MOVING AVERAGE PROCESSES IN QUEUEING THEORY

by

C. E. M. PEARCE

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

THE AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA

December, 1965.

Except as stated herein, this thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of the author's knowledge, contains no copy or paraphrase of material previously published or written by another person, except when due reference is made in the text of the thesis.

C. Pearce

(C.E.M. Pearce)

PREFACE

The research embodied in this thesis was done in the Mathematical Statistics Department of the Australian National University during the years 1963-65. Though the techniques involved are standard, the material presented is original throughout. The second part of Chapter Six, the substance of Chapter Two, and a special case of Chapter Four have been accepted for publication, and the simplest case of Chapter Two, that of the second order moving average, has already been published as a joint paper with Professor P. D. Finch. It is hoped to submit several other parts of the thesis for publication shortly.

I should like to take this opportunity to thank Professor Finch for his help in the preparation of our joint paper, and also to thank him and Professor P. A. P. Moran for their stimulus, encouragement and help during my three years at the Australian National University.

I should also like to thank Mrs. B. Cranston for her patient work in typing very trying manuscripts of my general work at the University, and Mrs. H. West for her preparation of the excellent final copy of this thesis.

I should finally like to thank Miss F. O'Connor for a meticulous eye given in exacting checking of the final copy.

SUMMARY.

Whilst there are ergodic theorems and some results are on the waiting times of queues with stationary but not necessarily recurrent inputs, there have been almost no explicit results on the queue lengths in such systems. Once inter-arrival intervals are no longer independently distributed, it becomes unclear just how the input is to be classified and characterized in a mathematically convenient manner.

This thesis attempts to gain some explicit results on queue lengths in queueing systems with identically but not independently distributed inter-arrival intervals. We set up a moving average model for the input and examine its implications for single server queueing systems, for queues with batch arrivals, and for many server queues. Because of the mathematical complications we deal primarily with negative exponential and Erlang services. We deal more briefly with single server queues with Poisson inputs and moving average service processes. We also treat the queue with infinitely many servers for general recurrent services and a completely arbitrary input.

Using only a limited number of techniques widespread in queueing theory, we are able to investigate the systems that we consider in some detail, including a study of both their equilibrium behaviour and their transients.

The moving average input or service includes the standard general recurrent process as a special case, and we derive a number of results which are simple generalisations of the corresponding results for recurrent queues.

TABLE OF CONTENTS.

Page

CHAPTER ONE

INTRODUCTION

CHAPTER TWO

THE EQUILIBRIUM BEHAVIOUR OF THE QUEUE WITH GENERAL ORDER MOVING AVERAGE INPUT AND NEGATIVE EXPONENTIAL SERVICES

CHAPTER THREE

VARIETY OF EQUILIBRIUM QUEUE SIZE DISTRIBUTIONS OCCURRING IN G(p)/M/1 SYSTEMS

CHAPTER FOUR

THE TIME DEPENDENT BEHAVIOUR OF THE
QUEUE $G(p+1)/E_r/1$

CHAPTER FIVE

THE TIME DEPENDENT BEHAVIOUR OF G(2)/M/1

CHAPTER SIX

CHAPTER ONE

1. *Survey of the literature.*

It is now sixty years since A. K. Erlang began the work from which queueing theory was to develop, and we are in a position to survey the growth of the subject. We presume the standard terms, postponing a definitive statement until section three.

Erlang dealt with the Danish telephone system, and considered Poisson inputs into one or more channels with negative exponential or constant holding times. He was concerned with problems such as the probability of loss of a call or of a prescribed number of channels being occupied and, in systems in which a call may wait for a channel to become available, with the waiting times.

Negative exponential and constant holding time distributions are extreme in that the former is 'completely' random and the latter 'completely' deterministic. In his later work Erlang made use of an intermediate family of distributions which now bear his name. A number of others have extended Erlang's results. Pollaczek introduced the general *recurrent* service in 1930 and Palm the general recurrent input in 1943. The study of systems with arbitrary recurrent input or service processes was, however, considerably hindered by the absence of the characteristic 'lack of memory' property of the negative exponential distribution. Rapid advances followed only after the inception in the early 1950's of three new and general methods:-

In 1951 D. G. Kendall introduced the concept of the 'imbedded Markov chain'. It was shown that whilst the complete behaviour in time of a queue might be extremely difficult to determine, much information could often readily be obtained for a special sequence of instants, the 'state' of the system at any one of which depended only on that at the previous such instant. The times of arrival form such a set of 'regenerative points' for the single server queue with general independent inter-arrival intervals and negative exponential service distributions. A corresponding set of regenerative points is provided for the single server queue with Poisson input by the departure instants.

Cox (1955) was able, at the cost of mathematical complexity, to consider the queue at an arbitrary instant of time (and thus lose no information) by introducing supplementary variables. The purpose of these variables was to augment the state space of the queueing process so that not only the queue size at a given instant was specified, but also sufficient additional information to make the future probabilistic behaviour of the system predictable purely from its present state. Cox's method thus essentially recovers the 'memory-less' Markovian property by incorporating some of the past history of the process in an augmented state space.

The other contribution was by L. Takács, also in 1955. Takács introduced as an explicit variable the virtual waiting time of a hypothetical arrival joining the queue (for which a 'first come, first serve' priority is preserved) at any instant of time. The virtual waiting time can be regarded more directly without the service priority restriction as the time until the server has dealt with all the customers at present in the queue. This work is restricted to the extent that it deals only with waiting times and these only for (inhomogeneous) Poisson inputs, but has proved of considerable value in studies of the transient behaviour of queueing systems.

In 1952 Lindley determined under very mild conditions stability criteria for general recurrent input and service distributions by working in terms of the waiting time of the nth customer. His results have been extended to similar systems with a general number of servers by Kiefer and Wolfowitz (1955). Smith (1953) makes use of Lindley's analysis to give a more detailed examination of the general recurrent single server queue. An apparently little known paper by Šahbazov (1962) generalises Lindley's work to cover arrivals in batches whose size follows a general probability distribution

The waiting time process has, in general, proved much easier to investigate than the queueing process. Lindley established ergodicity in terms of waiting times,and these were also used in the above mentioned generalisations of his work. In fact Lindley derived an integral equation for the equilibrium distribution of waiting time. Kiefer and Wolfowitz (1955) have given what are essentially simultaneous integral equations for the equilibrium waiting time distributions in Gl/G/n. It was assumed that an arriving customer takes the vacant server with the lowest serial number if more than one is available,otherwise the first available server. Pollaczek((1953), (1954)) has simultaneous integral equations covering the transient behaviour of waiting time and queue length.

It was not until 1962, however, that a full treatment of queue length for the general recurrent single server queue was finally given by Keilson and Kooharian. The basis of their procedure using Cox's technique, was to adjoin to the state space the expended times since the last arrival and the commencement of the last service. Whilst they give the complete time behaviour of the general recurrent single server queueing system, the information contained in the solution is partly locked in generating functions, integral transforms and contour integrals from which explicit probabilities for particular systems are often difficult to obtain. Many results on this and more specialised systems are supplied in simpler terms elsewhere in the literature, to some of which we shall make reference in the text. A bibliography by Alison Doig in

1957 lists over sever hundred papers on queueing theory, and many papers have been written since. We do not attempt a complete discussion.

It is of interest to compare the difficulties in waiting time and queue length problems. Let us denote the random variables virtual waiting time and queue length at time t by $\eta(t)$, $n(t)$ respectively. Beneš (1960, 1960a) has shown that even for non-recurrent systems $\eta(t)$ can be derived from $K(t)$, the sum of the service times of all the arrivals in a system before time t, which is a readily found quantity. The problem of determining $\eta(t)$ is thus in principle solved. No correspondingly simple analogue to K for the determination of $n(t)$ is apparent. Clearly $n(t)$ cannot be derived from $K(t)$, as this contains no information on departure epochs. On the other hand, $\eta(t)$ can be derived from $n(t)$ as the sum of the unexpended portion of the service time of the customer in service and the service times of the remaining customers waiting.

The difference between the queueing process and the waiting time process is clear with ordinary recurrent queueing systems. The process giving the waiting time of the nth customer in GI/G/1 is a simple Markov process, studied by Lindley in 1952. The time dependent queue length process, on the other hand, is a one dimensional projection of a three dimensional Markov process. The solution was provided by Keilson and Kooharian in 1962. It is therefore hardly surprising that the behaviour of queue length in all but the simplest queueing systems is neither intuitively clear nor mathematically simple.

The queueing process can be regarded as a combinatorial problem, and as such has been tackled using results from fluctuation theory. In particular, Prabhu and Narayan Bhat (1963) have treated the queue with recurrent services and general bulk Poisson arrivals and the queue with recurrent arrivals and general bulk negative exponential services. Narayan Bhat has extended these results to cover arrivals and services being simultaneously bulk. Finch (1961) has studied the busy period of GI/G/1.

No general treatment has yet been given for queue length probabilities in many server queueing systems of either a parallel or series type. In series queues the complexity of the queueing process is increased since the input process to the second and successive stages is not in general a renewal process. Finch (1959) has shown that the output of a single server first stage is Poisson if the input is Poisson and the service negative exponential, but that, under even slight generalisations to the system this ceases to be true. Several writers have derived explicit results for tandem queues with Poisson inputs and all the services negative exponential. In particular Finch (1959a) considers such a system with positive feedback, i.e., where there is a probability that a customer may on completing service at one service point return to join the queue at an earlier server. Rényi treats a system with a sequence of service points, each of which filters off a constant proportion of arrivals and transmits the remainder instantaneously after scaling up their density in time. He shows that for a general recurrent input of intensity λ the resultant stream leaving the system has a Poisson distribution of intensity λ as the number of servers tends to infinity provided that the sequence $\{p_1 \, p_2 .. p_n\}$ of partial products of fractions transmitted tends to zero. It is also of note that the output of a first stage with general recurrent input and general recurrent service time distribution but with infinitely many servers is of a not too complex form, since the effect of the first stage is simply to superpose a general delay distribution for arrivals to the second stage.

Queue length distributions in series queues are, with a few exceptions, undetermined. In sharp contrast, very general ergodicity results have been established through the consideration of waiting times. Sacks establishes the existence of a unique limiting waiting time distribution for a general finite number of queues in tandem for a general recurrent input to the first under the restriction that the mean inter-arrival time exceeds the mean service time of each of the queues. In 1962 Loynes proved ergodicity results for the many server queue and for queues with several single server stages in series under very general conditions. The only restrictions made were that the traffic intensity was less than unity and that the input and services had stationary distributions. The treatment was completed by Loynes in 1964 in a second paper. Loynes also gives some techniques for solving for the equilibrium waiting time distributions (1962a). Belyaev has extended the results of Rényi mentioned above to a general stationary input.

Until recently, no explicit results existed on queue lengths for queues with stationary but non-recurrent inputs. This was partly because of the mathematical difficulties involved and partly because non-recurrent inputs do not seem to lend themselves to any natural classification or characterisation.

The restriction to recurrent inputs is quite severe. In many physical processes successive inter-arrival intervals will be correlated. A prime instance of this is the input to the second stage of a series queueing system whose first stage has a general recurrent input. This input is the output of the first stage of the system. When a departure from the first stage occurs after a comparatively long service there will be a corresponding long inter-arrival interval for the second stage, but there will also be a greater likelihood that the server in the first stage will not undergo an idle period but will have a further customer already waiting to commence service. The long inter-arrival period will thus have a corresponding greater probability of being followed by a short inter-arrival interval. A short inter-arrival interval will similarly tend to be followed by a long inter-arrival interval, manifesting a negative correlation between successive inter-arrival intervals. This argument holds regardless of the past history of the first queue. In particular, it is not invalidated by the fact that a very long service time for a customer is likely to produce a sutstantial queue at the service point, resulting in an increased likelihood of a whole ensuing sequence of short inter-arrival periods for the second queue. It is clear that the output of the first stage of a queue will not, in general, give a sequence of identically and independently distributed random variables.

That the restriction to recurrent inputs is infelicitous can be seen more generally:

If the input arises from independent arrivals from a non-interacting population, one might expect the probability of an arrival during any small element of time Δt to be expressible as

$$\lambda(t) \Delta t + o(\Delta t),$$

i.e., for arrivals to form an inhomogeneous Poisson input. Stable equilibrium queue behaviour would hardly be expected from such an input unless $\lambda(t)$ was a constant. Hasofer has demonstrated that when $\lambda(t)$ varies sinusoidally this sinusoidal behaviour is reflected in the limiting behaviour of the system.

Suppose, on the other hand, that the input is provided or controlled by some unspecified mechanism. We can then regard the input as the output of some process, and we would expect the inter-arrival intervals to reflect the not too distant history of this process, probably with a heavier weighting on the most recent history. There will thus be a correlation between successive inter-arrival intervals. A simple model incorporating such a history dependence is provided by having the inter-arrival intervals moving averages of a sequence of identically and independently distributed random variables. The classical literature on queueing theory deals with the simplest case of a moving average of order 1, i.e., an ordinary renewal sequence. Yarovitsky (1962) has considered an input process similar to a second order moving average. The nth inter-arrival interval ξ_n^{**} is of the form

$$\xi_n^{**} = \xi_n^{*} + \eta_n,$$

where $\{\xi_n^{*}\}$ is a sequence of identically and independently distributed non-negative random variables but η_n depends on ξ_{n-1}. Yarovitsky is concerned with the loss in a system without waiting room.

Finch (1963) has given closed expressions (in terms of the inter-arrival intervals) of a determinantal form for the queue length as found by the nth arrival to a queue with Erlang service times and a completely arbitrary input. Finch showed for a stationary input, that, when the traffic intensity is less than unity, an equilibrium queue size distribution exists. In an attempt to determine the form of the equilibrium distribution he developed a heuristic symbolic method in which this distribution is expressed as a formal Taylor series. The symbolic method provides easy access to a whole new range of queueing problems, but seems difficult to justify. Finch made use of the symbolic method to deal with inputs which can be expressed as a second order moving average of independently and identically distributed random variables.

In this thesis we give rigorous treatment of queueing systems with general order moving average inputs and negative exponential services. We verify the form of the limiting queue length distribution obtained by Finch for the special case of a moving average of order two but find that the value of a constant characterizing the solution is incorrect.

Using a very limited number of techniques current in queueing theory we are able to make a fairly full investigation into queues with this very general class of inputs. Because of the mathematical difficulties involved we restrict ourselves to negative exponential and Erlang services. This is not as severe a restriction as it is for inputs, as most practical situations involve either negative exponential or deterministic services. We note that a deterministic probability distribution is approached as the limit of a kth order Erlang distribution as $k \to \infty$.

In Chapter Three we shall consider the generality of moving average inputs and gain some idea of their limitations.

Since the equilibrium behaviour of a queueing system is time independent, it is usually easier to determine than the transient behaviour and most of the literature on queueing theory deals with this limiting behaviour only. On the other hand, a system may be very slow in approaching equilibrium in the time scale in which we are working, or, if the traffic intensity is greater than unity, may not reach it at all. In such circumstances we are interested in the transient behaviour of the system. This gives us full information about the system, and indeed suffices for a determination of the equilibrium behaviour when the latter exists. A knowledge of the transient behaviour of a queueing system is necessary for a full discussion of the stability of that system. The transient behaviour of the systems GI/M/1 and M/G/1 which we are generalising is of particular importance for considering more complicated systems. That of M/G/1 can be used for a study of priority queues with general recurrent services (as is done by Keilson (1962b) and Gaver (1962)), whilst the analytical treatment required for GI/M/1 can be made use of for a study of the system M/G/1 with finite waiting room (Keilson, 1964). We have therefore given some attention to the transient behaviour of our moving average systems.

Treatments of the dependence of a system have largely sprung from the realization by Borel (1942) of the significance of the busy period. The busy period loses some of its importance in our work since for moving averages of order higher than one successive busy periods are not independently distributed as they are for a simple recurrent input, although they will still be identically distributed. We are, however, able to make some use of busy periods in our treatment of the time dependence of queue sizes.

Although moving average services do not appear to arise as naturally as with inputs, they can still be given physical interpretations. An instance is the server whose serving efficiency exhibits good and bad periods which manifest themselves despite the fluctuations in the work load provided by a customer. Moving average services can also be regarded as reflecting some similarity between consecutive items in the input.

We deal also with a common class of queueing processes in which the service facility operates whether or not customers are available, such as in a public transport system. Such queues seem to have been largely neglected in the literature, although they have been considered by Bailey (1954), Downton (1955 and 1956), and more recently and in greater detail by Keilson (1962a). We show that the basic general moving average problem can be solved through a knowledge of the corresponding standard system in which the service facility operates only when a customer is present. A fuller treatment could follow along the lines of the rest of our work.

Our study concludes with a brief look at moving average input queues with general recurrent service time distributions.

2. *General outline of the thesis.*

In the first section of this chapter we saw that one would naturally expect successive inter-arrival intervals in queueing systems to be correlated. Whilst ergodic theorems allow this possibility, under stationarity restrictions, explicit results on queue sizes in the literature have been for recurrent queueing systems exhibiting only zero correlations between successive inter-arrival intervals, and even in this case mathematical difficulties delayed a full treatment of the general recurrent single server queue until 1962.

The succeeding chapters investigate a moving average model for inter-arrival intervals which automatically incorporates both positive and negative correlations. The standard identically and independently inter-arrival intervals occur as the simplest case, and we generalise a number of known results. We deal with negative exponential and Erlang services.

In Chapter Two we use the method of supplementary variables to consider the basic question of the equilibrium behaviour of a queue with moving average input and negative exponential services. We find that the form of solution is a simple generalisation of a well known result. The complete explicit determination of the solution seems difficult, and we give a (finite) recursive proceedure which we illustrate with moving averages of orders two and three. We use the former to check a conjecture of Finch (1963). We look at the relation between our results and a paper of Loynes (1962a) on waiting times in queues with stationary but non-recurrent inputs. Finally we consider an alternative approach to our problem using techniques of Beneš, but find that these are somewhat less tractable.

In Chapter Three we review the scope of moving average inputs as a subclass of the class of stationary inputs. Using two families of distribution functions (Erlang and deterministic) from which a large class of distributions can be built up, we make a detailed investigation of limiting queue length behaviour for second order moving average inputs. The equilibrium queue length distribution for GI/M/1 (as found by arrivals) is geometric with parameter T the inner root of the characteristic equation

$$T = \psi(1-T) \equiv \int_0^\infty \exp[-\mu(1-T)\, u] \quad dU(u),$$

where μ is the service time distribution parameter. The corresponding result for a (p+1)th order moving average input, apart from the first p probabilities, is geometric with the same parameter. As mentioned above, the limiting queue length distribution for arrivals to GI/M/1 is of the form

$$\{(1-T)T^j \quad , \quad j \geq o\} \quad ,$$

so that

$$P_o = 1 - T \ ,$$

where P_o is the equilibrium probability that an arrival finds the queue empty. We find that for our second order moving averages for which successive inter-arrival intervals are positively correlated, we have

$$P_o \geq 1 - T,$$

whilst with negative correlations

$$P_o \leq 1 - T.$$

This is a simple generalisation of the result for the standard uncorrelated case.

Using the family of Erlang distributions we are also able to show that a given limiting distribution $\{(1-T)\ T^j\}$ can arise for arbitrarily small traffic intensities. This is perhaps a surprising result, since for stationary inputs ergodic behaviour is controlled rigidly by the traffic intensity. Loynes (1962) has shown that there is stable limiting behaviour if and only if the traffic intensity is less than unity.

We show in Chapter Four that the transient behaviour of general moving average systems can be derived much as the equilibrium behaviour in Chapter Two. With only moderately more involved equations we are able to treat Erlang services and batch arrivals. With the more general Erlang services we take up again the question of Chapter Three on the variety of limiting distributions arising in systems with moving average inputs.

Chapter Five gives a fairly full study of the transient behaviour in queues with second order moving average inputs. We find that in the case of second order inputs there exist regenerative points which enable us to deal much more freely and readily with transients than we could have by the methods of Chapter Four.

Conolly (1960) derives for $GI/E_k/1$ a simple relation between the limiting queue length distribution in continuous time and that on the imbedded Markov chain. For a general second order moving average the imbedded chain and the arrival instants of the queue no longer coincide. We find that Conolly's result still holds good for second order inputs if we deal with arrival instants. Denoting the continuous time and arrival instant equilibrium queue length distributions by $\{q_j, \ j \geq o\}$ and $\{p_j, j \geq o\}$, the result is

$$q_j = \begin{cases} (\mu m)^{-1} \quad (p_{j-k}+\ldots+p_{j-1}), & j > k, \\ (\mu m)^{-1} \quad (p_o \ +\ldots+p_{j-1}), & o < j \leq k, \\ 1 - \text{traffic intensity}, & j = 0, \end{cases}$$

where μ is the parameter of the (kth order) Erlang service and m the mean inter-arrival time.

Chapter Six, which deals with many server queues, consists of two parts. In the first of these we use the supplementary variables method of Chapter Two and the imbedded Markov chain method of Chapter Five to extend our results to cover queues with a finite number of identical servers in parallel. In particular, we find that Kendall's (1953) result that the equilibrium queue length distribution is of a delayed geometric form holds good, though the delay in the distribution is greater than just the number of servers.

As well as treating the second order moving average in detail for the usual unrestricted waiting room, we give simple algebraic equations sufficient to find the equilibrium queue length distribution for a restricted waiting room.

We also find the equilibrium queue length distribution for a second order moving average for the case where the number of servers is infinite

The second part of Chapter Six is concerned with the transient behaviour of a queueing system with infinitely many servers and general recurrent batch services. Using simple probabilistic reasoning, we find for a completely unrestricted input the generating function, mean and variance distribution of the number of customers waiting at an arbitrary instant of time. We specialise to stochastic inputs, i.e., inputs for which the inter-arrival intervals have a joint probability distribution. In the case of general recurrent inputs we show our results reduce to those of Finch (1961).

Chapter Seven considers the single server queue with Poisson arrivals (parameter λ) and moving average services, again using the techniques of supplementary variables. The solution for a general moving average seems somewhat involved, and we give only a(finite) recursive proceedure. This we illustrate with the case of moving averages of orders two and three. The generating function of the limiting queue length distribution for the former generalises a result of Kendall (1953). We find that Kendall's result

$$P_o = 1 - \lambda \times \text{mean inter-arrival time},$$

where P_o refers to the probability of zero queue length as left by a departure, holds for moving average services of all orders.

A more detailed examination of the queue length and busy period is made for moving averages of order two, using an imbedded Markov chain.

Chapter Eight, in two brief parts, goes on to further problems.

In Part I we show how the results obtained in Chapter Two can be employed to provide an easy solution to a queueing system with moving average inputs but a different type of service. The service facility operates whether or not customers are present, so that an arrival at an empty queue may have to wait before he can begin service.

In the second part we derive equations for the single server queue with general order moving average input and general recurrent services. We consider the possibility of solving these equations.

3. Basic Notions.

Queueing theory concerns itself with stochastic processes in which discrete units undergo a delay in a physical system before passing out and being lost to that system. The most common interpretation is that the units are customers, who arrive for service at one or more counters, although the original context for the subject was the handling of telephone calls by an exchange. There is also a close association with storage and inventory theory. To prescribe a queueing system we need to specify the following:

(i) *The serving mechanism.* There may be one or more servers (machines, channels, counters) in parallel or in series stages. Servers in parallel are normally identical whereas this restriction is not usually made with servers in series. If an arrival to an arrangement in parallel finds a server idle he will begin service immediately. With a series arrangement a customer proceeds to each stage only after completing service at the previous stage. Occasionally it is specified that an arrival at an idle server suffers a delay before commencing service.

When the service or holding times are identically distributed independently of one another and of the input, the service process constitutes a renewal process. For convenience we shall abbreviate identically and independently to I.I.D. in the text. Servicing may be of individuals or of groups (batch service).

(ii) *The input.* Customers (items, demands) arrive singly or in batches at the instants of a stochastic process termed the input process. When the inter-arrival intervals are I.I.D. the input process is a renewal process and is characterized by the probability distribution function of the inter-arrival intervals.

(iii) *The queue discipline,* i.e., the selective procedure for determining in what order customers are to be served. Most of the literature on queueing is to do with a first come, first served discipline, i.e., customers finding all servers occupied on their arrival wait for service in the order of their arrival, although there has been work done on inputs consisting of more than one class of customer and an indexing of priority for service. Also, there may be limited waiting room, or customers may balk if the queue present on their arrival is beyond a certain size.

The notation almost universally employed in queueing problems is an elaboration of that introduced by D. G. Kendall in 1953. A queueing system with n identical servers in parallel and renewal input and service processes characterized by probability distribution functions $A(x)$, $B(x)$ respectively is represented by $A/B/n$. The types of distribution A,B considered in the literature are of fairly well defined classes. A general distribution function is represented by G. Apart from G, the distribution functions are D, a distribution having all its probability concentrated in a single saltus, and t the family $\{E_k : k \geq 1, k \text{ integral}\}$ of Erlang distributions. The Erlang distribution of order k is

$$E_k(x) \equiv \begin{cases} 0, & x < 0, \\ 1 - \exp(-\mu x) \sum_{i=o}^{k-1} (\mu x)^i / i!, & x \geq 0, \ \mu > 0. \end{cases}$$

We remark that GI (standing for general independent input) is frequently used in place of G when referring to inputs. M(Markov) is also used in place of of E_1.

The above notation is readily extended to series (queues in tandem), which can be represented as

$$| A_1 | B_1 | n_1 \quad .. | B_2 | n_2 \quad .. , \text{etc.}$$

This shorthand notation is not, of course, sufficient to specify any queueing system. It makes no statement about queue discipline, about batch sizes or about the possibility of a serviced customer returning to an earlier stage of a system to recommence service, for example. It is normal to presume unless stated to the contrary that arrivals and services are of individuals, that the queue discipline is first come, first served, and that there are no restrictions on waiting room.

Once a system has been specified, queueing theory concerns itself with the determination of the temporal statistical behaviour of physical characteristics of that system. We may deal with characteristics of interest to the arriving customers or the servers. Of predominant interest are:

(i) The queue length (size). We adopt the convention that this includes the number of customers undergoing service as well as those waiting to be served. Because of the equivalence of servers at a given stage of a system, it suffices to give the number of customers at each stage of that system.

(ii) The waiting time of a customer, i.e., the length of the period between the moment a customer arrives and the moment he commences service. We define $\eta_n(t)$ as the probability that the nth arrival has to wait a time $\leq t$ before commencing service. $\eta.(\cdot)$ is a mapping from the direct product $\{n:n \geq 0\} \times \{t:t \geq 0\}$ of a discrete and a continuous space into the non-negative half axis. It is often more convenient to work in terms of the virtual waiting time, $\eta(t)$, defined as the time a hypothetical customer would have to wait for service were he to arrive at time t, presuming a first come, first served, queue discipline. The virtual waiting time can, for a single server queue, be regarded without the service priority restriction as the occupation time of the server, i.e., the time until the server finishes serving the last of the customers in the queue at time t.

(iii) The busy period. A busy period is a period throughout which the service facility is continuously occupied and which is not a proper subset of another such period. The convention is adopted that if a customer arrives just as a departure is about to deplete the queue, the busy period is deemed to be still on throughout the service of the new arrival. It is of interest to determine the joint probability that a busy period be of any given duration and involve any given number of customers.

CHAPTER TWO.

The equilibrium behaviour of the queue with general order moving average input and negative exponential services.

1. Introduction.

In this chapter we follow up the observations of the introduction about the nature of queueing system inputs. We propose a new input model and investigate the equilibrium behaviour of the single server queue with negative exponential services. In section 7, Finch's conjecture (1963) on the form of the equilibrium queue length distribution for a moving average of order two is examined.

The standard assumption that individual arrivals are mutually independent can be prescribed by having the probability of an arrival during a small time element Δt about t given by

$$\lambda(t)\ \Delta t + o(\Delta t),$$

independently of the past history of arrivals to the system. Stable equilibrium behaviour would not be expected for a general function $\lambda(t)$, and a large part of queueing literature deals with the simplest case where λ is a constant (the poisson input). Stable behaviour occurs for such a λ and also for the other well studied case where λ has its mass uniformly distributed between equi-spaced points, i.e. the deterministic input. Hasofer has shown that if λ varies sinusoidally then this is reflected in the limiting probability that the server is idle and in the Laplace transform of the waiting time distribution.

The more usual definition of an input in terms of I.I.D. inter-arrival intervals removes the mutual independence of arrivals. The arrival probability density at any instant now depends on the past history of arrivals in that it involves a knowledge of when the last arrival occurred. Such a prescription suggests some mechanism regulating admission of arrivals or the existence of some other system whose output is the present input.

As we saw in the last chapter with tandem queues, it will rarely happen that the output of such an earlier mechanism constitutes a recurrent process, so that the limitation of inputs to recurrent processes is quite severe. Inter-arrival intervals will be identically distributed but in general consecutive intervals will be (positively or negatively) correlated. One would expect closely consecutive inter-arrival intervals to be more highly correlated than more separated intervals.

We set up a model incorporating a positive correlation between consecutive and near consecutive intervals as follows:

Customers arrive singly at the instants

$$0 = A_o < A_1 < A_2 < \ldots\ldots,$$

where the intervals separating A_m, A_{m+1} are such that

(1.1) $$A_{m+1} - A_m = b_o U_{m+p-1} + \ldots + b_p U_m, \quad m \geq o,$$

the $b_i, o \leq i \leq p$, being non-negative constants and $\{U_m\}$ a sequence of I.I.D. non-negative random variables. For convenience we take

$$\sum_{i=o}^{p} b_i = 1 .$$

We denote the common distribution function of the U_m by

$$U(x) = P(U_m \leq x), \quad m \geq o, \ x \geq o,$$

and take the mean of $U(x)$ to be finite.

We refer to

(1.2) $$\{b_o U_{m+p} + b_1 U_{m+p-1} + \ldots + b_p U_m\}$$

as a moving average of order $p + 1$. An ordinary I.I.D. sequence of random variables is thus a moving average of order one.

By writing $G(p)$ for a p th order moving average process we can conveniently extend Kendall's notation for queueing systems. The standard GI or G is identical to our $G(1)$ and we can freely interchange these symbols.

It is clear from (1.1) that if the b_i are taken to be equal, the correlation between successive inter-arrival intervals can be made arbitrarily close to +1 by increasing p.

If S_m is the service time of the (m + 1)th arrival, then $\{S_m\}$ is to be a sequence of I.I.D. random variables, with

$$P\ (S_m \leq x) = 1 - \exp(-\mu x),\ x \geq o,\ \mu > o.$$

It follows from the work of Finch (1963) and Loynes (1962) on queueing systems with stationary inputs, that if

$$P_j^{(n)}\ ,\ j \geq o,\ n \geq o\ ,$$

denotes the probability that the arrival at A_n finds exactly j customers already in the system, then

$$P_j = \lim_{n\to\infty} P_j^{(n)}\ ,\quad j \geq o,$$

exists provided the traffic intensity is less than unity. Finch in fact derives explicit expressions for the Q_j ,

$$Q_j = \sum_{i=j+1}^{\infty} P_i\ ,\quad j \geq o\ ,$$

corresponding to a completely general stationary input and a negative exponential service time distribution. The expressions obtained were

(1.3) $$Q_j = \sum_{s=j}^{\infty} [\ \Delta_j^s\ \phi_s^{s+1}]\ \alpha_o = \alpha_1 = .. = \alpha_s = o\ ,\ j \geq o.$$

For convenience the parameter μ of the service time distributions was taken as unity.

In (1.3), Δ_j^s is a differential operator with an expansion in terms of

$$\delta_j^r = (r!)^{-1} \partial^r / \partial \alpha_j^r ,$$

$$\Delta_j^s = (-)^{s-j} \begin{vmatrix} \delta_j^1 & \delta_j^2 & \ldots\delta_j^{s-j} \\ 1 & \delta_j^1 & \ldots\delta_{j+1}^{s-j-1} \\ o & 1 & \ldots\delta_{j+2}^{s-j-2} \\ - & - & - \\ o & o & \delta_{s-1}^1 \end{vmatrix} .$$

ϕ_s^{s+1} is given by

$$\phi_s^{s+1}(\alpha_o, \alpha_1, .., \alpha_s) = E\,[\exp\{-\alpha_o\theta_{s,o} - \alpha_1\theta_{s,1} - .. -(\alpha_s+1)\theta_{s,s}\}],$$

where

$$\theta_{s,i} = A_{s+1} - A_{s-i}.$$

(1.3) is so general that it could hardly be expected to provide the simplest expressions for the equilibrium queue length probabilities for any given form of input. We shall later examine the conjectured form of equilibrium queue length distribution that Finch's heuristic symbolic method suggested for the case of a moving average input of order two.

We find it more convenient to proceed *ab initio* than to try to simplify (1.3) for a moving average input.

Our starting point is the set of recurrence relations expressing the probabilities of the (n+1)th arrival finding a given number of customers already in the queue in terms of queue length as found by the preceding arrival. From these we obtain an equation relating the corresponding probability generating functions, but involving unwanted extra terms which we handle by a complex variable argument, working with Laplace-Stieltjes transforms of the quantities concerned. Having found the functional form of the limiting distribution of queue length by these means, we consider the determination of a finite number of particular constants involved from the initial recurrence relations. These constants do not seem to have a simple form and we do not obtain them explicitly in the general case, although equations are given sufficient to determine their values. The determining procedure is illustrated by a detailed treatment of G(3)/M/1.

2. *Definitions and Preliminaries.*

We shall employ the same notation in subsequent chapters. Capital letters are used to denote random variables and the corresponding lower case letters for particular values taken on by these variables. The (n+1)-tuple $(u_o,u_1,..,u_n)$ is represented by $u^{(n)}$ and the corresponding vector random variable $(U_o,U_1,...,U_n)$ by $U^{(n)}$.

$P_j(u^{(n+p-1)})$, $j \geq o$, is the conditional probability, given $U^{(n)}=u^{(n)}$, that the arrival at A_n finds exactly j customers already in the system. $EP_j(U^{(n+p-1)})$ is the (unconditional) probability that the (n+1)th arrival finds j customers in the system.

The probability, $k_j(x_o,x_1,..,x_p)$, of j departures from the queue during an inter-arrival interval $b_o x_p + b_1 x_{p-1}+..+b_p x_o$, given that at the beginning of the interval the queue length was at least $j+1$, is given by

$$k_j(x_o,x_1,..,x_p) = [\{\mu(b_o x_p+b_1 x_{p-1}+...+b_p x_o)\}^j/j\,!] \times$$
$$\exp\{-\mu(b_o x_p+b_1 x_{p-1}+...+b_p x_o)\},\ j \geq 0.$$

Suppose now that the queue has length $j > 0$ at the beginning of an inter-arrival interval. If the number of departures during the interval is not 0, 1, ..., or $j - 1$, then there must be j departures.

Since

$$\sum_{i=o}^{\infty} k_i(x_o,\ x_1,\ ..\ ,x_p) = 1,$$

it follows that the probability $K_j\ (x_o,\ x_1,..,x_p)$ of j departures during such an interval $b_o x_p + b_1 x_{p-1} + \ldots + b_p x_o$ must be

$$K_j(x_o,x_1,\ldots,x_p) = \sum_{i=j}^{\infty} k_i(x_o,x_1,\ldots,x_p).$$

It is customary to think of this event as decomposing into a countable set of mutually exclusive events, each with j real departures, but with a different number t = 0, 1, 2,... of virtual departures. The probability of j real and t virtual departures is then $k_{j+t}(x_o,x_1,\ldots,x_p)$.

The generating function of the k_i's is

$$k(x_o,x_1,\ldots,x_p;z) = \sum_{i=o}^{\infty} k_i(x_o,x_1,\ldots,x_p)z^i$$

$$= \exp\{-(1-z)\mu(b_o x_p + b_1 x_{p-1} + \ldots + b_p x_o)\}.$$

We denote the generating function of the $P_i(u^{(n+p-1)})$, by

$$P(u^{(n+p-1)};z) = \sum_{i=o}^{\infty} P_{,}(u^{(n+p-1)})z^i, \quad |z| \le 1,$$

and its integral transform by

$$P^*(s^{(p)};z;n) = E[P(U^{(n+p-1)};z) \exp(-s_p U_{n+p-1} - s_{p-1} U_{n+p-2} - \ldots - s_1 U_n)],$$

$$|z| \le 1, \ \text{Re.} s_i \ge 0, \ i=1,\ldots,p.$$

$P_i^*(s^{(p)}, \ , n)$ is defined as the coefficient of z^i in the power series of $P^*(s^{(p)};z;n)$.

We shall also need

$$(2.1) \qquad c_i(u^{(n+p-1)}) = \sum_{j=o}^{\infty} P_j(u^{(n+p-1)})k_{j+i+1}(u_n,u_{n+1},\ldots,u_{n+p}), \ i \ge 0,$$

and its integral transform

$$c_i^*(s^{(p)};n) = E[c_i(U^{(n+p-1)}) \exp(-s_p U_{n+p-1} - \ldots - s_1 U_m)],$$

$$\text{Re.} s_i = 0, \ 1 \le i \le p.$$

We presume that the traffic intensity is less than one, i.e., that

$$(b_o + b_1 + \ldots + b_p) \int_0^{\infty} x \, d\, U(x) > \mu^{-1}$$

or

$$(2.2) \qquad \int_0^{\infty} x \, d\, U(x) > \mu^{-1}.$$

since the b_i sum to unity. By Finch this ensures the existence of

$$P_j = \lim_{n\to\infty} E[P_j(U^{(n+p-1)})],\ j \geq o,$$

and that of

$$P(w_1,w_2,..,w_p;z) = \lim_{n\to\infty} E[P(U_o,U_1,-.,U_{n-1},u_n,u_{n+1},..,u_{n+p-1};z)], |z| \leq 1,$$

where the particular values of $u_n,u_{n+1},..,u_{n+p-1}$ are $w_1,w_2,..,w_p$.

We write for its integral transform

(2.3) $$P^*(s^{(p)};z) = E[P(W^{(p)};z)\exp(-s_pW_p-s_{p-1}W_{p-1}-...-s_1W_1)],$$
$$|z| \leq 1,\ \text{Re. } s_i \geq 0,\ 1\leq i \leq p,$$

where the W_i, $1 \leq i \leq p$, are identically and independently distributed random variables with common distribution function $U(x)$.

$c_i^*(s^{(p)})$, $c(s^{(p)};z)$, are defined by

$$c_i^*(s^{(p)}) = \lim_{n\to\infty} c_i^*(s^{(p)};n),$$

(2.4) $$c(s^{(p)};z) = \sum_{i=o}^{\infty}(1-z^{-i})c_i^*(s^{(p)}),\ |z| \leq 1,\ \text{Re. } s_i \geq 0,\ 1 \leq i \leq p.$$

The function

$$\psi(\alpha) = \int_0^\infty \exp(-\mu\alpha u)d\,U(u),\ \text{Re.}\alpha \geq o,$$

plays as important a rôle in our study of G(p)/M/1 systems as it does in the standard GI/M/1 and M/G/1. We are similarly interested in the root of

(2.5) $$z = \psi(1-z)$$

inside the unit circle. That there exists such a root T and that it is unique follow from Rouché's theorem by virtue of (2.2). The argument is identical to that used in the theory of recurrent queues as is given in Takács's book (1962). Since the complex conjugate of T will satisfy (2.5) if T does, T must be real. It is clear that T must also be positive.

We shall also require later the relation

(2.6) $$P^*(s^{(p)};1) = \psi(s_p/\mu)..\psi(s_1/\mu),\ \text{Re. } s_i \geq 0,\ 1\leq i \leq p.$$

This is a direct consequence of (2.3) since

$$P(W^{(p)};1) = 1.$$

We shall later refer to the following results:

Abel's theorem on the continuity of power series.[1]

If the power series $\sum_{n=o}^{\infty} a_n z^n$ *with radius of convergence unity converges at* $z = 1$*, then*

$$\sum_{n=o}^{\infty} a_n z^n \to \sum_{n=o}^{\infty} a_n$$

as $z \to 1$ *along any path within the circle of convergence which does not touch that circle.*

3. Fundamental Equations.

If the arrival at A_{n+1} is to find $j \geq 1$ customers in the queue, the previous arrival at A_n must find $j - 1$ or more customers in the queue. Suppose that the queue length at A_n is $j+i-1$, $i \geq o$. Then the queue length at A_{n+1} will be j if and only if there are i departures after the arrival at A_n and before the arrival at A_{n+1}. Hence

$$(3.1)\quad P_j(u^{(n+p)}) = \sum_{i=o}^{\infty} P_{j+i-1}(u^{(n+p-1)})k_i(u_n,..,u_{n+p}),\ n \geq 0,\ j \geq 1.$$

Similarly

$$P_o(u^{(n+p)}) = \sum_{i=o}^{\infty} P_i(u^{(n+p-1)})\ K_{i+1}(u_n,u_{n+1},..,u_{n+p}),\ n \geq 0.$$

We note from the definition of the $c_i(u^{(n+p)})$ (2.1)

$$(3.2)\quad \sum_{i=o}^{\infty} c_i(u^{(n+p)}) = P_o(u^{(n+p)}).$$

Forming the product of the power series $k(u_n,u_{n+1},..,u_{n+p};z)$, $P(u^{(n+p-1)};z)$ and using the equations above, we obtain

1. *E.T.COPSON: An introduction to the theory of functions of a complex variable, Oxford Univ. Press (1948), p.100*

$$P(u^{(n+p)};z) + \sum_{i=o}^{\infty} (1-z^{-i})c_i(u^{(n+p)}) + zP(u^{(n+p-1)};z) \times$$

$$\exp [-(1-z^{-1})\mu(b_o u_{n+p}+b_1 u_{n+p-1}+\ldots+b_p u_n)]$$

for $|z| \leq 1$, $z \neq 0$. Hence

(3.3) $$P^*(s^{(p)};z;n+1) = \sum_{i=o}^{\infty} (1-z^{-i})c_i^*(s^{(p)};n+1)$$

$$+ \quad zP^*[(1-z^{-1})\mu b_1+s_{p-1},\ldots,(1-z^{-1})\mu b_{p-1}+s_1,(1-z^{-1})\mu b_p;z;n] \times$$

$$\psi[b_o(1-z^{-1}) + s_p/\mu]$$

for $|z| \leq 1$, $z \neq 0$, Re. $s_i \geq 0(i = 1,2,..,p)$, Re. $[(1-z^{-1})\mu b_j+s_{p-j}] \geq 0$ $(0 \leq j < p)$, Re. $[(1-z^{-1})\mu b_p] \geq 0$. These conditions are satisfied if z lies both in or on the unit circle and outside or on the circle with centre (½, 0) and radius ½, with the origin deleted. We denote by R this domain of the the z-plane.

