On Discrete Time Markovian N-policy Queues involving Batches

Böhm, Walter; Mohanty, Sri Gopal

DOI:
10.57938/1e4d1684-4ca6-41e7-bacc-db4a0b97d9ab

Published: 01/01/1991

Document Version:
Publisher's PDF, also known as Version of record

On Discrete Time Markovian N-policy Queues involving Batches

Walter Böhm, Sri Gopal Mohanty

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 24
1991

http://statmath.wu-wien.ac.at/

On discrete time Markovian N-policy queues involving batches

Walter Böhm
University of Economics
Vienna

S.G. Mohanty
McMaster University
Hamilton, Ontario

Abstract

Consider two Markovian N-policy queueing models in discrete time, one with batch arrival, the other with batch service. In this paper the transient behaviour of both models is studied and the analogous continuous time results are achieved by a limiting process. The steady state solution for the model with batch arrival is derived.

Keywords: N-policy, transient behaviour, $M/E_k/1$, $E_k/M/1$, limiting process, steady state, batch, discrete time queues

1 Introduction

Discrete time queueing models arise in a natural way in modern systems, for example in computers and communication channels (Falin (1989)), because time is treated as a discrete variable in many practical situations. For mathematical elegance (in particular by the use of methods in analysis) such models are approximated by continuous time models, which are treated extensively in the literature. However, discrete time models are conceptually simpler and often computationally more convenient (particulary by the use of high speed computers) than continuous time models. Furthermore in the recent past a steady progress has been made in discrete mathematics due to the advances in computer science, which may provide powerful tools to analyse these types of problems.

The traditional analysis of queueing models in continuous time heavily relies on transformation of the system equations at the outset without probing into the structure of the system. This is what we may call "top-to-bottom" analysis. On the other hand, for discrete time situations the tools in discrete mathematics (in particular enumerative combinatorics) can be used to examine the structure of the system and to split it into appropriate subsystems, which are mathematically easier to handle. This approach in contrast to the former may be referred to as "bottom-to-top" analysis.

In certain situations it is possible that solutions for continuous time problems may be obtained from the corresponding discrete time results by a limiting process. In these cases it is indeed more advantageous to start with the discrete time model rather than its approximated continuous time model.

In the past discrete time models have been dealt with sporadically (Meisling (1958), Bhat (1964, 1968), Dafermos and Neuts (1973), Neuts (1973), Kobayashi (1983), Hunter (1983), Mohanty and Panny (1990a), (1990b), Kanwar Sen and Jain (1990)). None-the-less the need for such models and their strength have been recognized by all of these authors.

In this paper we discuss two discrete time models, one involving batch arrival, the other batch departure. In addition we consider N-policy operating on these models, i.e. the service does not start unless there are N customers present in the system, except at the beginning. Because of the importance of transient behaviour, especially when the systems are not stable the transient solutions are derived in section 2 and 3. In section 4 the steady state solution for the queue with batch arrivals is given. Section 5 is devoted to the transient solutions in the continuous time case.

Queues involving batches have been treated earlier (Takács (1962), Bhat (1964), (1968), Dafermos and Neuts (1973), Chaudhry and Templeton (1983)). However, our method varies from others and is somewhat combinatorial in nature, which is based on the decomposition of the sample paths.

For earlier work on N-policy queues one may refer to Hayman (1968), Neuts (1989), Böhm and Mohanty (1990) and Kanwar Sen, Jain and Gupta (1990). In none of these papers transient solutions for systems involving batches have been dealt with. It may be seen that our combinatorial approach through decomposition seems to be simpler to apply to N-policy queues, than the usual solution techniques.

2 Batch arrivals

Consider the discrete time queueing process by specifying its arrival patterns, the service mechanism and the queue discipline in the following way:

- arrival process: customers arrive in batches of a fixed size $R \geq 1$ according to a Bernoulli process with success probability α. Thus the interarrival times are i.i.d. geometric.
- service mechanism: a single server serves customers one at a time. The service times, which are independent of arrival times are i.i.d. random variables having a geometric distribution with parameter γ.
- queue discipline: the customers are served on first-come-first-served basis.

