) = 1, n = 1, 2, \dots$. Then every arrival is registered. We then obtain the known result $E(\eta_t) = \frac{\lambda E(X)}{k}\left(1 - e^{-kt}\right), t > 0$ (e.g., p. 131 in [99]).

When $D = 0$, if $t \to \infty$, then (11.79) reduces to $\lim_{t\to\infty} E(\eta_t) = \frac{\lambda E(X)}{k}$.

Remark 11.8 When there is no locked time ($D = 0$), the foregoing type-1 and type-2 counter models coincide with an M/G/r(·) dam with efflux rate proportional to content. Thus, results for a dam with $r(x) = kx$, $x > 0$, can be derived as a special case of either counter model. (See Sect. 6.4 for a related analysis in the M/M/r(·) dam where $r(x) = kx$.)

11.8 Dam with Alternating Influx and Efflux

Consider a dam in which the content alternates between random periods of continuous influx and continuous efflux, when nonempty. We arbitrarily classify periods of emptiness as being parts of periods of efflux, for notational convenience. Periods of *influx* (inflow) are $\underset{dis}{=} \text{Exp}_{\lambda_1}$ and periods of *efflux* (outflow) are $\underset{dis}{=} \text{Exp}_{\lambda_2}$. Let $W(t) \ge 0$ denote the content of the dam at time $t \ge 0$. Assume that during an influx period, the rate of *increase* of content is $dW(t)/dt = q(W(t))$, where $q(x) > 0, x > 0$. Assume that during an efflux period, the rate of *decrease* of content is $dW(t)/dt = -r(W(t))$,

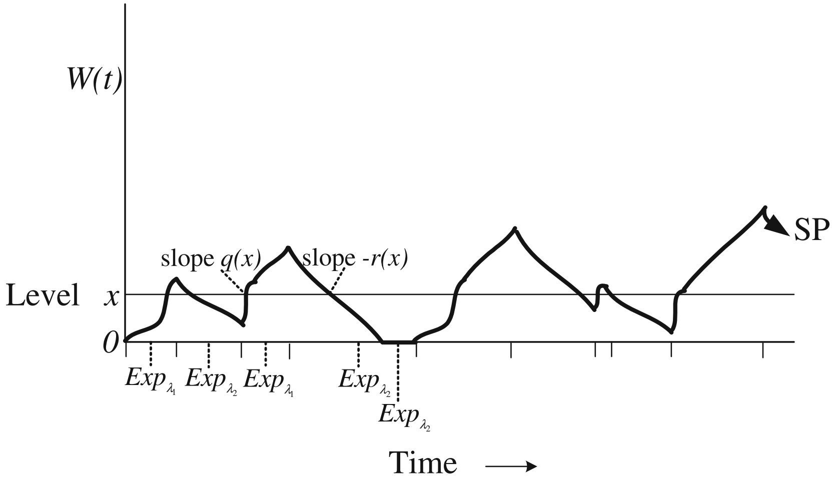


Fig. 11.5 Sample path of dam with alternating periods of continuous influx and efflux. Slope at level x: during influx is $\frac{d}{dt}W(t) = q(x)$; during efflux is $-r(x)$. Slope at level 0 is $\frac{d}{dt}W(t) = 0$. Influx times are $\underset{dis}{=}$ Exp_{λ_1}, efflux and empty times are $\underset{dis}{=}$ Exp_{λ_2} (memoryless property)

where $r(x) > 0$, $x > 0$. In addition, we assume that $r(0^+) > 0$, i.e., there exists $m > 0$ such that $\lim_{x\downarrow 0} r(x) = m$, which guarantees that the dam will reach emptiness. Whenever the dam is empty (i.e., $W(t) = 0$), $dW(t)/dt = 0$. (By contrast, in the model of a dam in Sect. 6.4, where $r(0^+) = 0$ so that emptiness is never achieved theoretically. This is also the case in the pharmaceutical kinetics model in Sect. 11.6.1.). By the memoryless property of Exp_{λ_2}, sojourns at level 0 are also distributed as Exp_{λ_2} (see Fig. 11.5). The empty period is analogous to an idle period in an M/G/1 queue, or empty period in an M/G/r($\cdot$) dam (Sect. 6.2). The efflux rate $r(x)$ is similar to that of the M/G/r($\cdot$) dam. We also assume, in the present model, that the influx rate is $q(x) > 0$, $x \geq 0$, so that $q(0^+) = q(0) > 0$.

11.8.1 Analysis of the Dam Using Method of Sheets

Consider the stochastic process $\{W(t), M(t)\}_{t\geq 0}$ where $W(t)$ denotes the content at instant t, and the system configuration is $M(t) \in \boldsymbol{M} = \{0, 1, 2\}$. The state space is $\boldsymbol{S} = [0, \infty) \times \boldsymbol{M}$. (See Sects. 4.4–4.5 for discussions on system configuration.) The meaning of $M(t)$ is given in the following table.

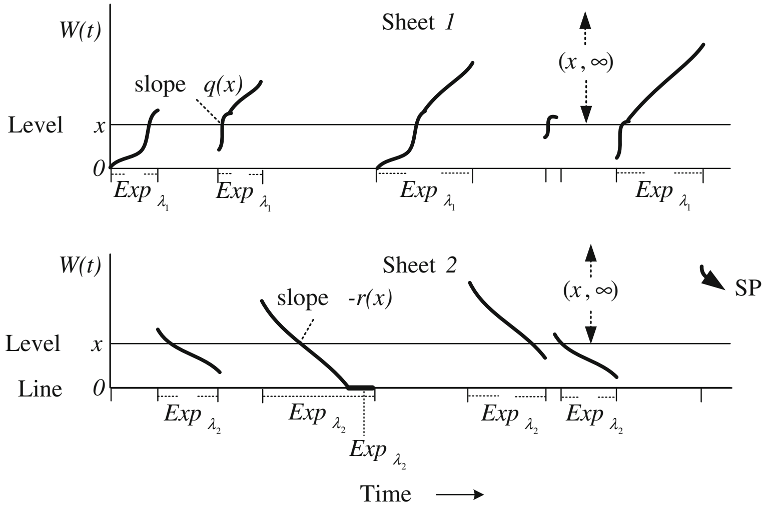


Fig. 11.6 Sample path of dam with continuous influx and efflux, showing line 0 and 2 sheets (pages). Sheet 1 $\leftrightarrow M(t) = 1$, influx phase. Sheet 2 $\leftrightarrow M(t) = 2$, efflux phase. Line 0 $\leftrightarrow W(t) = 0$, empty phase, bottom of Sheet 2. Also indicates composite states $\langle(x, \infty), i\rangle, i = 1, 2$

M(t)	Meaning
0	Empty period.
1	Influx phase: content is increasing.
2	Efflux phase: content is decreasing or at level 0.

A sample path of $\{W(t), M(t)\}_{t\geq 0}$ evolves over two sheets (i.e., *pages*) corresponding to system configurations 1 and 2, and on one line corresponding to an empty period ($W(t) = 0$) (Fig. 11.6).

11.8.2 Steady-State PDF of Content

Denote the 'partial cdfs' of content by

$$F_i(x) = \lim_{t\to\infty} P(W(t) \leq x, M(t) = i), x > 0, i = 1, 2.$$

Denote the steady-state 'partial' pdf of content by

$$f_i(x) = \frac{d}{dx}F_i(x), i = 1, 2, x > 0,$$

wherever the derivative exists.

The *total* pdf of content (marginal pdf) is

$$f(x) = f_1(x) + f_2(x), x > 0. \tag{11.80}$$

Let $P_0 = \lim_{t\to\infty} P(W(t) = 0)$. We shall derive: $f_i(x), i = 1, 2; f(x); P_0$; $F(x) = P_0 + \int_{y=0}^{x} f(y)dy$, in terms of the input parameters λ_1, λ_2, $q(\cdot)$, $r(\cdot)$. The steady-state probability that the dam is in the influx phase ($i = 1$) or efflux phase ($i = 2$) is $F_i(\infty) = \int_{x=0^-}^{\infty} f_i(x)dx, i = 1, 2$.

11.8.3 Equations for PDFs

Consider composite state $((x,\infty), 1)$, $x > 0$, on sheet 1. The SP rate *out* of $((x,\infty), 1)$ is $\lambda_1 \int_{y=x}^{\infty} f_1(y)dy$, since the end of an influx period signals an instantaneous SP *page* 1 → *page* 2 transition from $((x,\infty), 1)$ to $((x,\infty), 2)$ ***at the same level***.

The SP rate *into* $((x,\infty), 1)$ is

$$q(x)f_1(x) + \lambda_2 \int_{y=x}^{\infty} f_2(y)dy,$$

since: (1) the SP upcrosses level x on sheet 1 at rate $q(x)f_1(x)$; (2) the SP enters $((x,\infty), 1)$ from $((x,\infty), 2)$ (*page* 2 → *page* 1 transition) at the same level (the rate at which efflux periods end when the SP is in $((x,\infty), 2)$ is λ_2). Set balance, namely

$$\textbf{\textit{SP rate out of}}\ ((x,\infty), 1) = \textbf{\textit{SP rate into}}\ ((x,\infty), 1),$$

gives an integral equation relating $f_1(x)$ and $f_2(x)$,

$$\lambda_1 \int_{y=x}^{\infty} f_1(y)dy = q(x)f_1(x) + \lambda_2 \int_{y=x}^{\infty} f_2(y)dy. \tag{11.81}$$

Similarly, balancing SP rates out of, and into $((x,\infty), 2)$, $x > 0$, on sheet 2 yields the integral equation

$$\lambda_2 \int_{y=x}^{\infty} f_2(y)dy + r(x)f_2(x) = \lambda_1 \int_{y=x}^{\infty} f_1(y)dy. \tag{11.82}$$

In (11.82), the left and right sides are the SP exit and entrance rates respectively, of $((x,\infty), 2)$.

