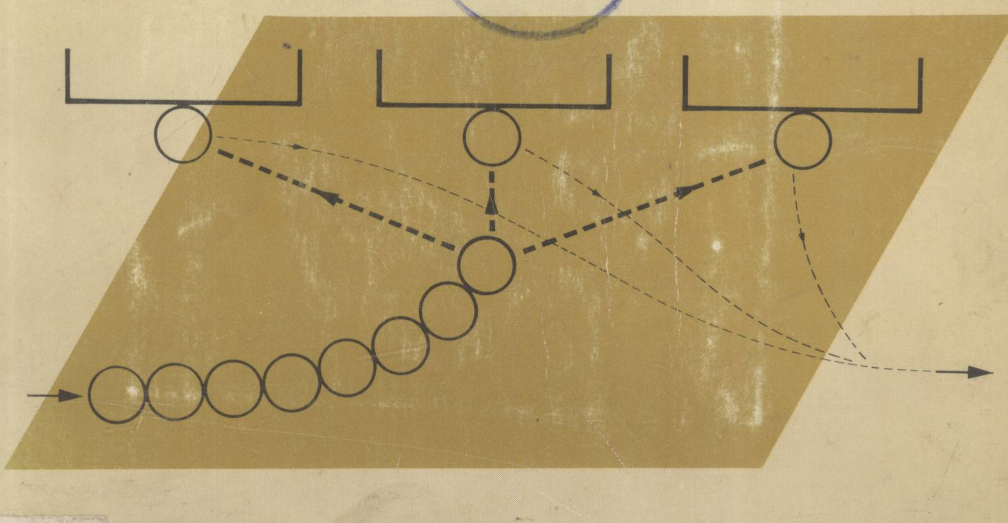
Queueing Theory

WORKED EXAMPLES AND PROBLEMS

J Murdoch



QUEUEING THEORY

Worked examples and problems

J. Murdoch

Head of Statistics and Operational Research Unit,
School of Production Studies,
Cranfield Institute of Technology

M

First published 1978 by
THE MACMILLAN PRESS LTD
London and Basingstoke
Associated companies in New York Dublin
Melbourne Johannesburg and Madras

ISBN 0 333 21702 0

Printed in Hong Kong by Wing King Tong Co., Ltd.

PREFACE

The basic concepts and an understanding of modern queueing theory are requirements not only in the training of operational research staff, management scientists, etc., but also as fundamental concepts in the training of managers or in mangement development programmes.

The efficient design and operation of 'service functions' is one of the main problems facing management today and the understanding obtained from a study of queueing theory is essential in the solution of these problems.

Industry and commerce have for too long concentrated their main resources on designing and operating the 'production units' and little attention has been paid until recently to the 'service units'. Basic concepts such as 'increased efficiency is achieved when the utilisation of service units is reduced' are still hard for practical personnel to understand, brought up as they are on the concept of 'maximising the utilisation' of their facilities. The ancient Chinese civilisation had a system based on queueing theory: 'Pay your doctor only when you are well'. Thus in industry if a system is correctly designed, management should be happy when 'its maintenance gang is playing cards' since there are no breakdowns to be repaired!

This book, by concentrating on problems with their fully worked-out solutions, gives students of queueing theory not only a chance to test their understanding of the theory, but also illustrates the wide range of application of the theory.

The book covers the steady-state solutions of random-arrival queueing systems. It is designed to meet the needs not only of management science training programmes, but also of mangement teaching programmes.

Cranfield, 1976 J. Murdoch

GLOSSARY OF SYMBOLS

D	deterministic distribution
G	general distribution
M	negative exponential distribution
M_n	negative exponential distribution with mean dependent on the number in the system n
C	number of channels
N	maximum system size in finite queues
λ	average arrival rate
$1/\lambda$	average inter-arrival time
μ	average service rate
$1/\mu$	average service time
$\rho = \frac{\lambda}{\mu}$ (or $\frac{\lambda}{C\mu}$)	intensity of traffic for single and multi-channel queues
or $\varepsilon = \lambda/\mu$	traffic offering (multi-channel queues)
$\sigma_s{}^2$	variance of service time
d(t)dt	distribution of the time in the system (steady state)
n	number in the system
$\overline{n}$	average number in the system
$P_n(t)$	transient state probabilities of n in the system
P_n	steady-state probabilities of n in the system
q	number in the queue
$\overline{q}$	average number in the queue
w(t)dt	distribution of the waiting time in the queue in the steady state
$\overline{w}$	average waiting time of all customers in the queue in the steady state.