Letting $n \to \infty$ in (3.3) gives us

(3.4) $$P^*(s^{(p)};z) = c(s^{(p)};z) + z\,\psi[b_o(1-z^{-1}) + s_p/\mu] \times$$

$$P^*[(1-z^{-1})\mu b_1+s_{p-1},\ldots,(1-z^{-1})\mu b_{p-1}+s_1,(1-z^{-1})\mu b_p;z].$$

$$z\epsilon R, \quad \text{Re. } s_i \geq 0, \; 1\leq i \leq p.$$

4. *Solution for $P^*(s^{(p)};z)$*

Substitution of $s_1 = (1-z^{-1})\mu b_p$, $s_2 = (1-z^{-1})\mu(b_{p-1}+b_p)$,.., $s_p = (1-z^{-1})\mu(b_1+..+b_p)$ in (3.4) yields

(4.1) $$P^*[(1-z^{-1})\mu(b_1+..+b_p),(1-z^{-1})\mu(b_2+..+b_p),\ldots,(1-z^{-1})\mu b_p;z]$$

$$= c^*[(1-z^{-1})\mu(b_1+..+b_p),\ldots,(1-z^{-1})\mu b_p:z]\,[1-z\psi(1-z^{-1})]^{-1},$$

$$z\epsilon R.$$

Also, if we replace $s_p,s_{p-1},\ldots,s_1$ by $(1-z^{-1})\mu b_1+s_{p-1}$, $(1-z^{-1})\mu b_2+s_{p-2}$,..,$(1-z^{-1})\mu b_p$ respectively and substitute in (3.4), we obtain

$$P^*[(1-z^{-1})\mu b_1+s_{p-1},..,(1-z^{-1})\mu b_p;z]$$

$$= c[(1-z^{-1})\mu b_1+s_{p-1},..,(1-z^{-1})\mu b_p;z]+z\psi[(1-z^{-1})(b_o+b_1)+s_{p-1}/\mu] \times$$

$$P^*[(1-z^{-1})\mu(b_1+b_2)+s_{p-2},..,(1-z^{-1})\mu(b_{p-1}+b_p),(1-z^{-1})\mu b_p;z].$$

$$z\varepsilon R, \text{ Re. } s_i \geq 0, \ i= 1,2,..,p.$$

By making substitutions in this equation analogous to those in (3.4), and proceeding recursively in this manner, we find that

(4.2) $$P^*(s^{(p)};z) = c^*(s^{(p)};z) + z\psi[b_o (1-z^{-1}) +s_p/\mu] \times$$

$$[c^*\{(1-z^{-1}) \mu b_1+ s_{p-1},......,(1-z^{-1})\mu b_{p-1} +s_1, (1-z^{-1})\mu b_p:z\}$$

$$+ z\psi\{(1-z^{-1}) (b_o+b_1)+s_{p-1}/\mu\} \times$$

$$[c^*\{(1-z^{-1})\mu(b_1+b_2)+s_{p-2},...,(1-z^{-1})\mu(b_{p-1}+b_p),(1-z^{-1})\mu b_p;z\}$$

$$+ z\psi\{(1-z^{-1})(b_o + b_1+b_2) +s_{p-2}/\mu\} \times$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\times [c^*\{(1-z^{-1})\mu(b_1+..+b_p),..,(1-z^{-1})\mu b_p;z\} \ [1-z\psi(1-z^{-1})]^{-1}]]..],$$

$$z\varepsilon R, \text{ Re. } s_i \geq 0, \ i= 1,2,..,p,$$

the last term arising from use of (4.1).

Consider the function $F(s^{(p)};z)$ defined by

$$F(s^{(p)};z) = \begin{cases} (1-Tz)P^*(s;z), & |z| \leq 1, \text{ Re. } s_i \geq 0, \ 1\leq i\leq p, \\ (1-Tz)D(s^{(p)};z), & |z| \geq 1, \text{ Re. } s_i \geq 0, \ 1\leq i\leq p, \end{cases}$$

where $D(s^{(p)};z)$ denotes the right hand side of (4.2).

Since $P^*(s^{(p)};z)$ is the generating function of a probability distribution $P^*(s^{(p)};z)$, and therefore $F(s^{(p)};z)$ must be a regular function of z for $|z| \leq 1$, Re. $s_i \geq 0$, $i = 1,2,...,p$.

Also, as the only zero of $1-z\psi(1-z)$ outside the unit circle is that of $1-Tz$, $F(s^{(p)};z)$ must be a regular function of z for $|z| \leq 1$, Re. $s_i \geq 0, i=1,2..p$.

Hence, by analytic continuation, $F(s^{(p)};z)$ is a regular function of z for all finite z for Re. $s_i \geq 0$, $1 \leq i \leq p$.

From (2.4) and (3.2) it follows by Abel's theorem on the continuity of power series that $c(s^{(p)};z)$ converges to $\sum_{i=1}^{\infty} c_i^*(s)$ as $z\to\infty$, so from (4.2) $\lim F(s^{(p)};z)/z^p$ exists, and is given by

$$\lim_{z\to\infty} \frac{F(s^{(p)};z)}{z^p} = -T\psi(b_o+s_p/\mu)\psi(b_o+b_1+s_{p-1}/\mu)..\psi(b_o+b_1+..+b_{p-2}+s_2/\mu) \times$$

$$[\sum_{i=1}^{\infty} c_i^*\{\mu(b_1+..+b_{p-1})+s_1,\ldots,\mu b_p\}$$

$$-\psi(b_o+b_1+..+b_{p-1}+s_1/\mu)[\sum_{i=1}^{\infty} c_i^*\{\mu(b_1+..+b_p),..,\mu b_p\}]$$

$$[\psi(1)]^{-1}],$$

$$\text{Re. } s_i \geq 0,\ 1 \leq i \leq p.$$

Since a function $\theta(z)$ which is analytic for all finite z and $O(|z|^k)$, k a non-negative integer, as $z\to\infty$ is a polynomial of degree at most k,[1] $F(s^{(p)};z)$ must be of the form

$$F(s^{(p)};z) = F_p(s^{(p)})z^p+F_{p-1}(s^{(p)})z^{p-1}+..+F_o(s^{(p)}),\ \text{Re. } s_i \geq 0,\ 1 \leq i \leq p,$$

where the $F_j(s^{(p)})$ are functions of the s_i alone. Consequently

$$P^*(s^{(p)};z) = [F_p(s^{(p)})z^p+..+F_o(s^{(p)})][1-Tz]^{-1},\ \text{Re. } s_i \geq 0,\quad |z| \leq 1,$$

or, more conveniently,

$$(4.3)\quad P^*(s^{(p)};z) = B_{p-1}(s^{(p)})z^{p-1}+..+B_o(s^{(p)})+B(s^{(p)})(1-zT)^{-1},$$

$$\text{Re. } s_i \geq 0,\quad |z| \leq 1.$$

When we substitute $s_p = s_{p-1} = .. = s_1 = 0$, $P^*(s^{(p)};z)$ becomes the generating function $\sum_{i=o}^{\infty} P_j z^i$ of the limiting distribution of queue size and the functions $B_j(s^{(p)})$ reduce to constants B_j. The generating function of the limiting queue length distribution is thus given by

$$(4.4)\quad \sum_{i=o}^{\infty} P_i z^i = B_{p-1}z^{p-1}+..+ B_o + B.\ (1-zT)^{-1}.$$

1. *E.C.TITCHMARSH: The theory of functions*, Oxford Univ.Press (1932),P.87

This is a probability distribution which assumes a geometric form from P_p onwards, the common ratio being T.

Our result is a natural generalisation of the well known result (e.g. D. G. Kendall (1954)) for the recurrent queueing system GI/M/1, which has a purely geometric limiting distribution $\{(1-T)\,T^j,\ j \geq o\}$. For GI/M/1 also T is the (unique) solution inside the unit circle of

$$z = \psi(1-z)$$

5. *Determination of the $B_j(s^{(p)})$*

From (3.1)

$$P_u(u^{(n+p)}) = \sum_{i=o} \sum_{\ell_o,\ell_1,\ldots,\ell_p=0,1,\ldots i} P_{j+1-1}(u^{(n+p-1)}) \prod_{k=0,1,\ldots,} \left[\exp(-\, b_i u_{n+k}) \frac{(\mu b_k u_{n+k})^{\ell_k}}{\ell_k!} \right] \qquad j \geq 1,$$

where the summation on the ℓ_k is over non-negative integers subject to the restriction $\sum_{k=o}^{p} \ell_k = i$.

Hence

$$P^*_j(s^{(p)},\ ,n+1) = \sum_{i=o}^{\infty} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [P^*_{j+i-1}(\sigma b_1+s_{p-1},\sigma b_2+s_{p-2},\ldots,\sigma b_{p-1}+s_1,\sigma b_p] \times$$

$$\psi[(b_o\ \sigma+s_p)/\mu]_{\sigma=\mu}, \qquad \text{Re. } s_i \geq 0.$$

Letting $n \to \infty$ and using (4.3), we see that for $j \geq p+1$

$$(5.1)\quad B(s^{(p)})T^j = \sum_{i=o}^{\infty} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [B(\sigma b_1+s_{p-1},\sigma b_2+s_{p-2},\ldots,\sigma b_p)T^{j+i-1} \times$$

$$\psi\{(b_o\sigma+s_p)/\mu\}]_{\sigma=\mu}$$

$$= B\{\mu(1-T)b_1+s_{p-1},\mu(1-T)b_2+s_{p-2},..,\mu(1-T)b_p\}T^{j-1} \times$$

$$\psi\{b_o(1-T) + s_p/\mu\}, \quad \text{Re. } s_i \geq 0,$$

whence

(5.2) $$B(s^{(p)}) =$$

$$T^{-p}\psi\{(1-T)b_o+s_p/\mu\}\psi\{1-T)(b_o+b_1)+s_{p-1}/\mu\}\psi\{(1-T)(b_o+b_1+..+b_{p-1})+s_1/\mu\} \times$$

$$B\{\mu(1-T)(b_1+..+b_p),\mu(1-T)(b_2+..+b_p),..,\mu(1-T)b_p\}, \quad \text{Re. } s_i \geq 0.$$

Working similar to the above for $1 \leq j \leq p$ yields

$$B_{p-1}(\mu b_1+s_{p-1},\mu b_2+s_{p-2},..,\mu b_{p-1}+s_1,\mu b_p)\psi(b_o+s_p/\mu) = 0, \quad \text{Re. } s_i \geq 0,$$

$$B_{p-1}(s^{(p)}) = B_{p-2}(\mu b_1+s_{p-1},..,\mu b_p)\psi(b_o+s_p/\mu)$$
$$+(-\mu)\frac{\partial}{\partial\sigma}[B_{p-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]_{\sigma=\mu}, \quad \text{Re.}s_i \geq 0,$$

$$B_{p-2}(s^{(p)}) = [B_{p-3}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}$$
$$+(-\mu)\frac{\partial}{\partial\sigma}[B_{p-2}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]$$
$$+\frac{(-\mu)^2}{2!}\frac{\partial^2}{\partial\sigma^2}[B_{p-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]]_{\sigma=\mu} \text{Re.}s_i \geq 0,$$

. . . .

$$B_1(s^{(p)}) = [B_o(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}$$
$$+ \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ .$$
$$+ \frac{(-\mu)^{p-1}}{(p-1)!}\frac{\partial^{p-1}}{\partial\sigma^{p-1}}[B_{p-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]]_{\sigma=\mu},$$
$$\text{Re. } s_i \geq 0.$$

The now familiar recursive substitution procedure when applied to the second of these equations provides an expression for $B_{p-1}(s^{(p)})$ in terms of B_{p-2} and its derivatives, evaluated at various arguments involving the s_i, similar functions of ψ, and B_{p-1} evaluated at a constant argument. If $B_{p-2}(s^{(p)})$ is known this suffices for the determination of $B_{p-1}(s^{(p)})$

Substituting for $B_{p-1}(s^{(p)})$ (known in terms of B_{p-2}, known functions, and a constant) in the third equation gives an expression for $B_{p-2}(s^{(p)})$ in terms of B_{p-3} and its derivatives, known functions, and a constant.

Proceeding in this fashion expressions are provided for $B_{p-1}(s^{(p)})$, $B_{p-2}(s^{(p)}),\ldots, B_1(s^{(p)})$ in terms of $B_o(s^{(p)})$ and its derivatives, known functions, and a set of constants. Use of these expressions, (5.2) and

(5.3) $$\psi(s_p/\mu)\psi(s_{p-1}/\mu)\ldots\psi(s_1/\mu) = \sum_{i=o}^{p-1} B_i(s^{(p)})+B(s^{(p)})(1-T)^{-1}, \qquad \text{Re. } s_i \geq 0,$$

an equation which results directly from (4.3) and (2.6), leads to a solution for the $B_i(s^{(p)})$ and $B(s^{(p)})$.

Putting $s_i = 0$, $i= 1,2,\ldots,p$, in (4.3) then gives directly the limiting queue length distribution as found by customers entering the system. There does not seem to be a simple general form of solution, but it can be seen that the solution will normally involve the derivatives of ψ as well as ψ itself. We illustrate the solution procedure for G(3)/M/1.

6. Moving average of order three.

In this case $P^*(s^{(2)};z)$ is of the form

$$P^*(s^{(2)};z) = B_1(s^{(2)})z+B_o(s^{(2)})+B(s^{(2)})(1-Tz)^{-1}, \quad \text{Re. } s_i \geq 0,$$

where T is the unique root within the unit circle of

$$T = \psi(1-T).$$

The equations determing the solution become

(6.1) $$B_1(\mu b_1+s_1,\mu b_2) = 0,$$

(6.2) $$B_1(s^{(2)})=\psi(b_o+s_2/\mu)[B_o(\sigma b_1+s_1,\sigma b_2)+(-\mu)\frac{\partial}{\partial\sigma}B_1(\sigma b_1+s_1,\sigma b_2)]_{\sigma=\mu}, \qquad \text{Re. } s_i \geq 0,$$

(6.3) $$B_1(s^{(2)})+B_o(s^{(2)})+B(s^{(2)})(1-T)^{-1} = \psi(s_2/\mu)\psi(s_1/\mu), \quad \text{Re. } s_i \geq 0,$$

(6.4) $B(s^{(2)}) = T^{-2}\psi\{(1-T)b_o+s_2/\mu\}\psi\{(1-T)(b_o+b_1)+s_1/\mu\}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}$,

Re. $s_i \geq 0$.

From (6.3) and (6.4),

(6.5) $$B_1(s^{(2)}) = \psi(s_2/\mu)\psi(s_1/\mu)-B_o(s^{(2)}) - (1-T)^{-1}T^{-2}\psi\{(1-T)b_o+s_2/\mu\}\psi\{(1-T)(b_o+b_1)+s_1/\mu\} \times B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\},\quad \text{Re. } s_i \geq 0.$$

Recursive substitution for B_1 in (6.2) shows us that $B_1(s^{(2)})$ is of the form

(6.6) $$B_1(s^{(2)}) = \psi(b_o+s_2/\mu)[B_o(\mu b_1+s_1,\mu b_2)+a\psi(b_o+b_1+s_1/\mu)],$$

where a is a constant. Substituting $s_1 = \mu b_2$ in (6.5) and making use of (6.1) and (6.6), we find that

(6.7) $$B_o(s^{(2)}) = \psi(s_2/\mu)\psi(s_1/\mu)-(1-T)^{-1}T^{-2}\psi\{(1-T)b_o+s_2/\mu\}\psi\{(1-T)(b_o+b_1)+s_1/\mu\} \times B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}-\psi(b_o+s_2/\mu)a\psi(b_o+b_1+s_1/\mu) - \psi(b_o+s_2/\mu)[\psi(b_1+s_1/\mu)\psi(b_2)-(1-T)^{-1}T^{-2}\psi\{(1-T)b_o+b_1+s_1/\mu\} \times \psi\{(1-T)(b_o+b_1)+b_2\}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}],$$

Re. $s_i \geq 0$,

(6.8) $$B_1(s^{(2)}) = \psi(b_o+s_2/\mu)[a\psi(b_o+b_1+s_1/\mu)+\psi(b_1+s_1/\mu)\psi(b_2) + (1-T)^{-1}T^{-2}\mu\{(1-T)b_o+b_1+s_1/\mu\}\psi\{(1-T)(b_o+b_1)+b_2\}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}],$$

Re. $s_i \geq 0$.

A little algebraic manipulation now enables us to find the two constants a and $B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}$ required for our solutions (6.5), (6.7) and (6.8) for $B(s^{(p)})$, $B_o(s^{(p)})$ and $B_1(s^{(p)})$ to be completely in terms of known quantities. (6.8) and (6.1) yield

(6.9) $$a = -[\psi(1)]^{-1}[\psi(b_1+b_2)\psi(b_2) +(1-T)^{-1}T^{-2}\psi\{(1-T)b_o+b_1+b_2\}\psi\{(1-T)(b_o+b_1)+b_2\}B\{\mu(1-T)(b_1+b_2)\mu(1-T)b_2\}],$$

and using this expression to simplify the working, we derive from (6.8) that

$$(-\mu)[\frac{\partial}{\partial\sigma}B_1(\sigma b_1+s_1,\sigma b_2)]_{\sigma=\mu} = -b_2\psi(b_o+b_1+s_1/\mu)[a\psi'(1)$$
$$+\psi'(b_1+b_2)\psi(b_2)$$
$$+(1-T)^{-1}T^{-2}\psi'\{(1-T)b_o+b_1+b_2)\psi\{(1-T)(b_o+b_1)+b_2\}B\{\mu(1-T)(b_1+b_2)$$
$$\mu(1-T)b_2\}], \qquad \text{Re. } s_1 \geq 0,$$

From (6.6)

$$a = [1+b_2\psi'(1)]^{-1}(-b_2)[\psi'(b_1+b_2)\psi(b_2)$$
$$+(1-T)^{-1}T^{-2}\psi'\{(1-T)b_o+b_1+b_2\}\psi\{(1-T)(b_o+b_1)+b_2\}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}],$$

and so by (6.9)

(6.10)

$$a = [1+b_2\psi'(1) - b_2\psi'\{(1-T)b_o+b_1+b_2\}\psi(1)]^{-1} \times$$
$$[-b_2\psi'(b_1+b_2)\psi(b_2)-\psi'\{(1-T)b_o+b_1+b_2\}\psi(b_1+b_2)\psi(b_2)\psi\{(1-T)b_o+b_1+b_2\}],$$

(6.11) $$T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}$$
$$= (1-T)[\psi'\{(1-T)b_o+b_1+b_2\}\psi\{(1-T)(b_o+b_1)+b_2\}]^{-1} \times$$
$$[-\psi'(b_1+b_2)\psi(b_2)-b_2^{-1}\{1-b_2\psi'(1)\} \times$$
$$\{1+b_2\psi'(1)-b_2\psi'\{(1-T)b_o+b_1+b_2\}\psi(1)\}^{-1} \times$$
$$\{-b_2\psi'(b_1+b_2)\psi(b_2)-\psi'\{(1-T)b_o+b_1+b_2\}\psi(b_1+b_2)\psi(b_2)\psi\{(1-T)b_o+b_1+b_2\}\}]$$

The limiting queue distribution is thus

$$\sum_{i=o}^{\infty}P_iz^i = 1-(1-T)^{-1}\psi\{b_o(1-T)\}\psi\{(1-T)(b_o+b_1)\}T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}$$
$$-a\psi(1)\psi(b_o+b_1)-\psi(1)[\psi(b_1)\psi(b_2)-(1-T)^{-1}\psi\{(1-T)b_o+b_1\} \times$$
$$\psi\{(1-T)(b_o+b_1)+b_2\}T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}]$$
$$+z\psi(1)[a\psi(b_o+b_1)+\psi(b_1)\psi(b_2)+(1-T)^{-1}\psi\{(1-T)b_o+b_1\} \times$$
$$\psi\{(1-T)(b_o+b_1)+b_2\}T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}]$$
$$+(1-Tz)^{-1}\psi(1-T)\psi\{(1-T)(b_o+b_1)\}T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\},$$
$$|z| \leq 1,$$

where the constants a and $T^{-2}B\{\mu(1-T)(b_1+b_2),\mu(1-T)b_2\}$ are given by (6.10) and (6.11).

7. Comparison with Finch's conjectured form of solution for G(2)/M/1.

As we noted before, Finch's result (1.3) determining the equilibrium queue length distribution arising from a general stationary input is not a simple one from which to deduce the distribution corresponding to any particular input.

Finch developed his heuristic symbolic method in an effort to evaluate (1.3) for prescribed inputs. The symbolic method gives the Q_j's as formal Taylor series.

(7.1) $$Q_j = \phi^{j+1}(o, o, .., 1-T_j),$$

where the T_j are operators for which

(7.2) $$T_j Q_r = Q_{r+1}, \qquad j \le k,$$

and

(7.3) $$\phi^{m+1}(\alpha_o, \alpha_1, .., \alpha_m) = E[\exp\{-\alpha_o \theta_{m,o} - \alpha_1 \theta_{m,1} - \dots - \alpha_m \theta_{m,m}\}].$$

For moving averages of orders one and two, the T_j appear as a simple multiplier T satisfying

$$T = \psi(1-T)$$

(in our notation).

The method gives Kendall's (1954) well known geometric limiting distribution for GI/M/1. The solution for G(2)/M/1 is

(7.4) $$Q_o = \psi\{b_o(1-T)\}\,\psi\{b_1(1-T)\},$$

$$Q_j = T^j Q_o, \qquad j \ge 1.$$

By making use of the relations

$$Q_j = \sum_{i=j+1}^{\infty} P_j,$$

(7.4) can be expressed in terms of the P_j as

(7.5) $$P_j = \begin{cases} 1-\psi\{b_o(1-T)\}\psi\{b_1(1-T)\}, & j=o \\ T^{j-1}(1-T)\psi\{b_o(1-T)\}\psi\{b_1(1-T)\}, & j\ge 1 \end{cases}$$

Thus the symbolic method suggests a limiting distribution which is geometric apart from the first term, as we have shown to hold. The values of P_o given by the two approaches are, however, as we shall see, different, so that whilst the symbolic method gives the form of solution correctly, it predicts the constant P_o incorrectly.

The correct limiting queue length distribution of G(2)/M/1 can be derived as follows:

For p = 1, the equations for the unknown functions B(s), $B_o(s)$ reduce to

$$B(s) = T^{-1}\psi\{b_o(1-T)+s/\mu\}\ B\{\mu b_1(1-T)\}, \quad \text{Re. } s \geq 0,$$

$$\psi(s/\mu) = B_o(s) + B(s)\ (1-T)^{-1}, \quad \text{Re. } s \geq 0,$$

$$B_o(\mu b_1) = 0.$$

The solution for B(s), $B_o(s)$ is

$$\left.\begin{aligned} B(s) &= (1-T)\psi(b_1)\psi\{(1-T)b_o+s/\mu\}\ /\ \psi\{(1-T)b_o+b_1\} \\ B_o(s) &= \psi(s/\mu) - \psi(b_1)\psi\{(1-T)b_o+s/\mu\}\ /\ \mu\{(1-T)b_o+b_1\} \end{aligned}\right\} \quad \text{Re. } s \geq 0.$$

The equilibrium queue length distribution is now obtained on setting s = o.

$$(7.6) \qquad P_j = \begin{cases} 1-T\psi(b_1)\psi\{(1-T)\ b_o\}\ /\ \psi\{(1-T)b_o+b_1\}, & j = o \\ T^j(1-T)\psi(b_1)\psi\{(1-T)b_o\}\ /\ \psi\ \{(1-T)b_o+b_1\}, & j > o. \end{cases}$$

The expressions for P_o in (7.5), (7.6) are apparently different. That they are actually different can be verified by taking a particular case.

If

$$U(x) = 1-\exp(-\lambda x),\ x \geq 0,\ \lambda > 0$$

we have

$$\psi(\alpha) = \lambda/(\lambda+\alpha\mu)$$

so that (2.5) yields immediately $T = 1$ or λ/μ, of which only the latter can lie within the unit circle. With $T = \lambda/\mu$, (7.5), (7.6) yield respectively

$$P_o = 1 - \lambda^2 (\lambda b_1+\mu b_o)^{-1} (\lambda b_o+\mu b_1)^{-1} ,$$

$$P_\infty = 1 - \lambda^2 (\lambda b_1+\mu) [\mu(\lambda+b_1\mu)(\lambda b_1+\mu b_o)]^{-1} .$$

It is easily seen that these two values of P_o differ unless either one of b_o, b_1 vanishes or $\lambda=\mu$. The first possibility corresponds to a general recurrent input. $\lambda=\mu$ is inadmissible since for ergodicity the traffic intensity λ/μ must be less than one (Loynes, 1962).

The identification of the T_j as a simple scalar multiplier T is not available to us with higher order moving averages since (7.1), (7.3) give

(7.7)
$$Q_j = \begin{cases} \psi[b_o(1-T_0)]\psi[b_1(1-T_o)]..\psi[b_p(1-T_o)] & , j = o, \\ \psi[b_o(1-T_1)]\psi[(b_o+b_1)(1-T_1)]\psi[(b_1+b_2)(1-T_1)]..\psi[(b_{p-1}+b_p)(1-T_1)] \times & \\ \qquad \psi[b_{p-1}(1-T_1)] & , j= 1, \\ \psi[b_o(1-T_2)]\psi[(b_o+b_1)(1-T_2]\psi[(b_o+b_1+b_2)(1-T_2]..\psi[b_p(1-T_2)], & \\ & j = 2, \\ \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . & \\ \psi[b_o(1-T_j)]..[\psi(1-T_j)]^{j+1-p}\psi[(b_1+..+b_p)(1-T_j)]..\psi[b_p(1-T_j)], & \\ & j \geq p. \end{cases}$$

Such an identification seems plausible for $j \geq p$, and on comparison of (7.7) with (7.2) we see that the scalar T would satisfy

$$T = \psi(1-T).$$

From the relation

$$Q_j = \sum_{i=j+1}^{\infty} P_j$$

we readily derive that $\{P_j\}$ would then have the known ultimately geometric form with common ratio T the solution of

$$T = \psi(1-T)$$

inside the circle. That we must choose this solution arises from the necessity for convergence of the Q_j.

It seems that any interpretation for $T_o,..,T_{p-1}$ would be rather less simple since, as we have seen in our discussion of the full solution of $G(p+1)/M/1$, the probabilities $P_o,..,P_{p-1}$ will in general involve the derivatives of ψ as well as ψ itself.

8. Waiting Time Distribution.

Loynes (1962a) has considered the possibility of determining the stationary waiting time distribution of single server queues in which inter-arrival intervals and service times are not necessarily independently distributed, and under mild restrictions found techniques applicable to a wide class of queueing systems.

In this section we deduce the form of the limiting waiting time distribution for the general moving average queue with negative exponential service and compare this with Loynes's results.

We denote by S_n, T_n, W_n, respectively the service time of the arrival at A_n, the length of the interval (A_n, A_{n+1}), and the waiting time (excluding service) of the arrival at A_n.

Loynes (1962) has shown that under the conditions that $\{S_n - T_n\}$ is a strictly stationary process and

(8.1) $$E(S_n - T_n) < 0,$$

the existence of a unique limiting distribution of waiting time is ensured. In the present problem this condition becomes (2.2), our condition for the existence of a unique limiting distribution of queue length, as one would intuitively expect.

The class of systems dealt with in Loynes (1962a) consists of queues for which :

$\mathcal{H}$; *There exists a sequence* $\{z_n\}$ *of random vectors defined in finite-dimensional Euclidean space with the following properties:*

(i) $\{ z_n, T_n, S_n\}$ *is a strictly stationary process,*

(ii) $S_n, T_n W_n$ *are conditionally independent given* z_{n-1}, z_n,

(iii) W_n, z_n *are conditionally independent given* z_{n-1}.

(One can regard the components of the z's as being of the nature of the additional variables introduced in a queueing problem to recover the Markovian property, as in D. G. Kendall (1954).)

We introduce

$$\phi(s, z_n) = \int_0^\infty \exp(-sx) d_x \, pr(W_{n+1} \leq x | z_n),$$

and similarly $\psi(s,z_n)$, $H(s,z_n,z_{n-1})$, $G(s,z_n,z_{n-1})$ corresponding to $W_n + S_n + T_n - W_{n+1}$, S_n, T_n, respectively.

Loynes shows that the Laplace-Stieltjes integral form of the equation here corresponding to the ordinary stationary waiting time integral equation is

$$(8.2) \quad 1 - \psi(s,z_n) = \phi(s,z_n) - E[\phi(s,z_{n-1}) H(s,z_n, z_{n-1}) \times G(-s,z_n,z_{n-1}) | z_n].$$

This equation is set up only for s on the imaginary axis, but it is often possible to continue H and ϕ analytically into the left half plane. Presuming H can be so continued to give a single valued function analytic everywhere in the left half plane, except for isolated singularities, the following theorem is derived:

If (8.2) has a solution $\beta(s,z_n)$ *such that*

(i) $\beta(s,z_n)$ *is, for fixed* z_n, *the analytic continuation of* $\beta(s,z_n)$,

(ii) $\exists\, a(z_n)$ *such that, for fixed* z_n, $\lim_{s\to\infty} \exp(as)\beta(s,z_n)/s$ *exists with value zero (in the left half plane),*

and

(iii) for fixed z_n, *the analytic function composed of* $\phi(s,z_n)$ *and* $S(s,z_n)$ *is regular everywhere except for poles, then for* $x \geq a$, $pr\,(W_{n+1} \leq x | z_n) - 1$ *is a*

finite sum of terms of the form

(8.3) $\sum_{r=0}^{k-1} g_r(z_n)\, x^r \exp(-bx)$,

where $-b$ *is a pole of* β *of order* k. *These poles may depend on* z_n, *but in any case* $Re.\ b \geq 0$.

It is readily verified that $z_n = (u_{n+p}, u_{n+p-1}, \ldots, u_n)$ suffices for $\mathcal{H}$ to be satisfied.

With negative exponential service of parameter μ and the above choice of the z's, $H(s, z_n, z_{n-1})$ becomes $\mu(\mu+s)^{-1}$, independent of z_n, z_{n-1}.

A subsidiary result of Loynes (1962a) gives that the conditions (i) and (ii) of the main theorem are satisfied with $a = 0$ when H is a rational function of s and is independent of the z's.

We now derive the form of the (unconditional) limiting waiting time distribution directly from (4.4).

If an arrival finds the queue empty, he begins service immediately. If on arriving he finds $j > 0$ customers already in the queue, then

$$\begin{aligned} \text{pr (waiting time} \leq x) &= \text{Pr (j services completed in the time} \leq x) \\ &= 1 - \exp(-\mu x) \sum_{i=o}^{j-1} (\mu x)^i / i!, \quad x \geq 0. \end{aligned}$$

Hence using (4.4), the (unconditional) waiting time distribution for an arrival is

$$\begin{aligned} (8.4) \quad Pr(W \leq x) &= P_o + \sum_{j=1}^{\infty} P_j [1 - \exp(-\mu x) \sum_{i=o}^{j-1} (\mu x)^i / i!] \\ &= 1 - \exp(-\mu x) \sum_{j=o}^{p-2} \left(\sum_{i=j+1}^{p-1} B_i \right) (\mu x)^j / j! \\ &\quad - BT(1-T)^{-1} \exp\{-\mu x(1-T)\}, \quad p \geq 2, \quad x \geq 0. \end{aligned}$$

This is the sort of expression that would arise from (8.3) on integrating out z_n if $\beta(s, z_n)$ were in fact analytic everywhere except for poles at $-\mu$, $-\mu(1-T)$ or orders $p-1$ and 1 respectively, both independent of z_n.

That $-\mu$ should be a pole seems natural from (8.2), since

$$\phi(s,z_{n-1}) = \mu(\mu+s)^{-1}$$

has a pole at $s = -\mu$. The possibility is left as a hypothesis.

We observe that when $p = 1$, i.e., when we have a general recurrent input, the terms in (8.4) involving the B_i 's do not appear, and the distribution becomes negative exponential together with a weight at the origin, a fact noted by Smith (1953).

9. Approach through waiting times.

We noted in the previous chapter that the queueing process is more complex than the waiting time process and cannot, in general, be deduced from it. Such a deduction can, however, be made when the service time distribution is negative exponential, thanks to the peculiar memory-less property with this distribution.

Suppose the limiting queue length distribution of a queueing system with negative exponential service to be $\{P_j, j \geq 0\}$. Equation (8.4) gives for the waiting time distribution $\Pr(W \leq x)$.

$$\Pr(W \leq x) = P_o + \sum_{j=1}^{\infty} P_j \left[1 - \exp(-\mu x) \sum_{i=o}^{j-1} (\mu x)^i / i!\right]$$

$$= 1 - \exp(-\mu x) \sum_{j=o}^{\infty} \left(\sum_{i=j+1}^{\infty} P_j\right)(\mu x)^j / j!.$$

Therefore

$$\sum_{j=o}^{\infty} \left(\sum_{i=j+1}^{\infty} P_j\right)(\mu x)^j / j! = [1 - \Pr(W \leq x)] \exp(\mu x).$$

If $\Pr(W \leq x)$ is a known function, $F(x)$ say, this relation will enable us to find the distribution $\{P_j\}$. Cauchy's theorem gives

$$\sum_{i=j+1}^{\infty} P_i = \frac{j!}{2\pi i} \mu^{-j} \oint z^{-(j+1)} [1-F(z)] \exp(\mu z)\, dz, \quad j \geq 0,$$

where the integration is performed around a small closed contour about the origin. Hence

$$P_j = \begin{cases} 1 - \dfrac{\mu}{2\pi i} \oint (\mu z)^{-1} [1-F(z)] \exp(\mu z)\, dz, & j = o, \\ \dfrac{\mu(j-1)!}{2\pi i} \oint (\mu z)^{-j} [1-F(z)] \exp(\mu z)\, dz & \\ \qquad - \dfrac{\mu j!}{2\pi i} \oint (\mu z)^{-(j+1)} [1-F(z)] \exp(\mu z)\, dz, & j \geq 1. \end{cases}$$

The determination of the stationary waiting time distribution $F(x)$ could be carried out by the techniques of Beneš shows that sufficient information to determine the waiting time distribution at time t is contained in $K(t)$, the sum of the service times of all the arrivals to the system before t. He gives forms of solution which are integral equations in the functions $Pr(K(t) \leq w)$ and $R(t,u,w)$, defined by

$$R(t,u,w) = Pr[\{K(t) - t\} - \{K(u) - u\} \leq w \mid W(u) = o].$$

The handling of these functions seems substantially harder than the procedure we have adopted in this chapter, and we shall not pursue these possibilities further.

* * * * * *

CHAPTER THREE.

Variety of equilibrium queue size distribution occurring in $G(p')/M/1$ Systems.

In a paper reviewing the then current work in queueing theory, D. G. Kendall (1964) draws attention to the problem of identifying a queueing system from its output.

In chapter II we saw that for single-server queues with negative exponential services and individual arrivals it is possible to obtain limiting distributions of queue size which are not purely geometric.

It is thus natural to consider a problem similar to Kendall's, namely: what is the range of equilibrium distributions that can arise in single-server queues with negative exponential services and individual arrivals? Because of the difficulty of deriving the limiting distribution explicitly this question does not appear to admit of an easy solution. We are, however, able to gain some limited insight as to the answer.

1. *The queue $G/M/1$.*

We consider first the distributions arising from general recurrent inputs. We know that if

$$\mu m > 1 \quad (m \text{ finite}),$$

where m is the mean inter-arrival time, then the limiting queue length distribution as found by arrivals is given by

$$P_j = \Pr(\text{queue length} = j) = (1-T)T^j, \quad j \geq 0.$$

T is the (demonstrably unique and positive) root of

$$z = \psi(1-z)$$

inside the unit circle, and ψ the Laplace - Stieltjes transform of the inter-arrival interval probability distribution function.

The distribution $\{P_j\}$ will define the input uniquely only if the inner root T arises from but one inter-arrival interval distribution function. This is never the case, as we shall show. It will also appear that every geometric probability distribution

$$\{(1-T)T^j\} \quad , \qquad 0 < T < 1,$$

is the limiting distribution of queue length as found by arrivals for some general recurrent input.

We note that if $\{A_i(x)\}$ is a set of (proper) non-negative distribution functions and $\{c_i\}$ a set of non-negative constants with sum unity, then $\Sigma_i\ c_i\ A_i(x)$ is also a non-negative distribution function. If m_i are the (finite) means of the A_i, the mean of $\Sigma_i\ c_i\ A_i(x)$ is $\Sigma_i\ c_i m_i$ which will also be finite.

For $\mu > 0$ and T satisfying $0 < T < 1$ given, define

$$b_i = \int_0^\infty \exp\,[-\mu(1-T)x]\ d\,A_i(x)\,.$$

$\Sigma c_i A_i(x)$ will be an inter-arrival interval distribution function giving rise to the equilibrium distribution $\{(1-T)T^j\}$ provided both

$$\Sigma c_i m_i > \mu^{-1} \tag{1.1}$$

and

$$\Sigma c_i b_i = T. \tag{1.2}$$

It seems plausible that if $\{A_i\}$ contains many members and these exhibit a wide variety of distribution of mass, then (1.1), (1.2) can be simultaneously satisfied by some set of non-negative constants $\{c_i\}$ for which

$$\sum_i c_i = 1. \tag{1.3}$$

In fact, it would seem that for a large and varied collection A_i the c's could be chosen in many ways giving many different distributions $\Sigma_i\ c_i\ A_i(x)$.

Such a set $\{A_i\}$ is provided by the Erlang distributions of different orders associated with a given parameter μ. The Erlang distribution of order i is

$$A_i(x) = \begin{cases} 0, & x \leq 0 \\ 1 - \sum_{\ell=o}^{i-1} \exp[-\mu x](\mu x)^{\ell}/\ell! \end{cases}, \quad i \geq 1,$$

and the corresponding probability density,

$$a_i(x) = \begin{cases} 0, & x \leq 0 \\ \exp[-\mu x](\mu x)^{i-1}/(i-1)! \end{cases},$$

is unimodal with a peak at $(i-1)/\mu$. There is thus a peak at the origin (for $i = 1$) and peaks extending out to infinity with i increasing. For this set $\{A_i\}$ we have

(1.4) $$m_i = i/\mu,$$

(1.5) $$b_i = (2-T)^{-i}.$$

For m_i of this form, it is evident that (2.1) will automatically be satisfied if at least one A_i other than A_1 occurs in $\Sigma_i\, c_i A_i$.

Since $\{(2-T)^{-n}\}$ is strictly monotone decreasing and bounded below by zero for $0 < T < 1$, $(2-T)^{-n} < T$ for all sufficiently large n, for $n \geq N$, say. Also, $(2-T)^{-1} > T$ for $0 < T < 1$, by elementary algebra, so there is always at least one value of n for which

$$(2-T)^{-n} > T .$$

If

(1.6) $$T = (2-T)^{-n}$$

for some $n \geq 2$, then clearly A_n suffices for $\Sigma c_i A_i$. More generally we will have

$$(2-T)^{-n} > T$$

for $1 \leq n \leq N$, and the inequality is reversed for $n > N (n > N + 1)$ when (1.6) can be satisfied).

Consider any ordered pair (i,j) of positive integers, $i \le N < j$ ($j > N + 1$ when (1.6) can be satisfied). By an elementary intermediate value theorem there is precisely one c, $0 < c < 1$ for which

$$c(2-T)^{-i} + (1-c)(2-T)^{-j} = T.$$

(1.2) is thus satisfied when we take

$$c\,A_i + (1-c)\,A_j$$

as our distribution function. From the freedom of choice of i, j, it is clear that there is a countable infinity of distribution functions formed as a linear combination of the $\{A_i\}$ which suffice for (1.1) - (1.3) to be satisfied, i.e., which can give rise to the prescribed geometric probability distribution.

In any case we have open to us the possibility of augmenting our set $\{A_i\}$ with further distribution functions. We could, in particular, make use of sets of Erlang distributions with parameters other than μ. Discussions of approximating general distributions by linear superposition of Erlang distributions are given by Jensen (1954) and others.

A salient characteristic of the distribution functions formed from Erlang distributions in this way is that their Laplace-Stieltjes transforms are the reciprocals of polynomials (meromorphic functions if we allow combinations of infinitely many of our basic distributions) and have their zeros on the negative real axis of the complex s-plane. It is possible to generalise to the reciprocals of polynomials with pairs of complex conjugate pseudo-negative zeros (i.e., zeros with negative real parts) if we allow complex probabilities. Such a possibility has been considered in some detail and validated by Cox (1955) and (1955a). By virtue of the partial fraction expansion this in fact includes rational functions. The numerator of a fraction whose denominator is a polynomial of degree k cannot, however, be a polynomial of degree greater than k, since we would then have a Laplace transform which was unbounded on the positive real axis.

Necessary conditions for a rational function to be the Laplace transform of a random variable have been derived by Lukács and Szász in a number of papers (see Lukács and Szász (1952) and (1954)).

2 *The queue G(p)/M/1.*

Suitable combinations $\Sigma c_i A_i$ using more than two A's can also be formed, whereby $\Sigma c_i A_i$ can be made to accommodate further conditions. In particular, by choosing sufficient A's we can make $\Sigma c_i A_i$ as smooth as we please, and, although we shall not attempt a proof, it seems reasonable to suspect that we can approximate a distribution function lacking finite jumps with an arbitrarily 'close' fit in some sense.

That we should be able to so approximate is natural in that A_1, the negative exponential distribution, is 'completely random' or 'memory-less', while, as $i \to \infty$, A_i approaches a distribution function which is a delta-measure, i.e., a completely deterministic distribution function. The profiles of the functions A_i show a regular gradation between these extremes.

Observing in the discussion on G(p+1)/M/1 that the limiting distributions obtained depend on only a finite number of ψ and its derivatives evaluated for particular arguments, we are inclined to believe that if a given limiting distribution which is geometric apart from the first few terms (we shall term this a 'delayed' geometric distribution) occurs it will arise from many different moving average inputs (of a particular order). On the other hand it will not, in general, be true that a delayed geometric distribution will necessarily arise from a moving average input. This we shall illustrate for the case of a moving average input of order two, where we see that some geometric distributions delayed by one term do not arise from second order moving averages of functions $\Sigma c_i A_i$.

A more complete discussion appears impracticable in view of the difficulty of an explicit determination of the limiting distribution arising from a general order moving average.

One would perhaps be surprised if there were not some restriction on the limiting distributions occurring in G(p)/M/1 systems. Clearly any queueing distribution can be approximated to as closely as desired by a sufficiently delayed geometric distribution, and it is hardly to be expected that any queueing distribution can be simulated by the limiting behaviour of a system with random services and non-negatively correlated inter-arrival intervals. We shall return to this in section six of chapter four.

Our solution for the equilibrium distribution of G(2)/M/1 was

$$(2.1) \quad P_j = \begin{cases} 1-T\psi(b_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o+b_1\}, & j = 0, \\ T^{j-1}(1-T)\psi(b_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o+b_1\}, & j \geq 1. \end{cases}$$

We find that for a given T, a moving average of functions $\Sigma c_i A_i$ will give

$$(2.2) \qquad P_o \geq 1 - T.$$

We begin with the following lemma.

Lemma: *Suppose the quantities* a_i, a_i', $i = 1,2,3$, *satisfy*

$$(2.3) \quad \begin{cases} 0 < a_i \;, \\ 0 < a_i' \;, \\ a_i' < a_i, \\ a_1' \; a_2'/a_3' \leq a_1a_2/a_3. \end{cases}$$

Then for any constants c, c' *in* (0,1) *satisfying*

$$c + c' = 1$$

we have

$$(2.4) \qquad a_1a_2/a_3 > (ca_1+c'a_1')(ca_2+c'a_2')/(ca_3+c'a_3').$$

The result follows from elementary calculus by considering the variation of

$$[c+(1-c)a_1'/a_1][c+(1-c)a_2'/a_2]/[c+(1-c)a_3'/a_3]$$

for c in $(0,1)$.

