République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université de Saida - Dr Moulay Tahar.
Faculté des Sciences.
Département de Mathématiques.

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Filière : Mathématiques

Spécialité: Analyse stochastique, statistique des processus et applications (ASSPA)

par

Tadjri **B**aghdad [a]

Sous la direction de

Dr. Mokhtar **K**adi

Thème:

A Multi-Server Markovian Feedback Queue with Balking Reneging and Retention of Reneged Customers

Soutenue le 26/06/2024 devant le jury composé de

Dr. Lamia. Bousmaha	Université de Saïda Dr. Moulay Tahar	Présidente
Dr. Kadi. Mokhtar	Université de Saïda Dr. Moulay Tahar	Encadreur
Dr. Lahcene. Yahiaoui	Université de Saïda Dr. Moulay Tahar	Examinateur

Année universitaire: 2023/2024

[a] e-mail:kingdoubaghdado@gmail.com

Dedication

To my father and mother, who accompanied me from the first moment of my life. You were my best supporter on all the roads I went on, and you blocked me from all the difficulties that tried to exhaust me and stop me. To you is this simple work whose grace is due to your after ***God Almighty****; without you, I would not have reached what I am today.*

*For the friends who became family "****Ilias Affane****", "****Charafe Dine Benjallal****", To all the friends who supported me in my academic career.*

*To my professor, "****Kadi Mokhtar****," who taught me the value of critical thinking and the power of the pen. To all the professors who helped me in this work, "****Lahcene Yahiaoui****," and in my academic career*

To the ones reading this, may you find the courage, joy, and love contained within these pages.

With my graduation today, I come to the end of the nights and days when worry kept me up all night and all day.

I want to start by thanking my parents and all my family (small, big)

Above all, i would like to thank my supervisor, Dr. Kadi. Mokhtar, for the time he has set aside for me, for his patience with me in guiding me, and for the wealth of knowledge that he has shared with me during the preparation of this master thesis.

Also, i would like to thank the committee members, Dr. L. Bousmaha and Dr. Lahcene. Yahiaoui, for examining my work.

Then i want to thanking my classmates for their encouragement and support throughout this master thesis.

Contents

List of Tables

List of Figures

List Of Notations And Symbols

$\mathbb{P}$:	Probability
$Var(.)$:	Variance
$\mathbb{E}(.)$:	Expected value
pmf :	Probability mass function
PDF :	Probability density function
CDF :	Cumulative distribution function
MGF :	Moment generating function
$\mathcal{E}xp$:	Exponential distribution
$\Gamma(.,.)$:	Upper incomplete Gamma function
$\mathcal{P}$:	Poisson Distribution
λ :	Average number of arrivals (arrival rate)
μ :	Average number of customers served (service rate)
$\frac{1}{\lambda}$:	Average time separating two consecutive arrivals
$\frac{1}{\mu}$:	Average length of service (one customer)
λ_n :	Average arrival intensity (arrivals per time unit) at n customers in the system
μ_n :	Average service intensity for the system when there are n customers
ρ :	Expected fraction of the time that the service facility is being used
T_B :	Exactly one birth
T_D :	Exactly one death
$P_n(t)$:	Probability that there are n customers in the system at time t

$\overline{N}$:	Average number of customers in the system
$\overline{N}_Q$:	Average number of customers who are waiting in line
$\overline{N}_S$:	Average number of customers being served
$\overline{T}$:	Average time that an arbitrary customer spends in the system
$\overline{Q}$:	Average waiting time of an arbitrary customer
$\overline{S}$:	Average service time of an arbitrary customer
L_s :	Average number of customers in the system
L_q :	Average number of customers in the queue
W_s :	Average waiting time of a customer in the system
W_q :	Average waiting time of customers in the queue
R_{aband} :	Rate of abandonment
R_r :	Average reneging rate
R_b :	Average balking rate
ξ :	Reneging rate
$q1$:	Probability of leaving after service
$p1$:	Probability of rejoining after service
$p2$:	Probability of leaving after reneging
$q2$:	Probability of staying after reneging
δ :	Interval of time infinitesimally small
N :	Capacity of the system

General Introduction

The queuing theory deals with one of life's most unpleasant experiences: waiting. Queueing is quite common in many fields, for example, in telephone exchange, in a supermarket, at a petrol station, at computer systems, etc. I have mentioned the telephone exchange first because the first problems of queueing theory were raised by calls, and Erlang was the first to treat congestion problems in the beginning of the 20th century[13]. the work of the Danish engineer Agner Krarup Erlang (1878–1929). His studies of how to best manage Copenhagen's telephone traffic to determine the number of circuits needed to provide acceptable telephone service are considered the first building blocks of this theory [5].Queueing theory became a field of applied probability, and many of its results have been used in operations research, computer science, telecommunications, traffic engineering, and reliability theory, just to mention a few[13].

Since Erlang's work, a large number of applications in all domains have been implemented and published. In 1953, it subsequently developed, thanks in particular to the work of Palm, Kolmogorov, and Khintchine. Kendall[a] proposed, in a research paper published in 1953, a notation to classify the various queueing models [17].In 1957, in a particularly elegant and efficient way, Jackson dealt with certain queue networks.

In this memo, we are interested in the system of queues. we aim to improve the understanding and management of queuing systems in environments where customer impatience and feedback significantly impact service quality and efficiency.

My memo is divided into three chapters, the first of which contains the basic notions of the study of queuing systems, namely stochastic processes:

→ Counting Processes,

→ Renewal process,

→ Markov Process,

→ Poisson Process,

→ Birth and Death Processes.

In the second chapter, we introduce certain definitions and notations in queuing theory, such as Kendall's notation, Little's law, etc. Then we study some queue models with single servers ($M/M/1$, $M/M/1/N$) and with many servers ($M/M/c$, $M/M/\infty$) and evaluate their performance parameters.

[a]David George Kendall, retired professor of the University of Cambridge, in England.

Finally, in the third chapter, we will study a finite-capacity multi-server Markovian feedback queuing model with balking, reneging, and retention of reneged customers. We get the steady-state probability distribution of the system size as well as some measures of service quality, such as the average system size, the average number of people served, and so on.

Chapter 1

Stochastic processes

Stochastic processes are ways of quantifying the dynamic relationships of sequences of random events. Stochastic models play an important role in elucidating many areas of the natural and engineering sciences. They can be used to analyze the variability inherent in biological and medical processes, to deal with uncertainties affecting managerial decisions and with the complexities of psychological and social interactions, and to provide new perspectives, methodology, models, and intuition to aid in other mathematical and statistical studies[12].
stochastic processes, **renewal processes**, and **birth-death processes** are closely related concepts that play vital roles in the modeling and analysis of queueing systems. Stochastic processes provide a framework for understanding the random behavior of these systems over time, while renewal processes and birth-death processes offer specific modeling techniques for capturing arrival patterns and system dynamics, respectively, within the queueing context

Basic definitions and properties

Definition 1.0.1. A stochastic process is a family of random variables $X(t)$, $t \in T$ where each random variable $X(t)$ is indexed by the parameter $t \in T$. If T is a set of $\mathbb{R}_+$, then t means time.
Generally $X(t)$ represents the state of the stochastic process at time t,
► if T is countable, i.e. $T \subseteq N$, then we say that $X(t)$, $t \in T$ is a discrete-time process,
► if T is an interval of $[0; \infty)$, then the stochastic process is said to be a continuous-time process.

The set of values of $X(t)$ is called the state space, which can also be either discrete (finite or countably infinite) or continuous (a subset of $\mathbb{R}$ or $\mathbb{R}_n$), so we write $(X_n)_{n\geq 0}$ for discrete-time process and $(X_t)_{t\geq 0}$ for the continuous-time process.

1.1 Counting Processes

Definition 1.1.1. A stochastic process $(N(t))_{t\geq 0}$ is said to be a counting process if $N(t)$ counts the total number of (events) that have occurred up to time t. Hence, it must satisfy:

(i) $N(t) \geq 0$ for all $t \geq 0$,

(ii) $N(t)$ is integer-valued,

(iii) If $s < t$, then $N(s) \leq N(t)$,

(iv) For $s < t$, the increment $N((s,t]) \stackrel{def}{=} N(t) - N(s)$ equals the number of events that have occurred in the interval $(s,t]$.

A counting process is a continuous-time discrete process. A second process can be associated with the process of occurrence times, the interarrival times process $W_n, n \in N_0$ or $\forall n \in N_0$ the random variable W_n is the elapsed time between the $(n-1)^{\mathbf{st}}$and the $n^{\mathbf{th}}$ event,i.e.

$$W_n = T_n - T_{n-1} \qquad T_n : \text{ the time at which the } n^{th}\text{arrival occurs.}$$

Consider the first arrival T_1. Saying that $T_1 > t$ (the first arrival will occur in the future) is the same as saying that $N_t = 0$ (no arrivals have occured yet). Therefore[21].

$$\mathbb{P}\{T_1 > t\} = \mathbb{P}\{N_t = 0\} = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Proof. We need to prove that: $W_n = T_n - T_{n-1}$

$$\begin{aligned} W_1 + W_2 + ... + W_n &= T_1 - T_0 + T_2 - T_1 + T_3 - T_2 + ... + T_{n-1} - T_{n-2} + T_n - T_{n-1} \\ &= T_0 + T_n \\ &= T_n \quad car \quad T_0 = 0 \end{aligned}$$

□

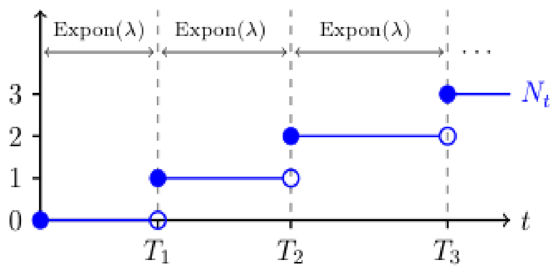

Figure 1.1: Counting process

1.2 Renewal process

Renewal theory began with the study of stochastic systems whose evolution through time was interspersed with renewals or regeneration times when, in a statistical sense, the process began anew [12]. Consider a device which starts to work at time 0 and works T_1 units of time. At time T_1 this device is replaced by another device which works for T_2 units of time. At time $T_1 + T_2$ this device is replaced by a new one, and so on. Let us denote the working time of the i-th device by T_i [25].