It may be realized that the batch arrival process as described is the discrete time analogue of $M/E_R/1$. In addition we assume that the system is operated according to N-policy strategy, which is explained as follows:

> whenever the server becomes empty the service is turned off and becomes active as soon as there are N ($N = cR, c$ being a positive integer) customers in the system.

In this section we derive the transient solution for the above N-policy queues with batch arrivals.

Let Q_n denote the number of customers in the system at time n and Z_n the state of the system, which may be either idle or busy, to be represented by $\{Z_n = \omega\}$ and $\{Z_n = \sigma\}$, respectively. The bivariate process $\{(Q_n, Z_n), n \geq 0\}$ is a Markov chain.

Let us represent Q_n by a lattice path in the (x, y)-plane with step set $S = \{(1, R), (1, 0), (1, -1)\}$, where $(1, R)$ represents a batch arrival, $(1, -1)$ a departure and $(1, 0)$ the event {no change}. A typical path in which the server is busy initially and at the end, is illustrated below:

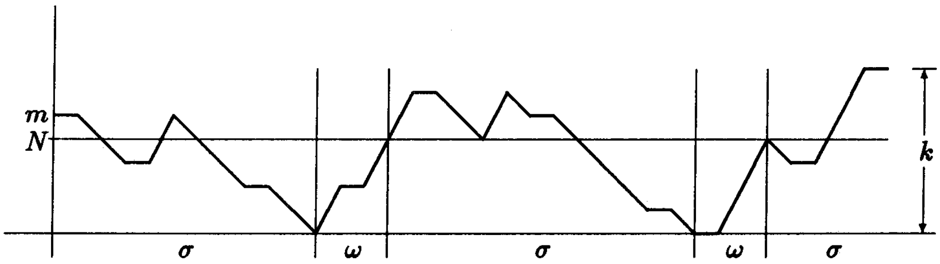

Figure 1.

Observe that the path demonstrates repetitive patterns consisting of idle period segments and busy period segments (excluding the first and last busy period). The segments due to busy periods may be described conveniently in terms of a random walk process X_n, defined on the set of all integers, having the following one-step transition probabilities

$$\begin{aligned} P(X_n = k + R|X_{n-1} = k) &= \alpha \\ P(X_n = k|X_{n-1} = k) &= \beta \\ P(X_n = k-1|X_{n-1} = k) &= \gamma, \end{aligned}$$

where $\beta = 1 - \alpha - \gamma$. For later reference let us denote

$$\begin{aligned} p_n(m,k) &= P(X_n = k|X_0 = m) \\ q_n(m) &= P(X_n = 0, X_i > 0, 0 \leq i < n|X_0 = m) \\ s_n(m,k) &= P(X_n = k, X_i > 0, 0 \leq i \leq n|X_0 = m). \end{aligned} \tag{1}$$

We set $q_n(0) = \delta_{n0}$, where δ_{n0} denotes the Kronecker delta and $s_n(m,k) = 0$, whenever either k or $m \leq 0$.

It is immediately verified that

$$s_n(m,k) = p_n(m,k) - q_n(m) * p_n(0,k), \tag{2}$$

where "*" denotes convolution.

Referring to Q_n let us denote the length of a busy period initiated by m customers by T_m, i.e.

$$T_m = \inf\{n : Q_n = 0|Q_0 = m\}. \tag{3}$$

Evidentely

$$P(T_m = n) = q_n(m), \tag{4}$$

and

$$s_n(m,k) = P(Q_n = k, T_m > 0 | Q_0 = m), \tag{5}$$

which gives the zero-avoiding transition probabilities of Q_n. The probabilities (1) may be determined either by analytical or combinatorial means. Each of these approaches has its merits and therefore we shall discuss both of them shortly (Takács has a similar treatment in Takács (1975)).

