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.

P.H. Brill, *Level Crossing Methods in Stochastic Models*,
DOI: 10.1007/978-0-387-09421-2_5,

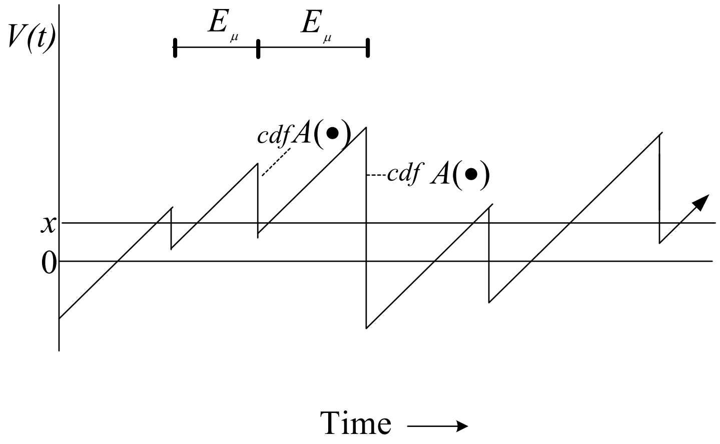

Figure 5.1: Sample path of extended age process $\{V(t)\}$ for G/M/1 queue. Inter-arrival times have cdf $A(\cdot)$ (cdf of downward jump sizes). Service times are $\underset{dist}{=}$ E_μ. Slope is $\frac{dV(t)}{dt} = +1$.

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).

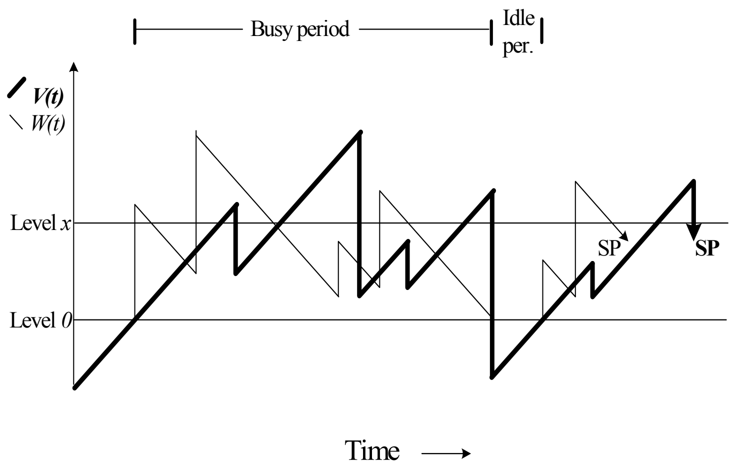

Figure 5.2: Sample path of extended age process "↗" compared with sample path of virtual wait process "↘" for G/M/1 queue. Illustarates duality properties. Corresponding peaks and corresponding troughs have equal ordinates and the same time order. Busy periods, idle periods, and busy cycles ar equal.

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

$$F(x) = \lim_{t \to \infty} P(V(t) \leq x), -\infty < x < \infty,$$

having pdf

$$\begin{aligned} f(x) &= \tfrac{dF(x)}{dx}, x \geq 0; \\ h(x) &= \tfrac{dF(x)}{dx}, x < 0, \end{aligned} \tag{5.2}$$

wherever the derivatives exist. The probability of an empty system is

$$P_0 = F(0) = \int_{y=-\infty}^{0} h(y)dy. \tag{5.3}$$

Then

$$F(x) = P_0 + \textstyle\int_{y=0}^{x} f(y)dy, x \geq 0,$$
$$F(x) = \textstyle\int_{y=-\infty}^{x} h(y)dy, x \leq 0,$$
$$F(0) = P_0,$$
$$F(\infty) = P_0 + \textstyle\int_{y=0}^{\infty} f(y)dy = 1.$$

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

$$\begin{aligned} \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x)) \cdot E(\mathcal{B})}{t} &= \lim_{t\to\infty} \frac{E(\mathcal{U}_t(x))}{t} \cdot E(\mathcal{B}) \\ &= f(x) \cdot E(\mathcal{B}) = 1 - F(x), \end{aligned} \tag{5.8}$$

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

$$\begin{aligned} F(x) &= 1 - (1 - P_0)e^{-\frac{1}{E(\mathcal{B})}x}, x \geq 0, \\ f(x) &= \frac{1 - P_0}{E(\mathcal{B})} e^{-\frac{1}{E(\mathcal{B})}x}, x > 0, \end{aligned} \tag{5.9}$$

where $F(0) \equiv P_0$.