Definition 1.2.1. Let us assume that $(T_1, T_2, ...)$ are independent and identically distributed random variables with $\mathbb{P}[T_i > 0] = 1$. The times:

$$S_n = T_1 + ... + T_n, \quad n \in \mathbb{N}$$

Are called renewal times because at time S_n some device is replaced by a new one. Note that $0 < S_1 < S_2 < ...$, The number of renewal times in the time interval $[0, t]$ is:

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{S_n \leq t} = \{n \in \mathbb{N} : S_n \leq t\}, t \geq 0$$

The process $\{Nt : t \geq 0\}$ is called a **renewal process** [25].

Example 1.2.1. Traffic flow the distances between successive cars on an indefinitely long single-lane highway are often assumed to form a renewal process. So also are the time durations between consecutive cars passing a fixed location [12].

Proposition 1.2.1. The following relationships are trivial such that $T_0 = 0$

1. $T_n = W_1 + W_2 + \cdots + W_n, \qquad \forall n \geq 1$ We proof this in 1.1;
2. $N(t) = \sup\{n \geq 0 : T_n \leq t\}$;
3. $\mathbb{P}[N(t) = n] = \mathbb{P}[T_n \leq t \leq T_{n+1}]$;
4. $\mathbb{P}[N(t) \geq n] = \mathbb{P}[T_n \leq t]$;
5. $\mathbb{P}[s < T_n < t] = \mathbb{P}[N(s) < n \leq N(t)]$.

Remark.

- A counting process is said to have independent increments if the numbers of events that occur in disjoint time intervals are independent, that is, the family $(N(I_k))_{1 \leq k \leq n}$ of independent random variables whenever I_1,...,I_n forms a collection of pairwise disjoint intervals. In particular, $N(s)$ is independent of $N(s+t) - N(s)$ for all $s, t \geq 0$.
- A counting process is said to have stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. In other words, the process has stationary increments if the

number of events in the interval $(s, s+t]$, i.e. $N((s, s+t])$ has the same distribution as $N((0,t])$ for all $s, t \geq 0$.

1.3 Markov Process

Definition 1.3.1. (stationary increments)

The process $\{X_t : t \geq 0\}$ has stationary increments if for all $n \in N$, $h \geq 0$ and $0 \leq t_0 \leq t_1 \leq ... \leq t_n$, we have the following equality in distribution:

$$(X_{t_1+h} - X_{t_0+h}, X_{t_2+h} - X_{t_1+h}, ..., X_{t_n+h} - X_{t_n-1+h}) \stackrel{d}{=} (X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}})$$

.

Definition 1.3.2. (independent increments)

The process $\{X_t : t \geq 0\}$ has independent increments if for all $n \in N$ and $0 \leq t_0 \leq t_1 \leq ... \leq t_n$,the random variables

$$X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$$

are independent.

Definition 1.3.3. Markov process[2]

A process $\{X(t)\}$ is said to be Markov process, if

$$\mathbb{P}\left(a < X(t) \leq b \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, ..., X(t_0) = x_0\right) = \\ \mathbb{P}\left(a < X(t) \leq b \mid X(t_n) = x_n\right)$$

for all n, where $t_0 \leq t_1 \leq ... \leq t_n \leq t$

Definition 1.3.4. A counting process $\{N_t\}_{t\geq 0}$ Poisson process with rate $\lambda > 0$ if it satisfies the following properties:

1. $N_t - N_s \perp\!\!\!\perp \{\mathrm{N}_r\}_{r\leq s}$ for $t \geq s$ (independent increments).
2. $N_t - N_s \sim Pois(\lambda(t-s))$ for $t \geq s$ (stationary increments).

1.4 Poisson Distribution and Poisson Process

1.4.1 Poisson Distribution

Definition 1.4.1. (**The Poisson Distribution**)

A recall that X has a Poisson distribution with mean μ, or $X = \mathcal{P}(\mu)$,for short, if

$$\mathbb{P}(X=n)=e^{-\mu}\frac{\mu^n}{n!} \text{ for } n=1,2,3,...$$

Proposition 1.4.1. Let X be a Poisson random variable $\mathcal{P}(\mu)$

→ The **MGF** a poisson distribution is

$$\varphi(t)=\mathbb{E}\left[e^{tX}\right]=e^{\mu\left(e^t-1\right)}$$

→ The mean and variance are

$$\mathbb{E}[X]=\mu \quad \mathrm{Var}(X)=\mu$$

proof

STEP 1 From the definition 1.4.1 of a **MGF**

$$\begin{aligned}\varphi(t)&=\sum_{k=0}^{\infty}e^{tk}\frac{\mu^k}{k!}e^{-\mu}\\&=e^{-\mu}\sum_{k=0}^{\infty}\frac{(\mu e^t)^k}{k!}\\&=e^{-\mu}e^{\mu e^t}\end{aligned}$$

STEP 2 Calculate the derivatives of the generating function

$$\begin{aligned}\varphi'(t)&=\mu e^t e^{\mu\left(e^t-1\right)}\\\varphi''(t)&=\left(1+\mu e^t\right)\mu e^{\mu}e^{\mu\left(e^t-1\right)}\end{aligned}$$

hence

$$\begin{aligned}\mathbb{E}[X]&=\varphi'(0)=\mu\\\mathrm{Var}(X)&=\varphi''(0)-\mathbb{E}[X]^2=\mu\end{aligned}$$

Definition 1.4.2. (**The Exponential Distribution**) A continuous random variable X is said to have an **exponential distribution** with parameter $\lambda>0$,shown as $X\sim\mathcal{E}(\lambda)$if its **PDF** is given by

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\geq 0\\ 0 & \text{otherwise}\end{cases}$$

Proposition 1.4.2. Suppose X follows an exponential distribution with rate λ; that is,$X\sim\mathcal{E}(\lambda)$. Then, the **CDF** is given by

→ We can express the **CDF** as $F(t)=\begin{cases}1-e^{-\lambda t} & \text{if } \quad t\geq 0\\ 0 & \text{otherwise.}\end{cases}$

→ The **MGF** of X is $\varphi(t)=\mathbb{E}\left[e^{tX}\right]=\begin{cases}\infty & \text{if } \quad t\geqslant\lambda\\ \dfrac{\lambda}{\lambda-t} & \text{if } \quad t<\lambda\end{cases}$

→ The **mean** and **variance** of an exponential distribution is $\mathbb{E}[X] = \frac{1}{\lambda}$, $\quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$

proof Suppose $X \sim \mathcal{E}(\lambda)$.

→ The CDF is 1.4.2 if $x < 0$ we have $(F_x(x) = \int_{-\infty}^{0} 0dx = 0)$ if $x \geq 0$ we have using (Definition 1.4.2):

$$
\begin{aligned}
F(t) &= \int_0^t f_X(x)dx \\
&= \int_0^t \lambda e^{-\lambda x}dx \\
&= 1 - e^{-\lambda t}
\end{aligned}
$$

→ The **MGF**is 1.4.2

$$
\begin{aligned}
\varphi(t) &= \mathbb{E}\left[e^{tX}\right] \\
&= \int_0^{+\infty} \lambda e^{tx}e^{-\lambda x}dx \\
&= \int_0^{+\infty} \lambda e^{(t-\lambda)x}dx \\
&= \left[\frac{\lambda}{t-\lambda}e^{(t-\lambda)x}\right]_0^{+\infty}
\end{aligned}
$$

→ After calculating the first and second derivative of a function 1.4.2

$$\varphi'(t) = \frac{\lambda}{(\lambda - t)^2} \quad \varphi''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

Now, substituting the mean and second moment of the exponential distribution, we get

$$\mathbb{E}[X] = \varphi'(0) = \frac{1}{\lambda} \text{ and } \mathrm{Var}(X) = \varphi''(0) - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$$

1.4.2 Relationship between the Exponential distribution and the distribution of Poisson

Theorem 1.4.1. Let $\{N(t), t \geq 0\}$ is a poisson process with rate λ.

Proof

Fix an integer $n \geq 0$.Then $S_n = T_1 + ... + T_n \sim \Gamma(n, \lambda)$ and it is independent of T_{n+1}
By definition of $N(t)$

$$\begin{aligned}
\mathbb{P}(N(t) = n) &= \mathbb{P}(S_n \leq t, S_n + T_{n+1} > t) \\
&= \int_0^t \int_{t-s}^{\infty} f_{S_n}(s) f_{T_{n+1}}(x) dx ds \\
&= \int_0^t \mathbb{P}(T_{n+1} > t - s) f_{s_n}(s) ds \\
&= \int_0^t e^{-\lambda(t-s)} \frac{\lambda(\lambda s)^{n-1} e^{-\lambda s}}{(n-1)!} ds \\
&= \frac{(\lambda t)^n e^{-\lambda t}}{n!}
\end{aligned}$$

Definition 1.4.3. (**Lack of memory property**)[22]

It is traditional to formulate this property in terms of waiting for an unreliable bus driver. In words, (if we've been waiting for t units of time then the probability we must wait's more units of time is the same as if we haven't waited at all.) In symbols

$$\mathbb{P}(T > t + s \mid T > t) = \mathbb{P}(T > s)$$

To prove this we recall that if $B \subset A$,then $\mathbb{B} \mid \mathbb{A} = \mathbb{P}(B)/\mathbb{P}(A)$,so

$$\mathbb{P}(T > t + s \mid T > t) = \frac{\mathbb{P}(T > t + s)}{\mathbb{P}(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(T > s)$$

where in the third step we have used the fact $e^{a+b} = e^a e^b$

1.5 Poisson Process

Poisson behavior is so pervasive in natural phenomena and the Poisson distribution is so amenable to extensive and elaborate analysis as to make the Poisson process a cornerstone of stochastic modeling[12].Point processes contribute components to a solution of many varied modelling problems. We may need to model any of the following:

(1) The times of arrivals (departures, service initiations, and so forth) in a queue;

(2) The breakdown times (repair times) of a machine or a group of machines;

(3) The positions and times of earthquakes in the next 50 years;

(4) The location of oil relative to a known deposit;

(5) The location of trees in a forest;

(6) The location of tanks in a battlefield [23].

