# Chapter 1
# Origin of Level Crossing Method

## 1.1 Prologue

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

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

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



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

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

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

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

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

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

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

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

## 1.2 Lindley Recursion for GI/G/1 Wait

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

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

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

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

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

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

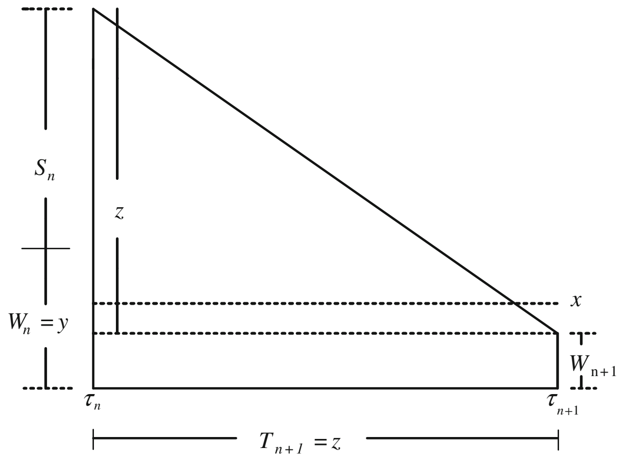


**Fig. 1.1** Lindley recursion for $W_n \to W_{n+1}$ geometrically

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

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

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

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

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

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

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

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

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

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

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

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

By definition,

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

Substituting into (1.6) yields

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

which simplifies to

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

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

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

The normalizing condition that all probabilities sum to 1, is

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

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

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

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

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

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

## 1.4 Observations and Questions

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

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

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

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

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

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

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

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

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

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

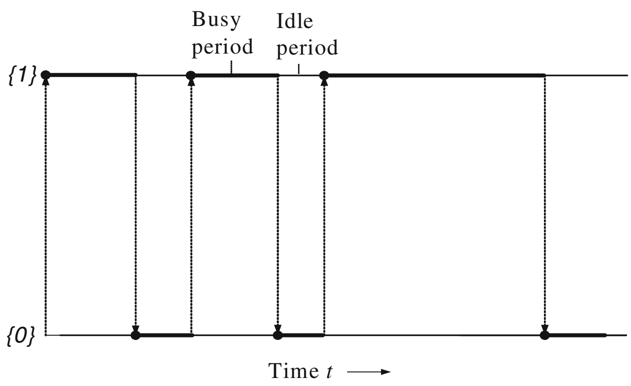


**Fig. 1.2** Sample path of alternating renewal process $\{A(t)\}_{t\geq 0}$

### *1.5.1 Connection with Virtual Wait Process*

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

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

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

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

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

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

The above discussion leads to the following observation.

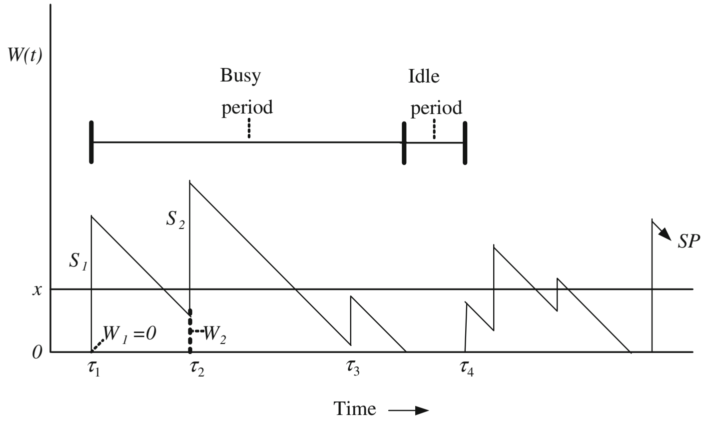


**Fig. 1.3** Sample path of virtual wait $\{W(t)\}_{t\geq 0}$ in M/G/1 showing: actual waits $\{W_n\}_{n=1,2,...}$; busy and idle periods; system point SP; fixed level $x$

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

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

**Observation:** *$f(0^+)$ = rate at which a typical sample path of $\{W(t)\}_{t\geq 0}$ hits level 0 from above at a 45° angle (Fig. 1.3). Hits of level 0 from above occur at the ends of busy periods.*
**Insight:** *Shift attention to sample paths of the virtual wait $\{W(t)\}_{t\geq 0}$! Focus on the geometry of a typical sample path of $\{W(t)\}_{t\geq 0}$!*

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

### *1.5.2 Looking Upward from Level Zero*

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

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

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

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

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

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

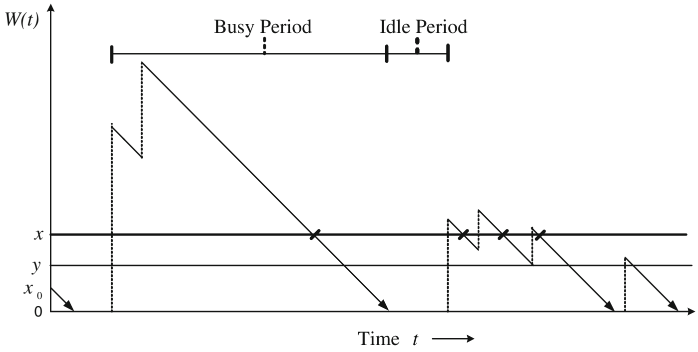


**Fig. 1.4** Sample path of $\{W(t)\}_{t\geq 0}$ indicating downcrossings of level $x$ and hits of level 0

## *1.5.3 Integral Equation in Light of the Sample Path*

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

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

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

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

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

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

## 1.6 Basic Level Crossing Theorem for M/G/1

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

### *1.6.1 Downcrossing and Upcrossing Rates*

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

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

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

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

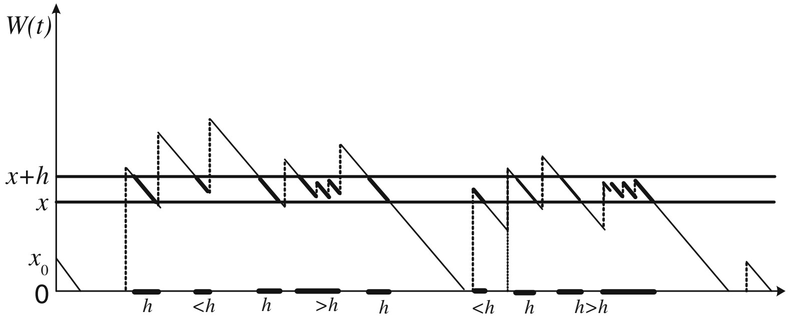


**Fig. 1.5** Sample path of virtual wait in M/G/1 queue. Shows levels $x$ and $x+h$ and various sojourn times in interval $(x, x+h)$, used in proof of Theorem 1.1

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Principle of Rate Balance for Level $x$**

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

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

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

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

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

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

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

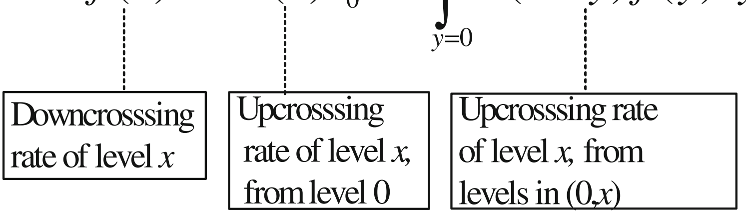


**Fig. 1.6** One-to-one correspondence between virtual-wait sample-path rates of crossing level $x$ and terms of integral equation (1.8) for $f(x)$

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

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

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