University of Adelaide
Department of Applied Mathematics
PhD Thesis

An Application of Martingales to Queueing Theory

Matthew Roughan B.Sc.(Ma)Hons.
19th July, 1995

Dedication

This thesis is dedicated to Sylvia Daina Luks
who has tolerated me these many years.

Acknowledgement

I would like to thank my supervisor Charles Pearce for his valuable suggestions and advice. I would further like to thank the numerous people who have at some stage in this thesis provided help or encouragement such as the staff and students of the Teletraffic Research Centre at the University of Adelaide and the staff of the Applied Mathematics Department at this same University. I would also like to thank Bong Dai Choi and Nick Duffield who on their brief visits supplied motivation for continuing.

Specifically I would like to thank Andrew Coyle for proof reading this thesis.

I would like also to acknowledge the help and support of my family.

This thesis was funded by a Australian Postgraduate Priority Research Award and a supplementary scholarship from the Teletraffic Research Centre at Adelaide University.

Statutory Declaration

I, Matthew Roughan, state that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and for photocopying.

SIGNED: ..... **DATE:** .....

Contents

List of Figures

Summary

Systems such as the M/G/1 queue are of great interest in queueing theory. Techniques such as M.Neuts's block matrix methodology have traditionally been used on the more complicated generalisations of this type of queue. In this thesis I develop an alternative method which uses martingale theory and some renewal theory to find solutions for a class of M/G/1 type queues.

The theory, originally applied by Baccelli and Makowski to simple queueing problems, derives its key result from Doob's Optional Sampling Theorem. To make use of this result some renewal theoretic arguments are necessary. This allows one to find the probability generating function for the equilibrium distribution of customers in the system.

Chapter 2 develops the renewal theoretic concepts necessary for the later parts of the thesis. This involves using the key renewal theorem on a modified type of Markov renewal process to obtain results pertaining to forward recurrence times.

Chapter 3 contains the martingale theory and the main results. The type of processes that can be dealt with are described in detail. Briefly these consist of processes where the busy period is broken into a series of phases. The transitions between phases can be controlled in a number of ways as long as they obey certain rules. Some examples are: phases ending when there are more than a certain number of customers in the system or when the busy period has continued for a certain number of services. The behaviour of the server can be different in each phase. For instance, the service-time distribution or the service discipline may change between phases. The main result uses Doob's Optional Sampling Theorem and so we must establish a number of conditions on the martingale used. We establish a simple criterion for these conditions to hold. Finally in this chapter we examine the simplest case, the M/G/1 queue.

The following chapters contain a number of examples. Standard probabilistic arguments are used to obtain the necessary conditions and results to use the theorems of Chapter 3. The examples considered include cases with two, three, four and an infinite number of phases. The theoretical results are supported by a number of simulations in the latter case.

Finally we have some suggestions for possible future work and the conclusion.

Chapter 1

Introduction

Aim

The aim of this thesis is to apply a martingale technique to queueing theory. Martingales have been used successfully in financial modelling and elsewhere and there is a well developed theory built around them.

One of probability's great achievements is in its use in the study of queues and so it is surprising that such a powerful area of probability such as martingale theory has only been slowly taken up in queueing theory.

In order to redress this we have demonstrated how a technique suggested by F. Baccelli and A.M. Makowski can be extended to solve some queueing problems of interest.

The Technique

The amount published on the use of martingales in queueing theory, compared to the total amount of literature on queueing theory, is minimal. Where martingales have been used they tend to be deployed on a particular part of a larger problem and not as a general technique for solving problems. Some examples where this is not true are Rosenkrantz (1983), Kella and Witt (1992), Ferrandiz (1993), Baccelli (1986) and Baccelli and Makowski (1985,1986,1989,1991). In these papers the approach is to solve a problem or problems using martingales as the fundamental means. We are particularly interested in the papers of Baccelli and Makowski.

Baccelli and Makowski's technique was first used, in the literature, to demonstrate stability conditions (1985,1986) and later to provide probability generating functions for the equilibrium number of customers in simple queueing systems, (1989,1991). The only problem with these was that the systems investigated were the well-known M/G/1 queue (1985,1989) and the single-server queue with Markov Modulated Poisson Process input (1986,1991). This was criticised in Mathematical Reviews (92k:60202) where the reviewer stated;

> A number of papers including this one *(Baccelli and Makowski, 1991)* have appeared in recent years wherein martingale methods have been used to derive reasonably well-known results derived initially by other, less abstract methods. It is not clear to the reviewer that our understanding of the related queueing systems has been advanced by these very technical papers.

The only other direct use of this technique, known to the author, is the master's thesis of Park (1990). This also examines a well-known system, the $M^X/G/1$ queue.

This thesis addresses this criticism. We use the four papers of Baccelli and Makowski as a basis for all of the theory herein. The idea is extended to cover a major group of useful and interesting problems.

The technique involves using Doob's Optional Stopping Theorem on the discrete-time process, embedded at departure epochs. This theorem allows one to relate the behaviour of the queue at arbitrary time points to the behaviour of a renewal process embedded at specific epochs of the queue's history. This is the most important part of all of the theory to follow, a large part of the remainder merely being support for this result. In Baccelli and Makowski's papers the ends of the busy periods form the renewal epochs of interest. In this extension other time points are allowed to have significance in the queue, and these are then included in the relationship. This considerably complicates the renewal process of interest and we require a new result because of this complication.

One final note to make about this technique is that it is underutilised in this thesis. Baccelli and Makowski derive stability criterion and transient results for the systems they investigate. I have chosen only to look at the equilibrium probability generating function of the number of customers in the systems. The stability criterion is omitted because in the general case examined here the stability criterion depend on the nature of the specific problem. The transient results are not provided in order to keep this thesis as concise as possible, as the transient results for many of the problems investigated are extremely complicated.

The problems

The class of systems to which this method is applicable is a generalisation of the M/G/1 queue. The single server is retained, as is the Poisson input, and what we call the *server's behaviour* is modified. The queue, during the course of a busy period, progresses through a series of *phases*. In each phase the server's behaviour may be different. For instance the service-time distribution, the service discipline or even the probability of blocking an arrival may vary between phases. (Note that our concept of a phase is different from the phase-type distribution of Neuts (1989).)

The phases must obey certain rules in order that the theory may be applied. In some cases, processes whose phase structure does not obey these rules may be remodelled so that the new phase structure does obey the rules. Through remodelling such as this a great number of problems are included in this class of queues.

Motivation

The M/G/1 queue and variants thereof form one of the largest fields of study in queueing theory. Many computing, communication and manufacturing systems can be modelled by M/G/1 models. There are too many articles on this subject to go into all of them here. The following brief list indicates some of the very recent papers in this area: Borst *et al* (1993), Ferrandiz (1993), Takine *et al* (1993), Schormans *et al* (1993) and Yashkov (1993). A good initial starting point for investigating all single-server queues is Cohen (1969).

Many interesting systems fit into the class of systems investigated. Later in this thesis a number of simple examples are presented. These, while still perhaps too simple

to be used directly, point quite clearly to some of the types of problems that can be solved. An example is a system with state dependent arrival and service processes as described in Courtois and Georges (1971). Numerous papers which deal with problems of this nature could be listed. A small sample of recent papers in this vein:

J. Dshalalow (1989),
W. Gong, A. Yan and C.G. Cassandras (1992),
O.C. Ibe and K.S. Trivedi (1990),
M. Kijima and N. Makimoto (1992a,b),
R.O. LaMaire (1992),
J.A. Morrison (1990),
J.A. Schormans and J.M. Pitts and E.M. Scharf (1993),
T. Takine, H. Takagi and T. Hesegawa (1993)
and H. Takagi (1992).

Furthermore there is an entire theory developed around precisely such problems. This is the block-matrix methodology of Neuts *et al* described in Neuts (1989). Given that this theory exists why have I presented an alternative? For a start many problems easily soluble by the martingale technique may be solvable in the Neutsian scheme, but not easily, and vice versa. Furthermore the martingale technique can provide elegant analytical results and in addition it can in some situations provide, as a byproduct, results in addition to those initially desired.

Further motivation is given by the examples presented in this thesis.

Outline

This thesis consists of two major parts. The general theory is covered in the first two chapters. Chapter 2 covers renewal theory. Because of the complex nature of the renewal processes which I use in this theory they need to be described in detail. Also a renewal result is given which to the author's knowledge is not to be found in the literature.

Chapter 3 gives the technical definition of the type of processes investigated and how these processes are modelled. The martingale (and related probabilistic elements) that are used are given, as is a set of stopping times which are crucial to the results. It is vital that these stopping times be regular with respect to the martingale and in condition $(*)$ (page 33) a condition for this is provided.

The martingale and stopping times are then used in Doob's Optional Stopping Theorem to provide a set of results which when used in conjunction with those of Chapter 2 provide the probability generating function of interest. The penultimate section of Chapter 3 deals with all of the possible generalisations of the basic model to which this technique can be applied, along with some of the potential pitfalls. Finally the technique is applied to the simple M/G/1 queue.

The second major part of this thesis consists of a number of examples. Chapter 4 contains the first set of examples, and for this reason is the most detailed. These are all two-phase examples, that is, M/G/1 queues where the busy period can be broken into two parts, with different server behaviour in each. We describe this by saying that the first phase ends when some threshold is crossed. Three major types of threshold are considered. The first is when there are more than a certain number of customers in the system, the second when a geometrically-distributed number of customers have been served in the busy period and the third when a fixed number of customers have been

served in the busy period.

The individual solutions for each of these cases requires a reasonable amount of work, using standard probabilistic techniques, before a result can be obtained. This involves showing that condition $(*)$ is satisfied and investigating the behaviour of the queue at the thresholds. Remarkably, all three solutions are in the same form. The form of the probability generating function for the number of customers in the systems is closely related to the form of that of the M/G/1 queue with a correction term that depends on the difference in behaviour in the two phases. A number of subsidiary results such as the equilibrium probability of each phase are calculated using Little's law.

Chapter 5 extends one of the examples of Chapter 4 to the case with three phases by including a second threshold. This chapter also introduces a new type of threshold.

Chapter 6 describes and solves a four phase model which can be used to model a M/G/1 queue which can break down. While this is an unsophisticated breakdown/repair model it demonstrates how a more sophisticated model could be formed.

Finally, the example in Chapter 7 considers what happens in a case with an infinite number of phases. This can happen if two service regimes may alternate an infinite number of times during the busy period. It is difficult in this case to prove condition $(*)$ and so a number of numerical examples are provided to support the result. This type of model is then used to investigate the M/G/1 queue with a finite waiting room.

The major part of the text is followed by Chapter 8, the conclusion, which discusses further work that could be conducted in this field and block-matrix methods and how they could be applied to some of the problems herein.

A number of appendices are included in this thesis. These cover some of the basic theory used in queueing theory and martingale analysis. A great deal of this will be known to the experienced reader but we include it for two reasons. Firstly it is desirable for this thesis to be as self contained as possible. To this end the major results which we call upon are included. Secondly, some of the results appear in several different forms in the literature. For instance, while the major reference on martingales is Neveu (1975) and we draw Doob's Optional Stopping Theorem from this reference, Neveu does not refer to this theorem as Doob's Optional Stopping Theorem. This name appears in relation to this theorem or related theorems in several places such as Williams (1991).

Chapter 2

Renewal Theory

2.1 Introduction

In this chapter we consider renewal processes. The results of this chapter provide the probability generating function for the forward recurrence times of a type of renewal process. By itself the result is of moderate interest. The results herein are more a means to an end than an end unto themselves. However it is important that the nature of the renewal process be understood before the more interesting parts of this thesis commence.

This chapter begins with a brief description of renewal processes, building up to Markov renewal processes. This basic work on renewal processes and Markov renewal processes is well known. Several references such as Wolff (1989), Cox (1962) and Pyke (1961a,b) cover this. We have included it in order to build up to the type of process of interest. Definitions such as that of forward renewal times are much easier in the simple renewal process. The concept is then extended to more complex systems. Further the arguments used to prove results may be clearer in the simple case. Thus it is to be hoped that when extended the arguments remain clear.

Two points to note are that we shall be concerned with non-delayed processes and also that we are interested in lattice processes. The reasons for this will become evident when the results are put to use.

Once the survey of Markov renewal processes is done we look at what we call generalised Markov renewal processes. By this we do not mean the same thing as Pyke. We mean a Markov renewal process in which not all of the states are renewal states. This concept is explained in Section 2.3.2. This describes a process more general than that needed for our results. Part of the reason for this is simply to provide a framework for the case of interest. However the main reason is that it is to be hoped that the results might be extended to the whole class of generalised Markov renewal processes. This would enable simplification of the technique for examining some processes.

The important part of this chapter is that which describes the multi-phase Markov renewal process. This is the motivating case and the process for which original results are produced. This can be thought of as a process in which the times between renewals have been broken into a number of phases. The times spent in each of the phases are not independent. Thus the recourse to the generalised Markov renewal process. Each phase is considered to be a separate state in the Markov renewal process with only one of the states being a renewal state. Entries to the renewal state correspond

to the renewals of the simple renewal process.

The two major results both connect the forward recurrence times of this process to the sojourn times of the process, one of the results through the probability density functions (Theorem 2.3) and the other through the probability generating functions (Theorem 2.4). It is the second which is of major interest.

The final part of this chapter considers what happens if the number of phases becomes infinite in such a process.

2.2 Simple renewal theory

A simple renewal process is a process

$$N_t = \sum_{n=1}^{\infty} I(T_n \leq t),$$

where $T_n = \sum_{i=1}^{n} X_i$ and the sequence (X_n) of non-negative, independent, identically distributed random variables have probability density function $F(\cdot)$. The times T_n are the times of the renewals and N_t counts the number of renewals to have occurred by time t. Note that

$$T_{N_t} \leq t < T_{N_t+1},$$

so that T_{N_t} is the epoch of the last renewal before t while T_{N_t+1} is the epoch of the next renewal after t. Also $N_{T_n} = n$. The renewal process is said to be *delayed* if X_1 has a different probability distribution function $F_1(\cdot)$ from $F(\cdot)$. We shall consider here only non-delayed renewal processes but note that the limiting results can all be extended to delayed processes. We take $m = E[X_n]$. The *forward* and *backward recurrence times* are defined by

$$\mu_t = T_{N_t+1} - t \quad \text{and} \quad \chi_t = t - T_{N_t},$$

respectively. They are the time until the next renewal and the time from the last renewal respectively.

We shall be concerned with the lattice or arithmetic case where all of the events in the process occur on a set of lattice points. This occurs if $F(x)$ is a step function with the steps on the lattice points kd, $k \in \mathbb{Z}^+$. If d is the largest such number then d is the span of the system. We let $d = 1$ by taking a change in time and then we can define the function $f(n) = p\{X_i = n\}$ for all i. We assume only one renewal can occur at a time so that $f(0) = 0$. We define the *renewal function*

$$H(n) = E[N_n].$$

Because we want a non-delayed renewal process we assume a dummy renewal that is not included in N_n occurs at time zero. We define $h(n) = p\{\text{a renewal occurs at time } n\}$ for $n \in \mathbb{Z}^+$ and thus

$$h(n) = \begin{cases} 0, & n = 0, \\ H(n) - H(n-1), & n > 0, \end{cases}$$

so that

$$H(n) = \sum_{i=1}^{n} h(i).$$

A number of limit theorems are associated with renewal theory. They can all be derived from the *Key Renewal Theorem*. This theorem can be found in many places in several forms. This version is the lattice version of Serfozo (1990). Once the renewal function is defined the Key Renewal Theorem for the lattice case is as follows.

Theorem 2.1 (Key Renewal Theorem) *If F is lattice with span d then*

$$\lim_{n\to\infty} H * g(x + nd) = m^{-1} \sum_{k=1}^{\infty} g(x + kd),$$

provided the sum is finite. In this $H * g$ is the convolution of H and g and m^{-1} is taken to be zero for $m = \infty$.

A renewal equation is an equation involving the renewal function. Such equations arise often in this context due to the regenerative nature of the process. For instance we can write a renewal equation for $h(n)$,

$$h(n) = f(n) + \sum_{l=1}^{n-1} h(n-l)f(l). \tag{2.1}$$

This equation arises from the two possible ways in which a renewal can occur at time n. The first renewal subsequent to time zero could occur at time n with probability $f(n)$. Secondly the first renewal could occur at time $l < n$ which happens with probability $f(l)$. If this is the case we consider the process to have started again at this time and there will be a renewal at time n with probability $h(n-l)$. Summing over these possibilities gives the result.

The Key Renewal Theorem can be applied to this renewal equation as follows. The first term in this sum tends to zero as n tends to infinity and the second sum is the convolution $H * f(k)$ so that as n tends to infinity

$$\begin{aligned} h(n) &\rightarrow m^{-1} \sum_{i=1}^{\infty} f(i) \\ &= m^{-1}. \end{aligned}$$

Theorem 2.2 *Given the above definitions the renewal equation for the forward recurrence times in the renewal process is*

$$p_n(r) = f(r+n) + \sum_{l=1}^{n-1} h(n-l)f(r+l). \tag{2.2}$$

where $p_n(r) = p\{\mu_n = r\}$. Assuming aperiodicity, as n tends to ∞

$$p_n(r) \rightarrow \frac{1}{m} \sum_{l=1}^{\infty} f(r+l). \tag{2.3}$$

Proof: We get the renewal equation for $p_n(r)$ in exactly the same way as in (2.1). The forward renewal time from time n can be equal to r in one of two ways. Firstly the first renewal can occur at time $n + r$. Secondly if a renewal occurs before this time at time $n - l$ (probability $h(n-l)$) it can be followed $r + l$ later by a second renewal with probability $f(r + l)$. Summing gives (2.2). From the key renewal theorem we get the limit as n tends to infinity of $p_n(r)$ to be

$$p(r) = \frac{1}{m} \sum_{l=1}^{\infty} f(r+l).$$

□

2.3 Basic Markov renewal theory

The concept of renewal theory as outlined above is not sufficient for our purposes here. We need the more advanced concepts of Markov renewal theory. This is because the renewal process we shall investigate will have a number of stages through which it progresses. It is natural to call these states and define a process on them as below. We shall need to progress to a more general description again before we are ready to produce the results necessary for the later theory.

Although the definitions are quite general the motivation comes from a simple set of examples. The generality is preserved because it is to be hoped that in the future more difficult problems might be tackled using these concepts.

We must first define what we mean by a Markov renewal process (MRP). If we take a Markov chain on the countable set of states $\mathcal{S}$, which we shall label $1, 2, 3, \ldots$, with the probability transition matrix $\mathbf{P} = (p_{ij})$, and we take the set of probability distribution functions $F_{ij}(t)$ defined for $p_{ij} > 0$ (and arbitrary for $p_{ij} = 0$) then we can describe a Markov renewal process as follows. It is a process in which a certain time is spent in each state before the transition to another state. The choice of transitions between states is governed by the matrix $\mathbf{P}$ and the time spent in state i, conditional on a transition from this state to state j, is determined by probability distribution function F_{ij}. Formally we may consider the two dimensional process $\{(J_n, X_n) : n \geq 0\}$ where

$$\begin{aligned} J_n &= \text{the state after } n \text{ transitions,} \\ X_n &= \text{the time spent in state } J_{n-1} \text{ before the transition to } J_n, \end{aligned}$$

such that if the filtration $\mathcal{G}_n = \sigma((J_m, X_m) : 0 \leq m \leq n)$, (see page 127) and the vector of initial probabilities is $\mathbf{a} = \{a_i\}$

$$\begin{aligned} P\{J_0 = i\} &= a_i, \\ P\{J_{n+1} = i | \mathcal{G}_n\} &= P\{J_{n+1} = i | J_n\} \\ &= p_{J_n, i}, \\ P\{X_{n+1} \leq x | J_n = i, J_{n+1} = j\} &= F_{ij}(x), \end{aligned}$$

where $F_{ij}(x)$ is a probability distribution function such that

$$F_{ij}(x) = 0, \qquad \forall x \leq 0.$$

The reason for this being called a *renewal process* is simply that the future of the queue at a transition epoch, is independent of the history of the process, except through the current state. Thus at transitions the process *renews* itself. An alternative name is a regenerative process.

Pyke (1961a) approaches the definition of these processes in a slightly different manner. It can be shown that the two definitions are equivalent by considering the matrix valued function $\mathbf{Q} : I\!R \to I\!R^N \times I\!R^N$. $\mathbf{Q} = (Q_{ij})$ which is called a *matrix of transition distributions* if the Q_{ij} are non-decreasing functions satisfying

$$Q_{ij}(x) = 0, \qquad x \leq 0$$

and $H_i(x) = \sum_{j=1}^{N} Q_{ij}(x)$ is a probability distribution function for $1 \leq j \leq N$. This agrees with our definition if we take

$$Q_{ij}(x) = p_{ij} F_{ij}(x),$$

or alternatively,

$$\begin{aligned} p_{ij} &= \lim_{x \to \infty} Q_{ij}(x), \\ F_{ij}(x) &= \begin{cases} p_{ij}^{-1} Q_{ij}(x), & p_{ij} > 0, \\ \text{arbitrary}, & p_{ij} = 0. \end{cases} \end{aligned}$$

In Pyke's notation

$$P\{J_{n+1} = k, X_{n+1} \leq x | \mathcal{G}_n\} = Q_{J_n,k}(x), \ \ a.s.$$

for all $x \in I\!R$ and $1 \leq k \leq N$. Pyke defines the process in terms of $\mathbf{Q}(\cdot)$. For our purposes it is more natural to consider $\mathbf{P}$ and $\mathbf{F}(\cdot)$ due to the simplicity of the matrix $\mathbf{P}$ in the process we shall investigate.

The things that are normally investigated in Markov renewal theory are the counting processes defined by

$$N(t) = \sup\{n \geq 0 | T_n \leq t\}, \tag{2.4}$$

$$N_j(t) = \sum_{n=1}^{N(t)} I(J_n = j), \tag{2.5}$$

where $T_0 = 0$ and $T_n = \sum_{i=1}^{n} X_i$. Note that the counting functions $N_j(t)$ do not record the value of J_0. The process $\mathbf{N}(t) = (N_1(t), N_2(t), \ldots)$ will be referred to as the Markov renewal process determined by $(\mathcal{S}, \mathbf{a}, \mathbf{P}, \mathbf{F})$. Where no ambiguity exists we shall refer simply to this as the MRP. We can also define a process (Z_t) by

$$Z_t = J_{N(t)}.$$

This gives the state of the MRP at time t.

Just as in the theory of Markov chains the states of a MRP may be classified as to whether they communicate, are (positive or null) recurrent or transient. For classification we use the following, defined for all $i, j \in \mathcal{S}$,

$$G_{ij}(t) = \begin{cases} p\{N_j(t) > 0 | Z_0 = i\}, & \text{if } t \geq 0, \\ 0, & \text{if } t < 0. \end{cases}$$

Also we define μ_{ii} the mean time between transitions to state i

$$\mu_{ii} = \int_0^\infty t \, dG_{ii}(t).$$

The following classifications are used.

Definition 2.1 *The following definitions are used herein.*
(a) States i and j **communicate** *iff $G_{ij}(\infty)G_{ji}(\infty) > 0$, or $i = j$.*
(b) The disjoint classes of communicating states are denoted by $\{\mathcal{C}_i\}$.
(c) A MRP is said to be **irreducible** *iff there is only one class.*
(d) A state i is said to be **recurrent** *iff $G_{ii}(\infty) = 1$, and is said to be* **transient** *otherwise.*
(e) A state is said to be **null recurrent** *iff it is recurrent and $\mu_{ii} = \infty$ while it is said to be* **positive recurrent** *iff it is recurrent and $\mu_{ii} < \infty$.*
(f) If all of the states $i \in \mathcal{S}$ have one of the preceding properties then we say that the MRP has this property.

Theorem 5.1 of Pyke (1961a) means that the irreducibility and recurrence of the process can be based on the corresponding properties of the Markov chain defined by the transition matrix $\mathbf{P}$. We also define a *renewal function* for the Markov renewal process by $\mathbf{M}(t) = (M_{ij}(t))$ where $M_{ij}(t)$ is

$$M_{ij}(t) = E[N_j(t)|Z_0 = i].$$

Note that we can still use a version of the Key Renewal Theorem. However we shall not be directly interested in $N(t)$ and $\mathbf{M}(t)$, we shall be more interested in forward recurrence times, the previous theory being necessary for definitions and background. The epoch of the next recurrence of state j after time t is given by

$$\tau^j(t) = \inf\{s > t \,|\, N_j(s) > N_j(t)\}.$$

From this we define the forward recurrence times for state j by

$$\mu_i^j(t) = \begin{cases} 0, & \text{if } t = S_n \text{ and } J_n = j \text{ for some } n, \\ \displaystyle\int_t^{\tau^j(t)} I(Z_s = i)ds, & \text{otherwise.} \end{cases}$$

These are the total times spent in states $i \in \mathcal{S}$ before the next recurrence of state j.

2.3.1 Delayed Markov renewal processes

Delayed Markov renewal processes are called generalised Markov renewal processes by Pyke but we shall reserve this terminology for a later stage. These are processes with a different set of probability distribution functions describing the time spent in the initial state. Thus we get a new process which we call the delayed MRP determined by $(S, \mathbf{a}, \mathbf{P}, \tilde{\mathbf{F}}, \mathbf{F})$

$$\begin{aligned} P\{J_0 = i\} &= a_i, \\ P\{J_{n+1} = i|J_n\} &= p_{J_n,i}, \\ P\{X_1 \le x|J_0 = i, J_1 = j\} &= \tilde{F}_{ij}(x), \\ P\{X_{n+1} \le x|J_n = i, J_{n+1} = j\} &= F_{ij}(x), \quad \text{for } n > 1. \end{aligned}$$

Note that if we take $a_i = \eta_i\mu_{ii}^{-1}$ where η_i is the mean time spent in state i and

$$\tilde{F}_{ij}(t) = \eta_i^{-1}\int_0^t [1 - F_{ij}]dy,$$

we get a stationary process. (Pyke, page 1256). For the remainder of this thesis we shall be considering non-delayed renewal processes. It is important to note that our reason for doing this stems from our choice, in later chapters, of initial conditions for the queueing systems considered. An alternative choice might well result in a delayed renewal process but this will have no effect on the limiting behaviour of the systems.

2.3.2 Generalised Markov renewal processes

We now consider a process such as the one above with the modification that the epochs of entry to some states do not constitute renewals. We call this a generalised Markov renewal process (GMRP). If the state space of the process is $\mathcal{S}$ and the non-renewal states are $\mathcal{S}^* \subset \mathcal{S}$ then it is clear that we can no longer simply define Q_{ij} for $i \in \mathcal{S}^*$ because

$$P\{J_{n+1} = j, X_{n+1} \leq x | \mathcal{G}_n\} \neq P\{J_{n+1} = j, X_{n+1} \leq x | J_n\},$$

for $J_n \in \mathcal{S}^*$. There are a number of ways of approaching this. The approach taken by Nakagawa and Osaki (1976) follows from Pyke's definition of a MRP and is to define the functions

$$\begin{aligned} Q_{ij}^{(k_1,k_2,..,k_m)}(x) = P\{&\text{after entering } i \in \mathcal{S}\backslash\mathcal{S}^* \text{ the process next makes transitions through} \\ &\text{states } k_1, k_2, \ldots, k_m \in \mathcal{S}^* \text{ and finally enters state } j \in \mathcal{S}, \\ &\text{in a total amount of time } \leq x\}, \end{aligned}$$

We shall instead preserve the transition matrix $\mathbf{P}$ and define the functions

$$\begin{aligned} F_{ij}^{(k_1,\ldots,k_m)}(x_0, x_1, \ldots, x_m) &= P\{X_n \leq x_0, X_{n+1} \leq x_1, \ldots, X_{n+m} \leq x_m \\ &\qquad | J_n = i, J_{n+1} = k_1, \ldots, J_{n+m} = k_m, J_{n+m+1} = j\}, \end{aligned}$$

for $i, j \in \mathcal{S}\backslash\mathcal{S}^*$ and $k_1, .., k_m \in \mathcal{S}^*$. We call the $(x_0, x_1, \ldots, x_m)$ state sojourn lifetimes conditional on the states $(i, k_1, \ldots, k_m, j)$. $F_{ij}^{(k_1,\ldots,k_m)}$ is then the joint probability distribution function of the times spent in each state prior to a renewal, given the initial state i, the other states prior to the renewal $(k_1, \ldots, k_m)$ and final state j.

We can arrive at the notation of Nakagawa and Osaki (1976) by simply considering

$$Q_{ij}^{(k_1,k_2,..,k_m)}(x) = p_{ik_1} p_{k_1 k_2} \cdots p_{k_m j} \int_{x_1+\cdots+x_m=x} F_{ij}^{(k_1,\ldots,k_m)}(x_0, x_1, \ldots, x_m) dx_0\, dx_1 \cdots dx_m$$

This definition limits the dependence upon the history of the process (beyond dependence on the current state) to the times spent in states and not the actual states. Thus the process describing the series of state transitions is still a Markov chain.

In such a process Definition 2.1 still holds. In order to establish the relationship between the MRP and the Markov chain that describes the transition states we shall use Theorem 5.1 of Pyke. We do this by forming a Markov renewal process from the renewal states $\mathcal{S}\backslash\mathcal{S}^*$. When the non-renewal states and corresponding transitions are ignored we have a standard Markov renewal process and thus we can apply the theorem.

2.3.3 Discrete time

In most analyses of renewal processes it is assumed that the distribution functions are non-lattice. We shall be using renewal theory to examine the behaviour of a discrete-time process embedded in the queueing processes of interest. For this reason we shall only consider lattice Markov renewal processes. If all of the functions F_{ij} or alternately $F_{ij}^{(k_1,\ldots,k_m)}$ are lattice with integer spans we call the greatest common denominator of these spans the span of the process. Herein we assume that the span is 1, which precludes periodicity. We can then introduce the joint probability function

$$\begin{aligned} f_{ij}^{(k_1,\ldots,k_m)}(i_0,i_1,\ldots,i_m) &= P\{X_n = i_0, X_{n+1} = i_1, \ldots, X_{n+m} = i_m \mid \\ &\qquad J_n = i, J_{n+1} = k_1, \ldots, J_{n+m} = k_m, J_{n+m+1} = j\}. \quad (2.6) \end{aligned}$$

The process $Z_t = J_{N(t)}$ is replaced by $Z_n = J_{N(n)}$, the state of the process at time n.

2.4 Multi-phase Markov renewal process

The case now described is the simplest non-trivial case of the generalised Markov renewal process. It is called a Markov renewal process of type I in Nakagawa and Osaki (1976). It is simply the case in which there are N states labelled $1, \ldots, N$ with only state 1 a renewal state and the states visited in sequential order. Thus the probability transition matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{Bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \end{Bmatrix}.$$

Ignoring non-renewal states $2, \ldots, N$ there is only a single state and hence the Markov chain on this is obviously recurrent, so from Theorem 5.1 of Pyke the MRP must be recurrent. Thus the multi-phase Markov renewal process must be recurrent. Furthermore as all of the states communicate the process is irreducible.

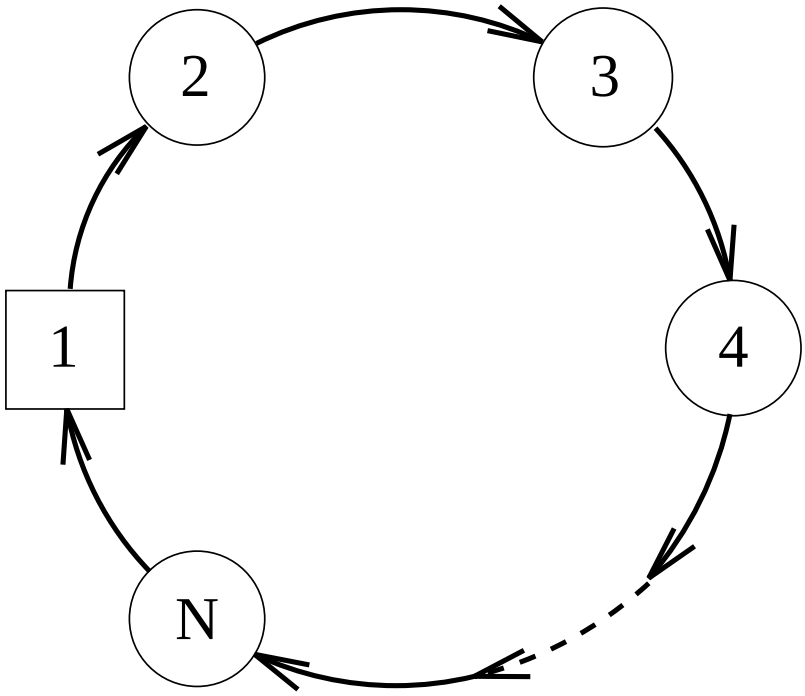

Figure 2.1: The Markov renewal process. □ denotes a renewal point while ◯ denotes a non-renewal point

We refer to this as a *multi-phase Markov renewal process*. The name refers to the fact that the epochs of transitions to state 1 form a simple renewal process which is then broken into phases, as we are interested in some of the internal behaviour of this process. Each of the phases is simply a state in the GMRP. Because of the simplicity of this case, much of the previous notation may be substantially abbreviated. Further this allows us to use the Key Renewal Theorem in its simple form later in this chapter.

Note that the results of this section seem to be novel.

2.4.1 Definitions

We define T_i^j to be the epoch of the ith departure from state j and take $T_m^0 = T_{m-1}^N$. As we consider the lattice case $T_i^j \in \mathbb{Z}^+$. We call the interval $[T_{m-1}^N, T_m^N)$ the mth cycle. The state sojourn times of the mth cycle are the times $\nu_m^j = T_m^j - T_m^{j-1}$, the times spent in each state during that cycle. We assume that the process is not delayed and the vector of initial probabilities is $\mathbf{a} = \mathbf{e}_1 = (1, 0, \ldots, 0)$. By this we mean that the process starts with a transition to state 1 (and a corresponding renewal). So $T_0^N = 0$ and the probability function for the initial transition times is the same as that for all of the other transition times. In terms of the notation of Section 2.3 we have

$$\begin{aligned} X_{(m-1).N+j} &= T_m^j - T_m^{j-1}, \\ J_{(m-1).N+j} &= j. \end{aligned}$$

Due to the nature of renewal processes the joint distributions of $\nu_m^1, \ldots, \nu_m^N$ are identical for all m. Thus we may drop the subscript and refer to the state sojourn times ν^j. We may from this define the joint probability function $f(i_1, i_2, \ldots, i_N)$ of the state sojourn times $\nu^1, \ldots, \nu^N$ by

$$\begin{aligned} f(i_1, i_2, \ldots, i_N) &= p\{\nu^1 = i_1, \nu^2 = i_2, \ldots, \nu^N = i_N\} \\ &= p\{X_{N.(m-1)+1} = i_1, X_{N.(m-1)+2} = i_2, \ldots, X_{N.(m-1)+N} = i_N\}, \\ &= f_{11}^{(2\cdots N)}(i_1, i_2, \ldots, i_N), \end{aligned}$$

with $f_{11}^{(2\cdots N)}$ as is defined by (2.6). We take $\rho = E\left[\nu^1 + \cdots + \nu^N\right]$ and require that $\nu^1 + \cdots + \nu^N \geq 1$ with probability one. The multi-phase MRP is positive recurrent when $\rho < \infty$ and null recurrent otherwise. This can again be seen by considering the renewal process formed by transitions to state 1.

The times of interest are the times spent in each state before the recurrence of state 1, the forward recurrence times μ_i^1. Mathematically these are as follows. Define the $N + 2$ sequences of random variables $(\tau_j(n))_{j=0}^N$ and $(\tau(n))$ for $n \in \mathbb{Z}^+$ as follows

$$\begin{aligned} \tau(n) &= \begin{cases} \inf\{k > n \,|\, k = T_m^N,\ m \in \mathbb{N}\}, & \text{if the set is non-empty,} \\ \infty, & \text{otherwise,} \end{cases} \\ \tau_j(n) &= \begin{cases} \sup\{n < k \leq \tau(n) \,|\, k = T_m^j,\ m \in \mathbb{N}\}, & \text{if the set is non-empty,} \\ n, & \text{otherwise.} \end{cases} \end{aligned}$$

(Note the difference between τ_j and τ^j. The former is defined immediately above while the latter is the epoch of the next recurrence of state j.) In this context we are only interested in the recurrence of state 1 and so we drop the superscript and refer to the following lemma. By the forward recurrence times we mean the forward recurrence times for state 1 and we label these times by $\mu_i(n)$.

Lemma 2.2.1 (Forward recurrence times) *For the multi-phase MRP the following holds.*

$$\mu_i(n) = \begin{cases} 0, & n = T_m^N, \text{ for some } m \in \mathbb{N}, \\ \tau_i(n) - \tau_{i-1}(n), & \text{otherwise,} \end{cases}$$

$\forall i \in \{1, \ldots, N\}$ *and* $n \in \mathbb{N}$.

Proof: We defined the forward recurrence times for state j by

$$\mu_i^j(n) = \begin{cases} 0, & \text{if } n = T_m \text{ and } J_m = j \text{ for some } m \in I\!N, \\ \sum_{k=n}^{\tau^j(n)} I(Z_k = i), & \text{otherwise,} \end{cases}$$

where, as before, Z_k is the state of the process at time k and

$$\tau^j(t) = \inf\{s > t \mid N_j(s) > N_j(t)\}.$$

Now for a multi-phase MRP the definitions of $\tau^1(n)$ and $\tau(n)$ are equivalent. Thus

$$\mu_i^1(n) = \begin{cases} 0, & \text{if } n = T_m \text{ and } J_m = 1 \text{ for some } m \in I\!N, \\ \sum_{k=n}^{\tau(n)} I(Z_k = i), & \text{otherwise.} \end{cases}$$

As T_m is the time of the mth transition and J_m is the state to which the mth transition is made we can see that $n = T_m$ and $J_m = 1$ for some $m \in I\!N$ is equivalent to $n = T_m^N$ for some $m \in I\!N$. Also for a multi-phase MRP

$$I(Z_k = i) = \begin{cases} 1, & \text{if } \tau_{i-1}(n) \le k < \tau_i(n), \\ 0, & \text{otherwise,} \end{cases}$$

so that (dropping the superscript) we get

$$\mu_i(n) = \begin{cases} 0, & \text{if } n = T_m^N \text{ for some } m \in I\!N, \\ \tau_i(n) - \tau_{i-1}(n), & \text{otherwise.} \end{cases}$$

□

Assuming that at time n, one is in the mth cycle and the current state is $Z_n > 1$, the forward recurrence times are as follows for $1 \le j \le N$,

$$\mu_j(n) = \begin{cases} 0, & j < k, \\ T_m^j - n, & j = k, \\ T_m^j - T_m^{j-1}, & j > k. \end{cases} \tag{2.7}$$

If the state at time n, $Z_n = 1$, then there are two possibilities. The first is simply that a renewal occurs at time n in which case $\mu_j(n) = 0$ for all $1 \le j \le N$. If this is not the case then we get

$$\mu_j(n) = \begin{cases} T_m^1 - n, & j = 1, \\ T_m^j - T_m^{j-1}, & j > 1. \end{cases} \tag{2.8}$$

Within the following sections we use the notation

Definition 2.2 *For $l = 1, \ldots, N$*

$$\sum_{\spadesuit l} = \sum_{i_l=0}^{\infty} \sum_{i_{l+1}=0}^{\infty} \cdots \sum_{i_N=0}^{\infty},$$

$$\sum_{\heartsuit l} = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_{l-1}=0}^{\infty},$$

with the empty set in the latter, for $l = 1$, being interpreted as zero.

This is used to simplify notation in later parts of this chapter.

We can define the joint probability functions

$$
\begin{aligned}
q^n(r_1, r_2, .., r_N) &= p\{\mu_i(n) = r_i,\ (1 \le i \le N)\}, && (2.9)\\
p_n^l(r_l, r_{l+1}, ..., r_N) &= p\{Z_{n-1} = l,\ \mu_i(n) = r_i,\ (l \le i \le N)\}, && (2.10)\\
f^l(i_l, i_{l+1}, ..., i_N) &= \sum_{\heartsuit l} f(i_1, i_2, ..., i_N), && (2.11)\\
g_n^l(i_l, i_{l+1}, ..., i_N) &= \sum_{j=1}^{n} \sum_{i_1+\cdots+i_{l-1}=n-j} f(i_1, .., i_{l-1}, i_l + j, i_{l+1}, ..., i_N), && (2.12)\\
f(i) &= \sum_{i_1+\cdots+i_N=i} f(i_1, i_2, ..., i_N). && (2.13)
\end{aligned}
$$

Note that $g_n^l(i_l, i_{l+1}, ..., i_N)$ is the probability that the first l transitions in the process take total time $n + i_l$, the lth transition takes time $> i_l$ and transitions $l + 1$ to N take times i_{l+1} to i_N respectively and $p_n^l(r_l, r_{l+1}, ..., r_N)$ is the probability that at time n the the last prior transition was to state l and the the forward recurrence times $\mu_l, \ldots, \mu_N$ are $r_l, \ldots, r_N$ respectively. Also from the theorem of total probability we get for $n > 1$

$$
q^n(r_1, r_2, .., r_N) = \sum_{l=1}^{N} p_n^l(r_l, r_{l+1}, .., r_N). \qquad (2.14)
$$

We can define the following probability generating functions

$$
\begin{aligned}
F^*(x_1, x_2, \ldots, x_N) &= \sum_{\spadesuit 1} \left(\prod_{k=1}^{N} x_k^{i_k} \right) f(i_1, i_2, .., i_N),\\
F_l^*(x_l, x_{l+1}, \ldots, x_N) &= \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{i_k} \right) f^l(i_l, i_{l+1}, .., i_N),\\
&= F^*(1, \ldots, 1, x_l, \ldots, x_N),\\
F_{N+1}^* &= 1,
\end{aligned}
$$

when these converge. We take $h(n)$ to be the probability of a renewal (subsequent to time 0) at time n. Thus

$$
h(n) = \sum_{m=1}^{\infty} p\{T_m^N = n\}.
$$

2.4.2 Results

Theorem 2.3 *Given the above definitions the renewal equation is*

$$p_n^l(r_1,\ldots,r_N) \;=\; g_n^l(r_l,\ldots,r_N) + \sum_{k=1}^{n-1} h(n-k) g_k^l(r_l,\ldots,r_N). \tag{2.15}$$

Assuming aperiodicity, for $l = 1,\ldots,N$, and $\rho < \infty$ as $n \to \infty$ we get

$$h(n) \;\to\; \frac{1}{\rho}, \tag{2.16}$$

$$p_n^l(r_l,\ldots,r_N) \;\to\; \frac{1}{\rho}\sum_{k=1}^{\infty} f^l(r_l+k,\ldots,r_N). \tag{2.17}$$

Proof: To obtain (2.16) we consider the simple renewal process formed by transitions to state 1. This has probability density function given by $f(\cdot)$ defined in (2.13). This will behave as in the simple renewal process of Section 2.2. Thus we get (2.16).

In order to obtain the renewal equation we follow the procedure of Theorem 2.2 and sum over all of the possibilities to get

$$p_n^l(r_l,\ldots,r_N) \;=\; g_n^l(r_l,\ldots,r_N) + \sum_{k=1}^{n-1} h(n-k) g_k^l(r_l,\ldots,r_N).$$

The lattice version of the Key Renewal Theorem gives

$$p^l(r_l,\ldots,r_N) \;=\; \frac{1}{\rho}\sum_{k=1}^{\infty} g_k^l(r_l,\ldots,r_N).$$

Now

$$\begin{aligned}
\sum_{k=1}^{\infty} g_k^l(r_l,\ldots,r_N) &= \sum_{k=1}^{\infty}\sum_{j=1}^{k}\sum_{i_1+\cdots+i_{l-1}=k-j} f(i_1,\ldots,i_{l-1},r_l+j,r_{l+1},\ldots,r_N) \\
&= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty}\sum_{i_1+\cdots+i_{l-1}=k-j} f(i_1,\ldots,i_{l-1},r_l+j,r_{l+1},\ldots,r_N) \\
&= \sum_{j=1}^{\infty}\sum_{k=0}^{\infty}\sum_{i_1+\cdots+i_{l-1}=k} f(i_1,\ldots,i_{l-1},r_l+j,r_{l+1},\ldots,r_N) \\
&= \sum_{j=1}^{\infty}\sum_{\heartsuit l} f(i_1,\ldots,i_{l-1},r_l+j,r_{l+1},\ldots,r_N) \\
&= \sum_{j=1}^{\infty} f^l(r_l+j,r_{l+1},\ldots,r_N)
\end{aligned}$$

and so

$$p^l(r_l,\ldots,r_N) \;=\; \frac{1}{\rho}\sum_{k=1}^{\infty} f^l(r_l+k,r_{l+1},\ldots,r_N). \qquad \square$$

Remark: Intuitively this can be explained in the following manner. We consider the process after a long time. We then consider all of the ways in which a set of forward recurrence times can occur. In order to have the forward renewal times $r_l,\ldots,r_N$,

given the last transition prior to time n was to state l, we must have the sojourn times $i_{l+1} = r_{l+1}, \ldots, i_N = r_N$. The the only restriction on i_l is that $i_l \geq r_l + 1$ and there will be no restrictions on i_1 to i_{l-1}. The probability of any particular set of sojourn times $i_1, \ldots, i_N$ will be $f(i_1, \ldots, i_N)$. As we consider the long time limit, the probability of a renewal at any particular time point is a constant $1/\rho$. Thus the probability of a set of sojourn times occurring in the context of our forward recurrence times will be $\frac{1}{\rho} f(i_1, \ldots, i_N)$ when $i_l \geq r_l + 1$ and $i_{l+1} = r_{l+1}, \ldots, i_N = r_N$ and zero otherwise. When we sum over all of these probabilities we get the result. This alternative explanation is more intuitive but lacks the rigour of the proof.

As we can now see that the limits exist we define

$$q(r_1, r_2, \ldots, r_N) = \lim_{n \to \infty} q^n(r_1, \ldots, r_N), \tag{2.18}$$

$$p^l(r_l, \ldots, r_N) = \lim_{n \to \infty} p_n^l(r_l, \ldots, r_N). \tag{2.19}$$

The corresponding probability generating functions are

$$Q^*(x_1, x_2, \ldots, x_N) = \sum_{\spadesuit 1} \left(\prod_{k=1}^{N} x_k^{r_k} \right) q(r_1, r_2, \ldots, r_N),$$

$$P_l^*(x_l, x_{l+1}, \ldots, x_N) = \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{r_k} \right) p^l(r_l, r_{l+1}, \ldots, r_N),$$

and from (2.14) we get

$$Q^*(x_1, x_2, \ldots, x_N) = \sum_{l=1}^{N} P_l^*(x_l, x_{l+1}, \ldots, x_N). \tag{2.20}$$

Theorem 2.4 *Given the above definitions, $\forall N \geq 1$ and $x_1, x_2, \ldots, x_N \in [0, 1)$,*

$$Q^*(x_1, x_2, \ldots, x_N) = \frac{1}{\rho} \sum_{l=1}^{N} \frac{F_{l+1}^*(x_{l+1}, \ldots, x_N) - F_l^*(x_l, \ldots, x_N)}{1 - x_l}. \tag{2.21}$$

Furthermore, if Q^* converges for some $x_1, x_2, \ldots, x_N$ not necessarily all in the interval $[0, 1)$, then it converges to the right-hand side of equation 2.21.