Three Ways To Define The Poisson Process

Definition 1.5.1. (**The Axiomatic Way**)

The counting process $(N(t)_{t\geq 0})$ is said to be a Poisson process with rate (or intensity) $\lambda, \lambda > 0$ if:

(PP1) $N(0) = 0$;

(PP2) The process has independent increments;

(PP3) The number of events in any interval of length t is Poisson distributed with mean λt. That is, for any $s, t \geq 0$:

$$\mathbb{P}(N((s,t]) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, n = \mathbb{N}_0$$

If $\lambda = 1$,then $(N(t))_{t\geq 0}$ is also called standard Poisson process.

Condition (PP1)1.5.1,indicates the start of event counting at t = 0, and condition (PP2)1.5.1 can usually be verified directly from our knowledge of the process. However, it is not at all clear how we could determine validity of condition (PP3)1.5.1, and for this reason an equivalent definition of a Poisson process would be useful.
A function $f : R \to R$ is said to be $o(h)$ (for $h \to 0$), if:

$$\lim_{h\to 0}\frac{f(h)}{h} = 0$$

Definition 1.5.2. (**By Infinitesimal Description**)

A counting process $(N(t)_{t\geq 0})$ is said to be a Poisson process with rate $\lambda, \lambda > 0$, if:

(PP1) $N(0) = 0$;

(PP4) The process has stationary and independent increments;

(PP5) $\mathbb{P}(N(h) = 1) = \lambda h + o(h)$;

(PP6) $\mathbb{P}(N(h) \geq 2) = o(h)$;

Definition 1.5.3. (**The Constructive Way**)

A counting process $(N(t)_{t\geq 0})$ is said to be a Poisson process with rate $\lambda, \lambda > 0$, if

$$N(t) = \sum_{n\geq 1} \mathbb{1}_{(0,t]}(T_n),\ t \geq 0$$

or a sequence $(Tn)_{n\geq 1}$ having i.i.d. increments $Y_1, Y_2, ...$, say, with an $\mathcal{E}(\lambda)$-distribution. The T_n are called jump or arrival epochs and the Y_n interarrival or sojourn times associated with $(N(t))_{t\geq 0}$.

1.6 Birth and Death Processes

Definition 1.6.1. (**Birth and Death Processes**)

We can realize out a process of birth and death in the following way:

→ The arrivals and departures of entities obey exponential laws with respective rates $\lambda(n)$ and $\mu(n)$;

→ Using regularity hypothesis: two events cannot occur at the same time, therefore the probability that two events occur in a time interval dt is negligible;

→ There is a transition to a neighboring state, either by the arrival of a client (birth) or by the departure of a client (death);

If $P_n(t)$ is the probability that there are n customers in the system at time t, the Kolomogorov equation [1] is written, for $n > 0$:

$$P_n(t + dt) = (1 - (\lambda_n + \mu_n)\, dt)\, P_n(t) + \mu_{n+1} P_{n+1}(t) dt + \lambda_{n-1} P_{n-1}(t) dt + o(dt)$$

We will tend dt towards 0, for $n > 0$:

$$\frac{d}{dt} P_n(t) = -\left(\lambda_n + \mu_n\right) P_n(t) + \mu_{n+1} P_{n+1}(t) + \lambda_{n-1} P_{n-1}(t)$$

we obtain for $n = 0$.

$$\frac{d}{dt} P_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t).$$

A particular case of the process of birth and death is the Poisson process with $\mu_n = 0$ and $\lambda_n = \lambda$ in this case we do not find a stationary regime [7] the differential equations are then written

$$P_0(t) = e^{-\lambda t} \qquad \frac{d}{dt} P_0(t) = -\lambda_0 P_0(t), \qquad \frac{d}{dt} P_n(t) = -\lambda \left(P_n(t) - P_{n-1}(t)\right)$$

The solution is $\quad P_n(t) = \dfrac{(\lambda t)^n e^{-\lambda t}}{n!}$

1.6.1 Birth Process

T_B are mutually independent stochastic variables and state transitions occur through exactly one Birth $(n \longrightarrow n + 1)$.

1.6.2 Death Processes

T_D are mutually independent stochastic variables and state transitions occur through exactly one Death ($n \longrightarrow n-1$).

1.6.3 Birth and Death Processes

The foundation of many of the most commonly used queuing models:

→ **Birth:** Equivalent to the arrival of a customer or job;

→ **Death:** Equivalent to the departure of a served customer or job;

Assumptions Give $N(t) = n$

- The time until the next birth (T_B) is exponentially distributed with parameter λ_n (Customers arrive according to a $\mathcal{P}$ process);
- The remaining service time (T_D) is exponentially distributed with parameter μ_n;

Transition graph of a birth and death process

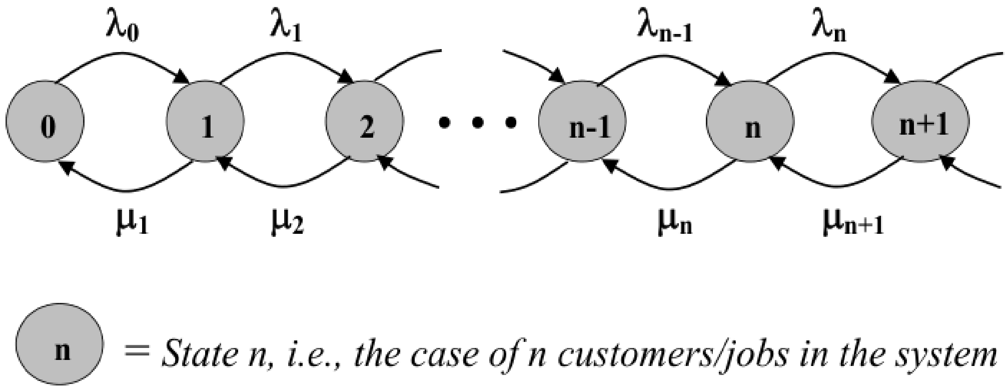

Figure 1.2: A Birth-and-Death Process Rate Diagram

→ λ_n: Birth rate when the number of population equals n

→ μ_n: Death rate when the number of population equals n.

Chapter 2

Queueing System

Queuing theory began in 1909 with Danish engineer Agner Krarup Erlang's (1878,1929) research into Copenhagen telephone traffic to determine the number of circuits needed to provide acceptable telephone service. Subsequently, the queues were integrated into modeling various fields of activity [10]. The phenomenon of queuing occurs naturally in most of the environments encountered in everyday life, including computer systems, production systems, and telecommunications networks. These phenomena can be modeled by an operational research technique called queue theory. This theory aims to optimize available resources and manage the waiting time of customers requesting a specific service for a specified period of time [8]. The purpose of a queueing system is to balance the demand for services with the capacity of the system to provide those services, ensuring that customers are served in a timely and organized manner. Queueing systems are essential for optimizing resource utilization, reducing waiting times, improving customer satisfaction, and enhancing operational efficiency in a wide range of industries.

In this chapter, you will be introduced to the basic structure, the terminology and the characteristics before embarking on the actual study of queueing [14].

Definition 2.0.1. (**Queue**)

A queueing system is a facility that consists of one or several servers designed to perform specific tasks or process jobs, along with a queue of jobs waiting to be processed. We suppose that the customers who arrive in the system come to receive some service or to perform a certain task (for example, to withdraw money from an automated teller machine). In a queueing system, jobs arrive at the system, wait for an available server, get processed by the server, and then leave. If we want to be precise, the queue should designate the customers who are waiting to be served, that is, who are queueing, while the queueing system includes all the customers in the system.Since queue is the standard expression for this type of process. may be, for example, airplanes that are landing or are waiting for the landing authorization, or machines that have been sent to a repair shop, etc [17].

2.1 The different types of queues

The following figures represent the different queuing systems according to the waiting space and the service space[9]

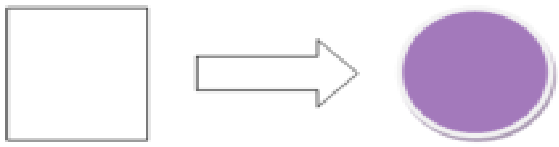

Figure 2.1: Queue with a single waiting area and a single server

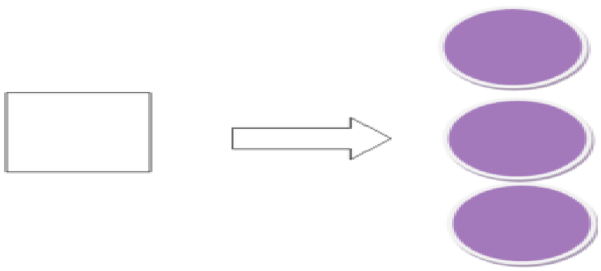

Figure 2.2: Queue with a single waiting area and multiple servers

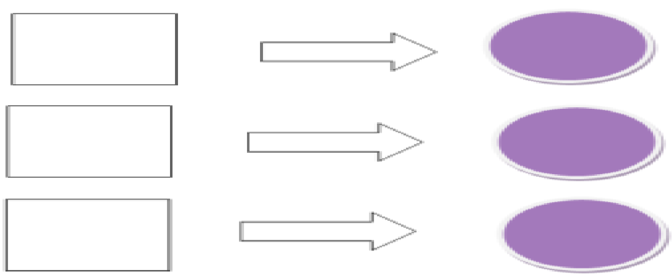

Figure 2.3: Queue with multiple waiting areas and multiple servers

2.1.1 Components of Queuing System

A queuing system typically includes the following elements:

1. **Arrival Process:** The arrival process describes how customers enter the system with rate λ.
2. **Service time distribution:** The amount of time required to serve a customer is described as service time distribution $\overline{S}$.
3. **Server:** The server c is the person who provides the service to the customers.
4. **Queue:** The customers who are waiting for service are held in a queue, as shown in the 2.4.
5. **Departure process:** The departure process describes how customers exit the system with rate μ once they have been served.
6. **Service discipline:** The order in which customers are served is determined by service discipline.
7. **System performance measures:** The System performance measures are used to analyze and evaluate the system's performance. Examples include the average wait time, the number of customers in the system, and the server's utilization.

