CHAPTER 3

M/G/1 QUEUES AND VARIANTS

3.1 Introduction

This chapter considers the virtual wait process in M/G/1 queues and model variants. It first develops relationships between sample-path level crossings and the time dependent (transient) distribution of wait. These relationships lead to a proof of the basic LC theorem for the steady-state pdf of wait in M/G/1 queues, including equation (1.8). The relationships are of inherent interest for time-dependent LC methods.

Next, alternative forms of the LC integral equation (1.8) are derived by using LC interpretations. The alternative forms are useful for analyzing certain variants of M/G/1 queues such as those with service times having discrete distributions.

LC analyses of several M/M/1 and M/G/1 models in the steady state are given which illustrate 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, 3.1). Let $\mathcal{D}_t(x)$, $\mathcal{U}_t(x)$ denote the number of down- and upcrossings of level $x \geq 0$ during $(0, t)$, respectively. Note that $\{\mathcal{D}_t(x), t \geq 0\}$ and $\{\mathcal{U}_t(x), t \geq 0\}$ are counting processes.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_3, © Springer Science+Business Media, LLC 2008

3.2.1 Differentiability and Downcrossings of Level x

The following lemma guarantees the existence of $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$, where $E(\mathcal{D}_t(x))$ is the expected value of $\mathcal{D}_t(x)$. 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 hits of 0 from above during $(0,t)$) $= \mathcal{I}_t(0)$ (number of SP entrances into $\{0\}$ during $(0,t)$) (see Subsection 2.4.10).

Lemma 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 $\{\mathcal{D}_t(x)\}$ is a delayed *renewal* process for each $x \geq 0$. 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 at time d_0, let the level-x inter-downcrossing times be $d_1, d_2, ...$ (Fig. 3.1). Let $H_{d_0}(\cdot)$, $h_{d_0}(\cdot)$ denote the cdf and pdf of d_0, respectively. We need only prove the result when $d_0 > 0$. If $d_0 = 0$, the proof is similar.

The following well known basic renewal relationship holds for $n = 1, 2, ...$ and $t > 0$,

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

Thus

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

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

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

where $F_{d_0+d_1+\cdots+d_{n-1}}(t)$ is the cdf of $d_0+d_1+\cdots+d_{n-1}$ and $F_{d_1}^{n-1}(\cdot)$ is the $(n-1)$-fold convolution of d_1. Taking $\frac{\partial}{\partial t}$ on both sides (differentiating under the integral) gives

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

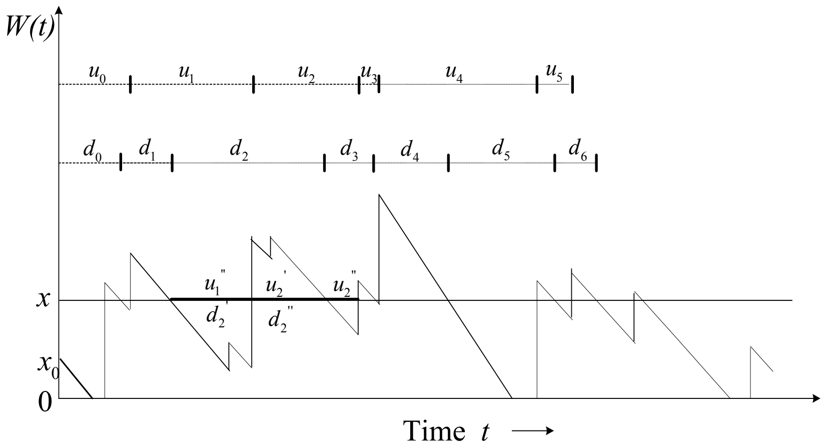

Figure 3.1: Sample path of virtual wait in M/G/1 showing inter down- and upcrossing times for level x, $\{d_n\}$, $\{u_n\}$, and their components, e.g., $d_2^{'}$, $d_2^{''}$, $u_2^{'}$, $u_2^{''}$, etc.

since $F_{d_1}(0) = 0$. The right side exists since $F_{d_1}^{n-1}(t-s)$ is the cdf of an $(n-1)$-fold sum of continuous random variables, each distributed as d_1. That is, $\frac{\partial}{\partial t} F_{d_1}^{n-1}(t-s) = f_{d_1}^{n-1}(t-s)$ exists; it is the pdf of a continuous r.v. Moreover, $\frac{\partial}{\partial t} E(\mathcal{D}_t(x)) > 0$ since both $f_{d_1}^{n-1}(t-s) > 0$, $h_{d_0}(s) > 0$. **Note**: Once existence of $\frac{\partial}{\partial t} E(\mathcal{D}_t(x))$ is established, positivity follows since $E(\mathcal{D}_t(x))$ is an increasing function of t. ■

3.2.2 Differentiability and Upcrossings of Level x

Consider a sample path of the virtual wait. The process $\{\mathcal{U}_t(x)\}$ is a "delayed" process. In general, however, $\{\mathcal{U}_t(x)\}$ is not renewal. The delay u_0, is the time from $t = 0$ to the first upcrossing of x after d_0. The level-x *inter-upcrossing* times starting at u_0 are denoted by $u_1, u_2, ...$ (Fig. 3.1). The random variables $\{u_i,\ i = 1, 2, ...\}$ are identically distributed (with the same distribution d_1). However, $\{u_i\}$ are not mutually independent. Successive pairs (u_i, u_{i+1}) are dependent.

Remark 3.1 *For an arbitrary typical sample path in general, successive pairs u_i, u_{i+1} are dependent. To see this, consider u_1, u_2 (Fig. 3.1). Let $d_i = d_i^{'} + d_i^{''}$, $u_i = u_i^{'} + u_i^{''}$, $i = 1, 2$. Note that $u_2^{'}$ $(= d_2^{''})$ is dependent on $u_1^{''}$ $(= d_2^{'})$, because the excess jump above x, say r_x^a, depends on $u_1^{''}$. If $u_1^{''}$ is small, r_x^a tends to be large. That is, $P(r_x^a > z|$jump starts at*

$y < x) = \frac{\overline{B}(x-y+z)}{\overline{B}(x-y)}$, *which depends on both* x *and* y. *Thus* u_2 *depends on* u_1. *Nevertheless* $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$ *exists (see the following lemma).*

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

Proof. The delay time u_0 is a continuous r.v. The process $\{\mathcal{U}_t(x)\}$ 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 relationship, usually applied for a renewal, process,

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

also holds for a general counting process even though the inter-arrival times are not independent. Thus

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

Summing on both sides over $n = 1, 2, ...$ 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 $u_1 + \cdots + u_{n-1}$. The sum $u_0 + u_1 + \cdots + u_{n-1}$ is a continuous r.v., since u_i is continuous for each $i = 1, 2, ...$ 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 right side is finite. Thus $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$ exists. Also, $\frac{\partial}{\partial t}E(\mathcal{U}_t(x)) > 0$, since $h_{u_0}(s) > 0$ and $f_{u_1+\cdots+u_{n-1}}(t-s) > 0$. Alternatively, positivity follows since $E(\mathcal{U}_t(x))$ is an increasing function of t. ■

The derivatives $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$, $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$ are fundamentally related (Theorem 3.1).

Remark 3.2 *If the service time is exponentially distributed with mean $\frac{1}{\mu}$, as in M/M/1, then for any sample path the excess jump above x, r_x^a, is exponentially distributed by the memoryless property, and*

$$\begin{aligned} P(r_x^a &> z|\textit{jump starts at } y < x) \\ &= \frac{\overline{B}(x-y+z)}{\overline{B}(x-y)} = \frac{e^{-(x-y+z)}}{e^{-(x-y)}} = e^{-\mu z}, \end{aligned}$$

independent of x and y. In that case, $\{u_n\}$ ***is a delayed renewal process.***

3.2.3 Level Crossings and Transient CDF of Wait

Denote the transient distribution of the virtual wait by

$$\begin{aligned} F_t(x) = &\quad P(W(t) \le x), x \ge 0, t \ge 0 \\ P_0(t) = &\quad P(W(t) = 0), F_t(x), t \ge 0, \\ f_t(x) = &\quad \tfrac{\partial}{\partial x} F_t(x), x > 0, t \ge 0, \end{aligned} \tag{3.1}$$

wherever $\frac{\partial}{\partial x} F_t(x)$ exists. Define the joint cdf of $(W(t_1),\, W(t_2))$ as

$$F_{t_1,t_2}(x_1, x_2) = P(W(t_1) \le x_1, W(t_2) \le x_2), t_1 \ne t_2 \ge 0, x_1, x_2 \ge 0. \tag{3.2}$$

Note that $\mathcal{D}_t(x) - \mathcal{U}_t(x) \in \{0, +1,\, -1\}$ for every $x \ge 0, t \ge 0$, since down- and upcrossings of a fixed level alternate in time (Proposition 2.3). The next lemma connects $E(\mathcal{U}_t(x))$, $E(\mathcal{D}_t(x))$ and the transient cdf $F_t(x)$, by using (3.2) with $t_1 = 0$, $t_2 = t$, $x_1 = x_2 = x$.

In M/G/1, $\mathcal{D}_t(x) = \mathcal{D}_t^c(x)$ (Subsection 2.4.4), since all downcrossings are left-continuous. Also $\mathcal{U}_t(x) = \mathcal{U}_t^j(x)$, since all upcrossings are jump upcrossings.

Theorem 3.1 *In the M/G/1 queue, for fixed $x \ge 0, t \ge 0$,*

$$E(\mathcal{D}_t(x)) = E(\mathcal{U}_t(x)) + F_t(x) - F_0(x). \tag{3.3}$$

Proof. The initial condition $\mathcal{D}_0(x) = \mathcal{U}_0(x) = 0$ implies (3.3) holds for $t = 0$. For $t > 0$, examination of possible sample paths $\{W(s)\}, 0 \le s \le t$, (Fig. 3.2) leads to the following values and probabilities for $\mathcal{D}_t(x) - \mathcal{U}_t(x)$:

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

(3.4)

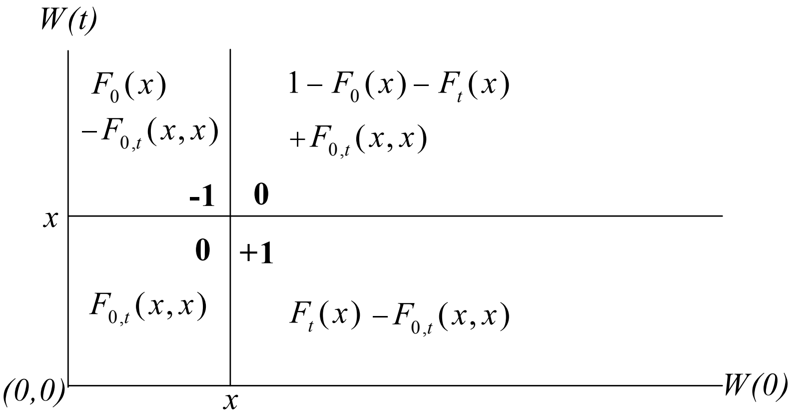

Figure 3.2: Values of $\mathcal{D}_t(x) - \mathcal{U}_t(x)$ are $+1$, 0, -1, with probabilities shown in areas of $(W(0), W(t))$ plane.

From (3.4) we obtain for fixed $x \geq 0$, the expected value

$$E(D_t(x)) - E(\mathcal{U}_t(x)) = F_t(x) - F_0(x), t \geq 0, . \tag{3.5}$$

identical to (3.3). ■

In (3.4) the term $D_t(x) - \mathcal{U}_t(x) = 0$ does not affect the expected value; it is included for completeness. In further similar computations of expected value, terms with value 0 may be omitted. Equation (3.5) leads to the following basic theorem relating the transient distribution of wait and sample-path properties.

Theorem 3.2 *In the M/G/1 queue*

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) = \frac{\partial}{\partial t}F_t(x) + \frac{\partial}{\partial t}E(\mathcal{U}_t(x)), t > 0, x \geq 0. \tag{3.6}$$

Proof. Differentiating (3.5) with respect to t gives formula (3.6). ■

Remark 3.3 *Theorem 3.2 is a special case of a general theorem connecting the marginal entrance and exit rates of an arbitrary measurable set* $\mathbf{A} \subset \mathbf{S}$ *(state space) to the transient probability of* $\mathbf{A}$*,* $P_t(\mathbf{A})$ *(see Theorems 4.1 and 4.1). In the present context,* $\mathbf{A} = [0, x]$*.*

3.2.4 Downcrossings and Transient PDF of Wait

The following theorem connects $\frac{\partial}{\partial t}E(\mathcal{D}_t(x))$ and $f_t(x)$,$x \geq 0$, the transient pdf, where $f_t(0) \equiv f_t(0^+)$.

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

$$\frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = f_t(x), x > 0, \tag{3.7}$$
$$\frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = f_t(0). \tag{3.8}$$

Proof. For the virtual wait, fix state-space level $x > 0$. Consider instants t and $t+h$, $t > 0$, and small $h > 0$. Examination of sample paths $W(s)$, $s \in (t, t+h)$ over the state space interval $(x, x+h)$, leads to the following values of $\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x)$ and probabilities (Fig. 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	$o(h)$

(3.9)

Taking the expected value of $D_{t+h}(x) - D_t(x)$ 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 (3.7); then letting $x \downarrow 0$ yields (3.8). (The value $\mathcal{D}_{t+h}(x) - \mathcal{D}_t(x) = 0$ does not affect the expected value). ■

Corollary 3.1 *For fixed $t > 0$,*

$$E(\mathcal{D}_t(x)) = \int_{s=0}^{t} f_s(x) ds, x > 0, t > 0. \tag{3.10}$$
$$E(\mathcal{D}_t(0)) = \int_{s=0}^{t} f_s(0) ds, t > 0. \tag{3.11}$$

Proof. In (3.7) and (3.8) change s to u and t to s. Then integrate both sides with respect to $s \in (0, t)$. The initial condition $E(\mathcal{D}_0(x)) \equiv 0, x \geq 0$, gives the result. ■

Let $\{P_0; f(x), x > 0\}$, $F(x), x \geq 0$ denote the *steady-state* pdf and cdf of the virtual wait, respectively.

Corollary 3.2 *If the steady state exists (stability), then*

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x), x > 0 \tag{3.12}$$
$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(0))}{t} = f(0^+) \equiv f(0). \tag{3.13}$$

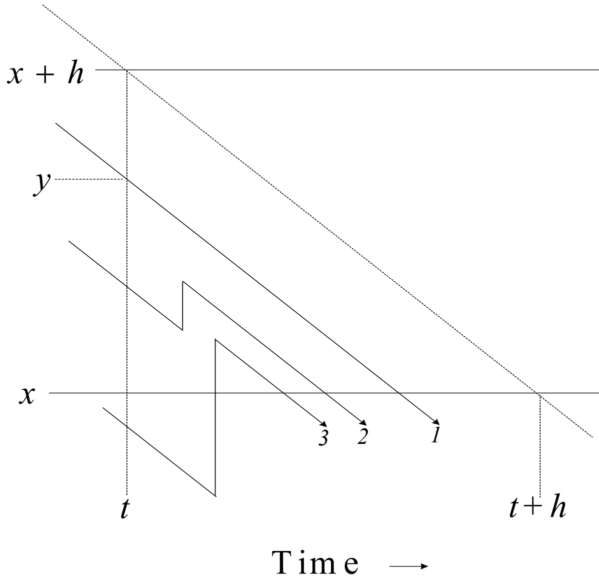

Figure 3.3: Sample path examples in time interval $(t, t+h)$ resulting in $D_t(x+h) - D_t(x) = 1$. Probabilities are: P(path type 1) $= 1 - \lambda(y - x) + o(y - x)$; P(path type 2) $\leq o(h)$; P(path type 3) $\leq o(h)$.

Proof. Let $t \to \infty$ in (3.7) and (3.8) giving

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

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

In (3.10) and (3.11) divide both sides by $t > 0$, and let $t \to \infty$. Since $\lim_{t\to\infty} f_t(x) = f(x), x \geq 0$, (3.12) and (3.13) follow. ■

Let "$\underset{a.s.}{=}$" mean "with probability 1" ($a.s. \equiv$ "almost surely").

Corollary 3.3 *If the steady state exists, then*

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

Proof. By the elementary renewal theorem,

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

The result follows from (3.12) and (3.13). ■

Corollary 3.4 ***Rate balance for level crossings:***

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

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

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 the virtual wait. 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). (see Subsection 2.4.6) ■

Remark 3.4 *Formulas (3.17), (3.18) are also statements of the principle of set balance, i.e.,* ***rate of sample-path exits from set*** $[0, x)$ = ***rate of sample-path entrances into*** $[0, x)$*. The same principle applies to set* $[x, \infty)$*. SP motion contains the sample path as a subset. Hence the same principle applies to SP exits and entrances.*

3.2.5 Upcrossings and Transient PDF of Wait

The next theorem connects $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$ to $P_0(t)$ and $f_t(y), 0 < y < x$.

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. Let $x > 0$, $t > 0$, be given, and let $h > 0$ be small. Observation of possible sample paths $\{W(s)\}$, $s \in (t, t+h)$ in the vicinity of state-space interval $(x, x+h)$ yields the following values of $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)$ and the corresponding probabilities.

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

(3.21)

the first $o(h)$ includes multiple jumps of which exactly one exceeds x.

In (3.21), the value $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) = 0$ is omitted since it does not affect the expected value. Negative values are not possible, since $\mathcal{U}_t(x)$ is a counting process (non-decreasing).
Taking the expected value in (3.21) yields

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

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

Corollary 3.5 *For fixed $t > 0$,*

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

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

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). ■

3.2.6 Equation for Transient PDF of Wait

We apply LC to derive a known integro-differential equation for the transient distribution of wait, by utilizing Theorems 3.2, 3.3 and 3.4.

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 (3.6), substituting from (3.7), (3.8), (3.19), (3.20), and using (3.1). Equation (3.28) is the normalizing condition. ■

Remark 3.5 *Minor extensions of the proofs in this section yield relationships and integro-differential equations for the transient pdf of wait when the arrival rate and probability distribution of the service time are time-dependent. That is, in the formulas of this section, we can replace λ by λ_t so that the arrival process is non-homogeneous Poisson. Also, we can replace $B(y)$ by $B_t(y)$.*

Remark 3.6 *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 distribution of wait** by simulation of multiple independent sample paths (see Remark 9.2). **LC estimation (computation, approximation) for steady-state distributions is discussed in Chapter 9.** LC estimation is a form of non-parametric distribution (or density) estimation.*

3.2.7 Steady-State Distribution of Wait

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*

$$f(x) = \lambda \overline{B}(x) P_0 + \lambda \int_0^x \overline{B}(x-y) f(y) dy, x > 0, \tag{3.29}$$

$$f(0) = \lambda P_0, \tag{3.30}$$

$$P_0 + \int_0^\infty f(y) dy = 1. \tag{3.31}$$

Proof. Since $\lambda E(S) < 1$, the transient distribution converges to the steady state 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

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

The result follows from Theorem 3.6, by letting $t \to \infty$.

Alternatively, the result follows from rate balance for level crossings, i.e., from (3.17), (3.18), and substituting from (3.12), (3.13), (3.24), (3.25). ■

Remark 3.7 *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* n^{th} *customer [99]. Hence equations (3.29) - (3.31) hold for the steady state distributions of both the customer wait and the virtual wait.*

Remark 3.8 *It is important to derive (3.29) - (3.31) for the steady state distribution of wait using LC, because each algebraic term corresponds to a unique down- or upcrossing rate of* $x \geq 0$. *This type of correspondence enables us to derive integral equations for steady state distributions of state variables in many complex stochastic models, intuitively and straightforwardly. The idea is to study a typical sample path of the stochastic model, and then write the integral equation(s) and any boundary conditions (e.g.,* $f(0) = \lambda P_0$*) by inspection using LC theorems and rate balance or set balance.*

Example 3.1 *Consider the* ***M/E$_k$/1 queue*** *with arrival rate* λ *and service time* S *having pdf*

$$b(x) = e^{-\mu x}\frac{(\mu x)^k \mu}{k!}, x > 0, \mu > 0, \text{ and } \lambda < \frac{\mu}{k}.$$

The cdf of the service time is

$$B(x) = \int_{y=0}^{x} e^{-\mu y}\frac{(\mu y)^k \mu}{k!} dy$$

and the complementary cdf is

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

Substituting into (3.29), the integral equation for the steady-state pdf of wait, $f(x)$*, is*

$$\begin{aligned} f(x) = \quad & \lambda P_0 e^{-\mu x}\left(\textstyle\sum_{i=0}^{k-1}\frac{(\mu x)^i}{i!}\right) \\ & + \lambda \textstyle\int_{y=0}^{x} e^{-\mu(x-y)}\left(\textstyle\sum_{i=0}^{k-1}\frac{(\mu(x-y))^i}{i!}\right) f(y)dy, \ x > 0. \end{aligned} \tag{3.32}$$

where $P_0 = 1 - \lambda E(S) = 1 - \frac{k\lambda}{\mu}$.

Case k = 2: *Setting* $k = 2$ *in (3.32) corresponds to the* $M/E_2/1$ *queue. The integral equation for* $f(x)$ *is then*

$$f(x) = \lambda P_0 e^{-\mu x}(1 + \mu x) + \lambda \int_{y=0}^{x} e^{-\mu(x-y)}(1 + \mu(x-y))f(y)dy, x > 0. \tag{3.33}$$

Differentiating (3.33) 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

$$f(x) = a_1 e^{r_1 x} + a_2 e^{r_2 x}, x > 0 \tag{3.34}$$

where a_1*,* a_2 *are constants to be determined and*

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

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*

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

3.2.8 Alternative Forms of the LC Integral Equation

We can write equation (3.29) for the steady-state pdf of wait as

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

The last two alternative forms of the LC equation,

$$f(x) = \lambda F(x) - \lambda \textstyle\int_{y=0}^{x} B(x-y)dF(y),\ x \geq 0; \tag{3.35}$$

$$f(x) = \lambda F(x) - \lambda \textstyle\int_{y=0}^{x} F(x-y)dB(y),\ x \geq 0. \tag{3.36}$$

have an intuitive interpretation in terms of level crossing dynamics, which enables them to be written down directly. Consider a sample path of the virtual wait (e.g., 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 (3.35) or (3.36) the left side is the SP downcrossing rate of level x, as usual (see 3.12). On the right side, the first term is the rate of *all* SP jumps that start in the state-space interval $[0, x]$. The second term subtracts 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 equations (3.35) and (3.36). Note that (3.35) yields (3.36) by using the transformation $z = x - y$, $dz = -dy$, and integrating by parts.

These alternative forms of the LC integral equation are useful when analyzing variants of M/D/1 and M/Discrete/1 queues (sections 3.8, 3.9), as well as other models. They are also useful in theoretical applications, such as in TAM (transform approximation method) [66], [93], [94]. The LC "intuitive" construction of (3.35) and (3.36), suggests how to use LC to develop integral equations for the pdf of wait in more general models.

Example 3.2 *Consider the M/Uniform/1 queue with arrival rate* λ. *Assume the service time is uniform on* $(0, c), c > 0$*, i.e.,*

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

Stability (steady state) exists provided $\lambda \frac{c}{2} < 1$. *Substituting the uniform* $B(\cdot)$ *into (3.35), gives an integral equation for the steady-state distribution of wait,*

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} \frac{(x-y)}{c} dF(y), 0 < x < c, \tag{3.37}$$

$$f(x) = \lambda F(x) - \lambda \int_{y=x-c}^{x} \frac{(x-y)}{c} dF(y) - \lambda F(x-c), x \geq c. \tag{3.38}$$

On the right side of equation (3.38), the difference $\lambda F(x) - \lambda F(x-c)$ *is the rate of jumps that start in state-space interval* $[x-c, x]$. *Jumps that start in* $[0, x-c)$ *cannot upcross* x.

Solution Approach for Example 3.2

We carry out only the first step of the solution by solving (3.37), to suggest a procedure applicable to many M/G/1 variants. We obtain $f(x)$,$x \in (0, c)$, and indicate the iteration on successive intervals of length c in the state space. Later we obtain an analogous complete solution for M/D/1 (Section 3.8).

Differentiating (3.37) twice with respect to x results in the second order differential equation

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

The solution is

$$
\begin{aligned}
f(x) \quad = \quad & a_1.e^{\frac{\lambda}{2}x}\cos\left(\frac{1}{2}\sqrt{\frac{4\lambda}{c}-\lambda^2}\cdot x\right) \\
& + a_2\cdot e^{\frac{\lambda}{2}x}\sin\left(\frac{1}{2}\sqrt{\frac{4\lambda}{c}-\lambda^2}\cdot x\right),
\end{aligned}
$$

where a_1, a_2 are constants. 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

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

Hence

$$
\begin{aligned}
f(x) = \quad & e^{\frac{\lambda}{2}x}[(\lambda(1-\tfrac{\lambda c}{2})\cos\left(\tfrac{1}{2}\sqrt{\tfrac{4\lambda}{c}-\lambda^2}\cdot x\right) \\
& + \tfrac{(1-\frac{\lambda c}{2})\lambda(\lambda-\frac{1}{c})}{\sqrt{\frac{4\lambda}{c}-\lambda^2}}\sin\left(\tfrac{1}{2}\sqrt{\tfrac{4\lambda}{c}-\lambda^2}\cdot x\right)], 0<x<c.
\end{aligned} \tag{3.39}
$$

We can iterate to solve for $f(x), x\in[c,2c)$, $x\in[2c,3c)$, etc., using (3.38). For $x\in[c,2c)$, we have

$$
\begin{aligned}
f(x) = \quad & \lambda F(x) - \lambda\textstyle\int_{y=c}^{x}\frac{(x-y)}{c}dF(y) \\
& -\lambda\textstyle\int_{y=x-c}^{c}\frac{(x-y)}{c}f(y)dy - \lambda F(x-c), c\le x<2c.
\end{aligned} \tag{3.40}
$$

We solve for $f(x)$,$x\in[c,2c)$ by substituting for $f(y)$ from (3.39) on the interval $(x-c,c)$ to evaluate the second integral in (3.40), and using continuity $f(c^-)=f(c)$. (Continuity can be proved similarly as for the M/D/1 queue in Section 3.10.) The procedure may be repeated recursively on intervals $[ic,(i+1)c)$, $i\ge 2$. When numerics are substituted for the parameters λ and c, the procedure can be readily programmed on a computer.

3.2.9 Equation for Distribution of System Time

This subsection uses LC to develop a relationship between the steady-state pdf of wait and the steady-state cdf of system time. Let σ denote

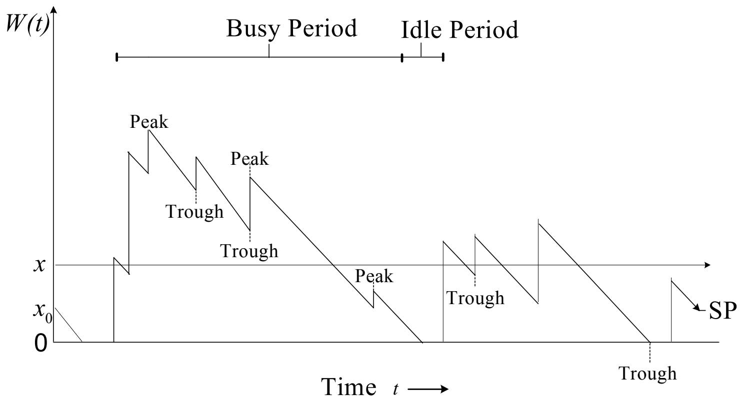

Figure 3.4: Sample path of virtual wait showing peaks and troughs, and a level x.

the *total time spent in the system* by an arbitrary arrival. Let the pdf and cdf of σ be $f_\sigma(x)$, $F_\sigma(x)$,$x > 0$, respectively. Then $\sigma = W_q + S$, where W_q is the wait before service and S is the common service time.

Consider a sample path of the virtual wait (Fig. 3.4). It has a sequence of peaks (relative maxima) and troughs (relative "minima" which are infima, due to sample-path *right continuity*). A trough at level 0 is considered to occur at an instant the SP hits 0 from above.

Fix level $x \geq 0$. Let $P_t^+(x)$, $T_t^+(x)$ denote the number of peaks and troughs, respectively, at levels strictly above level x during time interval $[0, t)$. Recall that $\mathcal{D}_t(x)$ is the number of SP downcrossings of x during $(0, t)$. It is straightforward to show that for fixed $t > 0$, $\mathcal{D}_t(x),$ is a step function in x, and

$$\mathcal{D}_t(x) = P_t^+(x) - T_t^+(x), t > 0. \tag{3.41}$$

Let $N_A(t)$ denote the number of arrivals during $(0, t)$. Assume $N_A(t) > 0$. Dividing (3.41) by $t > 0$, we obtain

$$\begin{aligned} \frac{\mathcal{D}_t(x)}{t} &= \frac{P_t^+(x)}{t} - \frac{T_t^+(x)}{t} \\ &= \frac{N_A(t)}{t} \cdot \frac{P_t^+(x)}{N_A(t)} - \frac{N_A(t)}{t} \cdot \frac{T_t^+(x)}{N_A(t)}, t > 0. \end{aligned} \tag{3.42}$$

Note that $P_t^+(x)$ represents the number of *system times* greater than x in $(0, t)$. Also $T_t^+(x)$ represents the number of *waiting times* greater than x in $(0,t)$. Also

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

Thus, letting $t\to\infty$ on both sides of (3.42) gives another alternative form of the M/G/1 "integral" equation for pdf of wait,

$$f(x)=\lambda(1-F_\sigma(x))-\lambda(1-F(x)), \tag{3.43}$$

or

$$f(x)=\lambda F(x)-\lambda F_\sigma(x). \tag{3.44}$$

The LC intuitive interpretation of these equations are as follows. On the right side of (3.43) the first term is the rate of all jumps that *end above* level x (system time $> x$). The second term subtracts 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 (3.44) is that the first term on the right side is the rate of all jumps that *start* at levels $\le x$ (wait $\le x$). The second term subtracts the rate of those jumps that *end* at levels $\le x$ (system time $\le x$). The right side is the rate of SP jumps that upcross x.