Addition of (11.81) and (11.82) yields

$$q(x) \cdot f_1(x) = r(x) \cdot f_2(x). \tag{11.83}$$

There is an easy alternative derivation of (11.83), which follows by viewing the sample-path via the "cover". That is, we *project* the segments of the sample path from sheets 1 and 2 (pages 1 and 2) onto a single t-$W(t)$ coordinate system (Fig. 11.5). Then we apply SP rate balance across level x:

$$\textbf{\textit{total upcrossing rate}} = \textbf{\textit{total downcrossing rate}},$$

which translates to formula (11.83).

Using (11.83), we substitute $f_2(x) = (q(x)/r(x)) f_1(x)$ into (11.81), and take d/dx in (11.81). Then we solve the resulting differential equation, applying the initial condition

$$r(0^+) f_2(0) = \lambda_2 P_0 = q(0^+) f_1(0).$$

These operations result in the formula

$$f_1(x) = \frac{\lambda_2 P_0}{q(x)} \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)}, x > 0. \tag{11.84}$$

Since $f_2(x) = (q(x)/r(x)) f_1(x)$, we have

$$f_2(x) = \frac{\lambda_2 P_0}{r(x)} \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)}, x > 0. \tag{11.85}$$

Since $f(x) = f_1(x) + f_2(x)$, adding (11.84) and (11.85) gives

$$\begin{aligned} f(x) &= \lambda_2 \left(\frac{1}{q(x)} + \frac{1}{r(x)} \right) P_0 \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)}, x > 0, \\ &= \lambda_2 \frac{q(x) + r(x)}{q(x) r(x)} \cdot P_0 \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)}, x > 0. \end{aligned} \tag{11.86}$$

The normalizing condition is

$$P_0 + \int_{x=0}^{\infty} f(x) dx = 1. \tag{11.87}$$

From (11.86) and (11.87)

$$P_0 = \frac{1}{1 + \lambda_2 \int_{x=0}^{\infty} \left(\frac{q(x)+r(x)}{q(x)r(x)} \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)} \right) dx}. \quad (11.88)$$

Remark 11.9 Formulas (11.84)–(11.88) are asymmetric with respect to λ_1 and λ_2, because empty periods are distributed as Exp_{λ_2} (classified as part of efflux phase).

Remark 11.10 The model can be generalized in various ways. There may be several different important state-space levels at which there is no change in content (no influx or efflux), other than at level 0. Such levels may be due to a control policy or due to natural phenomena. There would then be **more than one atom** in the state space. Also, the influx and efflux periods may have more general distributions. The content may be bounded above, resulting in an atom. Some of these variants are easy to analyze; others are more complex. We do not treat such variants here.

Stability Condition

A necessary condition for the pdf to exist is $f(\infty) = 0$. Thus, the exponent $\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)$ in (11.86) and (11.88) must be positive for all $\dot{x} > 0$. That is

$$\lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy < \lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy,$$

$$\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy > 0, \text{ for all } x > 0. \quad (11.89)$$

Remark 11.11 In (11.89), if $q(y) = r(y)$, $y > 0$, then the stability condition reduces to $1/\lambda_1 < 1/\lambda_2$ or E(nonempty period) $< E$(empty period). If $q(y) = q$ and $r(y) = r$, $y > 0$, then $1/\lambda_1 < (r/q)\,(1/\lambda_2)$ or E(nonempty period) $< (r/q)\,E$(empty period). Additionally, If $r/q < 1$ then E(nonempty period) $< E$(empty period).

11.8.4 Numerical Example

Let $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$, $x > 0$. Substituting into (11.89) gives, for $x > 0$,

$$\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy = 2\sqrt{x}\left(\lambda_1 - \frac{\lambda_2}{3}\right) = 2\sqrt{x}\left(1 - \frac{2}{3}\right) > 0,$$

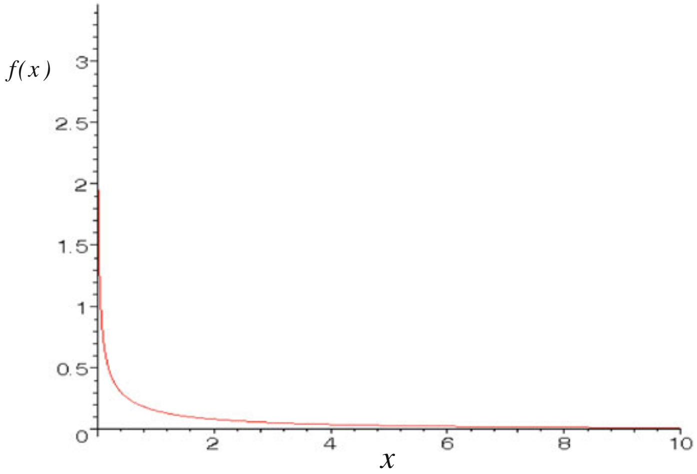


Fig. 11.7 Steady-state pdf $f(x) = \frac{8}{27\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}, x > 0$, in continuous dam with alternating influx/efflux periods: $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$

implying stability, and the steady-state pdf $f(x)$ exists. From (11.86), we obtain

$$f(x) = \frac{8}{3\sqrt{x}}P_0 \cdot e^{-\frac{2}{3}\sqrt{x}}, x > 0. \tag{11.90}$$

From the normalizing condition (11.87),

$$P_0 = \frac{1}{1 + \int_{x=0}^{\infty} \frac{8}{3\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}dx} = \frac{1}{9} = 0.111111. \tag{11.91}$$

Thus

$$f(x) = \frac{8}{27\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}, x > 0. \tag{11.92}$$

From (11.91) and (11.92), the cdf is (see Figs. 11.7, 11.8),

$$F(x) = P_0 + \int_{y=0}^{x} f(y)dy = 1 - \frac{8}{9}e^{-\frac{2}{3}\sqrt{x}}. \tag{11.93}$$

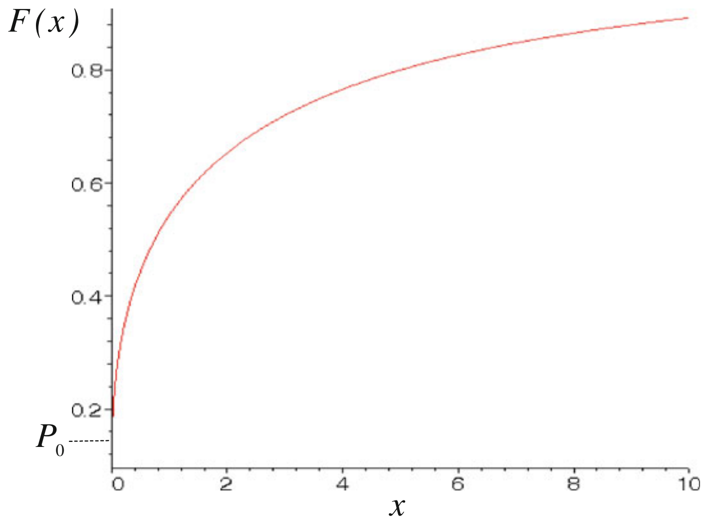


Fig. 11.8 Steady-state cdf $F(x) = 1 - \frac{8}{9}e^{-\frac{2}{3}\sqrt{x}}$, $x > 0$, $P_0 = 0.1111$, in continuous dam with alternating influx/efflux periods: $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$

Proportion of Time in Influx and Efflux Phases
From (11.83) and (11.80) we obtain

$$f_1(x) = \frac{2}{9\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}, \quad x > 0,$$
$$f_2(x) = \frac{2}{27\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}, \quad x > 0.$$

Hence the proportions of time the dam is in the influx, efflux phase respectively are

$$F_1(\infty) = \int_{x=0}^{\infty} \frac{2}{9\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}dx = 0.666667,$$
$$F_2(\infty) - P_0 = \int_{x=0}^{\infty} \frac{2}{27\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}dx = 0.222222.$$

These values are also the steady-state probabilities of the dam being in these phases at an arbitrary time point. A check on the normalizing condition is

$$P_0 + F_1(\infty) + F_2(\infty) = 0.111111 + 0.666667 + 0.222222 = 1.$$

11.9 Estimation of Laplace Transforms

We very briefly discuss a procedure for estimating the LST (Laplace-Stieltjes transform) of the state variable of a stochastic model. We shall use the virtual wait in a GI/G/1 queue as an example.

Suppose we want to estimate the LST of the steady-state pdf of the virtual wait in a GI/G/1 queue. Let the steady-state cdf of the virtual wait be $F(x)$, $x \geq 0$, having pdf $f(x)$, $x > 0$, and let $P_0 = F(0)$. The LST of the mixed pdf $\{P_0, f(x)\}_{x>0}$ is (see Sect. 3.4.4 in Chap. 3)

$$F^*(s) = \int_{x=0}^{\infty} e^{-sx} dF(x), s > 0. \tag{11.94}$$

11.9.1 Probabilistic Interpretation of LST

The probabilistic interpretation of the LST is as follows (see p. 264 and pp. 267–269 in [104]; and various papers, e.g., [41]). In formula (11.94), the right side is the probability that an independent "*catastrophe random variable*" $\underset{dis}{=}$ Exp_s is greater than the virtual wait having cdf $F(x)$, $x \geq 0$.