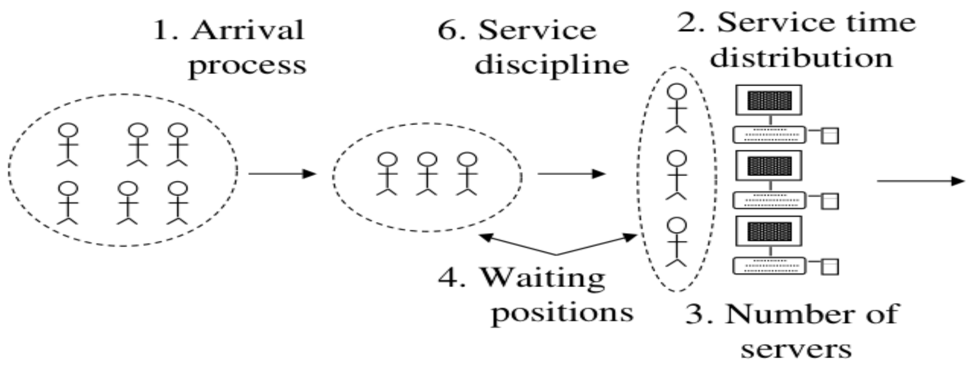

Figure 2.4: Basic Components of a Queue

2.2 The Simple Queue

A simple queue is a system consisting of one or more servers and a waiting area. customers arrive from outside, possibly wait in the queue, receive service, then leave the station [15]. In order to fully specify a simple queue, one must characterize the customer arrival process, service time, and the structure and service discipline of the queue.

Definition 2.2.1. (**The system capacity**)[14]

The system capacity refers to the maximum number of customers that a queueing system can accommodate, inclusive of those customers at the service facility

In a multi-server queueing system, as shown in Figure 2.4, the system capacity is the sum of the maximum size of the waiting queue and the number of servers. If the waiting queue can accommodate an infi nite number of customers, then there is no blocking, arriving customers simply joining the waiting queue. If the waiting queue is fi nite, then customers may be turned away. It is much easier to analyse queueing systems with infi nite system capacity as they often lead to power series that can be easily put into closed form expressions.

Arrival process

The arrival of customers at the station will be described using a stochastic counting process $(N_t)_{t\geq 0}$
If A_n designates the random variable measuring the time of arrival of the nth customer in the system, we will have:

$$A_0{=}0 \quad \text{and} \quad A_n = inf\{t; N_t = n\}$$

If T_n designates the random variable measuring the time separating the arrival of the $(n-1)$th and the nth customer [16] , we then have:

$$T_n{=}A_n - A_{n-1}$$

2.2.1 Structure of Queuing Systems

To study queues, we will use an abstract model called service station. This station is made up of one or more servers representing the resource and a queue containing at any time the clients waiting for service (busy servers). If a service station is free, the arriving customer immediately goes to this station where he is served, otherwise, he takes his place in a queue in which the customers line up according to their order of arrival[8].

2.3 Kendall Notation

Kendall proposed, in a research paper published in 1953, a notation to classify the various queueing models [17]:

$$\mathrm{A/B/c/N/p/D}$$

where:

1 → **A**: Denotes the distribution of the time between two successive arrivals. Again, we commonly use a single letter to indicate the type of service distribution [14]:

→ M:Markovian (or Memoryless) , imply exponential distributed service times;

→ D: Deterministic ; constant service times;

→ E_k:Erlang of order K service time distribution;

→ G:General service times distribution.

2 → **B**: Denotes the distribution of the service time of customers,

3 → **c**: Is the number of servers in the system,

4 → **N**: Is the capacity of the system,

5 → **p**: Is the size of the population from which the customers come,

6 → **D**: Designates the service policy, called the discipline, of the queue.

We suppose that the times W_n between the arrivals of successive customers are independent and identically distributed random variables. Similarly, the service times S_n of the customers are random variables assumed to be i.i.d. and independent of the $\mathcal{W}_n$. Actually, we could consider the case when these variables, particularly the S_n, are not independent among themselves.

Remark. [17]

→ We can use the notation GI for general independent, rather than G, to be more precise;

→ The number s of servers is a positive integer, or sometimes infinity. (For example, if the customers are persons arriving in a park and staying there some time before leaving for home or elsewhere, in which case, the customers do not have to wait to be served.);

→ By default, the capacity of the system is infinite. Similarly, the size of the population from which the customers come is assumed to be infinite. If c (or p) is not equal to infinity, its value must be specified. On the other hand, when $c = p = \infty$, we may omit these quantities in the notation;

→ Finally, the queue discipline is, by default, that of first-come, first-served, which we denote by **FCFS** or by **FIFO**, for first-in, first-out We may also omit this default discipline in the notation. In all other cases, the service policy used must be indicated. We can have LIFO, that is, last-in, first-out. The customers may also be served at random **RANDOM**. Sometimes one or more special customers are receiving priority service, etc.

Service discipline:[18]

The logical ordering of customers in a queue that determines which customer is chosen for service when a server becomes free, for example:

→ First-in-first-out(FIFO)

→ Last-in-first-out(LIFO)

→ Service in random order (SIRO)

→ Shortest processing time first (SPT)

→ Service according to prioroty (PR)

2.4 Little's Theorem

Before we examine the stochastic of a queueing system, let us first establish a very simple and yet powerful result that governs its steady-state performance measures Little's theorem, law or result are the various names. This result existed as an empirical rule for many years and was first proved in a formal way by Little in 1961 (Little 1961)[14]

Theorem 2.4.1. [14].

This theorem relates the average number of customers (L) in a steady-state queueing system to the product of the average arrival rate (λ) of customers entering the system and the average time (W) a customer spent in that system, as follows:

$$L = \lambda W$$

This result was derived under very general conditions. The beauty of it lies in the fact that it does not assume any specific distribution for the arrival as well as the service process, nor it assumes any queueing discipline or depends upon the number of parallel servers in the system. With proper interpretation of L, λ and W, it can be applied to all types of queueing systems, including priority queueing and multi-server systems.

2.5 Queues with a single server

The model M/M/1

We first consider a queueing system with a single server, in which the customers arrive according to a Poisson process with rate A, and the service times are independent exponential random variables, with mean equal to $1/\mu$. We suppose that the system capacity is infinite, as well as the population from which the customers come. Finally, the queue discipline is that of first-come, first-served. We can therefore denote this model simply by M/M/1 [17].

The balance equations of the system are the following:

$$\begin{cases} \quad 0 \quad , \quad \lambda P_0 = \mu P_1 \\ n(\geq 1) \quad , \quad (\lambda + \mu) P_n = \lambda P_{n-1} + \mu P_{n+1} \end{cases} \tag{2.0}$$

we have:

$$P_0 := 1 \text{ and } P_n = \underbrace{\frac{\lambda\lambda\cdots\lambda}{\mu\mu\cdots\mu}}_{\text{n*}} = \left(\frac{\lambda}{\mu}\right)^n \qquad \text{for } n = 1, 2, ...$$

If $\lambda < \mu$, the process $\{X(t), t > 0\}$ is positive recurrent, then enables us to write that:

$$\mathrm{P}_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k} = \frac{\left(\frac{\lambda}{\mu}\right)^n}{\left[1 - \frac{\lambda}{\mu}\right]^{-1}} \qquad \text{for } n = 0, 1, ...$$

That is,

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \quad \forall n \geq 0 \tag{2.1}$$

We can now calculate the quantities of interest. We aheady know that $\overline{S} = \dfrac{1}{\mu}$.Moreover, because here $\lambda_e = \lambda$,we may write that

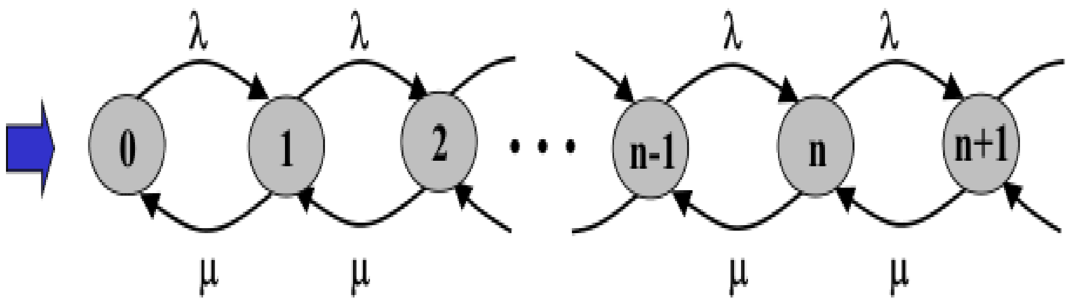

Figure 2.5: State-transition diagram for the model M/M/1.

$$\overline{N}_S = 1 - P_0 = 1 - \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu}$$

because the random variable N_S denoting the number of persons who are being served, when the system is in equilibrium,

$$p_0 := 1 - P_0$$

→ **The average number of customers in the system is $\overline{N}$:**

$$\begin{aligned}
\overline{N} &= \sum_{n=0}^{\infty} nP_n = \left(\frac{\lambda}{\mu}\right) \sum_{n=1}^{\infty} n \left(\frac{\lambda}{\mu}\right)^{n-1} \left(1 - \frac{\lambda}{\mu}\right) \\
&= \left(\frac{\lambda}{\mu}\right) \mathbb{E}\left[Z^{a}\right] \\
&= \left(\frac{\lambda}{\mu}\right) \frac{\mu}{\lambda - \mu} = \frac{\lambda}{\lambda - \mu} \qquad (2.2)
\end{aligned}$$

→ **The average number of customers in line $\overline{N}_Q$:**

$$\begin{aligned}
\overline{N}_Q &= \sum_{n \geq 1}^{\infty} (n-1) P_n = \overline{N} - \overline{N}_S = \frac{\lambda^2}{\mu(\mu - \lambda)} \\
&= \frac{\rho^2}{1 - \rho}
\end{aligned}$$

→ **The Average time a customer spends in the system $\overline{T}$:**
(using Little's law)

$$\begin{aligned}
\overline{T} &= \overline{N}/\lambda = \overline{Q} + \overline{S} \\
&= \frac{\rho}{\lambda(1 - \rho)} = \frac{\lambda}{\mu(\mu - \lambda)} + \frac{1}{\mu} \\
&= \frac{1}{\mu - \lambda} \qquad (2.3)
\end{aligned}$$

→ **The Average waiting time of customer $\overline{Q}$:**

$$\begin{aligned}
\overline{Q} &= \frac{\overline{N}_Q}{\lambda_e} = \overline{T} - S \\
&= \frac{\lambda}{\mu(\mu - \lambda)} \qquad (2.4)
\end{aligned}$$