Equation (3.43) can be rearranged as

$$\begin{aligned}\lambda(1-F_\sigma(x))= \quad & \lambda P_0\overline{B}(x)+\lambda\textstyle\int_{y=0}^{x}\overline{B}(x-y)f(y)dy\\ & +\lambda\textstyle\int_{y=x}^{\infty}f(y)dy,\end{aligned} \tag{3.45}$$

upon using (3.29) and

$$\lambda(1-F(x))=\lambda\int_{y=x}^{\infty}f(y)dy.$$

Remark 3.9 *Equation (3.42) combines sample-path peaks and troughs and the basic LC theorem* $\lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t}=f(x)$*, to provide a very simple derivation of the basic LC integral equation for the steady-state pdf of wait, since (3.43) and (3.44) are immediately transformable to (3.29.*

3.3 Waiting Time Properties

We derive several known properties of the waiting time using LC. (Note that (3.29) has been derived by LC.)

3.3.1 Probability of Zero Wait

In (3.29) integrate both sides with respect to x over $(0, \infty)$. This yields

$$
\begin{aligned}
1 - P_0 &= \lambda P_0 \int_{x=0}^{\infty} \overline{B}(x)dx + \lambda \int_{x=0}^{\infty} \int_{y=0}^{x} \overline{B}(x-y)f(y)dydx, \\
1 - P_0 &= \lambda P_0 E(S) + \lambda E(S)(1 - P_0), \\
P_0 &= 1 - \lambda E(S). \qquad (3.46)
\end{aligned}
$$

Formula (3.46) is the well known steady-state probability of a zero wait.

3.3.2 Pollaczek-Khinchin (P-K) Formula

In (3.29) multiply both sides by x and integrate with respect to x over $(0, \infty)$. We obtain

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

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

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

Thus we obtain the well known Pollaczek-Khinchin (P-K) formula

$$E(W_q) = \frac{\lambda E(S^2)}{2(1 - \lambda E(S))} = \frac{\lambda E(S^2)}{2P_0} = \frac{\lambda(Var(S) + (E(S))^2)}{2(1 - \lambda E(S))}. \qquad (3.47)$$

3.3.3 Expected Number in Queue

Let N_q denote the number of customers waiting before service, and L_q its expected value, in steady state. From Little's formula "$\boldsymbol{L = \lambda W}$" and (3.47),

$$
\begin{aligned}
E(N_q) \equiv L_q &= \lambda E(W_q) \\
&= \frac{\lambda^2 E(S^2)}{2(1 - \lambda E(S))} = \frac{\lambda^2 E(S^2)}{2(1 - \rho)}.
\end{aligned}
$$

The expected number in the system is

$$L = L_q + L_s$$

where L_s denotes the expected number in service. L_s is given by

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

Thus

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

3.3.4 Laplace-Stieltjes Transform

The Laplace–Stieltjes transform (LST) of the wait before service is

$$F^*(s) \equiv \int_{x=0}^{\infty} e^{-sx} dF(x) = P_0 + \int_{x=0}^{\infty} e^{-sx} f(x) dx, s > 0. \tag{3.48}$$

The LST of the service time is

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

Note that

$$\int_{x=0}^{\infty} e^{-sx} \overline{B}(x) dx = \int_{x=0}^{\infty} e^{-sx} (1 - B(x)) dx = \frac{1}{s}(1 - B^*(s)).$$

In (3.29) we multiply both sides by e^{-sx} and integrate with respect to x over $(0, \infty)$, and obtain

$$\begin{aligned} \int_{x=0}^{\infty} e^{-sx} f(x) dx = \quad & \lambda P_0 \int_{x=0}^{\infty} e^{-sx} \overline{B}(x) dx \\ & + \lambda \int_{x=0}^{\infty} \int_{y=0}^{x} e^{-sx} \overline{B}(x-y) f(y) dy dx. \end{aligned} \tag{3.49}$$

or

$$\begin{aligned} F^*(s) - P_0 = \quad & \lambda P_0 \int_{x=0}^{\infty} e^{-sx} \overline{B}(x) dx \\ & + \lambda \int_{x=0}^{\infty} \int_{y=0}^{x} e^{-sx} \overline{B}(x-y) f(y) dy dx. \end{aligned} \tag{3.50}$$

In the double integral, express $e^{-sx} = e^{-sy} e^{-s(x-y)}$, interchange limits of integration, and simplify to yield the well known formula (e.g., [63])

$$\begin{aligned} F^*(s) &= \frac{sP_0}{s - \lambda(1 - B^*(s))} = \frac{s(1 - \lambda E(S))}{s - \lambda(1 - B^*(s))} \\ &= \frac{1 - \lambda E(S)}{1 - \lambda E(S) \left(\frac{1 - B^*(s)}{sE(S)} \right)}, s > 0. \end{aligned} \tag{3.51}$$

Let $\rho = \lambda E(S)$. We can expand $F^*(s)$ as a series

$$\begin{aligned} F^*(s) &\equiv P_0 + \int_{x=0}^{\infty} e^{-sx} f(x)dx \\ &= 1 - \rho + (1-\rho)\sum_{k=1}^{\infty}\left(\frac{1-B^*(s)}{sE(S)}\right)^k . \end{aligned} \tag{3.52}$$

We can invert $F^*(s)$ to obtain

$$\begin{aligned} &P_0 = 1 - \rho, \\ &f(x) = (1-\rho)\textstyle\sum_{k=1}^{\infty} g^{*k}(x), x > 0, \end{aligned} \tag{3.53}$$

where $g^{*k}(x)$ is the k-fold convolution of the steady-state excess service time (see [78] pages 200-201, and our Subsection 10.2.2, or sections on renewal theory in, e.g., [74] or [91]). We shall see in Section 3.15, that the series (3.53) is a special case of a more general series having a level-crosssing interpretation.

Remark 3.10 *It is known that equations (3.49) and (3.51) can be interpreted as the probability that the waiting time in queue is less than an independent "catastrophe" random variable which is exponentially distributed with rate s. That is, the wait in queue finishes before the catastrophe occurs with probability $F^*(s)$. This* ***probabilistic interpretation*** *can often be used to derive Laplace transforms of random variables associated with stochastic models (e.g., [31], Section 3).*

3.3.5 System Time

Let σ denote the time spent in the system by an arbitrary arrival in steady state. Denote its pdf and cdf by $f_\sigma(x)$, $F_\sigma(x)$, $x > 0$, respectively. Let W_q be the wait before service and S the service time. Recall that $f(x)$, $F(x)$ are the pdf and cdf of W_q. For an arbitrary arrival, $\sigma > x$ iff the arrival waits in queue $y \leq x$ and the service time exceeds $x - y$, or, the arrival waits in queue $> x$. Thus

$$\begin{aligned} 1 - F_\sigma(x) &= P(\sigma > x) \\ &= P_0\overline{B}(x) + \int_{y=0}^{x} \overline{B}(x-y)f(y)dy + 1 - F(x) \\ &= \frac{f(x)}{\lambda} + 1 - F(x) \end{aligned} \tag{3.54}$$

and

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

which is the same as (3.44). 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.

3.3.6 PDF of System Time in Terms of PDF of Wait

We now give an LC equation for $f_\sigma(x)$ directly in terms of $f(x)$. 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 the 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 than $y - x$ later. Thus the time interval of duration $y - x$ is devoid of new arrivals and corresponding 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 interval (x, y). The end level y is the system time of the corresponding arrival.)

By the basic LC theorem for M/G/1 (Theorem 1.1), another expression for the SP downcrossing rate of x is $f(x)$ (equal to upcrossing rate). Hence we have the equation

$$\lambda \int_{y=x}^{\infty} e^{-\lambda(y-x)} f_\sigma(y) dy = f(x). \tag{3.55}$$

Multiplying both sides of (3.55) by $e^{-\lambda x}$ and differentiating with respect to x yields

$$f_\sigma(x) = f(x) - \frac{f'(x)}{\lambda}, x > 0, \tag{3.56}$$

wherever $f'(x)$ exists. Thus, if $f(x)$ is known, $f_\sigma(x)$ can be found directly using (3.56).

Example 3.3 *In $M_\lambda/M_\mu/1$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}$,$x > 0$ (see (3.86) below). Substituting into (3.56) yields*

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

(same as (3.90) below).

3.3.7 Number in System

We obtain the steady state probability distribution of the number in the system in two ways (for perspective), by conditioning on either W_q or on σ. Let $P_n, n = 0, 1, ...,$ denote the probability of n customers in the system at an arbitrary time point. Let a_n, $d_n, n = 0, 1, ...,$ denote the steady-state probability of n in the system just before an arrival, and just after a departure, respectively. 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., [91]).

Conditioning on W_q, we obtain

$$\begin{aligned} P_n = d_n &= \int_{y=0}^{\infty} P(n-1 \text{ arrivals during } y|W_q = y)f(y)dy \\ &= \int_{y=0}^{\infty} e^{-\lambda y}\frac{(\lambda y)^{n-1}}{(n-1)!}f(y)dy, n = 1, 2, ... \, . \end{aligned} \tag{3.57}$$

We can check that (3.57) is consistent with $P_0 + \int_{y=0}^{\infty} f(y)dy = 1$ since

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

Alternatively, conditioning on σ,

$$\begin{aligned} P_n = d_n &= \int_{y=0}^{\infty} P(n \text{ arrivals during } y|\sigma = y)f_\sigma(y)dy \\ &= \int_{y=0}^{\infty} e^{-\lambda y}\frac{(\lambda y)^{n}}{n!}f_\sigma(y)dy, n = 0, 1, ... \end{aligned} \tag{3.58}$$

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

$$\begin{aligned} \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} d_n &= \int_{y=0}^{\infty} e^{-\lambda y} \sum_{n=0}^{\infty} \frac{(\lambda y)^{n}}{n!} \cdot f_\sigma(y)dy \\ &= \int_{y=0}^{\infty} f_\sigma(y)dy = 1. \end{aligned}$$

If $f(\cdot)$, $f_\sigma(\cdot)$ are known for an M/G/1 model, equation (3.57) or (3.58) can yield$\{P_n\}$.

3.3.8 Expected Busy Period

Let $\mathcal{B}$ denote a busy period. Consider a sample path of the virtual wait. To gain insight and see connections among different approaches, we give three ways to derive the expected busy period $E(\mathcal{B})$.

(1) The long-run proportion of time that the sample path is in the state-space set $(0,\infty)$ is equal to $\lambda P_0 E(\mathcal{B})$ (SP rate out of $\{0\} \cdot E(\mathcal{B})$). It is also equal to $1 - P_0$. Hence

$$\begin{aligned} \lambda P_0 E(\mathcal{B}) &= 1 - P_0, \\ E(\mathcal{B}) &= \frac{1-P_0}{\lambda P_0} = \frac{E(S)}{1-\lambda E(S)}. \end{aligned} \tag{3.59}$$

Can the appearance of P_0 in the denominator of (3.59) be explained? We next give a derivation of (3.59) using the virtual-wait sample-path downcrossing rate of level 0 (hit rate of 0 from above), which provides intuitive insight.

(2) The long-run proportion of time that a sample path is in the state-space interval $(0,\infty)$ is $1 - P_0 = \rho = \lambda E(S)$. Successive busy *cycles* form a renewal process. There is one busy period embedded within each busy cycle. A sample path is in state-space interval $(0,\infty)$ only during busy periods. Busy periods are iid random variables. By the theory of regenerative processes (e.g., [96]) we obtain

$$\frac{E(\mathcal{B})}{E(\text{Busy cycle})} = \rho = 1 - P_0.$$

From renewal theory (e.g., [49], [74], [91]) and LC theory,

$$E(\text{Busy cycle}) = \frac{1}{(\text{Downcrossing rate of level 0})} = \frac{1}{f(0)} = \frac{1}{\lambda P_0}.$$

Hence $E(\mathcal{B})$ is the $(1-P_0)$ proportion of a busy cycle, i.e.,

$$E(\mathcal{B}) = (1-P_0)\cdot E(\text{Busy cycle}) = \frac{1-P_0}{f(0)} = \frac{1-P_0}{\lambda P_0} = \frac{E(S)}{1-\lambda E(S)}.$$

The key reason for P_0 appearing in the denominator is seen directly from Theorem 1.1, Corollary 1.1, namely $f(0)=\lambda P_0$! The expression

$$E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0} \tag{3.60}$$

appears to be more fundamental than the expression $E(\mathcal{B}) = \frac{E(S)}{1-\lambda E(S)}$, since in many model *variants* of the standard M/G/1 queue, $P_0 \neq 1 - \lambda E(S)$ (e.g., sections 3.7, 3.11)

(3) Busy periods and idle periods form an alternating renewal process. Hence

$$\frac{E(\mathcal{B})}{E(\mathcal{B}) + E(\text{Idle period})} = \frac{E(\mathcal{B})}{E(\mathcal{B}) + \frac{1}{\lambda}} = 1 - P_0,$$

which implies (3.60). This derivation is equivalent to (2), since (Busy cycle) = $\mathcal{B}$+(Idle period). However, it does not "explain" the appearance of λP_0 in the denominator. The LC derivation (2) does provide an explanation.

Remark 3.11 *Formula (3.60), $E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0}$, shows immediately that*

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

which is equivalent to

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

*The **stability condition** for the standard M/G/1 queue is $P_0 > 0$ (same as $\lambda E(S) < 1$). That is, the queue is stable iff state $\{0\}$ is positive recurrent, equivalently iff the expected busy period is finite.*

Remark 3.12 *Formula $E(\mathcal{B}) = \frac{1-P_0}{f(0)}$ is more fundamental than $E(\mathcal{B}) = \frac{1-P_0}{\lambda P_0}$, since in some M/G/1 variants $f(0) \neq \lambda P_0$. An example is M/G/1 with bounded virtual wait, as in Variant 2 of Subsection 3.14.3. In that model the upper bound is K. Then $f(0) = \lambda P_0(1 - \overline{B}(K))$ and*

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0(1 - \overline{B}(K))}.$$

3.3.9 Structure of Busy Period

Consider a busy period of the virtual wait (Fig. 3.5). We derive a property of the busy period from direct observation of the sample path. Suppose a customer arrives at t_A^- and must wait $y \geq 0$ before service. The SP then has coordinates (t_A^-, y). At t_A the sample path jumps an amount S, to level $y + S$. Let t_y be the first instant after t_A such that the sample path hits level y from above, i.e.,

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

A busy period may be defined as the interval length $t_y - t_A$. The time interval $t_y - t_A$ is independent of y, since the SP jump at t_A is a ***full service time distributed as*** S. We utilize this definition of a busy period to study the structure of a busy period. (The usual definition of busy period is made for $y = 0$ only, e.g., [99].)

Consider a busy period $\mathcal{B}$ during which at least one customer arrives after the start of the busy period. Denote their arrival times within $\mathcal{B}$ by $\tau_1 < \tau_2 < \cdots$. Then $0 < W(\tau_i^-), i = 1, 2, ...$. Define $\tau_1^* = \tau_1$ and $\tau_{n+1}^* = \min\{\tau_i | W(\tau_i^-) < W(\tau_n^*)\}, i > n = 1, 2, ...$. Due to the memoryless property of the inter-arrival times and since $\frac{d}{dt}W(t) = -1, W(t) > 0$, the waits $\{W(\tau_n^{*-})\}$ are distributed the same as the customer arrival times *during the first service time* S. We call the customers that arrive at time points $\{\tau_n^*\}$ *"tagged"* arrivals (see Fig. 3.5).

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. The tagged arrivals are those that initiate their own busy periods starting at $\{(\tau_n^{-*}, W(\tau_n^{-*}))\}$ in the time-state plane, similar to $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ depicted in Fig. 3.5. in Fig. 3.5, $\tau_1^* = \tau_1$, $\tau_2^* = \tau_4$, $\tau_3^* = \tau_6$. 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 $(\tau_n^{-*}, W(\tau_n^{-*})), n = 1, ..., N_S$ are strict descending ladder points ([56]) within $\mathcal{B}$. Then

$$\mathcal{B} \underset{dist}{=} S + \sum_{i=1}^{N_S} \mathcal{B}_i, \qquad (3.61)$$

where $\{\mathcal{B}_i\}$ are iid r.v.'s each distributed as $\mathcal{B}$ independent of N_S. Equation (3.61) is known, and is usually derived by different, but equivalent, reasoning (e.g., [78]). From (3.61), we obtain

$$\begin{aligned} E(\mathcal{B}) &= E(S) + E(N_S)E(\mathcal{B}) \\ &= E(S) + \lambda E(S)E(\mathcal{B}) \end{aligned}$$

which gives $E(\mathcal{B})$ as in (3.59).

Also, we can obtain (3.59) by recursively substituting for $\mathcal{B}_i$ in (3.61). This gives an infinite series of terms

$$\mathcal{B} \underset{dist}{=} S + \sum_{i=1}^{N_S} S_i + \sum_{i=1}^{N_S}\sum_{j=1}^{N_S} S_{ij} + \sum_{i=1}^{N_S}\sum_{j=1}^{N_S}\sum_{k=1}^{N_S} S_{ijk} + ...$$

where S_i, $S_{ij,j}$, $S_{ijk,}$ etc., are distributed as S. Assume $0 < \lambda E(S) < 1$, i.e., the steady state distribution of wait exists and $\mathcal{B} < \infty$ $(a.s.)$. Then

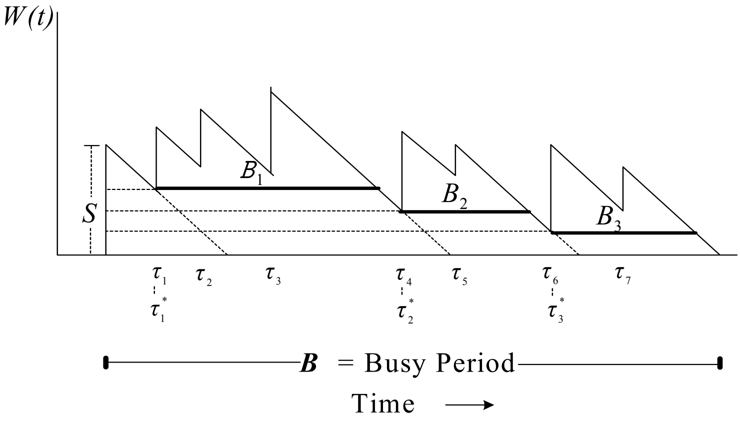

Figure 3.5: Busy period $\mathcal{B} \underset{dist}{=} S + \sum_{i=1}^{N_S} \mathcal{B}_i$. $\mathcal{B}_i \underset{dist}{=} \mathcal{B}, i = 1, ..., N_S$. N_S = number of "tagged" arrivals in $\mathcal{B}$. Here $N_S = 3$. $N_S \underset{dist}{=}$ number of arrivals during S. Tagged arrival times are $\tau_1^* = \tau_1$, $\tau_2^* = \tau_4$, $\tau_3^* = \tau_6$. Tagged arrivals 1, 4, 6 during $\mathcal{B}$ initiate $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$. (In figure symbol "B" represents "$\mathcal{B}$".)

expected value is

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

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

We compute the known formula (e.g., [91]) for the variance of $\mathcal{B}$ assuming it exists from (3.61) and the definition

$$Var(\mathcal{B}) = E(\mathcal{B}^2) - (E(\mathcal{B}))^2,$$

for completeness, and because we intend to use the result for $E(\mathcal{B}^2)$, e.g., when discussing M/G/1 priority queues in Section 3.12.

To compute $E(\mathcal{B}^2)$, we first obtain a formula for $\mathcal{B}^2$ from (3.61) 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. Thus

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

The third term on the right is

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

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 (3.59) and simplifying yields

$$\begin{aligned} E(\mathcal{B}^2) &= \frac{E(S^2)(1 + \lambda E(\mathcal{B}))^2}{1 - \lambda E(S)} \\ &= \frac{E(S^2)}{(1 - \lambda E(S))^3} = \frac{E(S^2)}{(1-\rho)^3}, \end{aligned} \tag{3.62}$$

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

Since $Var(\mathcal{B}) = E(\mathcal{B}^2) - (E(B))^2$, from (3.59) and (3.62)

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

3.3.10 Number Served in Busy Period

Notation 3.7 *Random variable* $X \underset{distr}{=} E_a$*: means that random variable* X *"is distributed as" an exponentially distributed r.v. with mean* $\frac{1}{a}, a > 0$. *(We will use this notation often for brevity.)*

Let $N_{\mathcal{B}}$ be the number of customers served in a busy period $\mathcal{B}$. Let S_i, T_i denote the i^{th} service and inter-arrival times during $\mathcal{B}$, respectively. Then $N_{\mathcal{B}} = \min\{n | \sum_{i=1}^{n}(S_i - T_i) \leq 0\}$ is a *stopping time* (e.g., [74], [91]) for the sequence $\{(S_i - T_i)\}$. Since $T_i \underset{distr}{=} \mathrm{E}_\lambda$, the remaining inter-arrival time at the end of $\mathcal{B}$ is also distributed as E_λ (memoryless property [91]). Hence $\sum_{i=1}^{N_{\mathcal{B}}}(S_i - T_i)$ ends a distance *below* 0, which is distributed as E_λ, and

$$E\left(\sum_{i=1}^{N_{\mathcal{B}}}(S_i - T_i)\right) = -\frac{1}{\lambda}, \tag{3.63}$$

$$E(N_{\mathcal{B}})\left(E(S) - \frac{1}{\lambda}\right) = -\frac{1}{\lambda}, \tag{3.64}$$

$$E(N_{\mathcal{B}}) = \frac{1}{1 - \lambda E(S)}. \tag{3.65}$$

We may also write $N_{\mathcal{B}} = \min\{n | \sum_{i=1}^{n} S_i \leq \sum_{i=1}^{n} T_i\}$. In this form it is seen that $N_{\mathcal{B}}$ is a stopping time for both sequences $\{S_i\}$ and $\{T_i\}$. That is, we observe the r.v.'s in the order $S_1, T_1, S_2, T_2,$... and stop at n in both sequences when the stopping criterion $(\sum_{i=1}^{n} S_i \leq \sum_{i=1}^{n} T_i)$ 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 $S_i \underset{dist}{\equiv} S$,

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

which yields (3.65).

Denote a busy cycle by d_0. Then $d_0 = \sum_{i=1}^{N_{\mathcal{B}}} T_i$, and

$$E(d_0) = E(N_{\mathcal{B}})E(T) = E(N_{\mathcal{B}})\frac{1}{\lambda} = \frac{1}{\lambda(1 - \lambda E(S))} \tag{3.66}$$

which also gives (3.65).

We may write

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

where $N_{\mathcal{B}_i} \underset{dist}{\equiv} N_{\mathcal{B}}$, and $N_S \underset{dist}{\equiv}$ number of arrivals in the first service time of a busy period (see Fig. 3.5). Taking expected values yields

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

again leading to (3.65).

Notably (3.65) is the same as $E(N_{\mathcal{B}}) = \frac{1}{P_0}$. 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 subsection, is that the downcrossing rate level 0 (SP hit rate of 0 from above) is $f(0)$, which implies $E(d_0) = \frac{1}{f(0)} = \frac{1}{\lambda P_0}$. Noting that d_0 is a busy cycle, and applying the stopping time definition of busy cycle as in (3.66), leads to (3.65).

3.3.11 Inter-Downcrossing Time of a Level

Consider a sample path of the virtual wait (Fig. 3.6). Let d_x represent the time between two successive downcrossings of level $x \geq 0$. Starting at the instant of the first downcrossing of level x, r.v. d_x is an interval of a renewal process $\{\mathcal{D}_t(x)\}$ due to exponential inter-arrival times. The renewal rate is $\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x)$ (Corollary3.2). Thus,

$$E(d_x) = \frac{1}{f(x)}, x \geq 0 \tag{3.67}$$

where $f(x)$ is the solution of (3.29) and (3.31).

A busy cycle $d_0 = \mathcal{B} + \mathcal{I}$ where $\mathcal{B}$, $\mathcal{I}$ represent the busy and idle periods, respectively. Letting $x \downarrow 0$ in (3.67) gives the expected busy cycle

$$\begin{aligned} E(d_0) &= \frac{1}{f(0)} = \frac{1}{\lambda P_0} = \frac{1}{\lambda(1 - \lambda E(S))} \\ &= E(\mathcal{B}) + E(\mathcal{I}) = E(\mathcal{B}) + \frac{1}{\lambda}. \end{aligned}$$

Thus we obtain the expected busy period as in (3.59),

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

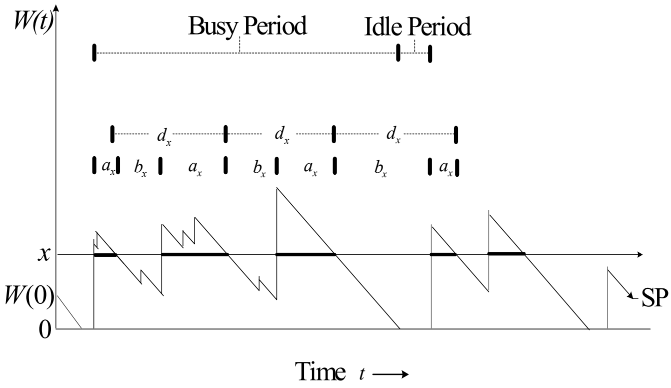

Figure 3.6: Sample path of virtual wait in M/G/1. Shows inter-downcrossing time d_x, sojourn a_x, sojourn b_x, busy and idle periods.

3.3.12 Sojourn Time Below a Level

Let b_x denote a virtual-wait sample-path sojourn time below, or at, level $x \geq 0$ (Fig. 3.6). Assuming the queue is stable ($\rho = \lambda E(S) < 1$), the proportion of time a sample path spends at or below x, is $f(x)E(b_x)$ *and is also equal to* $F(x)$. Hence

$$E(b_x) = \frac{F(x)}{f(x)}. \tag{3.68}$$

Letting $x \downarrow 0$, reduces (3.68) to the expected idle period

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

Also, from (3.68)

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

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

$$F(x) = P_0 e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x \geq 0, \tag{3.69}$$

$$f(x) = \frac{P_0}{E(b_x)} e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x > 0. \tag{3.70}$$

3.3.13 Sojourn Time Above a Level

Let a_x denote a virtual-wait sample-path sojourn time above level $x \geq 0$ (Fig. 3.6). Then $a_0 = \mathcal{B}$. By Theorem 1.1, for M/G/1 queues in equilibrium, the down- and upcrossing rates of x are both equal to $f(x)$. The proportion of time that a sample path spends above x is equal to $f(x)E(a_x)$ *and is also equal to* $1 - F(x)$. Therefore

$$E(a_x) = \frac{1 - F(x)}{f(x)}, x \geq 0. \tag{3.71}$$

Intuitively, $a_x \underset{stoch}{\leq} \mathcal{B}$ where " $\underset{stoch}{\leq}$ " means "stochastically less than or equal to", and $E(a_x) \leq E(\mathcal{B})$. Both inequalities seem to hold since the excess of an SP jump above x is, in general, stochastically less than a total service time. For $x = 0$, $E(a_0) = E(\mathcal{B})$. Proposition 3.1 below shows that if $E(a_x) = E(\mathcal{B})$ *for all* $x \geq 0$ then the absolutely continuous part of the pdf is exponentially distributed.

Proposition 3.1 *Assume* $\rho = \lambda E(S) < 1$.

(1) $E(a_0) = \frac{E(S)}{1-\lambda E(S)} = E(\mathcal{B})$.

(2) If $E(a_x) = E(\mathcal{B})) \equiv \frac{E(S)}{1-\lambda E(S)}$ for all $x \geq 0$, then the steady state cdf and pdf of wait are $F(x) = 1 - \rho e^{-\frac{x}{E(\mathcal{B})}}$ and

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

respectively.

Proof. (1) Letting $x \downarrow 0$ in (3.71) gives as in (3.59),

$$\begin{aligned} E(a_0) &= \frac{1 - F(0)}{f(0)} = \frac{1 - P_0}{\lambda P_0} \\ &= \frac{\lambda E(S)}{\lambda P_0} = \frac{E(S)}{1 - \lambda E(S)} = E(\mathcal{B}). \end{aligned}$$

(2) If $E(a_x) \equiv E(\mathcal{B}), x \geq 0$, then(from (3.71))

$$\frac{f(x)}{1 - F(x)} \equiv \frac{1}{E(\mathcal{B})}, x > 0, \tag{3.72}$$

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

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

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

Thus the cdf is

$$F(x) = 1 - \rho e^{-\frac{x}{E(\mathcal{B})}}, x \geq 0. \tag{3.73}$$

Differentiation of (3.73) with respect to $x > 0$ gives

$$f(x) = \lambda P_0 e^{-\frac{x}{E(\mathcal{B})}}, x > 0, \tag{3.74}$$

which is the absolutely continuous part of the pdf. ■

Remark 3.13 *The standard M/M/1 queue satisfies case (2) of Proposition 3.1. For M/M/1, the service time is exponentially distributed. Fix level $x > 0$. All jumps which start below level x and end above level x, have excess above x distributed as the exponential service time, by the memoryless property (discussed further in Section 3.4).*

Remark 3.14 *Note that (3.72) is the* **hazard rate** *(***failure rate***) of the pdf of wait.*

In addition to the two cases discussed in Proposition 3.1, we now show that $E(a_x) < E(\mathcal{B}), x > 0$, as intuitively expected. Note the role of the alternative form of the M/G/1 integral equation (3.35) in facilitating the proof.

