Level Crossing Methods in Stochastic Models

Percy H. Brill

Level Crossing Methods in Stochastic Models

Springer

Percy H. Brill
University of Windsor, ON
Canada
brill@uwindsor.ca

Series Editor
Frederick S. Hillier
Stanford University
Stanford, CA
USA

ISBN: 978-0-387-09420-5 e-ISBN: 978-0-387-09421-2
DOI: 10.1007/978-0-387-09421-2

Library of Congress Control Number: 2008926493

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

"Out of great complexity comes great simplicity."

adapted from Winston Churchill

To the memory of my parents

PREFACE

From 1972 to 1974, I was working on a PhD thesis entitled *Multiple Server Queues with Service Time Depending on Waiting Time.* The method of analysis was the embedded Markov chain technique, described in the papers [82] and [77]. My analysis involved lengthy, tedious derivations of systems of integral equations for the probability density function (pdf) of the waiting time. After pondering for many months whether there might be a faster, easier way to derive the integral equations, I finally discovered the basic theorems for such a method in August, 1974. The theorems establish a connection between sample-path level-crossing rates of the virtual wait process and the pdf of the waiting time. This connection was not found anywhere else in the literature at the time. I immediately developed a comprehensive new methodology for deriving the integral equations based on these theorems, and called it *system point theory.* (Subsequently it was called *system point method*, or *system point level crossing method: SPLC* or simply *LC*.) I rewrote the entire PhD thesis from November 1974 to March 1975, using LC to reach solutions. The new thesis was called *System Point Theory in Exponential Queues.* On June 12, 1975 I presented an invited talk on the new methodology at the Fifth Conference on Stochastic Processes and their Applications at the University of Maryland. Many queueing theorists were present. Ever since, LC has become an increasingly used technique for analyzing a large class of stochastic models. LC can be used to derive integro-differential equations for transient distributions, or integral equations for steady-state distributions.

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

To date, many researchers have applied LC. Applications have appeared in refereed journals, conference Proceedings, technical reports, Masters and PhD theses, and in chapters and sections of books, worldwide.

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

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

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

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

Chapter 1 outlines the original developmental ideas which led me to the discovery of LC. When combined synergistically, the basic ideas lead to a powerful modelling technique.

Chapter 2 defines and discusses basic concepts relevant to the method, such as: state space, sample path, system point (SP), SP jump, state-space level, boundary, downcrossing, upcrossing, tangent, etc.

Chapters 3, 4, and 5 analyze steady-state distributions in variants of M/G/1, M/M/c and G/M/c queues, respectively. Chapters 3 and 4 also provide some basic results for transient distributions.

Chapter 6 analyzes steady-state distributions in several basic dams, and in two inventory models. It also includes some transient results.

Chapter 7 demonstrates a multi-dimensional technique with applications to two 2-dimensional inventory models.

Chapter 8 explains the embedded level crossing technique with applications to dams and queues.

Chapter 9 gives an introduction to level crossing estimation, which uses simulation of sample paths to obtain solutions.

Chapter 10 applies LC to a variety of models including: a replacement model, renewal theory, Markov renewal theory, Markov chains, growth and counter models, a dam with alternating continuous influx and efflux, simple harmonic motion. It also illustrates some transient analyses.

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

Percy Brill

ACKNOWLEDGEMENTS

Many people supported the writing of this monograph. Dr. Mei Ling Huang provided helpful comments, advice and encouragement throughout. I am indebted to Dr. Myron Hlynka for many technical discussions and useful comments on theoretical and editorial points. The late Dr. Carl Harris deserves special credit for productive discussions and ideas. He originally convinced me to write this monograph. Dr. Donald Gross provided helpful discussions, comments and support. I am very grateful to Canadian queueing scholars for their encouragement of my research, in particular: Dr. David Stanford, Dr. Attahiru Alfa, Dr. Winfried Grassmann, Dr. Beth Jewkes, Dr. Steve Drekic and Dr. Marvin Mandelbaum. Also, Dr. Stanley Sclove contributed technical discussions and encouragement.

I appreciate the encouragement of Dr. F.S. Hillier, and the support of Gary Folven and Carolyn Ford of Springer in the production process. Thanks are due to my research-paper coauthors, postdoctoral fellows, and graduate students, for numerous technical discussions. Thanks to colleagues and staff at the University of Windsor for facilitating my research. I am thankful for the constant support of my family and friends, especially my sister Sally Levitan.

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

I am grateful to NSERC (Natural Sciences and Engineering Research Council of Canada) for long-term research support, which has been exceedingly helpful toward this project in many concrete ways.

CONTENTS

CHAPTER 1

ORIGIN OF LEVEL CROSSING METHOD

1.1 Introduction

This chapter presents a condensed version of the original development of the *level crossing method* for deriving probability distributions of state variables in stochastic models ***(LC)***. I developed LC concomitantly with the more general *system point* 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 to August 1974, while I was working on my PhD thesis of a different topic, namely *Multiple Server Queues with Service Time Depending on Waiting Time*. The work involved analyzing the steady-state distribution of customer wait in an M/M/c queue with service time depending on wait before service, since May 1972. This had been my original PhD thesis topic, suggested by my supervisor M.J.M. Posner. The goal had been to generalize to M/M/c queues, the (then) forthcoming paper [88] on M/M/1 queues, *using the method of embedded Markov chains*, a purely algebraic technique [77]. That analysis formulates Lindley recursions for successive customer waits and their probability distributions [82]. The approach utilizes inequalities, conditional probabilities, and the law of total probability. It also involves multiple integration, transformation of variable, differentiation, and limit operations.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_1,

The embedded Markov-chain analysis can be tedious and time consuming, especially for complex models. I worked for several thousand hours (about fifty hours 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 Section 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 Subsections 2.4.2, 2.4.3).

After my discovery in 1974, I completely rewrote my PhD thesis using LC, from November 1974 to March 1975. The new thesis was called *System Point Theory in Exponential Queues* [7]. This led to the subsequent publications [37], [38], [39]. Two years later in 1976, J.W. Cohen [45] discussed the same level crossing ideas, couched in terms of regenerative processes [96].

The following abridged version of the development of LC deals with the single server queue. (This preserves the main ideas, which originally evolved from analyzing complex M/M/c queues.) We first derive an integral equation based on the *classical* algebraic method for GI/G/1 and M/G/1 queues. This was the method used to analyze my original PhD thesis topic. (Due to multiple servers, that derivation started with a more general Lindley recursion [34], [35]. It ended with a system of integral equations for the steady-state pdf of wait. Working papers [34], [35] illustrate the original thesis using embedded Markov chains.)

1.2 Lindley Recursion for GI/G/1 Wait

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

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

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

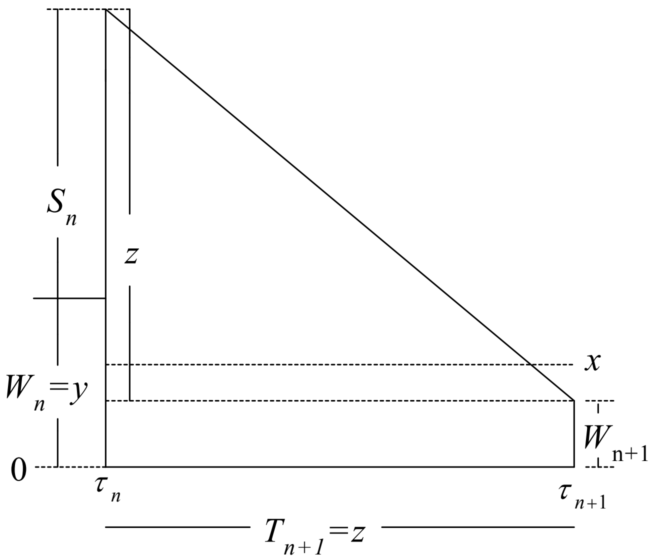

Figure 1.1: Lindley recursion for $\{W_n\}$ geometrically.

Thus $F_n(\cdot)$ 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 input parameters of the queue are such that the steady state cdf $F(\cdot)$ and pdf $\{P_0, f(\cdot)\}$ 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 Waiting Time Derived Using Lindley Recursion

Assume that the arrival process is Poisson at rate λ, and that the random variables $\cup_{n\in I^+}\{S_n, T_{n+1}\}$ are mutually independent (where $I^+ = \{1, 2, ...\}$). 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 (discrete state) $\{0\}$ is included. Substituting from (1.3), we obtain for $x \geq 0$,

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

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

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

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

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

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

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

which simplifies to

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

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

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

The normalizing condition that all probabilities sum to 1, is

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

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 ($\boldsymbol{L} = \boldsymbol{\lambda} \cdot \boldsymbol{W}$); 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; 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 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 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.

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 deriving integral equations for the pdf in complex state-dependent M/M/c models using the classical method. 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 equation (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. Persevering with this idea, while continuing to apply the classical method, ultimately led to the SPLC methodology. The developmental process is outlined in sections 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 $\left[\frac{1}{Time}\right]$. This follows because Δx has the same dimension as x, namely $[Time]$, in the defining formula $f(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x}$.

3. In integral equation (1.8), the dimension of both left and right hand sides is $\left[\frac{1}{Time}\right]$. Note that the parameter λ has dimension $\left[\frac{1}{Time}\right]$.

4. A number having dimension $\left[\frac{1}{Time}\right]$ is the measure of a *rate,* a notion from Physics.

5. Each side of integral equation (1.8), is the measure of some unknown *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 Section 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 Section 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 (if and only if) the server is idle at the arrival instant; (b) a positive time iff the server is busy at the arrival instant. Thus we may equivalently redefine the states from the viewpoint of the system (or server) as: $\{0\}$: *idle*; $\{1\}$: *busy*.

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

Figure 1.2 depicts the motion between the two states $\{0\}$,$\{1\}$. The sojourn times of visits to $\{0\}$ are iid (independently and identically distributed) random variables distributed as an idle period. An idle period is exponentially distributed with mean $\frac{1}{\lambda}$. The sojourn times of visits to $\{1\}$ are iid random variables distributed as a busy period. A sample

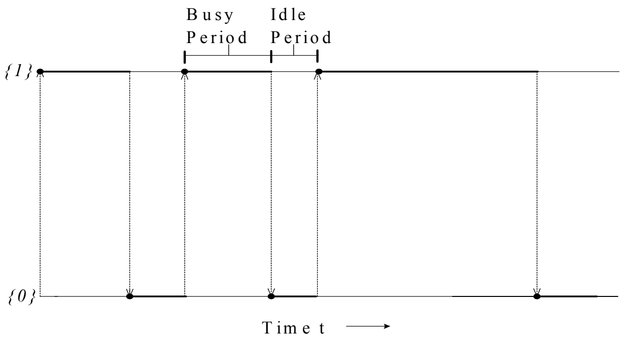

Figure 1.2: Sample path of alternating renewal process $\{A(t), t \geq 0\}$.

path corresponds to that of a two-state alternating renewal process. It is a special case of a Markov renewal or semi-Markov process with 2×2 Markov transition matrix $\|P_{ij}\|$ where $P_{01} = P_{10} = 1$. 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).

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 [99] which I had become aware of in 1964). The virtual wait represents how long a customer would wait for service if the customer arrived at time t. For the M/G/1 queue, the virtual wait $\{W(t), t \geq 0\}$ is a continuous-time, continuous-state process with state space $[0, \infty)$. Sample paths of $\{W(t), t \geq 0\}$ are real-valued, non-negative, right-continuous functions on $[0, \infty)$. Characteristically,

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

(Fig. 1.3). Jumps occur at Poisson rate λ. Jump sizes are distributed as the service time. Table 1.1 shows the correspondence between the two processes.

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.

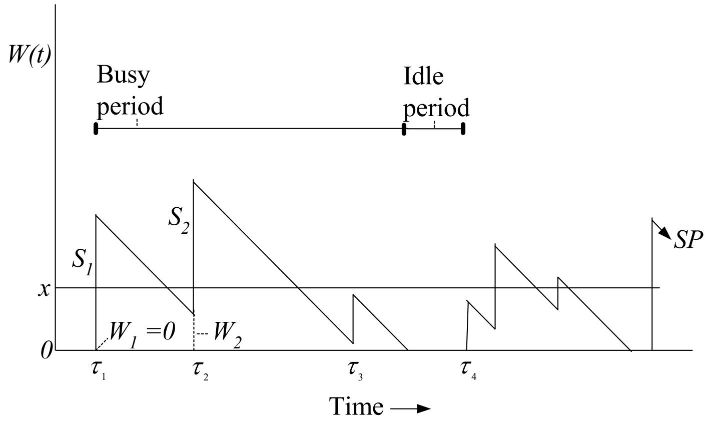

Figure 1.3: Sample path of virtual wait $\{W(t)\}$ in M/G/1 showing: actual waits $\{W_n\}$; busy, idle periods; system point SP; level x.

(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., [99]).

The above discussion leads to the following observation.

Observation: $f(0^+) =$ ***rate*** at which a typical sample path of $\{W(t)\}$ 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 $\{W(t)\}$!

The last observation provides an alternative interpretation of equation (1.10). In complex systems, this observation may lead to extra

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

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

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 subsection 1.5.2.

1.5.2 Looking Upward from Level Zero

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

To answer the key question, imagine, temporarily, that the M/G/1 model under consideration were really an M/M/1 model with service rate μ. The jump sizes of the virtual wait process (Fig. 1.3) would then be *exponentially* distributed with mean $\frac{1}{\mu}$. 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, $P(S_n > x - y + z \mid S_n > x - y) = e^{-\mu z}$, $n = 1, 2, ...$, 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 Figure 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 Subsection 3.3.12). 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 Subsection 1.1. 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, as occur in Operations Research. However, it is also applicable to processes with continuous sample paths, as in a dam with alternating influx and efflux analyzed in Chapter 10.

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 standard virtual wait of an 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 Chapter 2.

1.5.3 Integral Equation in Light of 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 equation (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 (equivalently rate balance).* Referring to $W(t), t \geq 0$, (Fig. 1.4), let $x_0 = W(0)$, and fix $x > 0$. The state space is $\boldsymbol{S} = [0, \infty) = [0, x] \cup (x, \infty)$ (union of two disjoint sets). The long-run sample-path *exit* and *entrance* rates of state-space set (x, ∞), are equal, independent of the initial state x_0. Intuitively, exits and entrances of (x, ∞) alternate in time and correspond to sample-path downcrossings and upcrossings of 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 is the correct interpretation.

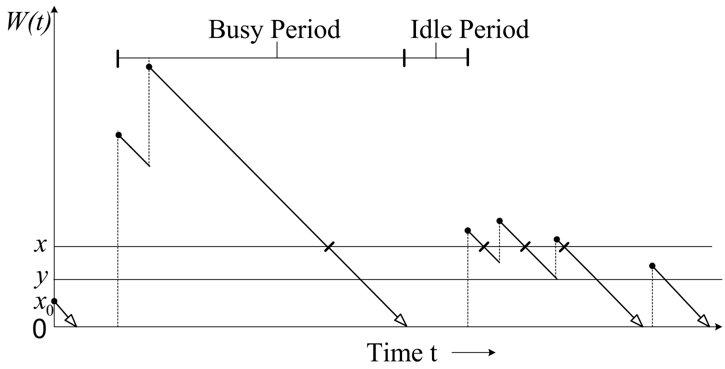

Figure 1.4: Sample path of $\{W(t), t \geq 0\}$ indicating crossings of level x and hits of level 0.

For the process $\{W(t)\}$, 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 a set of three mutually exclusive and exhaustive sample-path crossing rates of level x (see Fig. 1.6).

1.6 Basic Level Crossing Theorem for M/G/1

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

1.6.1 Downcrossing and Upcrossing Rates

For fixed $x > 0$ and $t > 0$, 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 $P_0 > 0$. 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 Chapter 2.

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

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

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

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

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

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

Proof. (Note: A different proof is given in a corollary of Theorem 3.6 for the transient pdf.)

Here we provide a proof which demonstrates simple intuition underlying the SPLC methodology. Consider a sample path of the virtual wait and levels $x > 0$ and $x + h$, where $h > 0$ is small (Fig. 1.5).

Just after each downcrossing of level $x+h$, the sample path spends a time h in the state-space interval $(x, x+h)$ with probability $1-\lambda h+o(h)$. It spends a shorter or longer time in $(x, x + h)$ with probability $o(h)$.

During a time interval $(0, t), t > 0$, the expected *proportion* of time spent in $(x, x + h)$ is $\frac{E(\mathcal{D}_t(x+h))\cdot h\cdot(1-\lambda h)+o(h)}{t}$. The limiting proportion of

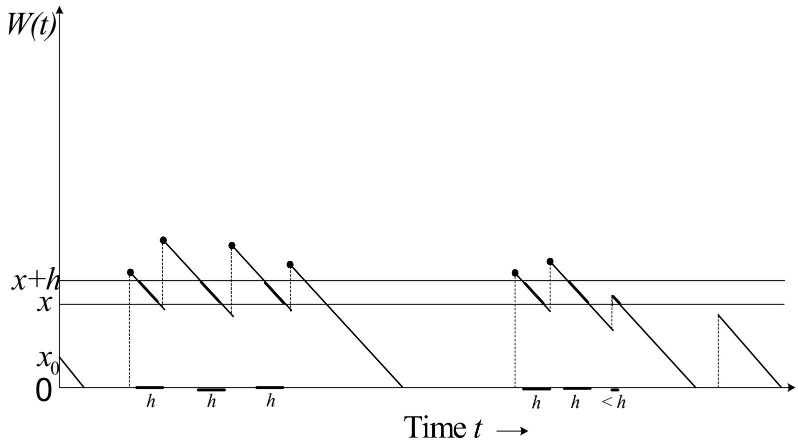

Figure 1.5: Sample path of virtual wait in M/G/1 queue. Shows levels x and $x+h$ and time spent in interval $(x, x+h)$, used in proof of Theorem 1.1.

time spent in $(x, x+h)$ is, by the definition of the steady-state cdf $F(\cdot)$ of wait,

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x+h)) \cdot h \cdot (1-\lambda h) + o(h)}{t} = F(x+h) - F(x).$$

Dividing by h and letting $h \downarrow 0$ then gives

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

Since all downcrossings are left-continuous and smooth (no jump downcrossings), $\mathcal{D}_t(x^+) = \mathcal{D}_t(x)$ and thus

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

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

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

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

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t(0))}{t} = f(0^+) = f(0) = \lambda P_0, \tag{1.16}$$

$$\lim_{t\to\infty}\frac{\mathcal{D}_t(0)}{t} \underset{(a.s.)}{=} f(0^+) = f(0) = \lambda P_0. \tag{1.17}$$

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

Note that (1.16) and (1.17) equate the sample-path: (1) downcrossing rate of level 0 (= *entrance rate into* discrete state $\{0\}$), (2) *exit rate from* $\{0\}$; and (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 (1.18) below, gives the principle of *set balance* for a state-space set $(x,\infty), x>0$, in terms of rate balance across level x.

1.6.2 Principle of Rate Balance for Level x

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

$$\left.\begin{array}{l}\lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t}, x>0,\ (a.s), \\ \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t} = \lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}, x>0.\end{array}\right\} \tag{1.18}$$

Formula (1.18) 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 Chapter 2, Subsection 2.4.6.

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 (1.18). *Start with a typical sample path of* $\{W(t)\}$. 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 $f(\cdot)$. Write integral equation (1.8) as a *rate-balance equation* using (1.18), *by inspection* (Fig. 1.6)!

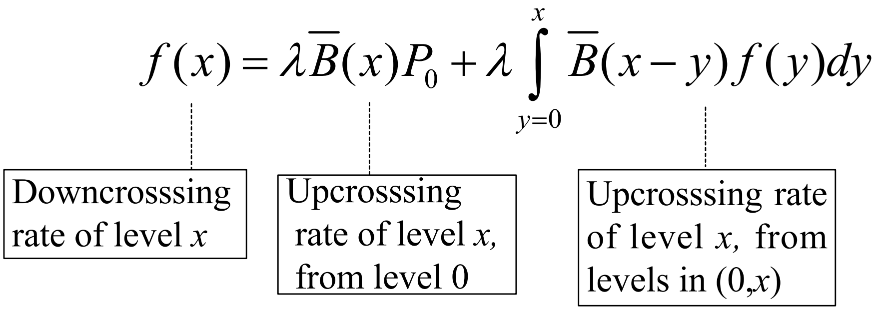

Figure 1.6: One-to-one correspondence between virtual-wait sample-path rates of crossing level x and terms of integral equation (1.8) for $f(x)$.

Note that starting from level 0, the upcrossing rate of level $x > 0$ is

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

Summary of Steps in LC Derivation of Integral Equation

1. Construct a sample path of $\{W(t)\}$ (Fig. 1.4).

2. Substitute from (1.12) and (1.14) term by term into (1.18).

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.

CHAPTER 2

SAMPLE PATH AND SYSTEM POINT

2.1 Introduction

When applying the system point level crossing method (abbreviated SPLC, or LC) to analyze stochastic models, *intuitive* notions of sample-path transitions often suffice For some models, however, *more precise* notions of such transitions are useful. Pertinent sample-path transitions include downcrossings, upcrossings, and tangents of state-space levels. This chapter presents definitions and examples which apply to a large class of stochastic models with continuous parameter sets. The last subsection summarizes various types of transitions geometrically.

2.2 State Space and Sample Paths in Continuous Time Stochastic Models

For each stochastic model considered, we will tacitly assume the existence of a basic probability space $(\mathbf{\Omega}, \mathcal{F}, \boldsymbol{P})$, where $\mathbf{\Omega}$ is the set of all possible outcomes of the associated random experiment, $\mathcal{F}$ is a σ-field of events, and $\boldsymbol{P}$ is a probability measure on $\mathcal{F}$. The LC method starts with a "typical" sample path of the underlying process constructed over Time, from the sequences of random variables defining the model. Examples of such sequences occur in: queues - inter-arrival and service times; inventories - inter-demand times and demand sizes; dams - inter-input times and input sizes; actuarial models - inter-claim times and claim amounts;

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_2,

pharmacokinetics - inter-dose times and dose amounts. The LC method emphasizes sample paths when analyzing stochastic models.

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

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

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

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

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

Sample paths are also called sample functions, realizations, trajectories, tracings, orbits. A sample path is a possible outcome $\omega \in \boldsymbol{\Omega}$ of the background random experiment associated with a model. For fixed t, $X(t)$ is a random variable with domain $\boldsymbol{\Omega}$ and *range* a subset of $\boldsymbol{S}$. If $\boldsymbol{S} \subseteq \boldsymbol{R}$ (set of real numbers), then $X(t)$ has cdf $P(X(t) \leq x), x \in \boldsymbol{S}$, and pdf $\frac{d}{dx}P(X(t) \leq x)$, where the derivative exists. If t_0 is not an instant of jump, then $X(\cdot)$ is continuous at t_0. If t_0 *is* an instant of "jump", it is still possible that $X(\cdot)$ is continuous at t_0 (see Example 2.3). For many models discussed in this monograph, sample paths are piecewise continuous and differentiable between jumps, and the slope, at a fixed state-space level x is independent of t (Figs. 2.2 - 2.5).

2.2.1 Sample-Path Properties and Jumps

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

Proof. The time domain $\boldsymbol{T} = \overset{\infty}{\underset{n=1}{\cup}} [n-1, n)$, is a countable union of disjoint finite intervals. Each interval contains at most a finite number of sample-path jumps and/or relative extrema (a.s.), by Definition 2.1. A countable union of countable sets is countable. ■

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

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

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

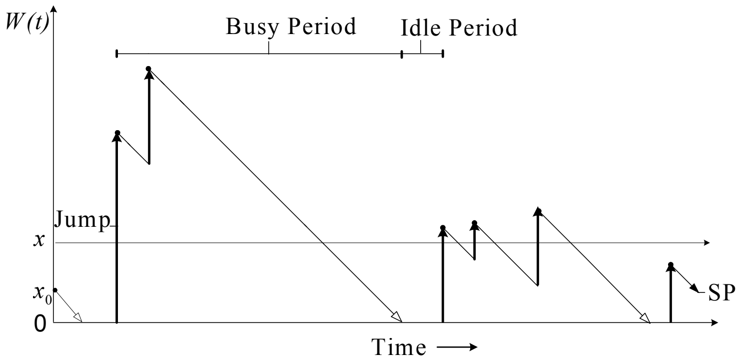

Figure 2.1: Sample path of virtual wait in $M/G/1$ queue. Emphasizes jumps and hits of level 0 from above.

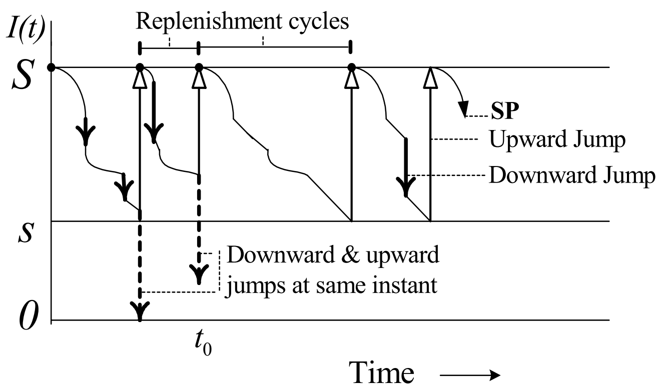

Figure 2.2: Sample path of net inventory in $\langle s, S \rangle$ model with product decay; no lead time.

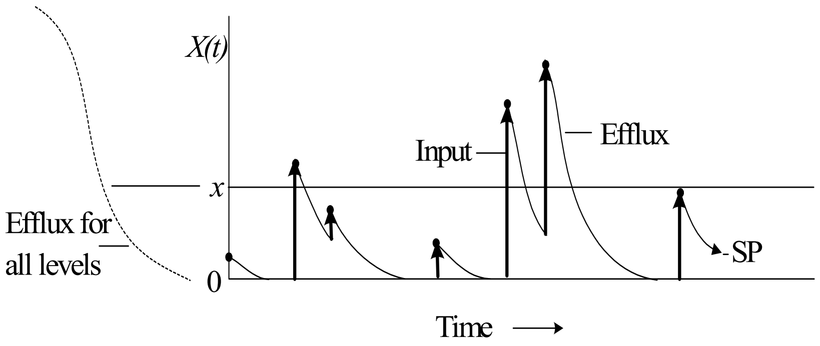

Figure 2.3: Sample path of content in dam with general release (efflux). Emphasizes jumps and right continuity of sample path.

Remark 2.1 *For: (1) the* ***virtual wait process*** *in a G/G/1 queue, upward jumps occur at arrival instants (Fig. 2.1); (2) the* ***content in a dam*** *with instantaneous inputs, upward jumps occur at input instants (Fig. 2.3); (3) the* ***extended age process*** *in a G/M/1 queue (time that a customer in service has been in the system, or, negative of remaining time until the next arrival to an empty system), downward jumps equal (in distribution) to inter-arrival times, occur at departure instants (Fig. 2.4); (4) the* ***risk reserve process*** *in a ruin-like model in Insurance, downward jumps occur at claim instants, and upward jumps may occur at ruin instants (epochs) (Fig. 2.5); (5) the* ***concentration of a drug*** *in a one-compartment pharmacokinetic model with multiple bolus dosing, upward jumps occur at instants of dosing.*

2.3 System Point Motion and Jumps

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

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

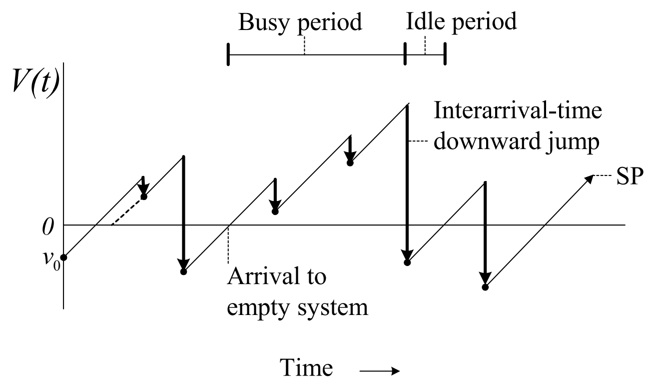

Figure 2.4: Sample path of extended age process in G/M/1 queue. $V(t)$ = time customer in service at t has been in system, if $V(t) > 0$. $-V(t)$ = remaining time at t until next arrival, if $V(t) < 0$. Emphasizes jumps.

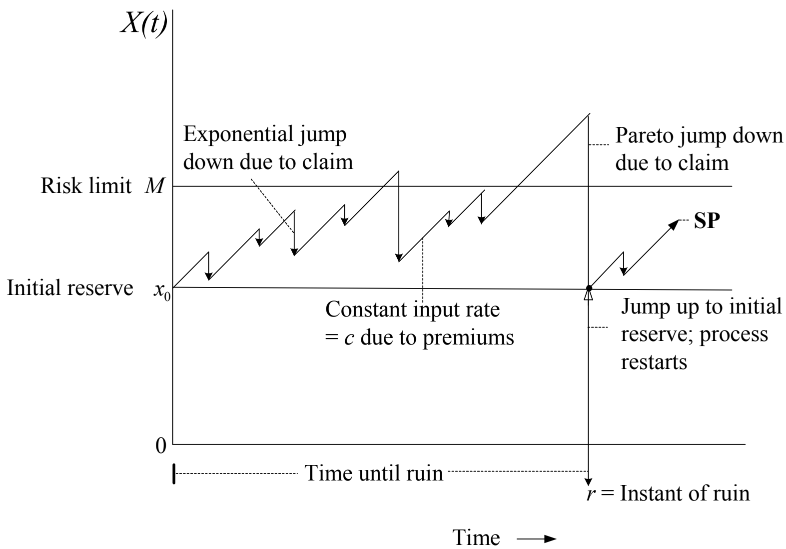

Figure 2.5: Sample path of risk reserve in ruin-like model in insurance.

If the state space is $\mathbf{S} \subseteq \mathbf{R}$, and if $t_0 > 0$ is a point of continuity of a sample path $X(t), t \geq 0$ then the SP has direction defined by $\frac{d}{dt}X(t)|_{t=t_0}$ assuming the derivative exists. If t_0 is an instant of jump, then the SP has direction up, down or *both, not in the direction of Time* (Figs. 2.2, 2.5, 2.6). Technically, the SP jumps are not part of the sample path, which is by definition a mathematical function.

In mathematical models, jumps occur at instants, i.e., not in Time. Given the positions of jumps, the continuous segments can be filled in provided their slopes at all levels are independent of time. Also, the continuous segments determine the *net* sample-path jumps. SP jumps are vectors having size, direction, and start and end points in the state space. While tracing the continuous segments of a sample path the SP has a finite velocity in Time. When the SP jumps, it has "infinite speed". A sample path is an inert graph. The SP is like the moving tip of a stylus that plots a graph. The sample path is the completed graph.

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

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

Possibly, both $u_{t_0} > 0$, $d_{t_0} > 0$. The *net* sample-path jump $X(t_0) - X(t_0^-)$ may be positive, negative or zero (Figs. 2.2, 2.5, 2.6). If it is zero, the sample path is continuous at t_0 although the SP may make two equal and opposite jumps at t_0, which correspond to real changes in the associated physical system (Fig. 2.6).

Example 2.2 *Consider the stock on hand in a continuous review $\langle s, S \rangle$* ***inventory*** *model with a single product, $0 \leq s < S < \infty$. Assume random demands, random demand sizes, no lead time, and* ***continuous product decay****. Downward jumps occur at demand instants (Fig. 2.2). Let t_0 be a demand instant with $I(t_0^-) = y, s < y < S$, and let the demand be $d_{t_0} > y - s$. The would-be resulting stock on hand at t_0, due to the demand, is $y - d_{t_0} < s$ and $y - d_{t_0} \notin (s, S]$. The unsatisfied demand at t_0 is $s - y + d_{t_0}$. The downward jump that ends below s is a signal, at t_0 to place an order and replenish the stock up to level S immediately (no lead time). There is an upward jump of stock at t_0 equal to $u_{t_0} = S - y + d_{t_0}$.*

This satisfies the deficit and restores the stock up to S, i.e., $I(t_0) = S$. The SP makes ***both*** *downward and upward jumps at t_0 (**double jumps**) which* ***result in a single net sample-path upward jump*** *of size $S - y = u_{t_0} - d_{t_0}$. The SP upward jump is a prescribed or policy jump. In summary,* ***at t_0 the SP makes two jumps in opposite directions; the sample path has one (net) upward jump.***

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

Remark 2.2 *The foregoing examples show that at least two SP jumps can occur at an instant. SP* ***multiple jumps*** *are compatible with a common assumption for continuous time stochastic models. That is,* ***multiple probabilistic events*** *cannot occur at the same instant. The latter assumption technically applies to sample paths and to the sequences of random variables defining the model. We usually prohibit more than one: arrival; service completion; demand; input; insurance claim; etc., at a particular instant. The LC method is based on the count or rate of SP transitions across levels or state-space boundaries, or between state-space sets. This is regardless of whether the transitions are due to SP jumps, or due to sample-path smooth descents or ascents. In the $\langle s, S \rangle$ inventory model, the LC method counts jumps both due to chance events like demands, and due to prescribed responses like replenishments, when computing rates of crossing state-space levels.*

Remark 2.3 *Consider Example 2.1. An observer of the* ***sample path*** *who is aware of the $\langle s, S \rangle$ policy, and knows that $X(t_0) = S$ due to a jump at t_0, cannot determine whether the SP made both a downward and upward jump, or a single upward jump at t_0 (Fig. 2.2). The jump resulting in $X(t_0) = S$ could have been caused by a signal of either type (a) (smooth descent) or type (b) (demand).*

Remark 2.4 *In Example 2.3 (Fig. 2.6), assume an observer of the sample path knows the policy is $\langle s, S \rangle$ and that $X(t_0) = S$. The observer*

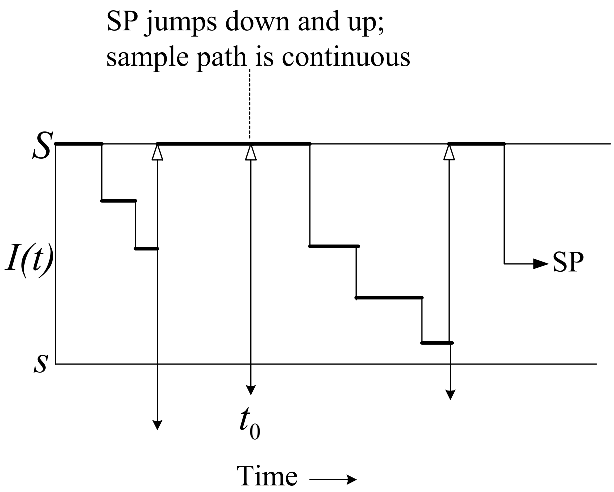

Figure 2.6: Sample path and system point motion in $\langle s, S\rangle$ inventory with no product decay. Sample path is continuous at t_0, and system point (SP) double jumps at t_0.

*cannot distinguish t_0 as being an instant of activity (placing an order and replenishing to S) or an instant of inactivity, since the SP motion is "invisible" at t_0. Knowledge of the sample path is sufficient to derive **probability distributions** of the net inventory. However, knowledge of the SP motion over Time **and** SP motion at instants of jump, implies knowledge of the sample path structure and of the ongoing actual activity of the real world system.*

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

To summarize, the *SP* is a point having motion over Time during continuous sample-path segments *or* motion in the state space at instants of jump. At instants of jump the SP "moves" only in the state space. The sample path is a mathematical function $X(\cdot)$: $\mathbf{T} \to \mathbf{S}$. (It is a coincidence that *sample path* and *system point* have the same initials.)

Stochastic Model	State Space **S**	Figure
Virtual wait in M/G/1 queue	$[0,\infty)$	2.1
Extended age process in G/M/1 queue	$(-\infty,+\infty)$	2.4
Stock on hand in $\langle s,S\rangle$ inventory	$(-\infty,S], S>0$	2.2, 2.6
Content in dam	$[0,\infty)$	2.3
Risk reserve in actuarial ruin-like model	$(-\infty,+\infty)$	2.5

Table 2.1: Examples of models with state space a subset of **R**; and corresponding ffgures.

2.3.1 State Space in the Wide Sense

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

Throughout this monograph whenever using the term "state space", we shall mean ***state space in the wide sense***, unless otherwise specified. The state space in the wide sense contains the range of all possible SP jumps.

2.4 State Space a Subset of R

In the models discussed so far, the state space is an interval subset of the real numbers. Most models in this monograph are in this category. We now discuss such models more formally, to develop intuitive background about the SPLC methodology. (In Chapter 7 we will discuss models with state space a subset of $\boldsymbol{R}^2$ or $\boldsymbol{R}^n$.)

Consider a stochastic model having state-space interval $\boldsymbol{S}\subseteq\boldsymbol{R}$**,** the set of real numbers**.** Set $\boldsymbol{S}$ is often an infinite interval. In Table 2.1 we assume that $\boldsymbol{S}$ is sufficiently large to contain end points of any SP jumps, even if sample paths do not extend over all of **S**. That is, **S** is the state space in the wide sense (see Subsection 2.3.1). In Example 2.2, using the state space in the wide sense has no effect on the values of the cdf and pdf of stock on hand. All probability is supported on the interval $(s,S]$. The same applies to all models in Table 2.1.

Let $X(t),t\geq 0$, denote a sample path in the Cartesian product space $\boldsymbol{T}\times\boldsymbol{S}=[0,\infty)\times\boldsymbol{S}$. Let $\boldsymbol{A}\subset\boldsymbol{S}$ be a proper *interval* subset of **S**. Then

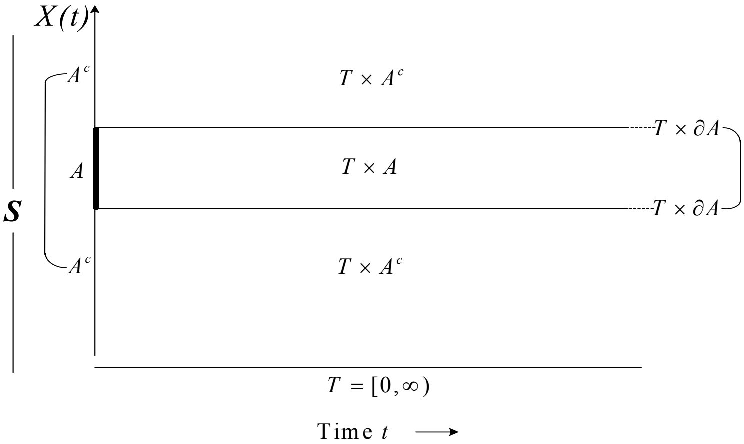

Figure 2.7: Sets $\boldsymbol{A}$, $\boldsymbol{A}^c$ and $\mathbf{T} \times \boldsymbol{A}$, $\boldsymbol{T} \times \boldsymbol{A}^{\mathbf{c}}$, $\boldsymbol{T} \times \partial \boldsymbol{A}$ when interval $\boldsymbol{A} \subset \boldsymbol{S} \subseteq \boldsymbol{R}$.

$\boldsymbol{T} \times \boldsymbol{A} \subset [0, \infty) \times \boldsymbol{S}$. For set $\boldsymbol{A}$, let $\boldsymbol{A}^c$ denote the complement in $\boldsymbol{S}$, $\partial \boldsymbol{A}$ the boundary, $\boldsymbol{A}^o$ the interior set , and $\boldsymbol{A}^e$ the exterior set (= interior of $\boldsymbol{A}^c$). Set $\boldsymbol{A}^c$ may be the union of two disjoint intervals. The two-dimensional Cartesian product sets $\boldsymbol{T} \times \boldsymbol{A}^c$, $\boldsymbol{T} \times \partial \boldsymbol{A}$, $\boldsymbol{T} \times \boldsymbol{A}^o$, $\boldsymbol{T} \times \boldsymbol{A}^e$, are proper subsets of $\boldsymbol{T} \times \boldsymbol{S}$ (Fig. 2.7). Sets $\boldsymbol{A}^o$, $\boldsymbol{\delta A}$, $\boldsymbol{A}^e$ are mutually disjoint, as are their respective Cartesian products with $\boldsymbol{T}$.

2.4.1 Levels in State Space

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

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

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

2.4.2 Sample Path Transitions

Consider an interval $\boldsymbol{A} \subseteq \boldsymbol{S}$ (Fig. 2.7). We first define the following transitions: *sample-path exits, entrances, tangents, boundary crossings, level crossings* with respect to $\boldsymbol{T} \times \boldsymbol{A}$. Let $X(t), t \geq 0$ denote a sample path. Assume $t_0 > 0$ is an instant of either continuity or jump. Let $X(t_0^-) = \lim_{t \uparrow t_0} X(t)$ (left limit at t_0 exists).

Definition 2.2 ***Sample-path Exit:*** *$X(\cdot)$ **exits A** at instant t_0 if*

$$\exists\, \varepsilon > 0 \ni X(t) \in \boldsymbol{T} \times \boldsymbol{A}, t \in (t_0 - \varepsilon, t_0)$$

and $X(t) \in \boldsymbol{T} \times \boldsymbol{A}^c$, $t \in (t_0, t_0 + \varepsilon)$.

Sample-path Entrance: *$X(\cdot)$ **enters A** at instant t_0 if $X(\cdot)$ exits $\boldsymbol{A}^c$ at t_0.*

Sample-path Interior Tangent: *$X(\cdot)$ is **interior tangent** to **A** at instant t_0 if $\exists\, \varepsilon > 0 \ni X(t) \in \boldsymbol{T} \times \boldsymbol{A}^o$, $t \in (t_0 - \varepsilon, t_0 + \varepsilon) \diagdown \{t_0\}$ and either $X(t_0^-) \in \boldsymbol{T} \times \partial \boldsymbol{A}$, or $X(t_0) \in \boldsymbol{T} \times \partial \boldsymbol{A}$.*

Sample-path Exterior Tangent: *$X(\cdot)$ is **exterior tangent** to **A** at instant t_0 if $X(\cdot)$ is interior tangent to $\boldsymbol{A}^c$ at instant t_0.*

Sample-path Boundary Crossing: *$X(\cdot)$ crosses boundary $\partial \boldsymbol{A}$ at instant t_0 if $X(\cdot)$ exits $\boldsymbol{A}^o$ and enters $\boldsymbol{A}^e$ ($A^o \to A^e$), or $X(\cdot)$ exits $\boldsymbol{A}^e$ and enters $\boldsymbol{A}^o$ ($A^e \to A^o$) at t_0.*

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

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

Definition 2.3 ***Sample-path Downcrossing:*** *$X(\cdot)$ **downcrosses** level x at instant t_0 if $X(\cdot)$ crosses boundary*

$$\boldsymbol{T} \times \{x\}\ (\boldsymbol{T} \times \boldsymbol{A}^o \to \boldsymbol{T} \times \boldsymbol{A}^e) \text{ at } t_0.$$

Equivalently, $X(\cdot)$ exits $\boldsymbol{T} \times ((x, \infty) \cap \boldsymbol{S})$ and enters $\boldsymbol{T} \times ((-\infty, x) \cap \boldsymbol{S})$ at $t_{0.}$.

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

Definitions 2.2 and 2.3 apply at an instant of either sample-path continuity or sample-path jump. System point (SP) transitions are defined identically as in Definitions 2.2 and 2.3 at instants of *continuity* of $X(t), t \geq 0$. However, at an instant of jump, an SP transition is defined differently, since the SP does not move in the direction of Time at an instant of jump; it moves either upward or downward in $\boldsymbol{S}$.

2.4.3 System Point Transitions

We now define SP transitions with respect to $\boldsymbol{T} \times \boldsymbol{A}$ *at an instant of jump*, say t_0. Assume that at t_0 the SP makes a *single* jump either of size d_{t_0} downward or size u_{t_0} upward. Let $\theta = 1$ or 0 as the direction of the jump is downward or upward respectively.

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

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

SP Entrance at Instant of Jump: *The SP **enters** $\boldsymbol{A}$ at t_0 if the SP exits $\boldsymbol{T} \times \boldsymbol{A}^c$ at t_0.*

SP Boundary Crossing: *The SP makes a **boundary crossing** of $\partial\boldsymbol{A}$ at t_0 if $X(t_0^-) \in \boldsymbol{T} \times \boldsymbol{A}^o$ and*

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

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

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

Fix $x \in \boldsymbol{S}$. Then $\{x\}$ is a boundary of both $(x, \infty) \cap \boldsymbol{S}$ and $(-\infty, x) \cap \boldsymbol{S}$.

Definition 2.5 ***SP Downcrossing:*** *The SP **downcrosses** level x at t_0 if the SP crosses boundary $\boldsymbol{T} \times \{x\}$ from $\boldsymbol{T} \times (x, \infty) \cap \boldsymbol{S}$ to $\boldsymbol{T} \times (-\infty, x) \cap \boldsymbol{S}$ at t_0 ($(x, \infty) \to (-\infty, x)$).*

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

To motivate the next definition consider Example 2.1 (see Fig. 2.2). Assume a demand for the product is placed at t_0^- causing the SP to jump downward to level $z < s$. The SP immediately rebounds with a prescribed upward jump to level S, to replenish the product. Thus the SP 'touches" level z from above and immediately makes an egress above, from level z. The SP has made a "passe-by" of level z but has not entered state $\{z\}$ at t_0. State $\{z\}$ is a boundary of the intervals (z, S) and $(-\infty, z)$. In order to make a pass-by of $\{z\}$ at t_0 the SP makes two jumps in opposite directions at t_0.

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

$$X(t_0^-) \in \boldsymbol{T}\times(\boldsymbol{A}^o \cup \boldsymbol{A}^e),\ X(t_0) \in \boldsymbol{T}\times(\boldsymbol{A}^o \cup \boldsymbol{A}^e)$$

and

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

if $\theta = 1$ *then* $X(t_0) = z + u_{t_0}$. *If* $\theta = 0$ *then* $X(t_0) = z - d_{t_0}$.

2.4.4 Continuous and Jump Crossings

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

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

Thus a continuous crossing is also a left-continuous crossing, but not vice versa.

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

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

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

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

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

Then (see Figs. 2.13 - 2.16),

$$\mathcal{D}_t(x) = \mathcal{D}_t^c(x) + \mathcal{D}_t^j(x),$$
$$\mathcal{U}_t(x) = \mathcal{U}_t^c(x) + \mathcal{U}_t^j(x).$$

Remark 2.6 *$\mathcal{D}_t^c(x)$, $\mathcal{U}_t^c(x)$ include down- and upcrossings that are continuous from the left but not from the right, or that that are continuous.*

2.4.5 Transitions in Finite Time Intervals

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

Notation 2.2 *$\mathcal{O}_t(\boldsymbol{A})$, $\mathcal{I}_t(\boldsymbol{A})$:* number of *SP exits and entrances of $\boldsymbol{T} \times \boldsymbol{A}$ during $(0, t)$, respectively.*

$\mathcal{T}_t^o(\boldsymbol{A})$, $\mathcal{T}_t^e(\boldsymbol{A})$: number of sample-path interior and exterior tangents of $\boldsymbol{A}$ during $(0, t)$, respectively.

$\mathcal{D}_t(x)$, $\mathcal{U}_t(x)$: combined number of sample-path and SP down- and upcrossings of a fixed level $x \in \boldsymbol{S}$ during $(0, t)$, respectively.

We extend the definition of relative maximum or minimum to mean the supremum or infimum of the sample path in a small neighborhood of t_0 if t_0 is an instant of jump.

Proposition 2.2 *The random variables*

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

and their expected values

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

are finite.

Proof. (1) **Exits and Entrances**: Consider a finite time interval $(t_1, t_2) \subset T$. At most one sample-path exit or entrance of $(t_1, t_2) \times \boldsymbol{A}$ can occur between successive instants of relative extrema on continuous segments during (t_1, t_2), due to monotonicity of the sample path between relative extrema. It is possible for both a relative infimum and relative maximum (or relative supremum and relative minimum) to occur at an instant of jump during (t_1, t_2) (see Fig. 2.8). At each such instant, at most one exit or entrance of $\boldsymbol{T} \times \boldsymbol{A}$ can occur.

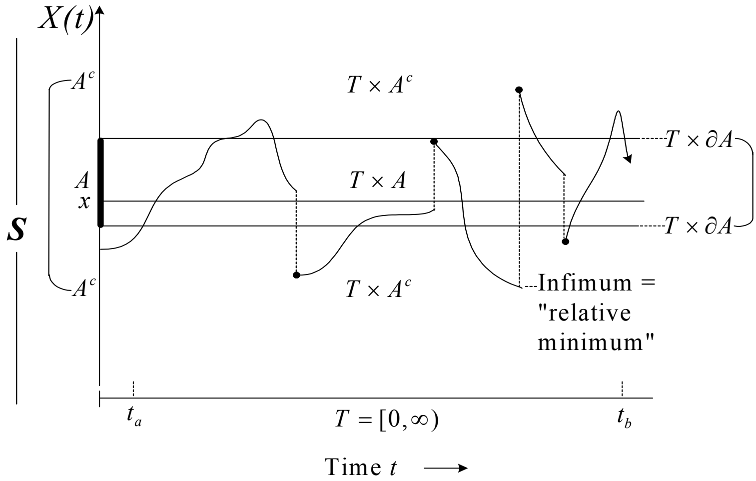

Figure 2.8: $X(t), t \geq 0$, during an arbitrary finite time interval (t_a, t_b) showing relative extrema, transitions with respect to set $\mathbf{T} \times \mathbf{A}$, and transitions with respect to a fixed level x.

(2) **Tangents**: Interior or exterior tangents can occur only at instants of relative extrema during (t_1, t_2).

(3) **Down- and upcrossings**: At most one SP down- or upcrossing of a fixed level x can occur between successive instants of relative extrema during (t_1, t_2).

From Definition 2.1, a sample path has at most a finite number of relative extrema (maxima, minima, infima (greatest lower bounds), suprema (least upper bounds)) during (t_1, t_2). Since $(0, t)$ is a finite interval for each fixed $t > 0$, the random variables in the hypothesis are discrete and finite. By Definition 2.1, their expected values are finite. ■

Corollary 2.1

$$\lim_{t\to\infty} (\mathcal{O}_t(\boldsymbol{A}) + \mathcal{I}_t(\boldsymbol{A}) + \mathcal{T}_t^o(\boldsymbol{A}) + \mathcal{T}_t^e(\boldsymbol{A}) + \mathcal{D}_t(x) + \mathcal{U}_t(x))$$

is countable.

Proof. The time axis $\boldsymbol{T} = [0, \infty) = \lim_{t\to\infty}[0, t)$. Countability follows since

$$\boldsymbol{T} = \cup_{n=0}^{\infty}[n, n+1)$$

which is a countable union of finite intervals. ■

2.4.6 Set and Rate Balance

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

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

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

From Proposition 2.3 for fixed $t > 0$

$$\mathcal{O}_t(\boldsymbol{A}) - \mathcal{I}_t(\boldsymbol{A}) = \begin{cases} -1 \\ 0 \\ +1 \end{cases} \tag{2.1}$$

depending on wether $X(0)$, $X(t)$ are in $\boldsymbol{A}$ or $\boldsymbol{A}^c$. Dividing both sides of (2.1) by t and letting $t \to \infty$, gives the *principle of set balance* for exits and entrances for set A, assuming the limits exist, as follows.

Principle of Set Balance

For every set $\boldsymbol{A} \subset \boldsymbol{S}$,

$$\left.\begin{array}{r} \lim_{t\to\infty} \frac{\mathcal{O}_t(\boldsymbol{A})}{t} \underset{(a.s.)}{=} \lim_{t\to\infty} \frac{\mathcal{I}_t(\boldsymbol{A})}{t}, \\ \lim_{t\to\infty} \frac{E(\mathcal{O}_t(\boldsymbol{A}))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{I}_t(\boldsymbol{A}))}{t}. \end{array}\right\} \tag{2.2}$$

When emphasizing entrance and exit rates of sets, we usually refer to (2.2).

2.4.7 Rate Balance for Down- and Upcrossings

Level crossings of $x \in \mathbf{S}$ are boundary crossings, where $\{x\}$ is a boundary of some subset of $\boldsymbol{S}$. Specifically, SP crossings occur at instants of exit from $\mathbf{T}\times((x,\infty)\cap\mathbf{S})$ and entrance into $\mathbf{T}\times((-\infty,x)\cap\mathbf{S})$**.** By Proposition 2.3, instants of down- and upcrossing alternate in time. Thus for each $t > 0$.

$$\mathcal{D}_t(x) - \mathcal{U}_t(x) = \begin{cases} -1 \\ 0 \\ +1 \end{cases} \tag{2.3}$$

depending on whether the values of $X(0)$, $X(t)$ are in (x, ∞) or in $(-\infty, x)$. Dividing (2.3) by t and letting $t \to \infty$, gives the *principle*

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

Table 2.2: Atoms (discrete states) in various models; and corresponding figures. Any other states are continuous.

of rate balance for down- and upcrossings across level $x \in \mathbf{S}$, assuming the limits exist,

$$\left.\begin{array}{r} \lim_{t\to\infty}\frac{\mathcal{D}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t}, \\ \lim_{t\to\infty}\frac{E(\mathcal{D}_t(x))}{t} = \lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}. \end{array}\right\} \tag{2.4}$$

When referring to strict level crossings we usually refer to (2.4) as *rate balance* across level x. However, occasionally we call (2.4) *set balance* if we emphasize that crossings are exits and entrances of the sets $\mathbf{T}\times((x,\infty)\cap\mathbf{S})$ and $\mathbf{T}\times((-\infty,x)\cap\mathbf{S})$.

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

2.4.8 Continuous and Discrete States

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

Continuous States

A continuous state $\{x\}$ is characterized by having probability 0. That is, $P(X(t) = x) = 0$, $t \geq 0$; and the steady state probability

$$\lim_{t\to\infty} P(X(t) = x) = 0.$$

The long-run proportion of time that $X(\cdot)$ spends in $\mathbf{T} \times \{x\}$ is 0.

The power of the LC method is largely due to the relationship between (1) rates of sample-path left-limit down- and upcrossing rates of a level x (i.e., at $\{t|X(t^-) = x\}$, and (2) the transient and/or steady-state pdf's of the state variable at level x.

Discrete States (Atoms)

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

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

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

Set Balance for Discrete States

Exits and entrances of a discrete state $\{x\} \subset \mathbf{S}$ alternate in time (Proposition 2.3). Finiteness of their numbers in finite time intervals yields

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

depending on the values of $X(0)$, $X(t)$ with respect to x. Dividing by t and letting $t \to \infty$ yields the principle of set balance for exits and entrances of atoms,

$$\left.\begin{array}{l} \lim_{t\to\infty} \frac{\mathcal{O}_t(\{x\})}{t} = \lim_{t\to\infty} \frac{\mathcal{I}_t(\{x\})}{t} \text{ (a.s.)}, \\ \lim_{t\to\infty} \frac{E(\mathcal{O}_t(\{x\}))}{t} = \lim_{t\to\infty} \frac{E(\mathcal{I}_t(\{x\}))}{t}. \end{array}\right\} \quad (2.5)$$

Equations (2.5) are precisely equivalent to the well known balance equations used for the rates into and out of states, in continuous-time discrete-state Markov chains (CTMC's). Conversely, the well known balance equations for CTMC's originally suggested to me, the notion of extending the "rate balance" idea to *continuous states* in continuous-time continuous-state Markov processes (or to mixed-state processes). (This was another sign post leading to the discovery of the LC methodology.)

2.4.9 Hits and Egresses of Levels

Hits

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

Egresses

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

Definition 2.9 ***Sample-path hit:*** $X(\cdot)$ ***hits*** *level x at instant t_0 if $X(t_0^-) = x$ (left limit) or if $X(t_0) = x$ and*

$$\exists \varepsilon > 0 \ni X(t) \neq x, t \in (t_0 - \varepsilon, t_0).$$

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

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

Definition 2.10 ***Sample-path egress:*** *A sample path makes an egress from level x at t_0 if $X(t_0^-) = x$ or if $X(t_0) = x$ and*

$$\exists \varepsilon > 0 \ni X(t) \neq x, t \in (t_0, t_0 + \varepsilon).$$

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

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

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

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

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

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

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

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

$$X(t_0) = x = \lim_{t \downarrow t_0} X(t).$$

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

Example 2.4 *Consider a typical sample path of the **virtual wait** $W(t)$, $t \geq 0$, for the **M/D/1 queue**. The state space is $\boldsymbol{S} = [0, \infty)$. Arrivals occur in a Poisson process at rate λ and every customer gets the same service time $D > 0$ (Fig. 2.9). All SP jumps are upward of size D. Consider level D, i.e., the line $\boldsymbol{T}\times\{D\}$. The SP hit rate of $\boldsymbol{T}\times\{D\}$ from*

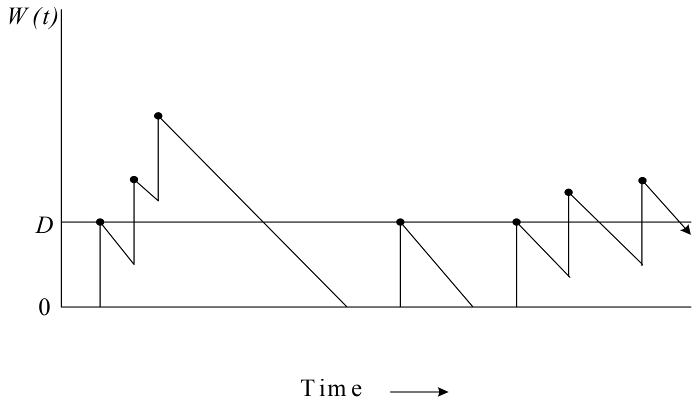

Figure 2.9: Sample path of virtual wait in $M/D/1$ queue. Service time $= D$ for all customers. Rate of hits of level D from above is less than rate of egress from level D below.

above is due exclusively to continuous left-limit downcrossings of level D. This rate is ***less*** *than the rate of egresses from level D below. The latter rate is due to continuous downcrossings of D* ***and*** *exterior, right-continuous (same as right-limit) tangents of the set* (D, ∞) *(tangents to D from below). The tangents touch level D at end points of jumps that start at level* 0*, at arrival instants when the system is empty. We show in Subsection 3.8 that singleton state* $\{D\}$ *is a* ***continuous*** *state (not an atom). If we assume* $\lambda D < 1$ *then the steady state pdf of wait exists. Level crossing theorems can be used to prove that the size of the discontinuity of the pdf of wait is given by* $f(D^-) - f(D) = \lambda P_0$*, where* $\{P_0; f(x), x > 0\}$ *is the steady state pdf of wait (Example 2.5).*

2.4.10 Rate Balance for Hits and Egresses

Notation 2.3 *Superscripts will have the following roles:*
"a": from above or (to) above;
"b": from below or (to) below;
"c": left-limit or left-continuous (e.g., $X(t_0^-) = x$*; same as continuous if* $X(t_0^-) = X(t_0) = x$*).*
"j": jump transition.

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

Notation 2.4 $\mathcal{H}_t^a(x)$, $\mathcal{H}_t^{a,c}(x)$: *number of sample-path hits and left-limit hits of level* x *from above during* $(0,t)$, *respectively.*

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

$\mathcal{T}_t^a(x)$, $\mathcal{T}_t^{a,c}(x)$: *numbers respectively of tangents and left-limit tangents of* x *from above during* $(0,t)$ *(interior tangents of*

$$\boldsymbol{T}\times((x,\infty)\cap\mathbf{S})).$$

$\mathcal{T}_t^b(x)$, $\mathcal{T}_t^{b,c}(x)$: *number respectively of tangents and left-limit tangents of* x *from below during* $(0,t)$ *(exterior tangents of*

$$\boldsymbol{T}\times((x,\infty)\cap\mathbf{S})).$$

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

Then

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

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

$$\mathcal{H}_t^{a,c}(D) = \mathcal{D}_t^c(D). \tag{2.6}$$

Also, since all hits of level D *from above are due to left-limit (same as continuous in this case) downcrossings,*

$$\begin{aligned}
\mathcal{E}_t^b(D) &= \mathcal{H}_t^{a,c}(D) + \mathcal{T}_t^b(D) \\
&= \mathcal{D}_t^c(D) + \mathcal{T}_t^b(D),
\end{aligned} \tag{2.7}$$

upon substitution from (2.6). In (2.7) dividing by t and letting $t \to \infty$ yields

$$\lim_{t\to\infty} \frac{\mathcal{D}_t^c(D)}{t} = \lim_{t\to\infty} \frac{\mathcal{E}_t^b(D)}{t} - \lim_{t\to\infty} \frac{\mathcal{T}_t^b(D)}{t}. \tag{2.8}$$

From Theorem 1.1

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

Substitution into (2.8) yields

$$f(D) = f(D^-) - \lambda P_0. \tag{2.9}$$

Equation (2.9) expresses an analytical property of the steady-state pdf of wait. The pdf has a jump discontinuity ***downward*** *at $x = D$ of size λP_0 (see Subsection 3.8.1). It has no other discontinuities for $x > 0$. In addition, every downcrossing and tangent from below of level D, has no motion in the direction of Time in $\boldsymbol{T} \times \{D\}$. The total number of such transitions of level D in $\boldsymbol{T} = [0, \infty)$ is countable. Thus the* ***proportion*** *of time spent at level D is 0. So $\{D\}$ is a continuous state.*

2.4.11 Hits and Egresses for Discrete States

A hit of a *discrete* state (atom) $\{x\} \subset \boldsymbol{S}$ may be an SP entrance into $\{x\}$, or instantaneous *left-limit* level crossing or *left-limit* tangent of level x. An egress out of a discrete state $\{x\}$ may be an SP exit from $\{x\}$ or instantaneous *right-continuous* level crossing or *right-continuous* tangent of level x. An SP jump may also be an instantaneous hit of, or egress from, level x.

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

The long-run proportion of time P_0 that the sample path spends in state $\{0\}$ is positive and

$$\lim_{t\to\infty} P(W(t) = 0) = P_0 > 0.$$

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

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

since every hit from above of each level $x > 0$ *is a continuous downcrossing of level* x.

The SP egress rate from level 0 *above is the exit rate from discrete state* $\{0\}$. *This is the rate at which customers arrive when the system is empty, namely* λP_0. *Set balance between the sets* $(0, \infty)$ *and* $\{0\}$, *equates entrance and exit rates of the atom* $\{0\}$. *Thus it yields yields the equation* $f(0) = \lambda P_0$, *which reveals a fundamental relation. It relates the continuous part of the pdf of a state variable "at" an atom, to the positive probability of that atom. Thus, the SP entrance rate into and exit rate out of the discrete state* $\{0\}$ *is* $f(0)$. *This type of relationship appears in different forms in various models, and is useful for computing steady-state distributions of state variables.*

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

Example 2.7 *Consider a* ***birth-death process*** *having states* $0, ..., N$ *(Fig. 2.10). Let the Poisson rate of jumps from* n *to* $n+1$ *be* λ_n, *and from* n *to* $n-1$ *be* μ_n, $n = 1, ..., N$. *The "usual" state space is the set of discrete states* $\boldsymbol{S} = \{0, 1, ..., N\}$ *having steady-state probabilities* $P_0, ..., P_N$ *respectively. Let* $\boldsymbol{S}$ *be* ***extended to the state space in the wide sense***, *i.e., he closed interval* $[0, N]$. *This extension does not change the probability distribution associated with the model. All probability is still concentrated on the discrete states* $0, ..., N$. *The totality of continuous states has probability* 0. *At instants of jump, the SP "moves" vertically in* $\boldsymbol{S}$, *not in Time.*

We derive the values of $P_0, ...P_N$ *using a level crossing argument. Consider a fixed level* x, $n < x < n+1$, $n = 0, ..., N-1$. *The down- and upcrossing rates of* x *are*

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{D}_t^j(x)}{t} = \mu_{n+1} P_{n+1}$$

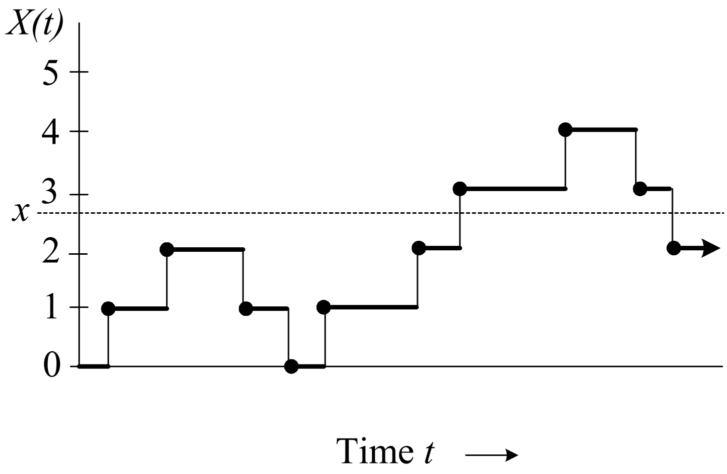

Figure 2.10: Sample path of birth-death process with $N = 5$ discrete states. State space (in the wide sense) is the interval $[0, 4]$.

and

$$\lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t^j(x)}{t} = \lambda_n P_n$$

respectively. By rate balance

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = \lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t}.$$

Thus we obtain the well known formulas

$$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n, n = 0, ..., N-1,$$

and

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

*Substituting into the normalizing condition $P_0 + ... + P_N = 1$ yields the solution for P_0. This leads to the values for $P_1, ..., P_N$. The above derivation appears to be identical to the standard "**rate in** = **rate out**" argument. However, the extension of the state space to the wide-sense state space, includes the continuous states. This allows us to use the LC method explicitly. The LC approach displays a subtle difference, which is prescient regarding solving more complex discrete-state continuous-time models.*

Example 2.8 *Consider an "**elevator-like" model** (Fig. 2.11) An elevator may stop at $N+1$ floors, $0,...,N$. Assume the elevator travels at constant speeds k and h meters per minute when moving respectively upward and downward between floors. We ignore the start-up acceleration and slow-down deceleration phases, for exposition. To fix ideas, assume the motion is in a semi-Markovian environment. From the instant the elevator stops at floor i, its sojourn time has mean μ_i minutes. Its next stop will be at floor j with probability $P_{ij}, i \neq j \in \{0,...,N\}$. The $(N+1) \times (N+1)$ matrix $\|P_{ij}\|$ is a Markov matrix. Assume the stationary probabilities are $\pi_i, i = 0,...,N$. Let the steady-state probability of the the elevator being at floor i be $P_i, i = 0,...,N$. Let the pdf of the position of the elevator when it is moving upward and downward between floors i and $i+1$ be respectively*

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

Let $f_i(x) = f_{i1}(x) + f_{i2}(x)$. The state space is $\boldsymbol{S} = [0, N]$. The discrete states (atoms) are $0,...,N$, representing the floors. The continuous states are points in the open intervals $(i, i+1), i = 0,...,N-1$. The total probability is concentrated on both the discrete and continuous states. Hence the total pdf of position will be "mixed", having piecewise continuous segments between the atoms and positive probabilities for the atoms. The normalizing condition is

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

The problem is to determine the values of the P_i's and the partial pdf's

$$f_i(x), x \in (i, i+1), i = 0,...,N = 1.$$

(We do not give the complete solution of this model. The machinery to solve it will be evident after perusing parts of chapters 4 and 6).

2.5 Transition Types Geometrically

In order to provide intuitive background, this section summarizes ***geometrically*** types of sample-path and SP transitions with respect to a level $x \in \boldsymbol{S}$ when $\boldsymbol{S} \subseteq \boldsymbol{R}$. We diagram *thirty-four different categories* of transitions that can occur in various models.

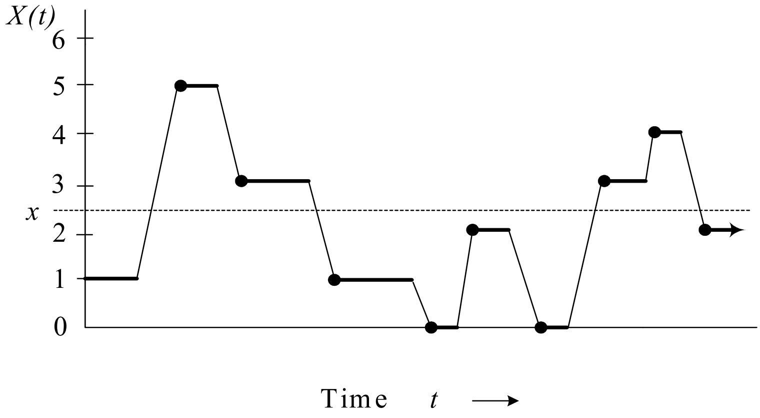

Figure 2.11: Sample path of elevator-like model, $N = 6$ floors. Discrete states are $0, ..., 5$. Continuous states are open intervals $(n, n+1), n = 0, ..., 4$.

Figs. 2.12, 2.13, 2.14, 2.15 illustrate four categories of transitions: SP hits from above and below; egresses above and below. In these figures, the instant of contact with level x is considered to be $t_0 > 0$.

In Figs. 2.12 and 2.13, "**left limit**" means $X(t_0^-) = x$; "**jump**" means $X(t_0^-) \neq x$ and $X(t_0) = x$.

In Figs. 2.14, 2.15, "**right limit**" means $X(t_0) = x$; "**jump**" means $X(t_0^-) = x$ and $X(t_0) \neq x$.

Fig. 2.16 illustrates level crossings that are not hits of or egresses from, level x.

Example 2.9 *In Fig. 2.12 consider the two sub-diagrams in position (Left Limit, Tangent from Above). The sub-diagrams represent geometrically the possible SP motion (left to right) with respect to level x; the SP makes a hit from above of level x which is a* ***left-limit tangent from above****. These two generic diagrams apply when $\{x\}$ is a continuous state or a discrete state (atom).*

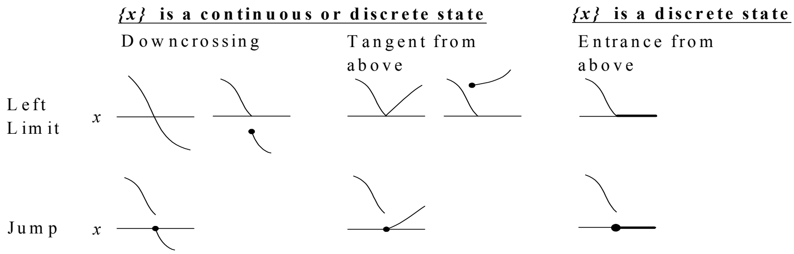

Figure 2.12: Sample-path hits of level $x \in S$ *from above*.

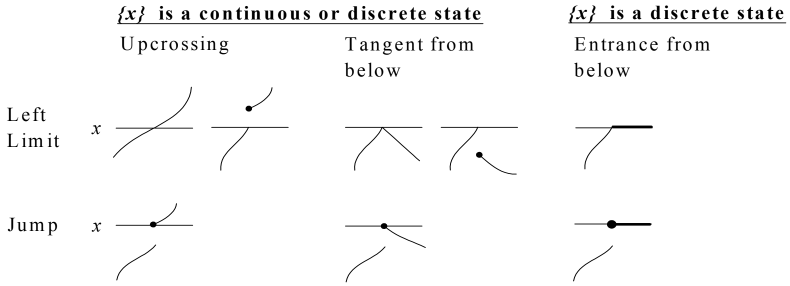

Figure 2.13: Sample-path hits of level $x \in S$ *from below*.

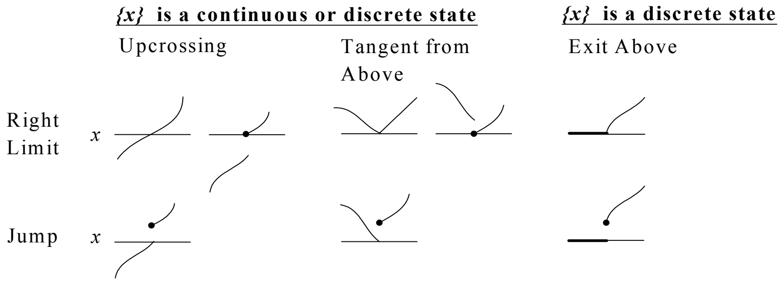

Figure 2.14: Sample-path egresses from level $x \in S$ *above*.

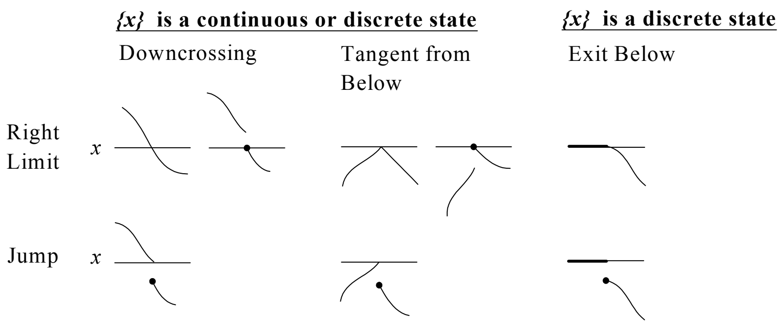

Figure 2.15: Sample-path egresses from level $x \in S$ *below.*

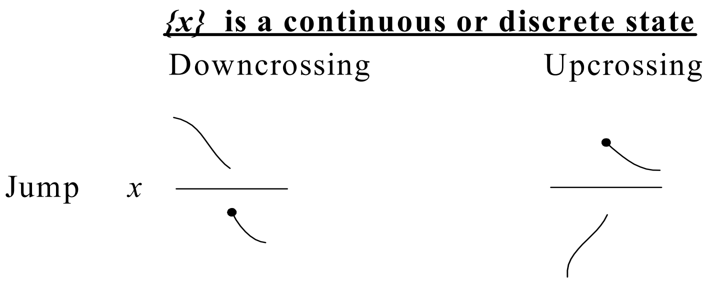

Figure 2.16: Jump downcrossing and upcrossing. No hit of, or egress from, level x.

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,

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

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

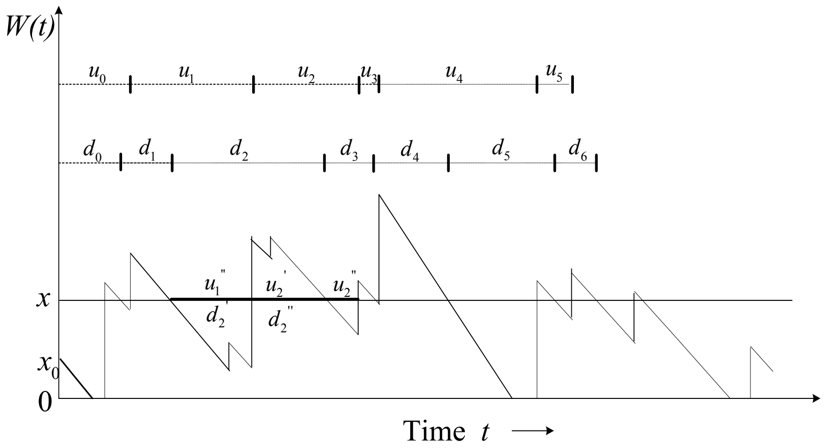

Figure 3.1: Sample path of virtual wait in M/G/1 showing inter down- and upcrossing times for level x, $\{d_n\}$, $\{u_n\}$, and their components, e.g., $d_2^{'}$, $d_2^{''}$, $u_2^{'}$, $u_2^{''}$, etc.

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

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

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

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

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)

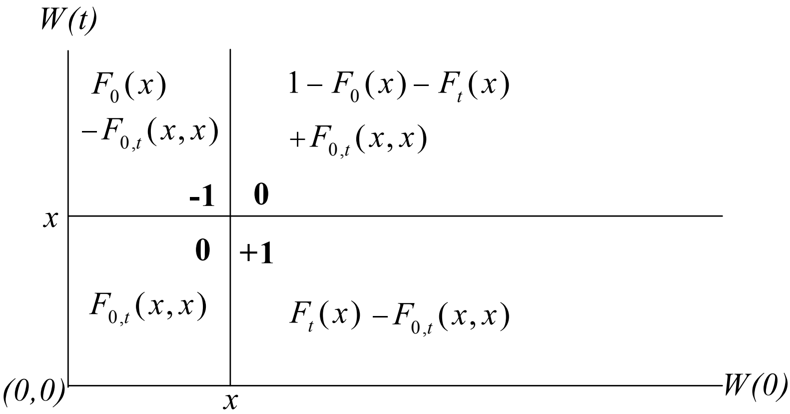

Figure 3.2: Values of $\mathcal{D}_t(x) - \mathcal{U}_t(x)$ are $+1$, 0, -1, with probabilities shown in areas of $(W(0), W(t))$ plane.

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

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

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*

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

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 $t > 0$,*

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

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

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*

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

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

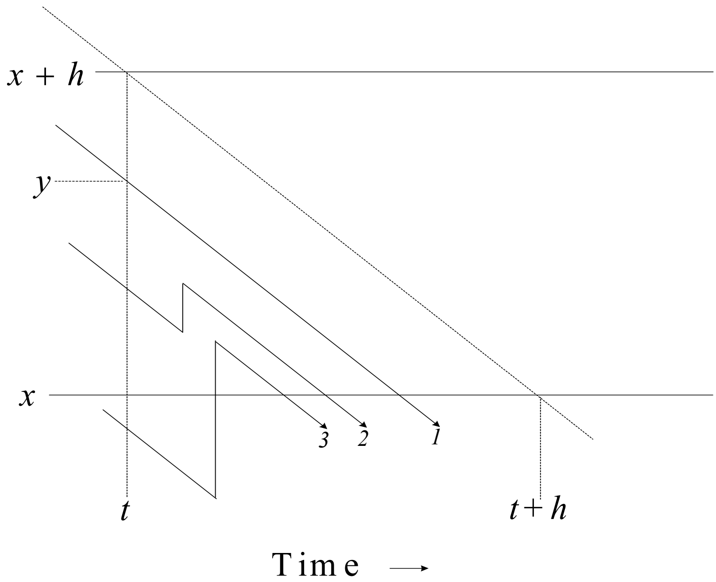

Figure 3.3: Sample path examples in time interval $(t, t+h)$ resulting in $D_t(x+h) - D_t(x) = 1$. Probabilities are: $P(\text{path type } 1) = 1 - \lambda(y - x) + o(y-x)$; $P(\text{path type } 2) \leq o(h)$; $P(\text{path type } 3) \leq o(h)$.

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
$+1$	$\lambda hP_0(t)\overline{B}(x)$ $+\lambda h\int_0^x\overline{B}(x-y)f_t(y)dy+o(h)$
≥ 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

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

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

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

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

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

Hence

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

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

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

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

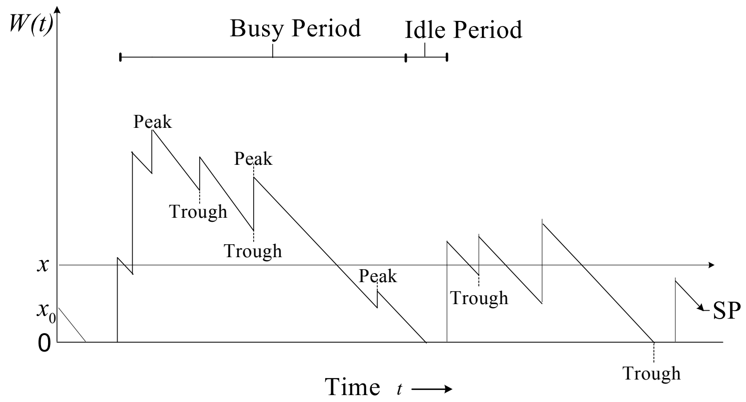

Figure 3.4: Sample path of virtual wait showing peaks and troughs, and a level x.

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

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

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

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

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),

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

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

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

or

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

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 $y \le x$ 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*

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

which is equivalent to

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

*The **stability condition** for the standard M/G/1 queue is $P_0 > 0$ (same as $\lambda E(S) < 1$). 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

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

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

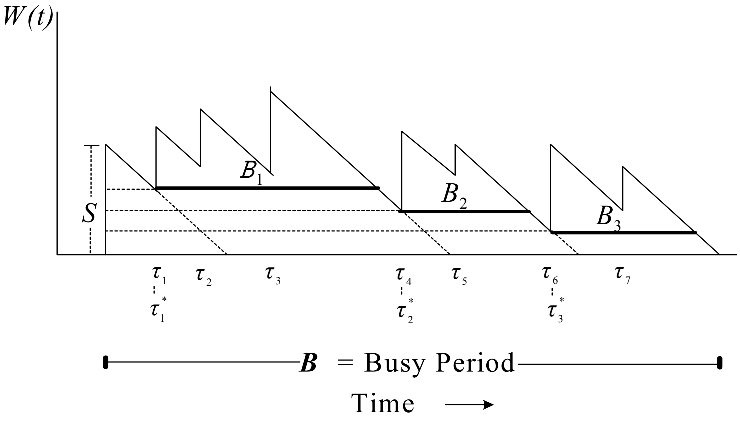

Figure 3.5: Busy period $\mathcal{B} \underset{dist}{=} S + \sum_{i=1}^{N_S} \mathcal{B}_i$. $\mathcal{B}_i \underset{dist}{=} \mathcal{B}, i = 1, ..., N_S$. N_S = number of "tagged" arrivals in $\mathcal{B}$. Here $N_S = 3$. $N_S \underset{dist}{=}$ number of arrivals during S. Tagged arrival times 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 $S_1, T_1, S_2, T_2, \ldots$ 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),

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

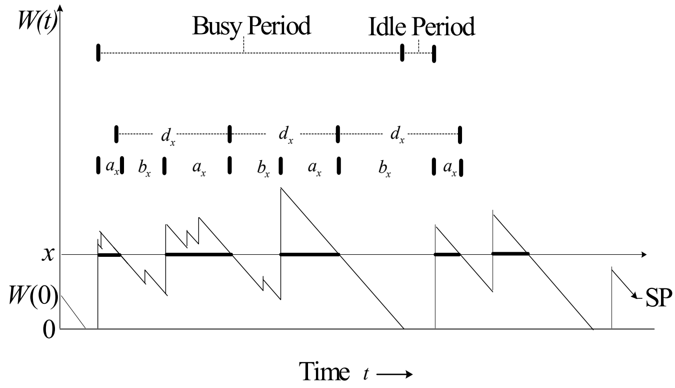

Figure 3.6: Sample path of virtual wait in M/G/1. Shows inter-downcrossing time d_x, sojourn a_x, sojourn b_x, busy and idle periods.

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 $E(b_y)$,$0 < y < x$,

$$\begin{aligned} F(x) &= P_0 e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x \geq 0, && (3.69) \\ f(x) &= \frac{P_0}{E(b_x)} e^{\int_{y=0}^{x} \frac{dy}{E(b_y)}}, x > 0. && (3.70) \end{aligned}$$

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

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

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.,

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

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 $P(N \geq n) = \rho^n, n = 0, ...$. 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

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

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

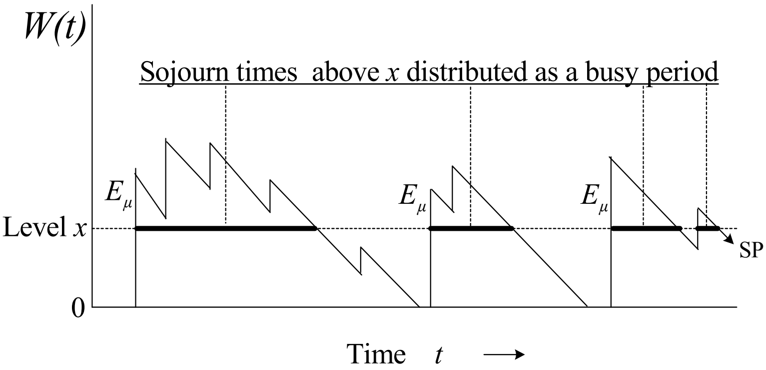

Figure 3.7: Sample path of virtual wait for $M_\lambda/M_\mu/1$ queue showing sojourns above level $x \underset{dist}{=} \mathcal{B}$ SP excess jumps above x are $\underset{dist}{\equiv} E_\mu$.

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).

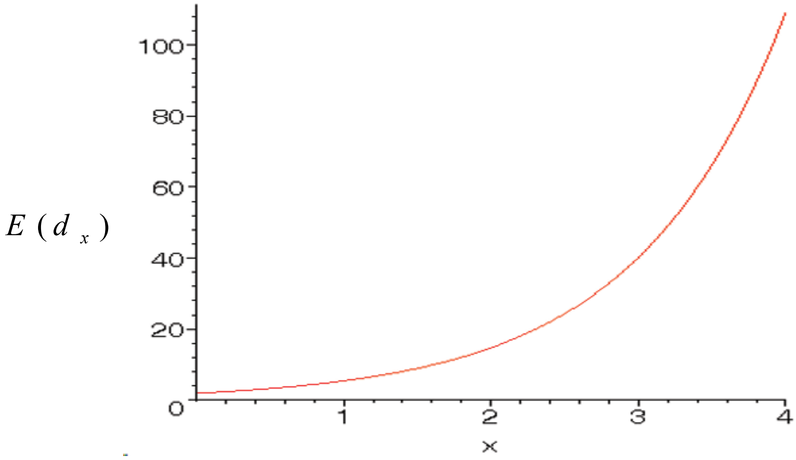

Figure 3.8: Expected inter-downcrossing (or inter-upcrossing) time of level x, $E(d_x)$ (or $E(u_x)$) in M/M/1: $\lambda = 1.0$, $\mu = 2.0$, $\rho = 0.5$.

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,

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

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 $\boldsymbol{S}$,*

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

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 $N_{\mathcal{B}} = \min(n | \sum_{i=1}^{n} S_i \le \sum_{i=1}^{n} T_i)$, 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*

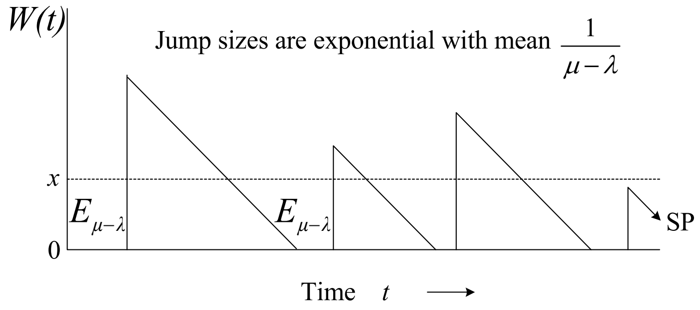

Figure 3.9: Sample path of workload for M/M/1/1 queue with arrival rate λ and service rate $\mu - \lambda$. Blocked customers are cleared.

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

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

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.

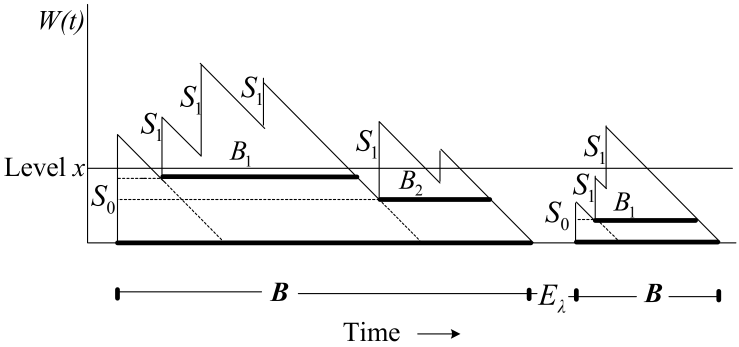

Figure 3.10: Busy periods $\{\mathcal{B}\}$ in $M_\lambda/G/1$ with zero-waits receiving service time $\underset{dist}{=} S_0$. $\{\mathcal{B}_i\}$ are busy periods of $M_\lambda/G/1$ with all service times $\underset{dist}{=} S_1$, generated by arrivals (as if) during S_0. In figure $B \equiv \mathcal{B}$.

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 $M_\lambda/G/1$ 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 $R_1R_2 = \mu_1\mu_2 - \mu_1\lambda_2 - \mu_2\lambda_1 > 0$ (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

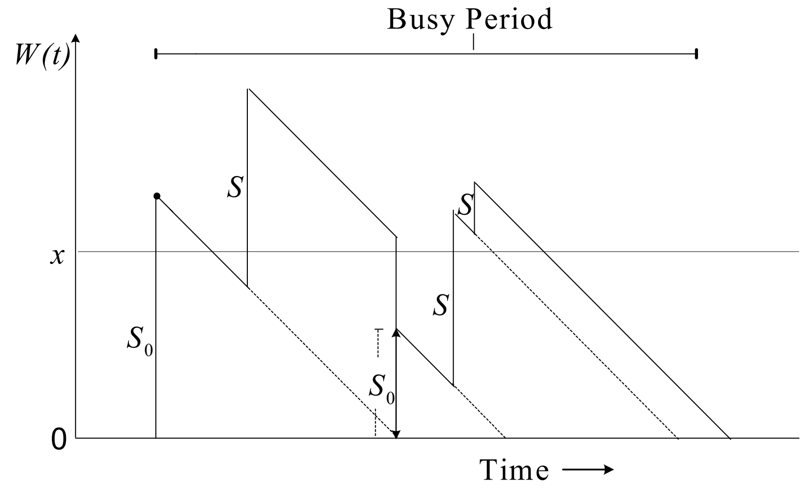

Figure 3.11: Sample path of virtual wait in M/G(N_q)/1 during a busy period. Shows jumps of size S_0 from level 0 and size S from positive levels. Illustrates possible downward jump in virtual wait.

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 $P_{0wk} + \int_{x=0}^{\infty} f_{wk}(x)dx = 1$.

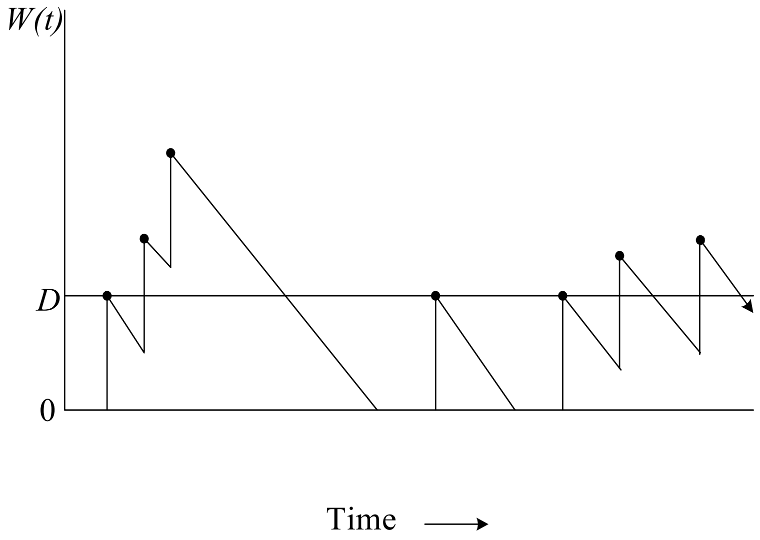

Figure 3.12: Sample path of virtual wait for M/D/1 queue.

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

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

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 $F(\cdot)$.

■

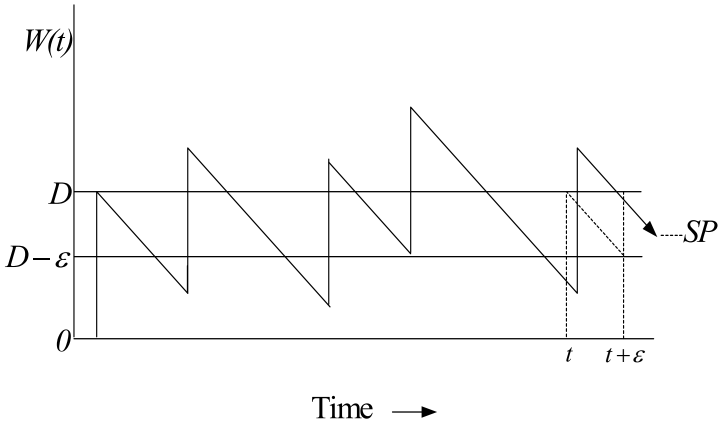

Figure 3.13: Sample path in M/D/1 showing levels D, $D-\varepsilon$ and instants t, $t+\varepsilon$. See Proposition 3.7, Proof, part (2).

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].

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

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

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

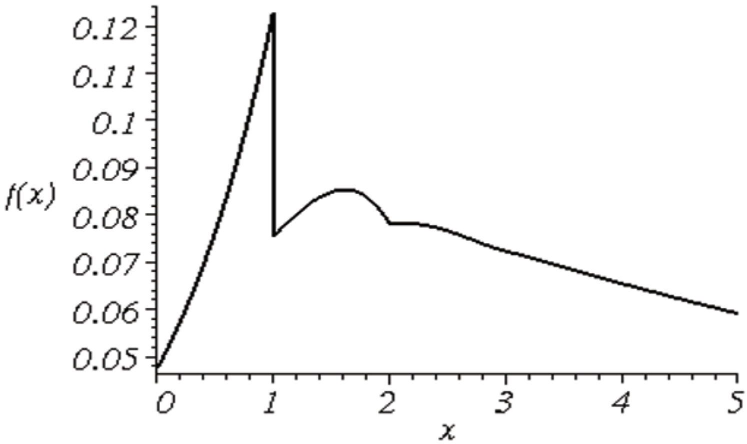

Figure 3.14: PDF $f(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$, $\rho = 0.95$ (high traffic). Shows discontinuity and downward jump of size λP_0 at $x = D$; and extreme waviness in right neighborhood $[D, 2D)$.

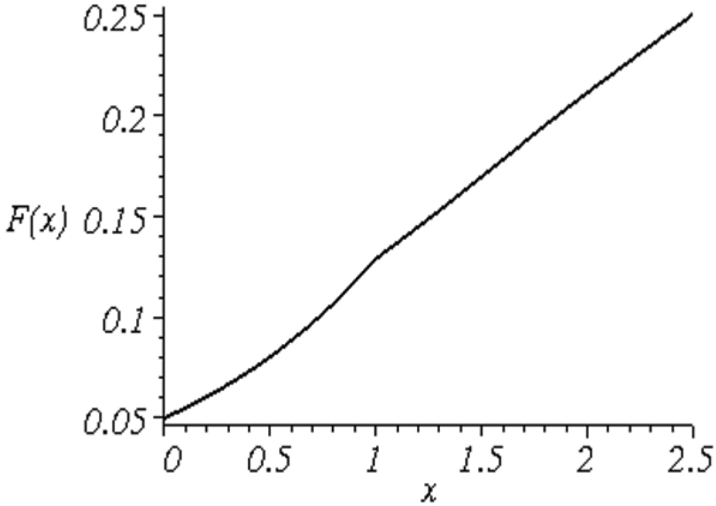

Figure 3.15: CDF $F(x)$ of wait in M/D/1: $\lambda = 0.95$, $D = 1$. Shows continuity of $F(x), x > 0$; and decrease in slope of $F(x)$ at $x = D$.

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)

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

The cdf of N is

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

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

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

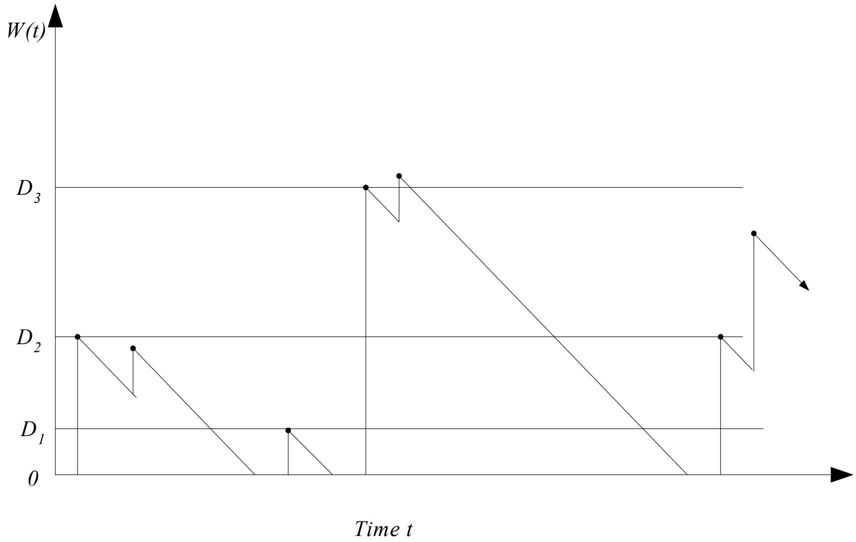

Figure 3.16: Sample path of virtual wait in $M/\{D_n\}/1$ queue with $N = 3$ service levels.

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., $x - \varepsilon < \min_{i=1,...,N}\{D_i - D_{i-1}\}$, 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. ■

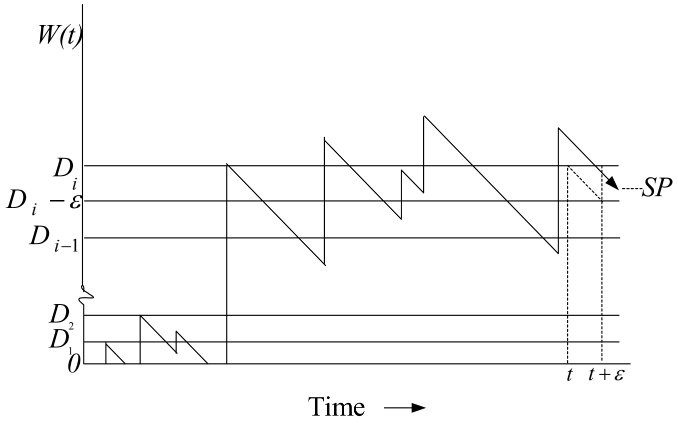

Figure 3.17: Sample path in M/{D$_n$}/1 showing levels $D_i, D_i - \varepsilon$ and instants t, $t + \varepsilon$. See Proposition 3.9, Proof, part (2).

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

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

$$= 1 - \lambda \sum_{i=1}^{N} p_i D_i.$$

3.9.3 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 $F(x)$),

$$f(x) = \lambda F(x) - \lambda \sum_{i=1}^{N} p_i F(x - D_i)$$

$$= \lambda F(x) - \sum_{i=1}^{N} \lambda_i F(x - D_i), x > 0. \tag{3.154}$$

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

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

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 $F(x) \equiv P(W_q \le x), x \ge 0$ 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

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

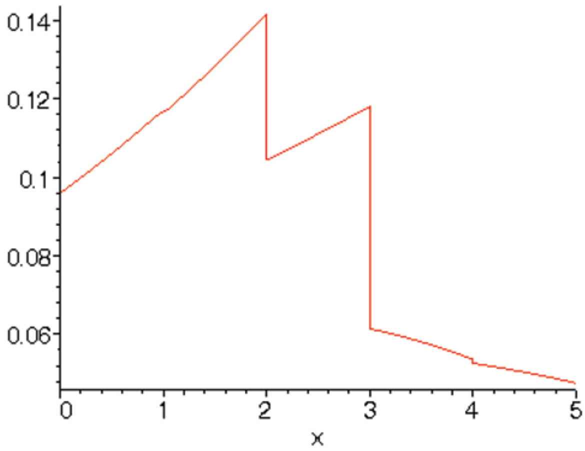

Figure 3.18: PDF of wait in $M/\{iD\}/1$ queue: four arrival types ($N = 4$), $\lambda = .2$, $p_1 = p_4 = .01$, $p_2 = .39$, $p_3 = .59$. Downward jumps at $x = 1, 2, 3, 4$.

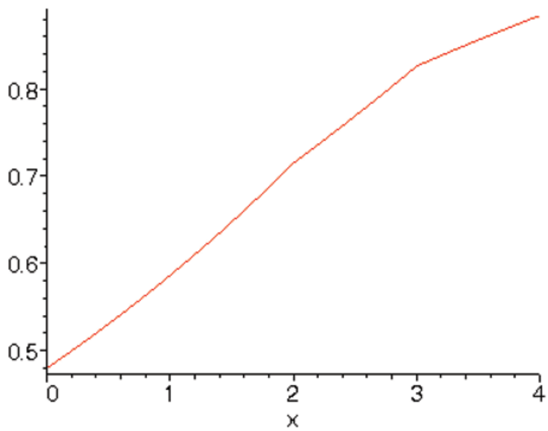

Figure 3.19: CDF of wait in $M/\{iD\}/1$ queue. $N = 4$, $\lambda = .2$, $p_1 = p_4 = .01$, $p_2 = .39$, $p_3 = .59$. Slope decreases abruptly at $x = 1, 2, 3, 4$.

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

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

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).

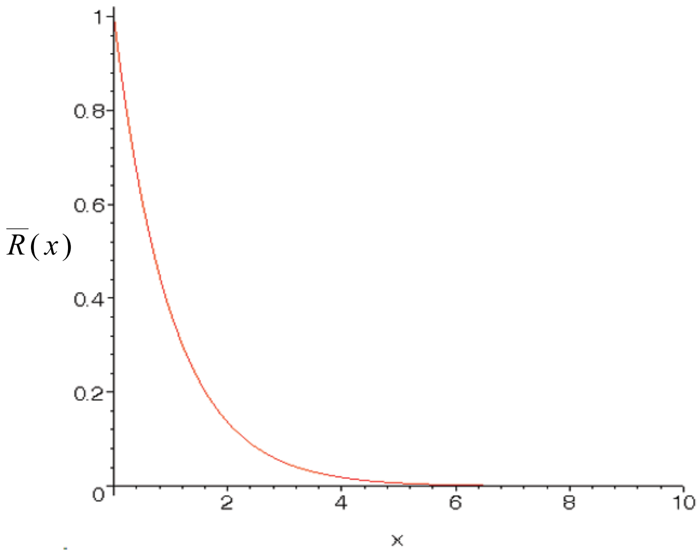

Figure 3.20: Staying function $\overline{R}(x) = e^{-rx.}$ $(r = 1)$; $L = 0$. $\overline{R}(x) = 1 - H(x)$ where $H(x)$ is a cdf.

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

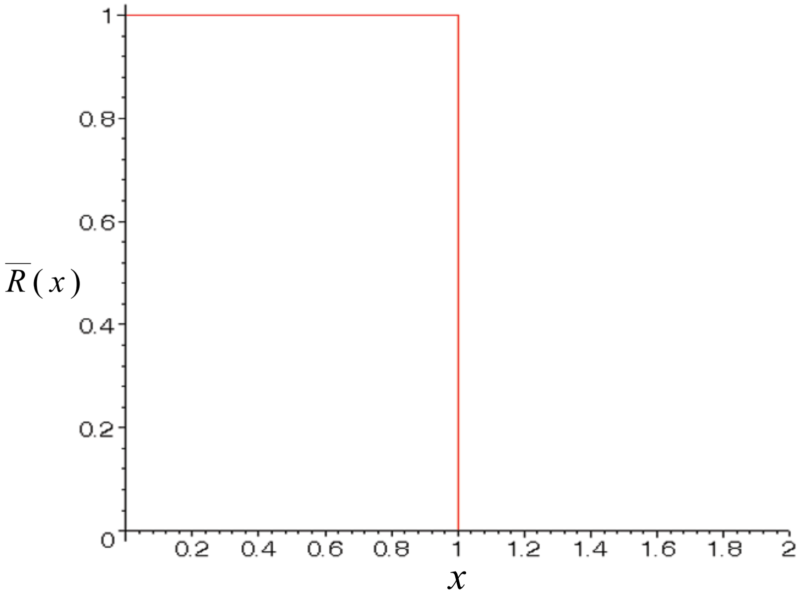

Figure 3.21: Staying function $\overline{R}(x) = 1, x < 1$, $\overline{R}(x) = 0, x \geq 1$. $L = 0$. $\overline{R}(x) = 1 - H(x)$, where $H(x)$ is a cdf.

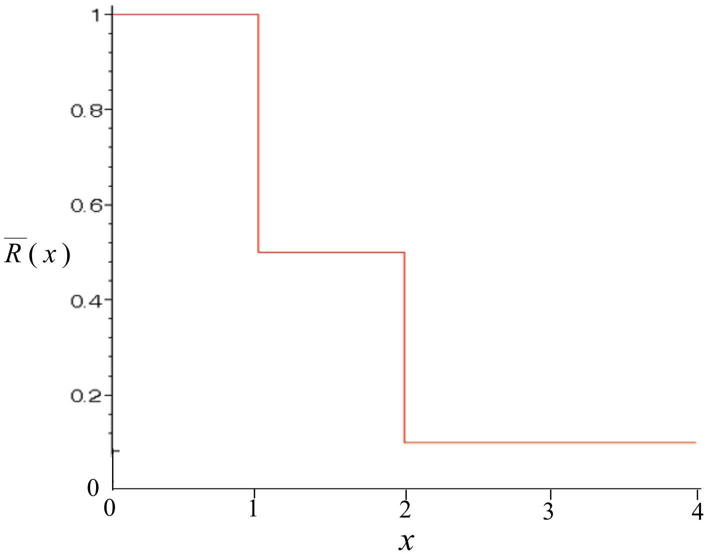

Figure 3.22: $\overline{R}(x) = 1, x < 1$, $\overline{R}(x) = 0.5, 1 \leq x < 2$, $\overline{R}(x) = 0.1, x \geq 2$. $\overline{R}(x) \neq 1 - H(x)$, where $H(x)$ is a cdf.

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

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

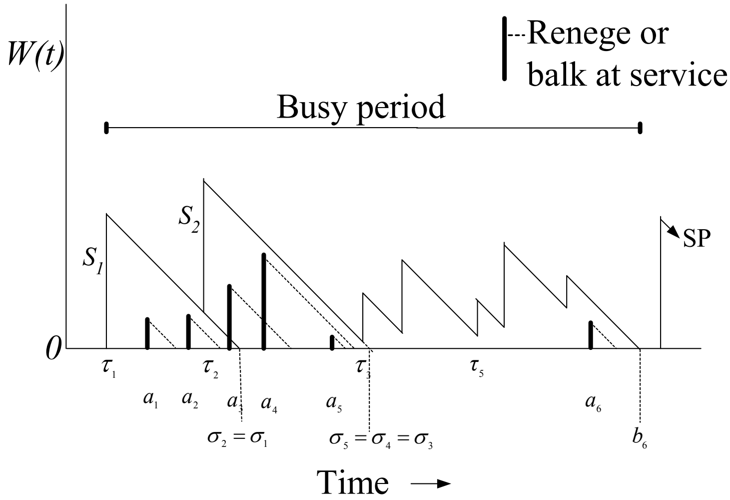

Figure 3.23: M/G/1 busy period showing stayers (τ_n), renegers (a_n) σ_n, and b_6 (end busy period), used to compute required waits of renegers.