2.5.1 The model M/M/l/N

Although the M/M/1 queue is very useful to model various phenomena, it is more realistic to suppose that the system capacity is an integer $N < \infty$. For $i = 0, 1, ..., N - 1$, the balance equations of the system remain the same as when $N = \infty$. However, when the system is in state N, it can only leave it because of the departure of the customer being served. In addition, this state can only be entered from $N-1$, with the arrival of a new customer. We thus have:

$$\begin{cases} 0 \quad , \quad \lambda P_0 = \mu P_1 \\ 1 \leq k \leq N-1 \quad , \quad (\lambda+\mu)P_k = \lambda P_{k-1} + \mu P_{k+1} \\ N \quad , \quad \mu P_N = \lambda P_{N-1} \end{cases} \tag{2.4}$$

As in the case **when the system capacity is infinite**, we find that:

$$\mathbb{P}_k = \left(\frac{\lambda}{\mu}\right)^k \quad \text{for } k = 0, 1, \cdots N$$

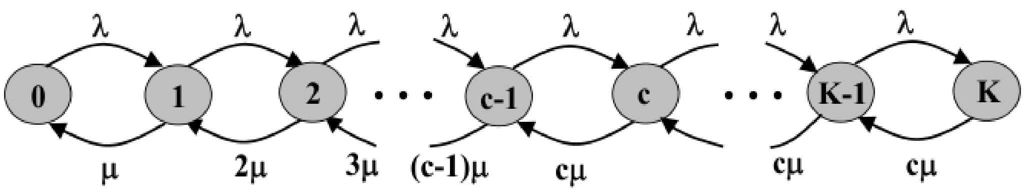

Figure 2.6: State-transition diagram for the model M/M/1/c

It follows, if $\rho := \dfrac{\lambda}{\mu} \neq 1$, that:

$$\sum_{k=0}^{N} P_k = \sum_{k=0}^{N} \left(\frac{\lambda}{\mu}\right)^k = \sum_{k=0}^{N} \rho^k = \frac{1-\rho^{N+1}}{1-\rho} \tag{2.5}$$

Theorem 2.5.1. When $\lambda = \mu$, we have that $\rho = 1, P_k = 1$, and $\sum_{k=0}^{N} \rho = N+1$, from which we find

$$P_i = \frac{P_i}{\sum_{k=0}^{N} P_k} = \begin{cases} \dfrac{\rho^i(1-\rho)}{1-\rho^{N+1}}, & \text{if } \rho \neq 1 \quad (2.6) \\ \dfrac{1}{N+1}, & \text{if } \rho = 1 \quad (2.7) \end{cases}$$

for $i = 0, 1, \cdots, N$.

→ **The average number of customers in the system is $\overline{N}$:**

$$\begin{aligned}\overline{N} &= \sum_{k=0}^{N} kP_k \\ &= \frac{\rho}{1-\rho}\frac{1-(N+1)\rho^N + N\rho^{N+1}}{1-\rho^{N+1}} \qquad (2.8)\end{aligned}$$

Remark. We easily in the case of the M/M/1 queue find that :

$$\lim_{N\to\infty} \overline{N} = \begin{cases} \dfrac{\rho}{1-\rho}, & \text{if } \rho < 1 \qquad (2.9) \\ \infty, & \text{if } \rho \geq 1 \qquad (2.10) \end{cases}$$

→ **The average number of customers in the queue is $\overline{N}_Q$:**

$$\begin{aligned}\overline{N}_Q &= \sum_{n=1}^{\infty}(n-1)P_n \\ &= \overline{N} - 1 + P_0\end{aligned}$$

→ **The Average time a customer spends in the system $\overline{T}$:**

$$\overline{T} = \frac{\overline{N}}{\lambda(1-P_N)}$$

If we consider an arbitrary customer arriving in the system, the average time that she will spend in it is then:

$$\mathbb{E}(T) = 0 \times P_N + \frac{\overline{N}}{\lambda(1-P_N)}(1-P_N) = \frac{\overline{N}}{\lambda} \qquad (2.11)$$

And allso we writ:

$$\overline{S} = \frac{1}{\mu}(1-P_N) \Longrightarrow \overline{Q} = \frac{\overline{N}}{\lambda} - \frac{1}{\mu}(1-P_N)$$

2.6 Queues with many servers

2.6.1 The model M/M/c

An important generalization of the M/M/1 model is obtained by supposing that there are c servers in the system and that they all serve at an exponential rate μ. the other basic assumptions that were made in the description of the M/M/1 model remain valid. thus, the customers arrive in the system according to a Poisson process with rate λ. The capacity of the system is infinite, and the service policy is that by default, namely, first-come, first-served.

$$\lambda_n = \lambda$$

$$\mu_n = \begin{cases} n\mu, & \text{For } i = 1, 2, ..., c-1; \\ c\mu, & \text{For } n \geq c; \end{cases}$$

$$\begin{cases} 0 & , \quad \lambda P_0 = \mu P_1 \\ 1 < k < c & , \quad (\lambda + k\mu)P_k = \lambda P_{k-1} + (k+1)\mu P_{k+1} \\ k \geq c & , \quad (\lambda + c\mu)P_k = \lambda P_{k-1} + c\mu P_{k+1} \end{cases}$$

with

$$\sum_{k=0}^{\infty} P = 1$$

Steady state condition:$\rho = \dfrac{\lambda}{N\mu} < 1$

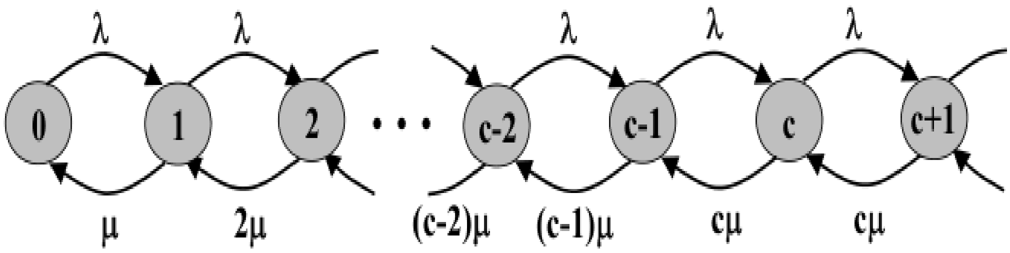

Figure 2.7: State-transition diagram for the model M/M/c

Once P_0 has been calculated, we se

$$P_k = \begin{cases} \dfrac{\rho^k}{k!} P_0, & \text{if } k = 0, 1, \cdots, c \quad (2.12) \\ \dfrac{\rho^k}{c!c^{k-c}} P_0, & \text{if } k = c+1, c+2, \cdots \quad (2.13) \end{cases}$$

→ **The average number of customers in line $\overline{N}_Q$:**

$$\overline{N}_Q = \frac{\rho^{c+1}}{c!c} \frac{c^2}{(c-\rho)^2} P_0 = \frac{\rho^{c+1}}{c!c} \frac{1}{(1-\xi)^2} P_0 \qquad (2.14)$$

with $\xi = \dfrac{\rho}{c}$, From $\overline{N}_Q$ and Little's formula.

→ **The average number of clients in the system $\overline{N}$:**

$$\begin{aligned}\overline{N} &= \lambda\overline{T} = \overline{N}_Q + \rho \\ &= \frac{\rho^{c+1}}{c!c}\frac{1}{(1-\xi)^2}P_0 + \rho\end{aligned}$$

→ **The Average time a customer spends in the system:**

$$\overline{Q} = \frac{\overline{N}_Q}{\lambda} \qquad \Longrightarrow \qquad \overline{T} = \frac{\rho^{c+1}}{\lambda c!}\frac{c}{(c-\rho)^2}P_0 + \frac{1}{\mu}$$

with $\overline{S} = \dfrac{1}{\mu}$

2.6.2 The model $M/M/\infty$

If the number c of servers tends to infinity, then we find that:

$$P_0 \longrightarrow e^{\frac{-\lambda}{\mu}} \quad and \quad P_k \longrightarrow \frac{(\frac{\lambda}{\mu})^k}{k!}e^{\frac{-\lambda}{\mu}}$$

with $\rho = \dfrac{\lambda}{\mu}$ That is, in the case of the $M/M/\infty$ model, we have:

$$P_k = \mathbb{P}[Y = k], \qquad \text{with Y} \sim \mathcal{P}\left(\frac{\lambda}{\mu}\right)$$

→ **The average number of customers in line:**

$$\overline{N}_Q = \overline{Q} = 0 \tag{2.15}$$

because there is no waiting time.

→ **The average number of customers in line:** we obtain

$$\overline{N} = \mathbb{E}(Y) = \frac{\lambda}{\mu} = \overline{N}_S \tag{2.16}$$

→ **The Average residence time:** using Little's law:

$$\overline{T} = \overline{S} = \frac{1}{\mu} \tag{2.17}$$

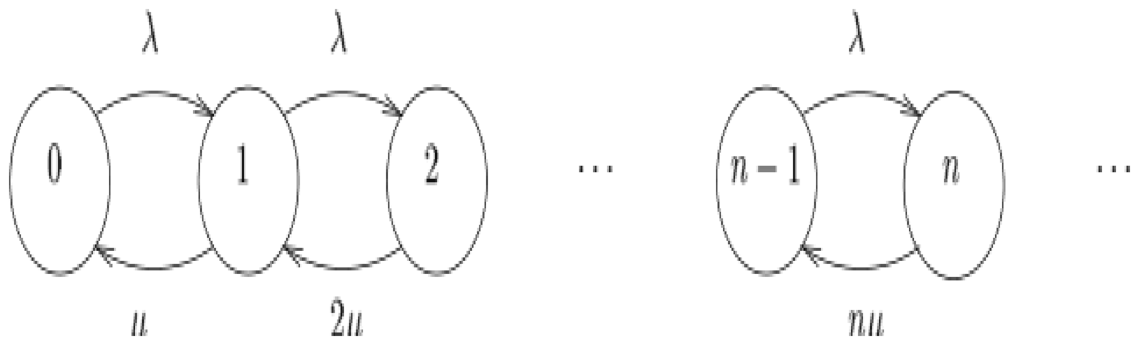

Figure 2.8: State-transition diagram for the model $M/M/\infty$

Chapter 3

A Multi-Server Markovian Feedback Queue with Balking Reneging and Retention of Reneged Customers

Queuing models have effectively been used in the design and analysis of telecommunication systems, traffic systems, service systems and many more. A number of extensions in the basic queuing models have been made and the concepts like vacations queuing, correlated queuing, retrial queuing, queuing with impatience, and catastrophic queuing have come up. Of these, queuing with customer impatience has special significance for the business world as it has a very negative effect on the revenue generation of a firm. A customer is said to be impatient if he tends to join the queue only when a short wait is expected and tends to remain in the line if his wait has been sufficiently small. Impatience generally takes three forms. The first is balking, deciding not to join the queue at all up on arrival, the second is reneging, the reluctance to remain in the waiting line after joining and waiting, and the third is jockeying between lines when each of a number of parallel service,[11]. An extensive review on queuing systems with impatient customers is presented by Wang [24]. They survey various queuing systems according to various dimensions like customer impatience behaviors, solution methods of queuing models with impatient customers, and associate optimization aspects.