Proof: From Theorem 2.3 and Definitions 2.19 and 2.11 we can see that

$$p^l(i_l, i_{l+1}, .., i_N) = \frac{1}{\rho} \left\{ f^{l+1}(i_{l+1}, \ldots, i_N) - \sum_{m=0}^{i_l} f^l(m, i_{l+1}, .., i_N) \right\}. \tag{2.22}$$

Consider first the case with $x_1, \ldots, x_N \in [0, 1)$. Multiplying (2.22) by $\left(\prod_{k=l}^{N} x_k^{i_k} \right)$ and summing over i_k for $k = l, \ldots, N$ we get the generating function P_l^* on the left-hand side and on the right-hand side we get the following

$$P_l^*(x_l, x_{l+1}, \ldots, x_N) = \frac{1}{\rho} \left\{ \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{i_k} \right) f^{l+1}(i_{l+1}, \ldots, i_N) \right.$$

$$
\begin{aligned}
& - \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{i_k} \right) \sum_{m=0}^{i_l} f^l(m, i_{l+1}, .., i_N) \Bigg\} \\
= \; & \frac{1}{\rho} \Bigg\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{i_l=0}^{\infty} x_l^{i_l} f^{l+1}(i_{l+1}, \ldots, i_N) \\
& - \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{i_l=0}^{\infty} x_l^{i_l} \sum_{m=0}^{i_l} f^l(m, i_{l+1}, .., i_N) \Bigg\} \\
= \; & \frac{1}{\rho} \Bigg\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) f^{l+1}(i_{l+1}, \ldots, i_N) \left(\sum_{i_l=0}^{\infty} x_l^{i_l} \right) \\
& - \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{m=0}^{\infty} f^l(m, i_{l+1}, .., i_N) \left(\sum_{i_l=m}^{\infty} x_l^{i_l} \right) \Bigg\} \\
= \; & \frac{1}{\rho} \Bigg\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) f^{l+1}(i_{l+1}, \ldots, i_N) \left(\frac{1}{1-x_l} \right) \\
& - \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{m=0}^{\infty} f^l(m, i_{l+1}, .., i_N) \left(\frac{x_l^m}{1-x_l} \right) \Bigg\} \\
= \; & \frac{1}{\rho} \frac{1}{1-x_l} \Bigg\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) f^{l+1}(i_{l+1}, \ldots, i_N) \\
& - \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{i_k} \right) f^l(i_l, .., i_N) \Bigg\},
\end{aligned}
$$

that is,

$$
P_l^*(x_l, x_{l+1}, \ldots, x_N) = \frac{1}{\rho} \left\{ \frac{F_{l+1}^*(x_{l+1}, \ldots, x_N) - F_l^*(x_l, \ldots, x_N)}{1 - x_l} \right\}. \tag{2.23}
$$

If Q^* converges for some set of $x_j > 1$ then P_l^* must also converge for this set. Note that if $x_l < 1$ the proof of (2.23) remains unchanged. However if $x_l > 1$ we must modify this proof as follows.

$$
\begin{aligned}
P_l^*(x_l, x_{l+1}, \ldots, x_N) &= \frac{1}{\rho} \left\{ \sum_{\spadesuit l} \left(\prod_{k=l}^{N} x_k^{i_k} \right) \sum_{m=1}^{\infty} f^l(i_l + m, \ldots, i_N) \right\} \\
&= \frac{1}{\rho} \left\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{i_l=0}^{\infty} \sum_{m=1}^{\infty} x_l^{i_l} f^l(i_l + m, \ldots, i_N) \right\} \\
&= \frac{1}{\rho} \left\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{m=1}^{\infty} x_l^{-m} \sum_{i_l=0}^{\infty} x_l^{i_l+m} f^l(i_l + m, \ldots, i_N) \right\} \\
&= \frac{1}{\rho} \Bigg\{ \sum_{\spadesuit l+1} \left(\prod_{k=l+1}^{N} x_k^{i_k} \right) \sum_{m=1}^{\infty} x_l^{-m} \Bigg[\sum_{i_l=0}^{\infty} x_l^{i_l} f^l(i_l, \ldots, i_N) \\
&\qquad - \sum_{i_l=0}^{m-1} x_l^{i_l} f^l(i_l, \ldots, i_N) \Bigg] \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty} x_l^{i_l} f^l(i_l,\ldots,i_N)\sum_{m=1}^{\infty} x_l^{-m}\right.\\
&\qquad\left.-\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{m=1}^{\infty} x_l^{-m}\sum_{i_l=0}^{m-1} x_l^{i_l} f^l(i_l,\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l}\left(\prod_{k=l}^{N} x_k^{i_k}\right) f^l(i_l,\ldots,i_N)\frac{x_l^{-1}}{1-x_l^{-1}}\right.\\
&\qquad\left.-\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty}\sum_{m=i_l+1}^{\infty} x_l^{i_l-m} f^l(i_l,\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l}\left(\prod_{k=l}^{N} x_k^{i_k}\right) f^l(i_l,\ldots,i_N)\frac{x_l^{-1}}{1-x_l^{-1}}\right.\\
&\qquad\left.-\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty} f^l(i_l,\ldots,i_N)\sum_{m=i_l+1}^{\infty} x_l^{i_l-m}\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l}\left(\prod_{k=l}^{N} x_k^{i_k}\right) f^l(i_l,\ldots,i_N)\frac{x_l^{-1}}{1-x_l^{-1}}\right.\\
&\qquad\left.-\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty} f^l(i_l,\ldots,i_N)\sum_{m=1}^{\infty} x_l^{-m}\right\}\\
&= \frac{1}{\rho}\frac{x_l^{-1}}{1-x_l^{-1}}\left\{\sum_{\spadesuit l}\left(\prod_{k=l}^{N} x_k^{i_k}\right) f^l(i_l,\ldots,i_N)\right.\\
&\qquad\left.-\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right) f^{l+1}(i_{l+1},\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\frac{F^*_{l+1}(x_{l+1},\ldots,x_N)-F^*_l(x_l,\ldots,x_N)}{1-x_l}\right\},
\end{aligned}
$$

which again gives us (2.23). The case when $x_l = 1$ can be seen to give

$$
\begin{aligned}
P^*_l(x_l,x_{l+1},\ldots,x_N) &= \frac{1}{\rho}\left\{\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty}\sum_{k=1}^{\infty} f^l(i_l+k,\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{i_l=0}^{\infty}\sum_{k=i_l+1}^{\infty} f^l(k,\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{k=1}^{\infty}\sum_{i_l=0}^{k-1} f^l(k,\ldots,i_N)\right\}\\
&= \frac{1}{\rho}\left\{\sum_{\spadesuit l+1}\left(\prod_{k=l+1}^{N} x_k^{i_k}\right)\sum_{k=1}^{\infty} k f^l(k,\ldots,i_N)\right\}\\
&= \left.\frac{dF_l}{dx_l}(x_l,...,x_N)\right|_{x_l=1},
\end{aligned}
$$

as we would expect from L'Hôpitals rule. From (2.23) and (2.20) we get (2.21), that is

$$Q^*(x_1, x_2, \ldots, x_N) \quad = \quad \frac{1}{\rho} \sum_{l=1}^{N} \frac{F^*_{l+1}(x_{l+1}, \ldots, x_N) - F^*_l(x_l, \ldots, x_N)}{1 - x_l},$$

which is the desired result. □

2.5 Multi-phase Markov renewal process with an infinite number of phases.

We consider here the case when $N = \infty$. These occur naturally in some circumstances. In Figure 2.2 we present a GMRP that represents a situation we shall see in Chapter 7.

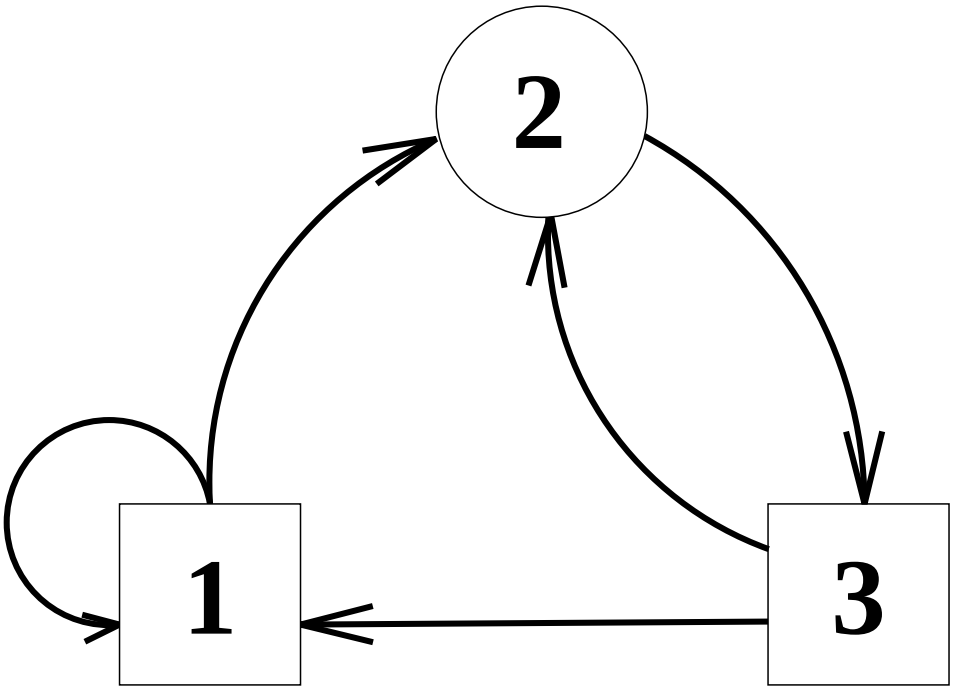

Figure 2.2: GMRP type II. □ denotes a renewal point while ◯ denotes a non-renewal point

This has two renewal states. Starting from state 1 we can return directly to state 1 or have a transition to state 2. From here the process may undertake any number of transitions between states 2 and 3 before returning through state 3 to state 1. We shall call this the GMRP type II.

We obtain a multi-phase MRP from this as follows. We start in state 1. If we return immediately to state 1 we consider the multi-phase process to have traversed all of the other states in zero time. If from state 1 we go to state 2 we proceed as follows. We consider each subsequent entry into states two and three, before returning to state 1, to be a new state. In this case we may alternate an infinite number of times between states 2 and 3, before the end of the busy period so there are an infinite number of states. The two phases alternate an infinite number of times with probability zero (as a transition from 3 to 1 occurs with positive probability and so the process's recurrence is not adversely effected by this. When the transition to state 1 finally occurs we consider the process to again traverse all of the unvisited states in zero time before returning to state 1.

Thus we get the multi-phase MRP of Figure 2.3 with the states defined as follows. States $2n+2$ in the multi-phase MRP corresponds to the nth entry (before returning to state 1) to state 2 of the GMRP type II while states $2n+3$ correspond to the nth entry

to state 3 in the GMRP type II and state 1 corresponds to state 1. Upon return to state one we begin this transition through states $2, 3, \ldots$, again. Note also that the positive recurrence (or null recurrence) of state 1 (and hence the other states) in the multi-phase MRP will be related directly to the positive recurrence (or null recurrence) of state 1 in the GMRP type II.

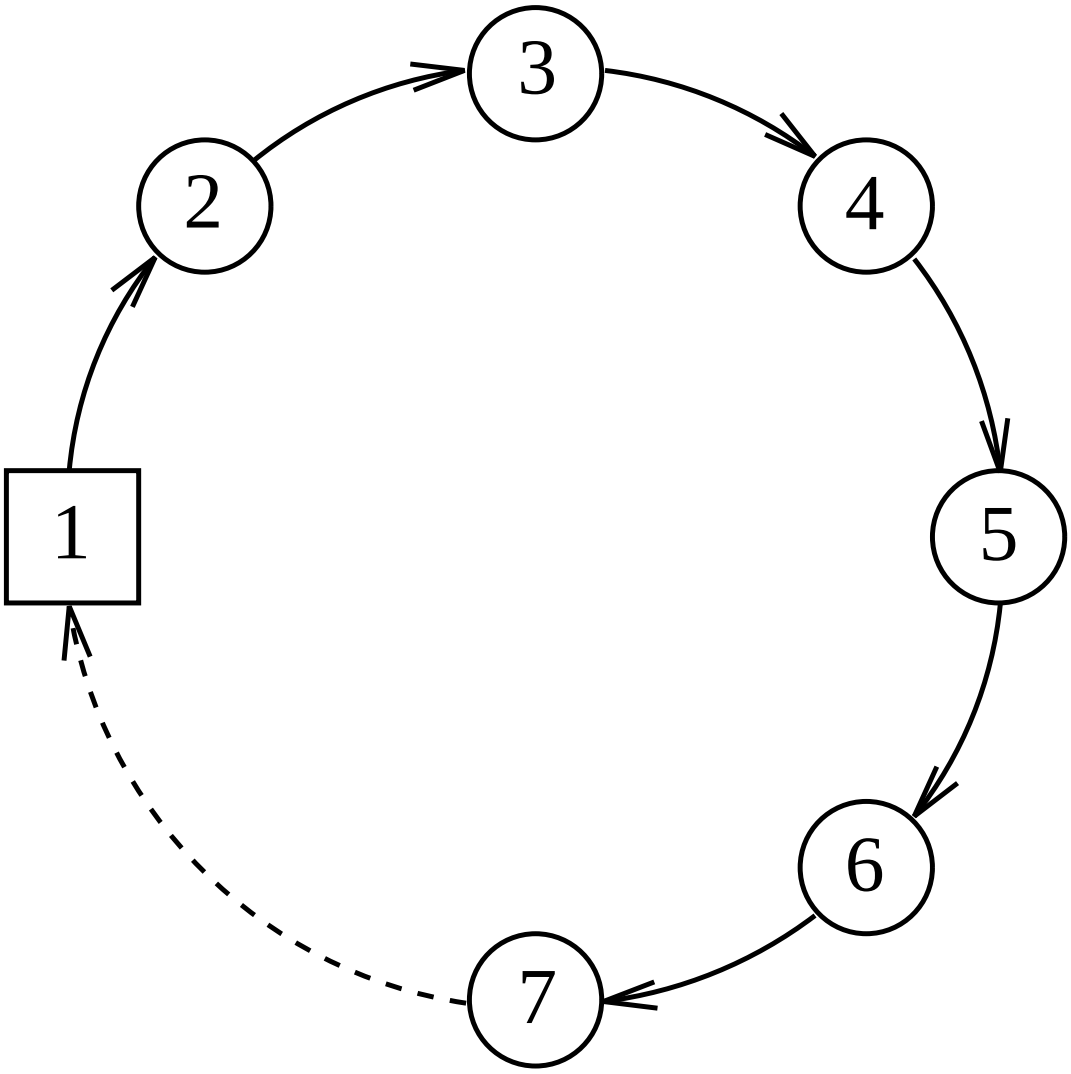

Figure 2.3: Multi-phase MRP with an infinite number of phases. □ denotes a renewal point while ◯ denotes a non-renewal point

Also we note that the odd numbered states would still appear to be renewal states in this process. They are not, but only because when we finally have a transition from state 3 to state 1 in the GMRP type II in the equivalent multi-phase MRP we get a transition through all of the remaining states in the process. This means that the times spent in these phases are no longer entirely independent.

Finally $N(t)$, the number of transitions that occur before time t, becomes infinite after only one cycle through the phases of the multi-phase MRP. Thus $p\{N(t) < \infty, \forall t \geq 0\} = 0$ and we can no longer define $N_j(t)$ as in (2.5). However if we define $C(t)$ the number of cycles that have occurred before time t we can see that $N_j(t)$ will simply be

$C(t)$ or $C(t)+1$ depending on the current state. Thus we can define all of the necessary quantities in a sensible fashion. The epochs of transitions T_m^j will still be defined and we can continue. We state without proof the following extension of Theorem 2.4.

Theorem 2.5 *Given the above definitions and* $x_1, x_2, \ldots \in [0,1)$,

$$Q^*(x_1, x_2, \ldots) = \frac{1}{\rho} \sum_{l=1}^{\infty} \frac{F_{l+1}^*(x_{l+1}, \ldots) - F_l^*(x_l, \ldots)}{1 - x_l}. \tag{2.24}$$

Furthermore, if Q^* *converges for some* $x_1, x_2, \ldots$ *not necessarily all in the interval* $[0,1)$, *then it converges to the right-hand side of equation 2.24.*

Chapter 3

The Multi-phase M/G/1 Queue

3.1 Introduction

We shall consider a single-server queueing process in which the arrivals form a homogeneous Poisson process with rate λ, and the service times are non-negative random variables each with probability distribution function given by one of a set of probability distribution functions $\{A^j(\cdot)\}_{j=1}^N$. The queue size is unlimited. The service discipline may be any non-preemptive discipline. The period during which $A^j(\cdot)$ is chosen is called *phase* j. Thus we have N phases labelled $1, \ldots, N$.

We assume that the phase changes occur at the end of services and are stopping times with respect to the filtration generated by the queueing process. This essentially means that the decision to change phase at time T is based only on the information up until the time T, and not on any information about the future behaviour of the queue. Also we assume that the times spent in two phases are independent if the two phases are not in the same busy period. These limitations are necessary for the analysis to follow, but are not unreasonable assumptions.

The motivating case and the case for which we calculate solutions is the case in which the phases occur in some specific pre-defined order. We shall label the phases in the order that they occur and call one transition through all of the phases a *cycle*.

It will also be convenient to consider each cycle of transitions between phases to occur during one busy period. That is, we start the cycle (in phase 1) when an arriving customer finds the system with no customers in it, and if the cycle is not complete by the time the system is again empty, we say the process spends zero time in the remaining phases. Thus we enter each phase exactly once during each busy period. This and the fact that the times spent in two phases in different busy periods are independent mean that the ends of busy periods are still renewal points of the process. We shall call a queue which satisfies all of the above the *multi-phase M/G/1 queue*.

This does not limit the systems considered as much as it may at first seem. For instance, if a particular phase is skipped over, we may insert a transition through the missing phase which takes zero time. Also if a phase may be visited more than once during a cycle we may consider the second entry to that phase to be a new phase, say $N+1$, and so on for future repetitions. This introduces the possibility of infinitely many phases, which we shall not consider until Section 3.6.

As is the case with the usual M/G/1 queue we consider the embedded, discrete-

time process formed when one observes the number of customers in the system after departures. In this case this embedded process does not form a Markov chain, as with the standard M/G/1 case, without the additional complexity of a supplementary variable to describe the current phase. We shall follow the approach of Baccelli and Makowski (1985,1989) in defining a martingale with respect to the embedded process, and from this we establish a relationship between the forward recurrence times in a multi-phase Markov renewal process and the system size.

Before the results can be obtained we need to prove the regularity of all the stopping times involved with respect to the martingale of interest. This is closely linked with the stability of the queueing process as we shall see. The primary condition of interest called condition $(*)$ is considered in Section 3.2.2.

Then we get the fundamental relationship of this paper which is expressed in Theorem 3.5. From an analysis of the multi-phase MRP, in Chapter 2, the limiting probability generating function for the number of customers in the system in equilibrium may be expressed in terms of that of the state sojourn times of the multi-phase MRP. Helpful results for the calculation of these probability generating functions are found in Corollary 3.6 and Theorem 3.7 using the martingale once again. A general form for the equilibrium probability generating function of the system size can then be found. This is expressed in Theorem 3.9.

The example of the standard M/G/1 queue is examined using this technique in Section 3.7. To obtain the final solution for a more complicated problem further work must be done using conventional probabilistic arguments. Further examples are examined in Chapters 4-7.

Note that in this chapter the exact nature of the transitions between phases is not specified. We provide the constraints on what types of transitions are allowed but say nothing about the actual way in which the transitions are governed. The theory allows a quite general approach to these transitions. In Chapters 4-7 we consider a number of possible cases. For instance in Chapter 4 we consider a process in which the transition occurs when some threshold is crossed. This threshold could be a physical limit on the queue size or a limit on the number of customers served after the busy period begins. The scope for choice of this threshold is quite large. We shall often refer to the points in the process at which transitions occur as a threshold. For more details the reader must consider the examples presented in following chapters. These, however, are by no means exhaustive.

We now describe the model used to examine these processes. The basic parts of this model are described in Appendix C.

3.1.1 The model

Take the number of customers in the system at time t to be $X(t)$. We consider the discrete-time embedded process X_n, where X_n is the number of customers seen by the nth departing customer. Formally, if the departure epochs are $t_1, t_2, \ldots$ then $X_n = X(t_n+)$. We shall assume that the queue starts at time 0 with a departure, thus $X_0 = 0$. The arrivals to the queue form a homogeneous Poisson process with rate λ. The service-time distribution is a general service-time distribution chosen from a set of general distributions according to the phase of the queue, where the phase is chosen from the set $\{1, \ldots, N\}$. The phases of the queue obey certain rules.

(i) The phase can change only on service completion.

(ii) The times at which phase changes occur are stopping times.

(iii) The times spent in two phases in different busy periods are independent.

For the examples we consider we also require the following three extra conditions on the phases.

(iv) At the start of busy periods we are always in phase 1.

(v) The phases occur in order, so phase i is followed by phase $(i+1) \bmod N$.

(vi) Each phase is entered exactly once during a busy period.

We refer to the time from the beginning of phase 1 to the end of phase N as a cycle. Thus a cycle corresponds to a busy period. We say $phase(n) = i$, if after the nth departure, the system is in phase i.

We define the epochs at which the phase changes $T_i^j \in \mathbb{Z}^+$, by

$$T_i^j = \text{the time of the } i\text{th transition out of phase } j,$$

with $T_i^0 = T_{i-1}^N$. We define C_n^j in terms of T_i^j by

$$\begin{aligned} C_n^j &= \bigcup_{i \in \mathbb{N}} \{\omega \,|\, T_i^{j-1}(\omega) \leq n < T_i^j(\omega)\} \\ &= \bigcup_{i \in \mathbb{N}} \{T_i^{j-1} \leq n < T_i^j\}, \end{aligned}$$

so that C_n^j is the event that at time n the queue is in phase j. We use the usual indicator notation

$$I_{C_n^j} = \begin{cases} 1, & phase(n) = j, \\ 0, & \text{otherwise.} \end{cases}$$

It is worth noting that the above does not preclude zero time being spent in a phase, as T_i^{j-1} might equal T_i^j. In this case we still say T_i^{j-1} occurs before T_i^j.

During phase j the service times are random variables with distribution function $A^j(\cdot)$. Service times are assumed to be independent of the arrival epochs. We take the number of arrivals during the nth service time, given the queue is in phase j, to be the random variable A_n^j. These random variables form N independent, identically distributed sequences of random variables (A_n^j). We define $a_i^j = p\{A_1^j = i\}$. We take the probability generating function

$$\begin{aligned} a_j(z) &= E\left[z^{A_1^j}\right] \\ &= \sum_{i=0}^{\infty} a_i^j \, z^i, \end{aligned}$$

and note that (from Theorem C.3) it is given in term of the Laplace-Stieltjes transform of $A^j(\cdot)$ by

$$a_j(z) = A^{j*}(\lambda - \lambda z).$$

We take $\rho_j = a_j'(1)$ which is the mean number of arrivals during a single service in phase j and we call this the *traffic intensity* during phase j. Note that $\rho_j = \lambda/\mu_j$ where $1/\mu_j$ is the mean service-time during phase j. In the following text we shall use $\xi_j(z)$ defined by

$$\xi_j(z) = \frac{z}{a_j(z)}.$$

Given this model we can define $N+2$ sequences of stopping times $(\tau_0(n)), (\tau_1(n)), ..., (\tau_N(n))$ and $(\tau(n))$ for $n \in \mathbb{Z}^+$ as follows

$$\begin{aligned}
\tau(n) &= \begin{cases} \inf\{m > n \,|\, X_m = 0\}, & \text{if the set is non-empty,} \\ \infty, & \text{otherwise,} \end{cases} \\
\tau_j(n) &= \tau(n) \wedge \inf\{m \geq n \,|\, phase(m) > j\} \\
&= \tau(n) \wedge \inf\left\{ m \geq n \,\middle|\, \sum_{i=j+1}^{N} I_{C_m^i} = 1 \right\},
\end{aligned}$$

where $\wedge$ denotes the minimum (and $\vee$ denotes the maximum). When $j = N$ the sum is empty and so $\tau_N(n) = \tau(n)$. Note then that

$$n = \tau_0(n) \leq \tau_1(n) \leq \cdots \leq \tau_i(n) \leq \cdots \leq \tau_N(n) = \tau(n),$$

with probability one. $\tau(n)$ is the epoch of the end of the current busy period at time n. $\tau_j(n) = n$ if the process has already been through phase j in the current busy period and otherwise it is the time of the next transition out of phase j. Note that when the busy period ends we consider the process to go through the remaining phases, spending zero time in each. That $\tau(n)$ and $\tau_j(n)$ are stopping times comes directly from the fact that we only allow phase transitions at stopping times. We can also define the following sequences of times

$$\nu_j(n) = \tau_j(n) - \tau_{j-1}(n), \tag{3.1}$$

$$\mu_j(n) = \begin{cases} \nu_j(n), & X_n \neq 0, \\ 0, & X_n = 0, \end{cases} \tag{3.2}$$

for $j = 1, \ldots, N$. We assume that there is a dummy service completion at time zero. Thus we can take X_0 to be some random variable Ξ. For our purposes here it is convenient to take $X_0 = 0$ a.s. and correspondingly $phase(0) = 1$. Hence $T_0^0 = 0$. One of the results of this is

$$\mu_j(\tau_{l-1}(0)) = \begin{cases} \nu_j(0), & j \geq l, \\ 0, & j < l. \end{cases}$$

3.1.2 Probabilistic elements

Of course the above model must be specified on some probability space $(\Omega, \mathcal{F}, P)$. We wish all of the random variables to be $\mathcal{F}$-measurable. The phase at a given time, and the

number of customers in the system at a given time are sufficient to generate this space. We may define the filtration $\mathcal{F}_n$ by

$$\mathcal{F}_n = \sigma(A_m^j | \, 0 \leq m \leq n, j = 1, .., N).$$

Clearly X_n is determined purely by A_m^j and $I_{C_m^j}$ at times $m \leq n$. We chose the ends of phases to be at stopping times, thus $\{T_i^j \leq n\} \in \mathcal{F}_n$ for all $j = 1, \ldots, N$ and $i \in I\!N$. As $\mathcal{F}_n$ is a σ-algebra we can see also that $\{T_i^j > n\} \in \mathcal{F}_n$ for all $j = 1, \ldots, N$ and $i \in I\!N$ (as $\mathcal{F}_n$ is closed under complements). Now from this $C_n^j = \bigcup_{i \in I\!N} \{T_i^{j-1} \leq n < T_i^j\} \in \mathcal{F}_n$ (as $\mathcal{F}_n$ is closed under intersections and countable unions) and hence $I_{C_n^j}$ is $\mathcal{F}_n$-measurable. Thus X_n is $\mathcal{F}_n$-measurable. Indeed we can see that for all $m \leq n$, X_m, $I_{C_m^j}$ and A_m^j are all $\mathcal{F}_n$-measurable. We then take

$$\mathcal{F} \;=\; \bigcup_{n=1}^{\infty} \mathcal{F}_n.$$

3.2 The martingale

We now define the martingale which will provide the majority of the results herein.

Theorem 3.1 *The following* $\Big(M_n(z)\Big)$ *is a non-negative integrable martingale for* $z \in (0, 1]$.

$$\begin{aligned} M_0(z) &= 1, \\ M_n(z) &= z^{X_n} \prod_{k=0}^{n-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^N I_{C_k^j} a_j(z)} \right), \quad n \geq 1. \end{aligned}$$

Proof: We have

$$\begin{aligned} E\left[M_{n+1}(z) | \, \mathcal{F}_n \right] &= E\left[z^{X_{n+1}} \prod_{k=0}^{n} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^N I_{C_k^j} a_j(z)} \right) \middle| \, \mathcal{F}_n \right] \\ &= \prod_{k=0}^{n} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^N I_{C_k^j} a_j(z)} \right) E\left[z^{X_{n+1}} \middle| \, \mathcal{F}_n \right], \; a.s., \end{aligned}$$

from the $\mathcal{F}_n$-measurability of C_k^j and $I(X_k \neq 0)$ for $k = 0$ to n. The following recurrence relation gives X_{n+1} in terms of X_n,

$$X_{n+1} = X_n + \sum_{j=1}^{N} I_{C_n^j} A_{n+1}^j - I(X_n \neq 0).$$

This simply states that during the busy period the number of customers left in the system after a service completion is the number in the system before the service begins, plus the number who arrive during the service, minus one for the customer who completed service. When the queue is empty it must wait for a customer to arrive before it begins

service and so there is one extra arrival to the system, hence the $I(X_n \neq 0)$ term. So

$$\begin{aligned} E\left[M_{n+1}(z)\middle|\mathcal{F}_n\right] &= \prod_{k=0}^{n}\left(\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)}\right) E\left[z^{X_n+\sum_{j=1}^{N} I_{C_n^j} A_{n+1}^j - I(X_n\neq 0)}\middle|\mathcal{F}_n\right] \quad a.s. \\ &= z^{X_n - I(X_n\neq 0)} \prod_{k=0}^{n}\left(\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)}\right) E\left[z^{\sum_{j=1}^{N} I_{C_n^j} A_{n+1}^j}\middle|\mathcal{F}_n\right] \quad a.s. \end{aligned}$$

Using the fact that the C_n^j are a disjoint, complete set for $j = 1, .., N$ we get

$$\begin{aligned} E\left[z^{\sum_{j=1}^{N} I_{C_n^j} A_{n+1}^j}\middle|\mathcal{F}_n\right] &= \sum_{i=1}^{N} I_{C_n^j} E\left[z^{A_{n+1}^j}\right] \quad a.s. \\ &= \sum_{i=1}^{N} I_{C_n^j} a_j(z) \quad a.s. \end{aligned}$$

and so

$$\begin{aligned} E\left[M_{n+1}(z)\middle|\mathcal{F}_n\right] &= z^{X_n} \prod_{k=0}^{n-1}\left(\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)}\right) \quad a.s. \\ &= M_n(z) \quad a.s. \end{aligned}$$

It is trivial to show that $M_n(z)$ is non-negative, therefore $E\left[|M_n(z)|\right] = E\left[M_n(z)\right] < \infty$ and hence the martingale is integrable. □

3.2.1 Stability and recurrence

In order for us to be able to find useful equilibrium results the processes investigated must be stable. Furthermore in order for the martingale results to be of use we shall relate them to the multi-phase MRP discussed in Section 2.4. To do this we require the process to be recurrent. Lemma 3.1.2 provides a useful result based on the stability of the M/G/1 queue and condition $(*)$ presented in the following section is a sufficient condition for stability but in general each situation must be considered on its merits.

The process should also be irreducible. That is, we want there to be no more than one communicating class. In some types of queueing process (see Section 3.4 on page 45) it is possible to have more than one communicating class. We wish to avoid these possibilities. More will be said about this in Section 3.3.1.

The obvious criterion of use here is simply to require that $\tau(n)$ be almost surely finite for all $n \in \mathbb{Z}^+$. This means that state 0 will recur within a finite time almost surely. Lemma 3.2.2 shows that it is sufficient to look at $\tau(0)$.

We need one further condition, that $\rho_1 > 0$. Simply stated, we require that the traffic intensity during the first phase to be positive. This is because the busy system enters phase one at the completion of each busy period. It cannot leave phase one until there has been at least one service completion. Thus, if ρ_1 were zero, there would be no arrivals and hence we would never leave state 0. There are other similar conditions which suffice, such as requiring that a maximum time can be spent in phase 1 before switching to a phase with positive traffic intensity but these violate some of our rules for phases, therefore we assume below that $\rho_1 > 0$.

Lemma 3.1.1 *Given $X_0 = 0$, if $\tau(0)$ is almost surely finite then $\tau(n)$ and $\tau_i(n)$ are also almost surely finite for $i = 1, \ldots, N$ and for all $n \in \mathbb{Z}^+$.*

Proof: When $X_0 = 0$, $\tau(0)$ being almost surely finite implies that the busy period is almost surely finite. From this $\tau(n)$ must be a.s. finite. This is because all of the states are reachable from state 0 in one transition. So, as $\tau(0)$ is almost surely finite, we return to state 0 in an almost surely finite time. Hence no matter what state the process is in it must return to state zero in an a.s. finite time. It is immediate that $\tau_i(n)$ must also be almost surely finite as $\tau_i(n) \leq \tau(n)$ almost surely. □

Lemma 3.1.2 *If $\rho_N > 1$ and $p\{\tau_N(0) - \tau_{N-1}(0) > 0\} > 0$ then the process is unstable.*

Proof: We take $\tau_{N-1}(0) < \infty$. In this case we have $p\{\tau_N(0) - \tau_{N-1}(0) > 0\} > 0$ so over a busy period there is a positive probability of spending time in phase N. Once in phase N the queue behaves as an M/G/1 queue. Thus the stability conditions of the M/G/1 queue apply. Hence for $\rho_N > 1$, $p\{\tau_N(0) < \infty\} < 1$ and hence the queue is unstable. □

This last condition arises from the fact that the last phase ends when the system empties. If $\rho_N > 1$ and the process is not already empty at the beginning of this phase then, with positive probability, the process may never become empty again, and hence the phase might not end.

3.2.2 Regularity of the stopping times

In order to uses Doob's Optional Sampling Theorem we must demonstrate that the stopping times involved are regular for the martingale. This would be trivial if the martingale were uniformly integrable, (from Neveu, IV-**3**-14) however it may not be. Thus we must investigate the conditions under which we can prove regularity.

Condition $(*)$ We take $\mathcal{S} = \{1, 2, \ldots, N\}$ and $\mathcal{S}^* \subset \mathcal{S}$ to be the set of all j with $\rho_j > 1$. The condition is that

$$E\left[\prod_{j\in\mathcal{S}^*} \xi_j(z)^{\tau_j(0)-\tau_{j-1}(0)}\right] < \infty,$$

for all $z \in [0,1]$. When $\mathcal{S}^* = \phi$ the condition is automatically satisfied.

When $\mathcal{S}^* = \{i\}$ so that $\rho_i > 1$ and $\rho_j \leq 1$ for all other j we can write this condition as

$$E\left[\alpha^{\tau_i(0)-\tau_{i-1}(0)}\right] < \infty,$$

where $\alpha = \sup_{z\in[0,1]} \xi_i(z)$. This will be the condition used in Chapter 4. Note that condition $(*)$ implies that $\tau(0)$ is almost surely finite, and hence the queue is stable.

Theorem 3.2 *If $(*)$ is satisfied then the stopping times $\tau_0(n), \ldots \tau_N(n)$ and $\tau(n)$ are regular for the martingale $M_n(z)$, $z \in [0,1]$, $n \in \mathbb{Z}^+$.*
Furthermore when $\tau(n) = \infty$

$$M_{\tau(n)}(z) = 0.$$

Proof: We wish the stopping times $\tau_0(n), \ldots \tau_N(n)$ and $\tau(n)$ to be regular for the martingale for $n \in \mathbb{Z}^+$. First we consider the case with $\mathcal{S}^* = \phi$. In this case Takács's lemma (C.4) implies that $\xi_j(z) \leq 1$ for all $j = 1, \ldots, N$. Hence

$$|M_n(z)| \leq 1,$$

which implies that $\left(M_n(z)\right)$ is uniformly integrable (page 134). As $\left(M_n(z)\right)$ is a positive integrable martingale, condition (a) of Neveu IV-**3**-14 is automatically satisfied. When $M_n(z)$ is uniformly integrable so too must $M_n(z)\, I(\tau > n)$ for all stopping times τ and so condition (b) of Neveu IV-**3**-14 is also satisfied. Hence any possible stopping time is regular in this case. Next we show that $\tau(0)$ is regular for the case when $\mathcal{S}^* = \{i\}$.

Lemma 3.2.1 *For $\rho_j \leq 1$, $\forall j \neq i$, and $\rho_i > 1$, if $E\left[\alpha^{\tau_i(0)-\tau_{i-1}(0)}\right] < \infty$ where $\alpha = \sup_{z\in[0,1]} \xi_i(z)$, then $\tau(0)$ is regular for the martingale $M_n(z)$, $z \in [0,1]$.*

Proof: We use Neveu IV-**3**-16. Condition (1) of this proposition,

$$\int_{\{\tau(0)<\infty\}} |M_{\tau(0)}(z)| dP < \infty,$$

is automatically satisfied for our martingale. Condition (2),

$$\lim_{n\to\infty} \int_{\{\tau(0)>n\}} |M_n(z)| dP = 0,$$

is satisfied as follows. Noting that the martingale is non-negative we start with $n > 0$ from

$$|M_n(z)| \;=\; z^{X_n} \prod_{k=0}^{n-1} \frac{z^{I\{X_k\neq 0\}}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)},$$

which because $\tau(0) > n$ gives

$$\begin{aligned}
|M_n(z)| &\leq \left(\prod_{k=0}^{(\tau_1(0)\wedge n)-1} \xi_1(z)\right)\left(\prod_{k=\tau_1(0)\wedge n}^{(\tau_2(0)\wedge n)-1} \xi_2(z)\right)\cdots\left(\prod_{k=\tau_{N-1}(0)\wedge n}^{(\tau_N(0)\wedge n)-1} \xi_N(z)\right) \\
&= \xi_1(z)^{(\tau_1(0)\wedge n)}\xi_2(z)^{(\tau_2(0)\wedge n)-(\tau_1(0)\wedge n)}\cdots\xi_N(z)^{n-(\tau_{N-1}(0)\wedge n)} \\
&\leq \alpha^{(\tau_i(0)\wedge n)-(\tau_{i-1}(0)\wedge n)} \\
&\leq \alpha^{\tau_i(0)-\tau_{i-1}(0)},\ a.s., \qquad (3.3)
\end{aligned}$$

as $\alpha = \sup_{z\in[0,1]} \xi_i(z) > 1$. Now due to the almost sure finiteness of $\tau(0)$ implied by $(*)$

$$\lim_{n\to\infty} I(\tau(0) > n) \;=\; 0 \;\; a.s. \qquad (3.4)$$

Thus (3.3) and (3.4) imply $|M_n(z)|\, I(\tau(0) > n)$ tends to 0 almost surely as n tends to infinity. Also from (3.3) we get

$$|M_n(z)|\, I(\tau(0) > n) \;\leq\; \alpha^{\tau_i(0)-\tau_{i-1}(0)},\ a.s.$$

the right-hand side of which has finite expectation by the assumption. Thus we can use the Dominated Convergence Theorem to show

$$\lim_{n\to\infty} E\left[|M_n(z)|\, I(\tau(0) > n)\right] = E\left[\lim_{n\to\infty} |M_n(z)|\, I(\tau(0) > n)\right] = 0,$$

from which we get condition (2) and thence the result. The latter part of the theorem follows also from Neveu IV-**3**-16. □

The generalisation of $\mathcal{S}^*$ is done in the same manner as the previous lemma with the substitution of the more general condition.

Lemma 3.2.2 *If $\tau(0)$ is regular for the martingale then $\tau(n)$ and $\tau_i(n)$ are also regular for the martingale for $i = 1, \ldots, N$ and for all $n \in \mathbb{Z}^+$..*

Proof: Given that $\tau(n)$ is regular, Neveu IV-**3**-13 implies that $\tau_i(n)$ must also be regular. All we need to show now is that the regularity of $\tau(0)$ implies the regularity of $\tau(n)$.

$$\begin{aligned} M_{\tau(n)}(z) &= z^{X_{\tau(n)}} \prod_{k=0}^{\tau(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \\ &= z^{X_{\tau(n)}} \prod_{k=0}^{\eta(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \prod_{k=\eta(n)}^{\tau(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right), \end{aligned}$$

where $\eta(n) = \sup\{m \leq n | X_m = 0\}$, the epoch of the beginning of the current busy period. Now the latter product in this equation is the product over one busy period. Due to the regenerative nature of this process at time $\eta(n)$, for $m \geq n$

$$\begin{aligned} & E\left[I(\tau(n) > m)\, z^{X_{\tau(n)}} \prod_{k=\eta(n)}^{\tau(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right] \\ &= E\left[I(\tau(0) > m - \eta(n))\, z^{X_{\tau(0)}} \prod_{k=0}^{\tau(0)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right] \end{aligned}$$

and also

$$\prod_{k=0}^{\eta(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \text{ and } I(\tau(n) > m)\, z^{X_{\tau(n)}} \prod_{k=\eta(n)}^{\tau(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right)$$

are independent. Thus for $m \geq n$ we can write

$$\begin{aligned} & E\left[|M_{\tau(n)}(z)| I(\tau(n) > m) \right] \\ &= E\left[\prod_{k=0}^{\eta(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right] E\left[I(\tau(n) > m)\, z^{X_{\tau(n)}} \prod_{k=\eta(n)}^{\tau(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right] \\ &= E\left[\prod_{k=0}^{\eta(n)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right] E\left[I(\tau(0) > m - \eta(n))\, z^{X_{\tau(0)}} \prod_{k=0}^{\tau(0)-1} \left(\frac{z^{I(X_k \neq 0)}}{\sum_{j=1}^{N} I_{C_k^j} a_j(z)} \right) \right]. \end{aligned}$$

Now for fixed, finite n the first expectation is clearly finite and as m tends to infinity the second equals $\lim_{n \to \infty} E\left[I(\tau(0) > m) M_{\tau(0)}(z) \right]$ which tends to zero as m tends to infinity from Lemma 3.2.1. So we have our result. □
This concludes the proof of Theorem 3.2. □

From here on we shall refer to the stopping time γ. By this we shall mean one of the stopping times n, $\tau_0(n), \ldots \tau_N(n)$ or $\tau(n)$ for $n \in \mathbb{Z}^+$. Theorems which are said to be true for stopping times γ are also true for each of these stopping times. This is expressed in the following definition.

Definition 3.1 *The stopping time γ will refer to each of the stopping times $\tau_0(n), \ldots \tau_N(n)$ or $\tau(n)$ for $n \in \mathbb{Z}^+$.*

Theorem 3.3 *If $(*)$ is satisfied then the stopping times $\tau_0(\gamma), \ldots, \tau_N(\gamma)$ and $\tau(\gamma)$ are regular for the martingale $M_n(z)$, $z \in [0,1]$ and γ defined in Definition 3.1. Furthermore when $\tau(\gamma) = \infty$*

$$M_{\tau(\gamma)}(z) = 0 \ \ a.s.$$

Proof: We proceed as in Lemma 3.2.1. We satisfy the conditions of Neveu's Proposition IV-**3**-16. As before Condition (1) is automatically satisfied. Condition (2),

$$\lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} |M_n(z)|\, dP = 0,$$

is satisfied as follows.

$$\begin{aligned}
\lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} |M_n(z)|\, dP &= \lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} \left[I(\gamma \le n) + I(\gamma > n)\right] M_n(z)\, dP \\
&= \lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} I(\gamma \le n) M_n(z)\, dP \\
&\quad + \lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} I(\gamma > n) M_n(z)\, dP \\
&= \lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} I(\gamma \le n) M_n(z)\, dP \\
&\quad + \lim_{n\to\infty} \int_{\{\gamma>n\}} M_n(z)\, dP. \qquad (3.5)
\end{aligned}$$

When $\gamma = \tau(m)$ the second term of (3.5) becomes

$$\lim_{n\to\infty} \int_{\{\tau(m)>n\}} M_n(z)\, dP = 0, \qquad (3.6)$$

since we know from Theorem 3.2 that $\tau(m)$ is regular for the martingale and must satisfy condition (2) of Neveu IV-**3**-16. Hence from (3.5) and (3.6) we get

$$\begin{aligned}
\lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} |M_n(z)|\, dP &= \lim_{n\to\infty} \int_{\{\tau(\gamma)>n\}} I(\gamma \le n) M_n(z)\, dP \\
&= \lim_{n\to\infty} \int_{\{\gamma \le n < \tau(\gamma)\}} M_n(z)\, dP,
\end{aligned}$$

when $\gamma = \tau(m)$. In this integral $\tau(m) \le n$ and so we can write

$$M_n(z) = \left[\prod_{k=0}^{\tau(m)-1} \left(\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^N I_{C_k^j} a_j(z)}\right)\right] \left[z^{X_n} \prod_{k=\tau(m)}^{n-1} \left(\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^N I_{C_k^j} a_j(z)}\right)\right].$$

As in Lemma 3.2.2 the two parts of this product are independent due to the regenerative nature of the process. Furthermore the regularity of $\tau(m)$ implies that the former term has finite expectation. The expectation of the second part can be shown to approach zero as n tends to infinity by exactly the same method as is used to demonstrate condition (2) of Neveu IV-**3**-16 in Lemma 3.2.1.

