Queueing Systems with Impatient Customers - Applications in Call Center

Magdalena Lalik

24.06.2002

Contents

1 Introduction **2**

2 Single Server Queues **4**

2.1 Assumptions and notation 4

2.2 Techniques for establishing ergodic properties for continuous-valued Markov chains 7

2.3 Stability conditions for the actual offered waiting time 8

2.4 Stationary distribution of $\{w_n\}_{n \in N}$ 15

2.5 On the virtual offered waiting time 16

2.6 Stability conditions for the complete convergence of the virtual offered waiting time distribution function 20

2.7 Coincidence of distribution functions V(x) and W(x) in case of Poisson arrivals 24

2.8 Density of the stationary v.o.w.t. distribution function 26

2.9 Quantities of practical interest 28

3 Multi Server Queues **32**

3.1 Fundamental relations for $M \mid M \mid m + GI$ queueing systems . . 32

3.2 Quantities of practical interest 35

3.3 On the $M \mid M \mid m + D$ queue 38

3.4 On the $M \mid M \mid m + M$ queue 39

A The Laplace-Stieltjes Transform **41**

A.1 The Stieltjes integral 41

A.2 The Laplace transform 44

B Moment Generating Functions **45**

Chapter 1

Introduction

Speaking in generalities, the theory of queues deals with the investigation of the stochastic law of different processes arising in connection with mass servicing in case random fluctuation occurs. Here are some examples of such processes.

Mass servicing process. Let us suppose that customers are arriving at a counter according to some probabilistic law. There are one or more servers, that serve the customers in the order of arrival. If every server is busy then the customer joins a queue or waiting line. Generally, the service times are random variables. We speak about a single server queue when there is only one server, and multi server queueing process if there is more than one server.

Several queueing processes belong to this category. For example, the process of calls in telephone exchange, systems the moving of equipment in production lines, the landing of airplanes at an airport, the arrival of ships in a harbor, railway traffic, road traffic, and many others. However, nowadays one of the most common examples of applications of the queueing theory can be found in telephone call centers.

A call center is a group of resources (typically agents and ICT equipment) capable of delivering services by telephone. One can find it as an integral part of many service companies such as airlines, hotels, retail banks, and credit card companies. Call centers and their contemporary successors, contact centers, are a preferred and prevalent means for these companies to communicate with their customers. The functions that they provide vary highly: from customer service, help desk, and emergency response services, to tele-marketing and order taking. Depending on the type of telephone traffic we can distinguish call centers into inbound or outbound ones. Inbound call centers handle incoming calls that are initiated by outside callers calling in to a center. An operational scheme of such a call center can be described as follows. Calls are arriving at a telephone exchange system accordingly to a certain stochastic law. There is a fixed number of available agents. If a call find a free agent then a connection is realized. The lengths of the holding times are random variables. If all agents are busy, then the incoming call is either lost or joins the queue and awaits its turn. Impatient customers may abandon the queue before the connection they ask for is completely established. Among callers that do not abandon the queue, the queueing discipline is first-come, first-served.

In this paper we shall consider queueing systems with limitation acting only on the customer's patience and we shall skip technical limitations, i.e., no customer is lost due to sufficiently large buffers. In Chapter 2 we shall study $GI \mid GI \mid 1 + GI$ queues (**notation;** the first three symbols have the same meaning as in Kendal's notation, i.e., they denote respectively: the type of arrivals, the service mechanism, the number of servers. The last one specifies the impatience law). We start with defining recurrence equations for the actual waiting time in case of abandonments. Then, in section (2.2) we give a brief introduction to techniques for establishing ergodicity of continuous valued Markov chains. This part is mainly based on ($[3, Tweedie]$) and we derive only formulas that are needed to show that the waiting time $\{w_n\}$ defined in section (2.1) is ergodic under some conditions. However, one can find a more detailed description of such techniques in ($[3, Tweedie]$). In the next sections we derive sufficient and necessary conditions for stability of the queueing system. Finally, in section (2.4) we derive formulas for the stationary distribution of $\{w_n\}$. In the next section we define the virtual offered waiting time (v.o.w.t.) and also derive conditions for its stability. Knowing stationary distributions of both actual and virtual waiting times, we show in section (2.7) that they coincide in case of Poisson arrivals. The last section of Chapter 2 is devoted to derive quantities of interest such as the mean waiting time, the mean queue length, the probability of rejection and the Pollaczek-Khinchin formula for queues with abandonments. In Chapter 3 we study multi server queueing systems with Poisson arrivals and exponential service times. In section (3.1) we derive fundamental relations for $M \mid M \mid m + GI$ queues and find results analogue to those in Chapter 2. The important part of this section is the explicit formula for the density of the virtual offered waiting time that allows us to compute quantities of interest, which is done in the next section. Finally, in sections (3.3) and (3.4) we give some exemplary formulas in case of respectively, deterministic and exponential patience distributions.

Chapter 2

Single Server Queues

2.1 Assumptions and notation

Input process

Throughout this paper we shall consider the following type of input process. We shall always suppose that the time t ranges over the interval $[0,\infty)$. Let us denote by $T_1, T_2, ..., T_n, ...$ the arrival instants customers. We will suppose that the inter-arrival times $t_n := T_n - T_{n-1}$ $(n = 1, 2, ...; T_0 = 0)$ are mutually independent, positive random variables with distribution function

$$P\left(t_n \leq x\right) = F\left(x\right), \qquad (n = 1, 2, ...).$$

If specifically we assume that

$$F\left(x\right) = \begin{cases} 1 - e^{-\lambda x} & if \quad x \geq 0, \\ 0 & if \quad x < 0, \end{cases}$$

then $\{T_n\}$ is said to be a Poisson process.

We will use the following notation. The expectation

$$\phi\left(s\right) = E\left[e^{-st_n}\right], \qquad (n = 1, 2, ...)$$

always exists if $\operatorname{Re}\left(s\right) \geq 0$. This can be written also as

$$\phi\left(s\right) = \int_0^\infty e^{-sx} dF\left(x\right),$$

and is called the Laplace-Stieltjes transform of distribution function $F\left(x\right)$.

The expectation

$$\frac{1}{\lambda} = E\left[t_n\right],$$

always exists but we will consider only distribution functions for which it is finite. $\frac{1}{\lambda}$ is the average inter-arrival time which can be written in the following form

$$\frac{1}{\lambda} = \int_0^\infty x dF\left(x\right).$$

Service mechanism

Generally, we shall consider the case "first come, first served". We shall denote by m the number of servers.

The service times are supposed to be identically distributed, independent, positive random variables, independent of the input process. The service time of the nth customer will be denoted by s_n. We shall define

$$P(s_n \leq x) = H(x),$$

as the distribution function of the service times. The Laplace-Stieltjes transform of $H(x)$ is denoted by

$$\psi(x) = E\left[e^{-s_n x}\right] = \int_0^\infty e^{-xu} dH(u),$$

which is convergent if $\mathrm{Re}(x) \geq 0$. The average service time will be denoted by

$$\frac{1}{\mu} = \int_0^\infty u dH(u).$$

We shall assume that $\frac{1}{\mu}$ is finite.

An important particular case is that in which the service time has an exponential distribution, i.e.,

$$H(x) = \begin{cases} 1 - e^{-\mu x} & if \;\; x \geq 0, \\ 0 & if \;\; x < 0. \end{cases}$$

Patience

A basic description of patience is the distribution of the time beyond which a customer is not willing to wait. Throughout this paper we will concern queueing systems with impatient customers, where a limitation acts on the waiting time. There could be two types of such systems

(a) aware customers:
The entering customer leaves immediately if he knows that his waiting time is above his patience.

(b) unaware customers:
The entering customers do not know anything about the system and leave the system when his time already spent in the queue reach the limit of his patience.

We shall mainly study the type (a) but by the following remark, it is possible to unify both systems in some cases.

Remark 2.1 *The unfinished work of the server is not modified by customers who finally leave impatiently, even if they stay in the queue.*

The correctness of this statement will be made clear in section (2.5).

Hence, as long as we are interested in the rejection probabilities, or in the waiting time distributions of successful customers, we can identify system (b) with system (a).

We shall denote by g_n the patience of the nth customer and define

$$P(g_n \leq x) = C(x),$$

where $C(x)$ may be defective, i.e., we may have

$$\lim_{x\to\infty} C(x) \neq 1,$$

but we assume $C(0) = 0$.
However, we will mainly use

$$G(x) := 1 - C(x).$$

The expected patience is given by

$$\gamma = \int_0^\infty x dC(x),$$

and we assume it to be finite.

Let $\{w_n\}_{n\in N}$ be the workload just before T_n (unfinished work). We assume the system to be of type (a), i.e., the nth customer enters the system only if the time to wait for accessing the server does not exceed his patience.
That is,

- If $g_n \leq w_n$ the nth customer is impatient and does not enter the system;
- If $g_n > w_n$ the nth customer joins a queue.

Recursive equations for the actual offered waiting time

Now, we derive a recursive equation for the sequence $\{w_n\}_{n\in N}$, where w_n is the time that the nth customer would have to wait for accessing the server if he were sufficiently patient. Hence, we call it the actual offered waiting time. Let $w_0 \in R^+$ be some initial condition, we have for $n \geq 0$:

$$\begin{cases} w_{n+1} = [w_n + s_n - t_{n+1}]^+ & if \;\; g_n > w_n \\ w_{n+1} = [w_n - t_{n+1}]^+ & if \;\; g_n \leq w_n \end{cases} \tag{2.1}$$

Remark 2.2 *It is important to note that above definition yields $\{w_n\}$ is a Markov chain with state space R^+.*

2.2 Techniques for establishing ergodic properties for continuous-valued Markov chains

Consider a time-homogeneous Markov chain $\{X_n\}$ with state space Υ, which is usually assumed to be a closed subset of R. The evolution of the chain is described by the collection of distribution functions

$$F_x(y) = P(X_{n+1} \leq y \mid X_n = x),$$

however it is easier to work with the corresponding measure:

$$P(x, A) = P(X_{n+1} \in A \mid X_n = x)$$

induced in one dimension by the distributions F_x;
that is, if $A = (a, b]$, then $P(x, A) = F_x(b) - F_x(a)$ and $P(x, \cdot)$ is then extended to all Borel subsets of R.
We assume, in order that the chain be well defined, that for each $A \in \mathcal{B}(\Upsilon)$ (where $\mathcal{B}(\Upsilon)$ denotes the $\sigma-$field of Borel subsets of Υ) the function $P(\cdot, A)$ is measurable, and for each x, $P(x, \cdot)$ is a probability measure on $\mathcal{B}(\Upsilon)$.

Definition 2.3 *We say $\{X_n\}$ is $\phi-$irreducible if there exists a nonzero measure ϕ on $\mathcal{B}(\Upsilon)$ such that, for any $x \in \Upsilon$ and $A \in \mathcal{B}(\Upsilon)$ with $\phi(A) > 0$ there is an $n \in N$ for which $P^n(x, A) > 0$.*

Definition 2.4 *We shall call a ϕ-irreducible chain $\{X_n\}$ ergodic if it has a unique stationary distribution, i.e., a probability measure π on $\mathcal{B}(\Upsilon)$ satisfying:*

$$\forall A \in \mathcal{B}(\Upsilon) \qquad \pi(A) = \int P(x, A) d\pi(x).$$

If $\{X_n\}$ is $\phi-$irreducible then (see $[3, Tweedie]$) there is at most one stationary distribution. Additionally, if $\{X_n\}$ is ergodic then the n-step transition probabilities converge to the stationary distribution π in the strong Cesaro sense that:

$$\sup_{A \in \mathcal{B}(\Upsilon)} \left| \frac{1}{n} \sum_{m=1}^{n} P^m(x, A) - \pi(A) \right| \to 0$$

for $\pi-$almost all x

There is a close relationship between ergodicity and the finiteness of the means of hitting times T_A, where:

$$T_A := \inf\{n > 0 \mid X_n \in A\}, \qquad A \in \mathcal{B}(\Upsilon).$$

It can be shown that, given certain conditions on the chain, ergodicity is a consequence of:

$$\sup E\{T_A \mid X_0 = x\} < \infty, \qquad \forall x \in \Upsilon. \tag{2.2}$$

provided A is one of a certain class of sets determined by preliminary conditions satisfied by the chain.