It is seen that the probabilities $s_n(m,k)$ satisfy the following system of partial difference equations:

$$\begin{aligned} s_n(m,k) &= 0 & (k < 1) \\ s_n(m,k) &= \beta s_{n-1}(m,k) + \gamma s_{n-1}(m,k+1) & (1 \le k \le R) \\ s_n(m,k) &= \alpha s_{n-1}(m,k-R) + \beta s_{n-1}(m,k) + \gamma s_{n-1}(m,k+1) & (k > R) \end{aligned} \tag{6}$$

Defining the generating functions $G_{mk}(t) = \sum_{n \ge 0} t^n s_n(m,k)$ and $H_m(t,z) = \sum_{i \ge 0} z^i G_{mi}(t)$, the solution of system (6) is found to be:

$$H_m(t,z) = \frac{z^{m+1} - t\gamma z G_{m1}(t)}{z - t(\gamma + \beta z + \alpha z^{R+1})}. \tag{7}$$

Note that the unknown function $t\gamma G_{m1}(t)$ is nothing else but the pgf. of $q_n(m)$, the probability function of the length of a busy period, since $t\gamma G_{m1}(t)$ is the generating function of the probability that the random walk X_n will move from m to state 1 in $n-1$ steps without touching the zero state, followed by a down step causing absorption at zero. To determine $G_{m1}(t)$ we observe that the denominator of (7) has exactly one positive real root inside the unit circle, say ξ. Since $H_m(t,z)$ is an analytic function in z for $|z| < 1$, we have

$$\xi^m = t\gamma G_{m1}(t).$$

The series expansion of ξ^m in powers of t is found by applying the Lagrange inversion theorem to

$$z = t(\gamma + \beta z + \alpha z^{R+1}), \tag{8}$$

in particular we obtain:

$$\begin{aligned} [t^n]\xi^m &= q_n(m) \\ &= \frac{m}{n} \binom{n; \alpha, \beta, \gamma}{n-m}_R, \end{aligned} \tag{9}$$

where $\binom{n;\alpha,\beta,\gamma}{n-m}_R$ denotes the coefficient of z^{n-m} in the expansion of $(\gamma+\beta z+\alpha z^{R+1})^n$. We will call these expressions "generalized R-trinomial coefficients" (see generalized trinomial coefficients in Panny (1984)).

Expanding $H_m(t,z)$ in powers of t and z we finally obtain:

$$\begin{aligned} s_n(m,k) &= \binom{n;\alpha,\beta,\gamma}{n+k-m}_R - \sum_{\nu\geq 1}\frac{m}{\nu}\binom{\nu;\alpha,\beta,\gamma}{\nu-m}_R \binom{n-\nu;\alpha,\beta,\gamma}{n-\nu+k}_R \\ &= p_n(m,k) - q_n(m) * p_n(0,k), \end{aligned}$$

since $p_n(a,b)$ is the coefficient of z^{n+b-a} in the series expansion of $(\gamma+\beta z+\alpha z^{R+1})^n$.

Alternatively the probabilities $p_n(m,k)$ and $q_n(m)$ may be determined by purely combinatorial arguments.

Let us consider first $p_n(m,k)$, or equivalently $p_n(0,a)$ because of spatial homogeneity of X_n. The sample paths of X_n having probability $p_n(0,a)$ may be constructed in the following way: join n_1 $(1,-1)$-steps with n_2 $(1,R)$-steps. This gives rise to a total of $\binom{n_1+n_2}{n_1}$ paths, each having probability $\alpha^{n_2}\gamma^{n_1}$. Any such path passes through n_1+n_2+1 lattice points and in each point we may now insert $(1,0)$-steps. If there are n_3 such steps, then this can be done in $\binom{n_1+n_2+n_3}{n_3}$ different ways. Thus if we set $n=n_1+n_2+n_3$ and $Rn_2-n_1=a$, then

$$\begin{aligned} p_n(0,a) &= \sum_{\substack{n_1+n_2+n_3=n \\ Rn_2-n_1=a}} \binom{n_1+n_2}{n_1}\binom{n}{n_3}\alpha^{n_2}\gamma^{n_1}\beta^{n_3} \\ &= \sum_{Ri\geq a}\binom{-a+i(R+1)}{i}\binom{n}{-a+i(R+1)}\alpha^i\gamma^{Ri-a}\beta^{n+a-i(R+1)} \\ &= \binom{n;\alpha,\beta,\gamma}{n+a}_R, \end{aligned} \tag{10}$$

which provides us with an explicit expression for generalized $R-$trinomial coefficients.