Thus the theorem is proven for $\gamma = \tau(m)$. When $\gamma = \tau_i(m)$ we can see that $\tau(\tau(m)) \ge \tau(\tau_i(m))$ almost surely and so from Neveu IV-**3**-13 we get the regularity of all $\tau(\tau_i(m))$. This is the result we need. The latter part of the theorem again comes directly from Neveu IV-**3**-16. □

3.2.3 Use of the Optional Sampling Theorem

Theorem 3.4 *For $z \in [0,1)$, γ as in Definition 3.1 and with $(*)$ satisfied*

$$E\left[\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma}z^{I(X_\gamma=0)},\;\; a.s.$$

Proof: Consider Doob's Optional Sampling Theorem (A.7) with stopping times γ and $\tau(\gamma)$. We know $\gamma \le \tau(\gamma)$ a.s. and condition $(*)$ gives the regularity of these stopping times through Theorem 3.3 and so the following is true

$$E\left[M_{\tau(\gamma)}(z)\middle|\mathcal{F}_\gamma\right] \;=\; M_\gamma(z),\;\; a.s.$$

Rewritten this is

$$E\left[z^{X_{\tau(\gamma)}}\prod_{k=0}^{\tau(\gamma)-1}\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma}\prod_{k=0}^{\gamma-1}\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)},\;\; a.s.$$

There are two possibilities: $\tau(\gamma)=\infty$ in which case $M_{\tau(\gamma)}(z)=0$ from Theorem 3.3 or $\tau(\gamma)<\infty$ in which case $X_{\tau(\gamma)}=0$. The former case can make no contribution to the expectation so by using the fact that $\prod_{k=0}^{\gamma-1}\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}$ is $\mathcal{F}_\gamma$-measurable we get

$$E\left[\prod_{k=\gamma}^{\tau(\gamma)-1}\frac{z^{I(X_k\neq 0)}}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma},\;\; a.s.$$

It is clear that $X_k\neq 0$ for $k=\gamma+1$ to $\tau(\gamma)-1$ from the definition of $\tau(\gamma)$. Thus

$$E\left[\frac{z^{I(X_\gamma\neq 0)}}{\sum_{j=1}^{N}I_{C_\gamma^j}a_j(z)}\prod_{k=\gamma+1}^{\tau(\gamma)-1}\frac{z}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma},\;\; a.s.$$

As $z^{I(X_\gamma\neq 0)}$ is also $\mathcal{F}_\gamma$-measurable we can write this as

$$z^{I(X_\gamma\neq 0)-1}E\left[\prod_{k=\gamma}^{\tau(\gamma)-1}\frac{z}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma},\;\; a.s.$$

$$E\left[\prod_{k=\gamma}^{\tau(\gamma)-1}\frac{z}{\sum_{j=1}^{N}I_{C_k^j}a_j(z)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{1-I(X_\gamma=0)}z^{X_\gamma},\;\; a.s.$$

We note that $1-I(X_\gamma\neq 0)=I(X_\gamma=0)$ and

$$I_{C_k^j}=\begin{cases}1, & \tau_{i-1}(\gamma)\le k<\tau_i(\gamma),\\ 0, & \text{otherwise,}\end{cases}$$

and so we get

$$E\left[\prod_{j=1}^{N}\left(\frac{z}{a_j(z)}\right)^{\tau_j(\gamma)-\tau_{j-1}(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{I(X_\gamma=0)}z^{X_\gamma},\;\; a.s.$$

and by substituting the definitions of $\nu_j(\gamma)$ and $\xi_j(z)$ this gives

$$E\left[\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{I(X_\gamma=0)}z^{X_\gamma},\;\; a.s.$$

which is the required result. □

Remark: Note that at this point we may multiply by any $\mathcal{F}_\gamma$-measurable random variable such as $I(\tau(0)>\gamma)$ to get

$$E\left[I(\tau(0)>\gamma)I(\gamma<\infty)\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; I(\tau(0)>\gamma)I(\gamma<\infty)z^{I(X_\gamma=0)}z^{X_\gamma},\;\; a.s. \tag{3.7}$$

Theorem 3.5 *For $z\in[0,1)$, γ as in Definition 3.1 and condition $(*)$ satisfied*

$$E\left[z^{X_\gamma}\right] \;=\; E\left[\prod_{j=1}^{N}\xi_j(z)^{\mu_j(\gamma)}\right].$$

Proof: From Theorem 3.4 we get

$$E\left[\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; z^{X_\gamma}z^{I(X_\gamma=0)},\;\; a.s.$$

Due to the $\mathcal{F}_\gamma$-measurability of X_γ we may multiply both sides of the equation by $I(X_\gamma\neq 0)$ and write

$$E\left[I(X_\gamma\neq 0)\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\middle|\mathcal{F}_\gamma\right] \;=\; I(X_\gamma\neq 0)z^{X_\gamma}z^{I(X_\gamma=0)},\;\; a.s.$$

We may then take expectations of this equation

$$E\left[I(X_\gamma\neq 0)\prod_{j=1}^{N}\xi_j(z)^{\nu_j(\gamma)}\right] \;=\; E\left[I(X_\gamma\neq 0)z^{X_\gamma}\right].$$

Noting that $X_\gamma\neq 0$ implies that $\mu_j(\gamma)=\nu_j(\gamma)$ for $j=1,\ldots,N$ and adding $p\{X_\gamma=0\}$ to both sides gives

$$p\{X_\gamma=0\}+E\left[I(X_\gamma\neq 0)\prod_{j=1}^{N}\xi_j(z)^{\mu_j(\gamma)}\right] \;=\; p\{X_\gamma=0\}+E\left[I\{X_\gamma\neq 0\}z^{X_\gamma}\right].$$

The right-hand side is equal to $E\left[z^{X_\gamma}\right]$. The events

$$[X_\gamma=0]\Leftrightarrow[\mu_j(0)=0,\; j=1,..,N],$$

are equivalent, by (3.2) and so

$$[X_\gamma \neq 0] \Leftrightarrow [\mu_j(0) \neq 0, \text{ for some } j],$$

which implies

$$p\{\mu_j(n) = 0, \ \forall j\} + E\left[I\{\exists j, \ st: \ \mu_j(n) \neq 0\} \prod_{j=1}^{N} \xi_j(z)^{\mu_j(\gamma)}\right] \quad = \quad E\left[z^{X_\gamma}\right],$$

and so finally

$$E\left[\prod_{j=1}^{N} \xi_j(z)^{\mu_j(\gamma)}\right] \quad = \quad E\left[z^{X_\gamma}\right],$$

which is the desired result. □

Corollary 3.6 *For $z \in [0,1)$, condition $(*)$ satisfied and $l > 1$*

$$E\left[\prod_{j=l}^{N} \xi_j(z)^{\nu_j(0)}\right] \quad = \quad E\left[z^{X_{\tau_{l-1}(0)}}\right].$$

Proof: The proof is simply a matter of putting $\gamma = \tau_{l-1}(0)$ in the preceding theorem and noting that we assume $X_0 = 0$ and so

$$\mu_j(\tau_{l-1}(0)) = \begin{cases} \nu_j(0), & j \geq l, \\ 0, & j < l. \end{cases}$$

□

Remark: Note that the case with $l = 1$ is excluded as when $l = 1$, $\tau_{l-1}(0) = \tau_0(0) = 0$ and $\mu_j(0) = 0$ because $X_0 = 0$. Thus the result would not hold. Instead we resort to the following theorem.

Theorem 3.7 *For $z \in [0,1)$ and condition $(*)$ satisfied*

$$E\left[\prod_{j=1}^{N} \xi_j(z)^{\nu_j(0)}\right] \quad = \quad z.$$

Proof: Simply taking $\gamma = 0$ in Theorem 3.4 and taking expectations gives us

$$E\left[\prod_{j=1}^{N} \xi_j(z)^{\nu_j(0)}\right] \quad = \quad E\left[z^{X_0} z^{I\{X_0=0\}}\right],$$

which, as we assume $X_0 = 0$ gives

$$E\left[\prod_{j=1}^{N} \xi_j(z)^{\nu_j(0)}\right] \quad = \quad z.$$

□

3.2.4 Martingale arguments for stability

In this section we examine martingale stability arguments. From Lemma 3.2.2 it is sufficient to consider $\tau(0)$ when $X_0 = 0$. As before it is sufficient to discuss the behaviour of $\tau(0)$ when $X_0 = 0$. When condition $(*)$ is satisfied we know that $p\{\tau(0) < \infty\} = 1$. Thus condition $(*)$ is a sufficient condition for the stability of the queueing process (and hence the recurrence of the MRP).

It would be nice to have necessary and sufficient conditions for stability as Baccelli and Makowski provide in the examples they have considered. Their results however revolve around the following type of technique. They provide a traffic intensity ρ for the system of interest and then the condition for stability is simply that $\rho \leq 1$. The traffic intensity in the M/G/1 case is the standard intensity λ/μ (Baccelli and Makowski, 1985). In the case with Markov modulated input the traffic intensity is a weighted sum of the intensities during each of the input states (Baccelli and Makowski, 1986).

We can see that this approach would give us a traffic intensity in our type of process as well. It is not, however, easy to see how this would benefit us in this case. In our type of process we could take the approach of setting

$$\rho = \sum_{j=1}^{N} \rho_j \phi_j,$$

where ϕ_j is the probability of finding the system in phase j during equilibrium. In order to have the probabilities p_j we must assume the existence of the equilibrium solution. Thus any argument based on this would be inherently circular. This occurs because the time spent in phase j may be strongly dependent on ρ_j. Thus we may have a situation where during one of the phases ρ_j is very large but a phase shift occurs if too many customers arrive during one service and so the length of the phase is very small, so the two balance. We may not however assume this balance.

It is thus not clear as yet how to provide the type of elegant stability criterion that is commonly used in many other situations. It is to be strongly suspected that condition $(*)$ is related to the necessary conditions. It would not be surprising if condition $(*)$ is in fact also a necessary condition for stability. One approach to this problem would be to use Rosenkrantz (1989) which deals with ergodicity conditions for two-dimensional Markov chains. Our queueing processes can be represented by a two-dimensional Markov chain by taking the number of customers in the system and the phase to be the two variables. If the conditions of Rosenkrantz could be related to condition $(*)$ this might provide the desired result.

Another related question is that of null recurrence. Even though $\tau(0) < \infty$ almost surely we may still have

$$E[\tau(0)] = \infty.$$

In other words the length of the busy period is almost surely finite but the mean length of the busy period is infinite, this is the null recurrent case. Although in the null recurrent case our martingale arguments will still work the equilibrium results will be inherently uninteresting. We shall not consider these cases in most examples.

3.3 Relationship with the MRP

The connection between the multi-phase MRP and the multi-phase M/G/1 queueing process will by now be obvious. Each phase of the queueing process is associated with a state in the multi-phase MRP. Transitions between states in the MRP are at the same times as changes in phase in the queueing process. Because phases change only at the end of services we use the embedded process and consequently the multi-phase MRP has discrete or lattice time. We must resort to a generalised multi-phase MRP because there is no requirement that the time spent in each phase be independent of the times spent in each of the other phases. However, we do assume that the times when the system is left empty do constitute renewals. This means the times spent in phases during different busy periods must be independent. Also because we assume the queue begins with a dummy service leaving it empty ($X_0 = 0$) we have a non-delayed renewal process.

Given this description it becomes clear that the $\mu_j(n)$ are forward recurrence times in the multi-phase MRP as can be seen in Lemma 2.2.1, and the $\nu_j(0)$ are sojourn lifetimes in the multi-phase MRP. With this relationship

$$\begin{aligned} Q^*(\xi_1(z), \xi_2(z), \ldots, \xi_N(z)) &= \lim_{n\to\infty} E\left[\prod_{j=1}^{N} \xi_j(z)^{\mu_j(n)}\right], \\ F^*(\xi_1(z), \xi_2(z), \ldots, \xi_N(z)) &= E\left[\prod_{j=1}^{N} \xi_j(z)^{\nu_j(0)}\right], \\ F_l^*(\xi_l(z), \xi_{l+2}(z), \ldots, \xi_N(z)) &= E\left[\prod_{j=l}^{N} \xi_j(z)^{\nu_j(0)}\right], \end{aligned}$$

where Q^* is defined in (2.20) and F^* and F_l^* are as defined in Section 2.4. Thus Theorems 3.5 and 3.7 and Corollary 3.6 imply respectively that

$$Q^*(\xi_1(z), \xi_2(z), \ldots, \xi_N(z)) = \lim_{n\to\infty} E\left[z^{X_n}\right], \tag{3.8}$$

$$F^*(\xi_1(z), \xi_2(z), \ldots, \xi_N(z)) = z, \tag{3.9}$$

$$F_l^*(\xi_l(z), \xi_{l+1}(z), \ldots, \xi_N(z)) = E\left[z^{X_{\tau_{l-1}(0)}}\right]. \tag{3.10}$$

From these we deduce the following

Theorem 3.8 *For $z \in [0,1)$ and $(*)$ satisfied*

$$\lim_{n\to\infty} E\left[z^{X_n}\right] = \frac{1}{m}\sum_{l=1}^{N}\left[\frac{F_{l+1}^*(\xi_{l+1}(z), \ldots, \xi_N(z)) - F_l^*(\xi_l(z), \ldots, \xi_N(z))}{1-\xi_l(z)}\right],$$

where m acts as a normalising constant.

Proof: The result comes directly by substituting (3.8) in Theorem 2.4. □

Theorem 3.9 *For $z \in [0,1)$ and $(*)$ satisfied*

$$E\left[z^X\right] = \frac{1}{m}\left[\sum_{l=2}^{N}\frac{E\left[z^{X_{\tau_l(0)}}\right] - E\left[z^{X_{\tau_{l-1}(0)}}\right]}{1-\xi_l(z)} + \frac{E\left[z^{X_{\tau_1(0)}}\right] - z}{1-\xi_1(z)}\right],$$

where m acts as a normalising constant and $X(t) \to X$ almost surely as $t \to \infty$.

Proof: We substitute (3.9) and (3.10) into the preceding theorem and then use the dominated convergence theorem (A.3), PASTA (page 143) and Theorem C.1 to see that

$$\lim_{n\to\infty} E\left[z^{X_n}\right] = \lim_{t\to\infty} E\left[z^{X(t)}\right].$$

From this we get the result. □

3.3.1 Discussion

A number of points deserve some further discussion before we continue on to some examples. The first point to note is that we have looked only at the system size distribution. From this we may use the arguments of Section C.5 to calculate the waiting time distributions for a first-in, first-out (FIFO) queue. For other service disciplines this may be more difficult.

Also it is of some interest to consider the origin of these results. We have considered three processes on different time scales: the queueing processes itself, a discrete-time queueing process embedded at departure epochs and a further process embedded in this at the epochs of phase changes.

Given sufficient conditions on the queueing process considered, one interpretation of the Optional Sampling Theorem is that the process $M_{Q_n}(z)$ (where $Q_{nN+j} = T_n^j$) is also a martingale and from this we derive our relationships. The epochs Q_n are the transition epochs of the multi-phase MRP of Chapter 2 and so we can obtain limiting formulae using the renewal techniques.

The results of this chapter are quite general but may be extended. The following section documents some of the ways in which the results can be generalised still further. Two major extensions are proposed. The first considers the server's behaviour. We have until now assumed that only the service-time distribution can change between phases but there are other types of behaviour that can be used. The second proposal (in Section 3.5) concerns the rules which limit the types of phase transitions allowed.

It is also worth noting here that many processes which might not appear to have the required phase structure actually can be considered in this mold. This is considered in Section 3.6.

3.4 Generalisations of server behaviour

In the processes considered so far we have considered service times. We have said that the service-time distribution changes between phases. There are, however, other plausible models with which this technique deals with equal facility.

We can vary a whole range of server behaviour between phases without affecting the results obtained thus far. The results we have obtained need only the probability generating function for the number of arrivals during a single service in each phase. So we may, for instance, change service discipline between phases with no alteration of the results (so long as the discipline remains non-preemptive).

In the following we elucidate a number of examples. We give the relevant probability generating function and some motivation for each example. Throughout this discussion we mean by blocked that a customer is either lost or rerouted. Customers who are blocked do not return for service at a later time.

(i)

During phase j a customer waits a random period of time before beginning service. If the extra time has probability distribution function $B^j(\cdot)$ then

$$a_j(z) = A^{j*}(\lambda(1-z))B^{j*}(\lambda(1-z)).$$

This is an example of the service time for a customer being extended and so the probability distribution function for the service time is a convolution of $A^j(\cdot)$ and $B^j(\cdot)$. Thus the Laplace-Stieltjes transform of the new service time is the product $A^{j*}(s)B^{j*}(s)$ and hence from Theorem C.3 the result.

This type of behaviour is likely to occur if the service time of a customer who arrives at an empty server is different from the service time of a customer who arrives at a busy server. This might occur if the server has some warmup time when it starts up at the beginning of the busy period or if the server can take vacations when unoccupied. In this case the modified behaviour occurs in the first phase which lasts but one service time.

(ii)

During phase j a customer waits until M_j arrivals have occurred before commencing service. In this case

$$a_j(z) = z^{M_j} A^{j*}(\lambda(1-z)).$$

This also might occur for one service at the start of the busy period but for a different reason. If the server has some overhead associated with starting and stopping service it is better to reduce the frequency of these events. To do this the busy period must be increased in length. One way to do this is to allow a backlog to build up before beginning service. This then minimises the number of times the server switches from idle to busy states. This is called the N-policy queue by Neuts (1989).

(iii)

During phase j only the first N_j arrivals during any particular service are allowed to join the queue, any further arrivals being blocked. In this case

$$\begin{aligned} a_j(z) &= \sum_{i=0}^{N_j-1} a_i^j z^i + z^{N_j} \sum_{i=N_j}^{\infty} a_i^j \\ &= \sum_{i=0}^{N_j-1} a_i^j z^i + z^{N_j} \left(1 - \sum_{i=1}^{N_j-1} a_i^j\right) \\ &= \sum_{i=0}^{N_j-1} a_i^j \left(z^i - z^{N_j}\right) + z^{N_j}, \end{aligned}$$

where a_i^j is the probability of i arrivals occurring during one service time in phase j.

This also might occur as part of a control strategy. However the reason for using such a strategy might be to limit the number of customers in the system (thus minimising waiting times).

(iv)

During phase j arrivals are blocked with probability p_j. The arrivals are still Poisson with new rate λp_j and so

$$a_j(z) = A^{j*}(\lambda p_j(1-z)).$$

This could also occur as part of a control strategy. For instance if the arrival process is the superposition of several independent Poisson streams with rates λ_i, it will itself be a Poisson stream with rate $\sum \lambda_i$. If these streams are then assigned different priorities then we can block some of the streams on the basis of their priority during phase j in order to limit congestion for the higher priority arrivals. This would give us the situation above.

(v)

Batch arrivals. If the batch size is given by the random variable B where b_i is defined by

$$b_i = p\{B = i\},$$

then

$$a_j(z) = A^{j*}(\lambda[1 - B(z)]),$$

where $B(z)$ is the probability generating function for the batch sizes. This is from Jun Huek Park (1990) in which Baccelli and Makowski's technique has been shown to work for the simple $M^B/G/1$ queue. The derivation from Jun Huek Park follows.

$$\begin{aligned} a_j(z) &= \sum_{k=0}^{\infty} P(A_n = k) z^k dA^j(t) \\ &= \int_0^{\infty} \sum_{k=0}^{\infty} \sum_{m=0}^{k} \frac{e^{-\lambda t}(\lambda t)^m}{m!} b_k^{(m)} z^k dA^j(t) \end{aligned}$$

$$= \int_0^\infty \sum_{k=0}^{\infty} \sum_{m=0}^{k} e^{-\lambda t} e^{\lambda t B(z)} dA^j(t)$$
$$= A^{j*}(\lambda(1 - B(z)),$$

where $b_k^{(m)}$ is the m-fold convolution of $b(k)$ with itself. Note that for such a system the initial service of the busy period will also have a slightly different form. Thus for the simple batch arrival queue

$$X_{n+1} = X_n + B_n I(X_n = 0) + A_{n+1} - 1,$$

where B_n is the batch size of the first arrival during a busy period after the nth service if $X_n = 0$.

(vi)

In all but (ii) of the above examples the service times are still independent of the arrival epochs. This is an important assumption and may only be relaxed with care. An example of a case in which this assumption does not hold would be if the server terminates service only after N arrivals have occurred during the service. This example could cause problems. For example a server that terminates service after exactly one arrival would result in the process sticking in a certain state. Other problems such as reducibility of the state space or periodicity of the states could occur given this type of service. We shall avoid these possibilities throughout with the exception of case (ii). We allow this in the situation described for its use, that is, for one service at the beginning of the busy period. In this case it cannot cause any problems for regularity. Further problems can occur in PASTA (page 143) if this independence is not maintained.

3.4.1 Blocking versus zero service time

We have considered the possibility that $\rho_j > 1$ in our discussion of regularity, and these cases must be dealt with on an individual basis. Mostly it will be obvious from the stability of the queue when the stopping times are regular. However we have not considered the case when $\rho_j = 0$. Given a normal service with probability distribution function $A^j(\cdot)$ this would imply that $\lambda \int_0^\infty (1 - A^j(t))dt = 0$. One way this can occur is if the mass of $A^j(\cdot)$ is concentrated at zero, that is the services take zero time with probability one. This could be the case if the server serves the customers in the queue instantaneously. We shall refer to this as discarding a customer. The customers are discarded in the order of service. Thus in a FIFO queue the customer at the front of the queue is discarded. This suggests that an alternative service discipline might be used in different phases. For example FIFO in the normal phases and LIFO in the phase with zero service time, in order to discard customers from the end of the queue.

This is different from the other situation in which $\rho_j = 0$. This is the case when $\lambda = 0$ or all arriving customers are blocked. If all of the customers arriving at the queue during a phase are blocked upon arrival, then the arrival rate becomes zero and traffic intensity also becomes zero. The service times, however, can still be positive.

It is noteworthy that the case with blocking does not immediately lead to the solution for the M/G/1 queue with a limited waiting room. This is the normal model

chosen for such a system but in our model we are limited to changing phase at the ends of services. Thus the limit of the waiting room might be reached and exceeded during a service without any blocking occurring. This could be overcome with difficulty by having a separate phase for each state of the queueing system with each phase allowing only a certain number of arrivals chosen to make sure the limit of the waiting room was not exceeded. We can however model the M/G/1 queue with a limited waiting room using the case with zero service times. This is discussed in Section 7.5.

We might note also that although there are subjective differences in what we may mean by having $\rho_j = 0$ the resulting solution is identical. Thus once we decide upon the model being used we may then ignore this for all further purposes.

These possibilities, with $\rho_j = 0$, may occur in a sensible fashion, as is described above, but there is a case where it does not. If $\rho_1 = 0$ and the queue starts empty the queue will remain empty forever. While this case is not impossible it is clearly trivial. Such trivial cases are easily avoided and so we shall say no more about them here. Throughout the rest of this we shall assume that they are excluded from any discussion.

Finally we make the observation that while customers who are blocked never enter the queue, they are still considered to have entered the system. Thus our results, which are arrived at from the equilibrium distribution that departing customers see, will include the number of customers left in the system by a departing customer that is blocked as well as the number left in the system by a customer who receives service.

3.4.2 Later modifications

Later in this text we shall present a number of examples. For the most part we shall assume the varying service-time description of the processes. We shall consider the other cases only briefly. However, once the modified form of the generating functions $a_j(z)$ is noted the only major difference in the computations occurs in the calculation $E\left[z^{\tau_j(0)}\right]$ for $j = 1, \ldots, N-1$ and then only in some cases. We shall give more details in the examples.

3.5 Generalisations of the phases

To understand possible generalisations of the phase structure of these processes we must first understand why we have chosen the restrictions on phase structure. We have chosen the rules governing the phases so that the phase transitions are as general as possible while still allowing us to work with our model. To this end the rules restrict the behaviour of the phases. Rule (i) is a result of considering the embedded process. We desire the restriction in order that the embedded process's behaviour characterises that of the queueing process. Rule (ii) is necessary in order that Doob's Optional Sampling Theorem can be applied at the relevant time points. Rules (iii) and (iv) are required so that the ends of busy periods are renewal points of the process. Finally we require rules (v) and (vi). These are required in order that the structure of the renewal process we consider is that of a multi-phase MRP.

Rule (ii) is therefore crucial to the whole idea of using a martingale argument. Rules (iii) and (iv) are requirements for the Markov renewal results. Thus we must retain these rules in all of the generalisations that can be considered.

Rules (v) and (vi) are required to enforce the multi-phase nature of the renewal process. If we could obtain an equivalent result to that of Theorem 2.4 for generalised Markov renewal processes that are not of the multi-phase form we could modify or remove these two rules. This is suggested as a direction for continued work in Section 8.1. However, in the next section we shall see that this is not a necessary area of expansion as equivalent multi-phase Markov renewal processes can be found for generalised Markov renewal process.

One possible area of expansion that has not yet been considered is processes which change behaviour between service completions. An example of this is the single server queue with Markov modulated Poisson input. This has been considered in Baccelli and Makowski (1986,1991) using a modification of their technique for the M/G/1 queue. Thus we can expect that it will be possible to modify the results herein to cover such cases. If we were to consider this type of process in our model it violates rules (i), (iv), (v) and (vi). However we have an extra condition which is simply that the time spent in each phase is independent of the times spent in all other phases. This will require some further work before it can be dealt with through the multi-phase method, if it can.

A more profitable approach would be to modify the multi-phase technique in the same way that Baccelli and Makowski modify the simple technique for the M/G/1 queue to cover Markov modulated arrivals. This also is mentioned in the section on possible further work.

We have mentioned a number of relaxations of the phase transition rules which can be considered. These have been left for future work as they require a great deal of theoretical work before becoming practical and because the resulting complexity of the theory might make this thesis somewhat unwieldy. The following section presents a more fruitful way of extending the systems that can be considered.

3.6 Infinitely-many phases

We have until now only considered systems with a finite number of phases. Now we consider some reasons for considering processes with an infinite number of phases. Some cases with more complex structure require that we modify how we consider them in order that they fit into the form of sequential phases each occurring once during the busy period. A simple example is one where one or more phases do not necessarily occur during every busy period. We deal with this by inserting transitions through the missed phase which spend zero time in the phase. An infinite number of phases can easily be dealt with in this case as long as the original process under consideration has only a finite number of phases occur during the busy period, with probability one.

A more difficult example is when a phase may be entered more than once during the busy period. This may be modelled by considering each subsequent entry into this phase during a single busy period to be a new phase. Clearly in cases where two phases may alternate an infinite number of times before the end of the busy period, this results in an infinite number of phases. In Chapter 7 we shall see an example of a situation in which this occurs. The two phases alternate for ever with probability zero and so the processes recurrence is not adversely effected by this.

Note that when we use this procedure the probability generating functions $a_j(z)$ (and hence $\xi_j(z)$) will be the same for a number of the new phases. This will allow a great simplification in the problems using this technique.

In all of these cases we simply generalise the results of this chapter replacing N with infinity.

3.7 A single-phase example

In this section we consider the simplest example of this type, the M/G/1 queue. This is one of the problems that this technique was originally applied to by Baccelli and Makowski (1989). Thus the results here are exactly the same as theirs with some slight modifications due to the notation. The solution to the ergodic M/G/1 queue is well known and can be derived by a number of means (Cooper (1972)). It is given by

$$E\left[z^X\right] = (1-\rho)\frac{a(z)(1-z)}{a(z)-z}, \tag{3.11}$$

where $a(z)$ is the probability generating function for the number of arrivals during a service.

Note that in our notation this is a single-phase M/G/1 queue and the result is obtained directly from Theorem 3.9 with $N = 1$. It is simply

$$\begin{aligned} E\left[z^X\right] &= \frac{1}{m}\left[\frac{1-z}{1-\xi_1(z)}\right] \\ &= \frac{1}{m}\left[\frac{a_1(z)(1-z)}{a_1(z)-z}\right], \end{aligned}$$

where $m = 1/(1-\rho_1)$, which is the expected answer. It is worth noting that Theorem 3.7 gives

$$F^*(\xi_1(z)) = z. \tag{3.12}$$

This may then be used both to calculate m and to provide the generating function for the number of customers served during the busy period (Baccelli and Makowski, 1989). Namely for each $\xi \in [0,1)$ the equation in the unknown variable z

$$z = \xi a(z),$$

has the unique solution $Z(\xi)$ in the interval $[0,1]$. From Baccelli and Makowski (2.14)

$$F^*(y) = Z(y). \tag{3.13}$$

Chapter 4

Two-phase examples

In this chapter we consider the simplest non-trivial case of the type of process described in Chapter 3. This is the case with only two phases and hence two possible service-time distributions $A^1(\cdot)$ and $A^2(\cdot)$. As in the general theory of Chapter 3 the times at which the service-time distributions switch must be stopping times. Also the two phases each occur exactly once during a busy period and always occur in the same order. We call the point at which the transition from phase 1 to 2 occurs a threshold. When the queue is empty we start with service-time distribution $A^1(\cdot)$. When the threshold is reached the server switches to distribution $A^2(\cdot)$ and as would be expected it switches back to the initial distribution when the system becomes empty.

As there are only two phases we shall use A and B instead of A^1 and A^2 with the corresponding changes in notation listed below.

$$\begin{array}{rclcrcl} A(t) & = & A^1(t), & & B(t) & = & A^2(t), \\ A_n & = & A_n^1, & & B_n & = & A_n^2, \\ a(z) & = & a_1(z), & & b(z) & = & a_2(z), \\ \xi_a(z) & = & \xi_1(z), & & \xi_b(z) & = & \xi_2(z), \\ a_i & = & a_i^1, & & b_i & = & a_i^2, \\ \rho_a & = & \rho_1, & & \rho_b & = & \rho_2. \end{array} \tag{4.1}$$

We assume that $\rho_a > 0$ throughout otherwise the solution is trivial. The results we use demonstrate the connection between the queueing process and a discrete-time two-phase MRP of the type described in Chapter 2. This process is illustrated in Figure 4.1.

We shall consider three major types of threshold in this chapter.

(i) We call the first a fixed upward threshold. This is when the phase change occurs at the first time immediately after a customer finishes service when there are more than a certain number of customers in the system. We shall label this critical number of customers by k.

(ii) The second is when the phase change occurs after a random number of customers have been served in the busy period. We will consider the case when the number of customers served before the phase change is geometric with parameter p and thus we call this the geometrically-distributed random-time threshold.

(iii) The third is a fixed-time threshold. This is when the phase change occurs after a set number of customers are served during the busy period. We shall label this number of customers by S.

In Section 4.5 we briefly consider some other random thresholds.

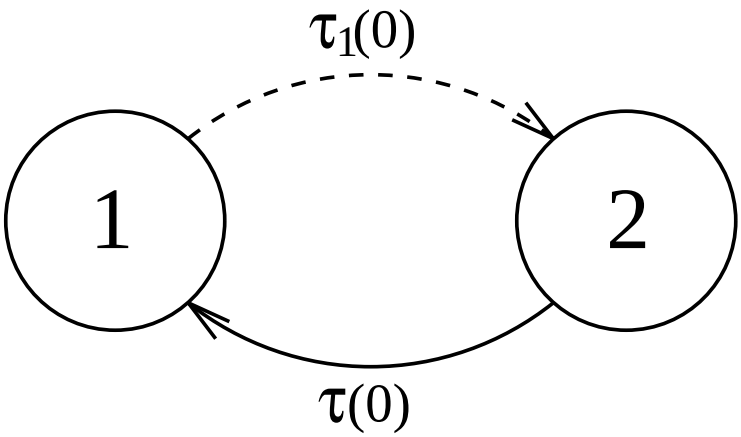

Figure 4.1: The two phase, Markov renewal process (state 2 is the non-renewal state).

4.0.1 Motivation

The motivation for each of these examples is slightly different. There are three fundamental reasons for considering problems of this sort. The first reason is simply to model systems which may have some peculiarity which fits this structure.

The second reason is control. By allowing some sort of control over the system through service times (or arrival blocking, etc) we can optimise a performance measure of the system. For instance we may wish to constrain the average waiting time for a customers while at the same time maximising the proportion of time in which server is busy in order to make the most of the server.

In the standard M/G/1 model these two objectives are not compatible. In the M/G/1 queue the probability of there being no customers in the system is $1-\rho$. The mean number of customer in the system has a term proportional to $1/(1-\rho)$ in it. Thus if we constrain this (and hence the mean waiting time) we may be forced to have an unacceptably high probability of the system being empty. Note also that the length of the busy period is 1 divided by the probability of the server being empty (from renewal theory). Thus the longer the busy period, the greater the server utilisation.

Because of this incompatibility we introduce elements such as we have described to give further control over the queue. This is the aim of a fixed upward threshold. The server may serve slowly during the first phase in order to lengthen the busy period and thus increase the utilisation of the server. However if the number of customers in the system becomes too large (and hence waiting time becomes too long) the queue switches to a faster service rate to remove the excess customers. In this case the queue is cleared before returning to the slower service rate. A more desirable situation is that the faster service continues only until enough customers are removed from the system to remove the problem. This will be considered in Chapter 7.

Such stochastic control over a queue is not an unusual idea, for instance Dshalalow (1989) uses this concept. In most problems the basis for the control is assumed to be state-dependent and otherwise independent of the history of the process. This is the novel part of the problem considered here. The phase depends on the history of the process, not just the current phase.

The fixed-time threshold is a cruder type of control. The server serves slowly

for the first S services in order to build up a backlog which will then increase the length of the busy period and thus server utilisation. This is mentioned in Neuts (1989) and is closely related to the E-limited service discipline of LaMaire (1992).

The final reason given here for considering such situations is to model a server which may breakdown (or exhibit some similar phenomena). In such a situation there are many possible descriptions for the way in which the breakdown occurs. The one we shall look at is when the server breaks down at the end of a service with probability p which is a constant. We assume in this particular model that the server must then alter its behaviour until the queue is emptied and the server can be repaired. For instance the queue might simply discard all of the customers present at the time of the breakdown.

This is not a very good model for breakdowns. For a start we have assumed that the repair takes zero time and requires the queue to be empty. Other descriptions of breakdowns include features such as the server breaking down during the service or even when the server is idle. Further the time until a breakdown might depend on the number of customers served since the last breakdown or since the last checkup of the server. We shall address some of these criticisms in Section 4.5 and other aspects of the problem in Chapter 5 where we consider problems with four phases. We might note also that these examples are provided to demonstrate the utility of this method, not to be an end unto themselves. With some further work a suitable model for breakdowns could be constructed but the specifics will depend on the mechanisms involved in the system.

We call this case the geometrically-distributed random-time threshold because the breakdown occurs after a geometrically distributed number of customers have been served in the busy period.

4.1 Results

Theorem 4.1 *The following results hold for the relevant thresholds described above.*

(i) For a fixed upward threshold condition $(*)$ *holds for all* $k \in \mathbb{N}$ *if* $\rho_b \leq 1$.

(ii) For a geometrically-distributed random-time threshold condition $(*)$ *holds for*

$$p \in \left(1 - \frac{1}{\alpha}, 1\right) \quad \textit{and} \quad \rho_b \leq 1,$$

where $\alpha = \sup_{z \in [0,1)} \xi_a(z)$.

(iii) For a fixed time threshold condition $(*)$ *holds for all* $S \in \mathbb{N}$ *if* $\rho_b \leq 1$.

Proof:

(i) See Lemma 4.2.1 in Section 4.2.

(ii) See Lemma 4.3.1 in Section 4.3.

(iii) See Section 4.4. □

For the rest of this chapter we shall use the following matrix defined for $k \in \mathbb{N}$.

$$\mathbf{P}_k = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{k-1} & a_k \\ a_0 & a_1 & a_2 & \cdots & a_{k-2} & a_{k-1} \\ 0 & a_0 & a_1 & \cdots & a_{k-3} & a_{k-2} \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & a_0 & a_1 \end{pmatrix}. \tag{4.2}$$

Theorem 4.2 *Given either* $k \in \mathbb{N}$, $p \in \left(1-\frac{1}{\alpha}, 1\right)$ *or* $S \in \mathbb{N}$ *for each type of threshold respectively the following results hold.*
(i) For $\rho_b > 1$ *the queue is transient.*
(ii) For $\rho_b = 1$ *the queue is null recurrent.*
(iii) For $\rho_b < 1$ *the queue is positive recurrent and the probability generating function for the equilibrium distribution of customers in the queue is given by*

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z) + \{b(z)-a(z)\}\, z R_q^{(F)}(z)}{b(z)-z}\right],$$

for $z \in [0,1)$ *and with the mean length of the busy period* m *given by*

$$m = \left[\frac{1 + \{\rho_a - \rho_b\}\, R_q^{(F)}(1)}{1-\rho_b}\right],$$

where $R_q^{(F)}(z)$ *is a non-negative function bounded above on the interval* $[0,1]$ *that is determined by the specific type of threshold between the phases.* F *specifies the type of threshold used and* q *is a parameter associated with that type of threshold. Thus we write*

$$F = \begin{cases} U, & \text{a fixed upward threshold at } k, & q = k \in \mathbb{N}, \\ G, & \text{a geometrically-distributed random-time threshold}, & q = p \in \left(1-\frac{1}{\alpha}, 1\right), \\ T, & \text{a fixed-time threshold at } S, & q = S \in \mathbb{N}. \end{cases}$$

The actual values for $R_q^{(F)}(z)$ *are given by*

$$R_k^{(U)}(z) = \frac{1}{z}\,\mathbf{e}_1\,(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{z}^t, \tag{4.3}$$

where $\mathbf{P}_k$ *is the* k x k *matrix defined in (4.2),*

$$R_p^{(G)}(z) = \frac{z - F^*(1-p)}{z - a(z)(1-p)}, \tag{4.4}$$

where $F^*(z)$ *is the probability generating function for the number of customers served during the busy period of the M/G/1 queue with service-time distribution* $A(\cdot)$ *(see 3.13 on page 49).*

$$R_S^{(T)}(z) = \left[\frac{\xi_a(z)^S - 1}{\xi_a(z)^{S-1}(\xi_a(z)-1)}\right] - \frac{(1-\delta_{S1})}{z}\sum_{i=1}^{S-1}\left[\frac{\xi_a(z)^{S-i}-1}{\xi_a(z)^{S-i-1}(\xi_a(z)-1)}\right] a^{(i)}, \tag{4.5}$$

where

$$a^{(i)} = \int_0^\infty \frac{e^{-\lambda x}(\lambda x)^{i-1}}{i!} dA^{(i)}(x),$$

$A^{(i)}(\cdot)$ *being the* i*-fold convolution of* $A(\cdot)$.

Proof: We use the notation from (4.1) and definitions from Chapter 3 so that

$$\begin{aligned} M_0(z) &= 1, \\ M_n(z) &= z^{X_n} \prod_{k=0}^{n-1} \left(\frac{z^{I(X_k \neq 0)}}{I_{C_k^1} a(z) + I_{C_k^2} b(z)} \right), \end{aligned}$$

is the martingale for $n \in I\!N$.

If $\rho_a > 0$ and $\rho_b > 1$, Lemma 3.1.2 demonstrates that $\tau(0) = \infty$ with positive probability. Thus the queue is unstable in the sense that $X_n \to \infty$ as $n \to \infty$.

If however $\rho_a > 0$ and $\rho_b \leq 1$ we must show that condition $(*)$ holds so that the stopping times we use are regular for the martingale. If $\rho_a \leq 1$ the condition is trivial and if $\rho_a > 1$ the condition becomes (noting that $\rho_b \leq 1$)

$$E\left[\alpha^{\tau_1(0)}\right] < \infty, \tag{4.6}$$

where $\alpha = \sup_{z \in [0,1)} \xi_a(z)$. This must be shown with respect to each specific threshold. Theorem 4.1 is provided to point to the relevant proofs as they are located in the sections dealing with their respective thresholds.

Theorem 3.9 gives the probability generating function for the number of customers in the system at equilibrium to be

$$\begin{aligned} E\left[z^X\right] &= \frac{1}{m}\left[\frac{1 - E\left[z^{X_{\tau_1(0)}}\right]}{1 - \xi_b(z)} + \frac{E\left[z^{X_{\tau_1(0)}}\right] - z}{1 - \xi_a(z)}\right] \\ &= \frac{1}{m}\left[\frac{1 - z}{1 - \xi_b(z)} + \frac{E\left[z^{X_{\tau_1(0)}}\right] - z}{1 - \xi_a(z)} - \frac{E\left[z^{X_{\tau_1(0)}}\right] - z}{1 - \xi_b(z)}\right] \\ &= \frac{1}{m}\left[\frac{1 - z}{1 - \xi_b(z)} + \frac{[\xi_a(z) - \xi_b(z)]\left[E\left[z^{X_{\tau_1(0)}}\right] - z\right]}{(1 - \xi_a(z))(1 - \xi_b(z))}\right] \\ &= \frac{1}{m}\left[\frac{b(z)(1 - z)}{b(z) - z} + \frac{\{b(z) - a(z)\}\, z\left[E\left[z^{X_{\tau_1(0)}}\right] - z\right]}{(a(z) - z)(b(z) - z)}\right]. \end{aligned} \tag{4.7}$$

In order to find the final result we must calculate

$$E\left[z^{X_{\tau_1(0)}}\right] - z,$$

but this will be different for each threshold considered and so we will consider each in a separate section below. However, we may note that this solution will exist for $\rho_a > 1$ and in this case there will be some $z_0 \in (0,1)$ such that $a(z_0) = z_0$ and hence the denominator in the second term of (4.7) will be zero. Therefore, in order that the generating function exist we must have the numerator equal to zero at this point as well. For this to be so we shall write

$$E\left[z^{X_{\tau_1(0)}}\right] - z = (a(z) - z)R(z),$$

for some function $R(z)$ bounded on $[0,1]$. Then we can write the solution as

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1 - z) + \{b(z) - a(z)\}\, zR(z)}{b(z) - z}\right]. \tag{4.8}$$

Note that we have yet to demonstrate that an $R(z)$ of this form can be found but we shall do this in the following sections. As the particular function $R(z)$ depends both on the type of threshold and its relevant parameter we shall write it as $R_q^{(F)}(z)$ where F gives the type of threshold and q gives the relevant parameter. For proofs of the expressions for $R_q^{(F)}(z)$ see Theorems 4.3, 4.4 and 4.6.

The value of m may be calculated in two ways. The first is to note that m is given by the renewal results to be the mean number of customers served in the busy period. This can be calculated through use of the generating function $F^*(x_1, x_2)$. The alternative, which we use here due to its ease, is to note that $E\left[z^X\right]$ is a probability generating function and in the limit as z tends up to 1 it must be 1. Hence m can be viewed as a normalising constant. Taking the limit as z tends up to 1 using L'Hôpital's rule the left-hand side is equal to one, and hence multiplying both sides by m gives

$$m = \left[\frac{-b(1) + \{b(1) - a(1)\} R_q^{(F)\prime}(1) + \{b(1) - a(1)\} R_q^{(F)}(1) + \{b'(1) - a'(1)\} R_q^{(F)}(1)}{b'(1) - 1}\right],$$

where noting that $a(1) = 1$, $a'(1) = \rho_a$, $b(1) = 1$ and $b'(1) = \rho_b$ gives

$$m = \left[\frac{1 + \{\rho_a - \rho_b\} R_q^{(F)}(1)}{1 - \rho_b}\right].$$

Note that as $\rho_b \uparrow 1$, m tends to infinity and so when $\rho_b = 1$ the process is null recurrent. □

Remarks: (i) The generating function in the solution is interesting in itself as it suggests that the solution can be written as the normal solution to the M/G/1 queue plus a correcting term that depends on the difference between $a(z)$ and $b(z)$ and the type of threshold.

(ii) Note that m is insensitive to the actual distribution $B(\cdot)$ except through ρ_b and when $\rho_1 = \rho_2$, m is insensitive to both distributions $A(\cdot)$ and $B(\cdot)$ except through ρ_a and ρ_b.

For each type of threshold considered we must now prove that condition $(*)$ is satisfied and calculate $R_q^{(F)}(z)$. It will be seen that conditions $(*)$ is satisfied without restriction except in the case of the geometrically-distributed random-time threshold.

4.2 Fixed upward threshold

In this case the threshold is a fixed number of customers, say k. If more than this number of customers are in the system at the end of a service then the queue switches from phase 1 to phase 2. We take

$$\begin{aligned}\tau_1(n) &= \begin{cases} \tau(n) \wedge \inf\{m > n | X_m > k\}, & \text{if phase(n)} = 1, \\ n, & \text{if phase(n)} = 2, \end{cases} \\ &= I_{C_n^1}\left[\tau(n) \wedge \inf\{m > n | X_m > k\}\right] + I_{C_n^2} n.\end{aligned}$$

If at time n the process is in phase 1 then $\tau_1(n)$ is the time of the next transition to phase 2. (Note that when the queue empties we assume a dummy transition through phase 2.) When at time n the process is in phase 2, $\tau_1(n)$ is defined to be equal to n in order to be consistent with our definitions. When $n = 0$ we have assumed the process to be in phase 1 and so

$$\tau_1(0) = \tau(0) \wedge \inf\{m > 0 | X_m > k\}. \tag{4.9}$$

In order to apply the results of Section 3.2.3 and 3.3 we must first prove the regularity of the stopping time $\tau(0)$. To do this we must satisfy condition $(*)$.

Lemma 4.2.1 *For a threshold as described above with $k \in \mathbb{N}$, $\rho_a > 0$ and $\rho_b \leq 1$ condition $(*)$ is satisfied. Furthermore*

$$E\left[\omega^{\tau_1(0)}\right] = 1 + (\omega - 1)\mathbf{e}_1 \left(\mathbf{I} - \omega \mathbf{P}_k\right)^{-1} \mathbf{1}^t,$$

for all $\omega \in [0, \alpha]$.

Proof: Condition $(*)$ is

$$E\left[\alpha^{\tau_1(0)}\right] < \infty,$$

for $\alpha = \sup_{z \in [0,1]} \xi_a(z)$. We write, for $\omega \in [0, \alpha]$ the expectation

$$\begin{aligned}E\left[\omega^{\tau_1(0)}\right] &= \sum_{i=1}^{\infty} \omega^i p\{\tau_1(0) = i\} \\ &= \sum_{i=1}^{\infty} \sum_{j=0}^{k} \omega^i p\{\tau_1(0) = i,\, X_{i-1} = j\} \\ &= \omega\, p\{\tau_1(0) = 1\} + \\ &\quad \sum_{i=2}^{\infty} \sum_{j=1}^{k} \omega^i\, p\{\tau_1(0) = i | X_{i-1} = j, \tau_1(0) \geq i\}\, p\{X_{i-1} = j, \tau_1(0) \geq i\}. \end{aligned} \tag{4.10}$$

Now we can see that $p\{\tau_1(0) = 1\} = a_0 + \sum_{l=k+1}^{\infty} a_l$ and for $i > 1$ and $j = 1, \ldots, k$

$$\begin{aligned}p\{\tau_1(0) = i | X_{i-1} = j, \tau_1(0) \geq i\} &= \left(a_0 \delta_{1j} + \sum_{l=k+1}^{\infty} a_{l-j+1}\right) \\ &= \left(1 - \sum_{l=(j-1)\vee 1}^{k} a_{l-j+1}\right),\end{aligned}$$

where $a_i = p\{A_1 = i\}$ as defined by (4.1). We shall define g_j by

$$g_j = \left(1 - \sum_{l=(j-1)\vee 1}^{k} a_{l-j+1}\right),$$

and from this we form the vector $\mathbf{g}$. Substituting these in (4.10) we arrive at

$$E\left[\omega^{\tau_1(0)}\right] = \omega g_1 + \omega \sum_{j=1}^{k} g_j \sum_{i=1}^{\infty} \omega^i p\{X_i = j, \tau_1(0) > i\}.$$

In order to find $\sum_{i=1}^{\infty} \omega^i p\{X_i = j, \tau_1(0) > i\}$, we define the vector

$$\mathbf{v}^i = (p\{X_i = 1, \tau_1(0) > i\}, p\{X_i = 2, \tau_1(0) > i\}, \ldots, p\{X_i = k, \tau_1(0) > i\}), \quad (4.11)$$

the sub-stochastic probability transfer matrix $\mathbf{P}_k$ as in (4.2) and $\mathbf{v}^1 = (a_1, a_2, \ldots, a_k)$, the probability vector of initial probabilities given a transition from $X_0 = 0$. Then

$$\mathbf{v}^i = \mathbf{v}^1 \mathbf{P}_k^{i-1}.$$

We seek conditions under which

$$\sum_{i=1}^{\infty} \omega^i \mathbf{P}_k^i$$

converges. From Property B.3 of norms and Theorems B.1 and B.2 the series converges if $|\omega| \, \|\mathbf{P}\| < 1$ for some matrix norm $\|\mathbf{P}\|$.