Proof: The second order moving average (b_o, b_1) of the distribution function A_ℓ has

$$(2.5)\qquad \begin{cases} \psi(b_1) & = (1+b_1)^{-\ell}, \\ \psi\{(1-T)b_o\} & = (1+(1-T)b_o)^{-\ell}, \\ \psi\{(1-T)b_o+b_1\} & = (1+(1-T)b_o+b_1)^{-\ell}. \end{cases}$$

Consider the moving average (b_o, b_1) of any finitely compounded distribution function $\Sigma c_\ell A_\ell$. This is the same as the sum of the moving averages (b_o, b_1) of the A_ℓ each weighted by c_ℓ, so that for the moving average (b_o, b_1) of $\Sigma c_\ell A_\ell$

$$\psi(b_1) = \Sigma c_\ell (1+b_1)^{-\ell},$$
$$\psi\{(1-T)b_o\} = \Sigma c_\ell (1+(1-T)b_o)^{-\ell},$$
$$\psi\{(1-T)b_o+b_1\} = \Sigma c_\ell (1+(1-T)b_o+b_1)^{-\ell}.$$

Suppose that for $\Sigma c_\ell A_\ell$

$$F(\Sigma c_\ell A_\ell) \equiv \psi(b_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o+b_1\}$$

is less than the corresponding function for A_m, where m is no greater than the least index ℓ in $\Sigma c_\ell A_\ell$. Since the terms of the right hand side of (2.5) are all strictly monotone decreasing with ℓ increasing, the inequalities (2.3) are satisfied when we take the $\psi(b_1), \psi\{(1-T)b_o\}, \psi\{(1-T)b_o+b_1\}$ associated with A_m as a_1, a_2, a_3 respectively and the corresponding functions for $\Sigma c_\ell A_\ell$ as the a''s. It follows from the result (2.4) of the lemma that (2.3) is also satisfied when we now take for the a''s the three ψ's associated with

$$c\,A_m + c'\Sigma c_\ell A_\ell \quad ,$$

where

$$c+c' = 1, \quad c, c' > 0.$$

This provides the basis for an inductive proof that the function F evaluated for any finite sum $\Sigma c_\ell A_\ell$ is less than the corresponding value for A_m, where m does not exceed the least index in $\Sigma c_\ell A_\ell$. We use for the induction the fact that (from (2.5))

$$(2.6) \qquad F(A_\ell) = [1+b_1+(1-T)b_0]^{\ell}[1+b_1+(1-T)b_0]^{-\ell},$$

so that $\{F(A_\ell)\}$ is a strictly monotone decreasing sequence.

Hence F evaluated for any finitely compounded sum $\Sigma c_\ell A_\ell$ (c_ℓ's positive with sum unity) is less than $F(A_1)$.

By taking the supremum on n of F evaluated for sums involving n components A_ℓ, the maximality of $F(A_1)$ is established for denumerably infinite sums $\Sigma c_\ell A_\ell$.

Since

$$F(A_1) < 1$$

(2.2) follows directly from (2.1).

The value obtained for A_1 itself is extreme in that a moving average of A_1 will not satisfy the stability criterion (1.1) for an equilibrium situation to exist.

The restriction that P_0 lie in $(1-TF(A_1),1)$ was obtained for a given moving average (b_0, b_1). We see from (2.6) that by allowing b_0 to range between 0 and 1 (keeping $b_0+b_1 = 1$ throughout) we can make only the weaker statement

$$P_0 \geq 1 - T$$

if we leave unprescribed the constants of the second order moving average.

Consider again (b_0, b_1) fixed. We observe that

$$F(A_1) < 1$$

and

$$F(A_\ell) = [F(A_1)]^\ell \quad ,$$

so that $F(A_\ell) \to 0$ as $\ell \to \infty$. By continuity,

$$F[cA_\ell + (1-c)A_m], \quad \ell \neq m,$$

will, as c varies between 0 and 1, take on every value between $F(A_\ell)$ and $F(A_m)$. It follows that F evaluated for sums $\Sigma c_\ell A_\ell$ of only two of the A_ℓ can take on every value in $(0, F(A_1))$. It seems reasonable to expect that analogously to the case of recurrent inputs, we can, by taking sufficiently many A_ℓ's, construct a sum $\Sigma c_\ell A_\ell$ with some second order moving average (b_o, b_1), which can give rise to a given delayed geometric distribution

$$P_j = \begin{cases} 1 - TK, & j = 0, \quad (0 < K < 1), \\ T^{j-1}(1-T)K, & j \geq 1. \end{cases}$$

The conditions we need to satisfy are the equilibrium condition (1.1) (which will be trivially satisfied as before if $\Sigma c_\ell A_\ell$ contains at least one A_ℓ other than A_1), and the conditions

(2.7) $$F(\Sigma c_\ell A_\ell) = K,$$

(2.8) $$T = \psi(1-T).$$

As before we expect many solutions for a given pair T, K. The lower bound 1-T for P_o can arise in an equilibrium distribution. The argument above considered only distributions that are saltus-free. Consider a process whose successively realised values constitute a moving average (b_o, b_1) of a sequence $\{S_n\}$ of random variables, the mass of each of which is concentrated at a single point. It is trivial that this process is the same as the renewal process $\{S_n\}$. The limiting distribution, when it exists, is then purely geometric, and if T is the common ratio, we have trivially that

$$P_o = 1 - T.$$

For any prescribed T in (0, 1), let us take a deterministic input for which the common inter-arrival d is

$$d = \mu^{-1} (1-T)^{-1} \log (T^{-1}).$$

By an elementary inequality

$$(1-T)^{-1} \log (T^{-1}) > 1$$

for $0 < T < 1$, so that $d > \mu^{-1}$ and an equilibrium distribution exists. Since

$$T = \exp [-(1-T)d\mu] = \int_0^\infty \exp [-\mu(1-T) x] \delta(d) dx,$$

the limiting distribution is, in fact, geometric with parameter T.

We have as a corollary that all purely geometric distributions as considered earlier can be produced by completely deterministic inputs consisting of regularly spaced arrival instants.

We observe that a process consisting of a moving average (b_o,b_1) of identical distribution functions in each of which the mass is distributed between several points differs from the completely deterministic case in that it is not the same as the corresponding renewal process. A similar argument holds good for the general order moving average with a deterministic input, so that the value $1-T$ for P_o can be realised for moving average inputs of every order.

We extend the result

$$P_o \geq 1-T$$

to systems whose inputs are a moving average (b_o, b_1) of random variables with distribution functions of the form

$$c\, A + (1-c)D, \quad 0 \leq c \leq 1,$$

where A can be written $\Sigma c_\ell A_\ell$ and D is a distribution function all of whose mass is at a (not necessarily finite) number of discrete points. Clearly a linear combination of such distribution functions is another of the same type.

The result is simply established. We deal first with inputs where $c = 0$ and D involves only a finite number of points.

D is itself a combination of distribution functions each with all its mass at a single point. We do not stipulate that each such component must necessarily give rise to an equilibrium distribution, or indeed that this need happen for any component. This result follows purely from the mathematical properties of F.

Select out the component distribution function D_1 with the least inter-arrival time d. Then

$$\phi(b_1) = \exp[-\mu b_1 d],$$

$$\phi\{b_o(1-T)\} = \exp[-\mu b_o(1-T)d],$$

$$\phi\{(1-T)b_o+b_1\} = \exp[-\mu\{(1-T)b_o+b_1\}d],$$

so that $F(D_1)$ has the value unity. Much as before, it follows from the lemma by an easy induction that

$$F(D) \leq F(D_1) = 1.$$

By considering the supremum of F(D) we can extend the result

$$F(D) \leq 1$$

to distributions D whose mass is distributed on an infinity of discrete points (by a standard result in probability theory the number of discontinuity points in D must be countable[1]).

A further simple application of the lemma finally gives the required result for the moving average (b_o,b_1) of $c\,A + (1-c)D$, as defined above.

The interpretation of this result is quite striking. Since the geometric distribution arising from a general recurrent input process and its moving average (b_o,b_1) are characterised by the same value of T, we see that in the moving average process, arrivals are in general more likely to find the queue empty than arrivals in the general recurrent process, and less likely to find any other given number of customers already in the queue. The probabilities coincide trivially for a deterministic input.

Since the waiting time distribution of an arrival finding $j > 0$ customers already in the queue is simply the jth iterated convolution of the

1. *M. Loève: Probability Theory, Van Nostrand Co., New York, (1955), ch. 4*

negative exponential distribution, i.e., an Erlang distribution of order j, it also follows that the probability that an arrival has to wait as long as $x(> 0)$ before commencing service is less for the (b_o,b_1) moving average process than for the corresponding uncorrelated input process. This is despite the fact that the traffic intensity, i.e., the ratio of the mean service time to the mean inter-arrival interval, is the same for both processes.

In calculating the ratio of mean waiting time to mean service time for the queues M/M/1, D/M/1 for various traffic intensities, Kendall (1953), found that much lower values were obtained from the deterministic input for a given traffic intensity. Since taking the moving average (b_o,b_1) introduces a positive correlation between lengths of successive inter-arrival intervals, one may think of the moving average as being in a sense closer to a deterministic input. The reduction in the mean waiting time/ mean service time ratio is thus not altogether unexpected, and although the result cannot be readily demonstrated, we would expect similar behaviour with moving averages of higher order.

3. *Traffic intensity and the equilibrium distribution.*

We derive in this section, a somewhat surprising result that is readily demonstrated by our approach of constructing inputs from a superposition of Erlang distributions.

Theorem: *A given geometric equilibrium queue-length distribution $\{(1-T)T^j\}$ can arise from recurrent inputs with arbitrarily large mean inter-arrival intervals.*

Proof: We have shown that a prescribed limiting distribution $\{(1-T)T^j\}$ can be produced by a recurrent input with an inter-arrival time distribution function

$$(3.1) \qquad c\,A_1 + (1-c)A_\ell, \quad 0 < c < 1. \quad \ell > 1.$$

We give an iterative procedure for producing a sequence of distribution functions, all associated with the same limiting distribution but each with a mean inter-arrival time exceeding that of the previous distribution by more than a fixed positive amount.

The first step involves replacing A_ℓ in (3.1) by

$$(3-T)^{-1} A_{\ell-1} + (2-T)(3-T)^{-1} A_{\ell+1}.$$

It is immediate that

$$(1-c)(3-T)^{-1} (2-T)^{-(\ell-1)} + (1-c)(2-T)(3-T)^{-1} (2-T)^{-(\ell+1)}$$
$$= (1-c)(2-T)^{-\ell},$$

so that by (1.2), (1.5) the new distribution gives rise to a limiting queue length distribution with the same value of T as the old.

It is also immediate that

$$(\ell-1)\mu^{-1} (1-c)(3-T)^{-1} + (\ell+1)\mu^{-1} (1-c)(2-T)(3-T)^{-1}$$
$$= \ell\mu^{-1} (1-c)+\mu^{-1}(1-T)(3-T)^{-1}(1-c),$$

so that the mean inter-arrival interval of the new process exceeds that of the old by

$$\mu^{-1}(1-c)(1-T)(3-T)^{-1}.$$

Suppose a sequence of further distribution functions is constructed by a like continued splitting at each stage of each A_i for which $i > 1$. If at any stage the distribution function

$$c_1 A_1 + \Sigma c_i A_i$$

is so split, where $\Sigma c_i A_i$ contains only A's for which $i > 1$, the new distribution function will have a mean increased by

$$\mu^{-1} \Sigma c_i (1-T)(3-T)^{-1}.$$

That Σc_i is bounded below by a positive constant, i.e., that not all the probability ultimately passes into A_1, can be seen as follows:

The splitting process in the coefficients c can be regarded as a random walk on the positive integers in discrete time. There is an absorbing state at one and the probabilities p, q of steps of one to the right and left at any integer are $(2-T)(3-T)^{-1}$, $(3-T)^{-1}$ respectively. Since $0 < T < 1$, the probability of a step to the right is the greater, and it is a standard result [2] that for such a random walk beginning at $\ell > 1$, only a fraction $(q/p)^{\ell-1}$ of the probability originally at ℓ is ultimately absorbed at one.

Since Σc_i is bounded below by the positive constant $(2-T)^{-(\ell-1)}$, the mean inter-arrival interval increases by at least

$$\mu^{-1}(2-T)^{-(\ell-1)}\ (1-T)(3-T)^{-1}$$

for each new distribution function of the sequence that we construct. The result follows.

A corollary is that even with a very large mean inter-arrival interval it is possible that there is only a low probability that an arrival finds the queue empty!

4. *Moving averages with the b's not all positive.*

In this section we consider the possibility of moving averages involving some negative b's. We find that this possibility entails restrictions both on the b's and on $U(\cdot)$.

It is a well known result [1] that a general stationary sequence $\{ v_n \}$ of random variables can be decomposed into a moving average form

$$v_n = \sum_{j=-\infty}^{\infty} b_j u_{n-j},$$

1. *E. J. Hannan: Time series analysis, Methuen monograph, Methuen and Co. Ltd., London (1960), Ch. 1, p.22.*

2. *W. Feller, An introduction to probability theory and its applications, Vol. 1, J. Wiley and Son, New York (1957), ch. 14.*

where the u_n are identically distributed. It is not, however, true that the u_n are necessarily mutually independent; nor that we can take a one-sided moving average

$$(4.1) \quad v_n = \sum_{j=o}^{\infty} b_j u_{n-j}.$$

Unfortunately, there is no theory dealing specifically with random variables which are constrained to be non-negative, and it is hard to assess the generality of the one-sided moving average for random variables of this type. We can gain some idea of the limitations as follows:

Let us write our moving average as

$$(4.2) \qquad v_n = \sum_j b_j u_{n-j} - \sum_k \beta_k u_{n-k},$$

where the v's, u's, b's and β's are all non-negative and the u's are I.I.D. We adopt the convention of writing $-\beta_k$ for any negative b_k. Since v_n is non-negative, we must have

$$(4.3) \qquad \sum_j b_j \inf(u_n) - \sum_k \beta_k \sup(u_n) \geq 0.$$

This relation is trivially satisfied if $\sum_k \beta_k$ is zero or u_n is identically zero. Unless one of these conditions is fulfilled, we must have that both

$$\inf(u_n) > 0$$

and

$$\sup(u_n) < \infty.$$

u_n is thus constrained to lie in a finite closed interval not containing the origin.

Therefore, unless $U(\cdot)$ is trivially zero, either all the coefficients b in (4.1) must be non-negative or v_n constrained to lie in a bounded interval.

We observe that when all the b's of (4.1) are non-negative, v_n and v_{n+1} will, in general, have positive correlation, although they may have zero correlation.

Consider the equilibrium input to the second stage of a stable series queueing system with recurrent first stage There will be a negative or zero correlation between successive inter-arrival intervals. Also, the service times of the first stage will, in general, take all values between zero and infinity, so that this will also be true of inter-arrival intervals of the second stage.

It follows that the inter-arrival intervals for the second stage of a series queueing system cannot, in general, be expressed as a moving average (4.1). In fact, an ordinary renewal process with zero correlation between successive lifetimes would be expected to offer a better approximation.

Although the reasoning in this thesis is formulated with non-negative b's in mind, the boundedness of U(·) when one or more b's are negative will ensure the convergence of ψ for negative arguments, and our working will still be valid in this extended case. Except where explicit comment is made to the contrary, as in Chapter Five, it is taken that the moving averages dealt with may be of either of the two forms.

5. Second order moving averages with b_0 positive, b_1 negative.

We now make an analysis similar to that of section two for a moving average (b_0 , b_1) with b_1 negative. In this case there will be a negative (or zero) correlation between the lengths of successive inter-arrival intervals.

We deal with distribution functions whose densities can be built up as a weighted sum of Dirac delta measures. These distributions are, of course, all step functions, but by having many component delta measures close together we can construct distributions which are reasonable approximations to a continuous distribution function. The advantage of using delta measures is that all the points of increase of the distribution functions can be confined to a finite domain.

We write U_d for the distribution whose density has its mass concentrated a distance d from the origin.

Theorem 1: *Suppose a queueing system with negative exponential services has an input which is a moving average* $(b_o, -\beta_1)$ *of a finite weighted sum* $U = \Sigma c_d U_d$ *with positive weights* c_d *with sum unity. If an equilibrium queue length distribution* $\{P_j\}$ *exists, then*

$$P_o < 1 - T,$$

where T *is the (unique) root inside the unit circle of*

$$T = \psi(1 - T).$$

As usual, the condition that the traffic intensity is less than unity:

$$\int_o^\infty u\,dU(u) > \mu^{-1}$$

suffices for the existence both of T and of $\{P_j\}$.

Since P_o, when it exists, is given by

$$P = 1 - T\,F(U),$$

where

$$F(U) \equiv \psi(-\beta_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o - \beta_1\},$$

we need only show that

$$F(U) > 1$$

to establish the theorem. To do that, we shall use the following lemma.

<u>*Lemma 1: Suppose the quantities* a_i, a_i', $i = 1,2,3$, *satisfy*</u>

$$(5.1) \qquad \begin{cases} a_1 > a_1' > o, \\ o < a_2 < a_2', \\ o < a_3 < a_3', \end{cases}$$

(5.2) $$a_1 a_2/a_3 \leq a_1' \; a_2'/a_3' \; .$$

Then for any constants c, c' in $(0,1)$ such that

$$c + c' = 1$$

we have

(5.3) $$a_1 a_2/a_3 < (ca_1 + c'a_1')(ca_2 + c'a_2')/(ca_3 + c'a_3').$$

Proof:

$$\frac{\partial}{\partial c}\{ [ca_1+(1-c)a_1'][ca_2+(1-c)a_2']/[ca_3+(1-c)a_3'] \}$$

$$= \{c(a_1-a_1')(a_2-a_2')[c(a_3-a_3') + 2a_3'] + a_2'a_3'(a_1 - a_1') + a_3'a_1'(a_2 - a_2') - a_1'a_2'(a_3 - a_3')\} \div [ca_3 + (1-c)a_3']^2 \; .$$

From (5.1), we see that both

$$c(a_1 - a_1')(a_2 - a_2') \; [c(a_3 - a_3') + 2a_3']$$

and

$$[\; ca_3 + (1-c)a_3' \;]^{-2}$$

are strictly monotone decreasing functions of c over $(0,1)$. Hence

$$\frac{\partial}{\partial c} \{[ca_1+(1-c)a_1'][ca_2+(1-c)a_2']/[ca_3+(1-c)a_3']\}$$

is strictly monotone decreasing for c in $(0,1)$, and

$$[ca_1 + (1-c)a_1'][ca_2 + (1-c)a_2']/[ca_3 + (1-c)a_3']$$

is concave downwards.

(5.2) establishes the lemma.

Lemma 2: Suppose that as an input to a queue with negative exponential service, the moving average $(b_o, -\beta_1)$ of a finite sum

(5.4) $$c_\ell U_\ell + (1-c_\ell) \; \Sigma \; c_d U_d$$

gives rise to an equilibrium queue length distribution, *and that* $\ell \leq \min(d)$.

If the moving average input corresponding to (5.4) is associated with a parameter T, *then*

$$(1 - T)b_o - \beta_1 > 0.$$

Also, the moving average input associated with $\Sigma c_d U_d$ *gives rise to an equilibrium queue length distribution.*

<u>*Proof:*</u> The second part is immediate, since the traffic intensity associated with $\Sigma c_d U_d$ is

$$\mu^{-1} (\Sigma c_d U_d)^{-1},$$

which is less than the corresponding intensity

$$\mu^{-1} [c_\ell \ell + (1-c_\ell) \Sigma c_d d]^{-1}$$

associated with

$$c_\ell U_\ell + (1-c_\ell) \Sigma c_d U_d .$$

The first part is also clear. Denote max(d) by p. Then by (1.2)

$$b_o \ell - \beta_1 p \geq 0$$

i.e.,

$$b_o \ell + (1-b_o)p \geq 0,$$

or

(5.5) $$b_o \leq p/(p-\ell).$$

The parameter T of the limiting distribution associated with (5.4) is given by

(5.6) $$T = c_\ell \exp[-\mu(1-T)\ell] + (1-c_\ell)\Sigma c_d \exp[-\mu(1-T)d],$$

so that

$$\exp[-\mu(1-T)p] < T < \exp[-\mu(1-T)\ell].$$

Thus

(5.7) $$T = \exp[-\mu(1-T) q]$$

for some q for which

$$\ell < q < p.$$

It must be the case that

$$\mu q \geq 1 ,$$

for if

$$\mu q < 1$$

we would have

$$\begin{aligned} \exp[-\mu(1-T)q] &> \exp(1-T) \\ &> [1+(1-T)]^{-1} \\ &> 1-(1-T) , \end{aligned}$$

i.e.,

$$T = \exp[-\mu(1-T)q] > T.$$

We observe that, because of (5.7),

$$\begin{aligned} \exp[-\mu(\ell/p)q] &\leq \exp[-\ell/p] \\ &< (1+\ell/p)^{-1} \\ &< 1-\ell/p, \end{aligned}$$

i.e.,

$$\exp\ [-\mu\{1-(1-\ell/p)\}q] < 1-\ell/p.$$

Since exp(·) is concave upwards, and (5.7) is satisfied by T and also when T is replaced by unity, we must have

$$T < 1-\ell/p \quad (<1).$$

Therefore, by (5.6),

$$b_o T < 1.$$

Since

$$b_o - \beta_1 = 1,$$

it follows that

$$(1-T)b_o - \beta_1 > 0,$$

thus establishing the first part of the lemma.

We now proceed on to the proof of the theorem.

Consider first a finite weighted sum $\sum_{d\varepsilon D} c_d U_d$. As the sum involves only a finite number of U's, min(d) exists. Choose some $\ell \leq \min(d)$.

Evaluating ψ for the distribution function $\Sigma c_d U_d$, we have

$$\psi(-\beta_1) = \Sigma c_d \exp(\mu\beta_1 d) ,$$

$$\psi[(1-T)b_o] = \Sigma c_d \exp[-\mu(1-T)b_o d],$$

$$\psi[(1-T)b_o-\beta_1] = \Sigma c_d \exp[-\mu\{(1-T)b_o-\beta_1\}d] .$$

$\mu\beta_1$ is positive, whilst $-\mu(1-T)b_o$, $-\mu\{(1-T)b_o-\beta_1\}$ are negative (the latter by the first part of lemma 2). Since exp (·) is concave upwards for real arguments:

$$(5.9)\quad \begin{cases} \exp(\mu\beta_1\ell) & > \Sigma c_d\exp(\mu\beta_1 d), \\ \exp[-\mu(1-T)b_o\ell] & < \Sigma c_d\exp[-\mu(1-T)b_o d], \\ \exp[-\mu\{(1-T)b_o-\beta_1\}\ell] & < \Sigma c_d\exp[-\mu\{(1-T)b_o-\beta_1\}d]. \end{cases}$$

Suppose that

$$(5.10)\quad [\Sigma c_d\exp(\mu\beta_1 d)][\Sigma c_d\exp[-\mu(1-T)b_o d]/[\Sigma c_d\exp[-\mu\{(1-T)b_o-\beta_1\}d] \equiv F(\sum_{d\,\varepsilon D} c_d U_d) \geq 1$$

for every set { c_d: all $d\,\varepsilon D$ } containing no zero element.

Then since

$$(5.11)\quad \exp(\mu\beta_1\ell)\exp[-\mu(1-T)b_o\ell]/\exp[-\mu\{(1-T)b_o-\beta_1\}\ell] = 1,$$

the three exponentials on the right and left hand sides of (5.9) suffice as the a_i', a_i respectively, of lemma 1.

The result of lemma 1 then extends (5.10) to hold for D augmented by ℓ.

The basis for this inductive procedure is supplied by the fact that (5.11) still holds if ℓ is replaced by any real number, in particular, by any one of the $d\varepsilon D$.

Our induction establishes that for any finite weighted sum $\Sigma c_d U_d$ (with positive weights)

$$F(\Sigma c_d U_d) > 1,$$

so that

$$P_o < 1 - T.$$

It is clear from continuity that much as in section two, we can extend our result to cover an infinite number of components U_d if we relax the result of the theorem to

$$P_o \leq 1 - T.$$

We shall not however, follow up this possibility.

Theorem 2: *If associated with the moving average input* $(b_o, -\beta_1)$ *of*

$$c_\ell U_\ell + (1-c_\ell) \; \Sigma c_d \, U_d \qquad (\ell \leq \min(d)),$$

we have an equilibrium distribution with parameter T, *then the moving average input* $(b_o, -\beta_1)$ *of*

$$\Sigma c_d \, U_d$$

also gives rise to an equilibrium distribution. If this has parameter T', *then*

$$T' < T.$$

<u>*Proof:*</u> The first part of the theorem is proved as the second part of lemma 2. That

$$T' < T$$

can be shown thus:

From (5.6)

$$T > \Sigma c_d \exp [- \mu(1-T)d],$$

and we know that T' is given by

$$T' = \Sigma c_d \exp [-\mu(1-T') \, d].$$

Since this latter equation is also satisfied when T' is replaced by unity, and exp (·) is concave upwards, the result follows.

CHAPTER FOUR.

The time dependent behaviour of the queue $G(p+1)/E_r/1$.

1. Introduction.

In his study of the single server queue with Erlang service time distribution and arbitrary (non-recurrent) input process, Finch (1963) derived determinantal expressions for the transient queue length distribution as found by the mth arrival and the probability that the first arrival begins a busy period of length m.

If, as before, $\{A_m\}$, $m \geq 0$, is the sequence of arrival instants and

$$\theta_{m,j} = A_{m+1} - A_{m-j}, \quad 0 \leq j \leq m,$$

Finch's expression for the transient queue length distribution from initial emptiness when services are negative exponential with parameter unity is

prob (queue length $\geq j+1$ at A_{m+1}),

$$Q_j^{m+1} = (-)^{m-j} \begin{vmatrix} \frac{-\theta_{m,j}}{1!}, & \frac{\theta^2_{m,j}}{2!}, & \ldots, & \frac{(-\theta_{m,j})^{m-j}}{(m-j)!} & , e^{-\theta_{m,j}} \\ 1 & , \frac{-\theta_{m,j+1}}{1!}, & \ldots, & \frac{(-\theta_{m,j+2})^{m-j-1}}{(m-j-1)!} & , e^{-\theta_{m,j+1}} \\ 0 & , 1 & , \ldots, & \frac{(-\theta_{m,j+2})^{m-j-2}}{(m-j-2)!} & , e^{-\theta_{m,j+2}} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & , 0 & , \ldots, & \frac{-\theta_{m,j-1}}{1!} & , e^{-\theta_{m,m-1}} \\ 0 & , 0 & , \ldots, & 1 & , e^{-\theta_{m,m}} \end{vmatrix}, 0 \leq j \leq m.$$

A similar expression can readily be derived by Finch's method for a queue where the negative exponential services are replaced by Erlang services.

The unconditional queue length distribution at arrival instants is given by the integration of Q_j^{m+1} under the joint distribution function of inter-arrival intervals, although this is not in general a simple procedure. Brockwell (1963) has carried out such an evaluation for GI/M/1, using an elegant inductive procedure depending on the independence of successive inter-arrival intervals. The expression he obtains for the unconditional value of Q_j^{m+1} is

$$(j+1)\sum_{n=j+1}^{m+1} \frac{1}{n} \sum_{\Sigma i\alpha_i = n-j-1} \binom{n}{\Sigma\alpha_i} \frac{(\Sigma\alpha_i)!}{\pi(\alpha_i!)} \psi_o^{n-\Sigma\alpha_i} \psi_1^{\alpha_1} \cdots \psi_{n-j-1}^{n-j-1} ,$$

where $\Sigma_{\Sigma i\alpha_i = n-j-1}$ is a summation over all j-tuples $(\alpha_1, \ldots, \alpha_j)$ of non-negative integers such that $\Sigma i\alpha_i = j$, and

$$\psi_k = \int_o^\infty \frac{x^k}{k!} e^{-x} \, dU(x),$$

$U(x)$ being the common inter-arrival interval distribution function.

For moving average inputs of order exceeding unity, it seems more convenient to adopt the approach of Chapter Two. In the next chapter we provide an alternative avenue for considering G(2)/M/1, which occurs here as a special case.

The Erlang distributions of orders exceeding one lack the memory-less property of the simple negative exponential distribution. We recover this property by imbedding $G(p+1)/E_r/1$ in a more complex system. Imagine the rth order Erlang services to be replaced by negative exponential services and the arriving units by batches each of r individuals. Since the rth order Erlang distribution is the rth iterated convolution of a negative exponential distribution, the Erlang servicing of a unit can be thought of as possessing r negative exponential phases. The successive phases in the servicing of a unit then correspond to the negative exponential servicings of of successive individuals in a batch. The queue length in the original system becomes the total number of batches (whether or not complete) present. The busy period distribution and virtual waiting time distribution will be the same in the two systems.

$GI/E_k/1$ or the more general GI/M/1 with batch arrivals has been considered by Takács (1961) (transient behaviour), Pollaczek (1957), Wishart (1956) and Foster (1961)(limiting queue size and waiting time distribution), and Conolly (1960) (the busy period).

Suppose that arrivals are in batches of size r and that the services are negative exponential.

For complete generality we assume that services begin at A_o with a queue length of i. We take the sequence $U^{(n+p-1)}$ to have a realization $u^{(n+p-1)}$, and denote by $P_{ij}(u^{(n+p-1)}$ the probability that the arrival at A_n finds j customers in the system.

With the minor modification of the i the notation is as in the earlier discussion of G(p+1)/M/1, and to obviate tedious repetition, we presume the corresponding preliminaries.

We shall make use of the following generalised form of Abel's theorem on the continuity of power series [1].

<u>*Theorem*</u>

If the sequence $\{a_n\}$ *has a finite limit* a *then*

$$\lim_{n\to\infty} (1-w) \sum_{\ell=o}^{n} a_\ell w^\ell$$

exists and has the value a.

<u>*2. The basic equations and the form of the solution.*</u>

The basic recurrence relations are

$$(2.1)\quad P_{ij}(u^{(n+p)}) = \begin{cases} \sum_{\ell=o}^{\infty} P_{i\ell}(u^{(n+p-1)}K_{\ell+r}(u_n,u_{n+1},..,u_{n+p}), & n \geq 0,\ j = 0, \\ \sum_{\ell=o}^{\infty} P_{i\ell}(u^{(n+p-1)})k_{\ell+r-j}(u_n,u_{n+1},..,u_{n+p}), & n \geq 0, 1\leq j\leq r-1, \\ \sum_{\ell=o}^{\infty} P_{i,j-r+\ell}(u^{(n+p-1)})k_\ell(u_n,u_{n+1},..,u_{n+p}), & n \geq 0,\ j \geq r. \end{cases}$$

1. *E.W.HOBSON: The theory of functions of a real variable, Camb.Univ.Press (1926) Vol.II, ch.3.*

The reasoning is much as in section three of Chapter Two. If the batch arriving at A_{m+1} finds $j \geq r$ individuals in the system, the batch arriving at A_n must have found at least $j - r$ customers already present. If there were $j - r+s$ $(s \geq o)$ customers present, there would need to be s departures in (A_n, A_{n+1}) for the queue length at $A_{n+1}-0$ to be j. A similar argument holds for $j < r$.

We form a generating function on j and integrate with respect to the u's, with the exception of $u_n,..,u_{n+p}$, on which we take Laplace transforms

$$(2.2)\quad P_i^*(s^{(p)};z;n+1) = \sum_{\ell=o}^{\infty} (1-z^{-\ell})c_{i\ell}^*(s^{(p)};n+1)$$
$$+ z^r P_i^*[(1-z^{-1})\mu b_1+s_{p-1},..,(1-z^{-1})\mu b_{p-1}+s_1,(1-z^{-1})\mu b_p;z;n]$$
$$\times \psi[(1-z^{-1})b_o+s_p/\mu], \quad z \in R, \quad Re.s_i \geq 0, \; n \geq 0,$$

where R is as before the domain consisting of the intersection of the interior and perimeter of the unit circle with the exterior and perimeter of the circle with centre $(\frac{1}{2}, 0)$ and radius $\frac{1}{2}$, with the origin deleted.

The dependence on n in this last relation is handled by taking generating functions on n, and we shall also take generating functions on i, i.e., we allow for a general probability distribution in the initial queue length.

We define

$$\pi(s^{(p)};z;w,y) = \sum_{i=o}^{\infty} \sum_{n=o}^{\infty} P_i^*(s^{(p)};z;n)w^n y^i,$$

and similarly

$$c_\ell(s^{(p)};w;y) = \sum_{i=o}^{\infty} \sum_{n=1}^{\infty} c_{i\ell}^*(s^{(p)};n)w^n y^i, \qquad \ell \geq 0,$$

for $Re.\, s_i \geq 0$, $z \in R$, $|w| < 1$, $|y| \leq 1$.

In terms of these functions, (2.2) becomes

$$\pi(s^{(p)};z,w,y) = (1-zy)^{-1} \prod_{\ell=1}^{p} \psi(s_\ell/\mu) + \sum_{\ell=o}^{\infty} c_\ell(s^{(p)};w;y)(1-z^{-\ell})$$
$$+ wz^r\psi[(1-z^{-1})b_o+s_p/\mu]\; \pi[(1-z^{-1})\mu b_1+s_{p-1},..,(1-z^{-1})\mu b_p\,;z,w,y],$$
$$z \in R,\; Re.s_i \geq 0, \quad |y| \leq 1, \quad |w| < 1.$$

The expression

$$(1-zy)^{-1}\prod_{\ell=1}^{p}\psi(s_\ell/\mu)$$

arises from the queue length at $A_o - 0$.

We employ the same recursive substitution procedure as was used for G(p+1)/M/1. This provides the equation

$$\begin{aligned}(2.3)\quad \pi(s^{(p)};z,w,y) &= (1-zy)^{-1}\prod_{\ell=1}^{p}\psi(s_\ell/\mu)+\sum_{\ell=o}^{\infty}c_\ell(s^{(p)},w,y)(1-z^{-\ell})\\
&+wz^r\psi[(1-z^{-1})b_o+s_p/\mu]\times\\
&(1-zy)^{-1}\psi[(1-z^{-1})b_p]\prod_{\ell=1}^{p-1}\psi[(1-z^{-1})b_\ell+s_{p-\ell}/\mu]\\
&+\sum_{\ell=o}^{\infty}c_\ell\{(1-z^{-1})\mu b_1+s_{p-1},\ldots,(1-z^{-1})\mu b_p;w,y\}\,(1-z^{-\ell})\\
&+wz^r\psi[(1-z^{-1})(b_o+b_1)+s_{p-1}/\mu]\times\\
&\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\\
&\times\;[(1-zy)^{-1}\prod_{\ell=1}^{p}\psi[(1-z^{-1})(b_\ell+..+b_p)]\\
&+\sum_{\ell=o}^{\infty}c_\ell\{(1-z^{-1})\mu(b_1+..+b_p),\ldots,(1-z^{-1})\mu b_p;w,y\}(1-z^{-\ell})]\times\\
&[1-wz^r\psi(1-z^{-1})]^{-1}]]..]\\
&\equiv D(s^{(p)};z,w,y),\ \text{say},\\
&z\in R,\ \text{Re. } s_i\ge 0,\quad |y|\le 1,\quad |w|<1.\end{aligned}$$

In Chapter Two we were able to use the relation

$$\sum_{i=o}^{\infty}c_i(u^{(n+p)})=P_o(u^{(n+p)})$$

to supply the boundedness condition

$$\sum_{i=o}^{\infty}|c_i^*(s^{(p)};n)|<1,\qquad n\ge 1,\text{Re.}s_i\ge 0,$$

which played an important rôle in our argument. Here we make use of the analogous relation

$$\sum_{\ell=o}^{\infty}|c_{i\ell}^*(s^{(p)};n)|<1,\quad n\ge 1,\quad \text{Re.}s_i\ge 0,$$

which gives

$$(2.4)\quad \sum_{\ell=o}^{\infty}|c(s^{(p)};w,y)|<|w|(1-|w|)^{-1}\,(1-|y|)^{-1},$$

$$\text{Re. } s_i\ge 0,\quad |y|\le 1,\quad |w|<1.$$

By its construction, $\pi(s^{(p)};z,w,y)$ is an analytic function of z for $|z| \leq 1$, Re. $s_i \geq 0$, $|y| \leq 1$, $|w| < 1$, so that

$$\pi(s^{(p)};z,w,y) = D(s^{(p)};z,w,y), \quad |z| \leq 1, \text{Re.} s_i \geq 0, \quad |y| \leq 1, |w| < 1,$$

by analytic continuation.

(2.4) enables a further analytic continuation of π defined by

$$\pi(s^{(p)};z,w,y) = D(s^{(p)};z,w,y)$$

for $|z| > 1$ (Re.$s_i \geq 0$, $|w| < 1$, $|y| \leq 1$). From (2.3) the only singularities of π so extended will be at $z = y^{-1}$ and at the zeros of $1 - wz^r\psi(1-z^{-1})$ outside the unit circle.

Takács (1961) shows that under the restriction

$$(2.5) \quad \mu \int_0^\infty udU(u) > r \text{ ,}$$

the equation

$$(2.6) \quad z^r = w\psi(1-z), \quad |w| < 1,$$

has exactly r roots $z = \gamma_i(w)$ $1 \leq i \leq r$, inside the unit circle.

(2.5) is, of course, simply the intuitive condition for a limiting distribution to exist. Provided this is satisfied, the singularities of π are just y^{-1}, $[\gamma_i(w)]^{-1}$ $(1 \leq i \leq r)$.

Also, by (2.3), (2.4),

$$(1-zy) \prod_{i=1}^{r} (z-[\gamma_i(w)]^{-1})\pi(s^{(p)};z,w,y) = O(|z^{rp+1}|)$$

for Re. $s_i \geq 0$, $|w| < 1$, $|y| \leq 1$, so that

$$(1-zy) \prod_{i=1}^{r} (z-[\gamma_i(w)]^{-1})\pi(s^{(p)};z,w,y)$$

is a polynomial in z of degree not greater than $rp+1$.

π can thus be expressed as

$$(2.7) \quad \pi(s^{(p)};z,w,y) = \sum_{\ell=0}^{r(p-1)} A_\ell(s^{(p)};w,y)z^\ell + \sum_{i=1}^{r} B_i(s^{(p)};w,y)[1-z\gamma_i(w)]^{-1}$$

$$+ B_0(s^{(p)};w,y)\,[1 - zy]^{-1},$$

provided that no two of the $\gamma_i(w)$ are coincident and that y is not permitted to take on any of the values $\gamma_i(w)$. This proviso is presumed throughout the next section.

L. Takács states in Takács (1961) and elsewhere that the $\gamma_i(w)$ must be distinct if (2.5) is satisfied. However, no proof seems to be given of this result. When dealing with queueing problems in which (2.6) occurs, some writers assume the roots $\gamma_i(w)$ are distinct, others (e.g. Wishart (1956)) give alternative procedures to deal with coincidences. The possibility of coincidences would seem to still be an open question.

3. *Determination of the functions* A_ℓ, B_i

$B_o(s^{(p)};w,y)$ can be found simply by multiplying throughout in (2.7) and (2.3) (with the extended domain of validity) by $1-zy$ and considering the limits as $z \to y^{-1}$. A simple comparison shows that

$$
\begin{aligned}
(3.1)\qquad B_o(s^{(p)};w,y) = {} & \prod_{\ell=1}^{p} \psi(s_p/\mu) \\
& +(wy^{-r})\psi[(1-y)b_o+s_p/\mu]\psi[(1-y)b_p]\prod_{\ell=1}^{p-1}\psi[(1-y)b_\ell+s_{p-\ell}/\mu] \\
& +(wy^{-r})^2\psi[(1-y)b_o+s_p/\mu]\psi[(1-y)(b_o+b_1)+s_{p-1}/\mu]\ \times \\
& \psi[(1-y)b_p]\psi[(1-y)(b_{p-1}+b_p)]\prod_{\ell=2}^{p-1}\psi[(1-y)(b_{\ell-1}+b_\ell)+s_{p-\ell}/\mu] \\
& + \cdots \\
& +(wy^{-r})^p\psi[(1-y)b_o+s_p/\mu]\ldots\psi[(1-y)(b_o+\ldots+b_{p-1})+s_1/\mu]\ \times \\
& \Big(\prod_{\ell=1}^{p}\psi[(1-y)(b_\ell+\ldots+b_p)]\Big)[1-wy^{-r}\psi(1-y)]^{-1},
\end{aligned}
$$

$$\text{Re. } s_i \geq 0,\quad |w| < 1,\quad |y| \leq 1.$$

Under our proviso $y \neq \gamma_i(w)$, for $1 \leq i \leq p$, no singularities of B_o can arise from the vanishing of $1 - wy^{-r}\psi(1-y)$.

The resolution of the other functions A_ℓ, B_i to reveal the nature of the dependence on $s^{(p)}$, w, and y is more complicated. We make use of (2.7) and the $s^{(p)}$ transforms of the original equations (2.1), in which we omit successively to take generating functions on z and n. The detail is heavy and we attempt only an outline.

