3-2-2021

# Probability Study of Medical Clinic Scheduling Procedures

Ranveer Kaur
*University of Windsor*

# PROBABILITY STUDY OF MEDICAL CLINIC SCHEDULING PROCEDURES

by

Ranveer Kaur

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2020

Probability Study of Medical Clinic Scheduling Procedures

by

Ranveer Kaur

APPROVED BY:

X. Yuan
School of Computer Science

M. Belalia
Department of Mathematics and Statistics

M. Hlynka, Co-Advisor
Department of Mathematics and Statistics

P.H. Brill, Co-Advisor
Department of Mathematics and Statistics

December 10, 2020

## Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.
I certify, to the best of my knowledge, that my thesis does not infringe upon anyone's copyright or violate any proprietary rights and that any ideas, techniques, quotations, or any other materials from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted materials that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission of the copyright owner(s) to include such materials in my thesis and have included copies of such copyright clearances to my appendix.
I declare that this is a true copy of my thesis, including any final revisions, as approved by committee and the Graduate Studies Office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

# Abstract

In this thesis, we consider the scheduling of patients in a single server medical clinic. We present the probability distribution for the number of patients in the system under certain settings using four different methods. The four methods used are theoretical calculations using convolution, simulation, probability generating functions, and Markov chains. Further, the best scheduling strategy is obtained on the basis of a minimum objective function in the case of fixed interval lengths (for service and interarrival times). Modified simulation annealing is used to aid in finding the best appointment strategy in the case of variable interval lengths.

## Acknowledgements

I would like to express my sincere appreciation for the invaluable supervision and support of Dr. Myron Hlynka during the course of my master's degree. His immense knowledge and plentiful experience have helped me with my academic research. He is an amazing supervisor. Without his guidance, this thesis would have never been accomplished. Next I am grateful to my co-advisor Dr. Percy Brill, for his academic and financial support. Also, special thanks to Dr. Xiaobu Yuan from the School of Computer Science and Dr. Mohamed Belalia from the Department of Mathematics and Statistics for giving their precious time to read my thesis and being a part of my defence committee.

I would also like to show my gratitude to Ms. Dina Labelle for being so helpful and updating me with all the useful information during my whole degree.

I am indebted to all my instructors for always lending a helping hand towards me whenever I needed them during my courses. I am lucky to have such accomplished and supportive people as my professors.

Getting through my thesis required more than academic support, and I have many people to thank for listening to me during the time I spent at the University. I am really grateful to all my friends.

Most importantly, none of this could have happened without both my family and my in-law's family. My grandfather, S. Muktiar Singh always encouraged me via telephone. I must express my very profound gratitude to my parents, sisters and brother for providing me with unfailing support and continuous encouragement throughout my years of study and research. My husband S. Navdeep Singh Matharu greatly supported me in everything during my degree. Thanks to all my family for their constant support.

# Table of Contents


Author's Declaration of Originality iii

Abstract iv

Acknowledgements v

List of Tables vii

List of Figures ix

Chapter 1. Introduction 1
1.1. Environmental Factors 1
1.2. Summary of the thesis 4

Chapter 2. Literature Review 6
2.1. List of Performance measurements in the literature 6

Chapter 3. Probability Distribution 10
3.1. Method 1: Theoretical calculations using convolution 10
3.2. Method 2: Simulation 16
3.3. Method 3: Probability generating function 17
3.4. Method 4: Markov Chain 20

Chapter 4. Appointment Strategies 24
4.1. Parameters 24

Chapter 5. Modified simulation annealing 45

Chapter 6. Conclusions and Future Questions 61
6.1. Conclusions 61
6.2. Future research 61

Bibliography 63

Appendix A 66
Codes and Programs 66

Vita Auctoris 71

# List of Tables


1.1 Clinic environmental factors 2

3.1 Probabilities for numbers of patients at $1d+$ 12

3.2 Probablity table for different numbers of patients in the system at $2d+$ 15

3.3 Observed frequency table at time 1d+ 16

4.1 Summary of Notation 24

4.2 Results for Perfect case 26

4.3 Results for Single Block strategy 27

4.4 Results for single Block strategy with some variablity 29

4.5 Results for individual block with initial block 31

4.6 Results for multiple block strategy 32

4.7 Results for multiple block with initial block strategy 34

4.8 Results for variable block strategy 36

4.9 Results for combination of variable block and individual block strategies for 15 patients 37

4.10 Average medical doctor salary in Canada by province 39

4.11 Results for different strategies after performing simulation 500 times 40

4.12 Results for different strategies after performing simulation 500 times, when physician is 5 mins late 41

5.1 Bad results 46

5.2 Best $a_0$ and $b_0$ for different strategies 47

5.3 Mean waiting time for each patient and mean idle times of physician for 4111 . . . 48

5.4 Mean waiting time for each patient and mean idle times of physician for 2222 . . . 49

5.5 Mean waiting time for each patient and mean idle times of physician for 4222 . . . 51

5.6 Mean waiting time for each patient and mean idle times of physician for 41313 . . . 52
5.7 Mean waiting time for each patient and mean idle times of physician for 412311 . . . 54
5.8 Mean waiting time for each patient and mean idle times of physician for 421311 . . . 55
5.9 Mean waiting time for each patient and mean idle times of physician for 431211 . . . 57
5.10 Objective function i.e total cost for different strategies 59
5.11 Objective function for different strategies, when physician is 5 mins late 60

# List of Figures


2.1 List of performance measurements 7

4.1 Different strategies in the literature 25

4.2 Perfect case with zero waiting and idle times 27

4.3 Single block with constant service times 28

4.4 Single block with variable service times 30

4.5 Individual block with initial block: 4111... 31

4.6 Multiple-block: 2222... 33

4.7 Multiple-block with an initial block: 4222... 35

4.8 Variable block: 41313... 36

4.9 Combination of variable and individual blocks: 412311... 38

4.10 Average salary of Canadian Physician by province 39

4.11 Plots for different probabilities for strategy 4111 42

4.12 Kernel density estimation for strategy 41111 with increasing probabilities 43

5.1 Mean waiting and idle time plot for 4111 ... 48

5.2 Mean waiting and idle time plot for 2222 ... 50

5.3 Mean waiting and idle time plot for 4222 ... 51

5.4 Mean waiting and idle time plot for 41313 ... 53

5.5 Mean waiting and idle time plot for 412311 ... 54

5.6 Mean waiting and idle time plot for 421311 ... 56

5.7 Mean waiting and idle time plot for 431211 ... 57

5.8 Mean waiting time plot for all strategies 58

5.9 Plot of mean waiting time for each patient for all strategies, when physician is late 59

CHAPTER 1

# Introduction

Outpatient clinics/departments are the main medical service providers for non-emergency patients. These clinics have a common problem of long waiting times. This is the primary complaint of patients in their experiences of visiting outpatient clinics. This topic has been of great interest for many years, with publications starting with the pioneering works of Bailey [2] and Lindley [19]. In our review of the literature, waiting times were studied but there did not appear to be a study on the distribution of the number of patients in the system. So, our research considered this topic.

The number of patients in the system is defined as the number of patients waiting in the queue plus the number of patients in the service. By knowing the probability distribution for the number of patients in the system, we can calculate the expected number of patients in the system at a particular time and appoint more patients accordingly. It is a kind of congestion measure, given in the literature. It could be useful for deciding appointments.

The work presented in this thesis is significant for the following reasons. We present new ways of computing the distribution of the number of patients at any given point in time. This allows us to better understand the degree of congestion and therefore make improvements to the system. The ultimate goal is to reduce the waiting time of the patients and the idle time of the physician. Our simulation of different strategies indicates what strategies are best and by how much.

A search for articles on appointment scheduling for outpatient clinics was done using the databanks of Google Scholar, Science Direct, Research Gate, using key-words such as outpatient scheduling, appointment systems, improving patient waiting time and doctor idle time, appointment scheduling in health care and many more. The articles found were evaluated on their relevance to the topics of this thesis.

## 1.1. Environmental Factors

Sometimes, scheduling techniques for medical clinics are not appropriate. The decisions to be made while designing the appointment system of a medical clinic are influenced by several factors, specific to the clinical environment for which the system is designed. Problems arise due to factors such as the physician being late, the physician being interrupted (phone calls, etc.), patients canceling or failing to arrive to see the doctor on the day of the appointment, emergency cases, patient unpunctuality (patients arriving earlier or later than the scheduled appointment time), etc. These factors are referred to as environmental factors. All these factors are summed up in Table 1.1:

TABLE 1.1. Clinic environmental factors

| |
|---|
| 1. Number of services |
| 2. Number of physicians |
| 3. Service times |
| 4. The process of Arrivals |
| 4.1 Unpunctuality of patients |
| 4.2 Presence of no-shows |
| 4.3 Presence of walk-ins |
| 4.4 Presence of companions |
| 5. Lateness and interruption (phone calls) of physician |

**1.1.1. Number of services.** On the basis of the number of services, we can divide the system into two parts:
1. Single-stage system: where the patients' queue for a single service.
2. Multi-stage system: where the patients' queue for multiple services such as registration, pre-examination, X-ray, etc as discussed in [6] (Cayirli et al.).
Most studies in the literature model a single-stage system.

**1.1.2. Number of Physicians.** The two types of system are based on the number of physicians:
1. Single-server system: a system where a single physician serves the patients.
2. Multi-server system: a system where more than one physician serve the patients.
From the literature, most studies have focused on single-server systems. The system with multiple physicians becomes more complicated.

**1.1.3. Service times.** The time for which a patient is in contact with the physician for consultation is referred to as the service or consultation time. As mentioned by Cayirli et al. ([9]), the majority of studies in the literature use independently and identically distributed (i.i.d) service times.

### 1.1.4. The arrival process.

1.1.4.1. *Unpunctuality of patients.* Patient unpunctuality is common in outpatient clinics. Some patients are late because they expect to wait a long time to get service. This may result in an increase in physician idle time and extend waiting times for other patients. This may also prolong the day's end time of the clinic because physicians and patients may not want to reschedule for another day.

Some patients arrive early because they hope to have their doctor consultation as soon as possible. Their usual understanding (often incorrect) is that the earlier they arrive, the sooner they finish. Early patients also create problems as they contribute to congestion in the waiting room which can lead to patient dissatisfaction and staff morale issues (Rohleder et al., [24]). In Leiba et al. [18], it is emphasized that accurate scheduling and low clinic waiting time are important factors that affect the satisfaction of young soldiers needing medical service.

1.1.4.2. *Presence of no-shows.* Patient no-shows are found to be a big problem in many health care settings, where no-show rates can vary from 3% to 80% (Rust et al.[26]). In Sharp and Hamilton ([30]) a 12% no-show rate was reported at outpatient clinics in the United Kingdom. Lack of transportation, scheduling problems, oversleeping or forgetfulness, and lack of child care are some reasons for no-shows (Campbell et al. [5]). Low-quality service either in terms of long wait times or inconvenient appointment systems causes frustration among patients. They feel that scheduling techniques should be designed from a customers' perspective.

1.1.4.3. *Presence of walk-ins.* Patients without any appointment fall into the "walk-in" category. They can be regular or emergency patients. To model walk-ins, Rising et al. ([22]) used an exponential distribution for inter-arrival times, with the mean value changed on an hourly basis to reflect a seasonal pattern.

1.1.4.4. *Presence of companions.* "Companions are those who accompany a patient to the clinic(e.g., a patient's child, husband etc.)" according to Cayirli et al. ([9]). The companions use the waiting room and thus one should consider this factor while determining the appropriate size of a waiting room (Swisher et al. ([32]).

**1.1.5. Lateness and interruption of physician.** Two factors are related to the physician. One is physician unpunctuality and another is interruption level (also called gap times: (Cayirli et al, [9]). Patient waiting times are highly dependent on these factors. Physician unpunctuality is referred to as the lateness for the first appointment.

Apart from this, due to the variation in patients' arrival times and patients' service times as mentioned in Noon et al. ([21]), patients have to wait even though they have reserved appointment slots. This issue has a negative impact on patient satisfaction in what they experience from healthcare facilities. Often the service time is only a few minutes, but patients have to wait for hours. The negative effects of waiting on patients and the methods for making waiting more tolerable are discussed in Katz et al. ([14]).

A well-designed appointment system can help in reducing waiting times for patients as well as increasing the utilization of expensive personnel and equipment based medical resources. Healthcare organizations can reduce waiting times by adopting several strategies to make it easier for patients to access healthcare services. It should be kept in mind that under perfect conditions (deterministic physician service times, patients arriving precisely on schedule, no drop-ins, no missing appointments, precise scheduling), the should be zero waiting time for patients. Although the perfect conditions do not happen often, there could and should be better approximations than currently exist.

As the demand for health care services increases, health care providers are faced with challenges that have enhanced the popularity of Operations Research in health care (Brailsford and Vissers, [4]). Some studies lack practical applicability or have a set of restrictive constraints (Cayirli et al, [9], Kuiper et al. [17]). Outpatient clinics are often considered as queueing systems, which include a unique set of conditions that must be considered for appointment scheduling. The main purpose of outpatient scheduling is to design an appointment technique for which a particular measure of performance is optimized in a clinical environment. Most studies assume that patients are scheduled on a first-call, first-appointment basis.

### 1.2. Summary of the thesis

The rest of the thesis is organized as follows. In Chapter 2, we give a literature review of the existing studies on the appointment scheduling problems and the relevant methodologies of this study. Chapter 2 also discusses the environmental factors used in the literature to describe the environment of the clinic. In Chapter 3, we calculate the probabilities of different numbers of patients at different times by theoretical calculations using convolutions of discrete random variables, for a particular multiple-block with an initial block strategy (4222...) assuming the probability of keeping an appointment is 0.85 (as opposed to missing an appointment). After that, we generalize the results for any number of patients at different times. Then we perform a simulation 100000 times and observe that the results are almost the same as those of the theoretical calculations. We also give a recursive relationship between the number of patients in the system and the number of arrivals at different times by the use of probability generating functions. Further, we discuss the last calculation method using Markov chains. We write an R program

for calculating the probabilities.
In chapter 4, we define some recursive relationships between different parameters for calculating waiting times and the idle times for the strategies from the literature. We provide an objective function on the basis of the salaries of the doctor and the patients, which is to be minimized for the best strategy. Then we compare all the strategies on the basis of the objective function. In chapter 5, modified simulated annealing is discussed for finding the best pair of $a$ and $b$ for variable interval lengths and then find the best strategy on the basis of the minimum objective function. Conclusions and the future questions are given in chapter 6.

CHAPTER 2

# Literature Review

Many analytical studies on scheduling propose algorithms or rules based on queueing theory or simulation. Most of the literature focuses on designing scheduling techniques to minimize the expected waiting and idle times. The question of how much waiting time is bearable by patients depends on various factors - for instance, the environment of the waiting area, facilities in the waiting area, etc. Huang ([12]) notes that patients arriving on time can tolerate a waiting time of 37 minutes or less and those who are late for appointments can accept a waiting time of 63 minutes or less.

## 2.1. List of Performance measurements in the literature

Major performance measures used in the literature include:
1. Cost-Based Measures
2. Time-Based Measures
3. Congestion Measures
4. Fairness Measures

These measures of performance are summarized in a table given in (Cayirli et al, [6]).

Scheduling techniques are usually determined by two components:
1) the number of patients assigned to each time block;
2) the time between two successive appointments.

Appointment systems can be divided into
a) Individual appointment systems: where each patient is given a particular appointment time;
b) Block appointment systems: where more than one patient is given the same appointment time;
c) Mixed appointment systems: this is a combination of individual and block appointment systems (Vissers [33]).

TABLE 2

*Performance Measurements Used in the Literature*

1. Cost-Based Measures
   Mean total cost calculated using relevant combinations of:
   1.1 Waiting time of patients
   1.2 Flow time of patients
   1.3 Idle time of doctor(s)
   1.4 Overtime of doctor(s)
2. Time-Based Measures
   2.1 Mean, maximum, and frequency distribution of patients' waiting time
   2.2 Mean, variance, and frequency distribution of doctor's idle time
   2.3 Mean, maximum and standard deviation of doctor's overtime
   2.4 Mean and frequency distribution of patients' flow time
   2.5 Percentage of patients seen within 30-minutes of their appointment time
3. Congestion Measures
   3.1 Mean and frequency distribution of number of patients in the queue
   3.2 Mean and frequency distribution of number of patients in the system
4. Fairness Measures
   4.1 Mean waiting time of patients according to their place in the clinic
   4.2 Variance of waiting times
   4.3 Variance of queue sizes
5. Other
   5.1 Doctor's productivity
   5.2 Mean doctor utilization
   5.3 Delays between requests and appointments
   5.4 Percentage of urgent patients served
   5.5 Likelihood of patients receiving the slots they requested
   5.6 Clinic effectiveness

FIGURE 2.1. List of performance measurements

The first appointment rule was purposed by Bailey ([2]). It schedules two patients at the start of the session and the rest individually at fixed intervals. This rule is usually known as "Bailey's Rule." This rule was restricted to the case when patients are punctual. Later research studies introduced multiple block rules (White et al. [3]) and also variable-block rules with fixed intervals (Rising et al. [22]).

Some studies have designed appointment rules with variable intervals. Ho and Lau ([11]) designed an appointment rule that allows patients to arrive with shorter interarrival times in the earlier part of the clinic session and larger interarrival times later on. These rules are further enhanced in Cayirli et al. ([9]) by including no-shows and walk-ins. Samoran et al. [27] showed that considering individual no-show predictions may significantly improve the performance of a schedule by strategically scheduling expected no-shows.

Vissers illustrated the use of the results in [33] with two examples. In one example, the number of appointments was 20 with a mean consultation time of five minutes and in another example, the number of appointments was 24 with a mean consultation time of 6.5 minutes. Blanco White and Pike [3] recommended an appointment scheme for unpunctual patients comprised of appointment-intervals of one-tenth of the average consultation-length, a fixed number of patients at each appointment-time with rate equal to the service rate, and physician time less than 5 min after the first appointment.

According to Ho and Lau ([11]), of three environmental factors (the probability of no-show, the coefficient of variation of service times, and the number of patients per clinical session) the probability of no-show is the dominant one that affects performance and the choice of an appointment system. In the literature, the case of no-shows is studied using no-show probabilities ($p$) that range from 5 to 30 percent. As per expectations, studies find that the doctor's idle time increases and the waiting time of the patients decrease with an increasing probability of no-shows.

Studies such as Wang et al. [35], Robinson et al. [23], Kaandorp et al. [13], Kuiper et al. [17] have used a "dome rule" in which the appointment intervals are of shorter length at the start and end of a session and increase in length towards the middle of a session to obtain optimal scheduling. Soriano [31] proposed a Two-at-a-Time appointment system with interval length equal to twice the average service time.

Unpunctuality makes the analysis of appointment scheduling more complicated. From their empirical study, Fetter and Thompson [10] found that approximately 81% of patients arrive early with an average of 17.2 minutes in one hospital clinic and with an average of 18.4 minutes in another. In some studies, theoretical distributions for patient unpunctuality have been considered using Pearson type VII (Blanco White and Pike [3]) and normal distributions (Cayirli et al. [7]).

Some studies have taken into account patient classification in the design of appointment systems. Walter [34] finds that the doctors' idle time in the radiology department is improved using a grouping of inpatients and outpatients. Simulation results by Cayirli et al. [8] note an improvement in doctors' idle time, doctors' overtime and patients' waiting times when appointment systems utilize interval adjustment for patient class.