Definition 2.5 *We call any set A such that (2.2) is a sufficient condition for ergodicity of $\{X_n\}$ a test set for that chain.*

Theorem 2.6 *Suppose $\Upsilon = [0, \infty)$ and there exist $\varepsilon > 0$, $M > 0$, and a bounded $A \in \mathcal{B}(\Upsilon)$ such that*

$$E\{X_1 \mid X_0 = x\} \leq x - \varepsilon, \qquad \forall x \in A^c, \tag{2.3}$$

and

$$E\{X_1 \mid X_0 = x\} \leq M, \qquad \forall x \in A. \tag{2.4}$$

Then

$$\sup E\{T_A \mid X_0 = x\} < \infty.$$

Proof. $[3, Tweedie]$ ■

Definition 2.7 *We say that ϕ has an atom at α when $\{\alpha\}$ can be reached from every point in the state space.*

Theorem 2.8 *If ϕ has an atom at α then the set B containing α is a test set if for some integer N and some $\delta > 0$*

$$\max_{n \leq N} P^n(y, \{\alpha\}) \geq \delta, \qquad \forall y \in B.$$

Proof. *[3, Tweedie]* ■

So to prove ergodicity for a given Markov chain $\{X_n\}$ carry out the following three steps:

STEP1 Identify a suitable ϕ and show $\{X_n\}$ is ϕ−irreducible for this ϕ.

STEP2 Identify possible test sets for the chain.

STEP3 Apply Theorem2.6 to one of these test sets to prove boundedness of the mean hitting times as specified by (2.2).

2.3 Stability conditions for the actual offered waiting time

Sufficient condition

In this section we will follow these three steps to show that Markov chain defined by (2.1) is ergodic.
Define:

$$a := \inf\{t : F(t) = 1\},$$

$$b := \sup\{t : H(t) = 0\}.$$

Notice that b is always well-defined, but a may not. The case when $a = \infty$ will be discussed separately.
We can interpret a as "the longest" time we have to wait for a customer to arrive and b as "the shortest" time in which we can complete the service of a customer.
Throughout this section we assume $b - a < 0$. This condition is necessary to avoid infinitely large queues.

STEP1

Lemma 2.9 *The Markov chain $\{w_n : n \in N\}$ is ϕ_0−irreducible, where ϕ_0 is a measure on R^+ with an atom at* 0.

Proof
Consider the sequence:

$$\begin{cases} z_0 = w_0, \\ z_{n+1} = \max(0, z_n + s_n - t_{n+1}) \end{cases} \tag{2.5}$$

comparing with (2.1) we get:

$$w_n \leq z_n, \qquad \forall n \in N.$$

Since

$$\begin{aligned} P(w_{n+1} = 0) &= P(w_n + s_n \chi_{(g_n > w_n)} - t_{n+1} \leq 0) \\ &\geq P(w_n + s_n - t_{n+1} \leq 0) = P(z_{n+1} = 0), \end{aligned}$$

where

$$\chi_{(g_n > w_n)}(\omega) = \begin{cases} 1 & \text{if } g_n(\omega) > w_n(\omega), \\ 0 & \text{if } g_n(\omega) \leq w_n(\omega), \end{cases} \tag{2.6}$$

we obtain:

$$P(w_n = 0) \geq P(z_n = 0), \qquad \forall n \in N. \tag{2.7}$$

By definition of a and b we get

$$\forall \varepsilon > 0 \; \exists \, p > 0 : \qquad P(b - a \leq s_n - t_{n+1} \leq b - a + \varepsilon) = p.$$

Fix $\varepsilon > 0$ and let $x \in R^+$ and $k := \left[\frac{x}{|b-a|}\right]$ (where $[y]$ for $y \in R$ denotes the smallest integer greater than y).
So, x satisfies the following inequalities:

$$-(k-1)(b-a) < x \leq -k(b-a) \tag{2.8}$$

Now, lets consider the events:

$$E := \bigcap_{0\le i\le k} \{b-a \le s_i - t_{i+1} \le b-a+\varepsilon\}$$

and

$$E^{'} := \{z_k = 0 \mid z_0 = x\}.$$

Claim 2.10 *The following inclusion holds*

$$E \subset E^{'}. \tag{2.9}$$

Proof of claim

Take $\omega \in E$, and observe that:

$$\begin{aligned} z_0(\omega) &= x, \\ z_1(\omega) &= \max\{0, z_0(\omega) + s_0(\omega) - t_1(\omega)\}, \\ z_2(\omega) &= \max\{0, z_1(\omega) + s_1(\omega) - t_2(\omega)\} \\ &= \max\{0, \max\{0, x + s_0(\omega) - t_1(\omega)\} + s_1(\omega) - t_2(\omega)\} \\ &= \max\{0, x + s_0(\omega) + s_1(\omega) - (t_1(\omega) + t_2(\omega))\}. \end{aligned}$$

Induction yields:

$$z_k(\omega) = \max\left\{0, x + \sum_{j=0}^{k-1}(s_j(\omega) - t_{j+1}(\omega))\right\}. \tag{2.10}$$

By definition of set E we have the following inequality:

$$\forall \varepsilon > 0, \qquad \sum_{j=0}^{k-1}(s_j(\omega) - t_{j+1}(\omega)) \le k(b-a) + k\varepsilon.$$

Using relation (2.8) we get:

$$\forall \varepsilon > 0, \qquad \sum_{j=0}^{k-1}(s_j(\omega) - t_{j+1}(\omega)) \le -x + k\varepsilon.$$

Letting $\varepsilon \searrow 0$ we obtain:

$$x + \sum_{j=0}^{k-1}(s_j(\omega) - t_{j+1}(\omega)) \le 0.$$

The above inequality and formula (2.10) eventually give:

$$z_k(\omega) = 0, \qquad \forall\, \omega \in E.$$

Hence, inclusion (2.9) is true. This ends the proof of claim.
Notice now that both events can happen with positive probability

$$P(E') \geq P(E) > p^k > 0.$$

Observe also that by (2.7) the following inequality holds:

$$P(w_n = 0 \mid w_0 = x) \geq P(z_n = 0 \mid z_0 = x), \qquad \forall n \in N.$$

Hence we have just shown that for any x there exists k such that:

$$P(w_k = 0 \mid w_0 = x) \geq P(z_k = 0 \mid z_0 = x) > 0$$

i.e.,

$$\forall\, x \geq 0\; \exists\, k \in N : P^k(x, \{0\}) > 0.$$

It thus becomes clear that the pair $\{w_n, \phi_0\}$ satisfies the Definition2.3.

<u>**Special case**</u> $(a = \infty)$
Observe that preceding proof is also valid when $a = \infty$. We define then $k = 1$ for every x. It becomes clear when we notice that the situation $a = \infty$ means that with positive probability no arrival occurs. Hence, there is also positive probability for the system to reach state zero when beginning at the initial state x.

STEP2

Lemma 2.11 *For every $\beta \in R^+$ the interval $B = [0, \beta]$ is a test set*

Proof. By Theorem2.8 we must find N and $\delta > 0$ such that

$$\max_{n \leq N} P^n(y, \{0\}) \geq \delta, \qquad \forall y \in B.$$

From STEP1 we know that for every $y \in B$,

$$P^n(y, \{0\}) = P(w_n = 0 \mid w_0 = y) \geq P(z_n = 0 \mid z_0 = y),$$

so any $N \geq [\frac{\beta}{|b-a|}]$ matches. ■

STEP3

Let T_B be the hitting time of B, i.e.,

$$T_B = \inf\{n \geq 0 \mid w_n \in B\},$$

where B is a test set as in STEP 2.
To show that

$$\sup E\{T_B \mid w_0 = x\} < \infty$$

we have to make an additional assumption:
Let $\rho := \frac{\lambda}{\mu}$, and assume

$$1 - \rho G(\infty) > 0. \tag{2.11}$$

Now we will apply Theorem2.6
At first derive from (2.1) the formula for the conditional expectation:

$$\begin{aligned} E\{w_1 \mid w_0 = x\} &= G(x)E\left\{[w_0 + s_0 - t_1]^+ \mid w_0 = x\right\} \\ &\quad + (1 - G(x))E\left\{[w_0 - t_1]^+ \mid w_0 = x\right\} \\ &= G(x)\iint_{R^+\times R^+}[x + s - t]^+ dH(s)dF(t) \\ &\quad + (1 - G(x))\int_{R^+}[x - t]^+ dF(t), \end{aligned}$$

and so

$$\begin{aligned} E\{w_1 \mid w_0 = x\} = &\int_{R^+}[x - t]^+ dF(t) \\ &+ G(x)\iint_{R^+\times R^+}([x + s - t]^+ - [x - t]^+)dH(s)dF(t) \end{aligned} \tag{2.12}$$

By the use of the formula for integration by parts for the Stieltjes integral (see AppendixA TheoremA.3) we have:

$$\int_0^\infty [x - t]^+ dF(t) = \int_0^x F(t)\,dt,$$

$$\int_0^\infty \int_0^\infty [x + s - t]^+ dH(s)dF(t) = -\int_0^\infty \int_0^\infty F(t)\,d\left([x + s - t]^+\right).$$

To simplify the last equation it is enough to notice that:

$$[x + s - t]^+ = x + s - t \quad \text{iff} \quad \left\{ \begin{array}{c} x - t > 0 \\ s > 0 \end{array} \right. \text{ or } \left\{ \begin{array}{c} x - t < 0 \\ s > t - x \end{array} \right. \quad \forall\, x, s, t \geq 0.$$

Hence,

$$\int_0^\infty \int_0^\infty [x + s - t]^+ dH(s)dF(t) = \int_0^x F(t)\,dt + \int_{t-x}^\infty \int_x^\infty F(t)\,dtdH(s).$$

Finally, some inversions in (2.12) give:

$$E\{w_1 \mid w_0 = x\} = \int_0^x F(t)dt + G(x)\int_x^\infty (1 - H(t-x))F(t)dt.$$

Now, we will show that $E\{w_1 \mid w_0 = x\}$ satisfies the conditions of Theorem2.6

Consider $x \in B$, i.e., $x \leq \beta$.
Since $F(t) \leq 1 \ \forall\ t \geq 0$ and $G(x) \leq 1 \ \forall\ x \geq 0$, it follows

$$\begin{aligned} E\{w_1 \mid w_0 = x\} &\leq x + \int_x^\infty (1 - H(t-x))dt \\ &= x + \int_0^\infty (1 - H(u))du \leq \beta + \frac{1}{\mu}. \end{aligned}$$

So the condition (2.4) is fulfilled with $M := \beta + \frac{1}{\mu} > 0$.