11.9.2 Estimation of LST

In order to estimate $F^*(s)$, we simulate a sample path of the virtual wait $\{W(u)\}_{u\geq 0}$, over a long period of simulated time $(0, t)$. Next, we *generate a sample path of a renewal process* $\{\mathcal{C}(u)\}_{u\geq 0}$ with inter-renewal times equal to the catastrophe r.v., and overlay it on the *same time-state coordinate system* (see Fig. 11.9). Fix $s > 0$. The SP jump sizes and inter-renewal times in the sample path of $\{\mathcal{C}(u)\}_{u\geq 0}$, are i.i.d. r.v.s $\underset{dis}{=}$ Exp_s. This is because the process $\{\mathcal{C}(u)\}_{u\geq 0}$ represents the excess life γ at time u (see Sect. 10.1.5 and Fig. 10.2). The steady-state pdf of the excess life is $f_\gamma(x) = se^{-sx}$, $x > 0$.

Now we observe the sample paths of $\{W(u)\}_{u\geq 0}$ and $\{\mathcal{C}(u)\}_{u\geq 0}$ on the fixed time interval $(0, t)$. We compute the **sum**, $T_s = \sum_i T_{si}$, of all time intervals such that $\mathcal{C}(u) > W(u)$, $u \in (0, t)$ (Fig. 11.9). An estimate of $F^*(s)$ is then $\widehat{F^*}(s) = T_s/t$, which is the proportion of time such that $\mathcal{C}(u) > W(u)$, $u \in (0, t)$. The probabilistic interpretation of the LST strongly suggests that T_s/t is an appropriate estimate of $F^*(s)$.

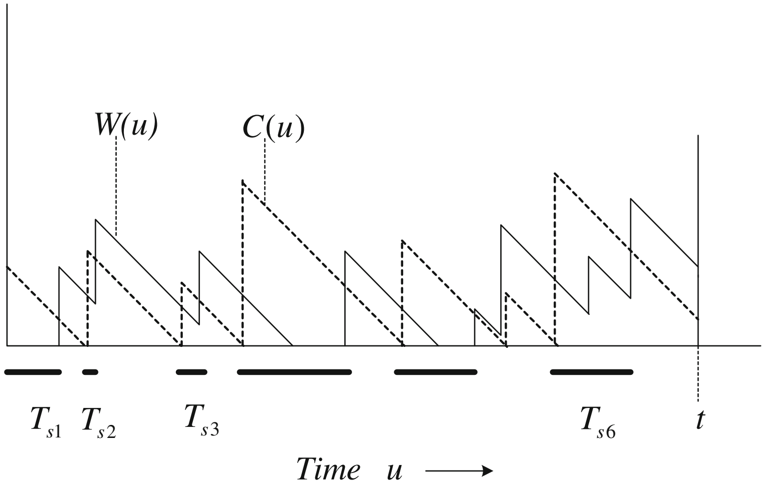


Fig. 11.9 Sample paths of virtual wait $\{W(u)\}_{u\geq 0}$ and renewal process with inter-arrival time $\underset{dis}{=}$ Exp$_s$, the catastrophe r.v., $\{\mathcal{C}(u)\}_{u\geq 0}$. $T_s = T_{s1} + T_{s2} + \cdots + T_{s6}$.

In order to estimate $F^*(s)$, $s > 0$, we repeat the simulation procedure using different values of $s > 0$. For example, we may choose a partition of N uniformly-spaced values for $s \in (0, t)$, such as $s = \Delta, 2\Delta, 3\Delta, ..., N\Delta$, where N is a large positive integer and Δ is a small positive number. (Different spacing for the partition may improve the estimates, e.g., if $F(\cdot)$ is known to have certain properties such as a long tail.) This procedure results in a set of estimates $\widehat{F^*}(n\Delta) = T_{n\Delta}/t$, $n = 1, ..., N$. (From (11.94), $\widehat{F^*}(0) = 1$, which is the normalizing condition.)

Finally, we can plot the points

$$\left(0, \widehat{F^*}(0)\right) = (0, 1) \text{ and } \left(n\Delta, \widehat{F^*}(n\Delta)\right), n = 1, ..., N,$$

on a two-dimensional $\left(s, \widehat{F^*}(s)\right)$ coordinate system. The $\{n\Delta\}_{n=1,...,N}$ grid is on the horizontal axis; the corresponding $\widehat{F^*}(n\Delta)$ terms are ordinates along vertical lines parallel to the $\widehat{F^*}(s)$-axis.

The plot will be a discrete estimate of the LST of the pdf of the virtual wait. It may be improved by smoothing techniques. In order to obtain an estimate of the pdf of the virtual wait from it, use numerical inversion of $\left\{\widehat{F^*}(n\Delta)\right\}_{n=1,...,N}$ (see, e.g., pp. 349–355 in [104]).

11.10 Simple Harmonic Motion

We analyze an elementary model of *deterministic* simple harmonic motion, using LC.

Consider a particle moving according to simple harmonic motion (SHM) (see, e.g., p. 133 in [10]; pp. 216–217 in [119]). Let $X(t)$ denote the position of the particle at instant $t \geq 0$, and $X(0) = 0$. Let the state space be the interval $\boldsymbol{S} = [-1, +1]$. In this version of the standard SHM model there is only one sample path, namely,

$$X(t) = \sin(t), t \geq 0.$$

We wish to determine the stationary pdf $f(x)$ and cdf $F(x)$ of $X(t)$ when the particle is observed at an arbitrary time point, as $t \to \infty$.

Consider the sample path $X(t)$, $t \geq 0$ (Fig. 11.10). The slope of the sample path at level x is

$$r(x) == \frac{d}{dt} \sin t|_{t=\sin^{-1} x} = \cos\left(\sin^{-1} x\right) = \sqrt{1 - x^2}, x \in [-1, +1]. \tag{11.95}$$

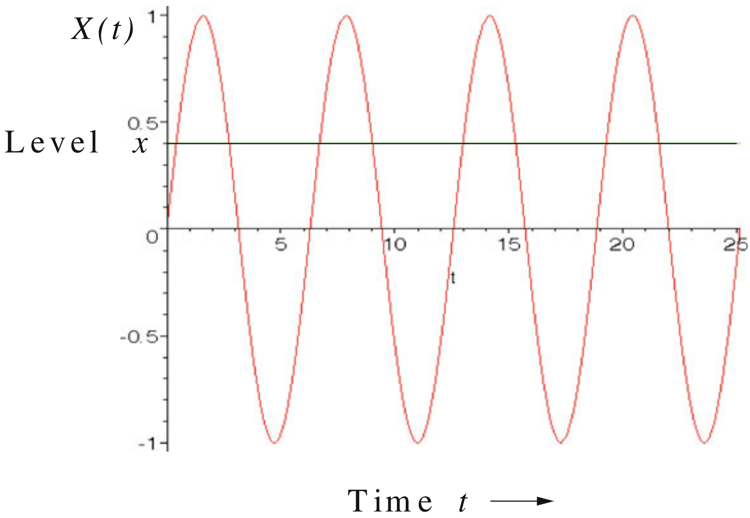


Fig. 11.10 Sample path of simple harmonic motion $X(t) = \sin t$. State space is $\boldsymbol{S} = [-1, +1]$. Shows level x in $\boldsymbol{S}$

Consider levels $x, x+h \in \boldsymbol{S}$, where $h > 0$ is small. The time required for the SP to ascend from level x to level $x+h$ is

$$\int_{y=x}^{x+h} \frac{1}{r(y)} dy = \int_{y=x}^{x+h} \frac{1}{\sqrt{1-y^2}} dy. \tag{11.96}$$

The symmetries of the sample path imply that the time required for the SP to descend from level $x+h$ to level x is also given by (11.96).

Applying (11.96), we see that the long-run *proportion of time* the SP spends in state-space interval $(x, x+h)$ in a cycle of length 2π time units is

$$\frac{2}{2\pi} \int_{y=x}^{x+h} \frac{1}{\sqrt{1-y^2}} dy = F(x+h) - F(x). \tag{11.97}$$

Formula (11.97) leads to

$$\frac{1}{\pi} h \frac{1}{\sqrt{1-(x^*)^2}} = F(x+h) - F(x) \tag{11.98}$$

where $x^* \in (x, x+h)$, by the definition of $F(x)$ as the long-run proportion of time the process is in state-space interval $[-1, x]$. Dividing both sides of (11.98) by h and letting $h \downarrow 0$, yields

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}}, x \in [-1, +1]\,. \tag{11.99}$$

The stationary pdf $f(x)$ in (11.99) is interesting and suggests intuitive insights (Fig. 11.11). Note that $\lim_{x \downarrow -1} f(x) = \lim_{x \uparrow +1} f(x) = \infty$. Also, $\min_{x \in \boldsymbol{S}} f(x) = \frac{1}{\pi}$, at $x = 0$. The pdf $f(x)$ is symmetric about $x = 0$, and is convex.