There are different choices of performance criteria in the literature to evaluate appointment systems. Many studies list results in terms of mean waiting time of patients and mean idle time of physician. Furthermore, team members involve patients, doctors, ... each of which attach varying importance to the different criteria. Some of the studies used cost-based measures in which relative weights in terms of the cost of patients' waiting time ($C_p$) and cost of physician idle time ($C_d$) were assigned. The waiting costs for all patients were assumed to be identical in the literature. Some studies also included the mean overtime of physicians with a special cost for it ($C_o$). Thus the objective was to minimize the expected total cost of the system given by:

$$Min\ E(TC) = E(W)C_p + E(I)C_d + E(O)C_o$$

where $E(W)$ is the mean waiting time of patients
$E(I)$ is the mean idle time of the physician
$E(O)$ is the mean over time of the physician

In order to estimate $C_d$, Keller and Laughhunn [15] divided the annual salary of the doctor by the hours worked per year and used the minimum wage to reflect the cost of the patients' waiting time.

CHAPTER 3

# Probability Distribution

Consider the simple case of multiple-block, fixed-interval scheduling model with a special initial block. Assume that an outpatient clinic has a single server over the course of a session. Assume that each patient has probability $p$ of keeping its appointment. Assume $n_0, n_1, n_2, n_3, \ldots$ patients or customers are scheduled at times 0,d,2d,3d,... respectively (here $d$ is the gap time between consecutive scheduling times). If a patient is available, one patient is served (and leaves the system) at the end of each time interval of length $d$. Let $X_0, X_1, X_2, X_3, \ldots$ be the number of patients remaining in the system at times $0^+, d^+, 2d^+, 3d^+, \ldots$ where $id^+$ refers to time $id + \epsilon$ for small $\epsilon > 0$, $i = 0, 1, 2, \ldots$.

## 3.1. Method 1: Theoretical calculations using convolution

For concreteness, we take $n_0 = 4$ and $n_1 = n_2 = n_3 = \cdots = 2$.

Property 3.1. Given the conditions above,

(1) $X_0 \sim Bin(n = 4, p)$
(2) $X_1$ has support on $i = 0, 1, 2, 3, 4, 5$ satisfying
$P(X_1 = i) = Bin(n = 4 + 2, p, i + 1)$ for $i = 5, 4, 3$ where $Bin(n, p, i) = \binom{n}{i} p^i (1 - p)^{n-i}$
(3) $X_2$ has a distribution on $i = 0, 1, 2, ..., 6$ satisfying
$P(X_2 = i) = Bin(n = 4 + 2 + 2, p, i + 2)$ for $i = 6, 5, 4$
(4) $X_3$ has a distribution on $i = 0, 1, 2, \ldots, 7$ satisfying
$P(X_3 = i) = Bin(n = 4 + 2 + 2 + 2, p, i + 3)$ for $i = 7, 6, 5$.

(The other probabilities for $X_1, X_2, X_3$ follow different patterns.)

Proof. At time 0, patients either arrive or not (some patients might miss an appointment for some reason). So, the possible number of arrivals is $0, 1, 2, 3, 4$ and clearly $X_0$ is binomially distributed. Thus at time $0^+$, $X_0 \sim Bin(n = 4, p)$.

At time d, we potentially add two more patients and complete service on a patient who arrived at time 0 (if there was such a patient). Because the sum of two independent binomial random variables $Y_1 \sim Bin(n_1, p)$ and $Y_2 \sim Bin(n_2, p)$ is binomial $Bin(n_1 + n_2, p)$, the total number of arrivals by time d will be the sum of two values, with a maximum of $4 + 2 = 6$. But we have potentially decreased by one customer because of service completion (if there was a customer at time 0 to service). So $X_1$ can take integer values from 0 to 4+2-1=5. But it is possible, for $i = 0, 1, 2$ that there were no arrivals at time 0 and hence no customer to serve at that time. It is also possible for there to be 1 arrival at time 0, and for that service to complete at time d. These two types of matching make the probabilities $P(X_1 = i)$ more complex for $i = 0, 1, 2$. For $X_i = 3, 4, 5$, there had to be at least one arrival at time 0. So for $X_i = 3, 4, 5$, the probabilities have to match binomial probabilities $P(X_1 = 3) = Bin(n = 6, p, i = 4)$, $P(X_1 = 4) = Bin(n = 6, p, i = 5)$, $P(X_1 = 5) = Bin(n = 6, p, i = 6)$.

At time 2d, we potentially add two more patients and complete service on a patient who was present at time d (if there was such a patient). Because the sum of three independent binomial random variables with a common probability of success $p$ is again binomial, the total number of arrivals by time 2d will be binomial with a maximum of $4 + 2 + 2 = 8$. But $X_2$ will be reduced by up to two patients because of service completion. So $X_2$ can take integer values from 0 to $4 + 2 + 2 - 2 = 6$. As in the $X_1$ case, if $i = 0, 1, 2, 3$, the computation of $P(X_2 = i)$ is more complex, whereas if $i = 4, 5, 6$, we have probabilities then $P(X_2 = i) = Bin(n = 8, p, i + 2)$.

At time 3d, we potentially add two more patients and complete service on a patient who was present at time 2d (if there was such a patient). Because the sum of four independent binomial random variables with a common probability of success $p$ is again binomial, the total number of arrivals by time 3d will be binomial with a maximum of $4 + 2 + 2 + 2 = 10$. But $X_3$ will be reduced by up to three patients because of service completion. So $X_3$ can take integer values from 0 to $4 + 2 + 2 + 2 - 3 = 7$. As in the $X_2$ case, if $i = 0, 1, 2, 3, 4$, the computation of $P(X_3 = i)$ is more complex, whereas if $i = 5, 6, 7$, we obtain probabilities $P(X_3 = i) = Bin(n = 8, p, i + 3)$. $\square$

We look at the particular case when $p = 0.85$.

(1) The probability of $x$ patients at time $0^+$ is given (using R notation) as $f_0(x) = dbinom(x, 4, 0.85)$ for $x = 0, 1, 2, 3, 4$

(2) At time $1d$, 2 patients have appointments. Therefore, the number of arriving patients at $d$ follows $Bin(2, 0.85)$. The total number of patients at time $d^+$ includes patients from 0 and patients from $d$ [perhaps] reduced by one patient, who completes service at time $d$ unless no patient arrived at 0. Thus $X_1$ has support $\{0, 1, 2, 3, 4, 5\}$.

The probabilities of $x$ patients in the system at $1d+$ are given by

$$\begin{aligned}
f_1(5) &= dbinom(4,4,0.85) * dbinom(2,2,0.85) \\
&= dbinom(6,6,0.85) \\
f_1(4) &= dbinom(4,4,0.85) * dbinom(1,2,0.85) \\
&+ dbinom(3,4,0.85) * dbinom(2,2,0.85) \\
&= dbinom(5,6,0.85) \\
f_1(3) &= dbinom(4,4,0.85) * dbinom(0,2,0.85) \\
&+ dbinom(3,4,0.85) * dbinom(1,2,0.85) \\
&+ dbinom(2,4,0.85) * dbinom(2,2,0.85) \\
&= dbinom(4,6,0.85) \\
f_1(2) &= dbinom(3,4,0.85) * dbinom(0,2,0.85) \\
&+ dbinom(2,4,0.85) * dbinom(1,2,0.85) \\
&+ dbinom(1,4,0.85) * dbinom(2,2,0.85) \\
&+ dbinom(0,4,0.85) * dbinom(2,2,0.85) \\
&= dbinom(3,6,0.85) \\
&+ dbinom(0,4,0.85) * dbinom(2,2,0.85) \\
f_1(1) &= dbinom(2,4,0.85) * dbinom(0,2,0.85) \\
&+ dbinom(1,4,0.85) * dbinom(1,2,0.85) \\
&+ dbinom(0,4,0.85) * dbinom(1,2,0.85) \\
&= dbinom(2,6,0.85) \\
&- dbinom(0,4,0.85) * dbinom(2,2,0.85) \\
&+ dbinom(0,4,0.85) * dbinom(1,2,0.85) \\
f_1(0) &= dbinom(1,4,0.85) * dbinom(0,2,0.85) \\
&+ dbinom(0,4,0.85) * dbinom(0,2,0.85) \\
&= dbinom(1,6,0.85) \\
&+ dbinom(0,6,0.85) \\
&- dbinom(0,4,0.85) * dbinom(1,2,0.85)
\end{aligned}$$

Upon simplifying, we obtain

| x | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Prob. | 0.00027 | 0.00525 | 0.04182 | 0.17618 | 0.39933 | 0.37715 |

TABLE 3.1. Probabilities for numbers of patients at $1d+$

From the above calculations it is clear that $X_1$ follows $Bin(x+1, 6, 0.85)$ for $x = 3, 4, 5$ and for lower values of $x$, probabilities have some additional terms.

(3) Similarly, at time $2d^+$, since we lose [up to 2] patients by time 2d, therefore $X_2$ has support $= \{0, 1, 2, 3, 4, 5, 6\}$.

The probabilities are given by

$$
\begin{aligned}
f_2(6) &= dbinom(4,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
&= dbinom(8,8,0.85)\\
f_2(5) &= dbinom(4,4,0.85) * dbinom(2,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(4,4,0.85) * dbinom(1,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
&= dbinom(7,8,0.85)\\
f_2(4) &= dbinom(4,4,0.85) * dbinom(2,2,0.85) * dbinom(0,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(2,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(1,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(2,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(4,4,0.85) * dbinom(0,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(4,4,0.85) * dbinom(1,2,0.85) * dbinom(1,2,0.85)\\
&= dbinom(6,8,0.85)\\
f_2(3) &= dbinom(4,4,0.85) * dbinom(1,2,0.85) * dbinom(0,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(2,2,0.85) * dbinom(0,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(1,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(2,4,0.85) * dbinom(2,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(2,4,0.85) * dbinom(1,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(1,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(3,4,0.85) * dbinom(0,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(4,4,0.85) * dbinom(0,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
&= dbinom(5,8,0.85)\\
&+ dbinom(0,4,0.85) * dbinom(2,2,0.85) * dbinom(2,2,0.85)\\
f_2(2) &= dbinom(2,4,0.85) * dbinom(2,2,0.85) * dbinom(0,2,0.85)\\
&+ dbinom(1,4,0.85) * dbinom(1,2,0.85) * dbinom(2,2,0.85)\\
&+ dbinom(2,4,0.85) * dbinom(1,2,0.85) * dbinom(1,2,0.85)\\
&+ dbinom(1,4,0.85) * dbinom(2,2,0.85) * dbinom(1,2,0.85)
\end{aligned}
$$

$$
\begin{aligned}
&+ dbinom(3,4,0.85)*dbinom(1,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(3,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(2,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(4,4,0.85)*dbinom(0,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(1,2,0.85)\\
&= dbinom(4,8,0.85)\\
&- dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(1,2,0.85)\\
f_2(1) &= dbinom(2,4,0.85)*dbinom(1,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(2,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(1,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(3,4,0.85)*dbinom(0,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(2,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&= dbinom(3,8,.85)\\
&- dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(2,2,0.85)\\
&- dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(2,2,0.85)\\
&- dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(0,2,0.85)*dbinom(1,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(2,2,0.85)*dbinom(0,2,0.85)\\
f_2(0) &= dbinom(1,4,0.85)*dbinom(1,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(2,4,0.85)*dbinom(0,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(1,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(1,4,0.85)*dbinom(0,2,0.85)*dbinom(0,2,0.85)\\
&+ dbinom(0,4,0.85)*dbinom(0,2,0.85)*dbinom(0,2,0.85)
\end{aligned}
$$

$$
\begin{aligned}
&= dbinom(2,8,0.85)\\
&+ dbinom(1,8,0.85)\\
&+ dbinom(0,8,0.85)\\
&- dbinom(1,4,0.85) * dbinom(0,2,0.85) * dbinom(1,2,0.85)\\
&- dbinom(0,4,0.85) * dbinom(1,2,0.85) * dbinom(1,2,0.85)\\
&- dbinom(0,4,0.85) * dbinom(0,2,0.85) * dbinom(1,2,0.85)\\
&- dbinom(0,4,0.85) * dbinom(2,2,0.85) * dbinom(0,2,0.85)\\
&- dbinom(0,4,0.85) * dbinom(0,2,0.85) * dbinom(2,2,0.85)
\end{aligned}
$$

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Prob. | 0.00012 | 0.00235 | 0.01862 | 0.08412 | 0.23760 | 0.38469 | 0.27249 |

TABLE 3.2. Probablity table for different numbers of patients in the system at $2d+$

From above calculations, $X_2 \sim Bin(x+2, 8, 0.85)$ for $x = 4, 5, 6$ and probabilities of $0, 1, 2, 3$ have some extra terms to $Bin(8, 0.85)$.

As we go further in time, the calculations of probabilities of numbers of patients in the system become even more complicated. We see that some of the probabilities (for high counts) take binomial values, and the probabilities for lower values are binomial probabilities with added correction terms.

PROPERTY 3.2. Generalization of property 3.1. Given the conditions in property 3.1,

(1) $X_0 \sim Bin(n = n_0, p)$.
(2) $X_1$ has support on $i = 0, \ldots, n_0 + n_1 - 1$ satisfying $P(X_1 = i) = Bin(n = n_0 + n_1, p, i + 1)$ for upper $n_0 - 1$ numbers where $Bin(n, p, i) = \binom{n}{i} p^i (1-p)^{n-i}$.
(3) $X_2$ has a distribution on $i = 0, \ldots, n_0 + n_1 + n_2 - 2$ satisfying $P(X_2 = i) = Bin(n = n_0 + n_1 + n_2, p, i + 2)$ for upper $n_0 - 1$ numbers.
(4) $X_3$ has a distribution on $i = 0, \ldots, n_0 + n_1 + n_2 + n_3 - 3$ satisfying $P(X_3 = i) = Bin(n = n_0 + n_1 + n_2 + n_3, p, i + 3)$ for upper $n_0 - 1$ numbers.

Researchers sometimes use models with zero inflated binomial distributions. This is not quite what appears here. Yet we do get a modified binomial type distribution, with larger probabilities at the lower values. This type of distribution could also appear in an inventory type model with regular increases in inventory and regular decreases in inventory.

Because of the complications in calculations, a reasonable procedure is to simulate the model. This can be done with less complication than continuing the exact calculation. So, we now introduce a simulation algorithm to obtain our desired probabilities.

## 3.2. Method 2: Simulation

Here we select $p = 0.85$. Our program is written in R.

```
set.seed(1)
v0 = c(); w1 = c(); w2 = c(); w3 = c();
for (i in 1:100000){
b0 = sum(1*(runif(4)>0.15)); v0 = c(v0,b0);
b1 = sum(1*(runif(2)>0.15)); w1 = c(w1,b1);
b2 = sum(1*(runif(2)>0.15)); w2 = c(w2,b2);
b3 = sum(1*(runif(2)>0.15)); w3 = c(w3,b3)
}
v1 = pmax(v0-1,0)+w1
v2 = pmax(v1-1,0)+w2
v3 = pmax(v2-1,0)+w3
v0; w1; w2;w3
v1; v2; v3
table(v0)
table(v1)
table(v2)
table(v3)
```

We present the output for $table(v1)$ to illustrate.

| x | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Obs. freq | 28 | 508 | 4165 | 17685 | 39997 | 37617 |

TABLE 3.3. Observed frequency table at time 1d+

### 3.2.1. Validation of Simulation.

To validate the above simulation we use the goodness of fit test. From Table 3.3, use observed frequency. From Table 3.1, we obtain true probabilities.

We are testing $H_0$ : good fit to theoretical outcomes
vs $H_1$ : poor fit.
From R, we obtain
obs = c(28,508,4165,17685,39997,37617)
p1=c(0.00027, 0.00525, 0.04182, 0.17618, 0.39933, 0.37715)
expt = 100000*p1
chisq.test(obs, p=p1)

Output:
Chi-squared test for given probabilities
data: obs
X-squared = 1.2686, df = 5, p-value = 0.9381

From the output, $p-value > 0.05$ so we do not reject the null hypothesis which means the observed frequencies are consistent with the expected or theoretical frequencies. Thus the simulation algorithm is [somewhat] validated.

We can present a description of $X_0, X_1, X_2, \ldots$ (number of customers at times $0^+, 1d^+, 2d^+, \ldots$) in a simpler manner as follows.

PROPERTY 3.3. Let $X_0, X_1, X_2, \ldots$ be the number of patients in the system at times $0^+, 1d^+, 2d^+, \ldots$ respectively. Let $Y_0, Y_1, Y_2, \ldots$ be the number of arrivals at times $0, 1d, 2d, \ldots$ respectively and $X^+ = max(X, 0)$. Then $Y_i \sim Bin(n_i, p)$ and

$$\begin{aligned}
X_0 &= Y_0 \\
X_1 &= (X_0 - 1)^+ + Y_1 \\
X_2 &= (X_1 - 1)^+ + Y_2 \\
X_3 &= (X_2 - 1)^+ + Y_3 \\
&\vdots \\
X_n &= (X_{n-1} - 1)^+ + Y_n, \quad n = 1, 2, 3, \ldots
\end{aligned}$$

PROOF. For $n = 1, 2, 3, \ldots$, if there were $X_{n-1}$ patients at time $(n-1)d^+$, then at time $nd$, we complete and subtract one patient (if the system is nonempty), and add $Y_n$ patients leaving $(X_{n-1} - 1)^+ + Y_n$ at time $nd^+$. □

Since we have a recursive relationship, involving sums of random variables, we are working with convolutions of discrete random variables. A common tool in such cases is the use of probability generating functions.

### 3.3. Method 3: Probability generating function

Recall that a probability generating function (pgf) of a discrete random variable $X$ is defined as

$$G(z) = E(z^X) = \sum_{x=0}^{\infty} z^x p(x)$$

where $X$ takes values on the non-negative integers and $p(x) = P(X = x)$ is the probability that the discrete variable X takes the value x. The following result is well known.

THEOREM 3.3.1. *Let $G_X(z)$ and $G_Y(z)$ be the probability generating functions of two independent discrete random variables $X$ and $Y$ respectively. Then the pgf of convolution $W = X + Y$ is the product of pgf of $X$ and pgf of $Y$ i.e $G_W(z) = G_X(z)G_Y(z)$.*

PROOF. Let $X$ and $Y$ be two independent random variables. The pgf of $X+Y$ is given by

$$
\begin{aligned}
G_{X+Y}(s) &= E(s^{X+Y}) \\
&= E(s^X s^Y) \\
&= E(s^X)E(s^Y) \quad \text{(Since X and Y are independent)} \\
&= G_X(s)G_Y(s) \\
\text{Hence,} \quad G_{X+Y}(s) &= G_X(s)G_Y(s)
\end{aligned}
$$

□

PROPERTY 3.4. Let the vectors $x_0, x_1, x_2, x_3 \ldots$ be the probabilities of the numbers of patients at times $0^+, 1d^+, 2d^+, 3d^+, \ldots$ and $y_0, y_1, y_2, y_3, \ldots$ be the vectors of the binomial probabilities of new arrivals at times $0, 1d, 2d, 3d, \ldots$. Let $u_0, u_1, u_2, u_3, \ldots$ be probability vectors created from $x_0, x_1, x_2, x_3, \ldots$ by reducing the vector length by 1 through summing the first 2 components. Then the pgf's of $X_i$ (for $i = 1, 2, 3, \ldots$) (numbers of remaining patients at time $id^+$) are obtained by multiplying pgf of $U_{i-1}$ (number of patients before new arrivals but after a service completion at time $id$) with the pgf of $Y_i$ (number of arrivals at time $id$).