Proposition 3.2 *Except for the two cases in Proposition 3.1,*

$$E(a_x) = \frac{1 - F(x)}{f(x)} < \frac{E(S)}{1 - \lambda E(S)} = E(\mathcal{B}), x > 0. \tag{3.75}$$

Proof. Cross multiplying in the inequality of (3.75) yields

$$1 - F(x) - \lambda E(S) + \lambda E(S)F(x) < E(S)f(x)$$

or

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

upon substituting for $f(x)$ from (3.35). Cancelling and rearranging terms, it is required to prove the following inequality holds:

$$1 + \lambda E(S) \int_{y=0}^{x} B(x-y)f(y)dy \quad < F(x) + \lambda E(S).$$

Note that $P_0 = 1 - \lambda E(S)$, $\lambda E(S) < 1$, $B(x-y) \leq 1$. Hence the left side

$$\begin{aligned} 1 + \lambda E(S) \int_{y=0}^{x} B(x-y)f(y)dy &< 1 + \int_{y=0}^{x} f(y)dy \\ &= 1 + (F(x) - P_0) \\ &= F(x) + \lambda E(S), \end{aligned}$$

as required. ■

3.3.14 Sojourn Above a Level and Distribution of Wait

The following relationship holds between the expected sojourn times $E(a_y), 0 < y < x$, and the steady-state cdf of wait $F(x)$. In general, $E(a_y)$ varies with y.

Proposition 3.3 *For the M/G/1 queue in equilibrium ($\rho = \lambda E(S) < 1$), the cdf of wait $F(x)$ is related to $E(a_y)$ the expected sojourn times of the virtual wait above level y,$0 < y < x$, by*

$$F(x) = 1 - \rho \cdot e^{-\int_{y=0}^{x} \frac{1}{E(a_y)} dy}, x \geq 0. \tag{3.76}$$

Proof. Consider a sample path of the virtual wait. The pdf of wait $f(x)$ is the SP upcrossing (and downcrossing) rate of level x. Hence the proportion of time the virtual-wait sample path spends above level x is

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

Thus (the hazard rate of $f(x)$ is)

$$\frac{f(x)}{1 - F(x)} = \frac{1}{E(a_x)}, x > 0. \tag{3.77}$$

Hence

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

Integrating with respect to x gives

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

Letting $x \downarrow 0$, the constant

$$A = 1 - F(0^+) = 1 - F(0) = \rho = \lambda E(S).$$

Hence we obtain (3.76). ■

Remark 3.15 *The term "hazard rate" is usually associated with positive continuous random variables (e.g., [50]). Here, we also use "hazard rate" for the non-negative waiting time (atom at* 0*).*

3.3.15 Hazard Rate of Steady-state Wait

Formula (3.77) is recognizable as the *hazard rate* of the steady-state random variable *wait* (see Remark 3.15). From it we can proceed in two different directions.

First, we may integrate with respect to x and get the expression for $F(x)$ given in (3.76).

Second, we may use simulation to estimate the hazard rate $\frac{f(x)}{1-F(x)}$ for various values of x with considerable accuracy. Fix $x > 0$. We simulate a single sample path of the virtual wait. Denote the successive sample-path sojourn times above level x by $a_{x1}, a_{x2}, ..., a_{xN}$. The simulated time is made sufficiently long such that N is "large". Then estimate $E(a_x)$ by the average simulated sojourn time

$$\widehat{E}(a_x) = \frac{1}{N} \sum_{j=1}^{N} a_{xj}.$$

Denote the hazard rate of wait at x by $\phi(x)$. From (3.77), a plausible estimate of $\phi(x)$ is

$$\widehat{\phi}(x) = \frac{1}{\widehat{E}(a_x)}. \tag{3.78}$$

By definition

$$\begin{aligned}\phi(x)dx &= P(W_q \in (x, x+dx) | W_q > x) \\ &= \frac{P(x < W_q < x + dx)}{P(W_q > x)},\end{aligned}$$

where W_q is the steady-state queue wait. Formula (3.77) suggests the following observation, which has an intuitive meaning. $\phi(x)$ varies inversely with $E(a_x)$. If the hazard rate at x is large then the $E(a_x)$ is small. If the hazard rate at x is small, then $E(a_x)$ is large.

The foregoing discussion suggests different avenues of investigation. One is an LC estimation method using simulated sample paths (see Chapter 9). Another is the relationship between hazard rates of state random variables and their sample-path expected sojourn times with respect to a level.

Example 3.4 *In the M_λ/Erlang-$(2,\mu)$/ 1 queue with arrival rate λ, expected service time $\frac{2}{\mu}$ and $\lambda \cdot \frac{2}{\mu} < 1$ (denoted by $M_\lambda/E_{2.\mu}/1$), consider a sample path of the virtual wait (see Example 3.1). The service time is distributed as an Erlang-$(2,\mu)$ random variable. The expected sojourn time above an arbitrary level $x > 0$ is equal to a busy period of the $M_\lambda/E_{2.\mu}/1$ queue, or to a busy period of the $M_\lambda/M_\mu/1$ queue, depending on the initial service-time phase that covers x, due to an SP jump upcrossimg of x. That is, the sojourn's initial SP upcrossing of x covers x either during phase 1 or during phase 2 of the Erlang-$(2,\mu)$ service time. If phase 1 covers x, then the excess jump above x is distributed as Erlang-$(2,\mu)$ (memoryless property of exponential). If phase 2 covers x, then the excess jump above x is distributed as an exponential r.v. with rate μ. Applying (3.60), for $M_\lambda/E_{k,\mu}/1$, we have $E(\mathcal{B}) = \frac{k}{\mu - k\lambda}$. For $M_\lambda/M_\mu/1$, $E(\mathcal{B}) = \frac{1}{\mu-\lambda}$ (formula (3.93) below). Thus,*

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

where $p_i(x) = P$(phase i of SP jump covers $x|SP$ upcrosses x),$i = 1, 2$. Thus from (3.76)

$$F(x) = 1 - \rho e^{-\left(\int_{y=0}^{x} \frac{1}{p_1(y)\left(\frac{2}{\mu-2\lambda}\right)+p_2(y)\left(\frac{1}{\mu-\lambda}\right)} dy\right)}. \tag{3.79}$$

In Example (3.1), equation (3.33) for $M/E_2/1$ yields

$$\begin{aligned} p_1(x) = &\quad \frac{\lambda\left(P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy\right)}{f(x)} \\ p_2(x) = &\quad 1 - p_1(x), \end{aligned} \tag{3.80}$$

in terms of $f(y)$ specified in (3.34).

We provide an LC intuitive interpretation of (3.80). Fix $x > 0$. Consider SP jumps that start below and end above x due to arrivals. The numerator of (3.80),

$$\lambda\left(P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy\right)$$

is the rate at which phase 1 of the service time covers x. From Theorem 1.1, the denominator $f(x)$ is the SP total upcrossing (and downcrossing) rate of level x. Thus $p_1(x)$ is the proportion of all upcrossings of x, which upcross x during phase 1.

Alternatively, we could estimate $p_1(x)$, $p_2(x), x > 0$, *from a simulated sample path of the virtual wait. Then substitute the estimated values into (3.79) to estimate $F(x), x > 0$. This **hybrid technique** combines estimated values from simulation and analytical results. Similar hybrid techniques may be applicable in various M/G/1 variants.*

3.3.16 Downcrossings During Inter-downcrossing Time

We state a proposition that gives the expected number of SP downcrossings of a level during an inter-downcrossing time of a different level, for a sample path of the virtual wait $\{W(t)\}$. The proof is given later in the discussion of M/M/1 queues in Section 3.4, as indicated in the "proof" part of the following proposition.

Proposition 3.4 *Consider the virtual wait $\{W(t), t \geq 0\}$ of an **M/G/1 queue** with $\lambda E(s) < 1$. Denote the steady-state pdf of wait by $f(x), x \geq 0$. Fix level $y \geq 0$ in the state space. Let $\mathcal{D}_{d_y}(x)$ denote the number of SP downcrossings of an **arbitrary** level x during a sample-path inter-downcrossing time of level y. Then*

$$E(\mathcal{D}_{d_y}(x)) = \frac{f(x)}{f(y)}. \tag{3.81}$$

Proof. The proof is given in Proposition 3.6, since it fits the context of Subsection 3.4.8 for M/M/1 queues. ■

3.3.17 Boundedness of Steady-state PDF

For M/G/1 with arrival rate λ and service time distribution $B(y)$,$y > 0$, assume the steady-state pdf of wait $f(x), x > 0$ exists.

Proposition 3.5

$$f(x) < \lambda, x > 0.$$

Proof. We present three proofs for perspective.

(1) In equation (1.8) (repeated here for convenience)

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

$\overline{B}(0) = 1$, $\overline{B}(x-y) = 1 - B(x-y) \leq 1, y > 0$. Assume $0 \leq F(x) < 1$. Then

$$\begin{aligned} f(x) &\leq \lambda P_0 + \lambda \int_{y=0}^{x} f(y)dy = \lambda\left(P_0 + \lambda \int_{y=0}^{x} f(y)dy\right) \\ &= \lambda F(x) < \lambda. \end{aligned}$$

If $F(x) = 1$ then $f(x) = \frac{d}{dx}F(x) = 0 < \lambda$ (In some models the wait will be concentrated on a finite interval $[0, M]$. Then $F(x) = 1, x \geq M$. Recall that $0 \leq F(x) \leq 1$, and $F(x)$ is right-continuous monotone non-decreasing.)

(2) Consider the alternative form of the LC integral equation (3.35) (repeated here)

$$f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} B(x-y) f(y)dy, x > 0. \tag{3.82}$$

On the right side of (3.82), the subtracted term is such that

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

From (3.82) $f(x) < \lambda F(x) < \lambda$.

(3) Consider a sample path of the virtual wait $\{W(t)\}$. Let $\mathcal{D}_t(x)$, $N_A(t)$ 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 implies $E(\mathcal{D}_t(x)) < E(N_A(t)), x \geq 0, t > 0$. Hence

$$f(x) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} < \lim_{t\to\infty} \frac{E(N_A(t))}{t} = \lambda,$$

since $\{N_a(t)\}$ is a Poisson process with rate λ. ■

Example 3.5 *In $M_\lambda/M_\mu/1$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0$, $P_0 = 1 - \frac{\lambda}{\mu} > 0$ (Subsection 3.4.1). Both $P_0 < 1$ and $e^{-(\mu-\lambda)x} < 1, x > 0$. Arrival rate λ is a conservative upper bound for $f(x)$ since*

$$f(x) < \lambda P_0,\ f(x) < \lambda e^{-(\mu-\lambda)x},\ f(x) < \lambda, x > 0$$

and $f(0) = \lambda P_0$.

3.4 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, μ = service rate, and traffic intensity $\rho = \frac{\lambda}{\mu} < 1$.

3.4.1 Waiting Time

Consider a sample path of the virtual wait (e.g., Fig. 3.4). From rate balance of SP down- and upcrossings of level x as in Fig. 1.6 (or (3.29)), we obtain

$$f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy, \ x > 0. \tag{3.83}$$

where $\{P_0; f(x), x > 0\}$ is the steady state pdf of wait and $B(x) = 1 - e^{-\mu x}, x \geq 0$.

Differentiating both sides with respect to x, yields the ordinary differential equation

$$f'(x) + (\mu - \lambda) f(x) = 0, x > 0, \tag{3.84}$$

with solution

$$f(x) = Ae^{-(\mu-\lambda)x}, x > 0; \tag{3.85}$$

constant A is determined by letting $x \downarrow 0$ in both (3.83) and (3.85). Thus $A = f(0^+) = \lambda P_0$. The pdf of wait is

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0, \tag{3.86}$$

where

$$P_0 = 1 - \lambda E(S) = 1 - \frac{\lambda}{\mu} = 1 - \rho, \tag{3.87}$$

(e.g., (3.46)). We may also compute P_0 by substituting (3.86) into the normalizing condition,

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

which yields $P_0 = 1 - \rho$ directly.

The cdf of wait is

$$\begin{aligned} F(x) &= P_0 + \textstyle\int_{y=0}^{x} \lambda(1-\rho)e^{-(\mu-\lambda)y}dy \\ &= 1 - \rho e^{-(\mu-\lambda)x}, x \geq 0. \end{aligned} \tag{3.88}$$

3.4.2 System Time

Let σ denote the steady-state system time, $f_\sigma(x)$ its pdf, $F_\sigma(x)$ its cdf, $x > 0$. Since $\sigma = W_q + S$, we obtain

$$\begin{aligned} P(\sigma > x) \quad &= P_0 e^{-\mu x} + \lambda P_0 \int_{y=0}^{x} e^{-(\mu-\lambda)y} e^{-\mu(x-y)} dy \\ &+\lambda P_0 \int_x^\infty e^{-(\mu-\lambda)y} dy, \ x \geq 0. \\ &= \frac{P_0}{1-\frac{\lambda}{\mu}} e^{-(\mu-\lambda)x} \\ &= e^{-(\mu-\lambda)x}. \end{aligned} \tag{3.89}$$

(We can obtain (3.89) using (3.54).)
Thus σ is exponentially distributed with mean $\frac{1}{\mu-\lambda}$, i.e.,

$$\begin{aligned} f_\sigma(x) &= (\mu-\lambda)\, e^{(\mu-\lambda)x}, x > 0 \\ F_\sigma(x) &= 1 - e^{(\mu-\lambda)x}, x \geq 0. \end{aligned} \tag{3.90}$$

We can also obtain $f_\sigma(x)$ directly in terms of $f(x)$ using LC, as in (3.55) and (3.56). Thus we obtain (3.90) as in Example 3.3 above.

3.4.3 Number in System

Let N denote the number of units in the system at an arbitrary time point in the steady state. Let $P(N = n) = P_n, n = 0, 1, \ldots$. Let $d_n = P(n$ units in system just after a departure). We obtain the distribution of N by conditioning on W_q, or on σ, providing two additional ways of deriving P_0 for M/M/1 (see Subsection 3.3.7). (Recall $\rho = \frac{\lambda}{\mu}$.)

First, conditioning on W_q,

$$\begin{aligned} P_n = d_n &= \int_{y=0}^\infty e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!} \lambda P_0 e^{-(\mu-\lambda)y} dy \\ &= P_0 \left(\frac{\lambda}{\mu}\right)^n \int_{y=0}^\infty e^{-\mu y} \frac{(\mu y)^{n-1}}{(n-1)!} \mu dy \\ &= P_0 \rho^n, n = 0, 1, \ldots . \end{aligned}$$

The normalizing condition $\sum_{n=0}^\infty P_n = 1$ yields

$$P_0(1 + \rho + \rho^2 \ldots) = 1,$$

whence $P_0 = 1 - \rho$, giving the well known geometric distribution

$$P_n = P_0\,(1 - P_0)^n = (1-\rho)\rho^n, n = 0, 1, \ldots . \tag{3.91}$$

Second, conditioning on σ,

$$\begin{aligned} P_n = d_n &= \int_{y=0}^{\infty} e^{-\lambda y} \frac{(\lambda y)^n}{(n)!} (\mu - \lambda) e^{-(\mu-\lambda)y} dy \\ &= \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n = (1-\rho)\rho^n, n = 0, 1, ..., \end{aligned}$$

(same as (3.91)).

Note that $P(N \geq n) = \rho^n, n = 0, \ldots$. Thus

$$E(N) = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} \rho^n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}. \tag{3.92}$$

3.4.4 Expected Busy Period

This subsection very brief, but important due to the key role of busy periods in queueing theory. The $M_\lambda/M_\mu/1$ queue is an $M_\lambda/G/1$ queue having exponential service S with $E(S) = \frac{1}{\mu}$. Substituting into (3.59) gives the well known result

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu\left(1 - \frac{\lambda}{\mu}\right)} = \frac{1}{\mu - \lambda}. \tag{3.93}$$

3.4.5 Geometric Derivation of CDF and PDF of Wait

Consider a sample path of the virtual wait of the M/M/1 queue. Given that the SP upcrosses level x, the resulting sojourn time above x is distributed as a busy period $\mathcal{B}$ independent of x, due to the memoryless property of the service time (Fig. 3.7). (See also Subsection 1.5.2, paragraph following "Key Question".) Therefore the long-run proportion of time that the sample path spends above x, is

$$\left(\lim_{t \to \infty} \frac{\mathcal{U}_t(x)}{t}\right) E(\mathcal{B}) = f(x) E(\mathcal{B}).$$

It is also equal to $1 - F(x)$. Thus

$$f(x)E(\mathcal{B}) = 1 - F(x), x > 0, \tag{3.94}$$

$$\frac{f(x)}{1 - F(x)} = \frac{1}{E(\mathcal{B})}. \tag{3.95}$$

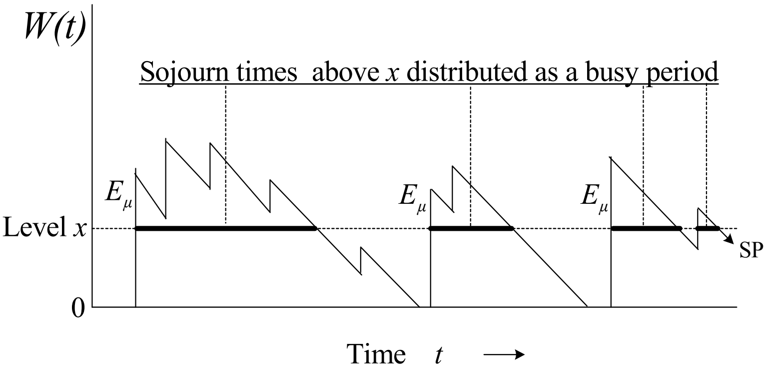

Figure 3.7: Sample path of virtual wait for $M_\lambda/M_\mu/1$ queue showing sojourns above level $x \underset{dist}{=} \mathcal{B}$ SP excess jumps above x are $\underset{dist}{\equiv} E_\mu$.

Hence

$$\frac{d}{dx}\ln(1 - F(x)) = -\frac{1}{E(\mathcal{B})}. \tag{3.96}$$

Integrating (3.96) with respect to x, letting $x \downarrow 0$ to compute the constant of integration, and using (3.93), gives the cdf of wait

$$F(x) = 1 - \rho e^{-\frac{x}{E(\mathcal{B})}} = 1 - \rho e^{-(\mu-\lambda)x}, x \geq 0. \tag{3.97}$$

Taking $\frac{d}{dx}$ in (3.97) gives the pdf of wait

$$f(x) = \lambda(1-\rho)e^{-(\mu-\lambda)x} = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0. \tag{3.98}$$

Note that (3.97) and (3.98) can be obtained immediately from Proposition 3.3. That is, for M/M/1, $E(a_{y)} \equiv E(\mathcal{B}), y \geq 0$. Thus

$$-\frac{1}{E(a_{y)}} = -\frac{1}{E(\mathcal{B})} = -(\mu - \lambda), y \geq 0,$$

and substituting into (3.76) yields (3.97). The M/M/1 model satisfies Case (2) of Proposition 3.1.

3.4.6 Inter-crossing Time of a Level

This subsection discusses the time between SP successive downcrossings (inter-downcrossing time) and between successive upcrossings (inter-upcrossing time) of a level. It also considers the expected number of SP crossings of a level during a busy cycle, and during sojourns above or below an arbitrary level.

Inter-downcrossing Time of a Level

Consider the virtual wait $\{W(t)\}$ and fix state-space level $x \geq 0$. Let d_x = SP inter-downcrossing time of level x, b_x = sojourn time at or below x, a_x = sojourn time above x. Then

$$d_x = b_x + a_x, \;\; E(d_x) = E(b_x) + E(a_x).$$

In M/M/1 both inter-arrival and service times are exponentially distributed. For fixed $x \geq 0$, successive triplets $\{d_x, b_x, a_x\}$ form a sequence of iid random variables. Thus $\{d_x\}$ forms a renewal process, $\{b_x, a_x\}$ form an alternating renewal process, and

$$\left.\begin{array}{rl} E(d_x) = & \frac{1}{f(x)}, \\ E(b_x) = & \frac{F(x)}{f(x)} \\ E(a_x) = & \frac{1-F(x)}{f(x)} \end{array}\right\} \tag{3.99}$$

For all $x \geq 0$, $a_x \underset{dist}{=} \mathcal{B}$. Thus

$$E(a_x) = \frac{1}{\mu - \lambda}, x \geq 0. \tag{3.100}$$

Hence,

$$E(d_x) = \frac{F(x)}{f(x)} + \frac{1}{\mu - \lambda}, x \geq 0. \tag{3.101}$$

Letting $x = 0$ in (3.101) gives the expected busy cycle

$$\begin{aligned} E(d_0) &= \frac{F(0)}{f(0)} + \frac{1}{\mu - \lambda} = \frac{P_0}{\lambda P_0} + \frac{1}{\mu - \lambda} \\ &= \frac{1}{f(0)} = \frac{1}{\lambda(1-\rho)}. \end{aligned} \tag{3.102}$$

We obtain the expected inter-downcrossing time of level x by substituting $f(x)$ from (3.98) into (3.101). Thus

$$E(d_x) = \frac{1}{f(x)} = \frac{e^{(\mu-\lambda)x}}{\lambda(1-\rho)}, x \geq 0. \tag{3.103}$$

Thus $E(d_x)$ increases exponentially with x (Fig. 3.8).

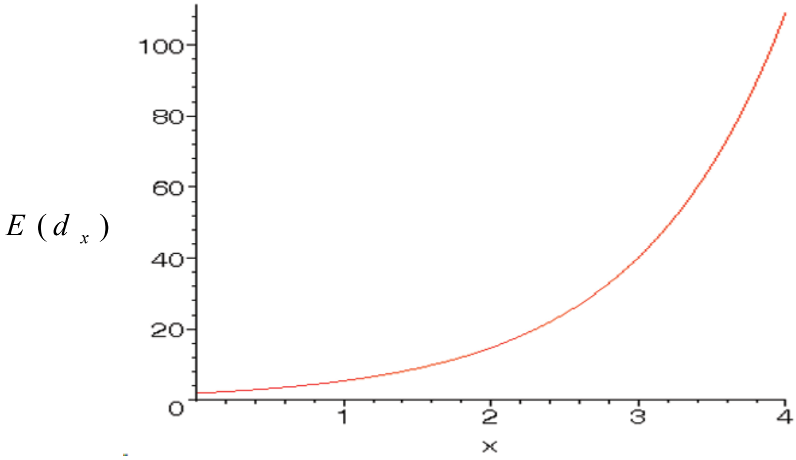

Figure 3.8: Expected inter-downcrossing (or inter-upcrossing) time of level x, $E(d_x)$ (or $E(u_x)$) in M/M/1: $\lambda = 1.0$, $\mu = 2.0$, $\rho = 0.5$.

Inter-upcrossing Time of a Level

Denote the inter-upcrossing time of level x by u_x. Inspection of sample paths of the virtual wait process, indicates that $u_x \underset{dist}{=} d_x$ due to the memoryless property of both the inter-arrival and service times in M/M/1. Hence $E(u_x)$ also increases exponentially with x, and the plot of $E(u_x)$ versus x is identical to that of $E(d_x)$ versus x (Fig. 3.8).

3.4.7 Number of Crossings of a Level in a Busy Cycle

Note that $d_0 =$ *busy cycle*..Denote the number of downcrossings of level $x \geq 0$ during d_0 by $\mathcal{D}_{d_0}(x)$. Since $\mathcal{D}_t(x)$ is the number of downcrossings of x during time interval $(0, t)$, from the theory of regenerative processes (e.g., [96])

$$\begin{aligned} \frac{E(\mathcal{D}_{d_0}(x))}{E(d_0)} &= \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} \\ &= f(x) = \lambda(1-\rho)e^{-(\mu-\lambda)x}, x \geq 0. \end{aligned}$$

Hence,

$$\begin{aligned} E(\mathcal{D}_{d_0}(x)) = &\quad \lambda(1-\rho)e^{-(\mu-\lambda)x} \cdot E(d_0) \\ = &\quad \lambda(1-\rho)e^{-(\mu-\lambda)x} \cdot \frac{1}{\lambda(1-\rho)} = e^{-(\mu-\lambda)x}, x \geq 0. \end{aligned} \quad (3.104)$$

Thus, $E(\mathcal{D}_{d_0}(x)) \leq 1$. From (3.104), $E(\mathcal{D}_{d_0}(x))$ decreases exponentially as x increases.

Let $\mathcal{U}_{d_0}(x)$ denote the number of upcrossings of level x during a *busy cycle*. Note that $\mathcal{D}_{d_0}(x) = \mathcal{U}_{d_0}(x), x \geq 0$. Thus from (3.104)

$$E(\mathcal{D}_{d_0}(0)) = E(\mathcal{U}_{d_0}(0)) = \lim_{x \downarrow 0} e^{-(\mu-\lambda)x} = 1. \tag{3.105}$$

Equation (3.105) is intuitive, since the SP hits level 0 from above and egresses from level 0 above (upcrosses 0) exactly once during a busy cycle. The SP hit occurs at the end of the embedded *busy period*. The SP egress occurs at the start of the embedded busy period.

3.4.8 Downcrossings at Different Levels

M/M/1

Consider a fixed level $y \geq 0$ and a fixed level $x > y$. SP downcrossings of x can occur only during an SP sojourn above y. In M/M/1 $a_y \underset{dist}{=} a_0 = \mathcal{B}, y \geq 0$. SP motion above level y is analogous to SP motion above level 0. Let $\mathcal{D}_{a_y}(x), x > y$ denote the number of downcrossings of x during an SP sojourn above y. We obtain an expression for $E(\mathcal{D}_{a_y}(x))$. Substituting y for 0 in (3.104) leads to

$$\begin{aligned} E(\mathcal{D}_{a_y}(x)) = e^{-(\mu-\lambda)(x-y)} &= \frac{e^{-(\mu-\lambda)x}}{e^{-(\mu-\lambda)y}} \\ &= \frac{E(\mathcal{D}_{d_0}(x))}{E(\mathcal{D}_{d_0}(y))}. \end{aligned}$$

Equivalently

$$E(\mathcal{D}_{d_0}(x)) = E(\mathcal{D}_{d_0}(y)) \cdot E(\mathcal{D}_{a_y}(x)). \tag{3.106}$$

Equation (3.106) can also be derived from

$$\mathcal{D}_{d_0}(x) = \sum_{i=1}^{\mathcal{U}_{d_0}(y)} \mathcal{D}^i_{a_y}(x), \tag{3.107}$$

where $\mathcal{D}^i_{a_y}(x) \underset{dist}{\equiv} \mathcal{D}_{a_y}(x)$ and $\{\mathcal{D}^i_{a_y}(x)\}$ are iid independent of $\mathcal{U}_{d_0}(y)$. In equation (3.107) $\mathcal{D}_{d_0}(x)$ is the total number of SP downcrossings of x during a busy cycle d_0. The upper limit of the sum $\mathcal{U}_{d_0}(y)$ is the number of SP sojourns above y during a busy cycle, since each upcrossing of y initiates a sojourn above y. The term $\mathcal{D}^i_{a_y}(x)$ is the number of SP downcrossings of x during the i^{th} sojourn above y in the busy cycle. Thus the sum is the total number of SP downcrossings of x during the busy cycle d_0. Each downcrossing of x can occur only during an SP sojourn above y. A sojourn time above y is distributed as a_y (same as $\mathcal{B}$).

Note that $\mathcal{U}_{d_0}(y) = \mathcal{D}_{d_0}(y)$ and $E(\mathcal{U}_{d_0}(y)) = E(\mathcal{D}_{d_0}(y))$, since *during a busy cycle* the number of SP down- and upcrossings of an arbitrary level y are equal. Thus, taking expected values in (3.107) gives

$$\begin{aligned} E(\mathcal{D}_{d_0}(x)) &= E(\mathcal{U}_{d_0}(y)) \cdot E(\mathcal{D}_{a_y}(x)) \\ &= E(\mathcal{D}_{d_0}(y)) \cdot E(\mathcal{D}_{a_y}(x)), \end{aligned}$$

which is the same as (3.106).

Generalization to M/G/1 Queues

We generalize the foregoing results for M/M/1 as follows, to **M/G/1** (see Subsection 3.3.16). Let $\mathcal{D}_{d_y}(x)$ denote the number of SP downcrossings of an **arbitrary** level x during a sample-path inter-downcrossing time of level y (may have $x \geq y$, or $x < y$ if $y > 0$).