From (11.99), the cdf is

$$\begin{aligned} F(x) &= \int_{y=-1}^{x} f(y) dy, = \frac{1}{\pi} \left(\sin^{-1}(x) - \sin^{-1}(-1)\right) \\ &= \frac{1}{\pi} \sin^{-1}(x) + \frac{1}{2}, x \in [-1, +1]. \end{aligned} \tag{11.100}$$

11.10.1 Inferences Based on PDF and CDF

From (11.95), the speed of the particle is $r(x) = \sqrt{1-x^2}$, implying $r(\pm 1) = 0$ and $r(0) = 1$, the maximum speed. Hence, at an arbitrary time point in

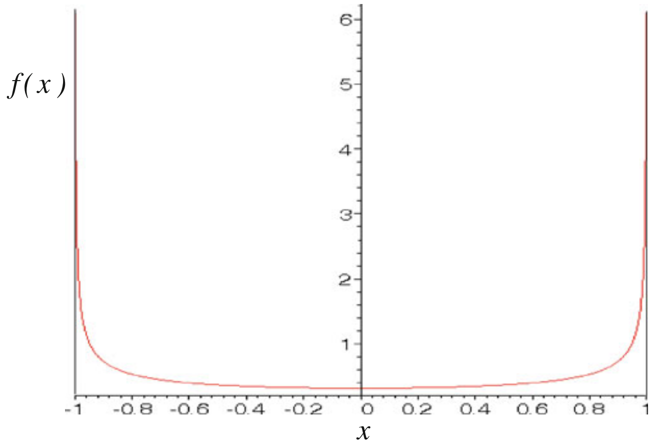


Fig. 11.11 Stationary pdf $f(x) = \frac{1}{\pi\sqrt{1-x^2}}, x \in [-1, +1]$, for particle moving in simple harmonic motion, $X(t) = \sin t, t \geq 0$

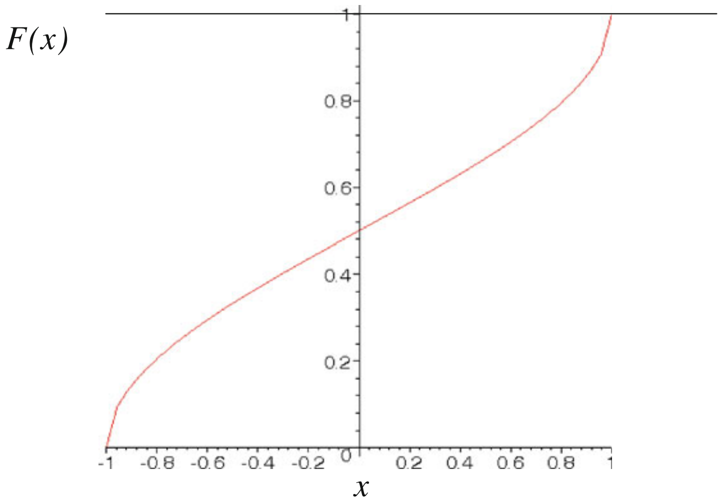


Fig. 11.12 Stationary cdf $F(x) = \frac{1}{\pi}\sin^{-1}(x) + \frac{1}{2}, x \in [-1, +1]$, for particle moving in simple harmonic motion, $X(t) = \sin t, t \geq 0$

the long run, it is much more likely to find the particle close to one of the boundaries of $\boldsymbol{S}$ ($x = \pm 1$), rather than close to the center of $\boldsymbol{S}$ (its range of motion). This fact suggests that the particle spends a much greater proportion of time near the boundaries $x = \pm 1$ than near the center $x = 0$ (Fig. 11.12).

From computations using formula (11.100), the proportion of time the SP (particle) spends in the central interval $[-0.5, +0.5]$ is equal to $F(0.5) -$

$F(-0.5) = 0.333$. The proportion of time the particle spends in the outer regions $[-1.0, -0.5] \cup [0.5, 1.0]$, is equal to $2 \cdot (F(1.0) - F(0.5)) = 0.667$. The "median" symmetric outer intervals of $\boldsymbol{S}$ with respect to the time spent by the particle, is $\boldsymbol{A}_{0.5} \equiv [-1.0, -0.707] \cup [0.707, 1.0]$, i.e., $P(\text{particle} \in \boldsymbol{A}_{0.5})$ = 0.5. This indicates that it is equally likely to find the particle in two bands of equal width 0.293 touching the edges ± 1.0 (total width 0.586), as it is to find it in a central interval of width 1.414 about 0. Arbitrary observations on operating pendulum clocks, readily corroborate these theoretical computations.

Remark 11.12 The type of LC analysis in this section, may be extended to analyze random trigonometric functions (e.g., like $A \sin(\theta t) + B \cos(\theta t)$, $t \geq 0$, where A, B are random variables and θ is a constant). Extensions may also be applicable in some models of physics, and in the analysis of roots of equations.

References

[1] Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions*. New York: Dover Publications Inc.

[2] Ancker, C. J., & Gafarian, A. V. (1961). Queueing with multiple inputs and exponential service times. *Operations Research*, *9*(3), 321–327.

[3] Azoury, K., & Brill, P. H. (1986). An application of the system-point method to inventory models under continuous review. *Journal of Applied Probability*, *23*(3), 778–789.

[4] Azoury, K., & Brill, P. H. (1992). Analysis of net inventory in continuous review models with random lead time. *European Journal of Operational Research*, *39*, 383–392.

[5] Baccelli, F., & Brémaud, P. (1991). *Elements of queueing theory*. New York: Springer.

[6] Bartle, R. G. (1976). *The elements of real analysis* (2nd ed.). New York: Wiley.

[7] Bartlett, M. S. (1978). *An introduction to stochastic processes* (3rd ed.). Cambridge University Press.

[8] Beneš, V. E. (1957). On queues with poisson arrivals. *Annals of Mathematical Statistics*, *28*, 670–677.

[9] Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics*. Oakland, California: Holden-Day Inc.

[10] Boyce, W. E., & DiPrima, R. C. (1969). *Elementary differential equations and boundary value problems* (2nd ed.). New York: Wiley.

[11] Brill, P. H. (1975). *System point theory in exponential queues*. Ph.D. Dissertation, University of Toronto. Available from University Microfilms International, Ann Arbor, Michigan, order number 8901129.

[12] Brill, P. H. (1976). A new methodology for modelling a broad class of exponential queues. *Advances in Applied Probability*, *8*(2), 242. (Abstract of presentation at the *Fifth Conference on Stochastic Processes and their Applications*, University of Maryland, College Park, MD, USA, June, 1975. Invited by J. Keilson (external examiner for the author's Ph.D. thesis) and R. Syski. First Conference Presentation on the Level Crossing Methodology.)

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5

[13] Brill, P. H. (1976). *Embedded level crossing processes in dams and queues*. Working #76-022, University of Toronto, Department of Industrial Engineering.

[14] Brill P. H. (1977). The system point process in queueing. *Advances in Applied Probability*, *9*(2), 216. Abstract of presentation at the *Sixth Conference on Stochastic Processes and their Applications*, Tel Aviv, Israel, June, 1976.

[15] Brill, P. H. (1979). An embedded level crossing technique for dams and queues. *Journal of Applied Probability*, *16*, 174–186.

[16] Brill, P. H. (1983a). *Queues with reneging depending on required wait*. Technical report STAT-83-07, Department of Statistics and Actuarial Science, University of Waterloo.

[17] Brill, P. H. (1983b). *System point Monte Carlo simulation of stationary distributions of waiting times in single server queues*. STAT 83–10, Department of Statistics and Actuarial Science, University of Waterloo.

[18] Brill, P. H. (1987). *System point computation in queues, dams and inventories*. Working paper No. W87-12, ISSN No. 0716-6301, Faculty of Business Administration, University of Windsor, Canada.

[19] Brill, P. H. (1988). Single-server queues with delay-dependent arrival streams. *Probability in the Engineering and Informational Sciences*, *2*, 231–247.

[20] Brill, P. H. (1988). *Waiting time in queues with state dependent bulk service*. Working Paper No. W88-02, ISSN No. 0714-6191, Faculty of Business Administration, University of Windsor, Canada.

[21] Brill, P. H. (1988). *A technique for transient distributions in stochastic models*. Working Paper W-88-14, ISSN No. 0714-6191, Faculty of Business Administration, University of Windsor.

[22] Brill, P. H. (1988). *The time dependent system point level crossing method for exponential queues*. Working Paper W-88-15, Faculty of Business Administration, University of Windsor, ISSN No. 0714-6191.

[23] Brill, P. H. (1988). *The time dependent system point level crossing method for queues, dams and inventories*. Working Paper W-88-16, Faculty of Business Administration, University of Windsor, ISSN No. 0714-6191.

[24] Brill, P. H. (1990). *Level crossing estimation of stationary distributions in storage processes* (Revision of Brill, P. H. (1987) *System point computation in queues, dams and inventories*. Working paper No. W87-12, University of Windsor, Canada, 64 pages.)

[25] Brill, P. H. (2015). Note on the service time in an M/G/1 queue with bounded workload. *Statistics and Probability Letters*, *96*, 162–169.

[26] Brill, P. H. (1990). Example of level crossing estimation in M/G/1 queues. In *Proceedings of the American Statistical Association, Statistical Computing Section, Anaheim, California* (pp. 151–154).

[27] Brill, P. H. (1991). Estimation of stationary distributions in storage processes using level crossing theory. In *Proceedings of the American Statistical Association, Atlanta, Georgia, Statistical Computing Section* (pp. 172–177).

[28] Brill, P. H. (1992). *A note on the age, excess life and total life in a renewal process* (p. 1992). June-July: Unpublished notes.

[29] Brill, P. H. (1994). *Level crossing estimation of stationary distributions in storage processes*. Paper Presented at CORS-94 Conference, Montreal, Canada, May 31.

[30] Brill, P. H. (1996). Level crossing methods. In S. I. Gass & C. M. Harris (Eds.), *Encyclopedia of operations research and management science* (pp. 338–340). Kluwer Academic Publishers.

[31] Brill, P. H. (2000). A brief outline of the level crossing method in stochastic models (Engish and French). CORS (Canadian Operational Research Society). *Bulletin*, *34*(4), 9–21.

[32] Brill, P. H. (2009). Note on a series for M/G/1 queues. *International Journal of Operational Research*, *5*(3), 363–373.