From (5.9) $f(x)$ has the exponential form

$$f(x) = Ke^{-\gamma x}, x \geq 0 \tag{5.10}$$

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 $F_\iota(x) = P(\text{actual wait} \le x), x \ge 0$; 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

$$f_\iota(x) = K_\iota e^{-\gamma x} = \gamma(1 - \frac{\gamma}{\mu})e^{-\gamma x}, x > 0, \tag{5.25}$$

$$F_\iota(x) = 1 - (1 - \frac{\gamma}{\mu})e^{-\gamma x}, x \geq 0. \tag{5.26}$$

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 $y > 0$,

$$\begin{aligned} \frac{1}{\mu} &< \int_{y=0}^{\infty} \overline{A}(y) dy = \frac{1}{\boldsymbol{a}} \\ &\Longrightarrow \boldsymbol{a} < \mu. \end{aligned}$$

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

$$\begin{aligned} P_0 &= \frac{\gamma}{\gamma + C_\gamma} = \frac{\mu - \lambda}{\mu - \lambda + \lambda} = 1 - \frac{\lambda}{\mu}. \checkmark \\ K &= \frac{\gamma \cdot C_\gamma}{\gamma + C_\gamma} = \lambda(1 - \frac{\lambda}{\mu}) = \lambda P_0, \checkmark \\ f(x) &= Ke^{-\gamma x} = \lambda P_0 e^{-(\mu - \lambda)x}, x > 0, \checkmark \\ E(\mathcal{B}) &= \frac{1}{\gamma} = \frac{1}{\mu - \lambda}, \checkmark \\ E(\mathcal{I}) &= \frac{1}{C} = \frac{1}{\lambda}. \checkmark \end{aligned}$$

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 $x < 0$

$$\begin{aligned} E(a_x) &= E(u_x) - E(b_x) = \tfrac{1}{h(x)} - E(b_x) \\ &= \tfrac{1}{\mu K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y}dy} \\ &\quad - \textstyle\int_{z=0}^{\infty}\int_{y=0}^{\infty} \tfrac{\overline{A}(y-x+z)}{\overline{A}(y-x)} K e^{-\gamma y}dydz, x < 0, \end{aligned} \tag{5.38}$$

where b_x is the sojourn time below x. The last term in (5.38),

$$E(b_x) = \int_{z=0}^{\infty}\int_{y=0}^{\infty} \frac{\overline{A}(y-x+z)}{\overline{A}(y-x)} K e^{-\gamma y}dydz$$

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

$$E(b_x) = E(u_x) - E(a_x) = \frac{e^{\gamma x}}{K} - \frac{1}{\gamma}, x \geq 0.$$

Levels < 0 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.

Number of System Times During a_x

Let $N^{\sigma}_{a_x}$ denote the number of customers *completing* service during a *sojourn* of $\{V(t)\}$ above level $x \geq 0$. Thus $N^{\sigma}_{a_x}$ 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)

$$N^{\sigma}_{a_x} = \min\left\{ n | \sum_{i=1}^{n} (S_i - T_i) \le 0 \right\}, x \ge 0.$$

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 $x \ge 0$

$$\begin{aligned} E(a_x) &= E(N^{\sigma}_{a_x})E(S_i), \\ E(N^{\sigma}_{a_x}) &= \frac{E(a_x)}{E(S_i)} = \frac{E(\mathcal{B})}{E(S)}. \end{aligned} \tag{5.40}$$

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, $N^{\sigma}_{\mathcal{B}} \underset{dist}{=} N^{\sigma}_{a_x}, x \ge 0$. 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 $x \ge 0$. Then $N^{w}_{a_x}$ is the number of customers that *wait* in line $> x$ (strictly) during $a_x, x \ge 0$. 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 $\le x$. Hence

$$E(N^{w}_{a_x}) = E(N^{\sigma}_{a_x}) - 1 = \frac{\mu}{\gamma} - 1 > 0, x \ge 0. \tag{5.43}$$