To treat the s dependence, we define the functions $(s^{(p)};w,y)$ by

$$\pi(s^{(p)};z,w,y) = \sum_{\ell=0}^{\infty} \pi_\ell(s^{(p)};w,y)\, z^\ell , \qquad |z| \le 1.$$

In terms of the π_ℓ's (2.1) can be written

$$(3.2)\quad \pi_j(s^{(p)};w,y) = y^j \prod_{\ell=1}^{p} \psi(s_p/\mu) + \sum_{i=0}^{\infty} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [\sigma_{j+i-r}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi[(\sigma b_o+s_p)/\mu]\,]_{\sigma=\mu},$$

$$j \ge r, \quad \text{Re. } s_\ell \ge 0, \quad |w|, |y| < 1.$$

Substitution for the σ_j's from (2.7) then gives, for $j > rp$,

$$(3.3)\quad \sum_{\ell=1}^{r} B_\ell(s^{(p)};w,y)[\gamma_\ell(w)]^j + B_o(s^{(p)};w,y)\, y^j$$

$$= y^j \sum_{\ell=1}^{p}\psi(s_\ell/\mu) + \sum_{\ell=1}^{r}[\gamma_\ell(w)]^{j-r}B_\ell[\mu(1-\gamma_\ell(s))b_1+s_{p-1},..\mu(1-\gamma_\ell(w))b_p;w,y]$$

$$\times\ \psi[(1-\gamma_\ell(w))b_o + s_p/\mu]$$

$$+\ B_o[\mu(1-y)b_1+s_{p-1},..,\mu(1-y)b_p;w,y]y^{j-r}\psi[(1-\gamma_\ell(w))b_o+s_p/\mu].$$

This equation is true for all $j > rp$, so that provided the $\gamma_\ell(w)$ and y are distinct we have the relations

$$(3.4)\quad B_\ell(s^{(p)};w,y) = B_\ell[\,(1-\gamma_\ell(w))b_1+s_{p-1},..,\mu(1-\gamma_\ell(w))b_p;w,y] \times [\gamma_\ell(w)]^{-r}\ \psi[(1-\gamma_\ell(w))b_o+s_p/\mu], \quad \ell = 1,2,..,r,$$

$$(3.5)\quad B_o(s^{(p)};w,y) = \sum_{i=1}^{p} \psi(s_i/\mu)$$

$$+\ B_o[\mu(1-y)b_1+s_{p-1},..,\mu(1-y)b_p;w,y]y^{-r}\psi[(1-\gamma_\ell(w))b_o+s_p/\mu].$$

It follows by a recursive sequence of substitutions that

$$(3.6)\quad B_\ell(s^{(p)};w,y) = [\gamma_\ell(w)]^{-rp}\psi[(1-\gamma_\ell(w))b_o+s_p/\mu]\psi[(1-\gamma_\ell(w))(b_o+b_1)+s_{p-1}/\mu] \times$$

$$\psi[(1-\gamma_\ell(w))(b_o+..+b_{p-1})+s_1/\mu]B_\ell[\mu(1-\gamma_\ell(w))(b_1+..+b_p),..,\mu(1-\gamma_\ell(w))b_p;w,y],$$

$$1 \le \ell \le r.$$

The functions $B_\ell(s^{(p)};w,y)$ are thus determined in terms of known functions and the $B_\ell[\mu(1-\gamma_\ell(w)(b_1+..+b_p),..,\mu(1-\gamma_\ell(w))b_p;w,y]$, which are functions of only two variables. We write

$$B_\ell(w,y) \equiv B_\ell[\mu(1-\gamma_\ell(w))(b_1+..+b_p),\ldots,\mu(1-\gamma_\ell(w))b_p;w,y], \quad 1 \le \ell \le r.$$

Utilizing (3.4), (3.5), we find for $j \le rp$ that corresponding to (3.3)

$$(3.7)\quad \begin{cases} [A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]]_{\sigma=\mu} = 0, \\ [A_{r(p-1)-1}(\sigma b_1+s_{p-1},..,\sigma b_p;w;y)\psi[(\sigma b_o+s_p)/\mu] \\ \quad + (-\mu)\frac{\partial}{\partial\sigma}\{A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}]_{\sigma=\mu} = 0, \\ \cdots\cdots \\ \cdots\cdots \\ [A_{r(p-1)-(r-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu] \\ \quad +(-\mu)\frac{\partial}{\partial\sigma}(A_{r(p-1)-r+2}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\} \\ \quad + \cdots\cdots \\ \quad + \frac{(-\mu)^{r-1}}{(r-1)!}\frac{\partial^{r-1}}{\partial\sigma^{r-1}}\{A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}]_{\sigma=\mu} = 0, \end{cases}$$

$$(3.8)\quad A_{r(p-1)}(s^{(p)};w,y)$$
$$= [A_{r(p-1)-r}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]$$
$$+..+ \frac{(-\mu)^r}{r!}\frac{\partial^r}{\partial\sigma^r}\{A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}]_{\sigma=\mu},$$

$$\cdots\cdots$$

$$A_r(s^{(p)};w,y)$$
$$= [A_o(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]$$
$$+..+ \frac{(-\mu)^{r(p-1)}}{(r(p-1))!}\frac{\partial^{r(p-1)}}{\partial\sigma^{r(p-1)}}\{A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p;w,y) \times$$
$$\psi[(\sigma b_o+s_p)/\mu]\}_{\sigma=\mu},$$

$$A_{r-1}(s^{(p)};w,y) = [(-\mu)\frac{\partial}{\partial\sigma}(A_o(\sigma b_1+s_{p-1},..,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}$$
$$+..+ \frac{(-\mu)^{r(p-1)+1}}{(r(p-1)+1)!}\frac{\partial^{r(p-1)+1}}{\partial\sigma^{r(p-1)+1}}\{A_{r(p-1)}(\sigma b_1+s_{p-1},..,\sigma b_p:w,y) \times$$
$$\psi[(\sigma b_o+s_p)/\mu]\}]_{\sigma=\mu}$$

$$- \Big[\sum_{i=1}^{r} B_i(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)[\gamma_i(w)]^{-1}$$

$$+B_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)y^{-1}\}\psi[(\sigma b_o+s_p)/\mu]\Big]_{\sigma=\mu},$$

$$A_{r-2}(s^{(p)};w,y) = \Big[\frac{(-\mu)^2}{2!} \frac{\partial^2}{\partial\sigma^2} \{A_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}$$

$$+\ldots+ \frac{(-\mu)^{r(p-1)+2}}{(r(p-1)+2)!} \frac{\partial^{r(p-1)+2}}{\partial\sigma^{r(p-1)+2}} \{A_{r(p-1)}(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y) \times$$

$$\psi[(\sigma b_o+s_p)/\mu]\}\Big]_{\sigma=\mu}$$

$$- \Big[\sum_{i=1}^{r} B_i(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)[\gamma_i(w)]^{-2}\psi[(\sigma b_o+s_p)/\mu]$$

$$+ (-\mu)\frac{\partial}{\partial\sigma} \{\sum_{i=1}^{r} B_i(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)[\gamma_i(w)]^{-1}\psi[(\sigma b_o+s_p)/\mu]\}$$

$$+ B_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)y^{-2}\psi[(\sigma b_o+s_p)/\mu]$$

$$+ (-\mu)\frac{\partial}{\partial\sigma} \{\sum_{i=1}^{r} b_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)y^{-1}\psi[(\sigma b_o+s_p)/\mu]\}\Big]_{\sigma=\mu},$$

. . . .

$$A_1(s^{(p)};w,y) = \Big[\frac{(-\mu)^{r-1}}{(r-1)!} \frac{\partial^{r-1}}{\partial\sigma^{r-1}} \{A_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu)\}$$

$$+\ldots+\frac{(-\mu)^{rp-1}}{(rp-1)!} \frac{\partial^{rp-1}}{\partial\sigma^{rp-1}} \{A_{r(p-1)}(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)\psi[(\sigma b_o+s_p)/\mu]\}\Big]_{\sigma=\mu}$$

$$- \Big[\sum_{i=1}^{r} B_i(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)[\gamma_i(w)]^{-(r-1)}\psi[(\sigma b_o+s_p)/\mu]$$

$$+\ldots+ \frac{(-\mu)^{r-2}}{(r-2)!} \frac{\partial^{r-2}}{\partial\sigma^{r-2}} \sum_{i=1}^{r} B_i(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)[\gamma_i(w)]^{-1} \times$$

$$\psi[(\sigma b_o+s_p)/\mu]\}$$

$$+B_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)y^{-r}\psi[(\sigma b_o+s_p)/\mu]$$

$$+ \frac{(-\mu)^{r-2}}{(r-2)!} \frac{\partial^{r-2}}{\partial\sigma^{r-2}} \{B_o(\sigma b_1+s_{p-1},\ldots,\sigma b_p;w,y)y^{-1}\psi[(\sigma b_o+s_p)/\mu]\}\Big]_{\sigma=\mu}.$$

We observe that all of the above equations occur only for $p \geq 2$. For $p = 1$ we have $r(p-1) = 0$ and there is only one A, namely A_o. The equations here become

$$
(3.9)\quad \begin{cases}
[A_o(\sigma b_1;w,y)\psi[(\sigma b_o+s)/\mu]]_{\sigma=\mu} = 0, \\
(-\mu)\dfrac{\partial}{\partial\sigma}[A_o(\sigma b_1;w,y)\psi[(\mu b_o+s)/\mu]]_{\sigma=\mu} \\
\quad = [\sum_{i=1}^{r} B_i(\sigma b_1;w,y)[\gamma_i(w)]^{-1}+B_o(\sigma b_1;w,y)y^{-1}]_{\sigma=\mu}, \\
\quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
\dfrac{(-\mu)^{r-1}}{(r-1)!}\dfrac{\partial^{r-1}}{\partial\sigma^{r-1}}[A_o(\sigma b_1;w,y)\psi(\sigma b_o+s/\mu)]_{\sigma=\mu} \\
\quad = [\sum_{i=1}^{r} B_i(\sigma b_1;w,y)[\gamma_i(w)]^{-(r-1)}\psi[(\sigma b_o+s)/\mu] \\
\quad +..+\dfrac{(-\mu)^{r-2}}{(r-2)!}\dfrac{\partial^{r-2}}{\partial\sigma^{r-2}}\{\sum_{i=1}^{r} B_i(\sigma b_1;w,y)[\gamma_i(w)]^{-1}\psi[(\sigma b_o+s)/\mu]\} \\
\quad +\dfrac{(-\mu)^{r-2}}{(r-2)!}\dfrac{\partial^{r-2}}{\partial\sigma^{r-2}}\{B_o(\sigma b_1;w,y)y^{-1}\psi[(\sigma b_o+s)/\mu]\}]_{\sigma=\mu}.
\end{cases}
$$

The procedure for solving the equations is similar to that for the functions B_i in our treatment of $G(p+1)/M/1$. (3.8) can be treated recursively to provide an expression for $A_{r(p-1)}(s^{(p)};w,y)$ in terms of A's with lower subscripts, known functions, and functions of w and y. Making use of the expression so obtained, we can treat $A_{r(p-1)-1}(s^{(p)};w,y)$ similarly, and ultimately obtain each $A_i(s^{(p)};w,y)$, $r(p-1) \geq i \geq 1$, in terms of known functions, functions of A_o, and functions of w and y. The resolution with respect to $s^{(p)}$ is completed by using conservation of probability. Since the system must be in some state at A_n-0,

$$\pi(s^{(p)};1,w,y) = (1-w)^{-1}(1-y)^{-1}\prod_{\ell=1}^{p}\psi(s_\ell/\mu),$$

or, by (2.7),

$$(3.10)\quad \sum_{\ell=o}^{r(p-1)} A_\ell(s^{(p)};w,y) + \sum_{i=1}^{r} B_i(s^{(p)};w,y)[1-\gamma_i(w)]^{-1}+B_o(s^{(p)};w,y)(1-y)^{-1} = (1-w)^{-1}(1-y)^{-1}\prod_{\ell=1}^{p}\psi(s_\ell/\mu).$$

This relation enables us to make the s dependence carried in known functions of the $s^{(p)}$. The subsequent resolution with respect to w and y will require the identities (3.7) in the $s^{(p)}$ and additional equations derived from (2.1) by taking s transforms. There do not seem to be any simple forms of solution.

The determination of the unconditional values $P_{ij}^{(n+1)}$ of the $P_{ij}(u^{(n+p)})$ can then be effected from

$$\pi(0,0,\ldots,0;z,w,y) = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} \sum_{n=1}^{\infty} P_{ij}^{(n+1)} y^i z^j w^n, \quad |z| \le 1, \quad |y| \le 1, \quad |w|<1.$$

For $p = 1$ there is only one A and the complete solution can be determined without further recourse to (2.1).

We have already from (3.1), (3.5)

$$B_o(s;w,y) = \psi(s/\mu)+wy^{-r}\psi[(1-y)b_o+s/\mu]\psi[(1-y)b_1+s/\mu][1-wy^{-r}\psi(1-y)]^{-1}$$

$$B_\ell(s;w,y) = [\gamma_\ell(w)]^{-r}\psi[(1-\gamma_\ell(w))b_o+s/\mu]\, B_\ell[\mu(1-\gamma_\ell(w))b_1;w,y], \quad 1 \le \ell \le r.$$

(3.10) becomes

$$A_o(s;w,y) + \sum_{i=1}^{r} B_i(s;w,y)[1-\gamma_i(w)]^{-1} + B_o(s;w,y)(1-y^{-1}) = (1-w)^{-1}(1-y)^{-1} \psi(s/\mu) .$$

This last relation provides an expression for $A_o(s;w,y)$ in terms of the known functions and the r unknowns $B_\ell[\mu(1-\gamma_\ell(w))b_1;w,y]$, $1 \le \ell \le r$. These unknowns are readily obtained from the r equations (3.9).

4. *The limiting queue length distribution.*

It can be shown by the methods of Finch (1963) that for a prescribed initial queue length the intuitive condition (2.5) for ergodic behaviour suffices for a limiting queue length distribution as found by arriving batches to exist. The intuitive result that the limiting distribution is independent of the initial queue length will be shown to follow from our equations.

By the extended form of Abel's theorem on the continuity of power series, the joint generating function of the initial queue length and the limiting queue length distribution is given by

$$\lim_{w\to 1} (1-w)\pi(0,0,..,0;z,w,y).$$

From (3.1) we see that

$$\lim_{w\to 1} (1-w)\, B_o(0,..,0;w,y) = 0$$

so that provided the limits

$$(4.1)\quad \lim_{w\to 1} (1-w) \sum_{\ell=o}^{r(p-1)} A_\ell(0,0,..,0;w,y)\ ,\ \lim_{w\to 1} (1-w)\, B_\ell(0,..,0;w,y)\ ,\ 1 \leq \ell \leq r,$$

exist, we have from (2.7) that for any given initial queue length, the equilibrium distribution of queue length will be a sum of r delayed geometric distributions with parameters $\gamma_\ell(1)$ $1 \leq \ell \leq r$. By (2.6), these parameters $\gamma_\ell(1)$ are the zeros inside the unit circle of

$$T^r = \psi(1-T).$$

This result can, however, be established without any problem about the existence of limits. One need consider only equilibrium behaviour as in Chapter Two without concern over transients.

That the limiting behaviour is independent of the initial queue size distribution is easily established. For convenience we shall work from the general equations presuming the limits (4.1) exist, although a precisely similar argument holds good without such a presumption if we deal with the corresponding equilibrium equations. We have seen that

$$\lim_{w\to 1} (1-w)\, B_o(s^{(p)};\ w,y)\ (1-zy)^{-1} = 0.$$

Referring to (2.7), we thus wish to show that

$$\lim_{w\to 1} (1-w)\{\sum_{\ell=o}^{r(p-1)} A_\ell(s^{(p)};w,y)z + \sum_{\ell=1}^{r} B_\ell(s^{(p)};\ w,y)[1-z\gamma_\ell(w)]^{-1}\}$$

is independent of the initial queue size. Consider the equations from which this expression is to be determined. We observe that in (3.6), and in (3.7), (3.8), *et seq.*, y occurs only implicitly and in functions A, B to be determined, and there not in combination with the other independent variables.

For the determination of the limiting distribution, the form of (3.10) we would use is

$$\lim_{w\to 1} (1-w)\{ \sum_{\ell=o}^{r(p-1)} A_\ell(s^{(p)};w,y) + \sum_{\ell=1}^{r} B_\ell(s^{(p)};w,y)[1-\gamma_\ell(w)]^{-1}\}$$
$$= (1-y)^{-1} \prod_{\ell=1}^{p} \psi(s_\ell/\mu) .$$

Whilst y does occur explicitly on the right hand side, the coefficient of each power of y in the power series expansion

$$(1-y) \prod_{\ell=1}^{p} \Psi(s_\ell/\mu) = \sum_{j=o}^{\infty} y^j \prod_{\ell=1}^{p} \psi(s_\ell/\mu)$$

is the same.

For p>1, we shall need further relations derived from (2.1). (2.1) can be written (*c.f.* (3.2))

$$\pi_j(s^{(p)};w,y) - \sum_{\ell=o}^{\infty} \frac{(-\mu)^{\ell+r-j}}{(\ell+r-j)!} \frac{\partial^{\ell+r-j}}{\partial\sigma^{\ell+r-j}} [\pi_\ell(\sigma b_1+s_{p-1},..,\sigma b_p)\psi[(\sigma b_o+s_p)/\mu]]_{\sigma=\mu},$$
$$= y^j \prod_{\ell=1}^{p} \Psi(s_p/\mu), \qquad 1 \le j \le r-1,$$

$$\pi_j(s^{(p)};w,y) - \sum_{\ell=o}^{\infty} \frac{(-\mu)^{\ell}}{\ell!} \frac{\partial^{\ell}}{\partial\sigma^{\ell}} [\pi_{j+\ell-r}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi[(\sigma b_o+s_p)/\mu]]_{\sigma=\mu} ,$$
$$= y^j \prod_{\ell=1}^{p} \psi(s_p/\mu) , \qquad j \ge r.$$

On multiplying these equations by 1-w and letting $w\to 1$ one finds that the right hand side forms vanish, so that as before y occurs only implicitly in the unknown functions π_j, and there uncombined with the other independent variables.

Since the known functions in our equations are independent of the initial queue size, it follows that the solutions for the unknown functions characterizing the equilibrium distribution must also display this independence. This gives our result.

When $p = 1$ the equations for the equilibrium queue length distribution take on a simple form. Because of the similarity between the arguments in this chapter and those of Chapter Two, we can write down the equilibrium distribution equations directly from the corresponding time dependent equations by omitting B_o, w, y and making appropriate corresponding adjustments.

In the notation of Chapter Two, the Laplace-Stieltjes transform of the generating function of the limiting distribution $\{P_i\}$ is

$$P^*(s;z) = A_o(s) + \sum_{j=1}^{r} B_j(s)\,(1-T_j z)^{-1} ,$$

and the limiting distribution

$$(4.2) \qquad P_i = \delta_{i,o} A_o + \sum_{j=1}^{r} B_j T_j^{\,i} , \qquad i \geq o,$$

where

$$A_o = A_o(0),$$

$$B_j = B_j(0), \qquad 1 \leq j \leq r$$

The equations determining $A_o(s)$, $B_j(s)$, $1 \leq j \leq r$, become

$$B_j(s) = T_j^{-r} \; B_j[\mu(1-T_j)b_1]\psi[(1-T_j)b_o+s/\mu]$$

$$= g_j\psi[(1-T_j)b_o+s/\mu] , \qquad 1 \leq j \leq r,$$

$$\psi(s/\mu) = A_o(s) + \sum_{j=1}^{r} B_j(s)\,(1-T_j)^{-1} ,$$

$$A_o(\mu b_1) = 0,$$

$$\left[\sum_{j=1}^{r} B_j(\sigma b_1)T_j^{-1} - (-\mu)\frac{\partial}{\partial\sigma} A_o(\sigma b_1)\right]_{\sigma=\mu} = 0 ,$$

$$\left[\sum_{j=1}^{r} B_j(\sigma b_1)T_j^{-2} + (-\mu)\sum_{j=1}^{r}\frac{\partial}{\partial\sigma} B_j(\sigma b_1)T_j^{-1} - \frac{(-\mu)^2}{2!} A_o(\sigma b_1)\right]_{\sigma=\mu} = 0 ,$$

. . . .

$$\left[\sum_{j=1}^{r} B_j(\sigma b_1)T_j^{-(r-1)} + \ldots + \frac{(-\mu)^{r-2}}{(r-2)!}\sum_{j=1}^{r}\frac{\partial^{r-2}}{\partial\sigma^{r-2}} B_j(\sigma b_1)T_j^{-1} - \frac{(-\mu)^{r-1}}{(r-1)!}\frac{\partial^{r-1}}{\partial\sigma^{r-1}} A_o(\sigma b_1)\right]_{\sigma=\mu} = 0.$$

(4.2) thus becomes

$$P_i = \begin{cases} 1-\sum_{j=1}^{r} T_j(1-T_j)g_j\ \psi\,(b_o(1-T_j)) , & i = o, \\ \sum_{j=1}^{r} T_j^{\,i}\ g_j\ \psi(b_o(1-T_j)) , & i \geq 1, \end{cases}$$

where the g_j are easily found explicitly from

$$\sum_{j=1}^{r} g_j\ (1-T_j)^{-1}\ \psi[b_o(1-T_j) + b_1] = \psi(b_1) ,$$

$$\sum_{j=1}^{r} g_j\{T_j^{-1}\psi[b_o(1-T_j)+b_1] + (-b_1)(1-T_j)^{-1}\psi^{(1)}[b_o(1-T_j)+b_1]\} = -b_1\psi^{(1)}(b_1),$$

$$\sum_{j=1}^{r} g_j\{ T_j^{-2}\psi[b_o(1-T_j)+b_1]+(-b_1)T_j^{-1}\psi^{(1)}[b_o(1-T_j)+b_1]+\frac{(-b_1)^2}{2!}\psi^{(2)}[b_o(1-T_j)+b_1]\} = \frac{(-b_1)^2}{2!}\ \psi^{(2)}\ (b_1) ,$$

. . . .

$$\sum_{j=1}^{r} g_j\{ T_j^{-(r-1)}\psi[b_o(1-T_j)+b_1] +\ldots+\frac{(-b_1)^{r-1}}{(r-1)!}\ (1-T_j)^{-1}\psi^{(r-1)}[b_o(1-T_j)+b_i]\} = \frac{(-b_1)^{r-1}}{(r-1)!}\ \psi^{(r-1)}\ (b_1) .$$

5. *Erlang services with batch arrivals.*

We now revert to our original problem involving unit arrivals to a queueing system with an Erlang service time distribution of order r. From our correspondence between the queue $G(p+1)/E_r/1$ with unit arrivals and the queue $G(p+1)/M/1$ with batch arrivals of size r we see that queue lengths of o, $pr+q$ $(p \geq o,\ 1 < q \leq r)$ in the batch arrival system correspond to queue lengths of o, $p+1$ in the unit arrivals system.

Accordingly, (2.7) and the results of section four reveal that the equilibrium queue length distribution for the queueing system $G(p+1)/E_r/1$ $(p \geq 1)$ with unit arrivals is, apart from $P_o,..,P_{p-1}$, the sum of r geometric distributions with parameters T_j^r satisfying

$$T^r = \psi(1-T)$$

In the simplest case $p = o$, i.e., a general recurrent input, the equilibrium distribution is the sum of r geometric distributions. This result has been derived by several workers, e.g. Foster (1961), Takács (1961).

The device of studying $G(p+1)/E_r/1$ via the more complex system $G(p+1)/M/1$ with batch arrivals can be extended to consider $G(p+1)E_r/1$ with batch arrivals of size k, say. We now work in terms of $G(p+1)/M/1$ with batch arrivals of size rk. A negative exponential servicing of r arrivals, one at a time, then corresponds to an Erlang servicing of one of a batch of k customers. The limiting distribution of $G(p+1)/E_r/1$ with batch arrivals of size k will be a sum of r geometric distributions from $P_{k(p-1)+1}$ onwards $(p \geq 1)$, the common ratios being T^r where

$$T^{rk} = \psi\,(1-T).$$

For $p = o$, the limiting distribution is a sum of r geometric distributions. This result does not appear to have previously been stated explicitly.

6. *The variety of limiting distributions occurring in $G(p+1)/E_r/1$ systems.*

We are now able to extend our work of Chapter Three to include $G(p+1)/E_r/1$ systems. We shall not consider the limiting distributions which can arise in such systems in detail. Instead, we follow up a question suggested by the form of the limiting distributions we have obtained.

The question is whether or not any discrete distribution can be expressed as the sum of a number (possibly infinite) of delayed geometric distributions.

If the delay is finite, i.e. we restrict ourselves to finite moving average inputs, it turns out that the answer is no. It suffices to show that there are distributions which cannot be built up from purely geometric components, for a delayed geometric distribution is *a fortiori* geometric from some fixed point onward. We wish to know whether there are proper (non-negative) integral valued random variables $\{P_j\}$ not permitting of an expansion.

$$(6.1) \qquad P_j = \int_0^1 (1-p)p^j \, dF(p),$$

where $F(p)$ is a distribution function on $(0,1)$.

This problem is treated by Widder, (1946). There are, in fact, quite stringent restrictions imposed on $\{P_j\}$ if (6.1) is to be satisfied. Widder gives the necessary and sufficient condition that each mth difference has sign $(-)^m$.

A discussion of the solution for $F(p)$ when the expansion is possible is given by Daniels (1961).

* * * * * * * * * * * *

CHAPTER FIVE

The Time Dependent Behaviour of G(2)/M/1.

1. *Introduction*

In the previous chapter we saw how the time dependent behaviour of $G(p+1)/E_r/1$ at arrival instants could be determined. By the use of a further supplementary variable we could investigate the complete time dependent behaviour of this system; the treatment would, however, be cumbersome, and we shall not proceed in this direction.

Instead, we shall adopt a new approach which, although it is applicable only for second (and first) order moving average inputs, enables us to give with great facility a much more detailed account of the transient behaviour of the systems we are studying.

The method is an extension of the standard imbedded Markov chain technique that has proved fruitful for $E_r/G/1$ and $Gl/E_r/1$ systems. In distinction to those of the usual $E_r/G/1$ and $Gl/E_r/1$, however, the regenerative points we use do not coincide with natural discontinuity points of the system such as arrival or departure instants. They arise as follows:

Consider first the case where b_o, b_1, are both non-negative, and

$$b_o + b_1 = 1.$$

since the interval (A_m, A_{m+1}) between the arrivals at A_m, A_{m+1} has length

$$A_{m+1} - A_m = b_o U_{m+1} + b_1 U_{m+1}, \qquad m \geq 0,$$

the point R_m occurring $b_1 U_m$ after A_m can be intuitively regarded as the instant at which the effect of U_m ceases and that of U_{m+1} commences. We illustrate this in figure one.

We note that there is a one-one correspondence between arrival instants $\{A_{m+1}\}$ and inter-regenerative point intervals (R_m, R_{m+1}), A_{m+1} segmenting

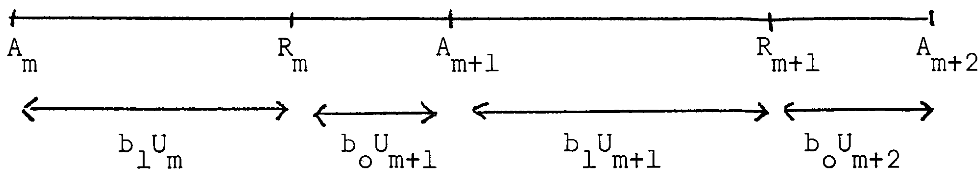

Figure One.

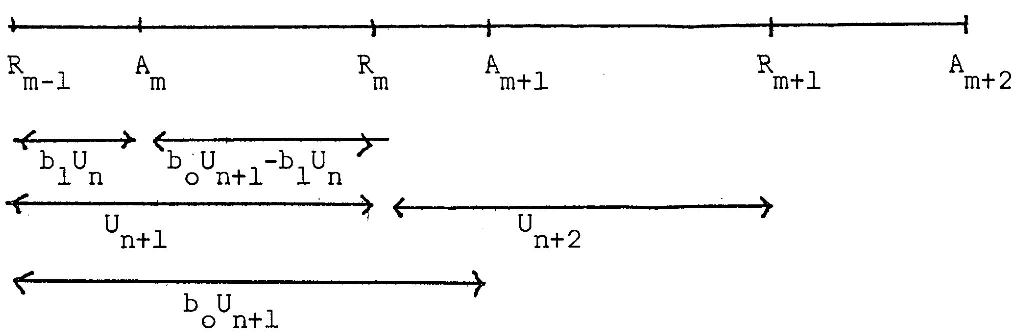

Figure Two.

(R_m, R_{m+1}) into portions of lengths $b_o U_m, b_1 U_{m+1}$. The points A_m, R_m occur alternately in time in an interlacing fashion.

Suppose now b_o is positive and b_1 negative, so that

$$A_{m+1} - A_m = b_o U_{m+1} - \beta_1 U_m, \qquad m \geq 0,$$

where b_o and β_1 are positive and

$$b_o - \beta_1 = 1.$$

As is shown in figure two, the chain $\{R_m\}$ is now more complex, asthere is some overlapping of the effects of U_{m+1}, U_{m+2} in the interval (R_m, A_{m+1}).

The analysis we carry out in this chapter for the simpler case could be performed with additional algebraic complexity for b_o positive, b_1 negative, but, in view of the restricted nature of moving averages with the latter property, we shall concern ourselves only with the more general problem.

A number of our results have been derived by Takács (1960) in the special case of a moving average of order one.

We define

(1.1) $$P_j = \sum_{\ell=o}^{j+1} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell (\mu b_1 u)^{j+1-\ell} [\ell!(j+1-\ell)!]^{-1} \, dU(u), \; j \geq -1,$$

as a basic set of quantities in terms of which to express unconditional queue length transition probabilities for the intervals between successive regenerative points.

Consider a typical such interval (R_m, R_{m+1}). If the queue length is $r \geq 0$ at R_m, then since the only arrival in (R_m, R_{m+1}) is that at A_{m+1}, the queue length at R_{m+1} must assume one of the values $0, 1, .., r+1$.

Suppose the queue lengths at the points R_m, R_{m+1} are r, s respectively. We first take the case $s \geq 2$. The realizations of the process giving rise to the prescribed pair r, s are those with ℓ departures from the queue in (R_m, A_{m+1}), $r \geq r-\ell \geq s-1$, an arrival at A_m, and a further $(r-s)+1-\ell$ departures in (A_{m+1}, R_{m+1}). Since the lengths of (R_m, A_{m+1}),

(A_{m+1}, R_{m+1}) are $b_o u_m$, $b_1 u_m$ respectively in any particular realization and the service is negative exponential, the unconditional probability of such an event for a given ℓ is

$$\int_o^\infty \exp[-\mu b_o u](\mu b_o u)^\ell (\ell!)^{-1} \exp[-\mu b_1 u](\mu b_1 u)^{r-s+1-\ell}[(r-s+1-\ell)!]^{-1} dU(u).$$

Summing over all permissible values of ℓ, we obtain that the unconditional probability of the transition $r \to s$ is P_{r-s}.

The corresponding results for transitions involving queue lengths of zero or unity can be most readily obtained through potential departures. If $r \geq 1$, the probability of fewer than r departures from the system in (R_m, A_{m+1}) is

$$\sum_{\ell=o}^{r-1} \exp[-\mu b_o u_m](\mu b_o u_m)^\ell/\ell!,$$

and so the probability of the system being found empty by the arrival at A_{m+1} is thus

$$1 - \sum_{\ell=o}^{r-1} \exp[-\mu b_o u_m](\mu b_o u_m)^\ell/\ell! = \sum_{\ell=r}^{\infty} \exp[-\mu b_o u_m](\mu b_o u_m)^\ell/\ell! .$$

By making use of virtual departures it is immediately found that the unconditional transition probabilities $0 \to 1$, $0 \to 0$ are $\int_o^\infty \exp[-\mu b_1 u]dU(u)$, $1 - \int_o^\infty \exp[-\mu b_1 u]dU(u)$, and that $r \geq 1$, $r \to 0$, $r \to 1$, are associated with the probabilities

$$1 - \sum_{\ell=-1}^{r-1} P_\ell - \sum_{\ell=r}^{\infty} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell/\ell! \quad dU(u)$$

$$= 1- \sum_{\ell=-1}^{r-1} P_\ell - \int_o^\infty \exp[-\mu b_1 u]dU(u) + \sum_{\ell=o}^{r-1} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell/\ell! dU(u),$$

and

$$P_{r-1} + \int_o^\infty \exp[-\mu b_1 u]dU(u) - \sum_{\ell=o}^{r-1} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell/\ell! \ dU(u).$$

2. *Transient behaviour of queue length on* $\{R_m\}$

If we define

$p_{ik}^{(m)}$ = probability of the transition $i \to k$ over an interval

(R_j, R_{j+m}),

$$p_{ik} = p_{ik}^{(1)} ,$$

then

$$p_{ik}^{(m+1)} = \sum_{j=o}^{\infty} p_{ij} p_{jk}^{(m)} ,$$

so that by the results of the last section

$$(2.1) \quad p_{ik}^{(m+1)} = P_{-1} p_{i+1,k}^{(m)} + P_o p_{i,k}^{(m)} + .. + P_{i-2} p_{2,k}^{(m)}$$
$$+ [P_{i-1} + \int_o^{\infty} \exp[-\mu b_1 u] dU(u) - \sum_{\ell=o}^{i-1} \int_o^{\infty} \exp[-\mu u](\mu b_o u)^{\ell}/\ell!\, dU(u)] p_{1,k}^{(m)}$$
$$+ [1 - \sum_{\ell=-1}^{i-1} P_{\ell} - \int_o^{\infty} \exp[-\mu b_1 u] dU(u) + \sum_{\ell=o}^{i-1} \int_o^{\infty} \exp[-\mu u](\mu b_o u)^{\ell}/\ell!\, dU(u)] p_{o,k}^{(m)},$$
$$i \geq 1,$$

$$(2.2) \quad p_{ok}^{(m+1)} = \int_o^{\infty} \exp[-\mu b_1 u] dU(u) \; p_{1k}^{(m)} + [1 - \int_o^{\infty} \exp[-\mu b_1 u] dU(u)] \; p_{ok}^{(m)} .$$

These relations can be expressed collectively by means of the generating function

$$P_{.k}^{(m)}(z) \equiv \sum_{i=o}^{\infty} p_{ik}^{(m)} z^i, \qquad |z| \leq 1,$$

as

$$zP_{.k}^{(m+1)}(z) = \sum_{i=o}^{\infty} P_{i-1} z^i P_{.k}^{(m)}(z) + [z - \sum_{i=o}^{\infty} P_{i-1} z^i] p_{ok}^{(m)} (1-z)^{-1}$$
$$- zP_{-1}(p_{1k}^{(m)} - p_{ok}^{(m)})$$
$$+ z(1-z)^{-1}(p_{1k}^{(m)} - p_{ok}^{(m)}) \int_o^{\infty} \exp[-\mu b_1 u] dU(u)$$
$$- z^2(1-z)(p_{1k}^{(m)} - p_{ok}^{(m)}) \int_o^{\infty} \exp[-\mu u(1-b_o z)] dU(u), \quad |z| < 1.$$

If we further define

$$P_{.k}(z,w) \quad \sum_{m=o}^{\infty} P_{.k}^{(m)}(z)w^{m}, \qquad |z| \le 1, \qquad |w| < 1,$$

then using

$$P_{.k}^{(o)}(z) = z^{k}$$

we find directly that

$$(2.3) \qquad P_{.k}(z,w) = [z^{k+1}+w\{(z-(\sum_{i=o}^{\infty} P_{i-1}z^{i})) \sum_{m=o}^{\infty} p_{ok}^{(m)} \; w^{m}(1-z)^{-1}$$
$$+ z\sum_{m=o}^{\infty} (p_{1k}^{(m)}-p_{ok}^{(m)})w^{m}[(1-z)^{-1}\int_{o}^{\infty}\exp[-\mu b_{1}u]dU(u)$$
$$- z(1-z)^{-1}\int_{o}^{\infty}\exp[-\mu u(1-b_{o}z)]dU(u) - p_{-1}]\}] \; [z-w\sum_{i=o}^{\infty} p_{i-1}z^{i}]^{-1},$$
$$|z| \;,|w| < 1.$$

From (1.1),

$$\sum_{i=o}^{\infty} p_{i-1}z^{i} = \int_{o}^{\infty}\exp[-\mu u(1-z)]dU(u), \qquad |z| \le 1,$$

and from (2.2)

$$(2.4) \qquad w\sum_{m=o}^{\infty} (p_{1k}^{(m)} - p_{ok}^{(m)})w^{m}\int_{o}^{\infty}\exp[-\mu b_{1}u]dU(u) = \sum_{m=o}^{\infty} p_{ok}^{(m)}w^{m}(1-w)-p_{ok}^{(o)}, \qquad |w| < 1.$$

Substituting these expressions in (2.3) gives

$$(2.5) \qquad P_{.k}(z,w) = [z^{k+1}+\sum_{m=o}^{\infty} p_{ok}^{(m)}(1-z)^{-1}(z-w\int_{o}^{\infty}\exp[-\mu u(1-z)]dU(u))$$
$$- wz^{2}(1-z)^{-1}\int_{o}^{\infty}\exp[-\mu u(1-b_{o}z)]dU(u)-z(1-z)^{-1}\delta_{o,k}$$
$$-zp_{-1}[\int_{o}^{\infty}\exp[-\mu b_{1}u]dU(u)]^{-1}[(1-w)\sum_{m=o}^{\infty} p_{ok}^{(m)}w^{m}-\delta_{o,k}]] \; \times$$
$$[z-w\int_{o}^{\infty}\exp[-\mu u(1-z)]dU(u)]^{-1}, \qquad |z|,\; |w| < 1.$$

For each w such that $|w| < 1$, $z-w\int_{o}^{\infty}\exp[-\mu u(1-z)]dU(u)$ has a unique zero $z = g(w)$ within the unit circle, by Rouché's theorem. Since $P_{.k}(z,w)$ is an analytic function of z for $|z| < 1$, $z = g(w)$ must also be a zero of the numerator of the right hand sice of (2.5). Therefore

$$(2.6) \qquad \sum_{m=o}^{\infty} p_{ok}^{(m)}w^{m} = (1-w)^{-1}[\delta_{o,k}+\int_{o}^{\infty}\exp[-\mu b_{1}u]dU(u)\{\int_{o}^{\infty}\exp[-\mu u]dU(u)\}^{-1} \times$$
$$[g(u)]^{k} - wg(w)\{1-g(w)\}^{-1}\int_{o}^{\infty}\exp[-\mu u(1-b_{o}g(w))]dU(u)$$
$$- (1-g(w))^{-1}\delta_{o,k}\}], \qquad |w| < 1.$$

Making use of (2.6), we can express (2.5) as

$$(2.7)\quad \sum_{m=o}^{\infty}\sum_{i=o}^{\infty} p_{ik}^{(m)} z^i w^m = (1-z)^{-1}(1-w)^{-1}[\delta_{o,k}+\psi(b_1)\{\psi(1)\}^{-1} \times$$

$$\{[g(w)]^k - wg(w)\{1-g(w)\}^{-1}\psi(1-b_o g(w)) - (1-z)^{-1}\delta_{o,k}\}]$$

$$+ [z-w\psi(1-z)]^{-1}[z^{k+1}-wz^2(1-z)^{-1}\psi(1-b_o z) - z(1-z)^{-1}\delta_{o,k}$$

$$- z\{[g(w)]^k-wg(w)\{1-g(w)\}^{-1}\psi(1-b_o g(w))-(1-g(w))^{-1}\delta_{o,k}\}],$$

$$|z|,\ |w| < 1.$$

When a power series expansion of $g(w)$ about the origin is derived for the determination of particular constants, we can make use of Lagrange's Theorem, which gives

$$g(w) = w\sum_{j=1}^{\infty}\frac{(-\mu w)^{j-1}}{j!}\left[\frac{d^{j-1}}{dx^{j-1}}\{\psi(y)\}^j\right]_{y=\mu}, \qquad |w| < 1.$$

(2.7) gives information about the queue length at regenerative points R_m, whereas one is more usually interested in the queue length as found by incoming customers. This is easy to deduce from the above results when we know the unconditional queue length transition probabilities between R_m and $A_{m+1}-o$ and the queue length distribution at R_o.

The unconditional queue length transition probabilities between R_m and $A_{m+1}-0$ are

$$P(r\to s)\begin{cases}\int_o^{\infty}\exp[-\mu b_o j](\mu b_o u)^{r-s}/(r-s)!\ dU(u), & 0 < s \le r,\\ \int_o^{\infty}[1-\exp[-\mu b_o]\sum_{\ell=o}^{r-1}(\mu b_o u)^{\ell}/\ell!]dU(u), & s = 0,\ r> 0,\\ 1, & s = 0 = r,\\ 0, & s > r.\end{cases}$$

If the queue size just after the commencement of the first service at A_o is $\ell \ge 1$, the queue size distribution $\{R_k^{(o)}\}$ at R_o is

$$R_k^{(o)} = \begin{cases} \int_o^\infty \exp[-\mu b_1 u](\mu b_1 u)^{\ell-k}/(\ell-k)!\,dU(u), & \ell \geq k > 0, \\ \int_o^\infty [1-\exp[-\mu b_1 u] \sum_{i=o}^{\ell-1} (\mu b_1 u)^i (i!)^{-1}]dU(u), & k = 0, \\ 0\,, & k > \ell. \end{cases}$$

3. *Transient behaviour of queue length at an arbitrary instant of time.*

A form of argument similar to that used in the last section gives the queue length probabilities at an arbitrary instant of time.

For convenience we take as our origin of time a regenerative point and label this R_o. We suppose that the initial queue length is i, and we write

$P_{ik}(t)$ = probability that queue size is k at time t, given that initial queue size is i, $t \geq 0$,

$\{j|i;u\}$= probability that the queue size changes from i to j in an interval of length u without arrivals.

We have that

$$\{j|i:u\} = \begin{cases} \exp[-\mu u](\mu u)^{i-j}/(i-j)!, & 0 < j \leq i, \\ 1-\exp[-\mu u] \sum_{\ell=o}^{i-1} (\mu u)^\ell/\ell!, & j = 0 < i, \\ 1\,, & j = 0 = i \\ 0\,, & j > i. \end{cases}$$

If R_1, the next regenerative point after R_o, occurs at a time $x \leq t$, then by the theorem of total probability

$$P_{ik}(t) = \sum_{r=o}^{i} \sum_{s=o}^{r+1} \{r|i;\ b_o x\}\{s|r+1;b_1 x\}\ P_{sk}(t-x).$$

If $t < x \leq t/b_o$, there is an arrival at $b_o x \leq t$ and so

$$P_{ik}(t) = \sum_{r=o}^{i} \{r|i;\, b_o x\}\, \{k|r+1;t-b_o x\}\ ,$$

and similarly for $x > t/b_o$ there is no arrival in $(0,t)$ and

$$P_{ik}(t) = \{k|i;t\}.$$

The unconditional value of $P_{ik}(t)$ is therefore

(3.1) $$P_{ik}(t) = \sum_{r=o}^{i} \sum_{s=o}^{r+1} \int_o^t \{r|i;b_o u\}\, \{s|r+1;b_1 u\}\, P_{sk}(t-u)dU(u)$$

$$+ \sum_{r=o}^{i} \int_t^{t/b_o} \{r|i;b_o u\}\, \{k|r+1;t-b_o u\}dU(u)$$

$$+ \{k|i;t\}\, (1-U(t/b_o)),\qquad t \geq 0.$$

Similarly to the last section we find it convenient to work in terms of families of basic quantities defined as follows:

$$P_j(u) = \sum_{\ell=o}^{j+1} \exp[-\mu u]\, (\mu b_o u)^{\ell}(\mu b_1 u)^{j+1-\ell}[\ell!(j+1-\ell)!]^{-1},\quad u \geq 0,\ j \geq -1,$$

$$q_j(u) = \exp\,[-\mu u](\mu b_o u)^j/j!\ ,\qquad\qquad u \geq 0,\ j \geq 0.$$

(3.1) can be rewritten as

(3.2) $$P_{ik}(t) = \begin{cases} \displaystyle\sum_{j=-1}^{i-1} \int_o^t p_j(u)\, P_{i-j,k}(t-u)\, dU(u) \\ \displaystyle + \int_o^t [\exp[-\mu b_1 u] - \sum_{\ell=o}^{i} q_i(u)]\, P_{1k}(t-u)\, dU(u), \\ \displaystyle + \int_o^t [1- \sum_{j=-1}^{i-1} p_j(u)-(\exp[-\mu b_1 u]- \sum_{\ell=o}^{i} q_i(u))]P_{ok}(t-u)dU(u) \\ + \begin{cases} \displaystyle\sum_{\ell=o}^{i-k+1} \int_t^{t/b_o} \exp[-\mu t](\mu b_o u)^{\ell}\{\mu(t-b_o u)\}^{i-\ell+1-k} \\ \qquad \{\ell!(1-\ell+1-k)!\}^{-1}dU(u), & i+1 \geq k > 0, \\ \displaystyle\int_t^{t/b_o} [1-\{1- \sum_{\ell=o}^{i} \exp[-\mu b_o u](\mu b_o u)^{\ell}(\ell!)^{-1}\}\exp[-\mu(t-b_o u)] \\ \displaystyle\qquad - \sum_{\ell=o}^{i} \exp[-\mu t](\mu t)^{\ell}/\ell!]dU(u), & k = 0, \\ 0, & i+1 < k, \end{cases} \\ + \begin{cases} \exp[-\mu t](\mu t)^{i-k}\{(i-k)!\}^{-1}[1-U(t/b_o)], & i \geq k, \\ 0, & i < k, \end{cases} \end{cases}$$

for $i \geq 1$,

and

$$P_{ok}(t) = \int_o^t \exp[-\mu b_1 u]\, P_{1k}(t-u)dU(u)$$

$$+ \int_o^t (1-\exp[-\mu b_1 u])P_{ok}(t-u)dU(u)$$

$$+\begin{cases} \int_t^{t/b_o} \exp[-\mu(t-b_o u)]dU(u), & k = 1, \\ \int_t^{t/b_o} \{1-\exp[-\mu(\ \ -b_o u)]\}dU(u), & k = 0, \\ 0, & k > 1, \end{cases}$$

$$+ \delta_{o,k}\,(1-U(t/b_o)).$$

These relations can be solved for the $P_{ik}(t)$, $i = 0,1,..$ through the Laplace transforms

$$P^*_{ik}(s) = \int_o^\infty \exp(-st)P_{ik}(t)dt, \quad Re.s \geq 0, \quad i = 0,1,..,$$

and the generating function

$$P^*_{.k}(s,z) = \sum_{i=o}^{\infty} P^*_{ik}(s)z^i, \qquad |z| \leq 1, \qquad Re.\ s \geq 0.$$

The solution for $P^*_{.k}(s,z)$ is, however, extremely complicated and we content ourselves with sketching the method.