BASIC DISTRIBUTIONS

$$P(x) = \frac{e^{-m} m^x}{x!}$$ *Poisson* distribution (mean = m)

$P(t) = \frac{1}{T}e^{-t/T}\,dt$ *Negative exponential* distribution (mean = T)

CLASSIFICATION OF QUEUEING SYSTEMS

Queues are classified in the book as follows.

(1) / (2) / (3) / (4)

(1)	*Input Distribution*	e.g. M, D, M_n, G, etc.
(2)	*Service Distribution*	e.g. G, D, M, etc.
(3)	*Number of Service Channels*	e.g. 1, 2, ... C, etc.
(4)	*Number in the System*	Unconstrained ∞, finite, maximum size = N

EXAMPLES OF USE OF CLASSIFICATION SYSTEM

Thus an $M/M/1/\infty$ system is random arrival, negative exponential service time distribution, single-channel, no constraint on queue size.

Again a G/M/C/N system is general arrival distribution, negative exponential service distribution, C service channels, maximum number in the system N.

The service mechanism in all problems, is service in order of arrival, or first-in, first-out (FIFO) system.

CONTENTS

1 BASIC CONCEPTS OF QUEUES

1.1 INTRODUCTION

Queueing situations arise in all aspects of work and life and are typified by the 'queueing for service'. The theory of queueing gives a basis for understanding the various aspects of the problems and enables a quantitative assessment to be made. Therefore the theory enables these 'service situations' to be more effectively designed and operated.

Understanding queueing theory and its concepts is thus basic to all personnel concerned with service situations. Since a large proportion of both capital and labour is tied up in service facilities, and these areas have in the past tended to be neglected for the direct productive units, there is clearly a large potential area of application of the theory and also large savings to be obtained.

This book, by giving a series of problems with their worked solutions, aims not only to teach understanding of the basic theory but also to give readers an insight into the potential of the theory and its wide field of application.

1.2 THE QUEUEING SITUATION

A situation in which queueing can occur may be typified by a shop where customers expect to be served by sales assistants. If all assistants are busy when a new customer enters, he has to wait and thus forms the beginning of a queue. In our discussion of queues in general, we shall call 'customer' the incoming unit, that is, the unit that enters into a situation in which a queue could form; such queues need not take the form of 'customers' actually lining up, all we need, to define a customer as queueing, is the fact that he has made clear his expectation of being served, and that the service is not available. By 'service' we shall mean any action necessary to allow the customer to leave the 'shop', or 'counter' - in general the situation where queueing had been possible. Thus there are three essential elements of any queueing situation: (1) input process - the manner in which customers arrive; (2) queue discipline - the manner in which customers wait for service after input; and (3) service mechanism - the manner in which customers are being served, or the way in which the queue is being resolved.

Figure 1.1 illustrates the queueing system for a shop

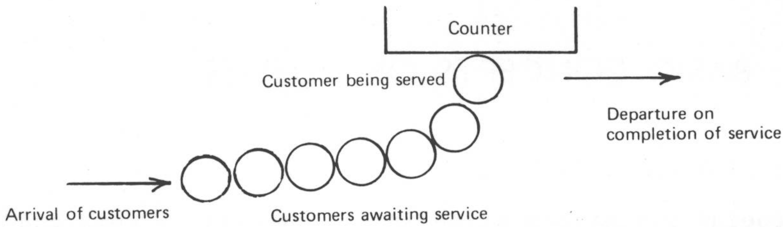


Figure 1.1 Diagrammatic Representation of Single-channel Queue.

with a single server or counter, while figure 1.2 illustrates two different queueing systems for a three-channel system (three counters in parallel).

