Preface

The first three chapters review all prerequisites from applied probability needed for a basic course in queueing systems. Yet, this book does not suit beginners in applied probability. It is recommended only to those who have taken a course that dealt at length with (discrete-and continuous-time) Markov processes, the Poisson process, and some basics in renewal theory. On the other hand, the mathematical level required here is not that advanced. There is no need for any knowledge in advanced probability or measure theory, or any skills in differential equations. There is only one instance in the last chapter where difference equations are utilized. Thus, this text suits an advanced undergraduate or first-year graduate program in operations research, statistics, computer science, electrical engineering, or industrial engineering.

A note on notation: I use the standard Kendall notation. Thus, service stations, i.e. stations that possess their own arrival process for customers (or jobs), are treated in isolation. The arrival process is always assumed to be a renewal process, as is the service process in each server (looked at when the server under consideration is busy). All servers are identical and service is provided on a first-come first-served basis. All random variables involved are independent. The notation contains four fields: $F_1/F_2/n/m$. The first entry denotes the family of distributions for which a result holds. The letter "G" (for "general") is used when any (continuous) distribution is assumed. The letter "M" (short for "Markov" or "memoryless") is used when exponential distributions are assumed. The same holds for the second field, yet now with regard to the service processes. The third field states the number of servers while the fourth states the maximum content of the system (including those who are served). The default value is an unlimited buffer so that when the fourth field is omitted, it means that no bound on the queue length is assumed.

Why do we need another book on queues? The main reason is the order of the chapters. Usually, the M/M/1 model is the first to be introduced and analyzed in detail. Here this is done only in Chap. 8 and as part of a more general treatment of memoryless queueing models. I tried to begin with what I think are the most important topics, while giving the students enough tools to start building themselves up as queueing theorists. I find M/G/1 to be *the* single most important model (and not M/M/1). The reason is twofold: first, the impact of the variability of service time is too implicit in the M/M/1 model. This variability is highlighted in the M/G/1 model (via the Khintchine–Pollaczek formula). Moreover, this result can be proved with basics in renewal theory and the introduction of the PASTA phenomenon. Second, in M/M/1 the residual service time of a customer in service and the queue length are independent. This is not the case in the M/G/1 queue (a phenomenon sometimes overlooked). In M/G/1 the connection between waiting times and queue length upon arrival is more involved. At the same time I tried to include in the book all material that I find important for anyone who considers a career in queuing theory.

The book is organized as follows. The first three chapters state many needed preliminaries to a study of exponential distributions, the Poisson process and generating functions (Chap. 1), renewal theory (Chap. 2), and Markov chains (Chap. 3). I do not claim to be comprehensive here and the content of these chapters is more for those who have seen it before in an applied or introductory course in probability than for those who are seeing it here for the first time. At the same time, it is recommended that you read through these three chapters as it is possible that some of their content was not covered (or was covered differently) in an earlier course. This is in particular true regarding Chap. 2 where the point of departure is an axiomatic statement of the length bias distribution and the distribution of the age given total longevity.

Most texts would switch here to first-come first-served memoryless queues, commencing with the M/M/1 model. I find this model too simple and some of the key phenomena associated with queues not sufficiently transparent. Thus, after dealing with the queueing property in Chap. 4, I discuss the M/G/1 model. Finding mean values for this model is quite straightforward and basically all that is needed here is the understanding of the Poisson process and results from renewal theory. Chapter 5 deals with priority queues and here, too, in spite of the supposedly advanced model assumed, no further prerequisites are needed in order to derive various mean values.

Chapter 6 examines the distributions of the queue length, of the waiting time and of the length of the busy period of the M/G/1 model. Chapter 7 does the same for the G/M/1 model. Only in Chap. 8 memoryless queues are introduced as a model in their own right (and not merely as special cases of the M/G/1 and G/M/1 models). I start by defining continuous time Markov processes and state 13 examples of such models, all dealing with various queueing models. I then define the limit probabilities and exemplify how they can be found using the balance equations. The concepts of time-reversed Markov processes and time-reversible Markov processes are introduced and utilized extensively.

Chapters 9 and 10 present two related important queueing models, namely, open and closed networks of memoryless queues. The product form limit probabilities are derived, customers' processes are considered, and conditional and unconditional mean waiting times are computed.

Chapter 11 deals with queueing regimes that lead to many of their parameters being insensitive, namely, those that are functions of the service times only through their mean value. Chapter 12 concludes with two-dimensional queueing models that are in fact quasi-birth-and-death processes. Such models lead to the powerful computation method of matrix-geometric. This technique is introduced and exemplified in various queueing models, among them parallel two-server queues.

I would like to thank all those without whose help this book would not have appeared, or at least would have been completely different and far worse. Offer Kella through a large number of discussions and e-mails taught me a lot. Much in Chaps. 2 and 5 I learned from him. With Yoav Kerner I consulted on too many points to be mentioned. Binyamin Oz made several inportant contributions. Moshe Zukerman kindly let me look into his electronic text from which I learned much. Ishai Menache read this text and made an enormous number of comments. I also got much technical assistance from David Azriel, Irena Kaplan, Uri Yovel, Kathy Huang and Maxim Bocharenko. Special thanks are due to Michael Borns who edited the book. I would like to thank the Discipline of Business Analytics at the University of Sydney, where much of this work was composed during a sabbatical. Finally, I would like to acknowledge financial assistance made through Israel Science Foundation grant no. 237/02.

Jerusalem, Israel Moshe Haviv

Springer

http://www.springer.com/978-1-4614-6764-9

Queues
A Course in Queueing Theory
Haviv, M.
2013, XIV, 221 p., Hardcover
ISBN: 978-1-4614-6764-9