Forming the generating function $P^*_{.k}(s,z)$ from the right hand side of (3.2) gives

$$(3.3) \qquad zP^*_{.k}(s,z) = P^*_{.k}(s,z)\int_o^\infty \sum_{\ell=o}^{\infty} z^\ell p_{-1}(t)\exp(-st)dU(t),$$

$$-z\, P^*_{1k}(s)\int_o^\infty p_{-1}(t)\exp(-st)dU(t)$$

$$+ z(1-z)^{-1}[P^*_{1k}(s)-P^*_{ok}(s)] \int_o^\infty \exp(-st)\exp(-\mu b_1 t)dU(t)$$

$$+ z(1-z)^{-1}P^*_{ok}(s) \int_o^\infty \exp(-st)dU(t)$$

$$- z(1-z)^{-1}[P^*_{1k}(s) - P^*_{ok}(s)]\int_o^\infty \sum_{\ell=o}^{\infty} q_\ell(t)\exp(-st)dU(t)$$

$$+ z[P^*_{1k}(s)-P^*_{ok}(s)]\int_o^\infty q_o(t)\exp(-st)dU(t)$$

$$- (1-z)^{-1}P^*_{ok}(s)\int_o^\infty \sum_{\ell=o}^{\infty} z^\ell P_{\ell-1}(t)dU(t)$$

$$+ P^*_{ok}(s)z\int_o^\infty P_{-1}(t)\exp(-st)dU(t)$$

$$+ \int_o^\infty \exp(-st)K(t,z)dt, \qquad |z| < 1, \quad Re.s \geq 0,$$

where $K(t,z)$ is an expression in terms only of known functions of t and z, namely,

$$K(t,z) = (1-\delta_{o,k})\, z^k \exp\{-\mu t(1-z)\}\,\{U(t/b_o)-U(t)\}$$
$$+\ \delta_{o,k}[z(1-z)^{-1}\{U(t/b_o)-U(t)\}$$
$$-\ z(1-z)^{-1} \int_t^{t/b_o} \exp\{-\mu(t-b_o u)\}dU(u)$$
$$+\ z(1-z)^{-1} \int_t^{t/b_o} \exp\{-\mu b_o u(1-z)\}\exp\{-\mu(t-b_o u)\}dU(u)$$
$$-\ z(1-z)^{-1} \int_t^{t/b_o} \exp\{-\mu t(1-z)\}dU(u)$$
$$-\ z \int_t^{t/b_o} \{\exp(-\mu b_o u)\ \exp(-\mu t)\}dU(u)]$$
$$+\ z^{k+1}\ \exp[-\mu t(1-z)]\ [1-U(t\ b_o)]$$
$$+\ z\ \delta_{o,k}(1-U(t/b_o))(1-\exp(-\mu t))$$
$$+\ z\ \delta_{1k} \int_t^{t/b_o} \exp\ \{-\mu(t-b_o u)\}dU(u), \qquad |z| < 1, \qquad t \geq 0.$$

(3.3) is simply solved for $P^*_{.k}(s,z)$ once $P^*_{ok}(s)$ and $P^*_{1k}(s)$ are known. One of these can be readily eliminated in terms of the other, since forming Laplace transforms in the second equation of (3.2) provides the relation

$$P^*_{ok}(s) = P^*_{1k}(s) \int_o^\infty \exp(-\mu b_1 t)\exp(-st)dU(t)$$
$$+\ P^*_{ok}(s) \int_o^\infty [1-\exp(-\mu b_1 t)]\ \exp(-st)dU(t)$$
$$+\ \delta_{1,k} \int_o^\infty \exp(-st)[\int_t^{t/b_o} \exp\{-\mu(t-b_o u)\}dU(u)]dt$$
$$-\ \delta_{o,k} \int_o^\infty \exp(-st)[\int_t^{t/b_o} \exp\{-\mu(t-b_o u)\}\ dU(u)]dt$$
$$+\ \delta_{ok} \int_o^\infty \exp(-st)[1-U(t)]dt,\ t \geq 0,\ \text{Re. } s \geq 0.$$

After such an elimination has been performed, (3.3) can be written as

(3.4) $$P^*_{.k}(s,z)[z-\int_o^\infty \sum_{\ell=o}^\infty P_{\ell-1}(t)\exp(-st)dU(t)] = D, \qquad |z| < 1, \quad \text{Re. } s \geq 0,$$

where D is a combination of known functions of s and z with one of

$P^*_{ok}(s)$, $P^*_{1k}(s)$. From the definition of the $p_\ell(t)$,

$$z - \int_0^\infty \sum_{\ell=0}^\infty z^\ell p_{\ell-1}(t)\exp(-st)dU(t)$$

$$= z - \int_0^\infty \exp[-\mu u(1-z)]\exp(-st)dU(t),$$

and by Rouché's theorem this expression has, for each s such that Re. $s \geq 0$, a unique zero $T(s)$ inside $|z| = 1$ if

$$\int_0^\infty t\,dU(t) > \frac{1}{\mu}$$

This last inequality is just the intuitive condition for a stable equilibrium queue length distribution to exist.

The vanishing of the left hand side of (3.4) for $z = T(s)$ implies that of the right hand side. This provides a relation whereby the remaining one of $P^*_{ok}(s)$, $P^*_{1k}(s)$ can be eliminated from (3.4), which is then directly solvable for $P^*_{.k}(s,z)$ in terms of known functions of s and z.

4. *The busy period.*

We wish to determine the probability that an arrival finding the queue empty starts a busy period of n services. The probability that an arrival at A_m initiates a busy period depends on the queue length distribution at R_{m-1} (the last point of the sequence $\{R_j\}$ before A_m) and in the length $b_o u_m$ of the interval (R_{m-1}, A_m) in which this queue is to dissipate. The probability that a busy period commencing at A_m persists to n services is also conditioned by $b_o u_m$, since the first service must certainly outlast the arrival-less period (A_m, R_m) of length $b_1 u_m$ if the busy period is to involve more than a single service.

The behaviour of a busy period is thus in two ways dependent on the value of m, where A_m is the instant of the arrival initiating the period.

This contrasts sharply with the usual situation with independently and identically distributed inter-arrival intervals, where the busy periods all have probabilistically equivalent structures.

Conditional on the initial interval (A_m, R_m) having a prescribed length $b_1 u_{m+1}$, however, a busy period has a determined probabilistic structure independent of m.

With probability

$$(4.1) \qquad 1- \exp(-\mu b_1 u_{m+1}) + \exp(-\mu b_1 u_{m+1}) \int_0^\infty [1- \exp(-\mu b_o u)] dU(u) = 1 - \exp(-\mu b_1 u_{m+1}) \int_0^\infty \exp(-\mu b_o u) dU(u)$$

the queue will be found empty by the next arrival (at A_{m+1}), i.e., the busy period will contain only one service.

The corresponding probability for a busy period of two services is

$$\begin{aligned}(4.2) \qquad & \exp(-\mu b_1 u_{m+1})[\int_0^\infty \exp[-\mu b_o u](1-\exp[-\mu b_1 u](\mu b_1 u))dU(u)] \\ & + \int_0^\infty \exp[-\mu b_o u] \exp[-\mu b_1 u](\mu b_1 u) dU(u) \int_0^\infty (1-\exp[-\mu b_o u])dU(u) \\ & + \int_0^\infty \exp[-\mu u] dU(u) \int_0^\infty (1-\exp[-\mu b_o u] \sum_{i=o}^{1} (\mu b_o u)^i) dU(u) \\ = & \exp(-\mu b_1 u_{m+1})[\psi(b_o)\{1- \int_0^\infty \exp[-\mu u]\mu b_1 u dU(u)\} \\ & + \psi(1) \int_0^\infty (1-\exp[-\mu b_o u] \sum_{i=o}^{1} (\mu b_o u)^i) dU(u).\end{aligned}$$

If the busy period involves more than two services, it will contain at least one interval (R_{m+i}, R_{m+1+i}), $i \geq 0$. Further analysis is conveniently done with working similar to that of section two in a system where we do not allow the queue to vanish at any instant.

We define

$P_{ik}^{(m)}$ = probability of the transition $i \to k$ over an interval (R_j, R_{j+m}), conditional on the queue never vanishing, $m \geq 1$,

$P_{ik} = P_{ik}^{(1)}$.

Analogously to section two, we have

$$(4.3)\quad P_{ik}^{(m+1)} = p_{-1}\, P_{i+1,k}^{(m)} + p_o\, P_{i,k}^{(m)} + \ldots + p_{i-1}\, P_{1,k}^{(m)}$$

$$- P_{ik}^{(m)} \int_o^{\infty} \exp[-\mu u](\mu b_o u)^i (i!)^{-1} dU(u), \quad k,i \geq 1,$$

where the p's are as defined by (1.1).

The subtracted term arises as follows:

If the queue length at some point R_t exceeds $j+1$, then p_j is the probability that at R_{t+1} the queue length is j fewer. p_j is a sum of the probabilities for the different ways the loss can be distributed between the subintervals (R_t, A_{t+1}), (A_{t+1}, R_{t+1}). One possibility is that $j + 1$ departures can be sustained by (R_t, A_{t+1}), and that there are no further departures after the arrival at A_{t+1}. If this possibility is removed, the modified p_j's are the appropriate transition probabilities between queue lengths $j+1$, 1 at successive R's when the queue is not permitted to become void.

Forming the generating function on i from (4.3):

$$z \sum_{i=1}^{\infty} P_{ik}^{(m+1)} z^i = \Big(\sum_{i=o}^{\infty} p_{i-1} z^i\Big) \sum_{i=1}^{\infty} P_{ik}^{(m)} z^i$$

$$- P_{ik}^{(m)} \int_o^{\infty} \exp[-\mu u(1-b_o z)] dU(u), \quad k \geq 1, \quad |z| \leq 1.$$

We now form the generating function on m

$$p_k(z,w) = \sum_{m=o}^{\infty} \Big(\sum_{i=1}^{\infty} P_{ik}^{(m)} z^i \Big) w^m, \qquad |z| \leq 1, \quad |w| < 1,$$

where we define

$$P_{ik}^{(o)} = \delta_{ik}\,.$$

We substitute

$$\sum_{i=o}^{\infty} p_{i-1} z^i = \int_o^{\infty} \exp[-\mu u(1-z)] dU(u), \qquad |z| \leq 1.$$

$$(4.4) \quad p_k(z,w) = [z^{k+1} - \sum_{m=o}^{\infty} w^m P_{1k}^{(m)} \int_o^{\infty} \exp[-\mu u(1-b_o z)]dU(u)] \times$$

$$[z - w\int_o^{\infty} \exp[-\mu u(1-z)]dU(u)]^{-1}, \quad k \geq 1, \quad |z|, \quad |w| < 1.$$

As noted in section two, for each w for which $|w| < 1$, $z - w \int_o^{\infty} \exp[-\mu u(1-z)]dU(u)$ has a unique zero $z = g(w)$ within the unit circle. $p_k(z,w)$ is an analytic function of z for $|z| < 1$, so that $z = g(w)$ must be a zero of the numerator of the expression for $p_k(z,w)$ in (4.4). Therefore

$$w \sum_{m=o}^{\infty} w^m P_{1k}^{(m)} = [g(w)]^{k+1}[\int_o^{\infty} \exp[-\mu u(1-b_o g(w))]dU(u)]^{-1}, \quad |w| < 1.$$

Thus (4.4) can be written

$$p_k(z,w) = [z^{k+1} - \psi(1-b_o z)[b(w)]^{k+1}[\psi(1-b_o g(w))]^{-1}][z-w\psi(1-z)]^{-1},$$

$$|z|, |w| < 1, \quad k \geq 1.$$

$P_k(z,w)$ is an analytic function in z,w for $|z|, |w| < 1$, so the probability of the transition $1 \to k$ over an interval (R_m, R_{m+j}), $j \geq 0$, can be determined as the coefficient of zw^j in $p_k(z,w)$ by, say, a repeated contour integration. Denote this probability by $q_k^{(j)}$.

We write $r_\ell^{(m)}$ for the probability that, at $R^{(m)}$, the queue length is ℓ, considered in section two.

We can now finally give an expression for the probability of a busy period of length n beginning at the arrival A_m.

If the queue is not empty at R_{m-1}, it has become so by A_m. It then remains non-empty until A_{m+n-1}, and becomes empty some time between A_{m+n-1} and A_{m+n}.

For $n = 1,2$, (4.1), (4.2) give

probability of a busy period of length one beginning at A_m

$$= r_o^{(m-1)}(1 - \psi(b_1)\psi(b_o))$$

$$+ \sum_{\ell=1}^{\infty} r_\ell^{(m-1)} \int_0^\infty \{1- \sum_{i=o}^{\ell-1} \exp[-\mu b_o u](\mu b_o u)^i (i!)^{-1}\} \times$$

$$\{1 - \exp[-\mu b_1 u]\psi(b_o)\}dU(u),$$

probability of a busy period of length two beginning at A_m

$$= [r_o^{(m-1)}\psi(b_1) + \sum_{i=1}^{\infty} r_i^{(m-1)} \int_0^\infty\{1 - \exp [- b_o u] \sum_{\ell=o}^{i-1} (\mu b_o u)^\ell (\ell!)^{-1}\} \times$$

$$\exp[-\mu b_1 u]dU(u)] \times$$

$$[\psi(b_o)\{1 - \int_0^\infty \exp[-\mu u]\mu b_1 u dU(u)\}$$
$$+ \psi(1) \int_0^\infty (1-\exp[-\mu b_o u] \sum_{i=o}^{1} (\mu b_o u)^i)dU(u)].$$

When a busy period of length $n > 2$ occurs the queue has size one between A_m and R_m and is non-empty throughout (R_m, R_{m+n-2}). The queue again becomes empty either during (A_{m+n-1}, R_{m+n-1}) or during (R_{m+n-1}, A_{m+n}). Compounding the probabilities of these events, we see that

probability of a busy period of length $n \geq 2$ beginning at A_m

$$= [r_o^{(m-1)}\psi(b_1) + \sum_{i=1}^{\infty} r_i^{(m-1)} \int_0^\infty\{1- \exp[-\mu b_o u] \sum_{\ell=o}^{i-1} (\mu b_o u)^\ell (\ell!)^{-1}\} \times$$

$$\exp[-\mu b_1 u]dU(u)] \times$$

$$[\sum_{k=1}^{n-1} q_k^{(n-2)} \int_0^\infty \exp[-\mu b_o u] \sum_{\ell=o}^{k-1} (\mu b_o u)^\ell \{1-\exp[-\mu b_1 u] \times$$

$$\sum_{j=o}^{k-\ell} (\mu b_1 u)^j (j!)^{-1}\}(\ell!)^{-1} dU(u)$$

$$+ \sum_{k=1}^{n} q_k^{(n-1)} \int_0^\infty \{ 1 - \exp[-\mu b_o u] \sum_{j=o}^{k-1} (\mu b_o u)^j (j!)^{-1}\}dU(u)].$$

We note that only more involved algebra is required to give a similar treatment to that of sections 1-4 for a queue in which the simple negative exponential services are replaced by general bulk negative exponential services, i.e., services in batches whose sizes follow a probability distribution

$$\text{prob (batch size} = r) = C_r, \quad r = 0,1,2,..,$$

where

$$\sum_{r=o}^{\infty} C_r = 1.$$

The generalisation to even batch arrivals of fixed size is, however, non-trivial.

The result for batch arrivals of size d corresponding to (2.1) is

$$p_{ik}^{(m+1)} = \begin{cases} p_{-d}p_{i+d,k}^{(m)} + p_{-(d-1)}p_{i+d-1,k}^{(m)} +..+ p_{i-(d+1)}p_{d+1,k}^{(m)} \\ \quad + \sum_{s=1}^{d} p_{s,k}^{(m)} [p_{i-s} - \sum_{\ell=i}^{i-s+d} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell (\mu b_1 u)^{i-\ell+d-s} \times \\ \quad\quad [\ell!(i-\ell+d-s)!]^{-1} dU(u) \\ \quad + \int_o^\infty \{1-\exp[-\mu b_o u] \sum_{\ell=o}^{i-1} (\mu b_o u)^\ell [\ell!]^{-1}\} \times \\ \quad\quad \exp[-\mu b_1 u](\mu b_1 u)^{d-s}/(d-s)! dU(u)] \\ \quad + p_{o,k}^{(m)} [1 - \sum_{\ell=-d}^{i-1} p_\ell - \int_o^\infty \{1-\exp[-\mu b_o u] \sum_{\ell=o}^{i-1} (\mu b_o u)^\ell [\ell!]^{-1}\} \times \\ \quad\quad \exp[-\mu b_1 u] \sum_{s=o}^{d-1} (\mu b_1 u)^s/s! \, dU(u) \\ \quad + \sum_{\ell=i}^{i+d-1} \int_o^\infty \exp[-\mu u](\mu b_o u)^\ell [\ell!]^{-1} \sum_{i=1}^{i+d-\ell} (\mu b_1 u)^{i+d-\ell-s} \\ \quad\quad [(i+d-\ell-s)!]^{-1} \, dU(u)], \qquad i > 0, \\ \sum_{s=1}^{d} p_{s,k}^{(m)} [\int_o^\infty \exp[-\mu b_1 u](\mu b_1 u)^{d-s}/(d-s)! \, dU(u)] \\ \quad + p_{o,k}^{(m)} [1 - \int_o^\infty \exp[-\mu b_1 u] \sum_{\ell=o}^{d-1} (\mu b_1 u)^\ell/\ell! \, dU(u)], \\ \qquad\qquad i = o. \end{cases}$$

For $d > 1$ we lack further simple results like (2.4) which we need to eliminate the unknowns $p_{s,k}^{(m)}$ $0 \leq s \leq d$, from the generating function formed on i.

We remark that when the inter-arrival intervals are independently and identically distributed, fluctuation theory makes available a full treatment with both general batch arrivals and general bulk services (Narayan Bhat (1964).

5. *Equilibrium behaviour of $G(2)/E_k/1$.*

As in chapter four we deal with the richer system G(2)/M/1 with batch arrivals of size k. Use of the regenerative points R considered in our discussion of G(2)/M/1 enables us to avoid the Laplace transforms that made our treatment of the general moving average cumbersome, and we obtain a fairly full picture of the equilibrium behaviour of our system.

We define $p_j^{(n)}(s)$ to be the limiting probability density that on the R_n the system contains j individuals when a time s has elapsed since the last regenerative point. As before we denote our moving average by (b_o, b_1).

It is convenient to determine first the distribution $\{p_j^{(n)}(0)\}$ of queue size at the regenerative points. We have for $j \geq k + 1$

$$(5.1) \quad p_j^{(n+1)}(0) = \sum_{r=o}^{\infty} p_{j-k+r}^{(n)}(0) \sum_{\ell=o}^{r} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^{\ell}}{\ell!} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{r-\ell}}{(r-\ell)!} dU(u) \,.$$

This follows from a consideration of the possible departures from the queue between a consecutive pair of regenerative points. Since there are k arrivals between these points, the queue size at the earlier must be at least $j-k$. There will, in general, be departures from the queue both before and after the arrival of the batch. If there are r departures altogether, we sum the probabilities corresponding to $\ell \le r$ of these occurring before the arrival instant and the remaining $r - \ell$ being between the arrival instant and the second regenerative point.

The relations corresponding for $0 \le j < k + 1$ are

$$(5.2 \qquad p_j^{(n+1)}(0) = \sum_{r=1}^{\infty} p_r^{(n)}(0) \sum_{\ell=o}^{r-1} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^\ell}{\ell!} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{r+k-\ell-j}}{(r+k-\ell-j)!} dU(u)$$

$$+ \sum_{r=o}^{\infty} p_r^{(n)}(0) \sum_{\ell=r}^{\infty} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^\ell}{\ell!} \times \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{k-j}}{(k-j)!} dU(u), \quad j > 0,$$

$$\sum_{r=1}^{\infty} p_r^{(n)}(0) \sum_{\ell=o}^{r-1} \exp(-\mu b_o u) \frac{(\mu b_o u)^\ell}{\ell!} \times \exp(-\mu b_1 u) \sum_{s=r+k-\ell}^{\infty} \frac{(\mu b_1 u)^s}{s!} dU(u)$$

$$+ \sum_{r=o}^{\infty} p_r^{(n)}(0) \sum_{\ell=r}^{\infty} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^\ell}{\ell!} \times (1 - \sum_{i=o}^{k-1} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^i}{i!}) dU(u), \quad j = 0.$$

The matrix P of transition probabilities will clearly be irreducible, since every state can be reached from every other, and aperiodic, since the diagonal elements are positive. Under these conditions it has been shown by Feller (1950), that a proper limiting distribution of queue length exists if we can construct a non-zero row vector $\underset{\sim}{x}$ for which

$$\underset{\sim}{x} P = \underset{\sim}{x}.$$

When normalised, $\underset{\sim}{x}$ will have as its components the limiting queue length probabilities.

We wish to find an $\underset{\sim}{x} = \{x_j\}$ for which

(5.3) $$x_j = \sum_{r=o}^{\infty} x_{j-k+r} \sum_{\ell=o}^{r} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^{\ell}}{\ell!} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{r-\ell}}{(r-\ell)!} \, dU(u),$$

$$j \geq k + 1,$$

(5.4) $$x_j = \sum_{r=1}^{\infty} x_r \sum_{\ell=o}^{r-1} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^{\ell}}{\ell!} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{r+k-\ell-j}}{(r+k-\ell-j)!} \, dU(u)$$

$$+ \sum_{r=o}^{\infty} x_r \sum_{\ell=r}^{\infty} \int_o^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^{\ell}}{\ell!} \exp(-\mu b_1 u) \frac{(\mu b_1 u)^{k-j}}{(k-j)!} \, dU(u),$$

$$0 < j < k + 1,$$

$$\sum_{r=1}^{\infty} x_r \sum_{\ell=o}^{r-1} \int_o^{\infty} \exp(-\mu b_1 u) \frac{(\mu b_o u)^{\ell}}{\ell!} \exp(-\mu b_1 u) \sum_{s=r+k-\ell}^{\infty} \frac{(\mu b_1 u)^{s}}{s!} \, dU(u)$$

$$+ \sum_{r=o}^{\infty} x_r \sum_{\ell=r}^{\infty} \exp(-\mu b_o u) \frac{(\mu b_o u)^{\ell}}{\ell!} (1 - \exp(-\mu b_1 u) \sum_{i=o}^{k-1} \frac{(\mu b_1 u)^{i}}{i!}) dU(u), j=0.$$

By direct substitution we can verify that (5.3) is satisfied by any linear combination

$$x_j = \sum_{i=1}^{k} \gamma_i T_k^j, \quad j \geq 1,$$

where T_i, $i = 1,..,k$, are the (presumed distinct) roots of

$$T^k = \int_o^{\infty} \exp(-\mu u(1-T)) \, dU(u)$$

inside the unit circle.

If we now put

$$x_o = \sum_{i=1}^{k} \gamma_i + A_o, \text{ say},$$

(5.4) provides k simultaneous inhomogeneous linear equations $(1 \leq j \leq k)$ for the unknowns γ_i/A_o. The equation for $j = 0$ will be consistent with the solution obtained since the elements of each row of the matrix P sum to unity.

The distribution

$$\{ p_j(0) \} = \lim_{n\to\infty} \{p_j^{(n)}(0)\}$$

thus exists, and, since

$$\sum_{j=o}^{\infty} p_j(0) = 1,$$

is of the form

$$\{p_j(0)\} = \{ \sum_{i=1}^{k} \beta_i T_i^j + \delta_{oj}[1-\sum_{i=1}^{k} \beta_i(1-T_i)^{-1}]\}, \qquad j \geq 0.$$

The queue length distribution as found by an arriving batch readily follows. Suppose that a batch arrives at time $b_o u$ after the last recurrence point R. Then the distribution $\{p_j(b_o u)\}$ at this arrival instant I is

$$p_j(b_o u) = \begin{cases} \sum_{r=o}^{\infty} p_r(0) \sum_{i=r}^{\infty} \exp(-\mu b_o u) \dfrac{(\mu b_o u)^i}{i!} , & j = 0 \\ \sum_{r=j}^{\infty} p_r(0) \exp(-\mu b_o u) \dfrac{(\mu b_o u)^{r-j}}{(r-j)!} , & j > 0 \end{cases},$$

so that using the known form of $\{p_r(0)\}$ we find that

$$p_j(b_o u) = \sum_{i=1}^{k} \beta_i T_i^j \exp(-\mu(1-T_i)b_o u), \quad j > 0.$$

Since

$$\sum_{j=o}^{\infty} p_j(b_o u) = 1 = \sum_{j=o}^{\infty} p_j(0),$$

we have also that

$$p_o(b_o u) = 1 - \sum_{i=1}^{k} \beta_i T_i (1-T_i)^{-1} \exp(-\mu(1-T_i)b_o u)$$

$$= \sum_{i=1}^{k} \beta_i \exp(-\mu(1-T_i)b_o u)+1-\sum_{i=1}^{k} \beta_i(1-T_i)^{-1} \exp(-\mu(1-T_i)b_o u).$$

The unconditional queue size distribution $\{p_j\}$ as found by an arbitrary arriving batch is thus

$$(5.5)\quad p_j = \sum_{i=1}^{k} \beta_i T_i^j \; \psi(b_o(1-T_i)) + \delta_{oj}\Big[1-\sum_{i=1}^{k} \beta_i (1-T_i)^{-1} \; \psi(b_o(1-T_i))\Big].$$

Consider an instant I occurring at time t after the last arrival instant A. Suppose A lies in an inter-regenerative point interval of length u. If $t < b_1 u$, I also lies in this interval. If $t > b_1 u$, a regenerative point R' divides internally the interval AI. We denote the next arrival instant by A'.

Since the queue size increases by k at A_1 the queue length distribution $\{q_j(t)\}$ at I is

$$(5.6)\quad q_j(t) = \begin{cases} \sum_{r=o}^{\infty} p_r(b_o u) \exp(-\mu t) \sum_{\ell=r+k}^{\infty} \frac{(\mu t)^{\ell}}{\ell!}, & j = 0, \\ \sum_{r=o}^{\infty} p_r(b_o u) \exp(-\mu t) \frac{(\mu t)^{r+k-j}}{(r+k-j)!}, & 0 < j \le k, \\ \sum_{r=j-k}^{\infty} p_r(b_o u) \exp(-\mu t) \frac{(\mu t)^{r+k-j}}{(r+k-j)!}, & j > k. \end{cases}$$

The limiting queue size distribution $\{q_j\}$ taken at an arbitrary time can be expressed as

$$q_j = \int_o^{\infty} dU(u) \int_o^{\infty} dU(v) \int_o^{b_1 u+b_o v} q_j(t)dt \Big/ \int_o^{\infty} dU(u) \int_o^{\infty} dU(v) \int_o^{b_1 u+b_o v} dt$$

$$= m^{-1} \int_o^{\infty} dU(u) \int_o^{\infty} dU(v) \int_o^{b_1 u+b_o v} q_j(t)dt, \quad j \ge 0,$$

where m is the mean of U.

Theorem: The limiting queue size distribution $\{q_j\}$ *is given by*

$$q_j = \begin{cases} (\mu m)^{-1}(p_{j-k} + \dots + p_{j-1}), & j > k \\ (\mu m)^{-1}(p_o + \dots + p_{j-1}), & 0 < j \le k \\ 1 - k/\mu m, & j = 0 \end{cases}.$$

Proof: We have determined the mean of $q_j(t)$ by first letting t vary over every moment in (AR') and (R'A') and then integrating over u and v (the length of the subsequent inter-regenerative point interval). Substituting for $q_j(t)$ gives

$$q_j = m^{-1} \sum_{i=1}^{k} \beta_i T_i^{j-k} [\mu(1-T_i)]^{-1} \psi((1-T_i)b_o)[1-\psi(1-T_i)]$$

$$= (\mu m)^{-1} \sum_{i=1}^{k} \beta_i T_i^{j-k} \psi(b_o(1-T_i)) (1+T_i+..+T_i^{k-1})$$

$$= (\mu m)^{-1} (p_{j-k}+..+p_{j-1}), \quad j > k.$$

Similarly, for $j \leq k$,

$$q_j = m^{-1}\int_o^\infty dU(u) \int_o^\infty \sum_{i=1}^{k} \beta_i T_i^{-(k-j)}[\mu(1-T_i)]^{-1}\exp[-\mu 1-T_i)b_o u] \times$$

$$[1-\exp(-\mu(1-T_i)(b_1u+b_ov)]$$

$$- \sum_{i=1}^{k} \beta_i T_i^{-(k-j)} \mu^{-1} \exp[-\mu(1-T_i)b_ou] \times$$

$$[(1+T_i+..+T_i^{k-j-1})-\exp(-\mu(b_1u+b_ov)) \times$$

$$\{(1+T_i+..+T_i^{k-j-1})$$

$$+ (\mu(b_1u+b_ov)) (T_i+..+T_i^{k-j-1})$$

$$+ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$+ \frac{(\mu(b_1u+b_ov))^{k-j-1}}{(k-j-1)!} T_i^{k-j-1} \}]$$

$$+ [1- \sum_{i=1}^{k} \beta_i(1-T_i)^{-1} \exp[-\mu(1-T_i)b_ou)]^{-1} \times$$

$$[- \exp(-\mu(b_1u+b_ov)) \sum_{\ell=o}^{k-j} \frac{(\mu(b_1u+b_ov))^\ell}{\ell!}] dU(v)$$

$$= m^{-1} \int_o^\infty dU(u)\int_o^\infty \sum_{i=1}^{k} \beta_i T_i^{-(k-j)}[\mu(1-T_i)]^{-1} \exp[-\mu(1-T_i)b_ou] \times$$

$$[1-\exp(-\mu(1-T_i)(b_1u+b_ov)]$$

$$-\sum_{i=1}^{k}\beta_iT_i^{-(k-j)}[\mu(1-T_i)]^{-1}\exp[-\mu(1-T_i)b_ou]\times$$

$$[\{1-\exp(-\mu(b_1u+b_ov))\sum_{\ell=o}^{k-j-1}\frac{(\mu T_i(b_1u+b_ov))^{\ell}}{\ell!}\}$$

$$-T^{k-j}\{1-\exp(-\mu(b_1u+b_ov))\sum_{\ell=o}^{k-j-1}\frac{(\mu(b_1u+b_ov))^{\ell}}{\ell!}\}]$$

$$+[1-\sum_{i=1}^{k}\beta_i(1-T_i)^{-1}\exp(-\mu(1-T_i)b_ou)]\mu^{-1}\times$$

$$[1-\exp(-\mu(b_1u+b_ov\sum_{\ell=o}^{k-j}\frac{(\mu(b_1u+b_ov))^{\ell}}{\ell!}]\quad dU(v)$$

$$=m^{-1}\int_o^{\infty}dU(u)\int_o^{\infty}\sum_{i=1}^{k}\beta_iT_i^{-(k-j)}[\mu(1-T_i)]^{-1}\exp[-\mu(1-T_i)b_ou]\times$$

$$[1-\exp(-\mu(1-T_i)(b_1u+b_ov)]$$

$$-\sum_{i=1}^{k}\beta_iT_i^{-(k-j)}[\mu(1-T_i)]^{-1}\exp[-\mu(1-T_i)b_ou]\times$$

$$\{1-\exp(-\mu(b_1u+b_ov))\sum_{\ell=o}^{k-j}\frac{(\mu T_i(b_1u+b_ov))^{\ell}}{\ell!}\}$$

$$+\mu^{-1}[1-\exp(-\mu(b_1u+b_ov))\sum_{\ell=o}^{k-j}\frac{(\mu(b_1u+b_ov))^{\ell}}{\ell!}]\quad dU(v)$$

$$=(\mu m)^{-1}\int_o^{\infty}dU(u)\int_o^{\infty}\exp(-\mu(b_1u+b_ov))\sum_{\ell=k-j+1}^{\infty}\frac{(\mu(b_1u+b_ov))^{\ell}}{\ell!}$$

$$-\sum_{i=1}^{k}\sum_{r=1}^{\infty}\beta_iT_i^{r}(1-T_i)^{-1}\exp(-\mu b_o(1-T_i)u)\exp(-\mu(b_1u+b_ov))$$

$$\frac{(\mu(b_1u+b_ov))^{k-j+r}}{r!}\quad dU(v)$$

$$=(\mu m)^{-1}\int_o^{\infty}dU(u)\int_o^{\infty}(1-\sum_{i=1}^{k}\beta_iT_i(1-T_i)^{-1}\exp(-\mu(1-T_i)b_ou))\exp(-\mu(b_1u+b_ov))$$

$$\sum_{\ell=k-j+1}^{\infty} \frac{(\mu(b_o u+b_o v))^{\ell}}{\ell!}$$

$$+ \sum_{r=1}^{\infty} \sum_{i=1}^{k} \beta_i T_i^r \exp(-\mu(1-T_i)b_o u) \exp(-\mu(b_1 u+b_o v))$$

$$\sum_{\ell=k-j+r+1}^{\infty} \frac{(\mu(b_1 u+b_o v))^{\ell}}{\ell!} \quad dU(u)$$

$$= (\mu m)^{-1}(p_o+..+p_{j-1}), \quad 0 < j < k + 1.$$

The equilibrium probability q_o that the queue is empty is now found from

$$1 - q_o = \sum_{i=1}^{\infty} q_i$$

$$= (\sum_{i=o}^{\infty} p_i)k/\mu m$$

as

$$q_o = 1 - k/\mu m$$

$$= 1 - \text{traffic intensity}.$$

q_o is thus independent of the functional form of U.

This establishes the theorem.

The same results hold in the special case of GI/M/1 with bulk arrivals, and can be readily deduced from equations of Conolly (1960), although his results were not expressed in this simple form.

For GI/M/1 the regenerative points and the arrival instants coincide and we have $b_o = 1, b_1 = 0$.

Conolly derived the unconditional queue length probabilities $\{q_n\}$ by integration over an inter-arrival interval;

$$q_n = \frac{bF(1)}{a} \sum_i \frac{\xi_i^{n-1}(1-\xi_i^c)}{(1-\xi_i)F'(\xi_i)}, \quad n \geq 1.$$

The ξ_i are the inner roots of the characteristic equation, c is the size of the arriving batches, corresponding to our k, and b is the mean service time, corresponding to our μ^{-1}.

$q_n(0)$ is the equilibrium probability density that the queue length is n and that an arrival has just occurred.

$$q_n(0) = \frac{F(1)}{a} \sum_i \frac{\xi_i^{n-1}}{F'(\xi_i)}, \quad n \geq 0.$$

$q_n(0)$ is, of course, zero for $0 \leq n \leq c-1$. $q_n(0)$ thus corresponds to our p_{n-k}, $n \geq k$.

In Conolly's notation, the result for GI/M/1 can be expressed as

$$q_n = b(q_n(0)+..+q_{n+c-1}(0)), \quad n \geq 1.$$

There is no factor a corresponding to our m appearing since Conolly's $q_n(0)$ is a joint probability density with $\sum_{n=o}^{\infty} q_n(0)$ normalised to a^{-1}. Our $\{p_n(0)\}$ is a conditional probability density normalised to unity.

The value for q_o for the system GI/M/1 with batch arrivals is as for G(2)/M/1, and is given by Conolly.

6. *Waiting Times.*

Consider first the waiting time distribution of the first member of an arriving batch. There is a probability p_o that such an arrival will not have to wait. If the arriving batch finds $r > 0$ customers already waiting or undergoing service, then by the lack of memory property of the negative exponential distribution, the waiting time of the first of the batch will be the iterated convolution of r service times. Thus

Pr (first of batch waits a time $\leq x$)

$$= 1 - \sum_{i=1}^{k} \beta_i (1-T_i)^{-1} \psi(b_o(1-T_i))$$

$$+ \sum_{r=1}^{\infty} \sum_{i=1}^{k} \beta_i T_i^r \psi(b_o(1-T_i))[1- \exp(-\mu x) \sum_{\ell=o}^{r-1} \frac{(\mu x)^\ell}{\ell!}]$$

$$= 1 - \sum_{i=1}^{k} \beta_i T_i (1-T_i)^{-1} \psi(b_o(1-T_i)) \exp(-\mu x(1-T_i)) .$$

Similarly the sth members of the batch, $1 < s \leq k$ has as his waiting time the iterated convolution of $s + r - 1$ service times if the batch finds r individuals already in the queue, so that

Pr(sth individual in batch waits a time $\leq x$)

$$= [1- \sum_{i=1}^{k} \beta_i T_i (1-T_i)^{-1} \psi(b_o(1-T_i))](1-\exp(-\mu x) \sum_{\ell=o}^{s-2} \frac{(\mu x)^\ell}{\ell!})$$

$$+ \sum_{i=1}^{k} \sum_{r=o}^{\infty} \beta_i T_i^r \psi(b_o(1-T_i))(1-\exp(-\mu x) \sum_{\ell=o}^{r+s-2} \frac{(\mu x)^\ell}{\ell!}), \; 1< s \leq r.$$

* * * * * * * * * * *

CHAPTER SIX

Many Server Queues.

In this chapter we extend our work to include queueing systems with two or more servers in parallel.

As we noted in Chapter One, no complete discussion of the queueing process in $Gl/G/k$ exists. The simple properties of the negative exponential distribution have, however, made $Gl/M/k$ mathematically accessible and an elegant imbedded Markov chain treatment has been given for the limiting queue length behaviour by Kendall (1953). The time dependent behaviour of $Gl/M/k$ was determined by Wu (1961).

This chapter consist of two parts. In part one, section one, we make use of the regenerative point technique of the last chapter to deal with the limiting behaviour of $G(2)/M/k$. We find the form of the solution and derive recurrence relations for the first few probabilities. These are solved for in section two. In section three we derive simple recurrence relations for the queueing system with limited waiting room. In section four we go on to consider $G(p+1)/M/k$ and in section six treat $G(2)/M/1$.

In part two we examine a queueing system with infinitely many servers, general recurrent services, and arrivals at completely arbitrary points of time. We specialise to stochastic inputs. In the case of a general recurrent input and a deterministic service time we obtain results agreeing with those of an inventory paper of Finch (1961).

PART 1.

1. The Queueing Process in G(2)/M/k.

As before, we make use of the regenerative points $\{R_m\}$, which we take as the points of the sequence $\{A_m+b_1 U_m\}$, $m \geq 0$. With the proviso that arrivals are never multiple, the sequences are disjoint and interlace, as in Chapter Five.

We define

p_{ij} = probability that queue length (i.e., number of individuals waiting or being served) at a regenerative point is j, conditional on the queue length at the previous regenerative point being i, $i, j \geq 0$,

P = the matrix with elements p_{ij}, the rows and columns of P being labelled $0,1,2,...$ instead of $1,2,..,$

$\{j|i;u\}$ = probability that the queue length is initially i and finally j in an interval of length u during which there are no arrivals.

From the theory of the simple death process or by considering virtual departures, we have

(1.1) $\{j|i;u\} = \binom{i}{j}(1 - \exp[-\mu u])^{i-j} \exp[-j\mu u], \quad j \leq i \leq k.$

If $k \leq j \leq i$, no server will be free during the interval u, so that $\{j|i;u\}$ will be the probability that $i-j$ services will be completed in u for a negative exponential service of mean service time $k\mu$, i.e.,

(1.2) $\{j|i;u\} = \exp[-k\mu u](k\mu u)^{i-j}/(i-j)!, \; k \leq j \leq i.$

If $j < k$, $i < k$, suppose that the last of the waiting customers commences service after a time $\theta < u$. Then by use of the last two results and integration, the unconditional probability $\{j|i;u\}$ is readily seen to be

$$(1.3)\quad \{j|i;u\} = \int_0^u \exp[-k\mu\theta]\,\frac{\theta^{i-k-1}}{(i-k-1)!}\,(\mu k)^{i-k}\,\{j|k;u-\theta\}\,d\theta, \qquad i > k,\ j < k.$$

Since we are taking u to be an interval during which there are no arrivals to the queue, it is immediate that

$$(1.4)\quad \{j|i;u\} = 0, \qquad j > i.$$

(1.1) - (1.4) define $\{j|i;u\}$ completely for all $i, j \geq 0$ if u is an interval of the type postulated.

As the only arrival to the queue between the regenerative instants R_m, R_{m+1} is at A_{m+1}, the subintervals (R_m, A_{m+1}), (A_{m+1}, R_{m+1}) of lengths $b_o U_m$, $b_1 U_m$ are of the type considered above, and we obtain directly the transition probabilities p_{ij} as

$$(1.5)\quad p_{ij} = \begin{cases} \int_0^\infty \sum_{\ell=o}^{i-j+1} \{i-\ell|i;b_o u\}\{j|i-\ell+1;b_1 u\}dU(u), & 0 < j \leq i+1, \\ \int_0^\infty \sum_{\ell=o}^{i} \{i-\ell|i;b_o u\}\{0|i-\ell+1;b_1 u\}dU(u), & j = 0, \\ 0, & j > i+1. \end{cases}$$

Having determined P, we establish ergodicity under the restriction

$$(1.6)\quad ka > \mu^{-1},$$

where a is the mean of $U(\cdot)$, by use of Feller's treatment of Markov chains (1950). The state i of the embedded Markov chain corresponds to a queue length i.

From (1.5), the transitions $i \to 0$, $i \to i+1$ have positive probability for each i, so that the embedded chain is irreducible, as a transition between any two states can occur in a finite number of steps. As the transition probabilities p_{ii} are all positive, all the states are aperiodic.

From Feller, we have that, unless p^n_{ij}, the probability of a transition from state i to state j in n steps, satisfies

$$p_{ij}^n \to 0 \quad \text{as} \quad n \to \infty \ , \quad \text{all} \quad i,j \geq 0,$$

then every state is ergodic and

$$p_{ij}^n \to \pi_j \quad \text{as} \quad n \to \infty, \quad \text{all} \quad i,j \geq 0,$$

where the π_j's are positive and

$$\sum_{j=0}^{\infty} \pi_j = 1.$$

If there exists a non-zero row vector $\underset{\sim}{x}$ whose components form an absolutely convergent series and for which $\underset{\sim}{x} P = \underset{\sim}{x}$, then, as $\underset{\sim}{x} = \underset{\sim}{x}P^n$, the possibility $p_{ij}^n \to 0$, all $i,j \geq 0$, as $n \to \infty$, implies

$$x_j = \sum_{i=0}^{\infty} x_i \, p_{ij}^n \to 0 \quad \text{all } j, \quad \text{as} \quad n \to \infty,$$

whereas $p_{ij}^n \to \pi_j$, all $i,j \geq 0$, as $n \to \infty$, implies

$$x_j = \sum_{i=0}^{\infty} x_i \, p_{ij}^n \to \left(\sum_{i=0}^{\infty} x_i \right) \pi_j \quad \text{as} \quad n \to \infty \ , \quad \text{all } j.$$

The existence of such a vector $\underset{\sim}{x}$ thus implies the ergodicity of the states, and when the sum of the components of $\underset{\sim}{x}$ is normalized to unity, $\underset{\sim}{x}$ becomes the limiting queue length distribution.

To establish ergodicity it thus certainly suffices to find a vector of the form

$$\underset{\sim}{x} = \{x_0, x_1, \ldots, x_{k-1}, 1, T, T^2, \ldots\}$$

such that

(1.7) $$x_j = \sum_{i=0}^{\infty} x_i \, p_{ij}, \qquad j \geq 0,$$

and with $|T| < 1$.

By (1.2), (1.5) we have that (1.7) reduces to

$$T = \int_0^{\infty} \exp\left[-(1-T)\,\mu u\right] dU(u)$$

for $j \geq k+1$, and by virtue of (1.6) Rouché's theorem yields that this equation has a unique solution inside the unit circle. As before, this solution will be real and positive.