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 $L \le \overline{R}(x), x \ge 0$

$$\begin{aligned}\lambda L x &= \lambda\int_{y=0}^{x} L dx \le \lambda\int_{y=0}^{x}\overline{R}(y)dy\\ &\Longleftrightarrow\ e^{-\mu x + \lambda L x} \le e^{\left(-\mu x + \lambda\int_{y=0}^{x}\overline{R}(y)dy\right)}\\ &\Longleftrightarrow\ \int_{x=0}^{\infty} e^{-(\mu-\lambda L)x}dx \le I.\end{aligned} \tag{3.170}$$

For given $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that $\overline{R}(x) < \varepsilon + L$ for $x > M_\varepsilon$. Thus

$$\begin{aligned}\lambda\int_{y=0}^{x}\overline{R}(y)dy &< \lambda\int_{y=0}^{M_\varepsilon}\overline{R}(y)dy + \lambda\int_{y=M_\varepsilon}^{x}(\varepsilon + L)\,dy\\ &= C_1 + \lambda(\varepsilon + L)\,x, x > M_\varepsilon\end{aligned}$$

$$\begin{aligned}&\Longrightarrow\ e^{\left(-\mu x + \lambda\int_{y=0}^{x}\overline{R}(y)dy\right)} < C_2 e^{-\mu x + \lambda(\varepsilon+L)x}, x > M_\varepsilon\\ &\Longrightarrow\ \int_{x=M_\varepsilon}^{\infty} e^{\left(-\mu x + \lambda\int_{y=0}^{x}\overline{R}(y)dy\right)}dx < C_2\int_{x=M_\varepsilon}^{\infty} e^{(-\mu+\lambda L+\lambda\varepsilon)x}dx\\ &\Longrightarrow\ I < C_3 + C_2\int_{x=M_\varepsilon}^{\infty} e^{(-\mu+\lambda L+\lambda\varepsilon)x}dx,\end{aligned} \tag{3.171}$$

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.*

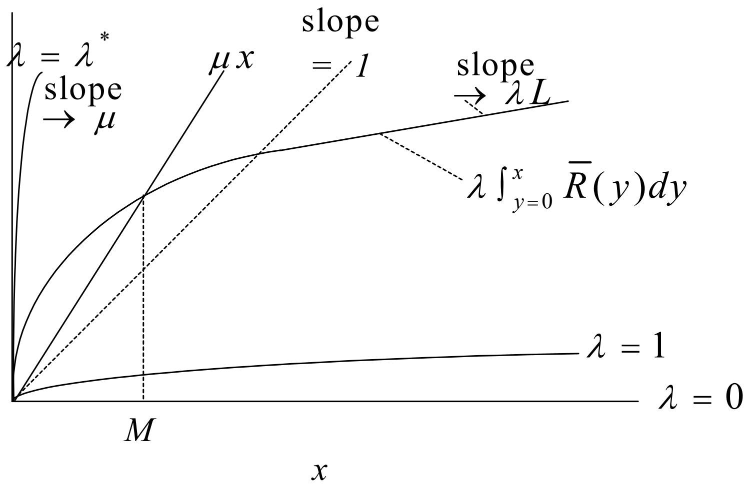

Figure 3.24: Functions μx and $\lambda \int_{y=0}^{x} \overline{R}(y)dy$, indicating M such that $\mu x - \lambda \int_{y=0}^{x} \overline{R}(y)dy > 0$ for $x \geq M$. Indicates range $0 < \lambda < \lambda^*$ such that stability holds. System is stable for λ if $\lambda \int_{y=0}^{x} \overline{R}(y)dy$ 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),

$$E(\mathcal{B}) = \frac{1-P_0}{f(0)} = \frac{1-P_0}{\lambda P_0}$$
$$= e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx = \int_{x=0}^{\infty} e^{-\mu x + \frac{\lambda}{r}(1-e^{-rx})} dx. \qquad (3.178)$$

(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

$$q_S = P_0 + \int_{x=0}^{\infty} e^{-rx} f(x)dx$$
$$= \frac{\left(1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx} - rx} dx\right)}{\left(1 + \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx\right)}. \quad (3.180)$$

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.

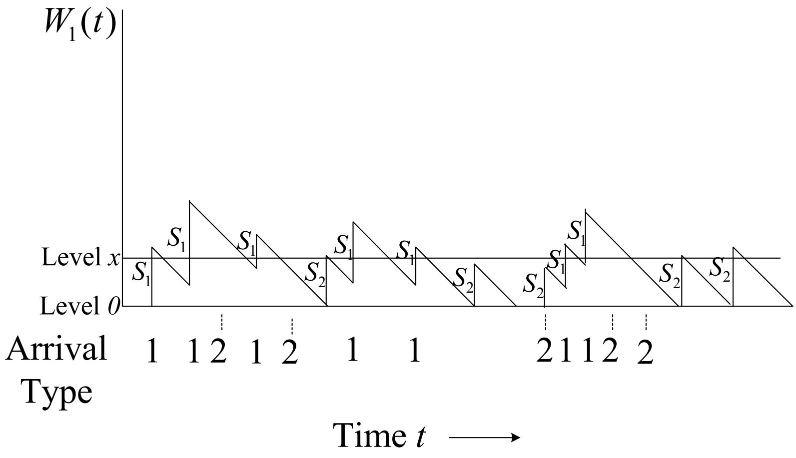

Figure 3.25: Sample path of virtual wait for high priority type-1 arrivals. Low priority type-2 arrivals that must wait, start service at the end of a $\mathcal{B}_1$ or a $\mathcal{B}_{21}$ (Fig. 3.26) busy period. All type 2 jumps start at level 0.

3.12.2 Equation for PDF of Wait of Type-1 Customers

From the sample path, we construct an integral equation for the pdf $f_1(x)$, $x > 0$,

$$\begin{aligned} f_1(x) = \quad & \lambda_1\overline{B}_1(x)P_0 + \lambda_2\overline{B}_2(x)P_0 + \lambda_1\int_{y=0}^{x}\overline{B_1}(x-y)f_1(y)dy \\ & +\lambda_2(1-P_0)\overline{B}_2(x). \end{aligned} \tag{3.181}$$

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

$$\begin{aligned} E(W_{q1}) = \quad & \left(\lambda_1 \tfrac{E(S_1^2)}{2} + \lambda_2 \tfrac{E(S_2^2)}{2}\right) P_0 + \lambda_1 E(S_1) E(W_{q1}) \\ & + \lambda_1 (1 - P_0) \tfrac{E(S_1^2)}{2} + \lambda_2 (1 - P_0) \tfrac{E(S_2^2)}{2}. \end{aligned}$$

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 $\mathcal{B}_{21}$.

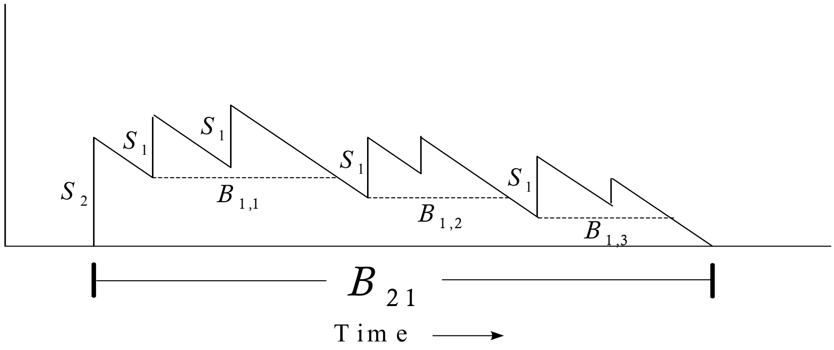

Figure 3.26: Busy period $\mathcal{B}_{21}$. Initial jump is a type 2 service S_2. Each subsequent jump is a type 1 service S_1. $\mathcal{B}_{1,j}$, $j = 1, 2, ...$, are M/G/1 type 1 busy periods.

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

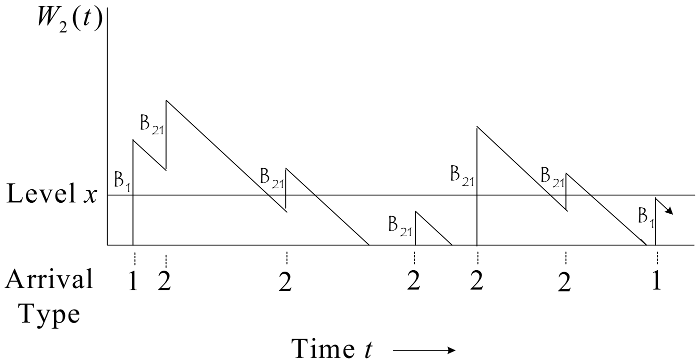

Figure 3.27: Sample path of virtual wait for *low priority, type 2 arrivals.* High priority type 1's that arrive when the system is empty generate jumps distributed as $\mathcal{B}_1$busy periods. All type 2 arrivals generate jumps distributed as $\mathcal{B}_{21}$ busy periods (see Fig. 3.26). All type 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

$$E(W_{q2}) = \frac{\left(\lambda_1 \frac{E(S_1^2)}{(1-\rho_1)^3}(1-\rho_1-\rho_2) + \lambda_2 E(\mathcal{B}_{21}^2)\right) \cdot (1-\rho_1)}{2\,(1-\rho_1-\rho_2)}. \tag{3.188}$$

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

$$\begin{aligned} \lambda_2 E(\mathcal{B}_{21}^2) = \quad & \lambda_2 E(S_2^2) + 2\lambda_2 E(S_2^2)\tfrac{\rho_1}{1-\rho_1} \\ & + \rho_2 \lambda_1 \tfrac{E(S_1^2)}{(1-\rho_1)^3} + \lambda_2 \tfrac{\rho_1^2}{(1-\rho_1)^2} E(S_2^2). \end{aligned} \tag{3.189}$$

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

$$\begin{aligned}E(W_{q2}) &= \frac{\frac{\lambda_1}{(1-\rho_1)}E(S_1^2) + \frac{\lambda_2}{(1-\rho_1)}E(S_2^2)}{2\,(1-\rho_1-\rho_2)}\\ &= \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1-\rho_1)\,(1-\rho_1-\rho_2)},\end{aligned} \tag{3.190}$$

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 $f_1(x)$,

$$\begin{aligned}f_1(x) = \quad & \lambda_1 e^{-\mu_1 x}P_0 + \lambda_2 e^{-\mu_2 x}P_0 + \lambda_1 \textstyle\int_{y=0}^{x} e^{-\mu_1(x-y)} f_1(y)dy\\ & +\lambda_2(1-P_0)e^{-\mu_2 x}.\end{aligned} \tag{3.191}$$

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: $P_0=1-\frac{\lambda_1}{\mu_1}$; $a=\lambda_1\left(1-\frac{\lambda_1}{\mu_1}\right)$; $b=0$ ✓.

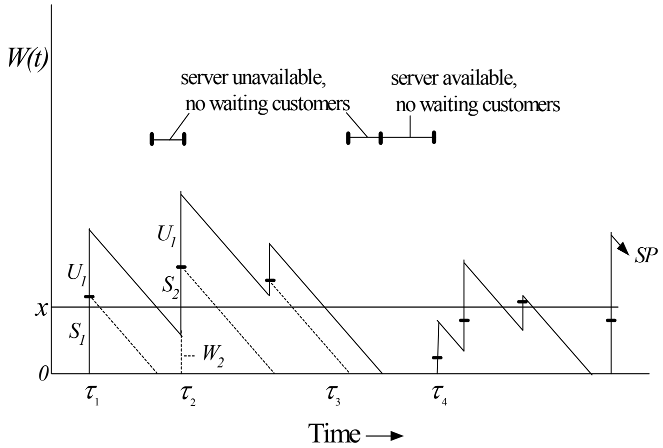

Figure 3.28: Sample path of virtual wait in M/G/1 queue with a server vacation after each service completion.

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$_\lambda$/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 $W(t) \leq K, t \geq 0$. 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

$$f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y) f(y) dy, 0 < x < K. \tag{3.202}$$

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.

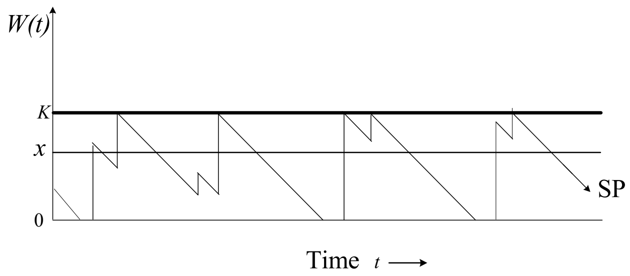

Figure 3.29: Variant 1. Sample path of virtual wait in M/G/1 with bounded virtual wait (bounded system time)

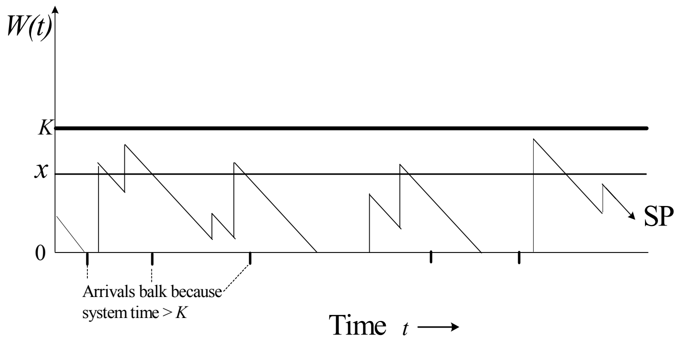

Figure 3.30: Variant 2. Sample path of virtual wait in M/G/1 with bounded virtual wait (boundrd system time)

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.

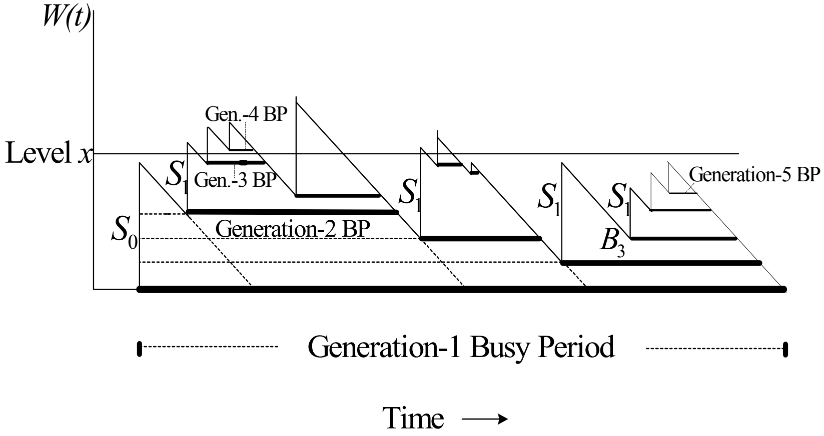

Figure 3.31: Multiplicative structure of busy-period. Each arrival generates an initial jump of a busy period of some generation. Initial busy periods of all generations account for all arrivals.

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

$$f(x) = P_0\rho_0 \textstyle\sum_{k=1}^{\infty} \rho_1^{k-1} \left(g_0 * g_1^{(k-1)}\right)(x), \tag{3.210}$$

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.

CHAPTER 4

M/M/C QUEUES

4.1 Introduction

In Section 4.2 we prove a useful general result about SP transitions, which we call Theorem B. This theorem facilitates the analysis of *transient* distributions of state variables, and will be applied in various sections of the sequel.

From Section 4.3 to Subsection 4.6.7 we describe a *generalized* M/M/c model, having a variety of service mechanisms. In this model the SP can *jump* between disjoint state-space sets (called *pages* or *sheets*). Geometrically, sheets are analogous to a package of sheets of paper, or to pages of a book. They are disjoint subsets of the state space that cannot be connected by a continuous segment of the sample path. (We can also model complex *single-server* queues using sheets, e.g., M/G/1 with Markov-generated server vacations [29], or M/G^{a,b}/1 bulk-server queues [71]). The concept of sheets can be applied to analyze other stochastic models as well, e.g., dams, inventories, production-inventory models, etc. Subsections 4.6.9 – 4.6.17 develop equations for transient and steady-state pdf's of wait in the generalized M/M/c model.

Section 4.8 derives known results for the standard M/M/c queue as a special case of the generalized model. Sections 4.7 – 4.12 provide steady-state analyses of M/M/c variants using LC. These analyses provide empirical background for potentially novel applications of LC.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_4,

4.2 Theorem B for Transient Analysis

We state and prove Theorem B. This straightforward theorem facilitates the transient analyses of a variety of stochastic models.

4.2.1 Theorem B

We first give a fundamental generalization of Theorems 3.1 and 3.2 of Chapter 3. It is useful for LC derivations of integro-differential equations for transient distributions in general.

Let $X(t), t \geq 0$, denote a sample path of a general stochastic process with state space $\boldsymbol{S}$. Let $\boldsymbol{A}$, $\boldsymbol{B}$ denote arbitrary measurable subsets of $\boldsymbol{S}$. Denote the transient probability $P(X(t) \in \boldsymbol{A})$ at instant t by $P_t(\boldsymbol{A}), t \geq 0$. Let $P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B})$ denote the joint probability $P(X(t_1) \in$ '$\boldsymbol{A}$, $X(t_2) \in \boldsymbol{B})$ at instants $t_1, t_2 \geq 0$. Let $\mathcal{I}_t(\boldsymbol{A})$ denote the number of SP entrances and $\mathcal{O}_t(\boldsymbol{A})$ the number of SP exits of $\boldsymbol{A}$, during $(0, t)$ (see Fig. 2.7). Assume the derivatives

$$\frac{\partial}{\partial t} P_t(\boldsymbol{A}), \frac{\partial}{\partial t} E(\mathcal{I}_t(\boldsymbol{A})), \frac{\partial}{\partial t} E(\mathcal{O}_t(\boldsymbol{A}))$$

exist for all $t > 0$. Note that both

$$\frac{\partial}{\partial t} E(\mathcal{I}_t(\boldsymbol{A})) > 0, \ \frac{\partial}{\partial t} E(\mathcal{O}_t(\boldsymbol{A})) > 0,$$

since $\mathcal{I}_t(\boldsymbol{A})$, $\mathcal{O}_t(\boldsymbol{A})$ are counting processes which increase (wide sense, i.e., not strictly) as t increases. The following useful result holds.

Theorem 4.1 Theorem B *(P.H. Brill, 1983)*

$$E(\mathcal{I}_t(\boldsymbol{A})) = E(\mathcal{O}_t(\boldsymbol{A})) + P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}) \tag{4.1}$$

$$\frac{\partial}{\partial t} E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t} E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t} P_t(\boldsymbol{A}) \tag{4.2}$$

Proof. We give two proofs in order to develop intuition about the result.

Proof 1: This proof is similar to that of Theorems 3.1 and 3.2. We make the correspondence:

$$\boldsymbol{A} \leftrightarrow (-\infty, x], \ \mathcal{I}_t(\boldsymbol{A}) \leftrightarrow \mathcal{D}_t(x), \ \mathcal{O}_t(\boldsymbol{A}) \leftrightarrow \mathcal{U}_t(x),$$
$$P_t(\boldsymbol{A}) \leftrightarrow F_t(x), t \geq 0, \ P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{A}) \leftrightarrow F_{t_1,t_2}(x, x), t_1, t_2 \geq 0.$$

SP down- and upcrossings of level x are entrances and exits of sets (Definitions 2.2, 2.3, 2.4, 2.5). Note that

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = +1 &\iff X(0) \in \boldsymbol{A}^c,\ X(t) \in \boldsymbol{A}, \\
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = -1 &\iff X(0) \in \boldsymbol{A},\ X(t) \in \boldsymbol{A}^c, \\
\mathcal{I}_t(\boldsymbol{A}) - \mathcal{O}_t(\boldsymbol{A}) = 0 &\iff X(0) \in \boldsymbol{A},\ X(t) \in \boldsymbol{A} \\
&\qquad \text{or } X(0) \in \boldsymbol{A}^c,\ X(t) \in \boldsymbol{A}^c.
\end{aligned}$$

We thus obtain the following values and corresponding probabilities:

$\mathcal{I}_t(\boldsymbol{A})-\mathcal{O}_t(\boldsymbol{A})$	Probability
$+1$	$P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) = P_t(\boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$
-1	$P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c) = P_0(\boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$
0	$1 - P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}) + 2P_{0,t}(\boldsymbol{A}, \boldsymbol{A})$

Taking the expected value $E(\mathcal{I}_t(A) - \mathcal{O}_t(A))$ and then the derivative $\frac{\partial}{\partial t}E(\mathcal{I}_t(A) - \mathcal{O}_t(A))$ yields (4.1) and (4.2).

Proof 2: Fix $t \geq 0$. The probability of the sure event $\boldsymbol{S}$ is

$$P_t(\boldsymbol{S}) = P_t(\boldsymbol{A} \cup \boldsymbol{A}^c) = P_t(\boldsymbol{A}) + P_t(\boldsymbol{A}^c) = 1.$$

Consider $P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{S})$. That is, events $\{X(t_1) \in \boldsymbol{A}\}$ and $\{X(t_2) \in \boldsymbol{S}\}$ are independent for every $0 \leq t_1 \neq t_2$. Knowledge that $\{X(t_1) \in \boldsymbol{A}\}$ has occurred, does not effect the probability of event $\{X(t_2) \in \boldsymbol{S}\}$, which is $P_{t_2}(\mathbf{S}) = 1$; and vice versa. Similarly, the events $\{X(t_1) \in \boldsymbol{S}\}$ and $\{X(t_2) \in \boldsymbol{B}\}$ are independent. Note that $\boldsymbol{S} = \boldsymbol{A} \cup \boldsymbol{A}^c = \boldsymbol{B} \cup \boldsymbol{B}^c$. Hence

$$\left.\begin{aligned}
P_{t_1}(\boldsymbol{A}) = P_{t_1,t_2}(\boldsymbol{A}, \mathbf{S}) = P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B} \cup \boldsymbol{B}^c), \\
P_{t_2}(\boldsymbol{B}) = P_{t_1,t_2}(\boldsymbol{S}, \boldsymbol{B}) = P_{t_1,t_2}(\boldsymbol{A} \cup \boldsymbol{A}^c, \boldsymbol{B}),
\end{aligned}\right\}$$

or

$$\left.\begin{aligned}
P_{t_1}(\boldsymbol{A}) = P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}) + P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}^c), \\
P_{t_2}(\boldsymbol{B}) = P_{t_1,t_2}(\boldsymbol{A}, \boldsymbol{B}) + P_{t_1,t_2}(\boldsymbol{A}^c, \boldsymbol{B}).
\end{aligned}\right\} \tag{4.3}$$

The possible values of $\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$ and corresponding joint probabilities at time points $t_1 = 0$ and $t_2 = t > 0$ are:

$\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$	Probability
0	$P_{0,t}(\mathbf{A}, \mathbf{A}) + P_{0,t}(\mathbf{A}^c, \mathbf{A}^c)$
$+1$	$P_{0,t}(\mathbf{A}^c, \mathbf{A})$
-1	$P_{0,t}(\mathbf{A}, \mathbf{A}^c)$

(4.4)

Taking the expected value of $\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})$ in (4.4) yields

$$\begin{aligned} E\left(\mathcal{I}_t(\mathbf{A}) - \mathcal{O}_t(\mathbf{A})\right) &= P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) - P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c) \\ &= P_{0,t}(\boldsymbol{A}^c, \boldsymbol{A}) + P_{0,t}(\boldsymbol{A}, \boldsymbol{A}) \\ &\quad - \left(P_{0,t}(\boldsymbol{A}, \boldsymbol{A}) + P_{0,t}(\boldsymbol{A}, \boldsymbol{A}^c)\right) \\ &= P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}), \end{aligned}$$

which gives (4.1). Taking $\frac{\partial}{\partial t}$ in (4.1) yields (4.2). ■

Remark 4.1 *Theorem B also applies to multi-dimensional processes with state space* $\boldsymbol{S} \subseteq \boldsymbol{R}^n, n = 2, \ldots$, *whose states are described by more than one continuous random variable. (We analyze two multi-dimensional inventory models in steady state, in Chapter 7.)*

4.3 Generalized M/M/c Model

Customers arrive at an M/M/c queue in a Poisson stream at rate λ. Arrivals start service from the first available server, in order of arrival. We assume that for each arrival, the service time is exponentially distributed with rate selected from a non-empty set $\boldsymbol{\mu} = \{\mu_0, ..., \mu_J\}$ of $J + 1$ positive constants. The rate selected depends on a server-assignment policy specified for the model.

We next define: virtual wait; server workload; system configuration; system point process (SP process).

4.4 Virtual Wait and Server Workload

Let C(t) be a *potential* time-t arrival to the system, $t \geq 0$. Let $R_i(t)$ denote the remaining workload (in time units) of server i, $(i = 1, ..., c)$, at instant $t \geq 0$. Let $W(t)$ be the virtual wait at instant t. Random variable $W(t)$ is the would-be wait required by C(t) measured from t until its start of service. Thus

$$W(t) = \min_{i=1,..,c} \left\{R_i(t)\right\}, t \geq 0.$$

We assume that sample paths of $W(t)$ and $\left\{R_i(t)\right\}, t \geq 0$ are right continuous and have left limits.

Let C_t be an *actual* "time-t" arrival to the system. We assume that C_t arrives at t^-. Then C_t waits an amount $W(t^-)$ before starting service from some server, say server i_t^*. If $R_i(t^-) = 0$ for some i, then $W(t^-) = 0$, and i_t^* is an idle server. Server i_t^* would be selected in accordance with a prescribed server-assignment policy. If $R_i(t^-) > 0, i = 1, ..., c$, then $W(t^-) > 0$ and C_t starts service from server i_t^* at instant $t + W(t^-)$.

4.4.1 Sample Path of Server Workload

In some models, sample paths of $R_i(t), i = 1, ..., c$, are useful for the overall analysis. We outline how to construct a sample path of $R_i(t), t \geq 0$, $i = 1, ..., c$. Without loss of generality, assume the system is empty at $t = 0$. Then $R_i(t) = 0, i = 1, ..., c$, from instant 0 until the first arrival time. When an arrival starts service from a server, that server's workload jumps by a service time. It then decreases steadily at time-rate 1 as service progresses.

Eventually at some instant after the system runs for a while, all c servers become occupied. Let $t_1 = \min\{t|\text{all } c \text{ servers are occupied}\}$. Suppose the next customer C_τ arrives at instant $\tau > t_1$, before any further service completions. Then C_τ is the sole customer waiting at time τ for service. Let C_τ's server be i_τ^*. The workload of server i_τ^* at τ^- is $R_{i_\tau^*}(\tau^-) = \min_{i=1,...,c}\{R_i(\tau^-)\} = W(\tau^-)$. The workload $R_{i_\tau^*}(\tau^-)$ jumps upward by the service time S_τ of C_τ. Thus $R_{i_\tau^*}(\tau) = R_{i_\tau^*}(\tau^-) + S_\tau = W(\tau^-) + S_\tau$. For *all other servers*, the workload is unchanged at τ. That is $R_i(\tau) = R_i(\tau^-), i \neq i_\tau^*$. The next arrival that finds all servers busy will be assigned to that server which has minimum workload, and so forth.

4.4.2 Distinguishable Servers

When tracking server workloads, we regard the servers as distinguishable. We are often interested in the statistical properties of the entire system, rather than the processing of each individual customer, or the action of a particular server. Hence, to analyze the system we may construct a sample path using randomly generated service times at arrival instants, in accordance with the prescribed probability laws.

Suppose we can keep track of the c server workloads in continuous time. Then we could assign a "ticket" to each arrival, which points to its target server, i.e., the one that will serve it. The target server is identified because it has the least workload at the arrival instant. This procedure distributes "theoretical" line-ups to the c servers, although there is only one physical line-up in the waiting room.

4.4.3 Indistinguishable Servers

In many models, it is not necessary to construct sample paths of $\{R_i(t)\}$ in order to construct a sample path of $W(t)$. It suffices to regard the servers as indistinguishable. Then we need not track each server workload explicitly. It is sufficient to track the virtual wait $W(t) = \min_{i=1,...,c}\{R_i(t)\}$ directly, in order to analyze the statistical properties of the model. For that purpose we utilize what we call the ***system configuration***, described in Section 4.5.

4.5 System Configuration

Suppose a "system manager" knows the target server i_t^* to be occupied at instant $t + W(t^-)$ by a would-be time-t arrival C(t). The *system configuration* at t, denoted by $\boldsymbol{M}(t)$, tracks the service rates of the *other* $c-1$ servers. We assume the model specifies $J+1$ possible exponential service rates in the set $\boldsymbol{\mu} = \{\mu_0, \mu_1, ..., \mu_J\}$ (J is a non-negative integer). Each arrival is assigned a service rate selected from the set $\boldsymbol{\mu}$. Recall that if t is not an arrival instant, sample-path right continuity implies $W(t^-) = W(t)$.

The system configuration $\boldsymbol{M}(t)$ is a $J+1$ vector of *server occupancy numbers* $m_j \geq 0$, namely

$$\boldsymbol{M}(t) = (m_0, ..., m_J).$$

Occupancy number m_j denotes the *number of servers* that will be occupied at $t + W(t^-)$ by customers having service rate $\mu_j \in \boldsymbol{\mu}$, among the $c-1$ servers other than i_t^*.

We denote the set of all possible configurations by $\boldsymbol{M}$. For each configuration $\boldsymbol{m} \in \boldsymbol{M}$, $0 \leq \sum_{j=0}^{J} m_j \leq c-1$. C$(t)$ would start service at instant $t + W(t^-)$ and would be assigned a service rate $\mu_t(x, \boldsymbol{m}) \in \boldsymbol{\mu}$, where $x = W(t^-)$, $\boldsymbol{m} = \boldsymbol{M}(t^-)$. Service rate $\mu_t(x, \boldsymbol{m})$ is a function of three variables, i.e.,

$$\mu_t(x, \boldsymbol{m}) : (t, x, \boldsymbol{m}) \to \mu_j \in \boldsymbol{\mu}, \text{ for some } j = 0, ..., J.$$

Remark 4.2 *In various models, the service rate $\mu_t(\cdot,\cdot)$ may also depend on other variables as well. It may be selected randomly from set $\boldsymbol{\mu}$. Also, the number of possible service rates may be infinite.*

4.5.1 Inter Start-of-service Departure Time

A *basic* random variable is the *inter start-of-service departure time at instant* t, denoted by $\mathcal{S}_t$. Let the state be $(W(t^-), M(t^-)) = (x, \boldsymbol{m})$ when an actual customer C_t arrives. The required wait before service is $x \geq 0$ and the configuration is $\boldsymbol{m}$. Assume $\boldsymbol{m}$ is such that $\sum_{i=0}^{J} m_j = c-1$. Thus, just after C_t starts service at $t + W(t^-)$, all c servers will be occupied.

Random variable $\mathcal{S}_t$ is the time measured from $t+x$ (start of service time) until the first departure from the system after $t+x$. In other words, $\mathcal{S}_t$ is the time from the start of service of C_t until the first departure thereafter. $\mathcal{S}_t$ is distributed as the minimum of $\sum_{i=0}^{J} m_j + 1 = c$ independent exponentially distributed r.v.'s. Among these, m_j have rate $\mu_j, j = 0, ..., J$, and one server has rate $\mu_t(x, \boldsymbol{m})$ (rate assigned to C_t). Thus $\mathcal{S}_t$ is exponentially distributed with rate $\nu_t = \sum_{j=0}^{J} m_j \mu_j + \mu_t(x, \boldsymbol{m})$.

4.5.2 Number of Configurations

Let $(W(t), M(t)) = (x, \boldsymbol{m})$. Assume that configuration $\boldsymbol{m} = (m_0, ..., m_J)$ is such that

$$\sum_{j=0}^{J} m_j = k, 0 \leq k \leq c-1.$$

The servers are considered to be indistinguishable. We track the *number of servers occupied* with service rates $\mu_j \in \boldsymbol{\mu}$, but not the identity of the servers having those service rates.

The number of possible configurations such that exactly k servers are occupied, is the number of non-negative integer solutions of the equation

$$m_0 + ... + m_J = k.$$

It is the same as the number of ways of distributing k *indistinguishable* balls in $J+1$ *distinguishable* cells, namely $\binom{J+k}{J} = \binom{J+k}{k}$ (e.g., [55], Ch. II). Thus, the total number of possible configurations is

$$\sum_{k=0}^{c-1} \binom{J+k}{J} = \binom{J+c}{J+1} = \binom{J+c}{c-1}. \tag{4.5}$$

The last equality is readily proved by induction.

Example 4.1 *Consider an M/M/3 queue with* $J = 2$. *The possible service rates are* μ_0, μ_1, μ_2. *If a potential arrival C(t) at t finds the* *system* ***empty****, the* ***one*** *possible configuration that C(t) would "see" is* $M(t^-) = (0,0,0)$*. The number of solutions of* $m_0 + m_1 + m_2 = 0$ *is* $\binom{J+0}{J} = \binom{2}{2} = 1$*. C(t) would wait* $W(t^-) = 0$ *and start service from one the three servers, per server-assignment policy.*

If C(t) would find ***one*** *customer in the system, the configuration that C(t) would see is one of* ***three*** *possible vectors*

$$M(t^-) \in \{(1,0,0),(0,1,0),(0,0,1)\}.$$

The number of servers occupied would be 1*. The number of solutions of* $m_0 + m_1 + m_2 = 1$ *is* $\binom{J+1}{J} = \binom{3}{2} = 3$*.* $C(t)$ *would wait* $W(t^-) = 0$ *and start service from one the two free servers, per server-assignment policy.*

If C(t) would find ***two*** *customers in the system, then the configuration that C(t) would see is one of* ***six*** *possible vectors*

$$M(t^-) \in \{(2,0,0),(0,2,0),(0,0,2),(1,1,0),(1,0,1),(0,1,1)\}.$$

The number of solutions of $m_0 + m_1 + m_2 = 2$ *is* $\binom{J+2}{J} = \binom{4}{2} = 6$*.* $C(t)$ *would wait* $W(t^-) = 0$ *and start service from the one free server.*

If C(t) would find ***three or more*** *customers in the system, then all* ***three*** *servers would be occupied at* t^-*. The configuration that C(t) would "see" just before start of service is also one of* ***six*** *possible vectors*

$$M(t) \in \{(2,0,0),(0,2,0),(0,0,2),(1,1,0),(1,0,1),(0,1,1)\}.$$

The six possible configurations are the same *as when C(t) sees two servers occupied. This is because a configuration tracks the service-rate occupancies of those servers* ***other than*** *C(t)'s target server. Customer* $C(t)$ *would wait a* ***positive time*** *and start service at* $t+W(t^-)$ *from some server* i_t^**. We "look ahead" to the start of service instant* $t + W(t^-)$ *and assign rate* $\mu_t(W(t^-), M(t^-))$ *to* i_t^**. The random variable* $M(t)$ *tracks the service-rate occupancies of the two servers other than* i_t^* *at* $t+W(t^-)$*. (The look-ahead idea is not new in queueing theory. For example, it is tacitly assumed for the virtual wait in the standard M/G/1 queue. In that case, we increase the virtual wait by a service time at an* arrival *instant, although the service is not started until the end of the waiting time.)*

At instant t*, the state* $(W(t), M(t))$ *specifies the position of the SP. The state conveys sufficient information to determine which active service time will be minimum among all the occupied servers at* $t + W(t^-)$*.*

In the present example, the total number of possible configurations is

$$\sum_{k=0}^{c-1}\binom{J+k}{J}=\sum_{k=0}^{2}\binom{J+k}{J}=\binom{J+c}{J+1}=\binom{5}{3}=10.$$

4.5.3 Border States

Definition 4.1 *In Example 4.1 we call a zero-wait state* $\{(0,\boldsymbol{m})\}$ *a* **border** *state if it is in the sextet*

$$\boldsymbol{m}\in\{(2,0,0),(0,2,0),(0,0,2),(1,1,0),(1,0,1),(0,1,1)\}.$$

Definition 4.2 *The set of* ***border states*** *comprises zero-wait states that form a* ***boundary*** *between other zero-wait states and positive wait states.*

In the above definition, the *other* zero-wait states are non-border states. Border states *communicate* in *one step* with the positive-wait states. When moving *from* a *non-border* zero-wait state *to* a positive-wait state, the SP must pass through a border state. In the opposite direction, *from* a positive-wait state to a non-border zero-wait state, the SP must pass through a border state. We denote the set of border states by $\boldsymbol{S}_b$; the set of border configurations by $\boldsymbol{M}_b$. Thus

$$\begin{array}{l}\boldsymbol{S}_b=\left\{(0,\boldsymbol{m})\,|\,\sum_{j=0}^{J}m_j=c-1\right\},\\ \boldsymbol{M}_b=\{\boldsymbol{m}|\boldsymbol{m}\in\boldsymbol{S}_b\}=\left\{\boldsymbol{m}|\sum_{j=0}^{J}m_j=c-1\right\}.\end{array}\quad (4.6)$$

4.5.4 The Next Configuration

Consider an actual arrival C_t at instant t. C_t "sees" configuration $M(t^-)$. Just *after* the arrival the configuration is $M(t)$. Either $M(t)=M(t^-)$ or $M(t)\neq M(t^-)$. Recall that $M(t)$ is right continuous and has left limits. We illustrate by example how to compute the probability mass function of $M(t)$.

Example 4.2 *Consider Example 4.1 for M/M/3. Then* $c=3$, $J=2$. *Suppose* C_t *arrives when the wait is* $W(t^-)$ *and the configuration is* $(m_0,m_1,m_2)=(1,1,0)$. *The state is*

$$(W(t^-),M(t^-))=(W(t^-),(1,1,0)).$$

Suppose that C_t *is assigned service rate* μ_0*, i.e.,*

$$\mu_t(W(t^-),(1,1,0)) = \mu_0.$$

At instant $t+W(t^-)$, ***just after*** C_t ***starts service****, there will be* two *servers with rate* μ_0 *since* $m_0 = 1$ *and* $\mu_t(W(t^-),(1,1,0)) = \mu_0$*. There will be* one *server with rate* μ_1*, since* $m_1 = 1$*. Then* $\mathcal{S}_t$ *will be distributed with rate* $2\mu_0 + \mu_1$*.*

We now compute the probability distribution of the ***next configuration*** *at instant* $t + W(t)$*. Thus,*

$$\begin{aligned} P(M(t) &= (2,0,0)) \\ &= P(\text{rate-}\mu_1 \text{ server finishes first}) \\ &= \frac{\mu_1}{2\mu_0+\mu_1}, \end{aligned}$$

$$\begin{aligned} P(M(t) &= (1,1,0)) \\ &= P(\text{rate-}\mu_0 \text{ server finishes first}) \\ &= \frac{2\mu_0}{2\mu_0+\mu_1}. \end{aligned}$$

Note that

$$P(M(t) = (2,0,0)) + P(M(t) = (1,1,0)) = 1.$$

The only two possible configurations for $M(t)$ *are* $(2,0,0)$ *and* $(1,1,0)$*, independent of whether* $W(t^-) = 0$ *or* $W(t^-) > 0$*. No other configuration is possible for* $M(t)$ *once the arrival at* t^- *has been assigned rate* μ_0*.*

Remark 4.3 *The service mechanism can be generalized considerably. We can expand the domain of* $\mu_t(w,\boldsymbol{m})$ *to include: type or priority class of* C_t*; type of customer replaced by* C_t *in server* i_t^**; type of any customer followed by* C_t *into service; identity of server* i_t^* *(e.g., server number or unique property); number of customers in the system or waiting for service at the arrival or start of service instant of* C_t*; various types of bounds on the virtual wait; reneging indices; blocked and cleared customers, etc.*

Other generalizations may incorporate: a non-homogeneous Poisson arrival process with intensity λ_t*, or a Poisson arrival rate* $\lambda(W(t),M(t))$ *which depends on the current state* $(W(t),M(t))$*; or various Markov arrival processes.*

4.6 System Point Process

We now discuss the ***system point process*** and the geometry of its state space.

We call $\{W(t), M(t)\}, t \geq 0$ the *system point (SP) process.* Its nomenclature derives from the fact that the SP traces out a sample path as the system evolves over time. The SP process for M/M/c queues is a generalization of the virtual wait process for M/G/1 queues (with exponential service times). State variable $W(t)$ represents the virtual wait. State variable $M(t)$ represents the system configuration. Random variable $M(t)$ is discrete. The SP process is a Markov process (discussed in Subsection 4.6.7).

We partition the state space $\boldsymbol{S}$ into three disjoint state-space sets $\boldsymbol{S}_0$, $\boldsymbol{S}_b$, $\boldsymbol{S}_1$. The states in $\boldsymbol{S}_0 \cup \boldsymbol{S}_b$ are atoms. The states in $\boldsymbol{S}_1$ are points in a continuum. That is,

$$\boldsymbol{S}_0 = \{(0, \boldsymbol{m})|0 \leq \sum_{j=0}^{J} m_j \leq c-2\},$$

$$\boldsymbol{S}_b = \{(0, \boldsymbol{m})|\sum_{j=0}^{J} m_j = c-1\}$$

$$\boldsymbol{S}_1 = \{(x, \boldsymbol{m})|x \in (0,\infty), \sum_{j=0}^{J} m_j = c-1\}.$$

Note that $\boldsymbol{S} = \boldsymbol{S}_0 \cup \boldsymbol{S}_b \cup \boldsymbol{S}_1$, and

$$\boldsymbol{S}_0 \cap \boldsymbol{S}_b = \boldsymbol{S}_0 \cap \boldsymbol{S}_1 = \boldsymbol{S}_b \cap \boldsymbol{S}_1 = \phi,$$

where ϕ is the empty set. Let

$$\boldsymbol{M}_0 = \{\boldsymbol{m}|(0, \boldsymbol{m}) \in \boldsymbol{S}_0) = \{\boldsymbol{m}|0 \leq \sum_{j=0}^{J} m_j \leq c-2\}.$$

Recall that $\boldsymbol{M}_b = \{\boldsymbol{m}|\sum_{j=0}^{J} m_j = c-1\}$. Let

$$\boldsymbol{M}_1 = \{m|(x, \boldsymbol{m}) \in \boldsymbol{S}_1\} = \{(x, \boldsymbol{m})|x \in (0,\infty), \sum_{j=0}^{J} m_j = c-1\}.$$

Thus $\boldsymbol{M}_b = \boldsymbol{M}_1$.

An arrival C(t) would "see" a state in $\boldsymbol{S}_0 \cup \boldsymbol{S}_b$ if and only if $W(t^-) = 0$. C(t) would wait zero and start service from some server i_t^*, at time t. Geometrically, we associate a distinct horizontal line $\boldsymbol{T} \times (0, \boldsymbol{m})$ with each state $(0, \boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b$. $\boldsymbol{T}$ is the time axis $[0, \infty)$. We call the line $\boldsymbol{T} \times (0, \boldsymbol{m})$ "*line* $\boldsymbol{m}$".

C(t) would "see" a state in $\boldsymbol{S}_1$ if and only if $W(t^-) > 0$. C(t) would wait a positive time $W(t^-)$ and start service from a target server i_t^*, at time $t + W(t^-)$. Geometrically, we associate the quadrant of a plane, $\boldsymbol{T} \times (0, \infty)$, with each set of continuous states $(x, \boldsymbol{m}) \in \boldsymbol{S}_1$. We call the positive quadrant $\boldsymbol{T} \times ((0, \infty), \boldsymbol{m})$ ***sheet*** $\boldsymbol{m}$ or ***page*** $\boldsymbol{m}$.

We imagine a plot of $W(t)$ versus t on page $\boldsymbol{m}$ while the system is in configuration $\boldsymbol{m}$.

We call the states in $\boldsymbol{S}_b$ "*border*" states. Geometrically, we may locate the "border" lines

$$\boldsymbol{T} \times (0, \boldsymbol{m}), (0, \boldsymbol{m}) \in \boldsymbol{S}_b,$$

alongside the lines for states $(0, \boldsymbol{m}) \in \boldsymbol{S}_0$, or at zero level of the corresponding sheets. There is a one-to-one correspondence between sheets and states in $\boldsymbol{S}_b$.

4.6.1 Sample Path of SP Process

A sample path of $\{W(t), M(t)\}$ is a piecewise right continuous function of t having left limits. It has a finite number of jumps during finite time intervals (see Section 2.2 and Definition 2.1). We plot a sample path within in a Cartesian product space $\boldsymbol{T} \times \boldsymbol{S} = \boldsymbol{T} \times (\boldsymbol{S_0} \cup \boldsymbol{S_b} \cup \boldsymbol{S_1})$. The direction of time is taken to be from left to right. It is useful to envisage each Cartesian product

$$\boldsymbol{T} \times (0, \boldsymbol{m}), (0, \boldsymbol{m}) \in \boldsymbol{S_0} \cup \boldsymbol{S_b}$$

as a ***line***; and each quadrant $\boldsymbol{T} \times ((0, \infty), \boldsymbol{m}), \boldsymbol{m} \subseteq \boldsymbol{M}_1$ as a sheet, or page in a ***book***.

Description of a Sample Path

Assume the system starts empty (see Fig. 4.1). The SP moves continuously among the zero-wait lines. It jumps from line to line at arrival and departure instants. Eventually the SP jumps from a *boundary* line "$\boldsymbol{m}$",$\boldsymbol{m} \in \boldsymbol{M}_b$, to some sheet "$\boldsymbol{k}$", at an arrival instant. It then moves with slope -1 on sheet $\boldsymbol{k}$. Either $\boldsymbol{m} = \boldsymbol{k}$, or $\boldsymbol{m} \neq \boldsymbol{k}$, depending on the probabilities governing the motion.

At an arrival instant while the SP is on sheet $\boldsymbol{k}$, the SP may jump to another sheet, say $\boldsymbol{m}'$, and move with slope -1 on sheet $\boldsymbol{m}'$ for a positive time. Otherwise the SP may jump, and stay on the same sheet $\boldsymbol{k}$. On each sheet it moves downward with slope -1. If the SP hits level 0 from above on page $\boldsymbol{k}$ before the next arrival (no customers waiting), it starts moving immediately on the border line $\boldsymbol{k}$.

If the SP is in a state of $\boldsymbol{S}_b \cup \boldsymbol{S}_1$ having configuration $\boldsymbol{m}$ at an arrival instant, it makes a jump ending either on page $\boldsymbol{m}$ or on some page $\boldsymbol{k} \neq \boldsymbol{m}$. That is, the SP can make an $\boldsymbol{m} \rightarrow \boldsymbol{k}$ *transition.* This may be an upward jump from a border line $\boldsymbol{m}$, or from sheet $\boldsymbol{m}$ to sheet $\boldsymbol{k}$ at an arrival instant. Generally, $\boldsymbol{m} \rightarrow \boldsymbol{k}$ transitions do not give rise to"typical" level crossings as in M/G/1 models that have one "page". However, $\boldsymbol{m} \rightarrow \boldsymbol{m}$ transitions from a border line $\boldsymbol{m}$ or from a point on sheet $\boldsymbol{m}$ to a higher point on sheet $\boldsymbol{m}$, are similar to SP jumps as discussed for models with a single sheet (Section 2.3).

Remark 4.4 *In some model variants, an* $\mathbf{m} \rightarrow \mathbf{k}$ *transition may be a* ***parallel*** *jump. That is, the SP jumps from a level* y *on page* $\mathbf{m}$ *to level* y *on page* $\mathbf{k} \neq \mathbf{m}$*, at an arrival instant. For example, in an M/G/1 queue, we may utilize a modified configuration* $M(t) = n$*, where* n *is the number of customers waiting for service. In some models the virtual wait may be unchanged at an arrival instant. Such parallel jumps occur in M/G/1 or M/M/c queues with bulk service [71].*

4.6.2 Metaphor for Sample Path and SP Motion

We can make a metaphor for the SP motion. It is like the motion of the tip of a pen writing out a *one-page-long* history of the system. The writing takes place in a book of transparent pages all of the same size. The cover is also transparent. The pen moves from left to right, and never overlays what has been written already. After writing lines on a page for a random amount of time, the pen jumps to a different page, and continues writing. The pen jumps in this manner at random time points from page to page. The next page is selected at random depending on where it is presently. The *entire* history up to an instant in time can be seen only by holding all the transparent pages one behind the other, like pages in a book, and viewing the projected history on the ***cover***. The projected history on the cover is invariant to shuffling of the pages. An analyst that views an arbitrary page in isolation, sees only local segments of the history specific to that page.

The global history is like the "total" sample path of the SP process over the state space $\boldsymbol{S_0} \cup \boldsymbol{S_b} \cup \boldsymbol{S_1}$. The local histories on various pages are like sample-path segments due to sojourns on the "lines" and "sheets" of the state space. SP motion on the lines occurs at level 0. When all the lines are projected onto the cover, they are placed at level zero – to form a single zero line.

We may think of the overall method as having several steps.

1. Partition the **Time-State space** into mutually exclusive and exhaustive lines and sheets.
2. Analyze the sample-path segments on the lines and sheets using LC.
3. Project the sample-path segments from the lines and sheets onto the "cover" of the "book". Analyze the projected path on the cover using LC.
4. Combine all the LC results with a normalizing condition. Construct the model equations and derive probability distributions of the model.

The LC method utilizes statistical properties of the local path segments on the lines and sheets. It also uses statistical properties of the projected path on the cover. It employs the one-step communication properties among the lines and sheets to construct a sample path. Basic LC theorems apply to each page $\boldsymbol{m} \in \boldsymbol{M}$. Jumps out of, and into lines and sheets, follow rate-conservation laws.

Equations

We use sample-path structure and transition rates in and out of state-space sets, to construct (by inspection) integro-differential and differential equations in a transient analysis. Similarly, we construct integral equations and algebraic equations in a steady-state analysis. These are equations for the joint pdf and/or cdf of *wait* and *configuration*. We can also derive equations for the marginal (total) pdf and cdf of wait, or for the probabilities of the system configurations.

Remark 4.5 *I originally had the idea for partitioning the state space, visualizing the positive-wait states over time as separate quadrants, and having a "system point" move on them, from an analogy with Riemann sheets and winding numbers in complex variable theory. My PhD thesis* *used the term "sheets". The term "pages" was introduced later. I also thought of using the term "cards", analogous to boxes of computer cards for data and programs in use in 1974. Then, the state space could be pictured like a box or deck of rectangular cards. Such cards had been ubiquitous until personal computers became common in the 1980's.*

4.6.3 Notation: Probabilities and Distributions

Transient Probabilities and Distributions

We denote the zero-wait probabilities by

$$P_t(0, \boldsymbol{m}) = P(W(t) = 0, \boldsymbol{M}(t) = \boldsymbol{m}), (0, \boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b.$$

We denote the mixed joint cdf of $(W(t), M(t))$ by

$$\begin{aligned} F_t(x, \boldsymbol{m}) &= P(W(t) \le x, \boldsymbol{M}(t) = \boldsymbol{m}) \\ &= P_t(0, \boldsymbol{m}) + P(0 < W(t) \le x, \boldsymbol{M}(t) = \boldsymbol{m}) \\ &= P_t(0, \boldsymbol{m}) + \int_{y=0}^{x} f_t(y, \boldsymbol{m}) dy,\ x \ge 0, t \ge 0, \\ &\qquad\qquad (0, \boldsymbol{m}) \in \boldsymbol{S}_b, (x, \boldsymbol{m}) \in \boldsymbol{S}_1, \end{aligned}$$

where $P(0 < W(t) \le x, \boldsymbol{M}(t) = \boldsymbol{m}) = P(\phi) = 0$ if $x = 0$.

The mixed joint pdf of $(W(t), M(t))$ is

$$f_t(x, m) = \frac{\partial}{\partial x} F_t(x, \boldsymbol{m}), x > 0, t \ge 0, (x, \boldsymbol{m}) \in \boldsymbol{S}_1,$$

wherever $\frac{\partial}{\partial x} F_t(x, \boldsymbol{m})$ exists.

We *assume*:

1. $F_t(x, \boldsymbol{m})$ and $f_t(x, \boldsymbol{m})$ are right continuous in x for every $t \ge 0, \boldsymbol{m} \in \boldsymbol{M}_1$
2. $\frac{\partial}{\partial t} F_t(x, \boldsymbol{m})$ and $\frac{\partial}{\partial t} f_t(x, \boldsymbol{m}), t > 0, x \ge 0$ exist and are finite for every $\boldsymbol{m} \in \boldsymbol{M}_1$.

Let $P_0(t) = P(W(t) = 0)$ be the marginal probability of a zero wait at t. Then

$$\begin{aligned} P_0(t) &= \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_0 \cup \boldsymbol{S}_b} P_t(0, \boldsymbol{m}) \\ &= \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_0} P_t(0, \boldsymbol{m}) + \sum_{(0,\boldsymbol{m}) \in \boldsymbol{S}_b} P_t(0, \boldsymbol{m}), t \ge 0. \end{aligned}$$

The transient marginal cdf of wait $P(W(t) \leq x)$ is

$$\begin{aligned} F_t(x) &= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0} P_t(0,\boldsymbol{m}) + \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_b} F_t(x,\boldsymbol{m}) \\ &= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0\cup \boldsymbol{S}_b} P_t(0,\boldsymbol{m}) + P(0 < W(t) \leq x) \\ &= P_0(t) + P(0 < W(t) \leq x) \\ &= P_0(t) + \int_{y=0}^{x} f_t(y)dy, \ \ x \geq 0, t \geq 0. \end{aligned}$$

Note that $P_t(0,\boldsymbol{m}) = F_t(0,\boldsymbol{m})$ for $(0,\boldsymbol{m}) \in \boldsymbol{S}_b$.

(Recall the definitions of $\boldsymbol{M}_b$ and $\boldsymbol{S}_b$ in (4.6), and $\boldsymbol{M}_b = \boldsymbol{M}_1$, which is the set of configurations for positive-wait states.)

The transient marginal pdf of $W(t)$ is

$$f_t(x) = \frac{\partial}{\partial x} F_t(x) = \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} f_t(x,\boldsymbol{m}), x > 0, t \geq 0.$$

A potential arrival $\mathrm{C}(t)$ would find the system configuration to be $\boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$ with probability $P_t(0,\boldsymbol{m})$. $\mathrm{C}(t)$ would find the configuration to be $\boldsymbol{m} \in \boldsymbol{M}_1$ with probability $F_t(\infty,\boldsymbol{m})$.The normalizing condition for fixed $t \geq 0$, is

$$\begin{aligned} F_t(\infty) &= \sum_{\boldsymbol{m}\in \boldsymbol{M}_0} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} F_t(\infty,\boldsymbol{m}) \\ &= \sum_{\boldsymbol{m}\in \boldsymbol{M}_0\cup \boldsymbol{M}_b} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} \int_{y=0}^{\infty} f_t(y,\boldsymbol{m})dy \\ &= \sum_{(0,\boldsymbol{m})\in \boldsymbol{S}_0\cup \boldsymbol{S}_b} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} \int_{y=0}^{\infty} f_t(y,\boldsymbol{m})dy = 1. \end{aligned}$$

Steady-State Probabilities and Distributions

We denote the steady-state zero-wait probabilities, pdf's and cdf's of wait by dropping the subscript t in the foregoing definitions for the transient quantities.

As in Notation 3.7, we use symbol E_a to denote an exponentially distributed random variable with rate $a > 0$ (mean $\frac{1}{a}$).

4.6.4 Configuration Just After an Arrival

Example 4.3 demonstrates the probability of a system configuration just after an arrival. Assume that an actual customer C_t arrives and finds the state to be $(W(t^-), \boldsymbol{M}(t^-)) = (x, \boldsymbol{m})$. The service rate assigned to C_t is $\mu_t(x, \boldsymbol{m}) \in \boldsymbol{\mu}$. Recall that sample paths are right continuous and have left limits.

Example 4.3 *Consider Example 4.2, where $c = 3$, $J = 2$. Let each arrival receive a service rate selected with equal probability from the set $\boldsymbol{\mu} = \{\mu_0, \mu_1, \mu_2\}$. Then*

$$\begin{aligned} &P(C_t \textit{ starts service at } t + W(t^-) \textit{ with service rate } \mu_i) \\ &\equiv \frac{1}{3}, i = 0, 1, 2, \end{aligned}$$

independent of t and $W(t^-)$. Assume $(W(t^-), M(t^-)) = (x, (2, 0, 0))$ just before *C_t arrives. Then C_t will wait time x, and two other service rates are equal to μ_0 when C_t starts service at $t + W(t^-)$. What is the configuration $M(t)$* just after *C_t arrives? It can be either $(2, 0, 0)$, $(1, 1, 0)$, or $(1, 0, 1)$. The probabilities for $M(t)$ are:*

$$\begin{aligned} P(M(t) = (2,0,0)) \quad &= P(\mu_t(x, (2,0,0)) = \mu_0) \cdot 1 \\ &\quad + P(\mu_t(x, (2,0,0)) = \mu_1) \cdot \tfrac{\mu_1}{2\mu_0 + \mu_1} \\ &\quad + P(\mu_t(x, (2,0,0)) = \mu_2) \tfrac{\mu_2}{2\mu_0 + \mu_2} \\ &= \tfrac{1}{3}\left(1 + \tfrac{\mu_1}{2\mu_0 + \mu_1} + \tfrac{\mu_2}{2\mu_0 + \mu_2}\right). \\ P(M(t) = (1,1,0)) \quad &= P(\mu_t(x, (2,0,0)) = \mu_1) \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_1} \\ &= \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_1}. \\ P(M(t) = (1,0,1)) \quad &= P(\mu_t(x, (2,0,0)) = \mu_2) \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_2} \\ &= \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_2}. \end{aligned}$$

Note that

$$\begin{aligned} &P(M(t) = (2,0,0)) + P(M(t) = (1,1,0)) + P(M(t) = (1,0,1)) \\ &= \tfrac{1}{3}\left(1 + \tfrac{\mu_1}{2\mu_0 + \mu_1} + \tfrac{\mu_2}{2\mu_0 + \mu_2}\right) + \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_1} + \tfrac{1}{3} \cdot \tfrac{2\mu_0}{2\mu_0 + \mu_2} = 1. \end{aligned}$$

Also the virtual wait at time t is

$$W(t) = W(t^-) + \mathcal{S}_t = x + \mathcal{S}_t,$$

where $\mathcal{S}_t$ is distributed as either $E_{3\mu_0}$, $E_{2\mu_0+\mu_1}$, or $E_{2\mu_0+\mu_2}$ (with probability $\frac{1}{3}$ each). The sample path will have a jump whose size is distributed as $\mathcal{S}_t$ at instant t.

4.6.5 Sample Path of SP Process Revisited

We first discuss the nature of a sample path in Example 4.4. Then we discuss a *specific* sample path. Consider the M/M/3 model in Example 4.3. We set $J = 1$ for exposition. Thus $\boldsymbol{\mu} = \{\mu_0, \mu_1\}$. If $J > 1$, sample-path construction would be similar. However there would be more lines and sheets in the product space $\boldsymbol{T} \times \boldsymbol{S}$.

Example 4.4 *Consider M/M/c with $c = 3$, $J = 1$ (see Fig. 4.1). Arrivals are assigned an exponential service rate from $\boldsymbol{\mu}$ $=\{\mu_0, \mu_1\}$ with equal probability $\frac{1}{2}$. (The probabilities could be general, e.g., p_0, $p_1 = 1 - p_0$.) The total number of possible configurations is $\binom{J+c}{J+1} = \binom{4}{2} = 6$. The full set of configurations is*

$$\boldsymbol{M} = \{00, 10, 01, 11, 20, 02\}.$$

We write $(2, 0)$ as 20 when $m_0 = 2$, $m_1 = 0$, indicating that two servers are occupied with rate μ_0; and similar notation for the other configurations.

The state space consists of: (1) six discrete points for the zero-wait states $(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M} = \boldsymbol{M}_0 \cup \boldsymbol{M}_b$, where $\boldsymbol{M}_0 = \{00, 10, 01\}$ and $\boldsymbol{M}_b = \boldsymbol{M}_1 = \{11, 20, 02\}$; (2) three intervals $((0, \infty), \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$. The three "border" states are $(0; \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_b$.

Arrival Waits Zero: *Assume an arrival "sees" state $(0, m), m \in \boldsymbol{M}_0$. The SP moves horizontally at time-rate 1 on a line $\boldsymbol{T} \times (0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0$. If the next arrival occurs before a departure, the SP jumps to a line $\boldsymbol{T} \times (0, \boldsymbol{m}'), \boldsymbol{m}' \in \boldsymbol{M}_0 \cup \boldsymbol{M}_b$, where*

$$m_0' + m_1' = m_0 + m_1 + 1.$$

If a departure occurs before an arrival, the SP jumps to a line $\boldsymbol{T} \times (0, \boldsymbol{m}''), \boldsymbol{m}'' \in \boldsymbol{M}_0$, where

$$m_0'' + m_1'' = m_0 + m_1 - 1.$$

If $\boldsymbol{m} = (0, 0)$, only an arrival is possible.

If an arrival finds the state to be $(0, m), m \in \boldsymbol{M}_b$ the SP jumps to a "sheet" $\boldsymbol{T} \times ((0, \infty), \boldsymbol{k}), \boldsymbol{k} \in \boldsymbol{M}_1$. (Recall that $\boldsymbol{M}_b = \boldsymbol{M}_1$.) Configuration

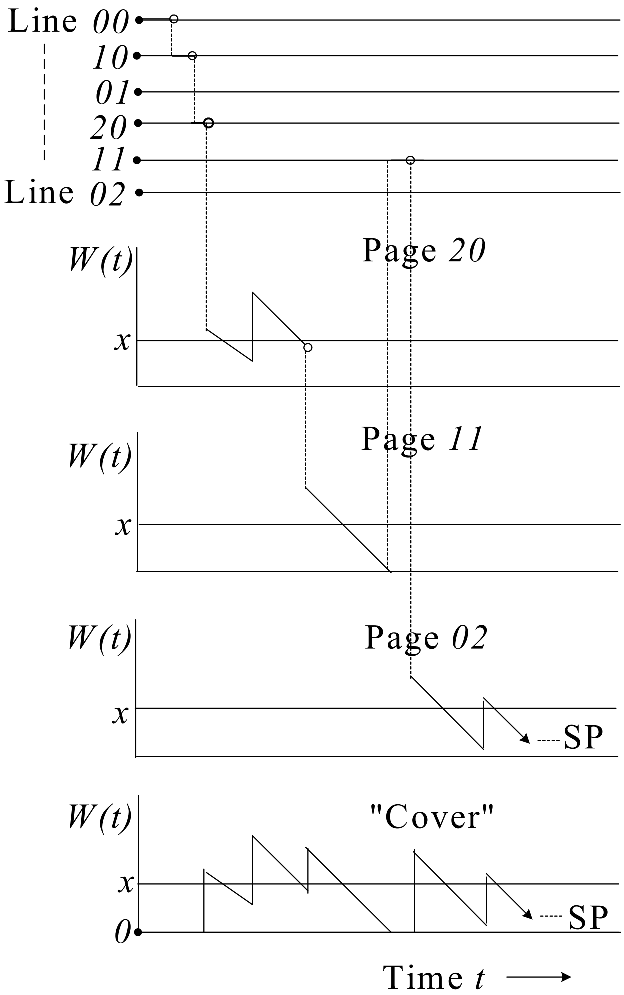

Figure 4.1: Sample path of SP process $\{W(t), M(t)\}$ for M/M/c example with $c = 3$, $J = 1$, and random assignment of service rates independent of t. The space $\boldsymbol{T} \times \boldsymbol{S}$ has six lines for zero-wait states, and three sheets for positive-wait states (pages 20, 11, 02) . The "cover" is the projection of the sample path from all lines and pages onto one non-negative planar quadrant.

k *is determined by the service rate assigned to the new arrival, and which server finishes first after the new arrival starts service. Denote the service time of an arrival* C_t *by* s_t*. Note that*

$$P(s_t \text{ has rate } \mu_i) = \frac{1}{2}, i = 0, 1$$

(see Fig. 4.1).

To fix ideas, let the SP be on "border" line $\boldsymbol{T} \times (0, 20)$ *at arrival instant* t*. Thus* $W(t^-) = 0$*,* $M(t^-) = 20$*. Then* C_t *starts service upon arrival in the single idle server and is assigned either rate* μ_0 *or* μ_1*. Let* $\mathcal{S}_t$ *denote the time from the start-of-service instant of* C_t *until the first departure from the system thereafter.*

Assume the service time s_t *has been assigned rate* μ_0*. Then*

$$\mathcal{S}_t \underset{dist}{=} E_{3\mu_0},$$

since $\mathcal{S}_t = \min\left\{3 \text{ iid } E_{\mu_0} \text{ r.v.'s}\right\}$*. (Recall that* $E_a \underset{dist}{=}$ *exponentially distributed r.v. with rate a.) The SP jumps upward an amount* $\mathcal{S}_t$*. The virtual wait at* t *is*

$$W(t) = W(t^-) + \mathcal{S}_t = \mathcal{S}_t.$$

At instant

$$t + W(t^-) + \mathcal{S}_t = t + \mathcal{S}_t$$

one of the three occupied servers completes service and becomes idle. The service rate of each of the remaining two occupied servers at $t + \mathcal{S}_t$ *must be* μ_0*. Thus, the configuration at* t *is* $M(t) = M(t^-) = 20$*. The configuration remains the same. Geometrically, at instant* t*, the SP jumps from* line 20 *to* page 20 *and enters page* 20 *at a height* $E_{3\mu_0}$ *(see Fig. 4.1).*

On the other hand, suppose s_t *has been assigned rate* μ_1*. Then* $\mathcal{S}_t \underset{dist}{=} E_{2\mu_0+\mu_1}$*. At* $t + \mathcal{S}_t$ *one of the three servers completes service and becomes idle. The service rates of the remaining two occupied servers at* $t + \mathcal{S}_t$ *are either: (1) both* μ_0 *with probability* $\frac{\mu_1}{2\mu_0+\mu_1}$ *(rate-*μ_1 *server finishes first), or (2)* μ_0 *and* μ_1 *with probability* $\frac{2\mu_0}{2\mu_0+\mu_1}$ *(a rate-*μ_0 *server finishes first).*

In case (1), at instant t *the SP jumps from line* 20 *to page* 20 *at a height* $E_{2\mu_0+\mu_1}$*. Thus* $W(t) \underset{dist}{=} E_{2\mu_0+\mu_1}$ *and* $M(t) = 20$*. The SP* ***height*** *on page* 20 *is distributed differently from when* $s_t \underset{dist}{=} E_{\mu_0}$*.*

In case(2) at instant t, the SP jumps from line 20 to page 11 at a height $E_{2\mu_0+\mu_1}$. Thus $W(t) \underset{dist}{=} E_{2\mu_0+\mu_1}$ and $M(t) = 11$.

Arrival Waits a Positive Time: *Suppose C_t arrives when the state is $(x, 20), x > 0$ (SP is at height x on page 20). If the service-rate assignment policy assigns $s_t \underset{dist}{=} E_{\mu_0}$, the SP jumps upward an amount $E_{3\mu_0}$, and remains and moves downward on page 20. If the service-rate assignment policy assigns $s_t \underset{dist}{=} E_{\mu_1}$, the SP can end up on either sheet 20 or sheet 11 at t. The SP jumps upward to $W(t) \underset{dist}{=} W(t^-)+E_{2\mu_0+\mu_1}$ and remains on page 20, with probability $\frac{\mu_1}{2\mu_0+\mu_1}$. The SP jumps upward to $W(t) \underset{dist}{=} W(t^-)+E_{2\mu_0+\mu_1}$ and simultaneously moves to the same height on page 11, with probability $\frac{2\mu_0}{2\mu_0+\mu_1}$.*

If the SP descends to the bottom of page 20 and hits level 0 from above in a continuous manner before a new arrival occurs, it immediately enters border line 20, and continues its motion along line 20.

4.6.6 Specific Sample Path

Consider a possible realization of the SP motion as it traces out the sample path depicted in Fig. 4.1. Assume that initially the system is empty. The SP moves on line 00. Arrival 1 (C_1) sees an empty system. The server-assignment policy assigns C_1 service rate μ_0. The SP jumps to, and moves on, line 10. C_2 arrives before C_1 completes service and is also assigned rate μ_0. At C_2's arrival the SP jumps to line 20. C_3 arrives while both C_1 and C_2 are in service. C_3 receives rate μ_1. The SP jumps to a height $E_{2\mu_0+\mu_1}$. Assume that the rate-μ_1 customer will finish first among the three customers in service. The resulting configuration is again 20. The probability of this event is $\frac{\mu_1}{2\mu_0+\mu_1}$, due to the memoryless property of exponential variates. This explains why at C_3's arrival the SP jumps to page 20.

Just before C_4 arrives the SP is descending at rate 1 on page 20. C_4 is assigned service rate μ_0. The SP jumps upward an amount $E_{3\mu_0}$. It remains on page 20. That is, whichever server finishes first, the two remaining active service rates will be μ_0, resulting in configuration 20. C_5 arrives when the SP is on page 20. C_5 is assigned rate μ_1. Suppose a server with rate μ_0 finishes first. The probability of this event is $\frac{2\mu_0}{2\mu_0+\mu_1}$. The SP jumps upward by $E_{2\mu_0+\mu_1}$. It simultaneously jumps from page 20 to page 11 at the increased height, since the two remaining busy servers have rates μ_0 and μ_1 when the first service ends (the SP makes a $20 \rightarrow 11$

transition). The configuration changes immediately from 20 to 11.

No new arrivals occur prior to the completion of the first rate-μ_0 customer. The SP descends on page 11 with slope -1 and hits level 0 from above, precisely when the first rate-μ_0 customer finishes service. The system now presents a zero wait to a potential arrival. When the SP hits level 0, it enters border line 11(in Fig. 4 it jumps to line 11). C_6 arrives, and starts service immediately. C_6 is assigned rate μ_1. The SP jumps to page 02, with probability $\frac{\mu_0}{\mu_0+2\mu_1}$ (μ_0-rate service finishes first). The configuration changes immediately from 11 to 02.

The system continues to evolve. The SP continues to trace a sample path on the lines and pages according to the probability laws of the model. The sample path gives us a precise picture of the evolving system over time. Construction of the path goes hand in hand with understanding the model dynamics.

Remark 4.6 *In Section 4.8 below we develop the* ***steady-state*** *theory. We will then return to Example 4.3. We will formulate the balance equations for the zero-wait probabilities* $P(0, \boldsymbol{m})$, $\boldsymbol{m} \in \boldsymbol{M} \equiv \boldsymbol{M}_0 \cup \boldsymbol{M}_b$*; integral equations for the "partial" pdf's of wait* $f(x, m), x > 0, \boldsymbol{m} \in \boldsymbol{M}_1$*; and for the total pdf* $\{P_0, f(x), x > 0\}$.

4.6.7 SP Process Is Markovian

We outline a proof that the SP process is a Markov process. Let $(x, \boldsymbol{m})_t$ denote the event $\{(W(t), M(t)) = (x, \boldsymbol{m})\}$. It is required to show that for $x, y \geq 0, \boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}$

$$\begin{aligned} &P((y, \boldsymbol{k})_{t+h} | (x, \boldsymbol{m})_t, (W(u), M(u)), 0 \leq u < t \\ &= P((y, \boldsymbol{k})_{t+h} \mid (x, \boldsymbol{m})_t), t \geq 0, h > 0. \end{aligned} \tag{4.7}$$

Formula (4.7) states that the probability of event $(y, \boldsymbol{k})_{t+h}$ given that event $(x, \boldsymbol{m})_t$ occurred, is independent of the history $(W(u), M(u)), 0 \leq u < t$. We sketch the proof in two steps: (1) **zero-wait states**; (2) **positive-wait states**.

Recall that for a Poisson (or non-homogeneous Poisson) arrival process, the probability of more than one event occurring in $(t, t + h)$ is $o(h)$.

Zero-wait States – Non-border

Assume state $(0, \boldsymbol{m})_t \in \left\{(0, \boldsymbol{m}) | 0 \leq \sum_{j=0}^{J} m_j \leq c - 2\right\}$ ($\boldsymbol{m} \in \boldsymbol{M}_0$, SP $\in \boldsymbol{S}_0$ at t).

No Departure or Arrival in $(t, t+h)$ The state remains $(0, \boldsymbol{m})$ in $(t, t+h)$ iff no arrival or departure occurs during $(t, t+h)$, or an event with probability $o(h)$ occurs. Thus

$$P((x, \boldsymbol{m})_{t+h} | (x, \boldsymbol{m})_t) = 1 - \left(\lambda + \sum_{j=0}^{J} m_j \mu_j \right) h + o(h),$$

which is independent of $(W(u), M(u)), 0 \leq u < t$.

Arrival in $(t, t+h)$ Possibly there is an arrival during $(t, t+h)$. The next configuration will have the form

$$\boldsymbol{m}_{L^+} = (m_0, ..., m_L + 1, ..., m_J)$$

for some $L \in \{0, ..., J\}$. Then

$$\begin{aligned} & P((0, \boldsymbol{m}_{L^+})_{t+h} | (0, \boldsymbol{m})_t) \\ & = (\lambda h + o(h)) \cdot P(\mu_t((0, \boldsymbol{m})) = \mu_L) \qquad (4.8) \\ & = \lambda h P(\mu_t((0, \boldsymbol{m})) = \mu_L) + o(h), L = 0, ..., J \cdot \end{aligned}$$

Formula (4.8) is the probability that there is an arrival during $(t, t+h)$, that is assigned service rate μ_L. That probability is independent of $(W(u), M(u)), 0 \leq u < t$. Note that

$$\sum_{L=0}^{J} P(\mu_t((0, \boldsymbol{m}) = \mu_L) = 1.$$

Departure in $(t, t+h)$ Possibly there is a departure during $(t, t+h)$. Let configuration

$$\boldsymbol{m}_{_{L^-}} = (m_0, ..., \theta_L \cdot (m_L - 1), ..., m_J), L \in \{0, ..., J\},$$

where

$$\theta_L = \begin{cases} 1 \text{ if } m_L \geq 1, \\ 0 \text{ if } m_L = 0. \end{cases}$$

Assume $\boldsymbol{m} \neq (0, ..., 0)$. Then

$$P((0, \boldsymbol{m}_{L-})_{t+h} | (0, \boldsymbol{m})_t) = (m_L \cdot \mu_L) h + o(h), \qquad (4.9)$$

which is the probability of a rate-μ_L departure during $(t, t+h)$ (*rate-μ_L service finishes first).* Expression (4.9) is independent of the history $(W(u), M(u))$,$0 \leq u < t$. Note that $\left(\sum_{L=0}^{J} m_L \mu_L\right) h + o(h)$ is the probability of a departure during $(t, t+h)$.

Zero-wait States – Border

Consider zero-wait border states $\left\{(0,\boldsymbol{m})_t | \sum_{j=0}^{J} m_j = c-1\right\}$ ($\boldsymbol{m} \in \boldsymbol{M}_b$, $(0,\boldsymbol{m}) \in \boldsymbol{S}_b$).

No Arrival in $(t,t+h)$ If no arrival or departure occur, or only a departure occurs, during $(t,t+h)$, the Markov property follows similarly as for the zero-wait non-border states given above.

Arrival in $(t,t+h)$ Possibly there is an arrival during $(t,t+h)$. In *this* case, the SP jumps to a *positive level* on a sheet (page). Let configuration

$$\begin{aligned}\boldsymbol{k} &= (m_0,...,m_L+1,...,m_R-1,...,m_J) \\ &= (k_0,...,k_J),\end{aligned}$$

for some L,R. Thus $\sum_{j=0}^{J} k_j = \sum_{j=0}^{J} m_j = c-1$. Let

$$\nu_L = \sum_{j=0}^{J} m_j \mu_j + \mu_L.$$

The probability that the SP jumps to sheet $\boldsymbol{k}$ during $(t,t+h)$ and is in state-space interval $((y,y+dy),\boldsymbol{k})\,, y>0$ at $t+h$, is

$$\begin{aligned}&P((W(t+h),M(t+h)) \in ((y,y+dy),\boldsymbol{k})|(0,\boldsymbol{m})_t) \\ &= (\lambda h + o(h)) \cdot P(\mu_t(0,\boldsymbol{m}) = \mu_L) \cdot \frac{m_R\mu_R}{\nu_L} \cdot \nu_L \cdot e^{-\nu_L \cdot y} dy \\ &= \lambda h \cdot P(\mu_t(0,\boldsymbol{m}) = \mu_L) \cdot m_R\mu_R \cdot e^{-\nu_L y} dy + o(h), L = 0,...,J,\end{aligned}$$

which is independent of the history $(W(u),M(u)), 0 \le u < t$. The right side is the probability that there is an arrival, it is assigned service rate μ_L, and a rate-μ_R *service* finishes first, at a time in $(y,y+dy)$.

Positive-wait States

Arrival In $(t,t+h)$ Given $(x,\boldsymbol{m})_t, x>0$, where $\sum_{j=0}^{J} m_j = c-1$, there may an arrival during $(t,t+h)$. Let

$$\boldsymbol{k} = (m_0,...,m_L+1,...,m_R-1,...,m_J).$$

Reasoning as for zero-wait border states, we obtain

$$P((W(t+h),M(t+h)\in((x+y,x+y+dy),\boldsymbol{k})|(x,\boldsymbol{m})_t))$$
$$=\lambda h\cdot P(\mu_t(0,\boldsymbol{m})=\mu_L)\cdot m_R\mu_R\cdot e^{-\nu_L(y-x)}dy+o(h),$$

which is independent of $(W(u),M(u)),0\leq u<t$.

Virtual Wait in $(0,h)$

Consider the case where all servers are occupied, there are no customers waiting and $W(t)\in(0,h)$, where h is "small". Assume a server completes service before a new arrival occurs. Thus, given $(x,\boldsymbol{m})_t, 0<x<h$, $\sum_{j=0}^{J}m_j=c-1$, we obtain

$$P((0,\boldsymbol{m})_{t+h}|(x,\boldsymbol{m})_t)$$
$$=1-\lambda x+o(x).$$

The SP hits level 0 from above in a continuous manner at $t+x$. It immediately enters border line $\boldsymbol{m}$ corresponding to the border state $(0,\boldsymbol{m})$, and continues its motion. This is independent of the past history prior to t.

The above cases cover all possible situations. Formula (4.7) follows in each case, implying that the SP process has the Markov property.

4.6.8 Departures from Positive-wait States

We examine the departure rates during a sojourn on a sheet (page).

The following table describes the symbols in Fig. 4.2.

Symbol	Description
τ_n	arrival instant
C_{τ_n}	customer that arrives at τ_n
σ_n	start of service instant of C_{τ_n}
$\mathcal{S}_{\tau_n}$	$\sigma_{n+1}-\sigma_n=$ inter start-of-service departure time

Suppose the SP is at a positive level on page $\boldsymbol{m}\in\boldsymbol{M}_1$ ($\sum_{j=0}^{J}m_j=c-1$ and all c servers are occupied). The occupancy number of service rate μ_j among the $c-1$ servers not occupied by the last arrival, is $m_j, j=0,...,J$.

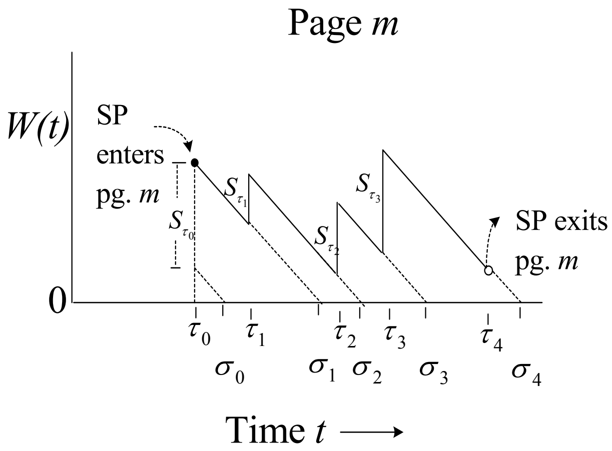

Figure 4.2: SP sojourn on page m. Departure rate may differ on intervals (τ_0, σ_0), (σ_0, σ_1), (σ_1, σ_2), (σ_2, σ_3), (σ_3, τ_4). At instants σ_0, σ_1, σ_2, arrivals C_{τ_0}, C_{τ_1}, C_{τ_2} start service. Just after departure instants $\sigma_0+\mathcal{S}_{\tau_0}$, $\sigma_1+\mathcal{S}_{\tau_1}$, $\sigma_2+\mathcal{S}_{\tau_2}$, the remaining $c-1$ servers will have server occupancies $\boldsymbol{m} = (m_0, ..., m_J)$.

The single remaining server, which is occupied by the last arrival, may have an arbitrary service rate $\mu^* \in \boldsymbol{\mu}$. Assume μ^* does not match a positive component in configuration $\boldsymbol{m}$. In order for the SP to *remain* on page $\boldsymbol{m}$ just after that arrival, the rate-μ^* server must finish first among the c busy servers (see Fig. 4.2).

While the SP is on page $\boldsymbol{m}$, the system departure rate will, in general, differ during inter-departure intervals. These possibly different departure rates have no effect on the *Markov property* of the SP process. The configurations are determined at arrival instants (earlier), when service rates are assigned (Fig. 4.2).

4.6.9 Transient PDF of Wait and Downcrossings

We next determine relationships between the transient pdf of wait and sample-path transitions. Let $\mathcal{D}_t(x, \boldsymbol{m})$ = number of sample-path downcrossings of level x on page $\boldsymbol{m} \in \boldsymbol{M}_1$ during $[0, t]$. Let

$$\mathcal{D}_t(x) = \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} \mathcal{D}_t(x, \boldsymbol{m})$$

denote the total number of downcrossings of level x on all pages during $[0,t]$. Theorem 4.2 connects the instantaneous rate of change of the expected number of downcrossings of level x in $[0,t]$, to the time-t transient pdf of wait at level x.

Theorem 4.2 *For each configuration $\boldsymbol{m} \in \boldsymbol{M}_1$,*

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(x,\boldsymbol{m})) = f_t(x,\boldsymbol{m}), x>0, t>0, \tag{4.10}$$

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(0,\boldsymbol{m})) = f_t(0^+,\boldsymbol{m}) \ (= f_t(0,\boldsymbol{m})), t>0, \tag{4.11}$$

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(x)) = f_t(x), x>0, t>0, \tag{4.12}$$

$$\frac{\partial}{\partial t}E(\mathcal{D}_t(0)) = f_t(0^+) \ (= f_t(0)), t>0. \tag{4.13}$$

Proof. Fix state-space level $x>0$. Consider instants t and $t+h$, where $t>0$, and $h>0$ is small. To prove (4.10) and (4.11) for page $\boldsymbol{m}$, we develop a table similar to (3.9) in Ch. 3 for the M/G/1 queue, and proceed as in the proof of (3.7) and (3.8). Formulas (4.12) and (4.13) follow from the definitions of $\mathcal{D}_t(x)$ and the total pdf $f_t(x), x>0$. ■

Corollary 4.1

$$E(\mathcal{D}_t(x,\boldsymbol{m})) = \int_{s=0}^{t} f_s(x,\boldsymbol{m})ds, \tag{4.14}$$

$$E(\mathcal{D}_t(0,\boldsymbol{m})) = \int_{s=0}^{t} f_s(0^+,\boldsymbol{m})ds, \tag{4.15}$$

$$E(\mathcal{D}_t(x)) = \int_{s=0}^{t} f_s(x)ds, \tag{4.16}$$

$$E(\mathcal{D}_t(0)) = \int_{s=0}^{t} f_s(0^+)ds. \tag{4.17}$$

Proof. Integrating both sides of (4.10), (4.11), (4.12) and (4.13) with respect to s over the interval $[0,t]$ and applying the initial conditions

$$E(\mathcal{D}_0(x,\boldsymbol{m})) = E(\mathcal{D}_0(x)) = 0, x \geq 0,$$

yield (4.14), (4.15), (4.16) and (4.17), respectively. ■

4.6.10 Steady-state PDF of Wait and Downcrossings

Corollary 4.2 below connects the SP limiting downcrossing rate as $t \to \infty$ and the steady-state pdf of wait, at a state-space level. It is analogous to Corollary 3.2 for M/G/1. It also demonstrates the equality of the limit of the instantaneous rate of change of the expected number of downcrossings in $[0,t]$, and the limit of of the average downcrossing rate over $[0,t]$.

Let $\boldsymbol{S_m} = ([0,\infty), \boldsymbol{m})\,, \boldsymbol{m} \in \boldsymbol{M}_1$. The results below apply to each page $\boldsymbol{T} \times \boldsymbol{S_m}, \boldsymbol{m} \in \boldsymbol{M}_1$ as well as to the the "book" $\boldsymbol{T} \times (\cup_{\boldsymbol{m}\in \boldsymbol{M}_1} \boldsymbol{S_m})$.

Corollary 4.2 *Assume the following limits exist*

$$\lim_{t\to\infty} f_t(x,\boldsymbol{m}) \equiv f(x,\boldsymbol{m}), x \in \boldsymbol{S_m}, \boldsymbol{m} \in \boldsymbol{M}_1.$$

Then

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x,\boldsymbol{m})) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x,\boldsymbol{m}))}{t} = f(x,\boldsymbol{m}), x > 0, \tag{4.18}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0,\boldsymbol{m})) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(0,\boldsymbol{m}))}{t} = f(0^+,\boldsymbol{m}) \equiv f(0,\boldsymbol{m}), \tag{4.19}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = f(x), x > 0, \tag{4.20}$$

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{D}_t(0)) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t(0))}{t} = f(0^+) \equiv f(0). \tag{4.21}$$

Proof. In (4.18), (4.19), (4.20) and (4.19), the equalities of the left-most terms to the pdf's on the right, follow by letting $t \to \infty$ in (4.10), (4.11), (4.12) and (4.13), respectively. The equalities of the middle terms to the pdf's on the right, follow by dividing both sides of (4.14), (4.15), (4.16) and (4.17) by $t > 0$ and letting $t \to \infty$. ■

4.6.11 SP $\boldsymbol{m} \to \boldsymbol{k}$ Transitions

Before discussing the relationship between the transient pdf of wait and SP upcrossings, we define SP $\boldsymbol{m} \to \boldsymbol{k}$ transitions. We say that the SP makes an $\boldsymbol{m} \to \boldsymbol{k}$ transition at instant t_0 if it *exits* state-space set $\boldsymbol{S_m}$ and *enters* state-space set $\boldsymbol{S_k}$ at t_0. That is, the SP exits $([0,\infty),\boldsymbol{m})$ and enters $([0,\infty),\boldsymbol{k})$ at t_0. If $\boldsymbol{m} = \boldsymbol{k}$, then an $\boldsymbol{m} \to \boldsymbol{k}$ transition maintains the SP on page $\boldsymbol{m}$ at t_0. Similar remarks apply to zero-wait lines $\boldsymbol{m}$, $\boldsymbol{k} \in \boldsymbol{M}_0$, or line $\boldsymbol{m} \in \boldsymbol{M}_b$ and $\boldsymbol{S_k}$ (see Subsections 2.4.2, 2.4.3 for definitions of entrance and exit).

$\boldsymbol{m} \to \boldsymbol{k}$ Upcrossing of a Level Consider $\boldsymbol{S_m}$, $\boldsymbol{S_k}$. Fix level $x > 0$. An $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing of level x occurs at instant t_0 if the SP exits set $([0,x),\boldsymbol{m})$ and enters set $((x,\infty),\boldsymbol{k})$ at t_0. That is, the SP makes both an $\boldsymbol{m} \to \boldsymbol{k}$ transition and an upcrossing of level x at t_0. Thus the SP moves instantaneously (not in Time) from page $\boldsymbol{m}$ to page $\boldsymbol{k}$ and from a level below x to a level above x. Viewed from the "cover" of the "book", the upcrossing of level x resembles an "ordinary" upcrossing of x by a sample path of the virtual wait in the M/G/1 queue (see Fig. 4.1). Similar definitions apply to line $\boldsymbol{m}$ and $\boldsymbol{S_k}$ (page $\boldsymbol{k}$).

$\boldsymbol{m} \to \boldsymbol{k}$ Parallel Transition In some variants of the M/M/c queue, the SP may make "parallel" transitions. The SP makes an $\boldsymbol{m} \to \boldsymbol{k}$ *parallel* transition at t_0 if it exits $\boldsymbol{S_m}$ ‘from a level y and enters $\boldsymbol{S_k}$ at the *same* level y, at t_0. SP "parallel" transitions can also occur in variants of *single-server* queues (e.g., queues with bulk service [16], [71]) and in other stochastic models. The concepts of configuration, pages (sheets), cover, $\boldsymbol{m} \to \boldsymbol{k}$ transitions, etc., are useful in analyzing many other stochastic models.

4.6.12 SP $\boldsymbol{m} \to \boldsymbol{k}$ Upcrossings Viewed from "Cover"

Let

$$\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}), \boldsymbol{m},\boldsymbol{k} \in \boldsymbol{M}_1$$

denote the number of SP $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings of level x during $[0,t]$. Denote the *total* number of upcrossings of level x during $[0,t]$ (as viewed from the "cover" of the "book") by

$$\mathcal{U}_t(x) = \sum_{\boldsymbol{m},\boldsymbol{k} \in \boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}). \tag{4.22}$$

In (4.22) $\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$ will be positive only if $\boldsymbol{m}$, $\boldsymbol{k}$ are such that page $\boldsymbol{k}$ is accessible from page $\boldsymbol{m}$ in one step at an arrival instant (considering lines $\boldsymbol{m}$ and $\boldsymbol{k}$ as zero-levels of pages $\boldsymbol{m}$, $\boldsymbol{k}$). For an $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing of level x to occur, the "target" page $\boldsymbol{k}$ can be either page $\boldsymbol{m}$ itself ($\boldsymbol{k} = \boldsymbol{m}$) or a different page ($\boldsymbol{k} \neq \boldsymbol{m}$).

4.6.13 Number of Types of $\boldsymbol{m} \to \boldsymbol{k}$ Upcrossings

A *type* of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossing is an ordered pair $(\boldsymbol{m},\boldsymbol{k})$. The *total* number of possible types of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings depends on how many pages communicate in one step at arrival instants. An upper bound on the total number of possible $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings is

$$\begin{aligned} &\text{number of ordered pairs } (\boldsymbol{m}, \boldsymbol{k}) \\ &= (\text{number of configurations in } \boldsymbol{M}_1)^2 \\ &= \binom{J+c-1}{c-1}^2 = \binom{J+c-1}{J}^2. \end{aligned}$$

This maximum number $\binom{J+c-1}{c-1}^2$ is realized only if *all* $\binom{J+c-1}{c-1}$ pages communicate in one step. In that case, there are $\binom{J+c-1}{c-1}$ ways to select the "source" page $\boldsymbol{m}$ and $\binom{J+c-1}{c-1}$ ways to select the "target" page $\boldsymbol{k}$ (with replacement).

Example 4.5 *Consider an M/M/c queue with $c = 3$ and $J = 1$, as in Example 4.4. The set of configurations corresponding to pages is $\boldsymbol{M}_1 = \{20, 11, 02\}$. Here $\binom{J+c-1}{c-1} = \binom{3}{2} = 3$. An upper bound on the number of types of $\boldsymbol{m} \to \boldsymbol{k}$ transitions (ordered pairs $(\boldsymbol{m}, \boldsymbol{k})$) is $3^2 = 9$. This maximum can be realized only if all configurations in $\boldsymbol{M}_1$ communicate with each other in one step. This will depend on the probabilities governing the evolution of the states over time. In the present example, configurations 20 and 02 do not communicate in one step (at an arrival instant). There are* seven *possible **types** of one-step transitions, namely,*

$$\{20 \to 20, 20 \to 11, 11 \to 20, 11 \to 11, 11 \to 02, 02 \to 11, 02 \to 02\}.$$

Transition types $20 \to 02$ and $02 \to 20$ are not possible.

We denote the probability that page $\boldsymbol{k}$ is accessible in one step from level z on page $\boldsymbol{m}$ at an arrival instant t, by $p_t(z, \boldsymbol{m} \to \boldsymbol{k})$. Thus for each $\boldsymbol{m} \in \boldsymbol{M}_1$

$$\sum_{\boldsymbol{k} \in \boldsymbol{M}_1} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) = 1.$$

Usually, for fixed z, there is some $\boldsymbol{k}$ for which $p_t(z, \boldsymbol{m} \to \boldsymbol{k}) = 0$. Then page $\boldsymbol{k}$ is not accessible from level z on page $\boldsymbol{m}$ in one step. If such inaccessibility applies for all $(z, \boldsymbol{m})$,$z \geq 0$, then page $\boldsymbol{k}$ is not accessible from page $\boldsymbol{m}$ in one step. This is the case in Example 4.3. That is, for $\boldsymbol{m} = 20$ and $\mathbf{k} = 02$,

$$p_t(z, 20 \to 02) = p_t(z, 02 \to 20) = 0, z \geq 0.$$

So pages $\boldsymbol{m}$, $\boldsymbol{k}$ are not accessible from each other in one step.

4.6.14 Transient PDF of Wait and Upcrossings

Recall that if a time-t arrival C_t finds the state to be $(z, \boldsymbol{m})$, then C_t is assigned a service rate $\mu_t(z, \boldsymbol{m}) \in \boldsymbol{\mu}$. We assume that $\mu_t(z, \boldsymbol{m})$ is a right continuous function of both z and t. Also recall that $\boldsymbol{M}_1 = \boldsymbol{M}_b = \left\{ \boldsymbol{m} | \sum_{j=0}^{J} m_j = c - 1 \right\}$.

Theorem 4.3 *For $\boldsymbol{m}, \boldsymbol{k} \in \boldsymbol{M}_1$, the instantaneous rate of change of the expected number of $\boldsymbol{m} \to \boldsymbol{k}$ upcrossings in $[0, t]$ is given by*

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})) = \lambda \int_{z=0}^{x} p_t(z, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(z,\boldsymbol{m})(x-z)} dF_t(z, \boldsymbol{m}), x \geq 0, t \geq 0, \quad (4.23)$$

where

$$\nu_t(z, \boldsymbol{m}) = \sum_{j=0}^{J} m_j \mu_j + \mu_t(z, \boldsymbol{m}).$$

Proof. Fix level $x > 0$ on page $\boldsymbol{m}$, and time $t > 0$. Examination of a sample path on page $\boldsymbol{m}$ over the time interval $(t, t + h)$ leads to the non-zero values of $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$, and corresponding probabilities in (4.24). We omit the value $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}) = 0$, since 0 does not contribute to the expected value. Also, we need not consider negative values, because $\{\mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})\}$ is a counting process (non-decreasing in time); thus $\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k}) - \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k}) \geq 0$.

$\mathcal{U}_{t+h}(x, \boldsymbol{m}, \boldsymbol{k})$ $- \mathcal{U}_t(x, \boldsymbol{m}, \boldsymbol{k})$	Probability
$+1$	$\lambda h P_0(t) p_t(0, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(0,\boldsymbol{m})x}$
	$+\lambda h \int_h^x p_t(z, \boldsymbol{m} \to \boldsymbol{k}) e^{-\nu_t(z,\boldsymbol{m})(x-z)} f_t(z) dy + o(h)$
≥ 2	$o(h)$.

(4.24)

In (4.24), taking the expected value of $\mathcal{U}_{t+h}(x,\boldsymbol{m},\boldsymbol{k})-\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$, dividing by $h>0$ and letting $h\downarrow 0$, yields

$$\begin{aligned}\frac{\partial}{\partial t}E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})) &= \lambda\cdot P_0(t)\cdot p_t(0,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_t(0,\boldsymbol{m})x}\\ &\quad+\lambda\int_{y=0}^{x} p_t(z,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_t(z,\boldsymbol{m})(x-z)\cdot} f_t(z)dy\\ &=\lambda\int_{z=0}^{x} p_t(z,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_t(z,\boldsymbol{m})(x-z)}\cdot dF_t(z,\boldsymbol{m}).\end{aligned} \tag{4.25}$$

Equation (4.25) is the same as (4.23). ■

Corollary 4.3 *For* $\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1$,

$$\begin{aligned}&E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}))\\ &=\lambda\int_{s=0}^{t}\int_{z=0}^{x} p_s(z,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_s(z,\boldsymbol{m})(x-z)}\cdot dF_s(z,\boldsymbol{m})ds, x\geq 0, t\geq 0.\end{aligned} \tag{4.26}$$

Proof. In (4.23) change the variable from t to s on both sides. Integrate with respect to s over the interval $[0,t]$. Apply the initial condition $E(\mathcal{U}_0(x,\boldsymbol{m},\boldsymbol{k}))=0$. This yields (4.26). ■

Corollary 4.4 *Consider the "cover". For* $x\geq 0, t\geq 0$,

$$\frac{\partial}{\partial t}E(\mathcal{U}_t(x))=\lambda\sum_{\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1}\int_{z=0}^{x} p_t(z,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_t(z,\boldsymbol{m})(x-z)}\cdot dF_t(z,\boldsymbol{m}) \tag{4.27}$$

$$\frac{\partial}{\partial t}E(\mathcal{U}_t(0))=\lambda\sum_{\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1} p_t(0,\boldsymbol{m}\to\boldsymbol{k})\cdot F_t(0,\boldsymbol{m}). \tag{4.28}$$

Proof. We define $\mathcal{U}_t(x), x\geq 0$ in (4.22). Equations (4.27) and (4.28) follow by setting $x>0$, and $x=0$, respectively, in (4.25), and applying (4.22). (The sample path viewed from the cover is the projection of the sample-path segments from all pages onto a single sheet.) ■

Corollary 4.5 *For* $\boldsymbol{m},\boldsymbol{k}\in\boldsymbol{M}_1$ *and* $x\geq 0, t\geq 0$,

$$E(\mathcal{U}_t(x))=\lambda\sum_{\boldsymbol{m},\boldsymbol{k}}\int_{s=0}^{t}\int_{z=0}^{x} p_s(z,\boldsymbol{m}\to\boldsymbol{k})\cdot e^{-\nu_s(z,\boldsymbol{m})(x-z)}\cdot dF_s(z,\boldsymbol{m})ds,$$

$$E(\mathcal{U}_t(0))=\lambda\sum_{\boldsymbol{m},\boldsymbol{k}}\int_{s=0}^{t} p_s(0,\boldsymbol{m}\to\boldsymbol{k})\cdot F_s(0,\boldsymbol{m})ds.$$

Proof. In (4.27) and (4.28) change t to s and integrate with respect to s on $[0,t]$. Then apply the initial condition $\mathcal{U}_0(x) = 0, x \geq 0$. ∎

4.6.15 Steady-State PDF of Wait and Upcrossings

The next corollary relates the two limits

$$\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})) \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}))}{t},$$

to the steady-state pdf and cdf of wait. Recall that

$$p(z,\boldsymbol{m}\to\boldsymbol{k}),\ \nu(z,\boldsymbol{m}),\ F(z,\boldsymbol{m}),\ \text{and } f(z,\boldsymbol{m})$$

are the limiting values of

$$p_t(z,\boldsymbol{m}\to\boldsymbol{k}),\ \nu_t(z,\boldsymbol{m}),\ F_t(z,\boldsymbol{m}),\ f_t(z,\boldsymbol{m}),$$

respectively, as $t \to \infty$.

Corollary 4.6 *For $\boldsymbol{m},\boldsymbol{k} \in \boldsymbol{M}_1$ and $x \geq 0$,*

$$\begin{aligned}
&\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})) \\
&= \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}))}{t} \\
&= \lambda \int_{z=0}^{x} p(z,\boldsymbol{m}\to\boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z,\boldsymbol{m}) \\
&= \lambda p(0,\boldsymbol{m}\to\boldsymbol{k}) \cdot e^{-\nu(0,\boldsymbol{m})x} P_0 \\
&\quad + \lambda \int_{z=0}^{x} p(z,\boldsymbol{m}\to\boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot f(z,\boldsymbol{m}) dz.
\end{aligned} \tag{4.29}$$

Proof. The equality

$$\begin{aligned}
&\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})) \\
&= \lambda \int_{z=0}^{x} p(z,\boldsymbol{m}\to\boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z,\boldsymbol{m}),
\end{aligned}$$

follows by letting $t \to \infty$ on both sides of (4.23). The equality

$$\begin{aligned}
&\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}))}{t} \\
&= \lambda \int_{z=0}^{x} p(z,\boldsymbol{m}\to\boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z,\boldsymbol{m})
\end{aligned}$$

is obtained upon dividing both sides of (4.26) by $t > 0$, letting $t \to \infty$, and using L'Hôpital's rule. Equation (4.29) then follows. ■

The next corollary relates the limits for the expected *total number* of upcrossings in $[0, t]$,

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

to the steady-state *total* probability distribution of wait.

Corollary 4.7 *For $x \geq 0$,*

$$\begin{aligned}
&\lim_{t\to\infty} \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) \\
&= \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} \\
&= \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in \boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot dF(z, \boldsymbol{m}) \\
&= \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in \boldsymbol{M}_1} p(0, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(0,\boldsymbol{m})x} P_{0,\boldsymbol{m}} \\
&\quad + \lambda \sum_{\boldsymbol{m},\boldsymbol{k}\in \boldsymbol{M}_1} \int_{z=0}^{x} p(z, \boldsymbol{m} \to \boldsymbol{k}) \cdot e^{-\nu(z,\boldsymbol{m})(x-z)} \cdot f(z, \boldsymbol{m}) dz. \quad (4.30)
\end{aligned}$$

Proof. *The result (4.30) follows from (4.29) and the definition of $\mathcal{U}_t(x)$ in (4.22).* ■

4.6.16 Equations for Transient PDF of Wait

We derive the transient model equations for the generalized M/M/c model. These equations comprise: (1) $\binom{J+c-1}{c-1}$ integro-differential equations for the "partial" pdf's $f_t(x, \boldsymbol{m}), x > 0, \boldsymbol{m} \in \boldsymbol{M}_1$; (2) $\binom{J+c-1}{c-1}$ differential equations for the zero-wait probabilities $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1 (= \boldsymbol{M}_b)$; (3) $\binom{J+c-1}{c-2}$ differential equations for the zero-wait probabilities $P_t(0, \boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0$; (4) one equation for the normalizing condition. (Recall that $\boldsymbol{M}_0 = \left\{ \boldsymbol{m} | 0 \leq \sum_{i=0}^{J} m_j \leq c - 2 \right\}$.)

We also derive the model equations for the *total* transient pdf of wait $f_t(x), x > 0$ (cover of book), and for the *total* zero-wait transient probability $P_0(t)$.

Formula (4.1) and especially (4.2) of Theorem B play important roles in these derivations. In Theorem B we take the set $\boldsymbol{A}$ to be an interval in the state space having one of its boundaries equal to x.

Equations for Partial Transient PDF's of Wait

Theorem 4.4 *(1) The integro-differential equations for $f_t(x,\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$, are*

$$\begin{aligned} & f_t(x,\boldsymbol{m}) + \lambda \textstyle\sum_{\boldsymbol{k}\neq\boldsymbol{m}} \int_{z=0}^{x} p_t(z,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu_t(z,\boldsymbol{m})(x-z)}) f_t(z,\boldsymbol{k})dz \\ & \quad +\lambda \textstyle\sum_{\boldsymbol{k}\neq\boldsymbol{m}} p_t(0,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu_t(0,\boldsymbol{k})(x-z)}) P_t(0,\boldsymbol{k}) \\ & = \tfrac{\partial}{\partial t} F_t(x,\boldsymbol{m}) - \tfrac{\partial}{\partial t} P_t(0,\boldsymbol{m}) + f_t(0,\boldsymbol{m}) \\ & \quad +\lambda \textstyle\int_{z=0}^{x} p_t(z,\boldsymbol{m},\boldsymbol{m}) e^{-\nu_t(z,\boldsymbol{m})(x-z)}) f_t(z,\boldsymbol{m})dz \\ & \quad +\lambda \textstyle\sum_{\boldsymbol{k}\neq\boldsymbol{m}} \int_{z=0}^{x} p_t(z,\boldsymbol{m},\boldsymbol{k}) f_t(z,\boldsymbol{m})dz, x \geq 0, t \geq 0, \end{aligned} \tag{4.31}$$

where configuration $\boldsymbol{k} \in \boldsymbol{M}_1$.
(2) The differential equation for $P_t(0,\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$, is

$$\begin{aligned} & f_t(0,\boldsymbol{m}) + \lambda \sum_{\boldsymbol{k}} p_t(0,\boldsymbol{k},\boldsymbol{m}) P_t(0,\boldsymbol{k}) \\ & = \frac{\partial}{\partial t} P_t(0,\boldsymbol{m}) + \left(\lambda + \sum_{j=0}^{J} m_j \mu_j\right) P_t(0,\boldsymbol{m}) \end{aligned} \tag{4.32}$$

where $\boldsymbol{k}$ is such that $\sum_{j=0}^{J} k_j = c-2$.
(3) The differential equations for $P_t(0,\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_0$, are

$$\begin{aligned} & \lambda \sum_{\boldsymbol{r}\neq\boldsymbol{m}} p_t(0,\boldsymbol{r},\boldsymbol{m}) P_t(0,\boldsymbol{r}) + \sum_{\boldsymbol{s}\neq\boldsymbol{m}} s_j \mu_j p_t(0,\boldsymbol{s},\boldsymbol{m}) P_t(0,\boldsymbol{s}) \\ & = \frac{\partial}{\partial t} P_t(0,\boldsymbol{m}) + \left(\lambda + \sum_{j=0}^{J} m_j \mu_j\right) P_t(0,\boldsymbol{m}), \end{aligned} \tag{4.33}$$

where state $(0,\boldsymbol{m})$ is accessible in one step from state $(0,\boldsymbol{r})$ at an arrival instant, and in one step from $(0,\boldsymbol{s})$ at a departure instant. That is,

$$\sum_{j=0}^{J} m_j = \sum_{j=0}^{J} r_j + 1 = \sum_{j=0}^{J} s_j - 1.$$

(4) The normalizing condition is

$$\sum_{\boldsymbol{m}\in\boldsymbol{M}_0\cup\boldsymbol{M}_1} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{x=0}^{\infty} f_t(x,\boldsymbol{m})dx = 1. \tag{4.34}$$

Proof. (1) We apply Theorem B to derive (4.31).

<u>**Choice of $\boldsymbol{A}$**</u>: In (4.1) and (4.2), substitute $((0,x],\boldsymbol{m})$ for set $\boldsymbol{A}$, (i.e., $\boldsymbol{A}$ is the left-open interval $(0,x]$ on page $\boldsymbol{m}$). The measure of set $\boldsymbol{A}$ at time t is

$$P_t(\boldsymbol{A}) = F_t(x,\boldsymbol{m}) - F_t(0,\boldsymbol{m}) = F_t(x,\boldsymbol{m}) - P_t(0,\boldsymbol{m}).$$

<u>**Entrance Rate**</u>: The SP can *enter* $\boldsymbol{A}$ by: (i) downcrossing level x on page $\boldsymbol{m}$; (ii) making a $\boldsymbol{k} \to \boldsymbol{m}$ $(\boldsymbol{k} \neq \boldsymbol{m})$ upward jump starting in $(0,x)$ on page $\boldsymbol{k}$ or from level 0 on page $\boldsymbol{k}$, that ends ***below*** x; (iii) making a jump that starts at level 0 on page $\boldsymbol{m}$, and ends ***below*** x on page $\boldsymbol{m}$.

The number of SP entrances into set $\boldsymbol{A}$ during $[0,t]$ is thus

$$\begin{aligned}\mathcal{I}_t(\boldsymbol{A}) = \mathcal{D}_t(x,\boldsymbol{m}) &+ \sum_{\boldsymbol{k}\neq\boldsymbol{m}} \mathcal{U}_t^{(0,x)}(\boldsymbol{k},\boldsymbol{m}) \\ &+ \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t^{0}(\boldsymbol{k},\boldsymbol{m}) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{k},\boldsymbol{m}),\end{aligned} \tag{4.35}$$

where $\mathcal{U}_t^{(0,x)}(\boldsymbol{k},\boldsymbol{m})$ is the number of $\boldsymbol{k} \to \boldsymbol{m}$ jumps that *start* in $(0,x)$ on page $\boldsymbol{k}$ in $[o,t]$; $\mathcal{U}_t^{0}(\boldsymbol{k},\boldsymbol{m})$ is the number of $\boldsymbol{k} \to \boldsymbol{m}$ jumps that *start* at level 0 on page $\boldsymbol{k}$ in $[o,t]$. Recall that $\mathcal{U}_t(x,\boldsymbol{k},\boldsymbol{m})$ is the number of SP $\boldsymbol{k} \to \boldsymbol{m}$ upcrossings of level x during $[0,t]$.

In (4.35) the algebraic sum

$$\left(\sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} \mathcal{U}_t^{(0,x)}(\boldsymbol{k},\boldsymbol{m}) + \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t^{0}(\boldsymbol{k},\boldsymbol{m})\right) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{k},\boldsymbol{m}) \tag{4.36}$$

is equal to

(number of SP jumps that *start below* x on pages or lines $\boldsymbol{k} \in \boldsymbol{M}_1$)
− number of such jumps that *end above* x on page $\boldsymbol{m}$ during $[0,t]$).

Thus, (4.36) is the number of SP entrances into $((0,x),\boldsymbol{m})$ during $[0,t]$, due to jumps that start below x on pages or lines outside of $\boldsymbol{T}\times((0,x),\boldsymbol{m})$ and end below x on page $\boldsymbol{m}$. Therefore $\mathcal{I}_t(\boldsymbol{A})$ is the *total* number of SP entrances into $((0,x),\boldsymbol{m})$ from all sources during $[0,t]$.