Customers are very hard pressed for time. Such constraints on time induce impatience on customers behaviour whenever they are required to wait in a service facility. In queuing parlance, such impatience may find reflection through the concepts of balking and reneging. Even though one can find queuing models of various types analysed in queuing literature, it is not often that reneging and balking have been analysed. Even if these have been dealt with, closed form expressions are not always available. This research is an attempt in this direction [6]

3.1 Model Assumptions

We study single channel multi-server queueing systems. The Poisson process with mean arrival rate λ governs the arrivals. The inter-arrival periods have a parameter λ-dependent, identical, exponential distribution. There exist c servers, and each one has an independent, identical, exponentially dispersed service time with parameter μ. When N is the system's capacity and n is the number of customers, the mean service rate for $n < c$ is $n\mu$, and the service rate for $c \leq n \leq N$ is $c\mu$. Customer can either join at the end of the line with probability p_1 or depart the system with probability q_1, where $p_1 + q_1 = 1$. Both newly arrived and feedback-filled customers are served in the sequence that they join the tail of the initial line. We make no differentiation between feedback arrival and regular arrival. Assumed is a finite (say, N) system capacity. Put otherwise, the system can support up to N customers. When there are more customers than servers, or when $n > c$, a queue forms. Upon entering the queue, every customer will have to wait a predetermined amount of time for his service to start. Should it not start by then, he will be reneged, with a probability of p_2 leaving the queue without receiving service and a probability of $q_2(= 1 - p_2)$ staying in the wait for his service. The exponential distribution with parameter ξ is followed by the reneging times. With a given amount of balking probability, the arriving client joins the system. N is the measure of the customer's willingness to join the line. It is assumed that an approaching consumer balks with probability (n/N), where n is the number of customers in the system and consequently joins the system with probability $1 - (n/N)$.[20]

Queue Behavior:[18]

The actions of customers while in a queue waiting for service to begin, for example:

→ **Balk:** leave when they see that the line is too long;

→ **Renege:** leave after being in the line when its moving too slowly;

→ **Jockey:** Move from one line to a shorter line.

3.2 Mathematical Formulation of the Model

We introduce a differential-difference equation based mathematical model in this part. We obtain these equations by application of the general birth-death arguments. Let $P_n(t)$, with $0 \leq n \leq N$, be the probability that the system has n customers at time t. In an incredibly short time $(t, t + \delta)$,

$$P_n(t + \delta t) = \text{Prob}\{\text{there are n customers in the system at time } (t + \delta t)\}$$

When $c + 1 \leq n \leq N - 1$,the equation derived as follows Here:

$$P_n(t+\delta t) = P_n(t)\left[\left\{1-\left(1-\frac{n}{N}\right)\lambda\delta t\right\}(1-c\mu q_1\delta t)\right]+P_n(t)\left[\left\{\left(1-\frac{n}{N}\right)\lambda\delta t\right\}(c\mu q_1\delta t)\right]+$$

$$\mathrm{P}_{n-1}(t)\left[\left\{\left(1-\frac{n-1}{N}\right)\lambda\delta t\right\}(1-c\mu q_1\delta t)\right]+P_{n+1}\left[\left\{1-\left(1-\frac{n+1}{N}\right)\lambda\delta t\right\}(c\mu q_1\delta t)\right]+$$

$$\mathrm{P}_n(t)\left[\left\{1-\left(1-\frac{n}{N}\right)\lambda\delta t\right\}(1-c\mu q_1\delta t)(n-c)\xi q_2\delta t\right]+$$

$$\mathrm{P}_{n+1}(t)\left[\left\{1-\left(1-\frac{n+1}{N}\right)\lambda\delta t\right\}(1-c\mu q_1\delta t)\{(n-c)\xi p_2\delta t\}\right]$$

The differential-difference equation 3.3 may be obtained by finding the difference $P_n(t+\delta t)-P_n(t)$, dividing both sides by δt, and taking the limit $\delta t \longrightarrow 0$. So does $o(\delta t)$ approach zero. The remaining formulas 3.1, 3.2, and 3.4 are

similarly derived as well. As a result, the model's differential-difference equations are:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu q_1 P_1(t) \tag{3.1}$$

$$\begin{aligned}\frac{dP_n(t)}{dt} = &-\left[\left(1-\frac{n}{N}\right)\lambda + n\mu q_1\right]P_n(t) + (n+1)\mu q_1 P_{n+1}(t)\\ &+\left(1-\frac{n-1}{N}\right)\lambda P_{n-1}(t),\ \ 1 \le n \le c\end{aligned} \tag{3.2}$$

$$\begin{aligned}\frac{dP_n(t)}{dt} = &-\left[\left(1-\frac{n}{N}\right)\lambda + \mu q_1 + (n-c)\xi p_2\right]P_n(t) + [c\mu q_1 + \{(n+1)-c\}\xi p_2]\\ &P_{n+1}(t) + \left(1-\frac{n-1}{N}\right)\lambda P_{n-1}(t),\ c+1 \le n \le N-1\end{aligned} \tag{3.3}$$

$$\frac{dP_N(t)}{dt} = \left(1-\frac{N-1}{N}\right)\lambda P_{N-1}(t) - [c\mu q_1 + (N-c)\xi p_2]\,P_N(t) \tag{3.4}$$

3.3 Steady-State Solution of the Model

We derive the model's steady-state solution repeatedly in this section. Since $\lim_{t\to\infty} P_n(t) = P_n$ in steady-state, the equations corresponding to equations 3.1 through 3.4 in steady-state are as follows:

$$0 = -\lambda P_0 + \mu q_1 P_1 \tag{3.5}$$

$$0 = -\left[\left(1 - \frac{n}{N}\right)\lambda + n\mu q_1\right] P_n + (n+1)\mu q_1 P_{n+1} + \left(1 - \frac{n-1}{N}\right)\lambda P_{n-1},\ 1 \le n \le c \tag{3.6}$$

$$\begin{aligned} 0 = -\left[\left(1 - \frac{n}{N}\right)\lambda + c\mu q_1 + (n-c)\xi p_2\right] P_n + \left[c\mu q_1 + \{(n+1) - c\}\xi p_2\right] P_{n+1} \\ + \left(1 - \frac{n-1}{N}\right)\lambda P_{n-1},\ c \le n \le N-1 \end{aligned} \tag{3.7}$$

$$0 = \left(1 - \frac{N-1}{N}\right)\lambda P_{N-1} - \left[c\mu q_1 + (N-c)\xi p_2\right] P_N,\ n = N \tag{3.8}$$

From equation 3.5, we have

$$\mu q_1 P_1 = \lambda P_0$$

$$P_1 = \frac{\lambda}{\mu q_1} P_0 \tag{3.9}$$

Substitute n = 1 in 3.6, we get:

$$2\mu q_1 P_2 = \left[\left(1 - \frac{1}{N}\right)\lambda + \mu q_1\right] P_1 - \lambda P_0$$

$$P_2 = \frac{N-1}{N}\frac{\lambda^2}{2!(\mu q_1)^2} P_0 \quad \{3.5\} \tag{3.10}$$

For n = 2 in 3.6, we have:

$$P_3 = \prod_{k=1}^{3} \frac{N-(k-1)}{N}\frac{\lambda^2}{3!(\mu q_1)^2} P_0$$

Proceeding in the same way, we get:

$$P_n = \prod_{k=1}^{n} \frac{N-(k-1)}{N}\frac{\lambda^n}{n!(\mu q_1)^n} P_0, \quad 1 \le n \le c \tag{3.11}$$

Now for n = c, 3.7 become:

$$(c\mu q_1 + \xi p_2)P_{c+1} = -\left[\left(1-\frac{c}{N}\right)\lambda + c\mu q_1\right]P_c - \left(1-\frac{c-1}{N}\right)\lambda P_{c-1}$$

$$c\mu q_1 P_c = \left(1-\frac{c-1}{N}\right)\lambda P_{c-1}$$

$$P_{c+1} = \prod_{k=c}^{c+1} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} P_c \tag{3.12}$$

Similarly, from 3.7 and 3.8, for $c+1 \leq n \leq N$, we get:

$$P_n = \prod_{k=c}^{n} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} P_c \tag{3.13}$$

Thus, the steady-state solution of the model is:

$$P_n = \begin{cases} \displaystyle\prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda^n}{n!(\mu q_1)^n} P_0, & 1 \leq n \leq c \quad (3.14) \\ \displaystyle\prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} P_0, & c \leq n \leq N \quad (3.15) \end{cases}$$

Using the normalization condition, $\sum_{n=0}^{N} P_n = 1$, we get

$$P_0 = \frac{1}{1 + Q_1 + Q_2} \tag{3.16}$$

So that all of Q_1 and Q_2 equals:

$$Q_1 = \sum_{n=1}^{c-1} \prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda^n}{n!(\mu q_1)^n} \tag{3.17}$$

and

$$Q_2 = \sum_{n=c}^{N} \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \tag{3.18}$$

As a result, the system size's steady-state probabilities are determined directly.

3.4 Performance Measures

In this section some various performance metrics are obtioned esealy.