For $x \notin B$, i.e., $x > \beta$ we have

$$\begin{aligned} E\{w_1 \mid w_0 = x\} &= x - \int_0^x (1 - F(t))dt + G(x)\int_x^\infty (1 - H(t-x))F(t)dt \\ &\leq x - \int_0^\infty (1 - F(t))dt + \int_x^\infty (1 - F(t))dt \\ &+ G(x)\int_0^\infty (1 - H(t))dt \\ &= x - \frac{1}{\lambda} + G(x)\frac{1}{\mu} + \int_x^\infty (1 - F(t))dt. \end{aligned}$$

Observe now that for every $\varepsilon > 0$ there exists an x_0 such that

$$\forall x > x_0 : \int_x^\infty (1 - F(t))dt < \varepsilon.$$

Hence,

$$E\{w_1 \mid w_0 = x\} \leq x - \frac{1 - \rho G(x)}{\lambda} + \varepsilon.$$

Now, having assumed that $1 - \rho G(\infty) > 0$ we can find $x_1 \in R^+$ such that

$$\forall x > x_1, \qquad 1 - \rho G(x) > 2\varepsilon.$$

Hence if $\beta > \max(x_0, x_1)$, then we obtain precisely condition (2.3), i.e.,

$$E\{w_1 \mid w_0 = x\} \leq x - \varepsilon$$

and this implies that $[0, \beta]$ is a test set with bounded mean hitting time.
This completes the proof of the ergodicity of $\{w_n\}_{n \in N}$.
Notice that (2.11) becomes then a sufficient condition for w_n to be ergodic.

Necessary condition

Lemma 2.12 *If w_n is ergodic, then*

$$0 \leq 1 - \rho G(\infty).$$

Proof. By the definition of w_n we can write

$$w_{n+1} = [w_n + s_n \chi_{(g_n > w_n)} - t_{n+1}]^+,$$

where $\chi_{(g_n > w_n)}$ is the indicator function defined by (2.6).
Observe now that the inclusion of the events

$$\{\omega : g_n(\omega) = \infty\} \subset \{\omega : g_n(\omega) > w_n(\omega)\}$$

yields:

$$w_{n+1} \geq [w_n + s_n \chi_{(g_n = \infty)} - t_{n+1}]^+$$

Hence,

$$w_{n+1} \geq w_0 + \sum_{j=0}^{n} b_i, \tag{2.13}$$

where

$$b_i := s_i \chi_{(g_i = \infty)} - t_{i+1}. \tag{2.14}$$

In the further part of the proof we will show that condition

$$0 > 1 - \rho G(\infty) \tag{2.15}$$

is in contradiction with ergodicity of w_n .

Suppose (2.15) holds, then it is equivalent to

$$\lambda G(\infty) > \mu \tag{2.16}$$

where $\frac{1}{\lambda}, \frac{1}{\mu}$ were assumed to be finite.
Since

$$Eb_i = \frac{G(\infty)}{\mu} - \frac{1}{\lambda},$$

then

$$0 < Eb_i < \infty. \tag{2.17}$$

Notice that b_i satisfies conditions of the Strong Law of Large Numbers, i.e.,

$$\frac{\sum_{i=0}^{n} b_i}{n} \to E[b_1] \qquad (n \to \infty) \qquad \textit{with probability } 1.$$

Suppose now that the series $\sum_{i=0}^{\infty} b_i$ converges, then

$$\frac{\sum_{i=0}^{n} b_i}{n} \to 0 \qquad (n \to \infty) \qquad \textit{with probability } 1$$

which is the contradiction with (2.17).
Hence

$$\sum_{i=1}^{n} b_i \to \infty \qquad (n \to \infty) \qquad \textit{with probability } 1.$$

Comparing now last statement with (2.13) gives the required contradiction with ergodicity of w_n. ■

2.4 Stationary distribution of $\{w_n\}_{n \in N}$

For the Markov chain defined by (2.1) we have

$P(w_{n+1} \in A \mid w_n = x) =$
$P(w_{n+1} \in A \mid w_n = x, g_n \le x) P(g_n \le x) + P(w_{n+1} \in A \mid w_n = x, g_n > x) P(g_n > x),$

where

$$P(w_{n+1} \in A \mid w_n = x,\ g_n \le x) = \int_{R^+} \chi_A([x - z]^+) dF(z)$$

and

$$P(w_{n+1} \in A \mid w_n = x,\ g_n > x) = \iint_{R^+ \times R^+} \chi_A([x + y - z]^+) dH(y) dF(z)$$

for $\chi_A : R \to \{0, 1\}$ - a characteristic function of set A.

Finally we can write

$$\begin{aligned} P(x, A) = G(x) &\int_{R^+ \times R^+} \chi_A([x + y - z]^+) dH(y) dF(z) \\ + (1 - G(x)) &\int_{R^+} \chi_A([x - z]^+) dF(z) \end{aligned} \tag{2.18}$$

where $P(x, A)$ is the measure defined in section (2.2)

Remark 2.13 *Let X, Y be random variables, then*

$$\begin{aligned} F_{X+Y}(a) &:= P(X + Y \le a) = \int_R F_X(a - y) dF_Y(y), \\ F_{X-Y}(a) &:= P(X - Y \le a) = \int_R F_X(a + y) dF_Y(y) \\ &= \int_R (1 - F_Y(x - a)) dF_X(x), \end{aligned}$$

where

$$\begin{aligned} F_X(a) &:= P(X \le a), \\ F_Y(a) &:= P(Y \le a). \end{aligned}$$

Let $W_n(x)$, $x \in R^+$ be the distribution function of w_n, then the following relation holds

$$W_{n+1}(x) = P(w_{n+1} \leq x) = P(w_{n+1} \leq x, g_n > w_n) + P(w_{n+1} \leq x, g_n \leq w_n).$$

By Remark (2.13) we get

$$\begin{aligned} W_{n+1}(x) = & \int_0^\infty G(u) \int_0^\infty H(x+y-u)\, dF(y)\, dW_n(u) \\ & + \int_0^\infty (1 - G(u))(1 - F(u-x))\, dW_n(u). \end{aligned} \tag{2.19}$$

We have shown in section (2.3) that Markov chain $\{w_n\}$ is ergodic provided that $1 - \rho G(\infty) > 0$. Hence when this condition is fulfilled there exists a non-defective distribution function $W(x)$ such that

$$\lim_{n \to \infty} W_n(x) =: W(x), \qquad \forall x \in R^+.$$

By the Theorem $A.2$ (AppendixA) we obtain for every positive R

$$\begin{aligned} & \lim_{n \to \infty} \int_0^R \int_0^\infty G(u) H(x+y-u)\, dF(y)\, dW_n \\ & = \int_0^R \int_0^\infty G(u) H(x+y-u)\, dF(y)\, dW_n \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} & \lim_{n \to \infty} \int_0^R (1 - G(u))(1 - F(u-x))\, dW_n(u) \\ & = \int_0^R (1 - G(u))(1 - F(u-x))\, dW(u) \end{aligned} \tag{2.21}$$

Hence, by definition of the improper Stieltjes integral and formulas (2.20) and (2.21) we derive the expression for stationary distribution of actual offered waiting time

$$\begin{aligned} W(x) = & \int_0^\infty \int_0^\infty G(u) H(x+y-u)\, dF(y)\, dW(u) \\ & + \int_0^\infty (1 - G(u))(1 - F(u-x))\, dW(u). \end{aligned} \tag{2.22}$$

2.5 On the virtual offered waiting time

Denote by $\eta(t)$ the virtual offered waiting time (v.o.w.t.) at the instant t ; i.e., $\eta(t)$ is the time that a test customer of infinite patience would wait if he joined

the queue at the time t.
The virtual offered waiting time can also be interpreted as follows: $\eta(t)$ is the time at the instant t needed to complete the serving of all those customers who joined the queue before t. Hence $\eta(t)$ can also be seen as "unfinished work" of the server.

Let $V(t,x)$ be the distribution function of $\eta(t)$ and $\Omega(t,s)$ - the Laplace Stieltjes transform of $V(t,x)$; *i.e.*,

$$\begin{aligned} V(t,x) &:= P(\eta(t) \leq x), \\ \Omega(t,s) &:= \int_0^\infty e^{-sx} d_x V(t,x). \end{aligned}$$

Assume Poisson arrivals:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & if \;\; x \geq 0, \\ 0 & if \;\; x < 0. \end{cases} \tag{2.23}$$

Functional equation for the workload distribution function

Denote by $\delta_{\Delta t}$ the number of customers arriving at the counter during the time interval $(t, \Delta t]$. By assumption (2.23)

$$P(\delta_{\Delta t} = j) = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^j}{j!}, \qquad (j = 0, 1, 2, ...).$$

Evaluating suitable functions in McLaurin series we derive:

$$\begin{aligned} &P(\delta_{\Delta t} = 0) = e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(t), \\ &P(\delta_{\Delta t} = 1) = e^{-\lambda \Delta t} \lambda \Delta t = \lambda \Delta t + o(t), \\ &P(\delta_{\Delta t} > 1) = 1 - (P(\delta_{\Delta t} = 0) + P(\delta_{\Delta t} = 1)) = o(t). \end{aligned} \tag{2.24}$$

Using the theorem of total probability we can write that:

$$\Omega\left(t + \Delta t, s\right) = E\left\{e^{-s\eta(t+\Delta t)}\right\} = \sum_{j=0}^{\infty} P\left(\delta_{\Delta t} = j\right) E\left\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = j\right\}. \tag{2.25}$$

The equations (2.24) simplify (2.25) to be of the form

$$\Omega(t,s) = (1 - \lambda \Delta t) E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 0\} + \lambda \Delta t E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 1\} + o(\Delta t). \tag{2.26}$$

Now we shall compute the conditional expectations $E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 0\}$ and $E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 1\}$.

Consider situation when $\delta_{\Delta t} = 0$.
It means that during interval $(t, \Delta t)$ there was no arrival, so the v.o.w.t. decreased and, at the moment $t + \Delta t$ is equal to

$$\eta(t+\Delta t) = \begin{cases} \eta(t) - \Delta t & if \;\; \eta(t) > \Delta t, \\ 0 & if \;\; \eta(t) \le \Delta t. \end{cases}$$

Hence,

$$E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 0\} = P(\eta(t) \le \Delta t) + P(\eta(t) > \Delta t) E\{e^{-s(\eta(t)-\Delta t)}\}.$$

Since $V(t, x)$ is right continuous in x

$$V(t, \Delta t) = V(t, 0) + O(\Delta t).$$

By the Law of the Mean for Stieltjes integrals (see Theorem$A.4$ in AppendixA)

$$0 \le \int_0^{\Delta t} x d_x V(t, x) = \Delta t[V(t, \Delta t) - V(t, 0)] = o(\Delta t).$$

Thus after some calculations,

$$E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 0\} = (1 + s\Delta t)\Omega(t, s) - s\Delta t V(t, 0) + o(\Delta t) \qquad (2.27)$$

Let now $\delta_{\Delta t} = 1$, i.e., in the interval $(t, \Delta t]$ one customer arrives at the counter. Because of the impatience there are two possibilities of his behavior:

1. Suppose he decides to join the queue at the moment t_1; $t < t_1 \le \Delta t$ (i.e., $g_{t_1} > \eta(t_1)$; where g_{t_1} is the patience of the customer)
   Then, his service time is added to the v.o.w.t. at time t_1.
   Hence at the instant $t + \Delta t$ the v.o.w.t. equals to:

   Let $\eta(t) = y$:

   $$\begin{array}{ll} \text{if} & y > \Delta t \\ \text{then} & \eta(t+\Delta t) = y + s_{t_1} - \Delta t, \end{array}$$

   where s_{t_1} is the service duration of the customer

   $$\begin{array}{ll} \text{if} & y \le \Delta t \\ \text{then} & \eta(t+\Delta t) = s_{t_1} - \varepsilon_t \Delta t \end{array}$$

   where $0 \le \varepsilon_t < 1$, and ε_t is the time that customer has been already served during the time interval $(\max(\Delta t - y, t_1), \Delta t]$.