Taking expected values and then $\frac{\partial}{\partial t}$ in (4.35) yields

$$\begin{aligned}\frac{\partial}{\partial t}E\left(\mathcal{I}_t(A)\right) = \frac{\partial}{\partial t}E\left(\mathcal{D}_t(x,\boldsymbol{m})\right) &+ \sum_{\boldsymbol{k}\neq\boldsymbol{m}} \frac{\partial}{\partial t}E\left(\mathcal{U}_t^{(0,x)}(\boldsymbol{k},\boldsymbol{m})\right) \\ &+ \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \frac{\partial}{\partial t}E\left(\mathcal{U}_t^{0}(\boldsymbol{k},\boldsymbol{m})\right) - \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} \frac{\partial}{\partial t}E\left(\mathcal{U}_t(x,\boldsymbol{k},\boldsymbol{m})\right).\end{aligned} \tag{4.37}$$

Exit Rate: The SP can *exit* set $\boldsymbol{A}$ by: (i) hitting level 0 on page $\boldsymbol{m}$ from above in a continuous fashion, ("downcrossing" of level 0 and simultaneously entering state $(0, \boldsymbol{m})$); (ii) making an $\boldsymbol{m} \to \boldsymbol{k}$ (including $\boldsymbol{m} \to \boldsymbol{m}$) upcrossing of level x at an arrival instant; (iii) making an $\boldsymbol{m} \to \boldsymbol{k}$ $(\boldsymbol{k} \neq \boldsymbol{m})$ transition that ends *below* x.

The number of exits from set $\boldsymbol{A}$ during $[0,t]$ is thus

$$\begin{aligned} \mathcal{O}_t(\boldsymbol{A}) &= \mathcal{D}_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{k}\in \boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}) \\ &\quad + \sum_{\boldsymbol{k}\neq \boldsymbol{m}\in \boldsymbol{M}_1} \mathcal{U}_t^{(0,x)}(\boldsymbol{m},\boldsymbol{k}) - \sum_{\boldsymbol{k}\neq \boldsymbol{m}\in \boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k}). \qquad (4.38) \end{aligned}$$

On the right side of (4.38) $\mathcal{D}_t(0,m)$ is the number exits out of $\boldsymbol{A}$ in $[0,t]$ due to downcrossings of (entrances into) level 0 on page $\boldsymbol{m}$. The term $\sum_{\boldsymbol{k}\in \boldsymbol{M}_1} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$ is the number of SP exits of $\boldsymbol{A}$ in $[0,t]$ due to *upcrossings* of level x at arrival instants (end *above* x on all pages). The difference

$$\sum_{\boldsymbol{k}\neq m} \mathcal{U}_t^{(0,x)}(\boldsymbol{m},\boldsymbol{k}) - \sum_{\boldsymbol{k}\neq m} \mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})$$

is the number of exits from set $\boldsymbol{A}$ during $[0,t]$ due to SP jumps out of $\boldsymbol{A}$ that end *below* level x (on pages outside $\boldsymbol{m}$).

Taking expected values and then $\frac{\partial}{\partial t}$ in (4.38) results in

$$\begin{aligned} &\frac{\partial}{\partial t}E\left(\mathcal{O}_t(\boldsymbol{A})\right) \\ &= \frac{\partial}{\partial t}E\left(\mathcal{D}_t(0,m)\right) + \sum_{\boldsymbol{k}\in \boldsymbol{M}_1} \frac{\partial}{\partial t}E\left(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})\right) \\ &\quad + \sum_{\boldsymbol{k}\neq m} \frac{\partial}{\partial t}E\left(\mathcal{U}_t^{(0,x)}(x,\boldsymbol{m},\boldsymbol{k})\right) - \sum_{\boldsymbol{k}\neq m} \frac{\partial}{\partial t}E\left(\mathcal{U}_t(x,\boldsymbol{m},\boldsymbol{k})\right). \qquad (4.39) \end{aligned}$$

Integro-differential Equation: We substitute in (4.39) from (4.10), (4.11), (4.23). This yields the integro-differential equation (4.31).

(2) We derive (4.32) by letting set $\boldsymbol{A} = (0,\boldsymbol{m})$ in Theorem B, and substituting formulas from Subsection 4.6.9 relating downcrossings and the transient distribution of wait, as in the proof of (1).

(3) We derive (4.33) in a similar manner as in (2).

(4) The final equation is the normalizing condition

$$\sum_{\boldsymbol{m}\in \boldsymbol{M}_0\cup \boldsymbol{M}_1} P_t(0,\boldsymbol{m}) + \sum_{\boldsymbol{m}\in \boldsymbol{M}_1} \int_{x=0}^{\infty} f_t(x,\boldsymbol{m})dx = 1.$$

■

Remark 4.7 *In practice we can derive an equivalent set of model equations by letting set* $\boldsymbol{A} = ((x,\infty),\boldsymbol{m}), x > 0$*, in Theorem B (instead of substituting* $((0,x],\boldsymbol{m})$*). This choice of* $\boldsymbol{A}$ *may simplify the derivation of the model equations for* $f_t(x,\boldsymbol{m})$*. We would then consider SP jumps that start below and end above level* x*. This would yield terms of the form* $e^{-\nu_t(z,\boldsymbol{m})(x-z)}$ *rather than* $(1-e^{-\nu_t(z,\boldsymbol{m})(x-z)})$ *in the integrands. In real-world applications, writing the integro-differential equations is much simpler than it may seem at this point. Some practice on a few simple models will quickly establish the method. It is very intuitive.*

Remark 4.8 *We can generalize the model upon replacing* λ *by* λ_t*, depending on* t*. The arrival stream would then be a non-homogeneous Poisson process. This generalization holds because the developments in the foregoing subsections involving* λ *are essentially the same if* λ_t *is substituted for* λ*.*

Model Equations for Total Transient PDF

In the following theorem, we utilize the previously defined equivalent notation $F_t(0,\boldsymbol{m}) \equiv P_t(0,\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$, $F_t(0) \equiv P_0(t)$, $f_t(0) \equiv f_t(0^+)$.

Theorem 4.5 *For the total pdf of wait* $\{P_0(t); f_t(x), x > 0\}$*, as viewed from the "cover", the following integro-differential and differential equations hold:*

$$\begin{aligned} f_t(x) &= \frac{\partial}{\partial t}F_t(x) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu_t(z,\boldsymbol{m})(x-z)} dF_t(z,\boldsymbol{m}) \\ &= \frac{\partial}{\partial t}F_t(x) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P_t(0,\boldsymbol{m}) e^{-\nu_t(z,\boldsymbol{m})x} \\ &\quad + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu_t(z,\boldsymbol{m})(x-z)} f_t(z,\boldsymbol{m}) dz, x > 0, t \geq 0, \end{aligned} \tag{4.40}$$

$$f_t(0) = \frac{\partial}{\partial t}P_0(t) + \lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P_t(0,\boldsymbol{m}), t \geq 0. \tag{4.41}$$

Proof. In Theorem B, consider the set

$$\boldsymbol{A} = \left(\cup_{\boldsymbol{m}\in\boldsymbol{M}_0\cup\boldsymbol{M}_1}(0,\boldsymbol{m})\right) \cup \left(\cup_{\boldsymbol{m}\in\boldsymbol{M}_1}((0,x]\,,\boldsymbol{m}), x > 0\right).$$

Set $\boldsymbol{A}$ includes all $\binom{J+c}{c-1}$ zero-wait states $\left\{(0,\boldsymbol{m})|0 \le \sum_{j=0}^{J} m_j \le c-1\right\}$, as well as all positive-wait states $\left\{(y,\boldsymbol{m})|\sum_{j=0}^{J} m_j = c-1, y \in (0,x]\right\}$.

Every SP *entrance* into $\boldsymbol{A}$ must occur from above at the border x. Therefore all entrances are due to (continuous) SP downcrossings of level x. Every *exit* out of $\boldsymbol{A}$ must be due to a jump starting below x on a page and ending at a level above x on some page. Therefore all SP exits from set $\boldsymbol{A}$ are due to upcrossings of level x.

Thus

$$\begin{aligned}
&\mathcal{I}_t(A) = \mathcal{D}_t(x),\ \mathcal{O}_t(A) = \mathcal{U}_t(x),\\
&E\left(\mathcal{I}_t(A)\right) = E\left(\mathcal{D}_t(x)\right),\ E\left(\mathcal{O}_t(A)\right) = E\left(\mathcal{U}_t(x)\right),\\
&\tfrac{\partial}{\partial t}E\left(\mathcal{I}_t(A)\right) = \tfrac{\partial}{\partial t}E\left(\mathcal{D}_t(x)\right),\ \tfrac{\partial}{\partial t}E\left(\mathcal{O}_t(A)\right) = \tfrac{\partial}{\partial t}E\left(\mathcal{U}_t(x)\right).
\end{aligned}$$

We then substitute these expressions into formulas (4.12), (4.13), (4.27) and (4.28). This substitution yields the integro-differential equation (4.40) and the differential equation (4.41). ■

The normalizing condition

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

is used along with (4.40), (4.41) to solve for the unknown zero-wait probabilities and positive-wait pdf's.

When it is not feasible to obtain an analytical solution, we can use numerical or approximation techniques to solve for the transient zero-wait probabilities and positive-wait pdf's.

4.6.17 Equations for Steady-state PDF of Wait

We obtain the model equations for the steady-state pdf of wait by letting $t \to \infty$ in (4.32) – (4.34). All quantities subscripted by t have limits as $t \to \infty$. We denote the limits utilizing the same notation, omitting subscript t. If stability holds, then

$$\lim_{t\to\infty}\frac{\partial}{\partial t}F_t(x,\boldsymbol{m}) = \lim_{t\to\infty}\frac{\partial}{\partial t}F_t(0,\boldsymbol{m}) = 0.$$

This corresponds to the cdf $F(x,\boldsymbol{m})$ being independent of t.

Theorem 4.6 *The integral equation for the steady-state pdf $f(x,m)$, $\boldsymbol{m} \in \boldsymbol{M}_1$, is*

$$
\begin{aligned}
& f(x,\boldsymbol{m}) + \lambda \textstyle\sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} p(z,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z,\boldsymbol{k})dz \\
& \quad +\lambda \textstyle\sum_{\boldsymbol{k}\in\boldsymbol{M}_1} p(0,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu(0,\boldsymbol{k})(x-z)})P(0,\boldsymbol{k}) \\
& = f(0,\boldsymbol{m}) \\
& \quad +\lambda \textstyle\int_{z=0}^{x} p(z,\boldsymbol{m},\boldsymbol{m}) e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z,\boldsymbol{m})dz \\
& \quad +\lambda \textstyle\sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} p(z,\boldsymbol{m},\boldsymbol{k}) f(z,\boldsymbol{m})dz, x \geq 0.
\end{aligned} \tag{4.42}
$$

Proof. We obtain (4.42) by letting $t \to \infty$ in (4.31). ■

Theorem 4.7 *The model equation for the total steady-state pdf is*

$$
\begin{aligned}
f(x) = \quad & \lambda \textstyle\sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P(0,\boldsymbol{m}) e^{-\nu(z,\boldsymbol{m})x} \\
& +\lambda \textstyle\sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu(z,\boldsymbol{m})(x-z)} f(z,\boldsymbol{m})dz, x > 0.
\end{aligned} \tag{4.43}
$$

Proof. Let $t \to \infty$ in (4.40). ■

4.6.18 Interpretation of Equations for Sheets

We now interpret (4.42) in terms of rate balance across levels and between pages. This interpretation gives the method power for deriving steady-state model equations by inspecting a typical sample path, in a vast array of complex stochastic models.

In (4.42) the *left* side is the SP *entrance rate* into $((0,x),\boldsymbol{m})$. The term $f(x,\boldsymbol{m})$ is the SP downcrossing rate of level x on page $\boldsymbol{m}$. The term

$$
\lambda \sum_{\boldsymbol{k}\neq\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} p(z,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu(z,\boldsymbol{m})(x-z)}) f(z,\boldsymbol{k})dz
$$

is the rate at which the SP enters composite state $((0,x),\boldsymbol{m})$ due to jumps that originate in $(0,x)$ on pages $\boldsymbol{k}\neq\boldsymbol{m}$. The term

$$
\lambda \sum_{\boldsymbol{k}\in\boldsymbol{M}_1} p(0,\boldsymbol{k},\boldsymbol{m})(1-e^{-\nu(0,\boldsymbol{k})(x-z)})P(0,\boldsymbol{k})
$$

is the rate at which the SP enters composite state $((0,x),\boldsymbol{m})$ due to jumps that originate at level 0 on all pages (i.e., from all lines $\boldsymbol{k}\in\boldsymbol{M}_1$). These three terms exhaust the possible pathways by which the SP can enter $((0,x),\boldsymbol{m})$.

The *right* side of (4.42) is the *exit rate* of $((0,x),\boldsymbol{m})$. The term $f(0,\boldsymbol{m})$ is the rate at which the SP exits $((0,x),\boldsymbol{m})$ and simultaneously enters the zero-wait boundary state $(0,\boldsymbol{m})$, due to downcrossings of level 0. The term

$$\lambda \int_{z=0}^{x} p(z,\boldsymbol{m},\boldsymbol{m})e^{-\nu(z,\boldsymbol{m})(x-z)})f(z,\boldsymbol{m})dz$$

is the rate at which the SP exits $((0,x),\boldsymbol{m})$ and simultaneously enters $([x,\infty),\boldsymbol{m})$ due to jumps at arrival instants. The term

$$\lambda \sum_{\boldsymbol{k}\neq\boldsymbol{m}} \int_{z=0}^{x} p(z,\boldsymbol{m},\boldsymbol{k})f(z,\boldsymbol{m})dz$$

is the rate at which the SP exits $((0,x),\boldsymbol{m})$ and simultaneously enters a page $\boldsymbol{k}\neq\boldsymbol{m}$. These three terms exhaust the possible paths by which the SP can exit $((0,x),\boldsymbol{m})$.

Thus equation (4.42) is a rate-balance equation of the form:

$$\textbf{\textit{Rate into}}\ ((0,x),\boldsymbol{m}) = \textbf{\textit{Rate out of}}\ ((0,x),\boldsymbol{m}).$$

4.6.19 Interpretation of Equation for Total PDF

We now provide an LC interpretation of (4.43). We may view the LC analysis of the sheets as a *dissection* of the states of the model (into a partition). The total equation is like a *synthesis*, i.e., reconstruction of the parts into a whole. This idea helps to derive model equations in complex models directly from sample-path considerations. It utilizes level crossing ideas for the sheets, and also for the "cover".

In (4.43) the *left* term $f(x)$ is the total downcrossing rate of level x on all pages. On the right side, the term $\lambda\sum_{\boldsymbol{m}\in\boldsymbol{M}_1} P(0,\boldsymbol{m})e^{-\nu(z,\boldsymbol{m})x}$ is the total rate at which the SP upcrosses level x, due to jumps starting at level 0. The term

$$\lambda \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} \int_{z=0}^{x} e^{-\nu(z,\boldsymbol{m})(x-z)} f(z,\boldsymbol{m})dz$$

is the total rate at which the SP upcrosses level x, starting from a level in $(0,x)$. We form equation (4.43) by rate balance, i.e., for level x

$$\textbf{\textit{Downcrossing rate}} = \textbf{\textit{Upcrossing rate}}.$$

The normalizing condition is

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

which too has an LC interpretation. That is, multiply both sides by λ. This yields

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

On the left side, λP_0 is the rate at which the SP makes jumps out of zero-wait states. The term $\lambda \int_{x=0}^{\infty} f(x)dx$ is the rate at which the SP makes jumps from positive-wait states. The right side λ is the total rate at which the SP makes jumps. The left and right sides are equal.

4.6.20 Discussion of Rate Balance

The rate-balance interpretation provides the analyst with a powerful technique for constructing model equations for steady-state distributions in very complex models. The method is straightforward, intuitive, and relatively easy.

1. Select a state-space interval with boundary x.

2. Express the SP entrance and exit rates of the interval algebraically in terms of the unknown probability of the interval and/or unknown pdf at x.

3. Apply rate balance to form an integral equation (or other type of balance equation) for the probability and/or pdf.

4. Repeat (1)-(3) for every sub-partition of the state space as required, to form a complete system of Volterra integral equations of the second kind (as above), plus other relevant equations, depending on the model.

5. Write the normalizing condition.

6. Solve the entire system of equations simultaneously for the probabilities and pdf's of the model. This can be done analytically, numerically, by approximation, or by system-point level-crossing estimation (see Chapter 9).

Remark 4.9 *I realized in 1974 that the steady-state model equations discussed here, are really* ***rate-balance equations****. Originally, these steady-state equations had been derived by starting with Lindley recursions, analogous to those described for M/G/1 in Chapter 1. The derivation for M/M/c queues started, however, with more complex Lindley recursions.*

4.7 Example of Steady-state Equations

We now derive steady-state model equations for the *specific* M/M/c queue with $c = 3$ and $J = 1$, discussed in Example 4.4. There are two possible service rates: $\boldsymbol{\mu} = \{\mu_0, \mu_1\}$. We make a slight generalization for the *service-rate* assignment policy. For each arrival, the rates $\{\mu_0, \mu_1\}$ are assigned with probabilities $\{\alpha_0, \alpha_1\}$,$\alpha_0 + \alpha_1 = 1$ (instead of $\frac{1}{2}$ each). Our present example reduces to Example 4.4 if $\alpha_0 = \alpha_1 = \frac{1}{2}$.

We use α_0, α_1 to make it easier to follow the intuitive derivation of the model equations. This is because α_0, α_1 appear explicitly in the equations.

The set of possible configurations is $\boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{(m_0, m_1)\}$, where m_j denotes the number of servers occupied by customers with service rate $\mu_j, j = 0, 1$. From the definition of configuration

$$0 \leq \sum_{j=0}^{1} m_j \leq c - 1 = 2.$$

We abbreviate (m_0, m_1) as $m_0 m_1$. There are *six* possible configurations:

$$\boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{00, 10, 01, 20, 11, 02\},$$

where

$$\boldsymbol{M}_0 = \{00, 10, 01\}, \quad \boldsymbol{M}_1 = \{20, 11, 02\}.$$

When an arrival finds more than one server idle, it immediately occupies one of them in accordance with a *server-assignment* rule.

We first derive the equations for the zero-wait states (atoms). These are represented in the virtual-wait diagram by the six lines $\boldsymbol{T} \times \boldsymbol{m}, \boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_1$ (Fig. 4.1).

Next we derive the integral equations for the pdf's of the positive-wait states (continuous states). These states are represented by pages $\boldsymbol{T} \times ((0, \infty), m), \boldsymbol{m} \in \boldsymbol{M}_1$ (Fig. 4.1). Fix level $x > 0$. For the equation corresponding to $\boldsymbol{m} \in \boldsymbol{M}_1$, the left side is the SP rate *out of* state-space interval $((x,\infty),\boldsymbol{m})$, and the right side is the SP rate *into* $((x,\infty),\boldsymbol{m})$. (Note that we use interval (x,∞) instead of $(0,x)$, since (x,∞) results in simpler equations.) Since $\boldsymbol{M}_1 = \{20, 11, 02\}$, there will be three pdf's and three corresponding model equations.

Remark 4.10 *To summarize, the zero-wait states are* $(0,m), \boldsymbol{m} \in \boldsymbol{M}_0 \cup \boldsymbol{M}_1$. *The positive-wait states we use for the derivation, are composite states* $((x,\infty),\boldsymbol{m}), \boldsymbol{m} \in \boldsymbol{M}_1$. *We could use alternative state-space intervals having a fixed level-x boundary, such as* $((0,x),\boldsymbol{m})$ *or* $((x,a),\boldsymbol{m})$, *where constant* $a > x$, *or* $((x,bx),\boldsymbol{m}), b > 1$, *etc. For different interval selections we would derive a different, but equivalent set of model equations. A creative choice of state-space interval may simplify the derivation and final form of the equations. It may lead to new identities or insights about the model. It may also suggest easier ways to obtain solutions of the equations.*

The configurations for the zero-wait states are

$$\boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{00, 10, 01, 20, 11, 02\}.$$

The configurations for the pages are $\boldsymbol{M}_1 = \{20, 11, 02\}$ (see Fig. 4.1).

We next derive the model equations. A detailed explanation then follows.

4.7.1 Equations for Zero-wait States

The model equations for the zero-wait states are:

State	Rate out		Rate in
$(0,00)$	λP_{00}	$=$	$\mu_0 P_{10} + \mu_1 P_{01}$
$(0,10)$	$(\lambda + \mu_0) P_{10}$	$=$	$\lambda\alpha_0 P_{00} + 2\mu_0 P_{20} + \mu_1 P_{11}$
$(0,01)$	$(\lambda + \mu_1) P_{01}$	$=$	$\lambda\alpha_1 P_{00} + 2\mu_1 P_{02} + \mu_0 P_{11}$
$(0,20)$	$(\lambda + 2\mu_0) P_{20}$	$=$	$\lambda\alpha_0 P_{10} + f_{20}(0^+)$
$(0,11)$	$(\lambda + \mu_0 + \mu_1) P_{11}$	$=$	$\lambda\alpha_1 P_{10} + \lambda\alpha_0 P_{01} + f_{11}(0^+)$

(4.44)

4.7.2 Equations for Positive-wait States

We now derive the model equations for pages $\boldsymbol{m} \in \boldsymbol{M}_1$. Using composite state $((x,\infty),20)\,, x>0$, we get:

$$\begin{aligned}
& f_{20}(x) + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=x}^{\infty} f_{20}(y)dy \\
= \; & \lambda\left(\alpha_0 e^{-3\mu_0 x} + \alpha_1 \frac{\mu_1}{2\mu_0+\mu_1} e^{-(2\mu_0+\mu_1)x}\right) P_{20} \\
& + \lambda\alpha_0 \frac{\mu_1}{2\mu_0+\mu_1} e^{-(2\mu_0+\mu_1)x} P_{11} \\
& + \lambda\alpha_0 \int_{y=0}^{x} e^{-3\mu_0(x-y)} f_{20}(y)dy \\
& + \lambda\alpha_0 \frac{\mu_1}{2\mu_0+\mu_1}\int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y)dy \\
& + \lambda\alpha_0 \frac{\mu_1}{2\mu_0+\mu_1}\int_{y=x}^{\infty} f_{11}(y)dy. \qquad (4.45)
\end{aligned}$$

The model equation for the composite states $((x,\infty),11)\,, x>0$, is:

$$\begin{aligned}
& f_{11}(x) + \lambda\alpha_1 \frac{\mu_0}{\mu_0+2\mu_1}\int_{y=x}^{\infty} f_{11}(y)dy + \lambda\alpha_0 \frac{\mu_1}{2\mu_0+\mu_1}\int_{y=x}^{\infty} f_{11}(y)dy \\
= \; & \lambda\left(\alpha_1 \frac{2\mu_1}{\mu_0+2\mu_1} e^{-(\mu_0+2\mu_1)x} + \alpha_0 \frac{2\mu_0}{2\mu_0+\mu_1} e^{-(2\mu_0+\mu_1)x}\right) P_{11} \\
& + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0+\mu_1} e^{-(2\mu_0+\mu_1)x} P_{20} + \lambda\alpha_0 \frac{2\mu_1}{\mu_0+2\mu_1} e^{-(\mu_0+2\mu_1)x} P_{02} \\
& + \lambda\alpha_1 \frac{2\mu_1}{\mu_0+2\mu_1}\int_{y=0}^{x} e^{-(\mu_0+2\mu_1)(x-y)} f_{11}(y)dy \\
& + \lambda\alpha_0 \frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y)dy \\
& + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{20}(y)dy \\
& + \lambda\alpha_0 \frac{2\mu_1}{\mu_0+2\mu_1}\int_{y=0}^{x} e^{-(\mu_0+2\mu_1)(x-y)} f_{02}(y)dy \\
& + \lambda\alpha_1 \frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=x}^{\infty} f_{20}(y)dy \\
& + \lambda\alpha_0 \frac{2\mu_1}{\mu_0+2\mu_1}\int_{y=x}^{\infty} f_{02}(y)dy. \qquad (4.46)
\end{aligned}$$

The model equation using composite state $((x,\infty),02)\,, x>0$, is:

$$\begin{aligned}
&f_{02}(x)+\lambda\alpha_0\frac{2\mu_1}{\mu_0+2\mu_1}\int_{y=x}^{\infty}f_{02}(y)dy\\
=\ &\lambda\left(\alpha_1e^{-3\mu_1x}+\alpha_0\frac{\mu_0}{\mu_0+2\mu_1}e^{-(\mu_0+2\mu_1)x}\right)P_{02}\\
&+\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}e^{-(\mu_0+2\mu_1)x}P_{11}+\lambda\alpha_1\int_{y=0}^{x}e^{-3\mu_1(x-y)}f_{02}(y)dy\\
&+\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}\int_{y=0}^{x}e^{-(\mu_0+2\mu_1)(x-y)}f_{11}(y)dy\\
&+\lambda\alpha_1\frac{\mu_0}{\mu_0+2\mu_1}\int_{y=x}^{\infty}f_{11}(y)dy. \qquad (4.47)
\end{aligned}$$

The normalizing condition is:

$$\begin{aligned}
&P_{00}+P_{10}+P_{01}+P_{20}+P_{11}+P_{02}\\
&+\int_{x=0}^{\infty}(f_{20}(x)+f_{11}(x)+f_{02}(x))dx=1. \qquad (4.48)
\end{aligned}$$

4.7.3 Explanation of Steady-state Equations

Discrete States

In (4.44) the first three equations are derived as in a "bubble" diagram for discrete-state continuous-time Markov chains, using *rate out* $=$ *rate in.* The last three equations are derived similarly, except for the terms $f_{20}(0^+)$, $f_{11}(0^+)$, $f_{02}(0^+)$. These are the exit rates from $((0,\infty),20)$, $((0,\infty),11)$, and $((0,\infty),02)$ into discrete states $(0,20)$, $(0,11)$, $(0,02)$ respectively. At instants of these exits, the SP simultaneously enters the corresponding line $\boldsymbol{T}\times(0,20)$, $\boldsymbol{T}\times(0,11)$, or $\boldsymbol{T}\times(0,02)$. It continues its motion.

Positive-wait States

Left Side of Equation (4.45) In (4.45), the left side represents the SP *exit* rate out of $((x,\infty),20)$. There are two routes by which the SP can exit this composite state: (1) downcrossing level x on page 20; (2) jumping to page 11 pursuant to an arrival that is assigned rate μ_1. The term $f_{20}(x)$ is the downcrossing rate of level x on page 20.

The term

$$\lambda\alpha_1\frac{2\mu_0}{2\mu_0+\mu_1}\int_{y=x}^{\infty}f_{20}(y)dy$$

is the rate at which the SP jumps to page 11 at arrival instants. In this expression, $\lambda f_{20}(y)dy$ is the rate at which arrivals find the SP in interval $(y, y+dy)$ on page 20. The term α_1 is the probability that an arrival gets assigned rate μ_1, resulting in two servers having rate μ_0 and one server having rate μ_1 just after the arrival starts service. The term $\frac{2\mu_0}{2\mu_0+\mu_1}$ is the probability that a rate-μ_0 customer finishes first, causing the SP to jump to page 11. The SP cannot jump to page 02 if an arrival finds the configuration to be 20. The sum of terms on the left side is the *exit* rate of the SP out of (x,∞) on page 20.

Right Side of Equation (4.45) The right side of (4.45) is the SP *entrance* rate into $((x,\infty),20)$. The term

$$\lambda\left(\alpha_0e^{-3\mu_0x}+\alpha_1\frac{\mu_1}{2\mu_0+\mu_1}e^{-(2\mu_0+\mu_1)x}\right)P_{20}$$

is the entrance rate into $((x,\infty),20)$ due to arrivals that find the state to be $(0,20)$. The term λP_{20} is the rate at which arrivals find the state to be $(0,20)$. The arrival does not wait, and immediately starts service from the one available server. The term $\alpha_0e^{-3\mu_0x}$ is the probability that the arrival is assigned rate μ_0 (probability α_0) and that the minimum of three independent service times each having rate μ_0 exceeds x (probability $e^{-3\mu_0x}$).

The term

$$\alpha_1\frac{\mu_1}{2\mu_0+\mu_1}e^{-(2\mu_0+\mu_1)x}$$

is the probability, α_1, that the arrival is assigned rate μ_1, the minimum of three service times (two having rate μ_0 and one having rate μ_1) is the rate-μ_1 service (probability $\frac{\mu_1}{2\mu_0+\mu_1}$), and the minimum exceeds x (probability $e^{-(2\mu_0+\mu_1)x}$). Both terms result in the SP landing above x on page 02. The entire term is the rate at which the SP moves from level 0 on page 20 to interval (x,∞) on page 20.

The term

$$\lambda\alpha_0\frac{\mu_1}{2\mu_0+\mu_1}e^{-(2\mu_0+\mu_1)x}P_{11}$$

is the rate at which arrivals find the state to be $(0,11)$ (rate λP_{11}), are assigned service rate μ_0 (probability α_0), the minimum service time is a rate-μ_1 service $\left(\text{probability} \frac{\mu_1}{2\mu_0+\mu_1}\right)$, and the minimum exceeds x (probability $e^{-(2\mu_0+\mu_1)x}$). This is the rate at which the SP moves from discrete level 0 on page 11 to (x,∞) on page 20.

The term

$$\lambda\alpha_0 \int_{y=0}^{x} e^{-3\mu_0(x-y)} f_{20}(y)dy$$

is the rate at which arrivals find the state to be $(y, 20), y \in (0, x)$, are assigned service rate μ_0 (probability α_0), and the minimum of three service times each having rate μ_0 exceeds $x - y$ (probability $e^{-3\mu_0(x-y)}$) integrated over all $y \in (0, x)$. This is the rate at which the SP moves from $(0, x)$ on page 20 to (x,∞) on page 20 (makes $20 \to 20$ upcrossings of x).

The term

$$\lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(2\mu_0+\mu_1)(x-y)} f_{11}(y)dy$$

is the rate at which arrivals find the state to be in $((y, y + dy), 11), y \in (0, x)$ (factor $\lambda f_{11}(y)dy$), are assigned service rare μ_0, the rate-μ_1 service ends first, and the minimum of three exponential r.v.'s (two having rate μ_0 and one rate μ_1) exceeds $x - y$, integrated over all $y \in (0, x)$. This is the rate at which the SP moves from $(0, x)$ on page 11 to (x,∞) on page 20 (makes $11 \to 20$ upcrossing of x).

The term

$$\lambda\alpha_0 \frac{\mu_1}{2\mu_0 + \mu_1} \int_{y=x}^{\infty} f_{11}(y)dy$$

is the rate at which arrivals find the state to be in $((y, y + dy), 11), y > x$, are assigned service rate μ_0, the rate-μ_1 service finishes first, and the minimum of three exponential service times (two having rate μ_0 and one having rate μ_1) has any value in (x,∞). This is the rate at which the SP moves from (x,∞) on page 11 to (x,∞) on page 20 (makes $11 \to 20$ transition, from and to, points above x).

Integral Equations (4.46) and (4.47)

We derive integral equations (4.46) and (4.47) for the pdf's $f_{11}(x)$ and $f_{02}(x)$ (pages 11 and 02), in a similar manner as for $f_{20}(x)$ above.

Normalizing Condition

The normalizing condition (4.48) ensures that the sum of all zero-wait and positive-wait probabilities is 1.

4.8 Standard M/M/c: Steady-state Analysis

We analyze the standard M/M/c queue as a special case of the generalized M/M/c queue developed in sections 4.3 – 4.7. It is instructive to derive known results for M/M/c using the level crossing method. *Moreover, M/M/c is one of the first models I analyzed in 1974, to obtain evidence that the LC method works [7].*

We assume the number of servers is $c \geq 2$; and each customer receives the same service rate. Thus, $J = 0$, and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_0\}$. A configuration has one component m_0 which can take values $0, 1, ..., c-1$. All arrivals are assigned service rate $\mu_0 \equiv \mu$.

In this model, a configuration is a scalar. It is the number of customers in the *other* servers just after an arrival starts service. Equivalently, it is the number of *other* occupied servers. The set $\boldsymbol{M} = \boldsymbol{M}_0 \cup \boldsymbol{M}_1$ of all configurations has size $\binom{J+c}{c-1} = \binom{0+c}{c-1} = \binom{c}{1} = c$. That is,

$$\boldsymbol{M}_0 = \{0, 1, ..., c-2\}, \ \boldsymbol{M}_1 = \{c-1\}.$$

(Recall that $\boldsymbol{M}_1 = \boldsymbol{M}_b$, the set of "border" configurations.)

A virtual wait diagram has c *lines* for the zero-wait states $(0, j)$, $j = 0, ..., c-1$, and *one page* (*sheet*) for the composite state $((0, \infty), c-1)$ (Fig. 4.3). Line $c-1$ is the "border" line corresponding to the one border state $(0, c-1)$.

Denote the zero-wait probabilities by $P_n, n = 0, ..., c-1$, the pdf of wait by $f(x), x > 0$, and the steady-state cdf of wait by $F(x), x \geq 0$. Then

$$\begin{aligned} F(x) &= \sum_{n=0}^{c-1} P_n + \int_0^x f(x)dx, x \geq 0, \\ F(0) &= \sum_{n=0}^{c-1} P_n. \end{aligned}$$

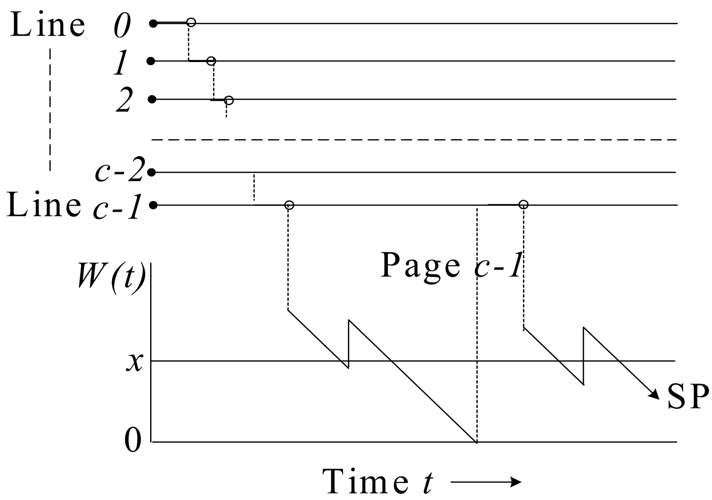

Figure 4.3: Sample path of virtual wait in *standard* M/M/c queue. There are c lines and *one* page. Line $c-1$ may be separate as depicted, or else line $c-1$ may be placed at level 0 of the page.

4.8.1 Equations for Steady-State PDF of Wait

Zero-wait States

For the zero-wait states (atoms) the model equations are (using *rate out = rate in*)

$$\begin{aligned} \lambda P_0 &= \mu P_1 \\ (\lambda+\mu)P_1 &= \lambda P_0 + 2\mu P_2 \\ (\lambda+2\mu)P_2 &= \lambda P_1 + 3\mu P_3 \\ &\cdots \\ (\lambda+(c-2)\mu)P_{c-2} &= \lambda P_{c-3} + (c-1)\mu P_{c-1} \\ (\lambda+(c-1)\mu)P_{c-1} &= \lambda P_{c-2} + f(0^+). \end{aligned} \quad (4.49)$$

Positive-wait States

For the positive-wait states $((0,\infty), c-1)$ (the single page) the model equation is

$$f(x) = \lambda P_{c-1} e^{-c\mu x} + \lambda \int_{y=0}^{x} e^{-c\mu(x-y)} f(y) dy, x > 0. \quad (4.50)$$

The normalizing condition is

$$F(\infty) = \sum_{n=0}^{c-1} P_n + \int_{y=0}^{\infty} f(x)dx = 1. \tag{4.51}$$

Explanation of Equations

We now explain the derivations of (4.49) and (4.50).

Equations (4.49) Equations (4.49) are rate-balance equations. They equate SP rates out of, and into, the discrete zero-wait states $(0, n), n = 0, ..., c-1$. The term $f(0^+)$ is the SP rate into state $(0, c-1)$ from above.

Equation (4.50) To derive the positive-wait model equation (4.50) consider composite state $((x, \infty), c-1)$ on the page (Fig. 4.3). We equate the SP exit rate (downcrossing rate of level x) to the entrance rate (upcrossing rate of level x). The downcrossing rate of level x is $f(x)$ (see Corollary 4.2).

The SP entrance rate into $((x, \infty), c-1)$ is from two sources:

(1) Entrances are generated by jumps due to arrivals when the state is border state $(0, c-1)$, starting from level 0 and ending above x. Since there is only one page, the only access to $((x, \infty), c-1)$ in one step from a zero-wait state is from $(0, (c-1))$. The SP entrance rate from this source is $\lambda P_{c-1} P(\mathcal{S} > x)$, where P_{c-1} is the probability of state $(0, c-1)$. Random variable $\mathcal{S}$ is the inter start-of-service departure time. It is distributed as $\mathrm{E}_{c\mu}$ (exponential r.v. with rate $c\mu$) since there would be c customers in service just after the arrival starts service, each with rate μ. Thus, $P(\mathcal{S} > x) = e^{-c\mu x}$. This gives the term $\lambda P_{c-1} e^{-c\mu x}$ in (4.50).

(2) Entrances into $((x, \infty), c-1)$ may also be generated by jumps due to arrivals that find the state to be $(y, c-1), y \in (0, x)$. Such jumps start at y and end above x. Just after such an arrival begins service (y later), all c servers will be occupied and $\mathcal{S}$ will be distributed as $\mathrm{E}_{c\mu}$. The SP will enter $((x, \infty), c-1)$ with probability $e^{-c\mu(x-y)}$. This leads to equation (4.50).

4.8.2 Solution of Equations

We first solve (4.50). Differentiating both sides with respect to x and solving the resulting first-order differential equation, gives

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

where A is a constant. Letting $x \downarrow 0$, we get the initial condition

$$f(0) = A = \lambda P_{c-1} \tag{4.52}$$

since $f(0)$ $(= f(0^+))$ is the SP downcrossing rate of level 0 and λP_{c-1} is the "upcrossing" rate of level 0 (rate of egress from $(0, c-1)$ above). (Equivalently, $f(0)$ is the exit rate out of $((0,\infty), c-1)$ and λP_{c-1} is the entrance rate into $((0,\infty), c-1)$.) Thus $A = \lambda P_{c-1}$ and

$$f(x) = \lambda P_{c-1} e^{-(c\mu-\lambda)x}, x > 0. \tag{4.53}$$

Note that the condition (4.52) is itself a rate-balance equation for the rates out of, and into, $((0,\infty), c-1)$.

Next, from (4.49) and (4.53) we obtain

$$\begin{aligned} P_n &= \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} P_0, n = 0, ..., c-1, \\ P_{c-1} &= \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0. \end{aligned} \tag{4.54}$$

Substituting (4.54) into (4.53) gives

$$f(x) = \lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0 \cdot e^{-(c\mu-\lambda)x}, x > 0.$$

The normalizing condition (4.51) is

$$\left(\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}\right) P_0 + \lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = 1.$$

This gives the well known value

$$P_0 = \frac{1}{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \left(\frac{\lambda}{\mu}\right)^c \frac{c\mu}{c!(c\mu-\lambda)}}. \tag{4.55}$$

The cdf of wait is

$$\begin{aligned} F(x) &= P_0 + \int_{y=0}^{x} \lambda P_{c-1} e^{-(c\mu-\lambda)y} dy \\ &= P_0 \left(1 + \lambda \left(\frac{\lambda}{\mu} \right)^{c-1} \frac{1}{(c-1)!(c\mu-\lambda)} \left(1 - e^{-(c\mu-\lambda)x} \right) \right), x \geq 0. \end{aligned} \tag{4.56}$$

Boundedness of PDF of Wait

From (4.53), $f(x) < \lambda$, since $P_{c-1} < 1$ and $e^{-(c\mu-\lambda)x} < 1$. Also, $f(0) < \lambda$, since $P_{c-1} < 1$.

4.8.3 CDF and PDF Geometrically

It is insightful and intuitive to derive the steady-state cdf and pdf of wait geometrically from sample path properties. This derivation bypasses the model equation (4.50). (We have given a similar geometric derivation for the cdf of wait in the M/M/1 queue Subsection 3.4.5).

Consider level $x > 0$ on the single page (Fig. 4.3). Rate balance across level x yields

$$\textit{Upcrossing rate of } x = \textit{Downcrossing rate of } x = f(x).$$

Equivalently

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

or

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

The sojourn time above x initiated by each upcrossing of x, is distributed as a *busy period of a standard* $M_\lambda/M_{c\mu}/1$ queue with arrival rate λ and *service rate* $c\mu$. That is, when the SP is on the page, all c servers are occupied. Thus the inter start-of-service departure time $\mathcal{S} \underset{dist}{=}$ each jump sizes on the page $\underset{dist}{=} E_{c\mu}$. By the memoryless property, excess jumps above level x are $\underset{dist}{=} E_{c\mu}$.

Let a_x denote an SP sojourn time above x. Then

$$E(a_x) = \frac{1}{c\mu - \lambda}, \tag{4.57}$$

independent of x, since the expected value of the $M_\lambda/M_{c\mu}/1$ busy period is $\frac{1}{c\mu-\lambda}$. The *proportion* of time that the SP spends above level x is

$$\lim_{t\to\infty}\frac{\mathcal{U}_t(x)\cdot\frac{1}{c\mu-\lambda}}{t} = \lim_{t\to\infty}\frac{\mathcal{U}_t(x)}{t}\cdot\frac{1}{c\mu-\lambda}$$
$$= f(x)\cdot\frac{1}{c\mu-\lambda}.$$

That proportion is also equal to the complementary cdf $1-F(x)$. Therefore the pdf and cdf satisfy

$$f(x)\cdot\frac{1}{c\mu-\lambda} = 1-F(x), x>0. \qquad (4.58)$$

This is equivalent to the differential equation

$$\frac{\frac{d}{dx}(1-F(x))}{1-F(x)} = -(c\mu-\lambda),$$
$$\frac{d}{dx}\ln(1-F(x)) = -(c\mu-\lambda),$$

with solution

$$1-F(x) = A\cdot e^{-(c\mu-\lambda)x},$$

where A is a constant. We evaluate A by letting $x\downarrow 0$, yielding the cdf of wait as

$$F(x) = 1-(1-F(0))e^{-(c\mu-\lambda)x}, x\geq 0,$$

where $F(0)=P(\text{zero wait})$. Taking $\frac{d}{dx}F(x)$,$x>0$, gives the pdf of wait

$$f(x) = (1-F(0))(c\mu-\lambda)e^{-(c\mu-\lambda)x}, x>0. \qquad (4.59)$$

We next employ the equations (4.54) to get

$$F(0) = \sum_{n=0}^{c-1}P_n = P_0\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^n\frac{1}{n!}. \qquad (4.60)$$

Note that $f(0)=\lambda P_{c-1}$. That is, the SP entrance rate into state $(0,c-1)$ from above (downcrossing rate of level 0) is equal to the SP exit rate from state $(0,c-1)$ due to arrivals.

Thus, letting $x\downarrow 0$ In (4.59) yields

$$f(0) = (1-F(0))(c\mu-\lambda) = \lambda P_{c-1}. \qquad (4.61)$$

From (4.61) and (4.54)

$$F(0) = 1 - \frac{\lambda}{c\mu - \lambda} P_{c-1} = 1 - \frac{\lambda}{c\mu - \lambda} \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} P_0. \qquad (4.62)$$

Substituting the value of $F(0)$ from (4.60) into (4.62) and solving for P_0 gives (4.55). Thus we have determined P_0 in two different ways.

Also, we now have two different, equivalent formulas for the pdf of wait, namely (4.53) and (4.59).

Remark 4.11 *Another way to obtain the second equality in (4.61) is to note that the SP expected sojourn time above* 0 *is*

$$E(a_0) = E(\textit{busy period of } M_\lambda/M_{c\mu}/1) = \frac{1}{c\mu - \lambda}.$$

The proportion of time the SP spends above level 0 *is therefore*

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(0))}{t} \cdot \frac{1}{c\mu - \lambda} = \lambda P_{c-1} \cdot \frac{1}{c\mu - \lambda} = 1 - F(0).$$

Remark 4.12 *Note that* a_0 *is equal to a* $[c-1, c]$ *busy period, denoted by* $\mathcal{B}_{c-1,c}$. *We define* $\mathcal{B}_{c-1,c}$ *as the time measured from an arrival instant when the state is* $(0, c-1)$ *until the first departure instant thereafter that leaves the system in state* $(0, c-1)$ *again. (The arrival increases the number in the system to c. The departure decreases the number to* $c-1$.*) Thus*

$$E(a_0) = E(\mathcal{B}_{c-1,c}) = E(a_x) = \frac{1}{c\mu - \lambda}, x > 0.$$

4.8.4 PMF of Number in the System

We use the foregoing pdf of wait (4.53) to derive $P_n, n = c, c+1, \ldots$.. This approach is the reverse order of the usual derivation, which first derives the pmf (probability mass function) of the number-in-system using a birth-death analysis. It then obtains the pdf of wait by conditioning on the number in the system when there is an arrival. The method that we apply here utilizes partly birth-death analysis and partly LC. It provides a different perspective on the M/M/c model.

Due to Poisson arrivals, $P_n = a_n = d_n$, where a_n, d_n are the steady-state probabilities of n units in the system just before an arrival, and just after a departure, respectively (in this subsection). That is, a_n and

d_n are the proportion of customers that "see" n in the system at arrival instants and "leave" n in the system at departure instants, respectively. Reasoning as for M/M/1 (Subsection 3.4.3), we get

$$P_n = d_n = P(n - c \text{ arrivals during a waiting time}), n = c, c+1, ...$$

Thus using (4.53) and (4.54)

$$\begin{aligned} P_n &= \int_{x=0}^{\infty} e^{-\lambda x} \frac{(\lambda x)^{n-c}}{(n-c)!} f(x) dx \\ &= \left(\frac{\lambda}{\mu}\right)^{n-c+1} \frac{1}{c^{n-c+1}} P_{c-1} \int_{x=0}^{\infty} c\mu e^{-c\mu x} \frac{(c\mu x)^{n-c}}{(n-c)!} dx \\ &= \left(\frac{\lambda}{\mu}\right)^{n} \frac{1}{c^{n-c} c!} P_0, n = c, c+1, ... \end{aligned}$$

In summary, we obtain the well known formulas

$$\left.\begin{aligned} P_0 &= \frac{1}{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} + \left(\frac{\lambda}{\mu}\right)^c \frac{c\mu}{c!(c\mu-\lambda)}} \\ P_n &= \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} P_0, n = 0, ..., c-1, \\ P_n &= \left(\frac{\lambda}{\mu}\right)^n \frac{1}{c^{n-c}} \frac{1}{c!} P_0, n = c, c+1, ... \end{aligned}\right\} \quad (4.63)$$

Also, the probability that all servers are occupied is

$$\begin{aligned} \sum_{n=c}^{\infty} P_n = P(\text{wait} > 0) &= \int_{x=0}^{\infty} f(x) dx \\ &= \lambda P_{c-1} \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = \frac{\lambda}{c\mu - \lambda} P_{c-1} \\ &= \frac{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{c!}}{c\mu - \lambda} P_0. \end{aligned} \quad (4.64)$$

The probability that there is at least one idle server is

$$\sum_{n=0}^{c-1} P_n = P(\text{wait} = 0) = 1 - \frac{\lambda \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{c!}}{c\mu - \lambda} P_0. \quad (4.65)$$

4.8.5 Inter-downcrossing and Sojourn Times

Let d_x, a_x, b_x denote respectively the time between successive SP downcrossings of level x, sojourn time above x initiated by an upcrossing of x, and sojourn time at or below x initiated by a downcrossing of x. Recall that $\mathcal{D}_t(x)$ is the number of downcrossings of x during time $(0, t)$.

Due to exponentially distributed inter-arrival times, $\mathcal{D}_t(x), t > 0$, is a renewal process for each $x \geq 0$. Random variable d_x represents the common inter-downcrossing time. Since

$$\lim_{t\to\infty} \frac{D_t(x)}{t} = \lim_{t\to\infty} \frac{E(D_t(x))}{t} = f(x),$$

we have

$$E(d_x) = \frac{1}{f(x)} = \frac{e^{(c\mu-\lambda)x}}{\lambda P_{c-1}}, x > 0. \tag{4.66}$$

From (4.57)

$$E(a_x) = \frac{1}{c\mu - \lambda},$$

independent of x.

The proportion of time that the SP is above x is equal to $1 - F(x)$. Thus

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

Therefore

$$\frac{1 - F(x)}{f(x)} = \frac{1}{c\mu - \lambda}.$$

We see this equality in (4.58) when solving for the pdf of wait geometrically using LC. Also, we can check (4.58) using

$$\begin{aligned}\frac{1 - F(x)}{f(x)} &= \frac{\int_{y=x}^{\infty} f(y)dy}{\lambda P_{c-1} e^{-(c\mu-\lambda)x}} \\ &= \frac{\int_{y=x}^{\infty} \lambda P_{c-1} e^{-(c\mu-\lambda)y} dy}{\lambda P_{c-1} e^{-(c\mu-\lambda)x}} = \frac{1}{c\mu - \lambda}. \checkmark\end{aligned}$$

The long-run proportion of time that the SP spends at or below x, is the steady-state cdf $F(x)$. Each instant that the SP downcrosses x is a regenerative point due to Poisson arrivals. From the theory of regenerative processes (e.g., [96])

$$\begin{aligned}
E(b_x) = \frac{F(x)}{f(x)} &= \frac{1-(1-F(0))e^{-(c\mu-\lambda)x}}{\lambda P_{c-1}e^{-(c\mu-\lambda)x}} \\
&= \frac{e^{(c\mu-\lambda)x}}{\lambda P_{c-1}} - \frac{(1-F(0))}{\lambda P_{c-1}} \\
&= \frac{1}{\lambda\left(\frac{\lambda}{\mu}\right)^{c-1}\frac{1}{(c-1)!}P_0}\left(e^{(c\mu-\lambda)x} - (1 - P_0\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^n\frac{1}{n!})\right) \\
&= \frac{e^{(c\mu-\lambda)x}-1}{\lambda\left(\frac{\lambda}{\mu}\right)^{c-1}\frac{1}{(c-1)!}P_0} + \frac{\sum_{n=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^n\frac{1}{n!}}{\lambda\left(\frac{\lambda}{\mu}\right)^{c-1}\frac{1}{(c-1)!}}. \qquad (4.67)
\end{aligned}$$

From (4.57), when the SP jumps above x, it next descends below x after a time having *constant* expected value independent of x. From (4.67), when the SP descends *below* level x, it next jumps above level x in a time having expected value that grows exponentially with increasing x. From (4.66), $E(d_x)$ grows exponentially as x increases.

The foregoing results for

$$d_x,\ a_x,\ b_x,\ E(d_x),\ E(a_x),\ E(b_x)$$

generalize analogous results for M/M/1 (Subsection 3.4.6).

4.9 M/M/c/c and Standard M/M/c Queues

We develop a relationship between M/M/c/c and the standard $M_\lambda/M_\mu/c$ queue. By a judicious choice of parameters, the pdf of the virtual wait in the two models have identical forms. However, the jump structure of the sample path of M/M/c/c is much simpler than that of the corresponding M/M/c model, for positive values of the virtual wait. This makes it much easier to derive the pdf of the virtual wait in M/M/c/c. The point of this exercise is that we have a model (M/M/c/c) where the pdf can be derived in one line, without having to solve an integral equation (as in M/M/c). This relationship gives rise to a broader question. For a given complex model, can we identify a related model having the same solution form, that can be solved more easily?

The M/M/c/c queue is usually analyzed using a birth-death analysis. Here, we employ an LC approach. Consider an M/M/c/c queue where the service time for each customer that enters the system has exponential rate $\mu - \frac{\lambda}{c} > 0$. (We choose $\lambda < c\mu$ because our related model is a standard $\mathrm{M}_\lambda/\mathrm{M}_\mu$/c queue in equilibrium.)

In M/M/c/c all *actual* waits are 0. In a queue where blocking is possible, we shall define the virtual wait as the time that a potential arrival *would* wait to start service, if it were not blocked. Thus the virtual wait is not always 0. In M/M/c/c, customers that arrive when the *virtual* wait is positive, are blocked and cleared from the system. In both models, the virtual wait is positive if and only if all c servers are occupied.

For M/M/c/c, consider the process $\{W(t), M(t)\}, t \geq 0$, where $W(t)$ is the virtual wait and $M(t) \in \{0, ..., c-1\}$ is the system configuration. That is, $M(t)$ is the number of occupied servers at instant t^-, if there is a vacant server at t^-. We denote the *discrete states* by $\{(0,0), ..., (0, c-1)\}$. Thus $M(t) = n$ if n *other* servers are occupied when a customer joins the system and starts service, $n = 0, ..., c-1$. Let the steady-state probability of $(0, n)$ be $P_n, n = 0, ..., c-1$. The *continuous* virtual-wait states are defined as $\{(x, c-1), x \in (0, \infty)\}$.

4.9.1 Sample Path

Consider a sample path of $\{W(t), M(t)\}$ (Fig. 4.4). Without loss of generality, assume the system starts empty. The SP is on line 0 at $t = 0$. As the system evolves, it moves among the lines until $c-1$ of the servers are occupied, just as in a standard $\mathrm{M}_\lambda/\mathrm{M}_{\mu-\frac{\lambda}{c}}$/c model. In Fig. 4.4 we situate line $c-1$ at level 0 of the page. This layout makes it easier to depict SP exchanges between line $c-1$ and the virtual-wait positive states.

Suppose a customer arrives when $c-1$ servers are occupied. The arrival joins the system and starts service. All c servers are busy just after the arrival starts service. The configuration is $c-1$, since $c-1$ *other* servers are occupied just after the arrival instant. Each of the c servers has service rate $\mu - \frac{\lambda}{c}$ once the arrival starts service, due to the memoryless property of exponential service times. The SP jumps to some level $y \in (0, \infty)$ on the page. Height y is distributed as the minimum of c iid exponential r.v.'s each distributed with rate $\mu - \frac{\lambda}{c}$. Thus $y \underset{dist}{=} \mathrm{E}_{c\mu-\lambda}$.

The SP descends at rate 1 (slope $= -1$), until it makes a continuous hit of level 0 from above. New arrivals have no effect on the sample path during this descent. Such arrivals are blocked and cleared from

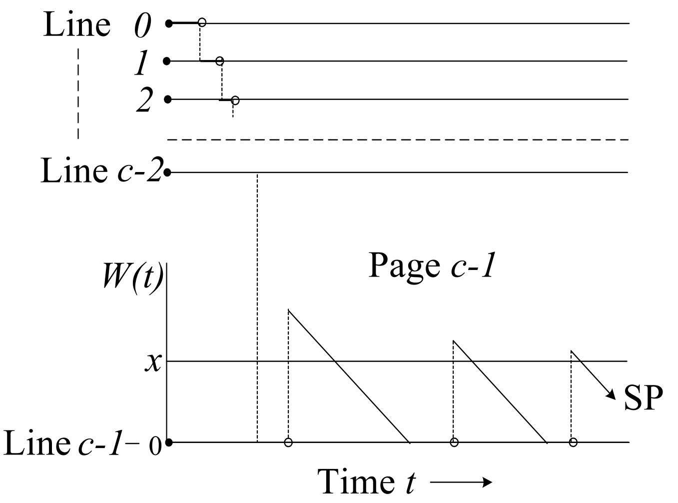

Figure 4.4: Sample path of virtual wait of $M/M/c/c$ queue. All upward jumps on the page start from level 0 and are $\underset{dist}{=}$ $\mathrm{E}_{c\mu-\lambda}$.

the system. Once the SP hits level 0, it continues its motion among the states $(0,0),...,(0,c-1)$, until it makes another jump out of state $(0,c-1)$ onto the page, etc.

A key point is that all upward jumps on the page start at level 0. Hence the jump structure for M/M/c/c is much simpler than that of the standard M/M/c queue.

4.9.2 PDF of Virtual Wait

Let the pdf of the virtual wait be $f_{c-1}(x) \equiv f(x), x > 0$. We derive the pdf of $f(x)$. Fix level $x > 0$. The SP downcrossing rate of x is $f(x)$. To obtain the upcrossing rate of x, note that an SP jump on the page is possible only when $c-1$ servers are occupied and there is an arrival. The upcrossing rate of x is $\lambda P_{c-1}e^{-(c\mu-\lambda)x}$ since all jumps start at level 0, and are $\underset{dist}{=}$ $\mathrm{E}_{c\mu-\lambda}$. Balancing SP rates out of and into set $((x,\infty),c-1)$ yields

$$f(x) = \lambda P_{c-1}e^{-(c\mu-\lambda)x}, x > 0. \tag{4.68}$$

Remark 4.13 *Formula (4.68) has precisely the same* ***form*** *as the steady-state pdf of wait in the $M_\lambda/M_\mu/c$ queue given by (4.53), except that P_{c-1} has a* ***different value****. For the $M_\lambda/M_{\mu-\frac{\lambda}{\mu}}/c/c$ queue, formula*

(4.68) is derived "instantly" from observing a sample path of the virtual wait. There is no need to solve an integral equation, as in M/M/c. In M/M/c/c, the pdf formula for $f(x)$ is inherently a model equation. This is the main relationship between the two models. The result for $M_\lambda/M_{\mu-\frac{\lambda}{\mu}}/c/c$ allows us to write the ***form*** *of the pdf of wait in $M_\lambda/M_\mu/c$ immediately.*

4.9.3 Non-blocking States

The rate-balance equations for the non-blocking states $(0,0), ..., (0,c-1)$ are the same as in (4.49) for $\mathrm{M}_\lambda/\mathrm{M}_\mu$/c with $\mu-\frac{\lambda}{c}$ substituted for μ. Thus in M/M/c/c

$$P_n = \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!} P_0, n = 0, ..., c-1,$$

so that

$$P_{c-1} = \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} P_0.$$

The normalizing condition is

$$\left(\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!}\right) P_0 + \int_{x=0}^{\infty} f(x)dx = 1.$$

Applying (4.68) gives

$$\left(\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!}\right) P_0 + \lambda \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x} dx = 1,$$

$$P_0 = \frac{1}{\sum_{n=0}^{c-1} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!} + \lambda \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} \frac{1}{c\mu-\lambda}}$$

$$= \frac{1}{\sum_{n=0}^{c} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^n \frac{1}{n!}}. \checkmark$$

4.9.4 Blocking Time

Let T_B denote the time from the instant the system gets blocked (all c servers occupied) until the first instant thereafter it becomes unblocked ($c-1$ servers occupied). We call T_B the blocking time.

The pdf of the virtual wait in $\mathrm{M}_\lambda/\mathrm{M}_{\mu-\frac{\lambda}{c}}/\mathrm{c}/\mathrm{c}$ is the same as the pdf of $\mathcal{S}$ (inter start-of-service departure time) when an arrival "sees" state $(0, c-1)$. Also, $\mathcal{S} \underset{dist}{=} T_B$.

Then $E(T_B)$ is the expected value of $\mathrm{E}_{c\mu-\lambda}$, i.e., $E(T_B) = \frac{1}{c\mu-\lambda}$. Let P_c denote the *proportion* of time the system is blocked. Then

$$
\begin{aligned}
P_c &= \int_{x=0}^{\infty} f(x)dx = \lambda P_{c-1} \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x}dx \\
&= \lambda \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c-1} \frac{1}{(c-1)!} P_0 \int_{x=0}^{\infty} e^{-(c\mu-\lambda)x}dx \\
&= \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c} \frac{1}{c!} P_0 \\
&= \frac{\left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{c} \frac{1}{c!}}{\sum_{n=0}^{c} \left(\frac{\lambda}{\mu-\frac{\lambda}{c}}\right)^{n} \frac{1}{n!}}.
\end{aligned}
$$

P_c is the probability that a right-truncated Poisson variate (truncated at c), has value c. It is the classical *Erlang loss formula* for the $\mathrm{M}_\lambda/\mathrm{M}_{\mu-\frac{\lambda}{\mu}}/\mathrm{c}/\mathrm{c}$ queue.

Note that the blocking time is a $[c-1, c]$ busy period, denoted by $\mathcal{B}_{c-1,c}$. Thus $T_B \underset{dist}{=} \mathcal{B}_{c-1,c}$. From Remark 4.12, $E(\mathcal{B}_{c-1,c}) = \frac{1}{c\mu-\lambda}$.

Remark 4.14 *Suppose that in the M/M/c/c model the servers were numbered* $1, ..., c$. *Let the service rates assigned to arrivals depend on which server is occupied, say rates* $\nu_i, i = 1, ..., c$. *Assume* $\sum_{i=1}^{c} \nu_i = c\mu - \lambda > 0$, *where* μ, λ *are the parameters of a stable* $M_\lambda/M_\mu/c$ *queue. Then the distribution of* T_B *would be the same as in (4.68). So this model can also be used as a "companion" model to obtain the pdf of wait in the* $M_\lambda/M_\mu/c$ *queue.*

4.9.5 Discussion

We can derive formula (4.68) for $f(x)$, the pdf of the virtual wait, geometrically as in Section 4.8.3. Let $F(x), x \geq 0$, be the cdf of the virtual wait. We get

$$\begin{aligned} \frac{d}{dx}\ln(1-F(x) &= \frac{-1}{E(B_{c-1,c})} = -(c\mu-\lambda), \\ F(x) &= 1-(1-F(0))e^{-(c\mu-\lambda)x}, x \geq 0, \\ f(x) &= (c\mu-\lambda)(1-F(0))e^{-(c\mu-\lambda)x}. \end{aligned} \tag{4.69}$$

Comparing (4.68) and (4.69) shows that

$$\lambda P_{c-1} = (c\mu-\lambda)(1-F(0)) = (c\mu-\lambda)P_c, \tag{4.70}$$

where P_c is the probability of c units in the system.

Note that an arrival enters the system iff the virtual wait is 0. Thus $F(0) = P$(an arrival enters the system). Hence $(1-F(0)) = P$(an arrival is blocked and cleared) $= P_c$. Equation (4.70) is precisely the balance equation that would appear in a birth-death analysis of the system.

4.10 M/M/c: Zero-waits Get Special Service

Consider an M/M/c ($c \geq 2$) queue with arrival rate λ. We assume that zero-wait customers are assigned service rate μ_0. Positive-wait customers are assigned service rate μ_1. Thus, the assigned service rate is state-dependent. We derive the steady-state pdf of wait, distribution of the number-in-system, and related model characteristics.

Denote the state of the system by $(W(t), \boldsymbol{M}(t)), t \geq 0$, where $W(t) \geq 0$ is the virtual wait and $M(t)$ is the system configuration. Thus

$$\boldsymbol{M}(t) = (m_0, m_1), 0 \leq m_0 + m_1 \leq c-1,$$

where m_j is the number of servers occupied with customers receiving rate $\mu_j, j = 0, 1$. In the notation of Subsection 4.5, integer $J = 1$. The number of zero-wait states is the total number of non-negative integer solutions for m_0, m_1 in the c equations

$$m_0 + m_1 = k, k = 0, ..., c-1.$$

That total is

$$\sum_{k=0}^{c-1} \binom{J+k}{J} = \binom{J+c}{J+1} = \binom{c+1}{2}$$
$$= \frac{c(c+1)}{2} = 1 + 2 + ... + c,$$

since $J = 1$. Recall that $\boldsymbol{M}_0 = \left\{(0, \boldsymbol{m}) | 0 \leq \sum_{j=0}^{J} m_j \leq c-2\right\}$, which contains $\frac{(c-1)c}{2}$ configurations. The set $\boldsymbol{M}_1 = \left\{\boldsymbol{m} | \sum_{j=0}^{J} m_j = c-1\right\}$, which contains $\binom{J+c-1}{J} = \binom{c}{1} = c$configurations. $\boldsymbol{M}_1$ comprises the discrete *boundary* states. The set of boundary states is specifically denoted by $\boldsymbol{M}_b$. Thus $\boldsymbol{M}_1 = \boldsymbol{M}_b$.

Let $P_{m_0 m_1}$ denote the steady-state probability that an arrival waits zero and "sees" m_0 rate-μ_0 customers and m_1 rate-μ_1 customers in service. Thus, $P_{m_0 m_1}$is the steady-state probability of state $(0, (m_0, m_1))$.

There are c positive-wait *pages (sheets)*, one for each configuration in $\boldsymbol{M}_1$. Note that

$$\boldsymbol{M}_1 = \{(c-1, 0), (c-2, 1), ..., (0, c-1)\}.$$

Let $f_{\boldsymbol{m}}(x), x > 0$, denote the steady-state pdf of the virtual wait when the occupancies of the *other* $c-1$ servers will be $\boldsymbol{m} \in \boldsymbol{M}_1$ at start of service (look-ahead idea of virtual wait).

4.10.1 Model Equations

We derive the model equations for the pdf of wait, upon considering the zero-wait and positive-wait states, and the normalizing condition.

4.10.2 Equations for Zero-wait States

The entire set of zero-wait states with configurations in $\boldsymbol{M}_0 \cup \boldsymbol{M}_1$ is

$$(0, (m_0, m_1)), 0 \leq m_0 + m_1 \leq c-1.$$

There are $\frac{c(c+1)}{2}$ linear balance equations for these states.

First consider states $(0, \boldsymbol{m})$ with $\boldsymbol{m} \in \boldsymbol{M}_0$. For $m_0 = m_1 = 0$, (empty system) there is one equation:

$$\begin{array}{rl} \textit{Rate out} & \textit{Rate in} \\ \hline \lambda P_{00} & = \mu_0 P_{10} + \mu_1 P_{01}. \end{array} \tag{4.71}$$

For states $(0,(m_0,m_1))$, $1 \le m_0+m_1 \le c-2$, there are $\frac{(c-1)c}{2}-1$ equations:

$$\begin{array}{ll} \textit{Rate out} & \textit{Rate in} \\ \hline (\lambda+m_0\mu_0+m_1\mu_1)P_{m_0m_1} & =\lambda P_{(m_0-1)m_1} \\ & \quad +(m_0+1)\mu_0 P_{(m_0+1)m_1} \\ & \quad +(m_1+1)\mu_1 P_{m_0(m_1+1)}. \end{array} \tag{4.72}$$

For states with configurations in $\boldsymbol{M}_1$, there are c equations:

$$\begin{array}{lcl} \textit{Rate out} & & \textit{Rate in} \\ \hline (\lambda+m_0\mu_0+m_1\mu_1)P_{m_0m_1} & = & \lambda P_{(m_0-1)m_1}+f_{m_0m_1}(0). \end{array} \tag{4.73}$$

These "zero-wait" equations for $P_{m_0m_1}$ are linear balance equations. In (4.73) the term $f_{m_0m_1}(0)$ $(=f_{m_0m_1}(0^+))$ is the rate at which the SP enters border state $(0,(m_0,m_1))$ due to continuous hits of level 0 from above on page m_0m_1.

4.10.3 Equations for Positive-wait States

There are c integral equations for the positive-wait states. Consider composite state $((x,\infty),\boldsymbol{m}),x>0$, on page $\boldsymbol{m}\in\boldsymbol{M}_1$. We first specify the SP exit and entrance rates of the pertinent states and sets of states in the state space. Then we write the equations.

Rate Out of $((x,\infty),m_0m_1)$

The SP rate out of $((x,\infty),m_0m_1)$ is

$$f_{m_0m_1}(x)+\lambda\frac{m_0\mu_0}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=x}^{\infty} f_{m_0m_1}(y)dy. \tag{4.74}$$

Explanation of Terms in (4.74)

In (4.74), the first term $f_{m_0m_1}(x)$ is the SP downcrossing rate of level x on page m_0m_1. The second term

$$\lambda\frac{m_0\mu_0}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=x}^{\infty} f_{m_0m_1}(y)dy$$

is the rate of arrivals when the state is $(y,m_0m_1),y>x$ (being assigned service rate μ_1 thereby adding one rate-μ_1 server upon start of service); and a rate-μ_0 service ends first. At the arrival instant the SP jumps to a page corresponding to *configuration* $(m_0-1, m_1+1) = (m_0-1, c-m_0)$. If $m_0 = 0$, the SP is on page $(0, c-1)$. The only exit from the page would be via a downcrossing of level 0. All arrivals would be assigned service rate μ_1 and cause the SP to jump upward but remain on page $(0, c-1)$. The second term in (4.74) would equal 0 if $m_0 = 0$.

Rate Into $((x,\infty), m_0m_1)$

The SP rate into $((x,\infty), m_0m_1)$ is

$$\begin{aligned}
&\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1}\\
&+\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}e^{-(m_0\mu_0+(m_1+1)\mu_1)x}P_{m_0-1,m_1+1}\\
&+\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=x}^{\infty}f_{m_0+1,m_1-1}(y)dy\\
&+\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=0}^{x}e^{-((m_0+1)\mu_0+m_1\mu_1)(x-y)}f_{m_0+1,m_1-1}(y)dy\\
&+\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=0}^{x}e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}f_{m_0m_1}(y)dy.
\end{aligned} \tag{4.75}$$

where we have inserted a comma in m_0-1, m_1+1, etc., in the subscripts, for clarity.

Explanation of Terms in (4.75)

The term

$$\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1}$$

is the rate at which the SP jumps at arrival instants from level 0 on page m_0m_1 into $((x,\infty), m_0m_1)$. At an arrival instant the customer is assigned service rate μ_0 (wait $= 0$), resulting in (m_0+1) rate-μ_0 and m_1 rate-μ_1 customers in service. If a rate-μ_0 service finishes first, the SP jumps to page m_0m_1; the probability is

$$\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}.$$

The SP jumps from level 0 over level x with probability

$$e^{-((m_0+1)\mu_0+m_1\mu_1)x}$$

since $\mathcal{S} \underset{dist}{=} \mathrm{E}_{(m_0+1)\mu_0+m_1\mu_1}$.

The term

$$\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}e^{-(m_0\mu_0+(m_1+1)\mu_1)x}P_{m_0-1,m_1+1}$$

is the rate at which the SP jumps at arrivals from level 0 on page (m_0-1, m_1+1) into $((x,\infty), m_0m_1)$. At an arrival the customer is assigned service rate μ_0 (wait $=0$), resulting in m_0 rate-μ_0 and (m_1+1) rate-μ_1 customers in service. If a rate-μ_1 service finishes first, the SP jumps to page m_0m_1; the probability is

$$\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}.$$

The SP jumps from level 0 over x with probability

$$e^{-(m_0\mu_0+(m_1+1)\mu_1)x}$$

since the inter-start-of-service departure time $\mathcal{S} \underset{dist}{=} \mathrm{E}_{m_0\mu_0+(m_1+1)\mu_1}$.

The term

$$\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=x}^{\infty} f_{m_0+1,m_1-1}(y)dy$$

is the rate at which the SP jumps at arrivals, out of (x,∞) on page (m_0+1, m_1-1) into $((x,\infty), m_0m_1)$. At an arrival the customer is assigned service rate μ_1 (wait >0) resulting in (m_0+1) rate-μ_0 and m_1 rate-μ_1 customers in service just after the start of service of the arrival. If a rate-μ_0 service finishes first, the SP jumps to page m_0m_1; this has probability

$$\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}.$$

A jump $\mathcal{S}$ of any size will cause such a jump to enter $((x,\infty), m_0m_1)$ since the start of the jump is already above level x.

The term

$$\lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=0}^{x} e^{-((m_0+1)\mu_0+m_1\mu()x-y)}f_{m_0+1,m_1-1}(y)dy$$

is the rate at which the SP jumps upward at arrivals, out of

$$((0,x),(m_0+1,m_1-1)) \text{ into } ((x,\infty),m_0m_1)\,.$$

That is, the SP upcrosses level x on page m_0m_1. An arrival is assigned service rate μ_1 (wait > 0). Just after the arrival starts service there are m_0+1 rate-μ_0 and m_1 rate-μ_1 customers in service. If a rate-μ_0 service finishes first, the SP jumps to page m_0m_1. This has probability

$$\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}.$$

If the SP starts at level $y<x$ it will upcross level x provided the r.v. $\mathcal{S}$ exceeds $x-y$. The probability of this event is

$$e^{-((m_0+1)\mu_0+m_1\mu_1)(x-y)}$$

since $\mathcal{S}$ is distributed as $\mathrm{E}_{(m_0+1)\mu_0+m_1\mu_1}$.

The term

$$\lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}f_{m_0m_1}(y)dy$$

is the rate at which the SP jumps at arrivals from $(0,x)$ on page m_0m_1 upward into $((x,\infty),m_0m_1)$. That is, it upcrosses level x on page m_0m_1. An arrival is assigned service rate μ_1 (wait >0). Just after the arrival starts service there are m_0 rate-μ_0 and (m_1+1) rate-μ_1 customers in service. If a rate-μ_1 service ends first, the SP jumps to page m_0m_1. This has probability

$$\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}.$$

If the SP starts at level y it will upcross level x provided $\mathcal{S}>x-y$. This has probability

$$e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}$$

since $\mathcal{S} \underset{dist}{=} \mathrm{E}_{m_0\mu_0+(m_1+1)\mu_1}$.

Equations for Positive-wait States

The model equation for the positive-wait states on page m_0m_1 is written by rate balance of set $((x,\infty),m_0m_1)$, *exit rate = entrance rate*. Equating (4.74) and (4.75) gives

$$\begin{aligned}
&f_{m_0m_1}(x) + \lambda\frac{m_0\mu_0}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=x}^{\infty} f_{m_0m_1}(y)dy \\
&= \lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1} \\
&\quad + \lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}e^{-(m_0\mu_0+(m_1+1)\mu_1)x}P_{m_0-1,m_1+1} \\
&\quad + \lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=x}^{\infty} f_{m_0+1,m_1-1}(y)dy \\
&\quad + \lambda\frac{(m_0+1)\mu_0}{(m_0+1)\mu_0+m_1\mu_1}\int_{y=0}^{x} e^{-((m_0+1)\mu_0+m_1\mu_1)(x-y)}f_{m_0+1,m_1-1}(y)dy \\
&\quad + \lambda\frac{(m_1+1)\mu_1}{m_0\mu_0+(m_1+1)\mu_1}\int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}f_{m_0m_1}(y)dy.
\end{aligned} \tag{4.76}$$

Equation for "Cover"

The total probability of a zero wait is

$$P_0 = \sum_{\boldsymbol{m}\in\boldsymbol{M}_0\cup\boldsymbol{M}_1} P_{\boldsymbol{m}} = \sum_{0\le m_0+m_1\le c-1} P_{m_0m_1}.$$

The total pdf of wait is

$$f(x) = \sum_{\boldsymbol{m}\in\boldsymbol{M}_1} f_{\boldsymbol{m}}(x) = \sum_{m_0+m_1=c-1} f_{m_0m_1}(x), x>0.$$

Let $x>0$ be fixed. The total SP downcrossing rate of x is $f(x)$. The total SP upcrossing rate of x due to jumps starting from level 0 at arrivals, is

$$\lambda\sum_{m_0+m_1=c-1} e^{-((m_0+1)\mu_0+m_1\mu_1)x}P_{m_0m_1}.$$

The total SP upcrossing rate of x due to jumps starting from levels in $(0,x)$ at arrivals, is

$$\lambda\sum_{m_0+m_1=c-1}\int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)}f_{m_0m_1}(y)dy.$$

Rate balance across level x gives the model equation for the cover,

$$f(x) = \lambda \sum_{m_0+m_1=c-1} e^{-((m_0+1)\mu_0+m_1\mu_1)x} P_{m_0m_1} + \lambda \sum_{m_0+m_1=c-1} \int_{y=0}^{x} e^{-(m_0\mu_0+(m_1+1)\mu_1)(x-y)} f_{m_0m_1}(y)dy. \tag{4.77}$$

Normalizing Condition

The normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ can be expressed as

$$\sum_{0\leq m_0+m_1\leq c-1} P_{m_0m_1} + \sum_{m_0+m_1=c-1} \int_{x=0}^{\infty} f_{m_0m_1}(x)dx = 1. \tag{4.78}$$

4.10.4 Solution of Model Equations

A solution procedure of (equivalent) model equations for $c \geq 2$ is detailed in [7]. In Section 4.11 below, we formulate and solve the model with $c = 2$ servers. That solution illustrates relevant SPLC ideas and related insights.

4.11 M/M/2: Zero-waits Get Special Service

M/M/2/$(\mu_0, \mu_1), (0, (0, \infty))$

To fix ideas and clarify the system dynamics of M/M/c with special service for zero-wait customers, we formulate the model with $c = 2$ servers. We discuss the solution for the zero-wait probabilities and the positive-wait pdf's. We denote the model by M/M/2/$(\mu_0, \mu_1), (0, (0, \infty))$. This notation indicates that 0-wait arrivals get service rate μ_0 and $(0, \infty)$-wait arrivals get service rate μ_1. Diagrammatically, $\mu_0 \leftrightarrow 0, \mu_1 \leftrightarrow (0, \infty)$.

There are now only three zero-wait states in $\boldsymbol{M}_0 \cup \boldsymbol{M}_1$ (see Section 4.10):

$$\{(0, m_0m_1)\} = \{(0, 00), (0, 10), (0, 01)\}.$$

Denote the steady-state probabilities of the zero-wait states by P_{00}, P_{10}, P_{01} respectively.

For example, state $(0, 10)$ indicates that an arrival would wait 0 and would "see" a rate-μ_0 customer being served by the *other* server. The

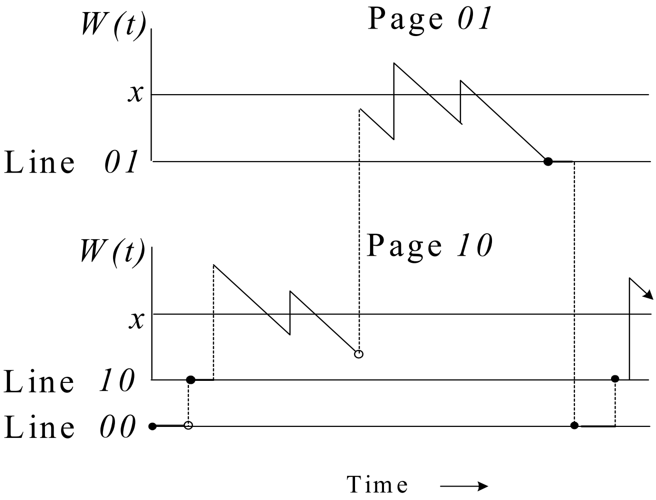

Figure 4.5: Sample path of virtual wait in $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$. Lines for states $(0, 10)$, $(0, 01)$ are at level 0 of corresponding pages. Line for state $(0, 00)$ is isolated. The SP can enter state $(0, 01)$ only by downcrossing level 0 on page 01.

arrival would be assigned rate μ_0 since it waits 0. There would then be two rate-μ_0 customers in service. The inter start-of-service departure time $\mathcal{S}$ would be distributed as $E_{2\mu_0}$.

There are only two zero-wait states such that $m_0 + m_1 = 1$ (border states). Denote the pdf's of the positive-wait states $(x, 10)$, $(x, 01), x > 0$, by $f_{10}(x)$, $f_{01}(x)$, respectively. A would-be arrival that finds the state $(x, 10), x > 0$, for example, would wait x before service. It would be assigned service rate μ_1 since its wait is positive. Just after its start of service, it would have a rate-μ_0 customer as neighbor in the other server. Random variable $\mathcal{S}$ would be distributed as $E_{\mu_0+\mu_1}$. The rate-μ_1 customer would finish service first with probability $\frac{\mu_1}{\mu_0+\mu_1}$, leaving the rate-$\mu_0$ customer in service ($m_0 m_1 = 10$). The rate-μ_0 customer would finish service first with probability $\frac{\mu_0}{\mu_0+\mu_1}$, leaving the rate-$\mu_1$ customer in service ($m_0 m_1 = 01$).

If an arrival "sees" state $(x, 01)$, $\mathcal{S}$ would be distributed as $E_{2\mu_1}$. The first customer to complete service would have rate μ_1 with certainty. The customer remaining in service just after that completion would have service rate μ_1 ($m_0 m_1 = 01$).

The virtual wait diagram has three lines and two pages (Fig. 4.5).

The total (marginal) probability of a zero wait is

$$P_0 = P_{00} + P_{10} + P_{01}.$$

The total pdf of wait is

$$f(x) = f_{10}(x) + f_{01}(x), x > 0.$$

4.11.1 Model Equations

Zero-wait States

Balancing the SP exit and entrance rates for the zero-wait states $(0, 00)$, $(0, 10)$, $(0, 01)$ gives

$$\begin{aligned} \lambda P_{00} &= \mu_0 P_{10} + \mu_1 P_{01}, \\ (\lambda + \mu_0) P_{10} &= \lambda P_{00} + f_{10}(0), \\ (\lambda + \mu_1) P_{01} &= f_{01}(0). \end{aligned} \tag{4.79}$$

In (4.79), the terms $f_{10}(0)$, $f_{01}(0)$ (same as $f_{10}(0^+)$, $f_{01}(0^+)$) are the rates at which the SP hits level 0 from above on pages 10 and 01 respectively. Immediately following such hits, the SP moves on lines 10 and 01 respectively.

Positive-wait States

Balancing the SP exit and entrance rates for $((x, \infty), 10)$ (on page 10) yields the integral equation

$$\begin{aligned} & f_{10}(x) + \lambda \int_{y=x}^{\infty} \frac{\mu_0}{\mu_0 + \mu_1} f_{10}(y) dy \\ & = \lambda P_{10} e^{-2\mu_0 x} + \lambda \frac{\mu_1}{\mu_0 + \mu_1} P_{01} e^{-(\mu_0 + \mu_1) x} \\ & \qquad + \lambda \frac{\mu_1}{\mu_0 + \mu_1} \int_{y=0}^{x} e^{-(\mu_0 + \mu_1)(x - y)} f_{10}(y) dy, x > 0. \end{aligned} \tag{4.80}$$

When formulating equation (4.80) note that the SP cannot jump from a positive-wait state on page 01 into set $((x, \infty), 10)$. An arrival that "sees" state $(y, 01), y > 0$, will be assigned rate μ_1 and start service after a wait y. Its neighbor in the other server will also have service rate μ_1 ($m_1 m_2 = 01$). Random variable $\mathcal{S}$ will be distributed as $\mathrm{E}_{2\mu_1}$, and the remaining customer in service just after the first departure thereafter will have rate μ_1. At the arrival instant, the SP will start a jump at level y on page 01, which ends at level $y + \mathcal{S} \underset{dist}{=} y + \mathrm{E}_{2\mu_1}$ on page 01. The configuration remains 01 just after the arrival. The only exit route from page 01 is via a downcrossing of level 0 (continuous hit of 0 from above)..

We balance the SP exit and entrance rates for $((x,\infty),01)$ (page 01). This gives integral equation

$$\begin{aligned} f_{01}(x) = \lambda \frac{\mu_0}{\mu_0+\mu_1} P_{01} e^{-(\mu_0+\mu_1)x} \\ + \lambda \int_{y=0}^{x} e^{-2\mu_1(x-y)} f_{01}(y)dy \\ + \lambda \frac{\mu_0}{\mu_0+\mu_1} \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} f_{10}(y)dy \\ + \lambda \frac{\mu_0}{\mu_0+\mu_1} \int_{y=x}^{\infty} f_{10}(y)dy. \end{aligned} \tag{4.81}$$

When formulating (4.81), note that the SP can exit $((x,\infty),01)$ only by downcrossing level x. The SP cannot enter $((x,\infty),01)$ from state $(0,10)$ at arrivals, since all jumps that start from line 10 (corresponding to state $(0,10)$) would be distributed as $\mathrm{E}_{2\mu_0}$, and must end on page 10 at a positive level.

The equation for the total pdf is

$$f(x) = f_{10}(x) + f_{01}(x),$$

as viewed from the "cover". The sample path as viewed from the cover is the result of sample-path segments on pages 10 and 01 projected onto a single sheet. An integral equation for $f(x)$ is obtained by balancing the SP *total* down- and upcrossing rates of level $x > 0$. This is equivalent to equating the exit and entrance rates for the state-space set

$$((x,\infty),10) \cup ((x,\infty),01).$$

The resulting equation is

$$\begin{aligned} f(x) = \lambda P_{10} e^{-2\mu_0 x} + \lambda P_{01} e^{-(\mu_0+\mu_1)x} \\ + \lambda \int_{y=0}^{x} e^{-(\mu_0+\mu_1)(x-y)} f_{10}(y)dy \\ + \lambda \int_{y=0}^{x} e^{-2\mu_1(x-y)} f_{01}(y)dy, \ x > 0. \end{aligned} \tag{4.82}$$

Equation (4.82) can also be derived by summing the corresponding sides of (4.80) and (4.81). However, it is intuitive and instructive to interpret equation (4.82) as total SP rate-balance across level $x > 0$.

The normalizing condition is

$$P_{00} + P_{10} + P_{01} + \int_{x=0}^{\infty} f_{10}(x)dx + \int_{x=0}^{\infty} f_{01}(x)dx = 1,$$

or

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{4.83}$$

4.11.2 Solution of Equations

Equation (4.80) is an integral equation in $f_{10}(x)$, which is not confounded by the presence of $f_{01}(x)$. Therefore we can utilize (4.80) directly to obtain the functional form of $f_{10}(x)$. Applying differential operator $\langle D\rangle\langle D + \mu_0 + \mu_1\rangle$ to both sides of (4.80) leads to the second order differential equation

$$\begin{aligned} &f''_{10}(x) + (\mu_0 + \mu_1 - \lambda)f'_{10}(x) - \lambda\mu_0 f_{10}(x) \\ &= 2\lambda\mu_0(\mu_0 - \mu_1)P_{10}e^{-2\mu_0 x}, \ x > 0. \end{aligned} \tag{4.84}$$

The solution of (4.84) is

$$f_{10}(x) = C_{10}e^{ax} + C_{10}^1 e^{bx} + \lambda K_{10}P_{10}e^{-2\mu_0 x}, \ x > 0$$

where

$$\begin{aligned} a &= \frac{1}{2}\left(\lambda - \mu_0 - \mu_1 - \sqrt{\lambda^2 + 2\lambda\mu_0 - 2\lambda\mu_1 + \mu_0^2 + 2\mu_0\mu_1 + \mu_1^2}\right) < 0, \\ b &= \frac{1}{2}\left(\lambda - \mu_0 - \mu_1 + \sqrt{\lambda^2 + 2\lambda\mu_0 - 2\lambda\mu_1 + \mu_0^2 + 2\mu_0\mu_1 + \mu_1^2}\right) > 0, \\ K_{10} &= \frac{2(\mu_0 - \mu_1)}{\lambda + 2\mu_0 - 2\mu_1}, \end{aligned}$$

and C_{10}, C_{10}^1 are constants of integration. A necessary condition for system stability is $f_{10}(\infty) = 0$, which implies that $C_{10}^1 = 0$ (since $b > 0$). Thus the functional form of $f_{10}(x)$ is

$$f_{10}(x) = C_{10}e^{ax} + \lambda K_{10}P_{10}e^{-2\mu_0 x}, x > 0, \tag{4.85}$$

where C_{10} is a constant to be determined.

The term K_{10} will be undefined if $\lambda+2\mu_0-2\mu_1=0$. If $\lambda+2\mu_0-2\mu_1\neq 0$ and $\mu_0-\mu_1\neq 0$, K_{10} may be positive or negative. If $\mu_0-\mu_1=0$ the model reduces to a standard M/M/c queue with $c=2$ (Section 4.8). The computed distribution of wait should then match that of standard M/M/2. (We will utilize this property later as a mild check on the correctness of the solution.)

We obtain the functional form of $f_{01}(x)$ by substituting the expression for $f_{10}(x)$ (4.85) into (4.82). Since

$$f_{01}(x)=f(x)-f_{10}(x),$$

this substitution gives the integral equation

$$\begin{aligned} f_{01}(x)=\lambda(1-K_{10})P_{10}e^{-2\mu_0 x}+\lambda P_{01}e^{-(\mu_0+\mu_1)x}-C_{10}e^{ax} \\ +\lambda\int_{y=0}^{x}e^{-(\mu_0+\mu_1)(x-y)}(C_{10}e^{ay}+\lambda K_{10}P_{10}e^{-2\mu_0 y})dy \\ +\lambda\int_{y=0}^{x}e^{-2\mu_1(x-y)}f_{01}(y)dy. \end{aligned} \tag{4.86}$$

The first integral term in (4.86) is

$$\begin{aligned} &\lambda\int_{y=0}^{x}e^{-(\mu_0+\mu_1)(x-y)}(C_{10}e^{ay}+\lambda K_{10}P_{10}e^{-2\mu_0 y})dy \\ &=\frac{\lambda C_{10}}{\mu_0+\mu_1+a}e^{ax}-\frac{\lambda^2K_{10}P_{10}e}{\mu_0-\mu_1}^{-2\mu_0 x} \\ &\quad-\left(\frac{\lambda C_{10}}{\mu_0+\mu_1+a}-\frac{\lambda^2K_{10}P_{10}}{\mu_0-\mu_1}\right)e_1^{-(\mu_0+\mu_1)x}. \end{aligned}$$

Thus (4.86) is equivalent to the integral equation

$$\begin{aligned} f_{01}(x)=&H_{01}C_{10}e^{ax}+\lambda B_{01}P_{10}e^{-2\mu_0 x} \\ &+D_{01}e^{-(\mu_0+\mu_1)x} \\ &+\lambda\int_{y=0}^{x}e^{-2\mu_1(x-y)}f_{01}(y)dy, \end{aligned} \tag{4.87}$$

where

$$\begin{aligned} H_{01}&=\frac{\lambda}{\mu_0+\mu_1+a}-1, \\ B_{01}&=1-K_{10}-\frac{\lambda K_{10}}{\mu_0-\mu_1}, \\ D_{01}&=\lambda P_{01}-\frac{\lambda C_{10}}{\mu_0+\mu_1+a}+\frac{\lambda^2K_{10}P_{10}}{\mu_0-\mu_1}. \end{aligned}$$

Applying the differential operator $\langle D+2\mu_1\rangle$ to both sides of (4.87) yields the differential equation for $f_{01}(x)$,

$$\begin{aligned} f'_{01}(x) + (2\mu_1 - \lambda) f_{01}(x) = {} & (2\mu_1 + a) H_{01} C_{10} e^{ax} \\ & + 2\lambda(\mu_1 - \mu_0) B_{01} P_{10} e^{-2\mu_0 x} \\ & + (\mu_1 - \mu_0) D_{01} e^{-(\mu_1+\mu_0)x}. \end{aligned} \tag{4.88}$$

The solution of (4.88) is

$$\begin{aligned} f_{01}(x) = {} & \frac{2\lambda(\mu_1 - \mu_0)}{2\mu_1 - \lambda - 2\mu_0} B_{01} P_{10} e^{-2\mu_0 x} \\ & + \frac{\mu_1 - \mu_0}{\mu_1 - \lambda - \mu_0} D_{01} e^{-(\mu_1+\mu_0)x} \\ & + \frac{2\mu_1 + a}{2\mu_1 - \lambda + a} H_{01} C_{10} e^{ax} \\ & + C_{01} e^{-(2\mu_1 - \lambda)x}, \end{aligned} \tag{4.89}$$

where C_{01} is a constant of integration to be determined (Subsection 4.11.4).

4.11.3 Stability Condition

Consider the functional forms of $f_{10}(x)$ and $f_{01}(x)$ in (4.85) and (4.89). In the exponents, all the coefficients of x are negative except possibly the term $-(2\mu_1 - \lambda)$ in (4.89). A necessary condition for stability is that

$$f_{10}(\infty) = f_{01}(\infty) = f(\infty) = 0.$$

This implies that $-(2\mu_1 - \lambda) < 0$, or $\lambda < 2\mu_1$. That is, the arrival rate must be less than the system departure rate when both servers are occupied by positive-wait customers. Thus, for stability, if the waiting time is large and customers are arriving, then the mean inter-arrival time should exceed the mean inter-departure time. This ensures that the waiting time will return to zero in a finite time.

4.11.4 Determination of Constants

A complete solution for the distribution of wait requires the values of five unknown constants

$$P_{00},\ P_{10},\ P_{01},\ C_{10},\ C_{01},$$

which we obtain from five independent equations.

In (4.85) letting $x \downarrow 0$ to obtain $f_{10}(0)$, and referring to (4.79) gives

$$C_{10} + \lambda K_{10} P_{10} = (\lambda + \mu_0) P_{10} - \lambda P_{00}. \tag{4.90}$$

In (4.89) letting $x \downarrow 0$ to obtain $f_{01}(0)$ gives

$$\begin{aligned} f_{01}(0) = & \frac{2\lambda(\mu_1 - \mu_0)}{2(\mu_1 - \mu_0) - \lambda} B_{01} P_{10} \\ & + \frac{\mu_1 - \mu_0}{\mu_1 - \mu_0 - \lambda} D_{01} \\ & + \frac{2\mu_1 + a}{2\mu_1 + a - \lambda} H_{01} C_{10} + C_{01}. \end{aligned} \tag{4.91}$$

Substituting $f_{01}(0)$ from (4.91) into (4.79) gives

$$\begin{aligned} C_{01} = (\lambda + \mu_1) P_{01} & - \frac{2\lambda(\mu_1 - \mu_0)}{2(\mu_1 - \mu_0) - \lambda} B_{01} P_{10} \\ & - \frac{\mu_1 - \mu_0}{\mu_1 - \mu_0 - \lambda} D_{01} \\ & - \frac{2\mu_1 + a}{2\mu_1 + a - \lambda} H_{01} C_{10}. \end{aligned} \tag{4.92}$$

We get another independent equation by substituting the functional form

$$f_{10}(x) = C_{10} e^{ax} + \lambda K_{10} P_{10} e^{-2\mu_0 x}$$

into the integral equation (4.80) and equating the coefficients of corresponding exponential terms on both sides (exponentials are linearly independent). The coefficient of $e^{-(\mu_0+\mu_1)x}$ on the right side must be 0. This yields the linear equation

$$\frac{\lambda \mu_1}{\mu_0 + \mu_1} P_{01} - \frac{1}{\mu_0 + \mu_1 + a} - \frac{\lambda K_{10}}{\mu_1 - \mu_0} P_{10} = 0. \tag{4.93}$$

The normalizing condition is

$$\begin{aligned} 1 = P_{00} + P_{10} + P_{01} + \frac{C_{10}}{(-a)} + \frac{\lambda K_{10} P_{10}}{2\mu_0} \\ + \frac{\lambda(\mu_1 - \mu_0)}{\mu_0(2(\mu_1 - \mu_0) - \lambda)} B_{01} P_{10} + \frac{\mu_1 - \mu_0}{(\mu_1 + \mu_0)(\mu_1 - \mu_0 - \lambda)} D_{01} \\ + \frac{2\mu_1 + a}{(-a)(2\mu_1 + a - \lambda)} H_{01} C_{10} + \frac{1}{2\mu_1 - \lambda} C_{01}. \end{aligned} \tag{4.94}$$

We now have a set of five equations to solve for the five constants: from (4.79)

$$\lambda P_{00} = \mu_0 P_{10} + \mu_1 P_{01},$$

and (4.90), (4.92), (4.93), (4.94).

Remark 4.15 *In the derivation of the functional forms of* $f_{10}(x)$, $f_{01}(x)$ *the expressions*

$$\mu_1 - \mu_0,\ 2\mu_1 - \lambda - 2\mu_0,\ \mu_1 - \lambda - \mu_0,\ 2\mu_1 - \lambda + a$$

appear in various denominators. If some of these four expressions were equal to 0*, the functional forms would have to be modified. There are possibly* $2^4 = 16$ *cases. The five equations used to solve for the constants would have to be modified accordingly. We must take computational care in such a situation. In this monograph we emphasize the system-point level-crossing approach to derive model equations. Techniques to solve systems of integral equations require additional study. Such solution techniques are of utmost importance. We give **numerical** solutions of the equations in several examples below.*

Remark 4.16 *It would be interesting to explain the appearance of the above expressions in the denominators. That is, does the system reduce to a particular queueing model when they are equal to* 0*? For example, when* $\mu_1 - \mu_0 = 0$*, the* $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$ *system reduces to a standard M/M/2 model. In* $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$ *the only criterion necessary for stability is* $\lambda < 2\mu_1$*. What do these exceptional denominators mean with regard to physical models?*

Another question is how to select a set of linearly independent equations to solve for the constants. Once a set of equations is derived, it can be checked for independence using matrix methods. But this amounts to trial and error. Is there a way to derive five independent equations directly? Taking derivatives may be the answer to this question.

Example 4.6 *We first give a mild numerical check on the five equations by letting* $\mu_1 - \mu_0 = 0$*. In this case* $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$ *reduces to a standard M/M/2 queue. We arbitrarily take*

$$\lambda = 1,\ \mu_0 = 1.5,\ \mu_1 = 1.5.$$

Then $a = -2.581139$*. The solution for the constants is*

$$C_{10} = 0.0,\ P_{01} = .133333,\ P_{10} = .20,\ C_{01} = .333333,\ P_{00} = .50.$$

We compare this solution with that of the standard M/M/2 queue with $\lambda = 1$, $\mu = 1.5$. *In M/M/2, the probability of an empty system is* $P_0 = 0.5$. *The probability of* 1 *customer in the system is* $P_1 = 0.33333$. *The values match* P_{00} *and* $P_{10} + P_{01}$ *in* $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$ *model, as expected.*

Also, in $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$, *we see from (4.89) that*

$$\begin{aligned} f_{01}(x) &= C_{01}e^{-(2\mu_1-\lambda)x} \\ &= \lambda P_1 e^{-(2\mu_1-\lambda)x} \\ &= 1 \cdot (0.33333)e^{-2x}, x > 0, \end{aligned}$$

since $\mu_1 - \mu_0 = 0$ *and* $C_{10} = 0$.

Example 4.7 *Let* $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$. *These values preclude that any of the four above-mentioned denominators are* 0. *We get* $a = -2.715136$. *We solve the equations and obtain*

$$\begin{aligned} &P_{00} = \ .417715, \ P_{10} = 0.339103, \ P_{01} = 0.0202270, \\ &C_{01} = 0.022818, \ C_{10} = -0.322655. \end{aligned}$$

The functions $f_{10}(x), x > 0$, *and* $f_{01}(x), x > 0$, *are linear combinations of exponentials,*

$$\begin{aligned} f_{10}(x) &= -0.322655e^{-2.715136x} + 0.617056e^{-2.2x}, \\ f_{01}(x) &= 0.505784e^{-2.2x} + 0.0678308e^{-3.31x} \\ &\quad - 0.531504e^{-2.715136x} + 0.0228180e^{-3.42x}. \end{aligned}$$

We substitute the values of P_{00}, P_{10}, P_{01}, $f_{10}(x)$, $f_{01}(x)$ *into the normalizer (4.83), and obtain* 1; *it checks* $\checkmark$.

The partial pdf's of wait $f_{10}(x)$, $f_{01}(x)$ *and total pdf of wait* $f(x)$ *are depicted in Figs. 4.6, 4.7, and 4.8 respectively.*

4.11.5 Expected Sojourn Time on a Page

Consider page 01. The SP can enter page 01 from discrete state $(0, 01)$ or from page 10, due to a jump at an arrival (Fig. 4.5). It cannot enter directly from state $(0, 10)$ at an arrival, since zero-wait arrivals are assigned rate μ_0 resulting in both servers being occupied with rate-μ_0 customers. Consequently, the SP must jump to a positive level on page 10.

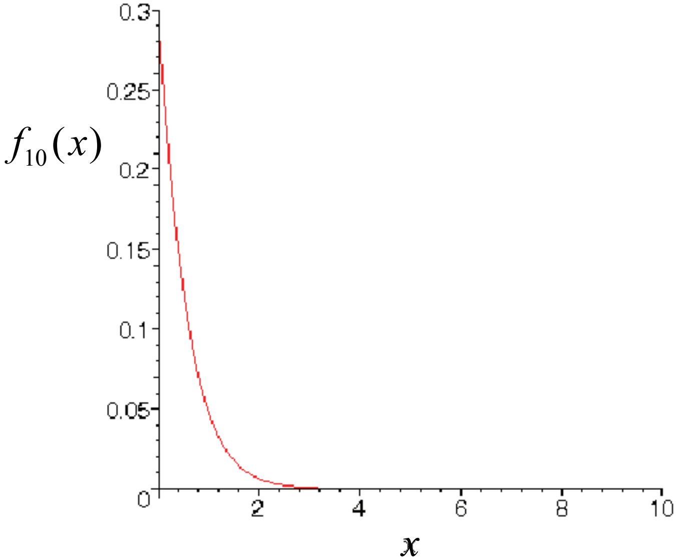

Figure 4.6: Partial pdf of wait $f_{10}(x)$ in $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$. $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$.

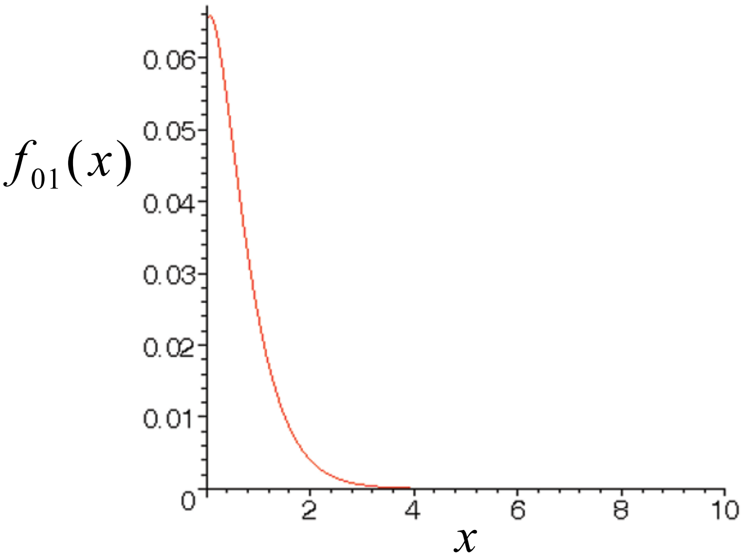

Figure 4.7: Partial pdf of wait $f_{01}(x)$ in $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$. $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$.

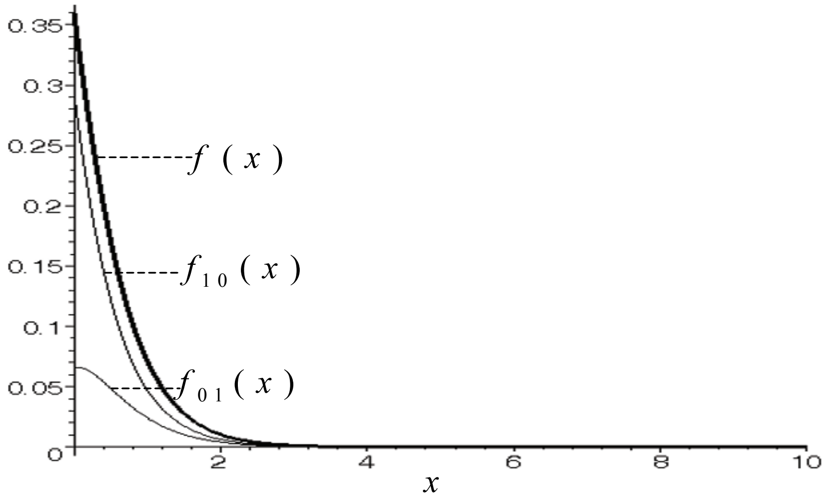

Figure 4.8: Total pdf of wait $f(x) = f_{10}(x) + f_{01}(x)$ in $M/M/2/(\mu_0, \mu_1), (0, (0, \infty))$. $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$.

In a sojourn on page 01, the first inter start-of-service departure time will be distributed as $E_{\mu_0+\mu_1}$; any other inter start-of-service departure times that follow will be distributed as $E_{2\mu_1}$. While on page 01, each departure will leave a rate-μ_1 customer in the busy server. Given that the SP enters page 01, its source was state $(0, 01)$ with probability

$$q = \frac{P_{01}}{P_{01} + \int_{y=0}^{\infty} f_{10}(y)dy}.$$

Its source was $((0, \infty), 10)$ with probability

$$1 - q = \frac{\int_{y=0}^{\infty} f_{10}(y)dy}{P_{01} + \int_{y=0}^{\infty} f_{10}(y)dy}.$$

Let H denote the height at which the SP enters page 01. Thus, a sojourn on page 01 starts at height H. Then

$$E(H|\text{source is } (0, 01)) = \frac{1}{\mu_0 + \mu_1},$$

and

$$E(H|\text{source is level } y \text{ on page } 10) = y + \frac{1}{\mu_0 + \mu_1}, y > 0,$$

since the size of a jump from either source onto page 01 is distributed as $\mathrm{E}_{\mu_0+\mu_1}$. Thus

$$E(H) = \frac{1}{\mu_0+\mu_1}q + \left(\int_{y=0}^{\infty}(y+\frac{1}{\mu_0+\mu_1})f_{10}(y)dy\right)(1-q).$$

From (4.85) $f_{10}(y)$ is given by

$$f_{10}(y) = C_{10}e^{ay} + \lambda K_{10}P_{10}e^{-2\mu_0 y}, y>0,$$

and thus

$$\begin{aligned} E(H) &= \frac{1}{\mu_0+\mu_1}q \\ &\quad + \left(\int_{y=0}^{\infty}\left(y+\frac{1}{\mu_0+\mu_1}\right)\left(C_{10}e^{ay}+\lambda K_{10}P_{10}e^{-2\mu_0 y}\right)dy\right)(1-q) \\ &= \frac{1}{\mu_0+\mu_1}\cdot q + \Bigg(\frac{1}{4}\Big(4C_{10}\mu_0^2\mu_1 + 4C_{10}\mu_0^3 \\ &\quad + 3\lambda K_{10}P_{10}a^2\mu_0 + \lambda K_{10}P_{10}a^2\mu_1 \\ &\quad - 4C_{10}a\mu_0^2\Big)/\left(a^2\mu_0^2(\mu_0+\mu_1)\right)\Bigg)\cdot(1-q). \end{aligned} \tag{4.95}$$

Let T_{01} denote a sojourn time on page 01, i.e., the time from SP entrance until the first exit from page 01 thereafter. The only possible exit is due to a downcrossing of level 0 (Fig. 4.5). Thus

$$T_{01} = H + \sum_{i=1}^{N_H}\mathcal{B}_i$$

where N_H is the number of arrivals during time H and $\mathcal{B}_i$ represents a busy period of an *M/M/1 queue* with service rate $2\mu_1$, since both servers are busy with rate-μ_1 customers. The expected busy period is obtained from (3.93) with $2\mu_1$ substituted for μ. Thus

$$E(\mathcal{B}_i) \equiv \frac{1}{2\mu_1-\lambda}.$$

The r.v.'s N_H and $\mathcal{B}_i, i=1,...,N_H$ are independent, since the $\mathcal{B}_i$'s are iid each distributed as an $\mathrm{M}_\lambda/\mathrm{M}_{2\mu_1}/1$ busy period. The expected sojourn time on page 01 is

$$\begin{aligned} E(T_{01}) &= E(H) + E\left(\sum_{i=1}^{N_H} \mathcal{B}_i\right) \\ &= E(H) + E(N_H)E(\mathcal{B}_i) \\ &= E(H) + \lambda E(H)\frac{1}{2\mu_1 - \lambda} \\ &= E(H)\left(1 + \frac{\lambda}{2\mu_1 - \lambda}\right), \end{aligned} \tag{4.96}$$

where $E(H)$ is given in formula (4.95).

Example 4.8 *In Example 4.7 with* $\lambda = 1$, $\mu_0 = 1.1$, $\mu_1 = 2.21$,*we obtain*

$$q = .111216, \;\; 1 - q = .888784,$$
$$E(H) = 0.151416.$$

The expected sojourn time on page 01 *is* $E(T_{01}) = 0.195689$.