Proposition 3.6 *Consider the virtual wait $\{W(t), t \geq 0\}$ of an **M/G/1 queue** with $\lambda E(S) < 1$. Denote the steady-state pdf of wait by $f(x), x \geq 0$. Fix level $y \geq 0$ in the state space. Then*

$$E(\mathcal{D}_{d_y}(x)) = \frac{f(x)}{f(y)}, x \geq 0. \tag{3.108}$$

Proof. Fix level $y \geq 0$. Due to system stability and Poisson arrivals, without loss of generality we may assume the sample-path inter-downcrossing times of level y, $\{d_{yi}, i = 1, 2, ...\}$ form a renewal process. The $\{d_{yi}\}$ are iid r.v.'s. Let $d_{yi} \underset{dist}{=} d_y, i = 1, 2, ...$. Fix arbitrary level $x \geq 0$, *independent* of y. The regenerative cycle of length d_y is a probabilistic replica of the process $\{W(t), t \geq 0\}$ at level y over the entire time line. Let $\mathcal{D}_{d_y}(x)$ denote the number of SP downcrossings of level x during d_y. From regenerative processes,

$$\frac{E(\mathcal{D}_{d_y}(x))}{E(d_y)} \equiv \lim_{t \to \infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x), x \geq 0 \tag{3.109}$$

for each level $y \geq 0$. From renewal theory and the basic LC theorem for M/G/1 (Theorem 1.1), $E(d_y) = \frac{1}{f(y)}$. Thus

$$\begin{aligned} E(\mathcal{D}_{d_y}(x)) &= E(d_y) \cdot f(x) \\ &= \frac{f(x)}{f(y)}, \end{aligned}$$

which is the same as (3.108). (If $y = 0$, (3.108) holds, except $x \geq 0$.) ■

Corollary 3.7 *For the M/M/1 queue,*

$$E(\mathcal{D}_{d(y)}(x)) = e^{-(\mu-\lambda)(x-y)}, \ x \geq 0, \ y \geq 0. \tag{3.110}$$

Proof. In M/M/1, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x \geq 0$. ■

From (3.110), in the M/M/1 queue

$$E(\mathcal{D}_{d_y}(x)) \begin{cases} < 1 \text{ if } x > y \\ = 1 \text{ if } x = y \\ > 1 \text{ if } x < y \end{cases}. \tag{3.111}$$

Setting $x = y$ in (3.111) shows that the expected number of SP downcrossings of x during an inter-downcrossing time of x is

$$E(\mathcal{D}_{d_x}(x)) = e^{-(\mu-\lambda)(x-x)} = 1, x \geq 0,$$

in agreement with intuition. Examination of a sample path of the virtual wait corroborates this fact.

Corollary 3.8 *For levels $x, y, y_1, y_2, ..., y_n$ in the state space $\boldsymbol{S}$,*

$$\begin{aligned} &E(\mathcal{D}_{d_y}(x)). \\ &= E(\mathcal{D}_{d_y}(y_1)) \cdot E(\mathcal{D}_{d_{y_1}}(y_2)) \cdots \ E(\mathcal{D}_{d_{y_{n-1}}}(y_n)) \cdot E(\mathcal{D}_{d_{y_n}}(x)) \end{aligned} \tag{3.112}$$

Proof. From (3.108) we obtain

$$\begin{aligned} E(\mathcal{D}_{d_y}(x)) &= \frac{f(x)}{f(y)} \\ &= \frac{f(y_1)}{f(y)} \cdot \frac{f(y_2)}{f(y_1)} \cdots \frac{f(y_n)}{f(y_{n-1})} \cdot \frac{f(x)}{f(y_n)} \end{aligned}$$

which is equivalent to (3.112). ■

Remark 3.16 *The results in (3.108) and (3.112) hold for the standard M/G/1 queue, since the proofs depend only on having a Poisson arrival process. In order to apply (3.108) and (3.112) to a specific M/G/1 queue, it is necessary to have a formula for $f(x)$. The pdf $f(x)$ is known in many M/G/1 models (e.g., M/D/1, $ME_k/1$ and variants); if necessary $f(x)$ can be approximated or estimated by a variety of means.*

3.4.9 Number Served in a Busy Period

Substituting $E(S) = \frac{1}{\mu}$ in (3.63), gives

$$\begin{aligned} E\left(\sum_{i=1}^{N_{\mathcal{B}}}(\frac{1}{\mu} - T_i)\right) &= -\frac{1}{\lambda}, \\ E(N_{\mathcal{B}})\left(\frac{1}{\mu} - \frac{1}{\lambda}\right) &= -\frac{1}{\lambda}, \end{aligned}$$

yielding

$$E(N_{\mathcal{B}}) = \frac{\mu}{\mu - \lambda} = \frac{1}{P_0}. \tag{3.113}$$

as in (3.65). (See also (3.64).)

Writing $N_{\mathcal{B}} = \min(n | \sum_{i=1}^{n} S_i \leq \sum_{i=1}^{n} T_i)$, shows that $N_{\mathcal{B}}$ is a stopping time for both sequences $\{S_i\}$ and $\{T_i\}$ as mentioned following (3.65). Then

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

and E(busy cycle) is

$$E(d_0) = E\left(\sum_{i=1}^{N_{\mathcal{B}}} T_i\right) = E(N_{\mathcal{B}})E(T) = E(N_{\mathcal{B}})\frac{1}{\lambda} = \frac{\mu}{\lambda(\mu - \lambda)}.$$

The last two equations both lead to (3.113).

The role of LC, is that the downcrossing rate of level 0 (left-continuous hit rate from above) is $f(0) = \lambda P_0$, and $E(d_0) = \frac{1}{f(0)}$. Using this fact and applying the stopping time criterion for a busy cycle, leads to the value of $E(N_{\mathcal{B}})$.

Remark 3.17 *Consider a sample path of the virtual wait for M/M/1. Subsection 5.1.14 discusses the number of system times above or below*

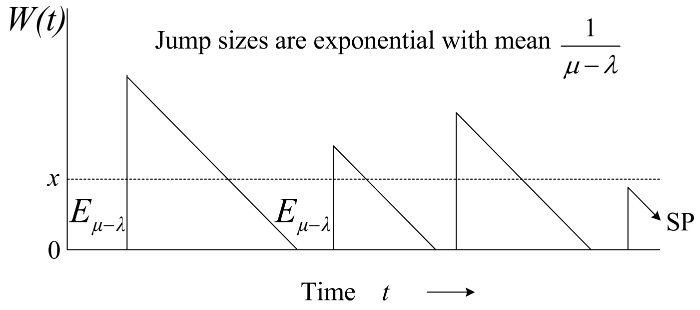

Figure 3.9: Sample path of workload for M/M/1/1 queue with arrival rate λ and service rate $\mu - \lambda$. Blocked customers are cleared.

a state-space level, during a sojourn time above or below that level. It also discusses the number of system times above or below a level, during a busy period. It similarly considers the number of waiting times. The results are presented in Subsection 5.1.14 because they follow as a special case of related results for G/M/1, given in subsections 5.1.12 and 5.1.13.

3.4.10 Relationship Between M/M/1 and M/M/1/1

The M/M/1/1 queue is an M/M/1 variant having capacity 1. Only one customer is allowed to be in the system. Customers that arrive when the server is busy, are blocked and cleared. Compare the virtual wait process for M/M/1 (Fig. 3.7) and the workload process for M/M/1/1 (Fig. 3.9). The LC approach immediately connects the two models in steady-state. The cdf (3.97) and pdf (3.98) of *wait* in the $M_\lambda/M_\mu/1$ (arrival rate λ, service rate μ), are respectively *identical to* the steady-state cdf and pdf of *workload* in the $M_\lambda/M_{\mu-\lambda}/1/1$ (arrival rate λ, ***service rate*** $\boldsymbol{\mu - \lambda}$).

This identicalness is evident from a sample path of the workload in $M_\lambda/M_{\mu-\lambda}/1/1$ (Fig. 3.9). 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 $\lambda P_0 e^{-(\mu-\lambda)x}$, since *all* SP jumps start at level 0, and are distributed as $E_{\mu-\lambda}$. In *both* M/M/1 and M/M/1/1, $E(\mathcal{B}) = \frac{1}{\mu-\lambda}$ and $P_0 = 1 - \frac{\lambda}{\mu}$. In $M_\lambda/M_{\mu-\lambda}/1/1$, the busy period $\mathcal{B}$ and the blocking time are identical, having exponential pdf $(\mu - \lambda)e^{-(\mu-\lambda)x}, x > 0$. The $M_\lambda/M_{\mu-\lambda}/1/1$ workload has the same distribution as the 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 of this subsection is that the pdf of workload for $M_\lambda/M_{\mu-\lambda}/1/1$

is derived *by inspection in one line*, since all SP jumps start at level 0.

The foregoing relationship suggests re-examining integral equation (3.83). We substitute the $\mathrm{M}_\lambda/\mathrm{M}_{\mu-\lambda}/1/1$ solution in the integral, namely

$$f(y) = \lambda P_0 e^{-(\mu-\lambda)y},$$

and simplify. The immediate result is the solution for the $\mathrm{M}_\lambda/\mathrm{M}_\mu/1$ model

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x},$$

obtained while bypassing differential equation (3.84). This solution for $\mathrm{M}_\lambda/\mathrm{M}_{\mu-\lambda}/1/1$ "solves" integral equation (3.83) for $\mathrm{M}_\lambda/\mathrm{M}_\mu/1$.

This solution procedure suggests exploring conditions that facilitate solving for the steady state 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 simpler sample-path jump structure.

3.5 M/G/1 with Service 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 the conditional cdf of $S(W_q)$ be $P(S(W_q) \leq x\,|W_q = y|) = B(x,y), x \geq 0, y \geq 0$, having pdf $b(x,y) = \frac{\partial}{\partial x}B(x,y), x > 0, y \geq 0$, 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 the virtual wait 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).

3.5.1 Integral Equation for PDF of Wait

Consider a fixed state-space level $x \geq 0$. The downcrossing rate of x is $f(x)$, by Theorem 1.1. The upcrossing rate of x is

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

the term $\lambda P_0 \overline{B}(x,0)$ is the upcrossing rate of x by SP jumps at arrival instants when the system is empty. The term $\lambda \int_{y=0}^{x} \overline{B}(x-y,y)f(y)dy$ is the upcrossing rate of x by SP jumps at arrival instants when the virtual wait is at state-space levels $y \in (0, x)$. Rate balance across level x yields the integral equation for $f(x)$,

$$f(x) = \lambda P_0 \overline{B}(x,0) + \lambda \int_{y=0}^{x} \overline{B}(x-y,y) f(y)dy,\ x \geq 0. \tag{3.114}$$

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 (3.114) with respect to x over $(0, \infty)$ gives

$$\begin{aligned} 1 - P_0 &= \rho_0 P_0 + \int_{y=0}^{\infty} \rho_y f(y)dy, \\ P_0 &= \frac{1 - \int_{y=0}^{\infty} \rho_y f(y)dy}{1 + \rho_0}, \end{aligned} \tag{3.115}$$

where $\rho_y \equiv \lambda E(S(y)), y \geq 0$. (Note that (3.115) is an implicit formula for P_0, since the integral contains P_0 implicitly. See (3.119) below.)

Consider a partition of the state space $\{x_i, i = 0, .., M+1\}$, where integer $M \geq 0$, and

$$0 \equiv x_0 < x_1 < x_2 < ... < x_M < x_{M+1} \equiv \infty.$$

Denote the service time of a zero-waiting customer by S_0, and of a y-waiting customer, $y \in (x_i, x_{i+1}]$, by S_i. Assume the service-time distribution is the same for all customers who wait zero; and the same for all customers that wait a time within the same state-space subinterval. Thus the cdf of service time is

$$\begin{aligned} &B_0(x) = B(x,0), x > 0 \\ &B_i(x) = B(x,y), x > 0, x_{i-1} < y \leq x_i, i = 1, ..., M+1. \end{aligned} \tag{3.116}$$

Integral equation (3.114) can be written

$$\begin{aligned} f(x) = \ & \lambda P_0 \overline{B_0}(x) + \lambda \textstyle\sum_{i=1}^{j-1} \int_{y=x_{i-1}}^{x_i} \overline{B_i}(x-y) f(y)dy \\ & + \lambda \textstyle\int_{y=x_{j-1}}^{x} \overline{B_j}(x-y) f(y)dy, x \in (x_{j-1}, x_j], j = 1, ..., M+1. \end{aligned} \tag{3.117}$$

where $\sum_{i=1}^{0} \equiv 0$. We have constructed integral equation (3.117) in an easy, intuitive, straightforward manner using LC.

Queues with service time depending on wait appear in [41]. A related theorem is given in [42]. The model was solved in the literature using Laplace transforms [81], and also by the embedded Markov chain technique using a Lindley recursion in [88].

Remark 3.18 *Deriving (3.117) using the embedded Markov chain technique is "relatively" tedious and purely algebraic (see Section 1.3). The model was generalized to multiple servers using the embedded Markov chain technique in [34] and [35] (*original *topic of my PhD thesis). After my discovery of LC in 1974, the model was re-solved using LC [7]. A two-server analysis is given in [39]; a revised version is given in Section 4.11 below.*

3.5.2 M/G/1: Zero-waits Receive Special Service

In the case where the first customer of every busy period receives specialized service, we have $M = 0$, $x_0 = 0$, $x_1 = \infty$ (M defined in 3.116) The integral equation (3.117) reduces to

$$f(x) = \lambda P_0 \overline{B_0}(x) + \lambda \int_{y=0}^{\infty} \overline{B_1}(x-y) f(y) dy, x \geq 0. \tag{3.118}$$

Integrating (3.118) with respect to x over $(0, \infty)$ and noting

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

gives

$$P_0 = \frac{1 - \lambda E(S_1)}{1 - \lambda E(S_1) + \lambda E(S_0)} = \frac{1 - \rho_1}{1 - \rho_1 + \rho_0}. \tag{3.119}$$

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 $P_0 > 0$. But $1 - \rho_1 + \rho_0 < 0$ would imply $P_0 > 1$, which is impossible. If $\rho_1 = 1$, then $P_0 = 0$,which would imply the queue is unstable.)

Multiplying both sides of (3.118) by x, and integrating for $x \in (0, \infty)$ gives a Pollaczek-Khinchin (P-K)-like result for the expected wait before service

$$E(W_q) = \frac{\lambda(E(S_0^2) + E(S_1^2))}{2(1 - \lambda E(S_1))}. \tag{3.120}$$

Expected Busy Period When $M = 0$

Customers that wait 0 have service time S_0. Customers that wait a positive time have service time S_1.

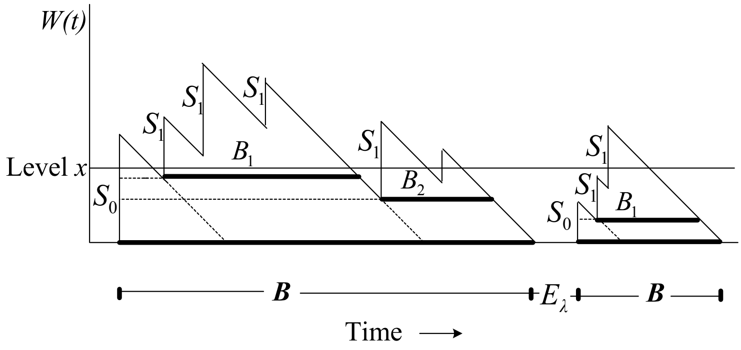

Figure 3.10: Busy periods $\{\mathcal{B}\}$ in M_λ/G/1 with zero-waits receiving service time $\underset{dist}{=} S_0$. $\{\mathcal{B}_i\}$ are busy periods of M_λ/G/1 with all service times $\underset{dist}{=} S_1$, generated by arrivals (as if) during S_0. In figure $B \equiv \mathcal{B}$.

Method 1 The busy period is

$$\mathcal{B} = S_0 + \sum_{i=1}^{N_{S_0}} \mathcal{B}_{1i} \tag{3.121}$$

where $N_{S_0} \underset{dist}{=}$ the number of arrivals during the first service time of a busy period, and the $\mathcal{B}_{1i}$'s are iid 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. 3.10). Taking the expected value in (3.121) gives

$$\begin{aligned} E(\mathcal{B}) &= E(S_0) + \lambda E(S_0) E(\mathcal{B}_1) \\ &= E(S_0) + \lambda E(S_0) \frac{E(S_1)}{1 - \lambda E(S_1)} \\ &= \frac{E(S_0)}{1 - \lambda E(S_1)}. \end{aligned} \tag{3.122}$$

Method 2 Applying the LC-based result for the expected busy period (3.60), we get using (3.119)

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = \frac{1 - \frac{1-\lambda E(S_1)}{1-\lambda E(S_1)+\lambda E(S_0)}}{\lambda \frac{1-\lambda E(S_1)}{1-\lambda E(S_1)+\lambda E(S_0)}} = \frac{E(S_0)}{1 - \lambda E(S_1)}.$$

Remark 3.19 *We may derive the expression for P_0 directly using the expression for $E(\mathcal{B})$. This serves as a further check on the solution.*

Thus

$$P_0 = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + E(\mathcal{B})} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + \frac{E(S_0)}{1-\lambda E(S_1)}}$$
$$= \frac{1 - \lambda E(S_1)}{1 - \lambda E(S_1) + \lambda E(S_0)}$$
$$= \frac{1 - \rho_1}{1 - \rho_1 + \rho_0}.$$

Example 3.6 *Let the service times be* ***exponentially distributed****, i.e.,* $B_0(x) = 1 - e^{-\mu_0 x}$, $B_1(x) = 1 - e^{-\mu_1 x}$. *Substitute for* $B_0(x)$, $B_1(x-y)$ *in (3.118) and apply differential operator* $\langle D + \mu_0 \rangle \langle D + \mu_1 \rangle$ *(equivalent to differentiating twice with respect to x, followed by some algebra) to yield a second order differential equation*

$$\langle D + \mu_1 - \lambda \rangle \langle D + \mu_0 \rangle f(x) = 0,$$

with solution

$$f(x) = ae^{-(\mu_1 - \lambda)x} + be^{-\mu_0 x}, \; x > 0,$$

provided $\mu_0 \neq \mu_1 - \lambda$ *(if* $\mu_0 = \mu_1 - \lambda$, $f(x)$ *has a different solution). 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

$$a = \frac{-\lambda^2 P_0}{(\mu_1 - \mu_0 - \lambda)}, \; b = \frac{\lambda(\mu_1 - \mu_0) P_0}{(\mu_1 - \mu_0 - \lambda)}, \; P_0 = \frac{(1 - \rho_1)}{(1 - \rho_1 + \rho_2)},$$

where $\rho_i = \frac{\lambda}{\mu_i}, i = 1, 2.$

Expected Busy Period *The expected busy period is, from (3.122),*

$$E(\mathcal{B}) = \frac{\frac{1}{\mu_0}}{1 - \frac{\lambda}{\mu_1}} = \frac{\mu_1}{\mu_0(\mu_1 - \lambda)}.$$

(If $\mu_0 = \mu_1 = \mu$, *then* $E(\mathcal{B}) = \frac{1}{\mu - \lambda}$, *as in the standard* $M_\lambda/M_\mu/1$ *queue.)*

3.6 M/G/1 with Multiple Poisson Inputs

Assume customers arrive at a single-server system in N independent Poisson streams at rates λ_i, $i = 1, ..., N$, $\sum_{i=1}^{N} \lambda_i = \lambda$. 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, S, of an arbitrary arrival is S_i with probability $\frac{\lambda_i}{\lambda}$. 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.

We may view the system as an M/G/1 queue with arrival rate λ and service time

$$S = \begin{cases} S_1 \text{ with probability } \frac{\lambda_1}{\lambda}, \\ S_2 \text{ with probability } \frac{\lambda_2}{\lambda}, \\ \cdots \\ S_N \text{ with probability } \frac{\lambda_N}{\lambda}. \end{cases}.$$

Hence $E(S) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i)$, $E(S^2) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i^2)$ and

$$P_0 = 1 - \lambda E(S) = 1 - \sum_{i=1}^{N} \lambda_i E(S_i) = 1 - \sum_{i=1}^{N} \rho_i. \tag{3.123}$$

where $\rho_i = \lambda_i E(S_i)$.

Stability

The system is stable iff every typical sample path of the virtual wait returns to state $\{0\}$; i.e., iff $P_0 > 0$ or

$$\sum_{i=1}^{N} \rho_i < 1. \tag{3.124}$$

3.6.1 Integral Equation for PDF of Wait

Consider the virtual wait process. Sample paths resemble those of the standard M/G/1 queue, except that each jump size depends on the arrival type. Jump sizes have cdf $B_i(\cdot)$ at Poisson rate $\lambda_i, i = 1, ..., N$. Consider a state-space level $x > 0$. By Theorem 1.1, the SP downcrossing rate is $f(x)$. The SP upcrossing rate due to type i arrivals is

$$\lambda_i P_0 \overline{B}_i(x) + \lambda_i \int_{y=0}^{x} \overline{B}_i(x - y) f(y) dy, i = 1, ..., N.$$

Balancing the *total* SP down- and upcrossing rates of level x for all customer types, yields the integral equation for $f(x)$,

$$f(x) = \sum_{i=1}^{N} \lambda_i \left(P_0 \overline{B}_i(x) + \int_{y=0}^{x} \overline{B}_i(x-y) f(y) dy \right),$$

or

$$f(x) = \lambda P_0 \left(\sum_{i=1}^{N} \frac{\lambda_i}{\lambda} \overline{B}_i(x) \right) + \lambda \int_{y=0}^{x} \left(\sum_{i=1}^{N} \frac{\lambda_i}{\lambda} \overline{B}_i(x-y) \right) f(y) dy. \tag{3.125}$$

Integral equation (3.125) is in the form of an integral equation for the pdf of wait in a *standard* M/G/1 queue with $\lambda = \sum_{i=1}^{N} \lambda_i$, and $\overline{B}(x) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} \overline{B}_i(x)$.

3.6.2 Expected Wait Before Service

Since $E(S^2) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda} E(S_i^2)$, the Pollaczek-Khinchin (P-K) formula (3.47) gives the expected wait before service as

$$\begin{aligned} E(W_q) &= \frac{\lambda E(S^2)}{2(1 - \lambda E(S))} = \frac{\sum_{i=1}^{N} \lambda_i E(S_i^2)}{2(1 - \sum_{i=1}^{N} \lambda_i E(S_i))} \\ &= \frac{\sum_{i=1}^{N} \lambda_i E(S_i^2)}{2(1 - \sum_{i=1}^{N} \rho_i)} = \frac{\sum_{i=1}^{N} \lambda_i E(S_i^2)}{2P_0}. \end{aligned} \tag{3.126}$$

Alternatively, $E(W_q)$ can be obtained by multiplying (3.125) through by x and integrating both sides with respect to $x \in (0, \infty)$.

3.6.3 Expected Number in Queue

Let L_q = expected number of units in the queue before service in the steady state. Then by $\boldsymbol{L = \lambda W}$and (3.126)

$$L_q = \lambda E(W_q) = \frac{\lambda \sum_{i=1}^{N} \lambda_i E(S_i^2)}{2(1 - \sum_{i=1}^{N} \rho_i)}. \tag{3.127}$$

Denote the steady-state expected number of type i units in the queue by L_{qi}. Let the wait of an arbitrary *type i customer* be W_{qi}, the wait of an *arbitrary customer* be W_q. Then $W_{qi} \underset{dist}{=} W_q$. Thus $E(W_{qi}) =$

$E(W_q), i = 1, ..., N$, and by $\boldsymbol{L = \lambda W}$,

$$\begin{aligned} L_{qi} &= \lambda_i E(W_{qi}) = \lambda_i E(W_q) \\ &= \frac{\lambda_i \sum_{i=1}^N \lambda_i E(S_i^2)}{2(1 - \sum_{i=1}^N \rho_i)}, i = 1, ..., N. \end{aligned} \tag{3.128}$$

3.6.4 Expected Busy Period

The expected busy period is, applying (3.60),

$$E(\mathcal{B}) = \frac{1 - P_0}{f(0)} = \frac{1 - P_0}{\lambda P_0} = \frac{\sum_{i=1}^N \rho_i}{\lambda \left(1 - \sum_{i=1}^N \rho_i\right)}. \tag{3.129}$$

As a mild check on (3.129), let $\lambda_i \equiv \frac{\lambda}{N}$ so that $\rho_i \equiv \frac{\lambda}{N} E(S_i)$ and $\sum_{i=1}^N \rho_i = \frac{\lambda}{N} \sum_{i=1}^N E(S_i)$. The model reduces to a standard M/G/1 queue with arrival rate λ and $E(S) = \frac{1}{N} \sum_{i=1}^N E(S_i)$. Then from (3.129)

$$E(\mathcal{B}) = \frac{\frac{\lambda}{N} \sum_{i=1}^N E(S_i)}{\lambda \left(1 - \frac{\lambda}{N} \sum_{i=1}^N \rho_i\right)} = \frac{E(S)}{1 - \lambda E(S)},$$

which is the result for the standard M/G/1 queue.

3.6.5 Exponential Service

To outline a solution technique for integral equation (3.125), assume the service times are exponential, i.e., $B_i(x) = 1 - e^{-\mu_i x}$, $i = 1, 2, .., N$. Then (3.125) becomes

$$f(x) = \sum_{i=1}^N \lambda_i \left(P_0 e^{-\mu_i x} + \int_{y=0}^x e^{-\mu_i (x-y)} f(y) dy \right), \ x > 0. \tag{3.130}$$

We may apply the differential operator

$$\langle D + \mu_1 \rangle \langle D + \mu_2 \rangle ... \langle D + \mu_N \rangle$$

to (3.130), to derive an N^{th} order differential equation with constant coefficients for $f(x)$, then solve for the constants of integration, giving $f(x)$ analytically.

Note that the differential operator $\langle D + constant\rangle$ is commutative, i.e., for any permutation $(i_1 i_2 ... i_{N)}$ of the numbers $(1, 2, ..., N)$

$$\begin{aligned}\langle (D+\mu_1)\cdots(D+\mu_N)\rangle f(x) &= \langle D+\mu_1\rangle \cdots \langle D+\mu_N\rangle f(x) \\ &= \left\langle D+\mu_{i_1}\right\rangle \cdots \left\langle D+\mu_{i_N}\right\rangle f(x)) \\ &= \langle (D+\mu_{i_1})\cdots(D+\mu_{_{i_N}})\rangle f(x).\end{aligned}$$

This commutativity property simplifies the transformation of an integral equation into a differential equation, when the kernel of any integral is an exponential function like $e^{-\mu_i(x-y)}$ in (3.130).

Expected Number in Queue

The expected total number of customers in the queue is, substituting into (3.127),

$$L_q = \frac{\lambda \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i^2}}{\left(1 - \sum_{i=1}^{N} \frac{\lambda i}{\mu_i}\right)}. \tag{3.131}$$

The expected number of type i customers in the queue is, substituting into (3.128),

$$L_{qi} = \frac{\lambda_i \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i^2}}{\left(1 - \sum_{i=1}^{N} \frac{\lambda i}{\mu_i}\right)}, i = 1, ..., N. \tag{3.132}$$

Two Customer Types

To illustrate the solution, we consider two distinct customer types, and compute the pdf $f(x)$. Setting $N = 2$ in (3.130) and 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

$$f(x) = ae^{R_1 x} + be^{R_2 x} \tag{3.133}$$

where R_i, $i = 1, 2$ are the roots for z of 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 $R_1 R_2 = \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1 > 0$ (stability condition), and $R_1 + R_2 = -(\mu_1 + \mu_2 - \lambda) < 0$. Constants a, b are determined by applying two independent initial conditions involving $f(0)$ and $f\prime(0)$ obtained from (3.133) and (3.130), resulting in two equations for a, b:

$$f(0) = a + b = \lambda P_0,$$

and

$$\begin{aligned} f'(0) &= R_1 a + R_2 b \\ &= -(\mu_1\lambda_1 + \mu_2\lambda_2)P_0 + \lambda f(0) \\ &= -(\mu_1\lambda_1 + \mu_2\lambda_2 - \lambda^2)P_0. \end{aligned}$$

Thus $f(x)$ is given by (3.133) and

$$\begin{aligned} a &= \frac{(-\lambda_1\mu_1 + \lambda^2 - \lambda_2\mu_2 - \lambda R_2)}{R_1 - R_2} P_0, \\ b &= \frac{(\lambda_1\mu_1 - \lambda^2 + \lambda_2\mu_2 + \lambda R_1)}{R_1 - R_2} P_0, \end{aligned} \tag{3.134}$$

where

$$\left.\begin{aligned} &P_0 = 1 - \tfrac{\lambda_1}{\mu_1} - \tfrac{\lambda_2}{\mu_2}, \\ &R_1 = \tfrac{-B}{2} + \tfrac{\sqrt{B^2-4AC}}{2}, \\ &R_2 = \tfrac{-B}{2} - \tfrac{\sqrt{B^2-4AC}}{2}, \end{aligned}\right\} \tag{3.135}$$

and

$$A = 1,\ B = \mu_1 + \mu_2 - \lambda,\ C = \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1.$$

Example 3.7 *Consider a simple numerical example with* $N = 2$, $\lambda_1 = 1$, $\lambda_2 = .5$, $\mu_1 = 3$, $\mu_2 = 2$. *Then* $P_0 = 0.4167$, $R_1 = -1.0$, $R_2 = -2.5$, $a = 0.1667$, $b = 1.3333$, *and*

$$f(x) = 0.1667e^{-1.0x} + 1.3333e^{-2.5x}, x > 0.$$

To check that $F(\infty) = 1$, *compute*

$$\begin{aligned} F(\infty) &= P_0 + \int_{x=0}^{\infty} f(x)dx \\ &= 0.4167 + \int_{x=0}^{\infty} (0.1667e^{-1.0x} + 1.3333e^{-2.5x})dx = 1. \end{aligned}$$

3.7 M/G/1: Wait-number Dependent Service

Arrivals occur at Poisson rate λ. The queue discipline is FCFS. The service time is denoted by $S(N_q)$ where $N_q =$ number of customers left waiting in the queue just after a start of service. Note that $N_q \in \{0, 1, ...\}$. For exposition, we assume two types of service. Let

$$S(N_q) = \begin{cases} S_0, N_q = 0, \\ S, N_q = 1, 2, ... \end{cases}$$

Let $P(S_0 \leq x) = B_0(x)$, $\overline{B_0}(x) = 1 - B_0(x)$; $P(S \leq x) = B(x)$, $\overline{B}(x) = 1 - B(x)$. Denote the steady-state wait before service by W_q having cdf $P(W_q \leq x) = F(x)$ and pdf $f(x) = \frac{d}{dx}F(x), x > 0$, wherever the derivative exists.

We represent this non-standard M/G/1 queue by M/G(N_q)/1. We construct a sample path of the ***virtual wait*** 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 would have to 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.