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 $N_{a_x}^{\mathrm{soj}}(\mathcal{C})$, $N_{a_x}^{\mathrm{soj}}(\mathcal{B})$ be the number of SP *sojourns* above level x during a busy *cycle* and busy period, respectively. Then $N_{a_x}^{\mathrm{soj}}(\mathcal{C})\underset{dist}{=}N_{a_x}^{\mathrm{soj}}(\mathcal{B})$, 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 $N_{a_x}^{\mathrm{soj}}(\mathcal{C})\underset{dist}{=}\mathcal{U}_{\mathcal{C}}(x)$. By the theory of regenerative processes, specific time averages in a busy cycle recapitulate the same specific limiting time averages (e.g., [96]). Thus

$$\frac{E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{C})\right)}{E(\mathcal{C})}=\frac{E(\mathcal{U}_{\mathcal{C}}(x))}{E(\mathcal{C})}=\lim_{t\to\infty}\frac{E(\mathcal{U}_t(x))}{t}=f(x),x\geq 0,\qquad(5.44)$$

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)

$$\begin{aligned}E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{B})\right)&=E\left(N_{a_x}^{\mathrm{soj}}(\mathcal{C})\right)=E(\mathcal{U}_{\mathcal{C}}(x))=E(\mathcal{C})\cdot f(x)\\&=\frac{1}{K}\cdot Ke^{-\gamma x}=e^{-\gamma x},x\geq 0.\end{aligned}\qquad(5.45)$$

Setting $x=0$ in (5.45) implies

$$E\left(N_{a_0}^{\mathrm{soj}}(\mathcal{B})\right)=E\left(N_{a_0}^{\mathrm{soj}}(\mathcal{C})\right)=1.$$

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

$$N^{\sigma}_{a_x}(\mathcal{C}) = \sum_{i=1}^{N^{\mathrm{soj}}_{a_x}(\mathcal{C})} N^{\sigma}_{a_x i} \tag{5.46}$$

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 $N^{\mathrm{soj}}_{a_x}(\mathcal{C})$ above x (memoryless property of service time). Taking expected values in (5.46) and using (5.45) gives

$$E\left(N^{\sigma}_{a_x}(\mathcal{B})\right) = E\left(N^{\sigma}_{a_x}(\mathcal{C})\right) = E\left(N^{\mathrm{soj}}_{a_x}(\mathcal{C})\right) \cdot E\left(N^{\sigma}_{a_x i}\right) = \frac{\mu}{\gamma} e^{-\gamma x}. \tag{5.47}$$

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 $M_\lambda/M_\mu/1$ (see Fig. 3.6). The expected number of system times completed in a sojourn above level x is

$$E\left(N^\sigma_{a_x}\right) = E\left(N^\sigma_{\mathcal{B}}\right) = \frac{\mu}{\mu-\lambda} = \frac{1}{P_0}, x \geq 0, \tag{5.51}$$

where $N^\sigma_{\mathcal{B}}$ is the number served in a busy period, independent of x. Equality $E\left(N^\sigma_{a_x}\right) = E\left(N^\sigma_{\mathcal{B}}\right)$ follows because in M/M/1, $a_x \underset{dist}{=} \mathcal{B}, x \geq 0$. Also, $E\left(N^\sigma_{a_x}\right) > 1$ since $\mu > \mu - \lambda$ for stability (i.e., $0 < \lambda < \mu$).