Remark 4.17 *Various questions arise regarding Example 4.8. What is the* proportion *of time that the SP spends circulating on page* 01, *page* 10, *or in the zero-wait states? Can this question be answered for a general* $M/M/c/(\mu_0, \mu_1), (0, (0, \infty))$ *queue with* $c > 2$*? If yes, then it would be straightforward to determine* P_{00}. *This would facilitate solving for all the zero-wait probabilities and the partial pdf's of wait.*

4.12 M/M$_i$/c with Reneging

Consider an M/M/c queue, $c \geq 2$, with *distinguishable* servers having fixed exponential service rates $\mu_i, i = 1, ..., c$. Thus, the queue has *heterogeneous servers.* (This model is denoted by M/M$_i$/c.) Using notation for the general M/M/c model (Subsection 4.5), the set of possible service rates is $\boldsymbol{\mu} = \{\mu_1, ..., \mu_c\}$. However, the rates are specific to servers. New arrivals receive those service rates, depending on which server they engage. We assume that the μ_i*'s* are distinct. When some or all of the μ_i*'s* are equal, the analysis is similar with slight modification.

Assume that zero-wait arrivals start service immediately (no balking). In general, the zero-wait server-assignment policy is arbitrary. However when formulating equations for the zero-wait probabilities in specific models, we must specify a zero-wait server-assignment policy (Subsection 4.12.7).

4.12.1 Staying Function

Let $\{\tau_n\}$ be the arrival times of customers C_n,$n = 1, 2, ...$ Then $W(\tau_n^-) \equiv W_n$ is the *required wait* before service of C_n.

Define

$$\theta_{W_n} = \begin{cases} 1 \text{ if } C_n \text{ stays for a full service,} \\ 0 \text{ if } C_n \text{ reneges while waiting for service or} \\ \qquad\qquad \text{waits and balks at service, } n = 1, 2, ... \; . \end{cases}$$

We define the staying function $\overline{R}(\cdot)$ as in Subsection 3.11.1. For each $y \geq 0$, define the *conditional* probabilities

$$\overline{R}(y) \equiv P(\theta_{W_n} = 1|W_n \ = y), \;\; R(y) \equiv P(\theta_{W_n} = 0|W_n \ = y),$$

independent of $n = 1, 2, ...$ Note that $\overline{R}(y) + R(y) = 1$.

Given $W_n = y$, θ_y has a Bernoulli distribution for each $y \geq 0$, which depends on the value of y. The staying function $\overline{R}(y)$ is the probability of staying for a full service. Its complement $R(y)$ is the probability of reneging while in the waiting line or balking when reaching service.

Using the foregoing definition, $1 - \overline{R}(y)$ is not necessarily a cdf.

We assume: $\overline{R}(0) = 1$, $\overline{R}(y), y \geq 0$, is monotone decreasing in the wide sense (i.e., non-increasing); $\overline{R}(y), y > 0$, is bounded below by 0. The function $\overline{R}(y)$ may be continuous or piecewise continuous; it may be a step function.

Due to boundedness and monotonicity, $\lim_{y\to\infty} \overline{R}(y)$ exists. Let

$$\lim_{y\to\infty} \overline{R}(y) = L, 0 \leq L \leq 1.$$

If $\overline{R}(y) \equiv 1$, the model reverts to a standard $M/M_i/c$; in that case $L = 1$ (see Section 3.11 and Theorem 3.8.)

4.12.2 System Configuration

The set of possible system configurations is

$$\boldsymbol{M} = \boldsymbol{M}_0 \cup \boldsymbol{M}_1 = \{(m_1, m_2, ..., m_c) | 0 \leq \sum_{i=1}^{c} m_i \leq c - 1\},$$

where $m_i = \begin{cases} 0 \text{ if server } i \text{ is idle,} \\ 1 \text{ if server } i \text{ is occupied,} \end{cases}$ *just after* a start of service in some server.

There are $\binom{c}{j}$ configurations in which exactly j servers occupied $(\sum_{i=1}^{c} m_i = j)$. The total number of configurations in $\boldsymbol{M}$ is

$$\sum_{j=0}^{c-1} \binom{c}{j} = 2^c - 1.$$

The number of configurations in $\boldsymbol{M}_0 = \{\boldsymbol{m} | 0 \leq \sum_{i=1}^{c} m_i \leq c-2\}$, is $2^c - 1 - c$. In $\boldsymbol{M}_1 = \{\boldsymbol{m} | \sum_{i=1}^{c} m_i = c-1\}$ (border configurations), the number is c. (Recall that $\boldsymbol{M}_1 = \boldsymbol{M}_b$.)

4.12.3 State of System and Sample Path

State of System

Denote the state of the system by $(W(t), M(t)), t \geq 0$, where $W(t) \geq 0$ is the virtual wait and system configuration $M(t) \in \boldsymbol{M}$ at instant t. $\{(W(t), M(t))\}$ is the system point process of the model (Section 4.6).

Sample Path

Consider a sample path of $\{(W(t), M(t))\}$. A sample-path diagram has $2^c - 1$ lines corresponding to the zero-wait states $(0, m), m \in \boldsymbol{M}$. $(W(t) = 0)$, and c sheets corresponding to the positive-wait states $(W(t) > 0)$.

Assume the system starts empty at $t = 0$. The dynamics over time proceeds like a multiple server queue with heterogeneous servers until $c-1$ servers are occupied. That is, arriving customers wait 0, get service and then depart; they accumulate in the system until the first instant such that $c-1$ servers are occupied. Correspondingly the SP moves within the $2^c - 1 - c$ lines for the non-border zero-wait states. It resides on the lines for exponentially distributed times, making jumps from line to line over time. The various states unfold until the SP ends up on one of the c *border* lines.

Recall that all zero-wait arrivals stay for full service (no balking). Assume that a new arrival C_τ finds $c-1$ servers occupied (SP on a border line). Then C_τ waits 0, and starts service in the single idle server. At τ^- the configuration is some $\boldsymbol{m} \in \boldsymbol{M}_1$. At instant τ all c servers are occupied. The SP jumps at instant τ to one of the c sheets, depending on which service will finish first. The probability that server k will finish first is $\frac{\mu_k}{\mu}$ where $\mu = \mu_1 + ... + \mu_c$. The SP will be at a height $\underset{dist}{=}$ E_μ, since the inter start-of-service departure time $\mathcal{S}$ is the *minimum* of c independent exponentially distributed r.v.'s with rates $\mu_1, ..., \mu_c$.

Let $\boldsymbol{m}_{\bar{i}}$ denote a *border* configuration such that the rate-μ_i server is *idle.* For configuration $\boldsymbol{m}_{\bar{i}}$, $m_j = 1, j \neq i$. That is,

$$m_1 + ... + m_{i-1} + 0 + m_{i+1} + ... + m_c = c - 1.$$

At τ the SP will end up at a positive height on page $\boldsymbol{m}_{\bar{k}}$ with probability $\frac{\mu_k}{\mu}$,$k = 1, ..., c$, since the rate-μ_k server will finish first with probability $\frac{\mu_k}{\mu}$.

4.12.4 Zero-wait Probabilities

Let $P_n, n = 0, ..., c-1$ denote the steady-state probability of n customers in the system at an arbitrary point in time. Let $P_{n,\boldsymbol{m}}$ denote the probability that there are n customers in the system and the configuration is $\boldsymbol{m} \in \boldsymbol{M}$. There are $\binom{c}{n}$ configurations such that $\sum_{i=1}^{c} m_i = n$. Let

$$\boldsymbol{M}_n = \{\boldsymbol{m} | \sum_{i=1}^{c} m_i = n\}.$$

Thus

$$P_n = \sum_{\boldsymbol{m} \in \boldsymbol{M}_n} P_{n,\boldsymbol{m}}, n = 0, ..., c-1.$$

Due to Poisson arrivals P_n is the probability that an arrival waits 0 and "sees" n other customers in service just before it starts service.

Remark 4.18 *For the* ***zero-wait*** *states, a configuration specifies the service rates in the servers at an arbitrary time point. Due to Poisson arrivals, this is the same as the service rates* ***just before*** *an arrival. It is also the same as the service rates in the* ***other*** *servers* ***just after*** *an arrival starts service in a free server.*

For the ***positive-wait*** *states, a configuration defines the service rates in the* ***other*** *servers* ***just after*** *start of service.*

The probability of a zero wait is denoted by $F(0)$. Then

$$F(0) = \sum_{n=0}^{c-1} P_n = \sum_{n=0}^{c-1} \sum_{\boldsymbol{m} \in \boldsymbol{M}_n} P_{n,\boldsymbol{m}}. \tag{4.97}$$

4.12.5 Positive-wait PDF and CDF

Let $f_{\boldsymbol{m}}(x), x > 0$, denote the "partial" pdf of wait for page $\boldsymbol{m} \in \boldsymbol{M}_1$. Let the marginal pdf for the *cover* be

$$f(x) = \sum_{\boldsymbol{m} \in \boldsymbol{M}_1} f_{\boldsymbol{m}}(x), x > 0.$$

The *total* density function of wait is $\{\boldsymbol{P}_0; f(x), x > 0\}$. Let the cdf of wait be $F(x), x \geq 0$. Then $F(x) = F(0) + \int_{y=0}^{x} f(y)dy$, where $F(0)$ is defined in (4.97).

4.12.6 Model Equations

A key assumption of this model is that positive-wait arrivals may renege from the waiting line or wait and balk at service (probability $R(\cdot)$). Otherwise they may stay for complete service (probability $\overline{R}(\cdot)$).

We derive an integral equation for $f(x)$, the pdf of wait of stayers (arrivals that wait and receive a full service), namely,

$$f(x) = \lambda P_{c-1} e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy, x > 0, \tag{4.98}$$

directly using the sample path, as follows.

In (4.98) the left side $f(x)$ is the total SP downcrossing rate of level x on all c sheets, projected on the "cover". On the right side, the term $\lambda P_{c-1} e^{-\mu x}$ is the SP upcrossing rate of x due to jumps originating from level 0 of all the sheets. Such jumps start from some state $\{(0, \boldsymbol{m}_{\bar{i}})\}, i = 1, ..., c$. For these jumps $\mathcal{S} \underset{dist}{=} \mathrm{E}_{\mu}$. The term $\lambda \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy$ is the rate at which the SP upcrosses level x due to jumps starting at level $y \in (0, x)$. The right side is therefore the total SP upcrossing rate of level x. Rate balance across x yields (4.98).

Comparison of (4.98) with equation (3.166) implies that the solution of (4.98) is

$$f(x) = \lambda P_{c-1} e^{-\left(\mu x - \lambda \int_{y=0}^{x} \overline{R}(y) dy\right)}, x > 0, \tag{4.99}$$

where

$$\mu = \sum_{i=1}^{c} \mu_i \text{ and } P_{c-1} = \sum_{i=1}^{c} P_{c-1, \boldsymbol{m}_{\bar{i}}}.$$

We next obtain integral equations for the pdf's $f_{\bar{i}}(x), x > 0$, on the c sheets:

$$\begin{aligned} & f_{\bar{i}}(x) + \lambda(1 - \frac{\mu_i}{\mu}) \int_{y=x}^{\infty} \overline{R}(y) f_{\bar{i}}(y) dy \\ & = \lambda \frac{\mu_i}{\mu} P_{c-1} e^{-\mu x} + \lambda \frac{\mu_i}{\mu} \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy \\ & \quad + \lambda \frac{\mu_i}{\mu} \int_{y=x}^{\infty} \overline{R}(y) \left(f(y) - f_{\bar{i}}(y) \right) dy, i = 1, ..., c. \end{aligned} \tag{4.100}$$

On the left side of (4.100) $f_{\bar{i}}(x)$ is the SP exit rate from $((x,\infty),\bar{i})$ due to downcrossings of level x; $\lambda(1 - \frac{\mu_i}{\mu}) \int_{y=x}^{\infty} \overline{R}(y) f_{\bar{i}}(y) dy$ is the SP rate of jumps out of $((x,\infty),\bar{i})$ to other sheets. On the right side, the first two terms are SP entrance rates into $((x,\infty),\bar{i})$ due to jumps starting at level 0 and jumps starting in $(0,x)$ on any sheet, respectively (recall $f(y) = \sum_{i=1}^{c} f_{\bar{i}}(y)$). The third term is the SP entrance rate into $((x,\infty),\bar{i})$ due to jumps starting in $\cup_{j\neq i}((x,\infty),\bar{j})$. Rate balance of SP exits and entrances of $((x,\infty),\bar{i})$ yields (4.100).

We obtain the solution of (4.100) directly using the following Proposition.

Proposition 4.1 *The partial pdf is given by*

$$f_{\bar{i}}(x) = \frac{\mu_i}{\mu} f(x), x > 0, i = 1, ..., c. \tag{4.101}$$

Proof. Substitute $f_{\bar{i}}(x) = \frac{\mu_i}{\mu} f(x)$ in (4.100). The proposition is correct if and only if the following is an identity:

$$\begin{aligned} & \frac{\mu_i}{\mu} f(x) + \lambda \int_{y=x}^{\infty} \frac{\mu_i}{\mu} \overline{R}(y) f(y)) dy \\ & = \lambda \frac{\mu_i}{\mu} P_{c-1} e^{-\mu x} + \lambda \frac{\mu_i}{\mu} \int_{y=0}^{x} e^{-\mu(x-y)} \overline{R}(y) f(y) dy \\ & \quad + \lambda \frac{\mu_i}{\mu} \int_{y=x}^{\infty} \overline{R}(y) f(y) dy, \end{aligned} \tag{4.102}$$

if and only if

$$f(x) = \lambda P_{c-1} e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)\overline{R}(y)} f(y) dy \tag{4.103}$$

is an identity. Equation (4.103) is identical to equation (4.98). Hence the proposition is true. ∎

Exponential Staying Function

Consider an exponential staying function, $\overline{R}(x) = e^{-rx}$,$r > 0$,$x \geq 0$. The total pdf $f(x)$ is obtained by substituting e^{-ry} for $\overline{R}(y)$ in (4.99).

Then, substituting (4.99) into (4.101) gives

$$f_{\bar{i}}(x) = \lambda \frac{\mu_i}{\mu} e^{\frac{\lambda}{r}} P_{c-1} e^{-\mu x - \frac{\lambda}{r} e^{-rx}}, x > 0, i = 1, ..., c. \qquad (4.104)$$

4.12.7 Equations for Zero-wait Probabilities

Assume that the zero-wait server assignment policy is: arrivals that find k free servers, $1 \leq k \leq c$, get served by a particular free server with probability $\frac{1}{k}$. (Other policies are viable, e.g., the arrival gets served by the lowest-numbered available server, or by the fastest-available service rate, etc.) Balancing the SP exit and entrance rates for the zero-wait states we obtain the equations (notation explained below)

$$\begin{aligned}
(\lambda + \mu - \mu_i) P_{c-1,\bar{i}} &= f_{\bar{i}}(0) + \tfrac{\lambda}{2} \textstyle\sum_{j \in \boldsymbol{J}_i}^{c} P_{c-2,\overline{ij}}, i = 1, ..., c, \\
(\lambda + \mu - \mu_i - \mu_j) P_{c-2,\overline{ij}} &= \mu_j P_{c-1,\bar{i}} + \mu_i P_{c-1,\bar{j}} \\
&\quad + \tfrac{\lambda}{3} \textstyle\sum_{k \in J_{ij}}^{c} P_{c-3,\overline{ijk}}, j = 1, ..., c, \\
&\cdots \\
(\lambda + \mu_i) P_{1,i} &= \textstyle\sum_{k \neq i=1}^{c} \mu_k P_{2,ik} + \dfrac{\lambda}{c} P_{00},\ i = 1, ..., c, \\
\lambda P_{00} &= \textstyle\sum_{i=1}^{c} \mu_i P_{1,i}.
\end{aligned} \qquad (4.105)$$

We explain the notation in the sets of equations in (4.105). In the first set of c equations, the index j of the sum takes values in $\boldsymbol{J}_i = \{j | j = 0, ..., c, j \neq i\}$. In the second set of $\binom{c}{2}$ equations, the index k of the sum takes values in $\boldsymbol{J}_{ij} = \{k | k = 0, ..., c, k \neq i, k \neq j\}$. The row of dots "$\cdots$" indicates similar balance equations for $P_{c-3,\overline{ijk}}$ to $P_{2,\cdot}$. In the second last equation, $P_{2,ik}$ denotes the probability of two units in the system having service rates μ_i, μ_k.

We solve equations (4.105) in Subsection 4.12.8 below for a model with $c = 2$, in order to convey some characteristics of the solution.

4.12.8 Solution for M/M/2 with Reneging

When $c = 2$, there are two sheets corresponding to configurations $\overline{1} \equiv (01)$, $\overline{2} \equiv (10)$. Applying (4.104), the joint mixed (partial) pdf's of wait are

$$f_{\bar{i}}(x) = \lambda \frac{\mu_i}{\mu} e^{\frac{\lambda}{r}} P_1 e^{-\mu x - \frac{\lambda}{r} e^{-rx}}, x > 0, i = 1, 2.$$

The marginal ("total") pdf of wait is

$$f(x) = f_{\bar{1}}(x) + f_{\bar{2}}(x) = \lambda e^{\frac{\lambda}{r}} P_1 e^{-\mu x - \frac{\lambda}{r} e^{-rx}}, x > 0.$$

The zero-wait probabilities are $P_{1,\bar{i}}, i = 1, 2$, and P_{00}; and

$$\begin{aligned} P_1 &= P_{1,\bar{2}} + P_{1,\bar{1}} \\ P_0 &= P_{00} + P_{1,\bar{2}} + P_{1,\bar{1}} \\ &= P_{00} + P_1. \end{aligned}$$

The balance equations for the zero-wait probabilities are

$$\begin{aligned} (\lambda + \mu_1) P_{1,\bar{2}} &= \frac{\lambda}{2} P_{00} + f_{\bar{2}}(0) \\ (\lambda + \mu_2) P_{1,\bar{1}} &= \frac{\lambda}{2} P_{00} + f_{\bar{1}}(0) \\ \lambda P_{00} &= \mu_1 P_{1,\bar{2}} + \mu_2 P_{1,\bar{1}}. \end{aligned} \tag{4.106}$$

Substituting for $f_{\bar{2}}(0)$, $f_{\bar{1}}(0)$ in (4.106) we rewrite the equations as

$$\begin{aligned} (\lambda + \mu_1) P_{1,\bar{2}} &= \frac{\lambda}{2} P_{00} + \lambda \frac{\mu_2}{\mu} P_1, \\ (\lambda + \mu_2) P_{1,\bar{1}} &= \frac{\lambda}{2} P_{00} + \lambda \frac{\mu_1}{\mu} P_1, \\ \lambda P_{00} &= \mu_1 P_{1,\bar{2}} + \mu_2 P_{1,\bar{1}}. \end{aligned} \tag{4.107}$$

The solution of (4.107) in terms of P_{00} is

$$\begin{aligned} P_{1,\bar{1}} &= \frac{\lambda}{2\mu_2} P_{00}, \\ P_{1,\bar{2}} &= \frac{\lambda}{2\mu_1} P_{00}, \\ P_1 &= \frac{\lambda(\mu_1 + \mu_2)}{2\mu_1\mu_2} P_{00} = \frac{\lambda\mu}{2\mu_1\mu_2} P_{00}. \end{aligned} \tag{4.108}$$

The normalizing condition

$$P_{00} + P_1 + \int_{x=0}^{\infty} f(x)dx = 1$$

yields

$$P_{00} = \left(1 + \frac{\lambda(\mu_1 + \mu_2)}{2\mu_1\mu_2} + \frac{\lambda(\mu_1 + \mu_2)}{2\mu_1\mu_2} \lambda e^{\frac{\lambda}{r}} \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx\right)^{-1} \tag{4.109}$$

Example 4.9 *We present a numerical example for the M/M/2 queue with reneging (see Fig. 4.9). Let*

$$\lambda = 5.2,\ \mu_1 = 2.4,\ \mu_2 = 1.1,\ \mu = \mu_1 + \mu_2 = 3.5,\ r = 2.1.$$

Then

$$\int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx = 0.074741,$$

and

$$\begin{aligned}
P_{00} &= 0.049059, P_{1\overline{1}} = 0.115958, \\
P_{1\overline{2}} &= 0.053147, P_1 = 0.169105, \\
P_{00} + P_1 &= 0.218164, \\
F(\infty) &= P_{00} + P_1 + \lambda e^{\frac{\lambda}{r}} P_1 \int_{x=0}^{\infty} e^{-\mu x - \frac{\lambda}{r} e^{-rx}} dx \\
&= 0.218164 + 0.781836 = 1.0, \\
f(x) &= \lambda e^{\frac{\lambda}{r}} P_1 e^{-\mu x - \frac{\lambda}{r} e^{-rx}} = 10.461 \cdot e^{-3.5x - 2.476 e^{-2.1x}}, \\
f_{\overline{1}}(x) &= \frac{\mu_1}{\mu} f(x) = 7.173 \cdot e^{-3.5x - 2.476 e^{-2.1x}}, \\
f_{\overline{2}}(x) &= \frac{\mu_2}{\mu} f(x) = 3.288 \cdot e^{-3.5x - 2.476 e^{-2.1x}}.
\end{aligned}$$

Remark 4.19 *In the $M/M_i/c$ model with reneging from the waiting line (or waiting and balking at service) allowed we can generalize the staying function $\overline{R}(x), x \geq 0$. For example, $\overline{R}(x)$ may depend on the server that would be occupied by an arrival, i.e., on the system configuration at the arrival instant. We may then use the notation $\overline{R}_{\overline{i}}(x)$. In this way $\overline{R}_{\overline{i}}(x)$ may depend not only on customer required wait for service, but also on customer attraction or aversion to the "target" server. A natural question arises. Can this model be modified to study attraction or aversion in natural processes such as: electrically charged particles approaching an electrically charged environment, asteroids approaching a planet, particles adhering or falling away from a surface; laser pulses affecting cells containing certain chemicals in biological or medical applications, etc.?*

4.12.9 Stability Condition

Theorem 3.8 also applies in the $M/M_i/c$ environment as follows.

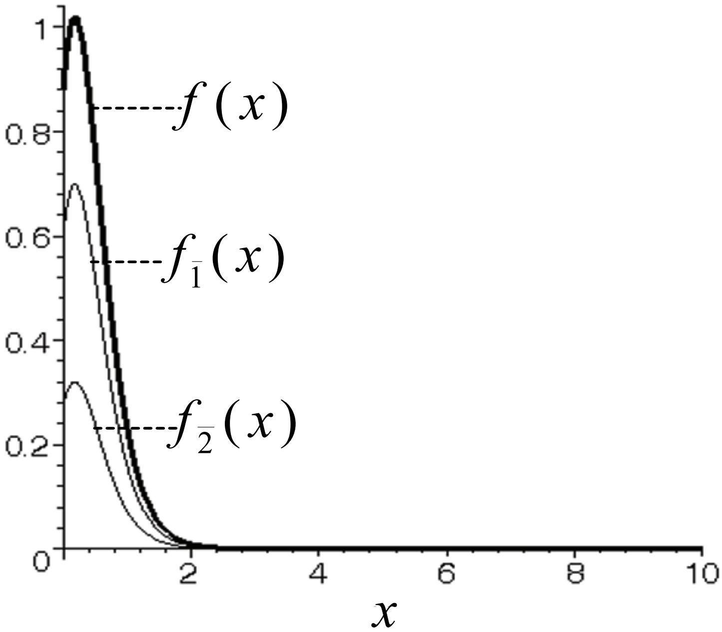

Figure 4.9: Plot of $f(x)$, $f_{\overline{1}}(x)$, $f_{\overline{2}}(x)$ in Example 4.9.

Theorem 4.8 *Consider the $M_\lambda/M_i/c$ ($c \geq 2$) queue with heterogeneous servers having rates $\mu_1, .., \mu_c$ in which reneging is allowed before service starts. Let the staying function be $\overline{R}(x), x \geq 0$, where $\overline{R}(x)$ is monotone non-increasing, $\overline{R}(0) = 1$ (no balking upon arrival), and $0 \leq \overline{R}(x) \leq 1$. Let $L = \lim_{x\to\infty} \overline{R}(x)$. A necessary and sufficient condition for stability is*

$$\lambda < \frac{\mu}{L} \text{ if } 0 < L \leq 1,$$
$$\lambda < \infty \text{ if } L = 0,$$

where $\mu = \sum_{i=1}^{c} \mu_i$.

Proof. The proof is similar to that of Theorem 3.8. The alternative proof given there, Remark 3.31 and Fig. 3.24 also apply for the M/M$_i$/c queue with reneging, upon substituting $\mu = \sum_{i=1}^{c} \mu_i$. ■

4.13 Discussion

We can use LC to analyze a vast array of additional M/M/c models. We mention only a few other examples.

LC has been applied to M/M/c queues in which customers receive simultaneous service from a random number of servers. The original source for such queueing models is [61]. An LC analysis is given in [28].

LC has been applied to M/M/c with bounded system time (wait + service). An arrival balks upon arrival if its system time would exceed an upper bound K [38]. This generalizes variant 2 of the M/G/1 model discussed above in Section 3.14. It is straightforward to apply LC to analyze a model analogous to variant 1 in Section 3.14. In that model customers renege from *service* if their *age* in the system is K. Similar remarks apply to M/M/c where the actual waits are bounded by K. In that case the workload can exceed K. We can develop an expression for the tail of the steady-state pdf of workload, from the integral equation for the pdf of workload.

LC can be used to analyze a variety of M/M/c queues with server vacations. It can be used to analyze M/M/c queues with priorities.

CHAPTER 5

G/M/c QUEUES

This chapter applies a level-crossing approach (SPLC, abbreviated LC) to derive the steady-state pdf of the virtual wait and the actual wait (arrival-point wait) in single-server G/M/1, and in multiple-server G/M/c queues. Section 5.1 treats G/M/1 and Section 5.2 treats G/M/c ($c = 2, 3, ...$). It is assumed that arrivals occur according to a renewal process and service times are exponentially distributed.

We will not derive transient distributions in this chapter. However, for G/M/c ($c = 1, 2, ...$), we could use LC to derive the transient distribution of *extended age*, which is related to the virtual wait (Subsection 5.1.1). We would then apply techniques similar to those utilized in sections 3.2, 4.3, Subsection 6.2.5, Section 10.9 and other sections of Chapter 10. Those analyses provide background for deriving transient distributions using LC in G/M/c queues, as well as in a great variety of stochastic models. (The extended age is utilized in [15].)

5.1 Single-server G/M/1 Queue

We analyze the single-server G/M/1 queue in steady state. Arrivals occur according to a renewal process. For the common inter-arrival time denote the cdf, complementary cdf, and pdf respectively by $A(x), x > 0$, $\overline{A}(x) = 1 - A(x), x \geq 0$ and $a(x) = \frac{d}{dx}A(x)$ wherever the derivative exists. Assume the service time of each customer has an exponential distribution with mean $\frac{1}{\mu}$ (denoted by E_μ). Using LC we derive the steady-state pdf and cdf of the *virtual* wait, the steady-state pdf and cdf of the *actual* (arrival-point) wait just before arrival instants, expressions for the expected busy and idle periods, and related results.

DOI: 10.1007/978-0-387-09421-2_5,

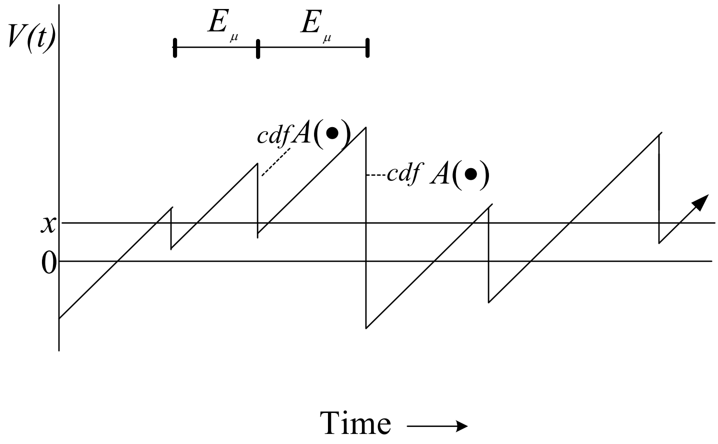

Figure 5.1: Sample path of extended age process $\{V(t)\}$ for G/M/1 queue. Inter-arrival times have cdf $A(\cdot)$ (cdf of downward jump sizes). Service times are $\underset{dist}{=} \mathrm{E}_\mu$. Slope is $\frac{dV(t)}{dt} = +1$.

5.1.1 Virtual Wait and Extended Age Processes

Let $\{W(t), t \geq 0\}$ denote the virtual wait process having state space $\boldsymbol{S} = [0, \infty)$ (e.g., similar to Fig. 3.4).

We consider the "*extended age*" process $\{V(t), t \geq 0\}$ having state space $\boldsymbol{S} = (-\infty, \infty)$, defined as follows. For $t > 0$,

$$V(t) = \begin{cases} \textit{age of customer in service at } t \text{ if } V(t) \geq 0, \\ -\ \textit{time from } t \textit{ until next arrival instant} \text{ if } V(t) < 0. \end{cases} \tag{5.1}$$

In (5.1) "age" means "time spent in the system" measured from the arrival instant. A sample path of $\{V(t)\}$ is depicted in Fig. 5.1. Extended-age sample-path jumps are *downward* in direction. All jumps start at positive levels. (All virtual-wait jumps are upward.)

5.1.2 Duality Between Extended Age and Virtual Wait

Consider a sample path of $\{V(t), t \geq 0\}$. Assume $V(t) \geq 0$. There is a one-to-one correspondence between the peaks (relative maxima) of $\{V(t)\}$ and peaks of $\{W(t)\}$, as well as between troughs (relative minima or infima) of $\{V(t)\}$ and troughs of $\{W(t)\}$. Corresponding peaks and troughs have equal ordinates and occur in the same time order in both processes (Fig. 5.2).

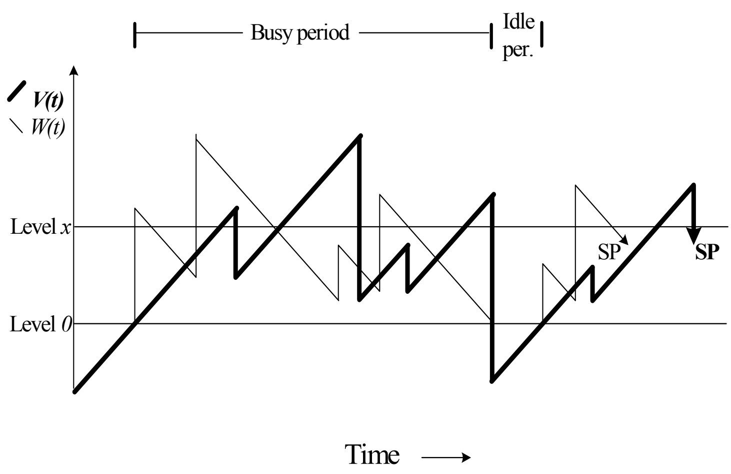

Figure 5.2: Sample path of extended age process "$\nearrow$" compared with sample path of virtual wait process "$\searrow$" for G/M/1 queue. Illustarates duality properties. Corresponding peaks and corresponding troughs have equal ordinates and the same time order. Busy periods, idle periods, and busy cycles ar equal.

The extended age process has slope $+1$ between SP downward jumps. The virtual wait has slope -1 between SP upward jumps within a busy period; the slope is 0 within an idle period. Busy periods are identical in both processes. These properties guarantee that the proportion of time that the SP spends in any state-space interval, is the same in both processes (see Proposition 5.1 below).

The sojourn time of $\{V(t)\}$ below level 0 is identical to an idle period in the $\{W(t)\}$ process (see Remark 5.2). Busy cycles are identical in both processes (Fig. 5.2).

The stability condition is $\dfrac{1}{E(\text{inter-arrival time})\cdot\mu} < 1$ (e.g., [63], p. 251). Intuitively, the expected number of arrivals in a service time is < 1. (See Proposition 5.4 below.)

Denote the steady-state cdf of the extended age by

$$F(x) = \lim_{t\to\infty} P(V(t) \le x), -\infty < x < \infty,$$

having pdf

$$\begin{aligned} f(x) &= \tfrac{dF(x)}{dx}, x \ge 0; \\ h(x) &= \tfrac{dF(x)}{dx}, x < 0, \end{aligned} \tag{5.2}$$

wherever the derivatives exist. The probability of an empty system is

$$P_0 = F(0) = \int_{y=-\infty}^{0} h(y)dy. \tag{5.3}$$

Then

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

Proposition 5.1 *The* ***steady-state*** *cdf of the* extended age *process* $\{V(t)\}$ *and of the virtual wait* $\{W(t)\}$ *as* $t \to \infty$*, are identical. That is,*

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

Proof. There is a one-to-one correspondence between sample paths of $\{V(t)\}$ and $\{W(t)\}$ because of the duality properties discussed above (see Fig. 5.2). The *proportion* of time spent in every state-space interval is the same in corresponding sample paths for every $\omega \in \Omega$, where Ω is the sample space of the "underlying experiment" and ω is a possible outcome.

For $\{V(t)\}$ a sojourn time below level 0 is the same as an idle period in $\{W(t)\}$. Thus $F(0) = P_0 = \lim_{t\to\infty} P_0(t)$ is the same for both processes ($P_0(t)$ is the probability of a zero wait at time t). ■

We employ $\{V(t)\}$ when analyzing G/M$_\mu$/1 using LC, because SP downward jumps occur at end-of-service instants at *Poisson rate* μ.

Remark 5.1 *We emphasize that the* transient probability distributions *of* $V(t)$ *and* $W(t)$ *are* ***not*** *equal. Proposition 5.1 holds for* steady-state *distributions only.*

Remark 5.2 *We may also define an* ***"extended virtual wait" process*** $\{W(t)\}$ *with state space* $(-\infty, +\infty)$*. If* $W(t) > 0$*, then* $W(t)$ *is the usual virtual wait. If* $W(t) < 0$*,* $-W(t)$ *is the time since the last departure of the immediately previous busy period. For the extended* ***virtual wait,*** *the slope is* -1 *between (upward) jumps. Sojourn times below level* 0 *are equal to idle periods. If arrivals are Poisson, an integral equation for the pdf of* $\{W(t)\}$ *when* $W(t) < 0$ *can be obtained by applying LC. All results for the usual virtual wait can be derived using the extended virtual wait. If arrivals are Poisson at rate* λ *the expected sojourn time below level* 0 *is* $\frac{1}{\lambda} = E($*idle period*$)$.

5.1.3 Equation for Steady-State PDF of Age

By Proposition 5.1 the steady-state pdf of the age process $f(x), x > 0$, is the same as the steady-state pdf of the virtual wait process. Thus, for G/M/1 we will obtain the steady-state pdf of $\{W(t)\}$ by deriving the steady-state pdf of $\{V(t)\}$.

Consider a sample path of $\{V(t)\}$ (Fig. 5.1). Fix level $x > 0$ in the state space. The SP *upcrossing* rate of x is

$$\lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = f(x), \tag{5.4}$$

(proved similarly as for the downcrossing rate in M/G/1, e.g., Theorem 1.1).

The SP *downcrossing* rate of x is

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \underset{(a.s.)}{=} \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \mu \int_{y=x}^{\infty} \overline{A}(y-x) f(y)dy, \tag{5.5}$$

(proved as for the upcrossing rate in M/G/1).

We give an LC interpretation of right-most term of (5.5). The SP rate of downward jumps staring from level $y > 0$ is the rate at which *service times end* when customers have been in the system for a time y, namely $\mu f(y)dy$. If $y > x$,

$$\begin{aligned} &P(\text{downward jump size} > y-x) \\ &\qquad = P(\text{inter-arrival time} > y-x) = \overline{A}(y-x). \end{aligned}$$

Summing over all $y > x$ gives the right-most term of (5.5).

The principle of rate balance across level x,

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

gives an integral equation for $f(x)$,

$$f(x) = \mu \int_{y=x}^{\infty} \overline{A}(y-x) f(y)dy. \tag{5.6}$$

5.1.4 Alternative Form of Equation for PDF of Age

An alternative form of integral equation (5.6) is

$$f(x) = \mu(1-F(x)) - \mu \int_{y=x}^{\infty} A(y-x) f(y)dy, x > 0. \tag{5.7}$$

The LC interpretation of (5.7) is as follows. The left side is the SP upcrossing rate of level x. On the rite side, the first term is the rate of service completions which generate SP downward jumps that start above level x. The second term is the rate of service completions that generate SP downward jumps that start above level x *and end above* level x. Thus the right side is the SP downcrossing rate of level x.

Note the similarity of the alternative LC equation (5.7) for G/M/1, and the alternative forms (3.35) for the M/G/1 queue, and (6.19) for the M/G/r(·) dam in Chapter 6.

5.1.5 PDF and CDF of Virtual Wait Geometrically

We demonstrate *geometrically* using LC, that the steady-state pdf of $\{V(t)\}$ (therefore of $\{W(t)\}$), as $t \to \infty$, has an *exponential form* over the state-space interval $(0, \infty)$, and an atom at 0.

Let $\mathcal{B}$ denote a busy period. Consider a sample path of $\{V(t)\}$. Due to the memoryless property of the service times, an SP sojourn time *above* an arbitrary level $x \geq 0$ is distributed the same as $\mathcal{B}$ *independent of* x (Figs. 5.1 and 5.2).

Thus the *proportion* of time spent above $x \geq 0$ is

$$\begin{aligned}\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x)) \cdot E(\mathcal{B})}{t} &= \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} \cdot E(\mathcal{B}) \\ &= f(x) \cdot E(\mathcal{B}) = 1 - F(x), \qquad (5.8)\end{aligned}$$

by (5.4), and the definition of $1 - F(x)$.

Equation (5.8) is equivalent to a differential equation

$$\begin{aligned}\frac{\frac{d}{dx}(1 - F(x))}{1 - F(x)} &= -\frac{1}{E(\mathcal{B})}, \\ \frac{d}{dx} \ln(1 - F(x)) &= -\frac{1}{E(\mathcal{B})},\end{aligned}$$

with solution

$$\begin{aligned}F(x) &= 1 - (1 - P_0)e^{-\frac{1}{E(\mathcal{B})}x}, x \geq 0, \\ f(x) &= \frac{1 - P_0}{E(\mathcal{B})} e^{-\frac{1}{E(\mathcal{B})}x}, x > 0,\end{aligned} \qquad (5.9)$$

where $F(0) \equiv P_0$.

From (5.9) $f(x)$ has the exponential form

$$f(x) = Ke^{-\gamma x}, x \geq 0 \qquad (5.10)$$

where

$$K = \frac{1 - P_0}{E(\mathcal{B})}, \ \gamma = \frac{1}{E(\mathcal{B})}. \tag{5.11}$$

Remark 5.3 *As a mild confirmation of the above results suppose the* $G/M_\mu/1$ *queue were an* $M_\lambda/M_\mu/1$ *queue. Then, in (5.10) we would have* $E(\mathcal{B}) = \frac{1}{\mu - \lambda}$. *Thus* $\gamma = \mu - \lambda$ *and*

$$\begin{aligned} K &= (1 - P_0)\gamma = (1 - P_0)(\mu - \lambda) \\ &= \left(1 - \left(1 - \frac{\lambda}{\mu}\right)\right)(\mu - \lambda) \\ &= \lambda\left(1 - \frac{\lambda}{\mu}\right) = \lambda P_0, \end{aligned}$$

giving $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0$ ✓. *This checks with the steady-state pdf of wait in M/M/1 (e.g., (3.86)).*

Substituting from (5.10) into (5.6) and cancelling K gives an equation for γ,

$$e^{-\gamma x} = \mu \int_{y=x}^{\infty} \overline{A}(x - y) e^{-\gamma y} dy.$$

Substituting $z = x - y$ results in

$$\int_{z=0}^{\infty} \overline{A}(z) e^{-\gamma z} dz = \frac{1}{\mu}. \tag{5.12}$$

Equation (5.12) for γ is a *fundamental* G/M/1 equation. The left side of (5.12) is the Laplace transform of $\overline{A}(z)$ evaluated with parameter γ.

Let $A^*(\gamma)$ denote the Laplace Stieltjes transform of $A(\cdot)$. Integrating (5.12) by parts gives

$$A^*(\gamma) = 1 - \frac{\gamma}{\mu}. \tag{5.13}$$

Thus γ is the solution of (5.12), or equivalently of (5.13). Some forms of $\overline{A}(\cdot)$ allow for an analytical solution for γ. Generally, however, γ is computed by numerical methods (e.g., by Newton's method or using computational software such as Maple).

Value of P_0

Consider a sample path of $\{V(t)\}$ on the state-space interval $(-\infty, 0)$, and fix level $x \in (-\infty, 0)$. The SP *upcrossing* rate of level x is equal to $h(x)$ (proved as for the downcrossing rate in M/G/1). The SP *down-crossing* rate of level x is

$$\mu \int_{y=0}^{\infty} \overline{A}(y-x)f(y)dy = \mu \int_{y=0}^{\infty} \overline{A}(y-x)Ke^{-\gamma y}dy,$$

since all downward jumps originate at end-of-service instants when the SP is in state-space set $(0,\infty)$. Rate balance across level x gives

$$h(x) = \mu \int_{y=0}^{\infty} \overline{A}(y-x)Ke^{-\gamma y}dy, x < 0. \tag{5.14}$$

Invoking (5.14) and (5.3) leads to

$$P_0 = \int_{x=-\infty}^{0} h(x)dx = K \int_{x=-\infty}^{0} \mu \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dydx.$$

Making the transformation $u = -x$, gives

$$P_0 = K \int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu.$$

Thus

$$P_0 = \frac{K}{C_\gamma}, \text{ or } K = P_0 C_\gamma \tag{5.15}$$

where

$$C_\gamma = \left(\int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y}dydu \right)^{-1}. \tag{5.16}$$

Note that $C_\gamma > 0$.

We evaluate P_0 from the normalizing condition and (5.15). Thus

$$\begin{aligned} P_0 + K \int_{y=0}^{\infty} e^{-\gamma x}dx &= 1, \\ P_0 + C_\gamma P_0 \int_{y=0}^{\infty} e^{-\gamma x}dx &= 1. \end{aligned}$$

These equations yield

$$P_0 = 1 - \frac{K}{\gamma}. \tag{5.17}$$

$$= \frac{\gamma}{\gamma + C_\gamma}. \tag{5.18}$$

From (5.15)

$$K = \frac{\gamma \cdot C_\gamma}{\gamma + C_\gamma}, \tag{5.19}$$

and $K < \gamma$.

Due to exponentially distributed service times, instants of SP egress from level 0 above, are regenerative points of $\{V(t)\}$ initiating busy cycles (see 2.4.9 for definitions of SP egresses). Thus, steady-state properties over busy cycles recapitulate limiting properties over the time axis as $t \to \infty$.

Let C represent a busy *cycle* and $\mathcal{I}$ an idle period. Then

$$\mathcal{C} = \mathcal{B} + \mathcal{I}$$

and

$$P_0 = \frac{E(\mathcal{I})}{E(\mathcal{C})} = \frac{E(\mathcal{I})}{E(\mathcal{B}) + E(\mathcal{I})}.$$

From (5.18)

$$\frac{E(\mathcal{I})}{E(\mathcal{B}) + E(\mathcal{I})} = \frac{\gamma}{\gamma + C_\gamma} = \frac{\frac{1}{C_\gamma}}{\frac{1}{\gamma} + \frac{1}{C_\gamma}}.$$

From (5.11) $E(\mathcal{B}) = \frac{1}{\gamma}$. Thus from (5.16)

$$E(\mathcal{I}) = \frac{1}{C_\gamma} = \int_{u=0}^{\infty} \mu \int_{y=0}^{\infty} \overline{A}(y+u) e^{-\gamma y} dy du. \tag{5.20}$$

5.1.6 PDF of Actual Wait

For G/M/1, generally the steady-state pdf of the actual wait (arrival-point wait) is not equal to the pdf of the virtual wait. In particular, these pdf's are equal when the arrival stream is Poisson. We can utilize results in subsections 5.1.1 - 5.1.5 to determine the pdf of the actual wait.

Form of PDF of Actual Wait

We use LC concepts to derive the *form* of the pdf of actual wait. The subscript "ι" (Greek iota) is used to signify actual wait. Let the steady-state cdf of actual wait be $F_\iota(x) = P(\text{actual wait} \leq x), x \geq 0$; and let the pdf be $\frac{d}{dx} F_\iota(x) = f_\iota(x), x > 0$. Recall that γ is the solution of (5.12).

Proposition 5.2 *The form of the steady-state pdf of* ***actual*** *wait is*

$$f_\iota(x) = K_\iota e^{-\gamma x}, x > 0, \tag{5.21}$$

where K_ι is a positive number.

Proof. The *proportion* of actual waits that are $> x$ is

$$1 - F_\iota(x) = \frac{\mu(1 - F(x)) - f(x)}{\mu\,(1 - F(0))}, \tag{5.22}$$

where $F\,(x)$, $f\,(x)$ denote the cdf and pdf respectively of the *virtual* wait.

We now explain (5.22). Consider the numerator. The term $\mu(1 - F(x))$ is the departure rate of customers that have been in the system $> x$ time units. Each such departure generates an SP *downward* jump of a sample path of the $\{V(t)\}$ process. The term $f(x)$ is the rate at which SP jumps start above x and end below (or at) x ($f(x)$ is the *downcrossing* rate, as well as the upcrossing rate, of level x). That is, $f(x)$ is the rate at which *next* actual waits are $\leq x$. Thus the numerator is the rate at which *next* actual waits are $> x$. The denominator $\mu\,(1 - F(0))$ of (5.22) is the *total* departure rate, which is the total rate of downward jumps.

From (5.9) $1 - F(x) = c_1 e^{-\gamma x}$, where c_1 is a positive constant, and from (5.10) $f(x) = Ke^{-\gamma x}$. We substitute these exponential terms on the right side of (5.22).

Then, taking $\frac{d}{dx}$ on both sides of (5.22) gives (5.21) for some positive constant K_ι. ■

Proposition 5.2 implies that the *form* of the pdf of actual wait $f_\iota(x)$, is the same as the *form* of the pdf of the virtual wait $f(x)$. Generally, the values of K_ι and K differ, except when the arrival stream is Poisson. The exponent γ is common to both $f_\iota(x)$ and $f(x)$.

Specification of PDF and CDF of Actual Wait

Denote the probability that an arrival waits zero by $P_{0\iota}$.

Proposition 5.3 *For the G/M/1 queue with service rate μ, the probability that an arrival waits zero time for service is*

$$P_{0\iota} = \frac{\gamma}{\mu} = \frac{K_\iota}{\mu - \gamma} \tag{5.23}$$

where

$$K_\iota = \gamma\left(1 - \frac{\gamma}{\mu}\right). \tag{5.24}$$

The steady-state pdf and cdf of the arrival-point wait are respectively

$$f_\iota(x) \quad = \quad K_\iota e^{-\gamma x} = \gamma(1 - \frac{\gamma}{\mu})e^{-\gamma x}, x > 0, \tag{5.25}$$

$$F_\iota(x) \quad = \quad 1 - (1 - \frac{\gamma}{\mu})e^{-\gamma x}, x \geq 0. \tag{5.26}$$

Proof. Probability $P_{0\iota}$ is the proportion of *arrivals* that wait zero before they start service. Thus

$$P_{0\iota} = \frac{f(0)}{f(0) + \mu \int_{y=0}^{\infty} A(y) f(y) dy}, \tag{5.27}$$

where $f(x)$ is the pdf of the virtual wait given in (5.10). We now explain (5.27). The term $f(0)$ is the *rate of arrivals* to the system that wait 0 (upcrossing rate of level 0). The term $\mu \int_{y=0}^{\infty} A(y) f(y) dy$ is the *rate* of arrivals to the system that wait a positive time, i.e., the rate at which SP downward jumps *start* and *end above* 0 (see Fig. 5.1). For such downward jumps, the end state-space level is the *actual wait* of the *next* arrival. Also, the rate at which *next* arrivals wait > 0 is the same as the *overall* rate at which arrivals wait > 0.

Substituting from (5.10) into (5.27) gives

$$P_{0\iota} = \frac{K}{K + (\frac{\mu K}{\gamma} - K)} = \frac{\gamma}{\mu}.$$

We ascertain K_ι from the normalizing condition for the arrival-point pdf,

$$\begin{aligned} P_{0\iota} + \int_{y=0}^{\infty} f_\iota(y) dy &= 1, \\ \frac{\gamma}{\mu} + \frac{K_\iota}{\gamma} &= 1. \end{aligned}$$

Thus we obtain (5.23), (5.24), (5.25) and (5.26) (from $F_\iota(x) = P_{0\iota} + \int_{y=0}^{x} f_\iota(y) dy$). ■

Remark 5.4 *Formula (5.23) for $P_{0\iota}$ matches the result derived later in formula (8.23) via the* **embedded** LC method. *The embedded LC result* ***is*** *indeed the value of $P_{0\iota}$, since it is the steady-state pdf of the actual wait W_n as $n \to \infty$. This match validates the standard "continuous" LC approach utilized in this section. In many models, it is easier to apply standard LC than* embedded LC. *We note, however, that embedded LC is useful in itself, for checking results obtained by other means, analyzing new models, and combining with continuous LC to obtain new results.*

Remark 5.5 *$P_{0\iota}$ and $f_\iota(x)$ in (5.23) and (5.25) correspond to results obtained by a different technique in [63], pages 250-254. In the present section, the constant $\gamma \equiv \mu(1 - r_0)$ where r_0 is the solution of $z = A^*(\mu(1 - z)), z \in (0, 1)$, in [63].*

5.1.7 Stability Condition for G/M/1

We develop the stability condition directly from equation (5.12). Stability occurs iff the solution of (5.12) for γ is positive and finite. That is, iff the "steady-state" pdf's $f(x) = Ke^{-\gamma x}$ (virtual wait) and $f_\iota(x) = K_\iota e^{-\gamma x}$ (arrival-point wait) exist. These pdf's exist provided γ is positive and finite, in which case K and K_ι are also positive and finite by (5.19) and (5.24) respectively.

Denote the expected inter-arrival time by $\frac{1}{\boldsymbol{a}}$ and the expected service time by $\frac{1}{\mu}$.

Proposition 5.4 *The G/M/1 queue is stable if and only if* $\boldsymbol{a} < \mu$.

Proof. The queue is stable iff the expected busy period $\frac{1}{\gamma}$ is positive and finite iff γ is positive and finite. Consider equation (5.12). Suppose that a positive finite number γ exists such that

$$\frac{1}{\mu} = \int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy.$$

Since $0 < e^{-\gamma y} < 1$ for all $y > 0$,

$$\begin{aligned} \frac{1}{\mu} &< \int_{y=0}^{\infty} \overline{A}(y)dy = \frac{1}{\boldsymbol{a}} \\ &\implies \boldsymbol{a} < \mu. \end{aligned}$$

Hence $\boldsymbol{a} < \mu$ is a *necessary* condition for stability.

Conversely, suppose $\boldsymbol{a} < \mu$. Then $\frac{1}{\mu} < \frac{1}{\boldsymbol{a}}$ and

$$\frac{1}{\mu} < \frac{1}{\boldsymbol{a}} = \int_{y=0}^{\infty} \overline{A}(y)dy.$$

Construct a function of γ, $\phi(\gamma) = \int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy, 0 < \gamma < \infty$. Then $\phi(\gamma) > 0$, $\lim_{\gamma\downarrow 0}\phi(\gamma) = \frac{1}{\boldsymbol{a}}$, $\lim_{\gamma\to\infty}\phi(\gamma) = 0$, $\phi'(\gamma) = -\gamma\phi(\gamma) < 0$, $\phi''(\gamma) = \gamma^2\phi(\gamma) > 0$. Thus $\phi(\gamma)$ is continuous, convex and monotone decreasing on $(0,\infty)$. Consequently $\phi(\gamma)$ assumes each value in its range $\left(0, \frac{1}{\boldsymbol{a}}\right)$. For each value of μ with the property $\frac{1}{\mu} \in (0, \frac{1}{\boldsymbol{a}})$, there is a unique value $\gamma \in (0,\infty)$ such that $\phi(\gamma) = \frac{1}{\mu}$. Hence for each such $\frac{1}{\mu}$ there exists exactly one *positive finite* root γ of equation (5.12). That is $\frac{1}{\mu} = \int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy$ has a unique positive finite solution for γ such that $\frac{1}{\mu} < \frac{1}{\boldsymbol{a}}$. Hence $\boldsymbol{a} < \mu$ is a *sufficient* condition for stability. ■

In conclusion $\boldsymbol{a} < \mu$ is a necessary and sufficient condition for stability.

5.1.8 Steady-state Distribution of System Time

Let W_q, S, σ denote respectively the steady-state actual wait before service, the service time, and the system time of a customer. Then $\sigma = W_q + S$. Note that the cdf of W_q is $P(W_q \leq x) = F_\iota(x), x \geq 0$ having pdf $f_\iota(x), x > 0$. Also $P(W_q = 0) = F_\iota(0) = P_{0\iota}$. Let $F_\sigma(x), x \geq 0$ denote the steady-state cdf of σ, and let $f_\sigma(x) = \frac{d}{dx}F_\sigma(x), x > 0$ be the pdf of σ, wherever the derivative exists. For the standard G/M/1 queue, S and W_q are independent.

Using the expressions in Proposition 5.3, the cdf of σ is the convolution

$$\begin{aligned} F_\sigma(x) &= P_{0\iota}P(S \leq x) + \int_{y=0}^{x} P(S \leq x - y) f_\iota(y) dy \\ &= \frac{\gamma}{\mu}(1 - e^{-\mu x}) \\ &\quad + \int_{y=0}^{x} \left(1 - e^{-\mu(x-y)}\right) \gamma \left(1 - \frac{\gamma}{\mu}\right) e^{-\gamma y} dy. \end{aligned} \tag{5.28}$$

The last integral in (5.28) is equal to

$$\frac{1}{\mu}(\mu e^{(\mu+\gamma)x} - \gamma e^{(\mu+\gamma)x} + \gamma e^{\gamma x} - \mu e^{\mu x})e^{-(\mu+\gamma)x}. \tag{5.29}$$

Summing (5.29) with $\frac{\gamma}{\mu}(1 - e^{-\mu x})$ simplifies to

$$F_\sigma(x) = 1 - e^{-\gamma x}, x \geq 0. \tag{5.30}$$

The pdf of σ is $\frac{d}{dx}F_\sigma(x)$, namely

$$f_\sigma(x) = \gamma e^{-\gamma x}, x > 0. \tag{5.31}$$

Remark 5.6 *The expressions for $F_\sigma(x)$ and $f_\sigma(x)$ in (5.30) and (5.31) for G/M/1 are analogous to those for the standard $M_\lambda/M_\mu/1$ queue given in (3.90), with $\gamma = \mu - \lambda$. Note that the coefficient of the exponent x in $F_\sigma(\cdot)$ is $\frac{-1}{E(\mathcal{B})}$ in both G/M/1 and M/M/1 ($\mathcal{B}$ = busy period).*

5.1.9 Arrival-point PMF of Number in System

This subsection derives the steady-state arrival-point pmf (probability mass function) of the number of units in the system. Let N_ι denote the number in the system *just before an arrival instant* in steady state. Then

$$P(N_\iota = 0) = P_{0\iota} = \frac{\gamma}{\mu}.$$

Let $P(N_\iota = n) = P_{n\iota}, n = 1, 2, \ldots$. Let d_n be the steady-state probability of n in the system *just after a departure instant.* Let $A^{(n)}(y)$ be the cdf of the n-fold convolution of the inter-arrival time evaluated at y.

Proposition 5.5 *For* $n = 1, 2, \ldots$,

$$\begin{aligned} P_{n\iota} &= d_n = \int_{y=0}^{\infty} \left(A^{(n)}(y) - A^{(n+1)}(y)\right) f_\sigma(y)dy \\ &= \gamma \int_{y=0}^{\infty} \left(A^{(n)}(y) - A^{(n+1)}(y)\right) e^{-\gamma y}dy. \end{aligned} \tag{5.32}$$

Proof. Let $N_A(t)$ be the number of arrivals in $(0, t)$ and let $\mathcal{S}_n$ be the time of the n^{th} arrival. A basic renewal equivalence relation is

$$N_A(t) \geq n \iff \mathcal{S}_n \leq t.$$

Thus

$$\begin{aligned} P(N_A(t) = n) &= P(N_A(t) \geq n) - P(N_A(t) \geq n + 1) \\ &= P(\mathcal{S}_n \leq t) - P(\mathcal{S}_{n+1} \leq t) \\ &= A^{(n)}(t) - A^{(n+1)}(t), t > 0 \end{aligned}$$

(see e.g., [74] or [91]). Also $d_n = P(n$ arrivals during a *system* time $\sigma)$. That is

$$\begin{aligned} d_n &= \int_{y=0}^{\infty} P(N_A(y) = n|\sigma = y)f_\sigma(y)dy \\ &= \int_{y=0}^{\infty} P(N_A(y) = n)f_\sigma(y)dy \\ &= \int_{y=0}^{\infty} \left(A^{(n)}(y) - A^{(n+1)}(y)\right) f_\sigma(y)dy. \end{aligned}$$

Since $d_n = P_{n\iota}$ (for any single-server queue), we obtain (5.32). ■

Compact Expression for PMF

Proposition 5.5 leads to a compact expression for $P_{n\iota}, n =$ 1,2,.... Integration by parts gives

$$\begin{aligned} \int_{y=0}^{\infty} A^{(n)}(y)e^{-\gamma y}dy &= \frac{1}{\gamma} \int_{y=0}^{\infty} a^{(n)}(y)e^{-\gamma y}dy \\ &= \frac{A^{n*}(\gamma)}{\gamma}, \end{aligned}$$

where $a^{(n)}(y)$ is the pdf of the n-fold convolution of inter-arrival times. Thus (5.32) becomes

$$P_{n\iota} = A^{n*}(\gamma) - A^{(n+1)*}(\gamma), n = 1, 2, ... \tag{5.33}$$

From Laplace-transform theory and (5.13)

$$A^{n*}(\gamma) = (A^*(\gamma))^n = \left(1 - \frac{\gamma}{\mu}\right)^n .$$

Substituting into (5.33) yields

$$\begin{aligned} P_{n\iota} &= \left(1 - \frac{\gamma}{\mu}\right)^n - \left(1 - \frac{\gamma}{\mu}\right)^{n+1} \\ &= \frac{\gamma}{\mu}\left(1 - \frac{\gamma}{\mu}\right)^n \\ &= P_{0\iota}\left(1 - P_{0\iota}\right)^n, n = 0, 1, 2, ... \end{aligned} \tag{5.34}$$

Formula (5.34) is analogous to the result for M/M/1 given in (3.91).

As a caveat to Proposition 5.5, the probabilities of n in the system *at an arbitrary time point* are *not* equal to $P_{n\iota}, n = 0, 1, 2, ...$(in general). Equality does hold if arrivals are Poisson.

5.1.10 G/M/1 with Poisson Arrivals

To enhance intuition, we specialize the foregoing G/M$_\mu$/1 results to M/M/1. When arrivals are Poisson at rate λ, the model reduces to an M$_\lambda$/M$_\mu$/1 queue.

Virtual Wait Assume $\overline{A}(x) = e^{-\lambda x}, x \geq 0$. Then $\gamma = \mu - \lambda$ is the solution of equation (5.12), $\int_{z=0}^{\infty} \overline{A}(z) e^{-\gamma z} dz = \frac{1}{\mu}$. Thus, $C_\gamma = \lambda$, where C_γ is defined in (5.16).

Hence

$$\begin{aligned} P_0 &= \frac{\gamma}{\gamma + C_\gamma} = \frac{\mu - \lambda}{\mu - \lambda + \lambda} = 1 - \frac{\lambda}{\mu}. \checkmark \\ K &= \frac{\gamma \cdot C_\gamma}{\gamma + C_\gamma} = \lambda(1 - \frac{\lambda}{\mu}) = \lambda P_0, \checkmark \\ f(x) &= K e^{-\gamma x} = \lambda P_0 e^{-(\mu - \lambda)x}, x > 0, \checkmark \\ E(\mathcal{B}) &= \frac{1}{\gamma} = \frac{1}{\mu - \lambda}, \checkmark \\ E(\mathcal{I}) &= \frac{1}{C} = \frac{1}{\lambda}. \checkmark \end{aligned}$$

These results check with the steady-state virtual wait for M/M/1.

Moreover, the part of the pdf of extended age for $x < 0$ is

$$h(x) = \mu \int_{y=0}^{\infty} e^{-\lambda(y-x)} K e^{-(\mu-\lambda)y} dy = K e^{\lambda x}, x < 0,$$

whence $P_0 = \int_{x=-\infty}^{0} h(x)dx = 1 - \frac{\lambda}{\mu}$.

Actual Wait For the *actual wait* in G/M/1, $\gamma = \mu - \lambda$, $P_{0\iota} = \frac{\gamma}{\mu} = 1 - \frac{\lambda}{\mu}$ and $K_\iota = \gamma\left(1 - \frac{\gamma}{\mu}\right) = \lambda(\frac{\mu-\lambda}{\mu}) = \lambda P_{0\iota}$; giving $f_\iota(x) = \lambda P_{0\iota} e^{-(\mu-\lambda)x}, x > 0$. These results agrees with P_0 and $f(x), x > 0$ in M/M/1 (see (3.86)).

For M/M/1, the Poisson arrival stream implies

$$\begin{aligned} P_{0\iota} &= P_0 = 1 - \frac{\lambda}{\mu},\ f_\iota(x) = f(x), x > 0, \\ \text{and } P_{n\iota} &= P_n = \left(\frac{\lambda}{\mu}\right)^n P_0, \end{aligned}$$

agreeing with PASTA [102].

5.1.11 Sojourn Time Above or Below a Level

We next determine the expected values of sojourn times above or below a state-space level.

Inter-upcrossing Time of a Level

Consider a sample path of the extended age process $\{V(t)\}$ (Fig. 5.1). Let u_x denote the inter-*upcrossing* time (between two successive upcrossings) of level x.

Levels ≥ 0 For $x \geq 0$, upcrossings of x are regenerative points due to exponentially distributed service times. Hence

$$E(u_x) = \frac{1}{\lim_{t\to\infty} \mathcal{U}_t(x)} = \frac{1}{f(x)}.$$

Therefore

$$E(u_x) = \frac{1}{f(x)} = \frac{e^{\gamma x}}{K}, x \geq 0, \tag{5.35}$$

where γ, K are given in (5.12), (5.19) respectively. (To compute K, we may use C_γ given in (5.16).)

Levels < 0 For $x < 0$, $-x$ is the time until the next arrival instant, at which a sample path of $\{V(t)\}$ hits level 0 from below. Upcrossings of x are regenerative points since the time to hit level 0 is $-x$, followed by a service time $\underset{dist}{=} \mathrm{E}_\mu$. From (5.14) we get

$$E(u_x) = \frac{1}{h(x)} = \frac{1}{\mu K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy}, x < 0. \tag{5.36}$$

Sojourn Time Above a Level

Let a_x denote the sojourn time of $\{V(t)\}$ above level x.

Levels ≥ 0 For $x \geq 0$, $E(a_x) = E(\mathcal{B})$ *independent* of x. By (5.11)

$$E(a_x) = \frac{1}{\gamma}, x \geq 0. \tag{5.37}$$

Levels < 0 For $x < 0$

$$\begin{aligned} E(a_x) &= E(u_x) - E(b_x) = \tfrac{1}{h(x)} - E(b_x) \\ &= \tfrac{1}{\mu K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy} \\ &\quad - \int_{z=0}^{\infty}\int_{y=0}^{\infty} \tfrac{\overline{A}(y-x+z)}{\overline{A}(y-x)} K e^{-\gamma y} dy dz, x < 0, \end{aligned} \tag{5.38}$$

where b_x is the sojourn time below x. The last term in (5.38),

$$E(b_x) = \int_{z=0}^{\infty}\int_{y=0}^{\infty} \frac{\overline{A}(y-x+z)}{\overline{A}(y-x)} K e^{-\gamma y} dy dz$$

is derived in Proposition 5.6 below.

Sojourn Time Below a Level

As noted previously, b_x is the sojourn time below level x.

Levels ≥ 0 We have

$$E(b_x) = E(u_x) - E(a_x) = \frac{e^{\gamma x}}{K} - \frac{1}{\gamma}, x \geq 0.$$

<u>**Levels < 0**</u> For $x < 0$ we have the following proposition.

Proposition 5.6 *The expected sojourn time of $\{V(t)\}$ below level x is*

$$E(b_x) = \int_{z=0}^{\infty}\int_{y=0}^{\infty} \frac{\overline{A}(y-x+z)}{\overline{A}(y-x)} Ke^{-\gamma y} dydz, x < 0. \tag{5.39}$$

Proof. Consider an SP downward jump that ends below $x < 0$ (all jumps start above level 0). Denote the *excess* of this jump below x by r_x. Since a sample path of $\{V(t)\}$ increases steadily at rate $+1$ and makes no jumps that start below 0, $E(b_x) = E(r_x)$. We have

$$\begin{aligned} P(r_x &> z|\text{jump starts at level } y > 0) \\ &= P(\text{inter-arrival time } > y-x+z|\text{inter-arrival time} > y-x) \\ &= \frac{\overline{A}(y-x+z)}{\overline{A}(y-x)}. \end{aligned}$$

Thus

$$\begin{aligned} E(b_x) &= E(r_x) = \int_{z=0}^{\infty} P(r_x > z)dz \\ &= \int_{z=0}^{\infty}\int_{y=0}^{\infty} P(r_x > z|\text{jump starts at level } y > 0)f(y)dydz \\ &= \int_{z=0}^{\infty}\int_{y=0}^{\infty} \frac{\overline{A}(y-x+z)}{\overline{A}(y-x)} Ke^{-\gamma y} dydz. \end{aligned}$$

■

5.1.12 Events During a Sojourn Above a Level

A system time = waiting time + service time. System times are realized at completions of service (instants of departure from the system). On the other hand, waiting times are realized at start of service instants.

<u>**Number of System Times During a_x**</u>

Let $N_{a_x}^{\sigma}$ denote the number of customers *completing* service during a *sojourn* of $\{V(t)\}$ above level $x \geq 0$. Thus $N_{a_x}^{\sigma}$ is the length of a run of system times $> x$. Let S_i, $T_i, i = 1, 2, ...$ denote the service and inter-arrival times, counting from the instant a sample path of $\{V(t)\}$ upcrosses level x (start of sojourn above x). If $x = 0$, S_1 is a full service time. If $x > 0$, S_1 is the remaining service time measured from the instant of upcrossing x. Thus S_1 is exponentially distributed with mean $\frac{1}{\mu}$ by the memoryless property. Then (Fig. 5.1)

$$N^{\sigma}_{a_x} = \min \left\{ n | \sum_{i=1}^{n} (S_i - T_i) \leq 0 \right\}, x \geq 0.$$

Thus $N^{\sigma}_{a_x}$ is a stopping time for $\{S_i - T_i\}$ and for $\{S_i\}$. The sojourn time of $\{V(t)\}$ above x is $a_x = \sum_{i=1}^{N^{\sigma}_{a_x}} S_i$. By Wald's equation and since $a_x \underset{dist}{=} \mathcal{B}$ for all $x \geq 0$

$$\begin{aligned} E(a_x) &= E(N^{\sigma}_{a_x})E(S_i), \\ E(N^{\sigma}_{a_x}) &= \frac{E(a_x)}{E(S_i)} = \frac{E(\mathcal{B})}{E(S)}. \end{aligned} \tag{5.40}$$

Substituting from (5.37) into (5.40) gives

$$E(N^{\sigma}_{a_x}) = \frac{\frac{1}{\gamma}}{\frac{1}{\mu}} = \frac{\mu}{\gamma}, \tag{5.41}$$

independent of x.

From (5.41) $E(N^{\sigma}_{a_x}) > 1$ since $\mu > \gamma$ (see Remark 5.7). This agrees with intuition, which suggests that there must be at least one departure instant in a sojourn above x (i.e., a sojourn ends at a departure instant).

Let $N^{\sigma}_{\mathcal{B}}$ denote the number of system-time realizations (number of customers served) in a busy period. Since $a_x \underset{dist}{=} \mathcal{B}$ and because of the memoryless property of the service time, $N^{\sigma}_{\mathcal{B}} \underset{dist}{=} N^{\sigma}_{a_x}, x \geq 0$. Therefore

$$E\left(N^{\sigma}_{\mathcal{B}}\right) = \frac{\mu}{\gamma}. \tag{5.42}$$

Number of Waiting Times During a_x

Let $N^{w}_{a_x}$ denote the number of customers that *start* service during a sojourn of $\{V(t)\}$ *above* level $x \geq 0$. Then $N^{w}_{a_x}$ is the number of customers that *wait* in line $> x$ (strictly) during $a_x, x \geq 0$. Examination of a sample path of $\{V(t)\}$ (Fig. 5.1) indicates that $N^{w}_{a_x} = N^{\sigma}_{a_x} - 1$. That is, the count of service starts during a_x is one less than the count of service completions during a_x, since the start of service initiating the sojourn corresponds to a wait $\leq x$. Hence

$$E(N^{w}_{a_x}) = E(N^{\sigma}_{a_x}) - 1 = \frac{\mu}{\gamma} - 1 > 0, x \geq 0. \tag{5.43}$$

Remark 5.7 *In (5.43) the inequality* $\frac{\mu}{\gamma} - 1 > 0$ *holds because of (5.12), i.e.,* $\int_{y=0}^{\infty} \overline{A}(y) e^{-\gamma y} dy = \frac{1}{\mu}$*; and* $\overline{A}(0) = 1$*,* $\overline{A}(y) = 1 - A(y)$ *is non-increasing with* $\lim_{y \to \infty} \overline{A}(y) = 0$*. Thus there exists finite* $M > 0$ *such that* $\overline{A}(y) < 1$ *(strictly) for* $y > M$*. Hence*

$$\frac{1}{\mu} = \int_{y=0}^{\infty} \overline{A}(y) e^{-\gamma y} dy < \int_{y=0}^{\infty} 1 \cdot e^{-\gamma y} dy = \frac{1}{\gamma} \implies \frac{\mu}{\gamma} > 1.$$

5.1.13 Events Above a Level During a Busy Period

We first obtain the expected *number of SP sojourns* above a level during a busy period.

Number of Sojourns in Busy Period Above Level $x > 0$

Let $\mathcal{C}$ denote a busy *cycle*. Let $N_{a_x}^{\text{soj}}(\mathcal{C})$, $N_{a_x}^{\text{soj}}(\mathcal{B})$ be the number of SP *sojourns* above level x during a busy *cycle* and busy period, respectively. Then $N_{a_x}^{\text{soj}}(\mathcal{C}) \underset{dist}{=} N_{a_x}^{\text{soj}}(\mathcal{B})$, since all such sojourns take place in an embedded busy period. Let $\mathcal{U}_{\mathcal{C}}(x)$ denote the number of SP *upcrossings* of level x during a busy cycle. Each sojourn above x starts with an upcrossing of x. Thus $N_{a_x}^{\text{soj}}(\mathcal{C}) \underset{dist}{=} \mathcal{U}_{\mathcal{C}}(x)$. By the theory of regenerative processes, specific time averages in a busy cycle recapitulate the same specific limiting time averages (e.g., [96]). Thus

$$\frac{E\left(N_{a_x}^{\text{soj}}(\mathcal{C})\right)}{E(\mathcal{C})} = \frac{E(\mathcal{U}_{\mathcal{C}}(x))}{E(\mathcal{C})} = \lim_{t \to \infty} \frac{E(\mathcal{U}_t(x))}{t} = f(x), x \geq 0, \tag{5.44}$$

where $\mathcal{U}_t(x)$ is the number of upcrossings of level x during $(0, t]$. Recall that $f(x) = Ke^{-\gamma x}$ and $E(\mathcal{C}) = \frac{1}{f(0)} = \frac{1}{K}$. Thus, from (5.44)

$$\begin{aligned} E\left(N_{a_x}^{\text{soj}}(\mathcal{B})\right) &= E\left(N_{a_x}^{\text{soj}}(\mathcal{C})\right) = E(\mathcal{U}_{\mathcal{C}}(x)) = E(\mathcal{C}) \cdot f(x) \\ &= \frac{1}{K} \cdot Ke^{-\gamma x} = e^{-\gamma x}, x \geq 0. \end{aligned} \tag{5.45}$$

Setting $x = 0$ in (5.45) implies

$$E\left(N_{a_0}^{\text{soj}}(\mathcal{B})\right) = E\left(N_{a_0}^{\text{soj}}(\mathcal{C})\right) = 1.$$

Note that the single sojourn above level 0 in a busy cycle and in the embedded busy period start simultaneously. In other words, a busy period consists of exactly one sojourn above level 0.

Number of System Times $> x$ in Busy Period

Let $N^{\sigma}_{a_x}(\mathcal{B})$, $N^{\sigma}_{a_x}(\mathcal{C})$ denote the number of completed system times $> x$ during a busy period and busy cycle respectively. Then $N^{\sigma}_{a_x}(\mathcal{B}) = N^{\sigma}_{a_x}(\mathcal{C})$ since all departures in a busy cycle occur during the contained busy period. Departures that correspond to system times $> x$ occur during a_x. Also

$$N^{\sigma}_{a_x}(\mathcal{C}) = \sum_{i=1}^{N^{\text{soj}}_{a_x}(\mathcal{C})} N^{\sigma}_{a_x i} \tag{5.46}$$

where $N^{\sigma}_{a_x i}$ is the number of system times $> x$ during the i^{th} *sojourn* above x in $\mathcal{C}$. The $N^{\sigma}_{a_x i}$'s are iid r.v.'s. with $E\left(N^{\sigma}_{a_x i}\right) = \frac{\mu}{\gamma}$ by (5.41) independent of the number of sojourns $N^{\text{soj}}_{a_x}(\mathcal{C})$ above x (memoryless property of service time). Taking expected values in (5.46) and using (5.45) gives

$$E\left(N^{\sigma}_{a_x}(\mathcal{B})\right) = E\left(N^{\sigma}_{a_x}(\mathcal{C})\right) = E\left(N^{\text{soj}}_{a_x}(\mathcal{C})\right) \cdot E\left(N^{\sigma}_{a_x i}\right) = \frac{\mu}{\gamma} e^{-\gamma x}. \tag{5.47}$$

Number of Waiting Times $> x$ in Busy Period

We obtain the expected number of *waiting times* $> x$ in $\mathcal{B}$, similarly as for the derivation of (5.43) (see Remark 5.7). Thus

$$E\left(N^{w}_{a_x}(\mathcal{B})\right) = \left(\frac{\mu}{\gamma} - 1\right) e^{-\gamma x}, x \geq 0. \tag{5.48}$$

Setting $x = 0$ in (5.48) gives $E\left(N^{w}_{a_0}(\mathcal{B})\right) = \frac{\mu}{\gamma} - 1$. $E\left(N^{w}_{a_0}(\mathcal{B})\right)$ is also the expected number of customers in a busy period that wait a *positive* time (same as (5.43)). Only the first customer in $\mathcal{B}$ waits 0.

Proportion that Wait > 0 We can connect this result with other parameters of the model. For example, the *proportion* of customers that wait > 0 in a busy period is

$$\frac{E\left(N^{w}_{a_0}(\mathcal{B})\right)}{E\left(N^{\sigma}_{\mathcal{B}}\right)} = \frac{\frac{\mu}{\gamma} - 1}{\frac{\mu}{\gamma}} = 1 - \frac{\gamma}{\mu} = 1 - P_{0\iota}. \tag{5.49}$$

In (5.49) the denominator $E\left(N^{\sigma}_{\mathcal{B}}\right)$ is the expected number of service completions in a busy period (equal to expected number of service starts in a busy period). Formula (5.49) is intuitive, as a busy cycle is a probabilistic microcosm of the evolution of the system over the entire time axis. The long-run proportion of customers that wait a positive time is $1 - P_{0\iota}$.

Number Served in a Sojourn Above Level $x < 0$

Fix a level in the state space $x < 0$. After upcrossing x, a sample path of $\{V(t)\}$ ascends steadily at rate +1 to level 0. Hence the *number* of service completions during a_x is

$$N_{a_x} = \min \left\{ n | -x + \sum_{i=1}^{n} (S_i - T_i) \le x \right\}.$$

Thus N_{a_x} is a stopping time for $\{S_i - T_i\}$ and for $\{S_i\}$. The *sojourn time* above x is $a_x = -x + \sum_{i=1}^{N_{a_x}} S_i$ implying that

$$E(a_x) = -x + E(N_{a_x}) \cdot E(S_i).$$

Thus

$$\begin{aligned} E(N_{a_x}) &= \frac{E(a_x) + x}{E(S_i)} \\ &= \mu \left(E(a_x) + x \right), \end{aligned} \tag{5.50}$$

where $E(a_x)$ is given in (5.38). Note that in (5.50) the numerator $E(a_x)+x$ is positive, since $a_x > -x$ (see Fig. 5.1).

5.1.14 Revisit of M/M/1

We revisit the M/M/1 model in the light of the results for G/M/1 in subsections 5.1.12 and 5.1.13.

Consider equation (5.48) for G/M/1. If arrivals are Poisson at rate λ then $\gamma = \mu - \lambda$. Thus

$$\begin{aligned} E\left(N_{a_0}^w(\mathcal{B})\right) &= \frac{\mu}{\gamma} - 1 = \frac{\mu}{\mu - \lambda} - 1 \\ &= \frac{1}{1 - \frac{\lambda}{\mu}} - 1 = \frac{1}{P_0} - 1. \end{aligned}$$

In M/M/1 (and M/G/1), the expected number of customers served in a busy period is $E\left(N_{\mathcal{B}}^g\right) = \frac{1}{P_0}$ (formula (3.65)). The customer that initiates $\mathcal{B}$ waits zero. Any other customer served in $\mathcal{B}$ waits a *positive* time. This explains intuitively why $E\left(N_{a_0}^w(\mathcal{B})\right) = E\left(N_{\mathcal{B}}^g\right) - 1$.

In M/M/1 (and M/G/1) the *proportion* of customers that wait a positive time in a busy period is

$$\frac{E\left(N_{a_0}^w(\mathcal{B})\right)}{E\left(N_{\mathcal{B}}^g\right)} = \frac{\frac{1}{P_0} - 1}{\frac{1}{P_0}} = 1 - P_0 = \frac{\lambda}{\mu} = \rho,$$

which agrees with the result for G/M/1 given in (5.49).

Related Results for $\mathbf{M}_\lambda/\mathbf{M}_\mu/\mathbf{1}$

In a similar manner to the analyses above for G/M/1, we obtain the following results for $\mathrm{M}_\lambda/\mathrm{M}_\mu/1$ (see Fig. 3.6). The expected number of system times completed in a sojourn above level x is

$$E\left(N_{a_x}^{\sigma}\right) = E\left(N_{\mathcal{B}}^{\sigma}\right) = \frac{\mu}{\mu - \lambda} = \frac{1}{P_0}, x \geq 0, \tag{5.51}$$

where $N_{\mathcal{B}}^{\sigma}$ is the number served in a busy period, independent of x. Equality $E\left(N_{a_x}^{\sigma}\right) = E\left(N_{\mathcal{B}}^{\sigma}\right)$ follows because in M/M/1, $a_x \underset{dist}{=} \mathcal{B}, x \geq 0$. Also, $E\left(N_{a_x}^{\sigma}\right) > 1$ since $\mu > \mu - \lambda$ for stability (i.e., $0 < \lambda < \mu$).

The expected number that wait $> x$ in a_x is

$$E(N_{a_x}^{w}) = \frac{\mu}{\mu - \lambda} - 1 = \frac{1}{P_0} - 1 = \frac{\lambda}{\mu - \lambda} \geq 0. \tag{5.52}$$

The expected number of sojourns above x in a busy period is

$$\begin{aligned} E(N_{a_x}^{\mathrm{soj}}(\mathcal{B})) &= E(N_{a_x}^{\mathrm{soj}}(\mathcal{C})) = E(\mathcal{C}) \cdot f(x) \\ &= \frac{1}{\lambda P_0} \cdot \lambda P_0 e^{-(\mu\, -\, \lambda)x} = e^{-(\mu - \lambda)x}, x \geq 0. \end{aligned} \tag{5.53}$$

If $x = 0$ then

$$E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{B})\right) \;=\; e^0 = \; E\left(N_{a_0}^{\mathrm{soj}}(\mathcal{B})\right) = 1.$$

In fact $\mathcal{B}$ has exactly one sojourn above level 0. In contrast, $\mathcal{B}$ may have a random number of sojourns above an arbitrary positive level.

The number of system times (service completions) above level x in a busy period is

$$N_{a_x}^{\sigma}(\mathcal{B}) = N_{a_x}^{\sigma}(\mathcal{C}) = \sum_{i=1}^{N_{a_x}^{\mathrm{soj}}(\mathcal{C})} N_{a_x i}^{\sigma}.$$

By (5.51) and (5.53),

$$\begin{aligned} E\left(N_{a_x}^{\sigma}(\mathcal{B})\right) &= E\left(N_{a_x}^{\sigma}\right) = E\left(N_{\mathcal{B}}^{\sigma}\right) \cdot E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{C})\right) \\ &= \frac{\mu}{\mu - \lambda} \cdot e^{-(\mu\, -\, \lambda)x}, \geq 0. \end{aligned} \tag{5.54}$$

The expected number of waiting times $> x$ in $\mathcal{B}$ is

$$\begin{aligned} E(N_{a_x}^{w}(\mathcal{B})) &= \left(\frac{\mu}{\mu - \lambda} - 1\right) \cdot e^{-(\mu\, -\, \lambda)x} \\ &= \frac{\lambda}{\mu - \lambda} \cdot e^{-(\mu\, -\, \lambda)x}, x \geq 0. \end{aligned} \tag{5.55}$$

If $x=0$, then $E(N_{a_x}^w(\mathcal{B}))=\frac{\lambda}{\mu-\lambda}$ = expected number that wait >0 in $\mathcal{B}$.

The *proportion* of customers that wait >0 in $\mathcal{B}$ is

$$\frac{E\left(N_{a_0}^w(\mathcal{B})\right)}{E\left(N_{\mathcal{B}}^\sigma\right)}=\frac{\frac{\lambda}{\mu-\lambda}}{\frac{\mu}{\mu-\lambda}}=\frac{\lambda}{\mu}=1-P_0,$$

where $N_{\mathcal{B}}^\sigma$ = number served in $\mathcal{B}$. The intuitive explanation of the last formula is that the long-run proportion of customers that wait >0 is $1-P_0$ ($\mathcal{C}$ is a probabilistic replica of the entire time line. All arrivals take place in the embedded $\mathcal{B}$).

Proposition 5.7 *For M/M/1 the expected number of system times* $\leq x$ *in a busy period* $\mathcal{B}$ *is*

$$\begin{aligned} E(N_{b_x}^\sigma(\mathcal{B})) &= \frac{\mu}{\mu-\lambda}-\frac{\mu}{\mu-\lambda}e^{-(\mu-\lambda)x} \\ &= \frac{\mu}{\mu-\lambda}(1-e^{-(\mu-\lambda)x}), x\geq 0. \end{aligned} \tag{5.56}$$

Proof. In $\mathcal{B}$, the number of customers with *system* times $\leq x$ plus the number with system times $>x$, is equal to the total number served in $\mathcal{B}$, namely $N_{\mathcal{B}}^\sigma$. Thus from (5.51)

$$E(N_{b_x}^\sigma(\mathcal{B}))+E(N_{a_x}^\sigma(\mathcal{B}))=E\left(N_{\mathcal{B}}^\sigma\right)=\frac{\mu}{\mu-\lambda}.$$

Then (5.56) follows from (5.51) and (5.54). ■

Proposition 5.8 *For M/M/1 the expected number of waiting times* $\leq x$ *in a busy period* $\mathcal{B}$ *is*

$$E(N_{b_x}^w(\mathcal{B}))=\frac{\mu}{\mu-\lambda}-\frac{\lambda}{\mu-\lambda}e^{-(\mu-\lambda)x}, x\geq 0. \tag{5.57}$$

Proof. In $\mathcal{B}$, the number of customers with *waiting* times $\leq x$ plus the number with waiting times $>x$, is equal to the number served in $\mathcal{B}$, namely $N_{\mathcal{B}}^\sigma$. By (5.51),

$$E(N_{b_x}^w(\mathcal{B}))+E(N_{a_x}^w(\mathcal{B}))=E\left(N_{\mathcal{B}}^\sigma\right)=\frac{\mu}{\mu-\lambda}.$$

Thus, (5.57) follows from (5.52) and (5.55). ■

Remark 5.8 *For* $M_\lambda/M_\mu/1$*, we have the following.*

If $x=0$ *then* $E(N_{b_x}^\sigma(\mathcal{B}))=0$.✓

If $x=\infty$ *then* $E(N_{b_x}^\sigma(\mathcal{B}))=\frac{\mu}{\mu-\lambda}$.✓

If $x=0$ *then* $E(N_{b_x}^w(\mathcal{B}))=1$ *(initiator of* $\mathcal{B}$ *waits* 0*).*✓

If $x=\infty$ *then* $E(N_{b_x}^w(B))=\frac{\mu}{\mu-\lambda}$.✓

5.1.15 Boundedness of Steady-state PDF of Wait

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

The pdf of the *virtual* wait is $f(x) = Ke^{-\gamma x}, x > 0$. From (5.19) $K < \gamma$. Also $\gamma < \mu$. This $f(x) < \mu, x > 0$.

The pdf of the *actual* wait is $f_\iota(x) = K_\iota e^{-\gamma x}, x > 0$. From (5.24) $K_\iota = \gamma\left(1 - \frac{\gamma}{\mu}\right)$. Since $\gamma < \mu$, we obtain $f_\iota(x) < \mu, x > 0$.

Proposition 5.9 below proves boundedness of the steady-state pdf or the virtual wait in several ways, from "first principles" without drawing on the result $f(x) = Ke^{-\gamma x}, x > 0$. We include it for ideas that may be useful to obtain bounds on the pdf of wait in variants of G/M/1 (or random variables in other models), from basic LC considerations.

Proposition 5.9

$$f(x) < \mu, x > 0. \tag{5.58}$$

Proof. We present three proofs for perspective.

(1) In the integral equation for G/M/1 (5.6) (repeated here)

$$f(x) = \mu \int_{y=x}^{\infty} \overline{A}(y-x) f(y)dy, x > 0,$$

we have $\overline{A}(z) < 1$ for $z > M$ sufficiently large, since $\lim_{z\to\infty} \overline{A}(z) = 0$. Thus,

$$f(x) < \mu \int_{y=x}^{\infty} 1 \cdot f(y)dy < \mu \left(P_0 + \int_{y=0}^{\infty} f(y)dy\right) = \mu,$$

since the normalizing condition is $P_0 + \int_{y=0}^{\infty} f(y)dy = 1$.

(2) An alternative form of the LC integral equation for G/M/1 (5.7) (repeated here for convenience)

$$f(x) = \mu(1 - F(x)) - \mu \int_{y=x}^{\infty} A(y-x) f(y)dy, x > 0. \tag{5.59}$$

The subtracted term is such that

$$0 < \mu \int_{y=x}^{\infty} A(y-x) f(y)dy < \mu \int_{y=x}^{\infty} 1 \cdot f(y)dy = \mu\left(1 - F(x)\right),$$

since $A(z) < 1$ for z in a positive neighborhood of 0.

Thus

$$f(x) < \mu(1 - F(x)) < \mu, x > 0.$$

(3) Consider a sample path of $\{V(t)\}$ (see (5.1) and Fig. 5.1). Let $\mathcal{U}_t(x)$, $N_{srv}(t)$ denote the number of SP upcrossings of level x and number of service completions during $(0, t)$ respectively. Assume t is larger than one busy cycle. Then $E(\mathcal{U}_t(x)) < E(N_{srv}(t)), x \geq 0$ because: (a) there is a one-to-one correspondence between upcrossings of x and the first service completions in the immediately ensuing sojourns above x (completions having system time $> x$); (b) there may be several service completions with system time $> x$ during a sojourn above x; (c) there may be service completions with system time $< x$, which do not correspond to an upcrossing of x during $(0, t)$. Hence

$$f(x) = \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} < \lim_{t\to\infty} \frac{E(N_{srv}(t))}{t} \leq \mu.$$

The last inequality $\lim_{t\to\infty} \frac{E(N_s(t))}{t} \leq \mu$ holds since $N_{srv}(t) \leq N_\mu(t)$ where $N_\mu(t)$ is a Poisson r.v. with rate μt, due to idle periods (see Fig. 5.1). ■

Example 5.1 ***$M_\lambda/M_\mu/1$*** *is a special case of G/M//1 in which $\lambda < \mu$ for stability. From Example 3.5, in M/M/1 $f(x) < \lambda < \mu$.*

5.2 Multiple-Server G/M/c Queue

The G/M/c ($c = 2, 3, ...$) queue generalizes G/M/1 of Section 5.1 to multiple parallel servers. The same symbols as in Section 5.1 specify the arrival stream: cdf $A(\cdot)$, pdf $a(\cdot)$, complementary cdf $\overline{A}(\cdot)$, mean $\frac{1}{\boldsymbol{a}}$. For each customer the service $\underset{dist}{=}$ E_μ. The service times in servers that are occupied simultaneously are assumed to be independent.

This section emphasizes the use of LC to analyze the steady-state pdf's of the *virtual wait* and of the *actual wait (arrival-point wait).* We derive explicit formulas for the pdf's in G/M/2, and check them against the pdf's in M/M/2; this mildly validates the LC approach. In addition we derive related properties of G/M/c using LC concepts.

5.2.1 Extended Age Process for G/M/c

For analyzing the multiple-server G/M/c queue, we employ the stochastic process

$$\{V(t), M(t), t \geq 0\}, -\infty < V(t) < \infty, M(t) \in \boldsymbol{M}.$$

Random variable $V(t)$ is the "extended age" at time t. For G/M/c, $V(t)$ is a slight generalization of $V(t)$ defined for G/M/1 in Subsection 5.1.1 (see next heading in the present subsection).

Random variable $M(t)$ is defined here as the number of customers in *service* at time t. Thus $M(t) \in \boldsymbol{M} = \{0, 1, 2, ..., c\}$. When $M(t) = c$ there are at least c customers in the system.

The state space of $\{V(t), M(t)\}$ is $\boldsymbol{S} = \boldsymbol{R} \times \boldsymbol{M}$ where $\boldsymbol{R} = (-\infty, +\infty)$. Random variable $M(t)$ is the "system configuration". Here, $M(t)$ is defined more simply than for the general M/M/c model in Chapter 4. This is because we are analyzing a *standard* G/M/c model without the generality of the M/M/c model of Chapter 4 (see Subsection 4.5).

The process $\{V(t), M(t)\}$ is a "system point" process. The state is two-dimensional. Random variable $V(t)$ is continuous; random variable $M(t)$ is discrete.

Remark 5.9 *The definition of* ***system configuration is flexible****. That is, an analyst utilizes a configuration that expedites the analysis of a model. We could define $M(t)$ for G/M/c as in Subsection 4.5 for M/M/c. However, we use a definition which is sufficient to examine a standard G/M/c model. If the objective were to analyze a more general G/M/c model, we would define $M(t)$ along the lines of Subsection 4.5. This would be the case in models with, for example: service time depending on wait; service time depending on the types of other customers in service at start of service times; service rate selected at random from a set of possible service rates; etc.*

Remark 5.10 *The definition of $M(t) \in \{0, 1, ..., c\}$, is a variation of the general definition in Subsection 4.5, which is appropriate for M/M/c. For G/M/c, if $M(t) \in \{0, 1, ..., c-1, c\}$, $M(t)$* ***is*** *the number of occupied servers "seen" by an arrival. This version of $M(t)$ encompasses a "sheet c" to denote "all servers are occupied" (instead of "sheet $c-1$" as for M/M/c), because sheet $c-1$ in the G/M/c model corresponds to arrivals that "see" $c-1$ units in service (Fig. 5.3).*

Extended Age and Inter Start-of-service Departure Times

Assume $M(t) = c$. When $M(t) = c$, $V(t)$ is the "*age*" (time already spent in the system) of the *last* customer to start service at or before t. Thus $V(t) > 0$. Let $\mathcal{S}$ denote the time from the instant a customer starts service until the first departure from the system thereafter. Random variable $\mathcal{S}$ is the *inter start-of-service departure time*. Then

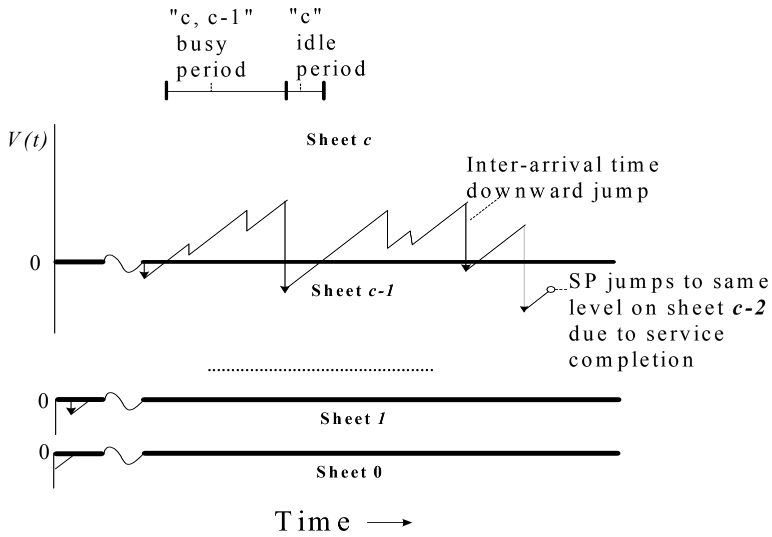

Figure 5.3: Sample path of $\{V(t), M(t)\}$ for G/M/c queue. There are $c+1$ sheets. Range of sheet c is $[0, \infty)$. Range of sheets $0, ..., c-1$ is $(-\infty, 0)$. Sheet $c-1$ abuts on sheet c for geometric convenience. Time between jumps originating on sheet $c = \mathrm{E}_{c\mu}$.

$\mathcal{S} = \min\{S_1, ..., S_c\}$ where $\{S_i\}$ are iid r.v.'s each $\underset{dist}{=}$ E_μ. One of the S_i's is a full service time; $c-1$ of the S_i's are *remaining* service times. Hence $\mathcal{S} \underset{dist}{=} E_{c\mu}$.

<u>**Relationship Between $V(t)$ and $M(t)$**</u>

When $M(t) \in \{0, 1, 2, ..., c-1\}$, random variable $-V(t)$ denotes the *remaining* inter-arrival time required until the next arrival joins the system. Thus (Fig. 5.3),

$$\text{if} \begin{cases} M(t) = c \text{ then } V(t) \geq 0; \\ M(t) \in \{0, 1, 2, ..., c-1\} \text{ then } V(t) < 0. \end{cases}$$

5.2.2 Steady-state PDF of Virtual Wait

Let $\boldsymbol{T} = [0, \infty)$ denote the time axis. Consider a sample path of the process $\{V(t), M(t)\}$. The *rate* at which the SP moves in $\boldsymbol{T} \times \boldsymbol{S}$ between downward jumps is

$$\frac{d}{dt}V(t) = +1, -\infty < V(t) < \infty,\ M(t) = 0, ..., c, t > 0.$$

The steady-state pdf of $V(t)$ as $t \to \infty$, is the same as that of the *virtual wait* $W(t)$ as $t \to \infty$ (proved similarly as in Proposition 5.1 for G/M/1).

Denote the steady-state cdf of the virtual wait by $F(x), x \geq 0$, having pdf $f(x) = \frac{d}{dx}F(x), x > 0$, wherever the derivative exists. The quantity $F(0)$ is the *proportion of time* there is fewer than c customers in service. That is, $F(0)$ is the probability that the system presents a zero wait to a potential arrival. Let P_i be the proportion of time that an arrival "sees" i customers in service, $i = 0, ..., c-1$. The P_i's are zero-wait probabilities. Then $F(0) = \sum_{i=0}^{c-1} P_i$.

Integral Equation for PDF of Wait

Consider a sample path of $\{V(t), M(t)\}$ (Fig. 5.3). The space $\boldsymbol{T} \times \boldsymbol{S}$ is partitioned into $(c+1)$ sheets (or pages). The sheets are planar subsets of $\boldsymbol{T} \times \boldsymbol{S}$. Sheets $0, ..., c-1$ can be thought of as being one behind the other like pages in a book, *below* the time axis. Only sheet c is *above* the time axis. Sheet c is pictured as being directly above, and contiguous to, sheet $c-1$.

Consider $M(t) = c$, and corresponding sheet c. Fix level $x > 0$. The SP *upcrossing* rate of level x is

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

(proved similarly as for the downcrossing rate in M/G/1, e.g., Theorem 1.1).

The SP *downcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} \underset{a.s}{=} \lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = c\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy.$$

The coefficient $c\mu$ of the integral, is the rate at which customers depart the system when all servers are occupied. Such departures generate SP downward jumps. Downward jump sizes are distributed as the *inter-arrival* time. The term $\overline{A}(y-x)$ in the integrand is the probability that an SP jump starts at level $y > x$ and downcrosses level $x \in (-\infty, y)$.

Rate balance across level x,

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

gives a basic LC integral equation for G/M/c

$$f(x) = c\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy, x > 0. \tag{5.60}$$

In contrast to (5.6) for G/M/1 where the SP downward jump rate is μ, in (5.60) for G/M/c the SP downward jump rate is $c\mu$.

Alternative Form of Integral Equation

An alternative form of (5.60) is

$$f(x) = c\mu\,(1 - F(x)) - c\mu \int_{y=x}^{\infty} A(y-x)f(y)dy, x > 0. \tag{5.61}$$

In (5.61) $c\mu\,(1 - F(x))$ is the rate at which downward jumps start in state-space set (x, ∞). The integral is the rate at which downward jumps start in (x, ∞) and end in (x, ∞); such jumps do not downcross x. Thus the right side is the downcrossing rate of level x.

5.2.3 Form of PDF of Wait in G/M/c Geometrically

Let $\mathcal{B}_{c-1,c}$ denote a $[c-1, c]$ busy period. Random variable $\mathcal{B}_{c-1,c}$ is the time from the instant the number of customers in service increases from $c-1$ to c until the first instant thereafter at which the number of customers in service decreases back to $c-1$ (Fig. 5.3). During $\mathcal{B}_{c-1,c}$ the number of customers in the system is $\geq c$. Thus $\mathcal{B}_{c-1,c}$ is equal to a sojourn time on sheet c, which starts by an SP upcrossing of level 0 (from top of sheet $c-1$). Let a_x denote a sojourn time above level $x \geq 0$ starting with an upcrossing of x (on sheet c). Then $\mathcal{B}_{c-1,c} = a_0$, and $E\,(\mathcal{B}_{c-1,c}) = E\,(a_0)$.

The memoryless property of $\mathcal{S}$ ($\underset{dist}{=}$ $\mathrm{E}_{c\mu}$) implies $E(a_x) = E(\mathcal{B}_{c-1,c})$ *independent of* $x \geq 0$. Thus the *proportion* of time the SP spends above an arbitrary level $x > 0$ is

$$\begin{aligned} \lim_{t\to\infty} \frac{E\,(\mathcal{U}_t(x)) \cdot E(a_x)}{t} &= \lim_{t\to\infty} \frac{E\,(\mathcal{U}_t(x))}{t} \cdot E(a_x) \\ &= f(x) \cdot E(\mathcal{B}_{c-1,c}), x \geq 0. \end{aligned}$$

Similarly as for G/M/1 in Subsection 5.1.5, we have

$$\begin{aligned} f(x)E(\mathcal{B}_{c-1,c}) &= 1 - F(x), \\ \frac{d}{dx} \ln(1 - F(x)) &= -\frac{1}{E(\mathcal{B}_{c-1,c})}. \end{aligned}$$

The solution of this differential equation is the cdf of wait

$$F(x) = 1 - (1 - F(0)) \cdot e^{-\frac{1}{E\left(\mathcal{B}_{c-1,c}\right)}x}, x \geq 0. \tag{5.62}$$

Taking $\frac{d}{dx}F(x)$ in (5.62) gives

$$f(x) = \frac{1 - F(0)}{E(\mathcal{B}_{c-1,c})} \cdot e^{-\frac{1}{E(\mathcal{B}_{c-1,c})}x}, x \geq 0. \tag{5.63}$$

Hence

$$f(x) = Ke^{-\gamma x}, x > 0, \tag{5.64}$$

where

$$K = \frac{1 - F(0)}{E(\mathcal{B}_{c-1,c})}, \ \gamma = \frac{1}{E(\mathcal{B}_{c-1,c})}. \tag{5.65}$$

Using (5.65) we have

$$E(\mathcal{B}_{c-1,c}) = \frac{1}{\gamma}. \tag{5.66}$$

Substituting $f(x)$ from (5.64) into (5.60) gives a transcendental equation for γ,

$$\int_{y=0}^{\infty} \overline{A}(y)e^{-\gamma y}dy = \frac{1}{c\mu}. \tag{5.67}$$

Note that the Laplace-Stieltjes transform of the inter-arrival distribution evaluated at γ, is $A^*(\gamma) = \int_{y=0}^{\infty} a(y)e^{-\gamma y}dy$. On the left side of (5.67) integration by parts gives an alternative equation for γ,

$$A^*(\gamma) = 1 - \frac{\gamma}{c\mu}. \tag{5.68}$$

To specify the mixed pdf of wait $\{F(0); f(x), x > 0\}$, it is required to solve for $F(0)$ in (5.63) or equivalently for K in (5.64). From (5.65) we obtain

$$F(0) = 1 - \frac{K}{\gamma}. \tag{5.69}$$

Note that once we know the form of $f(x)$, we could also obtain (5.69) from the normalizing condition

$$\begin{aligned}
F(0) + \int_{x=0}^{\infty} f(x)dx &= 1, \\
F(0) + \int_{x=0}^{\infty} Ke^{-\gamma x}dx &= 1, \\
F(0) + \frac{K}{\gamma} &= 1.
\end{aligned}$$

Remark 5.11 *Another way to obtain (5.69) is directly from the sample path of* $\{V(t)\}$ *and SP motion in the state space. We include this derivation because it highlights the close relationship between probabilities of the model and the motion of the SP. Note that* $F(0)$ *is the* ***proportion*** *of time that the system presents a zero wait. The expected time between successive SP upcrossings of level* 0 *due to arrivals that see* $c-1$ *customers in service, is* $\frac{1}{f(0)}$ *(starts of* $\mathcal{B}_{c-1,c}$ *busy periods). Also, since* $f(x) = Ke^{=\gamma x}$,

$$\lim_{t\to\infty} \frac{E\left(\mathcal{U}_t(0)\right)}{t} = f(0) = K.$$

After the SP moves on sheet c, it leaves sheet c when a departure propels it downward onto sheet $c-1$*. The SP then sojourns among some or all sheets* $0,...,c-1$*. During this SP sojourn, an arrival would wait zero. The sojourn continues until the SP next upcrosses level* 0 *from sheet* $c-1$ *to sheet c. From the theory of regenerative processes*

$$\begin{aligned} F(0) &= \frac{E(\textit{sojourn time among sheets } 0,...,c-1)}{E(\textit{time between entrances to sheet } c)} \\ &= \frac{\frac{1}{K} - E(\mathcal{B}_{c-1,c})}{\frac{1}{K}} = \frac{\frac{1}{K}-\frac{1}{\gamma}}{\frac{1}{K}} = 1 - \frac{K}{\gamma}. \end{aligned} \tag{5.70}$$

<u>**Value of K**</u>

At this point, we must solve for the value of K in order to specify $F(0)$ and $f(x), x > 0$ in terms of the model parameters. This requires a further analysis of sheets $0,...,c-1$.

Remark 5.12 *Applying the normalizing condition*

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

and using (5.63), does not give the value of $F(0)$ *in terms of the model parameters, since it yields the tautology* $1 = 1$*. In Subsection 5.2.4 below we develop integral equations for the steady-state partial pdf's of* $V(t)$ *on sheets* $0,...,c-1$*. These allow us to find an independent expression for* $F(0)$*, and then apply the normalizing condition to solve for* $F(0)$*. We shall not solve for* $F(0)$ *explicitly for the general G/M/c queue. However, we indicate the solution procedure by solving for* $F(0)$ *explicitly for G/M/2 in Section 5.3, below.*

5.2.4 Partial PDF's of Extended Age: Sheets 0 to $c-1$

Let $g_i(x), x<0$, denote the steady-state pdf of $V(t)$ when $M(t)=i, i=0,...,c-1$. In Fig. 5.3 the partial pdf's $\{g_i(x), x<0\}$ correspond to sheets $0,...,c-1$. We derive a set of integral equations for $g_i(x), x<0, i=0,...,c-1$, by applying rate balance of SP *exits* and *entrances* of state-space intervals $((-\infty,x),i), x<0$ on sheets $i=0,...,c-1$.

The probability $F(0)$ is the proportion of time that potential arrivals wait 0 for service. Thus

$$F(0)=\sum_{i=0}^{c-1}\int_{x=-\infty}^{0} g_i(x)dx=\sum_{i=0}^{c-1}P_i \tag{5.71}$$

where $P_i=\int_{x=-\infty}^{0} g_i(x)dx$ is the steady-state probability of i customers in service, $i=0,...,c-1$.

Integral Equation for PDF: Sheet $c-1$

First consider interval $((-\infty,x),c-1), x<0$, on sheet $c-1$.

Exit Rate The SP *exit* rate from $((-\infty,x),c-1)$ is

$$g_{c-1}(x)+(c-1)\mu\int_{y=-\infty}^{x} g_{c-1}(y)dy. \tag{5.72}$$

In (5.72) the first term is the SP (continuous) *upcrossing* rate of level x. The second term is the rate at which customers *depart* the system when $c-1$ servers are occupied *and* the *remaining time* until the *next* arrival to the system is $-y$, summed over all $y\in(-\infty,x)$. Departures occur at rate $(c-1)\mu$ since there are $c-1$ customers in service, and service times are independent of the remaining time until the next arrival. Such customer departures generate SP *parallel* jumps from sheet $c-1$ to sheet $c-2$ *at the same level.* That is, just after such departures there would be $c-2$ units in service and the remaining inter-arrival time would still be the same as just before the departure.

Entrance Rate The SP *entrance* rate into $((-\infty,x),c-1)$ is

$$c\mu\int_{y=0}^{\infty}\overline{A}(y-x)f(y)dy+g_{c-2}(0)\overline{A}(-x). \tag{5.73}$$

In (5.73) the first term is the rate at which the SP jumps downward from level $y > 0$ on sheet c into interval $((-\infty, x), c-1)$, due to customer departures that leave $c-1$ units in service. An inter-arrival time that is $> y - x$ causes the SP to jump downward below level x on sheet $c-1$ (probability is $\overline{A}(y-x)$). In the second term, factor $g_{c-2}(0)$ is the SP hit rate of level 0 from below ("upcrossing" rate), which is the *arrival* rate to the system when there are $c-2$ servers occupied. Such arrivals *increase* the number of occupied servers to $c-1$. The factor $\overline{A}(-x)$ is the probability that the *immediately following inter-arrival time* exceeds $-x$, thereby propelling the SP below level x on sheet $c-1$.

Equating (5.72) and (5.73) gives the integral equation for $g_{c-1}(x)$,

$$\begin{aligned} &g_{c-1}(x) + (c-1)\,\mu \int_{y=-\infty}^{x} g_{c-1}(y)dy \\ &= c\mu \int_{y=0}^{\infty} \overline{A}(y-x)f(y)dy + g_{c-2}(0)\overline{A}(-x), x < 0. \end{aligned} \tag{5.74}$$

Integral Equations for PDF: Sheets $1, ..., c-2$

Consider the state-space interval $((-\infty, x), i), x < 0$ on sheet i where $i \in \{1, ..., c-2\}$ (Fig. 5.3). Reasoning as in the derivation of (5.74) for sheet $c-1$, we obtain integral equations

$$\begin{aligned} &g_i(x) + i\mu \int_{y=-\infty}^{x} g_i(y)dy \\ &= (i+1)\mu \int_{y=-\infty}^{x} g_{i+1}(y)dy + g_{i-1}(0)\overline{A}(-x), \\ &\qquad i = 1, ..., c-2, x < 0. \end{aligned} \tag{5.75}$$

In (5.75) the left side is the SP *exit* rate from $((-\infty, x), i)$. The right side is the SP *entrance* rate into $((-\infty, x), i)$.