[33] Brill, P. H. (2009). Compound cycle of a renewal process and applications. *INFOR*, *47*(4), 273–281.

[34] Brill, P. H. (2014). Alternative analysis of finite-time probability distributions of renewal theory. *Probability in the Engineering and Informational Sciences*, *28*(2), 183–201.

[35] Brill, P. H. (2015). Note on the service time in an M/G/1 queue with bounded workload. *Statistics and Probability Letters*, *96*, 162–169.

[36] Brill, P. H., Huang, M. L., & Hlynka, M. (2009). Note on an $\langle s, S\rangle$ inventory system with decay. *IAENG International Journal of Applied Mathematics*, *39*(3), 171–174.

[37] Brill, P. H., & Chaouch, B. A. (1995). An EOQ model with random variation in demand. *Management Science*, *41*(5), 927–936.

[38] Brill, P. H., & Green, L. (1984). Queues in which customers receive simultaneous service from a random number of servers: A system point approach. *Management Science*, *30*(1), 51–68.

[39] Brill, P. H., & Harris, C. M. (1997). M/G/1 queues with Markov-generated server vacations. *Stochastic Models*, *13*(3), 491–521.

[40] Brill, P. H., & Hornik, J. (1984). A system point approach to nonuniform adversising insertions. *Operations Research*, *32*(1), 7–22.

[41] Brill, P. H., & Hlynka, M. (2000). An exponential queue with competition for service. *European Journal of Operational Research*, *126*, 587–602.

[42] Brill, P. H., & Hlynka, M. (2012). Server workload in an M/M/1 queue with bulk arrivals and special delays. *Applied Mathematics*, *3*(12A), 2174–2177 (http://www.SciRP.org/journal/am).

[43] Brill, P. H., Huang, M. L., & Hlynka, M. (2009). Note on an $\langle s, S\rangle$ inventory system with decay. *IAENG International Journal of Applied Mathematics*, *39*(3), 171–174.

[44] Brill, P. H., & Huang, M. L. (1993). System point estimation of the probability distribution of the waiting time in variations of M/G^B/1 queues. In *Proceedings of the American Statistical Association, San Francisco, California, Statistical Computing Section* (pp. 236–241).

[45] Brill, P. H., & Mandelbaum, M. (1989). On measures of flexibility in manufacturing systems. *International Journal of Production Research*, *27*(5), 747–756.

[46] Brill, P. H., & Mandelbaum, M. (1990). Measurement of adaptivity and flexibility in production systems. *European Journal of Operational Research*, *49*(3), 325–332.

[47] Brill, P. H., & Moon, R. E. (1980). Application of queueing theory to pharmacokinetics. *Journal of Pharmaceutical Sciences*, *89*(5), 558–560.

[48] Brill, P. H., & Posner, M. J. M. (1974a). *Two server queues with service time depending on waiting time* (pp. 1–61). Working Paper WP74-005, Department of Industrial Engineering, University of Toronto (uses Lindley recursions and an embedded Markov chain analysis).

[49] Brill, P. H., & Posner, M. J. M. (1974b). *A multiple server queue with service time depending on waiting time* (pp. 1–30). Working paper WP74-008, Department of Industrial Engineering, University of Toronto (uses Lindley recursions and an embedded Markov chain analysis).

[50] Brill, P. H., & Posner, M. J. M. (1974c). *On the equilibrium distribution for a class of exponential queues*. Working Paper WP74-012, Department of Industrial Engineering, University of Toronto (first paper using level crossing method).

[51] Brill, P. H., & Posner, M. J. M. (1977). Level crossings in point processes applied to queues: Single server case. *Operations Research*, *25*(4), 662–673.

[52] Brill, P. H., & Posner, M. J. M. (1981). The system point method in exponential queues: A level crossing approach. *Mathematics of Operations Research*, *6*(1), 31–49.

[53] Brill, P. H., & Posner, M. J. M. (1981). A two-server queue with non-waiting customers receiving specialized service. *Management Science*, *27*(8), 914–925.

[54] Brill, P. H., & Yu, K. (2011). Analysis of risk models using a level crossing technique. *Insurance: Mathematics and Economics*, *49*(3), 298–309.

[55] Brockmeyer, E., Halstrom, H. L., & Jensen, A. (1948). The life and works of A. K. Erlang. *Transactions of the Danish Academy of Technical Science*, *2*, 1–277.

[56] Burrill, C. W., & Knudsen, J. R. (1869). *Real variables, holt*. New York: Reinhart & Winston Inc.

[57] Callahan, J .R. (1971). *The nothing hot delay problem in the production of steel*. Ph.D. thesis, Department of Industrial Engineering, University of Toronto.

[58] Callahan, J. R. (1973). A queue with waiting time dependent service times. *Naval Research Logistics Quarterly*, *20*, 321–324.

[59] Chen, P. (1997). *Some topics on Markov chains and their applications*. Ph.D. dissertation, Colorado State University.

[60] Çinlar, E. (1975). *Introduction to stochastic processes*. Englewood Cliffs, N.J.: Prentice-Hall.

[61] Cohen, J. W. (1976). On regenerative processes in queuing theory. In M. Beckman & H. P. Kunzi (Eds.), *Lecture notes in economics and mathematical systems*. New York: Springer.

[62] Cohen, J. W. (1977). On up- and downcrossings. *Journal of Applied Probability*, *14*, 405–410.

[63] Cohen, J. W. (1982). *The single server queue* (Revised ed.). New York: North-Holland Publishing Co.

[64] Cooper, R. B. (1981). *Introduction to queuing theory* (2nd ed.). New York: North Holland.

[65] Cooper, R. B., & Niu, S.-C. (1986). Beneš's formula for M/G/1-FIFO 'explained' by pre-emptive resume LIFO. *Journal of Applied Probability*, *23*, 550–554.

[66] Cox, D. R. (1962). *Renewal theory*. London: Methuen & Co.

[67] Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Methuen.

[68] Cox, D. R., & Smith, W. L. (1967). *Queues*. London: Methuen and Co., Ltd.

[69] Cramer, H., & Leadbetter, M. R. (1967). *Stationary and related stochastic processes–Sample function properties and their applications*. Mineola, N.Y: Wiley, Dover Publications.

[70] Doshi, B. T. (1986). Queueing systems with vacations. *Queueing Systems: Theory and Applications*, *1*(1), 29–66.

[71] Embrechts, P., Klüppelberg, C., & Mikosch, T. (2003). *Modelling extremal events for insurance and finance*. Berlin, New York: Springer.

[72] Erlang, A. K. (1909). The theory of probabilities and telephone conversations. In E. Brockmeyer, H. L. Halstrøm, & A. Jensen (Eds.). (1948). The life and works of A.K. Erlang. *Transactions of the Danish Academy of Technical Sciences*, *2*, 131–137.

[73] Feller, W. (1950). *An introduction to probability theory and its applications* (Vol. I). New York: Wiley.

[74] Feller, W. (1966). *An introduction to probability theory and its applications* (Vol. II). New York: Wiley.

[75] Forbes, C., Evans M., Hastings, N., & Peacock, B. (2011). *Statistical distributions* (4th ed.): Wiley.

[76] Franx, G. J. (2001). A simple solution for the M/D/c waiting time distribution. *Operations Research Letters*, *29*(5), 221–229.

[77] Gaver, D. P., & Miller, R. G. (1962). Limiting distributions for some storage problems. In K. J. Arrow, S. Karlin, & H. Scarf (Eds.), *Studies in applied probability and management science* (pp. 110–126). Stanford Calif: Stanford University Press.

[78] Gavish, E., & Schweitzer, P. (1977). The Markovian queue with bounded waiting time. *Management Science*, *23*(12), 1349–1357.

[79] Gerber, H. U., & Shiu, E. S. W. (1997). The joint distribution of the time of ruin, the surplus immediately before ruin, and the deficit at ruin. *Insurance: Mathematics and Economics*, *21*, 129–137.

[80] Grassmann, W. K. (Ed.). (2000). *Computational probability*. Boston: Kluwer Academic Publishers.

[81] Green, L. (1978). *Queues which allow a random number of servers per customer*. Ph.D. dissertation, Yale University.

[82] Green, L. (1980). A queueing system in which customers require a random number of servers. *Operations Research*, *28*(6), 1335–1346.