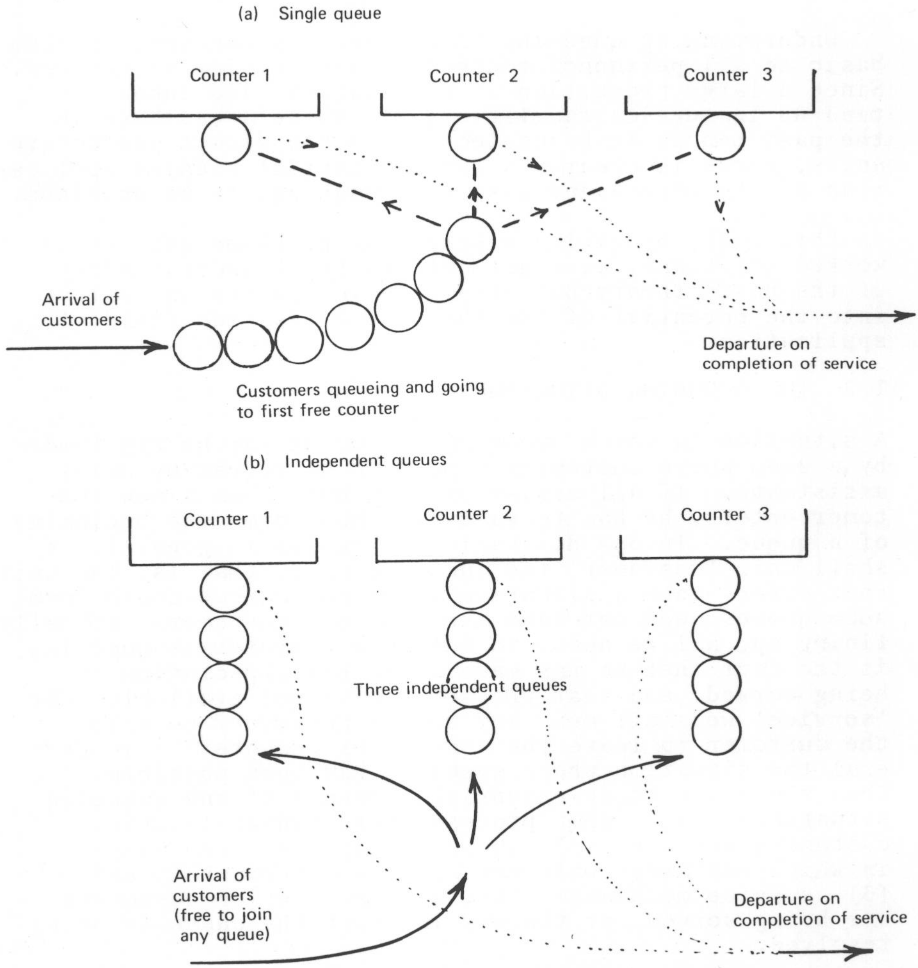


Figure 1.2 Diagrammatic Representation of Three-channel Queue.

Queueing situations are very widespread and many apparently quite different examples can be found in practice. Figure 1.3 shows a few of the more important ones, many of which have been the subject of published investigations.

1.3 TYPES OF QUEUEING PROBLEM

Although all queueing situations are basically similar, there are an almost infinite number of different situations that can arise in practice.

As previously stated, the three basic elements of a queueing problem are

(1) input process
(2) queueing discipline
(3) service mechanism.

These elements have within themselves a large number of possible variations, which give rise to a large number of different queueing situations. Figure 1.4 gives a list of possible variations, although this itself is not exhaustive.

1.4 THE BASIC THEORY

Queueing problems arise, as has been seen, in any activity where demands for service arise from a multiplicity of sources acting more or less independently of each other. Where customer arrivals, or demands, for service can be scheduled exactly, then it is relatively easy to provide appropriate service facilities, and this is really a trivial problem compared with the ones that occur more usually in practice and to which queueing theory give the basis for solution.

In developing the theory, it is convenient to imagine customers arriving at a counter and queueing for service, if the service mechanism is busy (see figures 1.1 and 1.2). This method can also cover situations where a physical queue does not in practice exist, for example, machines awaiting service from an overhead crane, callers waiting on different lines for a connection by the telephone operator, etc.

1.4.1 Measures of Effectiveness

It has been found possible to set up mathematical models to describe queueing situations specified by different forms of the three basic elements. These models can then be manipulated to show what the service system under investigation should be capable of achieving and how any two or more systems compare. In order to make a decision

Figure 1.3 Some Typical Situations for the Application of Queueing Theory

Situation	*Input to Queue*	*Queue*	*Service Mechanism*
Shop Booking office Post Office Bank Hairdressing salon	Customers or clients arriving for service	Waiting for counter to be free	Assistant, teller, etc., serving at counter
Traffic Bus stops Taxi ranks	Customers arriving	Customers queueing waiting for bus or taxi	Arrival of bus, taxi, etc.
Doctor's or hospital outpatients waiting room	Patients arrival for treatment	Patients waiting their turn	Treatment by doctor
Factory handling system	Jobs requiring movements	Jobs waiting at various points for movement	Actual movement of job by transport
Airport	Planes arriving to land	Planes circling overhead waiting for free runway	Planes landing on runway at airport
Stocking of goods	Arrival of batches of goods from supplier	Stocks of goods in store	Usage or purchases of goods from store