Integral Equation for PDF: Sheet 0

Consider state-space interval $((-\infty, x), 0), x < 0$.

Exit Rate The SP can exit $((-\infty, x), 0), x < 0$ only by means of a (left) continuous hit of level x from below (upcrossing). The system is empty and no customer departures can occur, when $M(t) = 0$. Therefore the exit rate of $((-\infty, x), 0)$ is $g_0(x)$.

Entrance Rate The SP can enter $((-\infty,x),0)$ only by a parallel jump from $((-\infty,x),1)$ on sheet 1. That is, there must be *one* customer in service, that customer departs before any arrivals occur, and the remaining inter-arrival time is some $y>-x$, so that $y\in(-\infty,x)$. The rate of this occurrence is $1\cdot\mu\int_{y=-\infty}^{x}g_1(y)dy$.

Rate balance of exits and entrances of set $((-\infty,x),0)$ gives an integral equation for sheet 0,

$$g_0(x)=\mu\int_{y=-\infty}^{x}g_1(y)dy. \tag{5.76}$$

Form of $F(0)$

The probability of a potential wait of zero is given in (5.71). Here we shall not detail a procedure to compute $F(0)$ for the virtual wait in G/M/c for general values of c. However, in Subsection 5.3.1 below we provide a detailed derivation of $F(0)$ for the virtual wait in G/M/2.

5.2.5 Stability Condition for G/M/c

The stability condition for G/M/c follows directly from (5.64) and (5.67). The system is stable iff the steady-state pdf in (5.64) exists iff there exists a positive finite solution γ for equation (5.67). Using an analysis similar to that given in Proposition 5.4 for G/M/1, we obtain a necessary and sufficient condition for stability in G/M/c, namely

$$\boldsymbol{a}<c\mu.$$

5.2.6 Form of PDF of Actual Wait

In the following proposition, we use the principle that the "long run" proportion of *next* arrivals that have a property, is the same as the "overall" proportion of arrivals that have the same property.

Proposition 5.10 *For the G/M/c queue, the form of the pdf of actual wait is*

$$f(x)=K_\iota e^{-\gamma x},x>0, \tag{5.77}$$

where $K_\iota>0$.

Proof. The proportion of arrivals that wait $>x$ is

$$1-F_\iota(x)=\frac{c\mu(1-F(x))-f(x)}{c\mu\,(1-F(0))+\sum_{i=1}^{c-2}g_i(0)},x>0. \tag{5.78}$$

In equation (5.78) the term $F_\iota(x) = P(\text{actual wait} \le x)$; terms $F(x), f(x)$ are respectively the cdf and pdf of the virtual wait; $F(0) = P(\text{virtual wait} = \text{zero})$; $g_i(0), i = 1, ..., c-1$ are respectively the arrival rates when i customers are in service (see Subsection 5.2.4).

In the numerator of (5.78), $c\mu(1 - F(x))$ is the rate of downward jumps that start at levels $> x$, i.e., in $((x, \infty), c)$ (on sheet c). Thus $c\mu(1 - F(x))$ is the rate at which customers are in the system $> x$. It is also the rate at which *next* customers *wait in line* less than levels where the jumps started. The term $f(x)$ is the rate of such downward jumps that *end below* x. Thus $f(x)$ is the rate at which *next* customers wait $< x$. (Recall that $f(x)$ is the SP upcrossing rate of x, and $f(x)$ is also the downcrossing rate of x.) Thus the numerator is the rate at which *next* customers wait $> x$.

In the denominator, $c\mu\,(1 - F(0))$ is the rate of downward jumps that start on sheet c; $\sum_{i=1}^{c-2} g_i(0)$ is the rate of downward jumps that start at level 0 on sheets $1, ..., c-2$, combined. Thus, the denominator is the total rate of all downward jumps, which is precisely the total rate at which *next* customers start service.

Thus the right side of (5.78) is the *proportion* of downward jumps that start above level x and end above level x on sheet c. This is the same as the proportion of *next* customers that wait $> x$. Note that the explanation of (5.78) is similar to that in the proof of Proposition 5.2.

From equations (5.62) and ((5.64), we have $1 - F(x) = c_2 e^{-\gamma x}$ where c_2 is a positive constant, and $f(x) = Ke^{-\gamma x}$. Also, $c\mu\,(1 - F(0)) + \sum_{i=1}^{c-2} g_i(0)$ is a positive constant. Substituting into the right side of (5.78) and taking $\frac{d}{dx}$ on both sides of (5.78) yields (5.77) where K_ι is a positive constant. ■

5.2.7 Steady-state PDF of Actual Wait

Let W_q be the actual wait in line before service (arrival-point wait), in steady state. Let $F_\iota(0) = P(W_q = 0)$, and let the pdf of W_q be $f_\iota(x), x > 0$. The total rate at which *zero-waiting* customers arrive is equal to the total rate at which the SP hits level 0 from below, namely $\sum_{i=0}^{c-1} g_i(0)$ (see definition of $g_i(\cdot), i = 0, ..., c-1$ in Subsection 5.2.4). That is, $g_i(0)$ is the rate at which customers arrive at the system (remaining inter-arrival time $= 0$), when there are i customers in service, $i = 0, ..., c-1$.

Let $N_t, N_t^0, N_t^{>0}$ denote the total number of arrivals during $(0, t)$, the number of arrivals that wait 0 during $(0, t)$, and the number of arrivals that wait > 0 during $(0, t)$, respectively.

Consider a sample path of $\{V(t)\}$. Let $\mathcal{U}_t^i(x)$ denote the number of SP upcrossings of level x on sheet i during $(0,t)$, $i=0,...,c-1$. Then

$$\lim_{t\to\infty}\frac{\mathcal{U}_t^i(x)}{t} \underset{a.s.}{=} \lim_{t\to\infty}\frac{E\left(\mathcal{U}_t^i(x)\right)}{t} = g_i(x), x\le 0, i=0,...,c-1.$$

Note that $N_t^0=\sum_{i=0}^{c-1}\mathcal{U}_t^i(0)$.

The *proportion* of arrivals that wait 0 is

$$\begin{aligned}\lim_{t\to\infty}\frac{N_t^0}{N_t} &= \lim_{t\to\infty}\frac{N_t^0}{N_t^0+N_t^{>0}}\\ &= \frac{\lim_{t\to\infty}\frac{N_t^0}{t}}{\lim_{t\to\infty}\frac{N_t^0}{t}+\lim_{t\to\infty}\frac{N_t^{>0}}{t}}\\ &= \frac{\sum_{i=0}^{c-1}\lim_{t\to\infty}\frac{\mathcal{U}_t^i(0)}{t}}{\sum_{i=0}^{c-1}\lim_{t\to\infty}\frac{\mathcal{U}_t^i(0)}{t}+\lim_{t\to\infty}\frac{N_t^{>0}}{t}}\\ &= \frac{\sum_{i=0}^{c-1}g_i(0)}{\sum_{i=0}^{c-1}g_i(0)+\lim_{t\to\infty}\frac{N_t^{>0}}{t}}.\end{aligned} \tag{5.79}$$

In the denominator of (5.79), the *rate* at which arrivals wait a positive time before service is

$$\begin{aligned}&\lim_{t\to\infty}\frac{N_t^{>0}}{t} \underset{a.s.}{=} \lim_{t\to\infty}\frac{E\left(N_t^{>0}\right)}{t}\\ &= c\mu\int_{y=0}^{\infty}A(y)f(y)dy\\ &= c\mu\int_{y=0}^{\infty}A(y)Ke^{-\gamma y}dy\\ &= c\mu\int_{y=0}^{\infty}\left(1-\overline{A}(y)\right)Ke^{-\gamma y}dy\\ &= \frac{c\mu}{\gamma}K-K,\end{aligned} \tag{5.80}$$

upon utilizing (5.64) and (5.67). That is, $c\mu\int_{y=0}^{\infty}A(y)f(y)dy$ is the rate at which customers depart after being in the system for a time y, and the immediately *next* inter-arrival time is $< y$, summed over all $y>0$. Then $c\mu\int_{y=0}^{\infty}A(y)f(y)dy$ is the rate at which *next* customers that enter service wait a positive time. Substituting from (5.80) into (5.79) gives

$$F_\iota(0)=\frac{\sum_{i=0}^{C-1}g_i(0)}{\sum_{i=0}^{C-1}g_i(0)+\frac{c\mu}{\gamma}K-K}. \tag{5.81}$$

In (5.74) let $x \uparrow 0$. Note that the SP exit rate from sheet $c-1$ across level 0 is equal to the SP entrance rate of interval $((0,\infty),c)$ (sheet c). Thus

$$\lim_{t\to\infty} \frac{E\left(\mathcal{U}_t^{c-1}(0)\right)}{t} = g_{c-1}(0) = f(0) = K.$$

Here we do not carry out the procedure to compute $F_\iota(0)$ for general values of c (equation (5.81)). In Subsection 5.3.2 below we derive $F_\iota(0)$ explicitly for G/M/2, to indicate the computational procedure.

5.3 G/M/2: PDF of Virtual and of Actual Wait

We derive the steady-state pdf of the virtual wait and of the actual wait for G/M/2. Consider the process $\{V(t), M(t)\}$. When $c = 2$, $M(t) \in \boldsymbol{M} = \{0,1,2\}$. Graphically, there are three corresponding sheets in $\boldsymbol{T} \times \boldsymbol{S}$ labeled $0,1,2$. (Fig. 5.3). The analyses below are examples of the type of solution approach that may be used for $c = 3,4,\ldots$. (The results for $c = 2$ are applied in [66].)

5.3.1 PDF of Virtual Wait

In G/M/2 the pdf of the virtual wait has the same form as in the general G/M/c model,

$$f(x) = Ke^{-\gamma x}, x > 0.$$

We repeat the integral equations for sheets 1 and 0 respectively for convenience,

$$\begin{aligned} g_1(x) + \mu \textstyle\int_{y=-\infty}^{x} g_1(y)dy = \quad & 2\mu K \textstyle\int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy \\ & + g_0(0)\overline{A}(-x), x < 0, \end{aligned} \tag{5.82}$$

and

$$g_0(x) = \mu \int_{y=-\infty}^{x} g_1(y)dy, \tag{5.83}$$

as in equations (5.74) and (5.76).

Also $g_1(0) = K$. The proportion of time that the system has less than 2 customers in service is

$$F(0) = \int_{x=-\infty}^{0} (g_1(x) + g_0(x))dx = 1 - \frac{K}{\gamma}, \tag{5.84}$$

as in (5.70).

Adding corresponding sides of (5.82) and (5.83) and integrating with respect to $x \in (-\infty, 0)$, gives

$$\begin{aligned} F(0) &\equiv \int_{x=-\infty}^{0} (g_1(x) + g_0(x))dx \\ &= 2\mu K \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y} dy dx + g_0(0)\frac{1}{\boldsymbol{a}}, \quad (5.85) \end{aligned}$$

where $\frac{1}{\boldsymbol{a}} = \int_{u=0}^{\infty} \overline{A}(u)du$ is the mean arrival time.

Taking $\frac{d}{dx}$ in (5.83) gives the relation

$$g_1(x) = \frac{g_0'(x)}{\mu}. \quad (5.86)$$

Substituting (5.86) and (5.83) into (5.82) gives a differential equation for $g_0(x)$

$$g_0'(x) + \mu g_0(x) = 2\mu^2 K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y} dy + +\mu g_0(0)\overline{A}(-x), x < 0. \quad (5.87)$$

The solution of (5.87) is

$$\begin{aligned} g_0(x) = 2\mu^2 K e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z} \int_{y=0}^{\infty} \overline{A}(y-z)e^{-\gamma y} dy dz \\ + \mu g_0(0)e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z}\overline{A}(-z)dz, x < 0, \quad (5.88) \end{aligned}$$

upon noting that the constant of integration is 0 because $\lim_{x\downarrow-\infty} g_0(x) = 0$ and $\lim_{x\downarrow-\infty} \int_{z=-\infty}^{x} (\cdots)\, dx = 0$.

Note that $\lim_{x\uparrow 0} e^{-\mu x} = e^0 = 1$. In (5.88) letting $x \uparrow 0$ gives an equation for $g_0(0)$ in terms of K (after making the transformation $u = -z$)

$$\begin{aligned} g_0(0) &= 2\mu^2 K \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y} dy du \\ &\quad +\mu g_0(0) \int_{u=0}^{\infty} e^{-\mu u}\overline{A}(u)du, \end{aligned}$$

or

$$\begin{aligned} g_0(0) &= \left(\frac{2\mu^2 \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y} dy du}{1 - \mu \int_{u=0}^{\infty} e^{-\mu u}\overline{A}(u)du,} \right) K \\ &\equiv H_0 \cdot K. \quad (5.89) \end{aligned}$$

Equation (5.89) defines the constant H_0, which is independent of K.

We now obtain an equation for K. From (5.84) and (5.82),

$$\begin{aligned} F(0) &= 1 - \frac{K}{\gamma} \\ &= 2\mu K \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x) e^{-\gamma y} dy dx + H_0 K \frac{1}{\boldsymbol{a}}. \end{aligned} \tag{5.90}$$

Solving (5.90) for K gives

$$K = \frac{1}{\frac{1}{\gamma} + 2\mu \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x) e^{-\gamma y} dy dx + H_0 \cdot \frac{1}{\boldsymbol{a}}}. \tag{5.91}$$

where H_0 is defined in (5.89).

Thus

$$\begin{aligned} F(0) &= 1 - \frac{K}{\gamma} \\ &= 1 - \frac{1}{1 + 2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x) e^{-\gamma y} dy dx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}} \\ &= \frac{2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x) e^{-\gamma y} dy dx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}{1 + 2\mu\gamma \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x) e^{-\gamma y} dy dx + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}} \\ &= \frac{2\mu\gamma \int_{u=0}^{\infty} \int_{y=0}^{\infty} \overline{A}(y+u) e^{-\gamma y} dy du + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}{1 + 2\mu\gamma \int_{u=0}^{\infty} \int_{y=0}^{\infty} \overline{A}(y+u) e^{-\gamma y} dy du + H_0 \cdot \frac{\gamma}{\boldsymbol{a}}}, \end{aligned} \tag{5.92}$$

upon making the transformation $u = -x$.

The pdf of the *virtual wait* is $\{F(0); f(x), x > 0\}$, where $f(x) = Ke^{-\gamma x}, x > 0$ and K is specified in (5.91). The probability of a zero wait $F(0)$, is given by (5.92).

5.3.2 PDF of Actual Wait

Equation (5.64) becomes

$$f_\iota(x) = K_\iota e^{-\gamma x}, x > 0,$$

where

$$K_\iota = \frac{1 - F_\iota(0)}{E(\mathcal{B}_{1,2})}, \; \gamma = \frac{1}{E(\mathcal{B}_{1,2})}, \; F_\iota(0) = P_{0\iota} + P_{1\iota}.$$

From (5.81) the *proportion of arrivals* that wait 0 is

$$F_\iota(0) = \frac{\sum_{i=0}^{1} g_i(0)}{\sum_{i=0}^{1} g_i(0) + \frac{2\mu}{\gamma}K - K}. \tag{5.93}$$

Taking $\frac{d}{dx}$ on both sides of (5.82) gives an ordinary differential equation for $g_1(x)$ with solution

$$\begin{aligned} e^{\mu x} g_1(x) &= 2\mu \int_{z=-\infty}^{x} e^{\mu z} \int_{y=0}^{\infty} a(y-x) K e^{-\gamma y} dy dz \\ &\quad + g_0(0) \int_{z=-\infty}^{x} e^{\mu z} a(-z) dz + H_1, \end{aligned} \tag{5.94}$$

where H_1 is a constant. Note that necessarily $\lim_{x\downarrow-\infty} g_1(x) = 0$; this helps to evaluate H_1 That is $\lim_{x\downarrow-\infty} e^{\mu x} g_1(x) = 0$. Also

$$\lim_{x\downarrow-\infty} \int_{z=-\infty}^{x} (\cdots) dz = 0.$$

Thus $H_1 = 0$.

Additionally $\lim_{x\uparrow 0} e^{\mu x} g_1(x) = g_1(0) = f(0) = K$. Letting $x \uparrow 0$ in (5.94) yields

$$g_0(0) = K \cdot B_0, \tag{5.95}$$

where

$$B_0 = \frac{1 - 2\mu \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} a(y+u) e^{-\gamma y} dy du}{\int_{u=0}^{\infty} e^{-\mu u} a(u) du}, \tag{5.96}$$

using the transformation $u = -z$.

Thus

$$g_1(0) + g_0(0) = K + KB_0,$$

with B_0 given in (5.96).

From (5.93)

$$\begin{aligned} F_\iota(0) &= \frac{\sum_{i=0}^{1} g_i(0)}{\sum_{i=0}^{1} g_i(0) + \frac{c\mu}{\gamma}K - K} \\ &= \frac{K + KB_0}{K + KB_0 + \frac{2\mu}{\gamma}K - K} = \frac{1 + B_0}{B_0 + \frac{2\mu}{\gamma}}, \end{aligned} \tag{5.97}$$

which is independent of K.

We then calculate K_ι from the normalizing condition

$$\begin{aligned} F_\iota(0)+\int_{x=0}^{\infty} f_\iota(x)dx &= 1, \\ F_\iota(0)+\int_{x=0}^{\infty} K_\iota e^{-\gamma x}dx &= 1. \end{aligned}$$

Applying (5.97) gives

$$\frac{1+B_0}{B_0+\frac{2\mu}{\gamma}}+\frac{K_\iota}{\gamma}=1$$

which yields

$$K_\iota=\gamma\left(\frac{2\mu-\gamma}{2\mu+\gamma B_0}\right)=\gamma\left(1-F_\iota(0)\right). \tag{5.98}$$

Thus

$$F_\iota(0)=1-\frac{K_\iota}{\gamma}=1-\left(\frac{2\mu-\gamma}{2\mu+\gamma B_0}\right)=\frac{\gamma\left(1+B_0\right)}{2\mu+\gamma B_0}. \tag{5.99}$$

5.3.3 Reduction of G/M/2 PDF to M/M/2 PDF

To enhance intuition, we check that the G/M/2 pdf for the actual wait, given above, reduces to the M/M/c pdf given in (4.53), (4.54) and (4.55) when $c=2$. In M/M/2 let P_0, P_1 be the steady-state probabilities of 0 units and 1 unit in the system, respectively. For M/M/2 the pdf's of the virtual wait and actual wait are the same, due to Poisson arrivals. We show that for G/M/2 with Poisson arrivals, $F_\iota(0)=P_0+P_1$.

From the standard formulas for M/M/c, we have the pdf of wait in M/M/2, namely

$$\begin{aligned} P_0 &= \frac{1}{1+\frac{\lambda}{\mu}+\frac{\lambda^2}{\mu(2\mu-\lambda)}} \\ P_1 &= \frac{\lambda}{\mu}P_0 \\ f(x) &= \lambda P_1 e^{-(2\mu-\lambda)x}, x>0. \end{aligned} \tag{5.100}$$

From (5.100), in M/M/2 P_0+P_1 simplifies to

$$P_0+P_1=\frac{(2\mu-\lambda)\left(\lambda+\mu\right)}{\lambda\mu+2\mu^2}. \tag{5.101}$$

To obtain these values from G/M/2, we first specialize the G/M/2 formula for B_0 in (5.96) to M/M/2, by letting $a(z)=\lambda e^{-\lambda z}, z>0$, and set $\gamma = 2\mu - \lambda$. This substitution yields $B_0 = \frac{\mu}{\lambda}$. Combining with (5.99) we get

$$F_\iota(0) = \frac{(2\mu - \lambda)(\lambda + \mu)}{\lambda\mu + 2\mu^2} \tag{5.102}$$

in agreement with (5.101).

The pdf is

$$\begin{aligned} f_\iota(x) &= K_\iota e^{-\gamma x} = \gamma(1 - F_\iota(0))\, e^{-(2\mu-\lambda)x} \\ &= \lambda P_1 e^{-(2\mu-\lambda)x}, x > 0, \end{aligned} \tag{5.103}$$

since $\gamma = 2\mu - \lambda$ and

$$\begin{aligned} \gamma(1 - F_\iota(0)) &= (2\mu - \lambda)\frac{(2\mu - \lambda)(\lambda + \mu)}{\lambda\mu + 2\mu^2} \\ &= \lambda\frac{\lambda(2\mu - \lambda)}{\lambda\mu + 2\mu^2} = \lambda P_1. \end{aligned}$$

Hence the G/M/2 pdf $\{F_\iota(0); f_\iota(x), x > 0\}$ in (5.102) and (5.103), when the arrival rate is Poisson at rate λ, agrees with the M/M/2 pdf.

5.3.4 Moments of Actual Wait for G/M/2

All statistical moments (about 0) of W_q can be found using

$$E(W_q^n) = \int_{y=0}^{\infty} y^n K_\iota e^{-\gamma y} dy = K_\iota \frac{n!}{\gamma^{n+1}}, n = 0, 1, 2, ...,$$

where K_ι is given in (5.98). In particular the mean and variance of the actual wait are

$$E(W_q) = \frac{K_\iota}{\gamma}, \quad Var(W_q) = \frac{K_\iota(2\gamma - K_\iota)}{\gamma^4}.$$

The Laplace-Stieltjes transform of the actual wait is

$$F_\iota(0)e^{-s.0} + \int_{y=0}^{\infty} e^{-sy} K_\iota e^{-\gamma y} dy = F_\iota(0) + \frac{K_\iota}{s + \gamma}, s > 0.$$

5.3.5 Discussion

Heavy-tailed Inter-arrivals

For the LC analysis of G/M/c the inter-arrival times may have a **heavy-tailed distribution**. For example, the inter-arrival times may have a

Pareto distribution with

$$A(x) = 1 - \frac{1}{(1+x)^{\beta}}, \overline{A}(x) = \frac{1}{(1+x)^{\beta}}, \; a(x) = \frac{\beta}{(1+x)^{\beta+1}}, x \geq 0,$$

where β is the shape parameter. All moments exist up to $\lceil \beta - 1 \rceil$, where $\lceil u \rceil$ denotes the smallest integer $\geq u$. The LC solution technique outlined in the present section applies because the solution for γ depends only on the complementary cdf $\overline{A}(\cdot)$, the probability of the *tail of distribution*, and not on whether the mean and variance exist.

Similar remarks apply to inter-arrival times which have a folded Cauchy, or inverse-log distribution, etc. Additional LC results for heavy-tailed inter-arrival times are given in [66].

Model Variants

The LC solution technique in this section is useful for analyzing models with state dependence. For example, inter-arrival times and/or service rates of arrivals, may depend on the number of customers in service, or on the system time of the last departure from the system. LC can be used to analyze other generalizations, e.g., bounded workload, or service rate depending on waiting time. In generalized models, we could derive integral equations for the pdf of wait in a similar manner as above for the standard G/M/c or G/M/1 queue, e.g., as in [15].

CHAPTER 6

DAMS AND INVENTORIES

6.1 Introduction

In this chapter we analyze several models of dams and inventories with state space $\boldsymbol{S} \subseteq \boldsymbol{R}$, using LC. When the content in a dam, or stock on hand in an inventory, is positive-valued, it can decline at varying instantaneous rates in accordance with a general release rule specified in the model. Thus the efflux differs from the virtual wait or workload in M/G/1 queues, which decreases at rate 1 when positive, or the extended age in G/M/c queues, which increases at rate 1.

Section 6.2 describes a model of a dam with general release rule, denoted by M/G/$r(\cdot)$ (or "M/G/1 dam"). The function $r(x), x \geq 0$, denotes the *efflux rate* when the content is at level x. We discuss sample-path and SP transitions in the time-state space, and derive *integro-differential* equations for the *transient* (time-dependent) distribution of the content. The subscript "t" is used to indicate transience. Integral equations for the *steady-state* distribution of content are then obtained by taking limits as $t \to \infty$.

Sections 6.3 – 6.9 apply SPLC to analyze several models of dams and inventories in steady state.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_6,

6.2 M/G/r(·) Dam

6.2.1 Model Description

Consider a dam with state space $\boldsymbol{S} = [0, \infty)$. Denote the content at instant t by $W(t), t \geq 0$. Assume inputs occur at a Poisson rate λ. Denote the instants of input by $\tau_n, n = 1, 2, ...$, where $0 \equiv \tau_0 < \tau_1 < \tau_2 < \cdots$. Denote the input size at τ_n by S_n,$n = 1, 2, ...$. We assume $\{S_n, n = 1, 2, ...\}$ are iid positive r.v.'s independent of n. Let $S \underset{dist}{\equiv} S_n$. Let $B(x) = P(S \leq x)$, $\overline{B}(x) = 1 - B(x)$.

In some variants, the input size may depend on the content $W(\tau_n^-)$ at input instant τ_n^- (denoted by $S(W(\tau_n^-))$, or on a Markovian environment (for example $S_{(i)}$ where i is a state of a continuous-time Markov chain). Other input-time dependencies are also possible.

If S depends on the current content only, the conditional cdf of $S(W(\tau_n^-))$, given $W(\tau_n^-)$, is denoted by

$$B_y(x) = P(S(W(\tau_n^-)) \leq x | W(\tau_n^-) = y), y \geq 0, n = 1, 2, ...$$

The efflux rate of content out of the dam, is denoted by $r(W(t))$, defined in Subsection 6.2.2 below. Generally, the efflux rate depends on the current content.

In M/G/r(·), we assume that the entire input amount goes into the dam instantaneously at an input instant. Under this assumption the model applies to some real-world situations, e.g., systems involving torrential rainfalls, repeated shocks, bolus injections of a drug in pharmacokinetics, instillment of certain eye drops, consumer response to a particular product when exposed to repeated non-uniform advertising in marketing-science models, etc.

We discuss variants and generalizations of this model in later subsections.

6.2.2 Efflux Rate

Let $r(W(t))$ denote the efflux rate at which the content decreases (flows out of the dam) at instant t, when the content is $W(t)$. Assume $r(W(t))$ is finite and

$$\begin{matrix} r(x) > 0 \\ r(x) = 0 \end{matrix} \quad \text{if} \quad \begin{cases} x > 0, \\ x = 0. \end{cases} \tag{6.1}$$

The rate of decline of $W(t)$ between input instants is

$$\frac{dW(t)}{dt} = -r(W(t)), \tau_n \leq t < \tau_{n+1}, n = 0, 1, 2, ... \tag{6.2}$$

The efflux rate $r(W(t))$ has physical dimension $\frac{[content\ unit]}{[Time]}$, e.g., $\frac{[liters]}{[Time]}$ if the content is measured in liters.

This section assumes that $r(x), x \in \boldsymbol{S}$ is a time-homogeneous piecewise right-continuous function, except at level 0. Usually $r(0) \neq r(0^+) = \lim_{x \downarrow 0} r(x)$. However, equality is possible in some models.

Example 6.1 *Suppose* $r(x) = (x+1)^2, x > 0,\ r(0) = 0$. *Then* $r(0^+) = 1 \neq r(0)$. *On the other hand, suppose* $r(x) = x^2, x > 0,\ r(0) = 0$. *Then* $r(0^+) = r(0)$.

Consider a state-space partition $\{x_j\}$ where $0 \equiv x_0 < x_1 < x_2 < \cdots < x_n < x_{n+1} \equiv \infty$. Let $\boldsymbol{I}_1 = (x_0, x_1)$, $\boldsymbol{I}_j = [x_{j-1}, x_j)\,, j = 2, ..., n+1$. Define $\{r_j(x)\}$ by

$$r(x) = \begin{cases} r_0(0) = 0 \\ r_1(x), x \in (0, x_1) \equiv \boldsymbol{I}_1 \\ r_2(x), x \in [x_1, x_2) \equiv \boldsymbol{I}_2 \\ \quad\quad \cdots \\ r_n(x), x \in [x_{n-1}, x_n) \equiv \boldsymbol{I}_n \\ r_{n+1}(x), x \in [x_n, \infty) \equiv \boldsymbol{I}_{n+1} \end{cases} \tag{6.3}$$

where $r_j(x), x \in \boldsymbol{I}_j$ is positive and continuous, $j = 1, 2, ..., n+1$.

Remark 6.1 *In model generalizations* $r(W(t))$ *may also depend on* t. *We would then append a subscript* t, *e.g., denote the efflux rate by* $r_t(W(t))$.

6.2.3 Sample Paths

We use the symbol "$W(t)$" to represent either the state random variable (content) or the value of a sample path at instant t (unless specified otherwise). This is for economy of notation, and because the usage will be clear from the context.

A sample path of $\{W(t), t \geq 0\}$ is a piecewise deterministic function plotted in the *time-state* plane $\boldsymbol{T} \times \boldsymbol{S}$, where $\boldsymbol{T} = \{t | t \geq 0\}$ (Fig. 6.1).

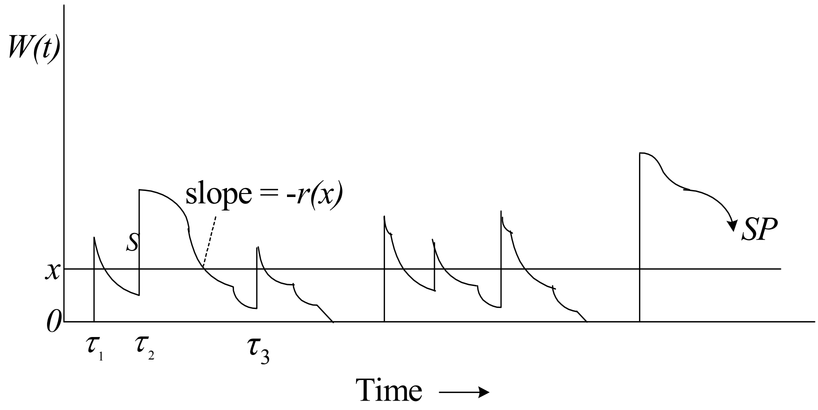

Figure 6.1: Sample path for $M/G/r(\cdot)$ dam in the time-state plane

6.2.4 Time for Content to Decrease to a Level

In equation (6.2), separating variables and integrating both sides, gives a measure of the *time* required for a sample path to decrease from a level $W(t_y) \equiv y$ at instant t_y to a lower level $W(t_x) \equiv x$ at instant t_x, if no inputs to the dam intervene. Necessarily, $t_x > t_y$. Fix n. Assume

$$W(\tau_n) > y > x > W(\tau_n^-) > 0.$$

The time to descend from level y to level x is

$$t_x - t_y = \int_{z=x}^{y} \frac{1}{r(z)} dz. \tag{6.4}$$

Formula (6.4) is useful when analyzing models of dams and inventories in continuous time (as in this chapter). Formula (6.4) is also useful when analyzing the model of a dam by the embedded level crossing method, which compares the state at successive input times (Chapter 8).

6.2.5 Transient Probability Distribution of Content

Transient Distribution

Denote the transient cdf of $W(t)$ by $F_t(x), x \geq 0$. Let $P_0(t) = F_t(0)$, and let $f_t(x), x > 0$, denote the transient pdf of $W(t)$, given by

$$f_t(x) = \frac{dF_t(x)}{dx}, x > 0, t > 0,$$

wherever the derivative exists. Assume $F_t(x)$, $f_t(x)$ are right continuous in x. We define $f_t(0^+) \equiv f_t(0)$ for notational convenience; $f_t(0)$ adds zero probability to $P_0(t)$. The pdf $f_t(x)$ may have jump discontinuities depending on the distribution of the input r.v. S (e.g., see sections 3.8, 3.9 regarding the pdf of wait in M/D/1 and M/Discrete/1 queues.)

For each $t \geq 0$,

$$F_t(x) = P_0(t) + \int_{y=0}^{x} f_t(y)dy, x \geq 0.$$

For each $t \geq 0$, the normalizing condition is

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

Steady-state Distribution

Assume the steady-state distribution of content exists. The steady-state cdf and pdf of content are respectively denoted by $F(x), x \geq 0$, and $\{P_0; f(x), x > 0\}$. Thus

$$F(x) = \lim_{t\to\infty} F_t(x), x \geq 0, \;\; f(x) = \lim_{t\to\infty} f_t(x), x > 0, \;\; P_0 = \lim_{t\to\infty} P_0(t).$$

6.2.6 Sample-path and SP Downcrossings

Consider a sample path of $\{W(t)\}$. Fix level $x \in \boldsymbol{S}$. Let $\mathcal{D}_t(x)$ denote the number of SP downcrossings of level x during $(0, t)$. SP downcrossings include left-continuous sample-path downcrossings. The SP traces the sample path during piecewise continuous segments between input instants. At sample-path discontinuities, the SP makes an upward jump, *not in Time* (see subsections 2.4.2, 2.4.3). Let $\mathcal{D}_t^c(x)$ and $\mathcal{D}_t^j(x)$ denote respectively the number of SP left-continuous downcrossings and SP jump downcrossings of level x during $(0, t)$. Then

$$\mathcal{D}_t(x) = \mathcal{D}_t^c(x) + \mathcal{D}_t^j(x), x \geq 0, t \geq 0.$$

In the basic M/G/r(·) dam of this section, $\mathcal{D}_t^j(x) \equiv 0, t \geq 0$. In variations of the basic model, however, SP downward jumps can indeed occur. Both SP left-continuous downcrossings and SP jump downcrossings also occur in a vast number of *inventory* and *production-inventory* models. Thus, we shall distinguish $\mathcal{D}_t(x)$ from $\mathcal{D}_t^c(x)$ in Theorem 6.1 below.

6.2.7 Level Crossings and Distribution of Content

Consider a sample path of $\{W(t)\}$. Fix level $x \in \boldsymbol{S}$. Let $\mathcal{U}_t(x)$ denote the number of SP *upcrossings* of level x during $(0,t)$. It can be shown, along the lines of subsections 3.2.1 and 3.2.2, that $\frac{\partial}{\partial t}E(\mathcal{D}_t^c(x))$, $\frac{\partial}{\partial t}E(\mathcal{U}_t(x))$ exist and are positive.

Theorems 6.1 and 6.2 were originally proved using LC in [19].

Downcrossings

Theorem 6.1 *For the $M/G/r(\cdot)$ dam*

$$\frac{\partial}{\partial t}E(\mathcal{D}_t^c(x)) = r(x)f_t(x), x > 0, \tag{6.5}$$

$$\frac{\partial}{\partial t}E(\mathcal{D}_t^c(0)) = r(0^+)f_t(0). \tag{6.6}$$

Proof. Consider a sample path of $\{W(t)\}$, and fix state-space level $x \in \boldsymbol{I}_j$ for some $j = 1,..,n+1$. Fix instant t. Consider $t+h$, $(h > 0)$ and define $\delta > 0$ by

$$\int_{z=x}^{x+\delta} \frac{1}{r(z)}dz = h. \tag{6.7}$$

Assume h is sufficiently small so that level $x + \delta \in \boldsymbol{I}_j$. That is, h is the time for the content to decrease from level $x+\delta$ to level x if there are no inputs during $(t, t+h)$ (see equation (6.4)). Applying the law of the mean value for integrals with continuous integrand to equation (6.7) yields

$$h = \frac{1}{r(z^*)}\delta \iff \delta = r(z^*)h \tag{6.8}$$

for some z^* such that $x < z^* < x + \delta$.

The event $\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x) = 1$ occurs iff $W(t) \in (x, x+\delta)$ and there is no input in a time subinterval $(t, t+\xi) \subseteq (t, t+h)$, or an event with probability $o(h)$ occurs. From (6.8)

$$\begin{aligned} P(\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x) = 1) &= f_t(x) \cdot \delta \cdot (1 - \lambda h) + o(h) \\ &= f_t(x) \cdot r(z^*) \cdot h \cdot (1 - \lambda h) + o(h). \end{aligned}$$

The value $\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x) = 0$ has no affect on $E(\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x))$. Also, due to the Poisson input stream, $P(\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x) \geq 2) = o(h)$. Hence the expected value

$$\begin{aligned} E(\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x)) &= 1 \cdot P(\mathcal{D}_{t+h}^c(x) - \mathcal{D}_t^c(x) = 1) + o(h), \\ E(\mathcal{D}_{t+h}^c(x)) - E(\mathcal{D}_t^c(x)) &= f_t(x) \cdot r(z^*) \cdot h \cdot (1 - \lambda h) + o(h). \end{aligned} \tag{6.9}$$

Dividing both sides of (6.9) by h and letting $h \downarrow 0$ gives (6.5) since $z^* \downarrow x$ and $r(z^*) \downarrow r(x^+) = r(x), x > 0$, as $h \downarrow 0$. Then letting $x \downarrow 0$ in (6.5) gives (6.6). ■

Corollary 6.1 *For each* $t \geq 0$,

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

Proof. In (6.5) and (6.6) set $t = s$, integrate with respect to $s \in [0, t]$, and apply the initial condition $E(\mathcal{D}_0^c(x)) \equiv 0, x \geq 0$. ■

Corollary 6.2

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

Proof. In Corollary 6.1 divide both sides by $t > 0$ and take limits as $t \to \infty$. ■

Upcrossings

Theorem 6.2 *For the M/G/r(·) dam*

$$\frac{\partial}{\partial t} E(\mathcal{U}_t(x)) = \lambda \int_{z=0}^{x} \overline{B}(x - z) dF_t(z)$$
$$= \lambda P_0(t) \overline{B}(x) + \lambda \int_{z=0}^{x} \overline{B}(x - z) f_t(z) dz, x > 0, \tag{6.10}$$
$$\frac{\partial}{\partial t} E(\mathcal{U}_t(0)) = \lambda P_0(t). \tag{6.11}$$

Proof. Fix instants t and $t + h, t \geq 0, h > 0$ (h small). Fix level $x > 0$. Then $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) = 1$ iff $W(s) = z < x$ at an instant $s \in (t, t + h)$ and there is an input of size $S > x - z$, or an event having probability $o(h)$ occurs. The value $\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) = 0$ does not contribute to

$E(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x))$. Also $P(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x) \geq 2) = o(h)$. Hence

$$\begin{aligned} E(\mathcal{U}_{t+h}(x) - \mathcal{U}_t(x)) &= E(\mathcal{U}_{t+h}(x)) - E(\mathcal{U}_t(x)) \\ &= \lambda \int_{z=0}^{x} \int_{s=t}^{t+h} \overline{B}(x-z) ds dF_{t+s}(z) + o(h) \\ &= \lambda h \int_{z=0}^{x} \overline{B}(x-z) dF_{t+s^*}(z) + o(h) \end{aligned} \tag{6.12}$$

where $t < s^* < t+h$. Dividing both sides of (6.12) by h and letting $h \downarrow 0$ gives (6.10) since $s^* \downarrow 0$ as $h \downarrow 0$, and $F_t(\cdot)$ is right-continuous in t. Then letting $x \downarrow 0$ in (6.10) gives (6.11). ■

Corollary 6.3

$$\begin{aligned} E(\mathcal{U}_t(x)) &= \lambda \int_{s=0}^{t} \int_{z=0}^{x} \overline{B}(x-z) dF_s(z) ds \\ &= \lambda \int_{s=0}^{t} P_0(s)\overline{B}(x) ds + \lambda \int_{s=0}^{t} \int_{z=0}^{x} \overline{B}(x-z) f_s(z) dz ds, x > 0, \\ E(\mathcal{U}_t(0)) &= \lambda \int_{s=0}^{t} P_0(s)\overline{B}(x) ds. \end{aligned}$$

Proof. Set $t = s$ in (6.10) and (6.11), integrate with respect $s \in [0, t]$ and apply the initial condition $E(\mathcal{U}_0(x)) = 0, x \geq 0$. ■

6.2.8 Equation for Transient Distribution of Content

The following theorem has been proved using classical methods by various authors. Here it is proved in an alternative manner using a level crossing method (based on [19]).

Theorem 6.3 *For the $M/G/r(\cdot)$ dam, the transient distribution of content satisfies the integro-differential **equation** for $f_t(x)$*

$$\begin{aligned} r(x) f_t(x) &= \tfrac{\partial}{\partial t} F_t(x) + \lambda \textstyle\int_{z=0}^{x} \overline{B}(x-z) dF_t(z) \\ &= \tfrac{\partial}{\partial t} F_t(x) + \lambda \overline{B}(x) P_0(t) \\ &\quad + \lambda \textstyle\int_{z=0}^{x} \overline{B}(x-z) f_t(z) dz, x > 0, \end{aligned} \tag{6.13}$$

*and **differential equation** for $P_0(t)$*

$$\frac{\partial}{\partial t} P_0(t) + \lambda P_0(t) = r(0^+) f_t(0). \tag{6.14}$$

Proof. In Theorem 4.1 (Theorem B), substitute set $[0,x] = \boldsymbol{A}$, $\mathcal{D}_t^c(x) = \mathcal{I}_t(x)$, $\mathcal{U}_t(x) = \mathcal{O}_t(x)$. This gives

$$\frac{\partial}{\partial t}E(\mathcal{D}_t^c(x)) = \frac{\partial}{\partial t}F_t(x) + \frac{\partial}{\partial t}E(\mathcal{U}_t(x)) \tag{6.15}$$

Substituting from (6.5) and (6.10) into (6.15) gives (6.13). Equation (6.14) then follows by letting $x \downarrow 0$ in (6.13), noting that $F_t(0) = P_0(t)$. ■

Remark 6.2 *The dimension of* $r(x)$ *is* $\left[\frac{content\ unitt}{Time}\right]$. *The dimension of* $f_t(x)$ *is* $\left[\frac{1}{content\ unit}\right]$. *The dimension of the left sides of (6.13) and of (6.14), is*

$$\begin{aligned}[r(x)f_t(x)] &= \left[\frac{content\ unitt}{Time}\right]\left[\frac{1}{content\ unit}\right]\\ &= \frac{1}{[Time]}, x \geq 0,\end{aligned}$$

which matches the dimensional unit of the right side.

6.2.9 Estimate of Transient Probability of Emptiness

This subsection briefly outlines a level crossing estimation procedure for the transient probability of the dam being empty. (We also call this procedure LC estimation or computation.) A description of LC estimation of the complete transient distribution would follow similar lines. (We do not expound on LC estimation of *transient* distributions in the present monograph. See Remark 9.2 in Chapter 9.)

Solving differential equation (6.14) yields a formula for $P_0(t), t \geq 0$, in terms of

$$r(0^+)f_t(0) = \frac{\partial}{\partial t}E(\mathcal{D}_t^c(0)).$$

Thus we obtain

$$\begin{aligned}P_0(t) &= e^{-\lambda t}\int_{s=0}^{t} e^{\lambda s} r(0^+)f_t(0)ds + P_0(0)e^{-\lambda t},\\ P_0(t) &= \left(\int_{s=0}^{t} e^{\lambda s}\frac{\partial}{\partial s}E(\mathcal{D}_s^c(0))ds + P_0(0)\right)e^{-\lambda t},\end{aligned} \tag{6.16}$$

,where

$$P_0(0) = \begin{cases} 1 \text{ if } W(0) = 0, \\ 0 \text{ otherwise.}\end{cases}$$

Formula (6.16) connects $P_0(t)$ and $\frac{\partial}{\partial s}E(\mathcal{D}_s^c(0)), 0 < s < t$, which appears as a factor in the integrand. This connection leads to an *estimation method* for $P_0(t)$. That is, we wish to estimate the integral in (6.16), resulting in an estimate of $P_0(t)$.

The idea is to first simulate N *independent* sample paths, $\{W_n(s), n = 1, ..., N\}$ on the same time interval $[0, t_M + r]$, where t_M is the maximum time of interest, r is an "extra" time which ensures that t_M is not the right end point of the simulated time interval, and N is a large positive integer. A reasonable value of N would be in the range $(400, 1000)$. Due to the high speed of today's computers, N may be considerably larger than 1000. Let $h = \frac{t_M}{m}$ be small, where m is a positive integer. We can use, e.g., $h = 0.01$ or 0.001, or any small value $h < r$. The accuracy of the estimated $P_0(\cdot)$ values improves with larger values of N combined with smaller values of h.

We then compute the number of SP left-continuous downcrossings (hits of level 0) $\mathcal{D}_{ih,n}^c(0), i = 0, ..., m$, for each sample path, $n = 1, ..., N$. For fixed i and n, the $\mathcal{D}_{ih,n}^c(0)$ values are independent since the N sample paths are independent. We compute point estimates of the true SP downcrossing rates at times ih and $(i+1)\,h$ by averaging over the N sample paths. Thus we compute

$$\widehat{E}(\mathcal{D}_{ih}^c(0)) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{D}_{ih,n}^c(0), \quad \widehat{E}(\mathcal{D}_{(i+1)h}^c(0)) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{D}_{(i+1)h,n}^c(0).$$

An estimate of the *derivative* $\frac{\partial}{\partial t}E(\mathcal{D}_{ih}^c(0))$ is then given by

$$\widehat{\frac{\partial}{\partial t}}E(\mathcal{D}_{ih}^c(0)) = \frac{\widehat{E}(\mathcal{D}_{(i+1)h}(0)) - \widehat{E}(\mathcal{D}_{ih}^c(0))}{h}, i = 0, ..., m.$$

Finally, we approximate the integral $\int_{s=0}^{kh} e^{\lambda s}\frac{\partial}{\partial s}E(\mathcal{D}_s^c(0))ds$ as a finite Riemann sum

$$h\sum_{i=0}^{k} e^{\lambda ih}\widehat{\frac{\partial}{\partial t}}E(\mathcal{D}_{ih}^c(0)).$$

A point estimate of $P_0(kh)$ is

$$\widehat{P}_0(kh) = \left(h\sum_{i=0}^{k} e^{\lambda ih}\widehat{\frac{\partial}{\partial t}}E(\mathcal{D}_{ih}^c(0)) + P_0(0)\right)e^{-\lambda t}. \tag{6.17}$$

We can thus compute $\widehat{P}_0(kh)$ for values of $k = 1, ..., m$. Note that $mh = t_M$. Interval estimates for $P_0(kh)$ can readily be developed. This technique results in estimates of $P_0(h), P_0(2h), ..., P_0(mh)$. Thus, we estimate $P_0(t)$, $t = 0, h, 2h, ..., t_M$. Smoothing techniques can be applied to estimate intermediate values. Then we can plot $\widehat{P}_0(t), 0 < t < t_M$.

Generalizations and variations of this technique can be used to estimate transient distributions of state variables in many stochastic models that have a continuous time parameter.

The foregoing is an example of **level crossing estimation (LCE)**, also called ***LC computation***. (Chapter *9* introduces LCE for Steady-state distributions. This has also been discussed in [13] and [20].) Also, see Remark 9.2.

Remark 6.3 *Computer speeds will undoubtedly increase in the future. Thus the computational method described above will achieve better and better accuracy. It will be possible to increase N and decrease h, while completing the computations in a shorter amount of real time.*

Remark 6.4 *In the $M/G/r(\cdot)$ dam, possibly $P_0(t) = 0$ for all $t \geq \tau_1$ (instant of first input). For example, if $r(x) = kx, x > 0, k > 0$, the decay of the sample-path has a negative exponential form between inputs. In theory the content will never reach level zero after τ_1. In practice, if there is a very long inter-input time, the content does reach level* 0 ***for all practical purposes***.

6.2.10 Equation for Steady-State PDF of Content

Assume the system is stable. Then

$$F(x) = \lim_{t\to\infty} F_t(x),\ f(x) = \lim_{t\to\infty} f_t(x),\ P_0 = F(0) = \lim_{t\to\infty} P_0(t)$$

exist. Also, $\lim_{t\to\infty} \frac{\partial}{\partial t} F_t(x) = 0$. In (6.13), taking limits of all terms as $t \to \infty$ yields

$$\begin{aligned} r(x)f(x) &= \lambda \int_{y=0}^{x} \overline{B}(x-y)dF(y), x > 0, \\ r(x)f(x) &= \lambda P_0 \overline{B}(x) + \lambda \int_{y=0}^{x} \overline{B}(x-y)f(y)dy, x > 0, \\ r(0^+)f(0) &= \lambda P_0. \end{aligned} \tag{6.18}$$

Alternative Forms of Equation for Steady-state PDF

Two alternative forms of the integral equation in (6.18) are

$$r(x)f(x) = \lambda F(x) - \lambda \int_{y=0}^{x} B(x-y)f(y)dy, x > 0, \tag{6.19}$$

$$r(x)f(x) = \quad \lambda F(x) - \lambda \int_{y=0}^{x} F(x-y)b(y)dy, x > 0, \tag{6.20}$$

where $b(y) = \frac{d}{dy}B(y)$.

In both (6.19) and (6.20), the left side is the SP *downcrossing* rate of level x. On the right side, the first term $\lambda F(x)$ is the rate of inputs when the content is $\leq x$; these inputs generate SP upward jumps that start in state-space interval $[0, x]$. The second term is the rate of such SP jumps that do not upcross level x. Hence the right side is the SP *upcrossing* rate of level x. Rate balance across level x gives the equations.

Equations (6.19) and (6.20) are analogous to equations (3.35) and (3.36) for the M/G/1 queue.

Stability'

A condition for stability of the M/G/r(·) dam is

$$\lambda E(S) < \lim_{x \to \infty} r(x). \tag{6.21}$$

Intuitively, formula (6.21) means that the rate at which the content increases is less than the efflux rate when the content is at high levels. Under condition (6.21) the content is prevented from increasing to indefinitely high amounts.

A condition that guarantees the content will return to level 0 is

$$\lim_{u \downarrow 0} \int_{y=u}^{x} \frac{1}{r(y)} dy < \infty \text{ for } \textit{every finite } x > 0. \tag{6.22}$$

Intuitively, from (6.4), formula (6.22) means that the time to return to level 0 from any positive content is finite if there are no intervening inputs.

Example 6.2 *The M/G/1 queue is a special case of the M/G/r(·) dam with* $r(x) \equiv 1, x > 0$*, and* $r(0) = 0$*. Stability holds iff* $\lambda E(S) < \lim_{x \to \infty} r(x) = 1$*, which is the well known stability condition for M/G/1 queues. If stability holds, the virtual wait returns to level* 0 *(a.s.) since for all finite* $x > 0$

$$\lim_{u \downarrow 0} \int_{y=u}^{x} \frac{1}{r(y)} dy = \lim_{u \downarrow 0} \int_{y=u}^{x} 1 \cdot dy = \lim_{u \downarrow 0} (x - u) = x < \infty.$$

Example 6.3 *For the $M/G/r(\cdot)$ dam with $\lambda > 0$, $E(S) < \infty$, and $r(x) = kx, k > 0$*

$$\lim_{u\downarrow 0}\int_{y=u}^{x}\frac{1}{ky}dy = \frac{1}{k}\lim_{u\downarrow 0}\left(\ln\left(\frac{x}{u}\right)\right) = \infty,$$

for every finite $x > 0$. Hence the content does ***not*** *return to level 0, which implies $P_0 = 0$.*

On the other hand, the dam is ***stable*** *for every $k > 0$ because $\lambda E(S)$ is finite, and*

$$\lambda E(S) < \lim_{x\to\infty} r(x) = \lim_{x\to\infty} kx = \infty.$$

6.2.11 Sojourn Times with Respect to a Level

Consider a sample path of $\{W(t)\}$. Fix level $x > 0$. Due to Poisson arrivals, $\mathcal{D}_t(x)$ ($= \mathcal{D}_t^c(x)$ in this model) is a renewal counting process. The times between successive downcrossings (renewals) are iid r.v.'s. Instants of SP downcrossings of level x are Markov points, at which the process starts over again independent of the past. Let d_x denote the time between successive downcrossings of level x. Let a_x, b_x denote sojourn times above and below level x respectively. A sojourn a_x starts with an upcrossing of x and ends with the next downcrossing of x. A sojourn b_x starts with a downcrossing of x and ends with the next upcrossing of x.

Inter-downcrossing Time

For the process $\{D_t(x), t \geq 0\}$ the renewal rate is

$$\lim_{t\to\infty}\frac{\mathcal{D}_t^c(x)}{t} \underset{a.s.}{=} \lim_{t\to\infty}\frac{E(\mathcal{D}_t^c(x))}{t} = r(x)f(x) = \frac{1}{E(d_x)}.$$

Hence

$$E(d_x) = \frac{1}{r(x)f(x)}, x > 0. \tag{6.23}$$

Sojourn Above a Level

From the theory of regenerative processes, $\frac{E(a_x)}{E(d_x)}$ is equal to the *proportion* of time that the sample path is above level x, namely $(1 - F(x))$. Thus

$$E(a_x) = (1 - F(x)) \cdot E(d_x) = \frac{(1 - F(x))}{r(x)f(x)}. \tag{6.24}$$

From (6.24)

$$\frac{f(x)}{1-F(x)} = \frac{1}{r(x)E(a_x)}, \tag{6.25}$$

$$\frac{d}{dx}\ln(1-F(x)) = \frac{-1}{r(x)E(a_x)}. \tag{6.26}$$

Integrating (6.26) on both sides with respect to x and computing the constant of integration by letting $x \downarrow 0$, gives

$$\begin{aligned} F(x) = \quad & 1-(1-P_0)e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x \geq 0, \\ f(x) = \quad & \tfrac{1-P_0}{r(x)E(a_x)}e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x > 0. \end{aligned} \tag{6.27}$$

The normalizing condition $F(\infty) = 1$, is

$$1-(1-P_0)e^{-\int_{y=0}^{\infty}\frac{1}{r(y)E(a_y)}dy} = 1,$$

which implies that $e^{-\int_{y=0}^{\infty}\frac{1}{r(y)E(a_y)}dy} = 0$, if $P_0 < 1$.

Note that the left side of (6.25) is the *hazard rate* of the steady-state content evaluated at x. There is an inverse relationship between the hazard rate at x and the product $r(x)E(a_x)$ (see Subsection 3.3.14).

Sojourn Below a Level

Similarly $\frac{E(b_x)}{E(d_x)}$ is the proportion of time the SP is at or below level x, which is $F(x)$. Thus

$$E(b_x) = F(x)\cdot E(d_x) = \frac{F(x)}{r(x)f(x)}, \tag{6.28}$$

or

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

This leads to

$$\begin{aligned} F(x) = \quad & P_0 e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy}, x \geq 0, \\ f(x) = \quad & \tfrac{P_0}{r(x)E(b_x)}e^{\int_{y=0}^{x}\frac{1}{r(y)E(b_y)}dy}, x > 0, \end{aligned} \tag{6.29}$$

using $F(0) = P_o$.

Interestingly, formulas (6.27) and (6.29) give two different expressions for $F(x)$ and $f(x)$, in terms of $E(a_x)$ and $E(b_x)$ respectively.

It is readily shown that when $r(x) \equiv 1, x > 0$, the right side of the second equation in (6.29) reduces to the pdf of wait in the M/G/1 queue. Thus, as in formula (3.69),

$$\begin{aligned}\frac{P_0}{1 \cdot E(b_x)} e^{\int_{y=0}^{x} \frac{1}{1 \cdot E(b_y)} dy} &= \frac{P_0 \cdot f(x)}{F(x)} e^{\int_{y=0}^{x} \frac{f(y)}{F(y)} dy} \\ &= \frac{P_0 \cdot f(x)}{F(x)} e^{(\ln F(x) - \ln F(0))} \\ &= \frac{P_0 \cdot f(x)}{F(x)} F(x) P_0^{-1} = f(x).\end{aligned}$$

As a mild check on (6.29), we compute $f(x)$ for the M/M/1 queue in which

$$E(b_x) = \frac{F(x)}{f(x)} = \frac{1 - (1 - (1 - \frac{\lambda}{\mu}))e^{-(\mu-\lambda)x}}{\lambda(1 - \frac{\lambda}{\mu})e^{-(\mu-\lambda)x}}, x \geq 0$$

and $F(0) = P_0 = 1 - \frac{\lambda}{\mu}$. That is, we substitute directly for $E(b_x)$, P_0 in (6.29). Some algebra yields $f(x) = \lambda(1 - \frac{\lambda}{\mu})e^{-(\mu-\lambda)x}, x \geq 0$, which is the steady-state pdf of wait in M/M/1 (see Section 3.4).

6.2.12 Expected Non-empty Period

Denote a non-empty period of the dam by $\mathcal{B}_D$. Then $\mathcal{B}_D = a_0$, the SP sojourn time above level 0. Generally, the structure of $\mathcal{B}_D$ is different from that of a busy period $\mathcal{B}$ for the M/G/1 queue given in (3.61), due to the variation of the efflux rate $r(x), x > 0$.

Constant Efflux Rate

In the particular case where there exists some constant $k > 0$, such that $r(x) \equiv k, x > 0$, the structure for $\mathcal{B}$ in (3.61) is preserved for $\mathcal{B}_D$. Let S denote the size of an input. In particular S is the size of the first input of a non-empty period. Let N_S denote the number of inputs occurring **during the time required for S to dissipate**, i.e., during a time of length $\int_{y=0}^{S} \frac{1}{r(y)} dy = \int_{y=0}^{S} \frac{1}{k} dy = \frac{S}{k}$ time units. Then

$$\mathcal{B}_D = \frac{S}{k} + \sum_{i=1}^{N_S} \mathcal{B}_{Di} \tag{6.30}$$

where $\mathcal{B}_{Di}, i = 1, ..., N_S$ are iid random variables distributed as $\mathcal{B}_D$, independent of N_S. Taking expected values on both sides of (6.30) gives

$$E(\mathcal{B}_D) = \frac{E(S)}{k} + E(N_S)E(\mathcal{B}_D) = \frac{E(S)}{k} + \lambda\frac{E(S)}{k}E(\mathcal{B}_D) \qquad (6.31)$$

since $E(N_S) = \lambda \cdot \frac{E(S)}{k}$. Solving (6.31) for $E(\mathcal{B}_D)$ gives

$$E(\mathcal{B}_D) = E(a_0) = \frac{E(S)}{k\left(1 - \frac{\lambda}{k}E(S)\right)}. \qquad (6.32)$$

Alternative Derivation of $E(\mathcal{B}_D)$

We can obtain P_0 directly when $r(x) \equiv k, x > 0$, by first integrating both sides of (6.18) with respect to $x \in (0, \infty)$. Note that $1 - P_0 = \int_{x=0}^{\infty} f(x)dx$. Thus we get directly

$$P_0 = 1 - \frac{\lambda}{k}E(S). \qquad (6.33)$$

We can now use P_0 in (6.33) for an alternative approach to compute $E(\mathcal{B}_D)$. Namely, $E(\mathcal{B}_D)$ is the proportion $1 - P_0$ of an expected non-empty *cycle* (theory of regenerative processes). Thus

$$\begin{aligned} E(\mathcal{B}_D) &= E(a_0) = (1 - P_0)\, E(d_0) \\ &= \frac{1 - P_0}{r(0^+)f(0)} = \frac{1 - P_0}{\lambda P_0}. \end{aligned}$$

Substituting for P_0 from (6.33) gives

$$E(\mathcal{B}_D) = E(a_0) = \frac{E(S)}{k\left(1 - \frac{\lambda}{k}E(S)\right)}. \qquad (6.34)$$

Formula (6.34) is derived entirely by LC. It illustrates the usefulness of the equation

$$E(a_0) = \frac{1 - P_0}{\lambda P_0}. \qquad (6.35)$$

Formula (6.35) applies to the busy period in M/G/1 queues, as well as to the non-empty period in M/G/r(·) dams.

6.3 M/M/r(·) Dam

Assume inputs occur in a Poisson process at rate λ and that input sizes are $\underset{dist}{=} \mathrm{E}_\mu$. Assume the dam is stable, i.e., $\lambda E(S) < \lim_{x\to\infty} r(x)$ (see equation (6.21). Then the steady-state distribution of content exists.

6.3.1 Equation for Steady-state PDF of Content

We substitute $\overline{B}(x-y) = e^{-\mu(x-y)}, 0 \leq y < x$, in equation (6.18). This results in an integral equation for the steady-state pdf of content $f(x), x > 0$,

$$\begin{aligned} r(x)f(x) &= \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy, x > 0, & (6.36) \\ f(x) &= \frac{\lambda}{r(x)} \left(P_0 e^{-\mu x} + \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy \right), x > 0. & (6.37) \end{aligned}$$

6.3.2 Solution of Equation for PDF of Content

Assume that $P_0 > 0$. Then $r(0^+)f(0) = P_0 > 0$. Applying differential operator $\langle D+\mu \rangle$ to both sides of (6.36), leads to the differential equation for $f(x)$,

$$\begin{aligned} \frac{f'(x)}{f(x)} &= -\frac{r'(x)+\mu-\lambda}{r(x)}, x > 0, \\ \frac{d}{dx} \ln f(x) &= -\frac{r'(x)+\mu-\lambda}{r(x)}, x > 0. & (6.38) \end{aligned}$$

The solution of (6.38) is

$$f(x) = \frac{\lambda P_0}{r(x)} e^{-(\mu-\lambda)\int_{y=0}^{x} \frac{dy}{r(y)}}, x > 0, \tag{6.39}$$

upon applying the initial condition $r(0^+)f(0) = \lambda P_0$.

Substituting for $f(x)$ from (6.39), the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ gives

$$P_0 = \frac{1}{\left(1 + \lambda \int_{x=0}^{\infty} \frac{1}{r(x)} e^{-(\mu-\lambda)\int_{y=0}^{x} \frac{1}{r(y)}dy} dx\right)}. \tag{6.40}$$

As a mild check, the M/M/1 queue is a special case of M/M/r(·) with $r(x) \equiv 1, x > 0$. Substituting $r(x) \equiv 1$ in (6.39) and (6.40) gives (3.86) and (3.87) respectively.

6.3.3 Sojourn Times and State-space Levels

Assume $P_0 > 0$. Note that

$$\begin{aligned} \text{SP entrance rate into } \{0\} &= \lim_{t\to\infty} \frac{D_t^c(0)}{t} \underset{a.s.}{=} r(0^+)f(0) = \lambda P_0 \\ &= \text{SP exit rate out of } \{0\}. \end{aligned}$$

Let us refer to (6.23) and (6.24) with $x = 0$. The expected value of the "non-empty" *cycle* d_0 and of the non-empty *period* a_0 $(= \mathcal{B}_D)$, are respectively

$$\begin{aligned} E(d_0) &= \frac{1}{r(0^+)f(0)} = \frac{1}{\lambda P_0}, \\ E(a_0) &\equiv E(\mathcal{B}_D) = (1 - P_0)E(d_0) = \frac{(1 - P_0)}{\lambda P_0}, \end{aligned}$$

with P_0 given in (6.40).

In M/M/r$(\cdot)$, all upward jumps are $\underset{dist}{=}$ E$_\mu$. By the memoryless property, the excess SP jump above any level x is also $\underset{dist}{=}$ E$_\mu$.

Nevertheless, in M/M/r$(\cdot)$, a_x (thus, also $E(a_x)$) generally depends on x. This differs from $E(a_x)$ in the M/M/1 queue, where $E(a_x)$ *is independent of* x. In M/M/1, $E(a_x) \equiv E(\mathcal{B})$, where $\mathcal{B}$ is a busy period. In the M/M/1 queue, $r(x) = 1, x > 0$; the structure of $\mathcal{B}$ results in this independence (see (3.61)). In the M/M/r$(\cdot)$ dam, generally $r(x)$ varies with x, and a_x depends on the values of $r(y), y > x, x \geq 0$.

Constant Efflux Rate

In the special case where $r(x) \equiv k, k > 0, x > 0$, the structure of $\mathcal{B}_D$ is similar to that of $\mathcal{B}$. Thus, from (6.33) and (6.34),

$$\begin{aligned} P_0 &= 1 - \frac{\lambda}{k\mu} \\ E(a_x) &= E(B_D) = \frac{\frac{1}{\mu}}{k\left(1 - \frac{\lambda}{k\mu}\right)} = \frac{1}{k\mu - \lambda}. \end{aligned}$$

6.4 M/M/r$(\cdot)$: Efflux Proportional to Content

When the efflux rate **varies directly with content**, $r(x) = kx, x > 0$, for fixed $k > 0$. The sample path has a negative exponential shape between input instants. In that case $P_0 = 0$ (see Example 6.3).

6.4.1 PDF and Laplace Transform of Content

Upon substituting $r(x) = kx$ in (6.37) with $P_0 = 0$, we solve for $f(x)$ in (6.37) using Laplace transforms.

Let the Laplace transform of $f(x)$ be

$$\widetilde{f}(s) \equiv \int_{x=0}^{\infty} e^{-sx} f(x)dx, s > 0.$$

In (6.37), multiplying both sides by e^{-sx}, and integrating on $x \in (0, \infty)$ yields

$$\widetilde{f}(s) = \lambda \int_{x=0}^{\infty} e^{-sx} \frac{1}{kx} \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dydx. \tag{6.41}$$

Taking $\frac{d}{ds}$ on both sides of (6.41) and interchanging the order of integration gives

$$\frac{d}{ds}\widetilde{f}(s) = -\frac{\lambda}{k} \int_{y=0}^{\infty} e^{-sy} f(y) \int_{x=y}^{\infty} e^{-(s+\mu)(x-y)} dxdy.$$

Some algebra on the right side leads to a differential equation in $\widetilde{f}(s)$,

$$\frac{d}{ds}\widetilde{f}(s) + \frac{\lambda}{k}\left(\frac{1}{\mu + s}\right)\widetilde{f}(s) = 0. \tag{6.42}$$

Separation of variables in (6.42), and integrating gives

$$\widetilde{f}(s) = A(\mu + s)^{-\frac{\lambda}{k}},$$

for some constant A. The identity $\widetilde{f}(s) \equiv \int_{x=0}^{\infty} e^{-sx} f(x)dx$ implies $\widetilde{f}(0^+) = \int_{x=0}^{\infty} f(x)dx = 1$ (normalizing condition since $P_0 = 0$). Thus

$$\widetilde{f}(0^+) = A\mu^{-\frac{\lambda}{k}} = 1 \text{ and } A = \mu^{\frac{\lambda}{k}}.$$

Hence

$$\widetilde{f}(s) = \left(\frac{\mu}{\mu + s}\right)^{\frac{\lambda}{k}} = \left(\frac{1}{1 + \frac{s}{\mu}}\right)^{\frac{\lambda}{k}} = \left(1 + \frac{s}{\mu}\right)^{-\frac{\lambda}{k}}, s > 0. \tag{6.43}$$

In (6.43) $\widetilde{f}(s)$ is the Laplace transform of a Gamma pdf,

$$f(x) = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} \mu(\mu x)^{(\frac{\lambda}{k}-1)} e^{-\mu x}, x > 0, \tag{6.44}$$

where $\Gamma(\cdot)$ denotes the Gamma function. There is a one-to-one correspondence, up to a set of measure 0, between a Laplace transform and its inverse. This guarantees the uniqueness of $f(x)$ in (6.44).

The statistical moments of the content about 0, are

$$E(W^n) = (-1)^n \left. \frac{d^n}{ds^n} \widetilde{f}(s) \right|_{s=0}, n = 1, 2, \ldots .$$

The first and second moments are

$$E(W) = \frac{\lambda}{k\mu}, \quad E(W^2) = \frac{\lambda}{k\mu^2} \left(\frac{\lambda}{k} + 1 \right).$$

The variance is

$$Var(X) = E(W^2) - (E(W))^2 = \frac{\lambda}{k\mu^2}.$$

Remark 6.5 *In the literature, the form of a Gamma pdf often appears as*

$$g(x) = \frac{1}{b\Gamma(c)} \left(\frac{x}{b} \right)^{c-1} e^{-\frac{x}{b}}, x > 0,$$

where $b > 0$, $c > 1$. The Laplace transform is

$$\widetilde{g}(s) = (1 + bs)^{-c}, s > -\frac{1}{b}.$$

Since $b > 0$, it is sufficient to take $s > 0$. Setting $b = \frac{1}{\mu}$, $c = \frac{\lambda}{k}$ gives $\widetilde{g}(s) = \widetilde{f}(s)$ in (6.43).

6.4.2 CDF of Content

The steady-state cdf of the content is

$$F(x) = \int_{y=0}^{x} f(y)dy = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)} \int_{y=0}^{x} \mu(\mu y)^{\left(\frac{\lambda}{k}-1\right)} e^{-\mu y} dy, x > 0, \quad (6.45)$$

where $\int_{y=0}^{x} \mu(\mu y)^{(\frac{\lambda}{k}-1)} e^{-\mu y} dy$ is the incomplete Gamma function (e.g., see [98]). Generally, $F(x)$ in (6.45) cannot be expressed in closed form, but can be evaluated numerically for each $x > 0$. Note that $F(\infty) = 1$.

6.4.3 Sojourns with Respect to a Level

We next examine the inter-downcrossing time d_x, and sojourns a_x and b_x , above and below level x. Consider a sample path of the process

$\{W(t)\}$. Referring to equations (6.23), and (6.24), the expected values of d_x, a_x and b_x are

$$E(d_x) = \frac{1}{r(x)f(x)} = \frac{1}{kxf(x)} = \frac{\Gamma\left(\frac{\lambda}{k}\right)}{kx\mu(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}}, x > 0, \tag{6.46}$$

$$\begin{aligned} E(a_x) = (1-F(x))E(d_x) &= \frac{\frac{1}{\Gamma\left(\frac{\lambda}{k}\right)}\int_{y=x}^{\infty}\mu(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}dy}{kx\frac{1}{\Gamma\left(\frac{\lambda}{k}\right)}\mu(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}} \\ &= \frac{\int_{y=x}^{\infty}\mu(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}dy}{kx\mu(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}}, \end{aligned} \tag{6.47}$$

$$E(b_x) = F(x)E(d_x) = \frac{\int_{y=0}^{x}\mu(\mu y)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu y}dy}{kx\mu(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}}. \tag{6.48}$$

Naturally, $E(a_x)+E(b_x) = E(d_{x)}$. The expected values $E(d_x)$, $E(a_x)$, $E(b_x)$ can be evaluated numerically and plotted over a range of x values in the state space for any valid triplet of model parameters $\{\lambda, k, \mu\}$.

Example 6.4 *Consider an M/M/r(·) dam with $r(x) = kx, x > 0$.)See Figs. 6.2 - 6.7.) Arbitrarily set $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Then the steady-state pdf of content is*

$$f(x) = 0.752253x^{1.5}e^{-x}, x > 0.$$

The cdf of content is, for $x > 0$,

$$F(x) = -0.188063\left(4.0x^{3/2} + 6.0x^{1/2} - 5.317362 \cdot erf(x^{1/2}) \cdot e^{x}\right) \cdot e^{-x},$$

*where **erf** is the **error function** defined by $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. In this example, since $\mu = 1$,*

$$E(W) = \frac{\lambda}{k\mu} = Var(W) = \frac{\lambda}{k\mu^2} = 2.5.$$

6.5 Generalization of M/G/r(·) Dam

This section presents a *moderate* generalization of the M/G/r(.) dam discussed in sections 6.2 and 6.3. The generalized model encompasses a

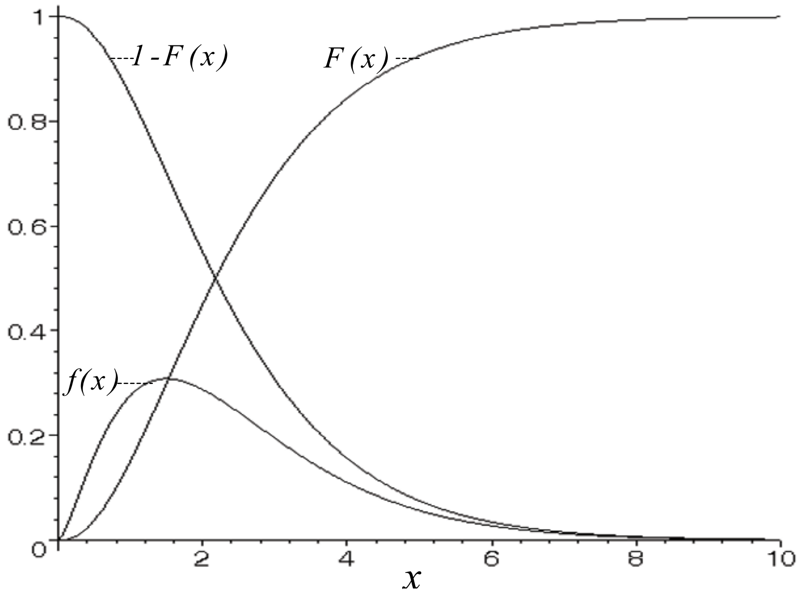

Figure 6.2: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Steady-state pdf $f(x)$, cdf $F(x)$, and complementary cdf $1 - F(x)$.

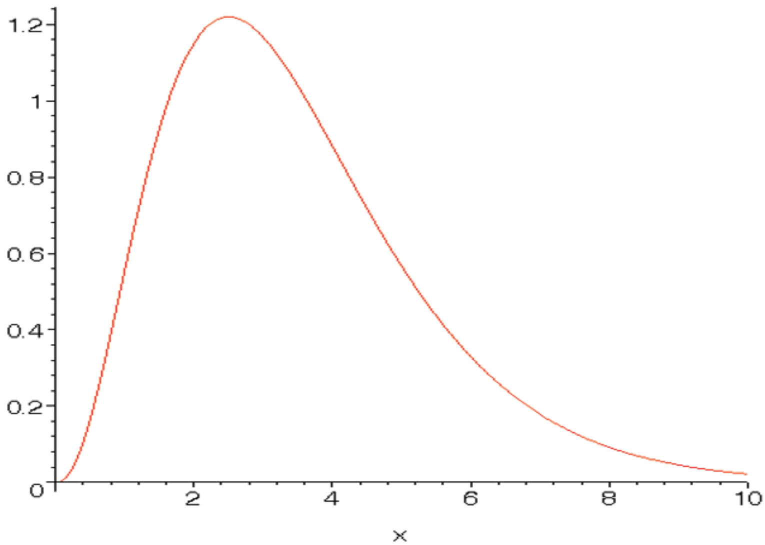

Figure 6.3: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. SP *downcrossing* (and upcrossing) ***rate*** of level x.

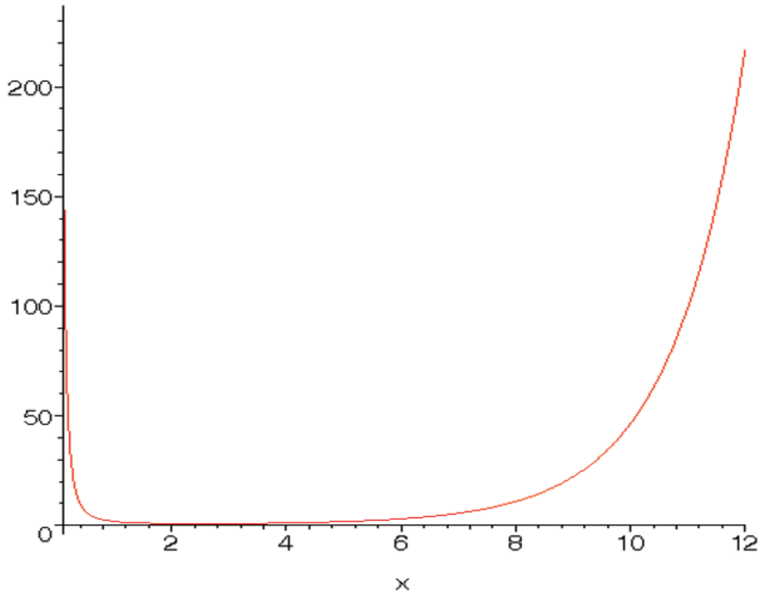

Figure 6.4: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Expected value of SP ***inter-downcrossing time*** $E(d_x), x > 0$. The "bathtub" shape of $E(d_x)$ is intuitive.

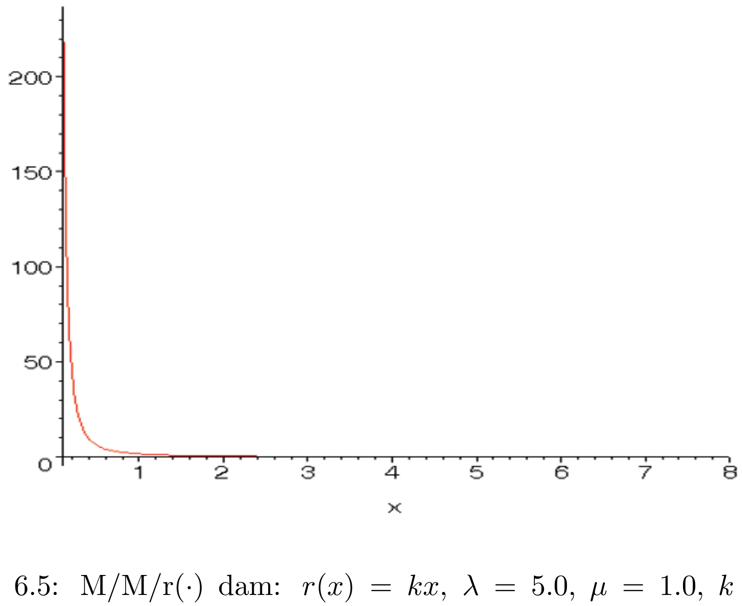

Figure 6.5: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Expected value of SP ***sojourn time above*** level x, $E(a_x), x > 0$.

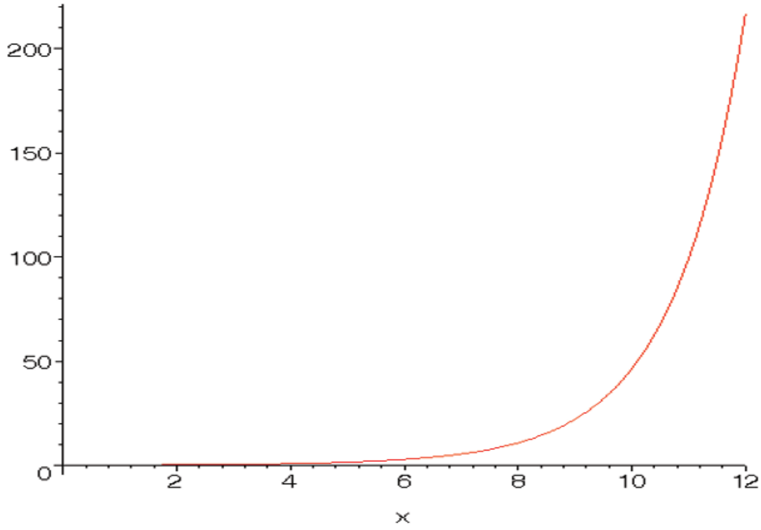

Figure 6.6: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. Expected value of SP ***sojourn time below*** level x, $E(b_x), x > 0$.

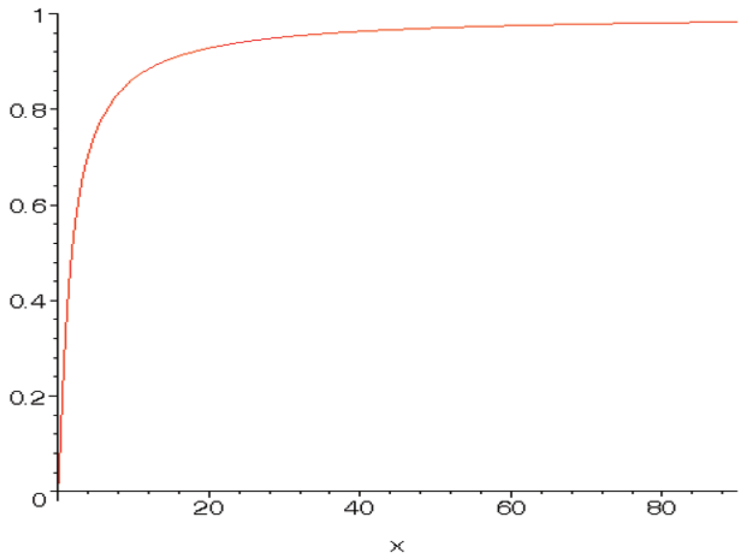

Figure 6.7: M/M/r(·) dam: $r(x) = kx$, $\lambda = 5.0$, $\mu = 1.0$, $k = 2.0$. ***Hazard rate*** for steady-state distribution of content $\frac{f(x)}{1-F(x)}, x > 0$. Note the inverse relation to $E(a_x)$.

large class of inventories and related models. This allows SP **downward jumps due to exogenous** events. It allows **prescribed jumps** when the SP **hits or jumps across designated state-space levels, e.g., barriers**. It allows **specialized jumps if an exogenous event occurs** when the SP is in **a designated state-space interval**.

For example in Marketing Science, a target population may develop a "rebound" effect against purchasing a particular product if repeated advertisements "oversell" the product. Suppose $W(t)$ represents the consumer response to the product. The SP may take a sudden jump downward if an ad occurs while the SP is above a threshold tolerance level. A sample path of the consumer response would increase in a "saw-tooth" pattern, and take a downward jumps from levels above the threshold.

An analogous model may apply in pharmacokinetics. Suppose that a patient's "health" is measured by blood pressure. Assume the patient is on a multiple-dosing regime for some illness related to blood pressure. The blood pressure may drop precipitously if the concentration of the drug in the blood stream breaches an upper threshold after multiple dosing. A sample path of drug concentration would increase in a "saw-tooth" manner, and take a downward jump when it got too high. Similar remarks apply to blood thinners.

6.5.1 Model and Steady-state Distribution of Content

Let $\{W(t), t \geq 0\}$ denote the content of a dam with continuous state space $\boldsymbol{S} \subseteq \boldsymbol{R}$. We assume that $\boldsymbol{S}$ is a "wide-sense" state space. That is, $\boldsymbol{S}$ may contain sets that have probability 0 in the model. (For example, in $\langle s, S \rangle$ inventory the "state space" is interval $(s, S]$; the "wide-sense" state space is $(-\infty, S]$; see Subsection 2.3.1.)

Consider a sample path $W(t), t \geq 0$. (We use "$W(t)$" to represent both the state random variable at instant t and a sample path, for economy of notation.) Assume that the SP makes upward and downward jumps at exogenous Poisson rates λ_u, λ_d respectively, which are independent of each other and of the current state of the system. Let the upward and downward jump magnitudes have cdf's $B_u(\cdot)$, $B_d(\cdot)$, and complementary cdf's $\overline{B_u}(\cdot)$, $\overline{B_d}(\cdot)$, respectively. Additional SP jump types may be allowed, depending on the system state in accordance with the specific model dynamics. For example, in the $\langle s, S \rangle$ inventory model *prescribed, state-dependent* upward jumps occur when the SP hits or jumps below the reorder point s (see Example 2.2 and Fig. 2.2). That is, the SP may move below s momentarily, outside the "state space", but within the wide-sense state space. A particular model may permit jumps due to exogenous events or by "prescription".

Let $F(\cdot)$, $f(\cdot)$ denote, respectively, the steady-state cdf and pdf of $W(t)$ as $t \to \infty$. Our immediate aim is to derive an integral equation for $f(x)$.

Let the downward and upward jumps occur at instants $0 \equiv \tau_{d0} < \tau_{d1} < \cdots$, and $0 \equiv \tau_{u0} < \tau_{u1} < \cdots$, respectively. It is possible that the SP makes both an upward and downward jump at the same instant (see Section 2.3). Without loss of consistency, we may assume what the value of the initial state $W(0) > 0$ is arbitrary. Let $\{\tau_n\} = \{\tau_{di}\} \cup \{\tau_{ui}\}$ be a partition of the time axis $\boldsymbol{T} = [0, \infty)$. Thus $\{\tau_n\}$ is a refinement of $\{\tau_{di}\}$ and $\{\tau_{ui}\}$. The SP jumps occur at instants $0 \equiv \tau_0 < \tau_1 < \cdots$.

Efflux Rate

The efflux rate is specified by equations (6.2) and (6.3).

Sample Path

A typical sample path of $\{W(t)\}$ is a piecewise continuous function in the time-state plane, which decreases continuously between jumps. (see Definition 2.1).

6.5.2 SP Downcrossings

Let $\mathcal{D}_t^c(x), \mathcal{D}_{td}^j(x)$ denote the number of SP *left-continuous and jump* downcrossings of level x *due to exogenous rate* λ_d during $(0, t)$, respectively. Let $\mathcal{D}_{tp}^j(x)$ denote the number SP downward *policy or prescribed* jumps during $(0, t)$. A prescribed downward jump may be due to hitting an upper threshold level. Denote the *total* number of SP downward jumps in $(0, t)$ by $\mathcal{D}_t^j(x)$. Then $\mathcal{D}_t^j(x) = \mathcal{D}_{td}^j(x) + \mathcal{D}_{tp}^j(x)$.

Theorem 6.4

$$\lim_{t\to\infty}\frac{\mathcal{D}_t^c(x)}{t} = r(x)f(x), x\in \boldsymbol{S} \ (a.s.), \tag{6.49}$$

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t^c(x))}{t} = r(x)f(x), x\in \boldsymbol{S}, \tag{6.50}$$

$$\lim_{t\to\infty}\frac{\mathcal{D}_t^j(x)}{t} = \lambda_d\int_{y=x}^{\infty}\overline{B_d}(y-x)f(y)dy \ \forall x\in \boldsymbol{S} \ (a.s.), \tag{6.51}$$

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_{td}^j(x))}{t} = \lambda_d\int_{y=x}^{\infty}\overline{B_d}(y-x)f(y)dy \ \forall x\in \boldsymbol{S}. \tag{6.52}$$

Proof. The proof is similar to the proofs of Theorems 6.1 and 6.2. ■

6.5.3 SP Upcrossings

Let $\mathcal{U}_{tu}^j(x)$ denote the number of SP jump upcrossings of level x during $(0,t)$ *due to the exogenous Poisson rate* λ_u. Let $\mathcal{U}_{tp}^j(x)$ denote the number of prescribed or policy SP jump upcrossings of level x during $(0,t)$. Let $\mathcal{U}_t^j(x)$ denote the total number of SP jump upcrossings during $(0,t)$. Then $\mathcal{U}_t^j(x) = \mathcal{U}_{tu}^j(x) + \mathcal{U}_{tp}^j(x)$. In this model, every upcrossing is a ***jump upcrossing***.

Theorem 6.5

$$\lim_{t\to\infty}\frac{\mathcal{U}_{tu}^j(x)}{t} = \lambda_u\int_{y=-\infty}^{x}\overline{B_u}(x-y)f(y)dy, x\in \boldsymbol{S} \ (a.s.), \tag{6.53}$$

$$\lim_{t\to\infty}\frac{E(\mathcal{U}_{tu}^j(x))}{t} = \lambda_u\int_{y=-\infty}^{x}\overline{B_u}(x-y)f(y)dy, x\in \boldsymbol{S}. \tag{6.54}$$

Proof. Similar to proof of Theorem 6.2. ■

Remark 6.6 *Both sides of the equations in Theorem 6.5 represent the long-run rate of SP upward jumps* ***due to Poisson rate*** λ_u*, from state-space set* $(-\infty,x]$ *into* (x,∞)*.*

6.5.4 Integral Equation for PDF of Content

Applying the principle of rate balance, the SP total downcrossing rate and total upcrossing rate of x, are equal. Thus

$$\lim_{t\to\infty}\frac{E(\mathcal{D}_t^c(x))}{t} + \lim_{t\to\infty}\frac{E(\mathcal{D}_t^j(x))}{t} = \lim_{t\to\infty}\frac{E(\mathcal{U}_{tu}^j(x))}{t} + \lim_{t\to\infty}\frac{E(\mathcal{U}_{tp}^j(x))}{t}. \tag{6.55}$$

Substituting from Theorems 6.4 and 6.5 gives for all $x \in \boldsymbol{S}$,

$$\begin{aligned} &r(x)f(x) + \lambda_d \int_{y=x}^{\infty} \overline{B_d}(y-x)f(y)dy + \lim_{t\to\infty} \tfrac{E(\mathcal{D}_{tp}^j(x))}{t} \\ &= \lambda_u \int_{y=-\infty}^{x} \overline{B_u}(x-y)f(y)dy + \lim_{t\to\infty} \tfrac{E(\mathcal{U}_{tp}^j(x))}{t}. \end{aligned} \tag{6.56}$$

In a particular model where equation (6.56) applies, the terms

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_{tp}^j(x))}{t} \text{ and } \lim_{t\to\infty} \frac{E(\mathcal{U}_{tp}^j(x))}{t}$$

may be expressed in terms of $f(x)$ or as constants. For example, in a standard $\langle s, S\rangle$ inventory model,

$$\lambda_u = 0, \lim_{t\to\infty} \frac{E(\mathcal{D}_{tp}^j(x))}{t} = 0,$$

and

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_{tp}^j(x))}{t} = r(s)f(s) + \lambda_d \int_{y=s}^{S} \overline{B}(y-s)f(y)dy,$$

where λ_d is the demand rate. (See Section 6.8, in which $\lambda_d \equiv \lambda$.)

Remark 6.7 *Integral equation (6.56) may serve as a* ***template*** *for various generalizations of the M/G/r(·) dam. We do not attempt to solve the equation at this point. In any particular model, equation (6.56) will have a particular form, depending on the model parameters. It can then be solved for $f(x)$ (see Section 6.8 below).*

$\langle s, S\rangle$ Inventory

The $\langle s, S\rangle$ continuous review inventory system is a special case of this model. Assume there is no lead time and no backlogging, to simplify the discussion. Then $r(x) > 0$ for all $x \in (s, S]$. Both lead time and backlogging can readily be incorporated into the model. If there is a lead time and backlogging is allowed, the regular state space and wide-sense state space are both equal to the interval $(-\infty, S]$; also $r(x) = 0$ for $x < s$. (See, e.g., [3])

In the $\langle s, S\rangle$ model, SP **prescribed or policy** jump upcrossings occur. Such jump upcrossings are due to placement of orders when the inventory jumps to or below level s or makes a left-continuous hit of level s from above.

6.6 r(·)/G/M Dam

Consider a dam in which there is a continuous *influx* when the content is positive. The influx is interrupted by "demands" for content (outputs), which occur in a Poisson process. The demand sizes are iid positive random variables, having a common general distribution. If a demand exceeds the current content, the dam becomes empty. Empty periods are exponentially distributed with a common mean, independent of other factors. We may regard the empty period as a "setup" time to start a new influx cycle. We shall call a dam having these properties, an **r(·)/G/M** dam. The r(·)/G/M dam is a generalization of the "extended age" process for a G/M/1 queue (Subsection 5.1.1). (We may also call this model a G/M/r(·) dam. However we use the nomenclature r(·)/G/M to emphasize the continuous influx rate $r(\cdot)$.)

The r(·)/G/M dam may be regarded as a template for a variety of production-inventory models where the production rate depends on the current stock level. There are many related variations. For example, we may include a fixed upper bound on content, several fixed levels at which production may pause, lost sales, backlogging, etc.

6.6.1 Model Specification and Notation

Let $W(t)$ denote the content of the dam at time $t \geq 0$. The influx goes on continuously at a positive rate $\frac{dW(t)}{dt} = r(W(t))$, when $W(t) > 0$. Demands for content occur at a Poisson rate μ, and are attended to instantaneously (e.g., a sudden demand for water from a reservoir, or oil from a storage tank; or a rush order for a product, etc.). The demand sizes are positive with common cdf $A(\cdot)$ and complementary cdf $\overline{A}(\cdot)$. If a demand at t_0^- exceeds the current content, the resulting "content" would be negative. The corresponding end point of the SP downward jump would be below level 0 (Fig. 6.8). In that case part of the demand is filled, and part unfilled. Various policies can be used regarding the excess demand (e.g., backlogging, lost sales, etc.). To focus on the LC analysis, we shall assume here that no backlogging is allowed. Then the content at t_0 would be $W(t_0) = 0$. It *remains at level* 0 for a time $\underset{dist}{=} \mathrm{E}_\theta$ independent of the excess demand below 0. During this empty period $\frac{dW(t)}{dt} = 0$. At the end of an empty period, the content begins to rise from level 0 at rate $r(0^+)$, and continues to rise until some future demand brings the content back to level 0. The content alternates between non-empty and empty periods (Fig. 6.8).

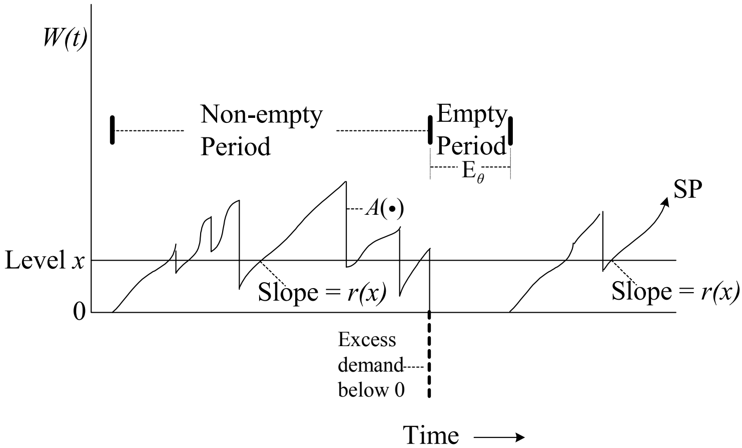

Figure 6.8: Sample path of standard r(·)/G/M dam.

Assume the dam is stable. Then the content will return to level 0, and state $\{0\}$ is positive recurrent. Denote the *transient* pdf and cdf of content by $\{P_0(t); f_t(x), x > 0\}$ and $F_t(x), x \geq 0$, respectively. Then $P_0(t) = F_t(0)$. Denote the *steady-state* pdf and cdf of content by $\{P_0; f(x), x > 0\}$ and $F(x), x \geq 0$, respectively.

6.6.2 Equation for Transient PDF of Content

Consider a sample path of $\{W(t)\}$. Let $\mathcal{U}_t(x)$, $\mathcal{D}_t(x)$ denote the number of up- and downcrossings of x during $(0,t)$, respectively. It can be shown along the lines of Theorems 3.3 and 3.4 that

$$\begin{aligned}
&\tfrac{\partial}{\partial t}E(\mathcal{U}_t(x)) = r(x)f_t(x), x > 0, t > 0,\\
&\tfrac{\partial}{\partial t}E(\mathcal{U}_t(0)) = r(0^+)f_t(0) = \theta P_0(t), t > 0,\\
&\tfrac{\partial}{\partial t}E(\mathcal{D}_t(x)) = \mu \textstyle\int_{y=x}^{\infty} \overline{A}(y-x)f_t(y)dy, x \geq 0, t > 0.
\end{aligned} \tag{6.57}$$

Consider set $\boldsymbol{A} = [0, x], x \geq 0$, in the state space. Recall from Theorem B (i.e., Theorem 4.2) that

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}), \tag{6.58}$$

where $\mathcal{I}_t(\boldsymbol{A})$, $\mathcal{O}_t(\boldsymbol{A})$ are the number of SP entrances and exits of $\boldsymbol{A}$ during $(0,t)$, respectively. In the present model $\mathcal{I}_t(\boldsymbol{A}) = \mathcal{D}_t(x)$, $\mathcal{O}_t(\boldsymbol{A}) = \mathcal{U}_t(x)$, $P_t(\boldsymbol{A}) = F_t(x)$. Substitution from (6.57) into (6.58) results in an *integro-differential* equation for $f_t(x)$ and a *differential* equation for $P_0(t)$, namely

$$\begin{aligned}\mu \int_{y=x}^{\infty} \overline{A}(y-x)f_t(y)dy = &\quad r(x)f_t(x) + \tfrac{\partial}{\partial t}F_t(x), x > 0,\\ \mu \int_{y=0}^{\infty} \overline{A}(y)f_t(y)dy = &\quad r(0^+)f_t(0) + \tfrac{\partial}{\partial t}P_0(t)\\ = &\quad \theta P_0(t) + \tfrac{\partial}{\partial t}P_0(t).\end{aligned} \tag{6.59}$$

The normalizing condition for each $t \geq 0$ is

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

Remark 6.8 *Note that in equations (6.59), the terms* $\frac{\partial}{\partial t}F_t(x)$, $\frac{\partial}{\partial t}P_0(t)$ *appear on the opposite side from the integrals. This placement is in contrast to equations (6.13) and (6.14). The reason for this, is that the sample path of content* ***increases*** *in the* $r(\cdot)/G/M$ *dam, whereas it* ***decreases*** *in the* $M/G/r(\cdot)$ *dam discussed in Subsection 6.2.8.*

6.6.3 Equation for Steady-State PDF of Content

If the dam is stable then the limiting distribution of content exists as $t \to \infty$. Thus

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

An integral equation for $f(x)$ is obtained by letting $t \to \infty$ in (6.59), namely,

$$\begin{aligned}r(x)f(x) = &\quad \mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy, x > 0,\\ r(0^+)f(0) = &\quad \mu \int_{y=0}^{\infty} \overline{A}(y)f(y)dy = \theta P_0.\end{aligned} \tag{6.60}$$

The normalizing condition is

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{6.61}$$

We may also derive (6.60) directly by considering a sample path of $\{W(t)\}$ (Fig. 6.8). Fix level $x > 0$. The SP upcrossing rate of level x is $r(x)f(x)$. The SP downcrossing rate of x is $\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy$. Rate balance across level x gives the first equation in (6.60); the second equation follows by balancing the SP entrance and exit rates of state $\{0\}$ (level 0).

Remark 6.9 *Note that the steady-state equation for the pdf of content in $r(\cdot)/G/M$ (equation (6.60)) is a generalization of the steady-state equation for the pdf of "extended" age· (same as steady-state pdf of virtual wait) in the $G/M/1$ queue. That is, the term $r(x)f(x)$ replaces $f(x)$ on the left side.*

6.6.4 Sojourn Times Above and Below a Level

Let a_x denote the sojourn time above level x, and b_x the sojourn time at or below level x.

Expression for $E(a_x)$

The proportion of time that the SP spends above level x is

$$\begin{aligned}\lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))E(a_x)}{t} &= \lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}\cdot E(a_x)\\ &= r(x)f(x)E(a_x) = 1-F(x), x>0.\end{aligned}$$

Also,

$$r(0^+)f(0)E(a_0) = 1-F(0) = 1-P_0.$$

Thus,

$$\begin{aligned} E(a_x) &= \tfrac{1-F(x)}{r(x)f(x)}, x>0,\\ E(a_0) &= \tfrac{1-F(0)}{r(0^+)f(0)} = \tfrac{1-P_0}{r(0^+)f(0)} = \tfrac{1-P_0}{\theta P_0},\end{aligned} \tag{6.62}$$

where $f(x)$, $f(0)$ are P_0 are the solutions of (6.60), and (6.61).

A Relationship Between $f(x)$ and $E(a_x)$

From (6.62), we obtain the hazard (or failure) rate of the content as

$$\begin{aligned}\frac{f(x)}{1-F(x)} &= \frac{1}{r(x)E(a_x)},\\ \frac{d}{dx}\ln(1-F(x)) &= \frac{-1}{r(x)E(a_x)},\\ 1-F(x) &= Ce^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy},\end{aligned} \tag{6.63}$$

where C is a constant. To evaluate C, let $x\downarrow 0$ in (6.63). This gives $C = 1-P_0$. Thus,

$$F(x) = 1-(1-P_0)e^{-\int_{y=0}^{x}\frac{1}{r(y)E(a_y)}dy}, x\ge 0. \tag{6.64}$$

Taking $\frac{d}{dx}$ in (6.64) gives the pdf

$$f(x) = \frac{(1-P_0)}{r(x)E(a_x)} e^{-\int_{y=0}^{x} \frac{1}{r(y)E(a_y)}dy}, x > 0. \tag{6.65}$$

From (6.62, $1 - P_0 = E(a_0)\theta P_0$. Thus

$$f(x) = \frac{E(a_0)\theta P_0}{r(x)E(a_x)} e^{-\int_{y=0}^{x} \frac{1}{r(y)E(a_y)}dy}, x > 0. \tag{6.66}$$

The normalizing condition gives

$$P_0 = \frac{1}{\left(1 + E(a_0)\theta \int_{x=0}^{\infty} \frac{1}{r(x)E(a_x)} e^{-\int_{y=0}^{x} \frac{1}{r(y)E(a_y)}dy} dx\right)}. \tag{6.67}$$

Another Look at P_0

The sojourn time a_0 represents the non-empty *period*. The length of a non-empty *cycle* is $a_0 + b_0$. Instants at which influx to the dam starts at level 0, are regenerative points due to exponentially distributed times between outputs and the common distribution of demand sizes. Thus

$$E(\text{non-empty cycle}) = E(a_0 + b_0) = \frac{1}{r(0^+)f(0)}.$$

By the theory of regenerative processes $\frac{E(a_0)}{E(a_0+b_0)}$ is the proportion of time the SP is above level 0 during a non-empty cycle. This ratio is equal to the long-run proportion of time that the dam is non-empty, namely $1 - P_0$. Hence

$$\begin{aligned} E(a_0) &= \tfrac{1-P_0}{r(0^+)f(0)} = \tfrac{1-P_0}{\theta P_0}, \\ \text{or} \qquad P_0 &= \tfrac{1}{1+\theta E(a_0)}. \end{aligned} \tag{6.68}$$

Expression for $E(b_x)$

Similarly, we obtain an expression for $E(b_x)$. That is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} E(b_x) = F(x),$$

which is the proportion of time the SP is at or below level x. Thus

$$\begin{aligned} \left(\mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy\right) \cdot E(b_x) &= F(x), \\ r(x)f(x) \cdot E(b_x) &= F(x). \end{aligned}$$

Hence,

$$\begin{aligned} &\tfrac{f(x)}{F(x)} = \tfrac{1}{r(x)E(b_x)}, \\ &\tfrac{d}{dx}\ln F(x) = \tfrac{1}{r(x)E(b_x)}, \\ &F(x) = C_1 e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy}, \\ &F(x) = P_0 e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy}, x \geq 0, \\ &f(x) = \tfrac{P_0}{r(x)E(b_x)} e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy}, x > 0. \end{aligned} \tag{6.69}$$

Note that in (6.69) $F(0) = P_0$ and $F(\infty) = 1$.

Also, we have

$$E(b_0) = \frac{F(0)}{r(0^+)f(0)} = \frac{P_0}{\theta P_0} = \frac{1}{\theta},$$

which is E(empty period), as expected.

Example 6.5 *As a mild check on (6.69) consider a model with* $r(x) \equiv 1, x > 0$, $\overline{A}(x) = e^{-\lambda x}, x \geq 0$, *inter-demand time* $\underset{dist}{=}$ *E, and* $\theta = \lambda$. *This corresponds to an M/M/1 queue in steady state (G/M/1 specialized to M/M/1).*

In M/M/1 $F(x) = 1 - \frac{\lambda}{\mu}e^{-(\mu-\lambda)x}, x \geq 0$, $f(x) = \lambda P_0 e^{-(\mu-\lambda)x}, x > 0$, $P_0 = 1 - \frac{\lambda}{\mu}$. *Then* $E(b_x) = \frac{F(x)}{f(x)} = \frac{1-\frac{\lambda}{\mu}e^{-(\mu-\lambda)x}}{\lambda P_0 e^{-(\mu-\lambda)x}}$. *Thus, in (6.69)*

$$\begin{aligned} \int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy &= \int_{y=0}^{x} \frac{f(y)}{F(y)}dy = \ln\left(\frac{F(x)}{F(0)}\right), \\ F(x) &= P_0 e^{\int_{y=0}^{x} \frac{1}{r(y)E(b_y)}dy} = P_0 e^{\ln\left(\frac{F(x)}{F(0)}\right)} = P_0 \frac{F(x)}{F(0)} \checkmark, \\ &= 1 - \frac{\lambda}{\mu}e^{-(\mu-\lambda)x}, x \geq 0 \checkmark. \end{aligned}$$

6.7 r(·)/G/M Dam: Constant Influx Rate

Suppose the influx rate $r(x) \equiv k, x > 0, k > 0$. In addition, assume the output sizes are exponentially distributed. Let $\overline{A}(x) = e^{-\lambda x}, x \geq 0, \lambda > 0$ (see Fig. 6.9). Since the inter-output times are exponentially distributed (with mean $\frac{1}{\mu}$), the sojourn time *above every* $x > 0$ is distributed the

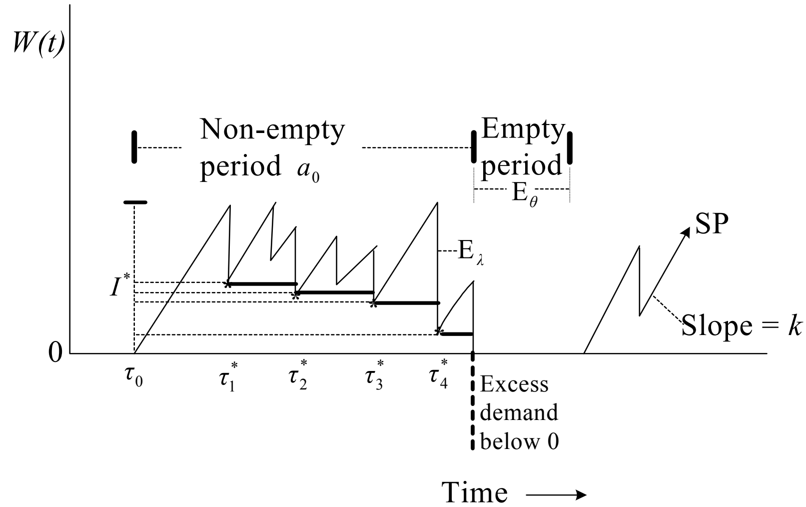

Figure 6.9: Sample path for r(·)/G/M dam with $r(x) \equiv k$, $\overline{A}(x) = e^{-\lambda x}$. Shows $I^* = k(\tau_1 - \tau_0)$, and descending ladder points at $\tau_1^*, ..., \tau_4^*$. The indicated ladder point ordinates are equivalent to four Poisson arrivals (rate λ) within length I^*.

same as a non-empty period a_0 (memoryless property). Thus, $a_x \underset{dist}{\equiv} a_0$ and $E(a_x) \equiv \frac{1-P_0}{\theta P_0}$. Hence, from (6.65)

$$f(x) = \frac{E(a_0)\theta P_0}{r(x)E(a_x)} e^{-\int_{y=0}^{x} \frac{1}{r(y)E(a_y)} dy}, x > 0.$$

$$f(x) = \frac{\theta P_0}{k} e^{-\frac{\theta P_0}{k(1-P_0)} x}, x > 0. \tag{6.70}$$

From (6.67)

$$P_0 = \frac{1}{\left(1 + \frac{\theta}{k} \int_{x=0}^{\infty} e^{-\int_{y=0}^{x} \frac{dy}{kE(a_0)}} dx\right)} = \frac{\mu - \lambda k}{\mu - \lambda k + \theta}, \tag{6.71}$$

where the last term $\frac{\mu - \lambda k}{\mu - \lambda k + \theta}$ in (6.71) is derived in Subsection 6.7.2 below.

6.7.1 Expected Non-empty Period

The non-empty period is a_0. Assume a_0 starts at time τ_0. Let $\tau_1 < \tau_2 < \cdots$, denote the times of successive decrements of content within a_0, that occur after τ_0. Let $\tau_1^* = \tau_1$ and

$$\tau_{n+1}^* = \min\{\tau_i > \tau_n^* | 0 < W(\tau_i) < W(\tau_n^*)\}.$$

The ordinates $\{W(\tau_n^*)\}$ form a set of *strictly descending ladder points* (Fig. 6.9) [55], [56]. Let I^* be the initial influx amount, up to the first output (decrement) at τ_1 (I^* = content at τ_1^-). Let N_{I^*} denote the number of descending ladder points during a_0. Due to exponentially distributed output sizes (mean $\frac{1}{\lambda}$), the memoryless property implies that these ladder points are distributed as Poisson "arrivals" in a length I^*. Note that $E(I^*) = E(\tau_1 - \tau_0) \cdot k = \frac{k}{\mu}$. If the output at τ_1 should empty the dam, then $a_0 = \tau_1 - \tau_0 = \frac{I^*}{k}$. In general,

$$a_0 = \frac{I^*}{k} + \sum_{i=1}^{N_{I^*}} a_{0i}, \tag{6.72}$$

where the a_{0i}'s are iid random variables distributed as a_0, independent of N_{I^*} (see Fig. 6.9 and Subsection 3.3.9).