[83] Gross, D., & Harris, C. M. (1998). *Fundamentals of queuing theory* (3rd ed.). New York: Wiley.
[84] Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2008). *Fundamentals of queueing theory* (4th ed.). New York: Wiley.
[85] Harris, C. M. (1966). *Queues with state dependent stochastic service rates*. Ph.D. dissertation, Polytechnic Institute of Brooklyn.
[86] Harris, C. M. (1967). Queues with state dependent stochastic service rates. *Operations Research*, *15*(1), 117–130.
[87] Harris, C. M., Brill, P. H., & Fischer, M. (2000). Internet-type queues with power-tailed interarrival times and computational methods for their analysis. *Informs Journal on Computing*, *12*(4), 261–271.
[88] Heyman, D. P., & Sobel, M. (1982). *Stochastic models in operations research, Volume I, Stochastic processes and operating characteristics*. New York: McGraw-Hill.
[89] Hlynka, M. (2007). *Myron Hlynka's queueing theory page*. http://www2.uwindsor.ca/~hlynka/queue.html.
[90] Hlynka, M., & Brill, P. H. (2007). *A note on stability in M/M/1 queues with reneging*. Technical report WMSR 07–09, Department of Mathematics and Statistics, University of Windsor.
[91] Hlynka, M., & Brill, P. H. (2008). *A result for a counter problem.* Technical report WMSR 08–01, Department of Mathematics and Statistics, University of Windsor.
[92] Hlynka, M., Brill, P. H., & Horn, W. (2010). A method for obtaining laplace transforms of order statistics of Erlang random variables. *Statistics and Probability Letters*, *80*(1), 9–18.
[93] Huang, M. L., & Brill P. H. (1993). System point estimation of the probability distribution of the waiting time in variations of M/G^B/1 queues. In *Proceedings of the Statistical Computing Section, American Statistical Association, San Francisco, August 8–12*.
[94] Huang, M. L., & Brill, P. H. (1999). A level crossing quantile estimation method. *Statistics and Probability Letters*, *45*, 111–119.
[95] Jaeger, J. C., & Newstead, G. (1968). *An introduction to the Laplace transform with engineering applications*. Methuen & Co. and Science Paperbacks.
[96] Jameson, G. J. O. (1970). *A first course in complex functions*. London: Chapman and Hall.
[97] Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics, continuous univariate distributions–1*. New York: Wiley.
[98] Huang, M. L., & Brill, P. H. (2004). A level crossing distribution estimation method. *Journal of Statistical Planning and Inference*, *124*(1), 45–62.
[99] Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes* (2nd ed.). New York: Academic Press.
[100] Katayamna, T. (2012). A note on M/G/1 vacation ststems with sojourn time limits. *Journal of Applied Probability*, *49*(4), 1194–1199.
[101] Keilson, J., & Servi, L. D. (1986). Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules. *Journal of Applied Probability*, *23*, 790–802.

[102] Keilson, J. (1965). *Green's function methods in probability theory*. London: Charles Griffin and Co., Ltd.

[103] Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of embedded markov chains. *Annals of Mathematical Statistics*, *24*(3), 338–354.

[104] Kleinrock, L. (1975) *Queueing systems volume I: Theory*. New York: Wiley.

[105] Leadbetter, M. R. (1972). Point processes generated by level crossings. In P. A. W. Lewis (Ed.), *Stochastic point processes: Statistical analysis, theory and applications* (pp. 436–467). New York: Wiley-Interscience.

[106] LeCorbeiller, P. (1966). *Dimensional analysis, appleton-century-crofts*. New York: Division of Meredith Publishing Co.

[107] Lehmann, E. L. (1991). *Theory of point estimation*. Belmont, California: Wadsworth Inc.

[108] Libura, M. (1971). On a one-channel queueing system with service time depending on waiting time. *Archiwum Automatuca Telemechaniki*, *10*, 279–286.

[109] Lindley, D. V. (1952). The theory of queues with a single server. *Proceedings of the Cambridge Philosophical Society*, *48*, 277–289.

[110] Little, J. D. C. (1961). A proof for the queuing formula L = λW. *Operations Research*, *9*(3), 383–387. doi:10.1287/opre.9.3.383.JSTOR167570.

[111] Lovitt, W. V. (1950). *Integral equations*. New York: Dover Publications Inc.

[112] Lucantoni, D. M., Meier-Hellstern, K. S., & Neuts, M. F. (1990). A single-server queue with server vacations and a class of nonrenewal arrival processes. *Advances in Applied Probability*, *22*(3), 676–705.

[113] Miyazawa, M. (1994). Rate conservation laws: A survey. *Queueing Systems*, *15*(1–4), 1–58.

[114] Nahmias, S. (2008). *Production and operations analysis* (6th ed.). New York: McGraw-Hill/Irwin.

[115] Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models. An algorithmic approach*. Baltimore: The John Hopkins university Press.

[116] Parzen, E. (1962). *Stochastic processes*. San Francisco: Holden-Day.

[117] Posner, M. J. (1973). Single server queues with service time depending on waiting time. *Operations Research*, *21*(2), 610–616.

[118] Prabhu, N. U. (1980). *Stochastic storage processes*. New York: Springer.

[119] Rainville, E. D., & Bedient, P. E. (1969). *A short course in differential equations* (4th ed.). New York: The Macmillan Co.

[120] Rangel, A. G. (1983). *Monte Carlo simulation for computing time-dependent and equilibrium probability densities of waiting times in M/G/1 queues*. Master's project: Department of Management Science, University of Waterloo.

[121] Reichenbach, H. (1956). *The direction of time*. New York: Dover Publications Inc.

[122] Resnick, S. (2005). *Adventures in stochastic processes* (4th ed.). Boston: Birkhauser.

[123] Ross, S. M. (1970). *Applied probability with optimization applications*. San Francisco: Holden Day.

[124] Ross, S. M. (1983). *Stochastic processes*. New York: Wiley.

[125] Ross, S. M. (2010). *Introduction to probability models* (10th ed.). Amsterdam: Elsevier.

[126] Roy, K. A. (1997). *Laplace transforms, probabilities and queues*. M.Sc. thesis, electronic theses and dissertations. Paper 2572. http://scholar.uwindsor.ca/etd/2572.

[127] Royden, H. L. (1968). *Real analysis* (2nd ed.). New York: MacMillan Publishing Co., Inc.

[128] Shortle, J. F., & Brill, P. H. (2005). Analytical distribution of waiting time in the M/iD/1 queue. *Queueing Systems*, *50*(2), 185–197.

[129] Shortle, J. F., Brill, P. H., Fischer, M. J., Gross, D., & Masi, D. M. B. (2004). An algorithm to compute the waiting time distribution for the M/G/1 queue. *INFORMS Journal on Computing*, *16*(2), 152–161.

[130] Shortle, J. F., Fischer, M., & Brill, P. H. (2007). Waiting-time distribution of $M/D_N/1$ queues through numerical Laplace inversion. *INFORMS Journal on Computing*, *19*(1), 112–120.

[131] Sigman, K. (1999). A primer on heavy-tailed distributions. *Queueing Systems*, *33*, 261–275.

[132] Sigman, K., & Wolff, R. W. (1993). A review of regenerative processes. *SIAM Review*, *35*(2), 269–288.

[133] Silverman, B. W. (1990). *Density estimation in statistics and data analysis*. New York: Chapman and Hall.

[134] Smith, W. L. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society A*, *232*, 6–31.

[135] Smith, W. L. (1958). Renewal theory and its ramifications. *Journal of the Royal Statistical Society, Series B* (Methodological), *20*(2), 243–302.

[136] Spivak, M. (1965). *Calculus on manifolds*. New York: W. A. Benjamin Inc.

[137] Spivak, M. (1967). *Calculus*. New York: W. A. Benjamin Inc.

[138] Srivastava, H. M., & Kashyap, B. R. K. (1982). *Special functions in queueing theory*. New York: Academic Press.

[139] Takács, L. (1955). Investigation of waiting time problems by reduction to Markov processes. *Acta Mathematica Academy of Science Hungary*, *6*, 101–129.

[140] Takács, L. (1962). *Introduction to the theory of queues*. New York: Oxford University Press.

[141] Takács, L. (1967). *Combinatorial methods in the theory of stochastic processes*. New York: Wiley.

[142] Taylor, H. M., & Karlin, S. (1998). *An introduction to stochastic modelling* (3rd ed.). San Diego: Academic Press.

[143] Tijms, H. C. (2003). *A first course in stochastic models*. Wiley.

[144] Welch, P. D. (1964). On a generalized M/G/1 queueing process in which the first customer of each busy period receives exceptional service. *Operations Research*, *12*(5), 736–752.

[145] Wolff, R. W. (1982). Poisson arrivals see time averages. *Operations Research*, *30*(2), 223–231.

Partial Bibliography

The foregoing *References* cite publications that the author referred to in some way when working on this monograph. The *Partial Bibliography* cites two types of other publications. One type consists of publications that, in some way, use the level crossing theory and methods elucidated in this monograph (SPLC). The second type consists of publications that discuss other theoretical or applied aspects of level crossings, or contain models and ideas that potentially can be analyzed using SPLC. Since there are very extensive literatures of both types, only a relatively small sample of each type is cited below.

Abramov, V. M. (2001). Some results for large closed queueing networks with and without bottleneck: Up- and down-crossings approach. *Queueing Systems*, *38*(2), 149–184.

Adler, R., & Samorodnitsky, G. (1997). Level crossings of absolutely continuous stationary symmetric “alpha”-stable processes. The *Annals of Applied Probability*, *7*(2), 460–493.

Adler, R. J. (1981). *The geometry of random fields*. New York: Wiley.

Afèche, P., Diamant, A., & Milner, J. (2014). Double-sided batch queues with abandonment: Modeling crossing networks. *Operations Research*, *62*(5), 1179–1201.

Ashour, O. M., & Okudan, G. E. (2011, January). Developing a correction factor to modify the FAHP and utility theory based triage algorithm. In *Proceedings of the IIE Annual Conference* (p. 1). Institute of Industrial Engineers-Publisher.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5

Ayesta, U., Jacko, P., & Novak, V. (2011, April). A nearly-optimal index rule for scheduling of users with abandonment. In *2011 Proceedings IEEE Conference on INFOCOM* (pp. 2849–2857). IEEE.

Azoury, K. S., Miyaoka, J., & Udayabhanu, V. (2012). Steady state analysis of continuous review inventory systems: A level crossing approach. *Journal of Supply Chain and Operations Management*, *10*(1), 66–86.

Bae, J., Kim, S., & Lee, E. Y. (2001). The virtual waiting time of the M/G/1 queue with impatient customers. *Queueing Systems*, *38*(4), 485–494.

Barakat, R. (1978). The distribution of values of trigonometric sums with linearly independent frequencies: Kac's problem revisited. *Stochastic Processes and their Applications*, *8*(1), 77–85.

Barron, Y. (2016). An (S, K, S) Fluid inventory model with exponential leadtimes and order cancellations. *Stochastic Models*, *32*(2), 301–332.