3.7.1 Sample Path of Virtual Wait

Consider Fig. 3.11. The first customer (C_1) arrives, initiates a busy period and receives a service time S_0, since zero customers are left behind it in queue when it starts service. Later C_2 arrives during C_1's service time and is allotted a "virtual" service time S, although C_2's true 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 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. For a would-be new arrival immediately after C_2's arrival, would force C_2 to start service with at least one customer left waiting behind C_2. 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. 3.11, *zero* customers arrive during C_2's wait. Then at C_2's start-of-service instant, C_2 must receive an actual

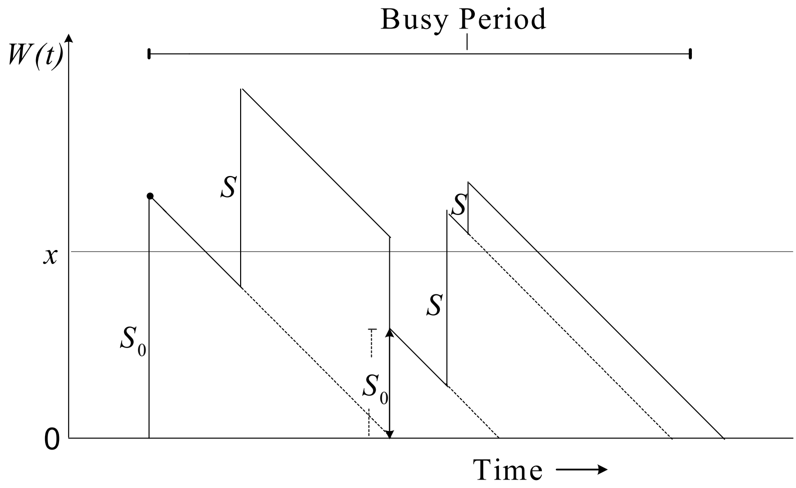

Figure 3.11: Sample path of virtual wait in M/G(N_q)/1 during a busy period. Shows jumps of size S_0 from level 0 and size S from positive levels. Illustrates possible downward jump in virtual wait.

service time S_0. This cancels S assigned at C_2's arrival epoch, and substitutes an actual service time S_0. The System Point (SP), jumps down to level 0, and up by an amount S_0, at the start-of-service instant of C_2. On reflection, all SP upward jumps from level 0 are of size S_0, and all SP upward jumps from positive levels are of size S.

At instants like the start-of-service instant of C_2 depicted in Fig. 3.11, the SP makes a double jump, one downward to level 0, and the other upward of size S_0 (see Examples 2.2 and 2.3.)

Remark 3.20 *In standard M/G/1, $W(t)$ is the same as the workload on the server 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 in queue. Next we discuss and derive the steady-state distribution of the* ***virtual wait*** *(in contrast to workload).*

3.7.2 Integral Equation for PDF of Virtual Wait

Consider a sample path of the virtual wait; fix level $x > 0$ in the state space (Fig. 3.11). The SP downcrossing rate of x has two components:

1. $f(x)$ by Theorem 1.1,

2. $\lambda\overline{B}(x)\mathcal{L}_f(\lambda)$ due to SP downward jumps similar to those at the start-of-service instant of C_2, where $\mathcal{L}_f(\lambda) = \int_{y=0}^{\infty} e^{-\lambda y} f(y)dy$ is the Laplace transform of $f(x)$.

In component 2, the rate of such downward jumps is

$$\begin{aligned}
&\lambda P(S > x,\ \text{and zero customers arrive in a waiting time}) \\
&\quad = \lambda P(S > x) P(\text{zero customers arrive in a waiting time}) \\
&\quad = \lambda P(S > x) \int_{y=0}^{\infty} e^{-\lambda y} f(y)dy = \lambda\overline{B}(x)\mathcal{L}_f(\lambda),
\end{aligned}$$

by independence of S and the arrival stream. The total downcrossing rate of x is

$$f(x) + \lambda\overline{B}(x)\mathcal{L}_f(\lambda), x > 0. \tag{3.136}$$

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. $\lambda\overline{B_0}(x)\mathcal{L}(\lambda)$, 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

$$\lambda\overline{B_0}(x)P_0 + \lambda \int_{y=0}^{x} \overline{B}(x - y)f(y)dy + \lambda\overline{B_0}(x)\mathcal{L}_f(\lambda). \tag{3.137}$$

SP rate balance across level x equates (3.136) and (3.137), leading to the integral equation for $f(x)$,

$$\begin{aligned}
f(x) = \quad & \lambda\overline{B_0}(x)P_0 + \lambda \int_{y=0}^{x} \overline{B}(x - y)f(y)dy \\
& +\lambda\left(\overline{B_0}(x) - \overline{B}(x)\right) \cdot \mathcal{L}_f(\lambda), x > 0.
\end{aligned} \tag{3.138}$$

3.7.3 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 (3.138) reduces to

$$\begin{aligned} f(x) = \quad & \lambda e^{-\mu_0 x} P_0 + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y) dy \\ & +\lambda \left(e^{-\mu_0 x} - e^{-\mu x}\right) \cdot \mathcal{L}_f(\lambda), x > 0. \end{aligned} \tag{3.139}$$

Applying differential operator $\langle D + \mu_0 \rangle \langle D + \mu \rangle$ to both sides of (3.139) yields the differential equation

$$\langle D^2 + (\mu_0 + \mu - \lambda) D + \mu_0 (\mu - \lambda) \rangle f(x) = 0,$$

with general solution

$$f(x) = a e^{-(\mu-\lambda)x} + b e^{-\mu_0 x}, x > 0, \tag{3.140}$$

assuming $\mu_0 \neq \mu - \lambda$. From the first term of 3.140, a necessary condition for stability is $\lambda < \mu$, since necessarily $f(\infty) = 0$.

Applying the initial condition $f(0) = \lambda P_0$, substituting

$$f(x) = a e^{-(\mu-\lambda)x} + b e^{-\mu_0 x} \tag{3.141}$$

from (3.140) into (3.139), and equating coefficients of common exponents, we obtain

$$P_0 = \frac{1 - \rho}{1 - \rho + \rho_0 + \rho_0^2 - \rho_0 \rho}, \tag{3.142}$$

and

$$a = \frac{-\lambda \rho_0^2 P_0}{\rho_0 - \rho - \rho_0 \rho}, \quad b = \frac{\lambda (1 + \rho_0)(\rho_0 - \rho) P_0}{\rho_0 - \rho - \rho_0 \rho}. \tag{3.143}$$

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. 3.11). Hence the expected busy period is, from (3.60),

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = \frac{\rho_0 + \rho_0^2 - \rho_0 \rho}{\lambda (1 - \rho)}. \tag{3.144}$$

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 (3.144) 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 (3.57) and using (3.141) and (3.143) we obtain the steady-state probability of n customers left in the system at departure instants,

$$\begin{aligned} d_n &= \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} f(x)dx \\ &= \frac{\rho_0 \cdot \left(\rho_0^{n-1} - \rho\rho_0^{n-2} - \rho^n(1+\rho_0)^{n-1}\right)}{(\rho_0 - \rho - \rho_0)(1+\rho_0)^{n-1}} P_0, \; n = 1, 2, ... \end{aligned} \tag{3.145}$$

where P_0 $(= d_0)$ is given in (3.142). The values in (3.145) agree with d_n in the literature, determined by different means (see [65]).

3.7.4 Workload

Consider the **workload process** $\{W_{wk}(t)\}$. Then $W_{wk}(t)$ = amount of remaining work in the system at time t. Let the steady-state pdf of $W_{wk}(t)$ as $t \to \infty$ be .$\{P_{0wk};\ f_{wk}(x), x > 0\}$

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 $(1 - e^{-\lambda y})$ (at least 1 arrival in time y), or size S_0 with probability $e^{-\lambda y}$ (no arrivals in time y). This leads to the upcrossing rate of level x to be the right side of (3.146) below. The downcrossing rate of x would be $f_{wk}(x)$. Rate balance across level x gives

$$\begin{aligned} f_{wk}(x) &= \lambda\overline{B_0}(x)P_{0wk} + \lambda \int_{y=0}^{x} \overline{B}(x-y)(1-e^{-\lambda y})f_{wk}(y)dy \\ &\quad + \lambda \int_{y=0}^{x} \overline{B_0}(x-y)e^{-\lambda y} f_{wk}(y)dy. \end{aligned} \tag{3.146}$$

We shall not develop the solution for the steady-state pdf of workload at this point, although it is interesting to compare with the pdf for the virtual wait. When service times are distributed as E_{μ_0} or E_{μ}, we would substitute $\overline{B_0}(x) = e^{-\mu_0 x}$, $\overline{B}(x) = e^{-\mu x}$ in (3.146) and solve with the normalizing condition $P_{0wk} + \int_{x=0}^{\infty} f_{wk}(x)dx = 1$.

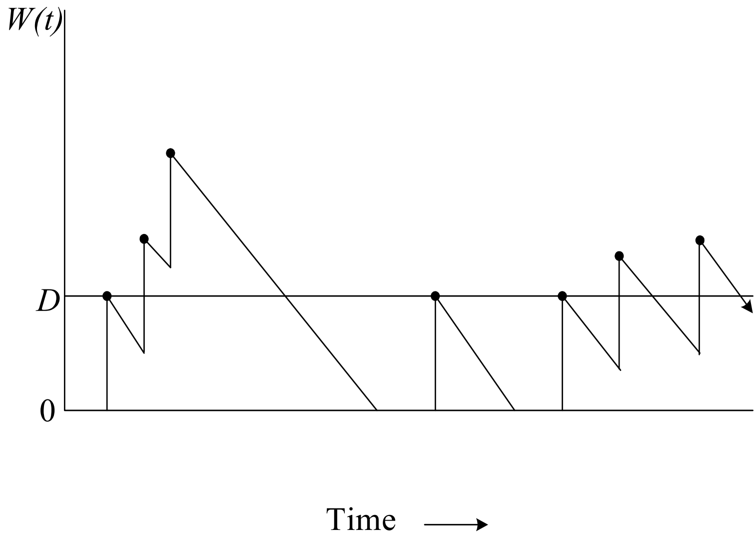

Figure 3.12: Sample path of virtual wait for M/D/1 queue.

3.8 M/D/1 Queue

The M/D/1 queue is a classical model in queueing theory, first solved by A.K. Erlang in 1909 [54].

Here we use LC to derive the steady-state cdf $F(x), x \geq 0$, pdf $f(x), x \geq 0$, of wait before service, the distribution of the number of customers in the system $P_n, n = 0, 1, 2, ...$, and related results.

The arrival stream is Poisson at rate λ. Denote the service time for each customer by $D > 0$. Let the traffic intensity be $\rho = \lambda D < 1$ implying stability. Consider the virtual wait $W(t), t \geq 0$, (Fig. 3.12) and the waiting time of the n^{th} arrival $W_n, n = 1, 2, ...$ Due to Poisson arrivals,

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

Also,

$$f(x) = \frac{d}{dx}F(x), x > 0,$$

wherever the derivative exists. We define $f(x), x > 0$, to be right continuous; and for notational convenience $f(0) \equiv f(0^+)$, which adds zero probability to $F(0)$. The probability of a zero wait is

$$P_0 \equiv F(0) = 1 - \rho = 1 - \lambda D.$$

The total pdf $\{P_0; f(x), x > 0\}$ is related to $F(x)$ by

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

3.8.1 Properties of PDF and CDF of Wait

Proposition 3.7 gives three properties of the steady-state pdf of wait in the M/D/1 queue.

Proposition 3.7 *For the M/D/1 queue, the steady-state pdf of wait* $\{P_0; f(x), x > 0\}$*: (1) has exactly one atom, which is at* $x = 0$*; (2) has a downward jump discontinuity of size* $\lambda(1-\rho) = \lambda P_0$ *at* $x = D$*; (3) is continuous for all* $x > 0$, $x \neq D$*.*

Proof. Consider a typical sample path of the virtual wait (Fig. 3.12).

(1) State $\{0\}$ is an atom since a sample path spends a positive proportion of time in $\{0\}$ (a.s.), namely $P_0 = (1-\rho) = 1-\lambda D$ (from (3.46)). The state space $\boldsymbol{S} = [0,\infty)$ has no other atoms, since the proportion of time a sample path spends in each state $x > 0$, is 0.

(2) Consider state-space levels D and $D-\varepsilon$, $0 < \varepsilon < D$ (Fig. 3.13). Fix $t > 0$. Recall that $\mathcal{T}_t^b(D)$ is the number of tangents to level D from below during $(0,t)$. Referring to Example 2.5 we have

$$\mathcal{D}_{t+\varepsilon}(D-\varepsilon) = \sum_{j=1}^{\mathcal{D}_t(D)+\mathcal{T}_t^b(D)} I_j(D), \tag{3.147}$$

where $I_j(D) = 1$ if the j^{th} downcrossing or tangent from below of level D, is followed by a downcrossing of level $D-\varepsilon$ exactly ε time units later (probability $e^{-\lambda\varepsilon}$); and $I_j(D) = 0$ otherwise. Note that $I_j(D)$ is independent of $\mathcal{D}_t(D)+\mathcal{T}_t^b(D)$ and $E(I_j(D)) = e^{-\lambda\varepsilon}, j = 1, 2, \ldots$. Taking expected values on both sides of (3.147) gives

$$E(\mathcal{D}_{t+\varepsilon}(D-\varepsilon)) = E(\mathcal{D}_t(D) + \mathcal{T}_t^b(D))e^{-\lambda\varepsilon} \tag{3.148}$$

By Corollary 3.2 the SP downcrossing rates of D and $D-\varepsilon$ are

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(D))}{t} = f(D) \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{D}_t(D-\varepsilon))}{t} = f(D-\varepsilon).$$

Also, $\lim_{t\to\infty} \frac{E\left(\mathcal{T}_t^b(D)\right)}{t} = \lambda P_0$. Dividing both sides of (3.148) by t, writing $\frac{1}{t} = \frac{1}{t+\varepsilon}\frac{t+\varepsilon}{t}$ on the left side, and letting $t \to \infty$ gives

$$f(D-\varepsilon) = (f(D) + \lambda P_0)e^{-\lambda\varepsilon}.$$

Then letting $\varepsilon \downarrow 0$ yields

$$f(\mathcal{D}^-) - f(D) = \lambda \mathcal{P}_0.$$

Hence the pdf has a *downward jump discontinuity* at D of size $\lambda P_0 = \lambda(1-\rho)$.

(3) Fix level $x > 0$, $x \neq D$. Sample paths are not tangent to level x with probability 1 due to continuous inter-arrival times (exponentially distributed). Let ε be small ($D \notin (x-\varepsilon, x)$ and $\varepsilon < \min(x, D)$). Then

$$\mathcal{D}_{t+\varepsilon}(x-\varepsilon) = \sum_{j=1}^{\mathcal{D}_t(x)} I_j(x) + o(\varepsilon) \cdot \theta_{x>D} \quad (a.s.) \tag{3.149}$$

where $\theta_{x>D} = 1$ if $x > D$ and $\theta_{x>D} = 0$ otherwise. (The term $o(\varepsilon)$ in (3.149) is the rate at which the SP jumps from the interval $(x-\varepsilon-D, x-D)$ into interval $(x-\varepsilon, x)$ at arrival instants.) Dividing both sides of (3.149) by t, letting $t \to \infty$ and noting that $\lim_{t\to\infty} \mathcal{D}_t(x) = \infty$ since $\{\mathcal{D}_t(x)\}$ is a renewal process, gives

$$\lim_{t\to\infty} \frac{\mathcal{D}_{t+\varepsilon}(x-\varepsilon)}{t+\varepsilon} \cdot \frac{t+\varepsilon}{t}$$
$$= \lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \cdot \lim_{t\to\infty} \frac{1}{\mathcal{D}_t(x)} \sum_{j=1}^{\mathcal{D}_t(x)} I_j(x) + \lim_{t\to\infty} \frac{o(\varepsilon)}{t} \quad (a.s.).$$

By the strong law of large numbers

$$\lim_{t\to\infty} \frac{1}{\mathcal{D}_t(x)} \sum_{j=1}^{\mathcal{D}_t(x)} I_j(x) = E(I_j(x)) = e^{-\lambda\varepsilon} \quad (a.s.).$$

Hence

$$f(x-\varepsilon) = f(x) \cdot e^{-\lambda\varepsilon} \quad (a.s.).$$

Letting $\varepsilon \downarrow 0$ yields $f(x^-) = f(x)$, so that x is a point of continuity.

■

Proposition 3.8 *The steady-state CDF of wait $F(x), x \geq 0$: (1) has a jump discontinuity at $x = 0$ of size $1-\rho$, (2) is continuous for all $\dot{x} > 0$.*

Proof. (1) $F(x)$ has a discontinuity at $x = 0$, since 0 is an atom having probability $F(0) = P_0 = 1 - \rho$.

(2) Fix $x > 0$ in the state space. Then x is not an atom by the previous proposition, and therefore $P(\{x\}) = 0$. That is, x is not a point of increase in probability. Thus x is a point of continuity of $F(\cdot)$.

■

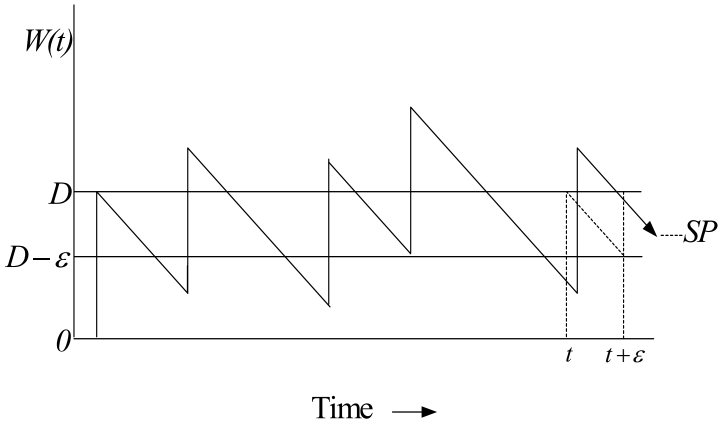

Figure 3.13: Sample path in M/D/1 showing levels D, $D-\varepsilon$ and instants t, $t+\varepsilon$. See Proposition 3.7, Proof, part (2).

3.8.2 Integral Equation for PDF of Wait

Applying the alternative form of the basic LC integral equation (3.36) with $B(x-y) = 0$ if $x-y < D$ and $B(x-y) = 1$ if $x-y \geq D$, we immediately write an integral equation for $f(x)$ (differential equation for the cdf $F(x)$) noting that $f(x) = F'(x)$,

$$f(x) = \lambda F(x) - \lambda F(x-D), x > 0. \tag{3.150}$$

To explain (3.150) in terms of LC, consider a virtual wait sample path (Fig. 3.12). In (3.150) 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 (3.150), 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 the rate of those jumps that start in $[0,x]$ and end below x. 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 (3.150).

Remark 3.21 *The properties in Proposition 3.7, and equation (3.150) are readily inferred intuitively upon considering a sample path (Fig. 3.12), and applying LC interpretations of transition rates. Such intuitive insights often lead to formal proofs as in Proposition 3.7.*

3.8.3 Analytic Solution for CDF and PDF of Wait

CDF of Wait

We give the classical solution of (3.150), for completeness. For $x \in (0, D)$, $F(x - D) \equiv 0$; thus $f(x) = \lambda F(x)$, or

$$F'(x) - \lambda F(x) = 0.$$

The solution of this differential equation is

$$F(x) = A_0 e^{\lambda x}.$$

Letting $x \downarrow 0$, gives the constant $A_0 = P_0 = 1 - \rho$. Thus

$$F(x) = (1 - \rho)e^{\lambda x}, x \in [0, D).$$

For $x \in [D, 2D)$, (3.150) is equivalent to

$$F'(x) - \lambda F(x) = -\lambda(1 - \rho)e^{\lambda(x-D)}.$$

Multiplying both sides by the integrating factor $e^{-\lambda(x-D)}$ and then integrating both sides from D to x yields the solution up to a constant

$$F(x) = -(1 - \rho)\lambda(x - D)e^{\lambda(x-D)} + A_1 e^{\lambda(x-D)}, x \in [D, 2D).$$

The constant A_1 is determined from the *continuity* of $F(x), x > 0$ (Proposition 3.7). Thus $F(D^-) = F(D)$, or $A_1 = (1 - \rho)e^{\lambda D}$ resulting in the solution

$$\begin{aligned} F(x) &= (1 - \rho)\left(-\lambda(x - D)e^{\lambda(x-D)} + e^{\lambda x}\right) \\ &= P_0\left(-\lambda(x - D)e^{\lambda(x-D)} + e^{\lambda x}\right), x \in [D, 2D). \end{aligned}$$

Mathematical induction on (3.150) yields the classical formula for the cdf of wait originally derived in [54].

$$F(x) = (1 - \rho)\sum_{i=0}^{m}(-\lambda)^i \frac{(x - iD)^i}{i!} e^{\lambda(x-iD)}, \quad x \in [m, (m + 1)D), \ m = 0, 1, 2, ... \tag{3.151}$$

PDF of Wait

The solution for the pdf $f(x)$ may be obtained by differentiating $F(x)$ with respect to x. We obtain $f(x)$ more simply by substituting (3.151) into (3.150) giving

$$f(x) = \lambda P_0 e^{\lambda x}, 0 < x < D$$

and for $x \in [mD, (m+1)D)$, $m = 0, 1, 2, ...,$

$$\begin{aligned} f(x) & \\ = \lambda P_0 & \left(\sum_{i=0}^{m} (-\lambda)^i \frac{(x-iD)^i}{i!} e^{\lambda(x-iD)} \right. \\ & \left. - \sum_{i=0}^{m-1} (-\lambda)^i \frac{(x-(i+1)D)^i}{i!} e^{\lambda(x-(i+1)D)} \right) \\ = \lambda P_0 & \left((-\lambda)^m \frac{(x-mD)^m}{m!} e^{\lambda(x-mD)} \right. \\ & \left. + \sum_{i=0}^{m-1} \frac{(-\lambda)^i}{i!} [(x-iD)^i e^{\lambda(x-iD)} - (x-(i+1)D)^i e^{\lambda(x-(i+1)D)}] \right). \end{aligned} \tag{3.152}$$

The pdf $f(x)$ in (3.152) has a discontinuity at $x = D$ (Proposition 3.7). That is $f(D^-) = \lambda(1-\rho)e^{\lambda D}$, and $f(D^-) - f(D) = \lambda(1-\rho)$, illustrating that $f(x)$ has a downward jump of size $\lambda(1-\rho) = \lambda P_0$ at $x = D$. Moreover $f(x)$ is continuous for all other $x > 0$ (see Fig. 3.14). Note the concave wave in $f(x)$ for $x \in [D, 2D) = [1, 2)$, and that the waviness dampens to the right of $x = 2$, in Fig. 3.14. The cdf $F(x)$, for the same example, is given in formula (3.151) and plotted in Fig. 3.15. Note the continuity of $F(x)$ and discontinuity of $f(x) = \frac{d}{dx}F(x)$ at $x = D$.

Remark 3.22 *LC indicates an isomorphism between sample-path properties of the virtual wait $W(t)$ and analytical properties of the functions $f(x)$ and $F(x)$.*

3.8.4 Distribution of Number in System

Let N be the number of customers in the system at an arbitrary time point and let W_q be the wait before service, in the steady-state. Then

$$\begin{aligned} &N \leq n \text{ iff } W_q \leq nD, \\ &N = n \text{ iff } (n-1)D \leq W_q < nD. \end{aligned}$$

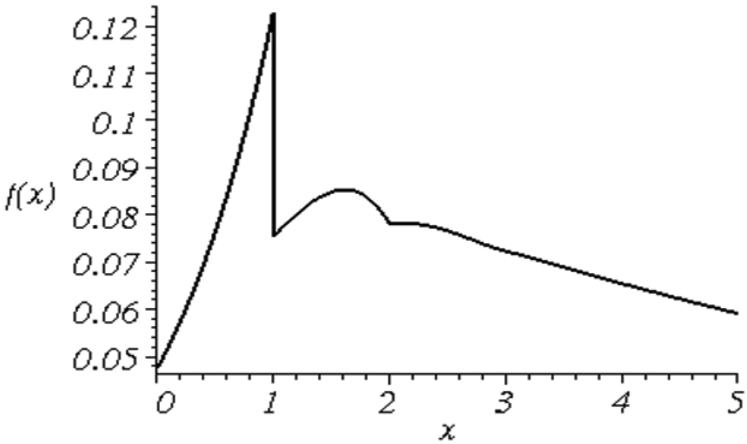

Figure 3.14: PDF $f(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$, $\rho = 0.95$ (high traffic). Shows discontinuity and downward jump of size λP_0 at $x = D$; and extreme waviness in right neighborhood $[D, 2D)$.

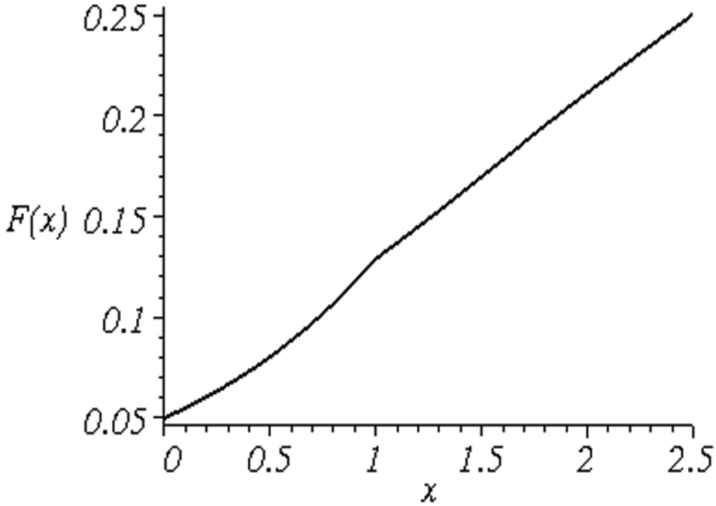

Figure 3.15: CDF $F(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$. Shows continuity of $F(x), x > 0$; and decrease in slope of $F(x)$ at $x = D$.

Let $P_n = P(N = n)$. Consider a_n, d_n, the steady state probabilities that the number of customers in the system is n just prior to an arrival, and just after a departure, respectively. Due to Poisson arrivals, $a_n = P_n = d_n, n = 0, 1, 2, ...$ Arrivals "see" n customers in the system iff their wait is in the time interval $((n-1)D,\, nD], n = 0, 1, 2, ...$. Thus

$$a_n = F(nD) - F((n-1)D) = P_n = d_n, n = 0, 1, 2, ...$$

From (3.151)

$$\begin{aligned}
P_0 &= F(0) - F(-D) = 1 - \rho \\
P_1 &= F(D) - F(0) = (1-\rho)e^{\lambda D} - (1-\rho) = (1-\rho)(e^{\lambda D} - 1) \\
P_2 &= F(2D) - F(D) = (1-\rho)e^{\lambda D}(-\lambda D + e^{\lambda D} - 1) \\
&\quad ... .
\end{aligned}$$

The cdf of N is

$$P(N \leq n) = \sum_{i=0}^{n} P_i = F(nD),\ n = 0, 1, 2, ...,$$

where $F(nD)$ is computed using (3.151).

3.9 M/Discrete/1 Queue

Consider the M/Discrete/1 queue, which we denote by M/{D$_n$}/1. This section derives analytical properties for the steady-state pdf and cdf of the wait before service, and suggests a technique for deriving analytical formulas for them. Consider a typical sample path of the virtual wait (Fig. 3.16).

In M/{D$_n$}/1, customers arrive in a Poisson stream at rate λ at a single server. Denote the service time by S. For each arrival,

$$P(S = D_i) = p_i, \quad \sum_{i=1}^{N} p_i = 1,$$

where D_i is a positive constant, $i = 1..., N$, and N is a positive integer. Then $E(S) = \sum_{i=1}^{N} p_i D_i$. Without loss of generality, let

$$0 \equiv D_0 < D_1 < ... < D_N < D_{N+1} \equiv \infty.$$

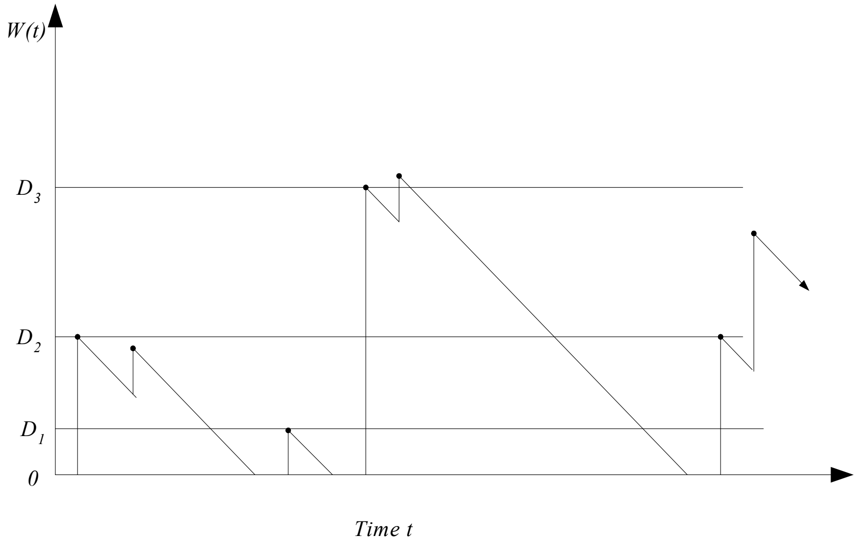

Figure 3.16: Sample path of virtual wait in $M/\{D_n\}/1$ queue with $N = 3$ service levels.