We use the matrix norm defined in (B.6), for $z \in (0, 1]$ by

$$\|\mathbf{A}\|_z = \max_{i=1,..,k} \left[\sum_{j=1}^{k} |a_{ij}| z^{j-i}\right],$$

for a matrix $\mathbf{A} = (a_{ij})$. Then

$$\begin{aligned}
\|\mathbf{P}_k\|_z &= \max\left\{\sum_{j=1}^{k} a_j z^{j-1}, \sum_{j=1}^{k} a_{j-1} z^{j-2}, \sum_{j=2}^{k} a_{j-2} z^{j-3}, \ldots\right\} \\
&= \max\left\{\sum_{j=1}^{k} a_j z^{j-1}, \sum_{j=0}^{k-1} a_j z^{j-1}\right\} \\
&< \sum_{j=0}^{k} a_j z^{j-1} \\
&< \frac{a(z)}{z},
\end{aligned}$$

so that

$$\frac{1}{\|\mathbf{P}_k\|_z} > \frac{z}{a(z)}.$$

Thus there exists a $z_0 \in [0, 1)$ such that

$$\sup_{z \in [0,1)} \frac{z}{a(z)} < \frac{1}{\|\mathbf{P}_k\|_{z_0}}.$$

Thus $\alpha\|\mathbf{P}_k\|_{z_0} < 1$ and hence the series converges for $\omega \in [0, \alpha]$. This proves that $E\left[\alpha^{\tau_1(0)}\right] < \infty$. For the second part we can note that we now have

$$E\left[\omega^{\tau_1(0)}\right] = \omega\left\{g_1 + \omega \mathbf{v}^1 \left(\sum_{i=0}^{\infty} \omega^i \mathbf{P}_k^i\right) \mathbf{g}^t\right\},$$

when the sum converges. The previous result means that this sum converges for all $\omega \in [0, \alpha]$ and we know from Theorem B.2 that it must converge to $(\mathbf{I} - \omega\mathbf{P}_k)^{-1}$ and so

$$E\left[\omega^{\tau_1(0)}\right] = \omega\left\{g_1 + \omega\ \mathbf{v}^1\ (\mathbf{I} - \omega\mathbf{P}_k)^{-1}\mathbf{g}^t\right\}.$$

Now $\mathbf{v}^1 = \mathbf{e}_1\,\mathbf{P}_k$ and so Lemma B.0.1 means that

$$\begin{aligned} \omega\ \mathbf{v}^1\ (\mathbf{I} - \omega\mathbf{P}_k)^{-1} &= \omega\mathbf{e}_1\,\mathbf{P}_k(\mathbf{I} - \omega\mathbf{P}_k)^{-1} \\ &= -\mathbf{e}_1\ + \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}, \end{aligned}$$

which gives

$$\begin{aligned} E\left[\omega^{\tau_1(0)}\right] &= \omega\left\{g_1 - \mathbf{e}_1\,\mathbf{g}^t + \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}\mathbf{g}^t\right\} \\ &= \omega\left\{g_1 - g_1 + \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}\mathbf{g}^t\right\}. \end{aligned}$$

Now $\mathbf{g}^t = (\mathbf{I} - \mathbf{P}_k)\mathbf{1}^t$ so that we get (again using Lemma B.0.1) that

$$\begin{aligned} E\left[\omega^{\tau_1(0)}\right] &= \omega\mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}(\mathbf{I} - \mathbf{P}_k)\mathbf{1}^t \\ &= \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}(\omega\mathbf{I} - \omega\mathbf{P}_k)\mathbf{1}^t \\ &= \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}(\mathbf{I} - \omega\mathbf{P}_k)\mathbf{1}^t + \mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}(\omega\mathbf{I} - \mathbf{I})\mathbf{1}^t \\ &= 1 + (\omega - 1)\mathbf{e}_1\,(\mathbf{I} - \omega\mathbf{P}_k)^{-1}\mathbf{1}^t. \end{aligned}$$

This is the desired result. □

Remark: It can be seen that this has the desirable properties of a probability generating function. When $\omega = 1$, $E\left[\omega^{\tau_1(0)}\right] = 1$ and when we take the derivative with respect to ω at $\omega = 1$ we get the mean time until the threshold is reached

$$E\left[\tau_1(0)\right] = \mathbf{e}_1\,(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{1}^t,$$

which agrees with (4.29) of Section 4.6.1.

Note that we now have the constraint necessary for the almost sure finiteness and regularity of $\tau(0)$. Next we provide the value of $R_k^{(U)}(z)$ from the value of $E\left[z^{X_{\tau_1(0)}}\right] - z$.

Theorem 4.3 *For $\rho_a > 0$, $z \in [0, 1)$, $X_0 = 0$ and the threshold $k \in I\!N$ we get*

$$E\left[z^{X_{\tau_1(0)}}\right] - z = [a(z) - z]\,R_k^{(U)}(z),$$

where

$$R_k^{(U)}(z) = \frac{1}{z}\,\mathbf{e}_1\,(\mathbf{I} - \mathbf{P}_k)^{-1}\,\mathbf{z}^t$$

and $\mathbf{P}_k$ is the k x k sub-stochastic matrix defined in (4.2).

Proof: We can write $E\left[z^{X_{\tau_1(0)}}\right]$ as

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right] &= \sum_{i=1}^{\infty}\sum_{j=0}^{k} E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right] p\{X_{i-1}=j, \tau_1(0)\geq i\} \\
&= E\left[z^{X_1} I(\tau_1(0)=1)\middle| X_0=0\right] p\{X_0=0\} \\
&\quad + \sum_{i=2}^{\infty}\sum_{j=1}^{k} E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right] p\{X_{i-1}=j, \tau_1(0)\geq i\} \\
&= E\left[z^{X_1} I(\tau_1(0)=1)\middle| X_0=0\right] \\
&\quad + \sum_{i=2}^{\infty}\sum_{j=1}^{k} E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right] p\{X_{i-1}=j, \tau_1(0)\geq i\}
\end{aligned}$$

as we have $X_0 = 0$. Now for $i = 1$ and $j = 0$, $E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right]$ is

$$\begin{aligned}
E\left[z^{X_1} I(\tau_1(0)=1)\middle| X_0=0\right] &= a_0 + \sum_{l=k+1}^{\infty} a_l z^l \\
&= a(z) - \sum_{l=1}^{k} a_l z^l.
\end{aligned}$$

For $i > 1$ and $j = 1$ the result is

$$\begin{aligned}
E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=1, \tau_1(0)\geq 1\right] &= a_0 + \sum_{l=k+1}^{\infty} a_l z^l \\
&= a(z) - \sum_{l=1}^{k} a_l z^l,
\end{aligned}$$

and for $j > 1$ and $i > 1$ it is

$$\begin{aligned}
E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right] &= \sum_{l=k+1}^{\infty} a_{l-j+1} z^l \\
&= a(z) z^{j-1} - \sum_{l=j-1}^{k} a_{l-j+1} z^l.
\end{aligned}$$

From the previous two equations we get for $j = 1, \ldots, k$ and $i > 1$ the following

$$E\left[z^{X_i} I(\tau_1(0)=i)\middle| X_{i-1}=j, \tau_1(0)\geq i\right] = a(z) z^{j-1} - \sum_{l=(j-1)\vee 1}^{k} a_{l-j+1} z^l,$$

which we shall call $g_j(z)$. Thus we arrive at the equation

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right] &= \left(a(z) - \sum_{l=1}^{k} a_l z^l\right) \\
&\quad + \sum_{i=1}^{\infty}\sum_{j=1}^{k} \left(a(z) z^{j-1} - \sum_{l=(j-1)\vee 1}^{k} a_{l-j+1} z^l\right) p\{X_i=j, \tau_1(0)\geq i\} \\
&= g_1(z) + \sum_{i=1}^{\infty}\left(\sum_{j=1}^{k} p\{X_i=j, \tau_1(0)\geq i\} g_j(z)\right).
\end{aligned}$$

In order to find $\sum_{i=1}^{\infty} p\{X_i = j, \tau_1(0) \geq i\}$ we define $\mathbf{v}^i$ and $\mathbf{P}_k$ as in (4.11) and (4.2) respectively and $\mathbf{v}^1 = (a_1, a_2, \ldots, a_k)$. These are respectively the probability vector after the ith transition in phase 1, the sub-stochastic probability transfer matrix and the vector of initial probabilities for the subset $\{1, 2, \ldots, k\}$, of the state-space of X_n. Immediately from this we get

$$\mathbf{v}^i = \mathbf{v}^1 \mathbf{P}_k^{i-1}.$$

Theorem B.2 shows that summing this from $i = 1$ to ∞ gives

$$\sum_{i=1}^{\infty} \mathbf{v}^i = \mathbf{v}^1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1}.$$

Now $\mathbf{v}^1 = \mathbf{e}_1 \mathbf{P}_k$ and so we can see (using Lemma B.0.1) that

$$\begin{aligned} \mathbf{v}^1 (\mathbf{I} - \mathbf{P}_k)^{-1} &= \mathbf{e}_1 \mathbf{P}_k (\mathbf{I} - \mathbf{P}_k)^{-1} \\ &= -\mathbf{e}_1 + \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1}. \end{aligned} \tag{4.12}$$

Take $\mathbf{g}(z) = (g_1(z), g_2(z), \cdots, g_k(z))$ and $\mathbf{g}(z)^t$ the corresponding column vector and we get

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] &= g_1(z) + \sum_{i=1}^{\infty} \left(\sum_{j=1}^{k} v_j^i g_j(z) \right) \\ &= \mathbf{e}_1 \mathbf{g}(z)^t + \sum_{i=1}^{\infty} \mathbf{v}^i \mathbf{g}(z)^t \\ &= \left[\mathbf{e}_1 + \mathbf{v}^1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1}\right] \mathbf{g}(z)^t, \end{aligned} \tag{4.13}$$

which from (4.12) gives

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] &= \left(\mathbf{e}_1 - \mathbf{e}_1 + \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1}\right) \mathbf{g}(z)^t \\ &= \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1} \mathbf{g}(z)^t. \end{aligned}$$

Now we can simplify $g_j(z)$ and hence $\mathbf{g}(z)^t$ as follows

$$\begin{aligned} g_j(z) &= \frac{a(z)}{z} z^j - \mathbf{e}_j \mathbf{P}_k \mathbf{z}^t, \\ \Rightarrow \mathbf{g}(z)^t &= \frac{a(z)}{z} \mathbf{z}^t - \mathbf{P}_k \mathbf{z}^t. \end{aligned}$$

Hence we can write $E\left[z^{X_{\tau_1(0)}}\right]$ as

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] &= \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1} \left(\frac{a(z)}{z} \mathbf{z}^t - \mathbf{P}\ \mathbf{z}^t\right) \\ &= \frac{a(z)}{z} \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1} \mathbf{z}^t - \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1} \mathbf{P}_k \mathbf{z}^t \end{aligned}$$

$$
\begin{aligned}
&= \frac{a(z)}{z}\,\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t + \mathbf{e}_1\,\mathbf{I}\,\mathbf{z}^t - \mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\,\mathbf{z}^t \\
&= \frac{a(z)}{z}\,\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t + z - \mathbf{e}_1\,(\mathbf{I}\text{-}\mathbf{P}\ \ )^{-1}\,\mathbf{z}^t \\
&= \left(\frac{a(z)}{z}-1\right)\,\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\,\mathbf{z}^t + z \\
&= \frac{1}{z}\,(a(z)-z)\,\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\,\mathbf{z}^t + z,
\end{aligned}
$$

using Lemma B.0.1. This leads easily to the desired result. □

4.2.1 Some limiting cases

There are several special cases of this system which have been examined in detail in the literature, the simplest of which is the M/G/1 queue. The probability generating function for the equilibrium number in the M/G/1 queueing system is expressed in (3.11). In the following special cases we expect the same solution as in the M/G/1 model. If $a(z) = b(z)$ the solution is immediate. If $k \to \infty$ and $\rho_a < 1$ we also expect (3.11). As $k \to \infty$, $F^*(1, y)$, which is the generation function for the time spent in phase 2, approaches 1. This is because the probability that zero time is spent in the second phase approaches one. Hence

$$E[z^X] = \frac{1}{m}\left[\frac{1-z}{1-\frac{z}{a(z)}}\right] = \frac{1}{m}\frac{a(z)(1-z)}{a(z)-z}.$$

We shall next consider what happens if a customer arriving at an empty server has a different service-time distribution from that of customers arriving when the server is busy. The solution to this type of problem can also be found in Yeo (1962). The result given by Yeo is

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)-za(z)}{b(z)-z}\right],$$

where $m = \frac{1+(\rho_a-\rho_b)}{1-\rho_b}$. This might occur if there was some overhead associated with restarting the server or if there are server vacations. The two-phase M/G/1 queue with a fixed upward threshold should be the same as this when $k = 0$. We have not included this in the previous results but it is an easy case since Corollary 3.6 gives $F^*(1, y)$ as

$$F^*\left(1, \frac{z}{b(z)}\right) = E\left[z^{X_{\tau_1(0)}}\right],$$

the right-hand side of which is $a(z)$ in this case. Once $F^*(1, y)$ is known we can write the solution using Theorem 3.8 as

$$\begin{aligned} E[z^X] &= \frac{1}{m}\left[\frac{(a(z)-z)(1-\frac{z}{b(z)}) + (1-a(z))(1-\frac{z}{a(z)})}{(1-\frac{z}{a(z)})(1-\frac{z}{b(z)})}\right] \\ &= \frac{1}{m}\left[\frac{b(z)-za(z)}{b(z)-z}\right]. \end{aligned}$$

4.2.2 Modifications

As noted in Section 3.4 some modifications may be needed to deal with other possible descriptions of the server's behaviour. We shall consider here some such modifications.
(I) Modifications during phase 1.
If we consider properties other than the service-time distribution of the server to vary between phases then we may need to modify the previous work slightly. We use the relevant expression for $a(z)$ in the solution. However we must, in some cases, make further modifications to the solution. Cases (i), (ii) and (iv) require no further modification to the solution. Case (v) is covered by Jun Huek Park (1990) for a queue with one phase and can be dealt with here by an obvious extension. Case (vi) will not be used. This leaves (iii) as the interesting example and we shall consider this here.
(iii) During phase 1 the server allows only the first $N_1 \geq 1$ arrivals during each service, the remaining arrivals (if there are any) are blocked. (We shall simply write $N = N_1$ here.) In this case

$$a(z) = \sum_{i=0}^{N-1} a_i (z^i - z^N) + z^N.$$

The only other modification necessary in the calculation of $E\left[z^{X_{\tau_1(0)}}\right]$. We define the sub-stochastic probability transfer matrix ${}_N\mathbf{P}_k$ for $1 \leq N \leq k$ by

$$ {}_N\mathbf{P}_k = \begin{pmatrix} a_1 & a_2 & \cdots & a_{N-1} & \sum_{i=N}^{\infty} a_i & 0 & \cdots & 0 \\ a_0 & a_1 & \cdots & a_{N-2} & a_{N-1} & \sum_{i=N}^{\infty} a_i & \cdots & 0 \\ 0 & a_0 & \cdots & a_{N-3} & a_{N-2} & a_{N-1} & \cdots & 0 \\ & & \vdots & & & & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & a_1 \end{pmatrix}. \tag{4.14}$$

Note that when $N > k$ this modification is trivial as ${}_N\mathbf{P}_k = \mathbf{P}_k$. Having done this, the same procedure as in Theorem 4.3 produces the result

$$E\left[z^{X_{\tau_1(0)}}\right] - z = \frac{1}{z}\left(a(z) - z\right)\mathbf{e}_1 \left(\mathbf{I} - {}_N\mathbf{P}_k\right)^{-1}\mathbf{z}^t,$$

which gives the solution

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z) + \{b(z) - a(z)\}\, z\ {}_N R_k^{(U)}(z)}{b(z) - z}\right],$$

where

$$m = \left[\frac{1 + \{\rho_a - \rho_b\}\, {}_N R_k^{(U)}(1)}{1 - \rho_b}\right]$$

and

$${}_N R_k^{(U)}(z) = \mathbf{e}_1 \left(\mathbf{I} - {}_N\mathbf{P}_k\right)^{-1}\mathbf{z}^t.$$

(II) Modifications during phase 2.
Modifications to the behaviour in phase two are relatively inconsequential. We take $b(z)$ as defined in the relevant part of Section 3.4 and then $\rho_b = b'(1)$. The condition for recurrence, $\rho_b \leq 1$, remains the same. With $b(z)$ given by the relevant function the solution remains the same as that in Theorem 4.2.

4.3 The geom.-distributed random-time threshold

This is the case in which the switch between the two service-time distributions occurs at a random time. We assume that the probability of the switch occurring at the end of a service is p. That is to say given the process is in phase 1 at time n and $X_{n+1} \neq 0$

$$p\{\tau_1(n) = n+1\} = p,$$

where p is a constant $0 < p < 1$. Thus we can write, at some time $n \in \mathbb{Z}^+$

$$\begin{aligned} \tau_1(n) &= \begin{cases} \tau(n) \wedge (n+R), & \text{if phase(n)} = 1, \\ n, & \text{if phase(n)} = 2, \end{cases} \\ &= I_{C_n^1}\left[\tau(n) \wedge (n+R)\right] + I_{C_n^2} n, \end{aligned}$$

where R is a random variable with the geometric distribution

$$p\{R = i\} = (1-p)^{i-1}p.$$

Thus we could consider the threshold to occur at a random time which is geometrically distributed. From this we can write

$$\tau_1(0) = \tau(0) \wedge R. \tag{4.15}$$

The following lemma provides the constraint necessary for regularity.

Lemma 4.3.1 *When $\rho_a > 0$, $\rho_b \leq 1$ and*

$$p > 1 - \frac{1}{\alpha},$$

condition $(*)$ is satisfied and furthermore for $\omega \in [0, \alpha]$

$$E\left[\omega^{\tau_1(0)}\right] = \frac{\omega p + (1-\omega)F^*(\omega(1-p))}{1-\omega(1-p)},$$

where $F^*(\xi)$ is the unique solution to the equation $z = \xi a(z)$ in the interval $[0,1)$.

Proof: When $\rho_a \leq 1$ and hence $\alpha = 1$, condition $(*)$ is trivially true. When $\rho_a > 1$ and hence $\alpha > 1$ the following is true

$$\alpha^{\tau_1(0)} \leq \alpha^R \ \ a.s.,$$

therefore

$$\begin{aligned} E\left[\alpha^{\tau_1(0)}\right] &\leq E\left[\alpha^R\right] \\ &= \sum_{i=1}^{\infty} \alpha^i p\{R = i\} \\ &= \sum_{i=1}^{\infty} \alpha^i (1-p)^{i-1} p \\ &= p\alpha \sum_{i=0}^{\infty} [\alpha(1-p)]^i \\ &= \frac{p\alpha}{1-\alpha(1-p)}, \end{aligned}$$

when this converges. It converges for $\alpha(1-p) < 1$ and so

$$p > 1 - \frac{1}{\alpha}.$$

For the second part of the proof we consider

$$\begin{aligned} E\left[\omega^{\tau_1(0)}\right] &= \sum_{n=1}^{\infty} E\left[\omega^{\tau_1(0)} \middle| R = n\right] p\{R = n\} \\ &= \frac{p}{1-p} \sum_{n=1}^{\infty} E\left[\omega^{\tau_1(0)} I(\tau(0) \geq R) \middle| R = n\right] (1-p)^n \\ &\quad + \frac{p}{1-p} \sum_{n=1}^{\infty} E\left[\omega^{\tau_1(0)} I(\tau(0) < R) \middle| R = n\right] (1-p)^n. \end{aligned}$$

When $\tau(0) \geq R$, $\tau_1(0) = R$ so that

$$\begin{aligned} E\left[\omega^{\tau_1(0)} I(\tau(0) \geq R) \middle| R = n\right] &= \omega^n p\{\tau(0) \geq R | R = n\} \\ &= \omega^n \left(1 - \sum_{i=1}^{n-1} p\{\tau(0) = i\}\right). \end{aligned}$$

When $\tau(0) < R$, $\tau_1(0) = \tau(0)$ and $I(\tau(0) < R) = \sum_{i=1}^{n-1} I(\tau(0) = i)$ so we get

$$\begin{aligned} E\left[\omega^{\tau_1(0)}\right] &= \frac{p}{1-p} \sum_{n=1}^{\infty} \left(1 - \sum_{i=1}^{n-1} p\{\tau(0) = i | R = n\}\right) \omega^n (1-p)^n \\ &\quad + \frac{p}{1-p} \sum_{n=1}^{\infty} \sum_{i=1}^{n-1} E\left[\omega^{\tau(0)} I(\tau(0) = i) \middle| R = n\right] (1-p)^n, \end{aligned}$$

which when we rearrange and swap the order of summands gives

$$\begin{aligned} E\left[\omega^{\tau_1(0)}\right] &= \frac{p}{1-p} \sum_{i=1}^{\infty} \omega^n (1-p)^n \\ &\quad - \frac{p}{1-p} \sum_{i=1}^{\infty} \sum_{n=i+1}^{\infty} \omega^n \, p\{\tau(0) = i | R = n\} (1-p)^n \\ &\quad + \frac{p}{1-p} \sum_{i=1}^{\infty} \sum_{n=i+1}^{\infty} \omega^i \, p\{\tau(0) = i | R = n\} (1-p)^n. \end{aligned}$$

When $n > i$, $p\{\tau(0) = i | R = n\}$ is simply $p\{\tau'(0) = i\}$ where $\tau'(0)$ is defined by

$$\tau'(0) = \inf\{m > 0 | X'_m = 0\},$$

where X'_m is the process formed by $X'_0 = 0$ and

$$X'_{m+1} = X'_m + A_{m+1} - I(X'_m \neq 0),$$

for all $m \in \mathbb{Z}^+$. Hence we get

$$E\left[\omega^{\tau_1(0)}\right] = \frac{\omega p}{1 - \omega(1-p)} - \frac{p}{1-p} \sum_{i=1}^{\infty} p\{\tau'(0) = i\} \sum_{n=i+1}^{\infty} \omega^n (1-p)^n$$

$$
\begin{aligned}
& + \frac{p}{1-p} \sum_{i=1}^{\infty} p\{\tau'(0) = i\} \omega^i \sum_{n=i+1}^{\infty} (1-p)^n \\
= \quad & \frac{\omega p}{1-\omega(1-p)} - \frac{\omega p}{1-\omega(1-p)} \sum_{i=1}^{\infty} p\{\tau'(0) = i\} \omega^i (1-p)^i \\
& + \sum_{i=1}^{\infty} p\{\tau'(0) = i\} \omega^i (1-p)^i \\
= \quad & \frac{\omega p \left[1 - E\left[(\omega(1-p))^{\tau'(0)}\right]\right]}{1-\omega(1-p)} + E\left[(\omega(1-p))^{\tau'(0)}\right].
\end{aligned}
$$

As X'_m is analogous to the embedded process of the M/G/1 queue we can see that $\tau'(0)$ is the number of customers served during the first busy period and so $E\left[\omega^{\tau'(0)}\right] = F^*(\omega)$, where $F^*(z)$ is the probability generating function for the number of customers served during the busy period of the M/G/1 queue with service-time distribution $A(\cdot)$ (see 3.13 on page 49). Thus we get

$$
\begin{aligned}
E\left[\omega^{\tau_1(0)}\right] &= \frac{\omega p \left[1 - F^*\left(\omega(1-p)\right)\right] + F^*\left(\omega(1-p)\right)\left[1-\omega(1-p)\right]}{1-\omega(1-p)} \\
&= \frac{\omega p + (1-\omega) F^*\left(\omega(1-p)\right)}{1-\omega(1-p)},
\end{aligned}
$$

which is the desired result. □

Note that the condition $p \in \left(1 - \frac{1}{\alpha}, 1\right)$ is a sufficient condition, not a necessary one. We have said nothing about the case when $p < 1 - \frac{1}{\alpha}$. It provides regularity and recurrence as does Lemma 4.2.1 in Section 4.2. It is important for us to note also that this is the only one of the three examples in which there is a constraint placed upon the threshold parameter p. Because, however, we have chosen this as a model for a server which may breakdown this causes no problems. In such a system ρ_a would be the normal traffic intensity of the system and as such would normally be chosen to be less than one. Hence $\xi_a(z) \leq 1$ for $z \in [0,1]$ and hence $\alpha = 1$. This means that for such an example there is only the restriction that p be positive. This is a trivial restriction as when $p = 0$ we have a degenerate case with only one phase. We next proceed to find the value of $R_p^{(G)}(z)$.

Theorem 4.4 *For $\rho_a > 0$, $p > 1 - \frac{1}{\alpha}$ with $\alpha = \sup_{z \in [0,1]} \xi_1(z)$ and $z \in [0,1)$,*

$$
E\left[z^{X_{\tau_1(0)}}\right] - z = \left[a(z) - z\right] R_p^{(G)}(z),
$$

where

$$
R_p^{(G)}(z) = \frac{z - F^*(1-p)}{z - a(z)(1-p)}
$$

and $F^*(\xi)$ is the unique solution $Z(\xi)$ to $z = \xi a(z)$, for $z \in [0,1)$.

Proof: First for $n \in I\!N$

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}} \middle| R=n\right] &= E\left[\left[I(\tau(0)\le n)+I(\tau(0)>n)\right] z^{X_{\tau_1(0)}} \middle| R=n\right] \\ &= E\left[I(\tau(0)\le n)\middle| R=n\right] + E\left[I(\tau(0)>n) z^{X_{\tau_1(0)}} \middle| R=n\right] \\ &= p\{\tau(0)\le n | R=n\} + E\left[I(\tau(0)>n) z^{X_n} \middle| R=n\right], \end{aligned}$$

because when $\tau(0) \le R$, $X_{\tau_1(0)} = 0$ and when $\tau(0) > R$, $X_{\tau_1(0)} = X_R$. As in the previous proof we use the process defined by $X'_0 = 0$ and

$$X'_{n+1} = X'_n + A_{n+1} - I(X'_n \neq 0),$$

where the random variables A_n are the number of arrivals during the nth service given that the process is in phase 1. We have $X'_n = X_n$ for $n \le \tau_1(0)$. Thus if we define $\tau'(n) = \inf\{m > n | X'_m = 0\}$ when this set is non-empty and $\tau'(n) = \infty$ when the set is empty we can see that $p\{\tau(0) \le n | R = n\} = p\{\tau'(0) \le n\}$ which gives

$$E\left[z^{X_{\tau_1(0)}} \middle| R=n\right] = p\{\tau'(0)\le n\} + E\left[I(\tau(0)>n) z^{X_n} \middle| R=n\right]. \qquad (4.16)$$

Now we consider X_n for $n \le R$ and $\tau(0) > n$. In this case $X_n = X'_n$ Thus we can use the following process slightly modified from Baccelli and Makowski (1989). We define for the standard M/G/1 queue (a single-phase queue with embedded process X'_n)

$$g(y,n) = E\left[y^{\mu_1(n)} I(\tau(0)>n)\right].$$

We can use the Remark on page 39 to show (in a similar manner to Theorem 3.5) that

$$g(\xi_a(z),n) = E\left[z^{X'_n} I(\tau(0)>n)\right].$$

Thus we get

$$\begin{aligned} \sum_{n=1}^{\infty} E\left[I(\tau(0)>n) z^{X_n} \middle| R=n\right] t^n &= \sum_{n=1}^{\infty} E\left[z^{X'_n} I(\tau(0)>n)\right] t^n \\ &= \sum_{n=1}^{\infty} g\left(\xi_a(z),n\right) t^n, \end{aligned}$$

which is defined to be $G^*\left(\xi_a(z),t\right) - 1$. Now Baccelli and Makowski also show (2.30) that

$$G^*(y,t) = 1 + \frac{tF^*(y) - yF^*(t)}{y-t}.$$

From this we can deduce that

$$\begin{aligned} \sum_{n=1}^{\infty} E\left[I(\tau(0)>n) z^{X_n} \middle| R=n\right] t^n &= \frac{tF^*(\xi_a(z)) - \xi_a(z)F^*(t)}{\xi_a(z)-t} \\ &= \frac{tz - \xi_a(z)F^*(t)}{\xi_a(z)-t}, \end{aligned} \qquad (4.17)$$

as $F^*(\xi_a(z)) = z$ from (3.12). We can also write

$$\begin{aligned}
\sum_{n=1}^{\infty} p\{\tau'(0) \leq n\}\, t^n &= \sum_{n=1}^{\infty}\sum_{i=1}^{n} p\{\tau'(0) = i\}\, t^n \\
&= \sum_{i=1}^{\infty} p\{\tau'(0) = i\}\, t^i \sum_{n=i}^{\infty} t^{n-i} \\
&= \frac{1}{1-t}\sum_{i=1}^{\infty} p\{\tau'(0) = i\}\, t^i \\
&= \frac{F^*(t)}{1-t}. \qquad (4.18)
\end{aligned}$$

We note that

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right] &= \sum_{n=1}^{\infty} E\left[z^{X_{\tau_1(0)}}\middle| R = n\right] p\{R = n\} \\
&= \sum_{n=1}^{\infty} E\left[z^{X_{\tau_1(0)}}\middle| R = n\right] p(1-p)^{n-1} \\
&= \frac{p}{1-p}\sum_{n=1}^{\infty} E\left[z^{X_{\tau_1(0)}}\middle| R = n\right] (1-p)^n \\
&= \frac{p}{1-p}\sum_{n=1}^{\infty} p\{\tau'(0) \leq n\}(1-p)^n + \frac{p}{1-p}\sum_{n=1}^{\infty} E\left[I(\tau(0) > n) z^{X_n}\middle| R = n\right](1-p)^n,
\end{aligned}$$

from (4.16). If we now substitute the results of (4.18) and (4.17) with $t = (1-p)$ we get

$$E\left[z^{X_{\tau_1(0)}}\right] = \frac{p}{1-p}.\frac{F^*(1-p)}{p} + \frac{p}{1-p}.\frac{(1-p)z - \xi_a(z)F^*(1-p)}{\xi_a(z) - (1-p)}.$$

This may be simplified in the following manner (where we have written for brevity $\xi = \xi_a(z)$).

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right] - z &= \frac{F^*(1-p)[\xi - (1-p)] + p(1-p)z - p\xi F^*(1-p)}{(1-p)(\xi - (1-p))} - z \\
&= \frac{F^*(1-p)(\xi-1)(1-p) + p(1-p)z - z(1-p)(\xi - (1-p))}{(1-p)(\xi - (1-p))} \\
&= \frac{F^*(1-p)(\xi-1)(1-p) + (1-p)z - z(1-p)\xi}{(1-p)(\xi - (1-p))} \\
&= \frac{F^*(1-p)(\xi-1) + z - z\xi}{\xi - (1-p)} \\
&= \frac{[z - F^*(1-p)](1-\xi)}{\xi - (1-p)} \\
&= \frac{z - F^*(1-p)}{\xi - (1-p)}\,(1-\xi) \\
&= \frac{z - F^*(1-p)}{z - (1-p)a(z)}\,(a(z) - z), \qquad (4.19)
\end{aligned}$$

which is the required result. □

Remark: From (3.13), $F^*(1-p)$ is the unique solution to

$$z = (1-p)a(z)$$

for unknown $z \in [0,1)$. Thus when the denominator of (4.19) is zero the numerator is also zero.

4.3.1 Some limiting cases

When $\rho_a \leq 1$ and hence $\alpha = 1$ the condition is simply that $p > 0$. If we then take the limit as $p \downarrow 0$ we should get the result for the standard M/G/1 queue with $\rho = \rho_a$. From the solution we get

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z) + z\left\{b(z) - a(z)\right\}\frac{z-F^*(1-p)}{z-(1-p)a(z)}}{b(z)-z}\right]$$

which when we take the limit as $p \downarrow 0$ gives

$$\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left[\frac{b(z)(1-z) + z\left\{b(z) - a(z)\right\}\frac{z-1}{z-a(z)}}{b(z)-z}\right] \\
&= \frac{1}{m}\left[\frac{b(z)(1-z)(a(z)-z) + z\left\{b(z) - a(z)\right\}(1-z)}{(a(z)-z)(b(z)-z)}\right] \\
&= \frac{1}{m}\left[\frac{\left\{b(z)(a(z)-z) + zb(z) - za(z)\right\}(1-z)}{(a(z)-z)(b(z)-z)}\right] \\
&= \frac{1}{m}\left[\frac{\left\{b(z)a(z) - za(z)\right\}(1-z)}{(a(z)-z)(b(z)-z)}\right] \\
&= \frac{1}{m}\left[\frac{(b(z)-z)a(z)(1-z)}{(a(z)-z)(b(z)-z)}\right] \\
&= \frac{1}{m}\left[\frac{a(z)(1-z)}{a(z)-z}\right],
\end{aligned}$$

which is the solution we expect for the M/G/1 queue. When $p = 1$ the system is the same as the M/G/1 queue with a different service-time distribution when a customer arrives at an empty server. From the result we get

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z) + z\left\{b(z) - a(z)\right\}\frac{z-F^*(0)}{z}}{b(z)-z}\right].$$

Now $F^*(0) = 0$ so that we get

$$\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left[\frac{b(z)(1-z) + z\left\{b(z) - a(z)\right\}}{b(z)-z}\right] \\
&= \frac{1}{m}\left[\frac{b(z) - za(z)}{b(z)-z}\right],
\end{aligned}$$

which is what we expect.

4.3.2 Modifications

No modification to the solution is necessary for any of the examples of modified server behaviour except using the correct generating functions for $a(z)$ and $b(z)$ from Section 3.4.

4.4 Fixed-Time Threshold

In this case the threshold is passed after a fixed number of customers have been served during the busy period. We label the number of customers served before the switch by $S \in I\!N$. Note that if the system clears before this threshold is reached then it is reset when the next busy period begins. This is so that times spent in phases in different busy periods will be independent. We can write

$$\begin{aligned} \tau_1(n) &= \begin{cases} \tau(n) \wedge \Big(n + S - \chi(n)\Big), & \text{if phase(n)} = 1, \\ n, & \text{if phase(n)} = 2, \end{cases} \\ &= \tau(n) \wedge \left[n + (S - \chi(n))^+\right], \end{aligned}$$

where $\chi(n)$ is the number of customers served since the beginning of the current busy period. The second equation occurs as $(S - \chi(n))$ will be negative when phase$(n) = 2$. Thus

$$\tau_1(0) = \tau(0) \wedge S. \tag{4.20}$$

It is worth noting at this point that condition $(*)$ is trivially satisfied for this type of threshold. This can be seen as when $\alpha > 1$ we have $\alpha^S \geq \alpha^{\tau_1(0)}$ with probability one.

Theorem 4.5 *For $\rho_a > 0$, $S \in I\!N$ and $z \in [0,1]$ we arrive at*

$$E\left[z^{X_{\tau_1(0)}}\right] - z = [a(z) - z] R_S^{(T)}(z),$$

where $R_S^{(T)}(z)$ can be defined recursively by

$$R_{S+1}^{(T)}(z) = \frac{1}{z}\left\{R_S^{(T)}(z) a(z) - a_0 R_S^{(T)}(0)\right\} + 1,$$

and $R_1^{(T)}(z) = 1$.

Proof: We use the stopped process Z_n defined in the following way.

$$Z_n = X_{n \wedge \tau(0)}.$$

Note that

$$E\left[z^{X_{\tau_1(0)}}\right] = E\left[z^{Z_S}\right],$$

so we shall consider Z_S rather than $X_{\tau_1(0)}$ throughout this proof. When $S = 1$

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] - z &= E\left[z^{Z_1}\right] - z \\ &= E\left[z^{A_1}\right] - z \\ &= a(z) - z, \end{aligned}$$

where A_1 is the number of arrivals during the first service. Clearly then $R_1^{(T)}(z) = 1$. This provides the starting point for an inductive proof. Assume the lemma is true for $S = n > 1$, so that

$$E\left[z^{Z_n}\right] - z = [a(z) - z]\, R_n^{(T)}(z).$$

Consider the case $S = n + 1$. We want to show that

$$E\left[z^{Z_{n+1}}\right] - z = [a(z) - z]\, R_{n+1}^{(T)}(z).$$

Now we can see for $S = n + 1$ that

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] - z &= E\left[z^{Z_{n+1}}\right] - z \\ &= \sum_{i=0}^{\infty} E\left[z^{Z_{n+1}} \middle| Z_n = i\right] p\{Z_n = i\} - z. \end{aligned}$$

We deduce that

$$E\left[z^{Z_{n+1}} \middle| Z_n = i\right] = \begin{cases} z^{i-1}a(z), & i > 0, \\ 1, & i = 0, \end{cases}$$

from the fact that when $Z_n = 0$, Z_{n+1} must also be zero and when $Z_n = i > 0$ then $Z_{n+1} = i + A_{n+1} - 1$ where A_{n+1} is the number of arrivals during the $(n+1)$th service. This means that

$$\begin{aligned} E\left[z^{Z_{n+1}}\right] - z &= p\{Z_n = 0\} + a(z)\sum_{i=1}^{\infty} z^{i-1} p\{Z_n = i\} - z \\ &= p\{Z_n = 0\} + \frac{a(z)}{z}\left\{E\left[z^{Z_n}\right] - p\{Z_n = 0\}\right\} - z. \end{aligned}$$

We can see that $p\{Z_n = 0\}$ will be $a_0 R_n^{(T)}(0)$ and we have assumed $E\left[z^{Z_n}\right] - z$ to be $[a(z) - z]\, R_n^{(T)}(z)$ so

$$\begin{aligned} E\left[z^{Z_{n+1}}\right] - z &= a_0 R_n^{(T)}(0) + \frac{a(z)}{z}\left\{[a(z) - z]R_n^{(T)}(z) + z - a_0 R_n^{(T)}(0)\right\} - z \\ &= [a(z) - z]\frac{R_n^{(T)}(z)a(z)}{z} + a_0 R_n^{(T)}(0) - \frac{a(z)a_0 R_n^{(T)}(0)}{z} + a(z) - z \\ &= [a(z) - z]\left\{\frac{R_n^{(T)}(z)a(z)}{z} - \frac{a_0 R_n^{(T)}(0)}{z} + 1\right\}, \end{aligned}$$

with some rearrangement. Thus we have inductively proved that $E\left[z^{X_{\tau_1(0)}}\right] - z$ contains a factor of $a(z) - z$ and also demonstrated the stated recursion. □

We still require a closed form for $R_S^{(T)}(z)$ for Theorem 4.2 equation (4.5). The following theorem provides it.

Theorem 4.6 *For the process discussed above*

$$R_S^{(T)}(z) = \left[\frac{1 - \xi_a(z)^S}{\xi_a(z)^{S-1}(1 - \xi_a(z))}\right] - \frac{(1 - \delta_{S1})}{z}\sum_{k=1}^{S-1}\left[\frac{1 - \xi_a(z)^{S-k}}{\xi_a(z)^{S-k-1}(1 - \xi_a(z))}\right] a^{(k)},$$

and

$$R_n^{(T)}(0) = \frac{1}{a_0}\sum_{k=1}^{n} a^{(k)},$$

where

$$a^{(k)} = \int_0^\infty \frac{e^{-\lambda x}(\lambda x)^{k-1}}{k!} dA^{(k)}(x),$$

$A^{(k)}(\cdot)$ *being the k-fold convolution of* $A(\cdot)$.

Proof: When we consider the geometrically-distributed random-time threshold the solution for $E\left[z^{X_{\tau_1(0)}}\right] - z$ could be written as

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] - z &= \sum_{n=1}^{\infty} \left[E\left[z^{X_{\tau_1(0)}} \middle| R = n\right] - z\right] p\{R = n\} \\ &= \sum_{n=1}^{\infty} \left[E\left[z^{X_{\tau_1(0)}} \middle| R = n\right] - z\right] p(1-p)^{n-1} \\ &= \frac{p}{1-p} \sum_{n=1}^{\infty} \left[E\left[z^{X_{\tau_1(0)}} \middle| R = n\right] - z\right] (1-p)^n. \end{aligned}$$

Now $\left[E\left[z^{X_{\tau_1(0)}} \middle| R = n\right] - z\right]$ is just $(a(z) - z)R_n^{(T)}(z)$ where $R_n^{(T)}(z)$ is defined for a fixed-time threshold at $S = n$. Thus we can write

$$\begin{aligned} R_p^{(G)}(z) &= \frac{p}{1-p} \left[a(z) - z\right] \sum_{n=1}^{\infty} R_n^{(T)}(z)(1-p)^n \\ &= \frac{p}{1-p} \left[a(z) - z\right] R(z, 1-p), \end{aligned} \tag{4.21}$$

where

$$R(z,p) = \sum_{n=1}^{\infty} R_n^{(T)}(z) p^n.$$

By comparing this with the answer in Theorem 4.4 which states

$$E\left[z^{X_{\tau_1(0)}}\right] - z = \frac{z - F^*(1-p)}{z - (1-p)a(z)} \left(a(z) - z\right),$$

we get

$$\begin{aligned} R(z, 1-p) &= \frac{1-p}{p} . \frac{z - F^*(1-p)}{z - (1-p)a(z)} \\ R(z,p) &= \frac{p}{1-p} . \frac{z - F^*(p)}{z - pa(z)}. \end{aligned}$$

This in itself is interesting but it also gives us a way of calculating $R_n^{(T)}(z)$ explicitly, namely by expanding the right-hand side in terms of p and equating coefficients of p. First we expand $1/(1-p)$ as $p < 1$ to get

$$\begin{aligned} R(z,p) &= \frac{p}{1-p} \frac{z - F^*(p)}{z - pa(z)} \\ &= p \sum_{i=0}^{\infty} p^i \frac{1 - \frac{1}{z}F^*(p)}{1 - p\frac{a(z)}{z}}. \end{aligned}$$

Next we expand $\frac{1}{1-pa(z)/z}$ which we can do for some interval $[z_0,1]$ as when z tends to one, $a(z)/z$ tends to one and $p<1$. This gives

$$\begin{aligned} R(z,p) &= p\sum_{i=0}^{\infty} p^i \sum_{j=0}^{\infty} \left(p\frac{a(z)}{z}\right)^j \left\{1-\frac{1}{z}F^*(p)\right\} \\ &= p\sum_{i=0}^{\infty} p^i \sum_{j=0}^{i} \left(\frac{a(z)}{z}\right)^j \left\{1-\frac{1}{z}F^*(p)\right\} \\ &= \sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] \left\{1-\frac{1}{z}F^*(p)\right\}. \end{aligned}$$

Now from Takács's lemma

$$F^*(p) = \sum_{k=1}^{\infty} p^k \int_0^{\infty} \frac{e^{-\lambda x}(\lambda x)^{k-1}}{k!} dA^{(k)}(x),$$

where $A^{(k)}(\cdot)$ is the k-fold convolution of $A(\cdot)$ with itself. We shall write

$$a^{(k)} = \int_0^{\infty} \frac{e^{-\lambda x}(\lambda x)^{k-1}}{k!} dA^{(k)}(x).$$

Thus

$$\begin{aligned} R(z,p) &= \sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] - \frac{1}{z}\sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] F^*(p) \\ &= \sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] - \frac{1}{z}\sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] \sum_{k=1}^{\infty} p^k a^{(k)} \\ &= \sum_{i=1}^{\infty} p^i \left[\frac{1-\left(\frac{a(z)}{z}\right)^i}{1-\frac{a(z)}{z}}\right] - \frac{1}{z}\sum_{i=2}^{\infty} p^i \sum_{k=1}^{i-1} \left[\frac{1-\left(\frac{a(z)}{z}\right)^{i-k}}{1-\frac{a(z)}{z}}\right] a^{(k)}. \end{aligned}$$

By equating coefficients we get

$$R_n^{(T)}(z) = \left[\frac{1-\left(\frac{a(z)}{z}\right)^n}{1-\frac{a(z)}{z}}\right] - \frac{(1-\delta_{n1})}{z} \sum_{k=1}^{n-1} \left[\frac{1-\left(\frac{a(z)}{z}\right)^{n-k}}{1-\frac{a(z)}{z}}\right] a^{(k)},$$

for $z \in (z_0,1]$ where $pa(z_0)/z_0 = 1$. Now as $R_n^{(T)}(z)$ should not in any way be dependent on p we can take the limit as p tends to zero. Thus the above is true for $z \in (0,1]$

To get $R_n^{(T)}(0)$ we can use the fact that $R_n^{(T)}(z)$ satisfies the recursive relationship

$$R_{n+1}^{(T)}(z) = \frac{1}{z}\left[R_n^{(T)}(z)a(z) - R_n^{(T)}(0)a_0\right] + 1,$$

with $R_1^{(T)}(z) = 1$. This can be used as follows. We set

$$R(z,p) = \sum_{n=1}^{\infty} R_n^{(T)}(z)p^n$$

$$
\begin{aligned}
&= p + \sum_{n=2}^{\infty} R_n^{(T)}(z) p^n \\
&= p + \sum_{n=2}^{\infty} \left\{ \frac{1}{z} \left[R_{n-1}^{(T)}(z) a(z) - R_{n-1}^{(T)}(0) a_0 \right] + 1 \right\} p^n \\
&= p + \frac{1}{z} \sum_{n=2}^{\infty} \left[R_{n-1}^{(T)}(z) a(z) - R_{n-1}^{(T)}(0) a_0 \right] p^n + \sum_{n=2}^{\infty} p^n \\
&= p + \frac{p^2}{1-p} + p \frac{1}{z} \sum_{n=1}^{\infty} \left[R_n^{(T)}(z) a(z) - R_n^{(T)}(0) a_0 \right] p^n \\
&= \frac{p}{1-p} + p \frac{1}{z} \left[R(z,p) a(z) - R(0,p) a_0 \right] p^n,
\end{aligned}
$$

which with some rearrangement gives

$$
\begin{aligned}
R(z,1-p) \left\{ 1 - p \frac{a(z)}{z} \right\} &= p \left[\frac{1}{p} - \frac{a_0}{z} R(0,1-p) \right], \\
R(z,1-p) &= p \frac{\left[\frac{1}{p} - \frac{a_0}{z} R(0,1-p) \right]}{1 - p \frac{a(z)}{z}} \\
&= p \frac{\left[z \frac{1}{p} - a_0 R(0,1-p) \right]}{z - p a(z)}.
\end{aligned}
$$

From the geometrically-distributed random-time threshold we have (4.21)

$$
\begin{aligned}
E\left[z^{X_{\tau_1(0)}} \right] - z &= \frac{p}{1-p} R(z,1-p)[a(z) - z] \\
&= \frac{z - p a_0 R(0,1-p)}{z - (1-p) a(z)} [a(z) - z]. \qquad (4.22)
\end{aligned}
$$

By comparing (4.22) and Theorem 4.4 we see that

$$
p a_0 R(0,1-p) = F^*(1-p),
$$

and so

$$
\begin{aligned}
R(0,p) &= \frac{1}{a_0(1-p)} F^*(p) \\
&= \frac{1}{a_0} \sum_{i=0}^{\infty} p^i \sum_{k=1}^{\infty} p^k a^{(k)} \\
&= \frac{1}{a_0} \sum_{i=1}^{\infty} p^i \sum_{k=1}^{i} a^{(k)}.
\end{aligned}
$$

By equating coefficients we get

$$
R_n^{(T)}(0) = \frac{1}{a_0} \sum_{k=1}^{n} a^{(k)},
$$

which is the required value. □

4.4.1 Some limiting cases

The case of the M/G/1 queue with a different service-time distribution for customers arriving at an empty server is the same as the case with $S = 1$. In this case we get $R_1^{(T)}(z) = 1$ so that the solution is simply

$$\begin{aligned} E\left[z^X\right] &= \frac{1}{m}\left[\frac{b(z)(1-z) + \{b(z) - a(z)\}\, z}{b(z) - z}\right] \\ &= \frac{1}{m}\left[\frac{b(z) - z a(z)}{b(z) - z}\right], \end{aligned}$$

which is as expected.