2. Suppose that the customer decides not to join the queue (i.e., $g_{t_1} \le \eta(t)$).

   Let $\eta(t) = y$,

   $$\begin{array}{ll} \text{if} & y > \Delta t, \\ \text{then} & \eta(t+\Delta t) = y - \Delta t. \end{array}$$

$$\begin{array}{ll} \text{if} & y \leq \Delta t, \\ \text{then} & \eta(t+\Delta t) = 0. \end{array}$$

After putting through all these possibilities we derive

$$E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 1, \eta(t) = y\} = G(y) \int_0^\infty e^{-s(y+x-\Delta t)} dH(x)$$
$$+ G(y) \int_0^\infty e^{-s(x-\varepsilon_t \Delta t)} dH(x) + (1-G(y))e^{-s(y-\Delta t)} + (1-G(y)).$$

Dropping the condition $\eta(t) = y$, we obtain

$$\begin{aligned} E\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 1\} &= \int_0^\infty G(y) \int_0^\infty e^{-s(y+x-\Delta t)} dH(x) d_y V(t,y) \\ &+ \int_0^\infty G(y) \int_0^\infty e^{-s(x-\varepsilon_t \Delta t)} dH(x) d_y V(t,y) \\ &+ \int_0^\infty (1-G(y)) e^{-s(y-\Delta t)} d_y V(t,y) \\ &+ \int_0^\infty (1-G(y)) d_y V(t,y) \end{aligned} \tag{2.28}$$

Let $\psi(s)$ be a Laplace-Stieltjes transform of the service time, i.e.,

$$\psi(s) = \int_0^\infty e^{-sx} dH(x) .$$

Then (2.28) can be written as

$$\begin{aligned} E\left\{e^{-s\eta(t+\Delta t)} \mid \delta_{\Delta t} = 1\right\} &= \int_0^\infty G(y) \psi(s) e^{-s(y-\Delta t)} d_y V(t,y) \\ &+ \int_0^\infty G(y) \psi(s) e^{s\varepsilon_t \Delta t} d_y V(t,y) \\ &+ \int_0^\infty (1-G(y)) e^{-s(y-\Delta t)} d_y V(t,y) \\ &+ \int_0^\infty (1-G(y)) d_y V(t,y) \end{aligned} \tag{2.29}$$

Finally, applying (2.27) and (2.29) to formula for $\Omega(t+\Delta t, s)$ given by (2.25) we obtain

$$\begin{aligned}\Omega(t+\Delta t, s) &= (1-\lambda\Delta t)[(1+s\Delta t)\int_0^\infty e^{-sy}d_yV(t,y) - s\Delta tV(t,0) + o(\Delta t)]\\ &+ \lambda\Delta t[e^{s\Delta t}\psi(s)\int_0^\infty G(y)e^{-sx}d_yV(t,y) + \psi(s)e^{s\varepsilon_t\Delta t}\int_0^\infty G(y)d_yV(t,y)\\ &+ e^{s\Delta t}\int_0^\infty(1-G(y))e^{-sy}d_yV(t,y) + \int_0^\infty(1-G(y))d_yV(t,y))]\\ &= \int_0^\infty e^{-sy}d_yV(t,y) + s\Delta t\int_0^\infty e^{-sy}d_yV(t,y) - s\Delta tV(t,0)\\ &+ \lambda\Delta te^{s\Delta t}(\psi(s)-1)\int_0^\infty G(y)e^{-sy}d_yV(t,y) + o(\Delta t).\end{aligned}$$

Denote

$$\Omega_G(t,s) := \int_0^\infty G(y)e^{-sy}d_yV(t,y).$$

Then

$$\Omega(t+\Delta t, s) = \Omega(t,s) + s\Delta t\Omega(t,s) - s\Delta tV(t,0) - \lambda\Delta te^{s\Delta t}(1-\psi(s))\Omega_G(t,s) + o(\Delta t).$$

By taking the limit for $\Delta t \to 0$, we obtain a very important differential equation, called the functional equation for workload distribution function

$$\frac{\partial\Omega(t,s)}{\partial t} = \lim_{\Delta t\to 0}\frac{\Omega(t+\Delta t,s)-\Omega(t,s)}{\Delta t} = s\Omega(t,s) - sV(t,0) - \lambda(1-\psi(s))\Omega_G(t,s),$$

i.e.,

$$\frac{1}{s}\frac{\partial\Omega(t,s)}{\partial t} = \Omega(t,s) - V(t,0) - \frac{\lambda(1-\psi(s))}{s}\Omega_G(t,s). \qquad (2.30)$$

2.6 Stability conditions for the complete convergence of the virtual offered waiting time distribution function

Necessary condition

Assume the existence of a limit

$$V(x) := \lim_{t\to\infty} V(t,x),$$

and let $\Omega(s)$ be a Laplace-Stieltjes transform of $V(x)$, i.e.,

$$\Omega(s) = \int_0^\infty e^{-sx}dV(x).$$

Then, by (2.30) $\Omega(s)$ will be the solution of :

$$\Omega(s) = V(0) + a(s)\,\Omega_G(s), \qquad \mathrm{Re}(s) \geq 0, \tag{2.31}$$

where

$$a(s) := \frac{\lambda\,(1-\psi(s))}{s}. \tag{2.32}$$

Observe now, that since $G(x) \leq 1$, $\forall x \geq 0$ we have

$$0 < \Omega_G(s) \leq \Omega(s) \leq 1, \qquad \forall s \geq 0.$$

One can also check that

$$a(s) < 1, \qquad \forall s \geq 0. \tag{2.33}$$

Hence, by formula (2.31) we get

$$V(0) \geq \Omega(s)\,(1 - a(s)),$$

and so, the condition (2.33) yeilds that $V(0)$ is strictly positive.

We now restrict s to take real values, hence Ω and Ω_G are also real and the following inequalities hold:

$$\Omega(s)\,G(\infty) = \int_0^\infty G(\infty)e^{-sx}dV(x) \leq \int_0^\infty G(x)e^{-sx}dV(x) = \Omega_G(s) \leq \Omega(s) \leq 1.$$

So multiplying the above expression by $a(s)$ we get

$$a(s)\,\Omega(s)\,G(\infty) \leq a(s)\,\Omega_G(s). \tag{2.34}$$

Moreover, observe that

$$V(0) = \Omega(s) - a(s)\,\Omega_G(s) \leq \Omega(s)\,[1 - a(s)\,G(\infty)].$$

We have shown that $V(0) > 0$, so the fact that $\forall s : 0 < \Omega(s) \leq 1$ implies

$$\forall s : 1 - a(s)\,G(\infty) > 0. \tag{2.35}$$

Observe now, that $a(s)$ by definition equals to

$$a(s) = \lambda\frac{1 - \int_0^\infty e^{-sx}dH(x)}{s} = \lambda\frac{\psi(0) - \psi(s)}{s}. \tag{2.36}$$

Hence

$$\lim_{s\to 0} a(s) = -\lambda\psi'(0).$$

Because $\psi(s)$ is the moment generating function (see AppendixB) for the service time distribution, we get

$$-\lambda\psi'(0) = \lambda E s_1 = \frac{\lambda}{\mu} = \rho.$$

The function $a(s)$ is continuous so by taking the limit for $s \to 0$ in inequality (2.35) we get

$$1 - \rho G(\infty) > 0. \tag{2.37}$$

So we have just shown that $1 - \rho G(\infty) > 0$ is the necessary condition for $V(x)$ to exist.

Sufficient condition

To derive the sufficient condition for existence of the steady-state distribution of the v.o.w.t. we use the fact that the discrete Markov chain $\{w_n : n \in N\}$ is imbedded in the continuous time Markov process $\{\eta(t) : t \in R\}$:

$$w_n = \eta\left(t_n^-\right). \tag{2.38}$$

The result will be a consequence of limit theorem for semi-regenerative processes.

Definition 2.14 *Consider a stochastic process $\{X(t) : t \geq 0\}$, and suppose that with probability 1, there exists a time T_1 such that continuation of the process beyond T_1 is a probabilistic replica of the whole process starting at 0. Such a stochastic process is known as a regenerative process.*
Note that this property implies the existence of further times $T_2, T_3, ...$ having the same property as T_1.
It may happen that distribution of T_1 differs from $T_2, T_3...$ then we call such a process semi-regenerative.

For both semi- and regenerative processes the following theorem is valid:

Theorem 2.15 *If F, the distribution of a cycle has a density over some interval and $ET_1 < \infty$*
then

$$\lim_{t\to\infty} P(X(t) = j) = \frac{E\{\textit{Amount of time in state } j \textit{ during a cycle}\}}{E\{\textit{Time of a cycle}\}}.$$

Proof. [5, Ross] ■
One can check that $w_n = \eta(t_n^-)$ is in fact semi-regenerative.

Now define the probability of being in set B during a cycle if we start from an arbitrary point $x \geq 0$.

$$K_t(x, B) := P\left(\eta(t) \in B, T_1 > t \mid \eta\left(0^+\right) = x\right), \qquad B \in \mathcal{B}\left(R^+\right).$$

This equation is equivalent to

$$\begin{aligned} K_t(x, B) &= P\left(\eta(t) \in B \mid T_1 > t, \eta\left(0^+\right) = x\right) P\left(T_1 > t \mid \eta\left(0^+\right) = x\right) \\ &= \delta_{(x-t)^+}(B)\, e^{-\lambda t}, \end{aligned}$$

where

$$\delta_y(B) = \begin{cases} 1 & if \;\; y \in B, \\ 0 & if \;\; y \notin B. \end{cases}$$

Now lets consider the distribution function $J(x)$ of $\eta(T_n^+)$.
It is easy to derive the appropriate formula for $J(x)$ when we notice that there are only two possible situations that may occur when we consider the event when the total amount of work in the system is smaller than x:

1. The customer arrives and if he decides to enter the system at time u, $(0 \le u \le x)$ his service time is less than $x - u$.
2. The customer arrives at the counter at instant u, $(0 \le u \le x)$ but the actual offered waiting time exceeds his patience and he decides not to join the queue.

The probability of both cases yield

$$J(x) = \int_0^x \left(G(u) H(x-u) + 1 - G(u)\right) dW(u) \tag{2.39}$$

From the Theorem 2.15 we get

$$\begin{aligned}\lim_{t\to\infty} P(\eta(t) \in B) &= \frac{1}{E\{\tau_1\}} \int_0^\infty \int_0^\infty K_t(x, B)\, dt dJ(x) \\ &= \lambda \int_0^\infty \int_0^\infty \delta_{(x-t)^+}(B)\, e^{-\lambda t} dt dJ(x) \\ &= \lambda \int_0^\infty \int_0^\infty (1 - F(t))\, \delta_{(x-t)^+}(B)\, dt dJ(x).\end{aligned}$$

Hence,

$$\begin{aligned} V(y) := \lim_{t\to\infty} P(\eta(t) \le y) &= \lambda \int_0^\infty \int_0^\infty (1 - F(t))\, \delta_{(x-t)^+}([0,y]) dt dJ(x) \\ &= \lambda \int_0^\infty \int_0^{y+t} (1 - F(t))\, dt dJ(x) \\ &= \lambda \int_0^\infty (1 - F(t))\, J(y+t)\, dt \end{aligned} \tag{2.40}$$

Applying now formula (2.39) to (2.40) we get:

$$\begin{aligned} V(x) &= \lambda \int_0^\infty (1 - F(t))\, W(t+x)\, dt \\ &\quad - \lambda \int_0^\infty \int_0^\infty (1 - F(t))\, G(u)\, (1 - H(x+t-u))\, dW(u)\, dt. \end{aligned} \tag{2.41}$$

It was shown that if $1 - \rho G(\infty) > 0$ then $\{w_n\}_{n\in N}$ is an ergodic Markov chain. So there exists a non-defective distribution function $W(x)$ such that

$$W(x) := \lim_{n\to\infty} W_n(x),$$

with

$$\lim_{x\to\infty} W(x) = 1.$$

Hence $V(x)$ is also proper distribution.