From (6.72)

$$\begin{aligned} E(a_0) &= \frac{E(I^*)}{k} + E(N_{I^*}) \cdot E(a_0) \\ &= \frac{1}{\mu} + \lambda E(I^*) \cdot E(a_0) \\ &= \frac{1}{\mu} + \lambda \frac{k}{\mu} \cdot E(a_0), \\ E(a_0) &= \frac{\frac{1}{\mu}}{1 - \frac{\lambda k}{\mu}} = \frac{1}{\mu - \lambda k}. \end{aligned} \tag{6.73}$$

6.7.2 Probability of Emptiness and PDF of Content

The expected empty period is $E(b_0) = \frac{1}{\theta}$. Hence, from (6.73)

$$P_o = \frac{E(b_0)}{E(b_0) + E(a_0)} = \frac{\frac{1}{\theta}}{\frac{1}{\theta} + \frac{1}{\mu - \lambda k}} = \frac{\mu - \lambda k}{\mu - \lambda k + \theta}. \tag{6.74}$$

Substituting for P_0 into (6.70) gives

$$f(x) = \frac{\theta(\mu - \lambda k)}{k(\mu - \lambda k + \theta)} e^{-\frac{(\mu - \lambda k)}{k} x}, x > 0. \tag{6.75}$$

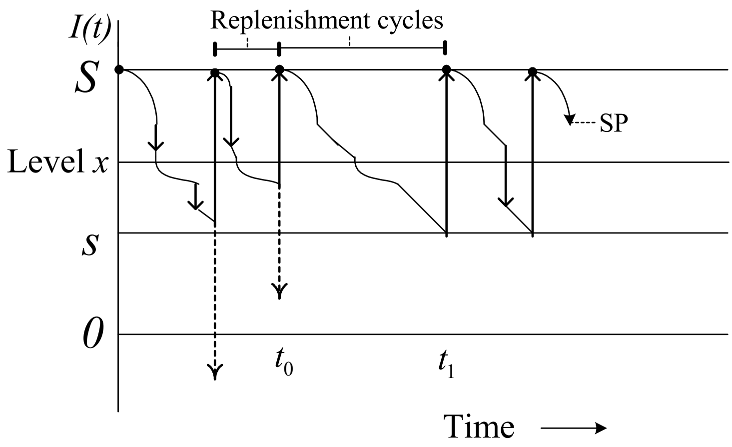

Figure 6.10: Sample path of $\langle s, S\rangle$ inventory with general deacy.

6.8 $\langle s, S\rangle$ Inventory Model: Decay

Consider a continuous review $\langle s, S\rangle$ inventory system with re-order point $s \geq 0$, and order-up-to level $S > s$. Assume that demands for stock occur at a Poisson rate λ. The demand quantities, $D_i, i = 1, 2, ...$, are iid random variables with common cdf $B(x)$; let $\overline{B}(x) = 1 - B(x), x \geq 0$. Denote the stock on hand at time t by $I(t), t \geq 0$. Assume that the stock decays at rate $\frac{dI(t)}{dt} = -r(I(t)) < 0$ when the stock is at level $I(t)$ in the state-space interval $(s, S]$. The ordering policy is as follows. If the stock either decays continuously from above to level s, or jumps downward to, or below level s due to a demand, then an order is placed and received immediately, replenishing the stock up to level S. All SP upward jumps end at level S. The *regular* state space is $(s, S]$ since all probability is concentrated in $(s, S]$. The *wide-sense* state space is $(-\infty, S]$, which accounts for SP downward jumps ending below s, and the subsequent upward jumps that start below s and end at S. The latter upward SP jumps correspond to replenishments (see Fig. 6.10, Fig. 2.2 and Example 2.2). In order to focus on the LC approach, we assume there is no lead time or backlogging. These extensions, as well as others, can be incorporated into the analysis (e.g., [3]).

Let $f(x), s < x \leq S$, $F(x), x \leq S$, denote, respectively, the steady-state pdf and cdf of $I(t)$ as $t \to \infty$. Assume each order size $\underset{dist}{=}$ E_{μ}.

6.8.1 PDF of Inventory: Decay Rate Constant

For elucidation, we take the rate of decay to be $r(x) = k > 0, x \in (s, S]$, unless otherwise specified. We assume the constant decay rate is independent of inventory level, and order sizes have a common exponential distribution, in order to focus on the LC solution technique. We derive an integral equation for $f(x), x \in (s, S]$ in the next subsection.

This $\langle s, S \rangle$ model is a special case of the generalized M/G/r(·) dam in Section 6.5, with $\lambda_d \equiv \lambda$, and where all downward jumps correspond to demands. Also, $B_d(x) \equiv B(x) = 1 - e^{-\mu x}, x > 0$. The upward jump rate due to *exogenous* factors is $\lambda_u \equiv 0$. That is, in $\langle s, S \rangle$ all upward jumps are due to replenishments, which are *prescribed* jumps. Other modifications include the boundaries s and S.

6.8.2 Integral Equation for PDF of Inventory

Consider a sample path of $\{I(t)\}$ (similar to Fig. 6.10 with slope $= -k$). Fix level $x \in (s, S)$. The rate at which the SP *decays* into level $x \in (s, S)$ from above (due to left-continuous strict downcrossings of x).is

$$\lim_{t \to \infty} \frac{E\left(\mathcal{D}_t^c(x)\right)}{t} = r(x) f(x) = k f(x).$$

(We use the terms "rate" and "expected rate" synonymously when they are equal.) The SP decay rate into level s is

$$\lim_{t \to \infty} \frac{E\left(\mathcal{D}_t^c(s)\right)}{t} = r(s^+) f(s^+) \equiv r(s) f(s) = k f(s).$$

The rate at which the SP *jumps* below level $x \in [s, S)$ due to demands is

$$\lim_{t \to \infty} \frac{E\left(\mathcal{D}_t^j(x)\right)}{t} = \lambda \int_{y=x}^{S} \overline{B}(y - x f(y) dy = \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y) dy$$

(jumps start at $y \in (x, S)$ and are greater than $y - x$). The *total* SP *downcrossing* rate of level $x \in [s, S]$ is

$$\begin{aligned} &\lim_{t \to \infty} \frac{E\left(D_t^c(x)\right)}{t} + \lim_{t \to \infty} \frac{E\left(\mathcal{D}_t^j(x)\right)}{t} \\ &= r(x) f(x) + \lambda \int_{y=x}^{S} \overline{B}(y - x f(y) dy \\ &= k f(x) + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y) dy, x \in (s, S]. \end{aligned}$$

The total "downcrossing" rate of the reorder point s is

$$\begin{aligned}\lim_{t\to\infty} \frac{E(\mathcal{D}_t(s))}{t} &= r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y-s)f(y)dy \\ &= kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy,\end{aligned}$$

where we have counted a left-continuous hit of level s from above as a downcrossing of s.

The SP downcrossing rate of level s is equal to the SP egress rate out of level S *below*. This is due to the ordering policy, which orders up to S with each continuous hit of level s. (There is a one-to-one correspondence between SP egresses from S below, and downcrossings of level s.) Rate balance into and out of state $\{S\}$ results in the equality

$$r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y-s)f(y)dy = r(S)f(S),$$

or

$$kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy = kf(S).$$

A simplifying feature of this model is that the total SP *upcrossing* rate of *every* level $x \in (s, S]$ *is equal to* the total *downcrossing* rate of level s. Applying rate balance across level x yields an integral equation for $f(x)$

$$\begin{aligned}& r(s)f(x) + \lambda \int_{y=x}^{S} \overline{B}(y-x)f(y)dy \\ &= r(s)f(s) + \lambda \int_{y=s}^{S} \overline{B}(y-s)f(y)dy \\ &= r(S)f(S), x \in (s, S], \end{aligned} \tag{6.76}$$

or

$$\begin{aligned} kf(x) + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y)dy &= kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy \\ &= kf(S), x \in (s, S]. \end{aligned} \tag{6.77}$$

In (6.77) the left side is the SP downcrossing rate of x. The right side is the SP upcrossing rate of every x.

The state space has no atoms, i.e., there is no state in which the SP spends a positive time. The probability distribution of stock on hand is concentrated on $(s.S]$. Thus the normalizing condition is

$$\int_{x=s}^{S} f(x)dx = 1. \tag{6.78}$$

In (6.77) differentiation with respect to x and carrying out some algebra gives the solution of (6.77) as

$$f(x) = A\left(1 + \frac{\lambda e^{\left(-\left(\frac{\lambda}{k}+\mu\right)(S-x)\right)}}{k\mu}\right), x \in (s, S] \tag{6.79}$$

where A is a constant and $f(x) = 0$ for $x \notin (s, S]$. Using (6.78) leads to

$$A = \left((S-s) + \frac{\lambda}{k\mu(\mu + \frac{\lambda}{k})}\left(1 - e^{\left(-\left(\frac{\lambda}{k}+\mu\right)(S-s)\right)}\right)\right)^{-1}. \tag{6.80}$$

Note that $f(x)$ in (6.79) is convex and increasing on (s, S) (i.e., $f'(x) > 0$, $f''(x) > 0$). This property agrees with intuition which 'suggests that the stock resides most of the time at high levels closer to S and less often near s, irrespective of the value of $k > 0$. This accumulation of inventory near S is a consequence of the re-order policy, which instantaneously replenishes the stock up to level S at each replenishment instant since there is no lead time. (See the numerical example in Subsection 6.8.8 and Figs. 6.11, 6.12.)

6.8.3 Sojourns Above and Below a Level

Let a_x and b_x denote a sojourn time of net inventory above and at or below level $x \in (s, S]$, respectively. Every instant $t \geq 0$ such that $I(t) = S$ is a regenerative point of the inventory process. The regenerative property holds whether replenishments up to S are due to SP smooth decays into level s from above, or due to SP jumps that end at or below level s as a result of a demand. For example, consider Fig. 6.10. At time points like t_1 the SP makes a left-continuous hit of level s from above, and jumps upward to level S. The Poisson arrival process for demands ensures that the excess time until the next demand $\underset{dist}{=}$ E_λ (memoryless property).

Inter-level-S Time

The times between successive instants at which $I(t) = S$, form a renewal process. Denote this "inter-level-S" time by d_S. The rate of replenishments is the total SP downcrossing rate of level s namely $kf(s) + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy$, which is the same as the egress rate from S below, namely $kf(S)$. The value of $f(S)$ is obtained from (6.79). Thus the expected time between two successive replenishment instants is

$$E(d_S) = \frac{1}{kf(S)} = \frac{\mu}{A(k\mu + \lambda)}, \tag{6.81}$$

where A is given by (6.80).

Expression for $E(a_x)$

The *proportion* of time the SP spends above level x is $1-F(x)$ and is also equal to $\frac{E(a_x)}{E(d_S)} = E(a_x)kf(S)$ (theory of regenerative processes). Thus

$$E(a_x) = \frac{1 - F(x)}{kf(S)} = \frac{\mu(1 - F(x))}{A(k\mu + \lambda)},$$

where $F(x) = \int_{y=s}^{x} f(y)dy$ is obtained from (6.79).

Expression for $E(b_x)$

Similarly,

$$E(b_x) = \frac{F(x)}{kf(S)} = \frac{\mu F(x)}{A(k\mu + \lambda)}.$$

with A given in (6.80). We may also obtain $E(b_x)$ using

$$E(a_x) + E(b_x) = E(d_S) = \frac{1}{kf(S)}.$$

We check on the formulas for $E(a_x)$, $E(b_x)$ when $x = s$. Note that $F(s) = 0$ and $1 - F(s) = 1$. Then

$$E(a_s) = \frac{1}{kf(S)} = \frac{\mu}{A(k\mu + \lambda)} = E(d_S),$$

the expected replenishment cycle, as intuitively expected. Also $E(b_s) = \frac{F(s)}{kf(S)} = 0$, which agrees with the SP spending no time below level s. (Recall that state-space jumps occur *not in Time*. This includes jumps that end below s.)

6.8.4 Total Ordering Rate

The *total* ordering *rate* is given by the total *downcrossing* rate of level s, which is the right hand side of (6.77), namely $kf(S) = \frac{A(k\mu+\lambda)}{\mu}$ (A is given by (6.80)).

6.8.5 Orders Due to Two Types of Signal

There are two types of signals that initiate an order. Define a **type-c** signal as an SP *left continuous decay into level s from above* (e.g., time point t_1 in Fig. 6.10). Define a **type-j** signal as an SP *downward jump that ends at or below s due to a demand* (e.g., time point t_0 in Fig. 6.10). Define an **order cycle** (same as **replenishment** cycle) as the time between two successive instants when an order is *received.* Thus an order cycle is the inter-level-S downcrossing time (inter-egress time from level S below, the same as d_S). Due to Poisson arrivals of demands, the sequence $\{d_{Si}\}$ with $d_{Si} \underset{dist}{=} d_S, i = 1, 2, ...,$ is a renewal process. The order initiating d_S is either type-c or type-j. Let $P_c(s) = P$(the order initiating d_S is type-c). Let $P_j(s) = P$(the order initiating d_S is type-j). Then $P_c(s) + P_j(s) = 1$.

We now determine $P_c(s)$ and $P_j(s)$. Note that the counting process $\{\mathcal{D}_t^c(s), t \geq 0\}$ is a renewal process due to Poisson arrivals of demand. Also,

$$E(\text{number of type-c orders in } d_S) = 1 \cdot P_c(s) + 0 \cdot P_j(s) = P_c(s).$$

By the theory of regenerative processes,

$$\frac{E(\text{number of type-c orders in } d_S)}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^c(s))}{t} = r(s)f(s)$$
$$\iff \frac{P_c(s)}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^c(s))}{t} = r(s)f(s).$$

Thus

$$P_c(s) = r(s)f(s) \cdot E(d_S) = \frac{r(s)f(s)}{r(S)f(S)}, \tag{6.82}$$

since $E(d_S) = \frac{1}{r(S)f(S)}$. Intuitively, in (6.82) the numerator is the rate of type-c orders; the denominator is the overall rate of orders.

Also, $\{D_t^j(s), t \geq 0\}$ is a renewal process (Poisson arrivals of demands). Therefore

$$\frac{E(\text{number of type-j orders in } d_S)}{E(d_S)} = \frac{1.P_j(s) + 0 \cdot P_c(s)}{E(d_S)} = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^j(s))}{t},$$

and

$$P_j(s) = \lim_{t\to\infty} \frac{E(\mathcal{D}_t^j(s))}{t} E(d_S) = \frac{\lambda \int_{y=s}^{S} \overline{B}(y-s) f(y)dy}{r(S)f(S)}. \tag{6.83}$$

(see Section 6.5). Intuitively, in (6.83) the numerator is the rate of type-j orders; the denominator is the overall rate of orders.

In the case where $r(x) \equiv k, x \in (s, S]$, and the demand sizes are distributed as E_μ, we have

$$P_c(s) = \frac{kf(s)}{kf(S)} = \frac{f(s)}{f(S)}, \quad P_j(s) = \frac{\lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy}{kf(S)}.$$

Then

$$P_c(s) = \frac{k\mu + \lambda e^{-(\frac{k}{\lambda}+\mu)(S-s)}}{k\mu + \lambda}, \tag{6.84}$$

$$P_j(s) = \frac{\mu\lambda A \int_{y=s}^{S} e^{-\mu(y-s)} \left(1 + \frac{\lambda e^{\left(-\left(\frac{\lambda}{k}+\mu\right)(S-y)\right)}}{k\mu}\right) dy}{k\mu + \lambda}, \tag{6.85}$$

with A given in (6.80).

6.8.6 Expected Order Size

Denote the order size by R.

If an order is caused by a left-continuous decay into level s, then $R = S - s$. If an order is caused by an SP downward jump below level s, then $R = S - s + r_s$ where r_s denotes the excess demand *below* s. If the order sizes are $\underset{dist}{=} \mathrm{E}_\mu$ then $r_s \underset{dist}{=} \mathrm{E}_\mu$ (memoryless property). Note that $P_c(s) + P_j(s) = 1$. Hence the expected order size is

$$\begin{aligned} E(R) &= (S-s)P_c(s) + \left(S - s + \frac{1}{\mu}\right) P_j(s) \\ &= S - s + P_j(s)\frac{1}{\mu}, \end{aligned} \tag{6.86}$$

where $P_j(s)$ is given in (6.85).

6.8.7 Cost Rate

Since there is no backlogging or lead-time costs in the $\langle s, S\rangle$ model described in this section, the cost function only accounts for setup of placing orders, and for holding inventory. Let $\mathcal{C}$, $\mathcal{C}_{Os}$, C_{Oj}, be the total average *cost rate*, the setup cost per order when initiated by product decay to level s, and the setup cost per order when initiated by a demand (SP jump) that propels the SP to or below level s, respectively. Let $\mathcal{C}_H$ be the holding cost per unit per unit time. Then

$$\begin{aligned} C &= \mathcal{C}_{Os}\cdot(\text{type-c ordering rate}) \\ &\qquad + \mathcal{C}_{Oj}\cdot(\text{type-j ordering rate}) + \mathcal{C}_H\int_{x=s}^{S} xf(x)dx \\ &= \mathcal{C}_{Os}kf(s) + \mathcal{C}_{Oj}\int_{x=s}^{S} e^{-(x-s)\mu}f(x)dx + \mathcal{C}_H\int_{x=s}^{S} xf(x)dx, \end{aligned} \tag{6.87}$$

where $f(x)$ is given by (6.79).

6.8.8 Numerical Example

In $\langle s, S\rangle$ with $r(x) \equiv k$ and all demand sizes $\underset{dist}{=}$ E_μ, assume $\lambda = 2$, $\mu = 0.10$, $k = 1$, $S = 2.5$, $s = 1$. Calculations give the constant in (6.80) to be $A = 0.094200$. The pdf of inventory is

$$f(x) = 0.094200 + 1.88400e^{(-5.250+2.10x)}, 1 < x \le 2.5.$$

Note that $f(1) = 0.1749$, $f(2.5) = 1.9782$.

The cdf of inventory is

$$F(x) = -0.132645 + 0.094200x + 0.897144e^{(-5.25+2.1x)}, 1 < x \le 2.5.$$

Note that $F(1) = 0$, $F(2.5) = 1.0$.

Functions $f(x)$ and $F(x)$ are plotted in figures 6.11 and 6.12, which demonstrate convexity and the probability massed towards S.

6.9 $\langle s, S\rangle$ Inventory Model: No Decay

Consider an $\langle s, S\rangle$ model as in Section 6.8, with demand sizes $\underset{dist}{=}$ E_μ and *no decay of inventory.*

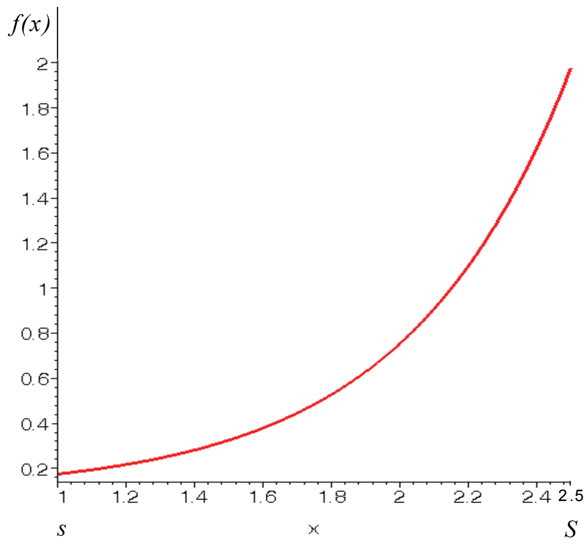

Figure 6.11: PDF $f(x)$ in $\langle s, S\rangle$ inventory with decay rate k. $\lambda = 2, \mu = 0.10$, $k = 1$, $S = 2.5$, $s = 1$. Note that $f(x) = 0 \begin{cases} x < 1, \\ x > 2.5. \end{cases}$

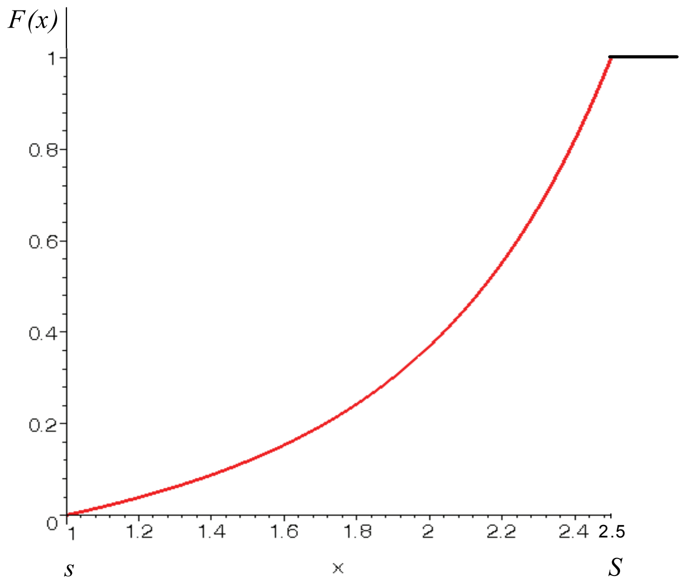

Figure 6.12: CDF $F(x)$ in $\langle s, S\rangle$ inventory with decay rate k. $\lambda = 2, \mu = 0.10$, $k = 1$, $S = 2.5$, $s = 1$.

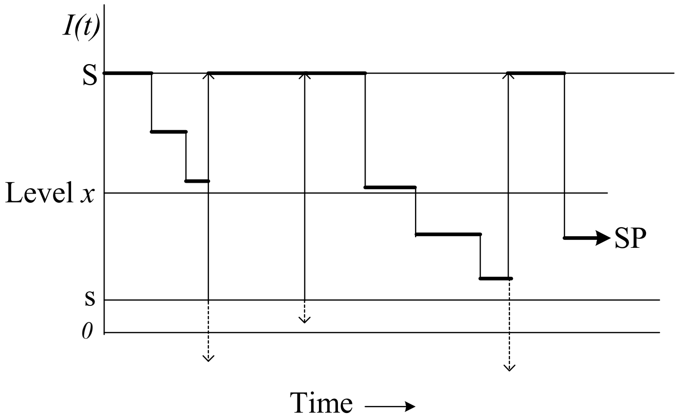

Figure 6.13: Sample path of $i(t), t \geq 0$ in $\langle s, S \rangle$ inventory with *no* decay. The SP stays at a level for a time distributed as the inter-demand time (exponentially distributed with mean $\frac{1}{\lambda}$in text).

Thus $r(x) \equiv 0$. Once the stock on hand enters a level in $(s, S]$, it remains at that level for a time $\underset{dist}{=} \mathrm{E}_\lambda$, until the next demand instant (see Fig. 2.6 and Fig. 6.13). The state space has an *atom* at level S (positive probability). Each SP sojourn time in state $\{S\} \underset{dist}{=} \mathrm{E}_\lambda$). Every state $\{x\} \in \{y|y \in (s, S)\}$ is continuous (not an atom). That is because the probability of entering and remaining in such $\{x\}$ for a positive time is 0, due to continuous demand sizes.

Let $\Pi_S = P(\text{inventory is at level } S)$ in the steady state. Equating the SP down- and upcrossing rates of level $x \in (s, S)$ we obtain an integral equation for $f(x)$,

$$\lambda \Pi_S e^{-\mu(S-x)} + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y) dy$$
$$= \lambda \Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y) dy,$$

or

$$\lambda \Pi_S e^{-\mu(S-x)} + \lambda \int_{y=x}^{S} e^{-\mu(y-x)} f(y) dy = \lambda \Pi_S, s < x < S. \tag{6.88}$$

In (6.88) we have used the fact that the SP rate into $\{S\}$, namely $\lambda \Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y) dy$ is equal to the SP rate out of $\{S\}$, which is $\lambda\Pi_S$. The normalizing condition is

$$\Pi_S + \int_{x=s}^{S} f(x)dx = 1. \tag{6.89}$$

6.9.1 PDF of Inventory

It is readily shown after some algebra using (6.88) and (6.89), that the pdf is a mixed density. That is, $f(x)$ is uniform on state-space interval (s, S) and there is an atom at S. The resulting mixed pdf is

$$\Pi_S = \frac{1}{1+\mu(S-s)}, \quad f(x) = \frac{\mu}{1+\mu(S-s)}, x \in (s, S). \tag{6.90}$$

6.9.2 Sojourn Times Above and Below a Level

The renewal rate of replenishments is the total SP downcrossing rate of level s, namely

$$\lambda\Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy = \lambda\Pi_S,$$

since the only signals to place orders are due to SP jumps ending at or below s. Recall that d_S is the time between two successive replenishments up to level S (same as an ordering cycle). Then

$$E\left(d_S\right) = \frac{1}{\lambda\Pi_S e^{-\mu(S-s)} + \lambda \int_{y=s}^{S} e^{-\mu(y-s)} f(y)dy} = \frac{1}{\lambda\Pi_S}. \tag{6.91}$$

Fix level $x \in (s, S]$. From the theory of regenerative processes

$$E(a_x) = (1-F(x)) \cdot E\left(d_S\right) = \frac{1-F(x)}{\lambda\Pi_S}, \tag{6.92}$$

$$E(b_x) = F(x) \cdot E\left(d_S\right) = \frac{F(x)}{\lambda\Pi_S}, \tag{6.93}$$

where

$$F(x) = \int_{y=s}^{x} f(y)dy = \frac{\mu(x-s)}{1+\mu(S-s)}, s < x < S.$$

Note that the sum of all probabilities is

$$\begin{aligned} F(S) &= \frac{\mu(S-s)}{1+\mu(S-s)} + \Pi_S \\ &= \frac{\mu(S-s)}{1+\mu(S-s)} + \frac{1}{1+\mu(S-s)} = 1. \end{aligned}$$

6.9.3 Ordering Characteristics

Ordering Rate

The ordering rate is the right hand side of (6.88) namely $\lambda\Pi_S = \frac{\lambda}{1+\mu(S-s)}$.

Expected Order Size

All orders are signalled by SP jumps ending at or below s. Thus $P_c(S) = 0$ and $P_j(S) = 1$ (see Subsection 6.8.5). Hence the expected order size is

$$E(R) = S - s + \frac{1}{\mu}. \tag{6.94}$$

Expected Number of Orders in an Ordering Cycle

Let N_{d_S} denote the number of orders in an ordering cycle $_{d_S}$. Then

$$N_{d_S} = \min\left\{ n | \sum_{i=1}^{n} D_i > S - s \right\}. \tag{6.95}$$

Random variable N_{d_S} is a stopping time for the sequence $\{D_i\}$. Random variable N_{d_S} is also equal to the number of inter-demand times during d_S. Thus

$$d_S = \sum_{i=1}^{N_{d_S}} T_i,$$

where $\{T_i\}$ are iid r.v.'s $\underset{dist}{=}$ E_λ. Also, N_{d_S} is a stopping time for the sequence $\{T_i\}$, signalling the end of an ordering cycle. Thus

$$E(d_S) = E(N_{d_S}) \cdot E(T),$$

and

$$E(N_{d_S}) = \frac{E(d_S)}{E(T)} = \frac{\frac{1}{\lambda\Pi_S}}{\frac{1}{\lambda}} = \frac{1}{\Pi_S} = 1 + \mu(S - s). \tag{6.96}$$

6.9.4 Cost Rate

Let $\mathcal{C}_O$, $\mathcal{C}_H$ denote the setup cost per order and holding cost per order per unit time, respectively. The total average cost rate is

$$C = \mathcal{C}_O \cdot (\text{ordering rate}) + \mathcal{C}_H \int_{x=s}^{S^+} x f(x) dx.$$

The ordering rate is $\lambda\Pi_S = \frac{\lambda}{1+\mu(S-s)}$. The average stock on hand is

$$\begin{aligned}\int_{x=s}^{S^+} xf(x)dx &= S\Pi_S + \int_{x=s}^{S} \frac{\mu x}{1+\mu(S-s)}dx \\ &= \frac{S}{1+\mu(S-s)} + \frac{\mu(S^2-s^2)}{2(1+\mu(S-s))} \\ &= \frac{2S+\mu(S^2-s^2)}{2(1+\mu(S-s))}.\end{aligned}$$

Thus

$$C = \frac{\lambda}{1+\mu(S-s)} \cdot C_O + \frac{2S+\mu(S^2-s^2)}{2(1+\mu(S-s))} \cdot C_H. \tag{6.97}$$

Remark 6.10 *In $\langle s, S\rangle$ with **no decay**, suppose the inter-demand times form a **renewal process (not necessarily a Poisson process)**. Then the results will be the same as in this section. The integral equation for the pdf $f(x)$ would be the same as (6.88), where λ represents the renewal rate of the demand process. The arrival rate λ cancels out of the equation. Thus the formulas for Π_S and $f(x), x \in (s, S)$ given in (6.90) are independent of λ. The underlying reason for this property is that all signals to place an order are SP jumps below s at the ends of inter-demand times. When the SP jumps up to level S, the time until the next demand is a **full inter-arrival time**. Hence $\{d_{Si}\}$ is a renewal process, where $d_{Si} \underset{dist}{\equiv} d_S$.*

Remark 6.11 *For exposition, we have applied LC to only two basic $\langle s, S\rangle$ inventory systems. We emphasize that LC equally applies to a vast array of other inventory systems as well, e.g., $\langle r, nQ\rangle$, variations of EOQ models, models with lead time and backlogging, production-inventory models of various complexity, models with a variety of state-dependent control policies, etc., (e.g., [2], [3]).*

CHAPTER 7

MULTI-DIMENSIONAL MODELS

For many stochastic models the state is described using *multiple continuous* random variables. These variables take values in a state space $\boldsymbol{S}$ which is a subset of multi-dimensional Euclidean space.

7.1 Models with State Space a Subset of $\boldsymbol{R}^2$

Suppose the state is described using *two continuous* random variables. Then $\boldsymbol{S} \subseteq \boldsymbol{R}^2$, where $\boldsymbol{R}^2 = \{(x, y)\,|x \in \boldsymbol{R}, y \in \boldsymbol{R}\}$, and $\boldsymbol{R}$ is the set of real numbers. We assume the time axis is $\boldsymbol{T} = [0, \infty)$.

Notation 7.1 *We denote an* ***n-dimensional*** *stochastic model by* $\boldsymbol{n_{(c,d)}}$***-dimensional*** *if c state variables have* ***continuous*** *components (and possibly some atoms, i.e., may be mixed r.v.'s), and d state variables are* ***discrete****, where* $c + d = n$.

In this chapter, we shall use the term "continuous" random variable to mean "continuous or mixed" random variable, for economy of notation.

Example 7.1 *Consider the* ***system point process*** $\{W(t), M(t)\}$ *for the general M/M/c queue (Section 4.3, subsections 4.5, 4.6). Random variable* $W(t)$ *is the waiting time, which is* ***continuous*** *on* $(0, \infty)$ *with an* ***atom*** *at* 0*. Thus r.v.* $W(t)$ *is* ***mixed****. Random variable* $M(t)$ *is the system configuration, which is* ***discrete****, taking values in a discrete set* $\boldsymbol{M}$*. Assume there is only one discrete state variable, e.g.,* $M(t) =$

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_7,

number of other servers occupied at a start of service instant of a time-t arrival. Then, the state would have two random variables. In M/M/c, the system point process is two-dimensional. We shall describe it as $\mathbf{2_{(1,1)}}$***-dimensional****. Then the state space* $\boldsymbol{S} \subseteq \boldsymbol{R} \times \boldsymbol{M}$*. (Similar remarks apply if* $M(t)$ *is a* **vector** *of discrete r.v.'s.)*

Remark 7.1 *If the state of a model with 2 continuous r.v.'s also has a discrete system configuration* $M(t) \in \mathbf{M}$*, then the state space* $\boldsymbol{S} \subseteq \boldsymbol{R}^2 \times \mathbf{M}$*. The model is described as being* $3_{(2,1)}$*-dimensional. Analogous descriptions apply to models with states consisting of c continuous r.v.'s,* $c = 3, 4, \ldots$ *. The model would be described as* $(c+1)_{(c,1)}$*-dimensional. Then the state space* $\boldsymbol{S} \subseteq \boldsymbol{R}^c \times \mathbf{M}$*.*

In the other chapters of this monograph, LC techniques are used for $1_{(1,0)}$-, or $1_{(0,1)}$-, or $2_{(1,1)}$-dimensional models, viz., queues, inventories, dams, renewal processes, counter models, etc. The same techniques can be applied to analyze $2_{(2,0)}$-dimensional models with state space $\boldsymbol{S} \subseteq \boldsymbol{R}^2$, or $3_{(2,1)}$-dimensional models with $\boldsymbol{S} \subseteq \boldsymbol{R}^2 \times \mathbf{M}$, etc. These techniques are also applicable to $(n+d)_{(n,d)}$-dimensional models with $\boldsymbol{S} \subseteq \boldsymbol{R}^n$ $(d = 0)$ or $\boldsymbol{S} \subseteq \boldsymbol{R}^n \times \mathbf{M}$,$n = 3, 4\ldots$, etc.

In this chapter we shall focus on two variants of a $2_{(2,0)}$-dimensional model with $\boldsymbol{S} \subseteq \boldsymbol{R}^2$. The idea is to fix $(x, y) \in \boldsymbol{S}$, and select a region $\mathcal{R}_{x,y} \subseteq \boldsymbol{S}$ depending on (x, y) having boundary $\partial\mathcal{R}_{x,y}$, such that *part of* $\partial\mathcal{R}_{x,y}$ is expressible as a function of x and y. That part of $\partial\mathcal{R}_{x,y}$ may be a "level" set $\phi(x, y) = constant$, for some function ϕ. Alternatively, that part of $\partial\mathcal{R}_{x,y}$ may be defined as the union of point sets, which describe a curve in $\boldsymbol{S}$; we call such a union a two-dimensional level. Suppose such a level is specified (see Subsection 7.1.2 below). We then apply SPLC methods to express the SP rates across it, in terms of the *joint pdf* of the *continuous* (or mixed) state variables. Specifically, we use the principle of rate balance across the level to formulate integral equations for the steady-state *joint pdf* of the state variables. The integral equations are solved using analytic, simulation or numerical methods.

We illustrate the technique by modelling a two-product inventory system in which the products share the same limited total storage space (Section 7.2 below). Before discussing the two-product inventory model, we very briefly review some properties of a rectangle (or interval) in $\boldsymbol{R}^2$.

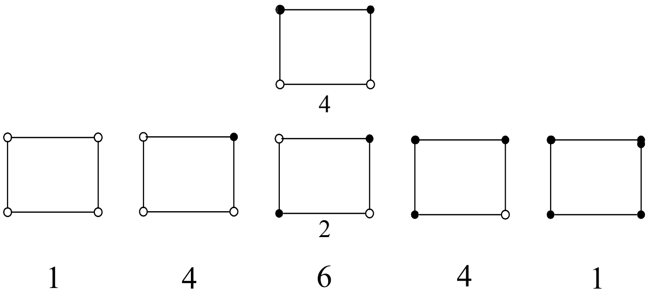

Figure 7.1: Rectangles in 2-dimensional space. Shows the number of combinations of open, half-open, and closed edges for each number of filled-in vertices, $0, ..., 4$. The total number of combinations is 16.

7.1.1 Rectangles in $\boldsymbol{R}^2$

Let $[x_1, y_1]$, $[x_2, y_2], x_i < y_i, i = 1, 2$, denote two finite closed *intervals* in $\boldsymbol{R}$. A closed *rectangle* (*interval)* in $\boldsymbol{R}^2$ is the cross product

$$[x_1, y_1] \times [x_2, y_2] = \{(\alpha, \beta) \,|\alpha \in [x_1, y_1], \beta \in [x_2, y_2]\} \,.$$

Similar definitions apply to open intervals, partially open intervals, etc. For an arbitrary finite interval in $\boldsymbol{R}$, there are two choices for *each* end point. Either the end point belongs to the interval, or does not. Since a rectangle in $\boldsymbol{R}^2$ has 4 edges and 4 vertices, there are $2^4 = 16$ possible combinations of open and closed edges (Fig. 7.1).

We may also determine the count of combinations of various types of edges by considering whether vertices are filled in. For example, if 0 vertices are filled-in (1 way), then all 4 edges are open sets. If 1 vertex is filled-in (4 ways), then 2 edges are half-open and 2 edges are open. If 2 vertices are filled-in (6 ways), then there are 2 distinct cases: if the filled-in vertices are adjacent, there are one closed edge, 2 half-open edges and 1 open edge (4 ways); if the filled-in vertices are kitty-corner then all 4 edges are half-open (2 ways). If 3 vertices are filled in (4 ways), then 2 edges are closed, and 2 edges are half open. If all 4 vertices are filled in (1 way), then all 4 edges are closed.

7.1.2 Two-dimensional Levels

In many models of Operations Research, the state space for mixed r.v.'s is $\boldsymbol{S} \subseteq \boldsymbol{R}^2$ *restricted to the non-negative quadrant*,

$$\boldsymbol{S} = \{(x, y) \,|x \geq 0, y \geq 0\} \,.$$

We may arbitrarily utilize a rectangle $\mathcal{R}_{x,y} \subseteq \boldsymbol{S}$ with a fixed corner $(x,y) \in \boldsymbol{S}$. Suppose we specify rectangle

$$\mathcal{R}_{x,y} = (a,x) \times (b,y), 0 < a < x, 0 < b < y,$$

as a region to be analyzed with regard to SP entrance and exit rates. Rectangle $\mathcal{R}_{x,y}$ is an open set. Since the state r.v.'s have continuous components, the analysis generally leads to identical results whether choosing an open or closed rectangle. Care must be taken in some models where the joint pdf takes different forms on subsets of $\boldsymbol{S}$ that are half open, etc.

Level as Boundary of Test Rectangle

Denote the "**north-east**" boundary of rectangle $\mathcal{R}_{x,y}$ by $\urcorner_{(a,b)}^{(x,y)}$. Then

$$\urcorner_{(a,b)}^{(x,y)} \equiv \{(x,\beta)\,|a < \beta < y\} \cup \{(\alpha,y)\,|b < \alpha < x\}.$$

The boundary $\urcorner_{(a,b)}^{(x,y)}$ consists of the union of two perpendicular edges of $\mathcal{R}_{x,y}$ shaped like the letter *el* rotated 180^o, namely "$\urcorner$" (Fig. 7.2). We call the set $\boldsymbol{T} \times \urcorner_{(a,b)}^{(x,y)}$ a level-$\urcorner_{(a,b)}^{(x,y)}$ "contour" in $\boldsymbol{T} \times \boldsymbol{S}$ (see Fig. 7.2, in which $a = b = 0$). We shall call the set $\boldsymbol{T}\times\urcorner_{(a,b)}^{(x,y)}$ a **2-dimensional level** in $\boldsymbol{T}\times\boldsymbol{S}$, or simply *a level in* $\boldsymbol{S}$. The context determines the dimension of the level.

Time Axis and Level

The time axis is $\boldsymbol{T} = \{t \in [0,\infty)\}$. The level $\boldsymbol{T}\times\urcorner_{(a,b)}^{(x,y)}$ is the union of two perpendicular planar strips of width $x-a$, $y-b$, extending to infinity with respect to $\boldsymbol{T}$, in the three-dimensional space $\boldsymbol{T} \times \boldsymbol{S}$. Pictorially, $\boldsymbol{T} \times \urcorner_{(a,b)}^{(x,y)}$ is a surface resembling an "edge guard" (or edge protector in carpentry), extending to infinity in the direction of $\boldsymbol{T}$. A plot of $\boldsymbol{T}\times\urcorner_{(a,b)}^{(x,y)}$ in the time-state space when $a = b = 0$ is given in Fig. 7.2.

Role of Level in Analysis

Assume system stability holds. Denote the steady-state *joint* pdf of the continuous r.v.'s by $f(x,y), x \geq 0, y \geq 0$. We compute SP transition rates across level $\urcorner_{(a,b)}^{(x,y)}$ in terms of values $\{f(\alpha,\beta)|$a crossing of $\urcorner_{(a,b)}^{(x,y)}$

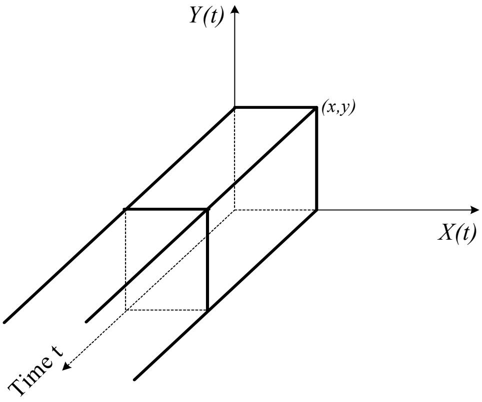

Figure 7.2: Assume $\{X(t), Y(t)\}, t \geq 0$ is a $2_{(2,0)}$-dimensional process with $(X(t), Y(t)) \in \boldsymbol{R}^2$ restricted to the non-negative quadrant. The level set $\rceil_{(0,0)}^{(x,y)} \equiv \{(x, \beta)\,|0 < \beta < y\} \cup \{(\alpha, y)\,|0 < \alpha < x\}$ is shown over time, and also projected on the $(X(t), Y(t))$ plane. Cross product $\boldsymbol{T} \times \rceil_{(0,0)}^{(x,y)}$ is a surface in 3-dimensional Euclidian space.

is possible starting from $(\alpha, \beta)\}$. These transition rates are determined using the probabilistic laws and any prescribed laws governing the model. Rate balance across level $\rceil_{(a,b)}^{(x,y)}$ results in one or more integral equations for $f(x, y)$.

7.2 Two Products Sharing Limited Storage

Consider an inventory system with two products that share a common limited storage facility having total capacity Q. Assume that product 1 is governed by an $\langle s, S\rangle$-like ordering policy, and product 2 by an EOQ-like ordering policy. Let $I_1(t), I_2(t), t \geq 0$ denote the stock on hand at time t of products 1 and 2 respectively. Assume that $I_1(t)$, $I_2(t)$ are continuous r.v.'s. Suppose that the parameters are such that the joint process $\{I_1(t), I_2(t)\}$ is stable. The state is classified as $2_{(2,0)}$-dimensional.

Denote the steady-state joint cdf of $I_1(t), I_2(t)$ as $t \to \infty$ by

$$F(x, y) = \lim_{t\to\infty} P(I_1(t) \leq x, I_2(t) \leq y), (x, y) \in \boldsymbol{S},$$

having joint pdf $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$, wherever the underlying partial derivatives exist.

We shall analyze two elementary versions of the model in order to demonstrate the use of LC. To further focus on LC, we assume no lead times or backlogging or product decay, etc. There are many plausible variations on the state space $\boldsymbol{S}$ and on the ordering policies. However, for elucidation, we analyze the model when the state space $\boldsymbol{S}$ is *relatively* simple, and the ordering policies are well known.

Assume the total available storage space is $Q > 0$ units. The units of the two products are assumed to be the same as the units of Q. For example, suppose Q is measured in cubic meters (m^3). If the products' units are not the same, we convert all units to cubic meters. Product 1 may consist of 2-meter (length) × 5-centimeter (outer diameter) plastic pipes which can be cut into continuously variable lengths. Product 2 may be 0.5-inch thick 8-feet by 4-feet ($8'$ by $4'$) plywood sheets, which can be cut into continuously variable rectangles in the $8'$ by $4'$ plane. We would convert all volumes to m^3.

An example where the units are the same as that of Q, is where the products are two different agricultural grains sharing a single storage space, such that they can be retrieved separately to satisfy demands. We will not address the accompanying "packing" problems here. We treat the inventory problem only, and assume that a model with continuous state variables is appropriate.

7.3 Two Products Sharing Storage: Model 1

7.3.1 Policies for Products

<u>**Product** 1</u>

Product 1 follows a modified $\langle s, S \rangle$ policy with *no decay.* For product 1 assume there is a Poisson demand rate $\lambda > 0$ per unit time and the demand sizes are $\underset{dist}{=} \mathrm{E}_\mu$. Product 1 does not decay. Its stock on hand remains constant until the next demand for it occurs.

If a demand for product 1 at instant t_0^- causes the stock-on-hand of product 1 to jump downward below a fixed level $s, 0 \leq s < Q$, an order is placed for product 1. The order is filled immediately. The amount received satisfies any shortage caused by the demand and replenishes the stock up to the *available space* $Q - I_2(t_0^-)$ at instant t_0.

Product 2

Product 2 follows a modified EOQ policy. For product 2 there is a *constant* demand rate equal to $k > 0$ units per unit time independent of the amount of stock on hand of product 2.

If the stock on hand of product 2 hits level 0 from above at t_0^- an order is placed and received immediately. The order replenishes the stock of product 2 up to the available space $Q - I_1(t_0^-)$ at t_0.

7.3.2 State Space $\boldsymbol{S}$

The *regular* state space $\boldsymbol{S}_r$ is a finite right-angled triangle with vertices at $(s, 0)$, $(Q, 0)$ and $(s, Q - s)$ (Fig. 7.3). We assume $s \geq 0$. The state space in the *wide sense* is

$$\boldsymbol{S} = \boldsymbol{S}_r \cup \{(\alpha, \beta) \, |\alpha < s, 0 \leq \beta \leq Q - s|\},$$

which appends an infinite rectagular region to the left of $\boldsymbol{S}_r$. We use $\boldsymbol{S}$ since the SP jumps into the infinite rectangular area

$$\{(\alpha, \beta) \, |\alpha < s, 0 \leq \beta \leq Q - s|\}$$

when an order for product 1 occurs (see Subsection 2.3.1 regarding a wide-sense state space).

Remark 7.2 *There will always be a positive amount of each product on hand, except possibly for an initial finite time period. For, suppose at* $t = 0$ *the state is* $(I_1(0), I_2(0)) = (Q, 0)$ *(all space used for product* 1*, and no product* 2 *present). The state will remain* $(Q, 0)$ *until a demand for product* 1 *of some size* $o_1 < Q - s$*, occurs for the first time. A storage space of size* $Q - o_1$ *will become available at that instant.*

The number of product-1 demands until such an o_1 *occurs is geometrically distributed. That is, product* 1 *may have successive demands of size* $> Q - s$ *before a demand of size* o_1 *occurs. The probability of "failure"* $= e^{-\mu(Q-s)}$*; the probability of "success"* $= (1 - e^{-\mu(Q-s)})$*. The expected number of demands until obtaining* $o_1 < Q - s$ *is* $\frac{1}{(1-e^{-\mu(Q-s)})}$*. The time between demands has mean* $\frac{1}{\lambda}$*. Thus, the expected* ***time*** *until a size* o_1 *demand occurs is* $\frac{1}{\lambda(1-e^{-\mu(Q-s)})}$*.*

*We assume that if a product-*1 *demand of size* o_1 *occurs, then an order for product* 2 *of size* o_1 *is placed and received immediately to fill the space. The resulting state becomes* $(Q - o_1, o_1)$*. From that instant*

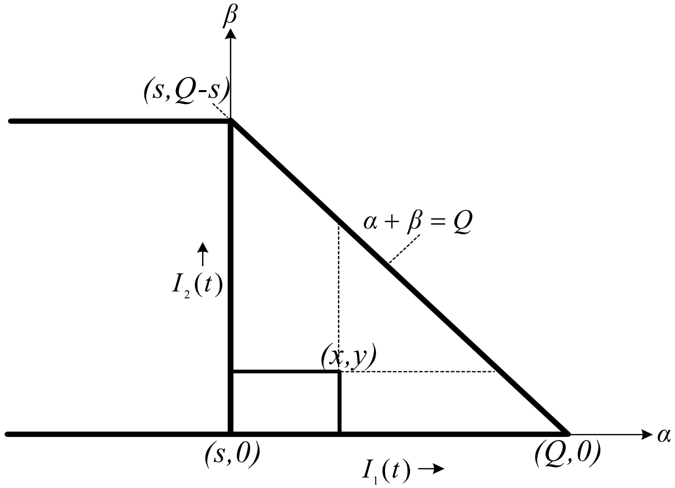

Figure 7.3: State space $\boldsymbol{S}$ in inventory model with two products sharing limited storage of capacity Q. Shows level $\rceil_{(0,0)}^{(x,y)}$. Indicates two trapezoidal regions of $\boldsymbol{S}$ from which SP can traverse $\rceil_{(0,0)}^{(x,y)}$ due to product demands. $\boldsymbol{S}$ includes infinite rectangle $\{(\alpha, \beta)\,|\alpha < s, 0 \le \beta \le Q - s|\}$ appended to $\boldsymbol{S}_r$ on the left.

on, the ordering policies preclude the system ever returning to the state $(Q, 0)$*, due to the continuity of the demand sizes of product* 1*. Thus state* $(Q, 0)$ *is not recurrent. Therefore* $(Q, 0)$ *is not an atom (has probability* 0*).*

7.3.3 Sample Path

A sample path consists partly of line segments with slope k in planes parallel to the $(t, I_2(t))$ plane, traced out by the SP for time intervals which are $\underset{dist}{=} \mathrm{E}_\lambda$. At the end of these time intervals the SP jumps due to product-1 demands, to planes closer to the $(t, I_2(t))$ plane, unless it jumps past the reorder plane $\alpha = s$. These jumps take place in planes parallel to the $(t, I_1(t))$ plane. The jump sizes are $\underset{dist}{=} \mathrm{E}_\mu$. If a jump crosses the plane $\alpha = s$, the SP 'double" jumps back to the right, to the order-up-to plane $\alpha + \beta = Q$, corresponding to a product-1 order being received and filling the available space.

If the SP makes a continuous hit of level 0 from above, it jumps upward parallel to the $(t, I_2(t))$ plane, to hit the plane $\alpha + \beta = Q$ from

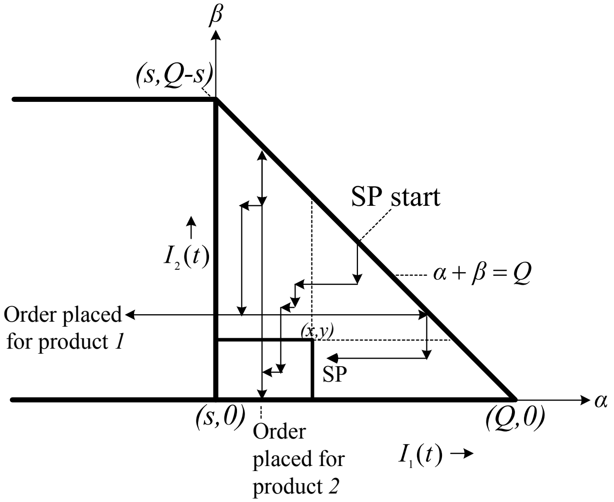

Figure 7.4: Sample path projected onto $(I_1(t), I_2(t))$ plane in Model 1 of inventory with two products sharing total space Q. The vertical line segments are projections of a line of slope k relative to the $(t, I_2(t))$ plane. The horizontal line segments are projections of horizontal line segments relative to the $(t, I_1(t))$ plane.

below, to fill the available space.

A possible sample path *projected onto the* $(I_1(t), I_2(t))$ *plane* is shown in Fig. 7.4. The projection of the line with slope k relative to the $(t, I_2(t))$ plane from $\boldsymbol{T} \times \boldsymbol{S}$ onto the $(I_1(t), I_2(t))$ plane, is a vertical line segment.

Perspectives of possible sample paths projected onto the $(t, I_1(t))$ and $(t, I_2(t))$ planes are given in figures 7.5 and 7.6 respectively.

7.3.4 Integral Equation for Steady-state Joint PDF

Fix $(x, y) \in \boldsymbol{S}, x > s, y > 0$. Consider level $\rceil_{(s,0)}^{(x,y)}$ (Fig. 7.4). The rate at which the SP crosses $\rceil_{(s,0)}^{(x,y)}$ from the right (east) is given by

$$\lambda \int_{\beta=0}^{y} \int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)} f(\alpha, \beta) d\alpha d\beta,$$

since all SP crossings from the right are due to jumps corresponding to demands for product 1. The demand rate is λ. A jump starting at $\alpha > x$ causes the SP to cross the vertical edge of level $\rceil_{(s,0)}^{(x,y)}$ with probability

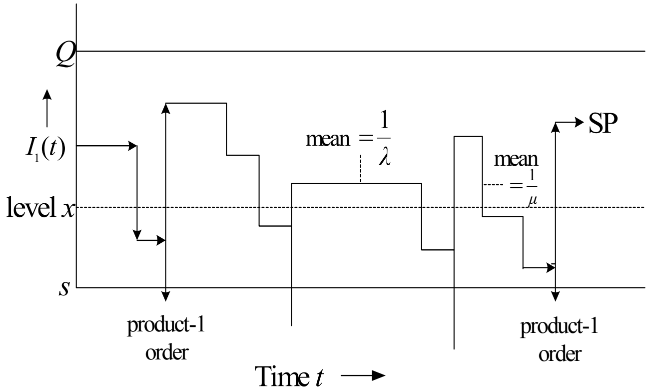

Figure 7.5: Possible sample path in Model 1 of two-product inventory with limited storage, projected onto the $(t, I_1(t))$ plane. Shows perspective of SP motion for product 1.

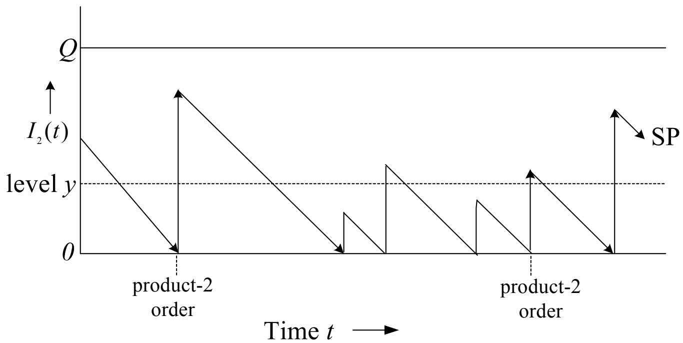

Figure 7.6: Possible sample path in Model 1 of two-product inventory with limited storage, projected onto the $(t, I_2(t))$ plane. Shows perspective of SP motion for product 2.

$e^{-\mu(\alpha-x)}$. SP jumps that cross level $\rceil_{(s,0)}^{(x,y)}$ from right to left must originate in the trapezoidal region $\{((\alpha,\beta))\,|x<\alpha<Q-\beta,0<\beta<y\}$.

The rate at which the SP crosses level $\rceil_{(s,0)}^{(x,y)}$ from above is given by

$$k\int_{\alpha=s}^{x} f(\alpha,y)d\alpha,$$

since the demand for product 2 is constant at rate k and the SP downcrossing rate of a point (α,y) is $kf(\alpha,y)$ (see Corollary 6.2). That the SP downcrossing rate is $kf(\alpha,y)$ at (α,y), can be proved by a slight modification of Theorem 1.1 where the SP declines at an arbitrary slope $k>0$.

Thus, the total SP crossing rate of level $\rceil_{(s,0)}^{(x,y)}$ from the right and from above is

$$\lambda\int_{\beta=0}^{y}\int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)}f(\alpha,\beta)d\alpha d\beta + k\int_{\alpha=s}^{x} f(\alpha,y)d\alpha. \tag{7.1}$$

Similarly the SP crossing rate of level $\rceil_{(s,0)}^{(x,y)}$ *from the left or from below* is

$$\lambda\int_{\beta=0}^{y}\int_{\alpha=s}^{Q-\beta} e^{-\mu(\alpha-s)}f(\alpha,\beta)d\alpha d\beta + k\int_{\alpha=s}^{x} f(\alpha,0)d\alpha, \tag{7.2}$$

by applying the ordering policies for product 1 and product 2. That is, a horizontal SP jump across s from the right due to a product-1 demand, double-jumps instantaneously (rebounds) to the right, ending at the plane $\{(\alpha,\beta)\,|\alpha+\beta=Q\}$ (first term in (7.2)). An SP continuous hit of level 0 from above signals an instantaneous SP vertical upward jump (rebound) to the plane $\{(\alpha,\beta)\,|\alpha+\beta=Q\}$ (second term in (7.2)).

SP rate balance across level $\rceil_{(s,0)}^{(x,y)}$ equates (7.1) and (7.2), giving integral equation

$$\begin{aligned}&\lambda\textstyle\int_{\beta=0}^{y}\int_{\alpha=x}^{Q-\beta} e^{-\mu(\alpha-x)}f(\alpha,\beta)d\alpha d\beta + k\int_{\alpha=s}^{x} f(\alpha,y)d\alpha\\ &\quad=\lambda\textstyle\int_{\beta=0}^{y}\int_{\alpha=s}^{Q-\beta} e^{-\mu(\alpha-s)}f(\alpha,\beta)d\alpha d\beta + k\int_{\alpha=s}^{x} f(\alpha,0)d\alpha.\end{aligned} \tag{7.3}$$

Form of Solution

Taking $\frac{\partial}{\partial y}$ once and $\frac{\partial}{\partial x}$ twice on both sides of (7.3) leads to a second order PDE (partial differential equation) for $f(x,y)$

$$\begin{aligned}&\tfrac{\partial^2}{\partial x\partial y}f(x,y)-\mu\tfrac{\partial}{\partial y}f(x,y)-\tfrac{\lambda}{k}\tfrac{\partial}{\partial x}f(x,y)=0,\\ &\qquad s<x<Q-y,0<y<Q-s.\end{aligned} \tag{7.4}$$

Applying *separation of variables* for PDE's to (7.4), let $f(x,y) = g(x)h(y)$ (e.g., [6]). Then (7.4) reduces to

$$\frac{g'(x)}{g(x)} = \frac{\mu h'(y)}{h'(y) - \frac{\lambda}{k}h(y)} = \sigma,$$

where the derivatives are taken with respect to the corresponding variables and σ is a constant to be determined. Thus

$$\frac{d\ln g(x)}{dx} = \sigma, \quad \frac{d\ln h(y)}{dy} = \frac{\lambda\sigma}{k(\sigma-\mu)},$$

with solutions

$$g(x) = Ae^{\sigma x}, \;\; h(y) = Be^{\frac{\lambda\sigma}{k(\sigma-\mu)}y}, \tag{7.5}$$

where A, B are constants. We next evaluate the constant σ.

Value of Constant σ

We utilize a "boundary" condition to evaluate σ. Consider a point $(x, Q-x)$ on the north-east boundary of $\boldsymbol{S}$, namely $\{(\alpha,\beta)\,|\alpha+\beta=Q\}$.

The SP total rate into $(x,Q-x)$ from the left and from below, is

$$\lambda\int_{\alpha=s}^{x} e^{-\mu(\alpha-s)}f(\alpha,Q-x)d\alpha + kf(x,0),$$

where the first term is due to product-1 demands that signal product-1 orders, given that product 2 is at level $Q-x$; and the second term is the rate of product 2 demands that signal product 2 orders, given that product 1 is at level x.

The SP rate out of $(x,Q-x)$ to the left and downward, is

$$\lambda f(x,Q-x) + kf(x,Q-x),$$

where the first term is due to product-1 orders when the state is $(x,Q-x)$ and the second term is the SP rate out of $(x,Q-x)$ due to the constant demand for product 2.

Equating the SP rates into and out of $(x,Q-x)$ gives

$$\lambda\int_{\alpha=s}^{x} e^{-\mu(\alpha-s)}f(\alpha,Q-x)d\alpha + kf(x,0) = \lambda f(x,Q-x) + kf(x,Q-x). \tag{7.6}$$

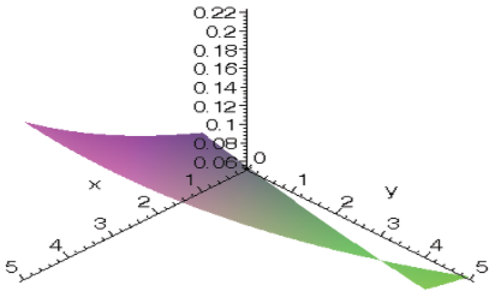

Figure 7.7: Joint pdf $f(x.y), s < x < Q - y, 0 < y < Q$ in model 1 of two product inventory with limited storage: example with $Q = 5$, $s = 1$, $\mu = 1$, $\lambda = 1.5$, $k = 2.5$.

From (7.5), substituting for $g(\cdot)$, $h(\cdot)$ in (7.6), simplifying and letting $x \downarrow s$ leads to

$$\sigma = \frac{k\ln(1+\frac{\lambda}{k})}{k\ln(1+\frac{\lambda}{k}) + \lambda(Q-s)}, \tag{7.7}$$

provided $\sigma \neq \mu$. The value $\sigma = \mu$ is impossible; otherwise $h(y)$ would be infinite for all y.

7.3.5 Solution for Joint PDF of Inventory

From (7.5), the steady-state joint pdf of inventory is

$$f(x,y) = ABe^{\sigma x} e^{\frac{\lambda\sigma}{k(\sigma-\mu)}y} = Ce^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y}$$

where σ is given in (7.7) and constant $C = AB$. The value of C is obtained from the normalizing condition

$$\begin{aligned} \int_{y=0}^{Q-s}\int_{x=s}^{Q-y} f(x,y)dxdy &= \int_{y=0}^{Q-s}\int_{x=s}^{Q-y} Ce^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y} = 1, \\ C &= \frac{1}{\int_{y=0}^{Q-s}\int_{x=s}^{Q-y} e^{\sigma x + \frac{\lambda\sigma}{k(\sigma-\mu)}y}}. \end{aligned} \tag{7.8}$$

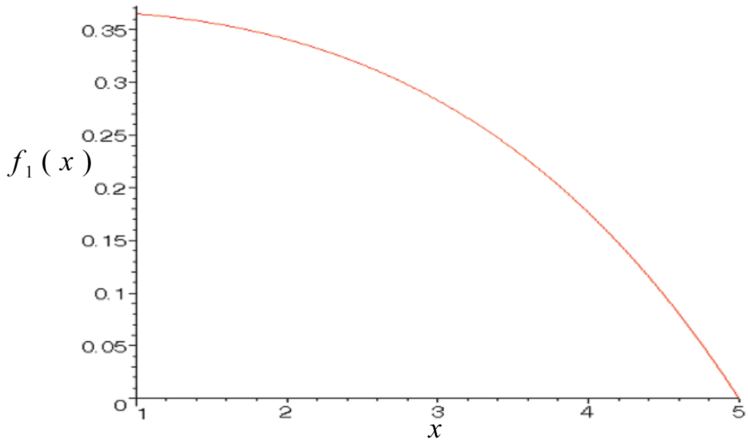

Figure 7.8: Marginal pdf $f_1(x), s < x < Q$ for product 1 in Model 1 of inventory with two products sharing limited storage: example with $Q = 5,\ s = 1,\ \mu = 1,\ \lambda = 1.5,\ k = 2.5$.

Example 7.2 *Consider Model 1 with arbitrary parameter values*

$$Q = 5, s = 1, \mu = 1, \lambda = 1.5, k = 2.5.$$

Then

$$f(x, y) = Ce^{\sigma x + \frac{0.6\sigma}{\sigma - 1} y}$$

and from (7.7) $\sigma = 0.1638$. *From (7.8)* $C = 0.0971$. *Thus (Fig. 7.7)*

$$f(x, y) = 0.0971e^{0.1638x - 0.1175y}, 1 < x < 5 - y, 0 < y < 4.$$

The marginal pdf of product 1 *is* $f_1(x) = \int_{y=0}^{Q-x} f(x, y)dy$ *or*

$$f_1(x) = 0.8266(e^{0.16376x} - e^{0.2813x - 0.5875)})$$

(Fig. 7.8). The marginal pdf of product 2 *is* $f_2(y) = \int_{x=s}^{Q-y} f(x, y)dy$ *or*

$$f_2(y) = -0.5931(e^{0.1638 - 0.1175y} + e^{0.8188 - 0.2813y})$$

(Fig. 7.9).

Let I_1 *and* I_2 *denote the limiting state variables for products 1 and 2 respectively, as* $t \to \infty$. *The expected values and variances of* I_1 *and* I_2 *are respectively*

$$E(I_1) = 2.5392,\ \ E(I_2) = 1.1617,$$

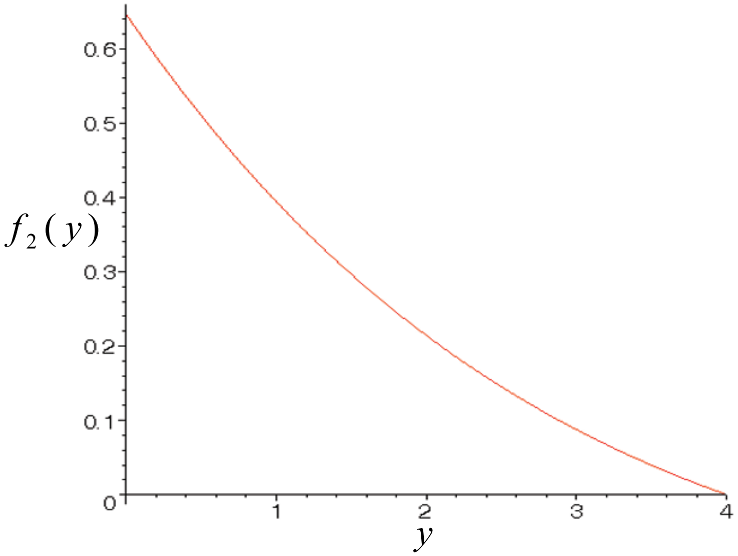

Figure 7.9: Marginal pdf $f_2(y), 0 < y < Q - s$ for product 2 in Model 1 of inventory with two products sharing limited storage: example with $Q = 5$, $s = 1$, $\mu = 1$, $\lambda = 1.5$, $k = 2.5$.

$$Var(I_1) = 0.9791, \; Var(I_2) = 0.7851.$$

The covariance is $Cov(I_1, I_2) = -0.4457$*. The correlation coefficient between* I_1 *and* I_2 *is*

$$\rho_{I_1,I_2} = \frac{Cov(I_1, I_2)}{\sqrt{Var(I_1)}\sqrt{Var(I_2)}} = -0.5084.$$

Intuitively, we expect ρ_{I_1,I_2} *to be negative. That is, if there is a high stock on hand of product* i*, then there is generally a low stock on hand of product* $3 - i, i = 1, 2$ *and vice versa, since the sum of the stocks on hand is bounded by* Q*.*

7.4 Two Products Sharing Storage: Model 2

We present a variant (Model 2) of the two-product inventory model in which the products share storage space. This variant has a $2_{(2,0)}$-dimensional state with a *probabilistic atom*. Model 2 places specific limits on the amounts of the two products in storage simultaneously. We analyze an "extreme" model where $\boldsymbol{S}$ is such that the resulting joint pdf

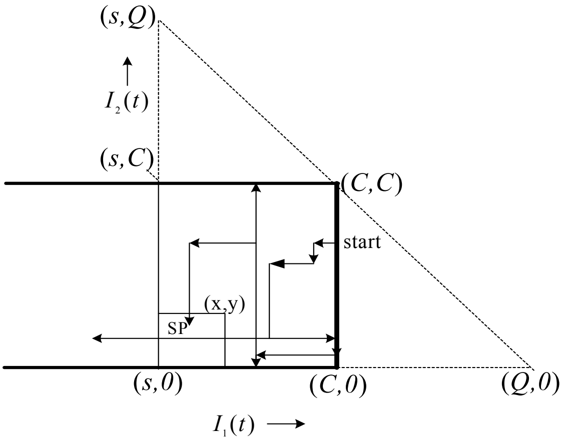

Figure 7.10: State space $\boldsymbol{S}$ for Model 2, with atom along right edge $\{(\alpha,\beta)\,|\alpha = C, 0 < \beta < C\}$. Shows possible sample path projected onto $(I_1(t), I_2(t))$ plane. Also shows level $\rceil_{(s,0)}^{(x,y)}$.

of stock on hand serves as a check on intuition and a mild check on the method of analysis.

7.4.1 Model 2 Description

Suppose that the amount of each product on hand is $\leq C$ $(= \frac{Q}{2})$. The "regular" state space is $\boldsymbol{S}_r = \{(\alpha,\beta)\,|s < \alpha \leq C, 0 < \beta \leq C\}$. The wide-sense state space is

$$\boldsymbol{S} = \boldsymbol{S}_r \cup \{(\alpha,\beta)\,|\alpha < s, 0 < \beta < C\}\,.$$

The ordering policies for products 1 and 2 are the same as in Model 1 (Subsection 7.3.1). However, in Model 2, product 1 can have C units in storage for a time period $\underset{dist}{=}$ E_λ. This period is followed by a demand which is $\underset{dist}{=}$ E_μ. Product 2 can have C units in storage only for an instant (at an order instant), as demand for it is continuous (at rate k) (see Fig. 7.10). In Model 2 the boundary $\{(C,\beta)\,|0 < \beta < C\}$ is an ***atom*** with positive probability.

Denote the steady-state joint pdf of inventory by

$$f(x,y), \{(x,y)\,|s < x < C, 0 < y \leq C\}\,,$$

and denote the pdf along edge $\{(C,\beta)\,|0 < \beta < C\}$ by $\Pi_C(y), 0 < y \leq C$.

The normalizing condition is

$$\int_{y=0}^{C}\int_{x=s}^{C} f(x,y)dxdy + \int_{y=0}^{C} \Pi_C(y)dy = 1. \tag{7.9}$$

7.4.2 Integral Equation for Joint PDF of Inventory

Fix the point $(x,y) \in \boldsymbol{S}$, $s < x < C$, $0 < y < C$ (Fig. 7.10). Reasoning as in Subsection 7.3.4 for Model 1, the SP rate from the right and from above across level $\rceil_{(s,0)}^{(x,y)}$ is

$$\begin{aligned}&\lambda \int_{\beta=0}^{y}\int_{\alpha=x}^{C} e^{-\mu(\alpha-x)} f(\alpha,\beta)d\alpha d\beta \\ &+\lambda \int_{\beta=0}^{y} e^{-\mu(C-x)}\Pi_C(\beta)d\beta + k\int_{\alpha=s}^{x} f(\alpha,y)d\alpha.\end{aligned}$$

The SP rate across level $\rceil_{(s,0)}^{(x,y)}$ from the left and from below, is

$$\begin{aligned}&\lambda \int_{\beta=0}^{y}\int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta)d\alpha d\beta \\ &+\lambda \int_{\beta=0}^{y} e^{-\mu(C-s)}\Pi_C(\beta)d\beta + k\int_{\alpha=s}^{x} f(\alpha,0)d\alpha.\end{aligned}$$

Rate balance across level $\rceil_{(s,0)}^{(x,y)}$ yields the integral equation

$$\begin{aligned}&\lambda \textstyle\int_{\beta=0}^{y}\int_{\alpha=x}^{C} e^{-\mu(\alpha-x)} f(\alpha,\beta)d\alpha d\beta \\ &\qquad +\lambda \textstyle\int_{\beta=0}^{y} e^{-\mu(C-x)}\Pi_C(\beta)d\beta + k\int_{\alpha=s}^{x} f(\alpha,y)d\alpha \\ &= \lambda \textstyle\int_{\beta=0}^{y}\int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta)d\alpha d\beta \\ &\qquad +\lambda \textstyle\int_{\beta=0}^{y} e^{-\mu(C-s)}\Pi_C(\beta)d\beta + k\int_{\alpha=s}^{x} f(\alpha,0)d\alpha.\end{aligned} \tag{7.10}$$

Note that

$$\begin{aligned}&\lambda \textstyle\int_{\beta=0}^{y}\int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} f(\alpha,\beta)d\alpha d\beta + \lambda\int_{\beta=0}^{y} e^{-\mu(C-s)}\Pi_C(\beta)d\beta \\ &\quad = \lambda \textstyle\int_{\beta=0}^{y} \Pi_C(\beta)d\beta,\end{aligned} \tag{7.11}$$

since the SP total crossing rate from the right of the re-order boundary $\{(s,\beta)|0 < \beta < y\}$, is equal to the SP jump rate to the left out of atomic boundary $\{(C,\beta)\,|0<\beta<y\}$ due to demands for product 1. Thus (7.10) simplifies to

$$\begin{aligned}&\lambda\int_{\beta=0}^{y}\int_{\alpha=x}^{C}e^{-\mu(\alpha-x)}f(\alpha,\beta)d\alpha d\beta\\&\quad+\lambda\int_{\beta=0}^{y}e^{-\mu(C-x)}\Pi_C(\beta)d\beta+k\int_{\alpha=s}^{x}f(\alpha,y)d\alpha\\&=\lambda\int_{\beta=0}^{y}\Pi_C(\beta)d\beta+k\int_{\alpha=s}^{x}f(\alpha,0)d\alpha,\, s<x<C,0<y\le C.\end{aligned}\tag{7.12}$$

7.4.3 Solution of Integral Equation

Taking $\frac{\partial}{\partial y}$ once and $\frac{\partial}{\partial x}$ twice in (7.12) leads to the second order PDE

$$\frac{\partial^2}{\partial x\partial y}f(x,y)-\mu\frac{\partial}{\partial y}f(x,y)-\frac{\lambda}{k}\frac{\partial}{\partial x}f(x,y)=0,$$
$$s<x<C,0<y\le C.$$

Applying the separation of variables technique for PDE's (as in Subsection 7.3.4) yields the solution

$$f(x,y)=ABe^{\sigma x}e^{\frac{\lambda\sigma}{k(\sigma-\mu)}y}=ABe^{\sigma x+\frac{\lambda\sigma}{k(\sigma-\mu)}y},\tag{7.13}$$

where the constant σ is to be determined.

Taking $\frac{\partial}{\partial y}$ in (7.11) gives

$$\Pi_C(y)(1-e^{-\mu(C-s)})=\int_{\alpha=s}^{C}e^{-\mu(\alpha-s)}f(\alpha,y)d\alpha,0<y\le C.\tag{7.14}$$

<u>**Solution for Constant σ**</u>

Consider the rectangular edge $\{(\alpha,C)\,,s<\alpha<C\}\in\boldsymbol{S}$. The SP rate *out* of the edge is $k\int_{\alpha=s}^{C}f(\alpha,C)d\alpha$ due to the constant ordering rate k of product 2. Any order for product 1 will not move the SP out of this edge. The SP rate *into* the edge is $k\int_{\alpha=s}^{C}f(\alpha,0)d\alpha$ due to orders for product 2 when product 2 becomes depleted to 0. Rate balance for this edge gives

$$k\int_{\alpha=s}^{C}f(\alpha,C)d\alpha=k\int_{\alpha=s}^{C}f(\alpha,0)d\alpha.\tag{7.15}$$

Substitute from (7.13) into (7.15) and cancel k, AB from both sides. Note that if $\sigma=0$ then $\int_{\alpha=s}^{C}e^{\sigma\alpha}da=C-s>0$. If $\sigma\neq0$ then $\int_{\alpha=s}^{C}e^{\sigma\alpha}da=\frac{e^{\sigma C}-e^{\sigma s}}{\sigma}\neq0$. Thus we may also cancel $\int_{\alpha=s}^{C}e^{\sigma\alpha}da$ from both sides. This leads to the equation for σ

$$e^{\frac{\lambda\sigma}{k(\sigma-\mu)}C}=1.\tag{7.16}$$

Solving (7.16) for σ gives the value

$$\sigma = 0. \tag{7.17}$$

Solution for Constant AB

From (7.13) and (7.17)

$$f(x,y) = AB, s < x < C, 0 < y \leq C.$$

Substituting into (7.14) gives

$$\begin{aligned} \Pi_C(y)(1 - e^{-\mu(C-s)}) &= \int_{\alpha=s}^{C} e^{-\mu(\alpha-s)} AB d\alpha, \\ \text{or} \qquad \Pi_C(y) &= \frac{AB}{\mu}, 0 < y \leq C. \end{aligned}$$

The normalizing condition (7.9) gives

$$AB \left(\int_{y=0}^{C} \int_{x=s}^{C} dx dy + \frac{1}{\mu} \int_{y=0}^{C} dy \right) = 1.$$

Hence

$$AB = \frac{\mu}{C(1 + \mu(C - s))}. \tag{7.18}$$

From (7.18)

$$f(x,y) = \frac{\mu}{C(1 + \mu(C - s))}, s < x < C, 0 < y \leq C, \tag{7.19}$$

and

$$\Pi_C(y) = \frac{1}{C(1 + \mu(C - s))}, 0 < y \leq C. \tag{7.20}$$

Let $\Pi_C = \int_{y=0}^{C} \Pi_C(y) dy$.

From (7.20)

$$\Pi_C = \frac{1}{(1 + \mu(C - s))}. \tag{7.21}$$

7.4.4 Marginal PDF's of Stock on Hand

From (7.19) the marginal pdf for product 1 in the interval $s < x < C$ is

$$\begin{aligned} f_1(x) &= \int_{y=0}^{C} f(x,y)dy \\ &= \frac{\mu}{(1+\mu(C-s))}, s < x < C. \end{aligned} \tag{7.22}$$

The complete mixed marginal pdf for product 1 is

$$\{f_1(x);\Pi_C\} = \left\{ \frac{\mu}{(1+\mu(C-s))}, s < x < C; \frac{1}{(1+\mu(C-s))} \right\}$$

Note that $\int_{x=s}^{C} f_1(x)dx + \Pi_C = 1$.

From (7.19) and (7.20) the marginal pdf for product 2 is

$$\begin{aligned} f_2(y) &= \int_{x=s}^{C} f(x,y)dx + \Pi_C(y) \\ &= \frac{\mu(C-s)}{C(1+\mu(C-s))} + \frac{1}{C(1+\mu(C-s))} \\ &= \frac{1+\mu(C-s)}{C(1+\mu(C-s))} = \frac{1}{C}, 0 < y \leq C. \end{aligned} \tag{7.23}$$

Note that (7.22) is identical to (6.90) with the order-up-to level S replaced by C. Intuitively, this result holds because the ordering policy for product 1 is of the $\langle s, S\rangle$ type with no decay, and the state space $\boldsymbol{S}$ is rectangular.

Similarly, (7.23) is uniform on $(0, C]$, which is a well known result for the stationary distribution in a standard EOQ model.

The motion of the SP in $\boldsymbol{T} \times \boldsymbol{S}$ is affected by orders of both product types. Nevertheless the stock on hand of products 1 and 2 in steady state are statistically independent, corroborated by the relationships between the joint pdf and marginal pdf's,

$$\begin{aligned} & f(x,y) = f_1(x) \cdot f_2(y), s < x < C, 0 < y \leq C, \\ & \Pi_C(y) = \Pi_C \cdot f_2 C), 0 < y \leq C. \end{aligned} \tag{7.24}$$

Remark 7.3 *Model 2 serves as a mild check on the LC method for analyzing $2_{(2,0)}$-dimensional models. Intuitively we expect statistical independence of the stock on hand of the two products. Indeed, the marginal pdf's turn out as expected for such independence. The stock on hand of each product is independent of the stock on hand of the companion product.*

7.4.5 Summary

In this chapter we have used LC to analyze two variants of a model in which two products share the same total storage space. There are many different ordering policies, different types of constraints, and modified state spaces possible for such variants. The model variants would have unique corresponding steady-state joint and marginal pdf's of stock on hand for the products.

We can analyze a vast array of other $2_{(2,0)}$-dimensional models by applying a similar LC technique. These include various types of inventory, production-inventory, queueing-network, natural-science models, etc. A similar remark applies to a vast array of $n_{(n,0)}$-dimensional models, $n = 3, 4, ...$ We can also extend the analysis to $n_{(c,d)}$-dimensional models where $c + d = n$ and both $c > 0$, $d > 0$.

CHAPTER 8

EMBEDDED LEVEL CROSSING METHOD

8.1 Dams and Queues

Consider a system modelled by $\{W(t), t \geq 0\}$, a continuous-parameter process with state space $\boldsymbol{S} = [0, \infty)$. (The state space can be extended to $\boldsymbol{S} \subseteq \boldsymbol{R^n}$ in more general models.) Let $\{\tau_n\}$ be an infinite set of embedded time points such that

$$0 \leq \tau_1 < \tau_2 < \cdots < \tau_n < \tau_{n+1} < \cdots .$$

Let $\{W_n, n = 1, 2, ...\}$ be the embedded discrete-parameter process, where $W(\tau_n^-) \equiv W_n$ and $W(\tau_n) \equiv W_n + S_n, n = 1, 2, ...$. Assume $W(t)$ is monotone in the interval $[\tau_n, \tau_{n+1})$.'Let

$$\frac{dW(t)}{dt} = -r(W(t)), t \in [\tau_n, \tau_{n+1}), n = 1, 2, ... ,$$

where $r(x) \geq 0$. Denote the cdf of $S_n, n = 1, 2, ...,$ by $B(x), x \geq 0$, with $B(0) = 0$, and pdf $b(x) = \frac{d}{dx}B(x), x > 0$, wherever the derivative exists. Denote the cdf of W_n by $F_n(x)$ with pdf $\frac{dF_n(x)}{dx} = f_n(x)$, wherever it exists.

Definition 8.1 *An* ***embedded downcrossing*** *of state-space level x occurs during the closed interval $[\tau_n, \tau_{n+1}]$ if $W_n > x \geq W_{n+1}$.*

An ***embedded upcrossing*** *of level x occurs during $[\tau_n, \tau_{n+1}]$ if $W_n \leq x < W_{n+1}$.*

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_8,

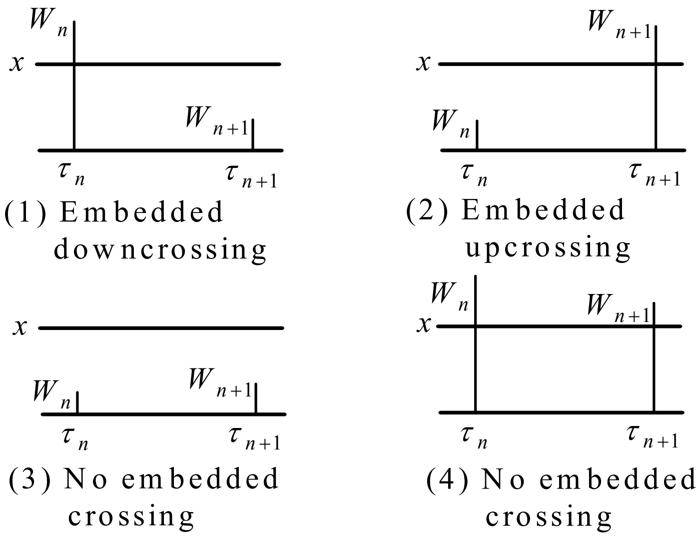

Figure 8.1: Embedded level crossings and non-crossings during time interval $[\tau_n, \tau_{n+1}]$.

Fix level $x \in \boldsymbol{S}$. Definition 8.1 classifies the set of intervals

$$\{[\tau_n, \tau_{n+1}], n = 1, 2, ...\}$$

into three mutually exclusive and exhaustive subsets with respect to level x (Fig. 8.1):

1. intervals that contain an embedded downcrossing,
2. intervals that contain an embedded upcrossing,
3. intervals that contain no embedded level crossing.

8.1.1 Rate Balance Across State-space Levels

Consider the time interval $[0, \tau_n], n \geq 2$ and a fixed level $x \in \boldsymbol{S}$. Let $\mathcal{D}_n(x)$, $\mathcal{U}_n(x)$ denote respectively the number of embedded down- and upcrossings of level x during $[0, \tau_n]$. Assume that the set of sample paths (sample functions) having an infinite number of embedded time points, has measure 1. The principle of rate balance across level x is

$$\begin{aligned} \lim_{n\to\infty} \frac{\mathcal{D}_n(x)}{n} &= \lim_{n\to\infty} \frac{\mathcal{U}_n(x)}{n} \ (a.s.), \\ \lim_{n\to\infty} \frac{E(\mathcal{D}_n(x))}{n} &= \lim_{n\to\infty} \frac{E(\mathcal{U}_n(x))}{n}. \end{aligned} \tag{8.1}$$

8.1.2 Method of Analysis

If the process is stable, the steady-state distribution of $W(t)$ as $t \to \infty$ and of W_n as $n \to \infty$, exist. Let $f(x) = \lim_{n\to\infty} f_n(x)$, $F(x) = \lim_{n\to\infty} F_n(x), x \in \boldsymbol{S}$. In the following sections, we shall derive an integral equation for $f(x)$ and $F(x)$ by using only:

1. the concept of embedded level crossings,
2. the principle of rate balance,
3. properties of the model,
4. knowledge of the efflux function $r(x), x \geq 0$.

8.2 GI/G/r(·) Dam

Assume that inputs to the dam occur in a renewal process with inter-input times having common cdf $A(\cdot)$. The model description is the same as for the M/G/r(·) dam in Subsection 6.2.1 except for the general renewal input stream.

The embedded process $\{W_n\}$ is a Markov chain, since

$$W_{n+1} = \max\{W_n + S_n - \Delta_n, 0\}$$

where S_n is the input amount at instant τ_n^- and Δ_n is the change in content during the time interval $[\tau_n, \tau_{n+1})$.

Define $\mathcal{G}(x)$ as the anti-derivative of $\frac{1}{r(x)}$ for $r(x) > 0$. Then $\mathcal{G}(x)$ is a continuous increasing function of x, since $\frac{d}{dx}\mathcal{G}(x) = \frac{1}{r(x)} > 0$. The time for the content to decline from state-space level v to level u, $v > u$, is

$$\int_u^v \frac{1}{r(x)} dx = \mathcal{G}(v) - \mathcal{G}(u).$$

A necessary and sufficient condition for the content of the dam to return to level 0 is: for every $v > 0$,

$$\begin{aligned}\lim_{u\downarrow 0} \int_{x=u}^{v} \frac{1}{r(x)} dx &= \lim_{u\downarrow 0}(\mathcal{G}(v) - \mathcal{G}(u)) \\ &= \mathcal{G}(v) - \lim_{u\downarrow 0} \mathcal{G}(u) < \infty. \qquad (8.2)\end{aligned}$$

For example, in a pharmacokinetic model (Section 10.8 below) with "first order" kinetics, $r(x) = kx, x > 0$. In theory the drug concentration never returns to level 0. In practice, the drug may be entirely removed from the body after some finite time.

8.2.1 Embedded Downcrossing Rate

Proposition 8.1 *The probability of an embedded downcrossing of level x occurring in $[\tau_n, \tau_{n+1}]$ is*

$$\begin{aligned} d_n(x) &= \int_{y=0}^{\infty}\int_{\alpha=x}^{\gamma(x,y)} B(\gamma(x,y)-\alpha)dF_n(\alpha)dA(y) \\ &= \int_{\alpha=x}^{\infty}\int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF_n(\alpha), n=1,2,\dots\ , \end{aligned} \quad (8.3)$$

where $\gamma(x,y) = \mathcal{G}^{-1}(\mathcal{G}(x)+y)$, and $\eta(\alpha,x) = \mathcal{G}(\alpha)-\mathcal{G}(x)$.

Proof. An embedded downcrossing occurs in $[\tau_n, \tau_{n+1}] \iff W_n > x$ and the time for $W(t)$ to descend from level $W_n + S_n$ to level x is $\leq (\tau_{n+1}-\tau_n) \iff$

$$\int_{z=x}^{W_n+S_n} \frac{1}{r(z)}dz = \mathcal{G}(W_n+S_n) - \mathcal{G}(x) \leq \tau_{n+1}-\tau_n. \quad (8.4)$$

Conditioning on $\tau_n - \tau_{n+1} = y$, (8.4) is equivalent to

$$\begin{aligned} \mathcal{G}(W_n+S_n)-\mathcal{G}(x) &\leq y, \\ \mathcal{G}(W_n+S_n) &\leq \mathcal{G}(x)+y. \end{aligned} \quad (8.5)$$

Note that $\mathcal{G}(\cdot)$ and its inverse $\mathcal{G}^{-1}(\cdot)$ are both continuous and increasing functions. Taking the inverse $\mathcal{G}^{-1}$ on both sides of (8.5) gives

$$S_n \leq \mathcal{G}^{-1}(\mathcal{G}(x)+y) - W_n = \gamma(x,y) - W_n.$$

Conditioning on $W_n = \alpha$, gives

$$\begin{aligned} &P(\text{embedded downcrossing in } [\tau_n,\tau_{n+1}] | \tau_n - \tau_{n+1} = y) \\ &\qquad = \int_{\alpha=x}^{\gamma(x,y)} B(\gamma(x,y)-\alpha)dF_n(\alpha). \end{aligned}$$

We obtain the unconditional probability of an embedded downcrossing of x during $[\tau_n, \tau_{n+1}]$ by integrating with respect to the inter-input time y having distribution $A(y)$. This yields $d_n(x)$ given in (8.3). ∎

Let

$$\delta_n(x) = \begin{cases} 1 \text{ if there is an embedded downcrossing of } x \text{ in } [\tau_n, \tau_{n+1}], \\ 0 \text{ if there is no embedded downcrossing of } x \text{ in } [\tau_n, \tau_{n+1}]. \end{cases}$$

Then $E(\delta_n(x)) = d_n(x)$. The number of embedded downcrossings of level x in $[0, \tau_{n+1}]$ is

$$\mathcal{D}_n(x) = \sum_{i=1}^{n} \delta_i(x).$$

Thus

$$E(\mathcal{D}_n(x)) = \sum_{i=1}^{n} d_i(x).$$

The long-run expected embedded downcrossing *rate* of level x is

$$\lim_{n\to\infty} \frac{E(D_n(x))}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(x).$$

From (8.3), since $\lim_{n\to\infty} F_n(x) \equiv F(x)$, then $\lim_{n\to\infty} d_n(x) = d(x)$ where

$$d(x) = \int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y) - \alpha) dA(y) dF(\alpha).$$

Also,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(x) = \lim_{n\to\infty} d_n(x) = d(x)$$

implies the expected embedded level downcrossing rate of level x is

$$\lim_{n\to\infty} \frac{E(\mathcal{D}_n(x))}{n} = \int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y) - \alpha) dA(y) dF(\alpha). \quad (8.6)$$

8.2.2 Embedded Upcrossing Rate

Proposition 8.2 *The probability of an embedded upcrossing of level x occurring in $[\tau_n, \tau_{n+1}]$ is*

$$\begin{aligned} u_n(x) &= \int_{y=0}^{\infty} \int_{\alpha=0}^{x} \overline{B}(\gamma(x,y) - \alpha) dF_n(\alpha) dA(y) \\ &= \int_{\alpha=0}^{x} \int_{y=0}^{\infty} \overline{B}(\gamma(x,y) - \alpha) dA(y) dF_n(\alpha), n = 1, 2, \ldots \,. \quad (8.7) \end{aligned}$$

Proof. An embedded upcrossing of level x occurs in $[\tau_n, \tau_{n+1}] \iff W_n \le x$, $W_n + S_n > x$, and the time for $W(t)$ to descend from level $W_n + S_n$ to level x exceeds $\tau_{n+1} - \tau_n$

$$\iff \int_{z=x}^{W_n+S_n} \tfrac{1}{r(z)} dz = \mathcal{G}(W_n + S_n) - \mathcal{G}(x) > \tau_{n+1} - \tau_n$$

$$\iff S_n > \mathcal{G}^{-1}(\mathcal{G}(x) + y) - W_n = \gamma(x,y) - W_n,$$

where we have conditioned on $\tau_n - \tau_{n+1} = y$. Therefore

$$P(\text{embedded upcrossing in } [\tau_n,\tau_{n+1}] | \tau_n - \tau_{n+1} = y) = \int_{\alpha=0}^{x} \overline{B}(\gamma(x,y)-\alpha)dF_n(\alpha),$$

where $\overline{B}(z) = 1 - B(z), z \geq 0$. The unconditional probability of an embedded upcrossing of x in $[\tau_n,\tau_{n+1}]$ is therefore given by (8.7). ■

As in the derivation of (8.4), it follows that the long-run expected embedded upcrossing rate of level x is

$$\lim_{n\to\infty} \frac{E(\mathcal{U}_n(x))}{n} = \int_{\alpha=0}^{x} \int_{y=0}^{\infty} \overline{B}(\gamma(x,y)-\alpha)dA(y)dF(\alpha). \tag{8.8}$$

8.2.3 Steady-state PDF of Content

We obtain an integral equation for the steady-state pdf of content. Applying rate balance (8.1) to formulas (8.6) and (8.8) gives an integral equation for $f(x)$ and $F(x)$, namely,

$$\int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha) - \int_{\alpha=0}^{x} \int_{y=0}^{\infty} \overline{B}(\gamma(x,y)-\alpha)dA(y)dF(\alpha) = 0, x \geq 0. \tag{8.9}$$

<u>**CDF Form of Integral Equation**</u>

In the second term of (8.9) write $\overline{B}(\cdot) = 1 - B(\cdot)$ and apply $F(x) = \int_{\alpha=0}^{x} dF(\alpha)$. This yields a ***cdf form*** with $F(x)$ on the left side explicitly,

$$F(x) = \int_{\alpha=0}^{x} \int_{y=0}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha) + \int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} B(\gamma(x,y)-\alpha)dA(y)dF(\alpha), x \geq 0. \tag{8.10}$$

<u>**PDF Form of Integral Equation**</u>

Differentiation of (8.10) with respect to $x > 0$, gives a **pdf form** with $f(x)$ explicitly on the left side,

$$f(x) = \int_{\alpha=0}^{x} \int_{y=0}^{\infty} \varrho(x,y) \cdot b(\gamma(x,y)-\alpha)dA(y)dF(\alpha) + \int_{\alpha=x}^{\infty} \int_{y=\eta(\alpha,x)}^{\infty} \varrho(x,y) \cdot b(\gamma(x,y)-\alpha)dA(y)dF(\alpha), x > 0, \tag{8.11}$$

where $\varrho(x,y) = \frac{\partial}{\partial x}\gamma(x,y) = \frac{r(\gamma(x,y))}{r(x)}$.

Probability of Zero Content

Letting $x \downarrow 0$ in (8.10) gives

$$F(0) = \frac{\int_{\alpha=0^+}^{\infty} \int_{y=\eta(\alpha,0)}^{\infty} B(\gamma(0,y)-\alpha)dA(y)dF(\alpha)}{\int_{y=0}^{\infty} \overline{B}(\gamma(0,y))dA(y)}. \tag{8.12}$$

The normalizing condition is

$$F(0) + \int_{\alpha=0}^{\infty} f(\alpha)d\alpha = 1 \tag{8.13}$$

If condition (8.2) does not hold, then $F(0) = 0$ (recall that $f(0) \equiv f(0^+)$).

Solution Method

The solution method in the following sections will be to obtain the functional form of $f(x)$ and $F(x)$ using (8.10) or (8.11), and applying the boundary conditions (8.12) and (8.13) to specify $f(x)$, $F(x), x \geq 0$.

8.2.4 M/G/r(·) Dam

In this model, $A(y) = 1 - e^{-\lambda y}, y \geq 0$. Note that

$$\frac{\partial\left(\gamma(x,y)\right)}{\partial y} = \frac{\partial(\mathcal{G}^{-1}(\mathcal{G}(x)+y))}{\partial y} = r(\gamma(x,y)) = r(\mathcal{G}^{-1}(\mathcal{G}(x)+y)).$$

Integrating (8.11) by parts, using parts

$$\frac{\lambda e^{-\lambda y}}{r(y)} \text{ and } r(\gamma(x,y)) \cdot b(\gamma(x,y)-\alpha)dy,$$

simplifying and substituting from (8.10) results in

$$r(x)f(x) = \lambda \int_{\alpha=0}^{x} \overline{B}(x-\alpha)dF(\alpha), x > 0. \tag{8.14}$$

Equation (8.14) is identical to the integral equation (6.18) for the steady-state pdf of content in the M/G/r(·) dam (derived using "continuous" LC)

Remark 8.1 *In equation (8.14) $f(x) = \lim_{n\to\infty} f_n(x)$ since (8.14) has been derived using* ***embedded*** *LC. In Chapter 6, equation (6.18), $f(x) = \lim_{t\to\infty} f_t(x)$ is the* ***time-average*** *steady-state pdf of content. The fact*

GI/G/r(·) Dam	Gi/G/1 Queue
Input instant τ_n^-	Customer arrival instant τ_n^-
Input amount at τ_n^-	Service time (jump size) S_n
Content at τ_n^-	Customer wait W_n in queue at τ_n^-
Content at instant τ_n	Virtual wait $W(\tau_n) = W_n + S_n$
Content at time $t \geq 0$	Virtual wait $W(t)$ at time $t \geq 0$
$r(x) > 0, x > 0$; $r(0) = 0$	$r(x) = 1, x > 0$; $r(0) = 0$
Distribution of content	Distribution of waiting time

Table 8.1: GI/G/r(.) dam versus GI/G/1queue.

that $\lim_{n\to\infty} f_n(x)$ *and* $\lim_{t\to\infty} f_t(x)$ *satisfy the same integral equation, demonstrates that the content of an M/G/r(·) dam satisfies the PASTA principle that Poisson arrivals "see" time averages [102]. Here we have derived PASTA for the M/G/r(·) dam by using continuous and embedded LC concepts only.*

8.3 GI/G/1 Queue

The GI/G/1 queue is closely related to the Gi/G/r(·) dam (Table 8.1). For the virtual wait of the GI/G/1 queue $r(x) = \begin{cases} 1, x > 0, \\ 0, x = 0. \end{cases}$

The anti-derivative of $\frac{1}{r(x)}, x > 0$, is

$$\mathcal{G}(x) = \int \frac{1}{r(x)} dx = \int 1 \cdot dx = x.$$

Thus,

$$\begin{aligned} \gamma(x,y) &= \mathcal{G}^{-1}(\mathcal{G}(x) + y)) = \mathcal{G}^{-1}(x+y)) = x+y \\ \eta(\alpha,x) &= \mathcal{G}(\alpha) - \mathcal{G}(x) = \alpha - x, \\ \varrho(x,y) &= \frac{r(\gamma(x,y))}{r(x)} = \frac{r(x+y)}{1} = \frac{1}{1} = 1. \end{aligned}$$

For the GI/G/1 queue, equations (8.10), (8.11) and (8.13) reduce respectively to

$$\begin{aligned} F(x) = &\textstyle\int_{\alpha=0}^{x} \int_{y=0}^{\infty} B(x+y-\alpha) dA(y) dF(\alpha) \\ &+ \textstyle\int_{\alpha=x}^{\infty} \int_{y=\alpha-x}^{\infty} B(x+y-\alpha) dA(y) dF(\alpha), x \geq 0, \end{aligned} \tag{8.15}$$

$$\begin{aligned} f(x) = &\textstyle\int_{\alpha=0}^{x} \int_{y=0}^{\infty} b(x+y-\alpha) dA(y) dF(\alpha) \\ &+ \textstyle\int_{\alpha=x}^{\infty} \int_{y=\alpha-x}^{\infty} b(x+y-\alpha) dA(y) dF(\alpha), x > 0, \end{aligned} \tag{8.16}$$

$$F(0) = \frac{\int_{\alpha=0^+}^{\infty} \int_{y=\alpha}^{\infty} B(y-\alpha)dA(y)dF(\alpha)}{\int_{y=0}^{\infty} \overline{B}(y))dA(y)}. \tag{8.17}$$

The normalizing condition is

$$F(0) + \int_{\alpha=0}^{\infty} f(\alpha)d\alpha = 1. \tag{8.18}$$

Applications

Some single-server queueing models can be solved using embedded LC, by applying equations (8.15) - (8.18). Other models are solved by deriving integral equations for the pdf of the state variables from first principles using embedded LC. The next four subsections illustrate some applications.

8.3.1 M/G/1 Queue

The M/G/1 queue is a special case of the M/G/r(·) dam, with $r(x) = 1, x > 0$ and $A(y) = 1 - e^{-\lambda y}, y \geq 0$. Substituting directly into equation (8.14) or into (8.16) followed by some algebra yields

$$\begin{aligned} f(x) &= \lambda \int_{\alpha=0}^{x} \overline{B}(x-\alpha)dF(\alpha) \\ &= \lambda P_0 \overline{B}(x) + \lambda \int_{\alpha=0}^{x} \overline{B}(x-\alpha)f(\alpha)d\alpha, x > 0, \end{aligned} \tag{8.19}$$

which is identical to equation (3.29). Remark 8.1 applies also to this model.

8.3.2 GI/M/1 Queue

The GI/M/1 queue is a special case of the GI/G/1 queue with

$$B(x) = 1 - e^{-\mu x}, x \geq 0, \quad b(x) = \mu e^{-\mu x} = \mu - \mu B(x), x > 0.$$

Substituting $b(x) = \mu - \mu B(x)$ into (8.16), simplifying and combining with (8.15) gives the integral equation

$$f(x) = \mu \int_{y=x}^{\infty} \overline{A}(y-x)f(y)dy, \ x > 0, \tag{8.20}$$

which is identical to equation (5.6).

Equation (8.19) for M/G/1	Equation (8.20) for G/M/1
λ	μ
x is upper bound of integral	x is lower bound of integral
$\overline{B}(x-y)$	$\overline{A}(y-x)$
P_0 appears explicitly	P_0 does not appear explicitly

Table 8.2: Interchanged roles of terms in integral equations for M/G/1 and G/M/1.

Duality of M/G/1 and GI/M/1 Queues

Upon comparing integral equations (8.19) and (8.20) it is evident that they are duals, in the sense that the roles of certain terms are interchanged (see Table 8.2). The significance of this "duality" is that we analyze the M/G/1 queue via LC using the virtual wait process. On the other hand, we are led to analyzing the G/M/1 queue via LC using the extended "age" process (see Subsection 5.1.1 and [11]).

Remark 8.1 applies also to GI/M/1, provided we analyze the extended age process, for which departures from the system occur in a Poisson process at rate μ conditional on the server being occupied. This implies that in (8.20), $f(x)$ on the left side (equal to time-average pdf of virtual wait) is the same function as $f(y)$ in the integrand on the right side (pdf of system time at departure instants).

Solution for Steady-state PDF of Wait in GI/M/1

Assume the solution for the pdf of wait has the form $f(x) = Ke^{-\gamma x}, x > 0$. Substituting into (8.20) yields the equation for γ

$$\int_{z=0}^{\infty} \overline{A}(z)e^{-\gamma z}dz = \frac{1}{\mu},$$

or

$$\frac{1}{\gamma} - \frac{1}{\gamma}A^*(\gamma) = \frac{1}{\mu}. \tag{8.21}$$

In (8.21) $A^*(\cdot)$ is the Laplace-Stieltjes transform of $A(\cdot)$ defined by

$$A^*(s) = \int_{y=0}^{\infty} e^{-sy}a(y)dy, s \geq 0,$$

and $a(y) = \frac{d}{dy}A(y)$, assuming the inter-arrival times are continuous r.v.'s. We obtain an expression for $P_0 = F(0)$ upon substituting $B(y) = 1 -$

$e^{-\mu y}$, $f(\alpha) = Ke^{-\gamma\alpha}$ in (8.17), namely

$$F(0) = [A^*(\mu)]^{-1}\left\lfloor \frac{\gamma - \mu + \mu A^*(\gamma) - \gamma A^*(\mu)}{\gamma(\gamma - \mu)} \right\rfloor \cdot K. \tag{8.22}$$

From (8.21)

$$\mu - \mu A^*(\gamma) = \gamma,$$

which substituted into (8.22) leads directly to

$$F(0) = \frac{K}{\mu - \gamma}. \tag{8.23}$$

The normalizing condition (8.18) gives

$$\frac{K}{\mu - \gamma} + \frac{K}{\gamma} = 1.$$

Then (8.23) implies

$$F(0) = \frac{\gamma}{\mu}. \tag{8.24}$$

Formula (8.24) is important because $F(0) = P_{0\iota}$ in (5.23) which was derived using "continuous" or "time-average" LC. This provides further evidence of the overall logical correctness of the LC methodology.

Check with M/M/1 Queue

It is instructive to check the result for the M/M/1 queue. Consider M/M/1 with arrival rate λ and service rate μ. Then $A^*(s) = \frac{\lambda}{\lambda+s}$. From (8.21) $\gamma = \mu - \lambda$, which substituted into (8.22), gives $F(0) = P_0 = \frac{K}{\lambda}$. Applying the normalizing condition $F(0)+\int_{y=0}^{\infty} f(y)dy = 1$, gives

$$\frac{K}{\lambda} + K\int_{y=0}^{\infty} e^{-(\mu-\lambda)y}dy = 1,$$

$$K = \lambda(1 - \frac{\lambda}{\mu}).$$

Thus

$$P_0 = \frac{K}{\lambda} = 1 - \frac{\lambda}{\mu}, \checkmark$$

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

which checks with the M/M/1 solution given in (3.86) and (3.87).

8.3.3 $E_k/M/1$ Queue

Assume the common pdf of the inter-arrival times $a(\cdot) \underset{dist}{=}$ Erlang-(k, λ). Thus for integer $k = 1, 2, \ldots$, $a(y) = e^{-\lambda y}\frac{(\lambda y)^{k-1}}{(k-1)!}\lambda, y > 0$. Let $A(\cdot)$ denote the cdf corresponding to $a(\cdot)$. Then the LST of $A(\cdot)$ is $A^*(\gamma) = \left(\frac{\lambda}{\lambda+\gamma}\right)^k$, which substituted into equation (8.21) gives an equation for γ,

$$\frac{1}{\gamma} - \frac{1}{\gamma}\left(\frac{\lambda}{\lambda+\gamma}\right)^k = \frac{1}{\mu}, k = 1, 2, \ldots . \tag{8.25}$$

We seek a unique positive solution of (8.25) for γ. Assume that $\lambda, \mu > 0$ and $\lambda < k\mu$ (stability condition for G/M/1 is $\boldsymbol{a} < \mu$, where $\boldsymbol{a} = \frac{k}{\lambda} =$ arrival rate). Then equation (8.25) has exactly one *real* positive root for γ (see [11]). If k is odd, all other roots are *complex*. If k is even, one other root is negative real and all other roots are complex. Thus the solution for γ is unique. Denote it by γ_k.

To solve for $K \equiv \eta_k$ we first substitute γ_k into (8.22) and use (8.25) to obtain

$$F(0) = \frac{\eta_k}{\mu - \gamma_k}.$$

(We use η_k instead of K_k in this subsection only, for notational contrast.) Then apply the normalizing condition (8.18) to obtain

$$\eta_k = \frac{\gamma_k(\mu - \gamma_k)}{\mu} = \gamma_k\left(1 - \frac{\gamma_k}{\mu}\right).$$

The steady-state pdf of wait is then given by

$$\begin{aligned} P_0 &= \frac{\eta_k}{\mu - \gamma_k} = \frac{\gamma_k}{\mu}, \\ f(x) &= \eta_k e^{-\gamma_k x} = \gamma_k(1 - \frac{\gamma_k}{\mu})e^{-\gamma_k x}, x > 0. \end{aligned}$$

Remark 8.2 *The solution of equation (8.25) can be readily obtained numerically for any specified values of* λ, μ, k.

8.3.4 D/M/1 Queue

Assume the common inter-arrival time is $D > 0$. Then $A^*(s) = e^{-sD}, s > 0$. Let the steady-state pdf of wait be $f(x) = Ke^{-\gamma x}, x > 0$. Substituting $A^*(\gamma) = e^{-\gamma D}$ into (8.21) gives the equation

$$\mu e^{-\gamma D} + \gamma - \mu = 0$$

for γ, whose solution we call $\gamma_{_D}$. From (8.22)

$$F(0) = \frac{K}{\mu - \gamma_{_D}}.$$

Let $K \equiv K_{_D}$. Substituting into (8.18) gives

$$\frac{K_{_D}}{\mu - \gamma_{_D}} + \frac{K_{_D}}{\gamma_{_D}} = 1,$$
$$K_{_D} = \gamma_{_D}\left(1 - \frac{\gamma_{_D}}{\mu}\right).$$

The steady-state pdf of wait is

$$\begin{aligned} P_0 &= \frac{K_{_D}}{\mu - \gamma_{_D}}, \\ f(x) &= K_{_D} e^{-\gamma_{_D} \cdot x}, x > 0. \end{aligned}$$

8.4 M/G/1 with Reneging

We apply the embedded LC method to an M/G/1 queue in which customers can either: (1) renege from the waiting line; (2) wait and balk at service; (3) wait and stay for a full service. Assume the *staying function* is $\overline{R}(y) = P(\text{arrival stays for service}|\text{required wait} = y)$. We verify that the pdf $f(\cdot)$ on the left side of (3.162) and the pdf $f(\cdot)$ on the right side of (3.162) are the same functions. In (3.162) the pdf on the left side is $\lim_{t\to\infty} f_t(x)$ (time-average pdf). The pdf on the right side is $\lim_{n\to\infty} f_n(x)$ (pdf at arrival instants, or arrival-point pdf). We now use embedded LC to derive an integral equation for $f(x) = \lim_{n\to\infty} f_n(x)$ and show that it is identical to equation (3.162).