(1) The Expected System Size (L_s):

$$L_s = \sum_{n=1}^{c-1} n \left\{ \prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda^n}{n!(\mu q_1)^n} \right\} P_0 +$$

$$\sum_{n=c}^{N} n \left\{ \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \right\} P_0$$

(2) The Expected Number of Customers Served (E(C.S)):

$$E(C.S) = \sum_{n=1}^{c} n\mu q_1 \left\{ \prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda^n}{n!(\mu q_1)^n} \right\} P_0 +$$

$$\sum_{n=c+1}^{N} c\mu q_1 \left\{ \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \right\} P_0$$

(3) Rate of Abandonment (R_{aband}):

$$R_{aband} = \lambda - \sum_{n=1}^{c} n\mu q_1 \left\{ \prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda^n}{n!(\mu q_1)^n} \right\} P_0 -$$

$$\sum_{n=c+1}^{N} c\mu q_1 \left\{ \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \right\} P_0$$

(4) Average Reneging Rate (R_r):

$$\mathrm{R}_r = \sum_{n=c}^{N} (n-c)\xi p_2 \left\{ \prod_{k=c}^{N} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \right\} P_0$$

(5) Average Balking Rate (R_b):

$$R_b = \sum_{n=1}^{c-1} \frac{n\lambda}{N} \left\{ \prod_{k=1}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{n!(\mu q_1)^n} \right\} P_0 +$$

$$\sum_{n=c}^{N} \frac{n\lambda}{N} \left\{ \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} \right\} P_0$$

(6) Average Retention Rate (R_R):

$$R_R = \sum_{n=1}^{N} (n-c)\xi q_2 \left\{ \prod_{k=c}^{n} \frac{N-(k-1)}{N} \frac{\lambda}{c\mu q_1 + (k-c)\xi p_2} \quad \left\{ \prod_{r=1}^{c-1} \frac{N-(r-1)}{N} \frac{\lambda^r}{r!(\mu q_1)^r} \right\} P_0$$

where P_0 is computed in 3.16.

3.5 Numerical Discussion

We wrote a program in R to simulate the results we obtained from this study. We also explain the simulations that supported our work.

3.5.1 The influence of service rate μ on the different performance parameters

In the tables 3.1 below, we have presented the numerical results of all measures of performance. Numerical results are obtained for various service and arrival rates. For the other parameters, we assume the following constant values:
$\lambda = 6$; $\xi = 5$; $p_1 = 0.5$; $q_2 = 0.5$; c = 5, and N = 10.

As can be seen from Table 3.1, The expected number of customers in the system $(E(C.S))$ and the probability of no customers in the system (P_0) increase with the increase in the service rate, while all the other performance measures decrease with the increase in the rate of service μ.

Table 3.1: Effect of μ on the various measures of performance

μ	P_0	L_q	L_s	$E(C.S)$	R_r	R_b
5	0.081445	0.075238	3.085926	4.152427	0.188094	1.851556
6	0.151300	0.015592	1.929184	4.850116	0.038980	1.157510
7	0.203469	0.003279	1.509939	5.097083	0.008198	0.905964
8	0.247240	0.000798	1.306188	5.217382	0.001996	0.783713
9	0.286552	0.000224	1.170238	5.298264	0.000560	0.702143
10	0.322513	0.000071	1.064963	5.361182	0.000177	0.638978
11	0.355545	0.000025	0.978363	5.413049	0.000062	0.587018
12	0.385917	0.000009	0.905089	5.456977	0.000024	0.543053
13	0.413867	0.000004	0.842058	5.494779	0.000010	0.505235
14	0.439618	0.000002	0.787201	5.527686	0.000004	0.472321

Remark. For the other parameters, in figure 3.1, 3.2,3.3, 3.4, 3.5 we assume the following constant values:
$\xi = 0.5$; $p_1 = 0.5$; $q_2 = 0.3$; c = 5, N = 10.

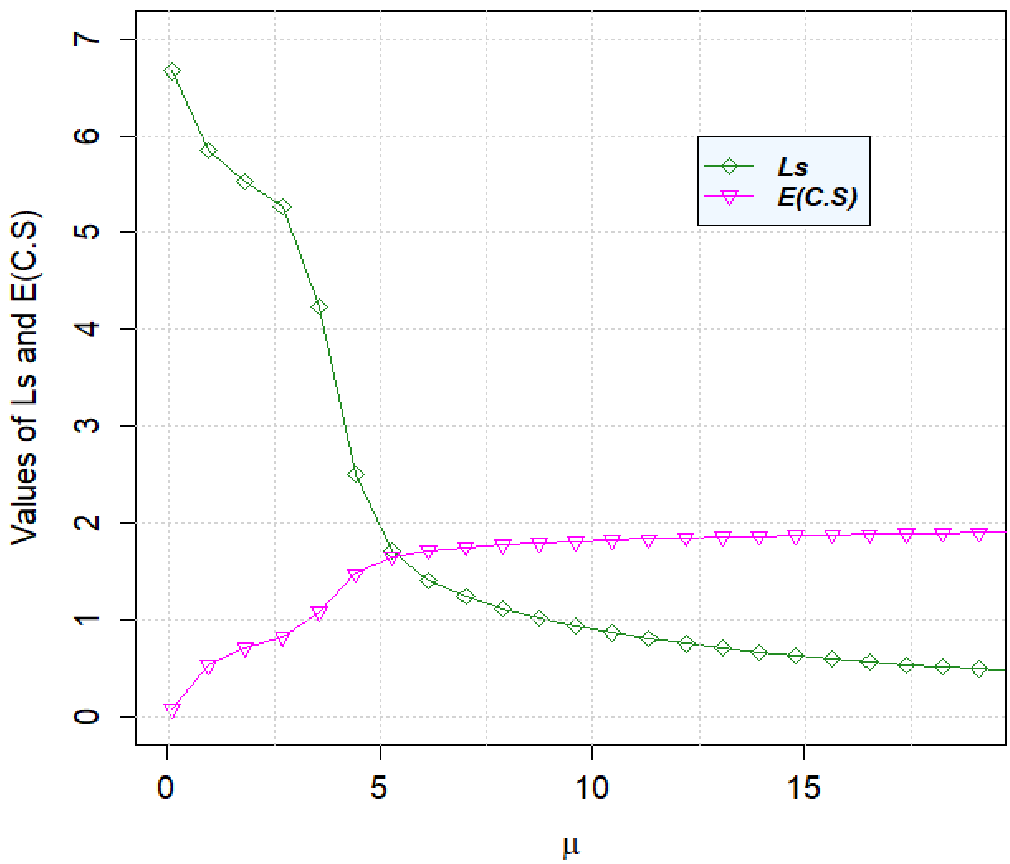

Figure 3.1: Effect of μ on the expected system size(L_s) and the expected number of customers served, E(C.S)

In the figure below, we have presented the graphical results of the average balking rate R_b with the rate of service μ.

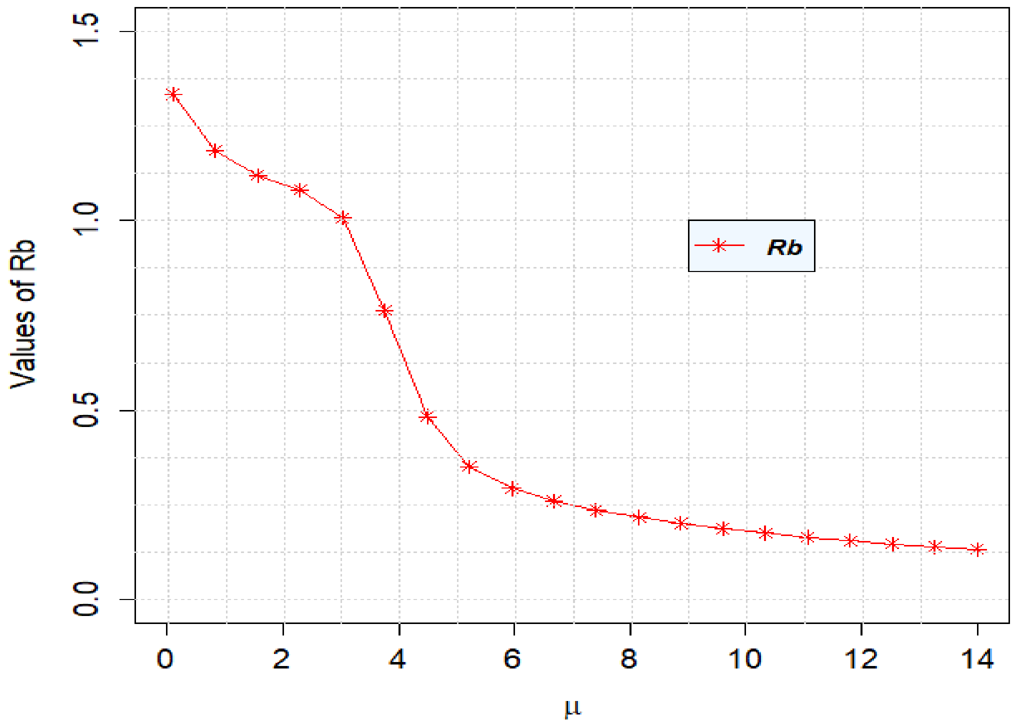

Figure 3.2: Effect of on the average rate of balking, R_b

The effect of the rate of service on the expected number of customers served and the length of the system is shown graphically in Figure 3.1. The average rate of balking is also shown to be a decreasing function of μ in Figure 3.2.

3.5.2 The influence of customer arrival rates λ on the different performance parameters

From Table 3.2, we see that with the increase in the rate of arrival λ, the size of the system increases rapidly. As a result, both the reneging rate R_r and the balking rate R_b increased due to the long queue. On the other hand, the expected number of customers served does not make a significant increase since it is mainly dependent on the service rate than the arrival rate of the system. Here also, we assume that:
$\mu = 5$; $\xi = 5$; c = 5; $p_1 = 0.5$; $q_2 = 0.5$; and N = 10.

Table 3.2: Effect of λ on the various measures of performance

λ	L_q	L_s	W_s	W_q	$E(C.S)$	R_r	R_b
10	0.397532	5.838082	0.583808	0.039753	4.148322	0.993831	5.838082
11	0.445605	5.920701	0.538246	0.040510	4.463747	1.114012	6.512771
12	0.491827	5.980571	0.498381	0.040986	4.788910	1.229567	7.176685
13	0.537445	6.032010	0.464001	0.041342	5.112161	1.343612	7.841613
14	0.582936	6.080159	0.434297	0.041638	5.428956	1.457340	8.512222
15	0.628479	6.126989	0.408466	0.041899	5.737445	1.571197	9.190483
16	0.674129	6.173295	0.385831	0.042133	6.036870	1.685322	9.877272
17	0.719889	6.219405	0.365847	0.042346	6.326950	1.799723	10.572989
18	0.765739	6.265449	0.348080	0.042541	6.607627	1.914347	11.277807
19	0.811646	6.311467	0.332182	0.042718	6.878954	2.029114	11.991786

The result that we extract from the table 3.2 is that the more customers enter the queue, the more it increases the number of customers in the queue L_q and in the system L_s, The customer's waiting time in the system also increases W_s because the average rate of balking R_b and reneging R_r increases.