2.7 Coincidence of distribution functions $\mathbf{V}(x)$ and $\mathbf{W}(x)$ in case of Poisson arrivals

Now we will prove that the stationary distribution function of $\{w_n\}_{n\in N}$ and $\eta(t)$, (respectively, W and V) coincide for $M \mid GI \mid 1 + GI$ queue.
For

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0, \\ 0 & t < 0, \end{cases}$$

we obtain $V(x)$ of the following form:

$$\begin{aligned} V(x) = {} & \lambda \int_0^\infty e^{-\lambda t} \int_0^{x+t} G(u)\, H(x+t-u)\, dW(u)\, dt \\ & + \lambda \int_0^\infty e^{-\lambda t} \int_0^{x+t} dW(u)\, dt \\ & - \lambda \int_0^\infty e^{-\lambda t} \int_0^{x+t} G(u)\, dW(u)\, dt. \end{aligned}$$

After changing variables we get: $\xi = x + t$

$$\begin{aligned} V(x) = {} & \lambda \int_x^\infty e^{-\lambda(\xi - x)} \int_0^{\xi} G(u)\, H(\xi - u)\, dW(u)\, d\xi \\ & + \lambda \int_x^\infty e^{-\lambda(\xi - x)} \int_0^{\xi} dW(u)\, d\xi \\ & - \lambda \int_x^\infty e^{-\lambda(\xi - x)} \int_0^{\xi} G(u)\, dW(u)\, d\xi \\ & =: (1) + (2) + (3). \end{aligned}$$

Consider also $W(x)$, for the Poisson arrivals

$$\begin{aligned} W(x) = {} & \lambda \int_0^\infty e^{-\lambda t} \int_0^\infty G(u)\, H(x+t-u)\, dW(u)\, dt \\ & + \int_0^\infty e^{-\lambda (u-x)^+} dW(u) \\ & - \int_0^\infty G(u)\, e^{-\lambda (u-x)^+} dW(u) \\ & =: (A) + (B) + (C). \end{aligned}$$

Now we will show that $(1), (2), (3)$ are equal to $(A), (B), (C)$ respectively.

- $\underline{(1) = (A)}$

  To prove that relation, it is enough to notice that

$$\int_0^\infty G(u)\, H(x+t-u)\, dW(u) = \int_0^{x+t} G(u)\, H(x+t-u)\, dW(u).$$

- $\underline{(2) = (B)}$

  Lets consider (B) :

$$\int_0^\infty e^{-\lambda(u-x)^+} dW(u) = \int_0^x dW(u) + \int_x^\infty e^{-\lambda(u-x)} dW(u).$$

By the Theorem$A.3$ (AppendixA) we get

$$\int_x^\infty e^{-\lambda u} dW(u) = -e^{-\lambda x} W(x) + \lambda \int_x^\infty e^{-\lambda u} W(u)\, du,$$

hence

$$\int_x^\infty e^{-\lambda(u-x)} dW(u) = -\int_0^x dW(u) + \lambda \int_x^\infty e^{-\lambda(u-x)} W(u)\, du,$$

and thus

$$\int_0^\infty e^{-\lambda(u-x)_+} dW(u) = \lambda \int_x^\infty e^{-\lambda(u-x)} W(u)\, du.$$

- $\underline{(3) = (C)}$

  Lets focus on (3).
  Suppose w is a density function of W, i.e.,

$$W(x) = \int_0^x w(u)\, du,$$

then:

$$\lambda \int_x^\infty e^{-\lambda(\xi-x)} \int_0^\xi G(u)\, dW(u)\, d\xi = e^{\lambda x} \lambda \int_x^\infty e^{-\lambda \xi} \int_0^\xi G(u)\, w(u)\, du d\xi.$$

Denote $h(u) := G(u)\, w(u)$ and integrate $\int_x^\infty e^{-\lambda u} h(u)\, du$ by parts, the result is

$$\int_x^\infty e^{-\lambda u} h(u)\, du = -e^{-\lambda x} \int_0^x h(u)\, du + \lambda \int_x^\infty e^{-\lambda u} h_1(u)\, du$$

where $h_1(u) = \int_0^u G(s)\, w(s)\, ds$.
Hence,

$$\begin{aligned} e^{\lambda x} \lambda \int_x^\infty e^{-\lambda \xi} \int_0^\xi G(u)\, w(u)\, du d\xi &= e^{\lambda x} \lambda \int_x^\infty e^{-\lambda \xi} h_1(\xi)\, d\xi \\ &= \int_0^x G(u)\, w(u)\, du + \int_x^\infty e^{-\lambda(u-x)} G(u)\, w(u)\, du \\ &= \int_0^x G(u)\, dW(u) + \int_x^\infty e^{-\lambda(u-x)} G(u)\, dW(u). \end{aligned} \quad (2.42)$$

It ends the proof, since the right hand side of equation (2.42) is precisely (C).

2.8 Density of the stationary v.o.w.t. distribution function

Lemma 2.16 *For ψ to be of the form*

$$\psi(s) = \int_0^\infty e^{-sx} f(x)\, dx, \qquad \textit{where } \ 0 \le f \le A,$$

it is sufficient and necessary that

$$0 \le \frac{(-s)^n}{n!} \psi^{(n)}(s) \le \frac{A}{s}, \qquad \forall s \ge 0,$$

where $\psi^{(n)}(s)$ denotes the nth derivative of $\psi(s)$.

Proof. $[2, Feller]$ ■

We will apply this criterion to $\Omega(s) - V(0)$.

Let $\phi(s)$ be LS transform of $V(x) - V(0)$. By the definition of the Stieltjes integral

$$\phi(s) = \Omega(s) - V(0),$$

and so from (2.31)

$$\phi(s) = a(s)\, \Omega_G(s).$$

One can check that $a(s)$ is the Laplace-Stieltjes transform of the "unfinished work" in the system.
Recall that $a(s)$ was defined as

$$a(s) = \lambda \frac{1 - \int_0^\infty e^{-sx} dH(x)}{s}.$$

Observe that following property of moment generating function $(B.2)$

$$Ee^{-sX} = 1 - s \int_0^\infty e^{-sx}(1 - F(x))dx,$$

where F is the d.f. of random variable X,
yields:

$$a(s) = \lambda \int_0^\infty e^{-sx}(1 - H(x))dx = \int_0^\infty e^{-sx} \alpha(x)\, dx,$$

where

$$\alpha(x) = \lambda(1 - H(x)).$$

It is clear that $a(s)$ satisfies the conditions of the above lemma.

Let

$$D := \max_x \alpha(x).$$

For $n = 0$, we shall use: $\Omega_G(s) \le \Omega(s) \le 1, \qquad \forall s > 0.$
Hence,

$$0 \le \phi(s) = a(s)\,\Omega_G(s) \le a(s)\,\Omega(s) \le a(s).$$

Observe now, that

$$a(s) = \int_0^\infty e^{-sx}\alpha(x)\,dx \le D\int_0^\infty e^{-sx}dx = \frac{D}{s}, \tag{2.43}$$

so the conditions of lemma are fulfilled when $n = 0$.

For $n \ge 1$ the nth derivative of $\phi(s)$ is of the form

$$\phi^{(n)}(s) = \sum_{j=0}^{n}\tbinom{n}{j}\Omega_G^{(j)}(s)\,a^{(n-j)}(s).$$

From (2.43) we derive

$$0 \le (-1)^k a^{(k)}(s) \le \frac{Dk!}{s^{k+1}},$$

Therefore

$$(-1)^n \phi^{(n)}(s) \le \sum_{j=0}^{n}\frac{n!\,(-1)^j}{j!\,(n-j)!}\Omega_G^{(j)}(s)\,\frac{D\,(n-j)!}{s^{n-j+1}} = \frac{n!D}{s^{n+1}}\sum_{j=0}^{n}\frac{(-s)^j}{j!}\Omega_G^{(j)}(s).$$

Define now

$$\delta := \sum_{j=0}^{n}\frac{(-s)^j}{j!}\Omega_G^{(j)}(s) = \sum_{j=0}^{n}\int_0^\infty \frac{(sx)^j}{j!}e^{-sx}G(x)\,dV(x).$$

By the definition of the exponential function we have

$$\delta \le \int_0^\infty \sum_{j=0}^{\infty}\frac{(sx)^j}{j!}e^{-sx}G(x)\,dV(x) = \int_0^\infty G(x)\,dV(x) \le 1,$$

hence

$$0 \le (-1)^n \phi^{(n)}(s) \le \frac{Dn!}{s^{n+1}}.$$

So $\forall n \ge 0$, $\phi(s)$ satisfies the assumption of our lemma.
Hence, there exist a function $v(x)$ such that

$$\begin{cases} \phi(s) = \int_0^\infty e^{-sx}v(x)\,dx, \\ 0 \le v(x) \le D, \end{cases}$$

and by definition of $\phi(s)$ we finally get

$$\begin{cases} \Omega(s) = V(0) + \int_0^\infty e^{-sx}v(x)\,dx \\ 0 \le v(x) \le D. \end{cases} \tag{2.44}$$

The latter relation shows that $V(x)$ is composed of an absolutely continuous part and a mass at the origin.
We can derive the explicit formula for $v(x)$ by considering the equation

$$\begin{aligned} P(\eta(t+\Delta) > x) = P(\eta(t) > x+\Delta) + \lambda\Delta \int_0^{x+\Delta} G(u)\left(1 - H(x-u)\right) dV(u) \\ + \lambda\Delta P(\eta(t) > 0)\left(1 - H(x+\Delta)\right) \end{aligned}$$

and now, some inversions in above formula yield

$$\begin{aligned} \frac{V(t+\Delta, x) - V(t, x+\Delta)}{\Delta} &= -\lambda \int_0^{x+\Delta} G(u)\left(1 - H(x-u)\right) dV(u) \\ &\quad - \lambda V(t,0)\left(1 - H(x+\Delta)\right) \end{aligned} \tag{2.45}$$

Since we can evaluate $V(t, x+\Delta)$,

$$V(t, x+\Delta) = V(t,x) + \frac{\partial V(t,x)}{\partial x}\Delta + o(\Delta)$$

so, after taking the limit for $\Delta \to 0$, the expression (2.45) can be rewritten as follows

$$\frac{\partial V(t,x)}{\partial t} = \frac{\partial V(t,x)}{\partial x} - \lambda \int_0^x G(u)\left(1 - H(x-u)\right) dV(u) - \lambda V(t,0)\left(1 - H(x)\right)$$

Suppose the limit $\lim_{t\to\infty} V(t,x) = V(x)$ exists. Then, since $v(x) = \frac{\partial V(t,x)}{\partial x}$ we can finally write

$$v(x) = \lambda V(0)\left(1 - H(x)\right) + \lambda \int_0^x v(u)\, G(u)\left(1 - H(x-u)\right) du. \tag{2.46}$$

2.9 Quantities of practical interest

In this section we shall study quantities such as: the probability of rejection, mean waiting time and mean queue length. In further part we also derive a Pollaczek-Khinchin mean value formula for queues with impatient customers.

Probability of rejection

Let Π be the probability of rejection, i.e., the probability that an arriving customer decides not to enter the system. Recall that such situations may happen when the actual waiting time offered to a customer at the moment when he arrives exceeds his patience.
Recall also that we defined $G(x)$ as

$$G(x) = 1 - C(x),$$

where $C(x)$ is a distribution function of the patience.
Hence, for $GI \mid GI \mid 1 + GI$ queues we have

$$\Pi = \int_{0-}^{\infty} (1 - G(u))\, dW(u), \tag{2.47}$$

where $W(x)$ is the limiting distribution of $W_n(x)$, which exists when $0 < 1 - \rho G(\infty)$.