4.4.2 Modifications

As before no modifications to the solution are necessary for any of the examples save using the correct generating functions for $a(z)$ and $b(z)$.

4.5 Other random thresholds

We would like to be able to find a solution if the random variable R is not geometrically distributed, for instance if the time of a breakdown in the system depends in some way on the number of customers served in the current busy period.

If the random variable R has probability function $h(\cdot)$ then the solution will be of the form

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z)+\{b(z)-a(z)\}\,zR_h^{(R)}(z)}{b(z)-z}\right],$$

for $z \in [0,1)$ where

$$R_h^{(R)}(z) = \sum_{n=1}^{\infty} h(n)R_n^{(T)}(z),$$

providing we can satisfy condition $(*)$. Condition $(*)$ is easily satisfied if there exists an N such that $h(n) = 0$ for all $n > N$. Otherwise it might be difficult to provide condition $(*)$. This would have to be dealt with on an individual basis.

4.6 The probability of a given phase

One thing that may be of use in these results is the probability of being in a given phase. For example in order to calculate the cost of running the queue, given the costs for running it in phase 1 and 2. This might be used to optimise a queueing system. Another situation in which this information might prove useful is the breakdown model, in which we may want to know how many customers are affected by a breakdown. In this section we shall use Little's law to calculate these probabilities. Little's law (1961) states

$$L = \lambda W, \tag{4.23}$$

where L is the mean number of customers in the system, λ is the arrival rate to the system and W is the mean time spent by a customer in the system. If we apply this to the server alone we can see that L is the probability that there is a customer in the system and W is the mean service time so

$$\begin{aligned} L &= p\{X \neq 0\} = 1 - \frac{1}{m}, \\ W &= \frac{1-\phi}{\mu_a} + \frac{\phi}{\mu_b}, \end{aligned}$$

where ϕ is the probability of of being in phase 2. We have from Theorem 4.2 that

$$m = \left[\frac{1 + \{\rho_a - \rho_b\} R_q^{(F)}(1)}{1 - \rho_b} \right],$$

where $R_q^{(F)}(z)$ is determined by the specific type of threshold between the phases. Thus

$$\begin{aligned} L &= \frac{\rho_b + (\rho_a - \rho_b) R_q^{(F)}(1)}{1 + (\rho_a - \rho_b) R_q^{(F)}(1)}, \\ \lambda W &= \phi(\rho_b - \rho_a) + \rho_b. \end{aligned}$$

Substituting in (4.23) we get an equation for ϕ (when $\rho_a \neq \rho_b$)

$$\begin{aligned} \phi &= \left[\frac{\rho_b + (\rho_a - \rho_b) R_q^{(F)}(1)}{1 + (\rho_a - \rho_b) R_q^{(F)}(1)} - \rho_a \right] \frac{1}{\rho_b - \rho_a} \\ &= \left[\frac{\rho_b - \rho_a + (1 - \rho_a)(\rho_a - \rho_b) R_q^{(F)}(1)}{1 + (\rho_a - \rho_b) R_q^{(F)}(1)} \right] \frac{1}{\rho_b - \rho_a} \\ &= \frac{1 + (\rho_a - 1) R_q^{(F)}(1)}{1 + (\rho_a - \rho_b) R_q^{(F)}(1)}. \end{aligned} \tag{4.24}$$

This is the probability of being in phase 2. It is worth noting that this is insensitive to the actual distribution $B(\cdot)$ except through ρ_b, the traffic intensity during phase 2. Because of this insensitivity we can calculate ϕ even when $\rho_a = \rho_b$ by taking a set a distributions such that $\lim \rho_b \to \rho_a$ and using L'Hôpital's rule in the previous working to get

$$\phi = 1 + (\rho_a - 1) R_q^{(F)}(1).$$

As an example we investigate the random threshold where

$$\begin{aligned} R_q^{(F)}(z) &= R_p^{(R)}(z) \\ &= \frac{z - F^*(1-p)}{z - (1-p)a(z)}, \end{aligned}$$

which when we take $z = 1$ gives

$$R_q^{(F)}(1) = \frac{1}{p}[1 - F^*(1-p)].$$

When we substitute this back into (4.24) we get

$$\phi = \frac{p + (\rho_a - 1)[1 - F^*(1-p)]}{p + (\rho_a - \rho_b)[1 - F^*(1-p)]}.$$

In the breakdown model where all of the customers in the system at the time of a breakdown are discarded $\rho_b = 0$. In this case ϕ is the probability that a customer is discarded because of a breakdown and it is given by

$$\begin{aligned} \phi &= \frac{p + (\rho_a - 1)[1 - F^*(1-p)]}{p + \rho_a[1 - F^*(1-p)]} \\ &= 1 - \frac{1 - F^*(1-p)}{p + \rho_a[1 - F^*(1-p)]}. \end{aligned}$$

4.6.1 The length of the phases

Another thing that might be of interest is the time spent in each of the two phases. As we have seen in the previous sections the generating functions $E\left[z^{\tau_1(0)}\right]$ are quite complex and depend on the specific threshold and we may suppose the same about $E\left[z^{\tau_2(0)}\right]$. We shall just consider the averages which are relatively easy to calculate. $\nu_1(0)$ and $\nu_2(0)$ give the number of customers served during phase 1 and 2 respectively of the busy period. We calculate the expected values which are given by equations (4.29) and (4.30). The calculation of these values is as follows

$$\begin{aligned} \frac{d}{dz}\left[F^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right)\right] &= \left(\frac{a(z) - za'(z)}{a(z)^2}\right) E\left[\nu_1(0)\left(\frac{z}{a(z)}\right)^{\nu_1(0)-1}\left(\frac{z}{b(z)}\right)^{\nu_2(0)}\right] \\ &\quad + \left(\frac{b(z) - zb'(z)}{b(z)^2}\right) E\left[\nu_2(0)\left(\frac{z}{a(z)}\right)^{\nu_1(0)}\left(\frac{z}{b(z)}\right)^{\nu_2(0)-1}\right], \\ \frac{d}{dz}\left[F^*\left(1, \frac{z}{b(z)}\right)\right] &= \left(\frac{b(z) - zb'(z)}{b(z)^2}\right) E\left[\nu_2(0)\left(\frac{z}{b(z)}\right)^{\nu_2(0)-1}\right], \end{aligned}$$

from which we get

$$\frac{d}{dz}\left[F^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right)\right]_{z=1} = (1-\rho_a)\,E\left[\nu_1(0)\right] + (1-\rho_b)\,E\left[\nu_2(0)\right], \tag{4.25}$$

$$\frac{d}{dz}\left[F^*\left(1, \frac{z}{b(z)}\right)\right]_{z=1} = (1-\rho_b)\,E\left[\nu_2(0)\right]. \tag{4.26}$$

Now we have, from previous working, the following two equations

$$\begin{aligned} F^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right) &= z, \\ F^*\left(1, \frac{z}{b(z)}\right) &= [a(z) - z]\, R_q^{(F)}(z) + z. \end{aligned}$$

Taking the derivatives of these and taking $\lim_{z \uparrow 1}$ we get from equations (4.25) and (4.26).

$$(1-\rho_a)E[\nu_1(0)] + (1-\rho_b)E[\nu_2(0)], = 1, \tag{4.27}$$

$$(1-\rho_b)E[\nu_2(0)] = (\rho_a - 1)R_q^{(F)}(1) + 1. \tag{4.28}$$

Substituting (4.28) into (4.27) and then using the resultant expression for $E[\nu_1(0)]$ in (4.27) we arrive at

$$E[\nu_1(0)] = R_q^{(F)}(1), \tag{4.29}$$

$$E[\nu_2(0)] = \frac{1 + (\rho_a - 1)R_q^{(F)}(1)}{1 - \rho_b}. \tag{4.30}$$

Note that by adding (4.29) and 4.30) together we get

$$E[\nu(0)] = \frac{1 + (\rho_a - \rho_b)R_q^{(F)}(1)}{1 - \rho_b},$$

which agrees with our value for m.

4.7 Summary

As this chapter contains solutions to problems using the technique described in Chapter 3 now would seem an appropriate moment to summarise what has been done so far. Chapters 2 and 3 provide a powerful result using Markov renewal theory and some martingale results. This has been applied to a number of different two-phase problems in this chapter. The equilibrium probability generating functions for three problems are given in Theorem 4.2. In this chapter, three examples, each with a different type of threshold between the phases, are considered and a general form of solution is found. The three thresholds considered are the fixed upward threshold, the fixed-time threshold and a geometrically-distributed random-time threshold. Each of these has a different technique for finding the final solution and so each is considered in its own section of this chapter. For each of the thresholds two major results must be obtained.

The first is simply to demonstrate when condition $(*)$ is satisfied. This condition being sufficient to use the results of Chapter 3. In conjunction with this result we have also calculate the values of $E\left[\omega^{\tau_1(0)}\right]$ the probability generating function for the length of the first phase.

The second result necessary for a useful solution is the value of $E\left[z^{X_{\tau_1(0)}}\right] - z$. This can be given in each case in the following form

$$E\left[z^{X_{\tau_1(0)}}\right] - z = [a(z) - z]\, R(z),$$

and so we have throughout used the notation $R_q^{(F)}(z)$ to denote $R(z)$ where F gives the type of threshold and q is replaced with the type of parameter relevant to the threshold. Both this and the previous result are proven using a number of standard probabilistic techniques.

The results we have obtained are then used to calculate a number of quantities of interest in the study of such systems, namely the probability of being in a particular phase during equilibrium and the mean number of customers served in each phase during a busy period.

Finally we shall comment upon the solutions obtained. The final result is in an elegant form. The equilibrium distribution of customers in the queue is given by the probability generating function

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z)+\{b(z)-a(z)\}\,zR_q^{(F)}(z)}{b(z)-z}\right],$$

for $z \in [0,1)$ and with

$$m = \left[\frac{1+\{\rho_a-\rho_b\}\,R_q^{(F)}(1)}{1-\rho_b}\right],$$

where $R_q^{(F)}(z)$ is a non-negative function bounded above on the interval $[0,1]$ and is determined by the specific type of threshold between the phases. The value F specifies the type of threshold used and q is a parameter associated with the type of threshold. A standard M/G/1 queue with service-time distribution given by $B(\cdot)$ would have the probability generating function for the equilibrium distribution of customers in the queue given by

$$E\left[z^X\right] = (1-\rho)\left[\frac{b(z)(1-z)}{b(z)-z}\right].$$

Thus we can see that our solution is a modification of this solution. The modification is proportional to the difference of the probability generating functions $b(z)$ and $a(z)$. This can be seen to make sense by considering the following. Our systems are all simply M/G/1 queues with modified initial behaviour. (The behaviour during phase 1.) Thus we should expect the solution to be that of a M/G/1 queue with some sort of modification. That this modification is proportional to $b(z)-a(z)$ is of some interest.

We should, at this time, note the work of Fuhrmann and Cooper (1985) which deals with the stochastic decomposition of the M/G/1 queue with generalised vacations. This is an M/G/1 queue in which the server is unavailable for certain periods of time. For certain of these systems they obtain a result which states;

> The (stationary) number of customers in the system at a random point in time is distributed as the sum of two or more independent random variables, one of which is the (stationary) number of customers present in the corresponding M/G/1 queue at a random point in time.

We might, using some imagination, rearrange the scheme we have used herein to describe multi-phase M/G/1 queues in terms of generalised vacations. To do this the queueing

system must satisfy Assumptions 1-5 of Fuhrmann and Cooper. However, Proposition 2 of Fuhrmann and Cooper's paper require a further assumption, Assumption 6, which our systems violate and it is not clear how the other propositions could be usefully applied. (This assumption requires the number of arrivals during a vacation to be independent of the number of customers in the system at the start of the vacation.) We have, in a sense, obtained our own decomposition result however. This is not in terms of independent random variables but dependent random variables as we sum two generating function, not multiply. This could in itself be of interest in future research.

The solutions may look complex but they are not computationally hard to calculate. For instance the solution for the fixed upward threshold requires the inversion of a k x k matrix, however, the matrix is already in lower Hessenburg form. Putting a matrix into Hessenburg form is a major part of one of the better computational procedures for inverting matrices (Golub and van Loan (1983)) and so this matrix inversion is roughly an order of magnitude easier than an ordinary inversion. The solution for the fixed-time threshold is not difficult either. It can be done through a set of discrete convolutions using the fact that $a^{(k)}$ is $1/k$ times the probability that there are $(k-1)$ arrivals during k services so that

$$\begin{aligned} a^{(1)} &= a_0, \\ a^{(2)} &= a_1 a_0, \\ a^{(3)} &= a_1^2 a_0 + a_2 a_1 a_0^2, \\ &\vdots \end{aligned}$$

Also the geometrically distributed random threshold relies on $F^*(z)$ which is a standard function for the M/G/1 queue, the probability generating function for the number of customers served during the busy period. Thus the results are in a useful form for calculation.

In the next chapter we shall consider the slightly more complex case of three phases.

Chapter 5

Three-phase examples

In this chapter we consider an example with three phases. This is a simple extension of the fixed upward threshold example considered in the previous chapter. In this case we have two fixed upward thresholds at k_1 and k_2. The ends of phases 1 and 2 correspond to the times at which the system has more than k_1 and k_2 customers in it respectively.

We also describe a new type of threshold in Section 5.2, the fixed downward threshold. This is the case when a phase ends if there are fewer than a certain number of customers in the system immediately after a service completion.

We use a three-phase MRP which is shown in Figure 5.1 and we use the standard notation defined in Chapter 3. Throughout this chapter we use the matrix

$$\mathbf{P}_{k_i} = \begin{pmatrix} a_1^i & a_2^i & a_3^i & \cdots & a_{k_2-1}^i & a_{k_2}^i \\ a_0^i & a_1^i & a_2^i & \cdots & a_{k_2-2}^i & a_{k_2-1}^i \\ 0 & a_0^i & a_1^i & \cdots & a_{k_2-3}^i & a_{k_2-2}^i \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & a_0^i & a_1^i \end{pmatrix}, \tag{5.1}$$

which is analogous to the matrix $\mathbf{P}_k$ in the previous chapter. Several of the proofs that follow are also analogous to those in the previous chapter because each of the thresholds is not qualitatively different from the single upward threshold in Section 4.2.

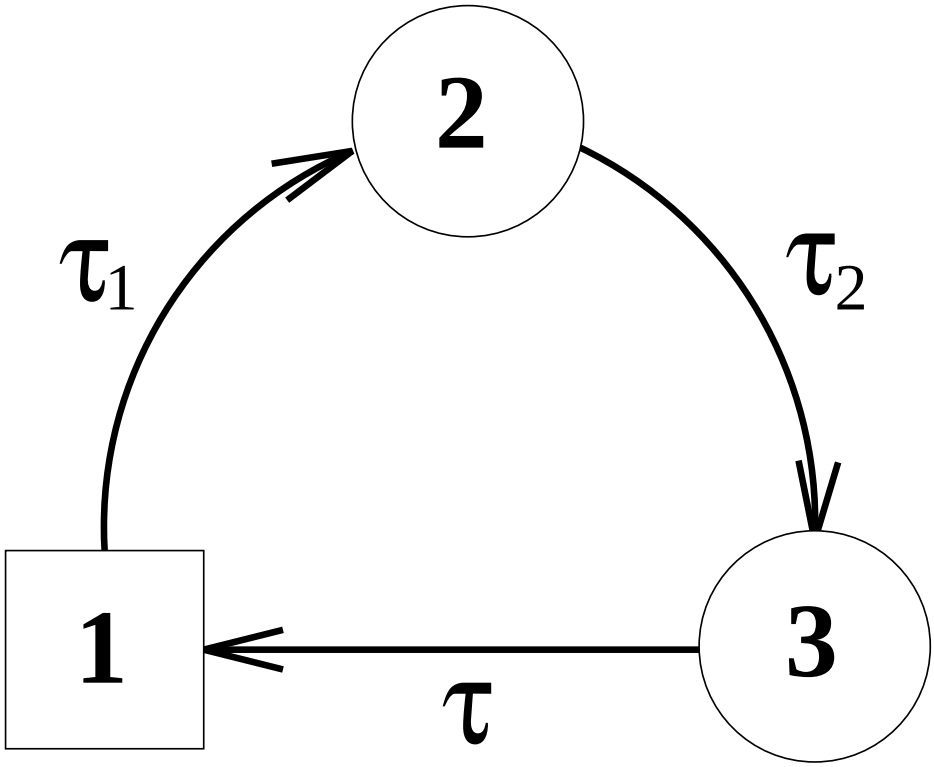

Figure 5.1: A three-phase MRP. □ denotes a renewal point while ◯ denotes a non-renewal point

5.1 Two fixed upward thresholds

In this case we take the end of phase i, (i=1,2) to be when the system first has more than k_i customers in it immediately after a service, or of course the end of the busy period. So we can write

$$\begin{aligned}\tau_1(n) &= I_{C_n^1}\left[\tau(n) \wedge \inf\{m > n | X_m > k_1\}\right] + \left(I_{C_n^2} + I_{C_n^3}\right) n, \\ \tau_2(n) &= \left(I_{C_n^1} + I_{C_n^2}\right)\left[\tau(n) \wedge \inf\{m > n | X_m > k_2\}\right] + I_{C_n^3} n.\end{aligned}$$

If $\rho_1 > 0$ and $\rho_3 > 1$, Lemma 3.1.2 shows that the queue is unstable. Thus we shall consider the case with $\rho_3 \leq 1$. We must show that condition $(*)$ is satisfied. We do this in the following lemma.

Lemma 5.0.1 *For $\rho_1 \geq 0$, $\rho_2 \geq 0$ and $\rho_3 \leq 1$ condition $(*)$ is satisfied.*

Proof: In this case condition $(*)$ is satisfied if $E\left[\alpha_1^{\tau_1(0)} \alpha_2^{\tau_2(0)-\tau_1(0)}\right] < \infty$, where $\alpha_i = \sup_{z\in[0,1]} \xi_i(z)$.

$$\begin{aligned}E\left[\alpha_1^{\tau_1(0)} \alpha_2^{\tau_2(0)-\tau_1(0)}\right] &= \sum_{i=0}^{\infty} E\left[\alpha_1^{\tau_1(0)} \alpha_2^{\tau_2(0)-\tau_1(0)} I(X_{\tau_1(0)} = i)\right] \\ &= \sum_{i=0}^{\infty} E\left[\alpha_1^{\tau_1(0)} I(X_{\tau_1(0)} = i)\right] E\left[\alpha_2^{\tau_2(0)-\tau_1(0)} I(X_{\tau_1(0)} = i)\right],\end{aligned}$$

as the times $\tau_1(0)$ and $\tau_2(0) - \tau_1(0)$ are independent given the value of $X_{\tau_1(0)}$. We know from Lemma 4.2.1 that $E\left[\alpha_1^{\tau_1(0)}\right]$ exists and is finite and therefore $E\left[\alpha_1^{\tau_1(0)} I(X_{\tau_1(0)} = i)\right]$ also exists and is finite. It can be shown in exactly the same fashion as in Lemma 4.2.1 that $E\left[\alpha_2^{\tau_2(0)-\tau_1(0)} I(X_{\tau_1(0)} = i)\right]$ exists and is finite and so we have condition $(*)$. □

Theorem 5.1 *For $\rho_1 > 0$, $\rho_2 \geq 0$ and $\rho_3 < 1$, $z \in [0, 1)$ the probability generating function for the equilibrium number of customer in the system is*

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{a_3(z)(1-z) + \{a_3(z) - a_1(z)\} z R_{k_1}^{(U)}(z) + \{a_3(z) - a_2(z)\} z R_{k_1,k_2}^{(UU)}(z)}{a_3(z) - z}\right],$$

where $R_{k_1}^{(U)}(z)$ is defined in Theorem 4.2 and

$$R_{k_1,k_2}^{(UU)}(z) = \mathbf{w}^1 (\mathbf{I} - \mathbf{P}_{k_2})^{-1} \mathbf{z}^t,$$

$\mathbf{w}^1 = \left(p\{X_{\tau_1(0)} = 1\}, p\{X_{\tau_1(0)} = 2\}, \ldots, p\{X_{\tau_1(0)} = k_2\}\right)$, $\mathbf{P}_{k_i}$ *is defined by (5.1) and*

$$m = \left[\frac{1 + \{\rho_1 - \rho_3\} R_{k_1}^{(U)}(1) + \{\rho_2 - \rho_3\} R_{k_1,k_2}^{(UU)}(1)}{1 - \rho_3}\right].$$

Proof: Note that as condition $(*)$ is satisfied (Lemma 5.0.1), Theorem 3.9 gives

$$\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left[\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{E\left[z^{X_{\tau_2(0)}}\right]-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}+\frac{1-E\left[z^{X_{\tau_2(0)}}\right]}{1-\xi_3(z)}\right] \\
&= \frac{1}{m}\left[\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}-\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_2(z)}\right. \\
&\quad \left.+\frac{E\left[z^{X_{\tau_2(0)}}\right]-z}{1-\xi_2(z)}-\frac{E\left[z^{X_{\tau_2(0)}}\right]-z}{1-\xi_3(z)}+\frac{1-z}{1-\xi_3(z)}\right] \\
&= \frac{1}{m}\left[\frac{\left[\xi_1(z)-\xi_2(z)\right]\left[E\left[z^{X_{\tau_1(0)}}\right]-z\right]}{(1-\xi_1(z))(1-\xi_2(z))}\right. \\
&\quad \left.+\frac{\left[\xi_2(z)-\xi_3(z)\right]\left[E\left[z^{X_{\tau_2(0)}}\right]-z\right]}{(1-\xi_2(z))(1-\xi_3(z))}+\frac{1-z}{1-\xi_3(z)}\right] \\
&= \frac{1}{m}\left[\frac{z\left\{a_2(z)-a_1(z)\right\}\left[E\left[z^{X_{\tau_1(0)}}\right]-z\right]}{(a_1(z)-z)(a_2(z)-z)}\right. \\
&\quad \left.+\frac{z\left\{a_3(z)-a_2(z)\right\}\left[E\left[z^{X_{\tau_2(0)}}\right]-z\right]}{(a_2(z)-z)(a_3(z)-z)}+\frac{a_3(z)(1-z)}{a_3(z)-z}\right]. \qquad (5.2)
\end{aligned}$$

The following lemma gives the important values of $E\left[z^{X_{\tau_1(0)}}\right]-z$ and $E\left[z^{X_{\tau_2(0)}}\right]-z$.

Lemma 5.1.1 *For $\rho_1 > 0$, $z \in [0,1]$, $X_0 = 0$ and the thresholds $1 \le k_1 < k_2$ we get*

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right]-z &= (a_1(z)-z)\,R_{k_1}^{(U)}(z), \\
E\left[z^{X_{\tau_2(0)}}\right]-z &= (a_1(z)-z)\,R_{k_1}^{(U)}(z)-(a_2(z)-z)\,R_{k_1,k_2}^{(UU)}(z),
\end{aligned}$$

where $R_{k_1}^{(U)}(z)$ is defined in Theorem 4.2,

$$R_{k_1,k_2}^{(UU)}(z) = \mathbf{w}^1(\mathbf{I}-\mathbf{P}_{k_2})^{-1}\mathbf{z}^t$$

and $\mathbf{w}^1 = \left(p\{X_{\tau_1(0)}=1\}, p\{X_{\tau_1(0)}=2\}, \ldots, p\{X_{\tau_1(0)}=k_2\}\right)$ and $\mathbf{P}_{k_i}$ is defined by (5.1).

Proof: The derivation of $E\left[z^{X_{\tau_1(0)}}\right]$ remains unchanged from Theorem 4.3 in Section 4.2 except for the slightly altered notation. Hence

$$E\left[z^{X_{\tau_1(0)}}\right]-z = (a_1(z)-z)\,R_{k_1}^{(U)}(z),$$

where $R_{k_1}^{(U)}(z) = \frac{1}{z}\,\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_{k_1})^{-1}\mathbf{z}^t$ and $\mathbf{P}_{k_1}$ is the $k_1 \times k_1$ sub-stochastic matrix defined in (5.1). However $E\left[z^{X_{\tau_2(0)}}\right]$ must be dealt with slightly differently.
There are three possible outcomes for $X_{\tau_1(0)}$.

(i) $X_{\tau_1(0)} = 0$ which implies $X_{\tau_2(0)} = X_{\tau_1(0)} = 0$.
(ii) $X_{\tau_1(0)} > k_2$ which implies $X_{\tau_2(0)} = X_{\tau_1(0)}$.
(iii) $X_{\tau_1(0)} = k_1 + 1, \ldots, k_2$.

Thus we arrive at the following result.

$$\begin{aligned} E\left[z^{X_{\tau_2(0)}}\right] - z &= E\left[z^{X_{\tau_1(0)}}\right] - z \\ &\quad + \sum_{i=k_1+1}^{k_2} p\{X_{\tau_1(0)} = i\}\left\{E\left[z^{X_{\tau_2(0)}}\middle| X_{\tau_1(0)} = i\right] - E\left[z^{X_{\tau_1(0)}}\middle| X_{\tau_1(0)} = i\right]\right\} \\ &= (a_1(z) - z)R_{k_1}^{(U)}(z) + \sum_{i=k_1+1}^{k_2} p\{X_{\tau_1(0)} = i\}\left\{E\left[z^{X_{\tau_2(0)}}\middle| X_{\tau_1(0)} = i\right] - z^i\right\}. \end{aligned}$$

Now for $i = k_1 + 1, \ldots, k_2$ we can write $E\left[z^{X_{\tau_2(0)}}\middle| X_{\tau_1(0)} = i\right]$ as

$$\sum_{m=\tau_1(0)+1}^{\infty} \sum_{j=1}^{k} E\left[z^{X_m} I(\tau_2(0) = m)\middle| X_{m-1} = j, \tau_2(0) > m-1, X_{\tau_1(0)} = i\right] \\ \text{x } p\{X_{m-1} = j, \tau_2(0) > m-1 | X_{\tau_1(0)} = i\}.$$

The value of $E\left[z^{X_m} I(\tau_2(0) = m)\middle| X_{m-1} = j, \tau_2(0) > m-1, X_{\tau_1(0)} = i\right]$ is given by

$$g_j^2(z) = a_2(z)z^{j-1} - \sum_{l=(j-1)\vee 1}^{k_2} a_{l-j+1}^2 z^l.$$

Thus we arrive at the equation

$$E\left[z^{X_{\tau_2(0)}}\middle| X_{\tau_1(0)} = i\right] = \sum_{m=\tau_1(0)}^{\infty}\left(\sum_{j=1}^{k} p\{X_m = j, \tau_2(0) > m | X_{\tau_1(0)} = i\} g_j^2(z)\right).$$

Now $p\{X_m = j, \tau_2(0) > m | X_{\tau_1(0)} = i\}$ is the probability of being in state j and still in phase 2 after the mth service, given that $m \geq \tau_1(0)$ and $X_{\tau_1(0)} = i$ for $i = k_1 + 1, \ldots, k_2$. In order to find $\sum_{m=\tau_1(0)}^{\infty} p\{X_m = j, \tau_2(0) > m | X_{\tau_1(0)} = i\}$ we define $\mathbf{v}_i^m$ as the row vector

$$\begin{aligned} \mathbf{v}_i^m = \quad & (\quad p\{X_m = 1, \tau_2(0) > m | X_{\tau_1(0)} = i\}, \\ & \quad p\{X_m = 2, \tau_2(0) > m | X_{\tau_1(0)} = i\} \\ & \quad , \ldots, p\{X_m = k_2, \tau_2(0) > m | X_{\tau_1(0)} = i\} \ ), \end{aligned} \tag{5.3}$$

which is the probability vector for phase 2 after the mth transition given that $X_{\tau_1(0)} = i$ and $m \geq \tau_1(0)$. The initial probability vector $\mathbf{v}_i^{\tau_1(0)} = \mathbf{e}_i$. If $\mathbf{P}_{k_2}$ is the sub-stochastic probability transfer matrix defined by (5.1) then

$$\mathbf{v}_i^{\tau_1(0)+m} = \mathbf{e}_i \mathbf{P}_{k_2}^m,$$

so that

$$\begin{aligned} \sum_{m=\tau_1(0)}^{\infty} p\{X_m = j, \tau_2(0) > m-1 | X_{\tau_1(0)} = i\} &= \mathbf{e}_i \sum_{m=0}^{\infty} \mathbf{P}_{k_2}^m \\ &= \mathbf{e}_i \left(\mathbf{I} - \mathbf{P}_{k_2}\right)^{-1}. \end{aligned}$$

Taking $\mathbf{g}(z) = (g_1(z), g_2(z), \cdots, g_{k_2}(z))$, we can see that

$$E\left[z^{X_{\tau_2(0)}} \middle| X_{\tau_1(0)} = i\right] \;=\; \mathbf{e}_i\,(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\,\mathbf{g}^2(z)^t. \tag{5.4}$$

We can simplify $\mathbf{g}^2(z)^t$ as in the proof of Theorem 4.3 to get

$$\mathbf{g}^2(z)^t \;=\; \frac{a_2(z)}{z}\mathbf{z}^t - \mathbf{P}_{k_2}\,\mathbf{z}^t.$$

Hence we can write $E\left[z^{X_{\tau_2(0)}} \middle| X_{\tau_1(0)} = i\right]$ as

$$E\left[z^{X_{\tau_2(0)}} \middle| X_{\tau_1(0)} = i\right] \;=\; \mathbf{e}_i\,(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\left(\frac{a_2(z)}{z}\mathbf{z}^t - \mathbf{P}_{k_2}\,\mathbf{z}^t\right),$$

which can be simplified as in Theorem 4.3 to get

$$E\left[z^{X_{\tau_2(0)}} \middle| X_{\tau_1(0)} = i\right] \;=\; \frac{1}{z}\,(a_2(z) - z)\,\mathbf{e}_i\,(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\mathbf{z}^t + z^i.$$

(Note that we get z^i not just z as the extra term.) Now using this result gives us

$$\begin{aligned} E\left[z^{X_{\tau_2(0)}}\right] &= (a_1(z) - z)R_{k_1}^{(U)}(z) \\ &\quad + \sum_{i=k_1+1}^{k_2} p\{X_{\tau_1(0)} = i\}\left[\frac{1}{z}(a_2(z) - z)\mathbf{e}_i\,(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\mathbf{z}^t + z^i - z^i\right] \\ &= (a_1(z) - z)R_{k_1}^{(U)}(z) + \frac{1}{z}(a_2(z) - z)\sum_{i=1}^{k_2} p\{X_{\tau_1(0)} = i\}\mathbf{e}_i\,(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\mathbf{z}^t, \end{aligned}$$

because for $i = 1, \ldots, k_1$ we get $p\{X_{\tau_1(0)} = i\} = 0$. From this we get

$$E\left[z^{X_{\tau_2(0)}}\right] \;=\; (a_1(z) - z)R_{k_1}^{(U)}(z) + (a_2(z) - z)\frac{1}{z}w^1(\mathbf{I} - \mathbf{P}_{k_2})^{-1}\mathbf{z}^t,$$

where $w^1 = \Big(p\{X_{\tau_1(0)} = 1\}, p\{X_{\tau_1(0)} = 2\}, \ldots, p\{X_{\tau_1(0)} = k_2\}\Big)$. Thus we get the result

$$E\left[z^{X_{\tau_2(0)}}\right] - z \;=\; (a_1(z) - z)R_{k_1}^{(U)}(z) + (a_2(z) - z)R_{k_1,k_2}^{(UU)}(z),$$

which proves Lemma 5.1.1. □

Now substituting the results of Lemma 5.1.1 into (5.2) we get

$$\begin{aligned} E\left[z^X\right] &= \frac{1}{m}\left[\frac{\Big[a_2(z) - a_1(z)\Big]\,z\,(a_1(z) - z)R_{k_1}^{(U)}(z)}{(a_1(z) - z)(a_2(z) - z)} + \frac{\Big[a_3(z) - a_2(z)\Big]\,z\,(a_1(z) - z)R_{k_1}^{(U)}(z)}{(a_2(z) - z)(a_3(z) - z)}\right. \\ &\qquad \left. + \frac{\Big[a_3(z) - a_2(z)\Big]\,z\,(a_2(z) - z)R_{k_1,k_2}^{(UU)}(z)}{(a_2(z) - z)(a_3(z) - z)} + \frac{a_3(z)(1 - z)}{a_3(z) - z}\right] \\ &= \frac{1}{m}\left[\frac{\Big[(a_2(z) - a_1(z))(a_3(z) - z) + (a_3(z) - a_2(z))(a_1(z) - z)\Big]\,z\,R_{k_1}^{(U)}(z)}{(a_2(z) - z)(a_3(z) - z)}\right. \end{aligned}$$

$$
\begin{aligned}
& + \frac{\left[a_3(z) - a_2(z)\right] z\, R^{(UU)}_{k_1,k_2}(z)}{a_3(z) - z} + \frac{a_3(z)(1-z)}{a_3(z) - z} \Bigg] \\
= \; & \frac{1}{m} \Bigg[\frac{\left[a_2(z)a_3(z) + z a_1(z) - a_2(z)a_1(z) - z a_3(z)\right] z\, R^{(U)}_{k_1}(z)}{(a_2(z) - z)(a_3(z) - z)} \\
& + \frac{\left[a_3(z) - a_2(z)\right] z\, R^{(UU)}_{k_1,k_2}(z)}{a_3(z) - z} + \frac{a_3(z)(1-z)}{a_3(z) - z} \Bigg] \\
= \; & \frac{1}{m} \Bigg[\frac{\left[(a_3(z) - a_1(z))(a_2(z) - z))\right] z\, R^{(U)}_{k_1}(z)}{(a_2(z) - z)(a_3(z) - z)} \\
& + \frac{\left[a_3(z) - a_2(z)\right] z\, R^{(UU)}_{k_1,k_2}(z)}{a_3(z) - z} + \frac{a_3(z)(1-z)}{a_3(z) - z} \Bigg] \\
= \; & \frac{1}{m} \left[\frac{\left[a_3(z) - a_1(z)\right] z\, R^{(U)}_{k_1}(z)}{a_3(z) - z} + \frac{\left[a_3(z) - a_2(z)\right] z\, R^{(UU)}_{k_1,k_2}(z)}{a_3(z) - z} + \frac{a_3(z)(1-z)}{a_3(z) - z} \right],
\end{aligned}
$$

which proves Theorem 5.1. □

Remark: In order to make use of this result we need to calculate w^1 in the above solution. Now $p\{X_{\tau_1(0)} = i\} = 0$ for $i = 1, \ldots, k_1$ and for $i = k_1 + 1, \ldots, k_2$ the following holds

$$
p\{X_{\tau_1(0)} = i\} \;=\; \frac{d^i}{dz^i} E\left[z^{X_{\tau_1(0)}}\right]_{z=0}.
$$

Now we know that

$$
E\left[z^{X_{\tau_1(0)}}\right] \;=\; (a_1(z) - z) R^{(U)}_{k_1}(z) + z.
$$

When we differentiate this i times we get

$$
\begin{aligned}
p\{X_{\tau_1(0)} = i\} &= \frac{1}{i!} \sum_{m=0}^{i} \frac{i!}{m!(i-m)!} \frac{d^m}{dz^m} \left(R^{(U)}_{k_1}(z)\right)_{z=0} \frac{d^{i-m}}{dz^{i-m}} \left(a_1(z) - z\right)_{z=0} + \delta_{i1} \\
&= \sum_{m=0}^{i} \frac{1}{m!(i-m)!} \frac{d^m}{dz^m} \left(R^{(U)}_{k_1}(z)\right)_{z=0} \left(a^1_{i-m}(i-m)! - \delta_{i-m,1}\right) + \delta_{i1}.
\end{aligned}
$$

Now $R^{(U)}_{k_1}(z) = \frac{1}{z}\mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_{k_1}\right)^{-1} \mathbf{z}^t$ so we get

$$
\frac{d^m}{dz^m} \left(R^{(U)}_{k_1}(z)\right)_{z=0} \;=\; \begin{cases} \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_{k_1}\right)^{-1} \mathbf{e}_{m+1}\, m!, & m = 0, \ldots, k_1 - 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}
$$

Thus as $i = k_1 + 1, \ldots, k_2$ we have $i > k_1 - 1$ and hence $i - m > 1$ whenever (5.5) is positive and also $i > 1$ (as we assume $k_1 \in I\!N$) and hence

$$
p\{X_{\tau_1(0)} = i\} \;=\; \sum_{m=0}^{k_1 - 1} \frac{1}{m!} \frac{d^m}{dz^m} \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_{k_1}\right)^{-1} \mathbf{e}_{m+1}\, m! \left(a^1_{i-m} - \delta_{i-m,1}\right) + \delta_{i1}
$$

$$
\begin{aligned}
&= \sum_{m=0}^{k_1-1} \mathbf{e}_1 \, (\mathbf{I} - \mathbf{P}_{k_1})^{-1} \mathbf{e}_{m+1} \, a^1_{i-m} \\
&= \sum_{m=1}^{k_1} \mathbf{e}_1 \, (\mathbf{I} - \mathbf{P}_{k_1})^{-1} \mathbf{e}_m \, a^1_{i-m+1}.
\end{aligned}
$$

This can then be used to perform the necessary calculations as we have already inverted $(\mathbf{I} - \mathbf{P}_{k_1})$. This could also be calculated directly using probabilistic arguments but as we need $R^{(U)}_{k_1}(z)$ in the solution it makes sense to use this.

The obvious extension to this chapters result, although not proved here, is

Proposition 5.1 *For n thresholds at $k_1 < k_2 < \cdots < k_n$ we find*

$$
E\left[z^X\right] = \frac{1}{m}\left[\frac{a_{n+1}(z)(1-z) + \sum_{i=1}^{n}[a_{n+1}(z) - a_i(z)]z \, R^{(iU)}_{k_1,..,k_i}(z)}{a_{n+1}(z) - z}\right],
$$

where

$$
R^{(iU)}_{k_1,..,k_i}(z) = \frac{1}{z}\mathbf{w}^{i-1}(\mathbf{I} - \mathbf{P}_{k_i})^{-1}\mathbf{z}^t
$$

and $\mathbf{w}^i = \left(p\{X_{\tau_i(0)} = 1\}, p\{X_{\tau_i(0)} = 2\}, \ldots, p\{X_{\tau_i(0)} = k_{i+1}\}\right)$ *and* $\mathbf{w}^0 = \mathbf{e}_1$.

5.2 Downward thresholds

One type of threshold has not been mentioned in previous examples, the fixed downward threshold. This threshold occurs when the number of customers in the system becomes less than or equal to say $l \in \mathbb{N}$.

We did not consider this type of threshold in Chapter 4 for the simple reason that it makes little sense to consider this type of threshold in isolation. The busy period begins with the number of customers in the system being zero and hence the threshold would automatically be passed before the process even began. Thus processes with two phases and hence only one threshold have little use for the fixed downward threshold.

We shall consider this type of threshold in the following context. The transition from phase 1 to phase 2 will have a threshold of the fixed upward type at k. The transition from phase 2 to phase 3 will occur at the first time subsequent to the process entering phase 2 at which there are no more than l customers in the system. The system then continues in phase three until the end of the busy period.

We shall limit our investigation to the case when $l \leq k$. This is because if $l > k$ the process would be equivalent to one with $l = k$ and so the case with $l > k$ is unnecessary. Further it makes little sense to consider the case with $l = 0$ as this will simply be the two-phase case because the third phase always takes zero time.

From the description above we are able to see that

$$\begin{aligned}
\tau_1(n) &= I_{C_n^1}\left[\tau(n) \wedge \inf\{m > n \mid X_m > k\}\right] + \left(I_{C_n^2} + I_{C_n^3}\right) n, \\
\tau_2(n) &= I_{C_n^1}\left[\tau(n) \wedge \inf\{m > n \mid X_m \leq l,\, I_{C_{m-1}^2} = 1\}\right] \\
&\quad + I_{C_n^2}\left[\tau(n) \wedge \inf\{m > n \mid X_m \leq l\}\right] + I_{C_n^3}\, n.
\end{aligned}$$

This type of process has an interesting feature. The GMRP that corresponds to the phase structure of the process is shown in Figure 5.2. In this GMRP state 3 is also a renewal state. This is because the embedded process obtained from considering the queueing process at departures is skip-free to the left. This means that if we consider the embedded process to be a random walk, the walk never skips a state while moving to the left. This is the result of only one customer being served at a time. If the process begins above state l it must pass through state l before entering any lower state and hence in this case $X_{\tau_2(0)} = l$. State 3 is then a renewal state because the time spent in this state will no longer be in any way dependent upon the time spent in states 1 or 2.

The alternative is that the system never passes the first threshold. If this is the case then it cannot pass the second threshold either and hence $X_{\tau_1(0)} = 0$ and $X_{\tau_2(0)} = 0$.

When we convert this GMRP to a three-phase MRP by considering transitions from state 1 to state 1 to pass through states 2 and 3 spending zero time in each we loose this renewal structure. However, it is still sufficient to enable some simplification of the results. The resulting three-phase MRP is shown in Figure 5.3

One further thing to note is that in this process neither ρ_2 nor ρ_3 can be greater than one if the queue is to be stable. Thus this is not perhaps a queue of great interest. What is more interesting is the process which may make repeated transitions between

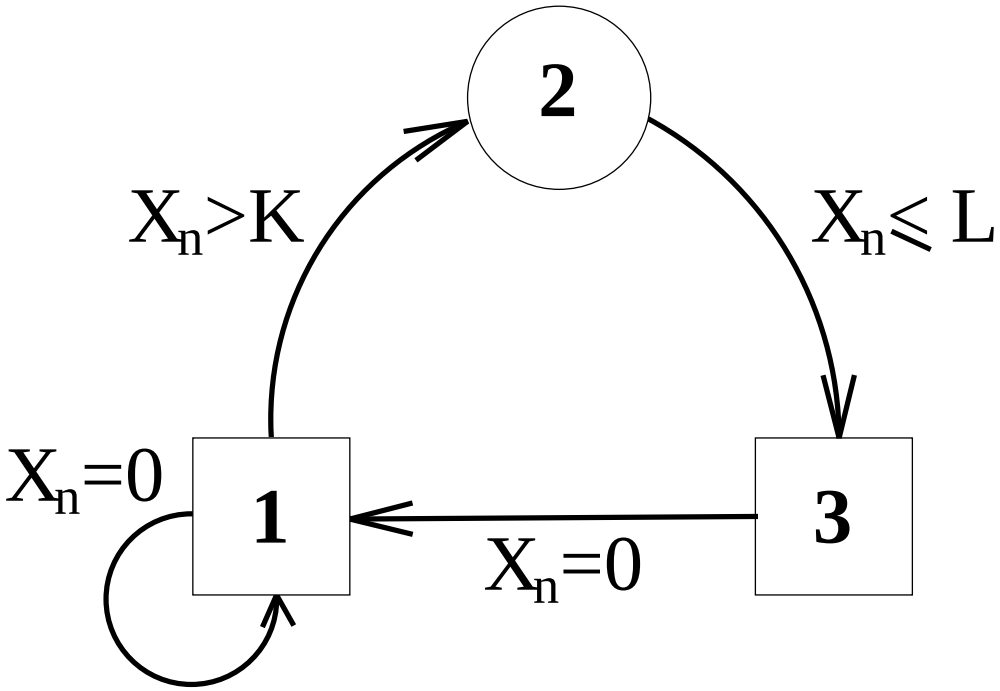

Figure 5.2: The MRP for one upwards and one downwards threshold.

phases as it passes above and below the thresholds. We shall consider this type of process in Chapter 7. Thus we shall consider a few results here which will contribute to that later problem but we shall not bother to obtain the probability generating function for the equilibrium behaviour of the system.

It is obvious that $X_{\tau_1(0)} = 0$ implies that $X_{\tau_2(0)} = 0$. Also when $X_{\tau_1(0)} \neq 0$ we can see from the skip-free to the left nature of the embedded process that $X_{\tau_2(0)} = l$. Hence we get

$$E\left[z^{X_{\tau_2(0)}}\right] = p\{X_{\tau_1(0)} = 0\} + z^l\left(1 - p\{X_{\tau_1(0)} = 0\}\right).$$

We can calculate $p\{X_{\tau_1(0)} = 0\}$ from the following

$$\begin{aligned} p\{X_{\tau_1(0)} = 0\} &= E\left[z^{X_{\tau_1(0)}}\right]_{z=0} \\ &= R_k^{(U)}(0) \\ &= a_0 \mathbf{e}_1 \left(\mathbf{I} - \mathbf{P}_k\right)^{-1} \mathbf{e}_1{}^t. \end{aligned}$$

From this we could obtain the solution but it would be of little use at this stage and so we shall adjourn this discussion until Chapter 7.

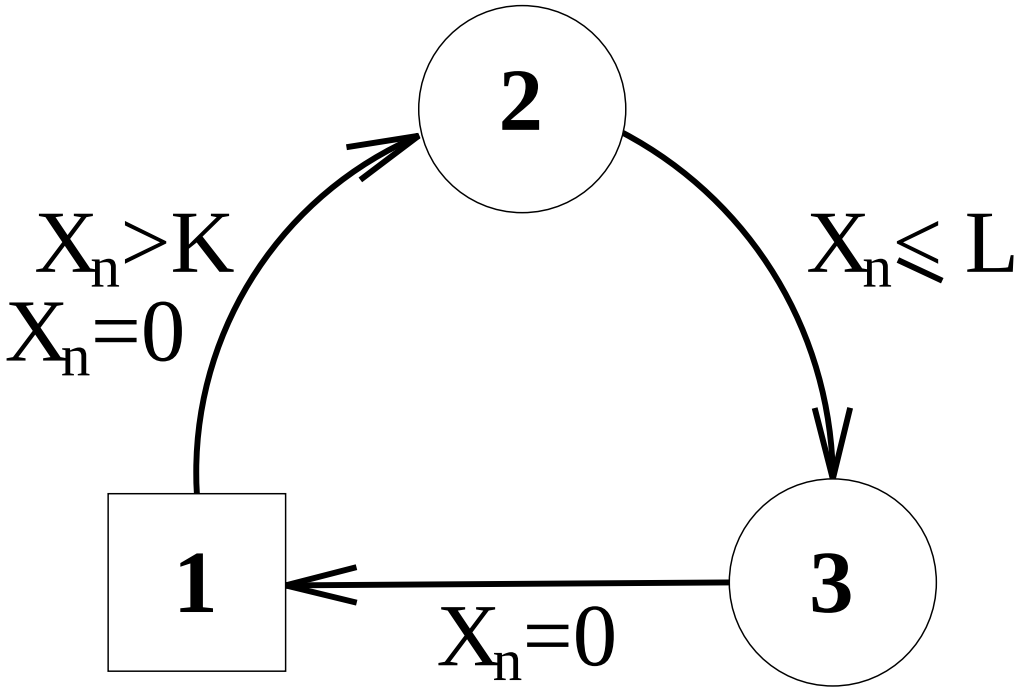

Figure 5.3: The three-phase MRP for one upwards and one downwards threshold. □ denotes a renewal point while ◯ denotes a non-renewal point

Chapter 6

A breakdown/repair model

In this chapter we consider a four-phase M/G/1 queue. This is intended to model a simple breakdown/repair queue. We have a single-server queueing process with generally-distributed service times, with probability distribution function $A(\cdot)$, and Poisson arrivals with rate λ. The process considered is one in which a breakdown occurs after a random number of customers has been served in the busy period. A major assumption we make here is that the server can only breakdown once per busy period. This may be a realistic approximation if the probability of a breakdown is small. If not, a more complex model must be considered.