To derive $q_n(m)$ we proceed as in the case of $p_n(0,a)$. Let us omit horizontal steps. Assuming that there are n_1 $(1,-1)$-steps and n_2 $(1,R)$-steps, we find by the generalzed ballot theorem (Mohanty (1979), p. 8), that these steps can be joined together in $\frac{n_1-Rn_2}{n_1+n_2}\binom{n_1+n_2}{n_1}$ different ways to yield a path which terminates with a first passage on the line $y=0$. Any such path passes through $n_1 + n_2 + 1$ lattice points and at those points, except for the last one, we insert horizontal steps, the number of which is say n_3. This can be done in $\binom{n_1+n_2+n_3-1}{n_3}$ different ways. Hence

$$\begin{aligned} q_n(m) &= \sum_{\substack{n_1+n_2+n_3=n \\ n_1-Rn_2=m}} \frac{n_1 - Rn_2}{n_1+n_2} \binom{n_1+n_2}{n_1} \binom{n-1}{n_3} \alpha^{n_2} \gamma^{n_1} \beta^{n_3} \qquad (11) \\ &= \sum_{i \geq 0} \frac{m}{m+i(R+1)} \binom{m+i(R+1)}{i} \binom{n-1}{n-m-i(R+1)} \alpha^i \gamma^{m+Ri} \beta^{n-m-i(R+1)} \\ &= \frac{m}{n} p_n(0,-m) \\ &= \frac{m}{n} \binom{n; \alpha, \beta, \gamma}{n-m}_R, \end{aligned}$$

which checks with the analytically derived result. For computational purposes it is important to note that the convolution of $q_n(m)$ and $p_n(0,k)$, which occurs in the definition of the zero avoiding probabilities $s_n(m,k)$, may be simplified further to give the following expression:

$$q_n(m) * p_n(0,k) = \qquad (12)$$

$$= \sum_{i \geq 0} \sum_{Rj \geq k} \frac{m}{m+i(R+1)} \binom{m+i(R+1)}{i} \binom{-k+j(R+1)}{j} \binom{n}{m-k+(i+j)(R+1)} \times$$

$$\alpha^{i+j} \gamma^{m_k+R(i+j)} \beta^{n-m+k-(i+j)(R+1)},$$

which follows from (10) and (11) and the identity

$$\sum_{\nu=1}^{n} \binom{\nu-1}{a} \binom{n-\nu}{b} = \binom{n}{a+b}.$$

During an idle period the process Q_n behaves like a Bernoulli process with success probability α, thus the duration of an idle period, which ends with an arrival, has a negative binomial distrtibution which we denote by

$$g_n(a) = \binom{n-1}{a-1} \alpha^a (1-\alpha)^{n-a}.$$

The number of arrivals during an idle period has a binomial distributions, in particular, we set

$$h_n(a) = \binom{n}{a}\alpha^a(1-\alpha)^{n-a}.$$

We may now prove the first main result:

Theorem 1. *For $m, k > 0$, we have:*

$$\begin{aligned} & P(Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \sigma) = \\ = \ & s_n(m,k) + \sum_{i\geq 1} g_n(ci) * [s_n(m+Ni,k) - s_n(m+N(i-1),k-N)]\,; \end{aligned} \tag{13}$$

For $0 \leq k < c$, we have:

$$P(Q_n = kR, Z_n = \omega | Q_0 = m, Z_0 = \sigma) \ = \ \sum_{i\geq 0} q_n(m+Ni) * h_n(k+ci); \tag{14}$$

For $0 \leq m < c$ and $k \geq 1$, we have:

$$\begin{aligned} & P(Q_n = k, Z_n = \sigma | Q_0 = mR, Z_0 = \omega) = \\ = \ & g_n(c-m) * s_n(N,k) + \\ + \ & \sum_{i\geq 1} g_n(c(i+1)-m) * [s_n(N(i+1),k) - s_n(Ni,k-N)]\,; \end{aligned} \tag{15}$$

For $0 \leq k, m < c$, we have:

$$P(Q_n = kR, Z_n = \omega | Q_0 = mR, Z_0 = \omega) \ = \ \sum_{i\geq 0} q_n(Ni) * h_n(ci+k-m). \tag{16}$$

Proof. The event $\{Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \sigma\}$ may arise in two mutually exclusive ways: (a) the server has never been idle in the interval $(0,n)$, (b) there has been at least one completed idle period. In case (a) the probability of this event is $s_n(m,k)$.