We notice in figures 3.3 and 3.4 that the greater the probability that a customer can either join at the end of the line with probability p_1, the greater the mean size of the system L_s and the queue L_q. Because they increase in size, in addition, the size of the balking R_b decreases.

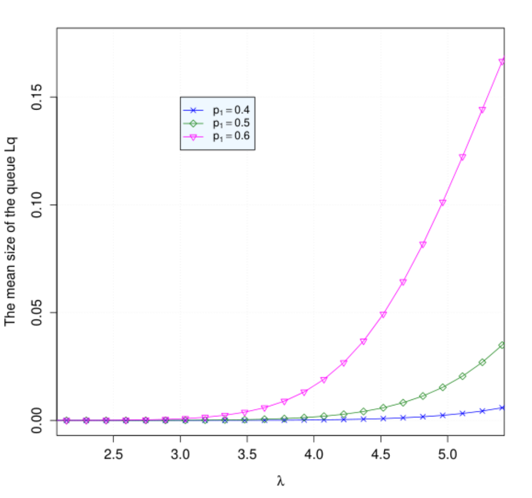

Figure 3.3: The mean size of the queue L_q vs λ and p_1

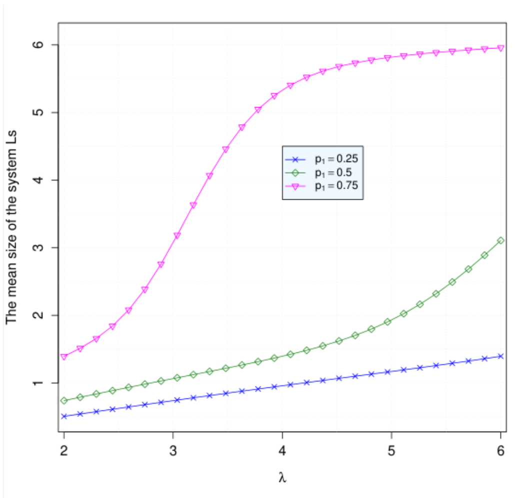

Figure 3.4: The mean size of the system L_s vs λ and p_1

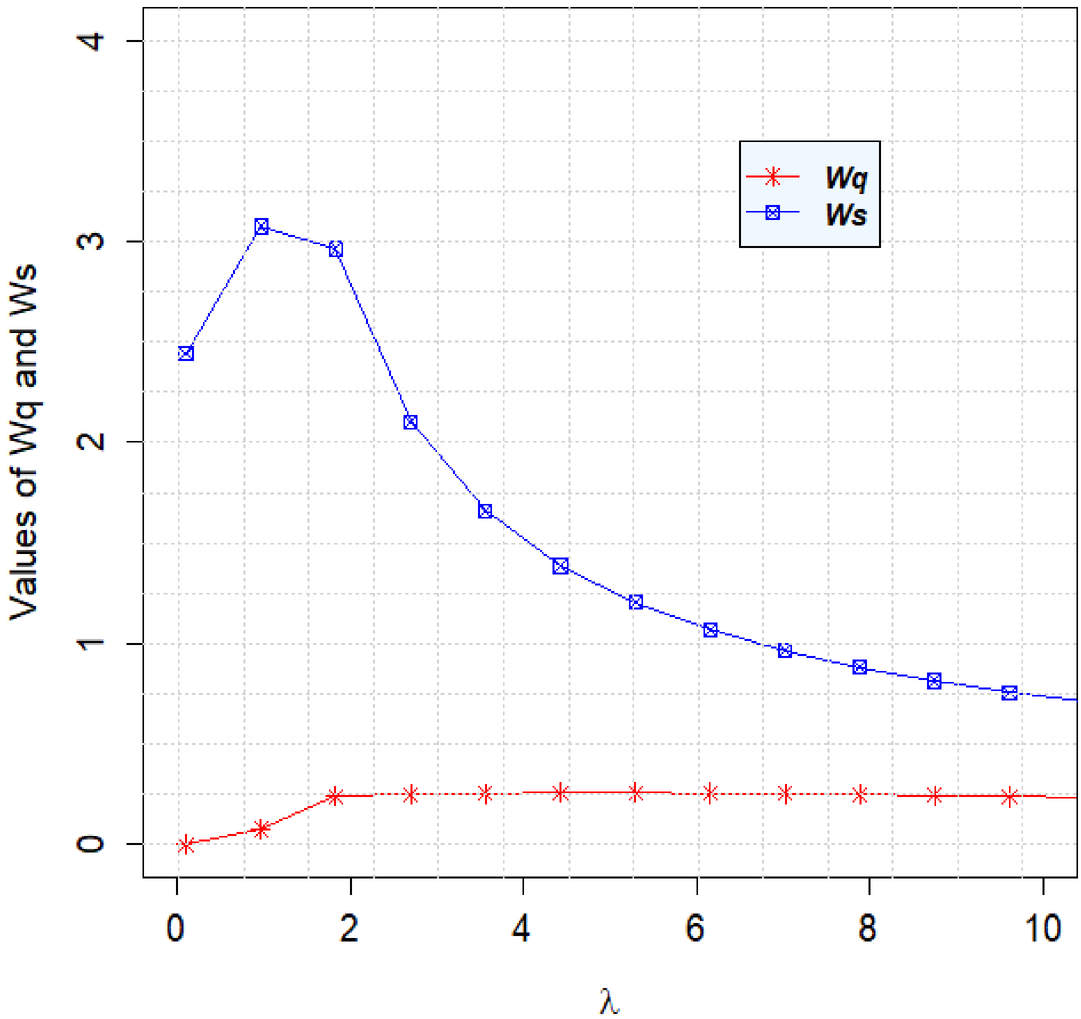

Figure 3.5: The average waiting time of a customer in the system W_s and in the queue W_q vs λ

We see in Figures 3.5, that as the number of customers that can join at the end of the queue increases, the waiting time in the system W_s decreases because both R_b and R_r increase. W_q increases because the size of the queue gets smaller.

Conclusion

We study a Markovian feedback queuing model with a finite capacity, multiple servers, balking, reneging, and keeping users who have reneged. We get the model's steady-state answer and also come up with some quality of service measures. The results of the model could be used to model different service and production processes that involve comments and customers who are impatient.

Bibliography

[1] A. Gomez-Corral,and M.F. Ramalhoto. On the waiting time distribution and the busy period of a retrial queue with constant retrial rate. Stochastic Modelling and Applications, 3, 37-47, (2000).

[2] Ana Aleksieva, University of Karlsruhe (TH) Department of Mathematics, Stochastic Methods in Industry, 2007/08.

[3] Anisimov, Zakusilo, O. K., and Donchenko, V. S. 1987. Elements of Queueing Theory and Asymptotic System Analysis. Vishcha Shkola, Kiev (in Russian).

[4] Athanasios Christou Micheas, Theory of Stochastic Objects, November 17, 2017.

[5] Chaudhry, M. L, and Templeton, J. G. C. A First Course in Bulk Queues. John Wiley and Sons, New York. (1983).

[6] Choudhury, A. and Medhi P. (2011), A simple analysis of customer impatience in multi-server queues, Int. J. Applied Management Science, Vol. 3, No. 3, pp.294ˇ2013315.

[7] Processus stochastiques et modélisation. (2012-2013).

[8] Elhaddad Meriem, Approximate analysis of M/M/c retrial queue, These De Doctorat, 2020.

[9] Fezza Kheira, Systˇ00e8me de files dˇ2019attente avec dˇ00e9robade, 2017.

[10] Gelenbe, E., and Pujolle, G. 1987. Introduction to Queuing Networks. John Wiley and Sons, Chichester, England (French Original : Introduction aux réseaux de files d'âttente, Edition Eyrolles ; Paris).

[11] Gross and Harris, C. M. (1985) Fundamentals of Queuing theory, Second Edition, Wiley, Hoboken, NJ.

[12] Howard M.Taylor, An Introduction to Stochastic Modeling Third Edition, Department of Mathematics Stanford University Stanford, California, 1998.

[13] Jénos Sztrik, University of Debrecen, Faculty of Informatics, Basic Queueing Theory, 2012.

[14] John Wiley Sons Ltd, Queueing Modelling Fundamentals (With Applications in Communication Networks), Second Edition, 2008.

[15] K.B, GK, Techniques de modŏ0e9lisation : Mŏ0e9thodes analytiques.

[16] Khintchine, A. Y 1969. Mathematical Methods in the Theory of Queueing. Second edition, Hafner Publishing Company, New York (First edition : Griffin, London, 1960; Russian original : 1955).

[17] Mario Lefebvre, Applied Stochastic Processes, New York, NY 10013, U.S.A, 2007.

[18] Mesut, Gŏ0fcnŏ0e9s, Queueing Models, Chapter 8, Freie Universitat Berlin.

[19] P. Vijaya Laxmi, T. Wondewosen Kassahun, Department of Applied Mathematics, Andhra University, Visakhapatnam ž013 530003, India, 2017.

[20] Rakesh Sumeet Kumar, A Multi-Server Markovian Feedback Queue with Balking Reneging and Retention of Reneged Customers, AMO-Advanced Modeling and Optimization, Volume 16, Number 2, 2014.

[21] Ramon van Handel, Probability and Random Processes, ORF 309/MAT 380 Lecture Notes Princeton University, February 22, 2016.

[22] Rick Durrett, Essentials of Stochastic Processes 2nd Edition, December, 2011.

[23] Sidney Resnick, Adventures in Stochastic Processes, School of Operations Research and Industrial Engineering, 1992.

[24] Wang, and Jiang, Queuing System with Impatient Customers: A Review, International Conference on Service Operations and Logistics and Informatics, 15-17 July, 2010, Shandong, pp. 82-87.

[25] Zakhar Kabluchko, Stochastic Processes (Stochastik II), University of Ulm Institute of Stochastics, April 29, 2014.