PROOF. The vectors $u_i$ $(i = 0, 1, 2, 3, \ldots)$ essentially correspond to the random variable $(X_i - 1)^+$ so the algorithm implements property 3.2. □

The R code implementation follows. We again use $p = 0.85$ for our example.

```
library(polynom)
# probability vector number of arrivals at time 0
y_0=dbinom(0:4,4,0.85)
# probability vector at time 0+
x_0=y_0
x_0
# probability vector generated from x_0
u_0=c(x_0[1]+x_0[2],x_0[-c(1:2)])
```

```
u_0
# probability vector at time d
y_1=dbinom(0:2,2,0.85)
y_1
# pgf of Y_1
pgf_y_1 = polynomial(y_1)
pgf_y_1
# pgf of U_0
pgf_u_0 = polynomial(u_0)
pgf_u_0
# pgf of X_1
pgf_x_1 = pgf_y_1*pgf_u_0
pgf_x_1
# probability vector at time d+
x_1 = coef(pgf_x_1)
x_1

# probability vector generated from x_1
u_1=c(x_1[1]+x_1[2],x_1[-c(1,2)])
# probability vector of number of arrivals at time d
y_2=dbinom(0:2,2,0.85)
# pgf of Y_2
pgf_y_2 = polynomial(y_2)
# pgf of U_1
pgf_u_1 = polynomial(u_1)
# pgf of X_2
pgf_x_2 = pgf_y_2*pgf_u_1
# probability vector at time 2d+
x_2 = coef(pgf_x_2)
```

Typical output shows

```
> x_0
[1] 0.00050625 0.01147500 0.09753750 0.36847500 0.52200625
> u_0
[1] 0.01198125 0.09753750 0.36847500 0.52200625
> y_1
[1] 0.0225 0.2550 0.7225
> pgf_u_0
0.01198125 + 0.0975375*x + 0.368475*x^2 + 0.5220062*x^3
> pgf_y_1
0.0225 + 0.255*x + 0.7225*x^2
```

```
> pgf_x_1
0.0002695781 + 0.005249813*x + 0.0418192*x^2 + 0.1761771*x^3 + 0.3993348*x^4 + 0.3771495*x^5
> x_1
[1] 0.0002695781 0.0052498125 0.0418192031 0.1761771094 0.3993347812 0.3771495156
```

An examination of pgf_x_1 shows $\sum_{i=0}^{5} P(X_1 = i)x^i$ so we can read off the probabilities $P(X_1 = i)$ as the coefficients given by x_1 in the output. It is simple to extend this program to find probabilities for $X_1, X_2, X_3, X_4, ....$ Compare this to the complicated structure that was presented earlier when all cases were done by hand. We note that this algorithm allows us to easily compute the probabilities of $x$ patients at time $id^+$ for any $i$. From these probabilities, we can find all moments of $X_i$.

### 3.4. Method 4: Markov Chain

As in property 3.3, let $X_0, X_1, X_2, \ldots$ be the number of patients in the system at times $0^+, 1d^+, 2d^+, \ldots$ respectively where $X_0, X_1, X_2, \ldots$ are clearly dependent. Further, let $Y_0, Y_1, Y_2, \ldots$ be the number of arrivals at times $0, 1d, 2d, \ldots$ respectively and these are independent of each other and of the lower indexed $X$'s. To find the probabilities of the number of patients at times $0^+, 1d^+, 2d^+, \ldots,$ in another way than used in Chapter 3, we use Markov chains. Seneta ([28]) states that Markov chains were introduced by Markov in 1906 [20]. A stochastic or random process is a Markov Chain if it satisfies the Markov property. If the Markov property/assumption (also called the memoryless property) holds, then the study of such a random process is easier.

**3.4.1. Markov Property:** Informally, the Markov property assumes

$$P(future|present, past) = P(future|present)$$

Let $X = \{X_n : n \geqslant 0\}$ be a random process on a countable set S. For any $i, j \in S$ and $n \geqslant 0$, the Markov property states

$$P\{X_{n+1} = j|X_0, \ldots, X_n\} = P\{X_{n+1} = j|X_n\}$$

It means at any time $n$, the conditional distribution of the future state $X_{n+1}$ given present and past states i.e $X_0, \ldots, X_n$ depends only on the present state $(X_n)$. Given the information in $X_n$, the information in the past states $(X_0, \ldots, X_{n-1})$ is no longer needed for computation of future probabilities. In other words, the future state $X_{n+1}$ is conditionally independent of the past $X_0, \ldots, X_{n-1}$ given the present state $X_n$. Define

$$P\{X_{n+1} = j|X_n = i\} = p_{ij}$$

where $p_{ij}$ is called the transition probability.

**3.4.2. Transition probability:** The transition probability $p_{ij}$ is the probability that the Markov chain moves from state $i$ to state $j$ in one step. The probability distribution of state transitions is typically represented as a Markov chain transition matrix i.e $P = (p_{ij})$. The transition probabilities satisfy $\sum_{j\in S} p_{ij} = 1, i \in S$. If the Markov chain has $m$ possible states, the transition matrix will be an $m \times m$ matrix, such that entry $(i, j)$ is the probability of transitioning from state $i$ to $j$.

**3.4.3. Random Walk:** We refer the book by Serfozo ([29]) for a discussion on random walks. Suppose that $W_1, W_2, \ldots$ are independent integer-valued random variables, and

$$X_t = X_0 + \sum_{m=1}^{t} W_m, \quad t \geqslant 1$$

The process $X_t$ is a random walk on the set of integers $S$, where $W_m$ is the step size at step $m$. A random walk represents a quantity that changes over time such that its increments (step sizes) are independent. Since $X_{t+1} = X_t + W_{t+1}$, then for any $i, j \in S$ and $t \geqslant 0$,

$$P\{X_{t+1} = j|X_1, \ldots, X_{t-1}, X_t = i\}$$

$$= P\{X_t + W_{t+1} = j|X_t = i\} = P\{W_{t+1} = j - i\}.$$

Therefore, the random walk $X_t$ is a Markov chain on the nonnegative integers $S$ with transition probabilities $p_{ij}$.

In our case, $X_0 = i$ with probability $P(X_0 = i) = Bin(n = 4, i, p = 0.85)$ and $Y_1, Y_2, \ldots$ are the number of arrivals at time $1d, 2d, \ldots$ respectively where $Y_{t+1} \sim Bin(2, 0.85)$ for $t = 0, 1, 2, 3, \ldots$. Further, $X_0, X_1, \ldots$ are the number of patients in the system at times $0^+, 1d^+, \ldots$ respectively. Then $X_{t+1} = X_t + W_{t+1}$ where :

$$W_{t+1} = \begin{cases} Y_{t+1} & \text{if } X_t = 0 \\ Y_{t+1} - 1 & \text{if } X_t \geqslant 1 \end{cases} \tag{3.1}$$

We generate a transition matrix P of order $\infty \times \infty$, but truncate it to size $n \times n$ where $n \geqslant 5$ is large enough so that the computations of interest will not be affected by the truncation. The transition probabilities are $p_{ij} = P(X_{n+1} = j|X_n = i)$. We define our initial vector to be

$\pi_0 = (P(X_0 = 0), P(X_0 = 1), P(X_0 = 2), P(X_0 = 3), P(X_0 = 4), 0, 0, 0, ...)$.

Then $\pi_1 = \pi_0 P$ will give the probabilities for $X_1$. Next $\pi_2 = \pi_0 P^2$ will give probabilities for $X_2$, etc. For computational purposes, we can truncate $\pi_0$, $\pi_1$, ..., and $P$ to be $n \times n$, where we choose $n$ larger as the time level increases, since the number of possible states that $X_i$ takes will increase with the step number.

```
# Initial State probabilities
p= function(n){
c(dbinom(0:4,4,0.85),rep(0,n-5))
}

# Transition probability matrix
P=function(n){
A=matrix(c(0.2550,0.0225,rep(0,n-2),0.7225),n,n);
A[1,1:3]=c(0.0225,0.2550,0.7225);A
}
P(6)
P(7)

# Probabilities at time d+
p_1=p(6)%*%P(6)
p_1

# Probabilities at time 2d+
p_2=p(7)%*%P(7)%*%P(7)
p_2
```

The truncated version (of length 6) of the initial probability distribution vector $\pi_0$ is given by:

$$\pi_0 = \begin{bmatrix} 0.00050625 & 0.01147500 & 0.09753750 & 0.36847500 & 0.52200625 & 0.00000000 \end{bmatrix}$$

The truncated versions P(6) and P(7) of $\infty \times \infty$ transition matrix P are given as:

$$P(6) = \begin{bmatrix} 0.0225 & 0.2550 & 0.7225 & 0.0000 & 0.0000 & 0.0000 \\ 0.0225 & 0.2550 & 0.7225 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0255 & 0.2550 & 0.7225 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0255 & 0.2550 & 0.7225 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0225 & 0.2550 & 0.7225 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0225 & 0.2550 \end{bmatrix}$$

$$P(7) = \begin{bmatrix} 0.0225 & 0.2550 & 0.7225 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0225 & 0.2550 & 0.7225 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0255 & 0.2550 & 0.7225 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0255 & 0.2550 & 0.7225 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0225 & 0.2550 & 0.7225 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0225 & 0.2550 & 0.7225 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0225 & 0.2550 \end{bmatrix}$$

Further, the probability vectors $\pi_1$ and $\pi_2$ at time $1d^+$ and $2d^+$ are as follows:

$$\pi_1 = \begin{bmatrix} 0.00027 & 0.00525 & 0.04182 & 0.17618 & 0.39933 & 0.37715 \end{bmatrix}$$

$$\pi_2 = \begin{bmatrix} 0.00012 & 0.00235 & 0.01862 & 0.08412 & 0.23760 & 0.38469 & 0.27249 \end{bmatrix}$$

The probability distributions of the number of patients in the system at times $1d^+$ and $2d^+$ obtained by using Markov chain match exactly with those calculated theoretically in Tables 3.1 and 3.2 respectively. This shows that the Markov chain approach is a valid and useful approach to determine the probability distribution for the number of patients in the system at various times.

CHAPTER 4

# Appointment Strategies

To better understand the situation in the outpatient clinic, we look at a perfect case when there is no waiting and idle time. Also, we try different strategies of appointment scheduling in the literature review given by Cayirli et al. ([6]) with different assumptions and then compare results and find the best for our assumptions. These strategies are given in Figure 4.1.

## 4.1. Parameters

Let n ($n_1$,$n_2$,...,$n_i$) be a vector of the number of patients arrived for their appointments at different time-slots in some outpatient clinic. Let index $i$ represents the element number from any vector and $i = 1, 2, 3, \ldots, t$. The arriving patients are examined by the physician in order of their appointment times. The vector ScT has the scheduled times for the appointments of each patient. The arrival times of different patients are given by the vector AT ($AT_1$,$AT_2$,...,$AT_i$) where $AT_i$ represents the arrival time of the ith patient. Furthermore, we assume that the physician arrival time is given by PA and the vector of the times when the physician starts consulting patients is represented by PST. PET is referred to as the vector of the physician end time of the service for each patient. Finally, W is the vector of the waiting times of the patients before the start of their service and PEmT represents the physician idle times before the start of the service of each patient. Below is the table of the notation for the ith element from each above-defined vectors.

Table 4.1. Summary of Notation

| | |
|---|---|
| $n_i$ | number of patients arrived at $i^{th}$ time-slot |
| $ScT_i$ | Scheduled time of the $i^{th}$ patient |
| $AT_i$ | Arrival time of the $i^{th}$ patient |
| $PST_i$ | Physician start time for the consultation of the $i^{th}$ patient |
| $PET_i$ | Physician end time of the service of the $i^{th}$ patient |
| $W_i$ | Waiting time of the $i^{th}$ patient before the start of the service |
| $PEmT_i$ | Physician empty time before the start of the service of the $i^{th}$ patient |

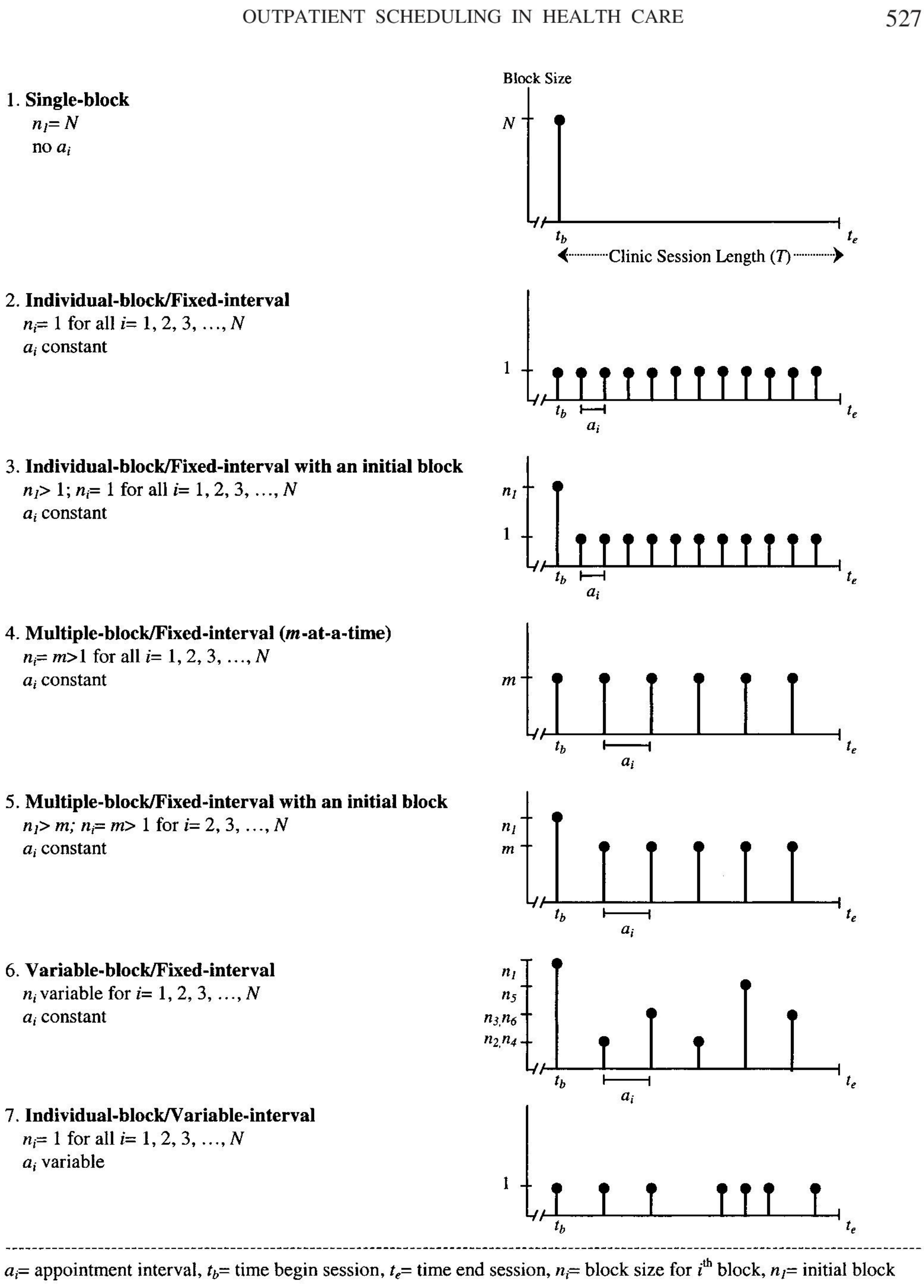


FIGURE 1. Appointment rules in the literature (adapted from Fries and Marathe 1981).

FIGURE 4.1. Different strategies in the literature

The recursive relationships between the parameters are given as:

$$\mathrm{PST}_i = \begin{cases} max(\mathrm{AT}_i, \mathrm{PA}), & \text{if } i = 1 \\ \mathrm{PST}_{i-1} + \mathrm{ST}_{i-1}, & \text{if } \mathrm{PST}_{i-1} + \mathrm{ST}_{i-1} > \mathrm{AT}_i \\ \mathrm{AT}_i & \text{otherwise} \end{cases} \tag{4.1}$$

$$\mathrm{PET}_i = \mathrm{PST}_i + \mathrm{ST}_i \tag{4.2}$$

$$\mathrm{W}_i = max(\mathrm{PST}_i - \mathrm{AT}_i, 0) \tag{4.3}$$

$$\mathrm{PEmT}_i = \begin{cases} max(\mathrm{PST}_1 - \mathrm{PA}, 0), & \text{if } i = 1 \\ \mathrm{PST}_i - \mathrm{PET}_{i-1}, & \text{otherwise} \end{cases} \tag{4.4}$$

where $i = 1, 2, \ldots, t$. These relations are used for R program given in Appendix A to calculate the parameters.

(1) Perfect Case with fixed service time: In the perfect case everything is perfect i.e $p = 1$ means every patient comes for the service and is punctual. The doctor also starts serving patients on time and complete the service on time. Each patient has given an individual appointment. Thus, when all the patients and the physician are on time, there is no waiting and idle time for patients and doctor respectively. This is the perfect case but more unrealistic. The results for all the parameters using recursive relationships in equations 4.1,4.2,4.3,4.4 are as follows:

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 7 | 0 | 7 | 0 | 0 |
| 2 | 7 | 7 | 7 | 7 | 14 | 0 | 0 |
| 3 | 14 | 14 | 7 | 14 | 21 | 0 | 0 |
| 4 | 21 | 21 | 7 | 21 | 28 | 0 | 0 |
| 5 | 28 | 28 | 7 | 28 | 35 | 0 | 0 |
| 6 | 35 | 35 | 7 | 35 | 42 | 0 | 0 |
| 7 | 42 | 42 | 7 | 42 | 49 | 0 | 0 |
| 8 | 49 | 49 | 7 | 49 | 56 | 0 | 0 |
| 9 | 56 | 56 | 7 | 56 | 63 | 0 | 0 |
| 10 | 63 | 63 | 7 | 63 | 70 | 0 | 0 |
| 11 | 70 | 70 | 7 | 70 | 77 | 0 | 0 |
| 12 | 77 | 77 | 7 | 77 | 84 | 0 | 0 |
| 13 | 84 | 84 | 7 | 84 | 91 | 0 | 0 |
| 14 | 91 | 91 | 7 | 91 | 98 | 0 | 0 |
| 15 | 98 | 98 | 7 | 98 | 105 | 0 | 0 |

TABLE 4.2. Results for Perfect case

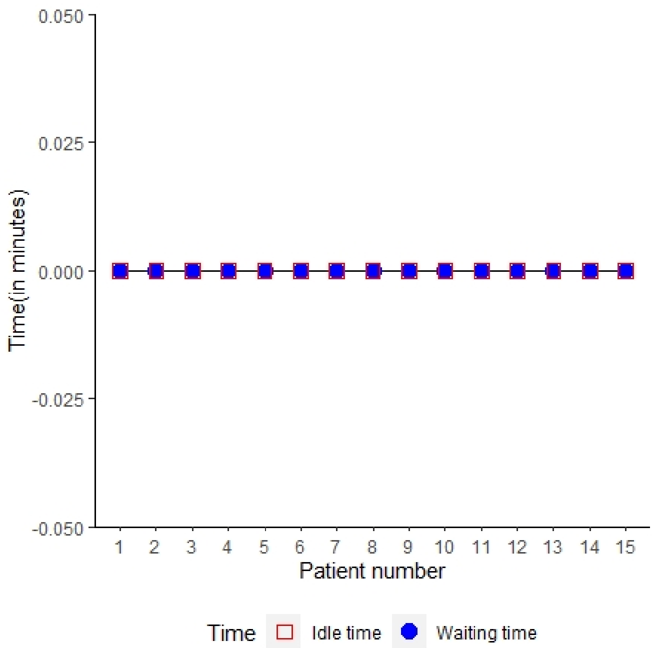


FIGURE 4.2. Perfect case with zero waiting and idle times

All the red squares in the Figure 4.2 indicate the idle times of the doctor and blue solid circles represent the waiting times of the patients.