To deal with case (b) assume that there have been $i \geq 1$ idle periods. Consider now the following mapping (see figure 2):

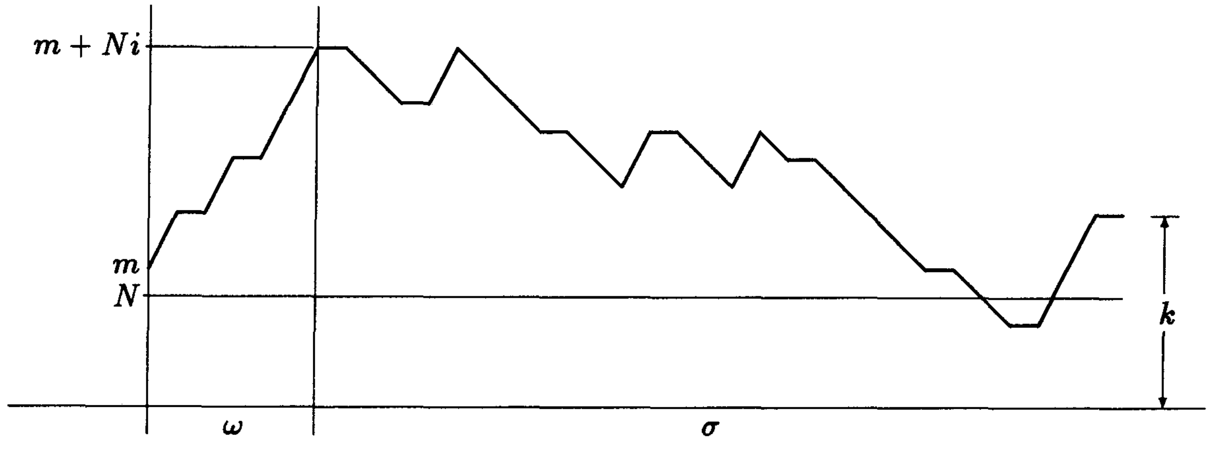

Figure 2.

In the path representing Q_n (see Figure 1), we cut out the i segments of idle periods. Next we form a new path by putting these segments of idle periods at the beginning and then concatenating the remaining segments of busy periods in their original order. The resulting path consists of two parts. The first part is a segment leading from height m to $m + Ni$ without going down. Clearly its probability is $g_n(ci)$. For evaluating the probability of the second part, we note that this path segment leads from height $m + Ni$ to a terminal point with height k, without touching or crossing the line $y = 0$. However, it has to touch or cross the line $y = N$, since the first point, where this happens, marks the beginning of the last busy period. This requirement will be automatically satisfied, if $k \leq N$. If $k > N$, observe that the set of all paths leading from height $m + Ni$ to height k, which do not touch or cross the line $y = 0$, is the disjoint union of the set of paths with same origin and terminus, which do not touch or cross $y = N$ and those paths, which do touch or cross this line. Hence the required probability is given by

$$s_n(m + Ni, k) - s_n(m + N(i - 1), k - N).$$

(It may be realized that the subscript "n" strictly speaking, is not appropriate. However, since the expression is used in a convolution formula, we intend to use the same in the rest of the paper without ambiguity.) It is easily verified that the mapping introduced above is a bijection, and therefore the proof of (13) is complete.

To prove (14) we observe that after removing the idle periods there are $i \geq 0$ completed busy periods and one left censored busy period with m initial customers. As before construct a new path by concatenating first the completed idle periods and then the busy periods without changing their order. In the resulting path the segment due to idle periods has probability $g_n(ci)$. However, the probability of the segment of busy periods, which starts at height $m+Ni$ and terminates with a first passage through $y=0$, is $q_n(m+Ni)$. During the right censored idle period there were k arrivals, the probability of this event being $h_n(k)$. The result follows by observing that $g_n(ci) * h_n(k) = h_n(k+ci)$.

Formula (15) follows by observing that

$$\begin{aligned} P(Q_n = k, Z_n = \sigma | Q_0 = mR, Z_0 = \omega) &= \\ = \quad g_n(c-n) * P(Q_n = k, Z_n = \sigma | Q_0 = N, Z_0 = \sigma). \end{aligned}$$

The proof of (16) follows the same arguments as we used in the proof of (14).