(1.5) gives $p_{ij} = 0$ for $i < j-1$, so that, as $p_{j-1,j} \neq 0$, $j \geq 1$, the equations

(1.8) $$x_j = \sum_{i=o}^{\infty} x_i p_{ij}, \qquad j = 1,2,\ldots,k,$$

now suffice for the successive determination of $x_{k-1}, x_{k-2},\ldots,x_o$. That

$$x_o = \sum_{i=o}^{\infty} x_i p_{io}$$

is not inconsistent with the values so obtained follows from the fact that each row of P sums to unity.

This establishes that there is a limiting queue length distribution as measured at regenerative points, and that this is of the form

(1.9) $$\text{prob (queue length} = j) = \begin{cases} \pi_j, & 0 \leq j \leq k \\ T^{j-k} \pi_k, & j \geq k \end{cases}.$$

T is the solution inside the unit circle of

$$T = \int_o^{\infty} \exp[-(1-T)\mu u] \, dU(u),$$

and the above working provides a (finite) recursive procedure for the determination of the π_j's, $i \leq j \leq k$.

The instants chosen as regenerative points have significance only in the process providing the arrivals, and it is of somewhat more interest in the queueing system to know the behaviour of queue length at arrival instants. For a moving average of order one, the two, of course, coincide.

The first arrival after a regenerative point R_m is at A_{m+1}, and the length $b_o U_m$ of the interval (R_m, A_{m+1}) is independent of any of the previous history of the queue. As the number of departures from the queue during (R_m, A_{m+1}) depends only on the queue length at R_m and on the time $b_o U_m$, the queue length as found by an arrival must possess a limiting distribution, since each consecutive pair of arrivals are separated by a regenerative point. In fact, if we write

P_j = limiting probability of an arrival finding a queue of length j, $\quad j \geq 0$,

then

$$P_j = \int_0^\infty \sum_{\ell=0}^{\infty} \pi_{j+\ell}\{j|j+\ell;b_0u\}\, dU(u), \qquad j \geq 0.$$

For $j \geq k$, P_j has a simple form, since

$$\pi_{j+\ell} = T^{j+\ell-k}\, \pi_k, \qquad j \geq k,$$

and

$$\{j|j+\ell;b_0u\} = \exp[-k\mu b_0u](k\mu b_0u)^{\ell}/\ell!, \qquad j \geq k,$$

and thus

$$P_j = T^j \int_0^\infty \pi_k T^{-k} \exp[-(1-T)k\mu b_0u]\, dU(u), \quad j \geq k.$$

The corresponding form of P_j for $j < k$ is easily found to be

$$P_j = \int_0^\infty \sum_{\ell=0}^{k-j} \pi_{j+\ell}\{j|j+\ell;b_0u\}\, dU(u)$$
$$+ \int_0^\infty \pi_k T\mu k\Big[\int_0^u \exp[-(1-T)k\mu\theta]\{j|k;u-\theta\}d\theta\Big]dU(u), \quad 0 \leq j < k.$$

2. *Solution for π_ℓ, $0 \leq \ell \leq k-1$.*

The recursive relations (1.8) for $\pi_0,\ldots,\pi_{k-1}$ can be solved explicitly in simple terms if we make use of known forms of the p_{ij}.

We introduce the generating function

$$\pi(z) = \sum_{j=0}^{k-1} \pi_j z^j .$$

By (1.9), (1.8), we have

$$(2.1)\quad \pi(z) = \int_0^\infty \sum_{=o}^{k-1} \pi_i \sum_{\ell=o}^{i} \sum_{j=o}^{k-1} \{i-\ell|i;b_o u\}\{j|i-\ell+1;b_1 u\}z^j \, dU(u)$$

$$+ \int_0^\infty \pi_k \sum_{\ell=o}^{k} \sum_{j=o}^{k-1} \{k-\ell|k;b_o u\}\{j|k-\ell+1;b_1 u\}z^j \, dU(u)$$

$$+ \int_0^\infty \sum_{m=1}^{\infty} \pi_k T^m \sum_{\ell=o}^{k+m} \sum_{j=o}^{k-1} \{k+m-\ell|k+m;b_o u\} \times$$

$$\{j|k+m-\ell+1;b_1 u\}z^j dU(u)$$

$$= \int_0^\infty [1+(z-1)\exp(-\mu b_1 u)]\pi[1+ \exp(-\mu b_o u)\,(z-1)\,\exp(-\mu b_1 u)]$$

$$- \pi_{k-1} \exp[-(k-1)\mu b_o u] \exp[-k\mu b_1 u]\, dU(u)$$

$$+\int_0^\infty [1+(z-1)\exp(-\mu b_1 u)]\pi_k(1+ \exp(-\mu b_o u)\,(z-1)\exp(-\mu b_1 u)]^k$$

$$- \pi_k \exp(-k\mu b_o u)\,[1+(z-1)\exp(-\mu b_1 u)]^{k+1}$$

$$+ \pi_k \exp(-k\mu b_o u) \int_0^{b_1 u} \exp(-k\mu\theta)\mu k \times$$

$$\{[1+(z-1)\exp\{-\mu(b_1 u-\theta)\}]^k - z^k \exp(-k\mu(b_1 u-\theta))\}d\theta\, dU(u)$$

$$+ \pi_k \int_0^\infty \sum_{m=1}^{\infty} \int_0^{b_o u} \exp(-k\mu\theta)\, \frac{\theta^{m-1}}{(m-1)!}\, (T\mu k)^m[1+(z-1)\exp(-\mu b_1 u)] \times$$

$$[\{1+(\exp(-\mu(b_o u-\theta)))(z-1)\exp(-\mu b_1 u)\}^k$$

$$- \{\exp(-\mu k(b_o u-\theta))\}\{1+(z-1)\exp(-\mu b_1 u)\}^k]d\theta$$

$$- \sum_{m=1}^{\infty} \int_0^{b_o u} \exp(-k\mu\theta)\, \frac{\theta^{m-1}}{(m-1)!}\, (T\mu k)^m \times$$

$$k\{1- \exp(-\mu(b_o u-\theta))\} \exp(-(k-1)\mu(b_o u-\theta))d\theta \times$$

$$z^k \exp(-k\mu b_1 u) \quad dU(u)$$

$$+ \pi_k \int_0^\infty \sum_{m=1}^{\infty} \sum_{i=o}^{m} (\exp(-k\mu b_o u))(k\mu b_o u)^{m-i}/(m-i)! \times$$

$$\int_0^{b_1u} \exp(-k\mu\theta) \frac{\theta^i}{i!} (T\mu k)^{i+1} \times$$

$$\{[1+(1-z) \exp(-\mu(b_1u-\theta))]^k - z^k \exp(-k\mu(b_1u-\theta))\} d\theta \quad dU(u),$$

making use of (1.1)-(1.4).

We now define

$$q_j = \frac{1}{j!} \left[\frac{d^j}{dz^j} \pi(z)\right]_{z=1}, \quad 0 \le j \le k-1.$$

By iterated differentiation of (2.1) we find

$$q_j = \int_0^\infty q_j \exp(-j\mu u)+q_{j-1}\exp(-\mu b_1 u)\exp(-(j-1)\mu u)-\binom{k}{j}\pi_{k-1}\exp(-\mu u(k-b_0)) \, dU(u)$$

$$+ \pi_k \int_0^\infty \binom{k}{j}\exp(-j\mu u)+j\binom{k-1}{j-1}\exp(-\mu b_1 u)\exp(-(j-1)\mu u)$$

$$+ \exp(-k\mu b_0 u) \int_0^{b_1u} \mu k \exp(-k\mu\theta)\binom{k}{j}[\exp(-j\mu(b_1u-\theta))$$

$$- \exp(-k\mu(b_1u-\theta))]d\theta \quad dU(u)$$

$$+ \pi_k \int_0^\infty \int_0^{b_0u} \exp(-k\mu\theta(1-T)) \times$$

$$\{[\binom{k}{j}\exp(-\mu j(u-\theta))+j\binom{k}{j-1}\exp(-\mu b_1 u) \exp(-\mu(j-1)(u-\theta))]$$

$$- \exp(-\mu k(b_0u-\theta))\binom{k+1}{j}\exp(-\mu b_1 u)\}d\theta$$

$$- \int_0^{b_0u} \exp(-k\mu\theta(1-T)) \, k\binom{k}{j}\{\exp(-\mu(k-1)(b_0u-\theta))\exp(-\mu b_1 u)$$

$$- \exp(-k\mu(u-\theta))\}d\theta \quad dU(u)$$

$$+ \pi_k \int_0^\infty T\mu k \int_0^{b_1u} \exp(-k\mu\theta(1-T))\binom{k}{j}[\exp(-\mu j(b_1u-\theta))-\exp(-\mu k(b_1u-\theta))]d\theta$$

$$dU(u), \qquad j > 0.$$

This is a linear first order difference equation for the quantities q_j. It is more naturally written in terms of

$$q_j' = q_j / \prod_{i=1}^{j} \psi(j-1+b_1)\ (1-\psi(j))^{-1}$$

as

(2.2) $$q_j' = q_{j-1}' + f_j \pi_k, \qquad j \geq 0,$$

where f_j is a known function of j and we take

$$q_o' = q_o.$$

The term in π_{k-1} has been expressed as a product of π_k with a known function which has been incorporated in f. The appropriate form for π_{k-1} readily follows from (1.9),(1.8):

$$\pi_k = \pi_{k-1} p_{k-1,k} + \sum_{m=o}^{\infty} \pi_k T^m p_{k+m,k},$$

on substitution for the transition probabilities.

By adding (2.2) for $j = i+1,..,k-1$ we see that

(2.3) $$q_i' = \begin{cases} g - \sum_{j=i+1}^{k-1} f_j \pi_k, & 0 \leq i \leq k-2, \\ g & i = k-1, \end{cases}$$

where

$$g = \pi_{k-1} \prod_{i=1}^{k-1} \psi(i-1+b_1)(1-\psi(i))^{-1}.$$

Since

$$\pi(1) = \sum_{j=o}^{k-1} \pi_j = 1 - \sum_{j=k}^{\infty} \pi_k T^{j-k}$$

we have

$$q_o = 1 - \pi_k(1-T)^{-1}$$

and

$$q_o' = 1 - \pi_k(1-T)^{-1}.$$

This relation suffices for the elimination of the unknown π_k from (2.3).

$$\pi_k = (g-1)/(\sum_{j=1}^{k-1} f_j + (1-T)^{-1})$$

The probabilities $\pi_o,\ldots,\pi_{k-1}$, can now be determined from

$$\pi_j = \frac{1}{j!} \left[\frac{d^j}{dz^j} \pi(z)\right]_{z=o}$$

$$= \sum_{i=j}^{k-1} (-)^{i-j} \binom{i}{j} q_i,$$

where

$$q_i = \begin{cases} g - \sum_{j=1}^{k-1} f_j(g-1)/(\sum_{j=1}^{k-1} f_j + (1-T)^{-1}), & i = 0, \\ [g - \sum_{j=i+1}^{k-1} f_j(g-1)/(\sum_{j=1}^{k-1} f_j + (1-T)^{-1})] \times \\ \quad \prod_{j=1}^{i} \psi(j-1+b_1)(1-\psi(j))^{-1}, & 1 \le i \le k-2, \\ g \prod_{j=1}^{i} \psi(j-1+b_1)(1-\psi(j))^{-1}, & i = k-1. \end{cases}$$

3. *Finite waiting room.*

Suppose that the maximum number of customers that can wait (including the k in service) is restricted to some finite number $m \geq k$, and that any arrivals finding m customers in the system are lost.

Then equations (1.1)-(1.4) are still valid, but (1.5) must be modified to

$$
p_{ij} = \begin{cases}
\int_0^\infty \sum_{\ell=0}^{i} \{i-\ell|i;b_0u\}\{0|i-\ell+1;b_1u\}dU(u), & j = 0,\ i \neq m, \\
\int_0^\infty \sum_{\ell=0}^{i-j+1} \{i-\ell|i;b_0u\}\{j|i-\ell+1;b_1u\}dU(u), & 0 < j \leq i+1, i\neq m, \\
0\ , & j > i+1, \\
\int_0^\infty \sum_{\ell=1}^{m} \{m-\ell|m;b_0u\}\{0|m-\ell+1;b_1u\}dU(u) & \\
\quad + \int_0^\infty \{m|m;b_0u\}\{0|m;b_0u\}dU(u), & i = m,\ j = 0, \\
\int_0^\infty \sum_{\ell=1}^{m-j} \{m-\ell|m;b_0u\}\{j|m-\ell+1;b_1u\}dU(u) & \\
\quad + \int_0^\infty \{m|m;b_0u\}\{j|m;b_1u\}dU(u), & i = m,\ 0 < j < m, \\
\int_0^\infty \{m|m;b_0u\}\{m|m;b_1u\}dU(u), & i = j = m.
\end{cases}
$$

The establishing of ergodicity follows as in section one. A non-zero row vector $\underset{\sim}{x} = \{x_o, x_1, .., x_m\}$ is required satisfying

(3.1) $$x_j = \sum_{i=o}^{m} x_i\, p_{ij}, \quad 0 \leq j \leq m.$$

As before $p_{ij} = 0$ for $i < j-1$, $j \geq 2$, so that (3.1) can be written as

(3.2) $$x_j = \sum_{i=j-1}^{m} x_i\, p_{ij}, \quad 1 \leq j \leq m,$$

(3.3) $$x_o = \sum_{i=o}^{m} x_i\, p_{io}.$$

Since $p_{j-1,j} \neq 0$, $1 \leq j \leq m$, if we assign to x_m an arbitrary non-zero real value, the equations (3.2) provide successively solutions for x_{m-1}, $x_{m-2}, .., x_o$. That the values thus obtained satisfy (3.3) also follows as before. Since the limiting probability distribution must also satisfy(3.1), this limiting distribution is obtained by normalisation of the constructed vector $\underset{\sim}{x}$.

4. *General moving average.*

The simple method used in the above sections does not readily generalise for moving averages of higher order, for which we shall work directly in terms of the more natural arrival instants using Laplace-Stieltjes transforms and complex variable theory as in Chapter Two. This method is, of course, available for the special case of G(2)/M/k, but is less convenient for an explicit determination of the P_j's for a particular $U(\cdot)$.

We give the argument in outline only.

The queue length transition probabilities $\{j|i;u\}$ for intervals without arrivals still possess the forms given by (1.1)-(1.4), so that, using (1.1), we have, from comparing the queue lengths at $A_m - 0$, $A_{m+1} - 0$, that

$$(4.1)\quad P_j(u^{(m+p)}) = \begin{cases} \sum_{i=o}^{\infty} P_i(u^{(m+p-1)})\{0|i+1;b_o u_{m+p}+..+b_p u_m\}, & j = 0, \\ \sum_{i=o}^{\infty} P_{j+i-1}(u^{(m+p-1)})\{j|j+i;b_o u_{m+p}+..+b_p u_m\}, & 0 < j < k, \\ \sum_{i=o}^{\infty} P_{j+i-1}(u^{(m+p-1)})\exp[-k\mu(b_o u_{m+p}+..+b_p u_m)] \times & \\ \qquad [k\mu(b_o u_{m+p}+..+b_p u_m)]^i/i!, & j \geq k. \end{cases}$$

Forming the generating function

$$P(u^{(m+p)};z) = \sum_{j=o}^{\infty} P_j(u^{(m+p)})z^j, \qquad |z| \leq 1,$$

from these relations leads to

$$\begin{aligned}(4.2)\quad P(u^{(m+p)};z) = {} & zP(u^{(m+p-1)};z)\exp[-(1-z^{-1})k\mu(b_o u_{m+p}+..+b_p u_m)] \\ & - z\sum_{j=o}^{\infty} P_j(u^{(m+p-1)})z^j \; B^{(m+p)}_{\max(0,j-k+2)}(z) \\ & + \sum_{i=o}^{\infty} P_i(u^{(m+p+1)})\{0|i+1;\, b_o u_{m+p}+..+b_p u_m\}\end{aligned}$$

$$+ \sum_{j=1}^{k-1} z^j \sum_{i=o}^{\infty} P_{j+i-1}(u^{(m+p-1)})\{j|j+i;b_o u_{m+p},..,b_p u_m\},$$

for $k \geq 2$, $|z| \leq 1$, $z \neq 0$,

where we define

$$B_i^{(m+p)}(z) = \sum_{\ell=i}^{\infty} \exp[-k\mu(b_o u_{m+p}+..+b_p u_m)](k\mu z^{-1}[b_o u_{m+p}+..+b_p u_m])^{\ell}/\ell!,$$

$i \geq 0$, $|z| \leq 1$, $z \neq 0$.

On taking Laplace-Stieltjes transforms, we note that the transform of the sum of the second, third and fourth expressions on the right hand side of (4.2)can be written in the form

$$\sum_{j=o}^{\infty} c_j(s^{(p)};m)z^{k-1-j} = c(s^{(p)};m;z).$$

We obtain

(4.3) $P^*(s^{(p)};z;m+1)$

$$= zP^*\{(1-z^{-1})k\mu b_1+s_{p-1},(1-z^{-2})k\mu b_2+s_{p-2},..,(1-z^{-1})k\mu b_{p-1}+s_1,(1-z^{-1}) \times k\mu b_p;z;m\}$$

$$\times\ \psi\{(1-z^{-1})kb_o+s_p/\mu\}+c(s^{(p)};m;z), \quad z\varepsilon R, \quad \text{Re. } s_i \geq 0,\ 1 \leq i \leq p,$$

where R is the subset of $|z| \leq 1$ defined in chapter two.

As in the case of a single server,

(4.4) $$\int_o^{\infty} u dU(u) > (\mu k)^{-1}$$

suffices for a proper queue length limiting distribution to exist (from Loynes (1962)). We write

$$P_j = \lim_{m\to\infty} E[P_j(U^{(m+p-1)})], \quad j \geq 0,$$

and we can justify the definition

$$P(w_1,w_2,..,w_p;z) = \lim_{m\to\infty} E[P(U_o,U_1,..,U_{m-1},u_m,u_{m+1},..,u_{m+p-1};z)], |z| \leq 1,$$

taking the expectation with respect to $U_o,..,U_{m-1}$, and where $w_1,w_2,..,w_p$ are the particular values u_m, $u_{m+1},..,u_{m+p-1}$. We denote its integral transform by

$$(4.5)\quad P^*(s^{(p)};z) = E[P(W^{(p)};z)\exp(-s_pW_p-s_{p-1}W_{p-1}-..-s_1W_1)], \quad |z|\le 1, \mathrm{Re}.s_i\ge 0,$$

where the W_i are identically and independently distributed random variables with common distribution function $U(\cdot)$. $c^*(s^{(p)};z)$ is defined by

$$c^*(s^{(p)};z) = \lim_{m\to\infty} c(s^{(p)};m;z), \quad |z| \le 1, \quad \mathrm{Re}.\ s_i \ge 0.$$

Letting $m \to \infty$ in (4.3) and making recursive substitutions gives

$$\begin{aligned}(4.6)\quad P^*(s^{(p)};z) = {} & c^*(s^{(p)};z) + z\psi\{(1-z^{-1})kb_o+s_p/\mu\} \times \\ & [c^*\{(1-z^{-1})\mu kb_1+s_{p-1},..,(1-z^{-1})\mu kb_{p-1}+s_1,(1-z^{-1})\mu kb_p:z\} \\ & + z\psi\{(1-z^{-1})k(b_o+b_1)+s_{p-1}/\mu\} \times \\ & [c^*\{(1-z^{-1})\mu k(b_1+b_2)+s_{p-2},..,(1-z^{-1})\mu k(b_{p-1}+b_p),(1-z^{-1})\mu kb_p;z\} \\ & + z\psi\{(1-z^{-1})k(b_o+b_1+b_2)+s_{p-2}/\mu\} \times \\ & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ & \times [c^*\{(1-z^{-1})\mu k(b_1+..+b_p),..,(1-z^{-1})\mu kb_p;z\} \times \\ & [1-z\psi\{(1-z^{-1})\ k(b_o+b_1+..+b_p)\}]^{-1}]]..], \\ & z\varepsilon R, \quad \mathrm{Re}.\ s_i \ge 0.\end{aligned}$$

Denote the expression on the right hand side of (4.6) by $D(s^{(p)};z)$. Since $P^*(s^{(p)};z)$ is the generating function of a probability distribution, $D(s^{(p)};z)$ must be a regular function of z for $|z| \le 1$, $\mathrm{Re}.\ s_i \ge 0$, by analytic continuation.

It can be shown that $D(s^{(p)};z)$ is a regular function of z for $|z| \ge 1$, $\mathrm{Re}.\ s_i \ge 0$, except where $[1-z\psi\{(1-z^{-1})k\}]$ vanishes, i.e., where $z = T^{-1}$, when it has a simple pole.

Also $D(s^{(p)};z) = O(|z|^{p+k-2})$ for $|z| \to \infty$.

Hence $P^*(s^{(p)};z)$ is of the form

(4.7) $$P^*(s^{(p)};z) = \sum_{\ell=o}^{p+k-2} B_\ell(s^{(p)})z^\ell + B(s^{(p)})(1-Tz)^{-1}, \quad \text{Re. } s_i \geq 0, \ |z| \leq 1,$$

where the $B_\ell(s^{(p)})$ and $B(s^{(p)})$ are functions of the s's alone.

When $s_i = 0$, $1 \leq i \leq p$, $P^*(s^{(p)};z)$ becomes the generating function $\sum_{i=o}^{\infty} P_i z^i$ of the limiting distribution of queue size as found by arrivals and the B's to constants, B and B_j, $0 \leq j \leq p+k-2$. The limiting distribution is thus geometric from P_{p+k-1} onwards, with common ratio T. This result was established by D.G.Kendall (1953) in the special case p = o, i.e., a general recurrent input. The B's can be determined much as in Chapter Two.

5. *Limiting distribution of waiting times.*

If an arrival finds k-1 or fewer customers already in the system, he does not have to wait to commence service. If he finds $j \geq k$ customers already present, he has to wait until j+1-k of these have been served before his service commences. As the probability of precisely j+1-k such services being completed in a time x is

$$\exp(-k\mu x)\,(k\mu x)^{j+1-k}/(j+1-k)!,$$

the waiting time for arrivals can readily be determined.

We write $pr(w \leq x)$ for the limiting cumulative probability of an arrival having to wait a time $\leq x$ before beginning service. Then

$$\begin{aligned} pr(w\leq x) &= \sum_{j=o}^{k-1} P_j + \sum_{j=k}^{\infty} P_j\Big(1-\sum_{i=o}^{j-k} \exp(-k\mu x)(k\mu x)^i/i!\Big) \\ &= 1 - \sum_{i=o}^{p-2}\Big(\sum_{j=i}^{p-2} B_{k+j}\Big)\exp(-k\mu x)(k\mu x)^i/i! \\ &\quad - T^k(1-T)^{-1} B \exp\{-k\mu x(1-T)\}, \quad x \geq 0,\ p \geq 2. \end{aligned}$$

When $p = 1$, i.e., when the input is of general recurrent form, it is readily verified that the terms involving the B_{k+j}'s do not appear in the result, so that the waiting time distribution reduces to a weight at the origin combined with a negative exponential distribution. This result was first noted by D. G. Kendall (1953).

6. *Infinitely many servers.*

The equilibrium queue length distribution of $G(2)/M/\infty$ can be obtained much as $\pi_o,\ldots,\pi_{k-1}$ in section two, working in terms of the regenerative points $\{R_n\}$.

Suppose that the equilibrium queue length distribution on the $\{R_n\}$ has the probability generating function

$$P(z) = \sum_{i=o}^{\infty} P_i z^i, \qquad |z| \leq 1.$$

The transition probabilities p_{ij} for the queue lengths at consecutive regenerative points R, R' have simpler expressions than before, since the fourfold form (1.1)-(1.4) simplifies to

$$\{j|i;u\} = \begin{cases} \binom{i}{j}((-\exp(-\mu u))^{i-j} \exp(-j\mu u), & 0 \leq j \leq i. \\ 0 & j > i. \end{cases}$$

The p_{ij} maintain the form

$$P_{ij} = \begin{cases} \int_0^\infty \sum_{\ell=o}^{i-j+1} \{i-\ell|i;b_o u\} \{j|i-\ell+1;b_1 u\} \ dU(u), & 0 < j \leq i+1, \\ \int_0^\infty \sum_{\ell=o}^{i} \{i-\ell|i;b_o u\} \{j|i-\ell+1;b_1 u\} \ dU(u), & j = 0, \\ 0 \ , & j > i+1. \end{cases}$$

(2.1) becomes

$$(6.1)\quad P(z) = \int_0^\infty [1+(z-1)\exp(-\mu b_o u)]P[1+(z-1)\exp(-\mu b_o u)\exp(-\mu b_1 u)] \quad dU(u).$$

$P(z)$ is no longer a polynomial and it is not immediate that the quantities

$$q_j = \frac{1}{j!}\left[\frac{d^j}{dz^j} P(z)\right]_{z=1}, \qquad j \geq 0,$$

still exist. We shall presume that they do, our *a posteriori* justification residing in the fact that our solution gives finite values for the q_j.

Differentiation of (6.1) gives

$$q_j = q_j \int_0^\infty \exp(-j\mu u)\, dU(u) + q_{j-1} \int_0^\infty \exp(-\mu u(j-b_1))\, dU(u), \quad j \geq 1,$$

that is

$$q_j = q_{j-1}\, \psi(j-b_1)/(1 - \psi(j)). \quad j \geq 1,$$

whence

$$q_j = q_o \prod_{i=1}^{j} [\psi(i-b_1)/(1-\psi(i))].$$

We note that since

$$\psi(\alpha) \to 0 \quad \text{as } \alpha \to \infty ,$$

$\psi(j-b_1)/(1-\psi(j))$ is ultimately less than 1/2, so that $\sum_{j=o}^{\infty} q_j$ is actually a convergent series. Our assumption is clearly justified.

Since

$$q_o = P(1) = 1,$$

we have

$$q_j = \prod_{i=1}^{j} [\psi(i-b_1)/(1-\psi(i))], \quad j \geq 0,$$

where we take the empty product as unity.

The probabilities $\{ P_j \}$ can now be determined from

$$P_j = \frac{1}{j!}\left[\frac{d^j}{dz^j} \sum_{i=o}^{\infty} q_i (z-1)^i\right]_{z=o}$$

$$= \sum_{i=j}^{\infty} (-)^{i-j} \binom{i}{j} q_i .$$

The equilibrium distribution $\{Q_j\}$ of the number of customers in the system as found by arrivals can now be found.

An arrival will find j customers in the system if and only if there were $j+m$ $(m \geq 0)$ customers in the system at the last regenerative point and there have subsequently been m departures. Thus

$$Q_j = \sum_{m=o}^{\infty} P_{j+m} \int_o^{\infty} \binom{j+m}{j} (1-\exp(-\mu b_o u))^m \exp(-j\mu b_o u) \quad dU(u), \quad j \geq 0.$$

PART 2

A queueing system with non-recurrent input and general batch servicing.

In this part we consider a system with an arbitrary input and general recurrent batch services. The number of servers is infinite. Section one gives the generating function and mean and variance of the queue length distribution at an arbitrary instant of time. In sections two and three we specialise to stochastic inputs, i.e., inputs for which the inter-arrival intervals possess a joint probability distribution. We verify that our results reduce to those of Finch (1961) in the special cases of a general recurrent input and a general recurrent input with constant service times.

Suppose that arrivals occur at times t_n, $n = 1,2,..$, and that after every kth arrival a servicing of k arrivals is begun. We assume that the number of servers is infinite. Initially, at $t_o = 0$, the system is empty and the arrival process $\{t_n\}$ is about to start. The batch service times are I.I.D. with distribution function $L(x)$, $x \geq 0$. No assumption is made about the process $\{t_n\}$. We define $\eta(t;t_1,t_2,..)$ to be the queue length, i.e., the number of individuals awaiting service or being served at time t.

The system can be regarded as a generalisation to bulk service of a telephone traffic problem considered by Finch (1963a) or as a generalisation of an inventory problem of Finch (1961). In the telephone traffic problem, calls arrive at prescribed times at a telephone exchange with infinitely many channels. The holding times of the calls are non-negative random variables distributed independently of one another, the channels concerned, and the arrival times.

Finch's inventory model is the infinite bin system of inventory control. Demands for an item occur at times t_n, $n = 1,2,..,$ separated by I.I.D. intervals, and after every kth demand an order for a replacement of k items is made. Initially, at $t_o = 0$, the bin is full of items held in stock and the demand process $\{t_n\}$ is about to start. The bin is taken to have infinite capacity and the lead times of the orders, i.e., the time intervals between the placing of the demands and the corresponding deliveries, are I.I.D.

Here we generalise Finch (1961) to non-recurrent demand times. The correspondence is obtained by replacing arrivals by demands and orders by services (the instants of arrival of deliveries corresponding to the instants of completion of calls).

In (1963a), Finch obtains the limiting value of the variance of η_{m+1} (the number of calls in the system at time t_m-0) for a holding time distribution $B(x)$ given by

$$B(x) = \begin{cases} 1-e^{-x}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$$

when the inter-arrival intervals, $t_{m+1}-t_m$, are given by

$$t_{m+1}-t_m = u_m+bu_{m-1} \ ,$$

b being positive and the u_m I.I.D. non-negative random variables. This is essentially our second order moving average.

We note that the value given for $\lim_{n\to\infty} \text{Var}\ (\eta_{m+1})$,

$$\lim_{n\to\infty} \text{Var}(\eta_{m+1}) = \psi(1)\psi(b)[1-\psi(1)\psi(b)]\ [1-\psi(1+b)]^{-1}$$
$$+ 2\psi(2)\psi(b)\psi(1+2b)[1-\psi(1+b)]^{-1}[1-\psi(2+2b)]^{-1},$$

where

$$\psi(s) = E[\exp(-su_m)],$$

is incorrect and should read

$$\lim_{n\to\infty} \mathrm{Var}(\eta_{m+1}) = \psi(1)\psi(b)[1-\psi(1)\psi(b) - \psi(1+b)]\,[1-\psi(1+b)]^{-2}$$
$$+ 2\psi(2)\psi(b)\psi(1+2b)[1-\psi(1+b)]^{-1}\,[1-\psi(2+2b)]^{-1}.$$

1. *Transient behaviour of* $\eta(t;t_1,t_2,..)$

After the batch beginning service at t_k has completed its service, the queue length is the same as if the arrivals at $t_1,t_2,..,t_k$ and the servicing at t_k had all not occurred, i.e., the queue length is the same as if the first arrival had been at t_{k+1} and that after that the process had proceeded as before.

When the servicing commencing at t_k has begun but not been completed, the queue length is precisely k greater than if the arrival service process had begun with a first arrival at t_{k+1} and first service at t_{2k}.

Hence if h is the duration of the service beginning at t_k, we have, for $t > t_k$

$$(1.1)\quad \eta(t;t_1,t_2,..) = \begin{cases} \eta(t;t_{k+1},t_{k+2},..) & \text{if } h \le t-t_k \\ \eta(t;t_{k+1},t_{k+2},..) & \text{if } h \ge t-t_k \end{cases}.$$

Define $P(z,t;t_1,t_2,..)$ by

$$(1.2)\quad P(z,t;t_1t_2,..) = \sum_{j=o}^{\infty} P_j(t;t_1,t_2,..)z^j, \qquad |z| \le 1,$$

where

$$(1.3)\quad P_j(t;t_1,t_2,..) = \Pr\,[\eta(t;t_1,t_2,..) = j\,|\,t_1,t_2,..].$$

On multiplication of (1.3) by z^j and summing, we obtain, from (1.1), (1.2), that for $t \ge t_k$

$$P(z,t;t_1,t_2,..) = [L(t-t_k)+z^k\,\{1-L(t-t_k)\}]P(z,t;t_{k+1},t_{k+2},..), \quad |z| \le 1,$$

where, as given in the previous section, $L(x)$ is the service time distribution function.

Since a similar argument can be applied to $P(z,t;t_{k+1},t_{k+2},..)$, we can make use of induction to express $P(z,t:t_1,t_2,..)$ in the form

(1.4) $$P(z,t;t_1,t_2,..) = z^u \prod_{s=1}^{r} [L(t-t_{sk})+z^k\{1-L(t-t_{sk})\}], \quad |z| \leq 1,$$

where r is the greatest non-negative integer for which $t-t_{rk} \geq 0$ (we take the right hand side of (1.4) to be z^u if $r = 0$), and u is the number of values of v for which $t_{rk} < t_v \leq t$. The term z^u arises from the arrivals in $(t_{rk},t]$, which must still be present in the system as their service has not commenced.

From the generating function $P(z,t;t_1,t_2,..)$ given by (1.4), we obtain in the usual way the following expressions for the mean and variance of the distribution of $\eta(t;t_1,t_2,..)$ at a given time t:

$$E[\eta(t;t_1,t_2,..)] = \begin{cases} u+k \sum_{s=1}^{r} [1-L(t-t_{sk})], & r > 0 \\ u, & r = 0 \end{cases},$$

$$Var[\eta(t;t_1,t_2,..)] = \begin{cases} k^2 \sum_{s=1}^{r} [\{1-L(t-t_{sk})\}L(t-t_{sk})], & r > 0 \\ 0, & r = 0 \end{cases}.$$

2. *Transient behaviour : stochastic arrival times.*

Define

$$\theta_{m,j} = t_{m+1} - t_{m-j}, \quad 0 \leq j \leq m, \quad t_o = 0,$$

and suppose that the arrival times form a stochastic process, with the non-negative random variables $\theta_{m,j}$ having joint distribution function

$$F_m(x_o,x_1,..,x_m) = Pr(\theta_{m,j} \leq x_j, \; 0 \leq j \leq m).$$

If $P_j^{(m+1)}$ denotes the unconditional probability that the (m+1)th arrival finds j individuals in the system, $j = 0,1,..,m$, then by writing r as $[m/k]$, and u as $m-[m/k]k$, substituting t_{m+1} for t in (1.4), and integrating, we obtain

$$P^{(m+1)}(z) = \int z^{m-[m/k]k} \prod_{s=1}^{[m/k]} [L(\phi_{m,m-sk})+z^k\{1-L(\phi_{m,m-sk})\}dH_{m,k}(\phi), \qquad |z| \leq 1,$$

where $H_{m,k}(\phi)$ is the joint probability distribution function

$$H_{m,k}(\phi) = Pr(\theta_{m,m-jk} \leq \phi_{m,m-jk}, \quad j = 1,2,..,[m/k]),$$

and

$$P^{(m+1)}(z) = \sum_{j=o}^{m} P_j^{(m+1)} z^j.$$

The mean value, $E(\eta_{m+1})$, of the queue length as found in the system by the (m+1)th arrival can be directly deduced to be

$$E(\eta_{m+1}) = k \int_{o-}^{\infty} [1-L(z)]dG_{m+1}(x) + m - [m/k]k,$$

and the corresponding variance, $Var(\eta_{m+1})$, to be

$$Var(\eta_{m+1}) = k^2 \int_{o-}^{\infty} \int_{o-}^{\infty} \{1-L(x)\}\{1-L(y)\}dG_{m+1}(x,y)-k^2[\int_{o-}^{\infty}\{1-L(x)\}dG_{m+1}(x)]^2,$$

where

$$G_{m+1}(x) = \sum_{s=1}^{[m/k]} F_{m,s}(x),$$

$$G_{m+1}(x,y) = \sum_{s=1}^{[m/k]} F_{m,s,t}(x,y),$$

$$F_{m,s}(x) = Pr(\theta_{m,m-sk} \leq x),$$

$$F_{m,s,t}(x,y) = Pr(\theta_{m,m-sk} \leq x, {}_{m,m-tk} \leq y), \quad s \neq t.$$

These results are extensions to bulk service of the corresponding results given in Finch (1963a).

From (1.4) also follow immediately expressions for the corresponding generating function (which we denote by $P(z,t)$), mean and variance for $\eta(t)$ (the unconditional queue length at time t), i.e., results analogous to the above for times no longer restricted to occur at the instant of a particular arrival. We express t_{sk}, $s = 1,..,[m/k]$ as $\theta_{m,m}-\theta_{m,m-sk}$.

$$P(z,t) = \sum_{m=o}^{\infty} \int z^{m-[m/k]k} \prod_{s=1}^{[m/k]} [L(t-\theta_{m,m}+\theta_{m,m-sk})+z^k\{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\}]dH_{m,k,o}(\theta),$$

$$|z| \leq 1,$$

$$E[\eta(t)] = \sum_{m=o}^{\infty} \int [k \sum_{s=1}^{[m/k]} \{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\}+m-[m/k]dH_{m,k,o}(\theta),$$

$$\begin{aligned}Var[\eta(t)] = \sum_{m=o}^{\infty} \int [(m-[m/k]k)^2+\{k^2+2k(m-[m/k]k)\} \sum_{s=1}^{[m/k]} \{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\} \\ + k^2 \sum_{\substack{s,q=1 \\ s\neq q}}^{[m/k]} \{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\} \{1-L(t-\theta_{m,m}+\theta_{m,m-qk})\}]dH_{m,k,o}(\theta) \\ - (E[\eta(t)])^2,\end{aligned}$$

where in each expression the integration is carried out for

$$\theta_{m,m}-\theta_{m,m-pk} < t,\ p = 1,2,.., [m/k] \text{ and } \theta_{m,m}-\theta_{m,o} < t;\ \theta_{m,m} \geq t,$$

and where

$$H_{m,k,o}(\phi) = Pr\{\theta_{m,o} \leq \phi_{m,o}, \theta_{m,m} \leq \phi_{m,m}, \theta_{m,m-jk} \leq \phi_{m,m-jk},\ j = 1,2,..,[m/k]\}.$$

3. *Transient behaviour : number of busy servers.*

Denote by $\xi(t;t_1,t_2,..)$ the number of servers occupied at time t and by $U(z,t;t_1,t_2,..)$ the corresponding generating function. Working similar to that used in considering $\eta(t;t_1,t_2,..)$ yields

$$(3.1)\quad U(z;t;t_1,t_2,..) = \prod_{s=1}^{r} [L(t-t_{sk})+z\{1-L(t-t_{sk})\}],\quad |z| \leq 1,$$

$$(3.2)\quad E[\xi(t;t_1,t_2,..)] = \begin{cases} \sum_{s=1}^{r} [1-L(t-t_{sk})], & r > 0 \\ 0 & , \quad r = 0 \end{cases},$$

$$(3.3)\quad \mathrm{Var}[\xi(t;t_1,t_2,..)] = \begin{cases} \sum_{s=1}^{r} [1-L(t-t_{sk})]L(t-t_{sk}), & r > 0 \\ 0 & , \quad r = 0 \end{cases},$$

where r is the greatest non-negative integer for which $t-t_{rk} \geq 0$ (we take the right hand side of (3.1) to be zero if $r = 0$).

(3.1), (3.2), (3.3) agree with the results obtained by Finch (1961) for the system D(G)k. In this notation the first letter refers to the demand process distribution, the second to the lead time distribution, and the third gives the re-order quantity. D denotes a deterministic distribution of a random variable which is constant with probability 1, G to a general distribution.

Let $U_j^{(m+1)}$ denote the unconditional probability that at time $t_{m+1}-0$ there are j busy servers, $j = 0,1,..,m$, and let $U^{(m+1)}(z) = \sum_{j=o}^{[m/k]} U_j^{(m+1)} z^j$. Then, analogously to the results in section 3, we obtain

$$(3.4)\quad U^{(m+1)}(z) = \int \prod_{s=1}^{[m/k]} [L(\theta_{m,m-sk})+z\{1-L(\theta_{m,m-sk})\}]dH_{m,k}(\theta),$$

$$(3.5)\quad E(\xi_{m+1}) = \int_{o-}^{\infty} [1-L(x)]dG_{m+1}(x),$$

$$\mathrm{Var}(\xi_{m+1}) = \int_{o-}^{\infty} [1-L(x)]L(x)dG_{m+1}(x),$$

where

ξ_{m+1} is $\xi(t;t_1,t_2..)$ evaluated at $t = t_{m+1} - 0$.

We have obtained generally as (3.4) the distribution for ξ_{m+1}, which is not derived by Finch. Finch gives $E(\xi_{m+1})$ for the case G(G)k and finds the distribution of ξ_{m+1} for G(D)k, i.e., when the lead time is constant. We verify below that in these cases (3.4), (3.5) simplify to yield the same results.

When the intervals between successive demands are identically and independently distributed with distribution function $A(x)$, say, $F_{m,s}(x)$ becomes $A^{(m+1-sk)}(x)$, where we write $A^{(r)}(x)$ for the rth-fold iterated convolution of $A(x)$, r a positive integer, and

$$E(\xi_{m+1}) = \sum_{s=1}^{[m/k]} \int_{0-}^{\infty} [1-L(x)]d[A^{(m+1-sk)}(x)]$$

$$= \sum_{s=1}^{[m/k]} \int_{0-}^{\infty} A^{(m+1-sk)}(x)dL(x),$$

agreeing with the result $\Sigma_{t=o}^{n-1}\ b_t(w)$ obtained by Finch, where

$b_t(w) = \int_0^\infty A^{(tk+w+1)}(x)dL(x)$ and $m = tk+w, t = [m/k]$, $w = 0,1,..,k-1$.

Suppose further that the lead times have constant length ℓ.

If for some s, $1 \le s \le [m/k]$, $\theta_{m,m-sk} < \ell$, then $\theta_{m,m-tk} < \ell$ for $s \le t \le [m/k]$.

$$L(\theta_{m,m-sk}) + z\{1-L(\theta_{m,m-sk})\} = \begin{cases} z, & \theta_{m,m-sk} < \ell \\ 1, & \theta_{m,m-sk} \ge \ell \end{cases}.$$

Hence

$$U_{[m/k]-s}^{(m+1)} = \int_{0-}^{\ell} [1-A^{(k)}(\ell-x)]d[A^{(m+1-\overline{s+1}\,k)}(x)]$$

$$= A^{(m+1-\overline{s+1}\,k)}(\ell)-A^{(m+1-sk)}(\ell), \quad s = 1,2,..,[m/k]-1,$$

and similarly

$$U_0^{(m+1)} = 1 - A^{(m+1-[m/k]k)}(\ell),$$

$$U_{[m/k]}^{(m+1)} = A^{(m+1-k)}(\ell),$$

which are the same results as those obtained by Finch.

If we denote by $\xi(t)$ and $U(z,t)$ the random variable giving the unconditional number of outstanding orders at time t and its generating function, then, reasoning as in section 2, we find that

$$U(z,t) = \sum_{m=o}^{\infty} \prod_{s=1}^{[m/k]} [1-L(t-\theta_{m,m}+\theta_{m,m-sk})+z\{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\}] \times$$

$$dH_{m,k,o}(\theta), \quad |z| \leq 1,$$

$$E[\xi(t)] = \sum_{m=o}^{\infty} \int \prod_{s=1}^{[m/k]} \{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\} \; dH_{m,k,o}(\theta),$$

$$Var[\xi(t)] = \sum_{m=o}^{\infty} \int \sum_{s=1}^{[m/k]} \{1-L(t-\theta_{m,m}+\theta_{m,m-sk})\} \; L(t-\theta_{m,m}+\theta_{m,m-sk})$$

$$dH_{m,k,o}(\theta) .$$

* * * * * * * * * * * * *

CHAPTER SEVEN

Queues with moving average service times.

1. Introduction.

In the preceding chapters, we observed that the usual assumption that successive inter-arrival intervals for a queue are identically and independently distributed is very restrictive, in that, where the input is provided by another process, one might well expect some time dependence on the history of that process. We have followed up the consequences of postulating a simple model incorporating such a time dependence, namely, an input whose inter-arrival intervals form a moving average of an I.I.D. sequence of non-negative random variables.