(2) Single Block with fixed service time: In this rule, all the patients are appointed in only one block at the beginning of the clinic session and let service time be 7 minutes, fixed for all patients. If all the patients arrive in the beginning, then the results are:

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 7 | 0 | 7 | 0 | 0 |
| 2 | 0 | 0 | 7 | 7 | 14 | 7 | 0 |
| 3 | 0 | 0 | 7 | 14 | 21 | 14 | 0 |
| 4 | 0 | 0 | 7 | 21 | 28 | 21 | 0 |
| 5 | 0 | 0 | 7 | 28 | 35 | 28 | 0 |
| 6 | 0 | 0 | 7 | 35 | 42 | 35 | 0 |
| 7 | 0 | 0 | 7 | 42 | 49 | 42 | 0 |
| 8 | 0 | 0 | 7 | 49 | 56 | 49 | 0 |
| 9 | 0 | 0 | 7 | 56 | 63 | 56 | 0 |
| 10 | 0 | 0 | 7 | 63 | 70 | 63 | 0 |
| 11 | 0 | 0 | 7 | 70 | 77 | 70 | 0 |
| 12 | 0 | 0 | 7 | 77 | 84 | 77 | 0 |
| 13 | 0 | 0 | 7 | 84 | 91 | 84 | 0 |
| 14 | 0 | 0 | 7 | 91 | 98 | 91 | 0 |
| 15 | 0 | 0 | 7 | 98 | 105 | 98 | 0 |

TABLE 4.3. Results for Single Block strategy

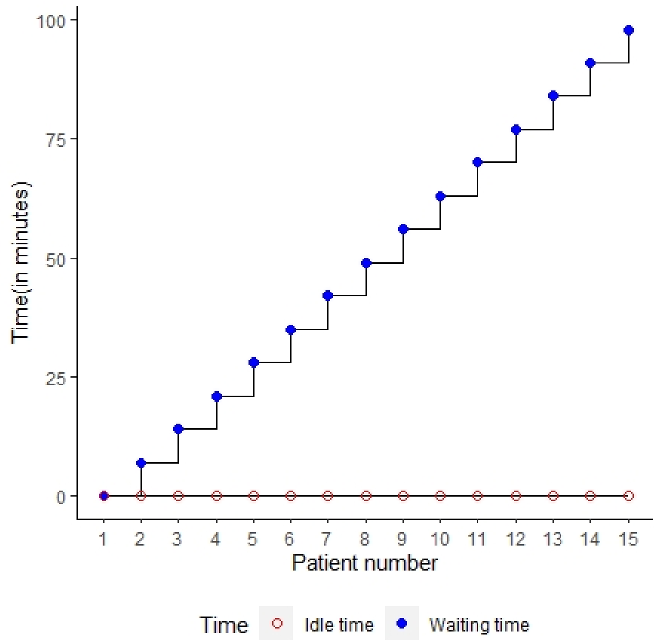


FIGURE 4.3. Single block with constant service times

It can be seen from Table 4.3 and in Figure 4.3 that the idle time of the doctor is zero because there are enough patients to serve for physician but the waiting time of the patients is increasing dramatically as all the patients arrive at the same single scheduled time.

**4.1.1. Assumptions.** We try strategies in the literature with assumptions:

1. Probability of keeping an appointment ($p$) is 0.85. It means some of the patients miss appointments.
2. Physician is punctual. Later we will see the results when the physician is late.
3. Patients are unpunctual
4. Varying service times for all patients.
5. Interval length is constant. Let it be 7 mins.
6. Suppose 15 patients are appointed for each strategy.

(3) Single block with variable service time: We will analyze how the addition of some variability in service times and arrival times affect the waiting times and idle times. Consider the single block rule with variable service time. The changes follow the same trend as in the case of fixed service time.

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | −4.382 | 6.027 | 0.000 | 6.027 | 4.382 | 0.000 |
| 2 | 0 | −3.744 | 6.765 | 6.027 | 12.792 | 9.771 | 0.000 |
| 3 | 0 | −3.234 | 7.739 | 12.792 | 20.531 | 16.026 | 0.000 |
| 4 | 0 | −2.940 | 6.681 | 20.531 | 27.212 | 23.471 | 0.000 |
| 5 | 0 | −2.879 | 6.964 | 27.212 | 34.176 | 30.090 | 0.000 |
| 6 | 0 | −1.200 | 7.199 | 34.176 | 41.375 | 35.375 | 0.000 |
| 7 | 0 | −1.159 | 6.987 | 41.375 | 48.362 | 42.534 | 0.000 |
| 8 | 0 | −0.023 | 6.372 | 48.362 | 54.734 | 48.385 | 0.000 |
| 9 | 0 | 1.517 | 7.655 | 54.734 | 62.389 | 53.218 | 0.000 |
| 10 | 0 | 1.870 | 7.337 | 62.389 | 69.726 | 60.519 | 0.000 |
| 11 | 0 | 2.176 | 7.588 | 69.726 | 77.315 | 67.550 | 0.000 |
| 12 | 0 | 2.698 | 6.216 | 77.315 | 83.530 | 74.616 | 0.000 |
| 13 | 0 | 2.774 | 7.447 | 83.530 | 90.978 | 80.756 | 0.000 |
| 14 | 0 | 4.347 | 6.823 | 90.978 | 97.800 | 86.631 | 0.000 |
| 15 | 0 | 4.919 | 7.642 | 97.800 | 105.442 | 92.881 | 0.000 |

TABLE 4.4. Results for single Block strategy with some variablity

By comparing Table 4.3 and Table 4.4, we can see that by adding some variability the waiting times of the starting half of the patients are increased by some amount from those without any variability whereas from 8th patient the waiting times decrease from the respective wait times in Table 4.3. The average of the waiting times in case of no variability is 49 minutes, and the mean wait time in case of some variability is 48.414 minutes. Somehow, the waiting times are improved by adding some variability in case of single block.

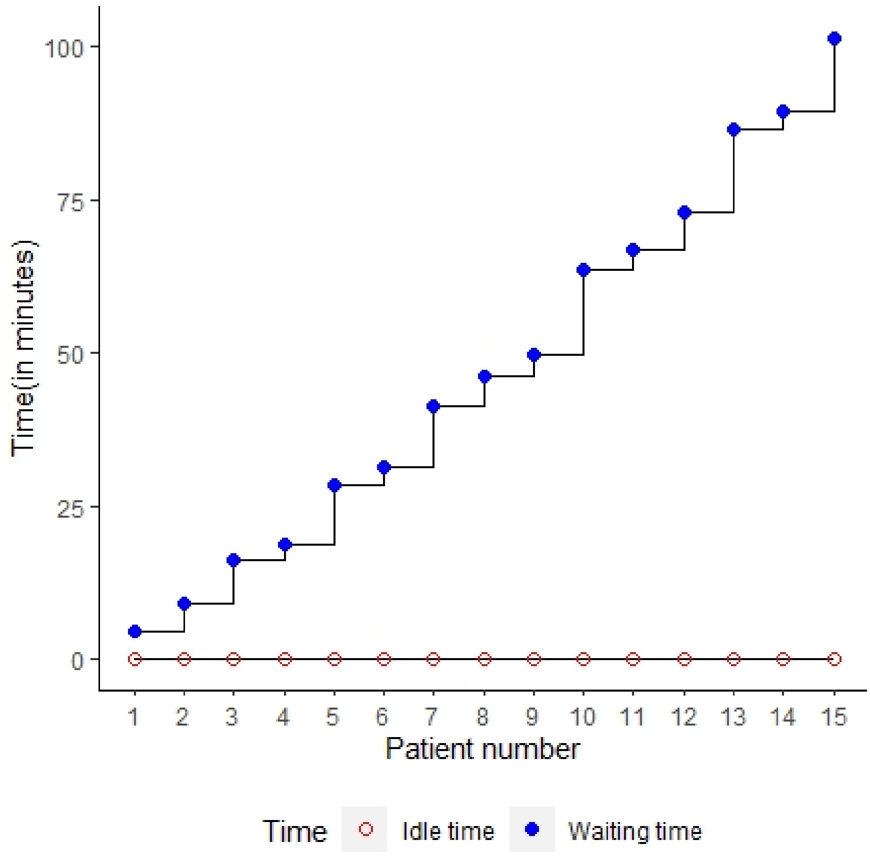


FIGURE 4.4. Single block with variable service times

(4) Individual block with initial block: Here 4 patients are appointed in the beginning slot and then each patient is given a unique appointment time. The main objective of an initial block in the starting was to keep an inventory of patients so that the chances of doctor staying idle were minimized if the first patient arrives late or fails to show up. Suppose that the patients are unpunctual and service times are varying for patients.

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | −3.200 | 6.987 | 0.000 | 6.987 | 3.200 | 0.000 |
| 2 | 0 | 0.176 | 6.372 | 6.987 | 13.360 | 6.811 | 0.000 |
| 3 | 0 | 0.774 | 7.655 | 13.360 | 21.014 | 12.585 | 0.000 |
| 4 | 0 | 2.919 | 7.337 | 21.014 | 28.351 | 18.095 | 0.000 |
| 5 | 7 | 9.347 | 7.588 | 28.351 | 35.940 | 19.004 | 0.000 |
| 6 | 14 | 9.121 | 6.216 | 35.940 | 42.156 | 26.818 | 0.000 |
| 7 | 28 | 27.517 | 7.447 | 42.156 | 49.603 | 14.639 | 0.000 |
| 8 | 49 | 43.256 | 6.823 | 49.603 | 56.426 | 6.347 | 0.000 |
| 9 | 56 | 51.672 | 7.642 | 56.426 | 64.067 | 4.753 | 0.000 |
| 10 | 63 | 59.861 | 7.294 | 64.067 | 71.362 | 4.206 | 0.000 |
| 11 | 70 | 63.134 | 7.566 | 71.362 | 78.927 | 8.228 | 0.000 |
| 12 | 77 | 73.824 | 7.106 | 78.927 | 86.033 | 5.104 | 0.000 |
| 13 | 84 | 85.697 | 7.059 | 86.033 | 93.093 | 0.337 | 0.000 |
| 14 | 91 | 87.403 | 7.579 | 93.093 | 100.672 | 5.689 | 0.000 |
| 15 | 98 | 95.821 | 6.047 | 100.672 | 106.718 | 4.851 | 0.000 |

TABLE 4.5. Results for individual block with initial block

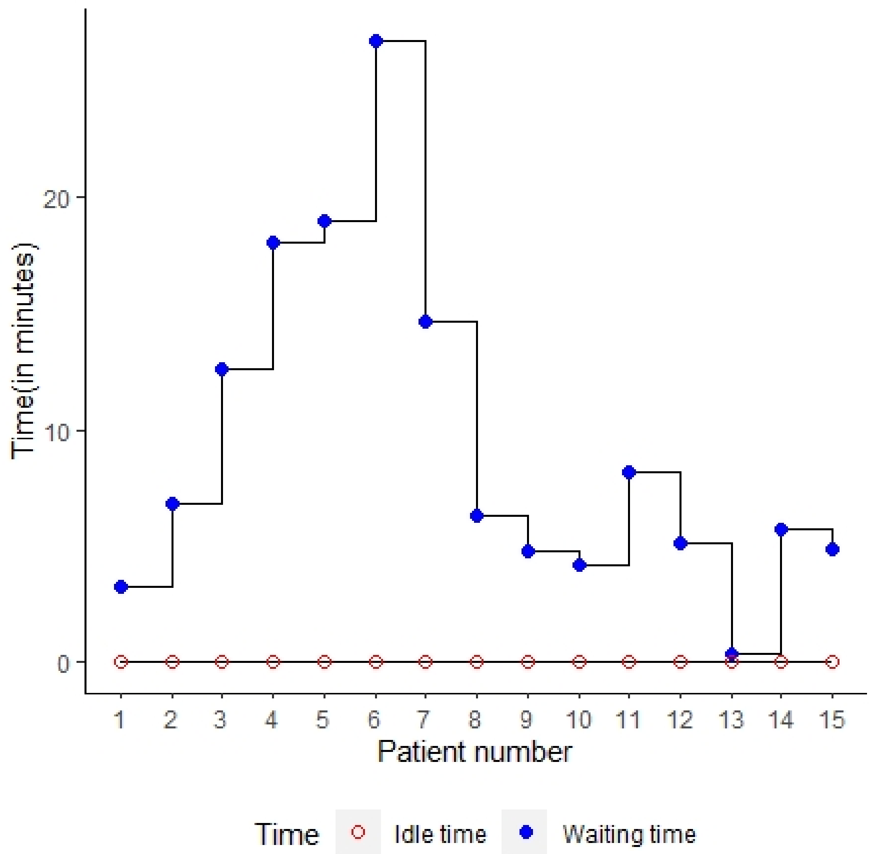


FIGURE 4.5. Individual block with initial block: 4111...

From Table 4.5 and Figure 4.5, we observe that the waiting times are increasing towards the middle. The patient number 6 waits for maximum i.e for 26.818 minutes. All the patients after the 7th patient wait are lesser than the patients before the 7th patient. The minimum waiting time is 0.337 minutes is for the 13th patient. Moreover, the physician is busy all the time because the idle time of the doctor is 0 min every time. The average waiting time for this rule is 9.378 minutes.

(5) Multiple-block: We appoint 2 patients for each appointment slot with appointment interval kept constant i.e after every 7 minutes, 2 patients are given the same appointment time. The results obtained for multiple block rule are:

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | $-5.234$ | 6.765 | 0.000 | 6.765 | 5.234 | 0.000 |
| 2 | 0 | $-4.940$ | 7.739 | 6.765 | 14.504 | 11.705 | 0.000 |
| 3 | 7 | 3.841 | 6.681 | 14.504 | 21.185 | 10.663 | 0.000 |
| 4 | 7 | 6.870 | 6.964 | 21.185 | 28.149 | 14.315 | 0.000 |
| 5 | 14 | 11.977 | 7.199 | 28.149 | 35.348 | 16.172 | 0.000 |
| 6 | 14 | 14.698 | 6.987 | 35.348 | 42.335 | 20.650 | 0.000 |
| 7 | 21 | 21.176 | 6.372 | 42.335 | 48.708 | 21.159 | 0.000 |
| 8 | 28 | 24.800 | 7.655 | 48.708 | 56.362 | 23.907 | 0.000 |
| 9 | 28 | 30.919 | 7.337 | 56.362 | 63.699 | 25.443 | 0.000 |
| 10 | 35 | 35.774 | 7.588 | 63.699 | 71.288 | 27.925 | 0.000 |
| 11 | 42 | 44.347 | 6.216 | 71.288 | 77.504 | 26.941 | 0.000 |
| 12 | 49 | 44.121 | 7.447 | 77.504 | 84.951 | 33.382 | 0.000 |
| 13 | 49 | 48.517 | 6.823 | 84.951 | 91.774 | 36.434 | 0.000 |
| 14 | 56 | 50.256 | 7.642 | 91.774 | 99.416 | 41.518 | 0.000 |
| 15 | 56 | 51.672 | 7.294 | 99.416 | 106.710 | 47.743 | 0.000 |

TABLE 4.6. Results for multiple block strategy

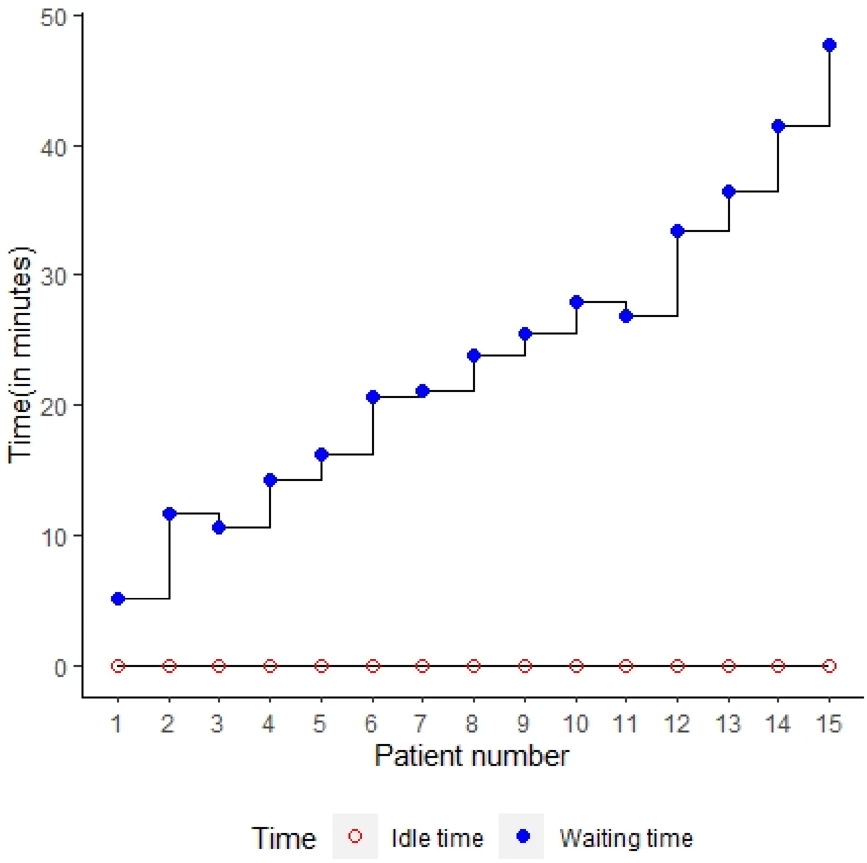


FIGURE 4.6. Multiple-block: 2222...

As in Figure 4.6, the waiting times are increasing except for two patients. For 3rd patient waiting time is less than that for 2nd patient. This is because the 2nd person comes earlier as compared to the arrival of the 3rd person. The 2nd person arrives 4.940 minutes before his/her scheduled appointment whereas the 3rd person comes 3.159 minutes. However, the decrease in wait time from 10th patient to 11th patient is because of missing an appointment by 1 patient scheduled at 35 minutes. The rest of the increase in wait times is because of assigning 2 to patients for each time-slot. The physician has enough patients to deal with. That is why idle time of the physician is zero. The average waiting time of the patients is 24.213 minutes.