Since we have proven that for $M \mid GI \mid 1 + GI$ queues the distribution of the actual offered waiting time-W and the virtual offered waiting time-V coincide, we obtain in this case:

$$\Pi = \int_{0}^{\infty} v(x)\,(1 - G(x))\, dx. \tag{2.48}$$

Now consider equation (2.31), i.e.,

$$\Omega(s) = V(0) + a(s)\,\Omega_G(s).$$

In case of Poisson arrivals the above formula is equivalent to

$$\int_{0}^{\infty} e^{-sx} v(x)\, dx = V(0) + a(s) \int_{0}^{\infty} e^{-sx} v(x)\, G(x)\, dx.$$

Note that for $s = 0$ we obtain

$$1 = V(0) + \rho \int_{0}^{\infty} v(x)\, G(x)\, dx,$$

and now, since

$$\int_{0}^{\infty} v(x)\, dx = 1,$$

some inversions in (2.48) give

$$(1 - \Pi)\,\rho = \rho \int_{0}^{\infty} v(x)\, G(x)\, dx = 1 - V(0). \tag{2.49}$$

Mean waiting time and mean queue length

Let EW_1 be the mean waiting time spent in the queue by patient customers, and EL_1 be the mean number of patient customers in the queue. Similarly, let EW_2 be the mean waiting time spent in the queue by all customers (the impatient ones, i.e., those who rejected after their time-out and the patient ones), and EL_1 be the mean number of such customers in the queue.

For the $GI \mid GI \mid 1 + GI$ queue we have:

$$EW_1 = \int_{0}^{\infty} uG(u)\, dW(u), \tag{2.50}$$

$$EW_2 = \int_0^\infty \int_0^u G(t)\, dt dW(u). \tag{2.51}$$

The expression for EW_2 can be derived from the following relation

$$\begin{aligned} EW_2 &= \int_0^\infty u dW(u) - \int_0^\infty \int_0^u C(t)\, dt dW(u) \\ &= \int_0^\infty \int_0^u dt dW(u) - \int_0^\infty \int_0^u C(t)\, dt dW(u). \end{aligned}$$

Applying Little's formula we derive

$$EL_1 = \lambda(1-\Pi)\, EW_1, \tag{2.52}$$

$$EL_2 = \lambda EW_2. \tag{2.53}$$

Examples

Let γ be the first moment of $C(x)$, then

- For the $GI \mid GI \mid 1 + D$ queue we have

$$\begin{aligned} EW_1 &= \int_0^\gamma u dW(u), \\ EW_2 &= \int_0^\gamma u dW(u) + \gamma \int_\gamma^\infty dW(u), \\ \Pi &= \int_\gamma^\infty dW(u). \end{aligned}$$

  Hence,

$$EW_2 = EW_1 + \gamma \Pi.$$

- For the $GI \mid GI \mid 1 + M$ queue we have

$$\begin{aligned} EW_1 &= \int_0^\infty u e^{-\gamma u} dW(u), \\ EW_2 &= \frac{1}{\gamma} \int_0^\infty \left(1 - e^{-\gamma u}\right) dW(u), \\ \Pi &= \int_0^\infty \left(1 - e^{-\gamma u}\right) dW(u). \end{aligned}$$

  Hence,

$$EW_2 = \frac{1}{\gamma}\Pi.$$

The average virtual offered waiting time in M|GI|1+GI case

Denote by $\bar{v}$ the mean v.o.w.t. By the result of section (2.7) we get

$$\bar{v} = \int_0^\infty xv(x)\,dx.$$

Consider again equation (2.31) and differentiate it with respect to s. This procedure yields

$$\int_0^\infty xe^{-sx}dW(x) = \int_0^\infty xe^{-sx}\lambda(1-H(x))\,dx \int_0^\infty e^{-sx}G(x)\,dW(x) + \int_0^\infty e^{-sx}\lambda(1-H(x))\,dx \int_0^\infty xe^{-sx}G(x)\,dW(x).$$

Since $V(x) = W(x)$ in our case, for $s = 0$ we obtain

$$\int_0^\infty xdV(x) = \int_0^\infty x\lambda(1-H(x))\,dx \int_0^\infty G(x)\,dV(x) + \lambda \int_0^\infty (1-H(x))\,dx \int_0^\infty xG(x)\,dV \tag{2.54}$$

Notice now, that

$$\int_0^\infty (1-H(x))\,dx = \frac{1}{\mu},$$

and also by another property of moment generating function (Appendix B), we know that

$$Es_1^2 = 2\int_0^\infty x(1-H(x))\,dx.$$

So

$$\int_0^\infty xdV(x) = \rho EW_1 + \frac{\lambda}{2}\int_0^\infty G(x)\,dV(x)\,Es_1^2.$$

Eventually, we obtain

$$\bar{v} = \rho EW_1 + \frac{\lambda}{2}(1-\Pi)Es_1^2,$$

which is Pollaczek-Khinchin formula for queues with impatient customers.

Chapter 3

Multi Server Queues

3.1 Fundamental relations for $M \mid M \mid m + GI$ queueing systems

Consider the $m-$server queueing system. In this section we assume Poisson arrivals and exponential service times for each server, i.e.,

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

$$H_i(t) = \begin{cases} 1 - e^{-\mu t} & t \geq 0, \\ 0 & t < 0, \end{cases} \qquad for \ \ i = 1, ..., m$$

where $H_i(t)$ denotes the service time distribution for the ith server.

We shall introduce the following process $\{(N(t), \eta(t)), t \geq 0\}$; where $\{\eta(t), t \geq 0\}$ is the v.o.w.t. and $\{N(t), t \geq 0\}$ equals to n when the number of customers in the system at time t is n $(0 \leq n \leq m-1)$, $N(t)$ is equal to m when the number of customers at time t is greater than $m-1$.

Observe that $\{(N(t), \eta(t)) : t \geq 0\}$ is then a Markov process with state space $[\{0\}, \{1\}, ... \{m-1\}, \{m\}] \times R^{+}$.

It is easy to see that the v.o.w.t. is equal to zero when $N(t) \neq m$ and is strictly positive otherwise. Hence only the following states can be obtained with positive probability:

$$\boxed{\{(0,0),(1,0),...,(m-1,0),(m,x), \qquad if \ \ x \geq 0\}}$$

Lets introduce the functions:

$$V(t,x) := P(N(t) = L,\ \eta(t) \le x),$$
$$P_j(t) := P(N(t) = j,\ \eta(t) = 0), \qquad 0 \le j \le m-1,$$

and

$$v(t,x) := \frac{\partial V(t,x)}{\partial x},$$
$$v(t,x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} P(N(t) = L,\ x < \eta(t) \le x + \Delta x).$$

Then we can write the normalizing condition for the process $\{(N(t), \eta(t)) : t \ge 0\}$ as:

$$\sum_{j=0}^{m-1} P_j + \int_0^\infty d_x V(t,x) = 1.$$

Observe now that the following relations hold:

<u>For state $(0,0)$</u>:

$$P_0(t) = P_0(t)(1 - \lambda\Delta) + \Delta\mu P_1(t) + o(\Delta). \tag{3.1}$$

<u>For state $(j,0), \qquad 0 < j < m-1$</u> :

$$P_j(t+\Delta) = P_j(t)(1 - \lambda\Delta - j\mu\Delta) + \lambda\Delta P_{j-1}(t) + \Delta\mu(j+1)P_{j+1}(t) + o(\Delta). \tag{3.2}$$

<u>For state $(m-1,0)$</u>:

$$P_{m-1}(t+\Delta) = (1 - \lambda\Delta - (m-1)\Delta\mu)P_{m-1}(t) + \lambda\Delta P_{m-2}(t)$$
$$+ (1-\lambda\Delta)\int_0^\Delta d_u V(t,u) + o(\Delta).$$

By the Law of the Mean for the Stieltjes integral this equation is equivalent to

$$P_{m-1}(t+\Delta) = (1 - \lambda\Delta - (m-1)\Delta\mu)P_{m-1}(t) + \lambda\Delta P_{m-2}(t)$$
$$+ V(t,\Delta) - V(t,0) + o(\Delta). \tag{3.3}$$

<u>For state $(L,x), \qquad x > 0$</u>:

$$P(N(t) = m, \eta(t+\Delta) > x) = P(N(t) = m, \eta(t) > x + \Delta)$$
$$+ \lambda\Delta \int_0^{x+\Delta} G(u)e^{-m\mu(x-u)} d_u V(t,u) + \lambda\Delta P_{m-1}(t)e^{-m\mu(x+\Delta)} + o(\Delta)$$

and so,

$$V(t+\Delta, x) = V(t, x+\Delta) - \lambda\Delta \int_0^{x+\Delta} G(u)e^{-m\mu(x-u)} d_u V(t,u)$$
$$- \lambda\Delta P_{m-1}(t)e^{-m\mu x+\Delta} + o(\Delta) \tag{3.4}$$

where $V(t, x+\Delta)$ can be evaluated as

$$V(t, x+\Delta) = V(t,x) + \frac{\partial V(t,x)}{\partial x}\Delta + o(\Delta).$$

After some calculus and taking the limit for $\Delta \to 0$, we can rewrite equations $(3.1), (3.2), (3.3), (3.4)$ as follows

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= -\lambda P_0(t) + \mu P_1(t), \\
\frac{dP_j(t)}{dt} &= -(\lambda + j\mu) P_j(t) + \lambda P_{j-1} + (j+1)\mu P_{j+1} \qquad 0 < j < m-1, \\
\frac{dP_{m-1}(t)}{dt} &= -(\lambda + (m-1)\mu) P_{m-1}(t) + \lambda P_{m-2} + v(t,0), \\
\frac{\partial V(t,x)}{\partial t} - \frac{\partial V(t,x)}{\partial x} &= -\lambda \int_0^x G(u) e^{-m\mu(x-u)} d_u V(t,u) - \lambda P_{m-1}(t) e^{-m\mu x}, \ x > 0.
\end{aligned}$$

Suppose now that the stationary solution exists and denote:

$$\begin{aligned}
V(x) &:= \lim_{t\to\infty} V(t,x), \\
v(x) &:= \lim_{t\to\infty} v(t,x).
\end{aligned}$$

Then we obtain Chapman-Kolomogorov equations for the process $\{(N(t), \eta(t)) : t \geq 0\}$:

$$\begin{aligned}
\lambda P_0 &= \mu P_1, \\
(\lambda + \mu j) P_j &= \lambda P_{j-1} + (j+1)\mu P_{j+1}, \qquad 0 < j < m-1, \\
(\lambda + \mu(m-1)) P_j &= \lambda P_{m-1} + v(0), \\
v(x) &= \lambda P_{m-1}(t) e^{-m\mu x} + \lambda \int_0^x G(u) e^{-m\mu(x-u)} d_u V(t,u), \qquad x > 0.
\end{aligned} \tag{3.5}$$

Solving the first three equations we get

$$\begin{cases} P_j = \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} P_0, & j = 0 \ldots m-1, \\ v(0) = \lambda P_{m-1}. \end{cases} \tag{3.6}$$