We assume that during each idle period the server is checked and any necessary repairs are made. Thus at the beginning of the busy period we can ignore how many customers have previously been served, making the beginning of the busy period a renewal point. This checkup takes a random time with distribution $S(\cdot)$. Checking the server can in many cases be considered to be a set task or sequence of tasks, such as checking certain components and replacing them if worn. Each task in this sequence takes either zero or a deterministic amount of time and so we may consider the checkup-time distribution to be a probabilistic mixture of deterministic distributions. That is $S(t) = \sum_i \rho_i \delta(x_i - t)$.

A breakdown occurs during a service with some probability which is dependent on how many customers have been served in the current busy period. When the server breaks down, the service currently in progress is interrupted and a repair is begun. The repair time has probability distribution function $R(\cdot)$. When the repair is completed the interrupted service may begin again in two ways. It may resume where it left off or it may have to repeat the work done before the interruption. In these cases the Laplace-Stieltjes transform of the total time spent during the service and repair is

$$\begin{aligned} \text{(i) resume:} \quad B^*(s) &= A^*(s)R^*(s), \\ \text{(ii) repeat:} \quad B^*(s) &= A^*(s)R^*(s)C^*(s), \end{aligned}$$

where $C(\cdot)$ is the probability distribution function for the amount of work done before a breakdown. The function $C(\cdot)$ will depend on the service time and where in this service the breakdown occurs. For our purposes here it is easier and more natural to use the resume model.

Once the breakdown and repair have occurred the process continues normally until the end of the busy period.

We use, for this process, a four-phase model. The first phase is only a single service long. A customer arriving at the empty queue must first wait until the server's check is completed before it can begin service. Thus we add to this customer's service time the remaining checkup-time. The Laplace-Stieltjes transform of this remaining time is

$$\hat{S}^*(s) = \sum_i \frac{\lambda \rho_i e^{-\lambda x_i}}{s - \lambda}.$$

From this we get $A^{1*}(s) = A^*(s)\hat{S}^*(s)$.

The second phase is the normal service period with a random threshold. During this phase customers are served normally. Thus $A^{2*}(s) = A^*(s)$ where $A(\cdot)$ is the service-time distribution. This phase ends at the breakdown. Phases must obey the rules of Chapter 3 but we have said that breakdowns may occur during services in this model. In this case phase rules (i) and (ii) appear to be broken. We can get around this by saying that when a breakdown occurs during a particular service, the decision that the breakdown occurs during that service is made before the service begins. Thus the phase change occurs before the service in which the breakdown occurs. With this model in mind we take the probability of the phase change occurring between the nth service and the $(n+1)$th service of a busy period to be $h(n)$.

The third phase is a repair phase, consisting of a single service in which the customer waits for the repair time and then has its normal service. Thus $A^{3*}(s) = B^*(s)$. As we are assuming the service resumption model of breakdowns it does not matter at what point the breakdown occurs during the service time.

The final phase is again a normal phase which ends when the system is empty. Hence $A^{4*}(s) = A^{2*}(s) = A^*(s)$. We define the traffic intensity during this phase and phase 2 to be ρ_a.

We shall consider only the case when $\rho_a < 1$. It makes sense to do this because we are considering an M/G/1 queue that may break down. For this to be stable the corresponding M/G/1 queue without breakdowns must also be stable.

One final assumption we make to simplify the problem is to assume that a breakdown will not occur during the first service of a busy period. (i.e. $h(0) = 0$) This is not an unreasonable assumption as we have assumed the server gets checked during the idle period. This is to avoid the possibility that the first service has extra time added to it from the left over checking and repair. Removing this assumption makes the problem harder but not insoluble.

If $\rho_a < 1$ we might still have a case when $\rho_1 > 1$ or $\rho_3 > 1$ in which case we must consider condition $(*)$. However as phases 1 and 3 last only one service each, condition $(*)$ is easily satisfied. The equilibrium distribution of customers in the system is then given by the following theorem.

Theorem 6.1 *For $\rho_a < 1$ and $z \in [0, 1)$*

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{a_2(z) - z a_1(z) + \Big(a_2(z) - a_3(z)\Big)\Big(z - a_0^2 R_h^{(H)}(0)\Big)}{a_2(z) - z}\right.$$

$$+ R_h^{(H)}(z)\left(a_2(z) - a_3(z)\right)\Big],$$

where $R_h^{(H)}(z)$ *is defined to be* $\sum_{n=1}^{\infty} h(n) R_n^{(T')}(z)$ *where* $R_n^{(T')}(z)$ *is defined by the recursive relationship*

$$R_{n+1}^{(T')}(z) = \frac{1}{z}\left\{R_n^{(T')}(z)a(z) - a_0 R_n^{(T')}(0)\right\} + 1$$

and $R_1^{(T')}(z) = \frac{a_1(z)-z}{a_2(z)-z}$. *Also the mean length of the busy period is given by*

$$m = \frac{1 + (\rho_1 - \rho_2) + (\rho_3 - \rho_2)\left[1 - a_0^2 R_h^{(H)}(0)\right]}{1 - \rho_2}.$$

Proof: From Theorem 3.9 we get the the equilibrium probability generating function to be

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{E\left[z^{X_{\tau_2(0)}}\right]-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}\right.$$
$$\left.+\frac{E\left[z^{X_{\tau_3(0)}}\right]-E\left[z^{X_{\tau_2(0)}}\right]}{1-\xi_3(z)}+\frac{1-E\left[z^{X_{\tau_3(0)}}\right]}{1-\xi_4(z)}\right].$$

Now because the first phase lasts for only one service

$$E\left[z^{X_{\tau_1(0)}}\right] = a_1(z).$$

Thus we get

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{a_1(z)-z}{1-\xi_1(z)}+\frac{E\left[z^{X_{\tau_2(0)}}\right]-a_1(z)}{1-\xi_2(z)}\right.$$
$$\left.+\frac{E\left[z^{X_{\tau_3(0)}}\right]-E\left[z^{X_{\tau_2(0)}}\right]}{1-\xi_3(z)}+\frac{1-E\left[z^{X_{\tau_3(0)}}\right]}{1-\xi_4(z)}\right].$$

Using $\xi_2(z)=\xi_4(z)$ and rearranging we get

$$E\left[z^X\right] = \frac{1}{m}\left[a_1(z)+\frac{E\left[z^{X_{\tau_2(0)}}\right]-E\left[z^{X_{\tau_3(0)}}\right]}{1-\xi_2(z)}\right.$$
$$\left.+\frac{E\left[z^{X_{\tau_3(0)}}\right]-E\left[z^{X_{\tau_2(0)}}\right]}{1-\xi_3(z)}+\frac{1-a_1(z)}{1-\xi_2(z)}\right],$$

which, as in similar arguments, gives

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{\left(E\left[z^{X_{\tau_3(0)}}\right]-E\left[z^{X_{\tau_2(0)}}\right]\right)z\left(a_2(z)-a_3(z)\right)}{(a_2(z)-z)(a_3(z)-z)}+\frac{a_2(z)-za_1(z)}{a_2(z)-z}\right]. \quad (6.1)$$

Now, since $X_{\tau_2(0)}=0$ implies that $X_{\tau_3(0)}=0$, we get

$$E\left[z^{X_{\tau_3(0)}}\right]-E\left[z^{X_{\tau_2(0)}}\right]=E\left[z^{X_{\tau_3(0)}}I(X_{\tau_2(0)}\neq 0)\right]-E\left[z^{X_{\tau_2(0)}}I(X_{\tau_2(0)}\neq 0)\right]. \quad (6.2)$$

Furthermore we spend exactly one service time in phase 3 and so $\tau_3(0)=\tau_2(0)+1$. This means that when $X_{\tau_2(0)}\neq 0$ we get $X_{\tau_3(0)}=X_{\tau_2(0)}+A^3_{\tau_2(0)+1}-1$ so that

$$E\left[z^{X_{\tau_3(0)}}I(X_{\tau_2(0)}\neq 0)\right] = E\left[z^{X_{\tau_2(0)}+A^3_{\tau_2(0)+1}-1}I(X_{\tau_2(0)}\neq 0)\right].$$

Now $A^3_{\tau_2(0)+1}$ is independent of $X_{\tau_2(0)}$ so that

$$E\left[z^{X_{\tau_3(0)}}I(X_{\tau_1(0)}\neq 0)\right] = E\left[z^{X_{\tau_2(0)}}I(X_{\tau_1(0)}\neq 0)\right]E\left[z^{A^2_{\tau_2(0)+1}}\right]z^{-1},$$

and $E\left[z^{A^3_{\tau_2(0)+1}}\right] z^{-1} = a_3(z)/z$ so we get

$$E\left[z^{X_{\tau_3(0)}} I(X_{\tau_2(0)} \neq 0)\right] = \frac{a_3(z)}{z} E\left[z^{X_{\tau_2(0)}} I(X_{\tau_2(0)} \neq 0)\right].$$

Using this result in (6.2) we get

$$E\left[z^{X_{\tau_3(0)}}\right] - E\left[z^{X_{\tau_2(0)}}\right] = \frac{1}{z}\left(a_3(z) - z\right) E\left[z^{X_{\tau_2(0)}} I(X_{\tau_2(0)} \neq 0)\right]. \qquad (6.3)$$

By the same arguments used in Section 4.5

$$E\left[z^{X_{\tau_2(0)}}\right] - z = (a_2(z) - z) \sum_{n=1}^{\infty} h(n) R_n^{(T')}(z),$$

where by the same recurrence arguments of Theorem 4.5 we can show that

$$R_{n+1}^{(T')}(z) = \frac{1}{z}\left\{R_n^{(T')}(z) a(z) - a_0 R_n^{(T')}(0)\right\} + 1,$$

and $R_1^{(T')}(z) = \frac{a_1(z)-z}{a_2(z)-z}$. The different initial value $R_1^{(T')}(z)$ arises from the different behaviour of the server during the first service of a busy period. We shall write $R_h^{(H)}(z) = \sum_{n=1}^{\infty} h(n) R_n^{(T')}(z)$ and so

$$E\left[z^{X_{\tau_2(0)}}\right] = (a_2(z) - z) R_h^{(H)}(z) + z.$$

From the fact that $p\{X_{\tau_2(0)} = 0\} = E\left[z^{X_{\tau_2(0)}}\right]_{z=0}$ we get

$$p\{X_{\tau_2(0)} = 0\} = a_0^2 R_h^{(H)}(0),$$

and so

$$E\left[z^{X_{\tau_2(0)}} I(X_{\tau_2(0)} \neq 0)\right] = (a_2(z) - z) R_h^{(H)}(z) - a_0^2 R_h^{(H)}(0) + z.$$

Substituting this into (6.3) and thence into (6.1) we get

$$
\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left[\frac{\frac{1}{z}\Big(a_3(z)-z\Big)\Big((a_2(z)-z)R_h^{(H)}(z)-a_0^2R_h^{(H)}(0)+z\Big)\,z\,\Big(a_2(z)-a_3(z)\Big)}{(a_2(z)-z)(a_3(z)-z)}\right. \\
&\qquad\qquad \left. +\frac{a_2(z)-za_1(z)}{a_2(z)-z}\right] \\
&= \frac{1}{m}\left[\frac{\Big((a_2(z)-z)R_h^{(H)}(z)-a_0^2R_h^{(H)}(0)+z\Big)\Big(a_2(z)-a_3(z)\Big)+a_2(z)-za_1(z)}{a_2(z)-z}\right] \\
&= \frac{1}{m}\left[\frac{a_2(z)-za_1(z)+\Big(a_2(z)-a_3(z)\Big)\Big(z-a_0^2R_h^{(H)}(0)\Big)}{a_2(z)-z}\right. \\
&\qquad\qquad \left. +R_h^{(H)}(z)\Big(a_2(z)-a_3(z)\Big)\right].
\end{aligned}
$$

If we now take the limit as $z \uparrow 1$ then we shall get an expression for m. This is

$$
m = \frac{1+(\rho_1-\rho_2)+(\rho_3-\rho_2)\left[1-a_0^2R_h^{(H)}(0)\right]}{1-\rho_2},
$$

as desired. □

Remarks:
(i) As in Theorem 4.6 we could obtain a closed form for $R_n^{(T')}(z)$ and thence write $R_h^{(H)}(z)$ explicitly. For brevity we have omitted such a derivation in this case.
(ii) It is noteworthy that if we take the repair times to be zero with probability one then $a_3(z) = a_2(z)$ and so the solution is

$$
E\left[z^X\right] = \frac{1}{m}\left[\frac{a_2(z)-za_1(z)}{a_2(z)-z}\right],
$$

which is the expected result for a queue with a different service-time distribution for a server arriving at an empty server. (See Section 4.2.1.)

Chapter 7

A two-threshold, infinite-phase example.

In Section 5.2 we described the downwards threshold. Used in conjunction with an upwards threshold this can produce an interesting example. We noted in Section 5.2 that the three-phase process with an upwards threshold at k and a downwards threshold at l is of limited interest. The process we are interested in is the one in which the server can switch between service-time distributions each time a threshold is crossed and not just the first time the thresholds are crossed each busy period. This presents a problem, the thresholds can be crossed an infinite number of times during one busy period, admittedly with probability zero. Thus the process we are interested in has an infinite number of phases. Another way of viewing this is given below.

The example can be considered as a process with two regimes. In regime one the service-time distribution is $A(\cdot)$ while in regime two the service-time distribution is $B(\cdot)$. If the process starts at time zero with no customers in the system and in regime one then we can describe the transitions between regimes in the following way. The first time the system has more than K customers in it at a service completion epoch it enters regime two. When next the queue has no more than L customers in it at a service completion epoch it returns to regime one. It makes sense to take only values of $L, K \in I\!N$ such that $L \leq K$. If $L = 0$ then we simply have the system of Chapter 4 with a fixed upwards threshold. If $L > K$ transitions to the second regime (which spend a positive time in that regime) would only occur when there were more than L customers in the system at the transition point and so we may as well increase K so that it equals L.

We can model this process by a three state GMRP as in Figure 7.1. This is a natural description of the process involved. This, however, would not fit the general theory we have developed. Thus we consider a different process in which each subsequent transition past a threshold is represented by a transition into a new phase. This procedure results in the multi-phase MRP of Figure 7.2. We obtain this multi-phase process by using the procedure of Section 3.6 on the GMRP of Figure 7.1.

Thus we consider an infinite-phase process. Clearly phases $1, 3, 5, \ldots$ will all correspond to the system being in regime 1 and hence $A^1(\cdot)$, $A^3(\cdot)$, $A^5(\cdot), \ldots = A(\cdot)$ while phases $2, 4, 6, \ldots$ correspond to regime two and so $A^2(\cdot)$, $A^4(\cdot)$, $A^6(\cdot), \ldots = B(\cdot)$.

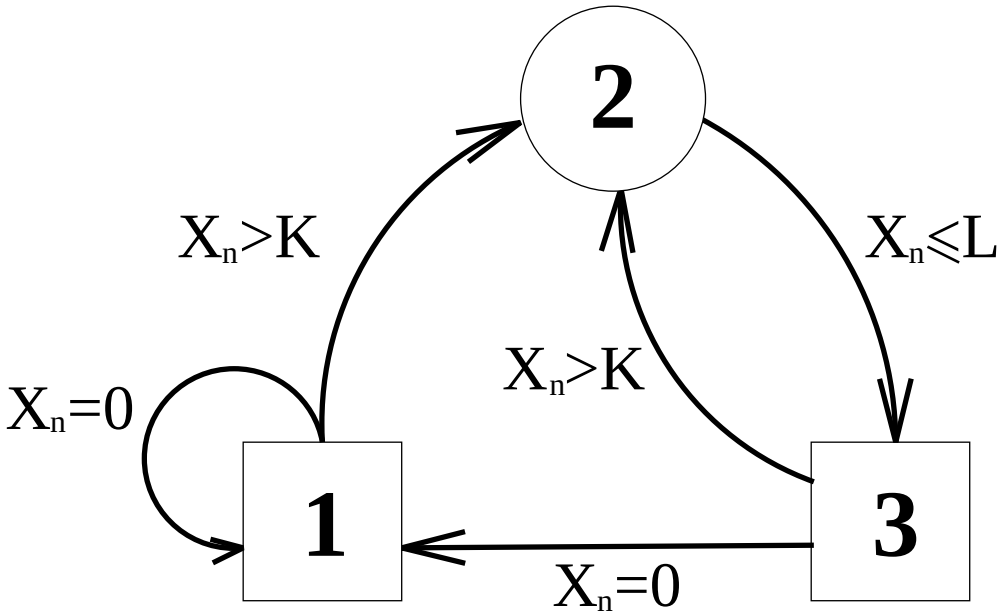

Figure 7.1: GMRP type II. □ denotes a renewal point while ◯ denotes a non-renewal point

Thus we can write, by extending the notation,

$$\begin{array}{rclrcl} A_n^1, A_n^3, A_n^5, \ldots &=& A_n, & A_n^2, A_n^4, A_n^6, \ldots &=& B_n, \\ a^1(z), a^3(z), a^5(z), \ldots &=& a(z), & a^2(z), a^4(z), a^6(z), \ldots &=& b(z), \\ \xi_1(z), \xi_3(z), \xi_5(z), \ldots &=& \xi_a(z), & \xi_2(z), \xi_4(z), \xi_6(z), \ldots &=& \xi_b(z), \\ a_i^1, a_i^3, a_i^5, \ldots &=& a_i, & a_i^2, a_i^4, a_i^6, \ldots &=& b_i, \\ \rho_1, \rho_3, \rho_5, \ldots &=& \rho_a, & \rho_2, \rho_4, \rho_6, \ldots &=& \rho_b. \end{array} \tag{7.1}$$

In this example we have not demonstrated that condition $(*)$ is satisfied. When $\rho_a \leq 1$ it is satisfied but when $\rho_a > 1$ there are problems. It is easy to prove the regularity of the stopping times $\tau_i(0)$ (when $\rho_b \leq 1$) using the same type of arguments as used in Section 4.2 but in this case it is hard to show that $\tau(0)$ is regular, due to the infinite number of phases. Thus we present the results we get for this problem in the proposition below. In Section 7.3 of this chapter we provide a number of numerical examples to support the result. While these do not prove the proposition it is to be hoped that they remove any immediate doubts about its veracity.

7.1 Motivation

There are several reasons for studying such systems. The case where $K = L$ appears in the literature. For instance, Morrison (1990) investigates a system in which the service times are negative exponentially distributed and instead of changing this between the two regimes he allows two servers to operate in the second regime. Because of the exponential service times, two servers are equivalent to one server with twice the service rate (for the purpose of queue length calculation). Another example is Gong *et al* (1992). In this case they consider the system with the arrival rate dependent on the number of

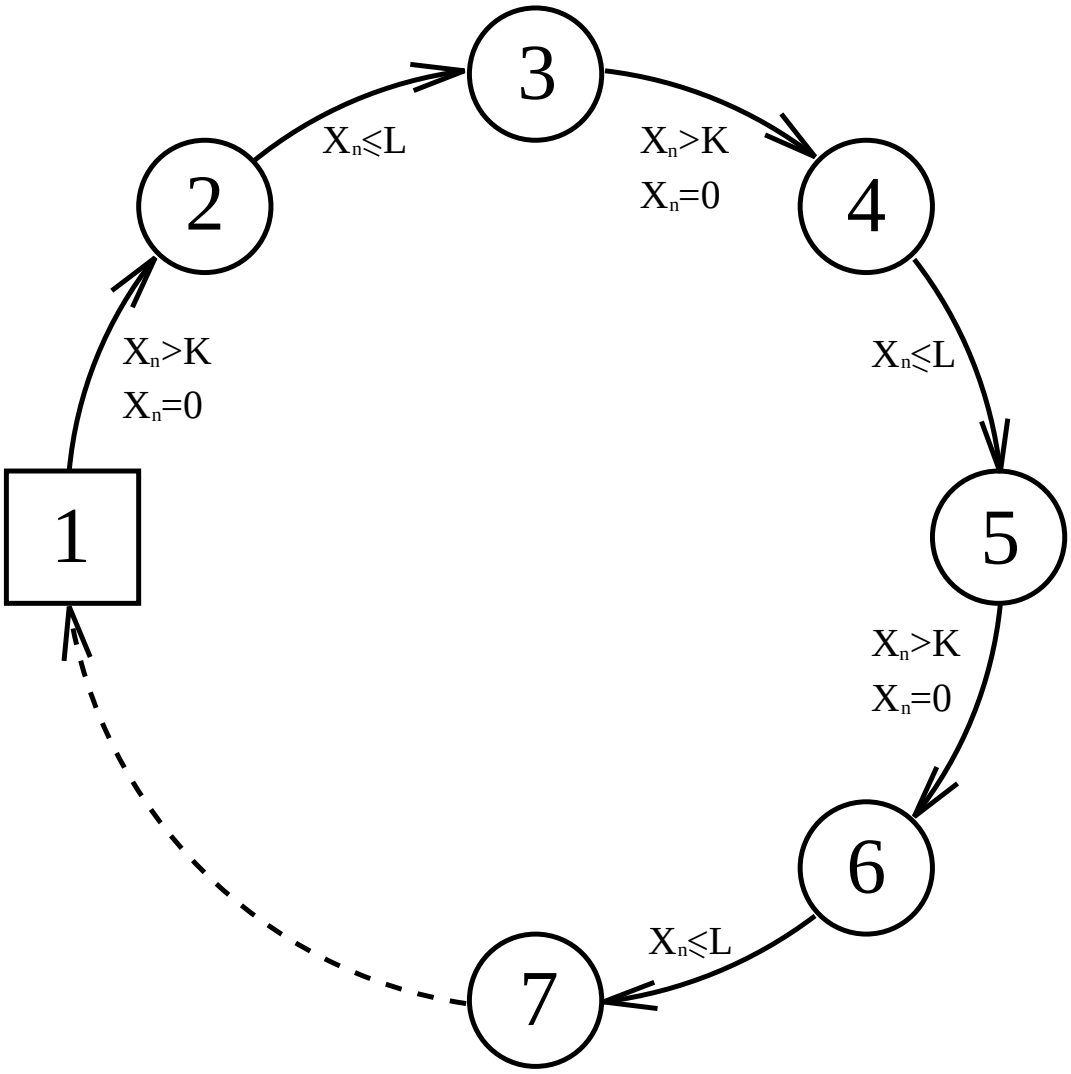

Figure 7.2: Multi-phase MRP with an infinite number of phases. □ denotes a renewal point while ◯ denotes a non-renewal point

customers in the queue. The example in this chapter, with $K = L$ is a particular case of this. A major reason why such models are considered is to allow customer balking. This is when a customer can refuse to enter a queue if it believes the queue is too long. A way of modelling this is by giving the customers a probability of balking when the queue has more than a set number of customers in it. Another reason for considering such a problem is if more than one type of customer arrives at a system. If these types have different priorities we may wish to block some types when the queue has more than a certain number of customers in it.

The unusual feature, in terms of queueing theory, of this chapter's example is that L can be less than K. What is particularly unusual about this is the fact that the embedded process is no longer a Markov chain as in most conventional examples. We might want $L < K$ because in the case with $K = L$ it is possible that the system will spend much of its time switching between the two regimes, for example if $\rho_a > 1$ and $\rho_b << 1$. If there is an overhead associated with swapping between regimes it is desirable

for frequent swapping to be avoided. An alternative to changing the traffic intensities is to have $L < K$ which will reduce the number of times that swaps between regimes occur.

This type of policy occurs in inventory problems, for example Morse (1967). We consider the queue to be a store of some resource and the service times to be the times between orders for the resource. The store does not want to run out of the resource, but neither does it want to store more than is necessary. A possible policy is to start ordering when the store falls below a certain level L, and stop ordering when above another level K. This is the type of process we consider although the exponential arrival times are not usual for ordered resources.

7.2 Results

Proposition 7.1 *For the process above the following hold*
(i) When $\rho_a > 0$ and $\rho_b > 1$ the process is transient.
(ii) When $\rho_a > 0$ and $\rho_b = 1$ the process is null recurrent.
(iii) When $\rho_a > 0$ and $\rho_b < 1$ the probability generating function for the equilibrium number of customers in the system is given by

$$E\left[z^X\right] = \frac{1}{m}\left\{\frac{b(z)(1-z)+\{b(z)-a(z)\}zR_{KL}^{(UD)}(z)}{b(z)-z}\right\},$$

for $z \in [0,1)$, where

$$\begin{aligned} R_{KL}^{(UD)}(z) &= \frac{1}{z}\left[\left(\mathbf{e}_1 + \left(\frac{h_1}{1-h}\right)\mathbf{e}_L\right)(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t\right], \\ h &= 1 - a_0\mathbf{e}_L\,(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{e}_1{}^t, \\ h_1 &= 1 - a_0\mathbf{e}_1\,(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{e}_1{}^t \end{aligned}$$

and m, the mean number of customers served in one busy period, is given by

$$m = \left[\frac{1+\{\rho_a-\rho_b\}R_{KL}^{(UD)}(1)}{1-\rho_b}\right].$$

Proof: (i) When $\rho_a > 0$ there is a positive probability that the process will get to the second regime at some stage in the busy period. While in regime two the traffic intensity is $\rho_b > 1$ and so the probability that the number of customers in the system goes below L again is less than one (from the behaviour of the standard M/G/1 queue). Thus in the long term the process will be unstable.
(ii) This follows from the value of m as $\rho_b \uparrow 1$.
(iii) From Theorem 3.9 we get the solution to be

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)} + \frac{E\left[z^{X_{\tau_2(0)}}\right]-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)} + \cdots\right].$$

We take $h_n = p\{X_{\tau_n(0)} > 0\}$, the probability that phase $n+1$ is reached before the end of the busy period. The process is skip-free to the left so that the transitions below L will always be to L and so

$$\begin{aligned} h_{2n} &= p\{X_{\tau_{2n}(0)} = L\}, \\ h_{2n+1} &= p\{X_{\tau_{2n+1}(0)} > K\}. \end{aligned}$$

It is easy to see that the thresholds must alternate. Thus a jump above K is always followed by a jump down to L at some time before the end of the busy period so that

$$h_{2n} = h_{2n-1}.$$

Also from the theorem of total probability we can write

$$\begin{aligned} h_{2n+1} &= p\{X_{\tau_{2n+1}(0)} > K | X_{\tau_{2n}(0)} = L\} p\{X_{\tau_{2n}(0)} = L\} \\ &\quad + \underbrace{p\{X_{\tau_{2n+1}(0)} > K | X_{\tau_{2n}(0)} = 0\}}_{\| \\ 0} p\{X_{\tau_{2n}(0)} = 0\} \\ &= p\{X_{\tau_{2n+1}(0)} > K | X_{\tau_{2n}(0)} = L\}\, h_{2n} \\ &= p\{X_{\tau_{2n=1}(0)} > K | X_{\tau_{2n}(0)} = L\}\, h_{2n-1}. \end{aligned}$$

Thus if we set $h = p\{X_{\tau_{2n+1}(0)} > K | X_{\tau_{2n}(0)} = L\}$ then

$$\begin{aligned} h_{2n+1} &= h_1 h^n, \\ h_{2n+2} &= h_{2n+1} \\ &= h_1 h^n. \end{aligned}$$

For $n > 1$ we get

$$\begin{aligned} E\left[z^{X_{\tau_n(0)}}\right] &= E\left[z^{X_{\tau_n(0)}} I(X_{\tau_{n-1}(0)} = 0)\right] + E\left[z^{X_{\tau_n(0)}} I(X_{\tau_{n-1}(0)} \neq 0)\right] \\ &= p\{X_{\tau_{n-1}(0)} = 0\} + + p\{X_{\tau_{n-1}(0)} \neq 0\} E\left[z^{X_{\tau_n(0)}} \middle| X_{\tau_{n-1}(0)} \neq 0\right] \\ &= (1 - h_{n-1}) + h_{n-1} E\left[z^{X_{\tau_n(0)}} \middle| X_{\tau_{n-1}(0)} > 0\right]. \end{aligned} \tag{7.2}$$

From (7.2) this gives

$$\begin{aligned} E\left[z^{X_{\tau_{2n}(0)}}\right] &= (1 - h_{2n-1}) + h_{2n-1} z^L, && n > 0, \\ E\left[z^{X_{\tau_{2n+1}(0)}}\right] &= \begin{cases} E\left[z^{X_{\tau_1(0)}}\right], & n = 0, \\ (1 - h_{2n}) + h_{2n} E\left[z^{X_{\tau_{2n+1}(0)}} \middle| X_{\tau_{2n}(0)} = L\right], & n > 0. \end{cases} \end{aligned}.$$

As in Lemma 4.3

$$E\left[z^{X_{\tau_1(0)}}\right] - z = \left[a(z) - z\right] R_k^{(U)}(z).$$

We can see that for $n > 0$

$$E\left[z^{X_{\tau_{2n+1}(0)}} \middle| X_{\tau_{2n}(0)} = L\right] = r(z),$$

where $r(z)$ is independent of n. Now this means that for $n > 0$

$$\begin{aligned} E\left[z^{X_{\tau_{2n+1}(0)}}\right] - E\left[z^{X_{\tau_{2n}(0)}}\right] &= (1 - h_{2n}) + h_{2n} r(z) - (1 - h_{2n-1}) - h_{2n-1} z^L \\ &= h_{2n} r(z) - h_{2n-1} z^L \\ &= h_1 h^{n-1} (r(z) - z^L), \\ E\left[z^{X_{\tau_{2n+2}(0)}}\right] - E\left[z^{X_{\tau_{2n+1}(0)}}\right] &= (1 - h_{2n+1}) + h_{2n+1} z^L - (1 - h_{2n}) - h_{2n} r(z) \\ &= -h_{2n+1} + h_{2n+1} z^L + h_{2n} - h_{2n} r(z) \\ &= h_1 h^{n-1} \left\{-h + h z^L + 1 - r(z)\right\} \\ &= h_1 h^{n-1} (1 - h) + h_1 h^{n-1} \left\{h z^L - r(z)\right\}. \end{aligned}$$

We can now write the probability generating function for the number of customers in the system as follows.

$$\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left\{\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{E\left[z^{X_{\tau_2(0)}}\right]-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}\right\} \\
&\quad +\frac{1}{m}\sum_{n=1}^{\infty}\left\{\frac{E\left[z^{X_{\tau_{2n+1}(0)}}\right]-E\left[z^{X_{\tau_{2n}(0)}}\right]}{1-\xi_1(z)}\right\} \\
&\quad +\frac{1}{m}\sum_{n=1}^{\infty}\left\{\frac{E\left[z^{X_{\tau_{2n+2}(0)}}\right]-E\left[z^{X_{\tau_{2n+1}(0)}}\right]}{1-\xi_2(z)}\right\} \\
&= \frac{1}{m}\left\{\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{(1-h_1)+h_1z^L-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}\right\} \\
&\quad +\frac{1}{m}\sum_{n=1}^{\infty}\left\{\frac{h_1h^{n-1}\left(r(z)-z^L\right)}{1-\xi_1(z)}\right\} \\
&\quad +\frac{1}{m}\sum_{n=1}^{\infty}\left\{\frac{h_1h^{n-1}(1-h)+h_1h^{n-1}\left\{hz^L-r(z)\right\}}{1-\xi_2(z)}\right\} \\
&= \frac{1}{m}\left\{\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{1-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}+\frac{h_1(z^L-1)}{1-\xi_2(z)}\right\} \\
&\quad +\frac{1}{m}\left\{\frac{h_1\left(r(z)-z^L\right)}{1-\xi_1(z)}\right\}\sum_{n=1}^{\infty}h^{n-1} \\
&\quad +\frac{1}{m}\left\{\frac{h_1(1-h)+h_1\left\{hz^L-r(z)\right\}}{1-\xi_2(z)}\right\}\sum_{n=1}^{\infty}h^{n-1} \\
&= \frac{1}{m}\left\{\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)}+\frac{1-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}+\frac{h_1(z^L-1)}{1-\xi_2(z)}\right\} \\
&\quad +\frac{1}{m(1-h)}\left\{\frac{h_1\left(r(z)-z^L\right)}{1-\xi_1(z)}\right\} \\
&\quad +\frac{1}{m(1-h)}\left\{\frac{h_1(1-h)+h_1\left\{hz^L-r(z)\right\}}{1-\xi_2(z)}\right\}.
\end{aligned}$$

Now the first two terms in this are the same as in the M/G/1 queue with a single fixed threshold at K (apart from the different normalising constant m). We shall replace these two terms with $Y(z)$ for the moment.

$$\begin{aligned}
E\left[z^X\right] &= \frac{Y(z)}{m}+\frac{h_1}{m(1-h)}\left\{\frac{r(z)-z^L}{1-\xi_1(z)}\right\} \\
&\quad +\frac{h_1}{m}\frac{z^L-1}{1-\xi_2(z)}+\frac{h_1}{m(1-h)}\left\{\frac{(1-h)+\left\{hz^L-r(z)\right\}}{1-\xi_2(z)}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{r(z)-z^L}{1-\xi_1(z)}\right\} \\
&\quad + \frac{h_1}{m(1-h)}\left\{\frac{(z^L-1)(1-h)+(1-h)+\left\{hz^L-r(z)\right\}}{1-\xi_2(z)}\right\} \\
&= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{r(z)-z^L}{1-\xi_1(z)}\right\} \\
&\quad + \frac{h_1}{m(1-h)}\left\{\frac{z^L(1-h)+hz^L-r(z)}{1-\xi_2(z)}\right\} \\
&= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{r(z)-z^L}{1-\xi_1(z)}\right\} + \frac{h_1}{m(1-h)}\left\{\frac{z^L-r(z)}{1-\xi_2(z)}\right\} \\
&= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{(r(z)-z^L)[\xi_1(z)-\xi_2(z)]}{(1-\xi_1(z))(1-\xi_2(z))}\right\} \\
&= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{z(r(z)-z^L)[b(z)-a(z)]}{(a(z)-z)(b(z)-z)}\right\}. \qquad (7.3)
\end{aligned}
$$

We must now calculate $r(z)$. As before we expect $r(z) = (a(z)-z)R(z) + z^L$ for some bounded function $R(z)$. We use the technique used in Section 4.2, Theorem 4.3 of expanding the expectation to get

$$
\begin{aligned}
r(z) &= E\left[z^{X_{\tau_{2n+1}(0)}} \middle| X_{\tau_{2n}(0)} = L\right] \\
&= \sum_{i=0}^{\infty} \mathbf{v}^i \mathbf{g}(z)^t,
\end{aligned}
$$

where we take $\mathbf{g}(z) = (g_1(z), g_2(z), \cdots, g_K(z))$,

$$
g_j(z) = \left(a(z)z^{j-1} - \sum_{l=(j-1)\wedge 1}^{K} a_{l-j+1} z^l\right),
$$

and $\mathbf{v}^i = (v_1^i, v_2^i, \ldots, v_K^i)$ where

$$
v_j^i = p\{X_{\tau_{2n}(0)+i} = j,\, \tau_{2n+1} > \tau_{2n} + i \,|\, X_{\tau_{2n}(0)} = L\}.
$$

Again we use the sub-stochastic transition matrix $\mathbf{P}_k$ defined in (4.2) with the initial vector $\mathbf{v}^0 = \mathbf{e}_L$. Now as before we can write

$$
\begin{aligned}
\sum_{i=0}^{\infty} \mathbf{v}^i &= \mathbf{v}^0(\mathbf{I}-\mathbf{P}_k)^{-1} \\
&= \mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}, \\
\mathbf{g}(z)^t &= \frac{a(z)}{z}\mathbf{z}^t + \mathbf{P}_k\mathbf{z}^t.
\end{aligned}
$$

Then we can see that

$$
\begin{aligned}
r(z) &= \frac{a(z)}{z}\left(\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\right)\mathbf{z}^t - \left(\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\right)\mathbf{P}_k\mathbf{z}^t \\
&= \frac{a(z)}{z}\left(\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\right)\mathbf{z}^t + \left(\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\right)(\mathbf{I}-\mathbf{P}_k)\mathbf{z}^t - \left(\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\right)\mathbf{z}^t \\
&= \frac{1}{z}\{a(z)-z\}\,\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t + z^L. \qquad (7.4)
\end{aligned}
$$

Substituting (7.4) into (7.3) we get

$$\begin{aligned} E\left[z^X\right] &= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{\{a(z)-z\}\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t\left[b(z)-a(z)\right]}{(a(z)-z)(b(z)-z)}\right\} \\ &= \frac{Y(z)}{m} + \frac{h_1}{m(1-h)}\left\{\frac{\{b(z)-a(z)\}\mathbf{e}_L(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t}{(b(z)-z)}\right\}. \end{aligned}$$

Now $Y(z)$ is given by

$$Y(z) = \left\{\frac{E\left[z^{X_{\tau_1(0)}}\right]-z}{1-\xi_1(z)} + \frac{1-E\left[z^{X_{\tau_1(0)}}\right]}{1-\xi_2(z)}\right\},$$

which we know, from the M/G/1 queue with a single upwards threshold at K, to be

$$Y(z) = \left\{\frac{b(z)(1-z)+\{b(z)-a(z)\}\,\mathbf{e}_1(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t}{b(z)-z}\right\},$$

and so

$$E\left[z^X\right] = \frac{1}{m}\left\{\frac{b(z)(1-z)+\{b(z)-a(z)\}\left[\left(\mathbf{e}_1+\left(\frac{h_1}{1-h}\right)\mathbf{e}_L\right)(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^t\right]}{b(z)-z}\right\}$$

m is calculated as before by taking $\lim_{z\uparrow 1}$ and using L'Hôpital's rule. From this we get

$$\begin{aligned} m &= \left[\frac{-b(1)+\{b'(1)-a'(1)\}R_{KL}^{(UD)}(1)}{b'(1)-1}\right] \\ &= \left[\frac{1+\{\rho_a-\rho_b\}R_{KL}^{(UD)}(1)}{1-\rho_b}\right]. \end{aligned}$$

Now from the definitions of h and h_1 we get

$$\begin{aligned} h_1 &= 1-E\left[z^{X_{\tau_1(0)}}\,\middle|\,X_0=0\right]_{z=0} \\ &= 1-\left[\frac{1}{z}\left(a(z)-z\right)\mathbf{e}_1\left(\mathbf{I}-\mathbf{P}_k\right)^{-1}\mathbf{z}^t+z\right]_{z=0} \\ &= 1-a_0\mathbf{e}_1\left(\mathbf{I}-\mathbf{P}_k\right)^{-1}\mathbf{e}_1{}^t, \\ h &= 1-E\left[z^{X_{\tau_{2n+1}(0)}}\,\middle|\,X_{\tau_{2n}(0)}=L\right]_{z=0} \\ &= 1-\left[\frac{1}{z}\left\{a(z)-z\right\}\mathbf{e}_L\left(\mathbf{I}-\mathbf{P}_k\right)^{-1}\mathbf{z}^t+z^L\right]_{z=0} \\ &= 1-a_0\mathbf{e}_L\left(\mathbf{I}-\mathbf{P}_k\right)^{-1}\mathbf{e}_1{}^t. \end{aligned}$$

Thus we have the proposition. □

Remark: Note that as condition $(*)$ is automatically true for $\rho_a \leq 1$ and $\rho_b \leq 1$ the proposition has been proved over this range. It would be nice to be able to prove it is true for $\rho_a > 1$. This, however, is somewhat elusive. If the connection between regularity and stability could be made into a necessary and sufficient relationship then this problem would be solved as this queue we have described above remains stable when $\rho_a > 1$ so long as $\rho_b \leq 1$.

7.3 Numerical examples

In this section we investigate a number of numerical examples. As was noted in the previous section the conditions necessary for the result have not all been proved. However, we shall support the conclusion with a number of examples. Obviously these do not form any sort of proof of the result but they lend support to our proposition.

We show in the graphs the probabilities of having zero to twelve customers in the system. The dots show the results as calculated using our method, while the bars show 95% confidence intervals for these probabilities, produced through simulations of the systems involved. We calculated the probabilities from the generating functions using Maple. This is a matter of expanding the generating function as a Taylor series about $z = 0$.

I have tried to present a spectrum of results. Each page has a different combination of service-time distributions and values of the thresholds K and L. Also within this I have varied the value of ρ_a the traffic intensity during the first regime. All of the examples are limited to $\rho_b = 0.5$ in order that comparisons between different systems might be made.

Of interest might be the fact that these show some nice behaviour in some cases. For instance the probability of having the system empty is quite small in some cases while the probability of having more than say ten customers is equally small. This type of control over the system was one of the stated aims of using this type of threshold and so these results are encouraging.

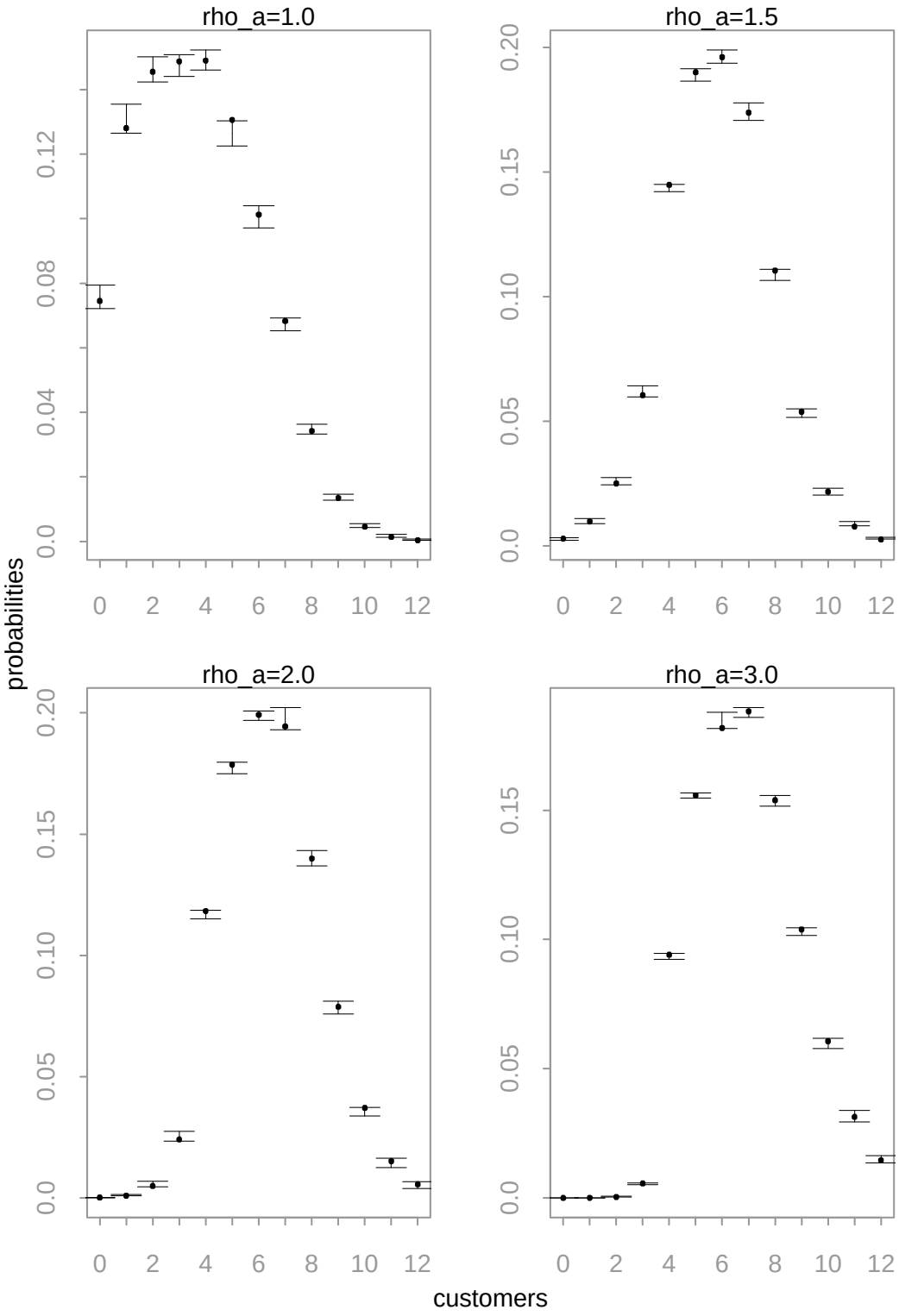

Figure 7.3: The two-regime process with K=7, L=4 and deterministic service times.

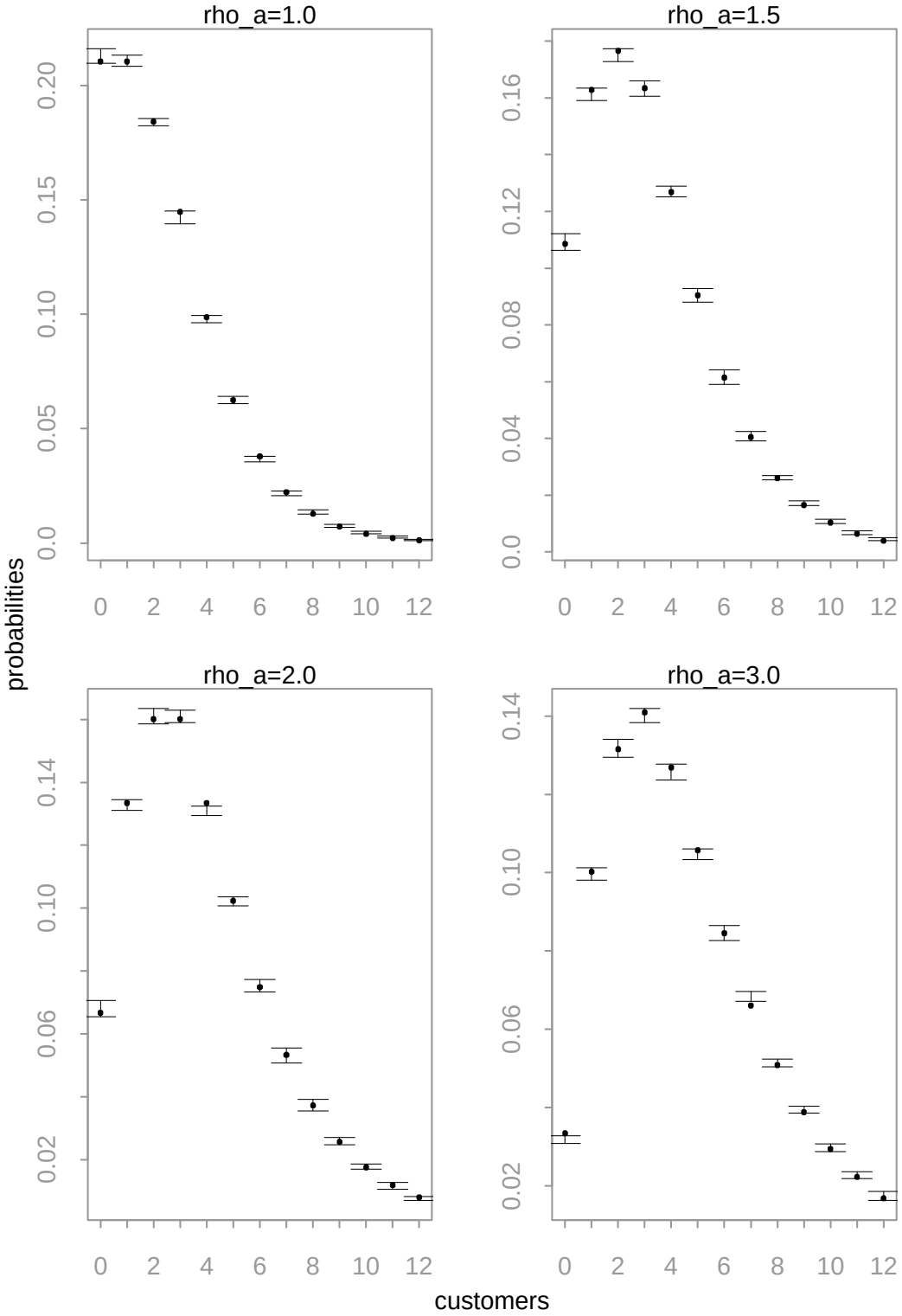

Figure 7.4: The two-regime process with K=3, L=1 and negative exponential service times.