In this chapter we give our attention to the possibility of moving average service times, i.e., service times which are moving averages of an I.I.D. sequence of non-negative random variables.

A situation for which moving average services might provide a better approximation to reality than the general recurrent service is where the server is a human operator. The operator's efficiency could well be expected to change only gradually with time, and this would be reflected in a positive correlation between the times taken for him to handle consecutive and similar tasks, whether handled immediately the one after the other or with a break between.

We deal with a single server only, since the interpretation for a single moving average is dubious when several customers are being served simultaneously, and there are mathematical complications concerning the non-equivalence of different servers when several distinct moving averages are involved.

In fact, even with only a single server the complete solution for the limiting queue size distribution appears to be difficult, and we derive it explicitly only for moving averages of order two and three. The difficulty in the general case is perhaps not too surprising in view of the awkwardness involved in the extraction of the individual probabilities from their generating function for a moving average of order one, i.e., the standard situation of the general recurrent service time. The generating function we obtain for the second order moving average is closely related to that for general recurrent services as derived by D.G. Kendall (1951).

The limiting probability that a departing customer leaves the queue empty is identical for first and second order moving averages, and we find that this result can be conveniently obtained for the general order also.

Consider, then, a single server queue for which

(i) arrivals occur individually in a Poisson stream with parameter λ,

and (ii) the service time S_n of the nth arrival has a distribution

$$S_n = b_o U_{n+p} + b_1 U_{n+p-1} + .. + b_p U_n, \quad n \geq 0,$$

where

$$\Sigma b_i = 1,$$

and $\{U_n\}$ is an I.I.D. sequence of non-negative random variables with distribution function $U(\cdot)$. As in chapters two and four, the b_i need not be non-negative, although both they and $U(.)$ are strongly constrained if they are not. Also, the working of sections five and six in which regenerative points are used will need to be modified in this case, as mentioned in Chapter Five.

We shall adopt the intuitive condition

$$\lambda^{-1} > \int_0^\infty u dU(u)$$

for the existence of a unique limiting distribution of queue length. The sufficiency of this condition follows from Loynes (1962).

The basic notation is that of Chapter Two, except that here $P_j(u^{(n+p-1)})$

represents the probability that, in the particular realization of the process in which $(U^{(n+p-1)})$ takes on the value $(u^{(n+p-1)})$, the nth customer on completing his *service* leaves j customers in the queue.

We define the associated generating function $P(u^{(n+p-1)};z)$ by

$$P(u^{(n+p-1)};z) = \sum_{i=o}^{\infty} P_i(u^{(n+p-1)})z^i, \quad |z| \le 1,$$

and its integral transform $P^*(s^{(p)};z;n)$ by

$$P^*(s^{(p)};z;n) = E[P(U^{(n+p-1)};z)\exp(-s_p U_{n+p-1} - s_{p-1} U_{n+p-2} - \ldots - s_1 U_n)],$$

$$|z| \le 1, \quad \text{Re. } s_i \ge 0, \; 1 \le i \le p.$$

Associated with the limiting distribution we have

$$P(w_1,\ldots,w_p;z) = \lim_{n\to\infty} E[P(U_o,\ldots,U_{n-1},u_n,u_{n+1},\ldots,u_{n+p-1};z)],$$

where $w_1,\ldots,w_p$, are the particular values assumed by $u_n,\ldots,u_{n+p-1}$, and the corresponding integral transform

$$P^*(s^{(p)};z) = E[P(W^{(p)};z) \exp(-s_p W_p - \ldots - s_1 W_1)], \quad |z| \le 1, \; \text{Re.} s_i \ge 0,$$

where the W_i are identically and independently distributed with common distribution function $U(\cdot)$.

2. *The limiting probability that a departing customer leave the queue empty.*

Since the input stream is Poisson, the queue length distributions at successive departure instants in any particular realization of the process are related by the particularly simple relations

$$\begin{aligned} P_j(u^{(n+p)}) = {} & P_{j+1}(u^{(n+p-1)}) \exp[-\lambda S_{n+1}] \\ & + P_j(u^{(n+p-1)}) \exp[-\lambda S_{n+1}](\lambda S_{n+1}) \\ & + \ldots + P_2(u^{(n+p-1)}) \exp[-\lambda S_{n+1}](\lambda S_{n+1})^{j-1}/(j-1)! \\ & + [P_1(u^{(n+p-1)}) + P_o(u^{(n+p-1)})] \exp[-\lambda S_{n+1}](\lambda S_{n+1})^j/j! \end{aligned}$$

$$n \ge 1, \; j \ge 0.$$

These relations can be collectively represented in terms of generating functions as

$$zP(u^{(n+p)};z) = P(u^{(n+p-1)};z)\exp[-\lambda S_{n+1}(1-z)]$$
$$- (1-z)P_o(u^{(n+p-1)})\exp[-\lambda S_{n+1}(1-z)], \quad n \geq 1, \quad |z| \leq 1.$$

If we take integral transforms and let $n\to\infty$ we find that

$$zP^*(s^{(p)};z) = [P^*\{(1-z)b_1+s_{p-1},..,(1-z)\lambda b_{p-1}+s_1,(1-z)\lambda b_p;z\}$$
$$- (1-z)P_o^*\{(1-z)\lambda b_1+s_{p-1},..,(1-z)\lambda b_{p-1}+s_1,(1-z)\lambda b_p\}]$$
$$\psi\{(1-z)b_o+s_p/\lambda\}, \qquad |z| \leq 1, \text{ Re. } s_i \geq 0.$$

By a recursive substituting of the arguments of P^* on the right hand side of this equation into the left hand side we derive

$$z^pP^*(s^{(p)};z) = \psi\{(1-z)b_o+s_p/\lambda\}\psi\{(1-z)(b_o+b_1)+s_{p-1}/\lambda\}$$
$$..\psi\{(1-z)(b_o+..+b_{p-1})+s_1/\lambda\} \times$$
$$P^*\{(1-z)\lambda(b_1+..+b_p),(1-z)\lambda(b_2+..+b_p),..,(1-z)\lambda b_p;z\}$$
$$-z^{p-1}(1-z)\psi\{(1-z)b_o+s_p/\lambda\}P_o^*\{(1-z)\lambda b_1+s_{p-1},..,(1-z)\lambda b_p\}$$
$$- \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$$
$$- (1-z)\psi\{(1-z)b_o+s_p/\lambda\}\psi\{(1-z)(b_o+b_1)+s_{p-1}/\lambda\}$$
$$\times..\times\ \psi\{(1-z)(b_o+..+b_{p-1})+s_1/\lambda \ \times$$
$$P_o^*\ (1-z)\lambda(b_1+..+b_p),..,(1-z)\lambda b_p\}, \qquad |z| \leq 1, \quad \text{Re.}s_i \geq 0.$$

The P^* generating function on the right hand side can now be eliminated through the substitutions

$$s_1 = (1-z)\lambda b_p,\ s_2 = (1-z)\lambda(b_{p-1}+b_p),..,\ s_p = (1-z)\lambda(b_1+..+b_p),$$

giving

$$(2.1) \quad z^pP^*(s^{(p)};z) = -[z-\psi(1-z)]^{-1}P_o^*\{(1-z)\lambda(b_1+..+b_p),..,(1-z)\lambda b_p\} \times$$
$$(1-z)\psi\{1-z\}\psi\{(1-z)b_o+s_p/\lambda\}..\psi\{(1-z)(b_o+..+b_{p-1})+s_1/\lambda\}$$
$$-z^{p-1}(1-z)\psi\{(1-z)b_o+s_p/\lambda\}P_o^*\ (1-z)\lambda b_1+s_{p-1},..,(1-z)\lambda b_p\}$$
$$- \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$$
$$-(1-z)\psi\{(1-z)b_o+s_p/\lambda\}..\psi\{(1-z)(b_o+..+b_{p-1})+s_1/\lambda\} \times$$
$$P_o^*\{(1-z)\lambda(b_1+..+b_p),..,(1-z)\lambda b_p\}, \quad |z| \leq 1, \text{ Re. } s_i \geq 0.$$

(We note that by virtue of the condition

$$\lambda^{-1} > \int_0^\infty u\,dU(u),$$

Rouché's theorem ensures that $z-\psi(1-z)$ has no zeros inside the unit circle.)

Since

$$P(u^{(n+p)};1) = 1,$$

it follows that

$$P^*(s^{(p)};1) = E[\prod_{i=1}^{p} \exp(-s_i W_i)]$$

$$= \prod_{i=1}^{p} \psi(s_i/\mu), \qquad \text{Re. } s_i \geq 0.$$

Using this result and L'Hôpital's rule, it follows immediately from letting $z \to 1$ in (2.1) that

$$P_o^*(0,0,..,0) = 1 - [\frac{d}{dz}\psi(1-z)]_{z-1}$$

$$= 1 - \lambda \int_0^\infty u\,dU(u).$$

As $P_o^*(0,..,0)$ is simply the limiting probability P_o that a departure leaves the queue empty, we have

$$P_o = 1 - \lambda \int_0^\infty u\,dU(u).$$

This result was obtained by D. G. Kendall (1951) for a general recurrent service time distribution.

3. *Limiting distribution for a moving average of order two.*

For $p = 1$, (2.1) becomes

$$(3.1) \quad zP^*(s_1;z) = -[z-\psi(1-z)]^{-1}P_o^*\{(1-z)\lambda b_1\}(1-z)\psi\{1-z\}\psi\{(1-z)b_o+s_1/\lambda\}$$

$$- (1-z)\psi\{(1-z)b_o+s_1/\lambda\}P_o^*\{(1-z)\lambda b_1\}$$

$$= -z(1-z)\psi\{(1-z)b_o+s_1/\lambda\}P_o^*\{(1-z)\lambda b_1\}[z-\psi(1-z)]^{-1},$$

$$|z| \leq 1, \text{ Re. } s_i \geq 0.$$

By dividing by z and letting $z \to 0$ in this equation we obtain

$$P_o^*(s_1) = \psi(b_o+s_1/\lambda)P_o^*(\lambda b_1)[\Psi(1)]^{-1}, \qquad \text{Re. } s_1 \geq 0,$$

or

$$P_o^*(s_1) = \psi(b_o+s_1/\lambda)[\psi(b_o)]^{-1}[1-\lambda \int_0^\infty udU(u)], \qquad \text{Re.} s_1 \geq 0,$$

from the result of the last section. This equation enables us to finally eliminate the unknown $P_o^*\{(1-z)\lambda b_1\}$ from (3.1).

$$P^*(s_1;z) = (1-z)\psi\{(1-z)b_o+s_1/\lambda\}\psi(1-b_1z)[\psi(b_o)]^{-1} \times$$
$$(1-\lambda \int_0^\infty udU(u))[\psi(1-z)]^{-1}, \qquad |z| \leq 1, \quad \text{Re.} s_1 \geq 0.$$

The generating function $P(z)$ of the stationary queue length distribution is found by setting $s_1 = 0$.

$$P(z) = (1-z)\psi\{(1-z)b_o\}\psi(1-b_1z)[\psi(b_o)]^{-1} \times$$
$$[1-\lambda \int_0^\infty udU(u)] [\psi(1-z)-z]^{-1}, \quad |z| \leq 1.$$

For $b_o = 1$, $b_1 = 0$, i.e., the general recurrent service time distribution, this reduces to

$$P(z) = (1-z)\psi(1-z)[1-\lambda \int_0^\infty udU(u)] [\psi(1-z)-z]^{-1}, \qquad |z| \leq 1,$$

a result obtained previously by D. G. Kendall (1951).

4. *Limiting distribution for a moving average of order three.*

Our starting point is equation (2.1), which for $p = 2$ becomes

(4.1) $$P^*(s^{(2)};z) = -z^{-2}(1-z)[\{z-\psi(1-z)\}^{-1}P_o^*\{(1-z)\lambda(b_1+b_2), (1-z)\lambda b_2\} \times$$
$$z\psi\{(1-z)b_o+s_2/\lambda\}\psi\{(1-z)(b_o+b_1)+s_1/\lambda\}$$
$$+ zP_o^*\{(1-z)\lambda b_1+s_1,(1-z)\lambda b_2\} \psi\{(1-z)b_o+s_2/\lambda\}].$$

$P^*(s^{(2)};z)$ is analytic at $z = 0$ whilst the expression on the right hand side would appear to have a pole. We shall use the known analyticity of the right hand side to furnish us with relations between values of P_o^* evaluated for different arguments. This will lead to an expression for $P_o^*(s^{(2)})$ in terms of known functions. (4.1) will give finally the limiting queue size distribution.

We obtain at once from (4.1) that as $z \to 0$,

$$zP^*(s^{(2)};z) \to \psi(b_o+s_2/\lambda)[\{\psi(1)\}^{-1}P_o^*\{\lambda(b_1+b_2),\lambda b_2\}\ \psi(b_o+b_1+s_1/\lambda) - P_o^*(\lambda b_1+s_1,\lambda b_2)].$$

Because of the continuity of $P^*(s^{(2)};z)$ at the origin, it follows that

$$P_o^*(\lambda b_1+s_1,\lambda b_2) = \{\psi(1)\}^{-1}\psi(b_o+b_1+s_1/\lambda)P_o^*\{\lambda(b_1+b_2),\lambda b_2\},$$

and thus

(4.2) $$P_o^*\{(1-z)\lambda b_1+s_1,\lambda b_2\} = \psi\{b_o+b_1(1-z)+s_1/\lambda\}[\psi\{b_o+(b_1+b_2)(1-z)\}]^{-1} \times P_o^*\{(1-z)\lambda(b_1+b_2),\ \lambda b_2\}.$$

We now introduce the notation

$$P_{o2}^*(a,b) = [\frac{\partial}{\partial s_2}\ P_o^*(s_1,s_2)]_{s_1=a,\ s_2=b}.$$

From (4.1), we find that for z small,

$$\begin{aligned}P^*(s^{(2)};z) = -z^{-1}(1-z)\psi\{(1-z)b_o+s_2/\lambda\}[-\{\psi(1)\}^{-1}\{1+z[1+\psi(1)][\psi(1)]^{-1}\} \times \\ \{P_o^*\{(1-z)\lambda(b_1+b_2),\lambda b_2\}-\lambda z b_2 P_{o2}^*\{(1-z)\lambda(b_1+b_2),\lambda b_2\}\} \times \\ \{\psi(b_o+b_1+s_1/\lambda)-(b_o+b_1)z\psi'(b_o+b_1+s_1/\lambda)\} \\ + \{P_o^*\{(1-z)\lambda(b_1+b_2),\lambda b_2\} - \lambda b_2 z P_{o2}^*\{(1-z)\lambda(b_1+b_2),\lambda b_2\} \times \\ \{\psi(1)\}^{-1}\{\psi(b_o+b_1+s_1/\lambda)-zb_1\psi'(b_o+b_1+s_1/\lambda)\} \times \\ \{1+z(b_1+b_2)\psi'(1)[\psi(1)]^{-1}\} + o(z^2)],\end{aligned}$$

where we have made use of the first two terms of Taylor expansions about the origin for the functions concerned and utilised (4.2).

Letting $z \to 0$, we find that

(4.3) $$P_o^*(s^{(2)}) = \psi(b_o+s_2/\lambda)[\psi(1)]^{-1}P_o^*\{\lambda(b_1+b_2),\lambda b_2\} \times [\{1+b_o\psi'(1)\}\{\psi(1)\}^{-1}\psi(b_o+b_1+s_1/\lambda)-b_o\psi'(b_o+b_1+s_1/\lambda].$$

The determination of the unknown constant $P_o^*\{\lambda(b_1+b_2),\lambda b_2\}$ is effected by use of (2.2), which can be expressed as

$$P_o^*(0,0) = 1 + \psi'(0).$$

We obtain

$$P_o^*(s^{(2)}) = \psi(b_o+s_2/\lambda)[\{1+b_o\psi'(1)\}\ \{\psi(1)\}^{-1}\psi(b_o+b_1+s_1/\lambda)-b_o\psi'(b_o+b_1+s_1/\lambda)]\times$$
$$[1+\psi'(0)]\ [\psi(b_o)]^{-1}[\{1+b_o\psi'(1)\}\ \{\psi(1)\}^{-1}\psi(b_o+b_1)-b_o\psi'(b_o+b_1)]^{-1}.$$

Reverting to (4.1), we find on substituting for the arguments of the P_o^* terms that

$$P^*(s^{(2)};z) = -z^{-1}(1-z)\psi\{(1-z)b_o+s_2/\lambda\}[1+\psi'(0)]\ [\psi(b_o)]^{-1}\ \times$$
$$[\{(1+b_o\psi'(1)\}\ \{\psi(1)\}^{-1}\psi(b_o+b_1)-b_o\psi'(b_o+b_1)]^{-1}\ \times$$
$$[\{z-\psi(1-z)\}^{-1}\psi\{(1-z)(b_o+b_1)+s_1/\lambda\}\psi\{b_o+(1-z)(b_1+b_2)\}$$
$$+\psi\{\ b_o+(1-z)b_1+s_1/\lambda\}].$$

The generating function $P(z) = P^*(0,0;z)$ of the limiting queue length distribution is therefore given by

$$P(z) = -z^{-1}(1-z)\psi\{(1-z)b_o\}\ [1^{+1}\psi'(1)]\ [\psi(b_o)]^{-1}$$
$$[\{1+b_o\psi^{1}(1)\}\ \{\psi(1)\}^{-1}\psi(b_o+b_1)-b_o\psi'(b_o+b_1)]^{-1}\ \times$$
$$[\{z-\psi(1-z)\}^{-1}\psi\{(1-z)(b_o+b_1)\}\psi\{b_o+(1-z)(b_1+b_2)\}$$
$$+\ \psi\{b_o\ +\ (1-z)b_1\}]\ , \qquad |z|\ \leq\ 1.$$

Beginning with (2.1), a similar procedure making use of Taylor expansions can be employed for moving averages of higher orders. The right hand side of (2.1) has an apparent pole of order $p-1$. The fact that this pole cannot occur (because of the analyticity of $P^*(s^{(p)};z)$) gives us a sequence of relations between values of P^* with different arguments involving the variables s_i. We simply consider the coefficients of like powersof z arising from $\{z-\psi(1-z)\}^{-1}$ and Taylor expansions in z. The only difficulty that can arise is in the possible non-cancelling of terms involving derivatives of P_o in the working leading to an equation corresponding to (4.3). Such derivatives can be evaluated by a recursive procedure by differentiating the $P_o^*\ (s^{(p)})$ in the left hand side of such an equation. The corresponding differentiation of the left hand side will be of known functions involving ψ or of a P_o^* term in which the corresponding s_i occupied a position further to the left in the sequence of arguments of P_o^*. A finite procedure would thus lead to the evaluation of such a differentiated P_o^* term by substitution.

Judicious utilisation of the condition obtained from each order in z can abbreviate the working considerably. We note in the example just considered that the immediate deduction (4.2) from the first order result enabled us to obviate a Taylor expansion about the first argument of P_o^*.

5. *Transient behaviour of queue length for second order moving average.*

M/G(2)/1, like its input counterpart G(2)/M/1, possesses a set of regenerative points that we can use to facilitate the investigation of transient behaviour. They are the points dividing the service times internally in the ratio $b_o:b_1$. The point in the service time of the nth customer we denote by R_n.

An arrival finding the queue empty enters immediately into service, so only arrivals occurring during a service time contribute to the probability distribution of queue size at departure instants. We therefore work in terms of an associated system consisting of a renewal process with lifetimes $\{S_n\}$, the nth lifetime ending at A_n. We again introduce points R_{n+1} intercepting the corresponding intervals (A_n, A_{n+1}) in the ratio $b_o:b_1$. As the length of (A_n, A_{n+1}) has a distribution given by

$$A_{n+1} - A_n = b_o U_{n+1} + b_1 U_n \quad , \quad n \geq 0,$$

the points $\{R_n\}$ constitute a regenerative sequence, and (R_n, R_{n+1}) has length u_{n+1}.

Departures occur at the points $\{A_n\}$, and arrivals are Poisson. Should a departure leave this associated system void, there is also an arrival at this precise instant.

It turns out that the results for Poisson arrivals can be extended to general bulk Poisson arrivals, i.e., an input where the arrival instants are Poisson but the arrivals are in batches whose size follows a general probability distribution. Suppose that the probability that an arriving batch be of size $j \geq 0$ is c_j, so that the probability of $i > 0$ arrivals to the queue in an interval of length t is

$$\sum_{\ell=o}^{i} \exp[-\lambda t](\lambda t)^{\ell}[\ell!]^{-1} \; c_i^{(\ell)} \quad , \qquad i > 0,$$

where $\{c_i^{(\ell)}\}$ is the ℓth iterated convolution of $\{c_i\}$. Then, in the associated system, should a departure leave the queue void, there is also a batch arrival at this instant with size distribution

$$\text{probability (batch size = j)} = \begin{cases} 0 & , \; j = 0, \\ c_j(1-c_o)^{-1}, & j > 0. \end{cases}$$

The corresponding expression for $i = o$ is

$$\sum_{\ell}^{\infty} \exp[-\lambda t] \; c_o^{(\ell)} = \exp[-\lambda t] \; (1-c_o)^{-1}.$$

Denote by $P(i \to j)$ the unconditional probability that the queue lengths at R_n, R_{n+1} are i, j respectively. Then

$$P(i\to j) = \begin{cases} \int_o^\infty \sum_{\ell=o}^{j+1-i} \left[\sum_{m=o}^{\ell} \exp[-\lambda b_o u](\lambda b_o u)^m (m!)^{-1} c_\ell^{(m)} \right] \times \\ \quad \left[\sum_{n=o}^{j+1-i-\ell} \exp[-\lambda b_1 u](\lambda b_1 u)^n (n!)^{-1} c_{j+1-i-\ell}^{(n)} \right] dU(u), \quad j+1 \geq i > 1, \\ \int_o^\infty \sum_{\ell=1}^{j} \left[\sum_{m=o}^{\ell} \exp[-\lambda b_o u](\lambda b_o u)^m (m!)^{-1} \; c_\ell^{(m)} \right] \times \\ \quad \left[\sum_{n=o}^{j-\ell} \exp[-\lambda b_1](\lambda b_1 u)^n (n!)^{-1} c_{j-\ell}^{(n)} \right] \\ \quad + \sum_{k=o}^{\infty} \exp\,[-\lambda b_o u] \; c_o^{(k)} \sum_{m=1}^{j} c_m (1-c_o)^{-1} \times \\ \quad \sum_{n=o}^{j-m} \exp[-\lambda b_1 u](\lambda b_1 u)^n (n!)^{-1} \; c_{j-m}^{(n)}] \; dU(u), \quad i = 1, j \geq 1, \\ 0, \; \text{otherwise} \end{cases}$$

We note that for $j + 1 \geq i > 1$, $P(i \to j)$ depends only on the difference $j - i$, and we re-express $P(i \to j)$ in terms of the constants

$$(5.1)\quad k_j = \int_0^\infty \sum_{\ell=0}^{j+1} \left[\sum_{m=0}^{\ell} \exp[-\lambda b_0 u](\lambda b_1 u)^m (m!)^{-1} c_\ell^{(m)} \right] \times$$

$$\left[\sum_{n=0}^{j+1-\ell} \exp[-\lambda b_1 u](\lambda b_1 u)^n (n!)^{-1} c_{j+1-\ell}^{(n)} \right] dU(u), \quad j \geq -1,$$

$$P_j = \int_0^\infty - \sum_{m=1}^{j} c_m (1-c_0)^{-2} \sum_{n=0}^{j} \exp[-\lambda u](\lambda b_1 u)^n (n!)^{-1} c_{j-m}^{(n)} + \sum_{n=0}^{j} \exp[-\lambda u](\lambda b_1 u)^n (n!)^{-1} c_j^{(n)} \; dU(u), \quad j \geq 1,$$

as

$$P(i \to j) = \begin{cases} k_{j-i}, & j+1 \geq i > 1, \\ k_{j-1} - P_j, & i=1,\ j \geq 1, \\ 0, & \text{otherwise} \end{cases}$$

Our aim is to find the queue length distribution at departure points in terms of known quantities. To this end we make the following definitions.

$k_{\ell-1}^{(n)}$ = coefficient of z^ℓ in the power series expansion of the nth power of the power series $\Sigma_{\ell=0}^{\infty} k_{\ell-1} z^\ell$ (absolutely convergent for $|z| \leq 1$), $n = 0,1,2,..,$

$$\alpha_{1+j}(z) = \sum_{n=1}^{\infty} k_{n+j-1}^{(n)} z^n, \qquad |z| \leq 1,\ j \geq -1,$$

$$\alpha_{-j}(z) = \sum_{n=0}^{\infty} k_{n-1}^{(n+j+1)} z^{n+j+1}, \qquad |z| \leq 1,\ j > 0,$$

$$K_j(z) = \sum_{n=0}^{\infty} z^n \left(\sum_{\ell=0}^{n+j} P_{n+1+j-\ell}\, k_{\ell-1}^{(n)} \right), \qquad j \geq 1,\ |z| \leq 1.$$

Denote by $Q_j^{(n)}$ the unconditional probability that at R_n the queue length is j. Then

$$Q_j^{(n+1)} = Q_{j+1}^{(n)} k_{-1} + Q_j^{(n)} k_0 + .. + Q_2^{(n)} k_{j-2} + Q_1^{(n)} k_{j-1} - Q_1^{(n)} P_j, \quad j \geq 1.$$

Multiplication of the equation for $Q_j^{(n+1)}$ by z^{j+1} and summation over j gives

$$z \sum_{j=1}^{\infty} Q_j^{(n+1)} z^j = \Big(\sum_{j=1}^{\infty} Q_j^{(n)} z^j\Big)\Big(\sum_{\ell=o}^{\infty} k_{\ell-1} z^{\ell}\Big) - zQ_1^{(n)} \sum_{j=1}^{\infty} p_j z^j, \quad n \geq 1, \quad |z| \leq 1,$$

so that by iteration,

$$(5.2) \quad z^{n+1} \sum_{j=1}^{\infty} Q_j^{(n+1)} z^j = z\Big(\sum_{j=1}^{\infty} Q_j^{(1)} z^j\Big)\Big(\sum_{\ell=o}^{\infty} k_{\ell-1}^{(n)} z^{\ell}\Big) - \sum_{j=1}^{\infty} p_j z^j \sum_{m=1}^{n} Q_1^{(m)} z^m \Big(\sum_{\ell=o}^{\infty} k_{\ell-1}^{(n-m)} z^{\ell}\Big),$$

$$n \geq 1, \quad |z| \leq 1.$$

Equating coefficients of z^{n+1+j} in (5.2):

$$Q_j^{(n+1)} = \sum_{m=1}^{n+j} Q_m^{(1)} k_{n+j-m-1}^{(n)} - \sum_{m=1}^{n} Q_1^{(m)} \sum_{\ell=o}^{n+j-m} P_{n+1+j-m-\ell} k_{\ell-1}^{(n-m)},$$

$$j \geq 1, \; n \geq 1.$$

The generating function of $Q_j^{(n)}$ on n can now be formed:

$$\sum_{n=1}^{\infty} Q_j^{(n)} z^{n-1} = Q_j^{(1)} + \sum_{n=1}^{\infty} Q_n^{(1)} \alpha_{j-n+1}(z) - z \sum_{n=1}^{\infty} Q_1^{(n)} z^{n-1} K_j(z), \quad j \geq 1, \quad |z| < 1,$$

or, writing

$$Q_j(z) = \sum_{n=1}^{\infty} Q_j^{(n)} z^{n-1}, \qquad j \geq 1, \quad |z| < 1,$$

$$(5.3) \quad Q_j(z) = Q_j^{(1)} + \sum_{n=1}^{\infty} Q_n^{(1)} \alpha_{j-n+1}(z) - zK_j(z)Q_1(z), \qquad j \geq 1, \quad |z| < 1.$$

Substituting $j = 1$ gives

$$(5.4) \quad Q_1(z) = \Big[Q_1^{(1)} + \sum_{n=1}^{\infty} Q_n^{(1)} \alpha_{2-n}(z)\Big]\big[1+zK_1(z)\big]^{-1}, \qquad |z| < 1.$$

(5.3), (5.4) together implicitly express the unconditional queue length at each R_n in terms of known functions and the queue length distribution at R_1. The probability $Q_j^{(n)}$ that the queue length is j at R_n is thus

~~$Q_j^{(n+1)} = Q_{j+1}^{(n)} k_{-1} + Q_j^{(n)} k_o + \ldots + Q_2^{(n)} k_{j-2}$~~

$$(5.5)\quad Q_j^{(n)} = \frac{1}{2\pi i} \oint z^{-n}[Q_j^{(1)} + \sum_{m=1}^{\infty} Q_m^{(1)}{}_{j-m+1}(z) -zK_j(z)\{Q_1^{(1)} + \sum_{m=1}^{\infty} Q_m^{(1)} \alpha_{2-m}(z)\}\{1+zK_1(z)\}^{-1}]dz,$$

where the integration is carried out around a small loop surrounding the origin. Since the functions concerned all have known power series expansions about the origin, the value of $Q_j^{(n)}$ can be written down from (5.3), (5.4) just by picking out the coefficients of z^{n-1}. However, the expression obtained involves clumsy infinite series, and we do not write it down explicitly.

The queue length as left by the departure at A_n, i.e., the departure at the completion of the service associated with R_n, is therefore known in terms of the queue length distribution at R_1, since

$$(5.6)\quad \text{prob (queue length} = j \text{ at } A_n)$$

$$= \sum_{i=1}^{j} Q_i^{(n)} \int_0^{\infty} \exp[-\lambda b_o u] \sum_{k=o}^{j+1-i} (\lambda b_o u)^{j+1-i}[(j+1-i)!]^{-1} c_{j+1-i}^{(k)} \, dU(u)$$

$$+Q_{j+1}^{(n)} \int_0^{\infty} \exp[-\lambda b_o u](1-c_o)^{-1} \, dU(u), \quad j \geq 0,\ n \geq 1.$$

The procedure of determining the queue length as left by a departing customer is completed by finding the distribution $\{ Q_j^{(1)} \}$. Since the first customer does not have to wait for service, the queue extant at R_1 consists simply of those arrivals which occurred during the first fraction b_o of the first service plus the first customer himself, whence

$$(5.7)\quad Q_j^{(1)} = \begin{cases} \int_0^{\infty} \exp[-\lambda b_1 u](1-c_o)^{-1} dU(u), & j = 1, \\ \int_0^{\infty} \sum_{\ell=o}^{j-1} \exp[-\lambda b_1 u](\lambda b_1 u)^{\ell}(\ell!)^{-1} c_{j-1}^{(\ell)} \, dU(u), & j > 1. \end{cases}$$

A complete knowledge of the transient behaviour of queue length at departure points, as is provided jointly by (5.5), (5.6), (5.7), is, as one would expect, sufficient for determining the limiting queue length on departures, when it exists. In principle such a determination could be made through the use of Abel's theorem. If we denote by $P_j^{(n)}$ the probability that the queue length is j at A_n, then

(5.8) $$P_j = \lim_{n\to\infty} P_j^{(n)} = \lim_{z\to 1} \{(1-z)P_j(z)\}.$$

In practice, however, the calculations are quite difficult and Finch (1959) in a study of M/G/1, i.e., a queue with simple Poisson arrivals and uncorrelated service times, was unable to obtain the complete limiting solution through use of (5.8), although P_j was found for $j \geq 1$ in terms of P_o.

With only heavier algebra it is possible to generalise the above working to bulk services, of size k, say. Two natural service mechanisms are

(i) If the queue length at some epoch is less than the bulk the server can handle, the latter waits until further arrivals make up the deficit before he commences serving.

(ii) The server operates if there is even one customer available.

A third natural service mechanism, servicing even with an empty queue (a public bus system provides an instance of this type), we have excluded already because of its removing the correlation between successive 'genuine' services.

(i) and (ii) clearly have the same distribution of queue size at departure instants, since only arrivals during services contribute to the queue size left on a departure.

In the working above, an extra arrival was superposed on the Poisson stream in the associated system for points $\{A_n\}$ at which the departure left the queue void. For (ii) we similarly superpose k arrivals whenever the completion of a service leaves fewer than k individuals in the queue. The reason for the artificial superposing of the k arrivals is, of course, just to cause the number of departures at each A_{n+1} to depend only on the queue length at that instant and not also on that at the previous departure point A_n.

In (i) we superpose sufficient additional arrivals to ensure that the queue length in the associated system just after a departure point is always at least k. The end results for queue lengths after departures always correspond to those in the original system, but only in (i) do the queue lengths at points of the sequence $\{R_n\}$ correspond to the queue size in the original system the appropriate fraction of the way through a service.

In the case of general independent service times, it is possible to deal with a general bulk service simultaneously (see U. Narayan Bhat (1964), but the device we have employed to remove dependence between behaviour in successive intervals (R_n, R_{n+1}), (R_{n+1}, R_{n+2}) does not suffice for such a general situation. The problem does not, of course, arise with independent service times because the departure instants then coincide with the end points of such intervals.

E. Sparre Andersen((1953), (1954)), has made use of combinatorial techniques in the study of sums of random variables. Subsequent simplifications by Spitzer (1956) and Feller (1959), have made possible investigations into very general queueing systems, and as noted Narayan Bhat (1964) has been able to obtain information on transition probabilities and the busy period in the case of independent services even for general batch Poisson arrivals and general bulk services. The difficulties of handling the additional customers superposed on the Poisson stream and non-independence between successive service periods make a similar treatment impracticable here even when we restrict ourselves to individual arrivals and single servicings.

In the next section we shall obtain information about the busy period by the same methods as we have used to investigate transient queue lengths.

6. *The busy period.*

We consider the busy periods for the same queueing system as we dealt with in the last section, with general batch Poisson arrivals and individual services. We again make use of the associated system.

The algebra involved is now simpler, because there must be no arrivals superposed on the Poisson stream for any interval (R_n, R_{n+1}) included in the busy period, as this would imply that the queue became empty in the corresponding interval in the original system. The queue length must be maintained strictly positive without the support of such arrivals.

We investigate the probabilities of transitions between points R_n, R_{n+1} when such arrivals are excluded and the queue never becomes empty.

Denote by P_{ij} the probability that the queue lengths at R_n, R_{n+1} are i, j respectively, and that the queue does not become empty at A_n. Then

$$P_{ij} = \begin{cases} \int_0^\infty \sum_{\ell=o}^{j+1-i} \left[\sum_{m=o}^{\ell} \exp[-\lambda b_o u](\lambda b_o u)^m (m!)^{-1} c_\ell^{(m)}\right] \times \\ \left[\sum_{n=o}^{j+1-i-\ell} \exp[-\lambda b_1 u](\lambda b_1 u)^n (n!)^{-1} c_{j+1-i-\ell}^{(n)}\right] dU(u), & j+1 \geq i > 1, \\ \int_0^\infty \sum_{\ell=1}^{j} \left[\sum_{m=o}^{\ell} \exp[-\lambda b_o u](\lambda b_o)^m (m!)^{-1} c_\ell^{(m)}\right] \times \\ \left[\sum_{n=o}^{j-\ell} \exp[-\lambda b_1 u](\lambda b_1 u)^n (n!)^{-1} c_{j-\ell}^{(n)}\right] dU(u), & i = 1, j \geq 1, \\ 0, & \text{otherwise}, \end{cases}$$

or, more conveniently,

$$P_{ij} = \begin{cases} k_{j-i}, & j+1 \geq i > 1 \\ k_{j-1} - r_j, & i = 1, j \geq 1 \\ o, & \text{otherwise}, \end{cases}$$

where the k's are as defined by (5.1) and

$$r_j = \int_o^\infty \exp[-\lambda u] \sum_{n=o}^{j} (\lambda b_1 u)^n (n!)^{-1} c_j^{(n)} dU(u), \qquad j \geq 1.$$

Analogously to $Q_j^{(n)}$, we write $R_j^{(n)}$ to denote the unconditional probability that at R_n the queue length is j. We have

$$R_j^{(n+1)} = R_{j+1}^{(n)} \; k_{-1} + R_1^{(n)} k_{j-1} - k_1^{(n)} \; r_j, \qquad j \geq 1,\; n \geq 1.$$

The working clearly follows that of section 5 closely and we omit the detail. We derive

$$R_j(z) = R_j^{(1)} + \sum_{n=1}^{\infty} R_n^{(1)} \alpha_{j-n+1}(z) - zL_j(z) R_1(z), \qquad j \geq 1, \quad |z| < 1,$$

$$R_1(z) = [R_1^{(1)} + \sum_{n=1}^{\infty} R_n^{(1)} \alpha_{2-n}(z)][1+zL_1(z)]^{-1}, \qquad |z| < 1,$$

where

$$R_j(z) = \sum_{n=1}^{\infty} R_j^{(n)} z^{n-1}, \qquad j \geq 1, \quad |z| < 1,$$

$$L_j(z) = \sum_{n=o}^{\infty} z^n \left(\sum_{\ell=o}^{n+j} r_{n+1+j-\ell} \, k_{\ell-1}^{(n)} \right), \quad j \geq 1, \quad |z| \leq 1.$$

As for $Q_j^{(n)}$, $R_j^{(n)}$ can be written down explicitly by the picking out of coefficients, or compactly expressed as a contour integral

$$(6.1) \quad R_j^{(n)} = \frac{1}{2\pi i} \oint z^{-n} [R_j^{(1)} + \sum_{m=1}^{\infty} R_m^{(1)} \alpha_{j-m+1}(z)$$

$$- zL_j(z) \{R_1^{(1)} + \sum_{m=1}^{\infty} R_m^{(1)} \alpha_{2-m}(z)\} \{1+zL_1(z)\}^{-1}] dz,$$

where the integration is performed on a suitably small loop round the origin.

By suitable labelling, a busy period consisting of n services occurs when an arrival occurs at A_o to find the queue empty; the queue then remains full until A_n, when a departure leaves it empty. As departures occur only at the points $\{A_j\}$, when the queue length at A_n, the last point of the sequence $\{R_j\}$ before A_n must be unity. (6.1) with $j = 1$ gives (in terms of the queue length distribution at R_1) the probability that the queue has length 1 at R_n after a busy period extending from R_1, the first point of $\{R_j\}$ after A_o.

Since the arrival at A_o finds the queue empty, the distribution $\{R_m^{(1)}\}$ is given by

$$(6.2)\quad R_m^{(1)} = \begin{cases} \int_0^\infty \exp[-\lambda b_1 u](1-c_o)^{-1} dU(u), & m = 1, \\ \int_o^\infty \sum_{\ell=o}^{j-1} \exp[-\lambda b_1 u](\lambda b_1 u)^\ell (\ell!)^{-1} c_{j-1}^{(\ell)} \, dU(u), & m > 1. \end{cases}$$

For a queue length of unity at R_n, the probability that the departure at A_n leaves the queue empty is

$$\int_o^\infty \exp[-\ b_o u](1-c_o)^{-1}\ dU(u).$$

The unconditional probability of a busy period of exactly n customers is therefore

$$\int_o^\infty \exp[-\lambda b_o u](1-c_o)^{-1} dU(u) \quad \frac{1}{2\pi i} \oint z^{-n}[R_1^{(1)} + \sum_{n=1}^{\infty} \alpha_{2-n}(z)][1+zL_1(z)]^{-1} dz,$$

where the integration is around a suitably small loop enclosing the origin and the $R_n^{(1)}$ are given by (6.2).

As in section 5, the working can readily be extended to bulk services of fixed size.

* * * * * * * *

CHAPTER EIGHT.

PART I.

Queueing systems with transport service processes

We deal briefly with queueing systems in which the service facility operates regardless of whether or not customers are present. An instance of such a queueing system is provided by a bus service which operates even if there are no passengers available. A consequence of such a service mechanism is that a customer arriving at an empty queue will not in general be able to commence service immediately.

This service mechanism was considered by Bailey in 1954, who dealt with the equilibrium queue length behaviour in a system with Poisson arrivals and bulk service with a general recurrent service time distribution. The waiting time of this system was investigated by Downton in 1955. Downton has made a further study (1956) on the limiting behaviour of this system with increasing size of service capacity. Keilson (1962a) has given a very general treatment of this system, with both arrivals and services being in batches whose sizes have general probability distributions.

A comparable problem has been treated by Finch (1959) and extended by Ewens and Finch (1962). A queueing system is considered in which if the nth arrival finds the server idle he does not commence service until a time v_n after arriving. $\{v_n\}$ is a sequence of I.I.D. variables. Finch deals with the waiting time distribution for arbitrary I.I.D. service and inter-arrival time distributions and generalises a result of Pollaczek in the case of Poisson inputs. Ewens and Finch extend Finch's results to Erlang inputs.

We shall consider the equilibrium distributions produced by negative exponential services and general moving average inputs.

An arrival finding no other customer waiting for service must wait until the end of the current service before his own service can commence. This is true whether or not the service facility is actually occupied with a customer. We thus find it convenient to take the queue length as the number of customers waiting rather than the number waiting or being served. In view of the absence of true idle periods for the server we need not concern ourselves with whether or not the service facility is occupied at a given moment. This enables us to work with a structurally simpler system involving "lumped" states. We shall then split the lumped states and regain full information on queue lengths.

1. *Negative exponential services.*

Consider first the case of a general recurrent input. The characteristic lack of memory property of the negative exponential distribution enables us to handle the imbedded Markov chain formed on arrival instants with great ease.

We denote by $P_j^{(n)}$ the probability that the nth arrival finds j customers waiting in the queue. Then a consideration of the changes possible in the queue between the nth and (n+1)th arrivals instants gives

$$(1.1) \qquad P_j^{(n+1)} = \begin{cases} \sum_{i=o}^{\infty} P_i^{(n)} K_{i+1}, & j = 0, \\ \sum_{i=o}^{\infty} P_{j+i-1}^{(n)} k_i, & j \geq 1, \end{cases}$$

where k_i is the probability that i services terminate between the nth and (n+1)th arrival instants and

$$K_i = \sum_{j=i}^{\infty} k_j , \quad i \geq 1.$$

(1.1) is the same expression as arises for the queueing system ordinary services and $P_j^{(n)}$ denotes the probability that the nth arrival finds j customers already in the queue, including the customer, if any, actually being served. This equivalence can, in fact, be shown to be good for the complete time dependent behaviour of the system. In particular, the limiting distribution of the number of customers waiting as found by an arrival will be the geometric distribution

$$\{(1-T)T^j, \quad j \geq 0\}$$

where T is the unique root inside the unit circle of

$$T = \psi(1-T),$$

and the symbols have their customary meanings.

It is similarly shown that the usual delayed geometric distribution arises from a moving average input.

In the event of a very long service period, it is possible that many arrivals can occur and that the queue waiting can become quite long, even though the server is not dealing with a customer. If we wish to know the queue length probabilities including a possible customer in service or the probability that the server is or is not occupied when an arrival finds a given number of customers are waiting, we can find these through use of our known forms of limiting distribution.

We consider a $(p+1)$th order general moving average and without further comment make use of the notation introduced earlier.