■

Remark. In the derivation of Theorem 1 we only used the renewal properties of the process (Q_n, Z_n). If the service time has a general distribution, then the basic decomposition given above will continue to hold, provided the process X_n is suitably adapted (see for example Prabhu chapt. 3).

3 Batch departures

In this section we will discuss a discrete time queueing process specified in the following way:

- arrival process: customers arrive one at a time according to a Bernoulli process with success probability α.
- service mechanism: a single server serves customers in batches of a fixed size $R \geq 1$. The service times, which are independent of arrival times are i.i.d. random variables having a geometric distribution with parameter γ.
- queue discipline: the batches are served on first-come-first-served basis.

It may be realized that the above process is the discrete time analogue of $E_R/M/1$. As earlier we assume N-policy strategy, which in the present case is as follows:

whenever the queue length becomes less than R, the service is turned off and becomes active as soon as there are N ($N \geq R$) customers in the system.

Again Q_n denotes the number of customers in the system at time n, including those which are in service at time n. Z_n denotes the state of the system, which may be either idle, i.e. ($Z_n = \omega$) or busy, i.e. ($Z_n = \sigma$). It follows from the above assumptions, that the vector process (Q_n, Z_n) is a Markov chain.

The queue length process Q_n may be represented by a lattice path in the $(x,y)-$ plane, having step set $S = \{(1,1),(1,0),(1,-R)\}$, where $(1,1)$ represents an arrival, $(1,-R)$ a batch departure and $(1,0)$ the event {no change}. A typical path is illustrated in figure 3 below.

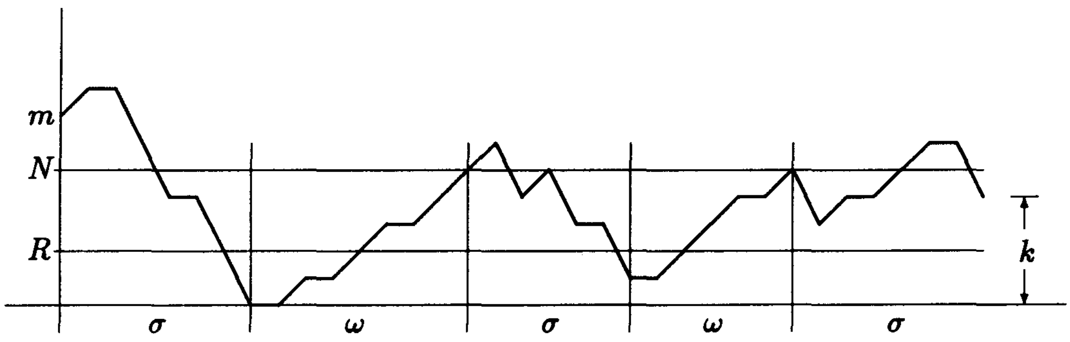

Figure 3.

We define, similar to the previous section, a discrete time random walk $X_n^\star$, which has state space the set of all integers and one-step transition probabilities

$$\begin{aligned} P(X_n^\star = k - R | X_{n-1}^\star = k) &= \gamma \\ P(X_n^\star = k | X_{n-1}^\star = k) &= \beta \\ P(X_n^\star = k + 1 | X_{n-1}^\star = k) &= \alpha. \end{aligned}$$

For the process $X_n^\star$, we retain all our previous notations, each with an asterisk. The random walk $X_n^\star$ is dual to the process X_n defined in section 2, in the sense that $X_i^\star = X_{n-i+1}, i = 0, 1, \ldots, n$, i.e. we are looking at the path from the end. In that case,

$$\begin{aligned} p_n^\star(m,k) &= P(X_n^\star = k | X_0^\star = m) \\ &= \bar{p}_n(k,m) \end{aligned} \tag{17}$$

and

$$\begin{aligned} s_n^\star(m,k) &= P(X_n^\star = k, X_i^\star > 0, 0 \le i \le n | X_0^\star = m) \\ &= \bar{s}_n(k,m) \\ &= \bar{p}_n(k,m) - \bar{q}_n(k) * \bar{p}_n(0,m), \end{aligned} \tag{18}$$

where $\bar{p}_n, \bar{q}_n$ and $\bar{s}_n$ are derived from p_n, q_n and s_n given in (1), by interchanging the probabilities α and γ.