8.4.1 Embedded Crossing Probabilities

The limiting probability of an SP *embedded upcrossing* of level x is

$$u = \int_{y=0^-}^{x} \int_{z=0}^{\infty} \overline{B}(x - y + z)\overline{R}(y) f(y) \lambda e^{-\lambda z} dz dy, \tag{8.26}$$

where the lower limit $y = 0^-$ means that the term $\overline{B}(x + z)P_0$ for the atom $\{0\}$ is included in the evaluation of u. The right side of (8.26) holds since an embedded upcrossing of x occurs iff $0 \le W_n = y < x$, the arrival at τ_n stays for service (probability $\overline{R}(y)$), and given that the time to the next arrival is z, the service time exceeds $x - y + z$.

The limiting probability of an SP *embedded downcrossing* of level x consists of two terms,

$$d = \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy + \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} R(y)f(y)\lambda e^{-\lambda z}dzdy. \tag{8.27}$$

The first term on the right of (8.27) is similar to (8.26), except that an SP jump starts at a level $y > x$ and the service time must be less than $x - y + z$ for an embedded downcrossing to occur. The second term is due to arrivals that *do not stay for service* (renege or balk at service); arrivals renege or balk at service with probability $R(y) = 1 - \overline{R}(y)$. We can assume that an SP "jump" is of size 0 (probability $R(y)$) when a reneger or service-balker arrives. Equivalently there is *no SP jump* when a reneger or service-balker arrives. In this case the SP makes an embedded downcrossing of level x provided the next inter-arrival time $z > y - x$. The second term in (8.27) simplifies to $\int_{y=x}^{\infty} R(y)f(y)e^{-\lambda(y-x)}dy$.

Since $\overline{B}(\cdot) \equiv 1 - B(\cdot)$, equation (8.26) can be written as

$$u = \int_{y=0^-}^{x} \overline{R}(y)f(y)dy - \int_{y=0^-}^{x} \int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy \tag{8.28}$$

8.4.2 Steady-State PDF of Wait of Stayers

Applying *embedded* rate balance across level x, we set $u = d$. This yields from equations (8.27) and (8.28), the integral equation

$$\int_{y=0^-}^{x} \overline{R}(y)f(y)dy = \int_{y=0^-}^{x} \int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy + \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z}dzdy + \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} R(y)f(y)\lambda e^{-\lambda z}dzdy. \tag{8.29}$$

We take $\frac{d}{dx}$ on both sides of (8.29). This involves differentiation under the integral sign. Some algebra including cancellation of terms and using

$R(y) + \overline{R}(y) = 1$ gives

$$\begin{aligned} f(x) = &\int_{y=0^-}^{x} \int_{z=0}^{\infty} b(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z} dz dy \\ &+ \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} b(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z} dz dy \\ &+ \lambda \int_{y=x}^{\infty} \int_{z=y-x}^{\infty} R(y)f(y)\lambda e^{-\lambda z} dz dy. \end{aligned} \tag{8.30}$$

Integrating each of the inner integrals

$$\int_{z=0}^{\infty} b(x-y+z)\lambda e^{-\lambda z} dz \quad \text{and} \quad \int_{z=y-x}^{\infty} b(x-y+z)\lambda e^{-\lambda z} dz$$

in (8.30) by parts, using parts $\lambda e^{-\lambda z}$ and $b(x-y+z)$, leads to the integral equation (assuming $B(0) = 0$)

$$\begin{aligned} f(x) = \quad & -\lambda \textstyle\int_{y=0^-}^{x} \overline{R}(y)f(y)B(x-y)dy \\ & +\lambda \textstyle\int_{y=0^-}^{x} \int_{z=0}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z} dz dy \\ & +\lambda \textstyle\int_{y=x}^{\infty} \int_{z=y-x}^{\infty} B(x-y+z)\overline{R}(y)f(y)\lambda e^{-\lambda z} dz dy \\ & +\lambda \textstyle\int_{y=x}^{\infty} \int_{z=y-x}^{\infty} R(y)f(y)\lambda e^{-\lambda z} dz dy. \end{aligned} \tag{8.31}$$

From (8.29) the sum of the last three terms on the right of (8.31) is

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

Hence

$$\begin{aligned} f(x) &= \lambda \int_{y=0^-}^{x} \overline{R}(y)f(y)dy - \lambda \int_{y=0^-}^{x} \overline{R}(y)f(y)B(x-y)dy, \\ f(x) &= \lambda \int_{y=0^-}^{x} \overline{B}(x-y)\overline{R}(y)f(y)dy. \end{aligned} \tag{8.32}$$

Equation (8.32) is *identical to* (3.162). Hence, in (3.162), the time-average pdf of stayers (left side) is equal to the arrival-point pdf of stayers (in integral on right side). The derivation of (3.162) using "continuous-time" LC is far simpler than that of (8.32). Nevertheless, the embedded LC method is very useful in this case, and elsewhere. It helps to confirm that "continuous" LC works in the reneging problem. The embedded LC method can often be applied to determine whether the time-average and arrival-point pdf's are equal. The embedded LC method is inherently very intuitive, and has additional applications as well.

CHAPTER 9

LEVEL CROSSING ESTIMATION

9.1 Introduction

This chapter describes a basic level crossing estimation method (LCE) for steady-state probability distributions in queues, storage processes and related stochastic models. LCE is also called: level crossing computation, system point estimation (or computation). LCE is related to non-parametric density estimation methods (e.g., [95]). In standard density estimation the data is assumed to be a random sample from an unknown pdf. The data is used to construct histograms, naive density estimators, kernel-density estimators, etc., for the unknown pdf, utilizing associated smoothing techniques.

In LCE we obtain the data from a simulated sample path of a stochastic process. We compute estimators of the pdf of the state variable from level-crossing time averages, or related averages. The estimators used in LCE can be combined with smoothing techniques to improve the estimates (e.g., [71], [72], [73]).

9.1.1 Main Steps of Level Crossing Estimation

The basic LCE procedure that we use here for steady-state distributions, has three main steps:

1. Simulate a single sample path of the process over a long simulated time period, say $[0, t]$.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_9,

2. From the simulated sample path, compute *point estimators* of the pdf and cdf of the state variable, in terms of level-crossing time averages calculated on a state-space partition. Compute point estimators of moments and of expected values of measurable functions of the state variable.

3. Obtain confidence limits for the estimates of the pdf, cdf, moments and expected values of measurable functions.

Remark 9.1 *Step 2 may also include a sensitivity analysis of the estimates. Thus, we may vary the simulated total time t, and/or the state-space partition norm size (fixed bin size, defined below in Subsection 9.4.1), to ensure that estimates remain within preassigned tolerances.*

In addition to the three main steps, we also characterize the steady-state pdf and cdf according to continuity, boundedness, convexity, differentiability, etc., by utilizing sample-path properties for the model. For example, in $M_\lambda/G/1$ and in $G/M_\mu/1$ queues, the steady-state pdf's of wait are bounded by λ and μ respectively (Propositions 3.5, 5.9).

I have carried out numerous LCE computational experiments using the procedure described herein, as well as other LCE procedures (e.g., [13], [21], [22], [32]). These experiments have detected all pdf discontinuities and intervals of convexity or concavity in benchmark models, where the pdf properties are known. For example, an M/Discrete/1 queue may serve as a benchmark. Proposition 3.9 specifies continuity/discontinuity properties of the pdf of wait. We can also apply LCE to estimate the pdf of wait in *variants* of M/Discrete/1 with state dependencies, etc., in which analytical results are tedious to obtain, or are not available.

9.2 Theoretical Basis for LC Estimation

LCE is based on level crossing theorems. Consider M/G/1. Theorem 1.1 implies that virtual-wait sample-path level-crossing time averages converge to the steady-state pdf of wait (*a.s.*) as time $t \to \infty$ (Subsection 9.2.2 below). This implies that time averages computed from a simulated sample path over a long simulated time t, should approximate the pdf accurately for all state-space values up to the maximum state-space level attained during $[0, t]$, say χ_t. Thus, the state-space interval $[0, \chi_t]$ will contain an increasing measure of the total probability as t increases (Subsection 9.2.4). The measure will grow to 1 as $t \to \infty$.

Remark 9.2 *The LCE method described here is one of several LC estimation methods. I have developed a version of LCE based on Theorems 3.2, 3.3 and related theorems, for estimating **transient distributions of state variables** (e.g., Remark 3.6). (I have discussed this technique at several conferences, e.g., P. H. Brill (1982), "System Point Monte Carlo Simulation of Time Dependent Probability Distributions of Waiting Times in Queues", TIMS/ORSA National Meeting, Chicago, April.)*

9.2.1 Boundedness of Steady-state PDF

A bound on the steady-state pdf of the virtual wait in $M_\lambda/G/1$ queues is given in Proposition 3.5, and on the steady-state arrival-point pdf of wait in $G/M_\mu/1$ queues in Proposition 5.9. In $M_\lambda/G/1$, $f(x) < \lambda, x > 0$. In $G/M_\mu/1$, $f_\iota(x) < \mu, x > 0$. Recall that $\mathcal{D}_t(x)$, $\mathcal{U}_t(x)$ are the numbers of SP down- and upcrossings of level x during $(0,t]$ respectively. Boundedness implies that for a *typical* sample path in $M_\lambda/G/1$,

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

In $G/M_\mu/1$,

$$f_\iota(x) = \lim_{t\to\infty} \frac{\mathcal{U}_t(x)}{t} < \mu, x > 0.$$

Similarly, we can develop bounds on $f(x)$ for other models, e.g., for M/M/c, G/M/c, etc. In $M_\lambda/G/r(\cdot)$ dams, boundedness follows from integral equation (6.18) for the steady-state pdf of content $f(x)$. If the efflux rate satisfies $r(x) > m > 0$, $x > 0$, then $f(x) < \frac{\lambda}{m}, x > 0$.

9.2.2 Role of Level Crossing Theorems in LCE

Consider $M_\lambda/G/1$. A sample path of the virtual wait is diagrammed in figures 3.4 and 9.1. Let $F(x)$, $f(x)$ be the steady-state cdf and pdf of wait respectively. Let $P_0 = F(0)$. Theorem 1.1 asserts

$$\lim_{t\to\infty} \frac{\mathcal{D}_t(x)}{t} = f(x), x \geq 0, \ \lim_{t\to\infty} \frac{\mathcal{D}_t(0)}{t} = f(0) = \lambda P_0 \ (a.s.)$$

(recall that $f(0) \equiv f(0^+)$). Hence, given $\varepsilon > 0$, for each $x > 0$ $\exists$ $t_{x\varepsilon}$ such that

$$t > t_{x\varepsilon} \implies \left|\frac{\mathcal{D}_t(x)}{t} - f(x)\right| < \varepsilon f(x) \ (a.s.), \tag{9.1}$$

since $f(x)$ is bounded, i.e., $0 < f(x) < \lambda < \infty, x \geq 0$ (Subsection 9.2.1). Also $\exists\ t_{0\varepsilon}$ such that

$$t > t_{0\varepsilon} \implies \left| \frac{\mathcal{D}_t(0)}{\lambda t} - P_0 \right| < \varepsilon P_0. \tag{9.2}$$

Choose an arbitrary "small" $\delta, 0 < \delta << 1$. Let W_q denote the steady-state queue wait. Define $z_\delta > 0$ by $P(W_q > z_\delta) = \delta$. Then δ is the probability of the right tail of the distribution of W_q, i.e., on the interval (z_δ, ∞). Thus

$$1 - F(z_\delta) \equiv \int_{y=z_\delta}^{\infty} f(y)dy = \delta. \tag{9.3}$$

Suppose we could determine (finite) $t_\delta^* = \max_x\{t_{x\varepsilon}|x \in [0, z_\delta)\}$, where $t_{x\varepsilon}, x > 0$ is defined in (9.1) and $t_{0\varepsilon}$ is defined in (9.2). Then

$$\begin{aligned} &t > t_\delta^* \implies \left|\tfrac{\mathcal{D}_t(x)}{t} - f(x)\right| < \varepsilon f(x) \text{ for all } x \in (0, z_\delta)\ (a.s.), \\ &\left|\tfrac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| < \varepsilon P_0\ (a.s.). \end{aligned} \tag{9.4}$$

By the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$, we have

$$\begin{aligned} P_0 + \int_{x=0}^{z_\delta} f(x)dx &= 1 - \int_{x=z_\delta}^{\infty} f(x)dx \\ &= 1 - \delta > 0. \end{aligned} \tag{9.5}$$

Summing over all $x \in [0, \infty)$ in (9.4) and using (9.5), yields

$$\begin{aligned} t > t_\delta^* \implies &\left|\tfrac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| + \int_{x=0}^{z_\delta} \left|\tfrac{\mathcal{D}_t(x)}{t} - f(x)\right| dx < \varepsilon P_0 + \varepsilon \int_{x=0}^{z_\delta} f(x)dx \\ &= \varepsilon(1-\delta) < \varepsilon\ (a.s.). \end{aligned} \tag{9.6}$$

Let $\left\{\widehat{P}_0; \widehat{f}(x)\right\}$ denote the estimate of $\{P_0; f(x)\}$. We assume that a sample path over a fixed simulated time interval $[0, t]$ is used to compute $\left\{\widehat{P}_0; \widehat{f}(x)\right\}$. (We omit subscript "$t$" in the symbols $\widehat{P}_0$ and $\widehat{f}(x)$, in order to distinguish $\widehat{P}_0$, $\widehat{f}(x)$ from estimators "$\widehat{P}_{0t}$, $\widehat{f}_t(x)$" for the *transient* pdf of wait, which we use outside this monograph.)

Assume we use the "natural" estimator based on the sample path, viz., $\widehat{P}_0 = \frac{\mathcal{D}_t(0)}{\lambda t}$, $\widehat{f}(x) = \frac{\mathcal{D}_t(x)}{t}, t > t_\delta^*$. Then (9.6) implies that the *total absolute error* of $\left\{\widehat{P}_0; \widehat{f}(x)\right\}$ in estimating $\{P_0; f(x), x \in (0, z_\delta)\}$ is less than ε.

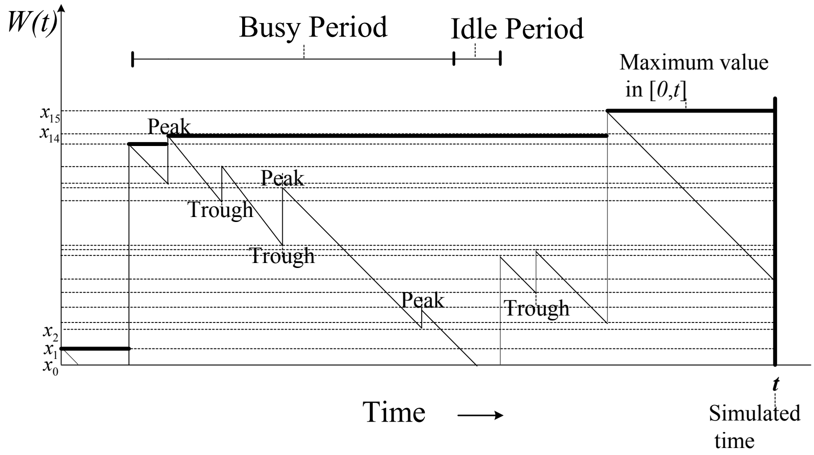

Figure 9.1: Sample path of virtual wait $\{W(t)\}$ in M/G/1. Shows peaks $\{W_n + S_n\}$, troughs $\{W_n\}$ and state-space partition $0 = x_0 < x_1 < x_2 \cdots < x_{15}$ in time interval $(0, t)$. Also shows maximum sample-path value attained in $[0, t]$.

Assume $\widehat{f}(x) = 0, x > z_\delta$. Then $t > t^*_\delta$ implies that the total absolute error in $\widehat{f}(x), x > z_\delta$ is equal to δ, i.e.,

$$t > t^*_\delta \implies \int_{x=z_\delta}^{\infty} \left|\widehat{f}(x) - f(x)\right| dx = \int_{x=z_\delta}^{\infty} f(x)dx = \delta, (a.s.). \qquad (9.7)$$

Suppose we could simulate a sample path over a sufficiently large time interval $(0, t), t > t^*_\delta$. Statements (9.6) and (9.7) imply that the total absolute error would be

$$\begin{aligned} &\left|\widehat{P}_0 - P_0\right| + \int_{x=0}^{\infty} \left|\widehat{f}(x) - f(x)\right| dx \\ &\quad = \left|\frac{\mathcal{D}_t(0)}{\lambda t} - P_0\right| + \int_{x=0}^{\infty} \left|\frac{\mathcal{D}_t(x)}{t} - f(x)\right| dx < \varepsilon + \delta, (a.s.). \end{aligned} \qquad (9.8)$$

In principle we can choose ε and δ arbitrarily small. Then we can simulate a sample path over a long simulated time $t > t^*_\delta$ and ensure that the total absolute error of $\left\{\widehat{P}_0; \widehat{f}(x)\right\}$ in estimating $\{P_0; f(x), x > 0\}$ is arbitrarily small. This procedure would amount to *computation* of the entire pdf $\{P_0; f(x), x > 0\}$ within a preassigned tolerance. The total error on $[0, z_\delta)$ is less than ε. The total error on (z_δ, ∞) is equal to δ.

9.2.3 Natural Partition of State Space

We illustrate a natural partition of the state space by means of an example.

Example 9.1 *Consider a* ***sample path of the virtual wait*** $\{W(t)\}$ ***in an M/G/1 queue*** *(Fig. 9.1). The state space is* $\boldsymbol{S} = [0, \infty)$*. For* ***fixed*** $x \in \boldsymbol{S}$, $\{\mathcal{D}_t(x)\}$ *is a counting process. For* ***fixed*** $t > 0$, $\mathcal{D}_t(x)$ *is a step function on* $\boldsymbol{S}$*. The jumps in the step function occur at the peaks* $\{W_n + S_n\}$ *and troughs* $\{W_n\}$*, where* W_n, $S_n, n = 1, 2, ...$ *are the customer waits and service times respectively. In Fig. 9.1 level* $W(0)$ *is a peak and level* $W(t)$ *is a trough. We* ***merge the peaks and troughs*** *to form a state-space partition*

$$\{x_i\} = W(0) \cup \{W_n\} \cup \{W_n + S_n\} \cup W(t),$$

arranged ***in ascending order of magnitude*** *in* $\boldsymbol{S}$*,*

$$0 = x_0 < x_1 < \cdots < x_{M(t)} < \infty.$$

The first partition point x_0 *corresponds to all troughs of* $W(0) \cup \{W_n\} \cup W(t)$ *such that the ordinate is* 0*. The second partition point is*

$$x_1 = \min_n \left\{ \begin{array}{c} W(0) \cup \{W_n\} \cup \{W_n + S_n\} \cup W(t) \\ \setminus \ \{troughs\ = 0\}\,. \end{array} \right\}$$

That is, $\min_n \{\cdot\}$ *excludes the troughs corresponding to* x_0 $(= 0)$*. The* j^{th} *partition point* x_j *is obtained similarly, excluding those troughs and/or peaks corresponding to* $\{x_0, x_1, ..., x_{j-1}\}$*. The number of subintervals of partition* $\{x_i\}$ *is* $M(t) \leq 2N_a(t)$*, where* $N_a(t)$ *is the number of arrivals during* $(0, t)$*. In Fig. 9.1,* $N_a(t) = 8$*,* $M(t) = 15$*.*

Note that t is fixed. Let

$$\mathcal{D}_t(x) = d_i, x \in [x_i, x_{i+1}), i = 0, 1, ..., M(t),$$

where $d_i \geq 0$ *is a constant. Then*

$$\frac{\mathcal{D}_t(x)}{t} = \frac{d_i}{t}, x \in [x_i, x_{i+1}), i = 0, 1, ..., M(t)$$

is a step function of $x \in \boldsymbol{S}$*. Suppose we can determine* t^*_δ *as in (9.4). Then, from (9.8)*

$$t > t^*_\delta \implies \left| \frac{d_0}{\lambda t} - P_0 \right| + \sum_{i=0}^{M(t)} \int_{x=x_i}^{x_{i+1}} \left| \frac{d_i}{t} - f(x) \right| dx < \varepsilon + \delta, (a.s.). \quad (9.9)$$

In Fig. 9.1

$$d_0 = 2, d_1 = 1, d_2 = 2, d_3 = 3, ..., d_{14} = 1, d_{15} = 0.$$

The recursion (9.11) below may simplify computation of $\{d_i\}$ using a computer program.

$$d_{i+1} = \begin{cases} d_i + 1 \text{ if } x_{i+1} \text{ is a trough,} \\ d_i - 1 \text{ if } x_{i+1} \text{ is a peak, } i = 0, ..., M(t) - 1, \end{cases} \quad (9.10)$$

$$d_{M(t)+1} = 0. \quad (9.11)$$

The sub-interval lengths of the partition $\{x_i\}$ are

$$\{x_{i+1} - x_i\}, i = 0, ..., M(t).$$

These lengths vary in a natural way (variable bin sizes).

9.2.4 Ladder Points and LCE Estimates

For the virtual wait, let χ_t denote the maximum sample-path level in $\boldsymbol{S}$ attained during $[0, t]$. For fixed t, $\chi_t = x_{M(t)}$, the greatest finite point of partition $\{x_i\}$. As t increases $\{\chi_t, t \geq 0\}$ is a non-decreasing step function with non-homogeneous inter-jump times. A sample path of $\{\chi_t\}$ is a non-decreasing right-continuous step function with upward jumps at embedded arrival instants $\tau_{ln}, n = 1, 2, ...$. The associated service-time jumps end strictly above $\chi_{\tau_{l(n-1)}} = \chi_{\tau_{ln}^-}$ (Fig. 9.1). Thus $\frac{d}{dt}\chi_t = 0, \tau_{l(n-1)} < t < \tau_{ln}, n = 0, 1, 2, ...,$ where $\tau_{l0} \equiv 0$. The increase in $\{\chi_t\}$ at arrival instant τ_{ln} is equal $\chi_{\tau_{ln}} - \chi_{\tau_{ln}^-}$ = excess service time above level $\chi_{\tau_{ln}^-}$. Random variables $\chi_{\tau_{ln}}, n = 1, 2, ...$ are ordinates of the "strict ascending ladder points" $\left\{\left(\tau_{ln}, \chi_{\tau_{ln}}\right)\right\}$ of the virtual wait process $\{W(s), s \geq 0\}$. The points $\left(\tau_{ln}, \chi_{\tau_{ln}}\right) \in \boldsymbol{T} \times \boldsymbol{S}, n = 1, 2, ...,$ are analogous to strict ascending ladder points for a random walk [56]. The LCE estimate of the pdf of wait $f(x), x \geq \chi_t$ is $\widehat{f}(x) = 0$. The number of strict ascending ladder points $\left(\tau_{ln}, \chi_{\tau_{ln}}\right)$ in time interval $[0, t]$ form a counting process as t increases. If the sample-path jump sizes are distributed as E_μ, then the n^{th} ascending ladder point is distributed as an Erlang-(n, μ) random variable. (We mention these ladder points because of their importance in the overall method. However, we shall not discuss them further in this introductory chapter on LCE.)

9.3 Computer Program for LCE

An LCE computer program can utilize different logical designs. Suppose we wish to estimate the steady-state pdf of wait. Assume that for fixed $t > 0$, we can simulate a sample path of the virtual wait over a simulated time interval $[0, t]$. We count the number of SP downcrossings of each state-space level $x \in \boldsymbol{S}$ during $[0, t]$. This is easier than it may seem at first glance, due to the step-function structure of $\mathcal{D}_t(x), x > 0$, for fixed $t > 0$.

9.3.1 Designs for Computer Program

We discuss two feasible designs for an LCE computer program.

State-space Partition with Variable Subintervals

One design is based directly on the discussion in Section 9.2, using partition $\{x_i\}$ having *variable* sub-interval lengths $\Delta_i = x_{i+1} - x_i$. The Δ_i's occur naturally in the simulated sample path (Fig. 9.1).

The embedded processes $\{W_n\}$ and $\{W_n + S_n\}$ are Markov processes. Thus, in a sample path the union $\{W_n\} \cup \{W_n + S_n\}$ of peaks and troughs, is everywhere dense in $\boldsymbol{S} = [0, \infty)$ as $t \to \infty$ (a.s.). That is, the entire state space will be covered *eventually* by the ordinates of the peaks and troughs.

An advantage of this design is that it takes every sample-path peak and trough during $[0, t]$ into consideration. In theory, any computed estimator will utilize all the information available in the sample path.

A possible disadvantage of this design is from a programming point of view. The points in $\{x_i\}$ become more dense as the sample path is generated over time. The Δ_i's in the region of higher probability, will become extremely small as simulated time t increases. The partition $\{x_i\}$ will contain on the order of $2N_a(t)$ distinct points, where $N_a(t)$ is the number of arrivals in time t (a peak and trough correspond to each arrival). If t is large, $N_a(t)$ will be large. Many Δ_i's will become less than a practical resolution size required for the estimation of the pdf of wait.

State-space Partition with Fixed Subintervals

A second design is to start with $x_0 = 0$ and a *fixed* partition norm size Δ. Thus $x_i = x_{i-1} + \Delta, i = 1, \ldots$. The program updates the count of SP downcrossings of each state-space level $x_i, i = 0, ..., M(t)$ as the sample path evolves over time interval $[0, t]$. We compute the maximum peak χ_s during $[0, s]$ as we generate a sample path over time. The state-space partition $\{x_i\}$ covers the state-space interval $[0, \chi_t]$. Generally the time intervals between successive ladder points of $\{W(s)\}$ increase. That is, $\tau_{l(n+1)} - \tau_{ln} > \tau_{ln} - \tau_{l(n-1)}$, after some $n \geq$ some integer $\in \{1, 2, ...\}$. Estimates of $\{P_0; f(x), x \geq 0\}$ that are computed using a fixed-Δ partition, very closely approximate estimates using a partition with variable Δ_i's, for most practical purposes. Moreover, the fixed-Δ design is easy to program.

9.4 LCE for M/G/1 Queue

This section describes LCE for the steady-state pdf of wait and related quantities for M/G/1 queues. A numerical example using this method is given in the next section. Let $\{W(t), t \geq 0\}$ denote the virtual wait. Without loss of generality assume $W(0) = 0$. The state space is $\boldsymbol{S} = [0, \infty)$. Let the arrival rate be λ. Let $S_n, n = 1, 2, ...$ denote the service times, which may be state dependent. Assume the parameters are such that the queue is stable, e.g., $\lambda E(S) < 1$. Assume $W(t) \underset{dist}{\rightarrow} W$ as $t \rightarrow \infty$ (weak convergence). Denote the cdf and pdf of W by $F(x), x \geq 0$, and $\{P_0; f(x), x > 0\}$ respectively. Here $P_0 = F(0) > 0$ and $f(x) = \frac{d}{dx} F(x)$ wherever the derivative exists. Denote the n^{th} moment of W by $m_n = \int_{x=0}^{\infty} x^n f(x) dx, n = 1, 2, ...$. Let $\psi(W)$ denote an arbitrary measurable function of W.

We use a computer program based on the fixed-norm size design of Subsection 9.3.1 to compute the estimators (fixed Δ). Definition 9.2 below incorporates minor modifications of the "basic" estimators, that retain theoretical consistency. The modified estimators are satisfactory for practical purposes.

9.4.1 Quantities Computed from a Sample Path

Fix finite time $t > 0$. Consider a simulated sample path of the virtual wait $\{W(s), 0 \leq s \leq t\}$. The SP is the leading point of a sample path when thought of as evolving over time (Section 2.3). In the fixed-Δ design, partition $\{x_i\}$ has a constant norm Δ. Define the following quantities.

Definition 9.1

$\mathcal{D}_t(x)$	number of SP downcrossings of level $x, x \geq 0$ during $[0, t]$,
χ_t	$\max\{W(s)\|0 \leq s \leq t\}$,
Δ	norm of preassigned uniform partition on $\boldsymbol{S}$,
ν	$\max\{n\|n\Delta \leq \chi_t, n = 0, 1, 2, ...\}$,
x_j	$x_j = j\Delta,\ j = 0, ..., \nu + 1;\ x_{\nu+2} \equiv \infty$,
$\{x_j\}$	preassigned uniform partition on $[0, (\nu + 1)\,\Delta]$ with norm Δ,
$\boldsymbol{J}_j$	interval $\boldsymbol{J}_j = [x_j, x_{j+1}), j = 0, 1, ..., \nu$,
d_j	$\mathcal{D}_t(x_j), j = 0, ..., \nu + 1$,
A_t	$A_t = \frac{1}{t}\left(\frac{d_0}{\lambda} + \Delta \sum_{j=0}^{\nu} d_j\right) = \frac{1}{t}\left(\frac{\mathcal{D}_t(0)}{\lambda} + \Delta \sum_{j=0}^{\nu} \mathcal{D}_t(x_j)\right)$.

Remark 9.3 *Definition 9.1 retains the argument "t" for $\mathcal{D}_t(x)$, χ_t and A_t. Both ν and d_j also depend on t. We omit subscript t for ν to simplify notation since ν often appears as a subscript or index. We omit the subscript t for d_j for computer-programming purposes. The quantities Δ, x_j and $\boldsymbol{J}_j$ are defined in the state space, and are generally independent of t. (However, we may vary t and Δ jointly for a **sensitivity analysis** in order to increase accuracy.)*

Remark 9.4 *Note the inequality $x_\nu = \nu\Delta \leq \chi_t < (\nu + 1)\Delta = x_{\nu+1}$. Also, for every $x \geq x_{\nu+1}$, $\mathcal{D}_t(x) = 0$, i.e., $d_{\nu+1} \equiv 0$.*

The term A_t is such that $A_t > 0, t > \tau_1$ (τ_1 = first arrival instant).

Proposition 9.1

$$\lim_{\substack{t\to\infty \\ \Delta\downarrow 0}} A_t = 1 \ (a.s.). \tag{9.12}$$

Proof. We *sketch* a proof of (9.12) in three steps.
(1) $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$ (normalizing condition).
(2)For the first term of A_t we have

$$\lim_{t\to\infty} \frac{d_0}{t\lambda} = \lim_{t\to\infty} \frac{\mathcal{D}_t(0)}{t\lambda} = \frac{f(0)}{\lambda} = \frac{\lambda P_0}{\lambda} = P_0 \ (a.s.). \tag{9.13}$$

(3) First assume the virtual wait $W(t) \leq K, t > 0$ for some upper bound $K > 0$. Evidence for the existence of such queueing models is demonstrated in Section 3.14. Then $\chi_t \leq K$ for all $t > 0$. Also $\nu \leq \left[\frac{K}{\Delta}\right]$ where $[z]$ denotes the greatest integer $\leq z, z \in \boldsymbol{R}$. Thus ν is finite and positive

for all values of t. For the second term of A_t we have

$$\begin{aligned}
\lim_{\substack{t\to\infty\\ \Delta\downarrow 0}} \left(\Delta \sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right) &= \lim_{\Delta\downarrow 0}\left(\lim_{t\to\infty}\left(\sum_{j=0}^{\nu}\frac{\mathcal{D}_t(x_j)}{t}\Delta\right)\right)\\
&= \lim_{\Delta\downarrow 0}\left(\sum_{j=0}^{\nu}\left(\lim_{t\to\infty}\frac{\mathcal{D}_t(x_j)}{t}\right)\Delta\right)\\
&= \lim_{\Delta\downarrow 0}\left(\sum_{j=0}^{\upsilon} f(x_j)\Delta\right)\\
&= \int_{x=0}^{K} f(x)dx \ (a.s.),
\end{aligned} \tag{9.14}$$

since $\lim_{t\to\infty}\frac{\mathcal{D}_t(x_j)}{t} = ff(x_j,(a.s.)$ by Theorem 1.1. In the last equality of (9.14), the expression $\sum_{j=0}^{\upsilon} f(x_j)\Delta$ is a Riemann sum. It converges to the definite integral $\int_{x=0}^{K} f(x)dx$ as $\Delta\downarrow 0$, since $K-\Delta < x_\nu \le K$.

The result (9.14) holds for every $K>0$. If $K\to\infty$, then

$$\lim_{\substack{t\to\infty\\ \Delta\downarrow 0}} \left(\Delta \sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right) = \int_{x=0}^{\infty} f(x)dx \ (a.s.). \tag{9.15}$$

Equation (9.12) then follows from (9.13), (9.14) and the normalizing condition. ■

9.4.2 Point Estimators

For fixed $t>0$ let

$$\widehat{f}(x), x>0,\ \widehat{F}(x), x\ge 0,\ \widehat{P}_0,\ \widehat{m}_n, n=1,2,...,\ \widehat{E}\left(\psi(W)\right),$$

denote *point estimators* of the corresponding quantities under the circumflexes. These point estimators are specified in Definition 9.2 below. Assume a "small" norm Δ is given (Δ = "bin size").

Definition 9.2 *For each fixed $t>0$, the **point estimators** are (see Definition 9.1):*

1. $\widehat{f}(x) \equiv \frac{d_j}{tA_t} = \frac{\mathcal{D}_t(x_j)}{tA_t}$, $x\in \boldsymbol{J}_j$, $j=0,...,\nu$,
2. $\widehat{P}_0 = \frac{d_0}{\lambda t A_t} = \frac{\mathcal{D}_t(0)}{\lambda t A_t}$,
3. $\widehat{F}(x) = \widehat{P}_0 + \Delta\sum_{i=0}^{j-1}\widehat{f}(x_i) + (x-x_j)\widehat{f}(x_j), x\in\boldsymbol{J}_j$, $j=0,...,\nu$,
4. $\widehat{m}_n = \Delta\sum_{i=0}^{\nu} x_i^n \widehat{f}(x_i)$,
5. $\widehat{E}(\psi(W)) = \psi(0)\widehat{P}_0 + \Delta\sum_{i=0}^{\nu}\psi(x_i)\widehat{f}(x_i)$.

Estimator of Laplace Stieltjes Transform

In Definition 9.2, set $\psi(W) = e^{-sW}, s > 0$. Then $E(\psi(W))$ is the Laplace-Stieltjes transform (LST) of W, namely

$$E(e^{-sW}) = \int_{x=0}^{\infty} e^{-sx} dF(x)dx.$$

The estimator of $E(e^{-sW})$ is

$$\widehat{E}(e^{-sW}) = \widehat{P}_0 + \Delta \sum_{i=0}^{\nu} e^{-sx_i} \widehat{f}(x_i), s > 0.$$

We may compute $\widehat{E}(e^{-sW})$ for $s = 0, h, 2h, ...$, where h is a small positive constant. Thus we can plot $\widehat{E}(e^{-sW})$ vs. s. Then we may substitute $\widehat{E}(e^{-sW})$ for the LST in formulas where it appears.

The value of Δ may be adjusted after a computer run, to increase accuracy or investigate an estimator's convergence rate with respect to Δ.

Remark 9.5 *In Definition 9.2 the quantities under the symbol "$\widehat{\ }$" omit the argument t, to distinguish them from* estimators of transient distributions. *(The latter estimators are not included in this monograph, but are discussed briefly in Remark 9.2 and remarks referred to therein.) The quantities also omit the argument Δ for notational simplicity.*

Remark 9.6 *For fixed $t > 0$, $\widehat{f}(x)$ is a step function of $x \in \cup_{j=0}^{\nu+1} \boldsymbol{J}_j$ having constant values on the intervals $\{\boldsymbol{J}_j\}$. The **term** A_t **is a normalizing constant** which guarantees that $\widehat{F}(x) = 1, x \geq x_{\nu+1}$, for any $t > 0$. Also, $\widehat{f}(x) = 0, x \in J_{\nu+1}$.*

Consistency of Estimators

An estimator $\widehat{\varphi}_t$ of quantity ϕ is *consistent* if $\lim_{t\to\infty} P(\widehat{\varphi}_t = \phi) = 1$. An estimator $\widehat{\varphi}_t$ of ϕ is *strongly consistent* if $P(\lim_{t\to\infty} \widehat{\varphi}_t = \phi) = 1$; equivalently $\lim_{t\to\infty} \widehat{\varphi}_t = \phi$ (*a.s.*).

The estimators

$$\widehat{f}(x), x > 0, \ \widehat{F}(x), x \geq 0, \ \widehat{P}_0, \ \widehat{m}_n, n = 1, 2, ..., \ \widehat{E}(\psi(W))$$

in Definition 9.2 are strongly consistent. The gist of the proofs utilizes level crossing theorems discussed in Subsection 9.2.2.

Proposition 9.2

1. (a). For each x_j, $\widehat{f}(x_j)$ is strongly consistent.
   (b). For each fixed $x \neq x_j$ $\lim_{\Delta\downarrow 0} \widehat{f}(x)$ is strongly consistent.

2. (a). For each fixed $t > 0$, $0 \leq \widehat{P}_0 \leq 1$.
   (b). $\widehat{P}_0$ is strongly consistent.

3. (a). For each fixed $t > 0$, $0 \leq \widehat{F}(x) \leq 1, x \geq 0$, and $\widehat{F}(\infty) = 1$.
   (b). For each fixed $x \geq 0$, $\lim_{\Delta\downarrow 0} \widehat{F}(x)$ is strongly consistent.

4. $\lim_{\Delta\downarrow 0} m_n$ is strongly consistent, $n = 1, 2, ...$

5. $\lim_{\Delta\downarrow 0} \widehat{E}(\psi(W))$ is strongly consistent.

Proof. 1(a).

$$\lim_{t\to\infty} \widehat{f}(x_j) = \lim_{t\to\infty} \frac{d_j}{tA_t} = \lim_{t\to\infty} \frac{D_t(x_j)}{tA_t} \underset{a.s.}{=} \frac{f(x_j)}{\lim_{t\to\infty} A_t} = f(x_j),$$

since $\lim_{t\to\infty} A_t = 1$ by formula (9.12).
1(b). Fix $t > 0$. Fix $x \in \boldsymbol{S}$. Let $\delta > 0$ be given. We can make the fixed norm size Δ arbitrarily small. There exists $\Delta > 0$ and x_j in the fixed norm partition such that $0 < x - x_j < \Delta$. Also we have $x - x_j < \Delta \implies |f(x) - f(x_j)| < \delta$, since $f(\cdot)$ is defined to be right continuous. Note that $\widehat{f}(x) \equiv \widehat{f}(x_j)$. Now let $t > t_{x_j\varepsilon}$, such that $t > t_{x_j\varepsilon}$ $\implies \left|\widehat{f}(x_j) - f(x_j)\right| < \varepsilon$. (Such $t_{x_j\varepsilon}$ exists by 1(a).) Hence for Δ sufficiently small and $t > t_{x_j\varepsilon}$,

$$\begin{aligned} \left|f(x) - \widehat{f}(x)\right| &= \left|f(x) - \widehat{f}(x_j)\right| = \left|f(x) - f(x_j) + f(x_j) - \widehat{f}(x_j)\right| \\ &\leq |f(x) - f(x_j)| + \left|f(x_j) - \widehat{f}(x_j)\right| \\ &< \delta + \varepsilon. \end{aligned}$$

As $t \to \infty$, $\left|f(x_j) - \widehat{f}(x_j)\right| \downarrow 0$. Thus $\left|f(x) - \widehat{f}(x)\right| < \delta$, implying that $\lim_{t\to\infty} \left(\lim_{\Delta\downarrow 0} \widehat{f}(x)\right) = f(x)$ $(a.s.)$.
2 (a). For fixed t, $\mathcal{D}_t(0) \geq 0$. hence

$$0 \leq \frac{D_t(0)}{\lambda t A_t} = \widehat{P}_0 = \frac{\frac{D_t(0)}{\lambda t}}{\left(\frac{\mathcal{D}_t(0)}{\lambda t} + \Delta \sum_{j=0}^{\nu} \frac{\mathcal{D}_t(x_j)}{t}\right)} \leq 1.$$

2 (b). For a stable queue, state $\{0\}$ is positive recurrent. Hence

$$\lim_{t\to\infty} \widehat{P}_0 = \lim_{t\to\infty} \frac{D_t(0)}{\lambda t A_t} = \frac{f(0)}{\lambda \lim_{t\to\infty} A_t} = \frac{\lambda P_0}{\lambda \cdot 1} = P_0 \ (a.s.).$$

3(a). This follows because the denominators of $\widehat{P}_0$ and $\widehat{f}(x_j), j = 1, ..., \upsilon$ contain the normalizing factor $A_t = \widehat{P}_0 + \Delta \sum_{j=0}^{\nu} \widehat{f}(x_j)$, which exceeds or equals the value of the total numerator.
3(b). This follows because $\lim_{t\to\infty} \widehat{P}_0 = P_0$. Also,

$$\begin{aligned}
&\lim_{t\to\infty}\left(\lim_{\Delta\downarrow 0}\left(\Delta\sum_{i=0}^{j-1}\widehat{f}(x_i) + (x - x_j)\widehat{f}(x_j)\right)\right)\\
&= \lim_{\Delta\downarrow 0}\left(\lim_{t\to\infty}\left(\Delta\sum_{i=0}^{j-1}\widehat{f}(x_i) + (x - x_j)\widehat{f}(x_j)\right)\right)\\
&= \lim_{\Delta\downarrow 0}\left(\Delta\sum_{i=0}^{j-1} f(x_i) + (x - x_j) f(x_j)\right),
\end{aligned}$$

since for fixed Δ, the values of the partition points $\{x_j\}$ are fixed (thus interchange of limits permitted). Hence

$$\lim_{t\to\infty}\left(\lim_{\Delta\downarrow 0}\widehat{F}(x)\right) = P_0 + \int_{y=0}^{x} f(x)dx = F(x) \ (a.s.).$$

4, 5. These follow using similar reasoning as in the proof of 3(b). ■

Remark 9.7 *In the estimation procedure of this section, we must make two important preset choices: (1) the value of simulated time t; (2) the value of Δ. Since t is finite and $\Delta > 0$, the estimators in Proposition 9.2 are* ***approximately*** *consistent. We consider the partition norm Δ to be sufficiently "small" if the following holds. We repeat the estimation procedure with a smaller Δ, say $\frac{\Delta}{10}$ or $\frac{\Delta}{100}$, etc.; this leaves the estimates within a preassigned tolerance.*

Similarly, we consider t to be sufficiently "large" if repeating the procedure with a larger t, say $10t$ or $100t$, etc., leaves the estimates within a preassigned tolerance (compare with Cauchy condition for convergence of series). The joint choice of (t, Δ) poses an interesting exercise. Experimentation may be informative. A discussion is given in [20]. Computational experimentation has shown that the estimation procedure is robust over a wide range of (t, Δ) values. With the advent of fast computer processors, fast random access memories, fast storage drives, etc., *a sensitivity analysis can be carried out very efficiently. Computer speeds will increase in the future. Sensitivity analyses of the estimates with respect to* (t, Δ) *will become ever more efficient.*

9.4.3 Statistical Properties and Confidence Limits

For an arbitrary sample path $W(s), 0 \leq s \leq t$, define the following quantities.

d_x	time between successive SP downcrossings of level x,
$Var(d_x)$	variance of d_x,
$\sqrt{Var(d_x)}$	standard deviation of d_x,
b_x	time SP is in state-space interval $[0, x]$ during d_x = sojourn time at or below level x,
$\mathcal{A}((W(\cdot))^n)$	area under the sample path of $(W(s))^n$ during a busy cycle of $W(s), 0 \leq s \leq t$,
λP_0	long-run rate at which arrivals initiate busy periods.

Asymptotic Normality of Estimators

The following proposition describes the asymptotic normality of the estimators. Let $N(0,1)$ denote a standard normal random variate with mean 0 and variance 1. Let $Var(Z)$ denote the variance of a generic random variable Z.

Proposition 9.3

1. For every $x_j, j = 0, ..., \nu$

$$\frac{\widehat{f}(x_j) - f(x_j)}{Var(d_{x_j})\left((tA_t)^{-1}\left(f(x_j)\right)^3\right)^{\frac{1}{2}}} \to N(0,1) \text{ as } t \to \infty.$$

2.

$$\frac{\widehat{P}_0 - P_0}{Var(d_0)\left((tA_t)^{-1}\lambda\left(P_0\right)^3\right)^{\frac{1}{2}}} \to N(0,1) \text{ as } t \to \infty.$$

3. If Δ is small then for every $x \geq 0$ approximately

$$\frac{\widehat{F}(x) - F(x)}{\left((tA_t)^{-1}Var(b_x - b_0)f(x)\right)^{\frac{1}{2}}} \to N(0,1) \text{ as } t \to \infty.$$

4. If Δ is small then approximately

$$\frac{\widehat{m}_n - m_n}{(t^{-1}Var(\mathcal{A}((W(\cdot))^n))\lambda P_0)^{\frac{1}{2}}} \to N(0,1) \text{ as } t \to \infty.$$

Proof. The proofs of statements 1 - 4 follow from the asymptotic normality of renewal processes (see e.g., [91] or [49]). This proposition is also discussed in Section 6 of [20], based on the same references. ■

Confidence Intervals for Estimators

Assume t is large and define $z_{\frac{\alpha}{2}}$ by $P(N(0,1) > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. The following $100(1-\alpha)\%$ confidence limits apply.

1. $f(x_j)$: $\quad \widehat{f}(x_j) \pm z_{\frac{\alpha}{2}} \cdot \widehat{Var}(d_{x_j}) \cdot \left((tA_t)^{-1}\widehat{f}(x_j)^3\right)^{\frac{1}{2}}$,

2. P_0: $\quad \widehat{P}_0 \pm z_{\frac{\alpha}{2}} \cdot \widehat{Var}(d_0) \cdot \left((tA_t)^{-1}\lambda\left(\widehat{P}_0\right)^3\right)^{\frac{1}{2}}$,

3. $F(x)$: $\quad \widehat{F}(x) \pm z_{\frac{\alpha}{2}} \cdot \left((tA_t)^{-1}\widehat{Var}(b_x - b_0)\widehat{f}(x_j)\right)^{\frac{1}{2}}$,

4. m_n: $\quad \widehat{m}_n \pm z_{\frac{\alpha}{2}} \cdot \left(t^{-1}\widehat{Var}(\mathcal{A}((W(\cdot))^n))\lambda\widehat{P}_0\right)^{\frac{1}{2}}$.

Proof. The profs are based on Proposition 9.3. ■

9.5 LCE Example: M/M/1 with Reneging

We consider an $M_\lambda/M_\mu/1$ queue in which customers may renege from the waiting line, or wait and balk at start of service (Section 3.11, Subsection 3.11.7 and equations (3.166), (3.167)). Alternatively customers may wait and stay for complete service. We compare LCE estimates of the steady-state pdf, cdf and mean wait of stayers with the analytical solutions for the same quantities.

We assume customers that wait less than 1 time unit stay ("reach" the server) and get complete service. Customers that are required to wait ≥ 1 time unit to reach the server, renege from the waiting line or wait the full time and then balk at service. In the notation of Section

3.11 the *staying function* $\overline{R}(x), x \geq 0$ has the same form as in Fig. 3.21, i.e.,

$$\overline{R}(x) = \begin{cases} 1, 0 \leq x < 1, \\ 0, x \geq 1. \end{cases} \tag{9.16}$$

The arrival rate λ and service rates μ may be arbitrary positive numbers since the queue is stable for all values of λ, μ (Theorem 3.8). We arbitrarily set $\lambda = 1$, $\mu = 5$.

Analytical Solution

We obtain the analytical solution for the pdf of the wait of stayers $\{P_0; f(x), x > 0\}$ from the model equations

$$f(x) = \begin{cases} \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y)dy, 0 < x < 1, \\ \lambda P_0 e^{-\mu x} + \lambda \int_{y=0}^{1} e^{-\mu(x-y)} f(y)dy, x \geq 1. \end{cases} \tag{9.17}$$

The solution of (9.17) is

$$f(x) = \begin{cases} \lambda P_0 e^{-(\mu-\lambda))x}, \ 0 < x < 1, \\ \lambda P_0 e^{\lambda} e^{-\mu x}, \ 1 \leq x < \infty. \end{cases} \tag{9.18}$$

We substitute (9.18) into the normalizing condition $P_0 + \int_{x=0}^{\infty} f(x)dx = 1$, yielding

$$P_0 = \frac{1}{1 + \frac{\lambda}{\mu-\lambda}(1 - e^{-(\mu-\lambda)}) + \frac{\lambda}{\mu} e^{-(\mu-\lambda)}}. \tag{9.19}$$

Substituting $\lambda = 1$, $\mu = 5$ in (9.19) and (9.18) results in (Fig. 9.3)

$$P_0 = 0.8006, \tag{9.20}$$

$$f(x) = \begin{cases} 0.8006 \cdot e^{-4.0x}, 0 < x < 1, \\ 2.1763 \cdot e^{-5.0x}, 1 \leq x < \infty. \end{cases} \tag{9.21}$$

From (9.21) the derivative is

$$f'(x) = \begin{cases} -3.2024 \cdot e^{-4.0x}, 0 < x < 1, \\ -10.8815 \cdot e^{-5.0x}, 1 \leq x < \infty. \end{cases}$$

The pdf $f(x)$ is continuous at $x = 1$. The derivative $f'(x)$ is discontinuous at $x = 1$. Thus $f'(1^-) = -0.058654$, $f'(1) = -0.073319$. The pdf is bounded above by the arrival rate λ, i.e.,

$$\max_{x \geq 0} f(x) = f(0) = 0.8006 < 1 = \lambda.$$

LC Estimation using $t = 3000, \Delta = 0.1$				
Estimated values			Analytical Values	
$\widehat{P}_0 = 0.7995$			$P_0 = .800587$	
x	$\widehat{f}(x)$	$\widehat{F}(x)$	$f(x)$	$F(x)$
0.1	.7995	.7995	.8006	.8006
0.2	.5265	.8652	.5366	.8666
0.3	.2447	.9395	.2411	.9403
0.4	.1602	.9591	.1616	.9603
0.5	.1142	.9734	1083	.9736
0.6	.0729	.9828	.0726	.9826
0.7	.0484	.9809	.0487	.9886
0.8	.0317	.9929	.0326	.9926
0.9	.0208	.9955	.0219	.9953
1.0	.0147	.9973	.0147	.9971
1.1	.0092	.9984	.0089	.9982
1.2	.0058	.9992	.0054	.9989
1.3	.0031	.9996	.0033	.9993
1.4	.0010	.9998	.0020	.9996
1.5	.0007	.9999	.0012	.9998
1.6	.0003	1.000	.0007	.9999
1.7	.0000	1.000	.0004	.9999

Table 9.1: Comparison of LC estimation with steady-state analytic values for M/M/1 with reneging or balking at service

LCE Estimates of PDF and CDF of Wait of Stayers

We present the LCE estimates of $f(x)$, $F(x)$ and P_0 in Table 9.1, using $t = 3000$, $\Delta = 0.1$.

LCE Estimates of Mean of Wait of Stayers and P_0

From (9.21), $E(W_q) \equiv \int_{x=0}^{\infty} xf(x)dx \equiv m_1 = 0.049$, where W_q denotes the required wait of stayers before service. Simulation of 10 independent sample paths using $t = 3000$, $\Delta = 0.1$, resulted in the sample-average point estimate $\overline{\widehat{m}_1} = 0.0489$. A 95% confidence interval for m_1 is obtained using $t_{9,0.025} \cdot s_{\widehat{m}_1}$ where $t_{9,0.025}$ is the right 2.5% tail of the Student "t" distribution with 9 degrees of freedom (Student "t" because 10 is a small sample size) and $s_{\widehat{m}_1}$ is the sample standard deviation of $\widehat{m}_1$. The value of $t_{9,0.025} \cdot s_{\widehat{m}_1}$ turned out to be 0.0013. Thus a 95% confidence interval is $m_1 = \overline{\widehat{m}_1} \pm t_{9,0.025} \cdot s_{\widehat{m}_1}$ or $m_1 = 0.0489 \pm 0.0013$, which covers the true mean wait.

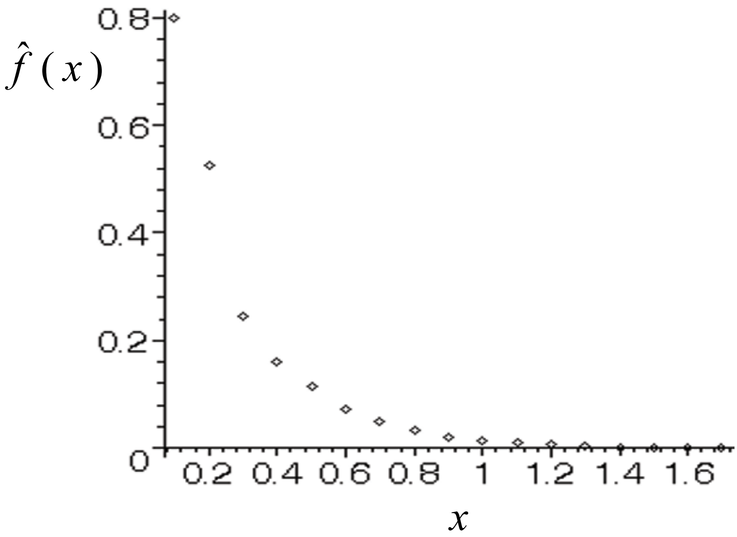

Figure 9.2: Point estimate $\widehat{f}(x)$ based on Table 9.1, for $f(x)$ in $M_\lambda/M_\mu/1$ queue with reneging or balking at service: $\lambda = 1$, $\mu = 5$. Compare with Fig. 9.3.

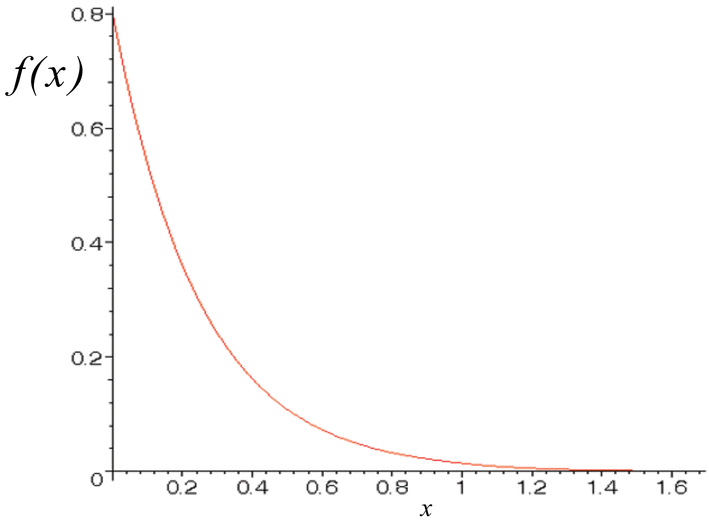

Figure 9.3: Analytical solution for $f(x)$ in $M_\lambda/M_\mu/1$ queue with reneging or balking at service: $\lambda = 1$, $\mu = 5$. See formulas (9.16), (9.18), (9.21). $f(x)$ is continuous at $x = 1$. $f'(x)$ is discontinuous at $x = 1$.

Similarly a 95% confidence interval for P_0 is $P_0 = \overline{\widehat{P}}_0 \pm t_{9,0.025} \cdot s_{\widehat{P_0}}$ or $P_0 = 0.7996 \pm 0.0025$, which covers the true value of P_0.

Discussion of Numerical Example

The probability that an arbitrary arrival stays and receives full service is

$$\begin{aligned} q_S &= P_0 + \int_{x=0}^{\infty} \overline{R}(x) f(x) dx \\ &= P_0 + \int_{x=0}^{1} f(x) dx \\ &= 0.8006 + \int_{x=0}^{1} 0.8006 e^{-4.0x} dx \\ &= 0.9971. \end{aligned}$$

For the particular choice of $(\lambda, \mu) = (1, 5)$ and $\overline{R}(\cdot)$ in the example, nearly all customers say, i.e., wait and get full service. Only $(1 - q_S) \cdot 100\% = 0.29\%$ either renege or balk at start of service. The reason is that the service rate is very fast relative to the arrival rate. The vast majority of arrivals (99.71%) are required to wait less than one time unit, and therefore stay for a full service.

The expected busy period is

$$E(\mathcal{B}) = \frac{1 - P_0}{\lambda P_0} = 0.24906.$$

The expected idle period is $E(\mathcal{I}) = \frac{1}{\lambda} = 1$. The proportion of time the server is idle is $\frac{E(\mathcal{I})}{E(\mathcal{I}) + E(\mathcal{B})} = 0.8006 = P_0$. Different values of (λ, μ) would, of course, give quite different results.

9.6 Discussion

LCE is useful for confirming theoretical results derived by various methods of analysis. LCE can be used to investigate the pdf of a state variable in a new model where the model equations are difficult to formulate, or' if formulated, are analytically intractable. It is an alternative approach for estimating pdf's, cdf's, moments, and expected values of functions of state variables (e.g., Laplace transforms) in stochastic models.

LCE for steady-state distributions has several advantages. It uses a single simulated sample path of the model. It requires the analyst to be sufficiently familiar with the model dynamics to construct a sample path using a computer program. It may help to uncover and explain subtleties about the pdf and cdf of the state variable, which enhance intuition about the model. It may help to discover unexpected properties about the pdf of the state variable.

LCE can be incorporated into a *hybrid* technique combining partially-known analytical solutions and statistical estimation. For example, in a single-server queue, the theoretical values of P_0 (probability of a zero wait) and $E(\mathcal{B})$ (expected busy period) may be known in terms of the model parameters. On the other hand, equations for the pdf of wait $f(x), x > 0$, may be analytically intractable. It may be possible to utilize the theoretical values of P_0 and $E(\mathcal{B})$ in the LCE computer program, to estimate $f(x), x > 0$.

LCE methods similar to that described here for M/G/1, have been applied to M/G/r($\cdot$) dams including cases where G is deterministic or discrete [22]; and to more complex models such as M/G^{a,b}/1 bulk-service queues [32]. The LCE technique is applicable in a vast array of other stochastic models as well.

We may classify the LCE method as an estimation method, or a *computational* method. With sensible values of the simulated time t and state-space partition norm size Δ, the technique gives almost-analytical values for the distribution of the state variable and related values, in many benchmark computational experiments already carried out.

CHAPTER 10

ADDITIONAL APPLICATIONS

10.1 Introduction

This chapter applies SPLC to a variety of stochastic models, in order to indicate the scope, applicability and flexibility of the methodology, and to suggest new applications. The chapter begins with the LC analysis of a replacement model, which is structured using renewal processes. In that model, we derive limiting pdf's of the excess life, age and total life of a renewal process, using LC. The chapter ends with the LC analysis of a classical renewal problem. The intervening sections analyze several models that suggest many additional potential applications of SPLC.

10.2 Renewal Processes

We shall derive steady-state pdf's of renewal processes in the context of a replacement model. This model is a variant of a GI/G/r(·) dam.

10.2.1 A Replacement Model

Consider a continuous-time stochastic process $\{X(t) \geq 0, t \geq 0\}$ having iid jumps of size $X_n > 0$ at τ_n, where $0 = \tau_0 < \tau_1 \cdots < \tau_n < \cdots$. Thus $X(\tau_n) = X_n$, $n = 0, 1, 2, ...$, (Fig. 10.1). Consider a sample path of $\{X(t)\}$ (we use $X(t)$ to denote the state variable and a sample path, for economy of notation). Assume $\frac{dX(t)}{dt} = -r(X(t))$, $t \in [\tau_n, \tau_{n+1})$, $n = 0, 1, ...$, where $r(x) > 0$, $x > 0$. Thus $X(t)$ is a piecewise deterministic

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_10,

function. Let the state space be $\boldsymbol{S} = [0, \infty)$. Assume that for all $v > 0$,

$$\lim_{u \downarrow 0} \int_{y=u}^{v} \frac{1}{r(y)} dy < \infty. \tag{10.1}$$

Condition (10.1) guarantees that a sample path $X(t), t \geq 0$, starting from any level $v > 0$, returns to level 0 in a finite time. The process $\{X(t)\}$ is a variant of the GI/G/r(·) dam such that inputs $\{X_n\}$ occur only at instants when the dam becomes empty. This mechanism can be thought of as that of a *replacement model.* New inputs replace the preceding inputs as soon as the latter become used up.

Denote the inter-replacement times by $\{Z_n\}$. The random variables Z_n and X_n are related by the equation

$$Z_n = \int_{y=0}^{X_n} \frac{1}{r(y)} dy, n = 0, 1, \ldots . \tag{10.2}$$

From (10.2), Z_n is the *time* required for $\{X(t)\}$ to descend from level X_n to level 0. The $\{Z_n\}$ are iid random variables.

Renewal Processes $\{Z_n\}$ and $\{X_n\}$

The sequence $\{Z_n\}$ is a renewal process synchronized with the sequence $\{X_n\}$ and with the piecewise deterministic continuous efflux rate $r(X(t))$. Due to the structure of the model, the sequence $\{X_n\}$ is also a renewal process.

Let $X_n \underset{dist}{\equiv} X$ and $Z_n \underset{dist}{\equiv} Z$.

Example 10.1 *Consider a newly-installed battery at τ_0 with initial electrical charge $X_0 \underset{dist}{\equiv} X$. Assume that the charge declines at a rate that depends on the present charge. That is, $\frac{dX(t)}{dt} = -r(X(t)) < 0, t \in [\tau_0, \tau_1)$. Assume the battery operates continuously. Its charge dissipates non-uniformly and descends to 0 after a time $\tau_1 = Z_0 \underset{dist}{\equiv} Z$. The battery is immediately replaced by a new fully-charged one. This procedure is repeated as batteries wear out. Thus $Z_n \underset{dist}{\equiv} Z$, $X_n \underset{dist}{\equiv} X$, $n = 0, 1, 2, \ldots.$ Then*

$$Z = \int_{y=0}^{X} \frac{1}{r(y)} dy, \tag{10.3}$$

is the inter-replacement time. The dimension of Z is $[Time]$. The dimension of X is $[Coulombs]$. The function $r(X(t))$ has dimension $[Coulomb][Time]^{-1}$.

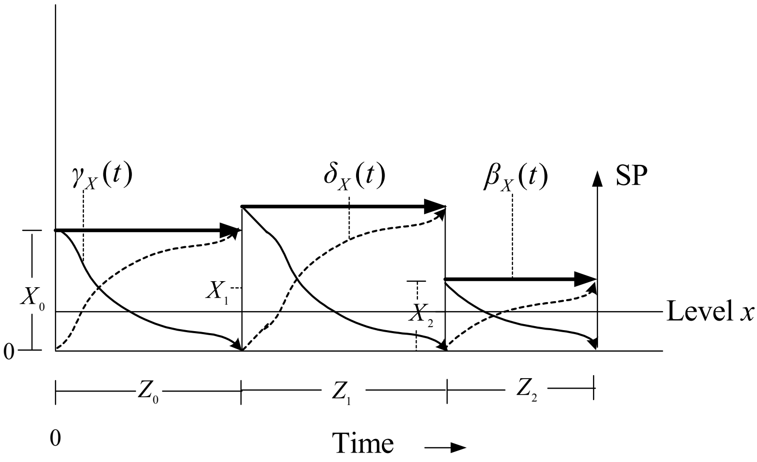

Figure 10.1: Sample path of excess life $\gamma_X(t)$, age $\delta_X(t)$, total life $\beta_X(t)$. Also shows a level x in the state space.

10.2.2 Renewal Process $\{X_n\}$

Excess Life, Age, Total Life

Let $\gamma_X(t)$ ($\equiv X(t)$) denote the *excess life* of *content* at instant $t \geq 0$. Then $\frac{d\gamma_X(t)}{dt} = -r(\gamma_X(t))$. Let $\delta_X(t)$ denote the *age* of the *content*, i.e., amount of content used up at instant t, from the latest renewed amount prior to t. Then $\frac{d\delta_X(t)}{dt} = +r(\delta_X(t))$. Let $\beta_X(t)$ denote the *total life* (*span*) of the latest renewed amount of *content* at t (Fig. 10.1). (In Example 10.1, $\gamma_X(t)$), $\delta_X(t)$, $\beta_X(t)$ are respectively to the remaining charge, the charge used up, and the total charge, of the battery in use at time t.)

In the sample paths of the processes $\{\gamma(t)\}$, $\{\delta(t)\}$, $\{\beta(t)\}$ all upward jumps start at level 0 and are $\underset{dist}{=} X$. All downward jumps start at a level X and end at level 0.

Limiting Distributions

We now apply LC to derive the limiting pdf's $f_{\gamma_X}(x)$, $f_{\delta_X}(x)$, $f_{\beta_X}(x), x > 0$, of r.v.'s $\gamma_X(t)$, $\delta_X(t)$, $\beta_X(t)$, as $t \to \infty$, assuming the limits exist. Consider sample paths of $\{\gamma_X(t)\}$, $\{\delta_X(t)\}$, $\{\beta_X(t)\}$, $t \geq 0$ (Fig. 10.1).

Let $F_X(x)$, $f_X(x)$, μ_X be the cdf, pdf and expected value respectively of r.v. X. Let $\overline{F_X}(x) \equiv 1 - F_X(x)$.

Limiting PDF of Excess Life

Consider a sample path of $\{\gamma(t)\}$. The long-run SP expected *downcrossing* rate of a *content* level $x > 0$, is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = r(x) f_{\gamma_X}(x). \tag{10.4}$$

(as in Corollary 6.2).

The long-run SP expected *upcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = \frac{1}{E(Z)} \cdot \overline{F_X}(x), \tag{10.5}$$

since the expected time between upward jumps starting from level 0 is $E(Z)$ $(= E(\tau_{n+1} - \tau_n), n = 0, 1, ...)$; also $\overline{F_X}(x) = P$(SP jump starting at level 0 is $> x$). In (10.3), substituting from (10.2), conditioning on $X = x$ gives

$$\begin{aligned} E(Z) &= \int_{x=0}^{\infty} \left(\int_{y=0}^{x} \frac{1}{r(y)} dy \right) f_X(x) dx \\ &= \int_{y=0}^{\infty} \int_{x=y}^{x} \frac{1}{r(y)} f_X(x) dx dy = \int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy. \end{aligned} \tag{10.6}$$

Equating (10.4) and (10.5) for rate balance across level x, and using (10.6), yields the equation

$$r(x) f_{\gamma_X}(x) = \frac{\overline{F_X}(x)}{E(Z)} = \frac{\overline{F_X}(x)}{\int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}, \tag{10.7}$$

$$f_{\gamma_X}(x) = \frac{\overline{F_X}(x)}{r(x) \int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}.. \tag{10.8}$$

The dimension of $f_{\gamma_X}(x)$ is $[content]^{-1}$ ($[Coulomb]^{-1}$ in Example 10.1).

Limiting PDF of Excess Life when $r(x) \equiv 1$

If the efflux rate $r(x) \equiv 1$, formula (10.8) reduces to

$$f_{\gamma_X}(x) = \frac{\overline{F_X}(x)}{\int_{y=0}^{\infty} \overline{F_X}(y) dy} = \frac{\overline{F_X}(x)}{\mu_X}, \tag{10.9}$$

since $\int_{y=0}^{\infty} \overline{F_X}(y) dy = E(X) = \mu_X$. (Note that γ_X represents the limiting excess life of *content* having pdf $f_{\gamma_X}(x)$.) Formula (10.9) is exactly the same as the well known limiting pdf of the excess life in a "standard" renewal process. However, here the dimension of $f_{\gamma_X}(x)$ is $[content]^{-1}$ instead of $[Time]^{-1}$.

Limiting PDF of Age

For the process $\{\delta_X(t)\}$, the long-run SP expected upcrossing rate of a content level $x > 0$, is

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = +r(x) f_{\delta_X}(x), \tag{10.10}$$

(as in Corollary 6.2). The long-run SP (expected) downcrossing rate of level x is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \frac{1}{E(Z)} \int_{y=x}^{\infty} f_X(y)dy = \frac{\overline{F_X}(x)}{E(Z)}, \tag{10.11}$$

since (1) downward jumps occur at rate $\frac{1}{E(Z)}$, (2) in order for the SP to downcross level x by a jump at some τ_n^-, the *upward* jump at τ_{n-1} from level 0 must have been such that $X_{n-1} > x$. Additionally, X_{n-1} is equal to the downward jump size at τ_n^- (Fig. 10.1).

Equating (10.10) and (10.11) for rate balance across level x, gives

$$\begin{aligned} r(x) f_{\delta_X}(x) &= \frac{\overline{F_X}(x)}{E(Z)} = \frac{\overline{F_X}(x)}{\int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}; \\ f_{\delta_X}(x) &= \frac{\overline{F_X}(x)}{r(x) \int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}. \end{aligned} \tag{10.12}$$

Comparison of (10.8) with (10.12) shows that $f_{\delta_X}(x) \equiv f_{\gamma_X}(x)$. The dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Age when $r(x) \equiv 1$

If $r(x) \equiv 1$, we obtain similarly as in (10.9), the limiting pdf

$$f_{\delta_X}(x) = \frac{\overline{F_X}(x)}{\mu_X}. \tag{10.13}$$

The dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$. It is well known that for a "standard" renewal process, the limiting distributions of the excess life and age are identical. In the variant of a GI/G/r(·) dam possessing the renewal structure here, these distributions are also identical with regard to the content, even when the efflux rate has a general form $r(x), x > 0$. That is, formulas (10.8) and (10.12) are identical.

Limiting PDF of Total Life

For the process $\{\beta_X(t)\}$, the long-run SP expected downcrossing rate of a *content* level $x > 0$, is

$$\lim_{t\to\infty} \frac{E(\mathcal{D}_t(x))}{t} = \int_{y=x}^{\infty} \left(\frac{1}{\int_{u=0}^{y} \frac{1}{r(u)} du} \right) f_{\beta_X}(y)dy. \tag{10.14}$$

In (10.14), we have conditioned on $\beta_X(t) = y > x$. The SP *downward* jump rate across level x starting at level y is $\frac{1}{\left(\int_{u=0}^{y} \frac{1}{r(u)} du\right)}$, which is *the reciprocal of the expected sojourn time of* $\{\beta_X(t)\}$ *at level* y (Fig. 10.1). At the end of a level-y ($y > x$) sojourn time, the SP jumps downward to level 0. It downcrosses every state-space level in $(0, y)$, including level x.

The SP long-run (expected) *upcrossing* rate of level x is

$$\lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} = \frac{1}{E(Z)} \int_{y=x}^{\infty} f_X(y)dy = \frac{\overline{F_X}(x)}{E(Z)}, \tag{10.15}$$

since the expected time between SP upward jumps out of level 0 is $E(Z)$, and the probability that such an SP jump exceeds level x is $\overline{F_X}(x)$. Note that the SP *double jumps* in opposite directions at each renewal instant of the sequence $\{Z_n\}$. One jump is downward ending at level 0; the "opposite jump" is upward starting at level 0.

Equating (10.14) and (10.15) for rate balance across level x, results in the integral equation for $f_{\beta_X}(\cdot)$,

$$\int_{y=x}^{\infty} \frac{1}{\left(\int_{u=0}^{y} \frac{1}{r(u)} du\right)} f_{\beta_X}(y)dy = \frac{\overline{F_X}(x)}{E(Z)}. \tag{10.16}$$

In (10.16), we differentiate with respect to x to yield

$$-\frac{1}{\left(\int_{u=0}^{x} \frac{1}{r(u)} du\right)} f_{\beta_X}(x) = -\frac{f_X(x)}{E(Z)}.$$

Hence

$$f_{\beta_X}(x) = \frac{\left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(x)}{E(Z)} = \frac{\left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(x)}{\int_{y=0}^{\infty} \frac{\overline{F}(y)}{r(y)} dy}. \tag{10.17}$$

The dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Total Life when $r(x) \equiv 1$

Assume $r(x) \equiv 1$. Then $Z_n = X_n$ and $E(Z_n) = E(X_n) = \mu_X$ in value. However, the dimensions differ: thus $[X_n] = [content]$ and $[Z_n] = [Time]$. Formula (10.17) resembles the well known limiting pdf of total life (span) for a *standard* renewal process,

$$f_{\beta_X}(x) = \frac{x f_X(x)}{E(Z)} = \frac{x f_X(x)}{\mu_X}, \tag{10.18}$$

except that the dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$ instead of $[Time]^{-1}$. That is, in the variant of the GI/G/r(·) dam described, the "life" is measured in content dimensions.

Remark 10.1 *This variant of GI/G/r(·) exhibits SP multiple jumps at the same instant (renewal instant). Recall that SP jumps in the state space* ***do not occur in Time****. (See Examples 2.2, 2.3 in Section 2.3, regarding SP multiple jumps.)*

Example 10.2 *Suppose $r(x) = kx, x > 0$, where $k > 0$ is a constant. Then the inequality (10.1) does not hold. However, the SP returns to every level $x > 0$, however small. We may select a small $\varepsilon > 0$, such that when the content hits level ε from above, a replenishment of new content is inserted (e.g., in Example 10.1, replace a battery with a new one when its charge decreases to ε Coulombs).*

Then for each positive $v > \varepsilon$,

$$\int_{y=\varepsilon}^{v} \frac{1}{kx} dx = \frac{1}{k} \ln \frac{v}{\varepsilon} < \infty,$$

so that the content returns to level ε in a finite time.

10.2.3 Renewal Process $\{Z_n\}$

Excess Life, Age, Total Life of $\{Z_n\}$ Process

Consider $\{Z_n\}$. Let $\gamma_Z(t)$, $\delta_Z(t)$, $\beta_Z(t)$ denote the excess life, age, total life respectively, at instant $t > 0$. Denote the limiting r.v.'s by γ_Z, δ_Z, β_Z respectively.

Define $\mathcal{G}(x) \equiv \int_{y=0}^{x} \frac{1}{r(y)} dy, x > 0$. Then $\mathcal{G}(x)$ is an increasing differentiable function of x (since $\frac{d}{dx}\mathcal{G}(x) = \frac{1}{r(x)}$). This implies $\mathcal{G}^{-1}(x)$ (inverse of $\mathcal{G}(x)$) exists, and

$$\frac{d}{dx}\mathcal{G}^{-1}(x) = \frac{1}{\frac{d}{dx}\mathcal{G}(x)} = \frac{1}{\frac{1}{r(x)}} = r(x), x > 0.$$

Thus $\mathcal{G}^{-1}(x)$ is also an increasing (differentiable) function of x. The quantity $\mathcal{G}(x)$ is the time required for the SP to descend from level x to level 0. The inverse $\mathcal{G}^{-1}(x)$ is the starting level of content, from which a descent to level 0 takes time x.

We may derive the pdf's of γ_Z, δ_Z, β_Z from the the results for the pdf's of γ_X, δ_X, β_X, respectively.

Limiting PDF of Excess Life of $\{Z_n\}$

The relation between Z_n and $X(t)$ implies

$$\gamma_Z \leq x \text{ iff } \gamma_X \leq \mathcal{G}^{-1}(x)).$$

Hence

$$F_{\gamma_Z}(x) = F_{\gamma_X}(\mathcal{G}^{-1}(x)). \tag{10.19}$$

(see Fig. 10.1).

Taking $\frac{d}{dx}$ on both sides of (10.19) and referring to (10.8) gives

$$\begin{aligned} f_{\gamma_Z}(x) &= f_{\gamma_X}(\mathcal{G}^{-1}(x)) \cdot \frac{d}{dx}\mathcal{G}^{-1}(x) \\ &= f_{\gamma_X}(\mathcal{G}^{-1}(x)) \cdot r(x) = \frac{r(x) \cdot \overline{F_X}(\mathcal{G}^{-1}(x))}{r(\mathcal{G}^{-1}(x)) \int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}. \end{aligned} \tag{10.20}$$

The dimension of $f_{\gamma_Z}(x)$ is $[Time]^{-1}$.

If $r(y) \equiv 1$ then $\mathcal{G}(x) = \mathcal{G}^{-1}(x) = x$. In that case $f_{\gamma_Z}(x) = \frac{\overline{F_X}(x)}{\int_{y=0}^{\infty} \overline{F_X}(y) dy} = f_{\gamma_X}(x)$, but the dimension of $f_{\gamma_Z}(x)$ is $[Time]^{-1}$, whereas the dimension of $f_{\gamma_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Age of $\{Z_n\}$

In a similar manner as for the excess life, the age satisfies

$$\delta_Z \leq x \text{ iff } \delta_X \leq \mathcal{G}^{-1}(x).$$

Thus, $F_{\delta_Z}(x) = F_{\delta_X}(\mathcal{G}^{-1}(x))$. Taking $\frac{d}{dx}$ then yields

$$f_{\delta_Z}(x) = \frac{r(x)\overline{F_X}(\mathcal{G}^{-1}(x))}{r(\mathcal{G}^{-1}(x)) \int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}. \tag{10.21}$$

Thus $f_{\delta_Z}(x) \equiv f_{\gamma_Z}(x)$. The dimension of $f_{\delta_Z}(x)$ is $[Time]^{-1}$.

If $r(y) \equiv 1$ then $\mathcal{G}(x) = \mathcal{G}^{-1}(x) = x$. Then $f_{\delta_Z}(x) = \frac{\overline{F_X}(x)}{\int_{y=0}^{\infty} \overline{F_X}(y)dy} = f_{\delta_X}(x)$. The dimension of $f_{\delta_Z}(x)$ is $[Time]^{-1}$, whereas the dimension of $f_{\delta_X}(x)$ is $[content]^{-1}$.

Limiting PDF of Total Life of $\{Z_n\}$

Note that $\beta_Z \leq x$ iff $\beta_X \leq \mathcal{G}^{-1}(x)$. Hence, as for $f_{\delta_Z}(x)$, $f_{\gamma_X}(x)$ above, we obtain

$$\begin{aligned} f_{\beta_Z}(x) &= f_{\beta_X}(\mathcal{G}^{-1}(x)) \cdot \frac{d}{dx}\mathcal{G}^{-1}(x) \\ &= f_{\beta_X}(\mathcal{G}^{-1}(x)) \cdot r(x). \end{aligned}$$

From (10.17) we get

$$f_{\beta_Z}(x) = \frac{r(x) \cdot \left(\int_{y=0}^{x} \frac{1}{r(y)} dy\right) f_X(\mathcal{G}^{-1}(x))}{\int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}. \tag{10.22}$$

The dimension of $f_{\beta_Z}(x)$ is $[Time]^{-1}$ whereas the dimension of $f_{\beta_X}(x)$ is $[content]^{-1}$. When $r(x) \equiv 1$, $f_{\beta_Z}(x) = \frac{x f_X(x)}{\int_{y=0}^{\infty} \frac{\overline{F_X}(y)}{r(y)} dy}$, having dimension $[Time]^{-1}$.

10.2.4 Standard Renewal Process

We now obtain the steady-state pdf's for the standard renewal process as a special case of those for the replacement model. In the *standard* renewal process, we have $X_n = Z_n, n = 0, 1, 2...$, and $r(X(t)) \equiv 1$. The dimensions of X_n and Z_n are the same, usually $[Time]$. The pdf's $f_{\gamma_Z}(x)$, $f_{\delta_Z}(x)$, $f_{\beta_Z}(x), x > 0$ are the same as (10.9), (10.13), (10.18) respectively, and all have dimension $[Time]^{-1}$.

Remark 10.2 *The LC derivations of the limiting pdf's of excess life, age and total life are* ***relatively*** *simple in the replacement model, and are much simpler for the standard renewal process. They are intuitive, and naturally suggest potential generalizations.*

Remark 10.3 *The derivations in this section are based directly on my unpublished notes of June 18-July 26, 1992 [23]. These notes were motivated by a talk at the* 21^{st} *conference on Stochastic Processes and their Applications, York University, Toronto, June 15-19, 1992 by van Harn and Steutel (see Partial Bibliography).(Their generalization differs conceptually from LC.) Results using LC for* ***standard*** *renewal processes were published independently by Katayama (2002) (see Partial Bibliography).*

10.3 A Technique for Transient Distributions

In this section we outline a technique for deriving transient distributions of processes with a continuous or discrete state, and a continuous parameter. The technique is based on the general version of Theorem B (Theorem 4.1). We repeat formulas (4.1) and (4.2) of Theorem B here for reference. For fixed $t > 0$

$$E(\mathcal{I}_t(\boldsymbol{A})) = E(\mathcal{O}_t(\boldsymbol{A})) + P_t(\boldsymbol{A}) - P_0(\boldsymbol{A}),\ t \geq 0, \tag{10.23}$$

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A})) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A})) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}), t > 0, \tag{10.24}$$

where $\mathcal{I}_t(\boldsymbol{A})$ is the number of SP entrances and $\mathcal{O}_t(\boldsymbol{A})$ is the number of SP exits, of state-space set $\boldsymbol{A}$ during $[0, t]$. Let the parameter set be $\boldsymbol{T} = [0, \infty)$

Remark 10.4 *If the* limiting *distribution of the state variable exists, it is obtained by taking the limit of the derived transient distribution as* $t \to \infty$.

10.3.1 State-space Set with Variable Boundary

State Space $\boldsymbol{S} \subseteq \boldsymbol{R}$

In formulas (10.23) and (10.24) assume set $\boldsymbol{A}$ depends on a continuous variable x and define $\boldsymbol{A} \equiv \boldsymbol{A}_x, x \in \boldsymbol{S}$. Thus x may be a state-space level, e.g., $\boldsymbol{T} \times \{x\}$ (a line in the $\boldsymbol{T}$-$\boldsymbol{S}$ coordinate system). For fixed x, replace formulas (10.23) and (10.24) by

$$E(\mathcal{I}_t(\boldsymbol{A}_x)) = E(\mathcal{O}_t(\boldsymbol{A}_x)) + P_t(\boldsymbol{A}_x) - P_0(\boldsymbol{A}_x) \tag{10.25}$$

$$\frac{\partial}{\partial t}E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial}{\partial t}E(\mathcal{O}_t(\boldsymbol{A}_x)) + \frac{\partial}{\partial t}P_t(\boldsymbol{A}_x). \tag{10.26}$$

Assume the following mixed partial derivatives exist and are equal, i.e.,

$$\frac{\partial^2}{\partial x \partial t}E(\mathcal{O}_t(\boldsymbol{A}_x)) = \frac{\partial^2}{\partial t \partial x}E(\mathcal{O}_t(\boldsymbol{A}_x)),$$
$$\frac{\partial^2}{\partial x \partial t}P_t(\boldsymbol{A}_x) = \frac{\partial^2}{\partial t \partial x}P_t(\boldsymbol{A}_x).$$

Taking $\frac{\partial}{\partial x}$ in (10.26) we obtain

$$\frac{\partial^2}{\partial x \partial t}E(\mathcal{I}_t(\boldsymbol{A}_x)) = \frac{\partial^2}{\partial t \partial x}E(\mathcal{O}_t(\boldsymbol{A}_x)) + \frac{\partial^2}{\partial t \partial x}P_t(\boldsymbol{A}_x). \tag{10.27}$$

State Space $\boldsymbol{S} \subseteq \boldsymbol{R}^n$

Let $\{\boldsymbol{X}(t), t \geq 0\}$ denote a continuous-time, continuous-state stochastic process with n-dimensional state space $\boldsymbol{S} \subseteq \boldsymbol{R}^n$. The state space may be discrete or continuous. Let vector $\boldsymbol{x} = (x_1, ..., x_n)$, and state-space set $\boldsymbol{A}_{\boldsymbol{x}} = \cap_{i=1}^n (-\infty, x_i] \subseteq \boldsymbol{S}$. Then $P_t(\boldsymbol{A}_{\boldsymbol{x}}) = F_t(\boldsymbol{x}) = F_t(x_1, ..., x_n)$ is the joint cdf of the n state variables at time $t \geq 0$.

From the general result (10.25) the joint cdf is given by

$$F_t(\boldsymbol{x}) = E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x)) + F_0(\boldsymbol{x})$$

where $F_0(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{X}(0) \in \boldsymbol{A}_x, \\ 0 & \text{if } \boldsymbol{X}(0) \notin \boldsymbol{A}_x. \end{cases}$

Provided the derivatives exist, we obtain

$$\frac{\partial F_t(\boldsymbol{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right], i = 1, ..., n,$$
$$\frac{\partial^n F_t(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n} = \frac{\partial^n}{\partial x_1 \cdots \partial x_n}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right],$$
$$\frac{\partial F_t(\boldsymbol{x})}{\partial t} = \frac{\partial}{\partial t}\left[E(\mathcal{I}_t(\boldsymbol{A}_x)) - E(\mathcal{O}_t(\boldsymbol{A}_x))\right].$$

If $\frac{\partial E(\mathcal{I}_t(\boldsymbol{A}_x))}{\partial t}$, $\frac{\partial E(\mathcal{O}_t(\boldsymbol{A}_x))}{\partial t}$ can be expressed as functions of $F_t(\boldsymbol{x})$ or $f_t(\boldsymbol{x})$, then we may be able to derive an integro-differential equation for $F_t(\boldsymbol{x})$ or $f_t(\boldsymbol{x})$.

If $n = 1$ the state space is one-dimensional. We get $\boldsymbol{A}_x = (-\infty, x]$. Thus

$$f_t(x) = \frac{\partial}{\partial x}\left[E(\mathcal{I}_t((-\infty, x])) - E(\mathcal{O}_t((-\infty, x]))\right]$$

where $f_t(x)$ represents the transient pdf of $\boldsymbol{X}(t)$.

LC Computation

The expressions in this subsection can aid in estimating or computing the transient cdf and pdf of an n-dimensional continuous-parameter process using level crossing estimation or computation (LCE) for transient distributions. We will not expound on this transient LCE technique further in this monograph. Remarks 3.6 and 9.2 briefly discuss the technique.

10.4 Discrete-Parameter Processes

Let $\{X_n, n = 0, 1, 2, ...\}$ denote a discrete-parameter process taking values in a state space $\boldsymbol{S}$, which may be discrete or continuous. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be (measurable) subsets of $\boldsymbol{S}$. Let $P_n(\boldsymbol{A}) = P(X_n \in \boldsymbol{A})$ and $P_{m,n}(\boldsymbol{B}, \boldsymbol{C}) = P(X_m \in \boldsymbol{B}, \boldsymbol{X}_n \in \boldsymbol{C})$.

Definition 10.1 *The SP* ***exits*** *set* $\mathbf{A}$ *at time* n *if* $X_n \in \mathbf{A}$ *and* $X_{n+1} \notin \mathbf{A}$.
The SP ***enters*** *set* $\mathbf{A}$ *at time* n *if* $X_{n-1} \notin \mathbf{A}$ *and* $X_n \in \mathbf{A}$.
$\mathcal{I}_n(\mathbf{A})) =$ ***number of*** *SP entrances into* $\mathbf{A}$ *during* $[0, n]$.
$\mathcal{O}_n(\mathbf{A}) =$ ***number of*** *SP exits out of* $\mathbf{A}$ *during* $[0, n]$.

We state a theorem for discrete-time processes which is analogous to Theorem B.

Theorem 10.1 *Let* $\{X_n, n = 0, 1, 2, ...\}$ *be a discrete-time process with state space* $\boldsymbol{S}$. *Let* $\mathbf{A} \subseteq \boldsymbol{S}$.

$$E(\mathcal{I}_n(\mathbf{A})) = E(\mathcal{O}_n(\mathbf{A})) + P_n(\mathbf{A}) - P_0(\mathbf{A}). \tag{10.28}$$

where $P_0(\mathbf{A}) = \begin{cases} 1 \text{ if } X_0 \in A, \\ 0 \text{ if } X_0 \notin A. \end{cases}$

Proof. The proof is similar to that of Theorem 4.1 in Chapter 4, upon replacing t by n. ∎

10.4.1 Application to Markov Chains

Let $\{X_n, n = 0, 1, ...\}$ be a Markov chain with the discrete state space $\boldsymbol{S}$. For example, let $\boldsymbol{S} = \{0, \pm 1, \pm 2, ...\}$. Let the set $\boldsymbol{A} = j \in \boldsymbol{S}$. Then

$$E(\mathcal{I}_n(j)) = \sum_{i \neq j} \sum_{m=0}^{n-1} P_i^m P_{ij}, \quad \text{and} \quad E(\mathcal{O}_n(j)) = \sum_{i \neq j} \sum_{m=0}^{n} P_j^m P_{ji},$$

where P_{ij} is the one-step transition probability from i to j and $P_j^m \equiv P_m(\boldsymbol{A}) = P_m(j)$. Substituting into (10.28) gives

$$P_j^n = \sum_{i \neq j} \sum_{m=0}^{n-1} P_i^m P_{ij} - \sum_{i \neq j} \sum_{m=0}^{n} P_j^m P_{ji} + P_j^0. \tag{10.29}$$

Assume the following limiting probabilities exist:

$$\lim_{n \to \infty} P_{ij}^n = \lim_{n \to \infty} P_{jj}^n = \lim_{n \to \infty} P_j^n \equiv \pi_j,$$

where P_{ij}^n is the n-step transition probability from i to j. That is, the chain is positive recurrent and aperiodic. Note that $\sum_{j \in \boldsymbol{S}} \pi_j = 1$. Dividing both sides of (10.29) by n and letting $n \to \infty$ yields

$$\begin{aligned}
\lim_{n \to \infty} \frac{P_j^n}{n} &= \sum_{i \neq j} \left(\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} P_i^m \right) P_{ij} \\
&\quad - \sum_{i \neq j} \left(\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n} P_j^m \right) P_{j,i} + \lim_{n \to \infty} \frac{P_j^0}{n}, \\
0 &= \sum_{i \neq j} \pi_i P_{ij} - \sum_{i \neq j} \pi_j P_{ji} + 0, \\
\sum_{i \neq j} \pi_j P_{ji} &= \sum_{i \neq j} \pi_i P_{ij}, \\
\pi_j (1 - P_{jj}) &= \sum_{i \neq j} \pi_i P_{ij}, \\
\pi_j &= \sum_{i \in \boldsymbol{S}} \pi_i P_{ij}, j \in \boldsymbol{S}.
\end{aligned}$$

Thus we have derived the classical equations for the limiting probabilities $\{\pi_j\}$ by using an LC method, namely

$$\begin{aligned}
&\pi_j = \textstyle\sum_{i \in \boldsymbol{S}} \pi_i P_{ij}, j \in \boldsymbol{S}, \\
&\textstyle\sum_{j \in \boldsymbol{S}} \pi_j = 1.
\end{aligned} \tag{10.30}$$

Remark 10.5 *We have applied the discrete-time analog of Theorem B to a standard Markov chain in order to demonstrate its applicability to discrete-time discrete-state models. Note that Theorem B emphasizes the* ***system point aspect*** *of the SPLC method. SPLC utilizes SP entrance/exit rates of state-space sets. (SP level crossings are special cases of SP entrances and exits.)*

10.5 Semi-Markov Process

Consider a semi-Markov process (SMP) $\{X(t), t \geq 0\}$, with discrete state space $\boldsymbol{S}$ (also called a Markov renewal process). Let the sojourn time in state $j \in \boldsymbol{S}$ have a general distribution with mean $\mu_j > 0$. The type of distribution of the sojourn time may differ from state to state; only the means are utilized in this analysis. At the end of a sojourn in state i, say instant τ^-, assume $P\{X(t) = j|X(t^-) = i) = P_{ij}, j \neq i,\ j \in \boldsymbol{S}$. The matrix $\|P_{ij}\|$ is a Markov matrix. Assume the Markov chain with transition matrix $\|P_{ij}\|$ is positive recurrent and aperiodic so that the limiting probabilities $\pi_j, j \in \boldsymbol{S}$ exist.

Let $P_j(t) = P(X(t) = j), t \geq 0$; $P_j = \lim_{t\to\infty} P_j(t), j, j \in \boldsymbol{S}$. We shall derive the probabilities $P_j, j \in \boldsymbol{S}$, by using SPLC.

Consider a sample path of $\{X(t)\}$. Let $T_t(i)$ denote the total time spent by the SP in state i during $(0, t)$. Then

$$E(T_t(i)) = \int_{s=0}^{t} P_i(s)ds. \tag{10.31}$$

The expected number of SP exits from state i during $(0, t)$ is $\frac{E(T_t(i))}{\mu_i}$ since the mean of each sojourn time in i is μ_i. The expected number of SP $i \to j$ transitions during $(0, t)$ is $\frac{E(T_t(i))}{\mu_i} P_{ij}$. The expected total number of SP transitions into (*entrances* into) state j during $(0, t)$ is

$$E(\mathcal{I}_t(j)) = \sum_{i \neq j} \frac{E(T_t(i))}{\mu_i} P_{ij}. \tag{10.32}$$

By a similar argument, the expected number of SP *exits* out of j during $(0, t)$ is

$$E(\mathcal{O}_t(j)) = \frac{E(T_t(j))}{\mu_j}. \tag{10.33}$$

Substituting from (10.32) and (10.33) into Theorem B (10.23) gives

$$\sum_{i\neq j} \frac{E(T_t(i))}{\mu_i} P_{ij} = \frac{E(T_t(j))}{\mu_j} + P_j(t) - P_j(0). \tag{10.34}$$

(We assume the interchange of summation and the limit operation is valid. This applies if, e.g., $\boldsymbol{S}$ is finite.)

From (10.31), the proportion of time the SP is in state i is

$$\lim_{t\to\infty} \frac{E(T_t(i))}{t} = P_i, i \in \boldsymbol{S}.$$

Also

$$\lim_{t\to\infty} \frac{P_j(t)}{t} = \lim_{t\to\infty} \frac{P_j(0)}{t} = 0,$$

since $0 \leq P_j(t) \leq 1, t \geq 0$. We divide (10.34) by $t > 0$ and let $t \to \infty$. This gives

$$\sum_{i\neq j} \frac{P_i}{\mu_i} P_{ij} = \frac{P_j}{\mu_j}, j \in \boldsymbol{S} \tag{10.35}$$

Suppose $\sum_{j\in S} \frac{1}{\mu_j} P_j = K > 0$. Then $\sum_{j\in S} \left(\frac{1}{K\mu_j} P_j\right) = 1$. Dividing (10.35) by K and transposing terms gives the system of equations for $\{P_i\}$,

$$\begin{aligned} \tfrac{1}{K\mu_j} P_j &= \textstyle\sum_{i\neq j} \left(\tfrac{1}{K\mu_j} P_i\right) P_{ij}, j \in \boldsymbol{S} \\ \textstyle\sum_{j\in S} \left(\tfrac{1}{K\mu_j} P_j\right) &= 1. \end{aligned} \tag{10.36}$$

The system of equations (10.36) for $\left\{\left(\frac{1}{K\mu_j} P_j\right)\right\}$ is identical to the system (10.30) for $\{\pi_j\}$. Thus

$$\begin{aligned} \frac{1}{K\mu_j} P_j &= \pi_j, j \in \boldsymbol{S}, \\ P_j &= (\pi_j \mu_j) K, j \in \boldsymbol{S}. \end{aligned} \tag{10.37}$$

We obtain K from the normalizing condition

$$\sum_{j\in S} P_j = K \sum_{j\in S} \pi_j \mu_j = 1,$$

namely

$$K = \frac{1}{\sum_{j\in S} \pi_j \mu_j}. \tag{10.38}$$

Substituting from (10.37) into (10.38) gives the well known formula

$$P_j = \frac{\pi_j \mu_j}{\sum_{j \in \boldsymbol{S}} \pi_j \mu_j}, \ j \in \boldsymbol{S}. \tag{10.39}$$

The key steps in this SPLC derivation of (10.39) are: (1) obtain expressions for the expected SP entrance and exit rates of a state; (2) apply formula (10.23) of Theorem B; (3) divide by t and take $\lim_{t \to \infty}$; (4) evaluate the constant K by recognizing the role of the **linear Markov-chain equations** (10.30) for $\{\pi_j\}$.

10.6 Non-homogeneous Pure Birth Processes

Let $\{X(t), t \geq 0\}$ denote the number of births during $(0,t), t > 0$. Let $X(0) = i$, where i is a non-negative integer. Consider the sequence of positive functions (birth rates) $\{\lambda_k(t), k = i, i+1, ...; i = 0, 1, ...\}$ such that

$$P(X(t+h) - X(t) = 1 | X(t) = k) = \lambda_t(k)h + o(h),$$
$$P(X(t+h) - X(t) = 0 | X(t) = k) = 1 - \lambda_t(k)h + o(h),$$

where $h > 0$.

Define $P_n(t) = P(X(t) = n)$. We shall compute $P_n(t), t > 0, n = 0, 1, 2, ...$; by utilizing Theorem B, i.e., (10.23) and (10.24).

The expected number of SP entrances into state i during $(0,t)$ is $E(\mathcal{I}_t(i)) = 0$, since $X(0) = i$, and $X(\cdot)$ never returns to i, once it increases from i to $i+1$. On the other hand the expected number of SP exits out of state i during $(0,t)$ is $E(\mathcal{O}_t(i)) = \int_{s=0}^{t} \lambda_s(i) P_i(s) ds$, since an SP $i \to i+1$ exit can occur at any instant $s \in (0,t)$. Note that $P_i(0) = 1$. Substituting $E(\mathcal{I}_t(i))$, $E(\mathcal{O}_t(i))$, $P_i(0)$ into (10.23), we obtain

$$0 = \int_{s=0}^{t} \lambda_s(i) P_i(s) ds + P_i(t) - 1. \tag{10.40}$$

Differentiating (10.40) with respect to t gives

$$\frac{d}{dt} P_i(t) + \lambda_t(i) P_i(t) = 0$$

having solution

$$P_i(t) = e^{-m_t(i)}, t \geq 0, \tag{10.41}$$

where $m_t(i) = \int_{s=0}^{t} \lambda_s(i) ds$, since $P_i(0) = 1$.

Next, consider an arbitrary state $j > i$. Then

$$E(\mathcal{I}_t(j)) = \int_{s=0}^{t} \lambda_{j-1}(s)P_{j-1}(s)ds, \tag{10.42}$$

$$E(\mathcal{O}_t(j)) = \int_{s=0}^{t} \lambda_j(s)P_j(s)ds. \tag{10.43}$$

Substituting from (10.42) and (10.43) into (10.23) gives

$$\int_{s=0}^{t} \lambda_s(j-1)P_{j-1}(s)ds = \int_{s=0}^{t} \lambda_s(j)P_j(s)ds + P_j(t) - 0. \tag{10.44}$$

Taking $\frac{d}{dt}$ in (10.44) yields

$$\frac{d}{dt}P_j(t) + \lambda_t(j)P_j(t) = \lambda_t(j-1)P_{j-1}(t),$$

with solution

$$P_j(t) = e^{-m_t(j)} \int_{s=0}^{t} e^{m_s(j)}\lambda_s(j-1)P_{j-1}(s)ds. \tag{10.45}$$

Formula (10.45) provides a recursive solution for $P_j(t), j = i, i+1, ...$.

10.6.1 Non-homogeneous Poisson Process

The non-homogeneous Poisson process is a special case of the pure growth process. Assume $X(0) = 0$, $\lambda_t(j) \equiv \lambda_t$ independent of the state, so that $m(t) = \int_{s=0}^{t} \lambda_s ds$. Setting $i = 0$ gives $P_0(t) = e^{-m(t)}$. From (10.45) we obtain (by induction) the well known formula

$$P_n(t) = e^{-m(t)}\frac{(m(t))^n}{n!}, n = 0, 1, 2, ... \tag{10.46}$$

Formula (10.46) is a Poisson distribution with mean $m(t)$. The $\{P_n(t)\}$ for the standard Poisson process are obtained by letting $\lambda_t \equiv \lambda$, so that $m(t) \equiv \lambda t$.

10.6.2 Yule Process

The Yule process is a special case of the pure growth process. Assume $X(0) = 1$ and $\lambda_t(i) = i\lambda, t \geq 0, i = 1, 2, ...$. Thus the growth rate is directly proportional to the current population, but independent of t. Then $P_1(t) = e^{-\lambda t}$ (= probability of no births in $(0, t)$). Using (10.45) and mathematical induction, we obtain the well known geometric distribution for the Yule process

$$P_n(t) = (1 - e^{-\lambda t})^{n-1} e^{-\lambda t}, n = 1, 2, ... \quad (10.47)$$

For completeness, we include the probability $P_{ik}(t)$ that i *independent* Yule processes with the same parameter λ, yield a sum equal to $k \geq i$ at time $t > 0$ (total number of individuals $= k$ at time t). Assume each process starts in state 1 at time 0. Since $P_n(t)$ in (10.47) is a geometric distribution, we obtain a negative binomial distribution

$$P_{ik}(t) = \binom{k-1}{i-1} e^{-i\lambda t}(1 - e^{-\lambda t})^{k-i}, k = i, i+1, ... \quad (10.48)$$

Formula (10.48) can be derived in several ways (e.g., [74], [91]). We shall outline a direct proof using LC.

We derive in a similar manner as for (10.45),

$$P_{ik}(t) = (k+1)\lambda e^{-k\lambda t} \int_{s=0}^{t} e^{k\lambda s} P_{i,k-1}(s) ds + C_k e^{-k\lambda t}, k \geq i, \quad (10.49)$$

where $C_k = \begin{cases} 1 \text{ if } k = i, \\ 0 \text{ if } k > i. \end{cases}$ Now, $P(\text{no births in } (0,t)) = P(\mathrm{E}_{i\lambda} > t)$ where $\mathrm{E}_{i\lambda}$ is an exponentially distributed r.v. with mean $\frac{1}{i\lambda}$. Hence

$$P_{ii} = e^{-i\lambda t}. \quad (10.50)$$

Thus (10.48) holds for $k = i$. From (10.50) and (10.49) with $k = i + 1$, we obtain

$$P_{i,i+1}(t) = ie^{-i\lambda t}\left(1 - e^{-\lambda t}\right) = \binom{i+1-1}{i-1} e^{-i\lambda t}\left(1 - e^{-\lambda t}\right). \quad (10.51)$$

Therefore (10.48) holds for $k = i + 1$.

Assume (10.48) holds for an arbitrary integer $k > i$. We then show using (10.49) that it holds for $k+1$. Hence it holds for all $k = i, i+1, ...$, by mathematical induction.

10.7 Revisit of Transient M/G/1 Queue

We very briefly revisit the transient M/G/1 queue of Section 3.2. It is readily proved by a slight generalization of the proofs in Section 3.2, that the theory holds for models where the arrival rate λ and cdf of service time $B(x)$ depend on time. Denote them by λ_t and $B_t(x), x \geq 0$, respectively. We obtain

$$
\begin{aligned}
f_t(x) &= \tfrac{\partial}{\partial t} F_t(x) + \lambda_t \overline{B}_t(x) P_0(t) \\
&\qquad + \lambda_t \int_{y=0}^{x} \overline{B}_t(x-y) f_t(y) dy, x > 0, \qquad (10.52) \\
f_t(0) &= \tfrac{\partial}{\partial t} P_0(t) + \lambda_t P_0(t).
\end{aligned}
$$

The solution of the differential equation for $P_0(t)$ in (10.52) is

$$P_0(t) = e^{-m(t)} \int_{s=0}^{t} e^{m(s)} f_s(0) ds + P_0(0) e^{-m(t)}, \qquad (10.53)$$

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

10.8 Pharmacokinetic Model

We outline an LC approach to pharmacokinetics with a brief discussion of a simplified one-compartment model. We assume bolus dosing, i.e., a full dose is absorbed into the blood stream immediately at a dosing instant. Also, inter-dose times are $\underset{dist}{=} \mathrm{E}_\alpha$. Thus doses occur in a Poisson process at rate λ. This assumption is valid outside of a controlled environment. Statistical tests have shown that many patients take certain medications non-uniformly over time in a Poisson process [33]. We first suppose the dose amounts are deterministic of size D.

We assume first-order kinetics. That is, the concentration of the drug in the blood stream decays at a rate which is proportional to the concentration. This is equivalent to a plot of the concentration over time having a negative exponential shape between doses (similar to Fig. 10.2).

This model is equivalent to an M/D/r($\cdot$) dam (Section 6.2). Let $W(t), t \geq 0$, denote the drug concentration at time t. Let the dose times be $\{\tau_n\}$. $\tau_n < \tau_{n+1}, n = 0, 1, 2, ...$, where $\tau_0 \equiv 0$. The decay rate is

$$\frac{dW(t)}{dt} = -kW(t), \tau_n \leq t < \tau_{n+1}, n = 0, 1, 2, ..., \qquad (10.54)$$

where $k > 0$. The dimension of the concentration $W(t)$ is $[W(t)] = \left[\frac{Mass}{Volume}\right]$; $\left[\frac{dW(t)}{dt}\right] = \left[\frac{Mass}{Volume}\right] \cdot \left[Time^{-1}\right]$; $[k] = [Time]^{-1}$.

Let $f(x), x > 0$ denote the steady-state pdf of concentration. The steady-state probability that the concentration is zero, is equal to 0. This is because a sample path never declines to level 0 once dosing begins, due to the negative exponential shape of the decay. In theory, the concentration of the drug never vanishes. In practice, it goes to 0 or is negligible. (We are not discussing the treatment effects of dosing; only the concentration dynamics.)

10.8.1 Equation for PDF of Concentration

Consider a sample path of $\{W(t)\}$. Fix level $x > 0$ (Fig. 10.2). The SP downcrossing rate of level x is $kxf(x)$. The SP upcrossing rate of x is equal to $\lambda F(x) - \lambda F(x - D)$ (see Section 3.8). Rate balance across x gives an equation for$f(x)$ and $F(x)$, namely

$$kxf(x) = \lambda F(x) - \lambda F(x - D), x > 0. \tag{10.55}$$

In integral equation (10.55) for the $F(\cdot)$, note that $F(x - D) = 0$ for $x \in (0, D)$. Also

$$\frac{f(x)}{F(x)} = \frac{d \ln F(x)}{dx} = \frac{\lambda}{kx},$$

with solution

$$F(x) = Ax^{\frac{\lambda}{\kappa}}, x \in (0, D), \tag{10.56}$$

where A is a positive constant. The solution for $F(x)$ on the state-space intervals $[iD, (i + 1)D), i = 1, 2, ...,$ can be obtained by an iteration procedure (not carried out here). We add that $F(x)$ is continuous for all $x > 0$. This continuity property helps to solve for $F(x)$ on successive state-space intervals $[iD, (i + 1)D), i = 1, 2, ...$, in terms of A. The constant A in (10.56) is then determined using the normalizing condition $F(\infty) = 1$. Once $F(x)$ is obtained, we can determine $f(x)$ by substituting into (10.55) (as in Section 3.8). Alternatively, we may solve for $f(x)$ using LC estimation, or a *hybrid* LC estimation procedure since we have a partial analytical solution in (10.56) (see Section 9.6).

10.8.2 Exponentially Distributed Doses

We may rationalize a model using exponentially distributed doses if the amount absorbed is affected by the dosing environment (e.g., acidity, presence of enzymes, interaction with other medications, etc.). Another

M/G/r(·) Dam	Pharmacokinetic Model
Input instant	Bolus dose instant
Input amount (jump size)	Dose amount (jump size)
Content $W(t), t \geq 0$	Concentration $W(t), t \geq 0$
Sample-path slope $-r(x), x > 0$	Sample-path slope $-r(x), x > 0$
CDF/PDF of content	CDF/PDF of concentration
Mean content	Average drug concentration
Variance of content	Variance of concentration

Table 10.1: M/G/r(·) Dam versus Pharmakokinetic model

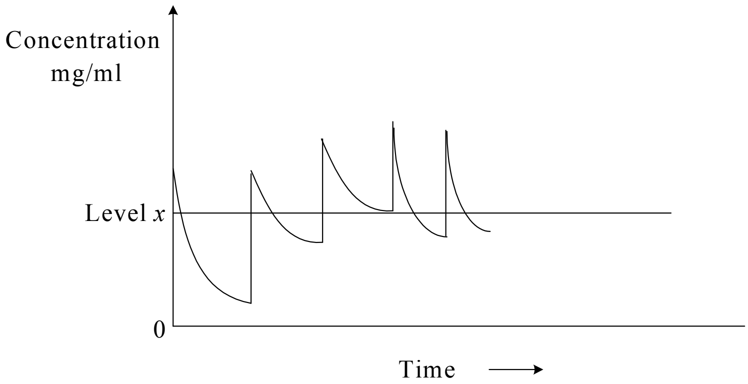

Figure 10.2: Sample path of drug concentration in one-compartment model with bolus dosing and first-order kinetics

instance could occur when eye drops are instilled by a patient, say approximately every six hours. The sizes of the individual drops may vary considerably, due to usually using a hand-squeezed container. The location on the cornea of the instillation may vary from dose to dose, thereby affecting absorption. This could create random increases in concentration with the successive doses during a dosing regime. Similar remarks apply to fast-acting sprays, such as nitrolingual pump sprays, or to nasal sprays. Also, for certain drugs it may be feasible to randomize dose sizes as an exponential random variable inherently in a prescription. Such randomization may tend to decrease variability in the long run concentration during the dosing regime.