We define $Q_j(u^{(n+p-1)})$ analogously to $P_j(u^{(n+p-1)})$, corresponding to the nth arrival finding j customers waiting and a further customer in service. Corresponding to j waiting customers but no customer in service is the probability

$$P_j(u^{(n+p-1)}) - Q_j(u^{(n+p-1)}).$$

We can write down equations for the Q's similar to (1.1)

$$Q_j(u^{(n+p)}) = \begin{cases} \sum_{i=o}^{\infty} P_i(u^{(n+p-1)})\, k_{i+1}(u_n,..,u_{n+p}), & j = 0, \\ \sum_{i=o}^{\infty} P_{j+i-1}(u^{(n+p-1)})\; k_i(u_n,..,u_{n+p}) \\ -\,[P_{j-1}(u^{(n+p-1)})-Q_{j-1}(u^{(n+p-1)})]\; k_o(u_n,..,u_{n+p}), & j \geq 1, \end{cases}$$

that is,

(1.2) $$Q_j(u^{(n+p)}) = \begin{cases} \sum_{i=o}^{\infty} P_i(u^{(n+p-1)})k_{i+1}(u_n,..,u_{n+p}), & j = 0, \\ P_j(u^{(n+p)}) - [P_{j-1}(u^{(n+p-1)})-Q_{j-1}(u^{(n+p-1)})]k_o(u_n,..,u_{n+p}), & j \geq 1. \end{cases}$$

Taking Laplace-Stieltjes transforms and letting $n \to \infty$ provides

(1.3) $$Q_j^*(s^{(p)}) = \begin{cases} \sum_{i=1}^{\infty} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [P_{i-1}^*(\sigma b_1+s_{p-1},..,\sigma b_{p-1}+s_1,\sigma b_p) \\ \quad \times\, \psi\{(\sigma b_o+s_p)/\mu\}]_{\sigma=\mu}, & j = 0, \\ P_j^*(s^{(p)}) \\ -[P_{j-1}^*(\mu b_1+s_{p-1},..,\mu b_p)-Q_{j-1}^*(\mu b_1+s_{p-1},..,\mu b_p)] \\ \quad \times\, \psi(b_o+s_p/\mu), & j \geq 1, \\ \text{for Re. } s_i \geq 0. \end{cases}$$

The expression for Q_o^* can be simplified by making use of the relation

$$P^*(s^{(p)};z) = B_{p-1}(s^{(p)})z^{p-1}+..+B_o(s^{(p)})+B(s^{(p)})(1-zT)^{-1}, \quad \text{Re. } s_i \geq 0.$$

We find that

$$Q_o^*(s^{(p)}) = \sum_{i=1}^{p} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [B_{i-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]_{\sigma=\mu}$$

$$+ \sum_{i=1}^{\infty} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [B(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}T^{i-1}]_{\sigma=\mu}$$

$$= \sum_{i=1}^{p} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [B_{i-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]_{\sigma=\mu}$$

$$+ T^{-1}[B\{\mu(1-T)b_1+s_{p-1},..,\mu(1-T)b_p\}\psi\{(1-T)b_o+s_p/\mu\}$$
$$-B(\mu b_1+s_{p-1},..,\mu b_p)\psi(b_o+s_p/\mu)]$$

$$= \sum_{i=1}^{p} \frac{(-\mu)^i}{i!} \frac{\partial^i}{\partial\sigma^i} [B_{i-1}(\sigma b_1+s_{p-1},..,\sigma b_p)\psi\{(\sigma b_o+s_p)/\mu\}]_{\sigma=\mu}$$

$$+ B(s^{(p)}) - B(\mu T b_o+s_p,\mu T b_1+s_{p-1},..,\mu T b_{p-1}+s_1).$$

$Q_o^*(s^{(p)})$ is thus determined since B is known and $B_o,..,B_{p-1}$ can be found in a finite number of steps. The (unconditional) equilibrium value Q_o of the probability that an arrival finds the system empty apart from a single customer in service is then given by

$$Q_o = Q_o^*(0,0,..,0).$$

The other unconditional values

$$Q_j = Q_j^*(0,0,..,0)$$

can now be found recursively from (1.3) using the known form of P. If we define

$$R_j(s^{(p)}) = P_j^*(s^{(p)}) - Q_j^*(s^{(p)}), \quad \text{Re. } s_i \geq 0, \quad j \geq 0,$$

then the second part of (1.2) can be expressed as

$$R_j(s^{(p)}) = \psi(b_o+s_p/\mu)R_{j-1}(\mu b_1+s_{p-1},..,\mu b_p), \qquad j \geq 1.$$

This set of relations admits of a neat treatment by use of generating functions. With an obvious notation

$$R(s^{(p)};z) = R_o(s^{(p)}) + z\psi(b_o+s_p/\mu)R(\mu b_1+s_{p-1},..,\mu b_p;z) \qquad |z| \leq 1,$$

a form to which we can apply our standard recursive sequence of substitutions in the arguments of R.

$$
\begin{aligned}
R(s^{(p)};z) = {} & R_o(s^{(p)}) + z\psi(b_o+s_p/\mu)R_o(\mu b_1+s_{p-1},\ldots,\mu b_p) \\
& + z^2\psi(b_o+s_p/\mu)\psi(b_o+b_1+s_{p-2}/\mu)R_o\{\mu(b_1+b_2)+s_{p-2},\ldots,\mu b_p\} \\
& + \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
& + z^{p-1}\psi(b_o+s_p/\mu)\ldots\psi(b_o+\ldots+b_{p-2}+s_2/\mu) \quad \times \\
& \qquad\qquad R_o\{\mu(b_1+\ldots+b_{p-1})+s_1,\ldots,\mu b_p\} \\
& + z^p\psi(b_o+s_p/\mu)\ldots\psi(b_o+\ldots+b_{p-1}+s_1/\mu) \quad \times \\
& \qquad\qquad R_o\{\mu(b_1+\ldots+b_p),\ldots,\mu b_p\}\ [1-z\psi(1)]^{-1}.
\end{aligned}
$$

We have, therefore, the simple results

$$
(1.4) \qquad Q_j = \begin{cases} P_j-\psi(b_o)\ldots\psi(b_o+\ldots+b_{j-1})R_o\{\mu(b_1+\ldots+b_j),\ldots,\mu b_p\}, & 1 \le j \le p-1, \\ \\ P_j-\psi(b_o)\ldots\psi(b_o+\ldots+b_{p-1})R_o\{\mu(b_1+\ldots+b_p),\ldots,\mu b_p\}[\psi(1)]^{j-p}, & j \ge p. \end{cases}
$$

$\{Q_j\}$ is thus the difference of two delayed geometric distributions, the two ratios being $T = \psi(1-T)$, $\psi(1)$.

(1.4) enables us to give the equilibrium queue length distribution in the normal sense, i.e., where we include in the queue length the customer, if any, in service at an arrival instant. If we denote this distribution by $\{p_j\}$ then, as the probabilities of a queue length j with or without a customer in service are Q_{j-1}, $P_j - Q_j$ respectively, $j \ge 1$,

$$
p_j = \begin{cases} R_o(0,0,\ldots,0), & j = 0 \\ P_o-R_o(0,0,\ldots,0) + \psi(b_o)R_o(\mu b_1,\ldots,\mu b_p), & j = 1, \\ P_{j-1}-\psi(b_o)\ldots\psi(b_o+\ldots+b_{j-2})[R_o\{\mu(b_1+\ldots+b_{j-1}),\ldots,\mu b_p\} & \\ \qquad -\psi(b_o+\ldots+b_{j-1})R_o\{\mu(b_1+\ldots+b_j),\ldots,\mu b_p\}], & 2 \le j \le p, \end{cases}
$$

$$P_{j-1}-\psi(b_o)\ldots\psi(b_o+\ldots+b_{p-1})R_o\{\mu(b_1+\ldots+b_p),\ldots,\mu b_p\} \quad \times$$

$$[\psi(1)]^{j-1}[\psi(1)-1], \qquad j\geq p+1.$$

(1.4) does not include the two simplest moving averages, of orders one and two.

The moving average of order one requires no supplementary variables s_i, and the relations corresponding to (1.3) are obtained directly from (1.2) by integration and letting $n \to \infty$:

$$Q_j = \begin{cases} \sum_{i=o}^{\infty} P_i\psi_{i+1}, & j = 0, \\ P_j - (P_{j-1} - Q_{j-1})\psi(1), & j \geq 1, \end{cases}$$

where

$$\psi_i = \int_o^{\infty} (\mu x)^i (i!)^{-1} \exp(-\mu x)\, dU(x), \qquad i \geq 0.$$

By virtue of the purely geometric form $\{(1-T)T^j\}$ of $\{P_j\}$, Q_o can be written as

$$Q_o = (1-T)T^{-1}[\psi(1-T)-\psi(1)]$$
$$= (1-T)[1-T^{-1}\psi(1)].$$

It follows readily as above that

$$Q_j = P_j - [\psi(1)]^j R_o, \qquad j \geq 1.$$

Thus

$$P_j = T^{-1}(1-T)(T^j - [\psi(1)]^j[1-\psi(1)]), \qquad j \geq 0.$$

The explicit solution for $p = 1$ is obtained equally readily. We have

$$R_o(s) = P_o^*(s) - Q_o^*(s)$$
$$+ B(s) + B_o(s) - \{ -\mu\frac{\partial}{\partial\sigma} [B_o(\sigma b_1)\Psi\{(\sigma b_o+s)/\mu\}]_{\sigma=\mu}$$
$$+ B(s) - B(\mu T b_o+s)\}$$
$$= \psi(s/\mu)-\psi(b_1)\psi\{(1-T)b_o+s/\mu\}/\psi\{(1-T)b_o+b_1\}$$

$$+ b_1[\psi'(b_1)-\psi(b_1)\psi'\{(1-T)b_o+b_1\}/\psi\{(1-T)b_o+b_1\}]\psi(b_o+s/\mu)$$

$$+ (1-T)\psi(b_1)\psi(b_o+s/\mu)/\psi\{(1-T)b_o+b_1\},$$

making use of our solution to G(2)/M/1. $\{P_j\}$ is now given in terms of R_o by

$$P_j = \begin{cases} R_o(0), & j = 0, \\ 1-T\psi(b_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o+b_1\}-R_o(0)+\psi(b_o)R_o(\mu b_1), & j = 1, \\ T^{j-1}(1-T)\psi(b_1)\psi\{(1-T)b_o\}/\psi\{(1-T)b_o+b_1\} & \\ \quad -\psi(b_o)R_o(\mu b_1)[\psi(1)]^{j-1}[\psi(1)-1], & j \geq 2. \end{cases}$$

PART 2.

2. *Moving average inputs and general recurrent service times.*

We now go on to consider queueing systems with moving average inputs and general recurrent service times. It is hardly to be expected that the equilibrium queue length distributions in such systems will have simple forms, since, indeed, even GI/G/1 does not exhibit such simplicity. In view of the mathematical complexity of G(p + 1)/G/1 and its dubious utility, we shall not dwell on it in any detail.

The simplest case of GI/G/1 was given a complete time dependent treatment by Keilson and Kooharian in 1962. By making use of the method of supplementary variables they were able to reduce the problem to one of solving a Hilbert problem. The method of supplementary variables is also the most natural to apply to the more general problem of G(p+1)/G/1, and, whilst the equations are much more complex than those for GI/G/1, it seems that the best chance of solution may again lie in attempting a reduction to a Hilbert problem.

We introduce the supplementary variables as follows. We denote by

$$P_j\ (u_1,..,u_p,x,y,t) \quad j \geq 1,$$

the joint probability and probility density that at time t there are j customers waiting together with an additional customer in service, that the times that have elapsed since the beginning of the current service and the occurrence of the last arrival are x,y respectively, and that the last p values of the process $\{u\}$ have been $u_1,.., u_p$.

We distinguish between two classes of state when a single customer is in service and no other customers waiting:

We write

$$F(u_1,.., u_p,x,t)$$

when the customer in service arrived to find the queue empty and so began service immediately. We are able to drop the variable y since under these circumstances x and y must be equal.

$$P_o(u_1,..,u_p , x,y,t)$$

is employed when the customer being served began his service on the departure of the previous customer.

We finally denote by

$$E(u_1,..,u_p,y,t)$$

the probability densities of the set of 'vacuous' states of the system.

The justification for presuming such densities exist is, as in Keilson and Kooharian's treatment of GI/G/1, best left to reside in the demonstration, through a constructive proceedure, that $C^{(1)}$ initial distributions give rise to unique $C^{(1)}$ solutions. Initial distributions with saltuses can be accommodated by extending our densities to include generalised functions.

For convenience we use lower case notation, $e(\underline{u}^{(p)}, y, t)$, etc., to refer to the states corresponding to the joint probabilities and probability densities defined above.

We denote by $\lambda(y)$, $\mu(x)$ the hazard functions of $U(y)$ and the service time distribution $D(x)$, so that

$$U'(y) = \lambda(y)\exp(-\int_0^y \lambda(u)du),$$

$$D'(x) = \mu(x)\exp(-\int_0^x \mu(u)du).$$

Consider how a state density

$$(2.1) \qquad p_j(u_1,..,u_p, x + \Delta, y + \Delta, t + \Delta),$$

$x, y, t > 0$, Δ a small positive increment, can occur. As $x, y > 0$, there cannot have been either an arrival or departure during the time interval

$(t, t+\Delta)$, or either x or y would be of order Δ. Hence the state density (2.1) can only arise from a state density

$$p_j(u_1,..,u_p, x, y, t)$$

at time t.

During interval $(t, t+\Delta)$, there are to be no departures or arrivals. The probability that there are no departures is simply

$$(1-\mu(x)\Delta) + o(\Delta).$$

Since the time between the last and the subsequent arrival is

$$b_o u_{p+1} + b_1 u_p + .. + b_p u_1,$$

of which by t an amount y has elapsed, the probability that there are no arrivals during $(t, t+\Delta)$ is

$$(1-\lambda[(y-b_1u_p-..-b_pu_1)/b_o]\Delta/b_o) + o(\Delta) .$$

Thus

$$P_j(u_1,..,u_p, x+\Delta, y+\Delta, t+\Delta)$$
$$= P_j(u_1,..,u_p, x, y, t)(1-\lambda[(y-b_1u_p-..-b_pu_1)/b_o]\Delta/b_o)$$
$$\times (1-\mu(x)\Delta) + o(\Delta), \qquad j \geq 0.$$

On dividing by Δ and letting $\Delta \to o$ we find

(2.2) $$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right) P_j + \{\lambda[(y-b_1u_p-..-b_pu_1)/b_o]b_o^{-1} + \mu(x)\}P_j = 0, \; j \geq 0.$$

Similarly

(2.3) $$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right) F + \{\lambda[(y-b_1u_p-..-b_pu_1)/b_o]b_o^{-1} + \mu(x)\} F = 0.$$

With $e(u_1,.., u_p, y, t)$ the considerations are slightly different. While

$$e(u_1,...,u_p, y+\Delta, t+\Delta)$$

cannot arise from $e(u_1,.., u_p, y, t)$ if there is an arrival in $(t, t+\Delta)$, it can arise from $f(u^{(p)}, y, t)$ or $p_o(u^{(p)}, x, y, t)$ (x arbitrary) from a departure. e thus has a different form of equation from either p_j or f.

$$(2.4)\quad \left(\frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right) E + \lambda\,[(y-b_1u_p-\ldots-b_pu_1)/b_o]b_o^{-1}\, E$$
$$= \mu(y)F(u^{(p)},y,t) + \int_o^\infty P_o(u^{(p)},x,y,t)\,\mu(x)dx.$$

To the basic equations (2.2) - (2.4) we need to add suitable initial and boundary conditions to complete the specification of the behaviour of the system.

Most simply we could adopt as initial conditions

$$(2.5)\quad \begin{cases} P(u^{(p)},x,\ y,\ o) & = o, \quad j \geq o, \\ F(u^{(p)},x,\ y,\ o) & = \delta(x-o) \prod_{i-1}^{p} (u_i-\alpha_i), \\ E(u^{(p)},y,\ o) & = o, \end{cases}$$

where $\alpha^{(p)}$ is a set of positive constants. These conditions have the merit of giving (2.2) a particularly simple form of general solution, as we shall see. By adding an extra term to the general solution we derive, we can easily accommodate more general initial conditinns.

We obtain boundary conditions by considering the system of arrival and departure instants.

First, take arrival instants. The system can enter $p_j(u^{(p)},\ x,\ o,\ t)$, $j > 1$, only from a p_{j-1} state. As there is a simple 'shift' in the u's at an arrival, that with lowest subscript being lost and a new u with highest subscript appearing, the set of u's just before such an arrival must have been of the form

$$v,\ u_1,\ldots,\ u_{p-1}.$$

The corresponding inter-arrival interval will be

$$b_ou_p + \ldots + b_{p-1}u_1 + b_pv.$$

We readily derive the boundary condition

$$(2.6)\quad P_j\ (u^{(p)},\ x,\ o,\ t)$$
$$= \int_o^\infty P_{j-1}(v,u_1,\ldots,u_{p-1},x,b_ou_p+\ldots+b_pv,t)\lambda(u_p)b_o^{-1}dU(v),\ j > 1,$$

on integrating to allow for all possible v's.

We similarly derive the further boundary conditions

(2.7)
$$\begin{cases} P_1(u_\infty^{(p)}, x, o, t) \\ \quad = \int_o^\infty P_o(v, u_1, .., u_{p-1}, x, b_o u_p + .. + b_p v, t)\lambda(u_p) b_o^{-1} dU(v) \\ \quad + F[(x - b_o u_p - b_o u_{p-1} - b_1 u_{p-2} - .. - b_{p-1} u_1)/b_p, u_1, .., u_{p-1}, x, t] \\ \qquad\qquad \lambda(u_p) b_o^{-1}, \\ P_o(u^{(p)}, x, o, t) = o \\ F(u_\infty^{(p)}, o, t) \\ \quad = \int_o E[(y - b_o u_p - b_o u_{p-1} - b_1 u_{p-2} - .. - b_{p-1} u_1)/b_p, u_1, .., u_{p-1}, y, t] \times \\ \qquad \lambda(u_p) b_o^{-1} d_y\, U[y - b_o u_p - b_o u_{p-1} - b_1 u_{p-2} - .. - b_{p-1} u_1], \\ E(u^{(p)}, o, t) = 0. \end{cases}$$

The departure instants give an additional relation

(2.8) $$P_j(u^{(p)}, o, y, t) = \int_o^\infty P_{j+1}(u^{(p)}, x, y, t)\mu(x)dx, \quad j \geq 0.$$

We now give some idea as to the connection between our equations (2.2)-(2.8) and Hilbert problems.

Suppose that L is the union of a set of smooth non-intersecting contours (of which one encloses all the others) bounding a connected region.

The basic non-homogeneous Hilbert problem is;[1]

To find a sectionally holomorphic function $\Phi(z)$, having finite degree at infinity and satisfying on L the boundary condition

(2.9) $$\Phi^+(t) = G(t)\Phi^-(t) + g(t),$$

where G, g are functions given on L satisfying the Hölder condition and $G(t) \neq o$ everywhere on L.

The suffices +, - refer to the half line decomposition

1. *MUSKHELISHVILI: Singular Integral Equations, P. Noordhoff, Groningen, Holland (1953), Ch. 5.*

$$\Phi^{+}(t) = \begin{cases} \Phi(t), & t \geq 0 \\ 0, & t < 0 \end{cases},$$

$$\Phi^{-}(t) = \begin{cases} 0, & t \geq 0 \\ \Phi(t), & t < 0 \end{cases}.$$

We observe that

$$\Phi(t) = \Phi^{+}(t) + \Phi^{-}(t) \tag{2.10}$$

When we say a function $f(t)$ satisfies the Hölder condition on L, we mean

There exist positive constants A, μ, *such that, for any two points* t_1, t_2 *of* L

$$|f(t_2) - f(t_1)| \leq A|t_2 - t_1|^{\mu}.$$

The basic Hilbert problem admits of a simple solution in terms of Cauchy integrals.

(2.9) may seem remote from our fundamental equations but a decomposition (2.10) in fact arises in a very natural way from (2.2).

If we make the substitution

$$Q_j(u^{(p)}, x, y, t) = \exp[b_o^{-1} \int_o^y \lambda[(u-b_1u_p-\ldots-b_pu_1)/b_o]du + \int_o^x \mu(u)du] \times P_j(u^{(p)}, x, y, t), \quad j \geq o,$$

(2.2) becomes

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right)Q_j = 0, \quad j \geq 0.$$

Following Keilson and Kooharian (1962), the general solutions to these equations with the initial conditions (2.5), i.e.,

$$Q_j(u^{(p)}, x, y, o) = o, \quad j \geq o,$$

are

$$Q_j(u^{(p)}, x, y, t) = Q_{j1}(u^{(p)}, x-y, t-y) + Q_{j2}(u^{(p)}, x-y, t-x), \quad x,y,t \geq o, \quad j \geq o, \tag{2.11}$$

where Q_{j1} vanishes for either $x < y$ *or* $t < y$ and Q_{j2} for either $x > y$ or $t < x$.

In the case of GI/G/1, (2.6), (2.8) have a particularly simple Wiener-Hopf form when we take generating functions on j, thanks to the decomposition (2.11). This enabled Keilson and Kooharian to reduce the equations to a Hilbert problem, using integral transforms. It seems possible that a similar though more involved treatment might succeed with the present equations.

* * * * * * * *

REFERENCES.

Andersen, E.S., (1953) *On sums of symmetrically dependent random variables, Skan. Aktuar.* 36, 123-38.

(1953) *On the fluctuations of sums of random variables I, Math. Scand.* 1, 263-85.

(1954) *On the fluctuations of sums of random variables II, Math. Scand.* 1, 195-223.

Arora, K.L., (1962) *Time-dependent solution of the 2-server queues fed by general arrival and exponential service time distributions, Opns Res.* 10, 327-34.

Bailey, N.T.J., (1954) *On queueing processes with bulk service, J.Roy. Statist. Soc. B* 16, 80-7.

(1954) *A continuous time treatment of a simple queue using generating functions, J.Roy. Statist. Soc. B* 16, 288-91.

Belyaev, U.K., *Predelnaya teorema dlya reduyušči potokov (rukopis!)*

Beneš, V. E., (1957) *On queues with Poisson arrivals, Ann. Math. Statist.* 28, 670-7.

(1960) *Combinatory methods and stochastic Kolmogorov equations in the theory of queues with one server, Trans. Amer. Math. Soc.* 94, 282-94.

(1960b) *General stochastic processes in traffic systems with one server, Bell System Tech. J.* 39, 127-60.

(1963) *General stochastic processes in the theory of queues, Addison-Wesley Publ. Co., Inc., Reading, Mass.*

Bhat, U. N., *(1963) Imbedded Markov chain analysis of bulk queues, J. Aust. Math. Soc. IV,244-63.*

Borel, E., *(1942) Sur l'emploi du théoreme de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au problème de l'attente à un guichet, C.R. Acad. Sci., Paris, 214, 452-6.*

Brockmeyer, E., Halstrøm, H.L. and Jensen, A., *(1948) The life and works of A.K.Erlang, Trans of the Danish Acad. of Sci. No. 2, 1-277.*

Brockwell, P. J., *(1963) The transient behaviour of the queueing system GI/M/1, J. Aust. Math. Soc. III, 249-56.*

Burke, P. J., *(1956) The output of a queueing system, Opns Res. 4, 699-704.*

Champernowne,D.G., *(1956) An elementary method of solution of the queueing problem with a single server and constant parameters, J. Roy. Statist. Soc. B 18, 125-8.*

Clarke, A.B., *(1953) The time-dependent waiting time problem, Univ. of Michigan Eng. Res. Inst. Report M 720-1, R39.*

(1956) A waiting time process of Markov type, Ann. Math. Statist. 27, 452-9.

Conolly, B.W., *(1958) A difference equation technique applied to the simple queue, J.Roy. Statist. Soc. B 20, 165-7.*

(1958) A difference equation technique applied to the simple queue with arbitrary arrival interval distribution, J. Roy. Statist. Soc. B 20, 168-75.

(1959) *The busy period in relation to the queueing process GI/M/1*, Biometrika 46, 246-51.

* (1960) *The busy period in relation to the single-server queueing system with general independent arrivals and Erlangian service-time*, J. Roy. Statist. Soc. B 22, 89-96.

(1960) *Queueing at a single serving point with group arrival*, J. Roy. Statist. Soc. B 22, 285-98.

Copson, E.T., (1935) *An introduction to the theory of functions of a complex variable*, Oxf. Univ. Press, London.

Cox, D. R., (1955) *The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables*, Proc. Camb. Phil. Soc. 51, 433-41.

(1955) *A use of complex probabilities in the theory of stochastic processes*, Proc. Camb. Phil. Soc. 51, 313-39.

Cox, D.R., and Smith, W.L., (1961) *Queues*, Methuen, London.

Daniels, H.E., (1961) *Mixtures of geometric distributions*, J. Roy. Statist. Soc. B 23, 409-13.

Doig, A. (1957) *A bibliography of the theory of queues*, Biometrika 44, 490-514.

Downton, F., (1955) *Waiting times in bulk service queues*, J. Roy, Statist. Soc. B 17, 256-61.

(1956) *On limiting distributions arising in bulk service queues*, J. Roy. Statist. Soc. B 18, 265-74.

Dunn, P.F., Flagle, C.D., and Hicks, P.A., (1956) *The Queuiac: An electromagnetic analog for the simulation of waiting time problems*, *Opns. Res.* 4,648-62.

Fabens, A.J., (1961) *The solution of queueing and inventory models by semi-Markov processes, J. Roy. Statist. Soc. B* 23, 113-27.

Feller, W., (1949) *Fluctuation theory of recurrent events, Trans. Amer. Math Soc.* 67, 98-119.

(1950) *An Introduction to Probability Theory and its Applications, Vol. 1, J. Wiley and Sons Inc., New York.*

(1959) *On combinatorial methods in fluctuation theory, The Harold Cramér Volume, 75-91, J. Wiley and Sons Inc., New York.*

Finch, P. D., (1958) *The effect of the size of the waiting room on a simple queue, J. Roy. Statist. Soc. B* 20, 182-6.

(1959a) *Cyclic queues with feedback, J.Roy.Statist.Soc B* 21, 153-7.

* (1959) *The output process of the queueing system M/G/1, J.Roy. Statist. Soc. B.* 21, 375-80.

(1959) *A probability limit theorem with application to a generalisation of queueing theory. Acta Math. Acad. Sci. Hung.* 10, 317-26.

(1959) *Balking in the queueing system GI/M/1. Acta Math. Acad. Sci. Hung.* 10,241-7.

(1959) *On the distribution of queue size in queueing problems, Acta Math. Acad. Sci. Hung.* 10,327-36.

(1960) *On the transient behaviour of a simple queue, J.Roy. Statist. Soc. B* 22, 277-84.

(1961) *The transient behaviour of a coincidence variate in telephone traffic, Ann. Math. Statist.* 32, 230-34.

* (1961) *Some probability problems in inventory control, Publ. Math. Debrecen* 8, 241-61.

(1961) *The order cycle system of a stock control, J. Aust. Math.*

Soc. 2, 209-16.

(1961) On the busy period in the queueing system GI/G/1, J. Aust. Math. Soc. 2, 217-28.

(1962) A generalised single-server queue with Erlang input (with Ewens, W.J.), Biometrika 49, 242-5.

* *(1963) The single server queueing system with non-recurrent input-process and Erlang service time, J. Aust. Math. Soc. III, 220-36.*

(1963) On partial sums of Lagrange's series with application to the theory of queues, J. Aust. Math. Soc. III, 488-490.

(1963a) A coincidence problem in telephone traffic with non-recurrent arrival process, J. Aust. Math. Soc. III, 237-40.

Finch, P.D., and Pearce, C., *(1965) A second look at a queueing system with moving average input process, J. Aust. Math. Soc. V, 100-106.*

Forsyth, A. R., *(1885) A treatise on differential equations, Macmillan and Co. Ltd, London.*

Foster, F.G., *(1953) On the stochastic matrices associated with certain queueing processes, Ann. Math. Statist. 24, 355-60.*

(1959) A unified theory for stochastic storage and queue control, Opns Res. Quart. 10, 121-30.

(1961) Queues with batch arrivals I, Acta Math. Acad. Sci. Hung. 12, 1 - 10.

Foster, F.G., and Nyunt, K.M., *(1961) Queues with batch departures 1, Ann. Math. Statist. 32, 1324-32.*

Gaver, D.P., *(1954) The influence of servicing times in queueing processes, Opns Res. 2, 139-49.*

(1959) Imbedded Markov chain analysis of a waiting time process in continuous time, Ann. Math. Statist. 30,698-720.

(1962) A waiting time with interrupted service including priorities, J. Roy. Statist. Soc. B. 24, 73-90.

Ghosal, A., *(1962) Queues in Series, J. Roy. Statist. Soc.B24,359-63.*

Hannan, E.J., *(1960) Time Series Analysis, Methuen Monograph, Methuen and Co. Ltd., London.*

Hasofer, A.M., *(1964) On the single-server queue with non-homogeneous Poisson input and general service time, J. Appl. Prob. 1, 369-84.*

Heathcote, C.R., *(1959) The random walk (in continuous time) and its application to the theory of queues, Biometrika 46, 400-11.*

(1961) On the queueing process M/G/1, Ann. Math. Stat. 32, 770-3.

Heyde, C.C., *(1964) On the stationary waiting time distribution in the queue GI/G/1, J. Appl. Prob. 1, 173-6.*

Hobson, E.W., *(1926) The Theory of Functions of a Real Variable, Camb. Univ. Press.*

Ibragimov, I.A., *(1962) Some limit theorems for stationary processes, T.V., 7, 4, 349-82.*

Jackson, R.R.P., and Nichols, D.G., *(1956) Some equilibrium results for the queueing process $E_k/M/1$, J. Roy.Statist. Soc. B 18,275-9.*

Jaiswal, N.K., *(1960)Bulk-service queueing problem, Opns Res. 8,139-43.*

(1960) Time-dependent solution of the bulk-service queueing problem, Opns Res. 8,773-81.

(1960) A bulk-service queueing problem with variable capacity, J. Roy. Statist. Soc. B 23,143-8.

(1962) Time-dependent solution of the head-of-the-line priority queue, J. Roy. Statist. Soc. B 24, 91-101.

Jensen, A., *(1954) A Distribution Model, Copenhagen.*

Karlin, S., and McGregor, J.,*(1958) Many server queueing process with Poisson input and exponential service times, Pacific J. Math. 8, 87-118.*

Keilson, J., *(1961) The homogeneous random walk on the half-line and the Hilbert problem, Bull.de l'Inst.Inter. de Statistique,33e session.*

(1962) The use of Green's functions in the study of bounded random walks with application to queueing theory, J. Math. Phys. 41, 42-52.

(1962) Non-stationary Markov walks on the lattice, J.Math. Phys. 41, 205-11.

(1962) A simple random walk and an associated asymptotic behaviour of the Bessel functions, Proc. Camb. Phil. Soc. 58, 708-9.

(1962b) Queues subject to service interruption, Ann. Math. Statist. 33, 1314-22.

(1962a) The general bulk queue as a Hilbert problem, J.Roy. Statist. Soc. 24, 344-58.

(1963) On the asymptotic behaviour of queues, J. Roy. Statist. Soc. B 25, 464-76.

(1964) Some comments on single-server queueing methods and some new results, Proc. Camb. Phil. Soc. 60, 237-51.

Keilson, J., and Kooharian, A., *(1960) On time-dependent queueing processes, Ann. Math. Statist. 31, 104-12.*

(1962) On the general time-dependent queue with a single server, Ann. Math. Statist. 33, 767-91.

Kemperman, J. H., *(1961) The first passage problem for a stationary Markov chain, Univ. of Chicago Press.*

Kendall, D.G., *(1951) Some problems in the theory of queues, J. Roy. Statist. Soc. B 13, 151-85.*

(1953) Stochastic processes occurring in the theory of queues and their analysis by means of the imbedded Markov chain, Ann. Math. Statist. 24, 338-54.

(1964) Some recent work and further problems in the theory of queues, T.V. 9, 3-15.

Khintchine, A.Y., *(1932) Mathematical theory of a stationary queue, Mat. Sbornik 39, 73-84.*

(1933) Über die mittlere Dauer des Stillstandes von Maschinen, Mat. Sbornik 40, 119-23.

(1960) Mathematical Methods in the Theory of Queueing (trans. by D.M.Andrews and M.H.Quenouille), Griffin & Co.Ltd, London.

Kiefer, J., and Wolfowitz, J., *(1955) On the theory of queues with many servers, Trans. Amer. Math. Soc. 78, 1-18.*

(1956) On the characteristics of the general queueing process, with applications to random walk, Ann. Math. Statist. 27, 147-61.

Kingman, J.F.C., *(1961) The single server queue in heavy traffic, Proc. Camb. Phil. Soc. 57, 902-4.*

(1962) On queues in heavy traffic, J. Roy. Statist. Soc. B 24, 383-92.

(1962) The use of Spitzer's identity in the investigation of the busy period and other quantities in the queue GI/G/1 J. Aust. Math. Soc. II, 345-56.

(1962) Some inequalities for the queue GI/G/1, Biometrika 49, 715-24.

Koenigsberg, E., *(1956) Queueing with special service, Opns Res. 4, 213-20.*

(1960) Finite queues and cyclic queues, Opns Res. 8, 246-53.

Kolmogorov, A.N., *(1931) Sur le problème d'attente, Recueil Mathématique, Mat. Sbornik 38, 101-6.*

Lederman, W., and Reuter, G.E.H., *(1954) Spectral theory for the differential equations of simple birth and death processes, Phil. Trans. A, 246, 321-69.*

Le Gall, P., *(1962) Les systèmes avec ou sans attente et les processus stochastiquesl, Paris, Dunod.*

Lindley, D.V., *(1952) The theory of queues with a single server, Proc. Camb. Phil. Soc. 48., 277-89.*

Loève, M., *(1955) Probability Theory, Van Nostrand Co., New York.*

Lovitt, W.V., *(1950) Linear Integral Equations, Dover Publications Inc., New York.*

Loynes, R. M., *(1962) The stability of a queue with non-independent inter-arrival and service times, Proc. Camb. Phil. Soc. 58, 497-520.*

(1962a) Stationary waiting-time distribution for single-server queues, Ann. Math. Statist. 33, 1323-39.

(1964) The stability of a system of queues in series, Proc. Camb. Phil. Soc. 60, 569-74.

(1964a) A continuous-time treatment of certain queues and infinite dams, J. Aust. Math. Soc II, 484-98.

Luchak, G., *(1956) The solution of the single-channel queueing equations characterized by a time-dependent Poisson-distribution arrival rate and a general class of holding times, Opns Res. 4, 711-32.*

(1957) The distribution of the time required to reduce to some preassigned level a single-channel queue characterized by a time-dependent Poisson distributed arrival rate and a general class of holding times, Opns Res. 5, 205-9.

(1958) The continuous-time solution of the equations of the single-channel queue with a general class of service-time distribution by the method of generating functions, J. Roy Statist. Soc. B 20. 176-81.

Lucács, E., and Szász, O., *(1952) Some non-negative trigonometric polynomials connected with a problem in probability, J. Res. Nat. Bur. Stand., 48, (2297).*

(1954) Non-regenerative trigonometric polynomials and certain rational characteristic functions, J. Res. Nat.

Bur. Stand., 52 (2484).

Mercer, A., (1960) *A queueing problem in which the arrival times of the customers are scheduled*, J. Roy. Statist. Soc. B 22, 108-13.

Miller, R. G., (1959) *A contribution to the theory of bulk queues*, J. Roy. Statist. Soc. B 21, 320-37.

Morse, P.M., (1958) *Queues, Inventories and Maintenance*, Wiley and Sons Inc., New York.

Muskhelishvili, N.T., (1946) *Singular Integral Equations*, Moscow. (Trans. by J.M.R. Radok, P. Noordhoff, Netherlands, 1953)

O'Brien, G.G., (1954) *The solution of some queueing problems*, J. Soc. Indust. Appl. Math. 2, 133-42.

Palm, C., (1943) *Intensitätsschwankungen im Fernsprechverker*, Ericsson Technics 44, 1-189.

Pearce, C., (1965) *A second look at a queueing system with moving average input process*, J. Aust. Math. Soc. (jointly with P.D. Finch).

(accepted for publication) *A queueing system with general moving average input and negative exponential service time*, J. Aust. Math. Soc.

(accepted for publication) *A queueing system with non-recurrent input and batch servicing*, J. Appl. Prob.

(accepted for publication) *A queueing system with moving average input process and batch arrivals*, J. Aust. Math. Soc.

Peck, L.G., and Hazelwood, R.N., *(1958) Finite queueing tables, ORSA publications in Operations Research No. 2, Wiley and Sons Inc., New York.*

Pinney, E., *(1959) Ordinary Difference-differential Equations, Univ. of Calif. Press.*

Pollaczek, F., *(1930) Über eine Aufgabe der Wahrscheinlich Keitsrechnung, Mat. Zeit. 32, 64-100, 729-750.*

(1952) Délais d'attente des avions atterrisant selon leur ordre d'arrivée sur un aéroport à s pistes, C. R. Acad. Sci. Paris, 235, 1246-8.

(1953) Sur une généralisation de la théorie des attentes, C. R. Acad. Sci., Paris, 236, 578-80.

(1954) Calcul des probabilités - développement de la théorie stochastique des lignes téléphoniques pour un état initial quelconque, C. R. Acad. Sci., Paris, 239, 1764-6.

(1957) Problèmes stochastique posés par le phénomène de formation d'une queue d'attente à un guichet et par des phénomènes apparentés, Mémorial des Sciences Mathématiques, fasc. 136, Gauthier-Villars, Paris.

(1963) Théorie analytique des problèmes stochastiques relatifs à un groupe de lignes téléphoniques avec dispostif d'attente, Gauthier-Villars, Paris.

Prabhu, N. U., *(1960) Some results for the queue with Poisson arrivals, J. Roy. Statist. Soc. B 22, 104-7.*

(1960) Application of storage theory to queues with

Poisson arrivals, Ann. Math. Statist. 31, 475-82.
(1965) Queues and inventories, Wiley and Sons, Inc., New York.

Prabhu, N.U., and Bhat, N.U., *(1963) Some first passage problems and their application to queues, Sankhya A 25, 281-92.*

Pyke, R., *(1959) The supremum and infinum of the Poisson process, Ann. Math. Statist. 30, 568-76.*

Reich, E., *(1956) Birth-death process and tandem queues, P. 863, the Rand Corp.*
(1957) Waiting times when queues are in tandem, Ann. Math. Statist. 29, 563-70.
(1958-9) On the Inte grodifferential equation of Takács I, II, Ann. Math. Statist 29, 30, 563-70, 143-8.
(1963) Notes on queues in tandem, Ann. Math. Statist. 34, 338-41.

Rényi, A., *(1956) A Poisson-folyamat egy jellemzése, Math.works Inst. A. N. Vingri I, 519-27.*

Runnenberg, J.T., *(1958) Probabilistic interpretation of some formulae in queueing theory, 31st session of the Int. Stat. Inst., Brussels.*
(1960) On the use of Markov processes in one-server waiting time problems and renewal theory, Klein Offsetdrukkerij Poortpers N.V., Amsterdam, Netherlands.
(1960) Probabilistic interpretation of some formulae in queueing theory, Bull. Inst. Int. Statist. 37, 405-14.

Rice, S.O., (1962) *Single server system II, busy periods, Bell System Tech. J. 41, 279-310.*

Riley, V., (1956) *Bibliography of queueing theory, Opns Res. for Management, John Hopkins Press, Baltimore, Vol. 2, 541-6.*

Saaty, T. L., (1957) *Résumé of useful formulas in queueing theory, Opns Res. 5, 161-200.*

(1960) *Time-dependent solution of the many-server Poisson queue, Opns Res. 8, 755-72.*

(1961) *Elements of Queueing Theory, McGraw-Hill Book Co. Inc., New York.*

Sacks, J., (1960) *Ergodicity of queues in series, Ann. Math. Statist. 31, 579-88.*

Šahbazov, A.A., (1962) *A service problem with unusual demand, Dokl. Akad. Nauk . S.S.R. 145, 289-92.*

Shu-teh Chen Moy, (1959) *Successive recurrence times in a stationary process, Ann. Math. Statist. 30, 1254-7.*

Smith, W.L., (1953) *On the distribution of queueing times, Proc. Camb. Phil. Soc. 49, 449-61.*

(1955) *Regenerative stochastic processes, Proc. Roy. Soc. Edinburgh Sec. A, 232, 6-31.*

(1958) *Renewal theory and its ramifications, J. Roy. Statist. Soc. B 20, 243-302.*

Smithies, F., (1958) *Integral Equations. Camb. Tracts in Maths and Math. Phys., No. 49, C.U.P.*

(1939) Singular integral equations, Proc. London Math. Soc. 46, 409-66.

Spitzer, F., *(1956) A combinatorial lemma and its applications to probability theory, Trans. Amer. Math. Soc. 82,323-39.*

(1957) The Wiener-Hopf equation whose kernel is a probability density, Duke J. Math. 24, 327-44.

Swensson, O., *(1958) An approach to a class of queueing problems, Opns Res. 6, 276-92.*

Syski, R., *(1960) Introduction to Congestion Theory in Telephone Systems, Oliver and Boyd, Edinburgh.*

Takács, L., *(1955) Investigation of waiting time problems by reduction to Markov processes, Acta Math. Acad. Soc. Hung. 6, 101-29.*

(1957) On a queueing problem concerning telephone traffic, Acta Math. Acad. Sci. Hung. 9, 45-81.

(1959) On the limiting distribution of the number of coincidences concerning telephone traffic, Ann. Math. Statist. 30, 134-42.

(1960) Transient behaviour of single-server queueing processes with recurrent input and exponentially distributed service times, Opns Res. 8, 231-45.

* *(1961) Transient behaviour of single-server queueing processes with queueing processes with Erlang input, Trans. Amer. Math. Soc. 100, 1-28.*

(1961) The transient behaviour of a single server queueing process with a Poisson input, Proc. Fourth Berkeley

(1958) On a coincidence problem concerning telephone traffic Acta Math. Acad. Sci. Hung. 9,45-81.

Symp. on Math. Stat. and Prob., Univ. of Calif. Press, 535-67.

* *(1961) The transient behaviour of a single server queueing process with recurrent input and gamma service time, Ann. Math. Statist. 32, 1286-98.*

(1962) A single-server queue with Poisson input, Opns Res. 10, 388-94.

(1962) A single-server queue with recurrent input and exponential distribution service time, Opns. Res. 10, 395-9.

(1962) The time dependence of a single-server queue with Poisson input and general service times, Ann. Math. Statist. 33, 1340-8.

(1962) Introduction to the Theory of Queues, Oxf. Univ. Press.

(1963) A single-server queue with feedback, Bell System Tech. J. 42, 505-19.

Tanner, J.C., *(1961) A derivation of the Borel distribution, Biometrika 48, 222-4.*

(1953) A problem of interference between two queues, Biometrika 40, 58-69.

Titchmarsh, E.C., *(1932) The Theory of Functions, Oxf. Univ. Press.*

Vere-Jones, D., *(1964) A rate of convergence problem in the theory of queues, T.V. 9, 104-112.*

Volberg, O., *(1939) Problème de la queue stationnaire et non-stationnaire, C.R.(Doklady) Acad.Sci.U.R.S.S. 24, 657-61.*

Widder, D.V., (1946) *The Laplace Transform, Princeton Univ. Press.*

Winsten, C.B., (1959) *Geometric distributions in the theory of queues, J. Roy. Statist. Soc. B 21, 1-35.*

Wishart, D.M.G., (1956) *A queueing system with* χ^2 *service-time distribution, Ann. Math. Statist. 27, 768-79.*

(1959) *A queueing system with service-time distribution of mixed Chi-squared type, Opns Res. 2, 174-9.*

Wu, Fang, (1962) *On the queueing process GI/M/n, Sci. Sinica 11, 1169-82.*

Yarovitsky, H.B., (1962) *On certain properties of singly connected dependent streams, Ukr. Math. J. 14, 170-9.*

Zygmund, A., (1951) *A remark on characteristic functions, Proc.Second Berkeley Symp. on Math. Stat. and Prob., Univ. of Calif. Press, 369-72.*

* * * *