Furthermore it is easy to see that $H(x) = e^{m\mu x} v(x)$ is the solution of the following equation:

$$H(x) = \lambda P_{m-1} + \lambda \int_0^x G(u) H(u) du, \qquad x > 0.$$

Basic theory of differential equations yields

$$H(x) = \lambda P_{m-1} e^{\lambda \int_0^x G(u) du},$$

hence, $v(x)$ is of the form

$$v(x) = \lambda P_{m-1} e^{\lambda \int_0^x G(u) du - m\mu x}. \tag{3.7}$$

Remark 3.1 *One can note the resemblance between* (3.5) *and the formula derived in section* (2.8) *for the density function of the v.o.w.t. in* $M \mid GI \mid 1+GI$ *case. Indeed, the situation in the* $m-$*server queue, where there are exactly* $m-1$ *servers occupied is equivalent to a single server queue. Hence,* v *given by* (3.7) *can be interpreted as a density function of the v.o.w.t. in the* $M \mid GI \mid m+GI$ *queue. However, it is important to note that the value of* $V(0)$ *in this case is not equal to the probability of having servers idle. That is why in the sequel part we will redefine this quantity.*

By the above Remark we can rewrite the normalizing condition as

$$\sum_{j=0}^{m-1} P_j + \int_0^\infty v(x)\,dx = 1. \tag{3.8}$$

Relation (3.6) and the above equation allow us to compute the exact formula for P_0 :

$$P_0 \left(\sum_{j=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} + \lambda \left(\frac{\lambda}{\mu}\right)^{m-1} \frac{1}{(m-1)!} \int_0^\infty e^{\lambda \int_0^x G(u)du - m\mu x} dx \right) = 1,$$

hence

$$P_0 = \left[\sum_{j=0}^{m-2} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} + \left(\frac{\lambda}{\mu}\right)^{m-1} \frac{1}{(m-1)!} (1 + \lambda J) \right]^{-1}, \tag{3.9}$$

where

$$J = \int_0^\infty e^{\lambda \int_0^x G(u)du - m\mu x} dx. \tag{3.10}$$

For stability one can check that the normalizing condition is feasible if and only if the integral in (3.10) converges,which is equivalent to the condition

$$\lambda G(\infty) < m\mu.$$

Remark 3.2 *The distribution of the v.o.w.t. is consistent with the distribution of the actual waiting time also in the multi server queue with Poisson input.*

3.2 Quantities of practical interest

Probability of rejection

Analogously to a single server case we will derive a formula for the probability that an arriving customer decides to reject the system due to his impatience. In a multi server case, such situation is possible when all servers are busy. Hence, throughout this section we shall assume that the service time distribution function for whole system is given by

$$H(t) = 1 - e^{-m\mu t}. \tag{3.11}$$

Furthermore observe that in case when there is a queue the formulas from section (2.9) are also valid but with $\rho^{'} = \frac{\rho}{m}$. It can be easily proved, since for $H(t)$ defined by (3.11) we have

$$a(s) = \lambda \int_0^{\infty} e^{-sx} (1 - H(x))\, dx,$$

hence,

$$a^{'}(0) = \frac{1}{m\mu},$$

and so, differentiating (2.31) in the point $s = 0$ yields

$$1 = V(0) + \frac{\rho}{m} \int_0^{\infty} v(x)\, G(x)\, dx.$$

Proceeding now as in a single server queue, we get

$$(1 - \Pi) \frac{\rho}{m} = 1 - V(0), \qquad (3.12)$$

where

$$V(0) := \sum_{j=0}^{m-1} P_j.$$

We will derive now an exact formula for Π.
Observe at first that the following equality holds:

$$P_{m-1} \lambda \int_0^{\infty} \left(G(x) - \frac{m\mu}{\lambda} \right) e^{\lambda \int_0^x (G(u) - \frac{m\mu}{\lambda}) du} dx = P_{m-1} \int_{-\infty}^{0} e^{\xi} d\xi = P_{m-1},$$

hence,

$$\begin{aligned} \int_0^{\infty} G(x)\, v(x)\, dx &= \lambda P_{m-1} \int_0^{\infty} G(x)\, e^{\lambda \int_0^x G(u) du - m\mu x} dx \\ &= \lambda P_{m-1} \int_0^{\infty} \frac{m\mu}{\lambda} e^{\lambda \int_0^x G(u) du - m\mu x} dx + P_{m-1} \\ &= \frac{m\mu}{\lambda} \int_0^{\infty} v(x)\, dx + P_{m-1}. \end{aligned} \qquad (3.13)$$

Since the probability of rejection was given by

$$\Pi = \int_0^{\infty} (1 - G(x))\, v(x)\, dx,$$

using formula (3.13) we can write

$$\Pi = \left(1 - \frac{m\mu}{\lambda}\right) \int_0^{\infty} v(x)\, dx.$$

Now, applying the normalizing condition (3.8) we finally obtain

$$\Pi = (1 - \frac{m\mu}{\lambda}) \left[1 - \sum_{j=0}^{m-1} P_j \right] + P_{m-1}.$$

Mean virtual offered waiting time

Proceeding analogously as in a single server case we will derive the Pollaczek-Khinchin formula for the mean waiting time in the $M \mid M \mid m + GI$ queue.
As we have previously shown

$$(1-\Pi)\frac{\rho}{m} = 1 - V(0).$$

Similar computation as in section (2.9) yields

$$\bar{v} = \frac{\rho}{m} EW_1 + \frac{\lambda}{2}(1-\Pi)Es_1^2,$$

where

$$Es_1^2 = \frac{2}{m^2\mu^2}.$$

Hence, we get:

$$\bar{v} = \frac{\rho}{m} EW_1 + \frac{1}{m\mu}\left[1 - V(0)\right].$$

Remark 3.3 *The formulas for the mean waiting time and the mean queue length remain the same as for one server queue.*

"Service level" measure

It may happen that one can be interested in defining a percentage α and a number a for which must hold: $\alpha\%$ of the customers must have a waiting time shorter than a seconds. We call this quantity the service level. In terms of the waiting time distribution it can be derived from

$$P(W\_time > a) = 1 - \alpha, \tag{3.14}$$

where for convenience we denoted by W_time the stationary distribution of waiting time given by (2.22), i.e. $P(W\_time \leq a) = W(a)$.
We shall call the left hand side of expression (3.14) the "service level" measure, and we will be mainly interested in this quantity.
One can note that

$$\begin{aligned} P(W\_time > a) &= \int_a^\infty dW(u) \\ &= \int_a^\infty dV(u) \qquad \text{in case of Poisson arrivals} \\ &= \int_a^\infty v(u)\,du \qquad \text{where } v(u) \text{ is a density of the v.o.w.t.} \end{aligned}$$

3.3 On the $M \mid M \mid m+D$ queue

Denote by γ the first moment of $C(x)$, then

$$\begin{aligned} G(x) &= 1, \qquad if \;\; x \leq \gamma, \\ G(x) &= 0, \qquad if \;\; x < \gamma. \end{aligned}$$

Using the results of preceding section we derive a formula for the density of the v.o.w.t.
In our case expression (3.10) takes form

$$J = \int_0^{\gamma} e^{\lambda x - m\mu x} dx + \int_0^{\infty} e^{\lambda\gamma - m\mu x} dx.$$

Hence, after some computation we get

$$J = \frac{1}{\mu(m-\rho)} \left(1 - \frac{\rho}{m} e^{\lambda x - m\mu x}\right). \tag{3.15}$$

Applying now J to (3.9) we obtain the expression for P_0

$$P_0 = \left(1 + \sum_{j=0}^{m-1} \frac{\rho^j}{j!} + \frac{\rho^m}{(\rho - m)\, m!} \left(\rho e^{\lambda\gamma - m\mu\gamma} - m\right)\right)^{-1},$$

and now we can compute the required density function $v(x) = \lambda P_{m-1} e^{\lambda \int_0^x G(u)du - m\mu x}$ which in our case takes form

$$v(x) = \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \exp(\lambda - m\mu)\gamma \qquad if \;\; x \leq \gamma, \tag{3.16}$$

$$v(x) = \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \exp(\lambda\gamma - m\mu x) \qquad if \;\; x < \gamma. \tag{3.17}$$

Knowing $v(x)$ we can also compute the following quantities

$$\begin{aligned} EW_1 &= \int_0^{\gamma} x v(x)\, dx, \\ \Pi &= \int_{\gamma}^{\infty} v(x)\, dx. \end{aligned}$$

and

$$EW_2 = EW_1 + \gamma\Pi.$$

EW_1 is then given by

$$\lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \int_0^{\gamma} x e^{(\lambda - m\mu)x} dx,$$

so

$$EW_1 = \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \frac{1}{(\lambda - m\mu)^2} \left[e^{(\lambda - m\mu)\gamma} (\gamma - 1) + 1 \right].$$

The probability of rejection is given by

$$\lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \int_\gamma^\infty e^{(\lambda\gamma - m\mu x)} dx,$$

so

$$\Pi = \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 e^{(\lambda - m\mu)\gamma}.$$

The "service level" measure is equal to

$$P(W\_time > t) = \begin{cases} \int_t^\gamma v_1(x)\,dx + \int_\gamma^\infty v_2(x)\,dx & if \;\; t < \gamma, \\ \int_t^\infty v_2(x)\,dx & if \;\; t \geq \gamma, \end{cases}$$

where $v_1(x)$, $v_2(x)$ are respectively given by (3.16) and (3.17).
Hence,

$$P(W\_time > t) = \begin{cases} \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \frac{1}{(\lambda - m\mu)} \left(e^{(\lambda - m\mu)\gamma} - e^{(\lambda - m\mu)t} \right) + \Pi & if \;\; t < \gamma, \\ \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 e^{(\lambda - m\mu)t} & if \;\; t \geq \gamma. \end{cases}$$

Remark 3.4 *It is easy to see that for a deterministic patience*

$$P(W\_time > t) < \Pi \qquad if t > \gamma$$

3.4 On the $M \mid M \mid m + M$ queue

Lets assume γ is the first moment of $C(x)$, then

$$G(x) = \begin{cases} e^{-\frac{x}{\gamma}} & if \;\; x \geq 0, \\ 1 & if \;\; x < 0. \end{cases}$$

Applying this function to the formula for the density of the v.o.w.t. we get

$$v(x) = \lambda P_{m-1} \exp\left(\lambda \int_0^x e^{-\frac{u}{\gamma}} du - m\mu x \right).$$

Denote by $J(z)$ the following integral

$$J(z) = \int_z^\infty \exp\left(\lambda \int_0^x e^{-\frac{u}{\gamma}} du - m\mu x \right) dx.$$

Integration by parts yields

$$J(z) = \gamma e^{\lambda\gamma\left(1 - e^{-\frac{z}{\gamma}}\right) - zm\mu} \sum_{j=1}^{\infty} \frac{e^{-\frac{z}{k} j} (\lambda\gamma)^{(j-1)}}{m\mu\gamma (m\mu\gamma + 1) \cdot \ldots \cdot (m\mu\kappa + j)}. \tag{3.18}$$

Observe now, that $J(0)$ is precisely formula (3.10) in case of exponential impatience.
Hence,

$$J = J(0) = \gamma\left(\frac{1}{m\mu\gamma} + \frac{\lambda\gamma}{m\mu\gamma(m\mu\gamma+1)} + \frac{(\lambda\gamma)^2}{m\mu\gamma(m\mu\gamma+1)(m\mu\gamma+2)} + ...\right)$$

Lets denote $\alpha = \frac{1}{m\mu\gamma}$, then

$$J = \frac{1}{m\mu}\left(1 + \frac{\rho/m}{1+\alpha} + \frac{(\rho/m)^2}{(1+\alpha)(1+2\alpha)} + ...\right).$$

Applying J to (3.9) we get the expression for P_0

$$P_0 = \left[\sum_{j=0}^{m-1}\frac{\rho^j}{j!} + \frac{\rho^m}{m!}\left(1 + \frac{\rho/m}{1+\alpha} + \frac{(\rho/m)^2}{(1+\alpha)(1+2\alpha)} + ...\right)\right]^{-1}.$$

Knowing P_0 we can derive formula for probability of rejection.
It was shown in section (3.2) that

$$\Pi = (1 - \frac{m\mu}{\lambda})\left[1 - \sum_{j=0}^{m-1} P_j\right] + P_{m-1},$$

where

$$P_j = \frac{\rho^j}{j!}P_0,$$

hence, after some calculus we get

$$\Pi = P_0\frac{\rho^{m-1}}{(m-1)!}\left[1 + \left(\frac{\rho}{m} - 1\right)\left(1 + \frac{\rho/m}{1+\alpha} + \frac{(\rho/m)^2}{(1+\alpha)(1+2\alpha)} + ...\right)\right].$$

We obtain also formula for EW_2

$$EW_2 = \gamma\Pi.$$

Finally we derive the "service level" measure.
Using the expression for $J(z)$, we can write

$$\begin{aligned} P(W\_time > z) &= \lambda P_{m-1}\int_z^\infty \exp\left(\lambda\int_0^x e^{-\frac{u}{\gamma}}du - m\mu x\right)dx \\ &= \lambda P_{m-1}J(z), \end{aligned}$$

where $J(z)$ is given by (3.18).