Assume the bolus dose amounts are random, distributed as E_μ. Then the equation for the pdf of concentration is

$$kxf(x) = \lambda \int_{y=0}^{x} e^{-\mu(x-y)} f(y) dy. \tag{10.57}$$

Equation (10.57) has the solution

$$f(x) = \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)}(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}\mu, x > 0. \tag{10.58}$$

where $\Gamma(\cdot)$ is the Gamma function (see Section 6.4). Let W denote the steady-state concentration. The mean and second moment of W are

$$E(W) = \frac{\lambda}{k\mu},\ E(W^2) = \frac{\lambda}{k\mu^2}\left(\frac{\lambda}{k}+1\right).$$

The variance of W is

$$Var(X) = E(W^2) - (E(W))^2 = \frac{\lambda}{k\mu^2}.$$

We can find the probability that the steady-state concentration is between two threshold limits, say $\alpha < \beta$, using

$$P(\alpha < \text{concentration} < \beta) = \int_{x=\alpha}^{\beta} \frac{1}{\Gamma\left(\frac{\lambda}{k}\right)}\mu(\mu x)^{\left(\frac{\lambda}{k}-1\right)}e^{-\mu x}dx. \tag{10.59}$$

The information in (10.59) may be useful when dosing continues for a long time, e.g., when administering the blood thinner coumadin. If the concentration is $< \alpha$ coumadin is not effective for the intended treatment. If the concentration is $> \beta$ the blood becomes too thin.

The type of analysis outlined briefly here can be extended to various pharmacokinetic models of varying complexity.

Remark 10.6 *We mention in passing that it is possible to apply Theorem B to compute the* ***time-dependent pdf and cdf of concentration*** *(see formulas (10.23) - (10.26)). Knowledge of transient distributions may be useful in dosing regimes where it important to estimate the concentration after a short dosing duration.*

Remark 10.7 *Some related stochastic models have characteristics in common with the pharmacokinetic model. One group of models involves consumer response (CR) to non-uniform advertisements [30]. Such models can be analyzed along similar lines, using LC.*

10.9 Counter Models

We consider the transient total output of type-1 and type-2 counters. We first treat a type-2 counter.

10.9.1 Type-2 Counter

Consider a type-2 counter. Electrical pulses arrive in a Poisson process at rate λ. Each *arriving* pulse is followed immediately by a fixed *locked* period of length $D > 0$, during which new arrivals cannot be detected by the counter. If a new arrival occurs at a time t when the counter is locked, then the locked period is extended to time $t + D$. Thus the locked time "telescopes". Assume the locked periods are $\underset{dist}{=} L$; note that $L \geq D$. Arrivals can be detected only when the counter is unlocked or *free*. Assume that the counter is free at time 0.

Let the amplitudes of the pulses be $\underset{dist}{\equiv} X$, having cdf $B(y), y > 0$. Let $\eta_i(t), t \geq \tau_i$, denote the output at time t due to the *detected pulse* X_i occurring at τ_i. Assume that $\eta_i(t)$ dissipates at rate

$$\frac{d\eta_i(t)}{dt} = -k \cdot \eta_i(t), t > \tau_i, \tag{10.60}$$

where the constant $k > 0$ is the same for all $i = 1, 2, ...$.

Let η_t denote the *total output* at time t, due to all *registered* pulses that arrive during $(0, t)$ (see Fig. 10.3). Then

$$\eta_t = \sum_{i=1}^{n} \eta_i(t), \tau_n \leq t < t_{n+1}, n = 1, 2, ..., \tag{10.61}$$

$$\frac{d}{dt}\eta_t = -k \sum_{i=1}^{n} \eta_i(t) = -k\eta_t, \tau_n \leq t < t_{n+1}, n = 1, 2, ... \tag{10.62}$$

Denote the cdf and pdf of η_t by $F_t(x)$ and $f_t(x) = \frac{d}{dt}F_t(x)$,$x > 0$, wherever the derivative exists.

10.9.2 Sample Path of Total Output

A sample path of the process $\{\eta_t, t \geq 0\}$ consists of segments that decay exponentially with decay constant k, between the τ_i's (Fig. 10.3). That is,

$$\eta_t = \sum_{i=1}^{n} X_i e^{-k(t-\tau_i)}, \tau_n \leq t < t_{n+1}, n = 1, 2, ... \tag{10.63}$$

Note that a sample path cannot descend to level 0 due to exponential decay.

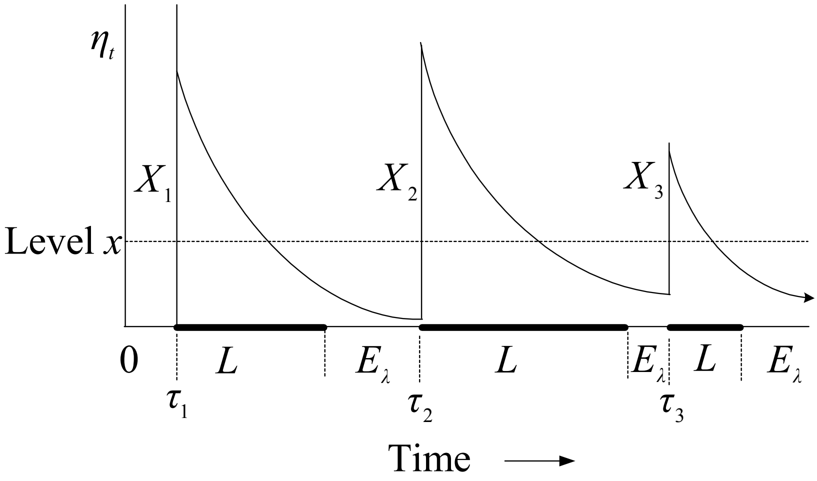

Figure 10.3: Sample path of total output η_t for type-2 counter model. Locked periods are each $\underset{dist}{=} L \geq D$. Arrivals during L are not detected, but extend the locked period. Arrival process of pulses is Poisson at rate λ.

Probability that the Counter is Free at Time t

Let $p(t) = P$(counter is free at time t). Then

$$p(t) = \begin{cases} e^{-\lambda t}, 0 < t < D, \\ e^{-\lambda D}, t \geq D. \end{cases} \tag{10.64}$$

The reason for (10.64) is that for $0 < t < D$, the counter is free at t iff there is no arrival in $(0, t)$, which has probability $e^{-\lambda t}$. For $t \geq D$, the counter is free at time t iff there has not been an arrival during the interval $(t - D, t)$. The probability of this event is $e^{-\lambda D}$, by the memoryless property of E_λ (see, e.g., [74]).

10.9.3 Integro-differential Equation for PDF of Output

Consider level $x > 0$ in the state space; and state-space set $A_x = (0, x]$. We can show as in Theorem 6.2.8, that for SP *entrances* into set A_x (downcrossings of level x)

$$\frac{\partial}{\partial t} E(\mathcal{I}_t(A_x)) = \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = kx f_t(x), t > 0. \tag{10.65}$$

For SP *exits* out of A_x (upcrossings of level x)

$$\begin{aligned}
&\tfrac{\partial}{\partial t}E(\mathcal{U}_t(x)) = \tfrac{\partial}{\partial t}E(\mathcal{O}_t(A_x)) \\
&= \begin{cases} \lambda e^{-\lambda t} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y)dy, x > 0, 0 < t < D, \\ \lambda e^{-\lambda D} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y)dy, x > 0, t \geq D. \end{cases}
\end{aligned} \tag{10.66}$$

Substituting (10.65) and (10.66) into Theorem B (noting that $\frac{\partial}{\partial t}F_t(x) = -\frac{\partial}{\partial t}(1 - F_t(x))$, we get integro-differential equations for the pdf $f_t(x)$,

$$\begin{aligned} kxf_t(x) = \lambda e^{-\lambda t} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y)dy - \tfrac{\partial}{\partial t}(1 - F_t(x)), \\ x > 0, 0 < t < D, \end{aligned} \tag{10.67}$$

$$\begin{aligned} kxf_t(x) = \lambda e^{-\lambda D} \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y)dy - \tfrac{\partial}{\partial t}(1 - F_t(x)), \\ x > 0, t \geq D, \end{aligned} \tag{10.68}$$

since the arrival rate is λ, and an arrival can be registered at time t iff the counter is unlocked or free at time t.

10.9.4 Expected Value of Total Output

We obtain the expected value of η_t by integrating both sides of (10.67) and (10.68) with respect to $x \in (0, \infty)$. (In (10.67) and (10.68), we assume that $\frac{\partial}{\partial t}F_t(x)$ is continuous with respect to $t > 0$. This condition is required to apply Fubini's Theorem on interchanging the operations $\int_{x=0}^{\infty}$ and $\frac{\partial}{\partial t}$.)

Upon integrating (10.67) we obtain

$$\begin{aligned}
kE(\eta_t) &= \lambda e^{-\lambda t}E(X) - \frac{\partial}{\partial t}E(\eta_t), \\
\frac{\partial}{\partial t}e^{kt}E(\eta_t) &= \lambda e^{(k-\lambda)t}E(X), \\
E(\eta_t) &= \frac{\lambda e^{-\lambda t}E(X)}{k-\lambda} + Ae^{-kt}, 0 < t < D, (A \text{ constant}), \\
E(\eta_t) &= \frac{\lambda E(X)}{k-\lambda}\left(e^{-\lambda t} - e^{-kt}\right), 0 < t < D,
\end{aligned} \tag{10.69}$$

since $E(\eta_0) = 0$.

Integrating (10.68), we obtain

$$\begin{aligned} kE(\eta_t) &= \lambda e^{-\lambda D}E(X) - \frac{\partial}{\partial t}E(\eta_t), \\ \frac{\partial}{\partial t}e^{kt}E(\eta_t) &= \lambda e^{-\lambda D}E(X)e^{kt}, \\ E(\eta_t) &= \frac{\lambda e^{-\lambda D}E(X)}{k} + Ae^{-kt}, t \geq D, \end{aligned} \tag{10.70}$$

where the constant A is given by

$$A = \lambda E(X)\left(\frac{e^{-(\lambda-k)D}-1}{k-\lambda} - \frac{e^{-(\lambda-k)D}}{k}\right).$$

To obtain the value of A, we have used the fact that $\eta_{D^-} = \eta_D$ (see Fig. 10.3), which implies continuity of $E(\eta_t)$ at $t = D$ $(a.s.)$. Thus, from (10.69), $E(\eta_D) = \frac{\lambda E(X)}{k-\lambda}\left(e^{-\lambda D} - e^{-kD}\right)$.

If $t \to \infty$, then (10.70) reduces to

$$\lim_{t\to\infty} E(\eta_t) = \frac{\lambda e^{-\lambda D}E(X)}{k}.$$

If $D = 0$, then $A = -\frac{\lambda E(X)}{k}$. We then obtain $E(\eta_t) = \frac{\lambda E(X)}{k}\left(1 - e^{-kt}\right)$ and $\lim_{t\to\infty} E(\eta_t) = \frac{\lambda E(X)}{k}$, as in [74].

10.9.5 Type-1 Counter

A type-1 counter differs from a typre-2 counter (Subsection 10.9.1) only in the locking mechanism. In a type-1 counter, only *registered* arrivals when the counter is free, generate locked periods. Arrivals when the counter is locked, have no effect on the locked period. Thus every locked period has length $D > 0$. Aside from the locking mechanism, we generally use the same notation and assumptions for type-1 and type-2 counters.

Thus equations (10.60) - (10.63) hold for type-1 counters.

10.9.6 Sample Path of Total Output

A sample path of the process $\{\eta_t, t \geq 0\}$ consists of segments that decay exponentially with decay constant k, between the τ_i's (Fig. 10.4).

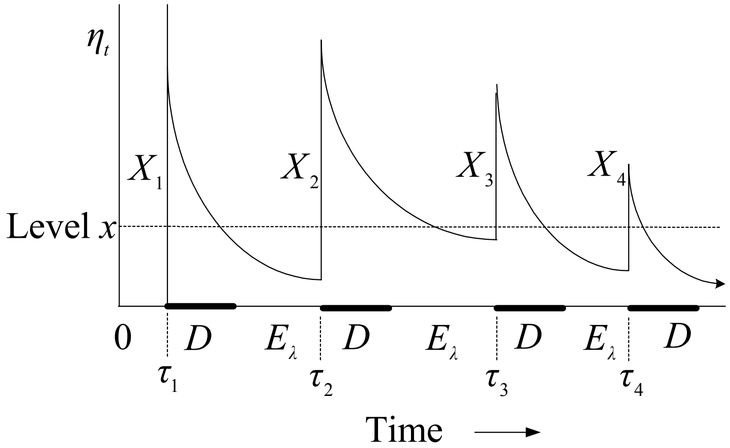

Figure 10.4: Sample path of total output η_t for type-1 counter model. Locked periods are each $= D$ (arrivals not detected therein, and have no effect on locked period) Arrival process of pulses is Poisson at rate λ.

Probability that the Counter is Free at Time t

The probability that the counter is free to register a newly arriving pulse at time t is given by the following recursion ([70]).

$$\begin{aligned}
p_1(t) &= e^{-\lambda t}, 0 < t < D, \\
p_2(t) &= e^{-\lambda(t-D)} p_1(D) + \frac{(\lambda(t-D))\, e^{-\lambda(t-D)}}{1!}, D \le t < 2D, \\
&\cdots \\
p_n(t) &= \sum_{j=1}^{n-1} \frac{(\lambda(t-(n-1)D))^{j-1} \cdot e^{-\lambda(t-(n-1)D)}}{(j-1)!} p_{n-j}((n-j)D) \\
&\quad + \frac{(\lambda(t-(n-1)D))^{n-1}\, e^{-\lambda(t-(n-1)D)}}{(n-1)!}, \\
&\qquad (n-1)\, D \le t < nD, n = 1, 2, \ldots\,, \qquad (10.71)
\end{aligned}$$

where $\sum_{j=1}^{0} \equiv 0$.

Remark 10.8 *Let $p(t) = P$(the counter is free at time t),$t \ge 0$. Then $\lim_{t\to\infty} p(t) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+D}$ (a known result for alternating renewal processes [49]). Hence we have proved using probability arguments that*

$$\lim_{n\to\infty} p_n(nD) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}+D},$$

where $p_n(nD)$ is the series obtained by substituting $t = nD$ in (10.71). More strongly, for every $\alpha \in [0,1]$,

$$\lim_{n\to\infty} p_n(\alpha(n-1)D + (1-\alpha)nD) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + D}.$$

10.9.7 Integro-differential Equation for PDF of Output

Consider level $x > 0$ in the state space; and state-space set $A_x = (0, x]$. We can show as in Theorem 6.2.8, that for SP entrances into set A_x,

$$\frac{\partial}{\partial t} E(\mathcal{I}_t(A_x)) = \frac{\partial}{\partial t} E(\mathcal{D}_t(x)) = kxf_t(x), t > 0. \tag{10.72}$$

For SP exits out of A_x,

$$\begin{aligned} \frac{\partial}{\partial t} E(\mathcal{O}_t(A_x)) &= \frac{\partial}{\partial t} E(\mathcal{U}_t(x)) \\ &= \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy, \\ & \qquad (n-1)\, D \le t < nD, n = 1, 2, \ldots\,. \end{aligned} \tag{10.73}$$

In (10.73), the factor $p_n(t)$ occurs because an arrival is registered iff it arrives when the counter is free.

Substituting (10.72) and (10.73) into Theorem B, we get an integro-differential equation for the pdf $f_t(x)$,

$$\begin{aligned} kxf_t(x) &= \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy + \frac{\partial}{\partial t} F_t(x), x > 0, \\ kxf_t(x) &= \lambda p_n(t) \cdot \int_{y=0}^{x} \overline{B}(x-y) f_t(y) dy, -\frac{\partial}{\partial t}(1 - F_t(x)), x > 0, \\ & \qquad (n-1)\, D \le t < nD, n = 1, 2, \ldots\,. \end{aligned} \tag{10.74}$$

10.9.8 Expected Value of Total Output

We obtain the expected value of η_t by integrating both sides of the integral equations (10.74) with respect to $x \in (0, \infty)$. We obtain

$$E(\eta_t) = \frac{\lambda E(X)}{k - \lambda} \left(e^{-\lambda t} - e^{-kt} \right), 0 < t < D \tag{10.75}$$

in the same manner as (10.69). Similarly, we can obtain $E(\eta_t), nD \le t < (n+1)\, D, n = 1, 2, \ldots$. (We shall not carry out this computation here.)

Remark 10.9 *If the locked period has value $D = 0$, then $p_n(t) = 1, n = 1, 2, \ldots$. Then every arrival is registered. We then obtain the known result $E(\eta_t) = \frac{\lambda E(X)}{k}\left(1 - e^{-kt}\right), t > 0$ (e.g., [74]).*

If $t \to \infty$, then (10.75) reduces to $\lim_{t\to\infty} E(\eta_t) = \frac{\lambda E(X)}{k}$.

Remark 10.10 *When there is no locked time ($D = 0$), the foregoing type-1 and type-2 counter models coincide with an M/G/r(·) dam with efflux rate proportional to content. Thus, results for a dam with $r(x) = kx, x > 0$, can be derived as a special case of either counter model.*

10.10 A Dam with Alternating Influx and Efflux

Consider a dam in which the content alternates between random periods of continuous influx and continuous efflux. We arbitrarily classify periods of emptiness as being parts of periods of efflux, for notational convenience. Periods of efflux are $\underset{dist}{=} \mathrm{E}_{\lambda_1}$ and periods of efflux are $\underset{dist}{=} \mathrm{E}_{\lambda_2}$. Let $W(t) \geq 0$ denote the content of the dam at time $t \geq 0$. Assume that during an influx period, the rate of *increase* of content is $\frac{dW(t)}{dt} = +q(W(t))$, where $q(x) > 0, x > 0$. Assume that during an efflux period, the rate of *decrease* of content is $\frac{dW(t)}{dt} = -r(W(t))$, where $r(x) > 0, x > 0$. When the dam is empty (i.e., $W(t) = 0$), $\frac{dW(t)}{dt} = 0$. By the memoryless property of E_{λ_2}, sojourns at level 0 are also distributed as E_{λ_2} (Fig. 10.5). The empty period is analogous to an idle period in an M/G/1 queue or empty period in an M/G/r(·) dam. The efflux rate $r(x)$ is similar to that of the M/G/r(·) dam (Section 6.2).

Consider the stochastic process $\{W(t), M(t)\}$ where $W(t)$ denotes the content at instant t, and the configuration $M(t) \in \boldsymbol{M} = \{0, 1, 2\}$. The state space is $\boldsymbol{S} = [0, \infty) \times \boldsymbol{M}$. The meaning of $M(t)$ is given in the following table.(See Subsections 4.5 – 4.6 for discussions on system configuration.)

$M(t)$	Meaning
0	Empty period.
1	Influx phase; content increasing.
2	Efflux phase; content decreasing or at level 0.

A sample path of $\{W(t), M(t)\}$ evolves on two sheets corresponding to configurations 1 and 2, and on one line corresponding to an empty period ($W(t) = 0$) (Fig. 10.6).

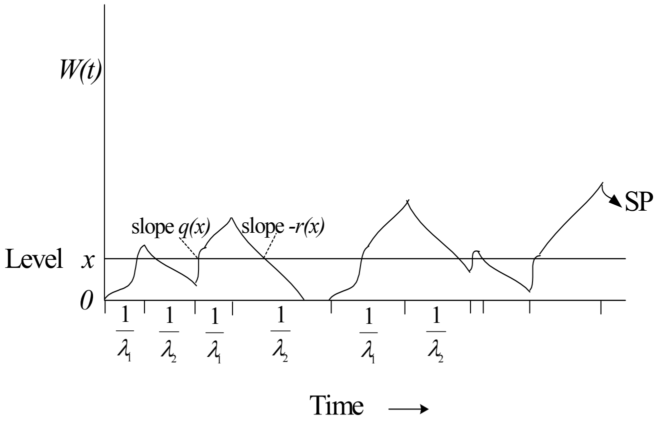

Figure 10.5: Sample path of dam with continuous influx and efflux. Slope at level x: during influx is $\frac{d}{dt}W(t) = q(x)$; during efflux is $-r(x)$. Slope at level 0 is $\frac{d}{dt}W(t) = 0$. Influx and efflux times are distributed as E_{λ_1}, E_{λ_2}, respectively.

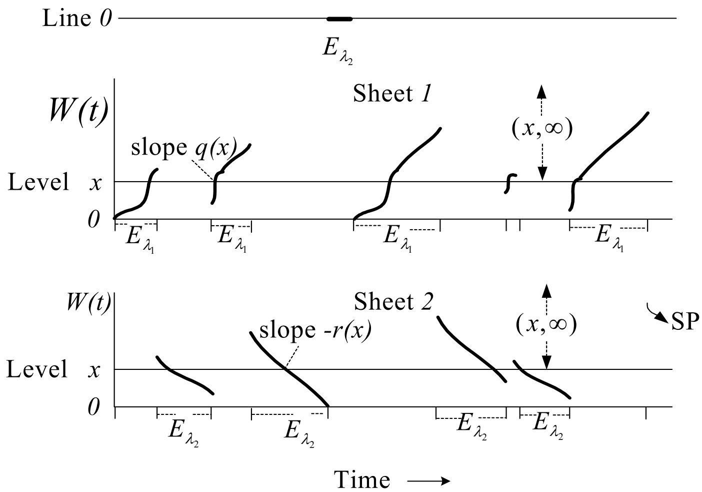

Figure 10.6: Sample path of dam with continuous influx and efflux, showing line and sheets (pages). Line $0 \leftrightarrow W(t) = 0$, dam empty. Sheet $1 \leftrightarrow M(t) = 1$, influx phase. Sheet $2 \leftrightarrow M(t) = 2$, efflux phase. Also indicates composite states $\langle(x,\infty), i\rangle, i = 1, 2$. Slope at level $x > 0$: during influx is $\frac{d}{dt}W(t) = q(x)$; during efflux is $-r(x)$. Slope at level 0 is $\frac{d}{dt}W(t) = 0$. Influx and efflux durations are distributed as E_{λ_1}, E_{λ_2}, respectively. Empty duration is distributed as E_{λ_2}.

10.10.1 Steady-state PDF of Content

Denote the "partial cdf's" of content by

$$F_i(x) = \lim_{t\to\infty} P(W(t) \leq x, M(t) = i), x > 0, i = 1, 2.$$

Denote the steady-state "partial" pdf of content by

$$f_i(x) = \frac{d}{dx}F_i(x), i = 1, 2, x > 0,$$

wherever the derivative exists.

The *total* pdf of content (marginal pdf) is

$$f(x) = f_1(x) + f_2(x), x > 0. \tag{10.76}$$

Let $P_0 = \lim_{t\to\infty} P(W(t) = 0)$. We shall derive: $f_i(x), i = 1, 2$; $f(x)$; P_0; $F(x) = P_0 + \int_{y=0}^{x} f(y)dy$, in terms of the input parameters λ_1, λ_2, $q(x)$, $r(x)$. The steady-state probability that the dam is in the influx phase $(i = 1)$ or efflux phase $(i = 2)$ is $F_i(\infty) = \int_{x=0}^{\infty} f_i(x)dx, i = 1.2$.

10.10.2 Equations for PDF's

Consider composite state $((x,\infty), 1)\,, x > 0$, on sheet 1. The SP rate *out* of $((x,\infty), 1)$ is $\lambda_1 \int_{y=x}^{\infty} f_1(y)dy$, since the end of an influx period signals an instantaneous SP $1 \to 2$ transition from $((x,\infty), 1)$ to $((x,\infty), 2)$ *at the same level*.

The SP rate *into* $((x,\infty), 1)$ is

$$q(x)f_1(x) + \lambda_2 \int_{y=x}^{\infty} f_2(y)dy,$$

since: (1) the SP upcrosses level x on sheet 1 at rate $q(x)f_1(x)$, (2) the SP enters $((x,\infty), 1)$ from $((x,\infty), 2)$ $(2 \to 1$ transition) at the same level (the rate at which efflux periods end when the SP is in $((x,\infty), 2)$ is $= \lambda_2$). Set balance, namely

$$\textbf{\textit{SP rate out of}}\ ((x,\infty), 1) = \textbf{\textit{SP rate into}}\ ((x,\infty), 1)\,,$$

gives an integral equation relating $f_1(x)$ and $f_2(x)$,

$$\lambda_1 \int_{y=x}^{\infty} f_1(y)dy = q(x)f_1(x) + \lambda_2 \int_{y=x}^{\infty} f_2(y)dy. \tag{10.77}$$

Similarly, balancing SP rates out of, and into $((x,\infty),2)$, $x>0$, on sheet 2 yields the integral equation

$$\lambda_2 \int_{y=x}^{\infty} f_2(y)dy + r(x)f_2(x) = \lambda_1 \int_{y=x}^{\infty} f_1(y)dy. \tag{10.78}$$

In (10.78), the left and right sides are the SP exit and entrance rates respectively, of $((x,\infty),2)$.

Addition of (10.77) and (10.78) yields

$$q(x)\cdot f_1(x) = r(x)\cdot f_2(x). \tag{10.79}$$

There is an easy alternative derivation 0f (10.79), which follows by viewing the sample-path via the "cover". That is, we *project* the segments of the sample path from sheets $1,2$ (pages) onto a single t-$W(t)$ coordinate system (Fig. 10.5). Then we apply SP rate balance across level x:

$$\boldsymbol{total\ upcrossing\ rate} = \boldsymbol{total\ downcrossing\ rate},$$

which translates to (10.79).

Using (10.79), we substitute $f_2(x) = \frac{q(x)}{r(x)} f_1(x)$ into (10.77), and take $\frac{d}{dx}$ in (10.77). Then we solve the resulting differential equation, and applying the initial condition

$$r(0^+)f_2(0) = \lambda_2 P_0 = q(0^+)f_1(0).$$

These operations result in the formula

$$f_1(x) = \frac{\lambda_2 P_0}{q(x)} \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)}dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)}dy\right)}, x>0. \tag{10.80}$$

Since $f_2(x) = \frac{q(x)}{r(x)} f_1(x)$, we have

$$f_2(x) = \frac{\lambda_2 P_0}{r(x)} \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)}dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)}dy\right)}, x>0. \tag{10.81}$$

The total pdf of content is $f(x) = f_1(x) + f_2(x)$. Adding (10.80) and (10.81) gives

$$\begin{aligned} f(x) &= \lambda_2 \left(\frac{1}{q(x)} + \frac{1}{r(x)}\right) P_0 \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)}dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)}dy\right)}, x>0, \\ &= \lambda_2 \left(\frac{q(x)+r(x)}{q(x)r(x)}\right) P_0 \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)}dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)}dy\right)}, x>0. \end{aligned} \tag{10.82}$$

The normalizing condition is

$$P_0 + \int_{x=0}^{\infty} f(x)dx = 1. \tag{10.83}$$

From (10.82) and (10.83)

$$P_0 = \frac{1}{1 + \lambda_2 \int_{x=0}^{\infty} \left(\left(\frac{q(x)+r(x)}{q(x)r(x)} \right) \cdot e^{-\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)} \right) dx}. \tag{10.84}$$

Remark 10.11 *Formulas (10.80)-(10.84) are asymmetric with respect to* λ_1 *and* λ_2*. This is because empty periods are distributed as* E_{λ_2} *(classified as part of efflux phase).*

Remark 10.12 *The model can be generalized in various ways. There may be several different important state-space levels at which there is no change in content (no influx or efflux), rather than only at level* 0*. Such levels may be due to a control policy or due to natural phenomena. There would then be* ***more than one atom*** *in the state space. Also, the influx and efflux periods may have general distributions. The content may be bounded above, resulting in an atom. Some of these variants are easy to analyze; others are more complex. We do not treat such variants here.*

Stability Condition

A necessary condition for the pdf to exist is $f(\infty) = 0$. Thus, the exponent $\left(\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy\right)$ in (10.84) must be positive for all $\dot{x} > 0$. That is

$$\lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy < \lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy,$$
$$\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy \quad > \quad 0, \text{ for all } x > 0. \tag{10.85}$$

10.10.3 Numerical Example

Let $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$. Substituting into (10.85) gives

$$\lambda_1 \int_{y=0}^{x} \frac{1}{q(y)} dy - \lambda_2 \int_{y=0}^{x} \frac{1}{r(y)} dy = 2\sqrt{x}\left(\lambda_1 - \frac{\lambda_2}{3}\right)$$
$$= 2\sqrt{x}\left(1 - \frac{2}{3}\right) > 0, x > 0,$$

implying stability. Thus the steady-state pdf $f(x)$ exists. From (10.82), we obtain

$$f(x) = \frac{8}{3\sqrt{x}} P_0 \cdot e^{-\frac{2}{3}\sqrt{x}}, x > 0. \tag{10.86}$$

From the normalizing condition (10.83),

$$P_0 = \frac{1}{1 + \int_{x=0}^{\infty} \frac{8}{3\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}} dx} = \frac{1}{9} = 0.111111. \tag{10.87}$$

Thus

$$f(x) = \frac{8}{27\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}}, x > 0. \tag{10.88}$$

From (10.87) and (10.88), the cdf is (see Figs. 10.7, 10.8),

$$F(x) = P_0 + \int_{y=0}^{x} f(y) dy = 1 - \frac{8}{9} e^{-\frac{2}{3}\sqrt{x}}. \tag{10.89}$$

Proportion of Time in Influx and Efflux Phases

From ((10.79)) and (10.76) we obtain

$$f_1(x) = \frac{2}{9\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}}, \ x > 0,$$
$$f_2(x) = \frac{2}{27\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}}, \ x > 0.$$

Hence the proportion of time the dam is in the influx, efflux phase respectively is

$$F_1(\infty) = \int_{x=0}^{\infty} \frac{2}{9\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}} dx = 0.666667,$$
$$F_2(\infty) = \int_{x=0}^{\infty} \frac{2}{27\sqrt{x}} e^{-\frac{2}{3}\sqrt{x}} dx = 0.222222.$$

These values are also the steady-state probabilities of the dam being in these phases at an arbitrary time point. A check on the normalizing condition is

$$P_0 + F_1(\infty) + F_2(\infty) = 0.111111 + 0.666667 + 0.222222 = 1.$$

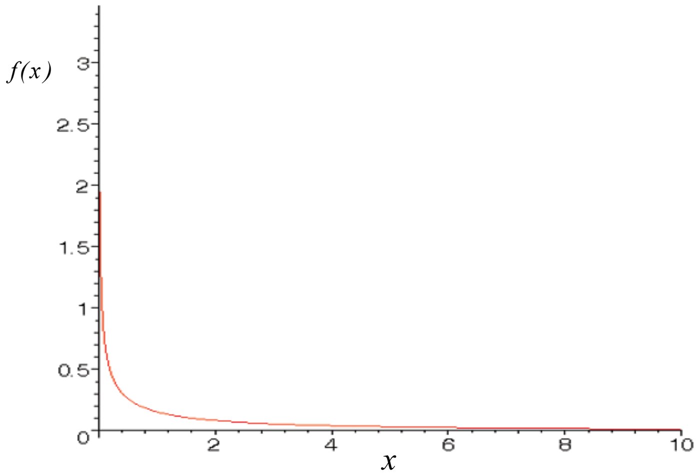

Figure 10.7: Steady-state pdf $f(x) = \frac{8}{27\sqrt{x}}e^{-\frac{2}{3}\sqrt{x}}, x > 0$, in continuous dam with alternating influx/efflux periods: $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$.

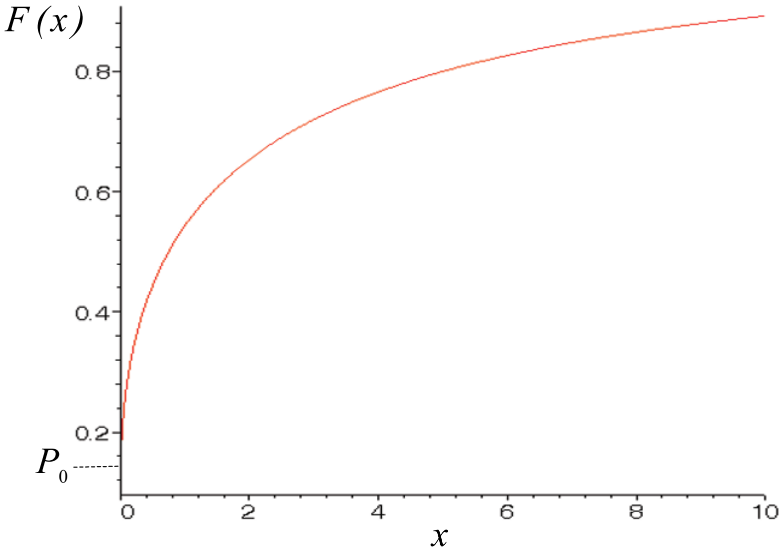

Figure 10.8: Steady-state cdf $F(x) = 1 - \frac{8}{9}e^{-\frac{2}{3}\sqrt{x}}$, $x > 0$, $P_0 = 0.1111$, in continuous dam with alternating influx/efflux periods: $\lambda_1 = 1$, $\lambda_2 = 2$, $q(x) = \sqrt{x}$, $r(x) = 3\sqrt{x}$.

10.11 Estimation of Laplace Transforms

We very briefly discuss a procedure for estimating the LST (Laplace-Stieltjes transform) of the state variable of a stochastic model. We shall use the virtual wait in a GI/G/1 queue as an example.

Suppose we want to estimate the LST of the steady-state pdf of the virtual wait in a GI/G/1 queue. Let the steady-state cdf of the virtual wait be $F(x), x \geq 0$, having pdf $f(x), x > 0$, and let $P_0 = F(0)$. The LST of the mixed pdf $\{P_0; f(x), x > 0\}$ is defined as

$$F^*(s) = \int_{x=0}^{\infty} e^{-sx} dF(x), s > 0. \tag{10.90}$$

10.11.1 Probabilistic Interpretation of LST

The probabilistic interpretation of the LST (10.90) is as follows ([78], and used in various papers, e.g., [31]). In (10.90), the right side is the probability that an independent "*catastrophe random variable*", distributed as E_s, is greater than the virtual wait having cdf $F(x), x \geq 0$.

10.11.2 Estimation of LST

In order to estimate $F^*(s)$, we can simulate a sample path of the virtual wait $W(u), u \geq 0$, over a long period of simulated time $(0, t)$. Next, we generate a sample path of a renewal process $\{\mathcal{C}(u), u \geq 0\}$ with inter-renewal times equal to the catastrophe r.v., and overlay it on the *same time-state coordinate system* (see Fig 10.9). Fix $s > 0$. The SP jump sizes and inter-renewal times in the sample path of $\{\mathcal{C}(u)\}$, are iid r.v.'s distributed as E_s. This is because the process $\mathcal{C}(u)$ represents the excess life γ at time u (see Subsection 10.2.4). The steady-state pdf of excess life is $f_\gamma(x) = s \cdot e^{-sx}, x > 0$.

Now we observe the sample paths of $\{W(u)\}$ and $\{\mathcal{C}(u)\}$ on the time interval $(0, t)$. We compute the **sum**, $T_s = \sum_i T_{si}$, of all time intervals such that $\mathcal{C}(u) > W(u), u \in (0, t)$ (Fig. 10.9). An estimate of $F^*(s)$ is then $\widehat{F^*}(s) = \frac{T_s}{t}$, which is the proportion of time that $\mathcal{C}(u)$ exceeds $W(u)$ during $(0, t)$. The probabilistic interpretation of the LST strongly suggests that $\frac{T_s}{t}$ is an appropriate estimate.

We repeat the procedure using different values of $s > 0$. For example, we may choose a partition of N uniformly-spaced values for s, such as $\Delta, 2\Delta, 3\Delta, ..., N\Delta$, where N is a large positive integer and Δ is a small positive number. (Different spacing for the partition may improve the

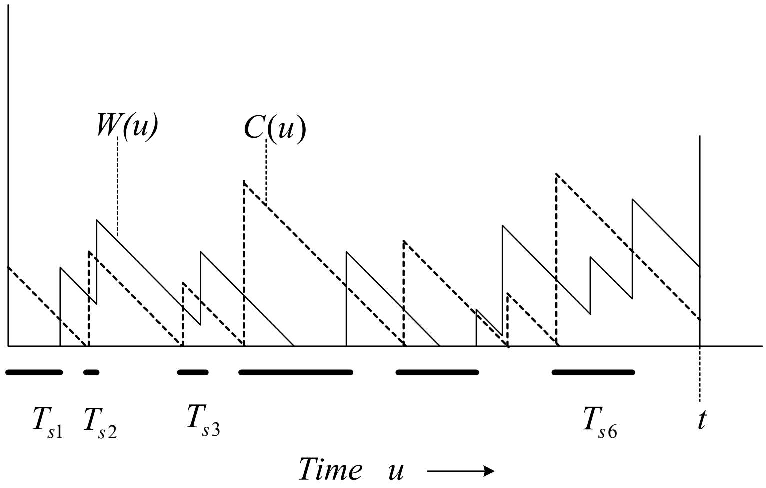

Figure 10.9: Sample paths of virtual wait $\{W(u), u \geq 0, \}$ and renewal process with inter-arrival time distributed as E_s, the catastrphe r.v., $\{\mathcal{C}(u), u \geq 0\}$. $T_s = T_{s1} + T_{s2} + \cdots + T_{s6}$.

estimates, e.g., if $F(\cdot)$ is known to have certain properties such as a long tail.) This procedure results in a set of estimates $\widehat{F^*}(n\Delta) = \frac{T_{n\Delta}}{t}, n = 1, ..., N$. (From (10.90), $\widehat{F^*}(0) = 1.$, which is the normalizing condition.)

Finally, we can plot the points

$$\left(0, \widehat{F^*}(0)\right) = (0, 1) \quad \text{and} \quad \left(n\Delta, \widehat{F^*}(n\Delta)\right), n = 1, ..., N,$$

on a two-dimensional $\left(s, \widehat{F^*}(s)\right)$ coordinate system. The $\{n\Delta\}$ grid is on the horizontal axis; the $\widehat{F^*}(n\Delta)$ terms are ordinates parallel to the vertical axis.

The plot will be a discrete estimate of the LST of the pdf of the virtual wait. It may be improved by smoothing techniques. In order to obtain an estimate of the pdf of the virtual wait from it, we can use numerical inversion of $\left\{\widehat{F^*}(n\Delta)\right\}$.

10.12 Simple Harmonic Motion

We analyze an elementary model of *deterministic* simple harmonic motion, using LC.

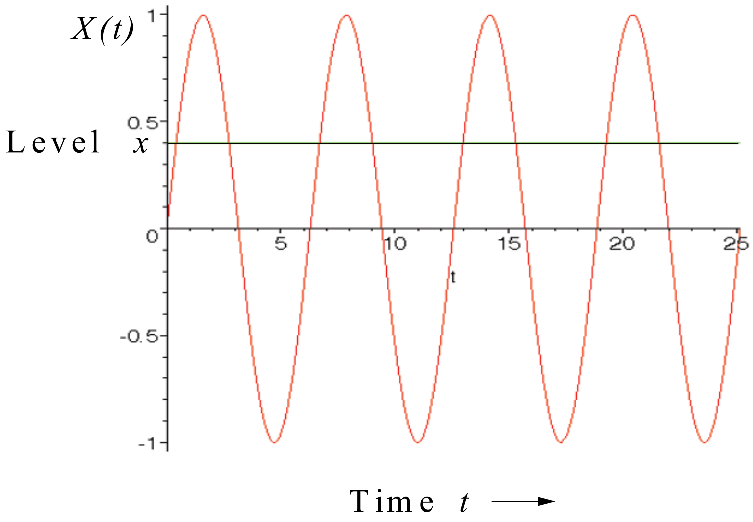

Figure 10.10: Sample path of simple harmonic motion $X(t) = \sin t$. State space is $\boldsymbol{S} = [-1, +1]$. Shows level x in $\boldsymbol{S}$.

Consider a particle moving according to simple harmonic motion (SHM) (see, e.g., [6]). Let $X(t)$ denote the position of the particle at instant $t \geq 0$, and $X(0) = 0$. Let the state space be the interval $\boldsymbol{S} = [-1, +1]$. In this version of the standard SHM model there is only one sample path, namely,

$$X(t) = \sin(t), t \geq 0.$$

We wish to determine the stationary pdf $f(x)$ and cdf $F(x)$ of $X(t)$ when the particle is observed at an arbitrary time point, as $t \to \infty$.

Consider the sample path $X(t), t \geq 0$ (Fig. 10.10). The slope of the sample path at level x is

$$r(x) == \frac{d}{dt} \sin t|_{t=\sin^{-1} x} = \cos\left(\sin^{-1} x\right) = \sqrt{1-x^2}, x \in [-1, +1]\,. \tag{10.91}$$

Consider levels x, $x + h \in \boldsymbol{S}$, where $h > 0$ is small. The time required for the SP to ascend from level x to level $x + h$ is

$$\int_{y=x}^{x+h} \frac{1}{r(y)} dy = \int_{y=x}^{x+h} \frac{1}{\sqrt{1-y^2}} dy. \tag{10.92}$$

The symmetries of the sample path imply that the time required for the SP to descend from level $x+h$ to level x is also given by (10.92).

Applying (10.92), we see that the long-run *proportion of time* the SP spends in state-space interval $(x, x+h)$ in a cycle of length 2π time units is

$$\frac{2}{2\pi}\int_{y=x}^{x+h}\frac{1}{\sqrt{1-y^2}}dy = F(x+h) - F(x). \tag{10.93}$$

Formula (10.93) leads to

$$\frac{1}{\pi}h\frac{1}{\sqrt{1-(x^*)^2}} = F(x+h) - F(x) \tag{10.94}$$

where $x^* \in (x, x+h)$, by the definition of $F(x)$ as the long-run proportion of time the process is in state-space interval $[-1, x]$. Dividing both sides of (10.94) by h and letting $h \downarrow 0$, yields

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}}, x \in [-1, +1]\,. \tag{10.95}$$

The stationary pdf $f(x)$ in (10.95) is interesting and suggests intuitive insights (Fig. 10.11). Note that $\lim_{x\downarrow(-1)} f(x) = \lim_{x\uparrow(+1)} f(x) = \infty$. Also, $\min_{x\in \boldsymbol{S}} f(x) = \frac{1}{\pi}$, at $x = 0$. The pdf $f(x)$ is symmetric about $x = 0$, and is convex.

From (10.95), the cdf is

$$\begin{aligned} F(x) &= \int_{y=-1}^{x} f(y)dy, \\ &= \frac{1}{\pi}\left(\sin^{-1}(x) - \sin^{-1}(-1)\right) \\ &= \frac{1}{\pi}\sin^{-1}(x) + \frac{1}{2}, x \in [-1, +1]. \end{aligned} \tag{10.96}$$

10.12.1 Inferences Based on PDF and CDF

From (10.91), the speed of the particle $r(x) = \sqrt{1-x^2} = 0$ at $x = \pm 1$. Hence, intuitively, it is much more likely to observe the particle close to the boundaries of $\boldsymbol{S}$ $(x = \pm 1)$, at an arbitrary time point in the long run. This fact implies that the particle spends a much greater proportion of time near the boundaries $x = \pm 1$ than near the center $x = 0$. At the center, the speed is $r(0) = 1$. This is the maximum speed.

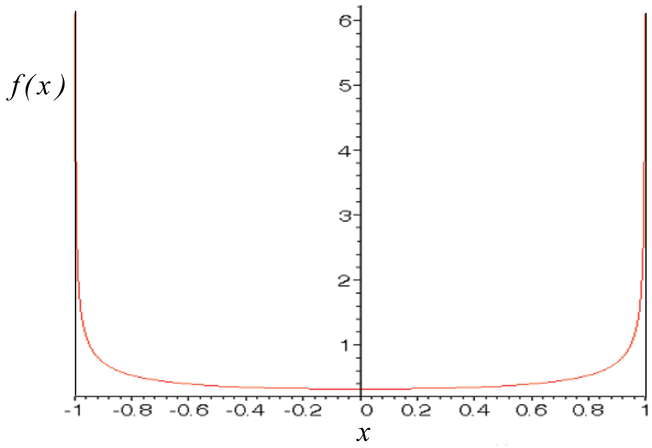

Figure 10.11: Stationary pdf $f(x) = \frac{1}{\pi\sqrt{1-x^2}}, x \in [-1, +1]$, for particle moving in simple harmonic motion, $X(t) = \sin t, t \geq 0$.

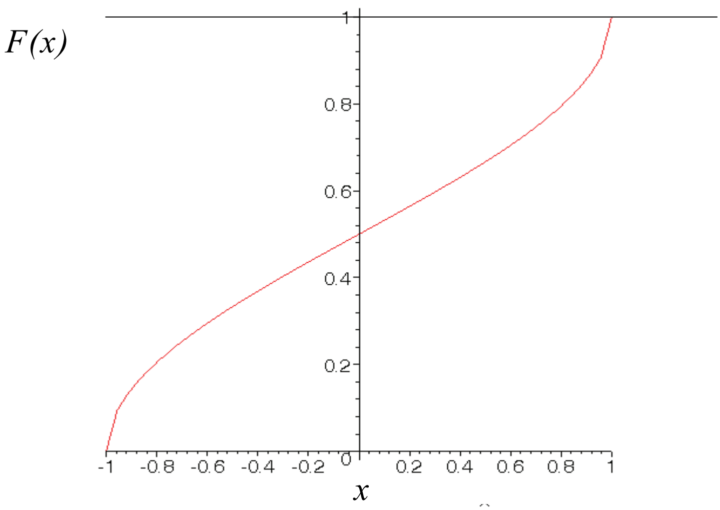

Figure 10.12: Stationary cdf $F(x) = \frac{1}{\pi}\sin^{-1}(x) + \frac{1}{2}, x \in [-1, +1]$, for particle moving in simple harmonic motion, $X(t) = \sin t, t \geq 0$.

From computations using (10.96), the proportion of time the SP (particle) spends in the central interval $[-.5, +.5]$ is equal to $F(.5)-F(-.5) = 0.333$. The proportion of time the particle spends in the outer regions $[-1.0, -.5] \cup [.5, 1.0]$, is equal to $2 \cdot (F(1.0) - F(.5)) = 0.667$. The "median" symmetric outer edges with respect to time spent by the particle, is $\boldsymbol{A}_{0.5} \equiv [-1.0, -.707] \cup [.707, 1.0]$. That is, $P(\text{particle} \in \boldsymbol{A}_{0.5}) = 0.5$. This indicates that it is equally likely to find the particle in two bands of equal width 0.293 touching the edges ± 1.0 (total width .586), as it is to find it in a central interval of width 1.414 about 0. Arbitrary observations on operating pendulum clocks, readily corroborate these theoretical computations.

Remark 10.13 *The type of LC analysis in this section, may be extendable to analyze random trigonometric functions (e.g., like $A \sin(\theta t) + B \cos(\theta t), t \geq 0$, where A, B are random variables and θ is a constant). Extensions may also be applicable in some models of physics, and in the analysis of roots of equations.*

10.13 Renewal Problem with Barrier

Consider a renewal process $\{Z_n\}, n = 1, 2....$. Assume $Z_n \underset{dist}{=} U_{(0,1)})$, a uniform random variable on $(0, 1)$ (Fig. 10.13). Let N_K denote the number of renewals required to *first exceed* a barrier $K > 0$. In this section we derive the expected value $E(N_K), K = 1, 2, 3, ...$, and related results. It is well known that $E(N_1) = e$, the base of natural logarithms. The general formula for $E(N_K)$ has not been reported previously in the literature or is not well known. It is usually shown that $E(N_1) = e$ by a standard renewal argument. That is, condition on the first renewal distance s (Fig. 10.13). Derive a renewal equation, and solve it.

In this section we derive $E(N_1)$ by an alternative method, which also leads to the values of $E(N_K), K = 1, 2, ...$. This alternative method facilitates finding the expected number of renewals required to exceed a barrier, in other (seemingly unrelated) models. The idea is to extend the one-dimensional renewal process to a two-dimensional *nested* renewal process. The new construct has applications in a variety of stochastic models.

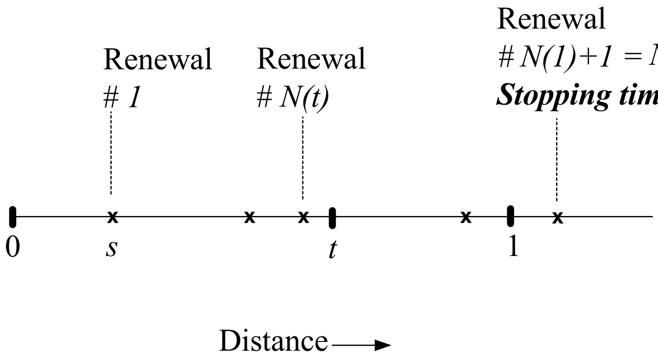

Figure 10.13: Renewal process $\{Z_n\}$ showing renewals. $N(t)$ is the number of renewals within $(0, t)$. $N_1 = N(1) + 1$ is number of renewals required to first exceed barrier $K = 1$. N_1 is a stopping time for the sequence $\{Z_n\}$ where $Z_n \underset{dist}{=} U_{(0,1)}$.

10.13.1 Alternative Solution Method

We construct a continuous-time continuous-state stochastic process

$$\{X(t), t \geq 0\}, X(0) = 0,$$

which is related to $\{Z_n\}$ (Fig. 10.14). A sample path of $\{X(t)\}$ is a non-decreasing step function. In sample paths of $\{X(t)\}$, SP *upward* jumps of size $\underset{dist}{=} U_{(0,1)}$, occur at an arbitrary Poisson rate λ. (We will select $\lambda = 1$ for convenience.) The upward jumps are denoted by

$$b_n \underset{dist}{\equiv} U_{(0,1)}, n = 1, 2, \ldots .$$

(Note that $Z_n \equiv b_n$. We replace symbol Z_n by b_n for generality beyond boundary $K = 1$, and because of applicability to other models.)

Let

$$N_K = \min\{n | \sum_{i=1}^{n} b_i > K\}, K = 1, 2, \ldots . \tag{10.97}$$

Random variable N_K is a *stopping time* for the sequence $\{b_n\}$.

Let random variable $a \underset{dist}{=} \mathrm{E}_\lambda = \mathrm{E}_1$. Thus $E(a) = 1$.

Define random variable c by

$$c = \sum_{i=1}^{N_K} a_i, \quad \text{where} \quad a_i \underset{dist}{\equiv} a. \tag{10.98}$$

Let $\{c_n\}$ be a renewal process where $c_n \underset{dist}{\equiv} c$. Then $\{c_n\}$ is a *nested renewal process* with components $\{c_n\}$ and sub-components $\{a_i\}$. Note that N_K is also a stopping time for the sequence $\{a_i\}$. Taking the expected value in (10.98) yields

$$E(c) = E(N_K)E(\alpha) = E(N_K), \qquad (10.99)$$

by Wald's equation (e.g., [91] or [101]).

At each instant when a sample path of $\{X(t)\}$ upcrosses level K, the SP jumps downward (rebounds) to level 0, and the process $\{X(t)\}$ starts over again at level 0. Our construction guarantees that the limiting distribution of $X(t)$ exists as $t \to \infty$. Random variable N_K equals the number of SP jumps required for $\{X(t)\}$ to first exceed level K. R.v. N_K is also equal to the number of subintervals which are $\underset{dist}{=} a$, that comprise a cycle c.

Relation to $\langle s, S\rangle$ with No Decay

It is notable that other stochastic models have a related structure. For example, the $\langle s, S\rangle$ inventory *with no decay* in Example 2.3 is the "flip" (like $\updownarrow$) of the $\{X(t)\}$ process, in which $K = S - s$, and the jump sizes are distributed as E_μ. In that $\langle s, S\rangle$ model $E(N_{S-s})$ is the expected number of orders in an ordering cycle.

10.13.2 Number of Renewals Required to Exceed 1

We first determine $E(N_1)$. Denote the limiting distribution of $\{X(t)\}$ as $t \to \infty$, by $\{\pi_0;\ f_0(x), 0 < x < 1\}$. Consider a sample path of $\{X(t)\}$. Fix level $x \in (0,1)$ (Fig. 10.14). SP upcrossings of level x are due to jumps starting at level 0 or at level $y, 0 < y < x$. Thus the SP upcrossing rate of level x is

$$1 \cdot \pi_0 \cdot P(b > x) + 1 \cdot \int_{y=0}^{x} P(b > x - y) \cdot f(y)dy, \qquad (10.100)$$

where r.v. $b \underset{dist}{\equiv} b_i$, and upward jumps occur at rate $\frac{1}{E(\alpha)} = \lambda = 1$.

The SP downcrossing rate of level x is equal to the *upcrossing rate of level* 1 *for all* $x \in (0,1)$. That is, the SP rebounds to level 0 at every instant it upcrosses level 1. (The SP makes a *double jump.* Compare with $\langle s, S\rangle$ inventory with no decay in Example 2.3.) The rate of SP downward jumps is also the rate of SP entrances into state $\{0\}$ from

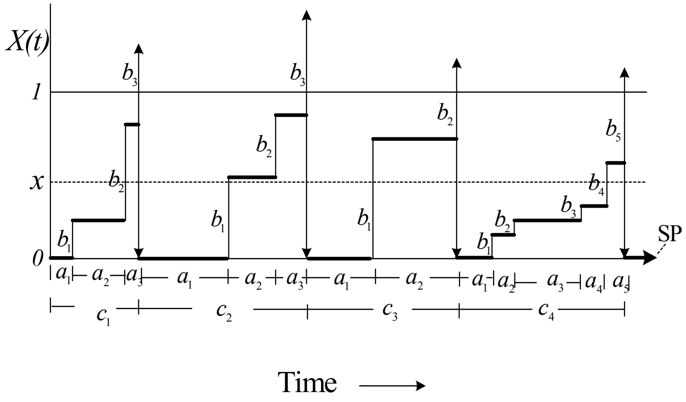

Figure 10.14: Sample path of $\{X(t), t \geq 0\}$, in renewal problem to determine $E(N_1)$ when renewal times $\underset{dist}{=} U_{(0,1)}$.

above. This rate is the same as the SP exit rate out of $\{0\}$, namely $\lambda\pi_0 = 1 \cdot \pi_0 = \pi_0$. Letting $x = 1$ in (10.100) we obtain

$$1 \cdot \pi_0 \cdot P(b > 1) + 1 \cdot \int_{y=0}^{1} P(b > 1 - y) \cdot f(y)dy = \pi_0. \tag{10.101}$$

Note that since $b \underset{dist}{=} U_{(0,1)}$,

$$P(b > x) = 1 - x, 0 < x < 1. \tag{10.102}$$

We substitute from (10.102) into (10.100). Then we apply rate balance across level x to equate (10.100) to the right-hand side of (10.101), resulting in

$$\pi_0(1 - x) + \int_{y=0}^{x} (1 - x + y)f(y)dy = \pi_0, 0 < x < 1. \tag{10.103}$$

Taking $\frac{d}{dx}$ twice in (10.103), and solving the resulting ordinary differential equation gives

$$f(x) = \pi_0 e^x, 0 < x < 1. \tag{10.104}$$

We substitute from (10.104) into the normalizing condition $\pi_0 + \int_{x=0}^{1} f(x)dx = 1$. This gives

$$\pi_0 = \frac{1}{e}. \tag{10.105}$$

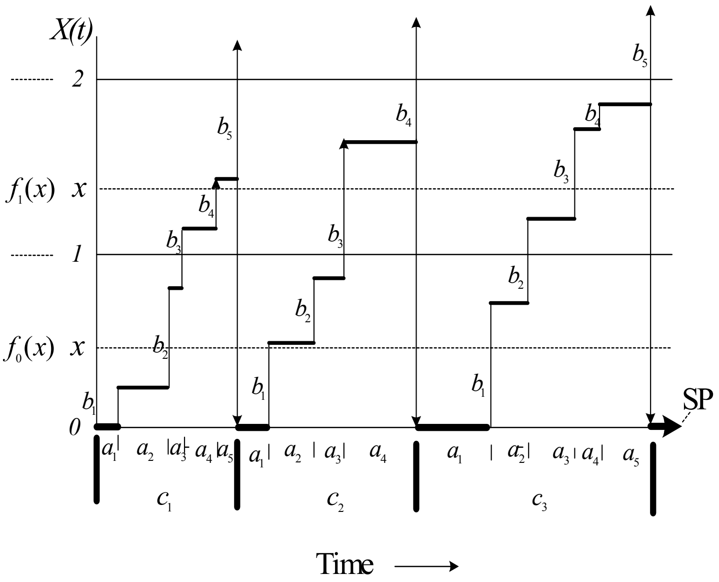

Figure 10.15: Sample path of $\{X(t)\}$ for renewal problem, with state space $\boldsymbol{S} = [0, 2)$. Facilitates solution for $E(N_2)$.

The renewal rate of $\{c_n\}$ is $\frac{1}{E(c)}$ = SP entrance rate into $\{0\} = \pi_0$. Thus $E(c) = \frac{1}{\pi_0}$. From (10.99) and (10.105),

$$E(N_1) = E(c) \cdot E(a) = \frac{1}{\pi_0} \cdot 1 = e = 2.71828. \tag{10.106}$$

We have derived $E(N_1)$ in detail using the nested renewal process structure, to fix ideas. The following results are new (or not well known).

10.13.3 Number of Renewals Required to Exceed 2

Next we determine $E(N_2)$. Let the steady-state PDF of $\{X(t)\}$ be

$$\{\pi_0;\ f_0(x), 0 < x < 1\};\ \{\ f_1(x), 1 \leq x < 2\}.$$

Consider a sample path of $\{X(t)\}$ (Fig. 10.15), where the state space is $\boldsymbol{S} = [0, 2)$. Balancing SP up- and downcrossing rates of $x \in (0, 1)$, as in the case $K = 1$, gives

$$\pi_0(1 - x) + \int_{y=0}^{x} (1 - x + y) f_0(y) dy = \pi_0, 0 < x < 1. \tag{10.107}$$

Fix $x \in [1,2)$. Balancing SP up- and downcrossing rates of x, gives

$$\int_{y=x-1}^{1}(1-x+y)f_0(y)dy + \int_{y=1}^{x}(1-x+y)f_1(y)dy = \pi_0. \quad (10.108)$$

The lower limit in the first integral of (10.108) is $y = x - 1$ because an SP jump upcrosses x only if it starts in interval $(x-1, x)$.

Taking $\frac{d}{dx}$ in (10.108) and solving in a similar manner as for $K = 1$, we obtain

$$\begin{aligned} f_0(x) = &\quad \pi_0 e^x, 0 < x < 1, \\ f_1(x) = &\quad \pi_0(1 - e^{-1}x)e^x, 1 \leq x < 2. \end{aligned} \quad (10.109)$$

The normalizing condition is

$$\pi_0 + \int_{x=0}^{1} f_0(x)dx + \int_{x=1}^{2} f(x)dx = 1. \quad (10.110)$$

Substituting from (10.109) into (10.110) gives

$$\pi_0 = \frac{1}{-e + e^{2.}}. \quad (10.111)$$

From (10.99),

$$E(N_2) = E(c)E(a) = \frac{1}{\pi_0} = -e + e^2 = 4.67077. \quad (10.112)$$

10.13.4 Number of Renewals Required to Exceed 3

To explore further the pattern of $\{E(N_K)\}, K = 1, 2, ...$ we derive $E(N_3)$. The state space is $\boldsymbol{S} = [0,3)$. Let the steady state PDF of $\{X(t)\}$ be

$$\{\pi_0;\ f_0(x), 0 < x < 1\};\ \ \{f_1(x), 1 \leq x < 2\};\quad \{f_2(x), 2 \leq x < 3\}.$$

We now balance SP up- and downcrossing rates across arbitrary levels $x \in (0,1)$; $x \in [1,2)$; $x \in [2,3)$. This gives respectively, integral equations

$$\pi_0(1-x) + \int_{y=0}^{x}(1-x+y)f_0(y)dy = \pi_0, \quad (10.113)$$

$$\int_{y=x-1}^{1}(1-x+y)f_0(y)dy + \int_{y=1}^{x}(1-x+y)f_1(y)dy = \pi_0, \quad (10.114)$$

$$\int_{y=x-1}^{2}(1-x+y)f_1(y)dy + \int_{y=2}^{x}(1-x+y)f_2(y)dy = \pi_0. \quad (10.115)$$

Solving integral equations (10.113), (10.114), (10.113) in a similar manner as for $K = 1, 2$ above, gives

$$\begin{aligned} f_0(x) = & \quad \pi_0 e^x, 0 < x < 1, \\ f_1(x) = & \quad \pi_0(1 - e^{-1}x)e^x, 1 \leq x < 2, \\ f_2(x) = & \quad \tfrac{1}{2}\pi_0(-2xe^{-2} + e^{-2}x^2 - 2xe^{-1} + 2)e^x, 2 \leq x < 3. \end{aligned} \tag{10.116}$$

The normalizing condition is

$$\pi_0 + \int_{x=0}^{1} f_0(x)dx + \int_{x=1}^{2} f_1(x)dx + \int_{x=2}^{3} f_2(x)dx = 1, \tag{10.117}$$

yielding

$$\pi_0 = \frac{1}{\frac{1}{2}e - 2e^2 + e^3}.$$

Substituting from (10.116) into (10.117) gives

$$E(N_3) = \frac{1}{\pi_0} = \frac{1}{2}e - 2e^2 + e^3 = 6.66656563. \tag{10.118}$$

10.13.5 Number of Renewals Required to Exceed K

After carrying out the procedure for several more steps, I hypothesized that the formula for general integer K is $E(N_K) = \sum_{i=1}^{K} \frac{(-i)^{K-i}}{(K-i)!} e^i$. This formula can be verified by mathematical induction. Thus

$$E(N_K) = \sum_{i=1}^{K} \frac{(-i)^{K-i}}{(K-i)!} e^i, K = 1, 2, \ldots . \tag{10.119}$$

The induction is carried out by assuming that the formulas for $f_i(x), i = 0, ..., K-1$ are similar to those in (10.116). Then we obtain (10.119) in a similar manner as for the derivation of (10.118).

10.13.6 Asymptotic Formula for $E(N_K)$

We can show that $E(N_K)$ given in (10.119) is asymptotic to $2K + \frac{2}{3}$. That is

$$\lim_{K \to \infty} \frac{E(N_K)}{2K + \frac{2}{3}} = 1. \tag{10.120}$$

For example, using (10.120), an approximation to $E(N_{20})$ is $2(20) + \frac{2}{3} = 40.6667$. The analytical value using (10.119) is 40.6667. The accuracy of the computation depends on the number of digits carried, and on the computational algorithm.

Remarkably, from the analytical values of $E(N_2)$ and $E(N_3)$ given in (10.112) and (10.118), the approximation (10.120) is very accurate for $K = 2, 3, \ldots$. Even for $K = 1$, we have $2K + \frac{2}{3} = 2.6666$, which is within 1.90% of $e = 2.71828$.

Derivation of Asymptotic Formula

We give a renewal-theoretic derivation of formula (10.120).

Let γ_x denote the excess life at a point $x \in \boldsymbol{S}$. The pdf of γ_x as $x \to \infty$ is given by $f_\gamma(y) = \frac{1}{\mu}(1 - B(y)), y > 0$ where $B(y)$ is the common cdf of the renewal r.v. having mean μ (formula (10.9)). In the present context, the renewal r.v. $\underset{dist}{=} U_{(0,1)}$. Thus $B(y) = y, 0 < y < 1$ and $\mu = \frac{1}{2}$. Hence $\lim_{x\to\infty} E(\gamma_x)$ is given by

$$\begin{aligned} \lim_{x\to\infty} E(\gamma_x) &= \frac{1}{\mu}\int_{y=0}^{\infty} y f_\gamma(y) dy \\ &= 2\int_{y=0}^{1} y(1-y)\, dy = \frac{1}{3}. \end{aligned} \tag{10.121}$$

Let γ_K denote the excess life at K; then $E(\gamma_K) \approx \frac{1}{3}$. Also,

$$K + \gamma_K = \sum_{j=1}^{N_K} Z_j, \tag{10.122}$$

where $\{Z_j\}$ are iid, $Z_j \underset{dist}{\equiv} U_{(0,1)}$, and N_K is a stopping time for $\{Z_j\}$. Taking expected values in (10.122) yields $K + \frac{1}{3} \approx E(N_K)\frac{1}{2}$. If $K \to \infty$, we obtain (10.120). Moreover, if $\alpha > 0$ is a real number, then $E(N_\alpha) \approx 2\alpha + \frac{2}{3}$, where N_α is the number of renewals required to first exceed α.

10.13.7 Number of Renewals Within an Interval

Let $N(a, b)$ denote the number of renewal instants occurring *within* interval (a, b), during a single cycle of $\{c_n\}$. Without loss of generality, $X(0) = 0$, and we stop after N_K renewals of $\{a_n\}$. Then

$$N(0, K) = N_K - 1, \quad \text{and} \quad E(N(0, K)) = E(N_K) - 1.$$

Thus the values of $E(N_1)$, $E(N_2)$, $E(N_3)$ lead to the expected number of renewal instants within intervals $(0,1)$, $(0,2)$, $(0,3)$, $(1,2)$, $(2,3)$, namely

$$\begin{aligned}
E(N(0,1)) &= E(N_1) - 1 = e - 1 = 1.7183,\\
E(N(0,2)) &= E(N_2) - 1 = -e + e^2 - 1 = 3.6708,\\
E(N(0,3)) &= E(N_3) - 1 = \frac{1}{2}e - 2e^2 + e^3 - 1 = 5.6666,\\
E(N(1,2)) &= E(N(0,2)) - E(N(0,1)) = E(N_2) - E(N_1) = 1.9525,\\
E(N(2,3)) &= E(N(0,3)) - E(N(0,2)) = E(N_3) - E(N_2) = 1.9958.
\end{aligned} \tag{10.123}$$

For large K,

$$\begin{aligned}
E(N(K, K+1) &= E(0, K+1) - E(0, K)\\
&= E(N_{K+1}) - E(N_K) \approx 2.
\end{aligned}$$

Note that in (10.123), the values of $E(N(1,2))$, $E(N(2,3))$ are already within 2.38% and 1.40% of the limiting value 2.0, respectively.

Suppose $0 < \alpha < \beta < 1$, where α, β are arbitrary real numbers. We obtain $E(N_\alpha) = e^\alpha$, and $E(N_\beta) = e^\beta$, analogously as for the solution for $E(N_1)$. Hence, $E(N(0,\alpha)) = e^\alpha - 1$, $E(N(0,\beta)) = e^\beta - 1$. Therefore, the expected number of renewals within (α, β) is

$$E(N(\alpha,\beta)) = E(N_\beta) - E(N_\alpha) = e^\beta - e^\alpha, 0 < \alpha < \beta < 1. \tag{10.124}$$

For example

$$\begin{aligned}
E\left(N\left(\tfrac{2}{3},1\right)\right) &= e - e^{\frac{2}{3}} = 0.77055,\\
E\left(N\left(\tfrac{1}{3},\tfrac{2}{3}\right)\right) &= e^{\frac{2}{3}} - e^{\frac{1}{3}} = 0.55212,\\
E\left(N\left(0,\tfrac{1}{3}\right)\right) &= e^{\frac{1}{3}} - e^0 = 0.39561.
\end{aligned}$$

Thus approximately 44.84% of the renewals occur in the top third, 32.13% in the middle third and 23.02% in the bottom third, of interval $(0,1)$. Hence, renewal instants tend to accumulate in the top portion of $(0,1)$. For a possible intuitive explanation of this phenomenon, fix the length of a "sliding interval" $\boldsymbol{I}_h$ to be $|\boldsymbol{I}_h| = h, 0 < h < 1$. As $\boldsymbol{I}_h$ slides from position $(0,h)$ to position $(1-h,1)$, the probability that $\boldsymbol{I}_h$ will contain n renewals increases for every $n = 1, 2, ...$.

We can extend the analysis to determine the expected number of renewals within an arbitrary interval $(\alpha,\beta), 0 \leq \alpha < \beta < \infty$.

10.13.8 Discussion

We can apply the nested renewal model of this section, to an arbitrary renewal process such that $\{b_n\}$ are non-lattice positive r.v.'s. The analysis can also be extended to models where $\{b_n\}$ are such that $-\infty < b_n < \infty$. In that case, $\{b_n\}$ is not a renewal process, but $\{c_n\}$ and $\{a_n\}$ are renewal processes, with $\{a_n\}$ nested in $\{c_n\}$.

Possible applications are to problems where it is required to determine the expected number of events until a stopping criterion is satisfied. Examples are the number of: customers served in a busy period of a queue; orders in an ordering cycle of an inventory; inputs until overflow of a dam; shocks until failure of a machine part; claims until ruin in an actuarial model; doses of a drug until an overdose; ads until a favorable consumer response to a product occurs.

REFERENCES

[1] Ancker, C.J., Gafarian, A.V. (1961). "Queueing with Multiple Inputs and Exponential Service Times", *Operations Research* **9**(3), 321-327.

[2] Azoury, K. and Brill, P.H. (1986). “An Application of the System-Point Method to Inventory Models Under Continuous Review”, *Journal of Applied Probability* **23**(3), 778-789.

[3] Azoury, K. and Brill, P.H. (1992). "Analysis of Net Inventory in Continuous Review Models with Random Lead Time", *European Journal of Operational Research*, **39**, 383-392.

[4] Baccelli, F., Brémaud, P. (1991). *Elements of Queueing Theory*, Springer, New York.

[5] Bartlett, M.S. (1978). *An Introduction to Stochastic Processes, Third edition, Cambridge University Press.*

[6] Boyce, W.E., DiPrima, R.C. (1969). *Elementary Differential Equations and Boundary Value Problems, Second Edition*, John Wiley & Sons, Inc., New York.

[7] Brill, P.H. (1975). *System Point Theory in Exponential Queues*, PhD Dissertation, University of Toronto. Available from University Microfilms International, Ann Arbor, Michigan, order number 8901129.

[8] Brill, P.H. (1976) "A New Methodology for Modelling a Broad Class of Exponential Queues", *Advances in Applied Probability*, **8**(2), p. 242. (Abstract of presentation at the *Fifth Conference on Stochastic Processes and their Applications*, University of Maryland, College Park, MD, USA, June, 1975. Invited by J. Keilson

(external examiner for my PhD thesis) and R. Syski. First conference presentation on the level crossing methodology.

[9] Brill, P.H., (1976). "Embedded Level Crossing Processes in Dams and Queues", Working Paper #76-022, Univ. of Toronto, Dept. of Industrial Engineering.

[10] Brill P.H. (1977) "The System Point Process in Queueing", *Advances in Applied Probability*, **9**(2), p.216. Abstract of presentation at the *Sixth Conference on Stochastic Processes and their Applications*, Tel Aviv, Israel, June, 1976.

[11] Brill, P.H. (1979). "An Embedded Level Crossing Technique for Dams and Queues", *Journal of Applied Probability* **16**, 174-186.

[12] Brill, P.H. (1983). "Queues with Reneging Depending on Required Wait", *Technical Report STAT-83-07*, Dept. of Statistics and Actuarial Science, University of Waterloo.

[13] Brill, P. H. (1983). "System Point Monte Carlo Simulation of Stationary Distributions of Waiting Times in Single Server Queues", STAT 83-10, Dept. of Statistics and Actuarial Sci., Univ. of Waterloo.

[14] Brill, P.H. (1987) "System Point Computation in Queues, Dams and Inventories", *Working paper No. W87-12*, ISSN No. 0716-6301, Faculty of Business Administaration, University of Windsor, Canada.

[15] Brill, P.H. (1988) "Single-server Queues with Delay-dependent Arrival Streams", *Probability in the Engineering and Informational Sciences*, **2**, 231-247.

[16] Brill, P.H. (1988), "Waiting Time in Queues with State Dependent Bulk Service". *Working Paper No. W88-02*, ISSN No. 0714-6191, Faculty of Business Administaration, University of Windsor, Canada.

[17] Brill, P.H. (1988) "A Technique for Transient Distributions in Stochastic Models", *Working Paper W-88-14*, , ISSN No. 0714-6191, Faculy of Business Administration, University of Windsor.

[18] Brill, P.H. (1988) "The Time Dependent System Point Level Crossing Method for Exponential Queues", *Working Paper W-88-15*, Faculy of Business Administration, University of Windsor, ISSN No. 0714-6191.

[19] Brill, P.H. (1988) "The Time Dependent System Point Level Crossing Method for Queues, Dams and Inventories", *Working Paper W-88-16*, Faculy of Business Administration, University of Windsor, ISSN No. 0714-6191.

[20] Brill, P.H. (1990) "Level Crossing Estimation of Stationary Distributions in Storage Processes", (Revision of Brill, P.H. (1987) "System Point Computation in Queues, Dams and Inventories", *Working paper No. W87-12,* University of Windsor, Canada, 64 pages.)

[21] Brill, P.H. (1990) "Example of Level Crossing Estimation in M/G/1 Queues", *Proceedings of the American Statistical Association, Statistical Computing Section*, Anaheim, California, 151-154.

[22] Brill, P.H. (1991) "Estimation of Stationary Distributions in Storage Processes Using Level Crossing Theory", *Proceedings of the American Statistical Association*, Atlanta, Georgia, Statistical Computing Section, 172-177.

[23] Brill, P.H. (1992) "A Note on the Age, Excess Life and Total Life in a Renewal Process", Unpublished notes. June – July 1992.

[24] Brill, P.H. (1994) "Level Crossing Estimation of Stationary Distributions in Storage Processes", Paper presented at CORS-94 conference, Montreal, Canada, May 31.

[25] Brill, P.H. (1996). "Level Crossing Methods", in *Encyclopedia of Operations Research and Management Science*, Gass, S.I. and Harris, C.M., editors, Kluwer Academic Publishers, 338-340.

[26] Brill, P.H. (2000). "A Brief Outline of the Level Crossing Method in Stochastic Models", (Engish and French), *CORS (Canadian Operational Research Society) Bulletin*, **34**(4), 9-21.

[27] Brill, P.H., Chaouch, B.A. (1995). "An EOQ Model with Random Variation in Demand", *Management Science*, **41**(5), 927-936.

[28] Brill, P. H., Green, L. (1984) "Queues in which Customers Receive Simultaneous Service from a Random Number of Servers: A System Point Approach", *Management Science*, 30, No. 1, 51-68.

[29] Brill, P. H., Harris, C. M. (1997). "M/G/1 Queues with Markov-generated Server Vacations", *Stochastic Models*, **13(3)**, 491-521.

[30] Brill, P.H., Hornik, J. (1984). "A System Point Approach to Nonuniform Adversising Insertions", *Operations Research*, **32** (1), 7-22.

[31] Brill, P. H., Hlynka, M. (2000). "An Exponential Queue with Competition for Service", *European J. of Operational Research*, **126**, 587-602.

[32] Brill, P.H., Huang, M.L. (1993) "System Point Estimation of the Probability Distribution of the Waiting Time in Variations of M/G^B/1 Queues", *Proceedings of the American Statistical Association*, San Francisco, California, Statistical Computing Section, 236-241.

[33] Brill, P. H., Moon,R.E. (1980). "Application of Queueing Theory to Pharmacokinetics", *Journal of Pharmaceutical Sciences,* **89**(5), 558-560.

[34] Brill, P.H., Posner, M.J.M. (1974a). "Two Server Queues with Service Time Depending on Waiting Time." Working Paper WP74-005, Department of Industrial Engineering, University of Toronto, pp. 1-61 (uses embedded Markov chain analysis).

[35] Brill, P.H., Posner, M.J.M. (1974b). "A Multiple Server Queue with Service Time Depending on Waiting Time." Working paper WP74-008, Department of Industrial Engineering, University of Toronto. pp 1-30 (uses embedded Markov chain analysis).

[36] Brill, P.H., Posner, M.J.M. (1974c). "On the Equilibrium Distribution for a Class of Exponential Queues", Working Paper WP74-012, Department of Industrial Engineering, University of Toronto (first paper using level crossing method).

[37] Brill, P.H., Posner, M.J.M. (1977). "Level Crossings in Point Processes Applied to Queues: Single Server Case", *Operations Research* **25(**4**)**, 662-673.

[38] Brill, P.H., Posner M.J.M. (1981). “The System Point Method in Exponential Queues: A Level Crossing Approach”, *Mathematics of Operations Research* **6(1)**, 31-49.

[39] Brill, P. H. and Posner M.J.M. (1981). "A two-server Queue with Non-waiting Customers Receiving Specialized Service", *Management Science*, **27**(8), 914-925.

[40] Brockmeyer, E., Halstrom, H.L., Jensen, A. (1948). "The Life and Works of A. K. Erlang", *Trans. of the Danish Acad. of Technical Sci., N0. 2*, 1-277.

[41] Callahan, J.R. (1971), "The Nothing Hot Delay Problem in the Production of Steel", PhD Thesis, Department of Industrial Engineering, University of Toronto.

[42] Callahan, J.R. (1973), "A Queue with waiting Time Dependent Service Times", *Naval Research Logistics Quarterly*, **20**, pp. 321 - 324.

[43] Chen, Peide (1997). *Some Topics on Markov Chains and theirApplications,* PhD Dissertation, Colorado State University.

[44] Çinlar, E. (1975). *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, N.J.

[45] Cohen, J.W. (1976). *On Regenerative Processes in Queueing Theory,* Lecture Notes in Economics and Mathematical Systems, Eds. Beckman, M., Kunzi, H. P., Springer-Verlag, New York.

[46] Cohen, J.W. (1977). "On Up-and Downcrossings", *J. Applied Probability,* **14***,* 405-410.

[47] Cohen, J.W. (1982). *The Single Server Queue*, Revised edition, North-Holland Publishing Co., New York.

[48] Cooper, R.B. (1981). *Introduction to Queueing Theory, 2^{nd} edition, North Holland, New York.*

[49] Cox, D. R. (1962). *Renewal Theory*, Methuen & Co., London.

[50] Cox, D.R., Miller, H.D. (1965). *The Theory of Stochastic Processes*, Methuen, London.

[51] Cox, D.R., Smith, W.L. (1967). *Queues*, Methuen and Co., Ltd., London.

[52] Cramer, H., Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes*, Dover Publications, Mineola, N.Y.

[53] Doshi, B.T. (1986). "Queueing Systems with Vacations", *Queueing Systems: Theory and Applications*, **1**(1), 29-66.

[54] Erlang, A.K. (1909). "The Theory of Probabilities and Telephone Conversations", in Brockmeyer, E., Halstrøm, H.L., Jensen, A. (1948). *The Life and Works of A.K. Erlang, Transactions of the Danish Academy of Technical Sciences,* **2**, 131-137

[55] Feller, W. (1950). *An Introduction to Probability Theory and its Applications, Volume I*, John Wiley, New York.

[56] Feller, W. (1966). *An Introduction to Probability Theory and its Applications, Volume II*, John Wiley, New York.

[57] Franx, G.J. (2001). "A Simple Solution for the M/D/c Waiting Time Distribution", *Operations Research Letters*, **29**(5), 221-229.

[58] Gaver, D.P., Miller, R.G. (1962). "Limiting Distributions for Some Storage Problems", in *Studies in Applied Probability and Management Science*, Arrow, K. J., Karlin, S., Scarf, H. (eds). Stanford University Press, Stanford Calif, 110-126.

[59] Gavish, E., Schweitzer, P. (1977). "The Markovian Queue with Bounded Waiting Time", *Management Science*, **23**(12), 1349-1357.

[60] Grassmann, W.K. (ed.) (2000). *Computational Probability*, Kluwer Academic Publishers, Boston.

[61] Green, L. (1978), *Queues Which Allow a Random Number of Servers Per Customer*, PhD Dissertation, Yale University.

[62] Green, L. (1980), "A Queueing System in Which Customers Require a Random Number of Servers", *Operations Research*, **28**(6), 1335-1346.

[63] Gross, D., Harris, C.M. (1998). *Fundamentals of Queueing Theory, 3^{rd} ed.*, Wiley, New York.

[64] Harris, C. M. (1966). *Queues with State Dependent Stochastic Service Rates*, PhD Dissertation, Polytechnic Institute of Brooklyn.

[65] Harris, C.M. (1967). "Queues with State Dependent Stochastic Service Rates", *Operations Research* **15**(1), 117-130.

[66] Harris, C.M., Brill, P.H., Fischer M. (2000). "Internet-type Queues with Power-tailed Interarrival Times and Computational Methods for their Analysis", *Informs Journal on Computing*, **12**(4), 261–271.

[67] Heyman, D.P., Sobel, M. (1982). *Stochastic Models in Operations Research, Volume I, Stochastic Processes and Operating Characteristics*, McGraw-Hill, New York.

[68] Hlynka, M. (2007). *Myron Hlynka's Queueing Theory Page,* at web site *http://www2.uwindsor.ca/~hlynka/queue.html.*

[69] Hlynka, M., Brill, P.H. (2007). "A Note on Stability in M/M/1 Queues with Reneging", Technical Report WMSR 07-09, Department of Mathematics and Statistics, University of Windsor.

[70] Hlynka, M., Brill, P.H. (2008). "A Result for a Counter Problem", Technical Report WMSR 08-01, Department of Mathematics and Statistics, University of Windsor.

[71] Huang, M.L., Brill P.H. (1993). "System Point Estimation of the Probability Distribution of the Waiting Time in Variations of M/G^B/1 Queues", American Statistical Association, Proceedings of the Statistical Computing Section, San Francisco, August 8-12.

[72] Huang, M.L., Brill P.H. (1999). "A Level Crossing Quantile Estimation Method", *Statistics and Probability Letters,* **45**, 111-119.

[73] Huang, M.L., Brill, P.H. (2004). "A Level Crossing Distribution Estimation Method", *Journal of Statistical Planning and Inference*, **124**(1), 45-62.

[74] Karlin, S., Taylor, H.M. (1975). *A First Course in Stochastic Processes, 2nd ed.*, Academic Press, New York.

[75] Keilson, J., Servi, L. D. (1986). "Oscillating Random Walk Models for GI/G/1 Vacation Systems with Bernoulli Schedules", *J. Applied Probability*, 23, 790-802.

[76] Keilson, J. (1965). *Green's Function Methods in Probability Theory*, Charles Griffen and Co. Ltd., London.

[77] Kendall, D.G. (1953). "Stochastic Processes Occurring in the theory of Queues and their Analysis by the Method of Embedded Markov Chains", *Annals of Mathematical Statistics*, **24**(3), 338-354.

[78] Kleinrock, L. (1975) *Queueing Systems Volume I: Theory*, John Wiley & Sons, New York.

[79] Leadbetter, M.R. (1972). "Point Processes Generated by Level Crossings", in *Stochastic Point Processes: Statistical Analysis, Theory and Applications, 436-467, Lewis, P. A. W. (ed.), Wiley-Interscience, New York.*

[80] LeCorbeiller P. (1966). *Dimensional Analysis*, Appleton-Century-Crofts, Div. Meredith Publishing Co., New York.

[81] Libura, M. (1971). "On a One-Channel Queueing System with Service Time Depending on Waiting Time", *Archiwum Automatuca Telemechaniki* **10**, 279-286.

[82] Lindley, D.V. (1952). "The Theory of Queues with a Single Server", *Proc. Cambridge Philosophical Soc.* **48,** 277-289.

[83] Lovitt, W.V. (1950). *Integral Equations*, Dover Publications Inc., New York.

[84] Lucantoni, D. M., Meier-Hellstern, K.S., Neuts, M.F. (1990). "A Single-server Queue with Server Vacations and a Class of Nonrenewal Arrival Processes", *Adv. in Appl. Probability*, **22**(3), 676-705.'

[85] Miyazawa, M. (1994). "Rate Conservation Laws: A Survey", *Queueing Systems,* **15**(1-4), 1-58.

[86] Neuts, M.F. (1981). *Matrix-geometric Solutions in Stochastic Models*, *An Algithmic Approach*, The John Hopkins university Press, Baltimore.

[87] Parzen, E. (1962). *Stochastic Processes*, Holden-Day, San Francisco.

[88] Posner, M. J., (1973). "Single Server Queues with Service Time Depending on Waiting Time", *Operations Research*, **21**(2), 610-616.

[89] Reichenbach, H. (1956), *The Direction of Time*, Dover Publications, Inc., New York.

[90] Ross, S.M. (1983). Stochastic Processes, Wiley, New York.

[91] Ross, S.M. (1997). Introduction to Probability Models, 6^{th}ed., Academic Press, San Diego.

[92] Shortle, J.F., Brill, P.H. (2005). "Analytical Distribution of Waiting Time in the M/{iD}/1 Queue", *Queueing Systems* ***50(2)***, 185-197.

[93] Shortle, J. F., Brill, P. H., Fischer, M. J., Gross, D., Masi, D. M. B. (2004), "An Algorithm to Compute the Waiting Time Distribution for the M/G/1 Queue", *INFORMS J. on Comput.* **16**(2), 152–161.

[94] Shortle, J.F., Fischer, M., Brill, P.H. (2007). "Waiting-Time Distribution of M/D_N/1 Queues Through Numerical Laplace Inversion", *INFORMS J. on Computing,* **19**(1), 112-120.

[95] Silverman, B.W. (1990), *Density Estimation in Statistics and Data Analysis*, Chapman and Hall, New York.

[96] Smith, W.L. (1955). "Regenerative Stochastic Processes", *Proc. Royal Soc. A*, **232**, 6-31.

[97] Spivak, M. (1‘965). *Calculus on Manifolds*, W. A. Benjamin, New York.

[98] Srivastava, H.M., Kashyap, B.R.K. (1982). *Special Functions in Queueing Theory*, Academic Press, New York.

[99] Takács, L. (1962). *Introduction to the Theory of Queues*, Oxford University Press, New York.

[100] Takács, L. (1967). *Combinatorial Methods in the Theory of Stochastic Processes*, John Wiley, New York.

[101] Taylor, H.M., Karlin S. (1998). *An Introduction to Stochastic Modelling, 3^{rd} ed.*, Academic Press, San Diego.

[102] Wolff, R.W. (1982), "Poisson Arrivals see Time Averages", *Operations Research*, **30(**2**)**, 223-231.

PARTIAL BIBLIOGRAPHY

The foregoing *References* lists publications that I referred to in some way when working on this monograph. The *Partial Bibliography* lists two types of publications. The first type involves those based on the level crossing theory and methods elucidated in this monograph (SPLC). The second involves publications that expound on other theoretical or applied aspects of level crossings; or contain models and ideas that can be potentially analyzed using SPLC. There are very extensive literatures on both types. Hence I have listed only a small sample of each.

Adler, R., Samorodnitsky, G. (1997). "Level Crossings of Absolutely Continuous Stationary Symmetric "alpha"-stable Processes", *The Annals of Applied Probability*, **7**(2), 460-493.

Adler, R.J., *The Geometry of Random Fields*, Wiley, New York, 1981.

Barakat, R. (1978). "The Distribution of Values of Trigonometric Sums with Linearly Independent Frequencies: Kac's Problem Revisited". *Stochastic Processes and their Applications*, **8**(1), 77-85.

Beekman, J.A. and Fuelling, C.P. (1990). "Interest and Mortality Randomness in Some Annuities", *Insurance: Mathematics & Economics*, **9**(2-3), 185-196.

Bekker, R., Borst, S.C., Boxma, O.J. and Kella, O. (2004). "Queues with Workload-dependent Arrival and Service Rates", *Queueing Systems: Theory and Applications*, **46**(3-4), 537-556.

Bekker, R. (2005). "Finite-buffer Queues with Workload-dependent Service and Arrival Rates", *Queueing Systems: Theory and Applications*, **50(**2-3**)**, 231-253.

Belyaev, Y.K. and Seleznjev, O. (2000). "Approaching in Distribution with Applications to Resampling of Stochastic Processes", *Scandinavian Journal of Statistics*, **27**(2), 371-384.

Berman, O., Parlar, M., Perry, P., Posner, M.J.M. (2005); "Production/Clearing Models Under Continuous and Sporadic Reviews" *Methodol. Comput. Appl. Probab.*, **7**(2), 203–224.

Berzin, C., Leon, J.R., Ortega, J. (1998). "Level Crossings and Local Time for Regularized Gaussian Processes", *Probability and Mathematical Statistics*, **18**(1), 39-81.

Besson, J.L. (1983). "Mean Number of the Level Crossings of Stochastic Processes with Absolutely Continuous Sample Paths", *Comptes Rendus de l'Académie des Sciences, Série I: Mathématique*, **297**, 635-637.

Chaouch, B.A. (2001). "Stock Levels and Delivery rates in Vendor-managed Inventory Programs", *Production and Operations Management*,**10**(1), 31-44.

De Boer, Pieter-Tjerk, Nicola, V. F. and Van Ommeren, Jan-Kees C.W. (2001). "The Remaining Service Time Upon Reaching a High Level in M/G/1 queues", *Queueing Systems: Theory and Applications*, **39**(1), 55-78.

Dshalalow, J.H. (1995). *Advances in Queueing: Theory, Methods, and Open Problems*, CRC Press.

Doshi, B. (1992). "Level-crossing Analysis of Queues" in *Queueing and Related Models*, Bhat, U. N., Basawa, I.V. (eds), Oxford Univ. Press, New York, 3–33.

El-Taha, M., Stidham Jr., S. (1998). *Sample-Path Analysis of Queueing Systems,* International Series in Operations Research and Management Science, Volume 11, Kluwer Academic Publishers, Boston.

Farahmand, K. (1990). "On the Average Number of Level Crossings of a Random Trigonometric Polynomial", *The Annals of Probability,* **18**(3), 1403-1409.

Farahmand, K. and Shaposhnikov, A. (2005). "On the Expected Number of Level Crossings of Random Trigonometric Polynomials", *Stochastic Analysis and Applications*, **23**(6), 1141-1147.

Feuerverger, A., Hall, P., Wood, A.T.A. (1994). "Estimation of Fractal Index and Fractal Dimension of a Gaussian Process by Counting the Number of Level Crossings", *Journal of Time Series Analysis*, **15**(6), 587-606.J.

Graves, S.C.; Keilson, J. (1981). "The Compensation Method Applied to a One-product Production/inventory Problem", *Mathematics of Oper. Res.*, **6**(2), 246-262.

He, Qi-Ming, Jewkes, E.M. (1997). "A Level Crossing Analysis of the MAP/G/1 Queue", *Matrix-Analytic Methods in Stochastic Models*, 107-116, Chakravarthy, Srinivas (ed.) and Alfa, Attahiru (ed.), Marcel Dekker Inc (New York).

Hébuterne, G., Hebuterne, G. and Rosenberg, C. (1999). "Arrival and Departure State Distributions in the General Bulk-service Queue", *Naval Research Logistics*, **46**(1), 107-118.‘

Hillier, F.S., Lieberman, G.J. (2004). *Introduction to Operations Research* (6^{th} ed.), McGraw Hill, New York.

Huang, M.L. (2001). "On a Distribution-free Quantile Estimator", *Computational Statistics & Data Analysis*, **37**(4), 477-486.

Illsley, R. (1998). "The Moments of the number of Exits from a Simply Connected Region", *Advances in Applied Probability*, **30**(1), 167-180.

Jewkes, E.M., Stanford, D.A. (2003). "A Two Priority Queue with Crossover Feedback", *Queueing Systems: Theory and Applications*, **43**(1-2), 129-146.

Kac, M. (1943). "On the Distribution of Values of Trigonometric Sums with Linearly Independent Frequencies", *American Journal of Mathematics*, **65**(4), 609-615.

Kaspi, H., Perry, D. (1989). "On a Duality Between a Non-Markovian Storage/production Process and a Markovian Dam Process with State-dependent Input and Output", *Journal of Applied Probability*, **26**(4), 835-844.

Katayama, T. (2002). "A Note on Level-Crossing Analysis for the Excess, Age, and Spread Distributions", Journal of Applied Probability, **39**(4), 896-900.

Katayama, T. (2005). "Level-crossing approach to a Time-limited Service System with Two Types of Vacations", *Operations Research Letters*, **33**(3), 295-300.

Kayama, T. (2007). "Analysis of a Nonpreemptive Priority Queue with Exponential Timer and Server Vacations", *Performance Evaluation,* **64**(6), 495-506.

Kim, S., Lee, E.Y. (2002). "A Level Crossing Approach to the Analysis of Finite Dam", *Journal of the Korean Statistical Society*, **31**(3), 405-413.

Lee, J., Kim, J. (2006). "A Workload-dependent M/G/1 Queue Under a Two-stage Service Policy", *Operations Research Letters,* **34**(5), 531-538

Liu, L., Kulkarni, V.G. (2006). "Explicit Solutions for the Steady State Distributions in M/PH/1 Queues with Workload Dependent Balking", *Queueing Systems,* **52**(4), 251-260.

Kroese, D.P. and Kallenberg, W.C.M. (1992). "Second-order Asymptotics in Level Crossing for Differences of Renewal Processes", *Stochastic Processes and their Applications*, **40**(2), 309-323.

Leadbetter, M.R. (1975). "Point Processes Generated by Level Crossings", in *Random Processes, II: Poisson and Jump-Point Processes*, Ephremides, A. (ed.), Dowden, Hutchinson & Ross Inc (Stroudsburg, PA), 109-140.

Leadbetter, M.R., Spaniolo, G.V. (2004). "Reflections on Rice's Formulae for Level Crossings – History, Extensions and Use", *Australian & New Zealand Journal of Statistics*, **46**(1), 173-180.

Lindgren, G., Rychlik, I. (1995). "How Reliable are Contour Curves? Confidence Sets for Level Contours", *Bernoulli*, **1**(4), 301-319.

Liu, B., Alfa, Attahiru S. (2002). "A Fluid Model with Data Message Discarding", *Advances in Applied Probability*, **34**(2), 329-348.

Mandelbaum, M. (1968). *Queueing with Splitting and Matching*, Masters Thesis, Technion, Israel Institute of Technology.

Mandelbaum, M., Avi-Itzhak, B. (1968). "Introduction to Queueing with Splitting and Matching", *Israel J. Technol.* **6**, 288-298.

Meyn, S.P„ Tweedie, R.L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag, New York.

Mohebbi, E., Posner, M.J.M. (1998). "A continuous-review Inventory System with Lost Sales and Variable Lead Time", *Naval Res. Logist,* **45**(3), 259–278.

Nahmias, S. (2004). *Production and Operations Analysis, (4*$^{\text{th}}$ ed.), Irwin Professional Pub, Burr Ridge, Illinois.

Neuts, M.F. (1989). *Structured Stochastic Matrices of M/G/1 Type and their Applications*, Probability: Pure and Applied, 5, Marcel Dekker, Inc., New York.

Posner, M.J.M., Berg, M. (1989), "Analysis of a Production-inventory System with Unreliable Production Facility", *Oper. Res. Lett.* **8**(6), 339–345.

Prabhu, N.U., Basawa, I.V. (eds.) (1990). *Statistical Inference in Stochastic Processes*, Marcel Dekker, Inc., New York.

Prabhu, N.U., (1965). *Stochastic Processes*, MacMillan, New York.

Rice, S. O. (1944). "Mathematical Analysis of Random Noise", *Bell Syst. Tech. J.*, **23**, 282-332.

Rice, S.O. (1945). "Mathematical Analysis of Random Noise", *Bell Syst. Tech. J.*, **24**, 46–156.

Rychlik, I. (2000). "On Some Reliability Applications of Rice's Formula for the Intensity of Level Crossings", *Extremes*, 3, 331-348.

Schagen, I. P. (1980). "A stochastic Model for the Occurrence of Oilfields and its Application to some North Sea Data", J. Roy. Statist. Soc. Ser. C **29**(3), 282–291.

Shanthikumar, J.G. (1981). "On Level Crossing Analysis of Queues", *The Australian Journal of Statistics*, **23**(3), 337-342.

Shanthikumar, J.G. (1988). "On Stochastic Decomposition of M/G/1 Type Queues with Generalized Server Vacations", *Operations Research*, **36**(4), 566-569.

van Harn, K., Steutel, F.W. (1992). "A Generalization of the Life -Time Pairs of Renewal Theory", Rapportnr WS-395, Vrieje Universiteit, Amsterdam, March.

Wu, P., Posner, M.J.M. (1997). "A Level-crossing Approach to the Solution of the Shortest-queue Problem", *Oper. Res. Lett.,* **21**(4), 181–189.

Zazanis, M.A. (1992), "Sample Path Analysis of Level Crossings for the Workload Process", *Queueing Systems Theory Appl.,* **11**(4), 419–428.

INDEX

Early Titles in the
INTERNATIONAL SERIES IN
OPERATIONS RESEARCH & MANAGEMENT SCIENCE
Frederick S. Hillier, Series Editor, *Stanford University*

Saigal/ *A MODERN APPROACH TO LINEAR PROGRAMMING*
Nagurney/ *PROJECTED DYNAMICAL SYSTEMS & VARIATIONAL INEQUALITIES WITH APPLICATIONS*
Padberg & Rijal/ *LOCATION, SCHEDULING, DESIGN AND INTEGER PROGRAMMING*
Vanderbei/ *LINEAR PROGRAMMING*
Jaiswal/ *MILITARY OPERATIONS RESEARCH*
Gal & Greenberg/ *ADVANCES IN SENSITIVITY ANALYSIS & PARAMETRIC PROGRAMMING*
Prabhu/ *FOUNDATIONS OF QUEUEING THEORY*
Fang, Rajasekera & Tsao/ *ENTROPY OPTIMIZATION & MATHEMATICAL PROGRAMMING*
Yu/ *OR IN THE AIRLINE INDUSTRY*
Ho & Tang/ *PRODUCT VARIETY MANAGEMENT*
El-Taha & Stidham/ *SAMPLE-PATH ANALYSIS OF QUEUEING SYSTEMS*
Miettinen/ *NONLINEAR MULTIOBJECTIVE OPTIMIZATION*
Chao & Huntington/ *DESIGNING COMPETITIVE ELECTRICITY MARKETS*
Weglarz/ *PROJECT SCHEDULING: RECENT TRENDS & RESULTS*
Sahin & Polatoglu/ *QUALITY, WARRANTY AND PREVENTIVE MAINTENANCE*
Tavares/ *ADVANCES MODELS FOR PROJECT MANAGEMENT*
Tayur, Ganeshan & Magazine/ *QUANTITATIVE MODELS FOR SUPPLY CHAIN MANAGEMENT*
Weyant, J./ *ENERGY AND ENVIRONMENTAL POLICY MODELING*
Shanthikumar, J.G. & Sumita, U./ *APPLIED PROBABILITY AND STOCHASTIC PROCESSES*
Liu, B. & Esogbue, A.O./ *DECISION CRITERIA AND OPTIMAL INVENTORY PROCESSES*
Gal, T., Stewart, T.J., Hanne, T. / *MULTICRITERIA DECISION MAKING: Advances in MCDM Models, Algorithms, Theory, and Applications*
Fox, B.L. / *STRATEGIES FOR QUASI-MONTE CARLO*
Hall, R.W. / *HANDBOOK OF TRANSPORTATION SCIENCE*
Grassman, W.K. / *COMPUTATIONAL PROBABILITY*
Pomerol, J-C. & Barba-Romero, S. */ MULTICRITERION DECISION IN MANAGEMENT*
Axsäter, S. */ INVENTORY CONTROL*
Wolkowicz, H., Saigal, R., & Vandenberghe, L. / *HANDBOOK OF SEMI-DEFINITE PROGRAMMING: Theory, Algorithms, and Applications*
Hobbs, B.F. & Meier, P. / *ENERGY DECISIONS AND THE ENVIRONMENT: A Guide to the Use of Multicriteria Methods*
Dar-El, E. / *HUMAN LEARNING: From Learning Curves to Learning Organizations*
Armstrong, J.S. / *PRINCIPLES OF FORECASTING: A Handbook for Researchers and Practitioners*
Balsamo, S., Personé, V., & Onvural, R./ *ANALYSIS OF QUEUEING NETWORKS WITH BLOCKING*
Bouyssou, D. et al. / *EVALUATION AND DECISION MODELS: A Critical Perspective*
Hanne, T. / *INTELLIGENT STRATEGIES FOR META MULTIPLE CRITERIA DECISION MAKING*
Saaty, T. & Vargas, L. / *MODELS, METHODS, CONCEPTS and APPLICATIONS OF THE ANALYTIC HIERARCHY PROCESS*
Chatterjee, K. & Samuelson, W. / *GAME THEORY AND BUSINESS APPLICATIONS*
Hobbs, B. et al. / *THE NEXT GENERATION OF ELECTRIC POWER UNIT COMMITMENT MODELS*
Vanderbei, R.J. / *LINEAR PROGRAMMING: Foundations and Extensions, 2nd Ed.*
Kimms, A. / *MATHEMATICAL PROGRAMMING AND FINANCIAL OBJECTIVES FOR SCHEDULING PROJECTS*
Baptiste, P., Le Pape, C. & Nuijten, W. / *CONSTRAINT-BASED SCHEDULING*
Feinberg, E. & Shwartz, A. / *HANDBOOK OF MARKOV DECISION PROCESSES: Methods and Applications*

Early Titles in the
INTERNATIONAL SERIES IN
OPERATIONS RESEARCH & MANAGEMENT SCIENCE
(Continued)

Ramík, J. & Vlach, M. / *GENERALIZED CONCAVITY IN FUZZY OPTIMIZATION AND DECISION ANALYSIS*
Song, J. & Yao, D. / *SUPPLY CHAIN STRUCTURES: Coordination, Information and Optimization*
Kozan, E. & Ohuchi, A. / *OPERATIONS RESEARCH/ MANAGEMENT SCIENCE AT WORK*
Bouyssou et al. / *AIDING DECISIONS WITH MULTIPLE CRITERIA: Essays in Honor of Bernard Roy*
Cox, Louis Anthony, Jr. / *RISK ANALYSIS: Foundations, Models and Methods*
Dror, M., L'Ecuyer, P. & Szidarovszky, F. / *MODELING UNCERTAINTY: An Examination of Stochastic Theory, Methods, and Applications*
Dokuchaev, N. / *DYNAMIC PORTFOLIO STRATEGIES: Quantitative Methods and Empirical Rules for Incomplete Information*
Sarker, R., Mohammadian, M. & Yao, X. / *EVOLUTIONARY OPTIMIZATION*
Demeulemeester, R. & Herroelen, W. / *PROJECT SCHEDULING: A Research Handbook*
Gazis, D.C. / *TRAFFIC THEORY*
Zhu/ *QUANTITATIVE MODELS FOR PERFORMANCE EVALUATION AND BENCHMARKING*
Ehrgott & Gandibleux/ *MULTIPLE CRITERIA OPTIMIZATION: State of the Art Annotated Bibliographical Surveys*
Bienstock/ *Potential Function Methods for Approx. Solving Linear Programming Problems*
Matsatsinis & Siskos/ *INTELLIGENT SUPPORT SYSTEMS FOR MARKETING DECISIONS*
Alpern & Gal/ *THE THEORY OF SEARCH GAMES AND RENDEZVOUS*
Hall/*HANDBOOK OF TRANSPORTATION SCIENCE - 2nd Ed.*
Glover & Kochenberger/ *HANDBOOK OF METAHEURISTICS*
Graves & Ringuest/ *MODELS AND METHODS FOR PROJECT SELECTION: Concepts from Management Science, Finance and Information Technology*
Hassin & Haviv/ *TO QUEUE OR NOT TO QUEUE: Equilibrium Behavior in Queueing Systems*
Gershwin et al/ *ANALYSIS & MODELING OF MANUFACTURING SYSTEMS*
Maros/ *COMPUTATIONAL TECHNIQUES OF THE SIMPLEX METHOD*
Harrison, Lee & Neale/ *THE PRACTICE OF SUPPLY CHAIN MANAGEMENT: Where Theory and Application Converge*
Shanthikumar, Yao & Zijm/ *STOCHASTIC MODELING AND OPTIMIZATION OF MANUFACTURING SYSTEMS AND SUPPLY CHAINS*
Nabrzyski, Schopf & Węglarz/ *GRID RESOURCE MANAGEMENT: State of the Art and Future Trends*
Thissen & Herder/ *CRITICAL INFRASTRUCTURES: State of the Art in Research and Application*
Carlsson, Fedrizzi, & Fullér/ *FUZZY LOGIC IN MANAGEMENT*
Soyer, Mazzuchi & Singpurwalla/ *MATHEMATICAL RELIABILITY: An Expository Perspective*
Chakravarty & Eliashberg/ *MANAGING BUSINESS INTERFACES: Marketing, Engineering, and Manufacturing Perspectives*
Talluri & van Ryzin/ *THE THEORY AND PRACTICE OF REVENUE MANAGEMENT*
Kavadias & Loch/*PROJECT SELECTION UNDER UNCERTAINTY: Dynamically Allocating Resources to Maximize Value*
Brandeau, Sainfort & Pierskalla/ *OPERATIONS RESEARCH AND HEALTH CARE: A Handbook of Methods and Applications*
Cooper, Seiford & Zhu/ *HANDBOOK OF DATA ENVELOPMENT ANALYSIS: Models and Methods*
Luenberger/ *LINEAR AND NONLINEAR PROGRAMMING, 2nd Ed.*

Early Titles in the
INTERNATIONAL SERIES IN
OPERATIONS RESEARCH & MANAGEMENT SCIENCE
(Continued)

Sherbrooke/ *OPTIMAL INVENTORY MODELING OF SYSTEMS: Multi-Echelon Techniques, Second Edition*
Chu, Leung, Hui & Cheung/ *4th PARTY CYBER LOGISTICS FOR AIR CARGO*
Simchi-Levi, Wu & Shen/ *HANDBOOK OF QUANTITATIVE SUPPLY CHAIN ANALYSIS: Modeling in the E-Business Era*
Gass & Assad/ *AN ANNOTATED TIMELINE OF OPERATIONS RESEARCH: An Informal History*
Greenberg/ *TUTORIALS ON EMERGING METHODOLOGIES AND APPLICATIONS IN OPERATIONS RESEARCH*
Weber/ *UNCERTAINTY IN THE ELECTRIC POWER INDUSTRY: Methods and Models for Decision Support*
Figueira, Greco & Ehrgott/ *MULTIPLE CRITERIA DECISION ANALYSIS: State of the Art Surveys*
Reveliotis/ *REAL-TIME MANAGEMENT OF RESOURCE ALLOCATIONS SYSTEMS: A Discrete Event Systems Approach*
Kall & Mayer/ *STOCHASTIC LINEAR PROGRAMMING: Models, Theory, and Computation*
Sethi, Yan & Zhang/ *INVENTORY AND SUPPLY CHAIN MANAGEMENT WITH FORECAST UPDATES*
Cox/ *QUANTITATIVE HEALTH RISK ANALYSIS METHODS: Modeling the Human Health Impacts of Antibiotics Used in Food Animals*
Ching & Ng/ *MARKOV CHAINS: Models, Algorithms and Applications*
Li & Sun/ *NONLINEAR INTEGER PROGRAMMING*
Kaliszewski/ *SOFT COMPUTING FOR COMPLEX MULTIPLE CRITERIA DECISION MAKING*

**** A list of the more recent publications in the series is at the front of the book ****