Beekman, J. A., & Fuelling, C. P. (1990). Interest and mortality randomness in some annuities. *Insurance: Mathematics & Economics*, *9*(2–3), 185–196.

Bekker, R. (2005). Finite-buffer queues with workload-dependent service and arrival rates. *Queueing Systems: Theory and Applications*, *50*(2–3), 231–253.

Bekker, R., Borst, S. C., Boxma, O. J., & Kella, O. (2004). Queues with workload-dependent arrival and service rates. *Queueing Systems: Theory and Applications*, *46*(3–4), 537–556.

Belyaev, Y. K., & Seleznjev, O. (2000). Approaching in distribution with applications to resampling of stochastic processes. *Scandinavian Journal of Statistics*, *27*(2), 371–384.

Berg, M., Posner, M. J. M., & Zhao, H. (1994). Production-inventory systems with unreliable machines. *Operations Research*, *42*(1), 111–118.

Berman, O., Parlar, M., Perry, D., & Posner, M. J. M. (2005). Production/clearing models under continuous and sporadic reviews. *Methodology and Computing in Applied Probability*, *7*, 203–224.

Berzin, C., Leon, J. R., & Ortega, J. (1998). Level crossings and local time for regularized Gaussian processes. *Probability and Mathematical Statistics*, *18*(1), 39–81.

Besson, J. L. (1983). Mean number of the level crossings of stochastic processes with absolutely continuous sample paths. *Comptes Rendus de l'Académie des Sciences, Série I: Mathématique*, *297*, 635–637.

Borovkov, K., & Last, G. (2008). On level crossings for a general class of piecewise-deterministic Markov processes. *Advances in Applied Probability*, 815–834.

Boucherie, R. J., & Boxma, O. J. (1996). The workload in the M/G/1 queue with work removal. *Probability in the Engineering and Informational Sciences*, *10*(02), 261–277.

Boxma, O., Löpker, A., & Perry, D. (2016). On a make-to-tock production/mountain model with hysteretic control. *Annals of Operations Research*, *241*(1–2), 53–82.

Boxma, O. J., David, I., Perry, D., & Stadje, W. (2011). A new look at organ transplantation models and double matching queues. *Probability in the Engineering and Informational Sciences*, *25*(02), 135–155.

Brémaud, P. (1993). A Swiss army formula of Palm calculus. *Journal of Applied Probability*, 40–51.

Chaouch, B. A. (2001). Stock levels and delivery rates in vendor-managed inventory programs. *Production and Operations Management*, *10*(1), 31–44.

Chaouch, B. A. (2011). A replenishment control system with uncertain returns and random opportunities for disposal. *Intrenational Journal of Inventory Research*, *1*(3–4), 221–247.

Chiu, Y. P. (2003). Determining the optimal lot size for the finite production model with random defective rate, the rework process, and backlogging. *Engineering Optimization*, *35*(4), 427–437.

Cooper, R. B. (1990). Queueing theory. *Handbooks in Operations Research and Management Science*, *2*, 469–518.

De Boer, Pieter-Tjerk, Nicola, V. F., & Van Ommeren, J.-K. C. W. (2001). The remaining service time upon reaching a high level in M/G/1 queues. *Queueing Systems: Theory and Applications*, *39*(1), 55–78.

Dickson, D. C. M. (2005). *Insurance risk and ruin*. Cambridge University Press.

Doshi, B. (1992). Level-crossing analysis of queues in queueing and related models. In U. N. Bhat, & I. V. Basawa (Eds.) (pp. 3–33). New York: Oxford University Press.

Doshi, B. (1992). Level-crossing analysis of queues. *Oxford Statistical Science Series*, 3–3.

Doshi, B., & Heffes, H. (1986). Overload performance of several processor queueing disciplines for the M/M/1 queue. *IEEE Transactions on Communications*, *34*(6), 538–546.

Doshi, B. T. (1986). Queueing systems with vacations—A survey. *Queueing Systems*, *1*(1), 29–66.

Drake, M. J., & Marley, K. A. (2014). A century of the EOQ. In *Handbook of EOQ inventory problems* (pp. 3–22). Springer.

Dshalalow, J. H. (1995). *Advances in queueing: Theory, methods, and open problems*. CRC Press.

El-Taha, M., & Stidham Jr., S. (2012). *Sample-path analysis of queueing systems* (Vol. 11). Springer Science & Business Media.

Fajardo, V. A., & Drekic, S. (2015). Controlling the workload of M/G/1 queues via the Q-policy. *European Journal of Operational Research*, *243*(2), 607–617.

Fajardo, V. A., & Drekic, S. (2015). Waiting time distributions in the preemptive accumulating priority queue. *Methodology and Computing in Applied Probability*, 1–30.

Fajardo, V. A., & Drekic, S. (2016). On a general mixed priority queue with server discretion. *Stochastic Models*, 1–31.

Farahmand, K. (1990). On the average number of level crossings of a random trigonometric polynomial. *The Annals of Probability*, *18*(3), 1403–1409.

Farahmand, K., & Shaposhnikov, A. (2005). On the expected number of level crossings of random trigonometric polynomials. *Stochastic Analysis and Applications*, *23*(6), 1141–1147.

Feuerverger, A., Hall, P., & Wood, A. T. A. (1994). Estimation of fractal index and fractal dimension of a Gaussian process by counting the number of level crossings. *Journal of Time Series Analysis*, *15*(6), 587–606.

Fondjo Fotou, F., Fujisaki, K., & Tateiba, M. (2008). Empirical analysis and modeling of combined long-term and short-term statistics on satellite links. In *26th International Communications Satellite Systems Conference (ICSSC).*

Graves, S. C., & Keilson, J. (1981). The compensation method applied to a one-product production/inventory problem. *Mathematics of Operation and Research*, *6*(2), 246–262.

Greenberg, I. (1997). Markov chain approximation methods in a class of level-crossing problems. *Operations Research Letters*, *21*(3), 153–158.

Guo, P., Lian, Z., & Wang, Y. (2011). Pricing perishable products with compound poisson demands. *Probability in the Engineering and Informational Sciences*, *25*(03), 289–306.

Guo, P., & Zipkin, P. (2007). Analysis and comparison of queues with different levels of delay information. *Management Science*, *53*(6), 962–970.

He, Q.-M., & Jewkes, E. M. (1997). A level crossing analysis of the MAP/G/1 queue. In S. Chakravarthy, & A. Alfa (Eds.), *Matrix-analytic methods in stochastic models* (pp. 107–116). New York: Marcel Dekker Inc.

Hébuterne, G., Hebuterne, G., & Rosenberg, C. (1999). Arrival and departure state distributions in the general bulk-service queue. *Naval Research Logistics*, *46*(1), 107–118.

Hillier, F. S., & Lieberman, G. J. (2004). *Introduction to operations research* (6th ed.). New York: McGraw Hill.

Huang, M. L., & Brill, P. H. (2004). A distribution estimation method based on level crossings. *Journal of Statistical Planning and Inference*, *124*(1), 45–62.

Huang, M. L., & Yuen, W. K. (2010). A bivariate density estimation method based on level crossings. *Statistics*, *44*(1), 31–55.

Huang, M. L. (2001). On a distribution-free quantile estimator. *Computational Statistics & Data Analysis*, *37*(4), 477–486.

Illsley, R. (1998). The moments of the number of exits from a simply connected region. *Advances in Applied Probability*, *30*(1), 167–180.

Iravani, F., & Balcıoğlu, B. (2008). Approximations for the M/GI/N+ GI type call center. *Queueing Systems*, *58*(2), 137–153.

Iravani, F., & Balcıoğlu, B. (2008). On priority queues with impatient customers. *Queueing Systems*, *58*(4), 239–260.

Jewkes, E. M., & Buzacott, J. A. (1991). Flow time distributions in a K-Class M/G/1 priority feedback queue. *Queueing Systems*, *8*(1), 183–202.

Jewkes, E. M., & Stanford, D. A. (2003). A two priority queue with crossover feedback. *Queueing Systems: Theory and Applications*, *43*(1–2), 129–146.

Jin, M., Roni, M., & Garcia-Diaz, A. (2012, January). A hybrid inventory policy responding to surge demand. In *Proceedings of the IIE Annual Conference Proceedings* (pp. 1–10). Orlando, Florida, USA: Institute of Industrial Engineers-Publisher.

Kac, M. (1943). On the distribution of values of trigonometric sums with linearly independent frequencies. *American Journal of Mathematics*, *65*(4), 609–615.

Kaspi, H., & Perry, D. (1989). On a duality between a non-Markovian storage/production process and a Markovian dam process with state-dependent input and output. *Journal of Applied Probability*, *26*(4), 835–844.

Katayama, T. (2002). A note on level-crossing analysis for the excess, age, and spread distributions. *Journal of Applied Probability*, *39*(4), 896–900.

Katayama, T. (2005). Level-crossing approach to a time-limited service system with two types of vacations. *Operations Research Letters*, *33*(3), 295–300.

Katayama, T. (2007). Analysis of a nonpreemptive priority queue with exponential timer and server vacations. *Performance Evaluation*, *64*(6), 495–506.

Kim, S., & Lee, E. Y. (2002). A level crossing approach to the analysis of finite dam. *Journal of the Korean Statistical Society*, *31*(3), 405–413.

Kolb, M., Stadje, W., & Wübker, A. (2016). The rate of convergence to stationarity for M/G/1 models with admission controls via coupling. *Stochastic Models*, *32*(1), 121–135.