(6) Multiple-block with an initial block: This rule is a variation of the above rule with an initial block. We appoint 4 patients at the beginning of the session and each time 2 patients for the rest of the slots in the session. As compare to the results for the previous strategy, we have an idea that in this rule also the waiting times would increase for patients because we have added a block of 4 patients in the beginning. The results are as:

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | −6.382 | 6.027 | 0.000 | 6.027 | 6.382 | 0.000 |
| 2 | 0 | −5.234 | 6.765 | 6.027 | 12.792 | 11.261 | 0.000 |
| 3 | 0 | −4.940 | 7.739 | 12.792 | 20.531 | 17.732 | 0.000 |
| 4 | 0 | −0.130 | 6.681 | 20.531 | 27.212 | 20.661 | 0.000 |
| 5 | 7 | 3.841 | 6.964 | 27.212 | 34.176 | 23.371 | 0.000 |
| 6 | 7 | 7.698 | 7.199 | 34.176 | 41.375 | 26.477 | 0.000 |
| 7 | 14 | 11.977 | 6.987 | 41.375 | 48.362 | 29.398 | 0.000 |
| 8 | 14 | 14.176 | 6.372 | 48.362 | 54.734 | 34.186 | 0.000 |
| 9 | 21 | 23.919 | 7.655 | 54.734 | 62.389 | 30.815 | 0.000 |
| 10 | 28 | 24.800 | 7.337 | 62.389 | 69.726 | 37.589 | 0.000 |
| 11 | 28 | 28.774 | 7.588 | 69.726 | 77.315 | 40.952 | 0.000 |
| 12 | 35 | 37.347 | 6.216 | 77.315 | 83.530 | 39.968 | 0.000 |
| 13 | 42 | 37.121 | 7.447 | 83.530 | 90.978 | 46.409 | 0.000 |
| 14 | 49 | 43.256 | 6.823 | 90.978 | 97.800 | 47.722 | 0.000 |
| 15 | 49 | 48.517 | 7.642 | 97.800 | 105.442 | 49.284 | 0.000 |

TABLE 4.7. Results for multiple block with initial block strategy

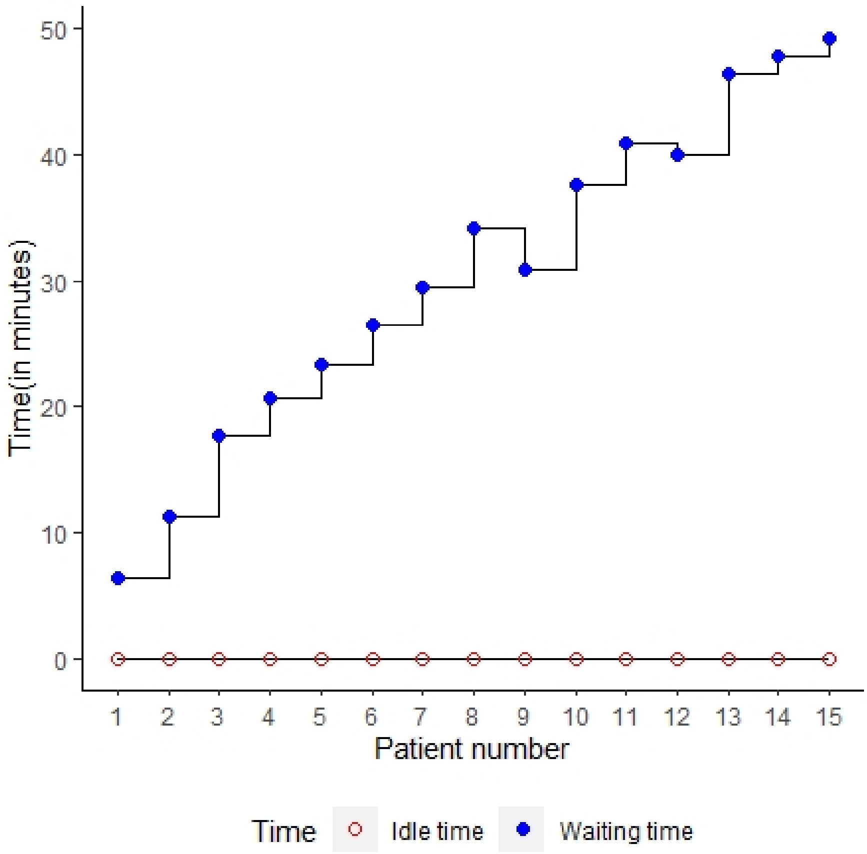


FIGURE 4.7. Multiple-block with an initial block: 4222...

There is also the same trend as in multiple block for waiting times of patients. For this rule, patients wait for more than in strategy multiple block without initial block. The mean of waiting times is 30.814 minutes.

(7) Variable block: This rule allows different block sizes during the clinic session with fixed appointment intervals. If 4 patients appoint in the first slot of the session, 1 patient in the second slot and 3 patients in the third slot and so on.

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | −5.234 | 6.964 | 0.000 | 6.964 | 5.234 | 0.000 |
| 2 | 0 | −3.159 | 7.199 | 6.964 | 14.163 | 10.123 | 0.000 |
| 3 | 0 | −0.130 | 6.987 | 14.163 | 21.150 | 14.293 | 0.000 |
| 4 | 0 | 0.698 | 6.372 | 21.150 | 27.523 | 20.452 | 0.000 |
| 5 | 7 | 4.977 | 7.655 | 27.523 | 35.178 | 22.546 | 0.000 |
| 6 | 14 | 10.800 | 7.337 | 35.178 | 42.514 | 24.377 | 0.000 |
| 7 | 14 | 14.176 | 7.588 | 42.514 | 50.103 | 28.338 | 0.000 |
| 8 | 14 | 16.919 | 6.216 | 50.103 | 56.319 | 33.184 | 0.000 |
| 9 | 28 | 23.121 | 7.447 | 56.319 | 63.766 | 33.197 | 0.000 |
| 10 | 28 | 28.774 | 6.823 | 63.766 | 70.589 | 34.992 | 0.000 |
| 11 | 28 | 30.347 | 7.642 | 70.589 | 78.231 | 40.242 | 0.000 |
| 12 | 42 | 41.517 | 7.294 | 78.231 | 85.525 | 36.714 | 0.000 |
| 13 | 49 | 43.256 | 7.566 | 85.525 | 93.091 | 42.269 | 0.000 |
| 14 | 56 | 51.672 | 7.106 | 93.091 | 100.197 | 41.418 | 0.000 |
| 15 | 56 | 52.861 | 7.059 | 100.197 | 107.256 | 47.336 | 0.000 |

TABLE 4.8. Results for variable block strategy

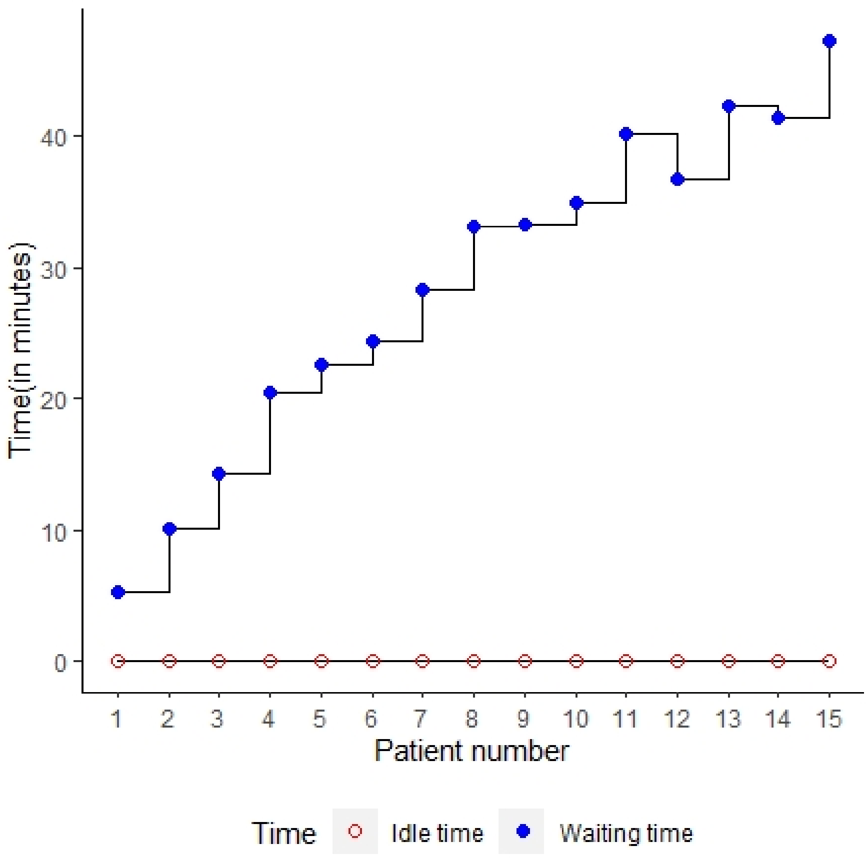


FIGURE 4.8. Variable block: 41313...

The average wait time of patients with this strategy is 28.981 minutes. Out of 15 patients, the maximum wait time is 47.336 minutes and it is for the 15th patient. The waiting times are increasing most of the time.

We will try the combination of some of the above strategies from the literature. We combine the variable block and individual block strategies. In this strategy, we have appointed 4 patients for 1st slot, 1 patient for 2nd slot, 2 patients for 3rd slot and 3 patients for the 4th slot. Further, it is then followed by a single appointment for each slot.

**4.1.2. Combination of Variable block and individual block:** Using the recursive relationships from equations 4.1, 4.2, 4.3, 4.4, the results for this new strategy are given in Table 4.9.

| n | ScT | AT | ST | PST | PET | W | PEmT |
|---|---|---|---|---|---|---|---|
| 1 | 0 | −3.159 | 6.681 | 0.000 | 6.681 | 3.159 | 0.000 |
| 2 | 0 | −2.023 | 6.964 | 6.681 | 13.645 | 8.704 | 0.000 |
| 3 | 0 | 0.176 | 7.199 | 13.645 | 20.844 | 13.469 | 0.000 |
| 4 | 0 | 0.698 | 6.987 | 20.844 | 27.831 | 20.146 | 0.000 |
| 5 | 7 | 9.919 | 6.372 | 27.831 | 34.204 | 17.912 | 0.000 |
| 6 | 14 | 10.800 | 7.655 | 34.204 | 41.858 | 23.403 | 0.000 |
| 7 | 14 | 14.774 | 7.337 | 41.858 | 49.195 | 27.084 | 0.000 |
| 8 | 21 | 16.121 | 7.588 | 49.195 | 56.784 | 33.074 | 0.000 |
| 9 | 21 | 23.347 | 6.216 | 56.784 | 63.000 | 33.437 | 0.000 |
| 10 | 28 | 27.517 | 7.447 | 63.000 | 70.447 | 35.483 | 0.000 |
| 11 | 49 | 43.256 | 6.823 | 70.447 | 77.270 | 27.191 | 0.000 |
| 12 | 56 | 51.672 | 7.642 | 77.270 | 84.911 | 25.597 | 0.000 |
| 13 | 63 | 59.861 | 7.294 | 84.911 | 92.206 | 25.050 | 0.000 |
| 14 | 70 | 63.134 | 7.566 | 92.206 | 99.771 | 29.072 | 0.000 |
| 15 | 77 | 73.824 | 7.106 | 99.771 | 106.877 | 25.948 | 0.000 |

Table 4.9. Results for combination of variable block and individual block strategies for 15 patients

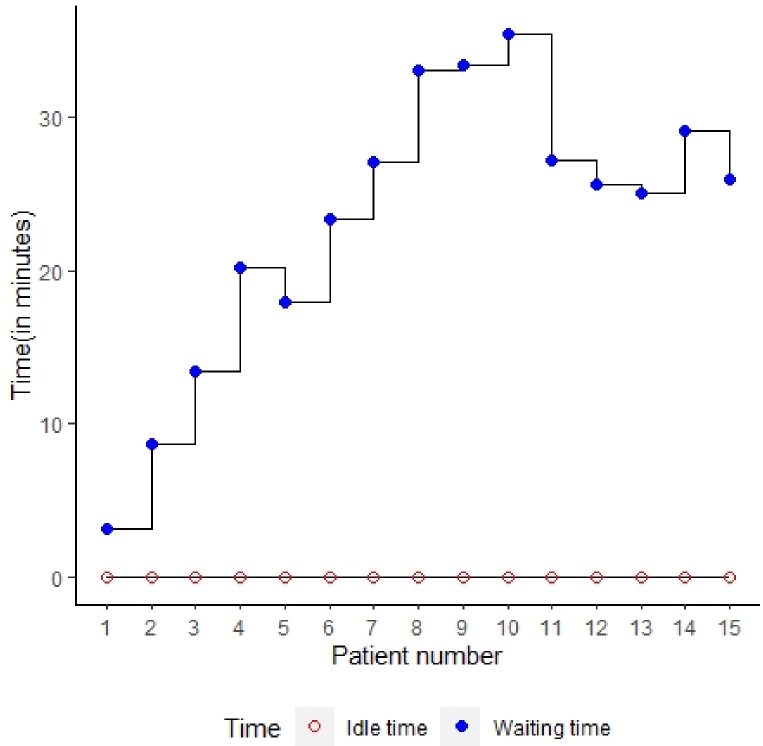


FIGURE 4.9. Combination of variable and individual blocks: 412311...

The mean waiting time of patients in the Table 4.9 is 23.248 mins.
On comparing the average waiting times from all the above-discussed appointment rules, 4111 ... has minimum mean wait. We have an idea from here that it can do best from all the mentioned strategies. Further, we define the objective function and we will decide the best rule by running simulation 100 times for each strategy. Then decide the best strategy on the basis of the minimum value of the objective function.

**4.1.3. Objective Function:** Let f be the objective function that needs to be minimized. It is given by the sum of the weighted average of patients' waiting times and physician idle time. The relative weights for patients' waiting time and physician idle time are denoted by $w_p$ and $w_i$ respectively. Thus f is given as:

$$f = w_p * mean(W) + w_i * mean(PEmT) \tag{4.5}$$

$W$ is the patient waiting time
$PEmT$ is the physician idle time

As figured in (Keller et al.[15]), for estimating the weight of doctor idle time, Keller and Laughhunn divided the annual salary of the doctor by the hours worked per year and used the minimum wage to reflect the opportunity cost of the patients' waiting time. In a report of Caribbean Medicine with the date June 20, 2018, they have mentioned that according to Statistics Canada, a Canadian physician's salary in Ontario is about $\$340,000$ per year. Below is the graph from this report that shows the average Canadian physician salary by province:

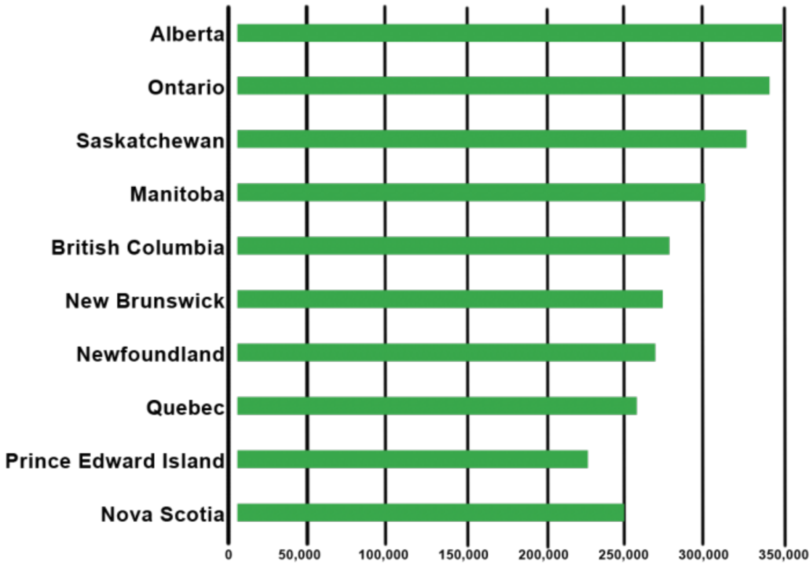


FIGURE 4.10. Average salary of Canadian Physician by province

Moreover, in the report of Medics Domain published on October 8, 2020, the average medical doctor salary in Canada by province is given as: Let average

| PROVINCE | SALARY PER MONTH | SALARY PER YER |
|---|---|---|
| Alberta | $\$32,031$ | $\$384,380$ |
| Ontario | $\$30,000$ | $\$360,000$ |
| Manitoba | $\$29,558$ | $\$354,705$ |
| Quebec | $\$27,083$ | $\$325,000$ |
| New Brunswick | $\$25,177$ | $\$302,123$ |
| Prince Edward Island | $\$25,424$ | $\$305,091$ |
| Saskatchewan | $\$24,083$ | $\$288,995$ |
| British Columbia | $\$22,750$ | $\$273,000$ |
| New Found Land | $\$22,470$ | $\$269,646$ |
| Nova Scotia | $\$21,614$ | $\$259,368$ |

TABLE 4.10. Average medical doctor salary in Canada by province

working hours of the physicians in Ontario, Canada is 40 hours per week. Let the physician works for 48 weeks excluding vacation weeks. Therefore approximate total working hours of a physician per year are $40 \times 48 = 1920$ hours. Thus, the average hourly salary of the physician is $350,000/1920 = \$182.292$. However, the minimum wage level by jurisdiction for employees in Ontario is \$14.25 per hour. Therefore we can have an idea from here for the relative weight of physician idle time as $182.292/14.25 = 12.79$, which is approximately 13. Since the physician decides the scheduling times so physicians could give more priority for his/her time. Therefore let the weight for physician idle time($w_i$) be 15 times the weight for patients' waiting time($w_p$). Therefore, $w_p = 14.25/60$ and $w_i = 15 \times 14.25/60$. Thus from equation 4.5, our objective function is:

$$f = 14.25 \times (PW + 15 \times DW)/60 \tag{4.6}$$

where $f$ is the expected cost per patient in the system. The expected total cost of the system with 15 patients is given by $f^* = 15 \times f$.

**4.1.4. Comparison of strategies:** Let each patient has a probability of 0.85 of keeping its appointment. Let v be the vector of mean waiting times for 500 runs of each strategy and t be the vector of mean physician idle time. We assume that waiting patients are seen in order of their appointments. Let f be the objective function. The results for different strategies with 15 patients are given below:

| strategy | mean(v) | mean(t) | $f^*$ | SD(v) | SD(t) |
|---|---|---|---|---|---|
| 41111... | 11.998 | 0.303 | 58.949 | 5.255 | 0.559 |
| 22222... | 23.833 | 0.025 | 86.248 | 3.879 | 0.073 |
| 42222... | 29.236 | 0.002 | 104.260 | 4.075 | 0.018 |
| 41313... | 28.049 | 0.003 | 100.074 | 4.013 | 0.018 |
| 412311... | 22.303 | 0.048 | 82.039 | 6.223 | 0.264 |
| 421311... | 23.056 | 0.027 | 83.584 | 6.192 | 0.150 |
| 431211... | 24.309 | 0.027 | 88.022 | 6.006 | 0.138 |

TABLE 4.11. Results for different strategies after performing simulation 500 times

Table 4.11 shows that the value of the objective function or the total cost is minimum for an appointment rule 41111 ... which is \$58.949. For the best strategy, the objective function should be minimum. However, 4222 ... strategy has a maximum value of objective function which is \$104.260. Also, the mean waiting time of the patients is minimum for 4111 ... which is 11.998 minutes. Hence, after performing simulation with 500 runs for each strategy when the physician is on time, we conclude that 41111 ... is the best strategy out of all the above strategies in Table 4.11.

**4.1.5. When a physician is 5 mins late:** Further, suppose that the physician is 5 minutes late. With the lateness of the physician, the waiting time of the patients' increases by some amount from the waiting times in table 4.11 whereas the idle time of physician decreases. The results are as follow:

| strategy | mean(v) | mean(t) | $f^*$ | SD(v) | SD(t) |
|---|---|---|---|---|---|
| 41111 ... | 17.023 | 0.149 | 68.602 | 5.879 | 0.384 |
| 22222 ... | 28.200 | 0.004 | 100.673 | 3.938 | 0.036 |
| 42222 ... | 34.829 | 0.000 | 124.078 | 3.752 | 0.000 |
| 41313 ... | 33.059 | 0.000 | 117.776 | 3.971 | 0.002 |
| 412311 ... | 28.148 | 0.012 | 100.910 | 6.154 | 0.079 |
| 421311 ... | 27.761 | 0.010 | 99.415 | 5.679 | 0.087 |
| 431211 ... | 29.138 | 0.010 | 104.344 | 6.093 | 0.098 |

TABLE 4.12. Results for different strategies after performing simulation 500 times, when physician is 5 mins late

Again, from the above results, it is clear that the objective function is minimum for strategy 41111 ... and maximum for strategy 42222 .... The explanation of the best strategy from the above table is the same as for table 4.11. Thus when the physician is late, for our assumptions the best strategy is still 41111 ....