The expected number that wait $> x$ in a_x is

$$E(N^w_{a_x}) = \frac{\mu}{\mu-\lambda} - 1 = \frac{1}{P_0} - 1 = \frac{\lambda}{\mu-\lambda} \geq 0. \tag{5.52}$$

The expected number of sojourns above x in a busy period is

$$\begin{aligned} E(N^{\text{soj}}_{a_x}(\mathcal{B})) &= E(N^{\text{soj}}_{a_x}(\mathcal{C})) = E(\mathcal{C}) \cdot f(x) \\ &= \frac{1}{\lambda P_0} \cdot \lambda P_0 e^{-(\mu\,-\,\lambda)x} = e^{-(\mu-\lambda)x}, x \geq 0. \end{aligned} \tag{5.53}$$

If $x = 0$ then

$$E\left(N^{\text{soj}}_{a_x}(\mathcal{B})\right) \;=\; e^0 = \; E\left(N^{\text{soj}}_{a_0}(\mathcal{B})\right) = 1.$$

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

$$N^\sigma_{a_x}(\mathcal{B}) = N^\sigma_{a_x}(\mathcal{C}) = \sum_{i=1}^{N^{\text{soj}}_{a_x}(\mathcal{C})} N^\sigma_{a_x i}.$$

By (5.51) and (5.53),

$$\begin{aligned} E\left(N^\sigma_{a_x}(\mathcal{B})\right) &= E\left(N^\sigma_{a_x}\right) = E\left(N^\sigma_{\mathcal{B}}\right) \cdot E\left(N^{\text{soj}}_{a_x}(\mathcal{C})\right) \\ &= \frac{\mu}{\mu-\lambda} \cdot e^{-(\mu\,-\,\lambda)x}, \geq 0. \end{aligned} \tag{5.54}$$

The expected number of waiting times $> x$ in $\mathcal{B}$ is

$$\begin{aligned} E(N^w_{a_x}(\mathcal{B})) &= \left(\frac{\mu}{\mu-\lambda} - 1\right) \cdot e^{-(\mu\,-\,\lambda)x} \\ &= \frac{\lambda}{\mu-\lambda} \cdot e^{-(\mu\,-\,\lambda)x}, x \geq 0. \end{aligned} \tag{5.55}$$

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

$$0 < \mu \int_{y=x}^{\infty} A(y - x) f(y) dy < \mu \int_{y=x}^{\infty} 1 \cdot f(y) dy = \mu (1 - F(x)),$$

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 ($c = 2, 3, \ldots$) 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

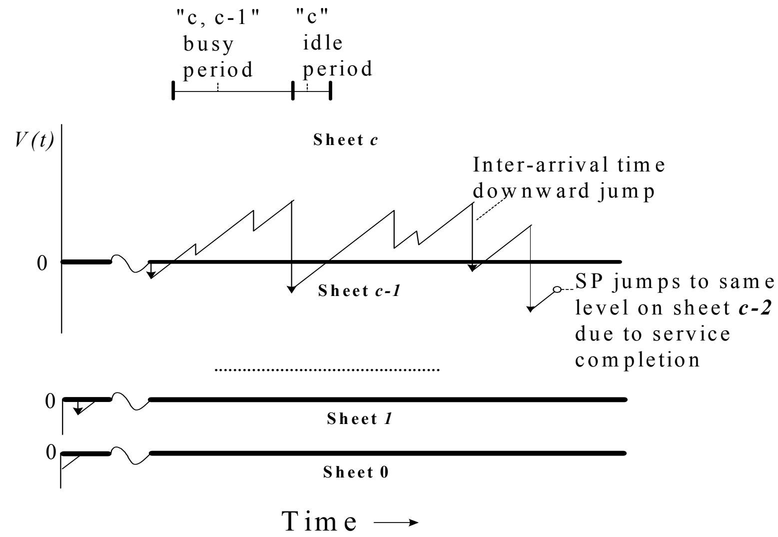

Figure 5.3: Sample path of $\{V(t), M(t)\}$ for G/M/c queue. There are $c+1$ sheets. Range of sheet c is $[0,\infty)$. Range of sheets $0,...,c-1$ is $(-\infty,0)$. Sheet $c-1$ abuts on sheet c for geometric convenience. Time between jumps originating on sheet $c = \mathrm{E}_{c\mu}$.

$\mathcal{S} = \min\{S_1,...,S_c\}$ 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 $\mathcal{S} \underset{dist}{=} E_{c\mu}$.

Relationship Between $V(t)$ and $M(t)$

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 $E\left(\mathcal{B}_{c-1,c}\right) = E\left(a_0\right)$.