Appendix A

The Laplace-Stieltjes Transform

A.1 The Stieltjes integral

Definition and properties

Let $\alpha(x)$ and $f(x)$ be real valued functions of the real variable x defined for $a \le x \le b$, $(a, b \in R)$. Denote by Δ a subdivision of the interval (a, b) by the points $x_0, x_1, \ldots, x_n$, where

$$a = x_0 < x_1 < \ldots < x_n = b.$$

By the norm δ of Δ we mean the largest of numbers

$$x_{i+1} - x_i, \qquad (i = 0, 1, \ldots, n-1).$$

Definition A.1 *If the limit*

$$\lim_{\delta \to 0} \sum_{i=0}^{n-1} f(\xi_i) [\alpha(x_{i+1}) - \alpha(x_i)]$$

where

$$x_i \le \xi_i \le x_{i+1}, \qquad (i = 0, 1, \ldots, n-1),$$

exists independently of the manner of subdivision and of the choice of the numbers ξ_i, then the limit is called the Stieltjes integral of $f(x)$ with respect to $\alpha(x)$ from a to b and is denoted by

$$\int_a^b f(x)\, d\alpha(x). \tag{A.1}$$

The definition is easily extended to include complex functions. Thus, if

$$f(x) = f_1(x) + i f_2(x),$$
$$\alpha(x) = \alpha_1(x) + i\alpha_2(x),$$

where $f_1(x), f_2(x), \alpha_1(x), \alpha_2(x)$ are real, we define the integral $(A.1)$ by the equation

$$\int_a^b f(x)\,d\alpha(x) = \int_a^b f_1(x)\,d\alpha_1(x) - \int_a^b f_2(x)\,d\alpha_2(x) + i\int_a^b f_2(x)\,d\alpha_1(x) + i\int_a^b f_1(x)\,d\alpha_2(x)$$

provided all of the integrals on the right exist.

Theorem A.2 *Let the sequence of functions* $\{\alpha_n(x)\}_{n=0}^{\infty}$ *be of uniformly bounded variation on* $[a,b]$. *Let*

$$\lim_{n\to\infty} \alpha_n(x) = \alpha(x), \qquad (a \le x \le b),$$

and let $f(x)$ *be continuous in* $a \le x \le b$. *Then*

$$\lim_{n\to\infty} \int_a^b f(x)\,d\alpha_n(x) = \int_a^b f(x)\,d\alpha(x).$$

Proof. [7, Widder] ■

Improper Stieltjes integrals

Let $f(x)$ be continuous on $[a,\infty)$ and let $\alpha(x)$ be of bounded variation and normalized on $[a,R]$ for every $R > 0$. Then we define the improper integral of $f(x)$ with respect to $\alpha(x)$ on the infinite interval (a,∞) by the equation

$$\int_a^\infty f(x)\,d\alpha(x) = \lim_{R\to\infty} \int_a^R f(x)\,d\alpha(x). \tag{A.2}$$

When the limit $(A.2)$ exists, the integral $(A.2)$ converges; otherwise it diverges. In a similar way, we define improper integrals over $(-\infty,a)$ and $(-\infty,\infty)$ by the equations

$$\int_{-\infty}^a f(x)\,d\alpha(x) = \lim_{R\to\infty} \int_{-R}^a f(x)\,d\alpha(x),$$
$$\int_{-\infty}^\infty f(x)\,d\alpha(x) = \int_0^\infty f(x)\,d\alpha(x) + \int_{-\infty}^0 f(x)\,d\alpha(x), \qquad b \in R.$$

Existence of Stieltjes integral

Theorem A.3 *If $f(x)$ is of bounded variation and $\alpha(x)$ is continuous on (a,b), then the Stieltjes integral of $f(x)$ with respect to $\alpha(x)$ from a to b exists and*

$$\int_a^b f(x)\,d\alpha(x) = f(b)\,\alpha(b) - f(a)\,\alpha(a) - \int_a^b \alpha(x)\,df(x). \tag{A.3}$$

Proof. $[7, Widder]$ ■

Laws of the Mean

Theorem A.4 *If $\alpha(x)$ is non-decreasing (or non-increasing) and $f(x)$ is a real valued continuous function on (a,b), then*

$$\textstyle\int_a^b f(x)\,d\alpha(x) = f(\xi)\,[\alpha(b) - \alpha(a)], \qquad (a \le \xi \le b). \tag{A.4}$$

Proof. The proof is similar to the classic proof of the first law of the mean for Riemann integrals. If

$$\begin{array}{ll} \alpha(x) = \int_a^x \phi(t)\,dt, & (a \le x \le b), \\ \phi(x) \in L^1, & (a \le x \le b), \end{array}$$

where $\phi(x)$ is non-negative, equation $(A.4)$ becomes

$$\int_a^b f(x)\,d\alpha(x) = \int_a^b f(x)\,\phi(x)\,dx = f(\xi)\int_a^b \phi(x)\,dx.$$

■

Theorem A.5 *If $\alpha(x)$ is real and continuous, and $f(x)$ is non-decreasing (or non-increasing) in $[a,b]$, then*

$$\textstyle\int_a^b f(x)\,d\alpha(x) = f(a)\int_a^\xi d\alpha(x) + f(b)\int_\xi^b d\alpha(x), \qquad (a \le \xi \le b). \tag{A.5}$$

Proof. By Theorem $(A.4)$

$$\begin{aligned} \int_a^b f(x)\,d\alpha(x) &= f(b)\,\alpha(b) - f(a)\,\alpha(a) - \int_a^b \alpha(x)\,df(x) \\ &= f(b)\,\alpha(b) - f(a)\,\alpha(a) - \alpha(\xi)\int_a^b df(x). \end{aligned}$$

This is precisely equation $(A.5)$ ■

A.2 The Laplace transform

Definitions

Let $\alpha(t)$ be a complex valued function of the real variable t defined on the interval $0 \leq t < \infty$. Denote its real and imaginary parts by $\alpha^{'}(t)$ and $\alpha^{''}(t)$ respectively, i.e.,

$$\alpha(t) = \alpha^{'}(t) + i\alpha^{''}(t).$$

Let $\alpha(t)$ be of bounded variation in the interval $0 \leq t \leq R$ for every positive R.

Let s be a complex variable with real and imaginary parts σ and τ respectively, i.e.,

$$s = \sigma + i\tau.$$

It follows from Theorem $A.3$ that integral

$$\int_0^R e^{-st} d\alpha(t)$$

exists for each positive R and for every complex s.

We now define the improper integral

$$\int_0^\infty e^{-st} d\alpha(t) = \lim_{R\to\infty} \int_0^R e^{-st} d\alpha(t) \tag{A.6}$$

When the integral $(A.6)$ converges it defines a function of s which we denote by $f(s)$. This function is called the Laplace-Stieltjes transform of $\alpha(t)$. If

$$f(s) = \int_0^\infty e^{-st} \phi(t)\, dt$$

we refer to $f(s)$ as the Laplace transform of $\phi(t)$. In either case $f(s)$ is called the generating function.

Appendix B

Moment Generating Functions

The moment generating function $\phi(t)$ of the random variable X is defined for all values t by

$$\begin{aligned}\phi(t) &= E\left[e^{tX}\right] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx,\end{aligned}$$

where $f(x)$ is the density function of X.
We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\begin{aligned}\phi'(t) &= \frac{d}{dt} E\left[e^{tX}\right] \\ &= E\left[\frac{d}{dt} e^{tX}\right] \\ &= E\left[X e^{tX}\right].\end{aligned}$$

Hence,

$$\phi'(0) = E[X]$$

Similarly,

$$\phi''(t) = E\left[X^2 e^{tX}\right]$$

and so

$$\phi''(0) = E\left[X^2\right].$$

In general, the nth derivative of $\phi(t)$ evaluated at $t = 0$ equals $E[X^n]$, that is,

$$\phi^{(n)} = E[X^n], \qquad n \geq 1$$

Moment generating function in term of Laplace-Stieltjes transform

Let X $(X \geq 0)$, be a continuous random variable and denote by F its distribution function. Then moment generating function $\phi(s)$ of X is defined as

$$\begin{aligned}\phi(s) &= E\left[e^{-sX}\right] \\ &= \int_0^\infty e^{-sx} dF(x). \qquad \text{(B.1)}\end{aligned}$$

Analogously as before, to obtain the nth moment of X we must evaluate the nth derivative of $\phi(s)$ at $s = 0$.

$$\begin{aligned}\phi'(s) &= E\left[\frac{d}{ds}e^{-sX}\right] \\ &= E\left[-Xe^{-sX}\right]\end{aligned}$$

Hence,

$$\phi'(0) = -E[X].$$

In general we obtain

$$E[X^n] = (-1)^n \frac{d^n\phi(s)}{ds^n}|_{s=0}$$

Recall that integration by parts in $(B.1)$ yields

$$\phi(s) = 1 - s\int_0^\infty e^{-sx}(1 - F(x))\,dx \qquad \text{(B.2)}$$

For example,
by $(B.2)$ we have

$$\phi'(s) = -\int_0^\infty e^{-sx}(1 - F(x))\,dx + s\int_0^\infty xe^{-sx}(1 - F(x))\,dx$$

Hence,

$$E[X] = -\phi'(0) = \int_0^\infty (1 - F(x))dx. \qquad \text{(B.3)}$$

In a similar way we can express the second moment of X.
Since

$$\phi''(s) = \frac{d}{ds}\phi'(s),$$

we get

$$\phi''(s) = 2\int_0^\infty xe^{-sx}(1 - F(x))\,dx - s\int_0^\infty x^2e^{-sx}(1 - F(x))\,dx,$$

and so

$$E\left[X^2\right] = \phi''(0) = 2\int_0^\infty x(1 - F(x))\,dx \qquad \text{(B.4)}$$

Bibliography

[1] F. Baccelli and G. Hebuterne. On queues with impatient customers. In *Performance'81,* pages 159-179. North-Holland, 1981

[2] W. Feller. *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons, 1970

[3] G.M. Laslett, D.B. Pollard, R.L. Tweedie. Techniques for establishing ergodic and recurrence properties of continuous-valued Markov chains. *Naval Research Logistics Quaterly,* Sept. 78, pages 455-472

[4] S.M. Ross. *Introduction to Probability Models.* Academic Press, 1993

[5] S.M. Ross. *Stochastic Processes.* John Wiley & Sons, 1983

[6] L. Tackas. *Introduction to the Theory of Queues.* Oxford University Press, 1962

[7] D.V. Widder. *The Laplace Transform.* Princeton University, 1964