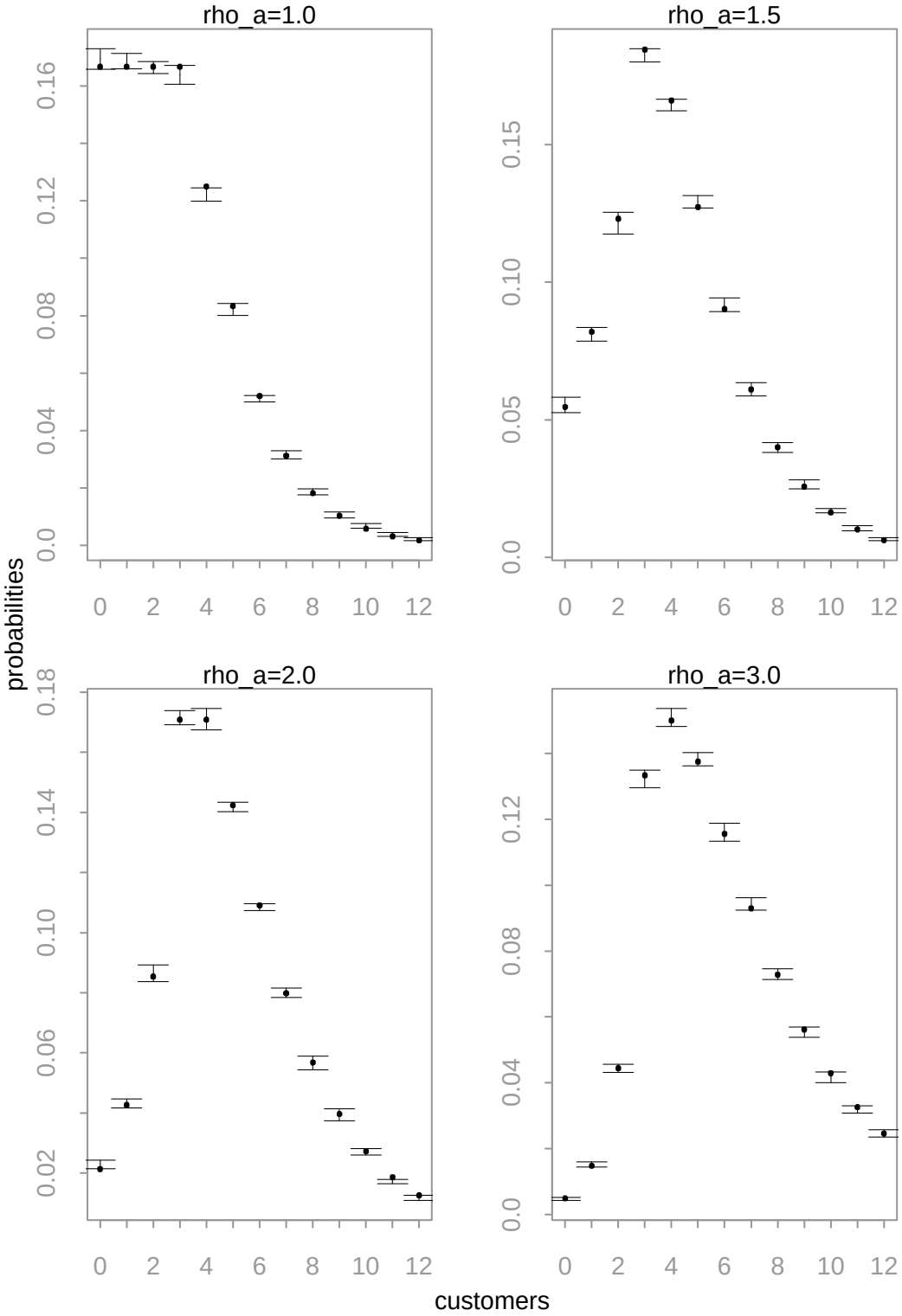

Figure 7.5: The two-regime process with K=3, L=3 and negative exponential service times.

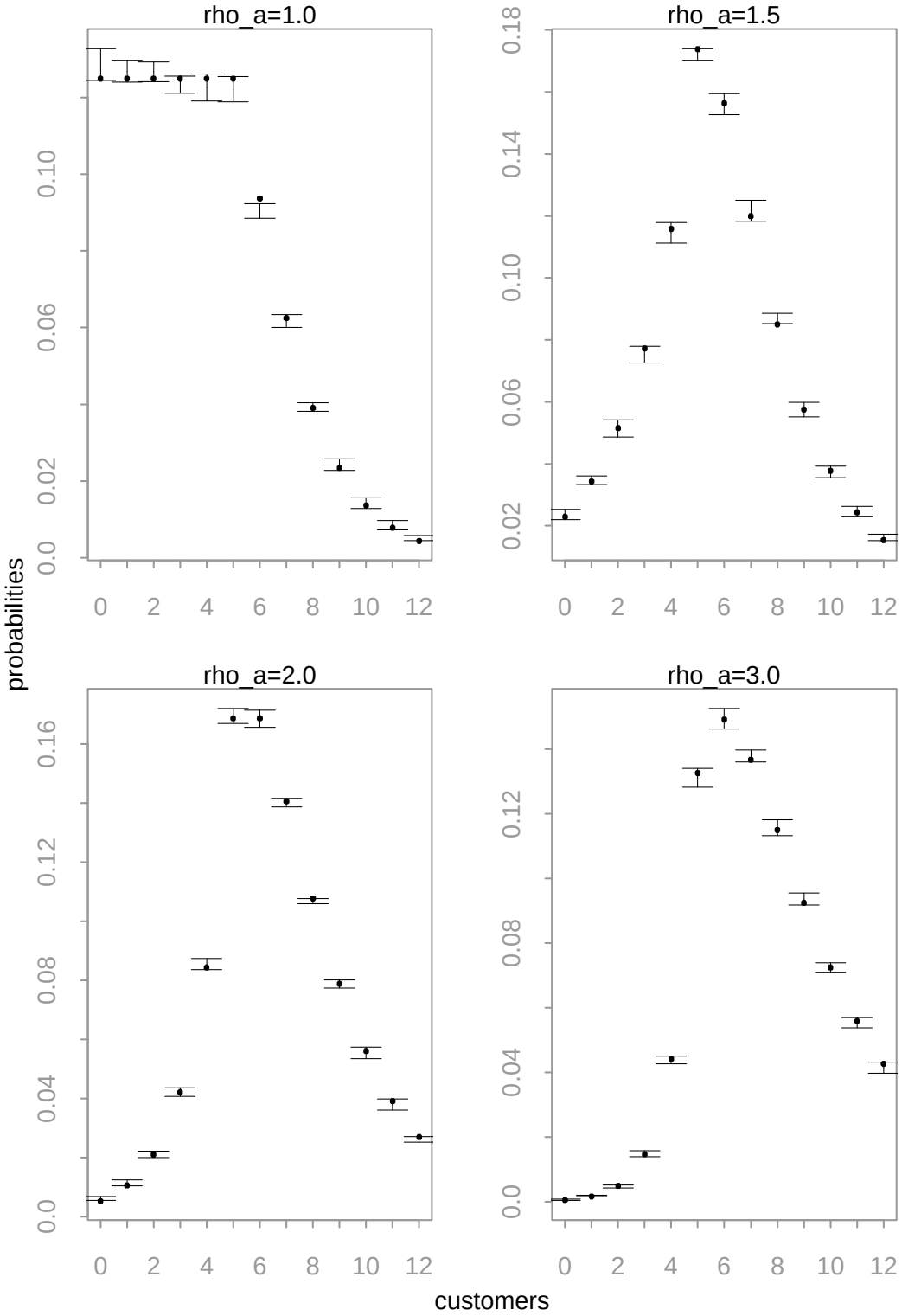

Figure 7.6: The two-regime process with K=5, L=5 and negative exponential service times.

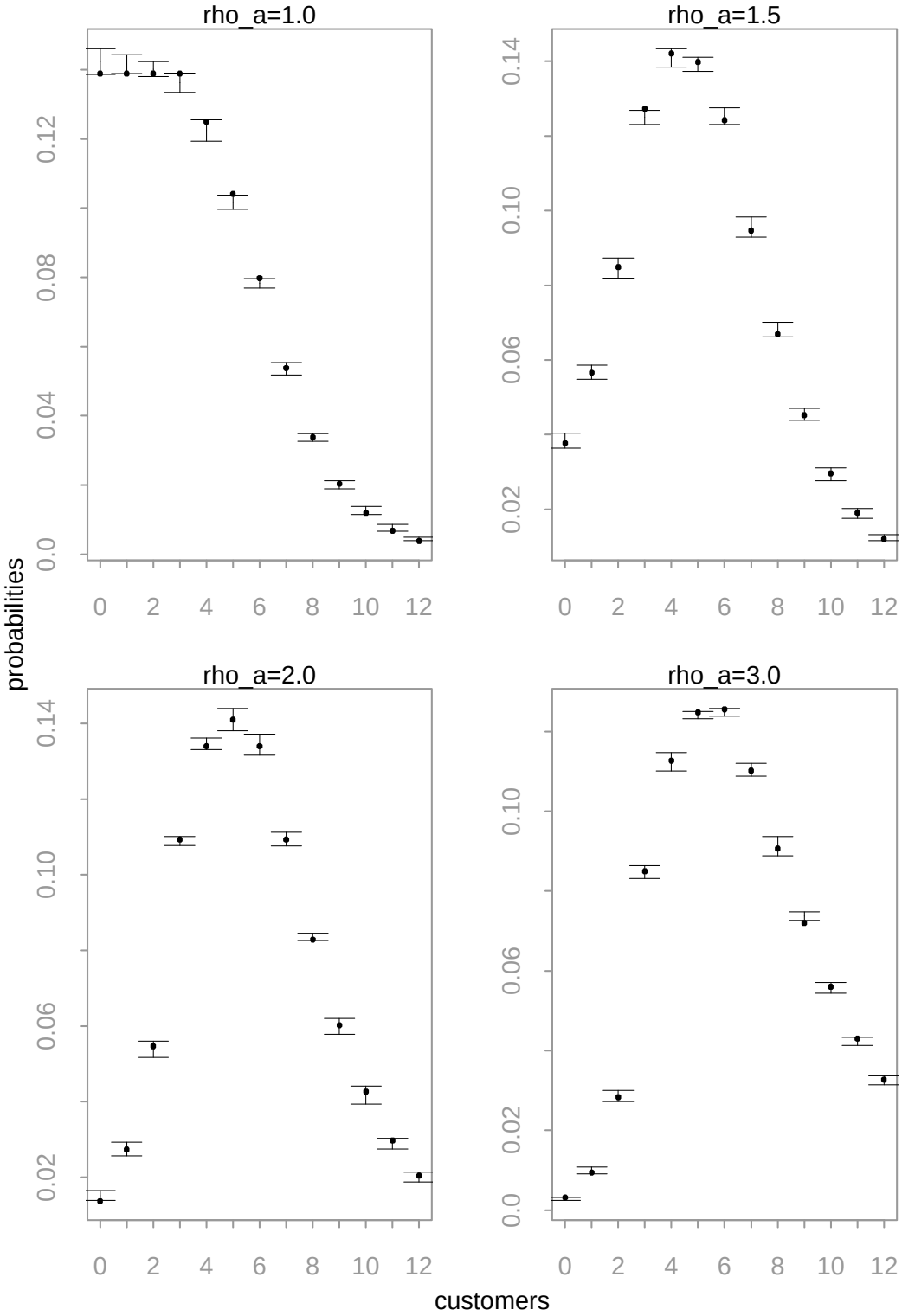

Figure 7.7: The two-regime process with K=6, L=3 and negative exponential service times.

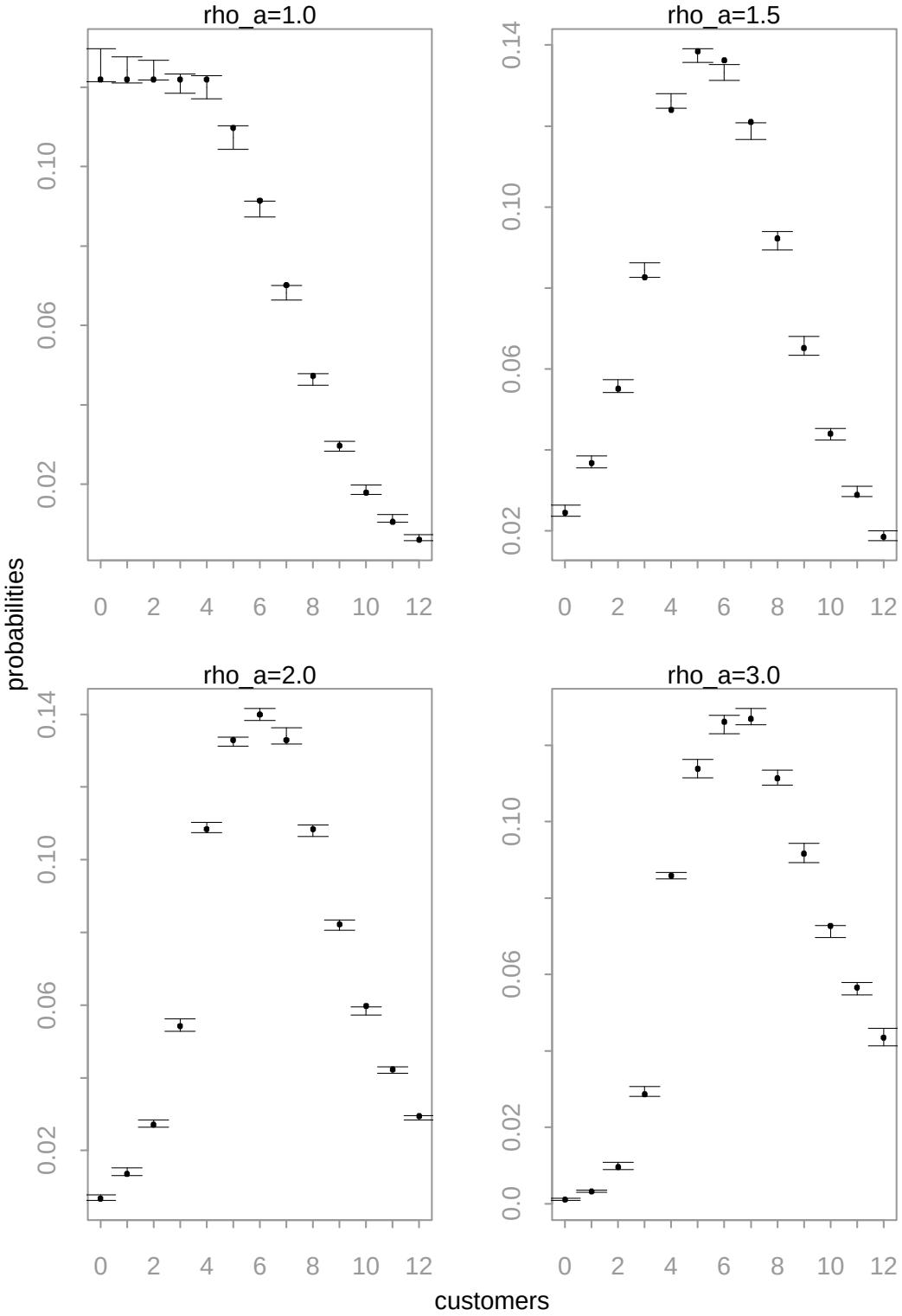

Figure 7.8: The two-regime process with K=7, L=4 and negative exponential service times.

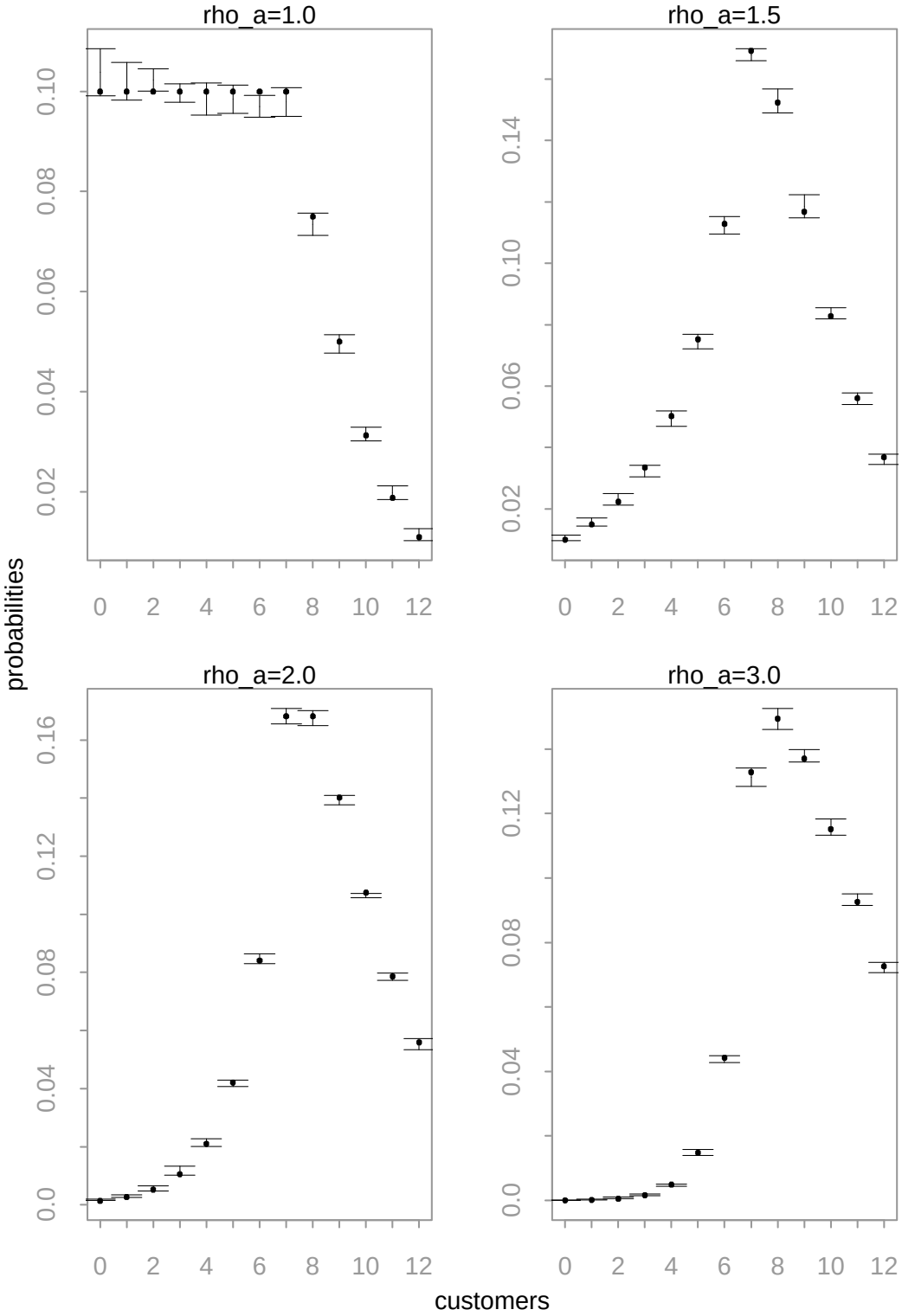

Figure 7.9: The two-regime process with K=7, L=7 and negative exponential service times.

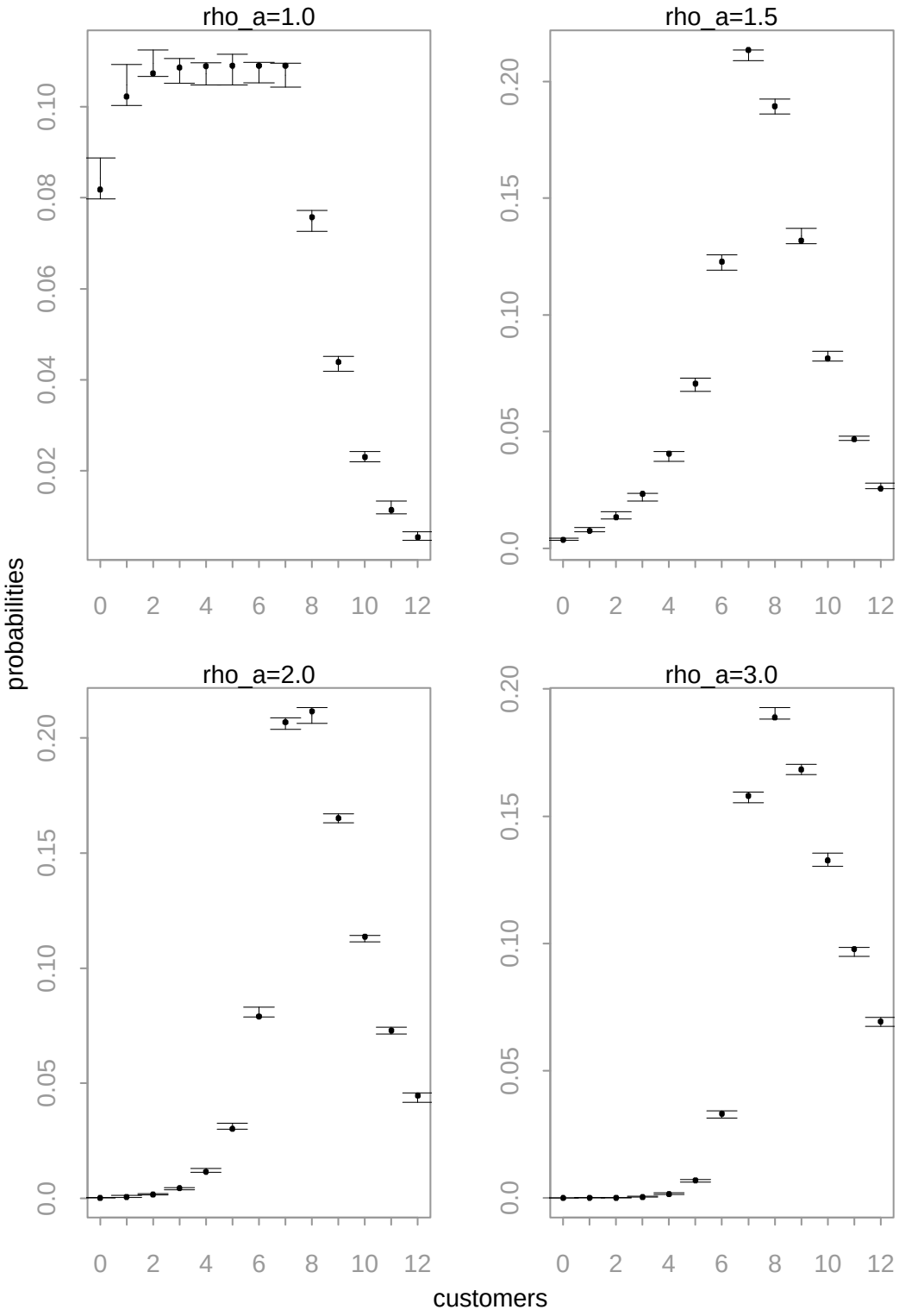

Figure 7.10: The two-regime process with K=7, L=7 and Erlang order 2 service times.

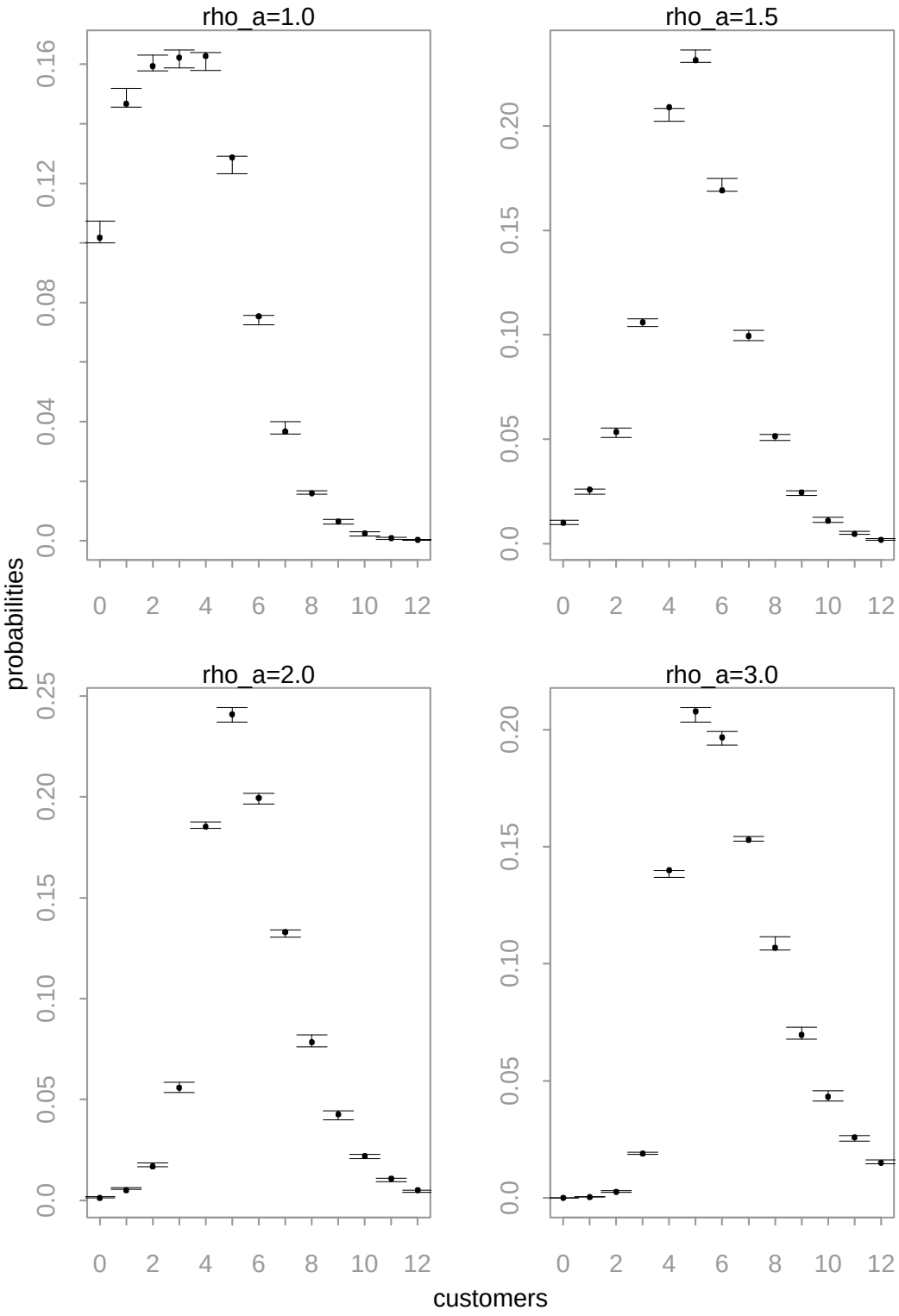

Figure 7.11: The two-regime process with K=5, L=4 and Erlang order 4 service times.

7.4 The probability of a given regime

We can extend the technique of Chapter 4 for finding the probability of a given phase to finding the probability of a given service regime, the technique is almost the same. We apply Little's law to the server so

$$\begin{aligned} L &= p\{X \neq 0\} = 1 - \frac{1}{m}, \\ W &= \frac{1-\psi}{\mu_a} + \frac{\psi}{\mu_b}, \end{aligned}$$

where ψ is the probability of of being in the second regime. For the case we have described

$$\begin{aligned} L &= \frac{\rho_b + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)}{1 + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)}, \\ \lambda W &= \psi(\rho_b - \rho_a) + \rho_a. \end{aligned}$$

Substituting in (4.23) we get an equation for ψ when $\rho_a \neq \rho_b$

$$\begin{aligned} \psi &= \left[\frac{\rho_b + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)}{1 + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)} - \rho_a\right]\frac{1}{\rho_b - \rho_a} \\ &= \left[\frac{(\rho_b - \rho_a) + (1-\rho_a)(\rho_a - \rho_b)R_{KL}^{(UD)}(1)}{1 + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)}\right]\frac{1}{\rho_b - \rho_a} \\ &= \frac{1 + (\rho_a - 1)R_{KL}^{(UD)}(1)}{1 + (\rho_a - \rho_b)R_{KL}^{(UD)}(1)}. \end{aligned} \tag{7.5}$$

This is the probability of being in regime 2.

When $K = L$ this gives the special result, the probability that there are more than K customers in the system. This can be used in the M/G/1 queue to calculate the equilibrium distribution of customers in the system. Note that we can consider $\psi = \psi(\rho_b)$ as the only dependence on $B(\cdot)$ is through ρ_b. If we then take a series of distributions such that $\rho_b \to \rho_a$ we get

$$\psi(\rho_a) = 1 + (\rho_a - 1)R_{KK}^{(UD)}(1).$$

When $B(t) = A(t)$ for all $t \in I\!R$ this simply gives the probability that there are more than K customers in the M/G/1 queue. Note that this is insensitive to the actual distribution $B(\cdot)$ except through its mean. We can then immediately deduce that w_n, the probability that there are n customers in the system, is given by

$$w_0 = (1 - \rho_a), \tag{7.6}$$

$$w_n = (1 - \rho_a)[R_{KK}^{(UD)}(1) - R_{K-1,K-1}^{(UD)}(1)], \tag{7.7}$$

with $R_{00}^{(UD)}(1) = 1$. (When $K = L = 0$ we have the situation where a customer arriving at the empty server has a different service-time distribution from all other customers. In Section 4.2.1 the solution is given for this and in order to be consistent with this we set $R_{00}^{(UD)}(1) = 1$.)

7.5 The M/G/1/N+1 queue

The M/G/1/N+1 queue is the M/G/1 queue with a finite waiting room. Any arrivals that occur when the waiting room is full are blocked. If we take the maximum queue size to be N then the maximum number of customers in the system is $N+1$. As in the previous examples we consider the process embedded at departure epochs. When we consider a departure we include customers who are blocked and leave the system immediately. After a customer that receives service leaves the queue there can be no more than N customers in the system and so the probability that there are $N+1$ customers in the system will give the blocking probability.

We shall not model this in the standard way. We take the system to be an multi-phase M/G/1 queue as described in this chapter with $L = K = N$. The probability generating function for the service times during the second regime is given by

$$B(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0. \end{cases}$$

This means that when the process is in regime two customers are served in zero time. This is equivalent to blocking the customers, except that the customer involved spends some time in the queue before being expelled.

Note that this would in fact give us a queue in which arrivals can stay in the system when there are more than N customers in the queue because the system must wait until the end of the current service before expelling any excess customers. Also the excess customers are be expelled in the order of service. Thus, as we want to expel the customers in the correct order (that is, remove those customers who arrived when the buffer size was exceeded and not those already in the system), the service discipline must be the non-preemptive last in first out discipline. We can however swap disciplines as well as service-time distribution at transition points and so we can always swap to this discipline when needed. Furthermore the service discipline does not effect the number of customers in the system only the waiting time distribution for those customers.

The extra arrivals during the service will affect the results. Obviously if we allow more than $N+1$ customers in the system this will affect the distribution for the number of customers in the system. However, if we note that when there are more than N customers in the queue the extra customers are all served in regime two and hence correspond to blocked customers, then we can see that the blocking probability will simply be the probability of being in regime two. We can get this probability from Section 7.4.

Proposition 7.2 *The blocking probability in the M/G/1/N+1 queue, ψ, is given by*

$$\psi = \frac{1 + (\rho_a - 1)R_{NN}^{(UD)}(1)}{1 + \rho_a R_{NN}^{(UD)}(1)},$$

where

$$R_{NN}^{(UD)}(1) = \left(\mathbf{e}_1 + \left(\frac{h_1}{1-h}\right)\mathbf{e}_N\right)(\mathbf{I} - \mathbf{P}_N)^{-1}\mathbf{1}^t.$$

Proof: From Proposition 7.1 for $0 \leq z < 1$ and $\rho_b < 1$ the generating function for the equilibrium behaviour of the queueing process described above is given by

$$E\left[z^X\right] = \frac{1}{m}\left\{\frac{b(z)(1-z) + \{b(z) - a(z)\}\, z R_{NN}^{(UD)}(z)}{b(z) - z}\right\},$$

where m, the mean number of customer's served in one busy period, is given by

$$m = \left[\frac{1 + \{\rho_a - \rho_b\} R_{NN}^{(UD)}(1)}{1 - \rho_b}\right].$$

Now $b(z) = 1$ so $\rho_b = 0$ and

$$E\left[z^X\right] = \frac{1}{1 + \rho_a R_{NN}^{(UD)}(1)}\left\{\frac{(1-z) + \{1 - a(z)\}\, z R_{NN}^{(UD)}(z)}{1 - z}\right\}. \tag{7.8}$$

Note that from (7.5) we get the probability of being in the second regime ψ by

$$\begin{aligned}\psi &= \frac{1 + (\rho_a - 1) R_{NN}^{(UD)}(1)}{1 + (\rho_a - \rho_b) R_{NN}^{(UD)}(1)} \\ &= \frac{1 + (\rho_a - 1) R_{NN}^{(UD)}(1)}{1 + \rho_a R_{NN}^{(UD)}(1)},\end{aligned}$$

where

$$R_{NN}^{(UD)}(1) = \left(\mathbf{e}_1 + \left(\frac{h_1}{1-h}\right)\mathbf{e}_N\right)(\mathbf{I} - \mathbf{P}_N)^{-1}\mathbf{1}^t.$$

This is the blocking probability. □

Now in our multi-phase M/G/1 queue we know that the probability that there are more than N customers in the system is the blocking probability, which in turn is the probability that there are more than N customers in the M/G/1/N+1 system. Also the probability that there are zero customers in the system is the same for both systems. Furthermore the behaviour of the two systems when the number of customers in the system is less than $N+1$ is identical. Thus the equilibrium distributions for the number of customers in the systems, given this number is less than $N+1$, are the same. Thus we have a successful model of the M/G/1/N+1 queue.

7.5.1 A check of the blocking probabilities

In order to check this result we use some results from Cooper (1972) and Cohen (1969). We define

$$\begin{aligned}p_n &= p\{\text{ a customer leaving the M/G/1/N+1 queue leaves n customers behind}\}, \\ q_n &= p\{\text{ a customer who receives service in the M/G/1/N+1 queue leaves n customers}\}, \\ w_n &= p\{\text{ a customer leaving the M/G/1 queue leaves n customers}\}.\end{aligned}$$

Note that p_n includes customers who are blocked and therefore leave $N+1$ customer behind. Cooper, pages 179-182 gives us

$$\begin{aligned} p_n &= \frac{q_n}{q_0+\rho_a}, \qquad 0 \le n \le N, \\ p_{N+1} &= \frac{q_0+\rho_a-1}{q_0+\rho_a}. \end{aligned}$$

Cohen, 6.26, page 560 gives for $0 \le n \le N$ that

$$q_n = \frac{w_n}{w_0+\cdots+w_N}.$$

From (7.6) and (7.7) we get

$$\begin{aligned} w_0 &= (1-\rho_a), \\ w_n &= (1-\rho_a)[R^{(UD)}_{nn}(1) - R^{(UD)}_{n-1,n-1}(1)], \end{aligned}$$

and so

$$\begin{aligned} \sum_{n=0}^{N} w_n &= (1-\rho_a) + \sum_{n=1}^{N}(1-\rho_a)[R^{(UD)}_{nn}(1) - R^{(UD)}_{n-1,n-1}(1)] \\ &= (1-\rho_a) + (1-\rho_a)[R^{(UD)}_{NN}(1) - R^{(UD)}_{00}(1)] \\ &= (1-\rho_a)R^{(UD)}_{NN}(1). \end{aligned}$$

Substituting back we get

$$\begin{aligned} q_0 &= \frac{(1-\rho_a)}{(1-\rho_a)R^{(UD)}_{NN}(1)} \\ &= \frac{1}{R^{(UD)}_{NN}(1)}, \end{aligned}$$

and from this we get from Cooper, 179-182,

$$\begin{aligned} p_0 &= \frac{\frac{1}{R^{(UD)}_{NN}(1)}}{\frac{1}{R^{(UD)}_{NN}(1)}+\rho_a} \\ &= \frac{1}{1+\rho_a R^{(UD)}_{NN}(1)}, \end{aligned}$$

which agrees with the value of p_0 derived from $p_0 = 1/m$. We also get

$$\begin{aligned} p_{N+1} &= \frac{q_0+\rho_a-1}{q_0+\rho_a} \\ &= \frac{\frac{1}{R^{(UD)}_{NN}(1)}+\rho_a-1}{\frac{1}{R^{(UD)}_{NN}(1)}+\rho_a} \\ &= \frac{1+(\rho_a-1)R^{(UD)}_{NN}(1)}{1+\rho_a R^{(UD)}_{NN}(1)}, \end{aligned}$$

which agrees with our blocking probability.

Chapter 8

Conclusion

8.1 Further work

There are a number of ways in which this work can be generalised. In this section we investigate some of the potential areas for future work. The most obvious is further use of the results of Chapter 3 to examine more problems. The examples of systems that we have considered here are, in some cases, a little too simple to model real situations. There are many possible systems which fit the multi-phase model and examination of these cases is not trivial as they require work to obtain condition $(*)$ and the values of $E\left[X^{\tau_j(0)}\right]$.

One of the examples that might prove interesting immediately are the extension of the results of Section 4.2 to deal with n upward thresholds at $k_1, \ldots, k_n$. A suggested result is given for this case in Proposition 5.1, but we have not proved this here.

Another example of interest would be that of Chapter 7 extended to multiple upward and downward thresholds at $k_1, \ldots, k_n$ and $l_1, \ldots, l_n$. This would link up nicely with the previous case.

Further examples could be considered with different thresholds. One possibility is the combination of two or more thresholds in some way. For instance if we take the first time either one of two different conditions is satisfied. One way of dealing with this might be to add in extra phases, with the same service-time distribution in order to take account of the two different thresholds.

Considering other types of system may provide motivations for various types of thresholds and blocking patterns. For example it may be useful to consider the systems with discrete time services and batch arrivals. This has been suggested as a model for some parts of ATM networks. For instance we consider packets arriving as a Poisson stream containing a number of cells, and the server takes unit time to serve each packet.

Another direction for continuation of this work is simply to consider processes based on queues other than the M/G/1 queue. This is more difficult as it requires more theoretical results. However, in simple cases they should be analogous to the results in this thesis.

A further area that needs work is condition $(*)$. We have shown in Section 3.2.2 that this is a sufficient condition for the regularity of the relevant stopping times. We have yet to find a necessary condition for the types of process considered. This would be a worthwhile task. Closely related to this is the issue of stability. When condition

$(*)$ is satisfied the queue is stable. It would be both elegant and useful to demonstrate that the reverse is also true, or to provide some other condition for which this holds.

One final and important way is to generalise the results of Chapter 2 from multi-phase MRPs to GMRPs. This would allow the calculation of results without the need to resort to modifying the phase structure to give a multi-phase MRP. It would also eliminate a large proportion of the infinite phase examples, making it easier to obtain results. This would be invaluable but requires some work. Namely the multi-phase MRP results must be extended and the results of Chapter 3 must also be extended. The hardest part of this seems to be finding an equivalent condition to condition $(*)$.

8.2 Block-matrix geometric techniques

The type of problem we have considered in this thesis has strong similarities with many of the systems discussed by Neuts. One important point to make is that we use 'phase' slightly differently herein. Phase often refers to phase-type distributions for the service times or arrival process. This is not the same as our use of phase which refers to the phase of the whole process.

Secondly we note that many of the problems considered herein could be examined through block-matrix type methods.

For example in the single fixed-upward threshold case we could write the transition matrix as in Table 8.1 where in this matrix the state is the number of customers in the system and the substates give the current phase. Even in this simple case this matrix is not trivial. Block matrix methods could be applied by some partitioning of the state space but this would be quite complex, especially in more complicated cases.

8.3 Dénoûement

In conclusion there are three main points I would like to make.

The first is that we have succeeded in our aim, which was to apply a martingale technique to queueing theory. The main part of the theory, developed in Chapter 3, and supported by Chapter 2, provides a major new technnique for investigating single server queues with some sort of phase structure.

The second point I would like to make is that we have proved the utility of this method. We have done this by considering a number of example in Chapters 4 to 7. Further we have mentioned several recent papers in which processes which fit our model are investigated.

Finally I would like to note that this is a fruitful area for further study. In the Section 8.1 we point out just a few of the areas in which this theory has potential for expansion.

States	0	1	2	$\cdots$	k	$k+1$	$k+2$	$\cdots$
0	$\begin{matrix} a_0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} a_1 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} a_2 & 0 \\ 0 & 0 \end{matrix}$	$\cdots$	$\begin{matrix} a_k & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & a_{k+1} \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & a_{k+2} \\ 0 & 0 \end{matrix}$	$\cdots$
1	$\begin{matrix} a_0 & 0 \\ b_0 & 0 \end{matrix}$	$\begin{matrix} a_1 & 0 \\ 0 & b_1 \end{matrix}$	$\begin{matrix} a_2 & 0 \\ 0 & b_2 \end{matrix}$	$\cdots$	$\begin{matrix} a_k & 0 \\ 0 & b_k \end{matrix}$	$\begin{matrix} 0 & a_{k+1} \\ 0 & b_{k+1} \end{matrix}$	$\begin{matrix} 0 & a_{k+2} \\ 0 & b_{k+2} \end{matrix}$	$\cdots$
2	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} a_0 & 0 \\ 0 & b_0 \end{matrix}$	$\begin{matrix} a_1 & 0 \\ 0 & b_1 \end{matrix}$	$\cdots$	$\begin{matrix} a_{k-1} & 0 \\ 0 & b_{k-1} \end{matrix}$	$\begin{matrix} 0 & a_k \\ 0 & b_k \end{matrix}$	$\begin{matrix} 0 & a_{k+1} \\ 0 & b_{k+1} \end{matrix}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	
k	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\cdots$	$\begin{matrix} a_1 & 0 \\ 0 & b_1 \end{matrix}$	$\begin{matrix} 0 & a_2 \\ 0 & b_2 \end{matrix}$	$\begin{matrix} 0 & a_3 \\ 0 & b_3 \end{matrix}$	$\cdots$
$k+1$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\cdots$	$\begin{matrix} 0 & 0 \\ 0 & b_0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & b_1 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & b_2 \end{matrix}$	$\cdots$
$k+2$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\cdots$	$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & b_0 \end{matrix}$	$\begin{matrix} 0 & 0 \\ 0 & b_1 \end{matrix}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	

Table 8.1: The transition matrix for the two-phase M/G/1 queue with a fixed-upward threshold at k.

Appendix A

Probability and Martingales

In this appendix we examine some of the basic concepts and theorems upon which the work in this thesis is based. The experienced reader will find this to a large degree uninteresting. It is included for completeness and to remove any possible doubt about the form of various theorems or notation.

In most cases the theorems herein are presented without proof but reference to the proof is provided.

The basic theory can be found in many places. We have used here as the chief reference Williams (1991) with smatterings from Neveu (1975) and Brémaud (1981).

A.1 Elementary probability theory and notation

Take a set Ω. A *sigma-algebra* $\mathcal{F}$ on this set is any collection of subsets of Ω which satisfies the following properties.

(i) $\Omega \in \mathcal{F}$.

(ii) $\mathcal{F}$ is closed under complements, that is, $A \in \mathcal{F} \Rightarrow \Omega \backslash A \in \mathcal{F}$.

(iii) $\mathcal{F}$ is closed under countable unions, that is, $A_n \in \mathcal{F}$, $n \in I\!N \Rightarrow \bigcup_n A_n \in \mathcal{F}$.

Note that property (ii) combined with properties (i) and (iii) implies

(i) $\phi \in \mathcal{F}$,

(ii) $\mathcal{F}$ is closed under countable intersections, that is , $A_n \in \mathcal{F}$, $n \in I\!N$
$\Rightarrow \bigcap_n A_n \in \mathcal{F}$,

respectively. A σ-algebra generated by a collection of subsets $\mathcal{C}$ of Ω is the smallest σ-algebra that includes all of the subsets in $\mathcal{C}$, and is written $\sigma(\mathcal{C})$.

We define a set-function $P : \mathcal{F} \rightarrow [0, 1]$ to be a *probability measure* if it satisfies the following properties.

(i) $P(\Omega) = 1$,

(ii) For all series A_n, $n \in I\!N$, of disjoint members of $\mathcal{F}$,

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n).$$

Note that these properties imply $P(\phi) = 0$. A $\mathcal{F}$-measurable subset of Ω, is any set $A \in \mathcal{F}$. The $\mathcal{F}$-measurable subsets of Ω are called *events*. $P(A)$ is called the probability of event A, $\forall A \in \mathcal{F}$. We shall also write this as $p\{\cdot\}$ where $\{\cdot\} \in \mathcal{F}$. The triple $(\Omega, \mathcal{F}, P)$

is then called a *probability space*. Throughout, where it is not explicitly stated, we assume that all stochastic elements have some underlying probability space. An event A, is said to be *almost sure* (a.s.), or to occur *with probability one* (w.p.1), if $P(A) = 1$. We assume also that all σ-algebras are completed, that is, if $A \in \mathcal{F}$ with $P(A) = 0$ and $A' \subset A$ then $A' \in \mathcal{F}$.

A.1.1 Random variables

A function $h : \Omega \to I\!R$ is called $\mathcal{F}$-measurable if $h^{-1} : \mathcal{B} \to \mathcal{F}$, where we define h^{-1} by

$$h^{-1}(B) = \{\omega \in \Omega | h(\omega) \in B\}, \quad \forall B \in \mathcal{B},$$

and $\mathcal{B}$ denotes the Borel sets of the real line. This may be simplified slightly to the condition that

$$h^{-1}(B) = \{\omega \in \Omega | h(\omega) \in B\}, \text{ for all intervals } B \text{ of the form } (-\infty, a],$$

as the mapping preserves intersections and unions and a σ-algebra is closed under such operations. A real-valued $\mathcal{F}$-measurable function X, on Ω is called a *random variable*. That is, a function $X : \Omega \to I\!R$ such that $\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}$. Given a family $(X_\gamma | \gamma \in \mathcal{C})$ of maps $X_\gamma : \Omega \to I\!R$ we define the σ-algebra generated by the family,

$$\mathcal{F} = \sigma(X_\gamma | \gamma \in \mathcal{C}),$$

to be the smallest σ-algebra $\mathcal{F}$ on Ω such that all of the maps X_γ $(\gamma \in \mathcal{C})$ are $\mathcal{F}$-measurable.