Customers that receive a service time D_i arrive at rate λp_i. The traffic intensity is $\rho = \lambda E(S)$. Assume $\rho < 1$ (stability). Due to Poisson arrivals

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

where $W_n, n = 1, 2, ...,$ is the actual wait of the n^{th} arrival (e.g., [99]).

Denote the steady-state cdf of wait by $F(x), x \geq 0$. The steady-state pdf of wait is $f(x) = \frac{d}{dx}F(x)$, $x > 0$, wherever the derivative exists. We define $f(x), x \geq 0$, to be 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 total pdf of wait is $\{P_0; f(x), x > 0\}$. A relationship between the cdf and pdf is given by

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

Remark 3.23 *The arrival stream may be viewed in two distinct ways:*

1. A homogeneous class of customers arrives at rate λ. Each arrival gets service time D_i with probability p_i, independently of other arrivals.

2. N separate classes of customers arrive at independent Poisson rates $\lambda_i \equiv \lambda p_i$ and receive service times $D_i, i = 1, ..., N$, respectively.

These two viewpoints yield the same steady state distribution of wait. This is reflected in the two equivalent forms for the traffic intensity $\rho = \lambda \sum_{i=1}^{N} p_i D_i = \sum_{i=1}^{N} \lambda_i D_i$.

Remark 3.24 *A similar analysis of the $M/\{D_n\}/1$ queue applies if $N = \infty$.*

3.9.1 Properties of PDF and CDF of Wait

The steady-state distribution of wait has analytical properties given in Proposition 3.9.

Proposition 3.9 *In the $M/\{D_n\}/1$ queue, the steady-state pdf of wait, $\{P_0; f(x), x > 0\}$: (1) has exactly one atom which is at $x = 0$ (state $\{0\}$ is atom); (2) has exactly $\mathbf{N}$ downward jump discontinuities of sizes $\lambda(1-\rho)p_i$ at $x = D_i$, $i = 1, ..., N$; (3) is continuous for all $x > 0, x \neq D_i, i = 1, ..., N$.*

Proof. Consider a typical sample path of the virtual wait process (Fig. 3.16).

(1) State $\{0\}$ is an atom since a sample path spends a positive proportion of time in $\{0\}$ (a.s.), namely $P_0 = (1-\rho) = 1 - \lambda \sum_{i=1}^{N} p_i D_i$. Each sojourn time in $\{0\} \underset{dist}{=} \mathrm{E}_\lambda$. 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) Fix $i \in \{1, ..., N\}$, and consider levels D_i and $D_i - \varepsilon$ in the state space, where $0 < \varepsilon < D_i - D_{i-1}$ and $\varepsilon < \min\{D_i\}$ (as in Fig. 3.13). By Corollary 3.2 of Theorem 3.3 the SP downcrossing rates of D_i and $D_i - \varepsilon$ are $\lim_{t\to\infty} \frac{E(\mathcal{D}_t(D_i))}{t} = f(D_i)$ and $\lim_{t\to\infty} \frac{E(\mathcal{D}_t(D_i-\varepsilon))}{t} = f(D_i - \varepsilon)$ respectively. Analogously to Example 2.5 we obtain

$$\mathcal{D}_{t+\varepsilon}(D_i - \varepsilon) = \sum_{j=1}^{\mathcal{D}_t(D_i)+\mathcal{T}_t^b(D_i)} I_j \tag{3.153}$$

where $I_j = 1$ if the j^{th} downcrossing of level D_i results in a downcrossing of level $D_i - \varepsilon$ exactly ε later, and $I_j = 0$ otherwise. In (3.153) the left side $\mathcal{D}_{t+\varepsilon}(D_i - \varepsilon)$ is the number of SP downcrossings of level $D_i - \varepsilon$ in $(0, t+\varepsilon)$. On the right side the sum's upper limit $\mathcal{D}_t(D_i) + \mathcal{T}_t^b(D_i)$ is the number of SP downcrossings of level D_i in $(0, t)$ (continuous downcrossings plus tangents from below). On the left side the subscript $t + \varepsilon$ accounts for the time taken for the SP to descend from D_i to $D_i - \varepsilon$. Taking expected values on both sides of (3.153) gives

$$E(\mathcal{D}_{t+\varepsilon}(D_i - \varepsilon)) = (E(\mathcal{D}_t(D_i)) + E(\mathcal{T}_t^b(D_i)))e^{-\lambda\varepsilon}$$

since $E(I_j) \equiv e^{-\lambda\varepsilon}$. Dividing by t and letting $t \to \infty$ (writing $\frac{1}{t} = \frac{1}{t+\varepsilon}\frac{t+\varepsilon}{t}$ on the left side) gives

$$f(D_i - \varepsilon) = (f(\mathcal{D}_i) + \lambda p_i P_0)e^{-\lambda\varepsilon},$$

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

$$f(\mathcal{D}_i^-) - f(\mathcal{D}_i) = \lambda p_i P_0.$$

Hence the pdf has a downward jump discontinuity at D_i of size $\lambda p_i P_0 = \lambda p_i(1 - \rho)$.

(3) Fix level $x > 0, x \neq D_i, i = 1, ..., N$. Sample paths are not tangent to level x (a.s.) due to continuous inter-arrival times (exponentially distributed). Let ε be small, i.e., $x - \varepsilon < \min_{i=1,\ldots,N}\{D_i - D_{i-1}\}$, no $D_i \in (x - \varepsilon, x)$ and $\varepsilon < x$. Then

$$\mathcal{D}_{t+\varepsilon}(x - \varepsilon) = \sum_{j=1}^{\mathcal{D}_t(x)} I_j.$$

On the left side the subscript $t + \varepsilon$ accounts for the time taken for the SP to descend from x to $x - \varepsilon$. Taking expected values gives

$$E(\mathcal{D}_{t+\varepsilon}(x - \varepsilon)) = E(\mathcal{D}_t(x))e^{-\lambda\varepsilon}.$$

Tandem downcrossings of x and $x - \varepsilon$ that happen more than ε apart require an arrival in time ε and a service time $< \varepsilon$, which is impossible by the choice of ε. Dividing by t and letting $t \to \infty$ (writing $\frac{1}{t} = \frac{1}{t+\varepsilon} \cdot \frac{t+\varepsilon}{t}$ on the left side) gives

$$f(x - \varepsilon) = f(x) \cdot e^{-\lambda\varepsilon}.$$

Letting $\varepsilon \downarrow 0$ yields $f(x^-) = f(x)$ so that x is a point of continuity. ∎

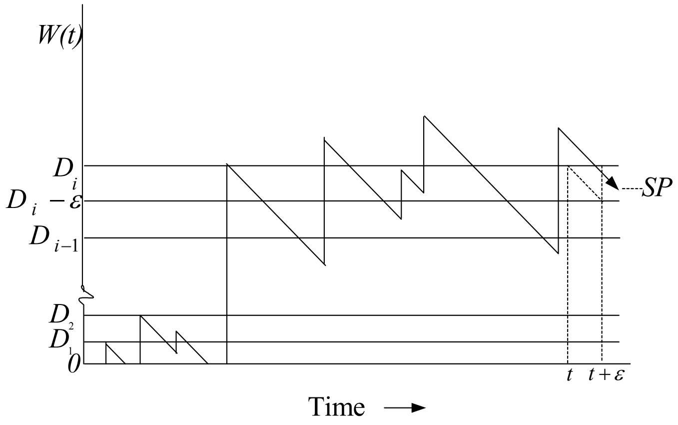

Figure 3.17: Sample path in M/{D$_n$}/1 showing levels $D_i, D_i - \varepsilon$ and instants t, $t + \varepsilon$. See Proposition 3.9, Proof, part (2).

Remark 3.25 *From part (2) of Proposition 3.9, the sum of the downward jumps at points of discontinuity of the pdf $f(x)$ is $\lambda(1-\rho)\sum_{i=1}^{N} p_i = \lambda(1-\rho) = \lambda P_0$. This sum is the same as the size of the single downward jump in the pdf of wait in the M/D/1 model!*

Proposition 3.10 *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 $1 - \rho$, and is continuous for all $x > 0$.*

Proof. $F(\cdot)$ has a jump discontinuity at level 0, since $\{0\}$ is an atom having probability $P_0 = F(0) = 1 - \rho$ (Proposition 3.9, part (2)). Fix $x > 0$ in the state space. Then x is not an atom (Proposition 3.9, part (3)). Hence x has probability 0. Thus x is a point of continuity of $F(\cdot)$. ■

3.9.2 Expected Busy Period

From (3.59) the expected busy period is

$$E(\mathcal{B}) = \frac{E(S)}{1 - \lambda E(S)} = \frac{1 - P_0}{\lambda P_0} = \frac{\sum_{i=1}^{N} D_i p_i}{1 - \lambda \sum_{i=1}^{N} p_i D_i}.$$

Let $\mathcal{I}$ denote an idle period. Another way to compute P_0 is

$$\begin{aligned} P_0 &= \frac{E(\mathcal{I})}{E(\mathcal{I})+E(\mathcal{B})} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+\frac{\sum_{i=1}^{N} p_i D_i}{1-\lambda\sum_{i=1}^{N} p_i D_i}} \\ &= 1-\lambda\sum_{i=1}^{N} p_i D_i. \end{aligned}$$

3.9.3 Integral Equation for PDF of Wait

The alternative form of the LC integral equation for M/G/1 (3.36) leads immediately to an integral equation for the pdf $f(x)$ (*differential equation* for cdf $F(x)$),

$$\begin{aligned} f(x) &= \lambda F(x) - \lambda\sum_{i=1}^{N} p_i F(x-D_i) \\ &= \lambda F(x) - \sum_{i=1}^{N} \lambda_i F(x-D_i), x > 0. \end{aligned} \qquad (3.154)$$

To verify (3.154) consider a virtual-wait sample-path (Fig. 3.16). In (3.154), 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 $p_i = \frac{\lambda_i}{\lambda}$. 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 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 (3.154).

3.9.4 Solution for CDF of Wait

Differential equation (3.154) for $F(x)$ is solvable. However the form of $F(x)$ differs in state-space state space intervals

$$\begin{aligned} &[0, D_1),\ [D_1, 2D_1), \\ &..., [j_{11}D_1, D_2),\ [D_2, (j_{11}+1)D_1),\ [(j_{11}+1)D_1, (j_{11}+2)D_1), \end{aligned}$$

etc., where $j_{11} = \left\lfloor \frac{D_2}{D_1} \right\rfloor$ (greatest integer $\leq \frac{D_2}{D_1}$). 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, ...,D_N.

3.9.5 Alternative Approach for CDF of Wait

An alternative way to obtain a solution for $F(x)$ is to derive the cdf of wait in a "specialized" M/{D$_n$}/1 queue. We can assume, without loss of *computational accuracy*, that all D_i*'s* are rational numbers. Let $D_1 = k_1 D, D_2 = k_2 D, ..., D_N = k_N D$, $D = \gcd\{D_1, ..., D_N\}$ and $0 < k_1 < k_2 < \cdots < k_N$ are positive integers (gcd denotes greatest common divisor).

To accomplish this, consider an M/{D$_n$}/1 queue where $D_i = iD$, $i = 1, ..., N$. 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. Arrival rates for intermediate service time values $\{iD | iD \neq D_i, i = 1, ..., N\}$ are set to 0 in the solution. The resulting cdf for M/{iD}/1 is equal to the cdf of wait for the original M/{D$_n$}/1 model (i.e., solution of (3.154)).

Thus M/{iD}/1 ($D = \gcd\{D_1, ..., D_N\}$) may be considered as equivalent M/{D$_n$}/1. Also, it is more amenable analytically and computationally.

3.10 M/{iD}/1 Queue

This section analyzes the M/{iD}/1 queue, keeping in mind its close relationship to M/{D$_n$}/1 (Subsection 3.9.5).

In M/{iD}/1 there are N types of arrivals at Poisson rates λ_i, $i = 1, ..., N$, where N is a positive integer. Customers of type i receive a service time iD,$D > 0$. Equivalently, customers arrive at Poisson rate λ and get a service time 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

$$P_0 = 1 - \lambda E(S) = 1 - \lambda \sum_{i=1}^{N} iDp_i = 1 - \sum_{i=1}^{N} iD\lambda_i.$$

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 $D_n = k_n D$, $D = \gcd\{D_1, ..., D_N\}$ and $k_n \in \{1, ..., N\}$. Paradoxically, M/{iD}/1 may also be considered as a *generalization* of M/{D_n}/1 (Subsection 3.9.5)!

3.10.1 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 equation (3.35) involving the pdf and cdf of wait to obtain

$$\begin{aligned} f(x) &= \lambda F(x) - \lambda \sum_{i=1}^{N} F(x - iD)p_i \\ &= \lambda F(x) - \sum_{i=1}^{N} \lambda_i F(x - iD), x > 0. \end{aligned} \tag{3.155}$$

Consider the virtual wait process (similar to Fig. 3.16). In (3.155) the left side is the virtual-wait 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 the rate of those 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 (3.155) results by rate balance across level x.

3.10.2 Recursion for CDF of Wait

This subsection outlines a procedure to solve (3.155) recursively for $F(x)$, $x \in [mD, (m+1)D)$, $m = 0, 1, 2, ...$ Let

$$F(x) \equiv F_m(x),\ f(x) \equiv f_m(x),\ x \in [mD, (m+1)D), m = 0, 1, 2, ...$$

and $F_{-k}(x) \equiv 0$ if k is a positive integer. Then write (3.155) as

$$\begin{aligned} f_m(x) &= \lambda F_m(x) - \sum_{i=1}^{N} \lambda_i F_{m-i}(x - iD), \\ &x \in [mD, (m+1)D),\ m = 0, 1, 2, ... \end{aligned} \tag{3.156}$$

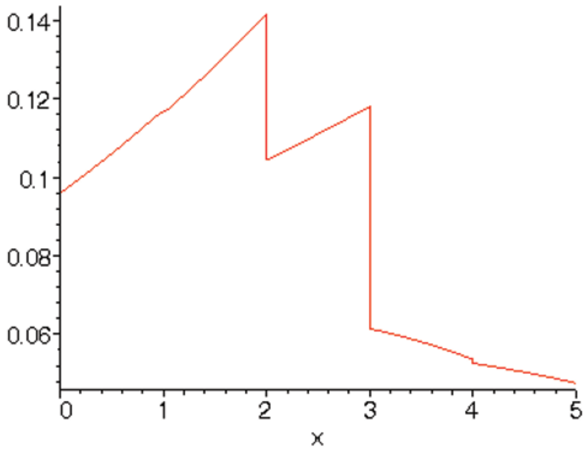

Figure 3.18: PDF of wait in $M/\{iD\}/1$ queue: four arrival types ($N = 4$), $\lambda = .2$, $p_1 = p_4 = .01$, $p_2 = .39$, $p_3 = .59$. Downward jumps at $x = 1, 2, 3, 4$.

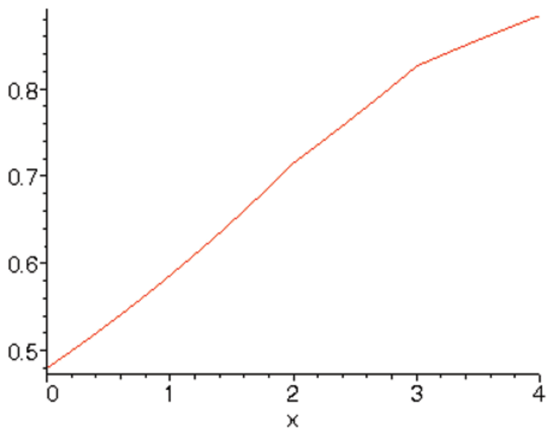

Figure 3.19: CDF of wait in $M/\{iD\}/1$ queue. $N = 4$, $\lambda = .2$, $p_1 = p_4 = .01$, $p_2 = .39$, $p_3 = .59$. Slope decreases abruptly at $x = 1, 2, 3, 4$.

Consider state-space interval $[0, D)$. Note that $F(x - iD) = 0$ if $x - iD < 0$. For $x \in [0, D)$, equation (3.156) reduces to

$$\begin{aligned} f_0(x) &= \lambda F_0(x), x \in [0, D), \\ \frac{dF_0(x)}{dx} &= \lambda F_0(x), x \in (0, D), \end{aligned}$$

with solution

$$F_0(x) = (1 - \rho)e^{\lambda x}, x \in [0, D).$$

Next, equation (3.156) reduces to

$$\begin{aligned} f_1(x) &= \lambda F_1(x) - F_0(x - D)\lambda_1, x \in [D, 2D), \\ f_1(x) &= \lambda F_1(x) - (1 - \rho)e^{\lambda(x-D)}\lambda_1, x \in [D, 2D). \end{aligned}$$

Substituting $f_1(x) = \frac{d}{dx}F_1(x)$ in the last equation makes it a differential equation in $F_1(x)$, which is readily solved up to a constant. The constant is evaluated using continuity $F_0(D^-) = F_1(D)$. The solution is

$$F_1(x) = (1 - \rho)\left(e^{\lambda x} + \lambda_1(D - x)e^{-\lambda(D-x)}\right), x \in [D, 2D),$$

which can be written as

$$F_1(x) = F_0(x) + (1 - \rho)\lambda_1(D - x)e^{-\lambda(D-x)}, x \in [D, 2D),$$

if we extend the domain of $F_0(x)$ to $[0, \infty)$.

In a similar manner, we obtain recursively

$$F_2(x), x \in [2D, 3D), .\ F_3(x), x \in [3D, 4D), \quad F_4(x), x \in [4D, 5D).$$

where we extend the domain of $F_m(x)$ to $[m, \infty)$. The recursive formulas in (3.157) below summarize the values of $F(x)$ on state-space interval $[0, 5D)$ by specifying the corresponding functions on intervals

$[0, D)$,...,$[4D, 5D)$.

$$
\begin{aligned}
F_0(x) = &\quad (1-\rho)e^{\lambda x}, \\
F_1(x) = &\quad F_0(x) + (1-\rho)\lambda_1(D-x)e^{-\lambda(D-x)}, \\
F_2(x) = &\quad F_1(x) + (1-\rho)\left(\lambda_2(2D-x) + \tfrac{\lambda_1^2(2D-x)^2}{2!}\right)e^{-\lambda(2D-x)}, \\
F_3(x) = &\quad F_2(x) + (1-\rho)(\lambda_3(3D-x) + \lambda_2\lambda_1(3D-x)^2 \\
&\quad + \tfrac{\lambda_1^3(3D-x)^3}{3!})e^{-\lambda(3D-x)}, \\
F_4(x) = &\quad F_3(x) + (1-\rho)(\lambda_4(4D-x) + \lambda_3\lambda_1(4D-x)^2 \\
&\quad + \tfrac{\lambda_2^2(4D-x)^2}{2!} + \tfrac{\lambda_2\lambda_1^2(4D-x)^3}{2!} + \tfrac{\lambda_1^4(4D-x)^4}{4!})e^{-\lambda(4D-x)}.
\end{aligned}
\tag{3.157}
$$

The recursion (3.157) can be continued. It can shown that the general form is (Shortle and Brill [92])

$$
F_m(x) = F_{m-1}(x) + (1-\rho)e^{-\lambda(mD-x)} \sum_{\mathcal{L}\in\mathcal{P}(m)} \frac{(mD-x)^{|\mathcal{L}|}}{H(\mathcal{L})} \prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j,
\tag{3.158}
$$

where $\mathcal{P}(m)$, $\mathcal{L}$, $H(\mathcal{L})$, and $\prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j$ are explained in the next subsection.

3.10.3 Solution for CDF and PDF of Wait

Using mathematical induction, it can be shown that an analytical solution of recursion (3.158) for the cdf of wait is

$$
\begin{aligned}
F_m(x) = &\quad (1-\rho)\textstyle\sum_{i=0}^{m} e^{-\lambda(iD-x)} \sum_{\mathcal{L}\in\mathcal{P}(i)} \frac{(iD-x)^{|\mathcal{L}|}}{H(\mathcal{L})} \prod_{j\in\mathcal{L}} \boldsymbol{\lambda}_j, \\
&\qquad x \in [mD, (m+1)D),\ m = 0, 1, ...,
\end{aligned}
\tag{3.159}
$$

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 $n_1, \cdots, n_d$, 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 $f_m(x) = \frac{d}{dx}F_m(x)$. Differentiating (3.159) with respect to x, gives for $x \in (mD, (m+1)D), m = 0, 1, 2, ...$,

$$f_m(x) = (1-\rho)\sum_{i=0}^{m} e^{-\lambda(iD-x)} \sum_{\mathcal{L}\in\mathcal{P}(i)} (\lambda(iD-x)-|\mathcal{L}|)\frac{(iD-x)^{|\mathcal{L}|-1}}{H(\mathcal{L})}\prod_{j\in\mathcal{L}}\boldsymbol{\lambda}_j.$$

As a mild check on (3.159), we obtain the cdf of wait for an M/D/1 queue from it, namely

$$\begin{aligned} F_m(x) &= (1-\rho)\sum_{i=0}^{m} e^{-\lambda(iD-x)}\frac{(iD-x)^i}{i!}\lambda^i \\ &= (1-\rho)\sum_{i=0}^{m}(-\lambda)^i\frac{(x-iD)^i}{i!}e^{-\lambda(iD-x)}, \\ & x \in [mD, (m+1)D), m = 0, 1, ... \end{aligned}$$

The latter M/D/1 formula 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 1's, $\{1, ..., 1\}$; (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 3.26 *Formula (3.159) can also be obtained by inversion of the Laplace transform of wait (see equation (3.51)) [92]. The inversion procedure is at least as involved as the LC derivation above. 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 directly in the time domain; (2) the effect on the solution, due to the discontinuities in $f(x)$ and continuity of $F(x)$, is clear; (3) because LC emphasizes sample paths, it enhances intuitive understanding of the model dynamics, and suggests new avenues for research.*

3.11 M/G/1 with Reneging

In this section we analyze an M/G/1 queue in which arrivals: (1) stay for full service if their wait is zero, (2) may renege from the waiting line, (3) may wait in line but balk at service, (4) may wait and receive full service if their required wait is positive.

Let the service time S having cdf $B(x)$ and $\overline{B}(x) = 1 - B(x), x \geq 0$. Let $W(t), t \geq 0$ denote the virtual wait. Let τ_n be the arrival time of customer C_n,$n = 1, 2, ...$. Then $W(\tau_n^-) \equiv W_n$ is the *required wait* before service of C_n, $n = 1, 2, ...$. Define for $n = 1, 2, ...$,

$$\theta_{W_n} = \begin{cases} 1 \text{ if } C_n \text{ waits and receives a full service,} \\ 0 \text{ if } C_n \text{ reneges while waiting or waits and balks at service.} \end{cases} \tag{3.160}$$

3.11.1 Staying Function

For each $y \geq 0$, define the common *conditional* probabilities

$$\begin{aligned} \overline{R}(y) &\equiv P(\theta_{W_n} = 1 | W_n = y), \\ R(y) &\equiv P(\theta_{W_n} = 0 | W_n = y), \end{aligned} \tag{3.161}$$

independent of $n = 1, 2, ...$ From (3.160) $\overline{R}(y) + R(y) = 1, y \geq 0$.

Random variable θ_y has a Bernoulli distribution for each required wait $y \geq 0$. The probability of staying for full service is $\overline{R}(y)$. The probability of reneging from the waiting line or balking at service is $R(y)$.

This section assumes $\overline{R}(y)$ is monotone non-increasing (decreasing in the wide sense), and bounded below by 0. Then $\lim_{y\to\infty} \overline{R}(y)$ exists. Let $\lim_{y\to\infty} \overline{R}(y) = L$. Then $0 \leq L \leq 1$. Let $H(y), y \geq 0$ denote a generic cdf.

$$\text{If} \begin{cases} L = 0 \text{ then } \overline{R}(y) = 1 - H(y), \\ L > 0 \text{ then } \overline{R}(y) \neq 1 - H(y). \end{cases}$$

Since no balking is allowed at an arrival instant, $\overline{R}(0) = 1$.

If $\overline{R}(y) \equiv 1, y \geq 0$, then $L = 1$. There would be no reneging from the waiting line and no balking at service. Then each $C_n, n = 1, 2, ...$ would wait and receive full service. The model would reduce to a standard M/G/1 queue.

Remark 3.27 *In a more general model,* $\overline{R}(y)$ *may be an arbitrary function such that* $\overline{R}(y) \in [0, 1], y \geq 0$, *not necessarily monotone. In that case, the presented analysis applies as well. However, the stability condition would be slightly modified (see Theorem 3.8 and Remark 3.31 below).*

We can use $\overline{R}(y)$ *to model balking upon arrival (e.g.,* $0 < \overline{R}(0) < 1$*) and/or reneging from service. The model may also incorporate state dependence (e.g., service time depending on wait).*

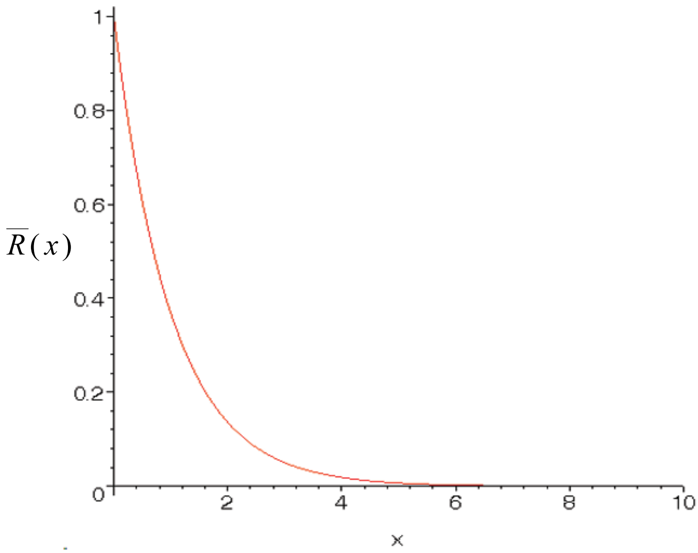

Figure 3.20: Staying function $\overline{R}(x) = e^{-rx.}$ $(r = 1)$; $L = 0$. $\overline{R}(x) = 1 - H(x)$ where $H(x)$ is a cdf.

Staying Function

We call $\overline{R}(y)$ the ***staying function***. $\overline{R}(y)$ is the probability that an arrival waits in line and stays for a full service, given that y is the required wait before service (see Figs. 3.20-3.22).

3.11.2 Reneging While Waiting or Balking at Service

We analyze the *required wait* before service. We may think of customers who renege from the waiting line *as if* they wait until start of service and then *balk at service*. (This makes no difference to the virtual wait for stayers.) Service-balkers receive zero service time. They are cleared from the system just before start of service. Thus they add zero to the required wait of any customer.

3.11.3 Sample Path of Virtual Wait for Stayers

The virtual wait $W(t)$ is the required wait of a would-be time-t arrival that stays for service. Consider a sample path of $\{W(t), t \geq 0\}$. If the actual wait is $W_n = 0$ then the SP jump size at τ_n has cdf $B(\cdot)$, starting

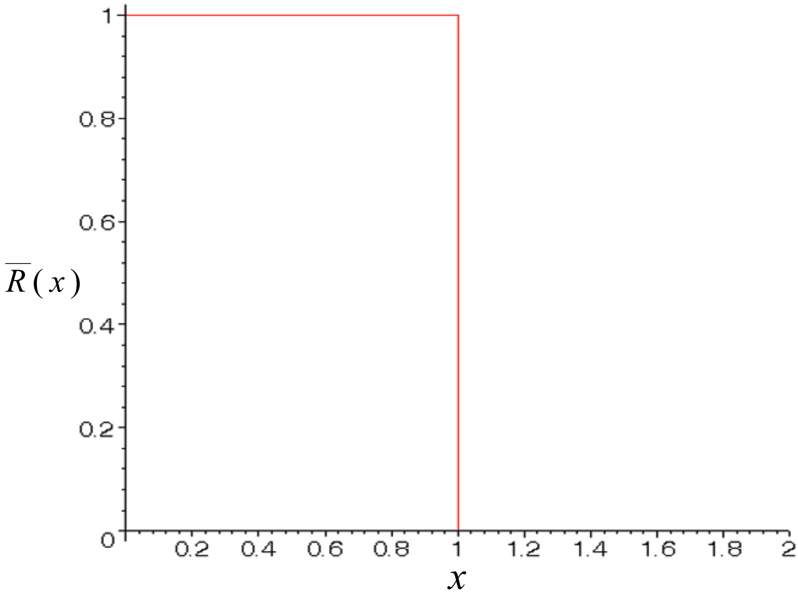

Figure 3.21: Staying function $\overline{R}(x) = 1, x < 1$, $\overline{R}(x) = 0, x \geq 1$. $L = 0$. $\overline{R}(x) = 1 - H(x)$, where $H(x)$ is a cdf.