Situation	*Input to Queue*	*Queue*	*Service Mechanism*
Telephone switch-board	Customers picking up phone	Customers awaiting telephone switch-board operator's response	Switching at switch-board or exchange
Estimating (tendering)	Possible contracts for pricing for tendering	Contracts awaiting pricing	Pricing by estimator and sending off tender
Harbour design	Ships arriving at port for loading or unloading	Ships awaiting berth	Ships being loaded or unloaded
Semi-skilled oper-ators in machine shop	Operators and/or machine breakage, etc., requiring skilled setter (or maintenance operator)	Operators and machines waiting for skilled setter (or mainten-ance)	Adjustment (or repair) by skilled setter, or maintenance)
Maintenance department	Plant breaks down	Plant awaiting repair	Plant repair

Input Process

Can vary as follows in

1. Number of potential customers
   - (a) Infinite
   - (b) Finite
2. Number arriving at one time
   - (a) Singly
   - (b) In batches of constant number
   - (c) In batches of variable number
3. Intervals between arrivals
   - (a) Constant
   - (b) Completely random (Poisson input)
   - (c) Other distribution of intervals

Queueing Discipline

Can vary as follows in

1. Number of queues
   - (a) Single queue
   - (b) Several queues
2. Queue discipline
   - (a) Service in order of arrival
   - (b) At random
   - (c) Priority
   - (d) Service in reverse order of queueing (unfair queue)
   - (e) Service time-dependent

Service Mechanism

Can vary as follows in

1. Number of service points or servers
   - (a) One
   - (b) Several
   - (c) Number variable
2. Number served
   - (a) One at a time
   - (b) In batches of constant number
   - (c) In batches of variable number
3. Service available
   - (a) Permanently
   - (b) Intermittently

Input Process	*Queueing Discipline*	*Service Mechanism*
4. Average rate of arrival (a) Constant (b) Varying with time (c) Influenced by a state of queue		4. Duration of service (a) Constant (b) Exponentially distributed (times of beginning and end of service distributed independently and at random) (c) Other distributions (d) Dependent on time customer has spent in queue
5. Outside influence (a) No outside influence (b) Input is the output of a previous queueing situation		5. Average rate of service (a) Constant (b) Varying with time or state

Figure 1.4 Queueing Systems: Types of Variation in System.

on which system is the 'best', certain 'measures of effectiveness' are required.

Useful measures of effectiveness have been found to be

(1) The probability of having n customers waiting at a time t, given the initial state of the system. Knowing this probability distribution, the size of the queue that will be exceeded for only 5 per cent, say, or 1 per cent of the time, can be determined. This could be useful, for example, in determining what size of waiting room needs to be provided for customers so that only rarely will there be an overflow of customers and possible loss of business if there is an alternative service point they can go to. Alternatively, with a given size of waiting space, the service facility required could be determined such that the waiting space will be adequate most of the time.

(2) The distribution of waiting time of customers. From this distribution can be found the average waiting time of customers and the proportion of customers who have to wait longer than a certain time t, say, If the probability of waiting longer than t is high, then customers may be discouraged from joining the queue, which would result in a loss of potential business in something like a petrol station or a supermarket. The cost of providing more or faster service facilities may be more than compensated for by the extra business produced by a reduction of customers' waiting time. Again, in the case of an internal stores in a factory, provision for an extra storekeeper, say, may pay handsome dividends in reducing the lost production time of skilled men who have to queue for a long time for service.

These measures of effectiveness (depending which, if any, is appropriate to the problem) can be used to decide, for instance (usually on a cost basis) whether to speed up the existing service rate of each channel or whether to provide extra channels working at the same rate as the present ones or whether even a reduction in service facility can be contemplated.

1.5 MATHEMATICAL SOLUTION OF QUEUEING PROBLEMS

1.5.1 Transient and Steady-state Solutions

The solution of queueing problems is considered in two parts, namely the transient or time-dependent solution, and the steady-state solution. Briefly, provided that the service channel is capable of serving at a faster