**4.1.6. Plots for the best strategy with different probabilities of keeping an appointment:** The plots of waiting times and idle times for different probabilities are given below:

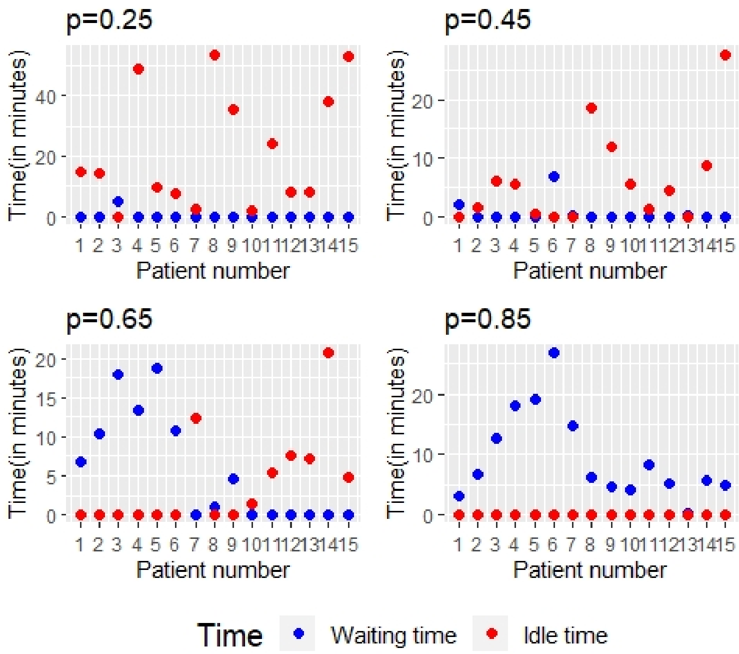


FIGURE 4.11. Plots for different probabilities for strategy 4111

In figure 4.11, red and blue dots represent the idle time of the physician and waiting time of the patients respectively. The x-axis represents the patient number and the y-axis shows the time in minutes. As the probability of attending an appointment increases, the waiting time of the patients also increases. Higher probability means more patients come for their appointments and they have to wait for more. On the contrary, the idle time of the doctor becomes zero, since the doctor has enough patients to serve. In figure 4.11, when $p = 0.25$ the idle time of the doctor is large because a lot of patients miss appointments and the physician has to wait for service whereas the waiting time of all the patients except one is zero. When $p = 0.85$ most of the patients attend their appointments and the physician needs not to wait. Hence the idle time of the doctor is zero. However, patients have to wait for more. At the beginning of the session, 4 patients are appointment at the same time and after that, only a single patient is called. Therefore the waiting time of the patients increases in the beginning. There is a trade-off between waiting times and idle times. Moreover, some patients miss appointment and the block of 4 patients overcome the situation of missing appointments that improves the idle time of the doctor. Thus, the waiting time in the middle of the session decreases. This will be the same case in other strategies that waiting times increase with increasing probability.

**4.1.7. Kernel density estimation.** The kernel density estimation (KDE) is a non-parametric method to estimate the probability density function of a random variable. It was given by Rosenblatt et al. ([25]). For best strategy 41111 . . . , the KDE plot for the waiting times is given in Figure 4.12. In our density plots, the kernel function we are using is Gaussian.

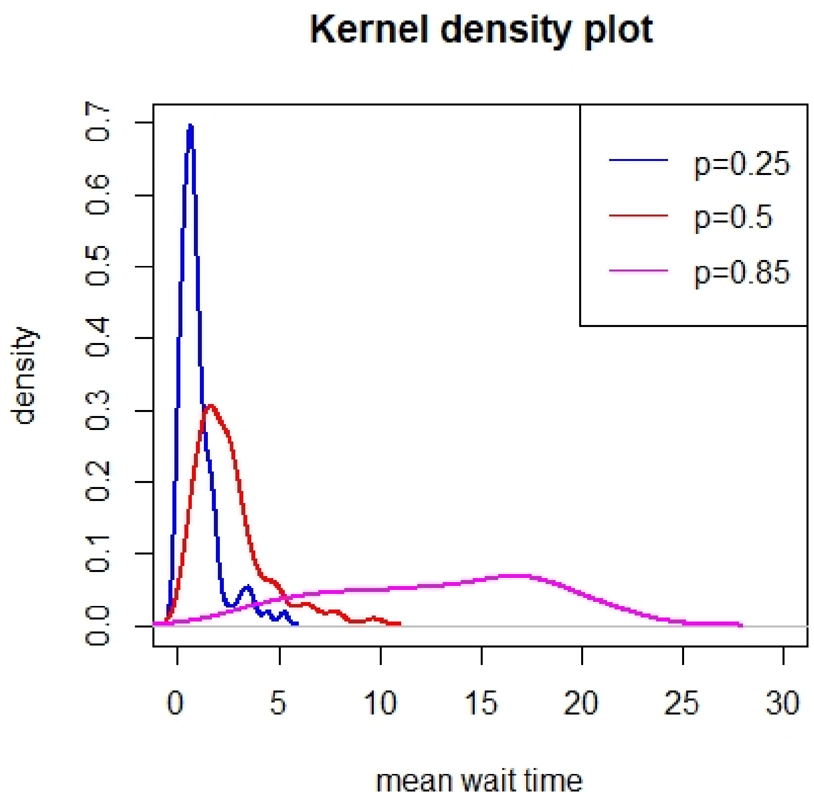


FIGURE 4.12. Kernel density estimation for strategy 41111 with increasing probabilities

The plot in Figure 4.12 gives the density plot for mean waiting times of the patients with increasing probabilities of keeping an appointment. In the plot N=100 i.e 100 observations and bandwidth is 0.2153. The plot with $p = 0.85$ has a wider range than the other two plots and represented by magenta colour. For the blue curve where $p = 0.25$, the mean of waiting times is 1.012 minutes and for the red curve $p = 0.5$ and the mean is 2.542 minutes. Further, the mean for plot

with $p = 0.85$ is 12.978 minutes. We can see from the curves that the mean and the variance are becoming bigger as the probability of keeping an appointment is increasing. Thus, the range of the waiting times increases with the increasing probability. Similarly, we can conclude the same result for the other strategies.
Further, we will see the results for different strategies with variable interval lengths. For this, we generate a sequence ($sq$) that is increasing towards the middle then it starts decreasing with a constant difference of 1. After that, we make some random changes to the sequence $sq$ to make the difference a variable. Let it call $new_{seq}$ and add this in original $sq$, where $new_{seq} = a * sq^b$. We then find the best $(a, b)$ that minimizes the objective function in equation 4.6. All the terms $sq, new_{seq}, a, b$ are taken from the R program provided in Appendix A. For this, we will use the concept of simulation annealing. We construct a program that performs like simulation annealing.

CHAPTER 5

# Modified simulation annealing

Finding an optimal solution for certain optimization problems can be a difficult task, often practically impossible. This is because we need to search through an enormous number of possible solutions to find the optimal one. In this case, we have to look for something that is close enough to the optimal one.

Simulation Annealing (SA) is one of the methods for solving optimization problems. It was introduced by Kirkpatrick et al.[16] in 1982. SA algorithm was originally inspired by the process of annealing in metalwork. Annealing involves heating a material to a specified temperature and cooling it at a very slow and controlled rate to alter its physical properties due to the changes in its internal structure. SA is a metaheuristic (local search algorithm) capable of escaping from local optima.

It helps us to reach a global optimum of a given function. Unlike traditional optimization techniques like a random walk or hill climbing it would not get stuck at a local optimum. In SA we even accept the worse solutions, this gives the algorithm the ability to jump out of any local optimum.

To minimize some function f$(a, b)$ or find $(a^*, b^*)$ to minimize given function f$(a, b)$, one numerical method is as follows:
Assume that $a_l < a^* < a_h$ where $(a_l, a_h)$ is an interval and $b_l < b^* < b_h$ where $(b_l, b_h)$ is an interval. Let $k_1$ and $k_2$ be two random Bernoulli values with high probability (0.9) of 0 and low probability (0.1) of 1. We start with a random guess $(a_0, b_0)$ and compute $f^*(a_0, b_0)$. After that we attempt to find a better solution by making an incremental change to the current solution. Let $a_1$ and $b_1$ be the values of $a$ and $b$ respectively after some incremental changes. Let $f_1^* = f^*(a_1, b_1)$ and $f_0^* = f^*(a_0, b_0)$. Some formulas are given as:

$$k_1 = rbinom(1, 1, 0.1) \tag{5.1}$$

$$k_2 = rbinom(1, 1, 0.1) \tag{5.2}$$

$$a_1 = a_0 + k_1 * runif(0, 1) + runif(0, 0.1) \tag{5.3}$$

$$b_1 = b_0 + k_2 * runif(0, 2) + runif(0, 0.1) \tag{5.4}$$

If $f_1^* < f_0^* + 0.5 * f_0^* * ((a_1 - a_0)^2 + (b_1 - b_0)^2 > 0.03) * (i < n/2)$ then we accept the new solution $(a_1, b_1)$, otherwise current solution. We repeat this 100 times.

**5.0.1. Why bad results.** For positive support of random uniform variables, we got increasing value of the expected total cost of the system which we call bad results. These are given in Table 5.1

| new $f^*$ (in \$) | $a_0$ | $b_0$ |
|---|---|---|
| 82.144 | 0.500 | 0.500 |
| 81.236 | 0.513 | 0.542 |
| 76.694 | 0.550 | 0.551 |
| 82.914 | 0.647 | 0.558 |
| 2008.684 | 0.647 | 0.558 |
| | ⋮ | |
| 90.395 | 0.731 | 0.657 |
| 94.702 | 0.788 | 0.717 |
| 525.426 | 0.788 | 0.717 |
| | ⋮ | |
| 173.273 | 0.991 | 0.911 |
| 168.374 | 0.991 | 0.911 |
| 191.527 | 0.991 | 0.911 |
| 183.408 | 0.991 | 0.911 |
| 189.878 | 0.991 | 0.911 |
| 213.718 | 0.991 | 0.911 |

TABLE 5.1. Bad results

From the table 5.1, we can see that the value for the objective function in equation 4.6 is becoming bigger most of the time. These are bad results since we want to minimize the objective function but the value of the objective function is becoming large with increasing iterations. This situation arises because we choose the positive support for randomly generated numbers in equations 4.6, 5.1, 5.2, 5.3. When the value of $a_0$ and $b_0$ is updated, it is always bigger than the previous values of $a_0$ and $b_0$ as some positive number is added every time. In our case, I am trying to find the value of $a$ and $b$ in $a * sq^b$ where $sq$ is an increasing decreasing sequence to minimize the objective function f. As $b_0$ becomes greater than 1 and $b_0$ is power then elements in $sq$ becomes large. This means my interval length becomes bigger than for the previous iteration and idle time for the doctor might increase. Thus the value of my objective function increases with increasing iterations.

We allow some negative support for the random variables in $a_1$ and $b_1$ in equations 5.3, 5.4. The remaining procedure is the same as defined in the beginning, keeping $a_1$ positive as $a_1$ is the multiple in the $new_{seq}$ and interval length cannot be negative.

$$a_1 = max(a_0 + k_1 * runif(1, -1, 1) + runif(1, -0.1, 0.1), 0.1) \tag{5.5}$$

$$b_1 = b_0 + k_2 * runif(1, -1, 1) + runif(1, -0.1, 0.1) \tag{5.6}$$

For different strategies, the minimum value of the objective function obtained by performing simulation annealing 100 times and their corresponding $a_0$ and $b_0$ are given in Table 5.2.

| strategy | $min(f_0^*)$ | $a_0$ | $b_0$ |
|---|---|---|---|
| 41111... | 73.940 | 1.072 | 0.010 |
| 22222... | 64.196 | 1.983 | 0.990 |
| 42222... | 91.435 | 1.111 | 1.671 |
| 41313... | 73.829 | 3.388 | 0.747 |
| 412311... | 99.529 | 0.969 | 0.785 |
| 421311... | 97.620 | 1.636 | 0.529 |
| 431211... | 102.746 | 0.703 | 1.044 |

TABLE 5.2. Best $a_0$ and $b_0$ for different strategies

Further, we will decide the best out of all the above appointment strategies with varying interval lengths by performing simulation 500 times for each strategy for optimal $a_0$ and $b_0$. Then the strategy with a minimum objective function will be the best one. We find the mean waiting time for each of 15 patients and try to plot them to figure out the results and conclusions.

(1) Individual block with an initial block(41111...): The results for mean waiting time for each patient and mean idle time of the physician before the start of the service of each patient are given in Table 5.3.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.628 | 0.022 |
| 2 | 9.182 | 0.000 |
| 3 | 14.200 | 0.000 |
| 4 | 19.671 | 0.002 |
| 5 | 24.400 | 0.002 |
| 6 | 26.916 | 0.030 |
| 7 | 27.376 | 0.078 |
| 8 | 27.093 | 0.293 |
| 9 | 25.318 | 0.347 |
| 10 | 22.851 | 0.436 |
| 11 | 19.709 | 0.602 |
| 12 | 17.172 | 0.660 |
| 13 | 16.525 | 0.841 |
| 14 | 16.523 | 0.567 |
| 15 | 18.168 | 0.290 |

TABLE 5.3. Mean waiting time for each patient and mean idle times of physician for 4111 . . .

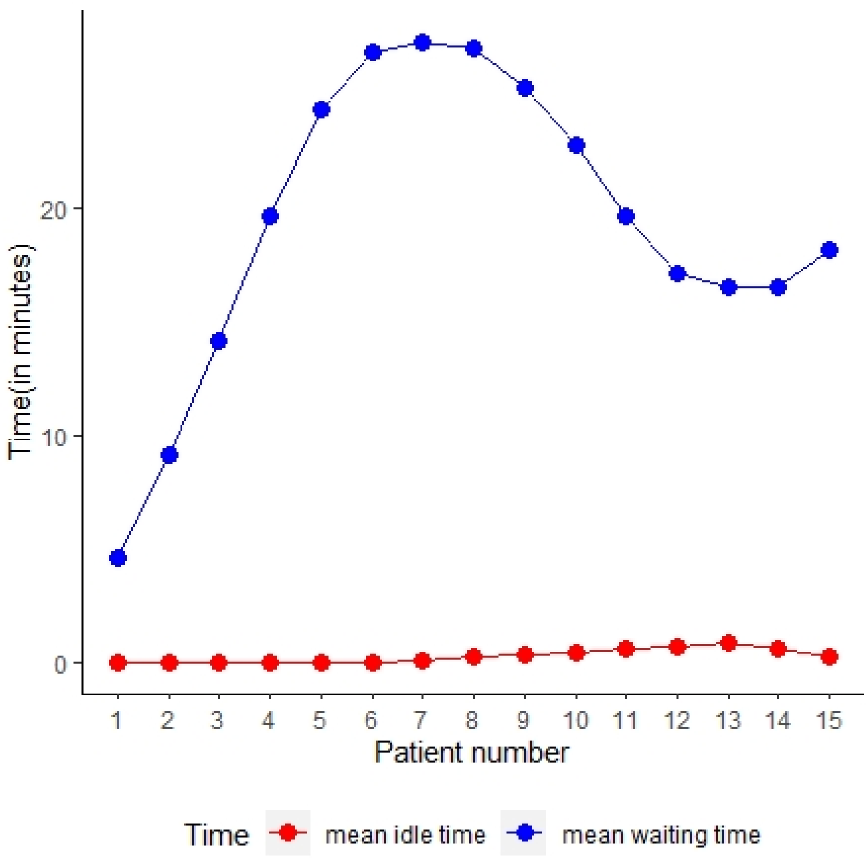


FIGURE 5.1. Mean waiting and idle time plot for 4111 . . .

It can be seen from Table 5.4 that the patient number 7 has a maximum mean waiting time of 27.376 mins and the physician mean idle time is 0.078 mins before the 7th patient, whereas the 1st patient wait on an average for only 4.628 mins. Furthermore, in Figure 5.1 the mean waiting times for patients increase till the 7th patient and then it starts decreasing up to the 14th patient. It again increases for the 15th patient and becomes 18.168 mins. The average waiting times set a dome-shaped trend. Apart from this, the mean idle time of the doctor is less than 1 min every time. The mean of mean waiting times is 19.315 mins and the mean of mean idle time is 0.278 mins. The value of the objective function is \$83.667.

(2) Multiple-block(2222...): Table 5.4 gives the average waiting time for an individual patient and average idle time for the physician.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 3.351 | 0.214 |
| 2 | 7.325 | 0.014 |
| 3 | 13.440 | 0.000 |
| 4 | 16.556 | 0.008 |
| 5 | 19.616 | 0.040 |
| 6 | 21.373 | 0.064 |
| 7 | 21.730 | 0.155 |
| 8 | 21.905 | 0.109 |
| 9 | 20.550 | 0.306 |
| 10 | 19.344 | 0.273 |
| 11 | 17.641 | 0.407 |
| 12 | 16.841 | 0.470 |
| 13 | 17.094 | 0.348 |
| 14 | 18.057 | 0.186 |
| 15 | 20.391 | 0.038 |

TABLE 5.4. Mean waiting time for each patient and mean idle times of physician for 2222...

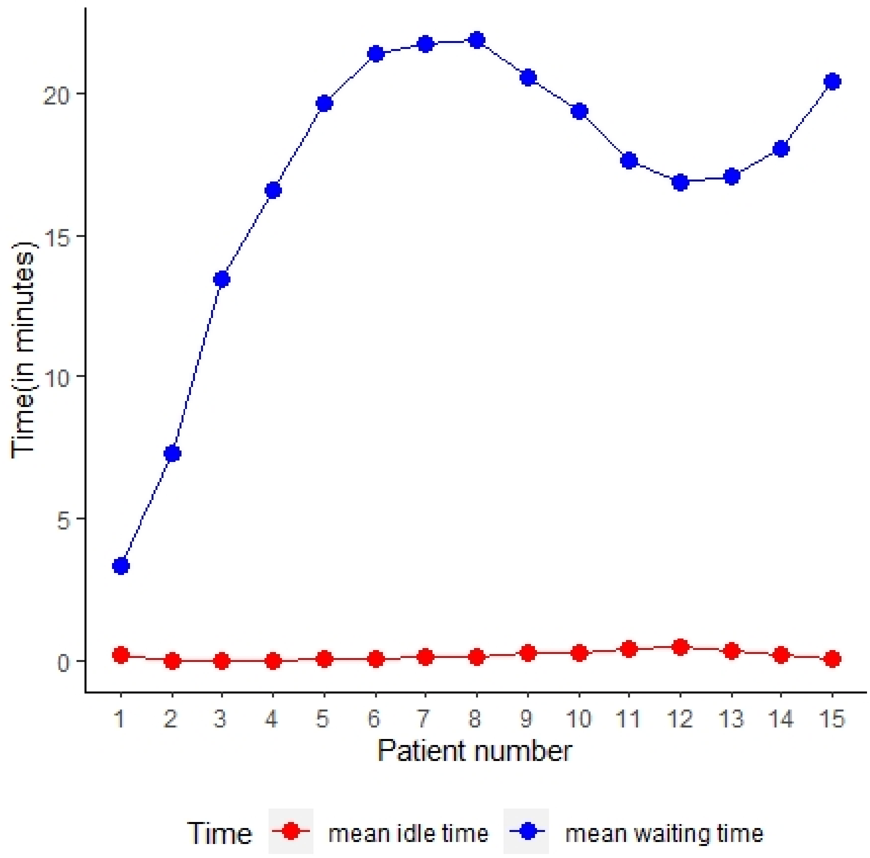


FIGURE 5.2. Mean waiting and idle time plot for 2222...

For strategy 2222... with variable interval length, the mean waiting times are less, relative to those for strategy 4111.... Here, the 8th patient waits a maximum of 21.905 mins. The trend for waiting times is the same as that for individuals with an initial block strategy. The average idle time of the physician is less up to the first 7 patients in strategy 4111... and, from the 8th patient, the idle time of the doctor is less for 2222.... The average of waiting times for each patient is 17.014 mins and the mean of mean idle times is 0.175 mins. The objective function is $69.99 for this strategy which is less than that of 4111.... Thus, out of these two strategies discussed so far, 2222... is much better than 4111....