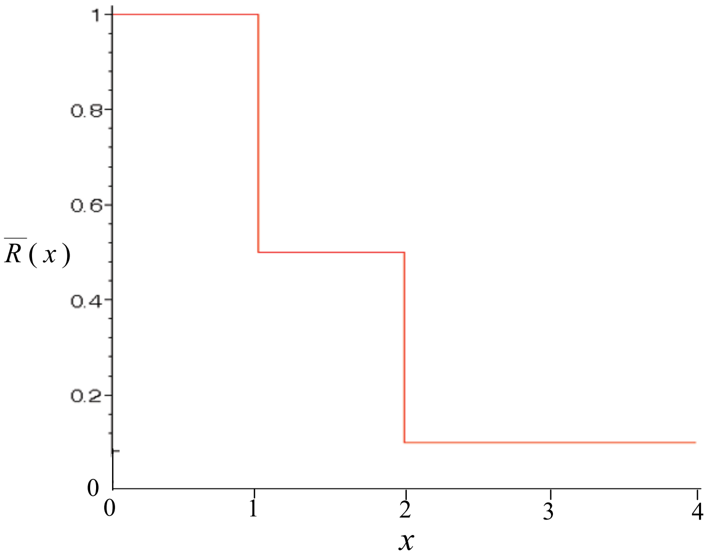

Figure 3.22: $\overline{R}(x) = 1, x < 1$, $\overline{R}(x) = 0.5, 1 \leq x < 2$, $\overline{R}(x) = 0.1, x \geq 2$. $\overline{R}(x) \neq 1 - H(x)$, where $H(x)$ is a cdf.

from level 0. If the event

$$\{W_n = y > 0, \theta_y = 1\}$$

occurs then the SP jump size at τ_n has cdf $B(\cdot)$, starting at level y. The probability that a jump occurs is $\overline{R}(y)$. If the event

$$\{W_n = y > 0, \theta_y = 0\}$$

occurs then C_n reneges or balks at service; the SP makes *no* jump at τ_n. The probability of no jump is $R(y)$.

A would-be arrival at τ_n^+ just after a reneger (or service-balker) C_n arrives, also would have a required wait y until service. This implies $W(\tau_n) = W(\tau_n^-) = y$. The sample path would be continuous with slope -1 at τ_n (Fig. 3.23).

Remark 3.28 *In Fig. 3.23 we consider a single busy period. Stayers arrive at τ_n; renegers arrive at $a_n, n = 1, 2, \ldots.$ If at least one stayer arrives after a_n, the start-of-service time of the first such stayer is denoted by σ_n. If zero stayers arrive after a_n, the end of the busy period is denoted by b_n. Knowledge of a_n, σ_n, b_n are sufficient to compute the required wait of the reneger arriving at a_n. If the reneger is cleared from the system prior to its required wait, the required wait is a "censored" variable. In order to compute the required wait we must observe the sample path until the end of the busy period in which the reneger arrives.*

The required waits of stayers and of renegers or service-balkers are useful quantities for a particular method of non-parametric estimation of the staying function from observations of the queue in continuous time.

3.11.4 Equation for PDF of Wait of Stayers

Denote the steady-state pdf of the *required wait* for stayers (virtual wait), by $\{P_0; f(x), x > 0\}$ where P_0 is the probability of a zero required wait. An LC-derived integral equation for $f(x)$ is

$$f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)\overline{R}(y) f(y) dy. \qquad (3.162)$$

In (3.162) the left side is the SP downcrossing rate of level x.

On the right side of (3.162), $\lambda P_0 \overline{B}(x)$ is the rate of SP jumps starting from level 0, that upcross level x (stayers). The term

$$\lambda \int_{y=0}^{x} \overline{B}(x-y)\overline{R}(y) f(y) dy$$

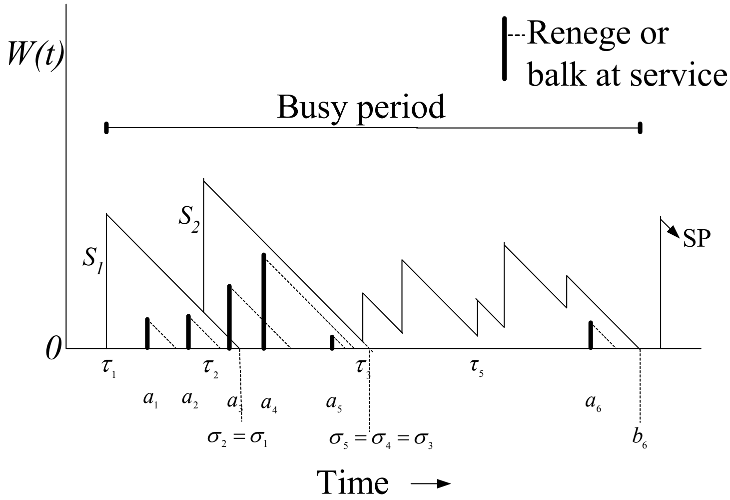

Figure 3.23: M/G/1 busy period showing stayers (τ_n), renegers (a_n) σ_n, and b_6 (end busy period), used to compute required waits of renegers.

is the rate of SP jumps starting at levels $y \in (0, x)$, that upcross level x. The right side is the SP total upcrossing rate of level x due to stayers. Rate balance across level x yields integral equation (3.162). The pdf on the left side is the *time-average* pdf. The pdf under the integral on the right side is the embedded pdf at arrival instants. Due to Poisson arrivals the two pdf's are equal. (We verify this claim by deriving integral equation (3.162) using the embedded LC method later in Subsection 8.4.2. In the embedded LC technique, $f(x) = \lim_{n\to\infty} f_n(x)$ inherently.)

Proportion of Customers That Get Full Service

Stayers are zero waiters or waiters that reach the server and receive full service. Denote by q_S, the proportion of arrivals that are stayers. Then q_S is the probability that an arbitrary arrival gets full service. Thus

$$q_S = P_0 + \int_{y=0}^{\infty} \overline{R}(y) f(y) dy. \tag{3.163}$$

The proportion of customers that renege while waiting, or balk at start of service, is

$$1 - q_S = \int_{y=0}^{\infty} R(y) f(y) dy.$$

3.11.5 M/M/1 with Reneging

Let $\overline{B}(x) = e^{-\mu x}, x \geq 0$ (service rate μ). Then (3.162) becomes

$$f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy. \tag{3.164}$$

Applying differential operator $\langle D + \mu \rangle$ to both sides of (3.164) yields the first order differential equation

$$\langle D + \mu \rangle f(x) = \lambda \overline{R}(x) f(x),$$
$$f'(x) + (\mu - \lambda \overline{R}(x)) f(x) = 0.$$

Separation of variables followed by integration gives the solution

$$f(x) = A e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)}, x > 0, \tag{3.165}$$

where A is a constant. Letting $x \downarrow 0$ in (3.164) and (3.165) implies

$$f(0) = A = \lambda P_0.$$

From LC, $f(0)$ is the SP entrance rate into $\boldsymbol{T} \times \{0\}$ (level 0) from above. The term λP_0 is the SP exit rate from level 0 into the state-space interval $(0, \infty)$. The resulting pdf of wait is

$$f(x) = \lambda P_0 e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)}, x > 0. \tag{3.166}$$

The normalizing condition $P_0 + \int_{x=0}^{\infty} f(x) dx = 1$ leads to

$$P_0 = \frac{1}{1 + \lambda \int_{x=0}^{\infty} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)} dx}. \tag{3.167}$$

3.11.6 Stability Condition for M/M/1 with Reneging

Theorem 3.8 gives a necessary and sufficient condition on the model parameters such that the steady-state distribution of required wait exists (stability).

Theorem 3.8 *Consider an $M_\lambda/M_\mu/1$ queue in which customers may renege before service, or wait the required time and then balk at service. Let the staying function be $\overline{R}(x), x \geq 0$, where $\overline{R}(x)$ is monotone non-increasing and $\overline{R}(0) = 1$. Let $L = \lim_{x \to \infty} \overline{R}(x)$. A necessary and sufficient condition for stability is*

$$\lambda < \begin{cases} \frac{\mu}{L} \text{ if } 0 < L \leq 1, \\ \infty \text{ if } L = 0. \end{cases} \tag{3.168}$$

Proof. (Adapted from [69]) Note that $\lim_{x\to\infty} \overline{R}(x) = L, 0 \le L \le 1$ exists. This is because $\overline{R}(0) = 1$, $\overline{R}(x)$ is monotone non-increasing and bounded below by 0. Stability holds iff the discrete state $\{0\}$ is positive recurrent iff $0 < P_0 \le 1$. Let $I \equiv \int_{x=0}^{\infty} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} dx$ in the denominator of (3.167). For stability, I is necessarily finite. That is we must have

$$I < \infty. \tag{3.169}$$

We show that (3.169) is equivalent to (3.168).

Since $L \le \overline{R}(x), x \ge 0$

$$\begin{aligned} \lambda L x &= \lambda \int_{y=0}^{x} L dx \le \lambda \int_{y=0}^{x} \overline{R}(y)dy \\ &\iff e^{-\mu x + \lambda L x} \le e^{\left(-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} \\ &\iff \int_{x=0}^{\infty} e^{-(\mu - \lambda L)x} dx \le I. \end{aligned} \tag{3.170}$$

For given $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that $\overline{R}(x) < \varepsilon + L$ for $x > M_\varepsilon$. Thus

$$\begin{aligned} \lambda \int_{y=0}^{x} \overline{R}(y)dy &< \lambda \int_{y=0}^{M_\varepsilon} \overline{R}(y)dy + \lambda \int_{y=M_\varepsilon}^{x} (\varepsilon + L)\, dy \\ &= C_1 + \lambda (\varepsilon + L)\, x, x > M_\varepsilon \end{aligned}$$

$$\begin{aligned} &\Longrightarrow \quad e^{\left(-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} < C_2 e^{-\mu x + \lambda(\varepsilon + L)x}, x > M_\varepsilon \\ &\Longrightarrow \quad \int_{x=M_\varepsilon}^{\infty} e^{\left(-\mu x + \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} dx < C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu + \lambda L + \lambda\varepsilon)x} dx \\ &\Longrightarrow \quad I < C_3 + C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu + \lambda L + \lambda\varepsilon)x} dx, \end{aligned} \tag{3.171}$$

where C_1, C_2, C_3 are positive constants. Combining inequalities (3.170) and (3.171) gives

$$\int_{x=0}^{\infty} e^{-(\mu - \lambda L)x} dx \le I < C_3 + C_2 \int_{x=M_\varepsilon}^{\infty} e^{(-\mu + \lambda L + \lambda\varepsilon)x} dx. \tag{3.172}$$

Consider (3.172). If $I < \infty$ then

$$\int_{x=0}^{\infty} e^{-(\mu - \lambda L)x} dx < \infty \iff \mu - \lambda L > 0. \tag{3.173}$$

If $\mu - \lambda L > 0$ then choose ε so that $-\mu + \lambda L + \lambda\varepsilon < 0$, i.e., choose $\varepsilon < \frac{\mu - \lambda L}{\lambda}$. Then

$$\int_{x=M_\varepsilon}^{\infty} e^{(-\mu+\lambda L+\lambda\varepsilon)x} dx < \infty \implies I < \infty. \tag{3.174}$$

The stability condition (3.168) is equivalent to (3.173) and (3.174). ■

Remark 3.29 *To shed additional perspective on the stability condition (3.168), consider the exponent in the integrand of*

$$I \equiv \int_{x=0}^{\infty} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} dx.$$

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

$$\frac{d}{dx}\int_{y=0}^{x} \overline{R}(y)dy = \overline{R}(x), x > 0.$$

Assume $\overline{R}(x), x > 0$, is strictly *decreasing and differentiable. Then $\int_{y=0}^{x} \overline{R}(y)dy$ is concave since*

$$\frac{d^2}{dx^2}\int_{y=0}^{x} \overline{R}(y)dy = \frac{d}{dx}\overline{R}(x) < 0, x > 0.$$

Also

$$\lim_{x\to\infty} \frac{d}{dx}\int_{y=0}^{x} \overline{R}(y)dy = \lim_{x\to\infty} \overline{R}(x) = L.$$

We compare the graphs of μx and $\lambda \int_{y=0}^{x} \overline{R}(y)dy, x > 0$ in Fig. 3.24. 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 $\mu > \lambda L$. 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 $\mu \geq \lambda \cdot 0$. Thus λ can assume any positive value, i.e., $0 < \lambda < \infty$.

Remark 3.30 *If $\overline{R}(x)$ is piecewise continuous, we can obtain similar perspective as in Remark 3.31.*

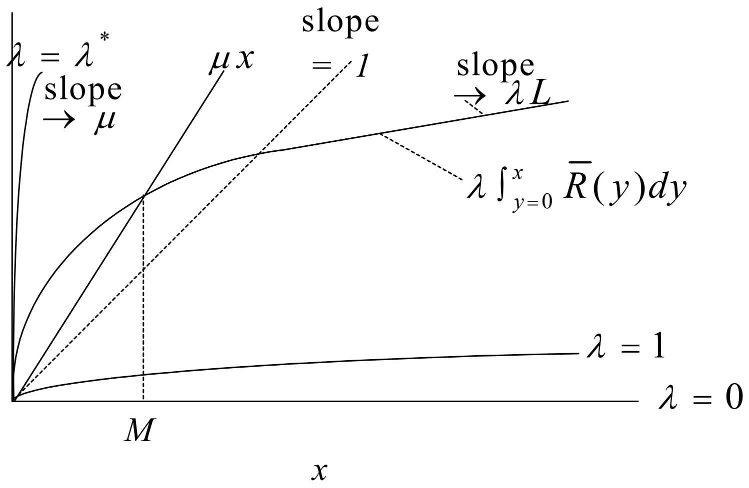

Figure 3.24: Functions μx and $\lambda \int_{y=0}^{x} \overline{R}(y)dy$, indicating M such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for $x \geq M$. Indicates range $0 < \lambda < \lambda^*$ such that stability holds. System is stable for λ if $\lambda \int_{y=0}^{x} \overline{R}(y)dy$ intersects and remains below μx thereafter.

Alternative Proof of Theorem 3.8

We provide an alternative proof of the stability condition, in order to clarify the intuition behind the result. Consider an *optimization problem* 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 $x \geq M$ (system is stable). 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 [12] together with a theorem in which the staying function may be other than monotone non-increasing. That proof is based on the fact that*

$$\int_{x=0}^{\infty} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy\right)} dx = \int_{x=0}^{\infty} e^{-\mu x} \cdot e^{\lambda \int_{y=0}^{x} \overline{R}(y)dy} dx$$

is the Laplace transform of $e^{\lambda \int_{y=0}^{x} \overline{R}(y)dy}$ *evaluated at 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*

$$\overline{L} = \lim \sup_{x\to\infty} \overline{R}(x).$$

A sufficient *condition for stability is*

$$\lambda < \frac{\mu}{\overline{L}} \text{ if } \overline{L} > 0,$$
$$\lambda < \infty \text{ if } \overline{L} = 0.$$

3.11.7 M/M/1 with Exponential Staying Function

Assume $\overline{B}(x) = e^{-\mu x}, x \geq 0$, 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 subsection 3.11.5.

Equation (3.162) becomes

$$f(x) = \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} e^{-ry} f(y)dy. \tag{3.175}$$

Substituting e^{-ry} for $\overline{R}(y)$ in (3.166) gives the pdf of wait for stayers,

$$\begin{aligned} f(x) &= \lambda P_0 e^{-\mu x + \frac{\lambda}{r}(1-e^{-rx})} \\ &= \lambda e^{\lambda/r} P_0 e^{-\mu x - \frac{\lambda}{r} e^{-rx}}, \ x > 0. \end{aligned} \tag{3.176}$$

We obtain

$$\begin{aligned} P_0 &= \frac{1}{1 + \lambda e^{\lambda/r} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx} \\ &= \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + e^{\lambda/r} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx}. \end{aligned} \tag{3.177}$$

In the denominator of P_0 the term $\int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx < \frac{1}{\mu} < \infty$ for *every* trio of positive numbers $\{\lambda, \mu, r\}$, since the integrand $e^{-\mu x - \frac{\lambda}{r} e^{-rx}} < e^{-\mu x}$. Thus $P_0 > 0$ for all positive $\{\lambda, \mu, r\}$. In particular $P_0 > 0$ for *every* arrival rate $\lambda > 0$. This corroborates Theorem 3.8 with $\lim_{x\to\infty} \overline{R}(x) = L = 0$.

Expected Busy Period

In the standard M/G/1 queue, $E(\mathcal{B}) = \frac{E(S)}{1-\lambda E(S)}$, where $\mathcal{B}$ is the busy period. In M/G/1 with reneging $P_0 \neq 1 - \lambda E(S)$. Hence, we use the more fundamental formula for $E(\mathcal{B})$ in terms of P_0. From (3.60) and (3.177),

$$E(\mathcal{B}) = \frac{1-P_0}{f(0)} = \frac{1-P_0}{\lambda P_0}$$
$$= e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx = \int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1-e^{-rx})} dx. \tag{3.178}$$

(Note that (3.178) is part of the denominator of (3.177). This infers (3.178.)

3.11.8 M/M/1 with Reneging and Standard M/M/1

We compare M/M/1 with reneging and the standard M/M/1 queue. Assume $\lambda < \mu$ (stability condition for standard M/M/1). In (3.178), $(1 - e^{-rx}) < rx \; \forall x > 0$ and $(1 - e^{-r\cdot 0}) = r \cdot 0 = 0$. Thus

$$E(\mathcal{B}_{\mathrm{r}}) < \int_{x=0}^{\infty} e^{-\mu x + \lambda x} dx = \frac{1}{\mu - \lambda} = E(\mathcal{B}_{\mathrm{s}}),$$

where subscript r represents M/M/1 with reneging, and subscript s represents standard M/M/1.

In (3.177), we again apply the inequality

$$\int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1-e^{-rx})} dx < \frac{1}{\mu - \lambda}.$$

This gives

$$P_{\mathrm{r}0} > \frac{1}{1 + \lambda \cdot \frac{1}{\mu - \lambda}} = 1 - \frac{\lambda}{\mu} = P_{\mathrm{s}0}.$$

The comparisons for $E(\mathcal{B})$ and P_0 are intuitive. The effective arrival rate of customers that increase workload on the server, is less in the reneging model than in the standard model.

3.11.9 Number in System for M/M/1 with Reneging

Let P_{sn}, a_{sn}, d_{sn} 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 $P_{sn} = a_{sn} = d_{sn}$, $n = 0, 1, 2, ...$, and $P_{s,0} = P_0$ given in (3.177). Furthermore

$$\begin{aligned} d_{sn} &= \int_{x=0}^{\infty} e^{-\lambda\overline{R}(x)x} \frac{(\lambda\overline{R}(x)x)^{n-1}}{(n-1)!} f(x)dx \\ &= \int_{x=0}^{\infty} e^{-\lambda e^{-rx}x} \frac{(\lambda e^{-rx}x)^{n-1}}{(n-1)!} \lambda e^{\frac{\lambda}{r}} P_0 e^{(-\mu x - \frac{\lambda}{r}e^{-rx})}, n = 1, 2, ... \end{aligned} \tag{3.179}$$

In formula (3.179), $\lambda\overline{R}(x)$ $(= \lambda e^{-rx})$ is the arrival rate of stayers when the required wait is x.

Remark 3.32 *We outline a derivation of (3.179) 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 n subintervals $\Delta_i = [x_i, x_{i+1}), i = 0, ..., m-1$, where $x_0 = 0$, $x_n = \Omega$. We approximate $\overline{R}(x)$ by $\overline{R}(x) \equiv \overline{R}(x_i), x \in \Delta_i$. Thus the arrival rate of stayers is a constant $\lambda\overline{R}(x_i)$ if the required wait $\in [x_i, x_{i+1})$. The probability that $n-1$ stayers arrive given the required wait $\in \Delta_i$ 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 unconditional 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 $n \to \infty$ and $|\Delta_i| \downarrow 0$. Then x_i, x_i', $x_i'' \to x$ and

$$\begin{aligned} &\lim_{m\to\infty, |\Delta_i|\downarrow 0} \sum_{i=0}^{m-1} e^{-\lambda\overline{R}(x_i)x_i'} \frac{(\lambda\overline{R}(x_i)x_i')^{n-1}}{(n-1)!} f(x_i'')\,|\Delta_i| \\ &= \int_{x=0}^{\Omega} e^{-\lambda\overline{R}(x)x} \frac{(\lambda\overline{R}(x)x)^{n-1}}{(n-1)!} f(x)dx. \end{aligned}$$

Letting $\Omega \to \infty$ implies (3.179).

3.11.10 Proportion of Customers Served

Consider M/M/1 with exponential reneging. From (3.163) the proportion of customers that get complete service is

$$q_S = P_0 + \int_{x=0}^{\infty} e^{-rx} f(x)dx$$
$$= \frac{\left(1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx} - rx} dx\right)}{\left(1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx\right)}. \tag{3.180}$$

The proportion of customers that renege while waiting, or reach the server and balk at service, is $1 - q_S$.

In the expressions for P_0, $E(\mathcal{B})$, q_S the integrals do not have closed forms. They can be evaluated readily using series expansion or numerical methods, for given values of λ, μ, r.

3.12 M/G/1 with Priorities

Assume N types of customers arrive at a single-server system in independent Poisson streams at rates λ_i, $i = 1, ..., N$. The respective service times S_i have cdf $B_i(x)$, $\overline{B}_i(x) = 1 - B_i(x), x \geq 0$, 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. Any customer that starts service is allowed to complete it. 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. Note that the probability of a zero wait P_0 is independent of type.

3.12.1 Two Priority Classes

For exposition we consider two priority classes. If there are two priority classes, $N = 2$. We confirm the well known stability condition, $\lambda_1 E(S_1) + \lambda_2 E(S_2) < 1$, using an LC approach. Consider sample paths of the virtual wait for *type*-1 *customers* (Fig. 3.25). Fix level $x > 0$ in the state space.

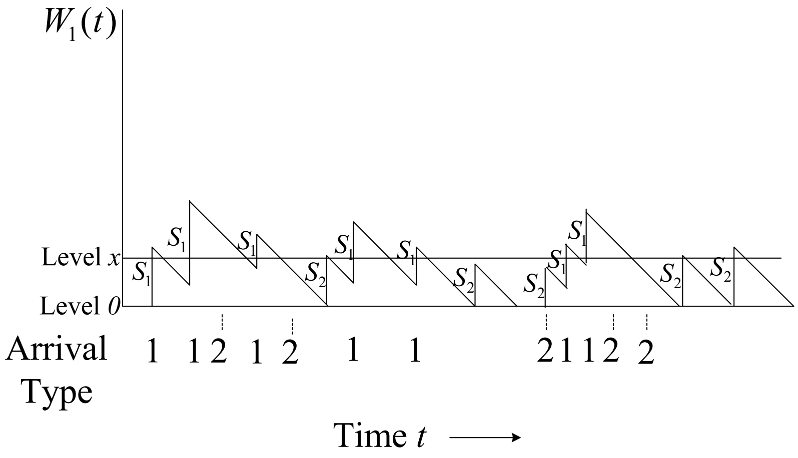

Figure 3.25: Sample path of virtual wait for high priority type-1 arrivals. Low priority type-2 arrivals that must wait, start service at the end of a $\mathcal{B}_1$ or a $\mathcal{B}_{21}$ (Fig. 3.26) busy period. All type 2 jumps start at level 0.

3.12.2 Equation for PDF of Wait of Type-1 Customers

From the sample path, we construct an integral equation for the pdf $f_1(x)$, $x > 0$,

$$f_1(x) = \lambda_1 \overline{B}_1(x)P_0 + \lambda_2 \overline{B}_2(x)P_0 + \lambda_1 \int_{y=0}^{x} \overline{B_1}(x-y)f_1(y)dy + \lambda_2(1-P_0)\overline{B}_2(x). \tag{3.181}$$

In (3.181) the left side $f_1(x)$ is the SP downcrossing rate of x (as in basic LC Theorem 1.1). On the right side terms $\lambda_1\overline{B}_1(x)P_0$, $\lambda_2\overline{B}_2(x)P_0$ are the SP upcrossing rates of x due to type-1 and type-2 arrivals, respectively, 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 2's must wait *until the end* of a type 1 busy period to start service. Any other such type 2's wait longer before they start service. Those type 2's 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 2's start service is $\lambda_2(1-P_0)$ since all type 2's must eventually get served in a finite time, due to stability.

3.12.3 Stability Condition

Integrate both sides of (3.181) with respect to x on $(0,\infty)$. Note that $\int_{x=0}^{\infty} f_1(x)dx = 1 - P_0$. Collect terms to yield

$$P_0 = 1 - \lambda_1 E(S_1) - \lambda_2 E(S_2) = 1 - \rho_1 - \rho_2, \tag{3.182}$$

where $\rho_i = \lambda_i E(S_i)$, $i = 1, 2$. For stability, we must have $0 < P_0 < 1$, or

$$0 < \rho_1 + \rho_2 < 1,$$

which implies both $\rho_1 < 1$ and $\rho_2 < 1$.

3.12.4 Expected Wait of High Priority Customers

We confirm the known formula for the expected wait of type-1 customers using (3.181). Denote the wait in queue before service of an arbitrary type-1 arrival by W_{q1}. Multiply both sides of (3.181) by x and integrate on $(0,\infty)$. The left side becomes $\int_0^{\infty} x f_1(x)dx = E(W_{q1})$. We obtain

$$\begin{aligned} E(W_{q1}) = & \left(\lambda_1 \tfrac{E(S_1^2)}{2} + \lambda_2 \tfrac{E(S_2^2)}{2}\right) P_0 + \lambda_1 E(S_1) E(W_{q1}) \\ & + \lambda_1 (1 - P_0) \tfrac{E(S_1^2)}{2} + \lambda_2 (1 - P_0) \tfrac{E(S_2^2)}{2}. \end{aligned}$$

or, the familiar result (e.g., [91])

$$E(W_{q1}) = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1 - \rho_1)}. \tag{3.183}$$

3.12.5 Equation for PDF of Wait of Type-2 Customers

Let $\{W_2(t)\}$ be the virtual wait process of type-2 customers. Let W_{q2} be the steady-state wait. Denote the pdf of W_{q2} by $f_2(x)$, $x > 0$. We now develop an integral equation for $f_2(x)$.

<u>**Preliminaries**</u>

Let $\mathbf{B}_1(x), x > 0$ denote the cdf of an M/G/1 type-1 busy period. Let $\overline{\mathbf{B}}_1(x) = 1 - \mathbf{B}_1(x)$. We use $\mathcal{B}_{21}$ to denote a busy period in which the first service is type 2. All linked subsequent services are type 1 (Fig. 3.26). Let random variable $N_{S_{21}}$ denote the number of strict descending ladder points that occur in a sample path of a $\mathcal{B}_{21}$ busy period. Then $N_{S_{21}}$ has the same distribution as the number of type-1 customers that arrive in a type-2 *service time* S_2. Thus we have

$$\mathcal{B}_{21} \underset{dist}{=} S_2 + \sum_{i=1}^{N_{S_{21}}} \mathcal{B}_{1i}, \tag{3.184}$$

where the $\mathcal{B}_{1i}$'s are iid random variables distributed as an M/G/1 type-1 busy period $\mathcal{B}_1$ independent of $N_{S_{21}}$. Equation (3.184) follows due to the memoryless property of the type-1 inter-arrival times (exponential with rate λ_1). (A related discussion of busy period structure is given above in Subsection 3.3.9.)

We illustrate the meaning of $N_{S_{21}}$ in Fig. 3.26. In that figure $N_{S_{21}} = 3$. There are three type-1 busy periods in B_{21}. There are four vertical gaps, each distributed as an inter-arrival time, separating and bordering on these three busy periods. The basic observation is that the sum of the four gaps is equal to S_2.

From (3.59)

$$E(\mathcal{B}_1) = \frac{E(S_1)}{1 - \lambda_1 E(S_1)}. \tag{3.185}$$

Taking expected values in (3.184) we obtain

$$\begin{aligned} E(\mathcal{B}_{21}) &= E(S_2) + \lambda_1 E(S_2) E(\mathcal{B}_1) \\ &= E(S_2) + \lambda_1 E(S_2) \frac{E(S_1)}{1 - \lambda_1 E(S_1)} \\ &= \frac{E(S_2)}{1 - \lambda_1 E(S_1)} = \frac{E(S_2)}{1 - \rho_1}. \end{aligned} \tag{3.186}$$

Remark 3.33 *Note that $E(\mathcal{B}_{21})$ is the same as the expected busy period in an M/G/1 queue in which zero-waiting customers receive specialized service. Thus we can obtain (3.186) immediately as a special case of (3.122).*

Let $\mathbf{B}_{21}(x)$ denote the cdf of $\mathcal{B}_{21}$, and $\overline{\mathbf{B}}_{21}(x) = 1 - \mathbf{B}_{21}(x)$, $x \geq 0$. Consider a sample path of the virtual wait of type-2 customers $\{W_2(t)\}$ (Fig. 3.27). The sample path illustrates that *type*-2 customers may view the model as a queue with server vacations. When a type 1 arrives to an empty system, the server vacation is a type-1 busy period. When a type 2 arrives, the server vacation consists of $N_{S_{21}}$ type-1 busy periods. By (3.184) type 2 generated SP jumps are distributed as $\mathcal{B}_{21}$.

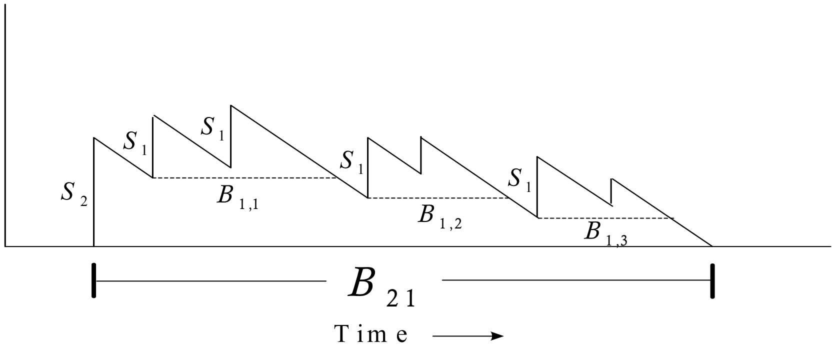

Figure 3.26: Busy period $\mathcal{B}_{21}$. Initial jump is a type 2 service S_2. Each subsequent jump is a type 1 service S_1. $\mathcal{B}_{1,j}$, $j = 1, 2, ...$, are M/G/1 type 1 busy periods.