König, D., & Schmidt, V. (1980). Imbedded and Non-imbedded stationary characteristics of queueing systems with varying service rate and point processes. *Journal of Applied Probability*, 753–767.

Kroese, D. P., & Kallenberg, W. C. M. (1992). Second-order asymptotics in level crossing for differences of renewal processes. *Stochastic Processes and their Applications*, *40*(2), 309–323.

Kurose, J. F., & Chipalkatti, R. (1987). Load sharing in soft real-time distributed computer systems. *IEEE Transactions on Computers*, *100*(8), 993–1000.

Larranaga, M., Ayesta, U., & Verloop, I. M. (2013). Dynamic fluid-based scheduling in a multi-class abandonment queue. *Performance Evaluation*, *70*(10), 841–858.

Leadbetter, M. R. (1975). Point processes generated by level crossings in random processes. In A. Ephremides (Ed.), *II: Poisson & jump-point processes* (pp. 109–140). Stroudsburg, PA: Dowden, Hutchinson & Ross Inc.

Leadbetter, M. R., & Spaniolo, G. V. (2004). Reflections on Rice's formulae for level crossings—History, extensions and use. *Australian & New Zealand Journal of Statistics*, *46*(1), 173–180.

Lee, E. Y. (2008). A new approach to an inventory with constant demand. *Journal of the Korean Data and Information Science Society*, *19*(4), 1345–1352.

Lee, J., & Kim, J. (2006). A workload-dependent M/G/1 queue under a two-stage service policy. *Operations Research Letters*, *34*(5), 531–538.

Lin, T. Y., & Hou, K. L. (2015). An imperfect quality economic order quantity with advanced receiving. *TOP*, *23*(2), 535–551.

Lindgren, G., & Rychlik, I. (1995). How reliable are contour curves? Confidence sets for level contours. *Bernoulli*, *1*(4), 301–319.

Liu, B., & Alfa, A. S. (2002). A fluid model with data message discarding. *Advances in Applied Probability*, 329–348.

Liu, J., Lee, T. T., Jiang, X., & Horiguchi, S. (2009). Blocking and delay analysis of single wavelength optical buffer with general packet size distribution. *Journal of Lightwave Technology*, *27*(8), 955–966.

Liu, L., & Kulkarni, V. G. (2006). Explicit solutions for the steady state distributions in M/PH/1 queues with workload dependent balking. *Queueing Systems*, *52*(4), 251–260.

Löpker, A., & Perry, D. (2010). The idle period of the finite G/M/1 queue with an interpretation in risk theory. *Queueing Systems*, *64*(4), 395–407.

Mandelbaum, M. (1968). *Queueing with splitting and matching*. Masters thesis, Technion, Israel Institute of Technology.

Mandelbaum, M., & Avi-Itzhak, B. (1968). Introduction to queueing with splitting and matching. *Israel Journal of Technology*, *6*, 288–298.

Manuel, U. R., & Offiong, A. (2014). Application of queuing theory to automated teller machine (ATM) facilities using Monte Carlo simulation. *International Journal of Engineering Science and Technology*, *6*(4), 162.

Mau, Y., Feng, X., & Porporato, A. (2014). Multiplicative jump processes and applications to leaching of salt and contaminants in the soil. *Physical Review E*, *90*(5), 052128.

Mazumdar, R., Badrinath, V., Guillemin, F., & Rosenberg, C. (1993). On pathwise rate conservation for a class of semi-martingales. *Stochastic Processes and their Applications*, *47*(1), 119–130.

Meyn, S. P., & Tweedie, R. L. (1993). *Markov chains and stochastic stability*. New York: Springer.

Mohebbi, E. (2004). A replenishment model for the supply-uncertainty problem. *International Journal of Production Economics*, *87*(1), 25–37.

Mohebbi, E., & Hao, D. (2006). When supplier's availability affects the replenishment lead time—An extension of the supply-interruption problem. *European Journal of Operational Research*, *175*(2), 992–1008.

Mohebbi, E., & Hao, D. (2008). An inventory model with non-resuming randomly interruptible lead time. *International Journal of Production Economics*, *114*(2), 755–768.

Mohebbi, E., & Posner, M. J. M. (1998). A continuous-review inventory system with lost sales and variable lead time. *Naval Research Logistics*, *45*(3), 259–278.

Movahed, M. S., & Khosravi, S. (2011). Level crossing analysis of cosmic microwave background radiation: A method for detecting cosmic strings. *Journal of Cosmology and Astroparticle Physics*.

Nahmias, S. (2004). *Production and operations analysis* (4th ed.). Burr Ridge, Illinois: Irwin Professional Publication.

Neuts, M. F. (1989). Structured stochastic matrices of M/G/1 type and their applications. *Probability: Pure and Applied*, *5* (New York: Marcel Dekker, Inc.).

Nezhadhaghighi, M. G., Movahed, S. M. S., Yasseri, T., & Allaei, S. M. (2015). *Crossing statistics of anisotropic stochastic surface*. arXiv:1508.01409.

Posner, M. J. M., & Berg, M. (1989). Analysis of a production-inventory system with unreliable production facility. *Operations Research Letters*, *8*(6), 339–345.

Prabhu, N. U. (1965). *Stochastic processes*. New York: MacMillan.

Prabhu, N. U., & Basawa, I. V. (Eds.). (1990). *Statistical inference in stochastic processes*. New York: Marcel Dekker, Inc.

Rege, K. (1993). On the M/G/1 queue with Bernoulli feedback. *Operations Research Letters*, *14*(3), 163–170.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell System Technical Journal*, *23*, 282–332.

Rice, S. O. (1945). Mathematical analysis of random noise. *Bell System Technical Journal*, *24*, 46–156.

Rolski, T. (2012). *Stationary random processes associated with point processes* (Vol. 5). Springer Science & Business Media.

Roni, M. S., Eksioglu, S. D., Jin, M., & Mamun, S. (2016). A hybrid inventory policy with split delivery under regular and surge demand. *International Journal of Production Economics*, *172*, 126–136.

Roni, M. S., Jin, M., & Eksioglu, S. D. (2015). A hybrid inventory management system responding to regular demand and surge demand. *Omega*, *52*, 190–200.

Rumyantsev, A., & Morozov, E. (2015). Stability criterion of a multiserver model with simultaneous service. *Annals of Operations Research*, 1–11.

Rychlik, I. (2000). On some reliability applications of rice's formula for the intensity of level crossing. *Extremes*, *3*, 331–348.

Salameh, M. K., & Jaber, M. Y. (2000). Economic production quantity model for items with imperfect quality. *International Journal of Production Economics*, *64*(1), 59–64.

Samal, N. K., & Pratihar, D. K. (2014). Optimization of variable demand fuzzy economic order quantity inventory models without and with backordering. *Computers & Industrial Engineering*, *78*, 148–162.

Sarhangian, V., & Balcioğlu, B. (2013). Waiting time analysis of multiclass queues with impatient customers. *Probability in the Engineering and Informational Sciences*, *27*(03), 333–352.

Schagen, I. P. (1980). A stochastic model for the occurrence of oilfields and its application to some North Sea data. *Journal of the Royal Statistical Society: Series C*, *29*(3), 282–291.

Shanthikumar, J. G. (1980). Some analyses on the control of queues using level crossings of regenerative processes. *Journal of Applied Probability*, 814–821.

Shanthikumar, J. G. (1981). On level crossing analysis of queues. *The Australian Journal of Statistics*, *23*(3), 337–342.

Shanthikumar, J. G. (1988). On stochastic decomposition of M/G/1 type queues with generalized server vacations. *Operations Research*, *36*(4), 566–569.

Sigman, K. (1991). A note on a sample-path rate conservation law and its relationship with H = λG. *Advances in Applied Probability*, 662–665.

Stadje, W. (1998). Level-crossing properties of the risk process. *Mathematics of Operations Research*, *23*(3), 576–584.

Stidham, S., Jr. & El Taha, M. (1989). Sample-path analysis of processes with imbedded point processes. *Queueing Systems*, *5*(1–3), 131–165.

Swensen, A. R. (1986). On a GI/M/C queue with bounded waiting times. *Operations Research*, *34*(6), 895–908.

Tan, B., & Gershwin, S. B. (2009). Analysis of a general Markovian two-stage continuous-flow production system with a finite buffer. *International Journal of Production Economics*, *120*(2), 327–339.

Turksen, I. B., & Berg, M. (1991). An expert system prototype for inventory capacity planning: An approximate reasoning approach. *International Journal of Approximate Reasoning*, *5*(3), 223–250.

van Harn, K., & Steutel, F. W. (1992). *A generalization of the life -time pairs of renewal theory*. Amsterdam: Rapportnr WS-395, Vrieje Universiteit.

Wang, L., Tachwali, Y., Verma, P., & Ghosh, A. (2009). Impact of bounded delay on throughput in multi-hop networks. *Journal of Network and Computer Applications*, *32*(5), 1031–1038.

Whitt, W. (1991). A review of L= λW and extensions. *Queueing Systems*, *9*(3), 235–268.

Wolff, R. W. (2011). *Little's law and related results. Wiley encyclopedia of operations research and management science*.

Wu, P., & Posner, M. J. M. (1997). A level-crossing approach to the solution of the shortest-queue problem. *Operations Research Letters*, *21*(4), 181–189.

Xiong, W., Jagerman, D., & Altiok, T. (2008). M/G/1 queue with deterministic reneging times. *Performance Evaluation*, *65*(3), 308–316.

Zazanis, M. A. (1992). Sample path analysis of level crossings for the workload process. *Queueing Systems*, *11*(4), 419–428.

Index

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
International Series in Operations Research & Management Science 250,
DOI 10.1007/978-3-319-50332-5