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

$$\begin{aligned}
\lim_{t\to\infty} \frac{E\left(\mathcal{U}_t(x)\right)\cdot E(a_x)}{t} &= \lim_{t\to\infty} \frac{E\left(\mathcal{U}_t(x)\right)}{t}\cdot E(a_x) \\
&= f(x)\cdot E(\mathcal{B}_{c-1,c}), x \geq 0.
\end{aligned}$$

Similarly as for G/M/1 in Subsection 5.1.5, we have

$$\begin{aligned}
f(x)E(\mathcal{B}_{c-1,c}) &= 1 - F(x), \\
\frac{d}{dx}\ln\left(1-F(x)\right) &= -\frac{1}{E(\mathcal{B}_{c-1,c})}.
\end{aligned}$$

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*

$$\begin{aligned} F(0) &= \frac{E(\textit{sojourn time among sheets } 0, ..., c-1)}{E(\textit{time between entrances to sheet } c)} \\ &= \frac{\frac{1}{K} - E(\mathcal{B}_{c-1,c})}{\frac{1}{K}} = \frac{\frac{1}{K} - \frac{1}{\gamma}}{\frac{1}{K}} = 1 - \frac{K}{\gamma}. \end{aligned} \tag{5.70}$$

Value of K

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 $F_\iota(x) = P(\text{actual wait} \leq x)$; 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

$$\begin{aligned}
&\lim_{t\to\infty} \frac{N_t^{>0}}{t} \underset{a.s.}{=} \lim_{t\to\infty} \frac{E\left(N_t^{>0}\right)}{t} \\
&= c\mu \int_{y=0}^{\infty} A(y) f(y) dy \\
&= c\mu \int_{y=0}^{\infty} A(y) K e^{-\gamma y} dy \\
&= c\mu \int_{y=0}^{\infty} \left(1 - \overline{A}(y)\right) K e^{-\gamma y} dy \\
&= \frac{c\mu}{\gamma} K - K,
\end{aligned} \tag{5.80}$$

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

$$\begin{aligned} F(0) &\equiv \int_{x=-\infty}^{0} (g_1(x) + g_0(x))dx \\ &= 2\mu K \int_{x=-\infty}^{0} \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y} dy dx + g_0(0)\frac{1}{\boldsymbol{a}}, \end{aligned} \tag{5.85}$$

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

$$g_1(x) = \frac{g_0'(x)}{\mu}. \tag{5.86}$$

Substituting (5.86) and (5.83) into (5.82) gives a differential equation for $g_0(x)$

$$g_0'(x) + \mu g_0(x) = 2\mu^2 K \int_{y=0}^{\infty} \overline{A}(y-x)e^{-\gamma y} dy + +\mu g_0(0)\overline{A}(-x), x < 0. \tag{5.87}$$

The solution of (5.87) is

$$\begin{aligned} g_0(x) = 2\mu^2 K e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z} \int_{y=0}^{\infty} \overline{A}(y-z)e^{-\gamma y} dy dz \\ + \mu g_0(0)e^{-\mu x} \int_{z=-\infty}^{x} e^{\mu z}\overline{A}(-z)dz, x < 0, \end{aligned} \tag{5.88}$$

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

$$\begin{aligned} g_0(0) &= \left(\frac{2\mu^2 \int_{u=0}^{\infty} e^{-\mu u} \int_{y=0}^{\infty} \overline{A}(y+u)e^{-\gamma y} dy du}{1 - \mu \int_{u=0}^{\infty} e^{-\mu u}\overline{A}(u)du,} \right) K \\ &\equiv H_0 \cdot K. \end{aligned} \tag{5.89}$$

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

$$P_0 + P_1 = \frac{(2\mu-\lambda)(\lambda+\mu)}{\lambda\mu+2\mu^2}. \tag{5.101}$$

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

$$F_\iota(0) = \frac{(2\mu - \lambda)(\lambda + \mu)}{\lambda\mu + 2\mu^2} \quad (5.102)$$

in agreement with (5.101).

The pdf is

$$\begin{aligned} f_\iota(x) &= K_\iota e^{-\gamma x} = \gamma(1 - F_\iota(0))\, e^{-(2\mu-\lambda)x} \\ &= \lambda P_1 e^{-(2\mu-\lambda)x}, x > 0, \end{aligned} \quad (5.103)$$

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].