Integral Equation for $f_2(x)$

We now construct an integral equation for $f_2(x)$, namely

$$f_2(x) = \lambda_1 \overline{\mathbf{B}}_1(x)P_0 + \lambda_2 \overline{\mathbf{B}}_{21}(x)P_0 + \lambda_2 \int_{y=0}^{x} \overline{\mathbf{B}}_{21}(x-y) f_2(y)dy. \quad (3.187)$$

In (3.187) the left side $f_2(x)$ is the sample-path downcrossing rate of level x (as in basic LC Theorem 1.1). On the right side of (3.187) the term $\lambda_1 \overline{\mathbf{B}}_1(x)P_0$ 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 $\lambda_2 \overline{\mathbf{B}}_{21}(x)P_0$ is the SP upcrossing rate of x due to type-2 arrivals when the system is empty. A potentially arriving type-2 customer, immediately after the type 2 starts service, would wait a busy period, $\mathcal{B}_{21}$, before starting service. It is possible that $\mathcal{B}_{21}$ consists of the initial type-2 service only. Possibly no type 1's arrive during the initial service time. Generally, $\mathcal{B}_{21}$ includes an additional run of $N_{S_{21}}$ M/G/1 type-1 busy periods (Fig. 3.26). The term $\lambda_2 \int_{y=0}^{x} \overline{\mathbf{B}}_{21}(x-y) f_2(y)dy$ 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 busy period $\mathcal{B}_{21}$, before starting service.

The three terms on the right of (3.187) account for all arrivals to the system. The type 2's are counted in the last two terms. These terms include all type 2's that wait ≥ 0. The type 1's are counted in all three

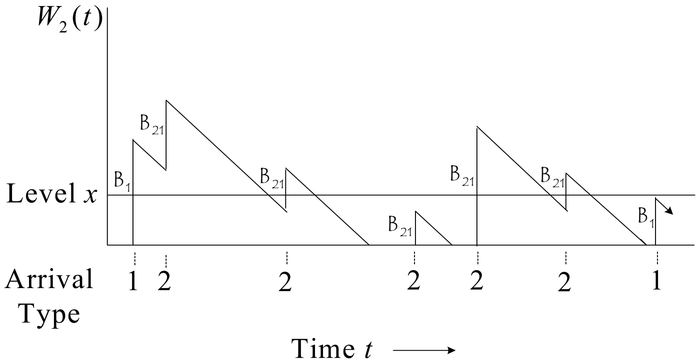

Figure 3.27: Sample path of virtual wait for *low priority, type 2 arrivals.* High priority type 1's that arrive when the system is empty generate jumps distributed as $\mathcal{B}_1$busy periods. All type 2 arrivals generate jumps distributed as $\mathcal{B}_{21}$ busy periods (see Fig. 3.26). All type 1's that must wait, are counted in the $\mathcal{B}_{21}$ jumps.

terms. The type 1's that wait zero are counted in the first term. The type 1's that wait a positive time are counted in all three terms.

Both Types Have Same P_0

We test for consistency of integral equations (3.187) and (3.181), by checking whether they give the same value of P_0. It is required to show that (3.182) results from (3.187). We integrate both sides of (3.187) 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 (3.185), (3.186) 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 (3.182); QED.

3.12.6 Expected Wait of Type-2 Customers

We obtain the expected wait $E(W_{q2})$ of *type*-2 *customers*. We multiply integral equation (3.187) by x on both sides and integrate with respect to x on $(0,\infty)$. Some algebra gives

$$\begin{aligned} E(W_{q2}) = \quad & \lambda_1 \tfrac{E(\mathcal{B}_1^2)}{2} P_0 + \lambda_2 \tfrac{E(\mathcal{B}_{21}^2)}{2} P_0 \\ & + \lambda_2 \tfrac{E(\mathcal{B}_{21}^2)}{2} (1-P_0) + \lambda_2 E(\mathcal{B}_{21}) E(W_{q2}) \end{aligned}$$

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 (3.62), (3.182) and (3.186) gives

$$E(W_{q2}) = \frac{\left(\lambda_1 \frac{E(S_1^2)}{(1-\rho_1)^3} (1-\rho_1-\rho_2) + \lambda_2 E(\mathcal{B}_{21}^2)\right) \cdot (1-\rho_1)}{2\,(1-\rho_1-\rho_2)}. \qquad (3.188)$$

The term $\lambda_2 E(\mathcal{B}_{21}^2)$ in the numerator of (3.188) is

$$\begin{aligned} \lambda_2 E(\mathcal{B}_{21}^2) &= \lambda_2 E\left(\left(S_2 + \sum_{i=1}^{N_{S_{2,1}}} \mathcal{B}_{1,i}\right)^2\right) \\ &= \lambda_2 E(S_2^2) + 2\lambda_2 E\left(S_2 \sum_{i=1}^{N_{S_{21}}} \mathcal{B}_{1,i}\right) + \lambda_2 E\left(\left(\sum_{i=1}^{N_{S_{21}}} \mathcal{B}_{1,i}\right)^2\right). \end{aligned}$$

We condition on $N_{S_{21}} = n, S_2 = s$ in the last two terms. Then $N_{S_{21}}$ is a Poisson random variable with parameter $\lambda_1 s$. We then carry out some algebra, and "uncondition". This procedure yields

$$\begin{aligned} \lambda_2 E(\mathcal{B}_{21}^2) &= \lambda_2 E(S_2^2) + 2\lambda_2 E(S_2^2) \frac{\rho_1}{1-\rho_1} \\ & \quad + \lambda_2(\lambda_1 E(S_2) E(\mathcal{B}_1^2) + \lambda_1^2 \left(E(\mathcal{B}_1)\right)^2 E(S_2^2)). \end{aligned}$$

Substituting from (3.62) into the last equation gives

$$\begin{aligned} \lambda_2 E(\mathcal{B}_{21}^2) = \quad & \lambda_2 E(S_2^2) + 2\lambda_2 E(S_2^2) \tfrac{\rho_1}{1-\rho_1} \\ & + \rho_2 \lambda_1 \tfrac{E(S_1^2)}{(1-\rho_1)^3} + \lambda_2 \tfrac{\rho_1^2}{(1-\rho_1)^2} E(S_2^2). \end{aligned} \qquad (3.189)$$

Substituting the expression in (3.189) for $\lambda_2 E(\mathcal{B}_{21}^2)$ in the numerator of (3.188) gives

$$\begin{aligned}
\text{coefficient of } (E(S_1^2)) &= \frac{\lambda_1}{(1-\rho_1)}, \\
\text{coefficient of } (E(S_2^2)) &= \frac{\lambda_2}{(1-\rho_1)}.
\end{aligned}$$

Hence

$$\begin{aligned}
E(W_{q2}) &= \frac{\frac{\lambda_1}{(1-\rho_1)} E(S_1^2) + \frac{\lambda_2}{(1-\rho_1)} E(S_2^2)}{2\,(1-\rho_1-\rho_2)} \\
&= \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1-\rho_1)\,(1-\rho_1-\rho_2)}, \qquad (3.190)
\end{aligned}$$

which agrees with the known result in the literature.

Remark 3.34 *We have used LC to derive $E(W_{q1})$ from the integral equation for $f_1(x)$, and $E(W_{q2})$ from the integral equation for $f_2(x)$. The importance of this approach is that we essentially have an analytic solution for the pdf's and cdf's of wait of both priority classes. The LC analysis is in the time domain without use of transforms. Integral equations (3.181), (3.187) 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.*

3.12.7 Exponential Service

We solve for the steady-state pdf of wait for high priority customers $\{P_0, f_1(x), x > 0\}$ when inter-arrival and service times are exponentially distributed. Assume the service times of type-1 and type-2 arrivals are exponentially distributed with rates μ_1 and μ_2, respectively. Substituting from the exponential cdf's into (3.181) gives an integral equation for $f_1(x)$,

$$\begin{aligned}
f_1(x) = \quad & \lambda_1 e^{-\mu_1 x} P_0 + \lambda_2 e^{-\mu_2 x} P_0 + \lambda_1 \int_{y=0}^{x} e^{-\mu_1 (x-y)} f_1(y) dy \\
& + \lambda_2 (1-P_0) e^{-\mu_2 x}.
\end{aligned} \qquad (3.191)$$

We apply differential operator $\langle D+\mu_1\rangle\langle D+\mu_2\rangle$ to both sides of (3.191). This operation gives the second order differential equation

$$\langle D+\mu_2\rangle\langle D+\mu_1-\lambda\rangle f_1(x)=0,$$

with solution

$$f_1(x)=ae^{-(\mu_1-\lambda_1)x}+be^{-\mu_2 x}, x\geq 0, \tag{3.192}$$

where constants a, b are to be determined.

Let $x\downarrow 0$ in (3.191) and (3.192). We get equation

$$a+b=\lambda_1 P_0+\lambda_2. \tag{3.193}$$

Take $\frac{d}{dx}$ on both sides of (3.191) and let $x\downarrow 0$. This gives

$$f_1'(0)=-\lambda_1\mu_1 P_0+\lambda_1^2 P_0+\lambda_1\lambda_2-\lambda_2\mu_2. \tag{3.194}$$

Take $\frac{d}{dx}$ in (3.192) and let $x\downarrow 0$. Equating to (3.194) we get

$$-(\mu_1-\lambda_1)a-\mu_2 b=-\lambda_1\mu_1 P_0+\lambda_1^2 P_0+\lambda_1\lambda_2-\lambda_2\mu_2. \tag{3.195}$$

We use (3.192) and the normalizing condition $P_0+\int_{x=0}^{\infty}f_1(x)dx=1$ to obtain

$$P_0+\frac{a}{\mu_1-\lambda_1}+\frac{b}{\mu_2}=1. \tag{3.196}$$

We now solve the system of three equations (3.193), (3.195), (3.196) for P_0, a, b to obtain

$$P_0 = \frac{(\mu_2\mu_1-\mu_2\lambda_1-\mu_1\lambda_2)}{\mu_2\mu_1}, \tag{3.197}$$

$$a = \frac{\lambda_1(\mu_2\mu_1^2+2\mu_2\mu_1\lambda_1+\mu_2^2\mu_1-\mu_2\lambda_1^2-\mu_2^2\lambda_1+\mu_1^2\lambda_2-\mu_1\lambda_2\lambda_1)}{(-\mu_1+\lambda_1+\mu_2)\mu_2\mu_1}, \tag{3.198}$$

$$b = \frac{\lambda_2(\mu_2-\mu_1)}{(-\mu_1+\lambda_1+\mu_2)}. \tag{3.199}$$

Check on Values

We conduct a mild check (indicated by ✓) on the values of P_0, a, b. Set $\lambda_2=0$. The model reverts to a standard $\mathrm{M}_{\lambda_1}/\mathrm{M}_{\mu_1}/1$ queue. In that model the steady-state absolutely continuous part of the pdf of wait $f(x)$, and P_0 are given in (3.86) and (3.87).

Substituting $\lambda_2=0$ in (3.197), (3.198), (3.199) respectively yields: $P_0=1-\frac{\lambda_1}{\mu_1}$; $a=\lambda_1\left(1-\frac{\lambda_1}{\mu_1}\right)$; $b=0$ ✓.

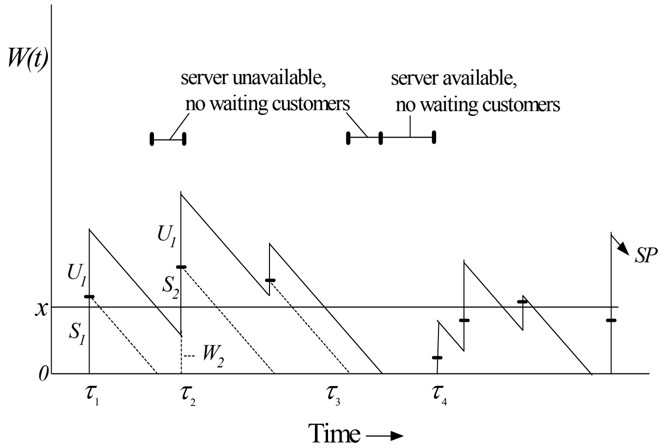

Figure 3.28: Sample path of virtual wait in M/G/1 queue with a server vacation after each service completion.

3.13 M/G/1 with Server Vacations

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 (Fig. 3.28).

Denote the complementary cdf of $S + U$ by $\overline{B * V}(x)$. An integral equation for the steady-state pdf of wait $f(x)$ is

$$f(x) = \lambda P_0 \overline{B * V}(x) + \lambda \int_{y=0}^{x} \overline{B * V}(x - y) f(y) dy, x \geq 0. \qquad (3.200)$$

In (3.200) 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 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 in state-space interval $(0, x)$.

Comparing (3.200) and (3.29) indicates that the server-vacation and standard M/G/1 models are equivalent with regard to the integral equation for the pdf of wait in the queue; only the "service time" cdf's differ.

3.13.1 Probability of Zero Wait

Let P_0 denote the steady-state probability that an arrival waits zero time for service. Since the queue behaves like an $M_\lambda/G/1$ queue with service time $S+U$, with respect to the customer wait for service, then

$$P_0 = 1 - \lambda E(S+U)$$

provided $\lambda E(S+U) < 1$.

3.13.2 Expected Busy and Idle Period

Define the idle period I as the time interval when the server is available to start service and no customers are waiting. Then $E(I) = \frac{1}{\lambda}$ (memoryless property). Let $\mathcal{B}_s$ = time that the server is busy serving a customer, $\mathcal{B}_u$ = 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_λ/G/1 queue with service time $S+U$. Hence

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

Given the server is "busy", the pairs $\{S_i, U_i\}, i = 1, 2, ...$, form an alternating renewal process (Fig. 3.28). 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

$$E(\mathcal{B}_s) = \frac{E(S)}{E(S)+E(U)} \cdot E(\mathcal{B}),\ E(\mathcal{B}_u) = \frac{E(U)}{E(S)+E(U)} \cdot E(\mathcal{B}),$$

or

$$E(\mathcal{B}_s) = \frac{E(S)}{1-\lambda E(S+U)},\ E(\mathcal{B}_u) = \frac{E(U)}{1-\lambda E(S+U)}.$$

3.13.3 Number in System

Let d_n denote the probability of n customers in the system *just after the server returns from vacation.* Then

$$d_n = \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} f(x)dx.$$

Let a_n denote the probability that an arrival "sees" n customers in the system. Then $a_n = d_n$ due to Poisson arrivals.

3.13.4 M/M/1 with Server Vacations

Let $\overline{V}(x) = e^{-\nu x}$, $\overline{B}(x) = e^{-\mu x}, x \geq 0$. Assume $\nu \neq \mu > 0$. Then

$$\overline{B * V}(x) = \frac{(\mu e^{-\nu x} - \nu e^{-\mu x})}{\mu - \nu}, x \geq 0,$$

and (3.200) reduces to

$$\begin{aligned} f(x) = \quad & \lambda P_0 \tfrac{\left(\mu e^{-\nu x} - \nu e^{-\mu x}\right)}{\mu - \nu} \\ & + \lambda \tfrac{1}{\mu-\nu} \int_{y=0}^{x} \left(\mu e^{-\nu(x-y)} - \nu e^{-\mu(x-y)}\right) f(y) dy, \ x \geq 0. \end{aligned} \tag{3.201}$$

In (3.201), applying differential operator $\langle D+\nu\rangle\langle D+\mu\rangle$ to both sides results in a second-order 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

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

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

$$\begin{aligned} c_1 &= \lambda P_0 \frac{\lambda - R_2}{R_1 - R_2}, \ c_2 = -\lambda P_0 \frac{-R_1 + \lambda}{R_1 - R_2}, \\ P_0 &= \frac{c_1 R_2 + c_2 R_1 + R_1 R_2}{R_1 R_2}. \end{aligned}$$

Busy Period

The expected values of $\mathcal{B}$, $\mathcal{B}_s$, $\mathcal{B}_u$ are

$$E(\mathcal{B}) = \frac{\frac{1}{\mu} + \frac{1}{\nu}}{1 - \lambda\left(\frac{1}{\mu} + \frac{1}{\nu}\right)},$$

$$E(\mathcal{B}_s) = \frac{\frac{1}{\mu}}{\frac{1}{\mu} + \frac{1}{\nu}} E(\mathcal{B}), \ E(\mathcal{B}_u) = \frac{\frac{1}{\nu}}{\frac{1}{\mu} + \frac{1}{\nu}} E(\mathcal{B}).$$

Number in System

The probability that the server finds n in the system just after a vacation is for $n = 1, 2, ...,$

$$\begin{aligned} d_n &= \int_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} \left(c_1 e^{R_1 x} + c_2 e^{R_2 x}\right) dx \\ &= \frac{1}{\lambda}\left(\left(\frac{\lambda}{\lambda - R_1}\right)^n c_1 + \left(\frac{\lambda}{\lambda - R_2}\right)^n c_2\right), \end{aligned}$$

where R_i, c_i, $i = 1, 2$ are given in Subsection 3.13.4. The probability that an arrival "sees" n customers in the system is $a_n = d_n$.

3.14 M/G/1 with Bounded System Time

We provide two M/G/1 variants having virtual wait bounded by a constant $K > 0$. These models are of inherent interest. Among other properties, they demonstrate the existence of models which are useful in the proof of Proposition 9.1 (Chapter 9). When $K \to \infty$, both variants become a standard M/G/1 queue. Let the arrival rate be λ and the cdf of service $B(\cdot)$ with $\overline{B}(\cdot) = 1 - B(\cdot)$.

3.14.1 Variant 1

Assume that for each customer, wait plus service $< K$. Thus all waiting times (before service) are $< K$. A customer **reneges from service** when its total system time reaches K. The virtual wait $W(t) \le K, t \ge 0$. Customers that complete their service have system times $< K$. Consider a sample path of $\{W(t)\}$ (Fig. 3.29). Using rate balance across level x we immediately obtain an integral equation for the steady-state pdf of wait, $f(x)$, as

$$f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y) dy, 0 < x < K. \qquad (3.202)$$

The normalizer is

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

The solution for $f(x)$ approaches that of a standard M/G/1 model as $K \to \infty$ (compare equations (3.29)-(3.31)).

3.14.2 Variant 1: M/M/1 Model

If the queue is an $M_\lambda/M/_\mu 1$ model, the solution of (3.202) together with the normalizer is

$$\begin{aligned} f(x) = & \quad \lambda P_0 e^{-(\mu-\lambda)x}, 0 < x < K, \\ P_0 = & \quad \frac{\mu - \lambda}{\mu + e^{-(\mu-\lambda)K}}. \end{aligned} \tag{3.203}$$

If $K \to \infty$ then $P_0 \to 1 - \frac{\lambda}{\mu}$ and the range of $f(\cdot) \to (0, \infty)$. This is the solution for a standard $M_\lambda/M/_\mu 1$ queue.

3.14.3 Variant 2

Assume customers **balk upon arrival** if their system time would be $\geq K$. We assume system time is known by a "system manager", at each arrival instant. The virtual wait $W(t) < K, t \geq 0$. Customers that wait, receive full service and depart *before* their system times reach K. Consider a sample path of $\{W(t)\}$ (Fig. 3.30). We obtain via LC analysis an integral equation for $f(x)$,

$$\begin{aligned} f(x) = & \quad \lambda P_0 \left(\overline{B}(x) - \overline{B}(K)\right) \\ & +\lambda \textstyle\int_{y=0}^{x} \left(\overline{B}(x-y) - \overline{B}(K-y)\right) f(y)dy, 0 < x < K, \end{aligned} \tag{3.204}$$

and normalizer

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

3.14.4 Variant 2: M/M/1 Model

If variant 2 is an $M_\lambda/M/_\mu 1$ model, we obtain the solution of (3.202) and the normalizer as a special case of the M/M/c queue with bounded system time given in Example 1 of [38], with the number of servers = 1. We get the solution

$$\begin{aligned} f(x) = & \quad \lambda e^{\rho b} P_0 e^{\mu(\rho-1)x}(1 - be^{\mu x})e^{-\mu b e^{\mu x}}, 0 < x < K, \\ P_0 = & \quad \frac{1}{1 + \lambda e^{\rho b} \int_{x=0}^{K} e^{\mu(\rho-1)x}(1 - be^{\mu x})e^{-\mu b e^{\mu x}} dx}, \end{aligned} \tag{3.205}$$

where $\rho = \frac{\lambda}{\mu}$, $b = e^{-\mu K}$. This single-server Markovian result is also obtained in [59]. The solution (3.205) is more complex than the solution (3.203) for variant 1.

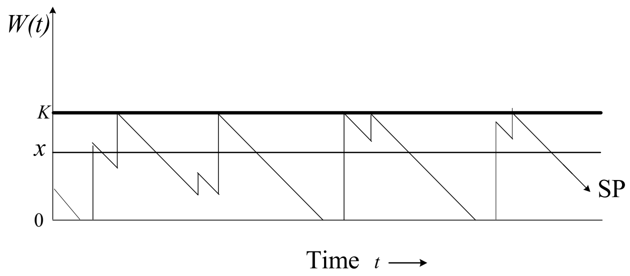

Figure 3.29: Variant 1. Sample path of virtual wait in M/G/1 with bounded virtual wait (bounded system time)

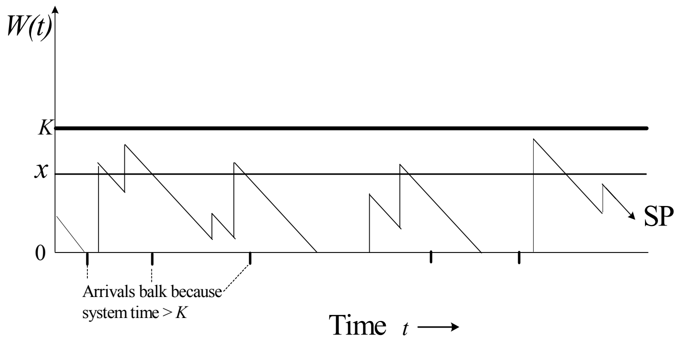

Figure 3.30: Variant 2. Sample path of virtual wait in M/G/1 with bounded virtual wait (boundrd system time)

If $K \to \infty$ then $b \downarrow 0$. We get

$$f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0, \quad P_0 = 1 - \frac{\lambda}{\mu}$$

as in the standard M/M/1 queue.

3.14.5 Convergence to Standard M/G/1

Variants 1 and 2 have different steady-state pdf's of wait when K is finite. Let $K \to \infty$. In variant 1 no one reneges from service. In variant 2 no one balks at arrival. Both variants "converge" to a standard M/G/1 queue as $K \to \infty$. We have given explicit examples of this convergence for M/M/1 with bounded system time.

3.15 PDF of Wait and Busy-period Structure

We shall utilize the busy-period structure of M/G/1 to write a series for the pdf of wait in the M/G/1 queue *by inspection*. This technique allows us to write an analogous series for model variants as well. We will illustrate the series for a model with balking and where zero-wait stayers receive special service.

3.15.1 Model Description

Let the arrival rate be λ. Arrivals balk with probability β_0 $(\overline{\beta_0} = 1-\beta_0)$ if their required wait is zero, and with probability β_1 $(\overline{\beta_1} = 1-\beta_1)$ if their required wait is positive. Joiners (stayers) that wait zero receive a service time $\underset{dist}{=} S_0$. Joiners that wait a positive time before service receive a service time $\underset{dist}{=} S_1$. Let the cdf of S_i be $B_i(x), x \geq 0, i = 0,1$ $(\overline{B_i}(x) = 1 - B_i(x))$. Define $\lambda_i = \lambda\overline{\beta_i}, i = 0,1$. Let $\rho_i = \lambda_i E(S_i), i = 0,1$. Denote the steady-state pdf of stayers by $\{P_0; f(x), x > 0\}$. An integral equation for $f(x)$ is

$$f(x) = \lambda_0 P_0 \overline{B_0}(x) + \lambda_1 \int_{y=0}^{x} \overline{B_1}(x-y) f(y) dy, x > 0. \tag{3.206}$$

Upon integrating both sides of (3.206) with respect to $x \in (0,\infty)$ we obtain

$$P_0 = \frac{1-\rho_1}{1-\rho_1+\rho_0}. \tag{3.207}$$

3.15.2 Busy Period Structure

Consider Fig. 3.31. Fix level $x > 0$. The SP upcrossing rate of level x due to arrivals that initiate generation-1 busy periods is $\lambda_0 P_0 \overline{B_0}(x)$. The SP upcrossing rate of x due to arrivals that initiate generation-2 busy periods is

$$\lambda_0 P_0 \lambda_1 E(S_0) \left(g_0 * \overline{B_1}\right)(x) = P_0 \rho_0 \rho_1 \left(g_0 * g_1\right)(x), \tag{3.208}$$

where $g_i(\cdot)$ is the pdf of the remaining service time of a type-$S_i, i = 0,1$, and "$*$" denotes convolution.

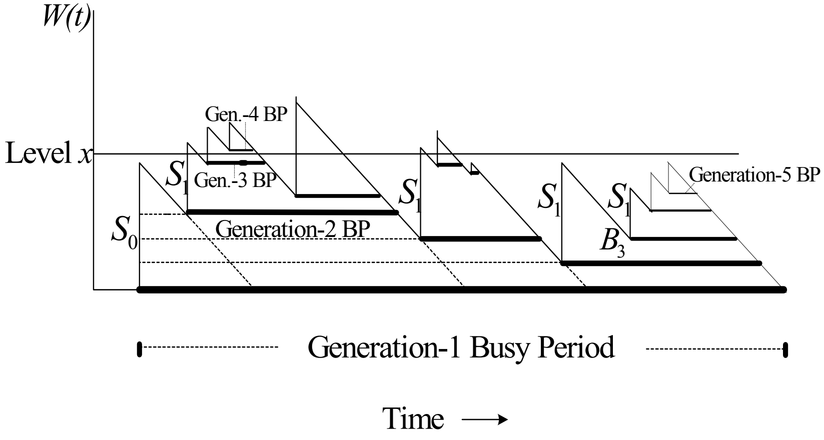

Figure 3.31: Multiplicative structure of busy-period. Each arrival generates an initial jump of a busy period of some generation. Initial busy periods of all generations account for all arrivals.

Explanation of (3.208)

Due to Poisson arrivals, the ordinates of the starts of the initial jumps of the generation-2 busy periods (their base ordinates) are distributed as independent Poisson arrivals at rate λ_1, in S_0. Thus the expected number of generation-2 busy periods within a type-1 busy period, is $\lambda_1 E(S_0)$. The generation-2 base ordinates are $\underset{dist}{=} g_0(\cdot)$ (PASTA principle). The initial jump of each generation-2 busy period is $\underset{dist}{=} S_1$. Hence the probability of an upcrossing of level x due to generation-2 initial jumps is $\left(g_0 * \overline{B_1}\right)(x)$. However, from renewal theory, $g_1(x) = \frac{1}{E(S_1)}\overline{B_1}(x)$. Therefore, multiplying and dividing the left side of (3.208) by $E(S_1)$ results in the right side of (3.208).

By similar reasoning, it is seen that the SP upcrossing rate of x due to arrivals that initiate generation-3 busy periods is

$$\lambda_0 P_0 \lambda_1 E(S_0)\left((g_0 * g_1) * \overline{B_1}\right)(x) = P_0 \rho_0 \rho_1^2 \left(g_0 * g_1^{(2)}\right)(x), \qquad (3.209)$$

where $g_1^{(2)}(\cdot)$ is the two-fold convolution of $g_1(\cdot)$.

3.15.3 Multiplicative Structure of PDF of Wait

By a recursive argument, it is seen that the pdf

$$f(x) = P_0\rho_0 \sum_{k=1}^{\infty} \rho_1^{k-1} \left(g_0 * g_1^{(k-1)}\right)(x), \tag{3.210}$$

where $g_1^{(k-1)}(\cdot)$ is the $(k-1)$-fold convolution of $g_1(\cdot)$.

In (3.210) the k^{th} term is the SP upcrossing rate of level x due to initial jumps of the generation-k busy periods. From Fig. 3.31 we see that every arrival is the first customer of some generation-k busy period. Hence, the initial jumps of the generation-k busy periods, $k = 1, 2, ...$, *account for all arrivals* to the system. In (3.210), the left side is the SP downcrossing rate of level x. Hence, (3.210) is an alternative way of writing the balance equation for $f(x)$. Due to the geometric factor ρ_1^{k-1}, the series converges rapidly to $f(x)$, in most situations. This series bypasses the standard Volterra integral equation for the pdf. In fact, the right side is a series expansion of the integral. By approximating the convolutions $\left(g_0 * g_1^{(k-1)}\right)(x), k - 1, 2, ...$, we can quickly arrive at an estimate of $f(x)$.

Note that for the standard M/G/1 queue, the series (3.210) reduces to (3.53).

3.16 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. We have emphasized steady-state results. Many of the derived properties have been obtained in the literature by different methods. Some properties and results given here are new. A vast array of additional models and variants have been analyzed using LC. We mention just a few.

M/G/1 with Markov-generated server vacations [29] 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. Such dependency arises in many situations. A teller in a bank may have to do paper work following a service. After the next service the amount of paper work may depend on how much was completed during the preceding vacation. Similar remarks apply to medical practitioners who fill out reports after seeing patients.

We have analyzed variants of the M/G$^{(a,b)}$/1 queue with bulk service in [16], [71] using LC. The model utilizes a two-dimensional state $(W(t), M(t))$ where $W(t)$ is the virtual wait. Random variable $M(t)$ is discrete. It represents the number of customers in the waiting line mod b (modulo b) where b is the quorum size. It is called a *system configuration*, which is explicated for M/M/c queues in subsections 4.5 – 4.6 below. *System configurations* are very useful in many stochastic models. They are akin to supplementary variables, and make a model Markovian.