(3) Multiple-block with an initial block(4222...): For results of this strategy, I refer Table 5.5.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.813 | 0.025 |
| 2 | 9.476 | 0.000 |
| 3 | 14.405 | 0.000 |
| 4 | 20.374 | 0.000 |
| 5 | 26.572 | 0.001 |
| 6 | 29.593 | 0.000 |
| 7 | 31.556 | 0.144 |
| 8 | 32.237 | 0.153 |
| 9 | 30.849 | 0.185 |
| 10 | 27.975 | 0.782 |
| 11 | 23.819 | 1.000 |
| 12 | 20.264 | 1.254 |
| 13 | 19.028 | 1.078 |
| 14 | 18.782 | 1.062 |
| 15 | 20.683 | 0.374 |

Table 5.5. Mean waiting time for each patient and mean idle times of physician for 4222...

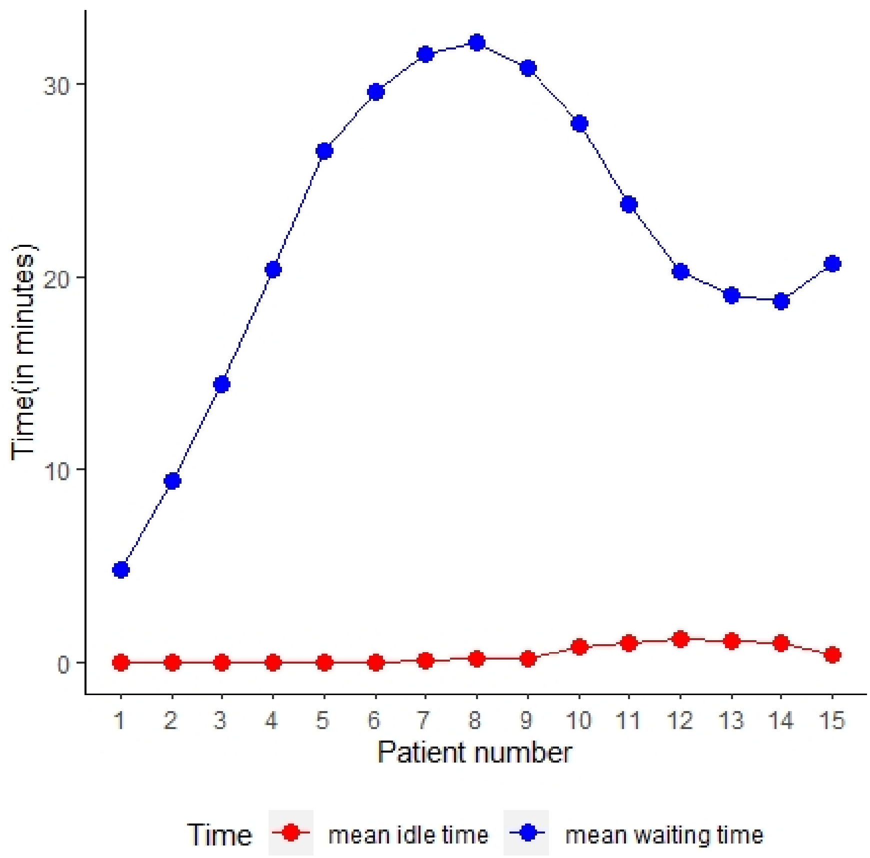


Figure 5.3. Mean waiting and idle time plot for 4222...

From Table 5.5, on average the 8th patient wait maximum i.e for 32.237 mins. The waiting times increase towards the middle and become maximum then followed by a decreasing trend. Same as in the above 2 strategies discussed the waiting time for the 15th patient again increases. The mean idle time of the physician increases towards the end till the 14th patient. The objective function has a value of \$100.058 which is bigger than in the previous two strategies.

(4) Variable block(41313...): The mean waiting time for each patient and mean idle time of the physician before the start of the service of each patient are given in Table 5.6.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.563 | 0.039 |
| 2 | 9.155 | 0.000 |
| 3 | 13.829 | 0.000 |
| 4 | 18.253 | 0.000 |
| 5 | 21.587 | 0.026 |
| 6 | 23.975 | 0.000 |
| 7 | 25.977 | 0.055 |
| 8 | 25.269 | 0.140 |
| 9 | 23.740 | 0.295 |
| 10 | 23.004 | 0.415 |
| 11 | 22.107 | 0.325 |
| 12 | 19.160 | 0.458 |
| 13 | 18.498 | 0.430 |
| 14 | 20.119 | 0.230 |
| 15 | 21.321 | 0.143 |

TABLE 5.6. Mean waiting time for each patient and mean idle times of physician for 41313...

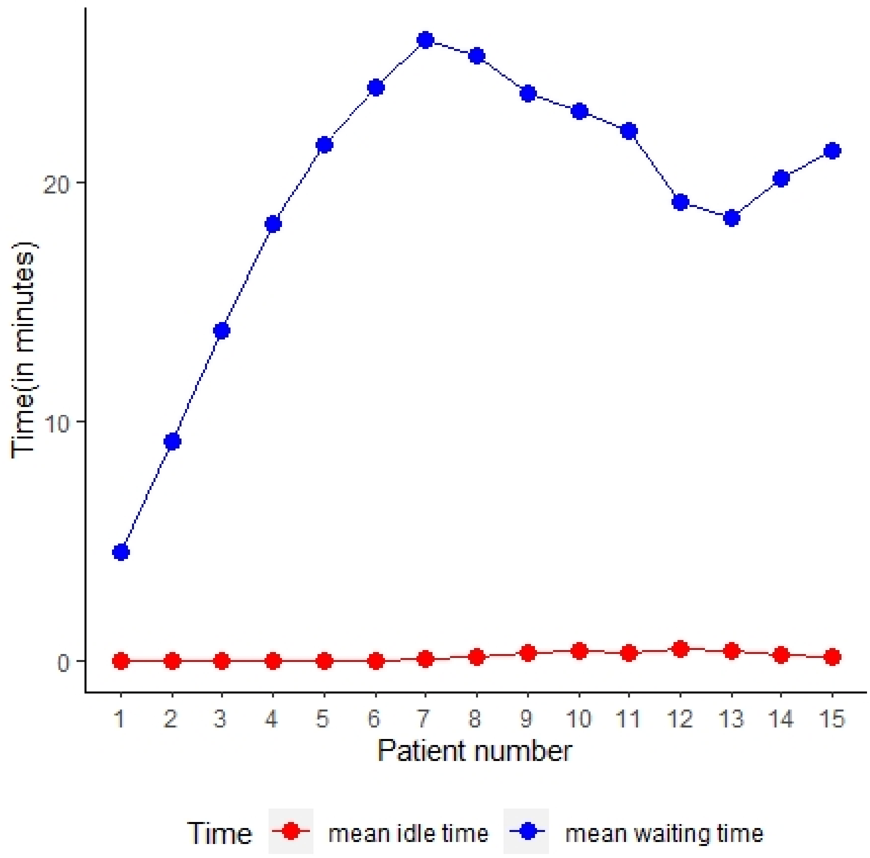


FIGURE 5.4. Mean waiting and idle time plot for 41313...

In this strategy, the 7th patient has a maximum mean wait time of 25.977 mins. Again for this strategy, the mean wait times form a dome-shape curve. The curve for mean idle times is almost a straight line about 0 min. For the 15th patient, the mean wait time increases in each of the above strategies. Here, the objective function takes values of \$78.113.

(5) Combination of Variable block and individual block(412311...): The mean waiting time for each patient and mean idle time of the physician before the start of the service of each patient are given in Table 5.7.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.607 | 0.022 |
| 2 | 9.210 | 0.000 |
| 3 | 14.150 | 0.000 |
| 4 | 19.954 | 0.000 |
| 5 | 25.951 | 0.000 |
| 6 | 29.952 | 0.000 |
| 7 | 33.743 | 0.070 |
| 8 | 37.371 | 0.167 |
| 9 | 38.817 | 0.238 |
| 10 | 38.348 | 0.376 |
| 11 | 34.670 | 0.604 |
| 12 | 29.926 | 0.630 |
| 13 | 25.812 | 0.659 |
| 14 | 23.067 | 0.695 |
| 15 | 23.170 | 0.572 |

TABLE 5.7. Mean waiting time for each patient and mean idle times of physician for 412311 . . .

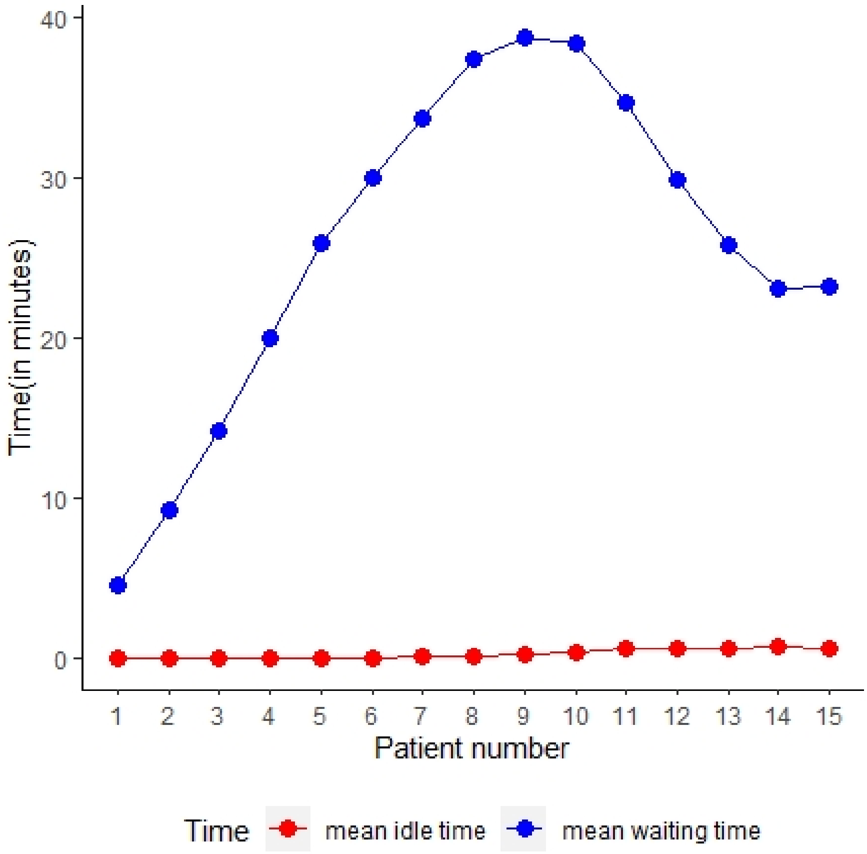


FIGURE 5.5. Mean waiting and idle time plot for 412311 . . .

In Table 5.7, the maximum mean wait is of 38.817 mins which is for the 9th patient. The value of the total cost is \$106.695 which is the maximum till now. It means this strategy doesn't do much better when the interval length is variable. The red line in the plot represents the mean idle time of the physician. For this strategy, the maximum mean idle time is 0.695 mins which is before the service of the 14th patient.

The results for strategies 421311... and 431211... are almost same as that for 412311.... These results are given in Table 5.8 and Table 5.9.

(6) Combination of Variable block and individual block(421311...): The

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.616 | 0.038 |
| 2 | 9.331 | 0.000 |
| 3 | 14.582 | 0.000 |
| 4 | 20.242 | 0.000 |
| 5 | 26.414 | 0.000 |
| 6 | 29.713 | 0.000 |
| 7 | 33.065 | 0.000 |
| 8 | 36.066 | 0.040 |
| 9 | 37.352 | 0.156 |
| 10 | 36.095 | 0.183 |
| 11 | 32.056 | 0.519 |
| 12 | 27.295 | 0.841 |
| 13 | 23.026 | 1.127 |
| 14 | 20.518 | 0.855 |
| 15 | 19.683 | 0.780 |

TABLE 5.8. Mean waiting time for each patient and mean idle times of physician for 421311...

objective function has a value of \$104.058. The 9th patient has a maximum mean waiting time of 37.352 mins. The maximum mean idle time is 1.127 mins before the start of the service of the 13th patient.

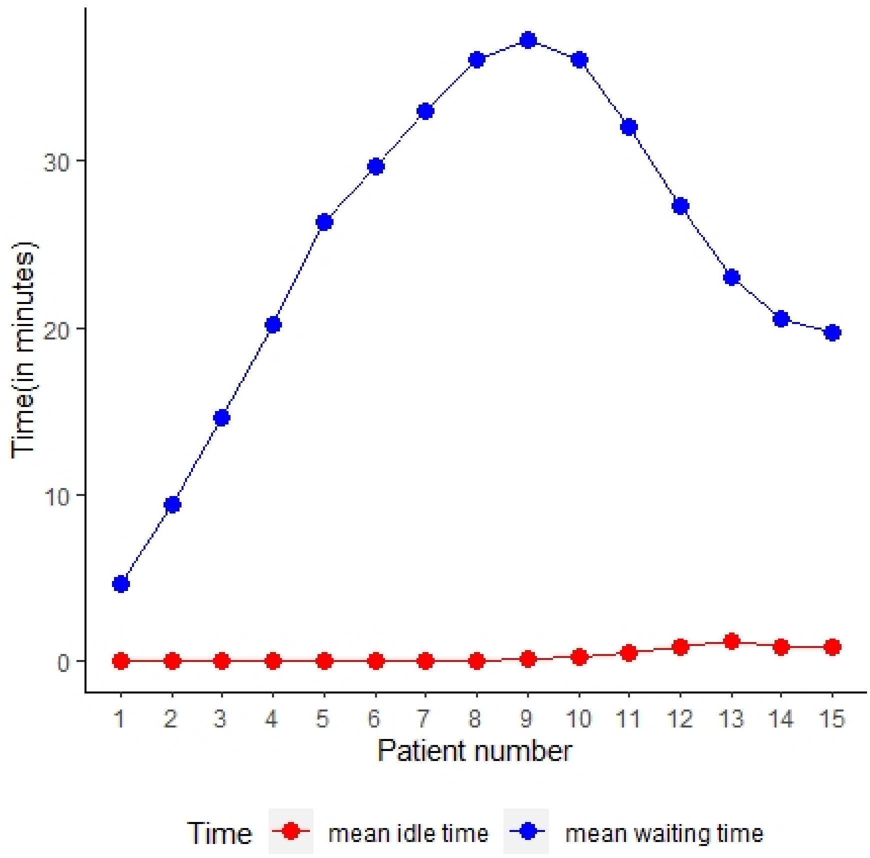


FIGURE 5.6. Mean waiting and idle time plot for 421311 . . .

(7) Combination of Variable block and individual block(431211 . . . ): The mean waiting time for each patient and mean idle time of the physician before the start of the service of each patient are given in Table 5.9.

| Patient number | mean wait time | mean idle time |
|---|---|---|
| 1 | 4.730 | 0.033 |
| 2 | 9.494 | 0.000 |
| 3 | 14.677 | 0.000 |
| 4 | 21.364 | 0.000 |
| 5 | 28.867 | 0.000 |
| 6 | 33.178 | 0.000 |
| 7 | 37.296 | 0.000 |
| 8 | 40.307 | 0.006 |
| 9 | 41.553 | 0.068 |
| 10 | 40.121 | 0.220 |
| 11 | 36.045 | 0.525 |
| 12 | 30.513 | 0.799 |
| 13 | 25.733 | 0.901 |
| 14 | 22.545 | 1.052 |
| 15 | 22.340 | 0.542 |

TABLE 5.9. Mean waiting time for each patient and mean idle times of physician for 431211...

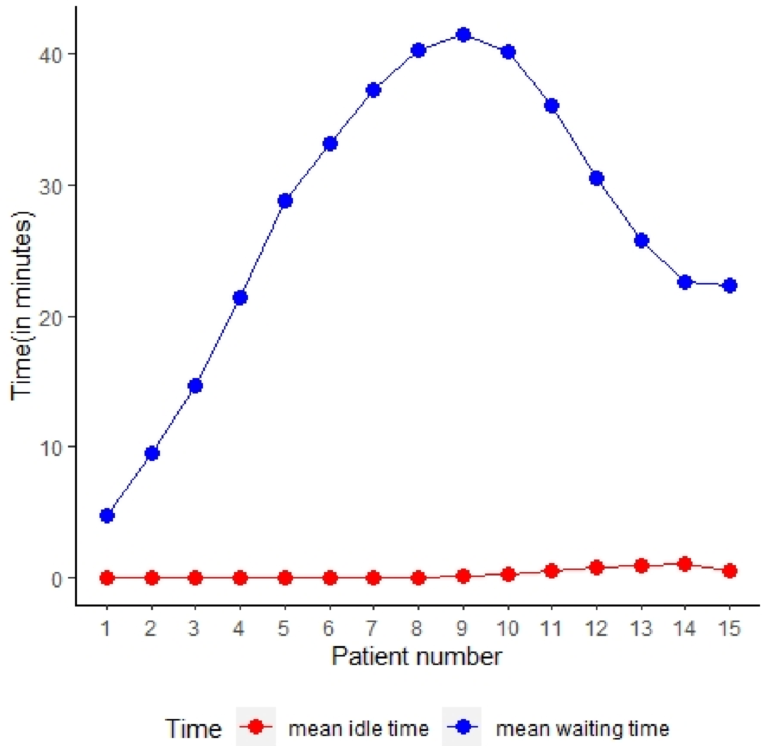


FIGURE 5.7. Mean waiting and idle time plot for 431211...

Out of all the strategies discussed above this strategy has big mean waiting times. The maximum value of the mean wait time is 41.553 mins for the 9th patient. The reason behind this big number is that in the beginning 3 time-slots, a block of 4, then 3 followed by 2 patients are appointed respectively. Due to this, the waiting times are much bigger than other strategies. The maximum mean idle time is 1.052 mins before the service of 14th patients. The total cost is $111.851

Figure 5.8 gives the curves of mean waiting times for all the above-mentioned strategies. In terms of mean waiting times, from all the curves in Figure 5.8 the

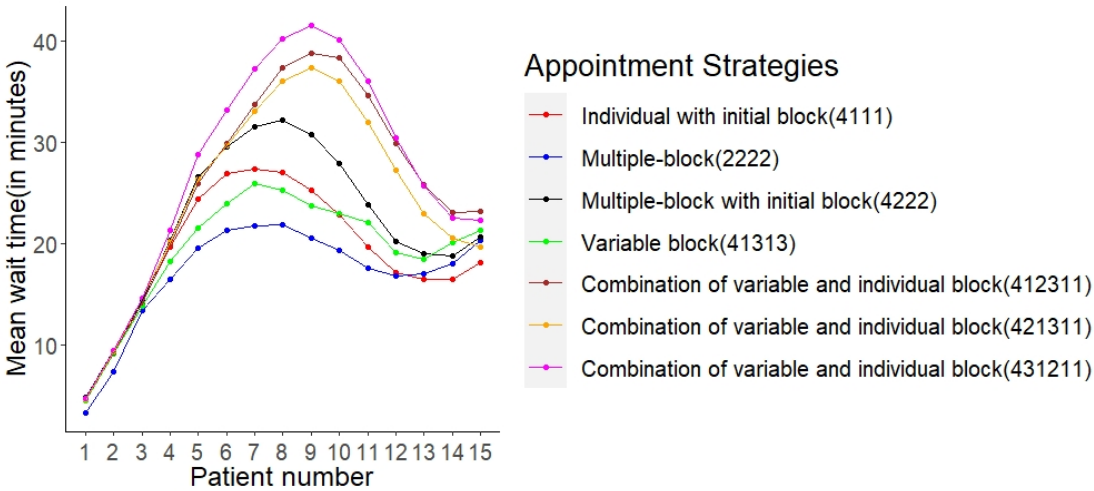


FIGURE 5.8. Mean waiting time plot for all strategies

curve with the blue colour is the best. It means each patient in the multiple-block strategy waits a minimum as compare to all other strategies plotted. The pink curve has the maximum waiting times for patients that implies the strategy 431211 . . . is worst in terms of waiting times. Since each patient has large waiting times. Moreover, the orange and brown curves are approximately the same that represent the combination of variable and individual blocks strategies 421311 . . . and 412311 . . . respectively. The mean waiting time of the 15th patient is less than the 14th patient only in strategy 421311 . . . whereas in all other strategies it is bigger than 14th patient.