A random variable X is called *lattice* if $p\{w | X(\omega) = b + kn,\, n \in Z\!\!\!Z\} = 1$. Written simply a random variable is lattice or discrete if $X = b + kn,\, n \in Z\!\!\!Z$ almost surely. If k is the largest number for which this is true then k is called the span of the random variable.

We define a random variable called the *indicator function* I, of an event $A \in \mathcal{F}$, by

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

In some cases we shall write $I(A)$ for I_A where $A \in \mathcal{F}$.

Given a random variable X on $(\Omega, \mathcal{F}, P)$, we can define the *probability distribution function* $F : I\!R \to [0, 1]$ of the random variable X by

$$F(x) = p\{X \leq x\} = p\{\omega | X(\omega) \leq x\}.$$

$F(x)$ must satisfy the following properties,

(i) $\lim_{x \to -\infty} F(x) = 0$.

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

(iii) $F(x)$ is non-decreasing.

(iv) $F(x)$ is right-continuous.

If $F(x)$ is differentiable we define $f(x) = \frac{dF}{dx}$ to be the *probability density function* of X. Where ambiguity exists we shall write $F_X(x)$ for the probability distribution function of X.

A.1.2 Independence

Independence of random variables and σ-algebras will be important in a number of places. Simply, we call two random variable's X and Y *independent*, if

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}.$$

More technical definitions are included here to cover events and σ-algebras.
(i) Events
Events $E_1, E_2, E_3, \ldots$ are *independent* if for $i_1, ..i_n \in I\!N$ distinct,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}) = \prod_{k=1}^{n} P(E_{i_k}).$$

(ii) σ-algebras
Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ of $\mathcal{F}$ are called *independent* if $\forall G_i \in \mathcal{G}_i$, $(i \in I\!N)$ and $i_1, ..i_n \in I\!N$ distinct, the events $G_{i_1}, G_{i_2}, \ldots, G_{i_n}$ are independent.
(iii) Random Variables
Random variables $X_1, X_2, \ldots$ are *independent* if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots$ are independent.
(iv) Combinations
Combinations of the three above, are said to be independent if the relevant combinations of σ-algebras are independent. For instance, a random variable X, and an event E, are said to be independent if $\sigma(X)$ and $\mathcal{E}$ are independent, where $\mathcal{E} = \{\phi, E, \Omega \backslash E, \Omega\}$, the σ-algebra generated by E.

A.1.3 Expectation

If we have probability space $(\Omega, \mathcal{F}, P)$ then the expectation of a $\mathcal{F}$-measurable random variable X, is simply the integral of X on Ω with respect to P. That is

$$E[X] = \int_\Omega X dP = \int_\Omega X(\omega) P(d\omega)$$

where this is defined.

As in all measure theory this integral must be defined through a series of extension from the basic measure $P(A)$ where $A \in \mathcal{F}$. Briefly this is done in the following way.
(i) Simple random variables
A simple random variable is a non-negative random variable which can be written

$$X = \sum_{i=1}^{m} \alpha_i I_{A_i},$$

for $\alpha_k \in [0, \infty]$ and $A_k \in \mathcal{F}$. The expectation of such a random variable is simply

$$E[X] = \sum_{i=1}^{m} \alpha_i P(A_i),$$

where $\infty.0$ is defined to be 0.

(ii) Non-negative random variables
The monotone convergence theorem is used here to show that

$$E[X] = \sup\{P(Y)|Y simple, Y \leq X\},$$

is a valid integral for non-negative random variables. This is done by showing that any non-negative $\mathcal{F}$-measurable random variable X is the limit of a series of simple random variables.

(iii) $\mathcal{F}$-measurable random variables
We now write a $\mathcal{F}$-measurable random variable X as $X = X^+ - X^-$ where $X^+(\omega) = max(X(\omega), 0)$ and $X^-(\omega) = max(-X(\omega), 0)$. Note that $|X| = X^+ + X^-$. A random variable for which $E[|X|] < \infty$ is said to be *integrable* and we denote the family of $\mathcal{F}$-measurable, integrable random variables by $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. For all random variables $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ we can write the expectation as follows

$$E[X] = E[X^+] - E[X^-].$$

Also even if only one of the conditions $E[X^+] < \infty$ or $E[X^-] < \infty$ is satisfied then the expectation is defined as above. Notationally we take, for $A \in \mathcal{F}$ and integrable random variable X,

$$E[XI_A] = \int_\Omega XI_A dP = \int_A X dP.$$

For a Borel measurable function $f : I\!R \to I\!R$ we define a new random variable $f(X) : \Omega \to I\!R$ by $f(X)(\omega) = f(X(\omega))$. When $f(X) \in \mathcal{L}^1$ we define the expectation as before.

$$E[f(X)] = \int_\Omega f(X) dP.$$

By Theorem **T15** of Brémaud (1981) we arrive at the Lebesgue-Stieltjes integral for the expectation in terms of the probability distribution function of a random variable Explicitly stated this gives,

$$E[g(X)] = \int_\Omega g(X(\omega)) P(d\omega) = \int_{I\!R} g(x) dF_X(x).$$

Theorem A.1 (Jensen's Inequality) *If $c : H \to I\!R$ is a convex function on an open sub-interval H of the real line, and X is an integrable random variable such that*

$$p\{X \in H\} = 1, \quad E[|c(X)|] < \infty,$$

then

$$E[c(X)] \geq c(E[X]).$$

Definition A.1 *For $1 \leq p < \infty$ we say that $X \in \mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, P)$ if*

$$E[|X|^p] < \infty,$$

and define

$$||X||_p = \{E[|X|^p]\}^{\frac{1}{p}}.$$

Definition A.2 (Laplace-Stieltjes transform) *The Laplace-Stieltjes transform of a random variable X, on the probability space $(\Omega, \mathcal{F}, P)$ is*

$$E\left[e^{-sX}\right] = \int_{-\infty}^{\infty} e^{-sx} dF_X(x).$$

Definition A.3 (Probability generating function) *The probability generating function of a lattice random variable X, is*

$$E\left[z^X\right] = \sum_{n=-\infty}^{\infty} z^n p\{X = nd\},$$

where d is the span of the random variable.

Given a probability distribution function $F(x)$ which is non-lattice or lattice we can define its Laplace-Stieltjes transform or probability generating function respectively, as above, and we denote this by $F^*(s)$ or $F^*(z)$ respectively. (The probability generating function is a special case of the Laplace-Stieltjes transform.)

If we take the random variable given by the sum of two independent random variables X and Y the the probability distribution function for this is given by

$$F_{X+Y}(x) = \int_{-\infty}^{x} F_X(x-y) dF_Y(y).$$

Definition A.4 *We define the* **convolution** *of two independent probability distribution functions $F_X(x)$ and $F_Y(y)$ by*

$$F_X * F_Y(x) = \int_{-\infty}^{x} F_X(x-y) dF_Y(y).$$

This is the probability distribution function for the sum of X and Y. The n-fold convolution of F_X is simply denoted $F_X^{(n)}(x)$.

Note that the Laplace-Stieltjes transform of the convolution $F_X * F_Y$ is simply the product of the respective transforms of F_X and F_Y.

A.1.4 Convergence

There are a number of different types of convergence. The two used herein are

(i) Almost sure convergence

Suppose that (X_n) is a sequence of random variables and X is a random variable then we say that $X_n \to X$ almost surely if, as $n \to \infty$,

$$P(X_n \to X) = 1.$$

(ii) Convergence in mean

Suppose that (X_n) is a sequence of random variables and X is a random variable then we say that $X_n \to X$ in mean (in expectation, in $\mathcal{L}^1$) if, as $n \to \infty$,

$$E[|X_n - X|] \to 0,$$

and thence as $n \to \infty$

$$E[|X_n|] \to E[|X|].$$

The following theorems relate the two types of convergence.

Theorem A.2 (Monotone Convergence Theorem) *If (X_n) is a series of random variables and X is a random variable such that $0 \leq X_n \uparrow X$ a.s. then*

$$E[X_n] \uparrow E[X].$$

Proof: Williams pages 211-213. □

Theorem A.3 (Dominated Convergence Theorem) *If (X_n) is a series of random variables and Y is a random variable such that $|X_n(\omega)| \leq Y(\omega)$ for all (n, ω) and $E[Y] < \infty$ then (X_n) converges to some random variable X in mean as $n \to \infty$.*

Proof: Williams page 55. □

Theorem A.4 (Bounded Convergence Theorem) *If (X_n) is a series of non-negative random variables bounded above by $M < \infty$, for all (n, ω) then (X_n) converges to some random variable X in mean as $n \to \infty$.*

Proof: This theorem is an immediate consequence of the Dominated Convergence theorem. □

A.1.5 Conditional expectation

We can define the conditional probability of an event $A \in \mathcal{F}$ given event $B \in \mathcal{F}$ of positive probability by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

This can be seen to form a new probability measure on $(\Omega, \mathcal{F})$. From this we may naively define a new expectation, the expectation of X given the event B, by

$$\begin{aligned} E[X|B] &= \int_\Omega X(\omega) P(d\omega|B) \\ &= \frac{1}{P(B)} \int_\Omega X(\omega) P(d\omega \cap B) \\ &= \frac{1}{P(B)} \int_B X(\omega) P(d\omega) \\ &= \frac{1}{P(B)} \int_\Omega I_B X(\omega) P(d\omega) \\ &= \frac{E[XI_B]}{P(B)}. \end{aligned}$$

A useful result for dealing with such expectations is as follows. If $A_0, A_1, A_2, \ldots$ is a complete system of events (i.e.: mutually exclusive and exhaustive) then for a random variable X

$$E[X] = \sum_{n=0}^{\infty} E[X|A_n] P(A_n).$$

This naive definition of conditional expectation does not in general suffice and so we must use a more technical definition.

Theorem A.5 (Conditional Expectation) *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and X an integrable random variable Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists an integrable $\mathcal{G}$-measurable random variable Y such that for every set $G \in \mathcal{G}$ we have*

$$\int_G Y dP = \int_G X dP.$$

Moreover if $\hat{Y}$ is another random variable with these properties then $\hat{Y} = Y$ a.s. The random variable Y is called a **version of the conditional expectation**, *$E[X|\mathcal{G}]$ of X given $\mathcal{G}$. We write*

$$Y = E[X|\mathcal{G}], \quad a.s.$$

Proof: Proof that such a random variable exists and is unique (except for a set of measure 0) can be found in Williams, page 85. □

Notationally we normally identify all of the versions of the conditional expectation. In general we use the notation $E[X|G]$ where X is a random variable to mean the naive concept of conditional expectation and $E[X|Y]$ where X and Y are random variables to mean the random variable $E[X|\sigma(Y)]$.

Properties of conditional expectation:

(i) Jensen's inequality, the monotonic convergence theorem and the dominated convergence theorem all hold for conditional expectations.

(ii) If Y is any version of $E[X|\mathcal{G}]$ then $E[Y] = E[X]$.

(iii) If X is $\mathcal{G}$-measurable then $E[X|\mathcal{G}] = X$, a.s.

(iv) linearity, $E[a_1X_1 + a_2X_2|\mathcal{G}] = a_1E[X_1|\mathcal{G}] + a_2E[X_2|\mathcal{G}]$, a.s.

(v) positivity, if $X \geq 0$ then $E[X|\mathcal{G}] \geq 0$, a.s.

(vi) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}], \quad a.s.$$

(vii) If Z is a $\mathcal{G}$-measurable, bounded random variable then,

$$E[ZX|\mathcal{G}] = ZE[X|\mathcal{G}], \quad a.s.$$

(viii) If X is independent of $\mathcal{H}$ then

$$E[X|\mathcal{H}] = E[X], \quad a.s.$$

A.2 Martingale theory

This is the essential part of the background theory, but still is only a very small subset of martingale theory. We shall only consider discrete-parameter martingales for a start. Although many theorems on discrete-parameter martingales can be generalised to continuous-parameter martingales they require more rigour. We take a probability space $(\Omega, \mathcal{F}, P)$ as before. We then take a set of increasing sub-σ-algebras of $\mathcal{F}$, $(\mathcal{F}_n : n \in I\!N)$. That is

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}.$$

We define

$$\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right) \subset \mathcal{F}.$$

We call (F_n) a *filtration* on the probability space and write $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$. A process $(X_n : n \in I\!N)$ is called *adapted* if for each n, X_n is $\mathcal{F}_n$-measurable. In most cases if we have a stochastic process $(X_n : n \in I\!N)$ we define $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ and $\mathcal{F} = \mathcal{F}_\infty$. Thus $\{\mathcal{F}_n\}$ is automatically a filtration and the sequence (X_n) is adapted to this filtration. We shall refer, in such a case, to $(\mathcal{F}_n)$, as the *history* of the process.

Definition A.5 (Martingale) *A process $X = (X_n)$ is called an integrable* **martingale** *with respect to $(\{\mathcal{F}_n\}, P)$ if*
(i) X is adapted.
(ii) $E(|X_n|) < \infty, \ \forall n \in I\!N$.
(iii) $E[X_{n+1}|\mathcal{F}_n] = X_n, \ a.s., \ \forall n \in I\!N$.

Submartingales and *supermartingales* are defined in the same way except that (iii) is replaced by $E[X_{n+1}|\mathcal{F}_n] \geq X_n$ and $E[X_{n+1}|\mathcal{F}_n] \leq X_n$ respectively.

Properties (of martingales)
(i) $E[X_n] = E[X_0]$ for all $n \in I\!N$.
(ii) $E[X_{n+m}|\mathcal{F}_n] = X_n, \ a.s., \forall n, m \in I\!N$.

Definition A.6 (stopping times) *A map $T : \Omega \to I\!N \cup \{\infty\}$ is called a* **stopping time** *if $\{T \leq n\}$ if $\mathcal{F}_n$-measurable. That is,*

$$\{T \leq n\} = \{\omega | T(\omega) \leq n\} \in \mathcal{F}_n, \ \forall n \in I\!N.$$

Intuitively the idea is that at some point in the process you stop. Your decision to stop (or not stop if $T = \infty$) can be based only on the information you have up to that time. Thus your decision to stop the process is made on the basis of the history of the process up until time n. Note that if S and T are stopping times then so also are $S + T$, $S \wedge T$ and $S \vee T$. For a stopping time T we define the σ-algebra $\mathcal{F}_T$ by

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty | A \cap \{T \leq n\} \in \mathcal{F}_n, \forall n \geq 0\}.$$

Now if S and T are two stopping times then

$$\{S < T\}, \{S \leq T\}, \{S = T\}, \{S \geq T\} \text{ and } \{S > T\} \in \mathcal{F}_T,$$

and by symmetry also in $\mathcal{F}_S$.

Definition A.7 *A class $\mathcal{C}$ of random variables is called* **uniformly integrable** *if given $\epsilon > 0$, there exists $K \in [0, \infty)$ such that*

$$E\left[|X|\, I\left(|X| > K\right)\right] < \epsilon, \;\; \forall X \in \mathcal{C}.$$

Sufficient Conditions (for the class to be uniformly integrable.)
(i) It is bounded in $\mathcal{L}^p$, $p > 1$.
(ii) It is dominated by another integrable random variable.

Theorem A.6 *Let M be a uniformly integrable martingale then*

$$M_\infty = lim_{n\to\infty} M_n \;\; exists \;\; a.s. \;\; and \;\; in \;\; \mathcal{L}^1.$$

Moreover, for every n,

$$M_n = E[M_\infty | \mathcal{F}_n], \;\; a.s.$$

Neveu uses the term *regular* to denote uniformly integrable martingales (Proposition IV **2**-3). We shall use the latter appellation here. However he also uses regular in the following way.

Definition A.8 (Regularity) *A stopping time τ, is* **regular** *with respect to a martingale M_n if the martingale formed by $M_{\tau\wedge n}$ is uniformly integrable.*

This definition is in Neveu, Proposition IV **3** 12. Also in this proposition is the following theorem which is the fundamental result used herein in order to obtain the result crucial to this thesis. It is a surprisingly elegant theorem referred to elsewhere as the optional stopping theorem, the optional sampling theorem, or either or these two versions prefaced by Doob's, after its originator. Versions of this theorem come in several form as well. Essentially they are either equivalent to or simple corollaries to the following version.

Theorem A.7 (Doob's Optional Sampling Theorem) *If the process M_n is an integrable martingale and τ is a regular stopping time then for every pair of stopping times τ_1 and τ_2 such that $\tau_1 \leq \tau_2 \leq \tau$ almost surely the random variables X_{τ_1} and X_{τ_2} both exist, are integrable and satisfy*

$$X_{\tau_1} = E\left[X_{\tau_2} | \mathcal{F}_{\tau_1}\right].$$

Proof: Neveu, Proposition IV-**3**-12. □
Essentially this means that given a sequence of stopping times (T_n) which satisfies certain properties on a martingale M_n then the process formed by M_{T_n} is also a martingale. Two results from Neveu that are of use are also given here.

Theorem A.8 (Corollary IV-3-13) *Let τ_1 and τ_2 be two stopping times such that $\tau_1 \leq \tau_2$ almost surely. For a given martingale (M_n), the stopping time τ_1 is regular whenever the stopping time τ_2 is regular.*

Theorem A.8 shows that for a uniformly integrable martingale all stopping times are regular.

Theorem A.9 (Proposition IV-3-16) *Let (M_n) be an integrable martingale. In order that the stopping time τ be regular for the martingale and that also $\lim_{n\to\infty} M_n = 0$ almost surely on $\{\tau = \infty\}$, it is necessary and sufficient that the following two conditions be satisfied:*

$$\begin{array}{lrcl} \textbf{(i)} & \int_{\{\tau<\infty\}} |X_\tau| dP & < & \infty; \\ \textbf{(ii)} & \lim_{n\to\infty} \int_{\{\tau>n\}} |X_n| dP & = & 0. \end{array}$$

Note that from Neveu, Proposition IV-**3**-14 condition **(i)** is satisfied for all positive integrable martingales.

Appendix B

Vectors and matrices

A certain amount of the work presented in this thesis relies on matrix notation and a number of theorems. For simplicity this section details those parts of matrix theory used herein. The work on norms presented here is based on Barnett and Storey (1970) and Householder (1964), while the probabilistic parts come from Gantmacher (1959). Other specifics are referred to within the text. The norm which we present in (B.6) and use in Lemma 4.2.1 is the only original part of this Appendix.

Throughout this thesis boldface capitals refer to matrices (eg: $\mathbf{A}$) while boldface lowercase letters denote row-vectors (eg: $\mathbf{v}$). We will be for the most part only concerned with real square matrices. The following common notational conventions are used.

$$\begin{aligned}
\mathbf{P}^i &= \mathbf{P} \text{ to the power of } i, \\
\mathbf{P}^0 &= \mathbf{I}, \text{ the identity matrix}, \\
\mathbf{P}^{-1} &= \text{the inverse of } \mathbf{P}, \\
\mathbf{P}^t &= \text{the transpose of } \mathbf{P}, \\
\mathbf{v}^t &= \text{the column vector corresponding to } \mathbf{v}, \\
\mathbf{0} &= \text{the zero vector or matrix depending on the context.}
\end{aligned}$$

Matrix products and determinants are standard and the scalar product of two vectors $\mathbf{v}$ and $\mathbf{w}$ is $\mathbf{v}.\mathbf{w} = \mathbf{v}\mathbf{w}^t$. We can specify a matrix by its elements in the following way

$$\mathbf{A} = \{a_{ij}\},$$

where a_{ij} is the element in the ith row and jth column. The following standard vectors will be used throughout.

$$\begin{aligned}
\mathbf{e}_i &= (0, 0, \cdots, 0, 1, 0, \cdots, 0) \\
&= (\delta_{i1}, \delta_{i2}, \cdots, \delta_{in}), \\
\mathbf{z} &= (z, z^2, z^3, \cdots, z^n), \\
\mathbf{1} &= (1, 1, \cdots, 1).
\end{aligned}$$

Also the following lemma will be of use later on.

Lemma B.0.1 *For any matrix* $\mathbf{P}$ *such that* $(\mathbf{I}-\mathbf{P})^{-1}$ *exists*

$$\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1} = (\mathbf{I}-\mathbf{P})^{-1}\mathbf{P} = -\mathbf{I} + (\mathbf{I}-\mathbf{P})^{-1}.$$

Proof: The proof is trivial. □

B.1 Eigenvalues and eigenvectors

For any $n \times n$ matrix $\mathbf{A}$ a scalar λ and vector $\mathbf{v}$ that satisfy the equation

$$\mathbf{vA} = \lambda\mathbf{v}, \tag{B.1}$$

are called the eigenvalue and corresponding eigenvector of $\mathbf{A}$, respectively. Note, because we are working with row vectors by default, we use the left-eigenvector of $\mathbf{A}$. This alters some of the later work from the texts but only notationally. It is easily seen that the values λ satisfy the equation

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0.$$

This is called the characteristic equation of $\mathbf{A}$. The left-hand side of this equation is clearly a polynomial of degree n in λ and thus a $n \times n$ matrix has at most n (possibly complex) eigenvalues. We define the *spectral radius* of the $n \times n$ matrix $\mathbf{A}$ to be

$$\rho(\mathbf{A}) = \max_{i=1,..,n} |\lambda_i(\mathbf{A})|,$$

where $|\lambda_i(\mathbf{A})|$ is the absolute value of the ith eigenvalue of $\mathbf{A}$. Note that $\lambda(\mathbf{A}^n) = \lambda(\mathbf{A})^n$ and hence the same is true for the spectral radius, namely $\rho(\mathbf{A}^n) = \rho(\mathbf{A})^n$. Also $\lambda(\mathbf{I} - \mathbf{A}) = 1 - \lambda(\mathbf{A})$.

B.2 Norms

A matrix norm is a single non-negative real-valued scalar that provides a measure of the magnitude of a matrix (or vector), in some sense of magnitude. In general we say that a real-valued function $\|\mathbf{A}\|$ of the elements of the matrix $\mathbf{A}$ is a norm if it satisfies the following four properties.

$$\mathbf{A} \neq \mathbf{0} \Rightarrow \|\mathbf{A}\| > 0, \tag{B.2}$$

$$\|\lambda\mathbf{A}\| = |\lambda|\|\mathbf{A}\|, \tag{B.3}$$

$$\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|, \tag{B.4}$$

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|, \tag{B.5}$$

for all $n \times n$ matrices and $\lambda \in I\!R$, while a vector norm is a real-valued function of the elements of a vector $\mathbf{x}$, that satisfies the first three properties above. A vector norm is said to be consistent with a matrix norm if

$$\|\mathbf{xA}\| \leq \|\mathbf{x}\|\|\mathbf{A}\|,$$

for any $\mathbf{A}$ and $\mathbf{x}$.

Theorem B.1 *If $\|\mathbf{A}\|$ is a matrix norm consistent with vector norm $\|\mathbf{x}\|$ then the spectral radius $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ satisfies the inequality*

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|.$$

A norm of later use is the norm $\|\mathbf{P}\|_z$ for $z \in (0,1)$ which is defined by

$$\|\mathbf{P}\|_z = \max_{i=1,..,n} \left[\sum_{j=1}^{n} |p_{ij}| z^{j-i} \right]. \tag{B.6}$$

It satisfies the properties as follows for $n \times n$ matrices $\mathbf{P}$ and $\mathbf{Q}$ and scalar λ.
(B.2) (this is obvious)
(B.3)

$$\begin{aligned} \|\lambda\mathbf{P}\|_z &= \max_{i=1,..,n} \left[\sum_{j=1}^{n} |\lambda p_{ij}| z^{j-i} \right] \\ &= |\lambda| \max_{i=1,..,n} \left[\sum_{j=1}^{n} |p_{ij}| z^{j-i} \right] \end{aligned}$$

(B.4)

$$\begin{aligned} \|\mathbf{P}+\mathbf{Q}\|_z &= \max_{i=1,..,n} \left[\sum_{j=1}^{n} |p_{ij}+q_{ij}| z^{j-i} \right] \\ &\leq \max_{i=1,..,n} \left[\sum_{j=1}^{n} |p_{ij}| z^{j-i} \right] + \max_{i=1,..,n} \left[\sum_{j=1}^{n} |q_{ij}| z^{j-i} \right] \end{aligned}$$

(B.5)

$$\begin{aligned} \|\mathbf{PQ}\|_z &= \max_{i=1,..,n} \left[\sum_{j=1}^{n} \left| \sum_{k=1}^{n} p_{ik} q_{kj} \right| z^{j-i} \right] \\ &\leq \max_{i=1,..,n} \left[\sum_{k=1}^{n} |p_{ik}| \left(\sum_{j=1}^{n} |q_{kj}| z^{j-k} \right) z^{k-i} \right] \\ &\leq \|\mathbf{Q}\|_z \max_{i=1,..,n} \left[\sum_{k=1}^{n} |p_{ik}| z^{k-i} \right]. \end{aligned}$$

In a similar manner it can be seen that the vector norm

$$\|\mathbf{x}\|_z = \sum_{i=1}^{n} |x_i| z^{i-1}, \tag{B.7}$$

satisfies properties (B.2), (B.3) and (B.4) for $z \in (0,1)$ and

$$\begin{aligned} \|\mathbf{xA}\|_z &= \sum_{i=1}^{n} \left| \sum_{k=1}^{n} x_k a_{ki} \right| z^{i-1} \\ &\leq \sum_{k=1}^{n} \sum_{i=1}^{n} |x_k| |a_{ki}| z^{i-1} \\ &\leq \sum_{k=1}^{n} |x_k| z^{k-1} \left(\sum_{i=1}^{n} |a_{ki}| z^{i-k} \right) \\ &\leq \sum_{k=1}^{n} |x_k| z^{k-1} \|\mathbf{A}\|_z, \end{aligned}$$

hence the norms are consistent.

B.3 Non-negative matrices

The following sections are from Gantmacher (1959). A matrix with real elements is called *non-negative* if and only if all of its elements are ≥ 0. We write this as $\mathbf{A} \geq \mathbf{0}$. A square matrix $A = \{a_{ij}\}$ is called *reducible* if the index set $1, 2, \ldots, n$ can be split into two complimentary sets $i_1, \ldots, i_\mu; k_1, \ldots, k_\nu$ $(\mu + \nu = n)$ such that

$$a_{i_\alpha k_\beta} = 0 \quad (\alpha = 1, \ldots, \mu;\ \beta = 1, \ldots, \nu).$$

otherwise the matrix is called *irreducible*.

This may also be expressed as follows. A matrix $\mathbf{A}$ is called reducible if there is a permutation (of the rows and columns) that puts it in the form

$$\mathbf{A}' = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{B}$ and $\mathbf{D}$ are square matrices. Otherwise A is called irreducible.

We shall for the most part only be concerned with irreducible matrices. It should be noted that most results can be generalised in some manner to reducible matrices. A theorem of Frobenius is used implicitly a great deal. The main result in this context is that the spectral radius of a non-negative matrix is one of the simple eigenvalues of that matrix, and that this eigenvalue has a corresponding eigenvector that is positive. We now state the most useful theorem of this section

Theorem B.2 *For any matrix* $\mathbf{P}$ *such that* $\rho(\mathbf{P}) < 1$

$$\sum_{i=0}^{\infty} \mathbf{P}^i = (\mathbf{I} - \mathbf{P})^{-1},$$

and if $\rho(\mathbf{P}) \geq 1$ *then the series does not converge.*

B.4 Matrices and probability

If we consider a system with transitions between states that occur at a countable set of times we may model this system by a series of random variables (X_n), $n \in I\!N$ where X_0 is the initial state of the system and X_n, for $n > 0$, describes the state of the system immediately after the nth transition. We take X_n to have values in Ω, where Ω is the state space of the process.

We can then make the further assumption that the process has the Markov property (or memoryless property). This is essentially

$$p\{X_{n+1} = x | X_0, \ldots, X_n\} = p\{X_{n+1} = x | X_n\},$$

for all $x \in \Omega$. Also we assume the process is homogeneous, that is

$$p\{X_{n+1} = x | X_n\} = p\{X_{n+2} = x | X_{n+1}\},$$

for all $n > 0$. The theorem of total probability (Takács, pg 230) essentially implies that for a system with a finite state space ($\Omega = \{1, \ldots, k\}$) the above equation reduces to the matrix equation.

$$\mathbf{p}_{n+1} = \mathbf{p}_n \mathbf{P}.$$

where $\mathbf{p}_n = \Big(p\{X_n = 1\}, p\{X_n = 2\}, \ldots, p\{X_n = k\}\Big)$ and $\mathbf{P} = \{p_{ij}\}$ where

$$p_{ij} = p\{X_{n+1} = j | X_n = i\}.$$

By extending this we arrive at

$$\mathbf{p}_n = \mathbf{p}_0 \mathbf{P}^n,$$

where $\mathbf{p}_0$ is the vector of initial probabilities of the system. We call $\mathbf{P}$ a probability transition matrix or stochastic matrix. Clearly its elements are all non-negative and its rows must each sum to 1. Hence a stochastic matrix will always have eigenvalue 1 with corresponding right-eigenvector $\mathbf{1}$. Also if we define the matrix norm

$$\|\mathbf{A}\|_r = \max_{i=1,..,n} \sum_{j=1}^{n} |a_{ij}|$$

we can see that $\|\mathbf{P}\|_r = 1$ for any stochastic matrix. Hence as $\rho(\mathbf{P}) \leq \|\mathbf{P}\|_r = 1$ and there is an eigenvalue at 1 we can see that $\rho(\mathbf{P}) = 1$.

If we consider the behaviour of the system on a subset $B = \{i_1, i_2, \ldots, i_\nu\},\ \nu < n$, of the state space we get

$$\mathbf{q}_n = \mathbf{q}_0 \mathbf{Q}^n,$$

where $\mathbf{q}_n = \Big(p\{X_n = i_1\}, p\{X_n = i_2\}, \ldots, p\{X_n = i_\nu\}\Big)$ and $\mathbf{Q}$ is the appropriate transition matrix. $\mathbf{Q}$ is now a sub-stochastic matrix. Its rows must each sum to less than 1. Clearly, using the same norm we can see that $\rho(\mathbf{Q}) < 1$.

Of interest in such systems is often equilibrium behaviour. This can be characterised by $\lim_{n\to\infty} \mathbf{P}^n$.

For our purposes later in this thesis we will be interested in the total time spent in a subset of the state space before leaving that subset. This may be written

$$\sum_{i=0}^{\infty} \mathbf{q}_n = \mathbf{q}_0 \left(\sum_{i=0}^{\infty} \mathbf{Q}^i \right).$$

Thus we need conditions for this to converge, and then determine what it converges to. We know $\rho(\mathbf{Q}) < 1$ and so we can see from Theorem B.2 that this converges and that it converges to $\mathbf{q}_0(\mathbf{I} - \mathbf{P})^{-1}$.

Appendix C

Queueing theory

One of the major uses for stochastic processes is in the study of queues. In this chapter we describe some of the basic queueing theory which is necessary for the work herein. In fact this is the principal use of this thesis. Parts of this section are drawn from Takács (1962), Cohen (1969), Wolff (1982), Kleinrock (1975), Neuts (1989). Again it is not intended to be anything like a complete survey. We shall concentrate here on the M/G/1 queue. This has been one of the most studied queues and consequently there are no shortage of references to it and many of its variants. We shall not attempt to document these. We devote our time to some standard techniques in order to lay the groundwork for the results of this thesis. Both in terms of supplying solutions to some problems and providing the basic models used throughout.

C.1 The Poisson process

Consider a counting process $A(t) : t \geq 0$ on $\mathbb{Z}^+$. We take $A(0) = 0$ and say that A can only increase by 1, at any particular time. If $A(t)$ increases at the times τ_i, $i > 0$, (with $\tau_0 = 0$) then

$$A(\tau_i) = i.$$

For technical reasons we desire A to be right-continuous and have left limits. We define $\theta_n = \tau_n - \tau_{n-1}$ for $n > 0$ to be a series of independent, identically distributed (i.i.d.) random variable with p.d.f. $F(x)$. If $F(x)$ is given by

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

then we say that $A(t)$ is a homogeneous Poisson process with rate λ. (We can define this more generally, see Brèmaud (1981), but this is unnecessary for our purposes). This is a very commonly used process. For our purposes we shall use it as an arrival stream to the queue. It is useful to note that if we remove the time epochs τ_n with probability p from the Poisson process of rate λ we are left with a Poisson process of rate λp. This is particularly useful when we consider a Poisson arrival process in which an arrival is blocked with probability p.

C.2 Queues

A queue, in general, is a stochastic process on the integers, which gives the number of customers in the system at a given time. The notion of a customer includes such concepts as a computer job. The process, which can be characterised by a series of arrivals and departures, usually has the following characteristics. Arrivals increase the number of customers in the system by one. (In some cases arrivals may come in batches which increase the number of customers by more than one.) Departures decrease the system size by one, except for batch departures. The most common reason for a departure is that a customer has received service although customers may depart for reasons such as not enough space in the queue. A general form of Notation can be used to describe a large number of queues. This is the Kendall notation (Kendall, 1951) which is as follows.

Arrival Process/ Service Times/ Number of servers/ Max no. of customers.

Where in the first two places we write , M for Poisson arrivals or Negative exponential services, G for generally distributed arrivals or services and D for deterministic arrivals or services. This is just the simplest form of this notation but it is all we use herein. Note that where a number is omitted it is assumed to be infinity. Some examples are

- GI/M/s - General independent arrivals, Negative exponential service times, s servers.
- M/G/1/n - Poisson arrivals, General service times, 1 server, maximum number of customers in the system n.

The system of primary interest to us here is the M/G/1 queue and its variants.

C.2.1 Queue discipline

The queueing discipline is the way in which customers get chosen for service from the queue. Possibilities include First-in First-out, Last-in First-out, and Random Order. We shall assume throughout that the queueing discipline is non-preemptive. That is, once a customer is in service it remains there until it is finished. Other customers do not *pre-empt* it.

Essential in the idea of waiting times is the concept of non-scheduled service disciplines. Basically these are disciplines in which the order of service does not depend on the amount of service each customer requires. Thus when it gets to the server each customers service time is taken as a random variable with whose service time is independent of the other customers in the queue. All of the above are of this type. An example of a service discipline that is scheduled is one in which the customer with the smallest service time in the queue is served first.

When considering queue lengths, the choice of discipline is irrelevant except for the proviso that it be non-scheduled and non-preemptive, which is necessary for the analysis of the embedded process.

C.3 The embedded process

We use one of the most common approaches to this type of problem which consists of considering the process at departures epochs. We say that we observe the system size distribution that the customers see on departure. The equilibrium number of customers

seen on departures in the stationary distribution can be seen by the following argument to be the same as the equilibrium number of customers in the system.

A very useful rule called PASTA, Poisson arrivals see time averages, can be used to see that the distribution seen by arrivals is the same as the stationary distribution of the queue. (A fact that is not true for general arrivals.) Note that a very neat proof of PASTA that uses martingales appears in Wolff (1989).

In Cooper (1972) pg 154 the following theorem appears.

Theorem C.1 *$\chi(t)$ is a stochastic process whose sample functions are (almost all) step functions with unit jumps. Let the points of increase after some time $t = 0$ be labelled consecutively t_α, and points of decrease t'_α, $\alpha = 0, 1, 2, \ldots$. Let $\chi(t_\alpha+)$ be denoted by ξ_α and $\chi(t'_\alpha-)$ be denoted by ζ_α. Then if either $\lim_{n\to\infty} P\{\xi_n \leq k\}$ or $\lim_{n\to\infty} P\{\zeta_n \leq k\}$ exists, then so does the other and they are equal.*

Because of PASTA and Theorem C.1 considering the queue immediately after departures is quite valid. By doing this we reduce the problem to the consideration of a discrete-time stochastic process. This process which we call (X_n) is in the case of the M/G/1 queue a Markov process. In the cases we consider it is not in general a Markov process although a Markov process can be constructed from it by adding 'supplementary' variables. That is by considering a new process (X_n, Y_n) where Y_n adds some information about the history of the process to the state. This is the normal approach in such modifications to the M/G/1 queue and results usually in the matrix geometric techniques of Neuts.

C.4 Useful theorems

There a number of useful theorems which we draw upon in this thesis.

Theorem C.2 (Little's Law) *For any queueing process the following relationship holds*

$$L = \lambda W,$$

where L is the mean number of customers in the system, λ is the arrival rate to the system and W is the mean waiting time of a customer.

Proof: Little (1961).

Theorem C.3 *If $a(z)$ is the probability generating function for the number of Poisson events of rate λ during a positive time interval with probability distribution function $F(x)$ then*

$$a(z) = F^*(\lambda - \lambda z),$$

where F^* is the Laplace-Stieltjes transform of F.

Proof:

$$a(z) = \sum_{i=0}^{\infty} z^i \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^i}{i!} dA(t)$$

$$
\begin{aligned}
&= \int_0^\infty \sum_{i=0}^{\infty} \frac{e^{-\lambda t}(\lambda t z)^i}{i!} dA(t) \\
&= \int_0^\infty e^{-\lambda t} \sum_{i=0}^{\infty} \frac{(\lambda t z)^i}{i!} dA(t) \\
&= \int_0^\infty e^{-\lambda t} e^{\lambda t z} dA(t) \\
&= \int_0^\infty e^{\lambda t (z-1)} dA(t).
\end{aligned}
$$

□

We use Lemma 1 from Takács (1962) page 47 at one point in this thesis. For brevity we do not include this here. We shall however include a direct result of this lemma.

Theorem C.4 *If $a(z)$ is as in theorem C.3 then $a(z) > z$ for all $z \in [0,1)$ if and only if $a'(1) \leq 1$.*

Proof: from Takács (1962) page 47. □
This could also be derived from convexity arguments.

C.5 Waiting times

In many applications it is useful to know the time spent by a customer before it receives service. We want to calculate the waiting-time distribution for the system. To do this we must specify the service discipline. For the First-in, First-out or order of arrival services there is a neat way of doing this. If we call this waiting-time distribution function $W(\cdot)$, then its Laplace-Stieltjes transform can be written as

$$W^*(s) = \frac{S^*(s)}{A^*(s)},$$

where $A(\cdot)$ is the service-time distribution and $S(\cdot)$ is the sojourn time distribution. (The sojourn time is the time spent by a customer in the system, clearly the sum of the waiting time and the service time.) From Theorem C.3 we can see that the probability generating function of the number of arrivals during the sojourn time of a customer, $s(z)$, is given by

$$s(z) = S^*(\lambda(1-z)).$$

For an order of arrival service discipline it is easy to see that the number of customers left in the queue by a customer who is departing is the number of customers that arrive during the customer sojourn time. This means that in $s(z) = g(z)$ where $g(z)$ is the p.g.f. for the number on the system in equilibrium. For the M/G/1 queue this is

$$g(z) = (1-\rho)\frac{A^*(\lambda(1-z))(1-z)}{A^*(\lambda(1-z)) - z}.$$

Thus we can easily get the Laplace-Stieltjes transform of the waiting time distribution by taking $s = \lambda(1 - z)$,

$$\begin{aligned} W^*(s) &= (1-\rho)\frac{1}{A^*(s)}\frac{A^*(s)\frac{s}{\lambda}}{A^*(s) - 1 + \frac{s}{\lambda}} \\ &= (1-\rho)\frac{\frac{s}{\lambda}}{A^*(s) - 1 + \frac{s}{\lambda}}. \end{aligned}$$

This technique can be used for all single server 'order of arrival processes'. In order to get the actual distribution the Laplace-Stieltjes distribution must be inverted.

This concludes our work on basic theory.

Bibliography

[1] F. Baccelli, (1986), Exponential martingales and Wald's formula for two-queue networks. *Journal of Applied Probability*, **23**, 812–819.

[2] F. Baccelli and A.M. Makowski, (1985), Direct martingale arguments for stability: the M/GI/1 case. *Systems Control Letters*, **6**, 181–186.

[3] F. Baccelli and A.M. Makowski, (1986), Stability and bounds for single server queues in a random environment. *Communications in Statistics-Stochastic Models*, **2** (2), 281–291.

[4] F. Baccelli and A.M. Makowski, (1989), Dynamic, transient and stationary behaviour of the M/GI/1 queue via martingales. *Annals of Probability*, **17** (4), 1691–1699.

[5] F. Baccelli and A.M. Makowski, (1991), Martingale relations for the M/GI/1 queue with Markov modulated Poisson input. *Stochastic Processes and their Applications*, **38**, 99–133.

[6] S. Barnett and C. Storey, (1970), *Matrix Methods in Stability.* Thomas Nelson and Sons Ltd..

[7] P. Brémaud, (1981), *Point Processes and Queues, Martingale Dynamics.* Springer-Verlag.

[8] S.C. Borst, O.J. Boxma and M.B. Combé, (1993), An M/G/1 queue with customer collection. *Communications in Statistics-Stochastic Models*, **9** (3), 341–371.

[9] J.W. Cohen, (1969), *The Single Server Queue.* North-Holland, Amsterdam.

[10] R.B. Cooper, (1972), *Introduction to Queueing Theory.* The Macmillian Company.

[11] P.F. Courtois and J. Georges, (1971), On a single-server finite queueing model with state-dependent arrival and service processes. *Operations Research*, **19**, 424–435.

[12] D.R. Cox, (1962), *Renewal Theory.* Spottiswoode Ballantyne and Co. Ltd.

[13] J. Dshalalow, (1989), Multi-channel queueing systems with infinite waiting rooms and stochastic control. *Journal of Applied Probability*, **26**, 345–362.

[14] J.M. Ferrandiz, (1993), The BMAP/GI/1 queue with server set-up times and server vacations. *Advances in Applied Probability*, **25**, 235–254.

[15] S.W. Fuhrmann and R.B. Cooper, (1985), Stochastic decomposition in the M/G/1 queue with generalised vacations. *Operations Research*, **33** (5), 1117–1129.

[16] F.R. Gantmacher, (1959), *The Theory of Matrices*. Chelsea Publishing Company.

[17] G.H. Golub and C.F. van Loan, (1983), *Matrix Computations*. North Oxford Academic.

[18] W. Gong, A. Yan and C. G. Cassandras, (1992) The M/G/1 queue with queue-length dependent arrival rate. *Communications in Statistics-Stochastic Models*, **8** (4), 733-741.

[19] A.S. Householder, (1964), *The Theory of Matrices in Numerical Analysis*. Blaisdell Publishing Company.

[20] O.C. Ibe and K.S. Trivedi, (1990), Two queues with alternating service and server breakdown. *Queueing Systems*, **7**, 253–268.

[21] O. Kella and W. Whitt, (1992), Useful martingales for stochastic storage processes with Lévy input. *Journal of Applied Probability*, **29**, 396–403.

[22] D.G. Kendall, (1951), Some problems in the theory of queues. *Journal of the Royal Statistical Society, Series B*, 151–185.

[23] M. Kijima and N. Makimoto, (1992a), Computation of the quasi-stationary distributions in M(n)/GI/1/K and GI/M(n)/1/K queues. *Queueing Systems*, **11**, 255–272.

[24] M. Kijima and N. Makimoto, (1992b), A unified approach to GI/M(n)/1/k and M(n)/G/1/K queues via finite quasi-birth-death processes. *Communications in Statistics-Stochastic Models*, **8** (2), 269–288.

[25] L. Kleinrock, (1975), *Queueing Systems*, Volume I: Theory. John Wiley and Sons, Inc..

[26] R.O. LaMaire, (1992), M/G/1/N vacation model with varying E-limited service discipline. *Queueing Systems*, **11**, 357–375.

[27] J.D.C. Little, (1961), A proof of the queueing formula: L = λ W. *Operations Research*, **9** (3), 383–87.

[28] J.A. Morrison, (1990), Two-server queue with one server idle below a threshold. *Queueing systems*, **7**, 325–336.

[29] P.M. Morse, (1967), *Queues, Inventories and Maintenance*. John Wiley and Sons, Inc..

[30] T. Nakagawa and S. Osaki, (1976), Markov renewal processes with some non-regeneration points and their applications to reliability theory. *Microelectronics and Reliability*, **15**, 633–636.

[31] M.F. Neuts, (1989), *Structured stochastic matrices of M/G/1 type and their applications*. Marcel Dekker.

[32] J. Neveu, (1975), *Discrete-Time Martingales*. North-Holland, Amsterdam.

[33] J.H. Park, (1990), The analysis of the $M^X/G/1$ queue by a martingale method, . Master's Thesis, Korea Advanced Institute of Science and Technology.

[34] R. Pyke, (1961a), Markov renewal processes: Definitions and preliminary properties. *Annals of Mathematical Statistics*, **32**, 1231–1242.

[35] R. Pyke, (1961b), Markov renewal processes with finitely many states. *Annals of Mathematical Statistics*, **32**, 1242–1259.

[36] W.A. Rosenkrantz, (1983), Calculation of the Laplace transform of the length of the busy period for the M/G/1 queue via martingales. *Annals of Probability*, **11** (3), 817–818.

[37] W.A. Rosenkrantz, (1989), Ergodicity conditions for two-dimensional Markov chains on the positive quadrant. *Probability Theory and Related Fields*, **83**, 309–319.

[38] M. Roughan, (1993a), Multi-phase discrete-time renewal processes. pre-print.

[39] M. Roughan, (1993b), An analysis of a modified M/G/1 queue using a martingale technique. pre-print.

[40] J.A. Schormans, J.M. Pitts, and E.M. Scharf, (1993), A priority queue with superimposed geometric batch arrivals. *Communications in Statistics-Stochastic Models*, **9** (1), 105–122.

[41] R.F. Serfozo, (1990), Point processes. In *Stochastic Models*, D.P.Heyman and M.J.Sobel, Editors, Volume 2, Chapter 1. North-Holland, Amsterdam.

[42] L. Takács, (1962), *Introduction to the Theory of Queues*. Oxford University Press.

[43] H. Takagi, (1992), Time dependent process of M/G/1 vacation models with exhaustive service. *Journal of Applied Probability*, **29**, 418–429.

[44] T. Takine, H. Takagi and T. Hesegawa, (1993), Analysis of an M/G/1/K/N queue. *Journal of Applied Probability*, **30**, 446–454.

[45] P.D. Welch, (1964), On a generalised M/G/1 queue in which the first customer in each busy period receives exceptional service. *Operations Research*, **12** (5), 736–752.

[46] D. Williams, (1991), *Probability with Martingales*. Cambridge University Press.

[47] R.W. Wolff, (1982), Poisson arrivals see time averages. *Operations Research*, **30**, 223–231.

[48] R.W. Wolff, (1989), *Stochastic Modelling and the Theory of Queues*. Prentice Hall International.

[49] S.F. Yashkov, (1993), On heavy traffic limit theorem for the M/G/1 processor-sharing queue. *Communications in Statistics-Stochastic Models*, **9** (3), 467–471.

[50] G.F. Yeo, (1962), Single server queue with modified service mechanisms. *Journal of the Australian Mathematical Society*, **2**, 499–507.