The values of objective function for all strategies are given in Table 5.10.

| strategy | $f_0^*$(in\$) |
|---|---|
| 41111 . . . | 83.667 |
| 22222 . . . | 69.990 |
| 42222 . . . | 100.058 |
| 41313 . . . | 78.113 |
| 412311 . . . | 106.695 |
| 421311 . . . | 104.058 |
| 431211 . . . | 111.851 |

TABLE 5.10. Objective function i.e total cost for different strategies

In above table, the minimum value of the objective function is \$69.99 for strategy 2222 . . . and maximum for strategy 431211 . . . which is \$111.851. Thus on the basis of minimum total cost of the system, multiple-block strategy with 2 patients for each block is the best under our assumptions for 15 patients.
The best strategy depends on the probability of missing an appointment. It can be different for different assumptions and different probabilities.
When physician is 5 mins late, the results for different strategies and the plots are given as:

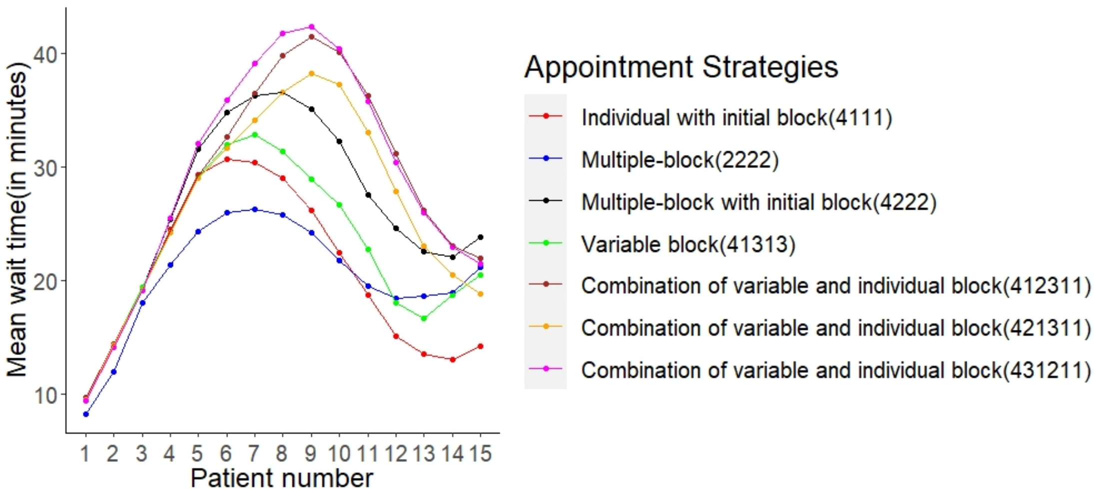


FIGURE 5.9. Plot of mean waiting time for each patient for all strategies, when physician is late

The results are similar to the case when physician is on time. With the lateness of the physician, the wait time increases for the patients. However, the idle times of the physician get lower. On comparing the red curve for strategy 411 . . . and the blue curve for strategy 222 . . . in Figure 5.9, the mean waiting times for first 10 patients are less for the blue curve than in the red curve whereas for last five patients the mean waits increase for blue curve and decrease for red curve. It seems blue curve is much better than red one. But this is not the final decision as these curves are only for mean waiting times. Each patient except the 15th patient waits maximum in the strategy 431211..., the pink curve in the plot. The strategies 412311... and 421311... perform almost the same represent by brown and orange curves respectively.

| strategy | $f_0^*$ (in \$) |
|---|---|
| 41111... | 97.065 |
| 22222... | 80.130 |
| 42222... | 108.405 |
| 41313... | 103.635 |
| 412311... | 112.965 |
| 421311... | 110.160 |
| 431211... | 117.570 |

TABLE 5.11. Objective function for different strategies, when physician is 5 mins late

The objective function has minimum value of \$80.130 for the strategy 222... and maximum for the strategy 431211... which is \$117.570. We can see in Table 5.11 that out of the individual block with initial block strategy (411...) and the multiple block strategy 222..., the second one performs better. In the case of the physician being late, on the basis of the minimum value of the objective function, the multiple block strategy 222... is the best out of all the mentioned strategies.

CHAPTER 6

# Conclusions and Future Questions

## 6.1. Conclusions

In this thesis, we gave the probability distribution of the number of patients or customers in the system by four different methods. Firstly, by theoretical calculations using convolutions of discrete random variables, secondly by performing simulation followed by third method using probability generating functions and the last method is Markov chains. We analyse that, if $n_0, n_1, n_2, \ldots$ patients are scheduled at times $0, 1d, 2d, \ldots$ respectively with probability of keeping an appointment $p$ and $X_0, X_1, X_2, \ldots$ are the number of patients remaining in the system at times $0^+, 1d^+, 2d^+ \ldots$ respectively then $X_0 \sim Bin(n_0, p)$ and $P(X_i = x) = Bin(\sum_{n=n_0}^{n_i} n, p, x+i)$ for upper $n_0 - 1$ numbers from the support of the random variables, where $i = 1, 2, 3 \ldots$.
Furthermore, we conclude that if the probability of missing an appointment is 0.15, then in the case of the physician being either punctual or 5 mins late, the individual block with an initial block strategy $(4111\ldots)$ is the best strategy on the basis of the minimum value of the total expected cost of the system, providing patients are unpunctual, fixed interval lengths and service times are varying for different patients.
Moreover, in the case of variable interval lengths under the same assumptions as in the case of fixed interval lengths, the multiple-block strategy $(222\ldots)$ performs best out of all the discussed strategies. Thus, in case of physician being punctual, one can use scheduling times (in mins) as,

```
> sct
[1]   0.000000   2.983000   8.921605  17.805606  29.628404  44.385105  62.071820
[8]  76.828522  88.651320  97.535321 103.473926 106.456926
```

for the best strategy $222\ldots$ where everytime 2 patients are given the same scheduling time. In case of physician being 5 mins late, the best pair is $(1.915, 1.062)$.

## 6.2. Future research

How good is our objective function? is the future question.
In this thesis, we have not considered the priority for the patients in the system. Some patients may have higher priorities than others, because their conditions may be more serious or they have special time constraints or they may be more important in some sense. If we consider priorities in our analysis the solution will be considerably more complex.
The initial system that was considered in this thesis had deterministic interarrival

times, with batch arrivals and deterministic service times. But since the batch size could be zero, this effectively means that the interarrival times are geometrically distributed. Since the deterministic service times can be represented by a discrete time Phase (PH) distribution, the system can be viewed as a special case of a discrete time $Geo^{[X]}/PH/1$ queueing system (Alfa [1]). A deeper study of this would be a subject of future research.

In this thesis, we always assumed there would be 15 customers who completed service. However, because patients miss appointments the actual number of customers served per session may vary. So we may wish to study the total number of customers served per session and how this affects our objective function.

# Bibliography

[1] Alfa, A. (2020). *Comment on Discrete Time Queues*, Personal Communication.

[2] Bailey, N. (1952). *A study of Queues and Appointment Systems in Hospital Outpatient Departments with Special Reference to Waiting Times*, Journal of the Royal Statistical Society 14, 185 – 199.

[3] Blanco White, MJ and Pike, MC (1964). *Appointment systems in out-patients' clinics and the effect of patients' unpunctuality*, Medical Care, 133 – 145.

[4] Brailsford, Sally and Vissers, Jan (2011). European journal of operational research 212(2), 223 – 234.

[5] Campbell, Jennifer D and Chez, Ronald A and Queen, Tina and Barcelo, Annette and Patron, Ellen (2000). *The no-show rate in a high-risk obstetric clinic*, Journal of women's health & gender-based medicine, 9(8), 891 – 895.

[6] Cayirli, Tugba and Veral, Emre(2003). *Outpatient scheduling in health care: a review of literature*, Production and operations management, 12(4), 519 – 549.

[7] Cayirli, Tugba and Veral, Emre and Rosen, Harry (2006). *Designing appointment scheduling systems for ambulatory care services*, Health care management science, 9(1), 47 – 58.

[8] Cayirli, Tugba and Veral, Emre and Rosen, Harry (2008). *Assessment of patient classification in appointment system design*, Production and Operations Management 17(3), 338 – 353.

[9] Cayirli, Tugba and Yang, Kum Khiong and Quek, Ser Aik (2012). *A universal appointment rule in the presence of no-shows and walk-ins*, Production and Operations Management 21(4), 682 – 697.

[10] Fetter, Robert B and Thompson, John D (1966). *Patients' waiting time and doctors' idle time in the outpatient setting*, Health services research, 1(1), 66.

[11] Ho, Chrwan-Jyh and Lau, Hon-Shiang (1992). *Minimizing total cost in scheduling outpatient appointments*, Management science 38(12), 1750 – 1764.

[12] Huang, Xiao-Ming (1994)' *Patient attitude towards waiting in an outpatient clinic and its applications*, Health Services Management Research, 7(1), 2 – 8.

[13] Kaandorp, Guido C and Koole, Ger (2007). *Optimal outpatient appointment scheduling*, Health care management science, 10(3), 217 – 229.

[14] Katz, K and Larson, B and Larson, R (2003). *Prescription for the waiting-in-line blues entertain, enlighten, and engage*, Oper Manag Crit Perspect Bus

Manag, 2, 160.

[15] Keller, TF and Laughhunn, DJ (1973). *An application of queuing theory to a congestion problem in an outpatient clinic*, Decision Sciences 4(3), 379 – 394.

[16] Kirkpatrick, Scott and Gelatt, C Daniel and Vecchi, Mario P (1983). *Optimization by simulated annealing*, science 220(4598), 671 – 680.

[17] Kuiper, Alex and Kemper, Benjamin and Mandjes, Michel (2015). *A computational approach to optimized appointment scheduling*, Queueing Systems 79(1), 5 – 36.

[18] Leiba, Adi and Weiss, Yuval and Carroll, Judith S and Benedek, Paul and Bardayan, Yaron (2002). *Waiting time is a major predictor of patient satisfaction in a primary military clinic*, Military medicine, 167(10), 842 – 845.

[19] Lindley, D.V. (1952). *The Theory of Queues with a Single Server*, Proceedings Cambridge Philosophy Society, 48, 277 – 289.

[20] Markov, Andrey Andreyevich (1906). *Extension of the law of large numbers to dependent quantities*, Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser) 15, 135 – 156.

[21] Noon, C. E., Hankins, C. T., & Cote, M. J. (2003). *Understanding the impact of variation in the delivery of healthcare services/practitioner application*, Journal of Healthcare Management 48(2), 82 – 98.

[22] Rising, Edward J and Baron, Robert and Averill, Barry (1973). *A systems analysis of a university-health-service outpatient clinic*, Operations Research 21(5), 1030 – 1047.

[23] Robinson, Lawrence W and Chen, Rachel R (2003). *Scheduling doctors' appointments: optimal and empirically-based heuristic policies*, Iie Transactions, 35(3), 295 – 307.

[24] Rohleder, T.R., Lewkonia, P., Bischak, D.P., Duffy, P., & Hendijani, R. (2011). *Using simulation modeling to improve patient flow at an outpatient orthopedic clinic*, Health care management science 14(2), 135 – 145.

[25] Rosenblatt, Murray and others (1956) *Remarks on Some Nonparametric Estimates of a Density Function*, The Annals of Mathematical Statistics 27(3), 832 – 837.

[26] Rust, Cynthia T and Gallups, Nancy H and Clark, W Scott and Jones, Donna S and Wilcox, Wallace D (1995). *Patient appointment failures in pediatric resident continuity clinics*, Archives of pediatrics & adolescent medicine, 149(6), 693 – 695.

[27] Samorani, Michele and LaGanga, Linda R (2015). *Outpatient appointment scheduling given individual day-dependent no-show predictions*, European Journal of Operational Research 240(1), 245 – 257.

[28] Seneta, Eugene (2006). *Markov and the creation of Markov chains*, Citeseer, $1-20$.

[29] Serfozo, Richard (2009). *Basics of applied stochastic processes.*

[30] Sharp, Deborah J and Hamilton, William (2001). *Non-attendance at general practices and outpatient clinics: Local systems are needed to address local problems*, British Medical Journal, 323(7321), $1081-1082$.

[31] Soriano, Alfonso (1966). *Comparison of two scheduling systems*, Operations Research, 14(3), $388-397$.

[32] Swisher, James R and Jacobson, Sheldon H and Jun, J Brian and Balci, Osman (2001). *Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation*, Computers & operations research, 28(2), $105-125$.

[33] Vissers, J (1979). *Selecting a suitable appointment system in an outpatient setting*, Medical Care, 17(12), $1207-1220$.

[34] Walter, SD (1973). *A comparison of appointment schedules in a hospital radiology department*, British journal of preventive & social medicine, 27(3), 160.

[35] Wang, P Patrick (1997). *Optimally scheduling N customer arrival times for a single-server system*, Computers & Operations Research, 24(8), $703-716$.

# Appendix A

## Codes and Programs

This section shows the R programming code and commands used in the study:

```
# Individual block rule with initial block
library(ggplot2)
library(dplyr)
library(tidyr)

# n is a vector of number of patients arrived at different time-slots
n=rbinom(1,4,0.85)
while(sum(n)<=15){
n=c(n,rbinom(1,1,0.85))
}

# AT is the vector of arrival times of the patients
AT=rep(-2,n[1])
for ( i in 1:(length(n)-1)){AT=c(AT, rep(-2+7*i,n[i+1]))}
u = -5+10*runif(n[1])
for (i in 1:(length(n)-1)){u=c(u,-5+10*runif(n[i+1]))}
D = data.frame(u,AT)
D = D[order(D$AT,D$u),]
D
AT = AT+(D$u)
AT=AT[1:15]

# ScT is the vector of the scheduling times of the patients
ScT=rep(0,n[1])
for ( i in 1:(length(n)-1)){ScT=c(ScT, rep(0+7*i,n[i+1]))}
ScT=ScT[1:15]

# ST is the vector of service times for 15 patients
ST = 6+2*runif(length(AT))
ST=ST[1:15]

# PA is the physician arrival time
PA=0

# PST is the vector of physician start times of the service
pst = function(i){
if (i==1) return(pst=max(AT[1],PA))
```

```
else if( pst(i-1)+ST[i-1]>AT[i]) return(pst=pst(i-1)+ST[i-1])
else return(pst = AT[i])
}
PST=sapply(1:length(AT), pst)
PST=PST[1:15]

# PET is the vector of the physician end times of the service
pet = function(i){
pet=pst(i) + ST[i]
return(pet)
}
PET=sapply(1:length(AT), pet)
PET=PET[1:15]

# W is the vector of the waiting times of the patients
w = function(i){
w=max(pst(i) - AT[i],0)
return(w)
}
W=sapply(1:length(AT), w)
W=W[1:15]

# PEmT is the vector of empty time of the physician before startig the service of a patient
pemt = function(i){

if (i==1) return(pemt = max(pst(1)-PA,0))
else return(pst(i) - pet(i-1))
}
PEmT=sapply(1:length(AT), pemt)
PEmT=PEmT[1:15]
x=1:15
df=round(data.frame(x, ScT, AT, ST,PST,PET,W,PEmT),3)

df %>% select(c(x,W,PEmT)) %>% pivot_longer(-x) %>%
ggplot(aes(x=x,y=value, color=name,shape=name,group=name,fill=name))+
geom_step(color='black',show.legend = F)+
geom_point(size=2)+
theme(axis.text.x = element_text( vjust = 0.5),
legend.position = "bottom",
axis.line = element_line(),
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
panel.background = element_blank(),
panel.border = element_blank())+
ylab("Time(in minutes)")+
scale_x_continuous("Patient number", labels = x, breaks = x)+
scale_color_manual('Time',values=c('red','blue'),
labels = c("Idle time","Waiting time"))+
```

```
scale_shape_manual('Time',values=c(21,21),
labels = c("Idle time","Waiting time"))+
scale_fill_manual('Time',values=c('transparent','blue'),
labels = c("Idle time","Waiting time"))

##############################################################################
#Strategy 411...
func = function(a,b){
L = list()
l = list()
for (i in 1:100) {
#nitr is iteration number
nitr=i
n=rbinom(1,4,0.85)
while(sum(n)<=15){
n=c(n,rbinom(1,1,0.85))
}
s_1 = seq(from=1, to=as.integer((length(n))/2), by=1)
s_2 = seq(from=as.integer((length(n))/2)+1, to=1, by=-1)
sq = c(s_1,s_2)
new_seq = a*sq^(b)
s = sq+new_seq
c = which(n!=0)

#Scheduling times
sct=c(0,cumsum(s))
ScT=rep(sct[c],n[c])
ScT=ScT[1:15]

#Arrival times
AT = function(i){
if(i==1)  return(AT =-2)
else   return(AT(i-1)+s[i-1])
}
AT=sapply(1:length(n),AT)
AT=rep(AT[c],n[c])
u = -5+10*runif(n[1])
for (i in 1:(length(n)-1)){u=c(u,-5+10*runif(n[i+1]))}
D = data.frame(u,AT)
D = D[order(D$AT,D$u),]
AT = AT+(D$u)
AT=AT[1:15]

#Service times
ST = 6+2*runif(length(AT))
ST=ST[1:15]

#Physician arrival time
```

```
PA=0

#Physician start time
pst = function(i){
if (i==1) return(pst=max(AT[1],PA))
else if( pst(i-1)+ST[i-1]>AT[i]) return(pst=pst(i-1)+ST[i-1])
else return(pst = AT[i])
}
PST=sapply(1:length(AT), pst)
PST=PST[1:15]

#Physician end time
pet = function(i){
pet=pst(i) + ST[i]
return(pet)
}
PET=sapply(1:length(AT), pet)
PET=PET[1:15]

#Waiting times of the patients
w = function(i){
w=max(pst(i) - AT[i],0)
return(w)
}
W1=sapply(1:length(AT), w)
W1=W1[1:15]
L[[nitr]] = W1

#Physician empty time
pemt = function(i){
if (i==1) return(pemt = max(pst(1)-PA,0))
else return(pst(i) - pet(i-1))
}
PEmT=sapply(1:length(AT), pemt)
PEmT=PEmT[1:15]
l[[nitr]] = PEmT
}
v = matrix(unlist(L), nrow = length(W1), ncol = nitr)
t = matrix(unlist(l), nrow = length(PEmT), ncol = nitr)

# Vector of mean wait times for each patient
M =rowMeans(v)

# Vector of mean idle time for physician before each patient
m =rowMeans(t)
PW=mean(M)
DW=mean(m)
```

```
#Objective function
f=function(PW,DW){
f=14.25*15*(PW+15*DW)/60
}
f(PW,DW)
}

#Modified Simulation Annealing
g=function(func,a0,b0,n){
for (i in 1:n) {
k1=rbinom(1,1,0.1)
k2=rbinom(1,1,0.1)
a1=max(a0+k1*runif(1,-1,1)+runif(1,-0.1,0.1),0.1)
b1=b0+k2*runif(1,-1,1)+runif(1,-0.1,0.1)
f_1=func(a1,b1)
f_0=func(a0,b0)
if(f_1<f_0+.05*f_0*((a1-a0)^2+(b1-b0)^2>.03)*(i<n/2)){
a0=a1
b0=b1
print(round(c(f_1,a0,b0),3))
}
else{
print(round(c(f_0,a0,b0),3))
}
}
}

g(func,0.5,0.5,100)
```

# Vita Auctoris

Ms. Ranveer Kaur was born in 1996 in Sangrur, Punjab, India. She graduated from National Institute of Technology, Jalandhar, India in 2018 with a Master of Science degree in Mathematics. For further education, she came to Canada. She is receiving the Master of Science degree in Statistics from the University of Windsor.