To derive the distribution of the length of a busy period, we note, that a busy period terminates whenever there are less than R customers in the queue. Let us denote by $T_{m,i}$ the length of a busy period initiated by $m \ge R$ customers, leaving $i, (0 \le i < R)$ customers behind. Then we have

Lemma 1.

$$P(T_{m,i} = n) = \gamma \bar{s}_{n-1}(i+1, m-R+1) \tag{19}$$

Proof. During a busy period the transitions of Q_n and $X_n^\star$ coincide. Moreover, for a busy period to terminate at the nth step leaving $i, (0 \le i < R)$ customers behind, it is necessary that $X_{n-1}^\star = R + i$, and the path has not crossed the line $y = R$ so far. Thus

$$\begin{aligned} P(T_{m,i} = n) &= \gamma P(X_{n-1}^\star = R+i, X_j^\star > R-1, 0 \le j < n | X_0^\star = m) \\ &= \gamma P(X_{n-1}^\star = i+1, X_j^\star > 0, 0 \le j < n | X_0^\star = m-R+1) \\ &= \gamma s_{n-1}^\star(m-R+1, i+1). \end{aligned}$$

The result follows from the duality result (18).

■

If we let T_m denote the duration of a busy period, irrespective of the number of customers left behind, then we obtain

$$P(T_m = n) = \gamma \sum_{i=0}^{R-1} \bar{s}_{n-1}(i+1, m-R+1),$$

the discrete time version of a result which may be found in Keilson (1964), who derived the continuous time result by solving a Hilbert problem.

We are now in a position to prove the main result of this section:

Theorem 2. *For $k, m \geq R$ we have:*

$$P(Q_n = k, Z_n = \sigma | Q_n = m, Z_0 = \sigma) = \tag{20}$$

$$= \bar{s}_n(k - R + 1, m - R + 1) + \gamma \bar{s}_n(k - R + 1, N - R + 1) *$$

$$* \sum_{a>0} g_n(a) * \bar{s}_{n-1}(R, m + a - N)$$

For $k \geq R$ and $m < N$ we have:

$$P(Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \omega) = \tag{21}$$

$$= g_n(N - m) * \bar{s}_n(k - R + 1, N - R + 1)$$

$$+ \gamma \bar{s}_n(k - R + 1, N - R + 1) * \sum_{a>0} g_n(a + N - m) * \bar{s}_{n-1}(R, a);$$

For $m \geq R, 0 \leq k < N$ we have:

$$P(Q_n = k, Z_n = \omega | Q_0 = m, Z_0 = \sigma) = \tag{22}$$

$$= \gamma \sum_{i=0}^{k} \sum_{a \geq 0} \bar{s}_{n-1}(i + 1, m + a - R + 1) * h_n(k - i + a);$$

For $0 \leq k, m < N$ we have:

$$P(Q_n = k, Z_n = \omega | Q_0 = m, Z_0 = \omega) = \tag{23}$$

$$= \gamma \sum_{i=0}^{k} \sum_{a \geq 0} \bar{s}_{n-1}(i + 1, N + a - R + 1) * h_n(k - i + a + N - m).$$

Proof. As in Theorem 1 we consider two cases: (i) the server is never idle, (ii) the server is idle at least once. Let a denote the total number of arrivals during all completed idle periods. Clearly there will be no idle period, if and only if $a = 0$. Since a busy period terminates whenever the path of Q_n crosses the line $y = R$ downwards, the probability of (i) is $P(X_n^\star = k, X_i^\star > R - 1, 0 \leq i \leq n | X_0^\star = m)$, which equals $s_n^\star(m - R + 1, k - R + 1)$ and this gives the first term in (20).

Consider now the case $a > 0$, which implies that there is at least one completed idle period. We will construct a bijective transformation which maps sample paths of Q_n into a set of paths, whose probabilities can be given in relatively simple form in terms of the probabilities $\bar{s}_n(m,k)$.

This mapping is illustrated in figure 4.

Figure 4.

First, we cut out the segments due to idle periods and concatenate them to a path leading from height m to height $m+a$. At the terminal point of this path we append the concatenation of successive busy periods, yielding a path segment with the following properties: it starts at height $m+a$ and ends at height k. The point at which the path coming from height $m+a$ reaches the line $y=N$ for the first time, marks the beginning of the last busy period. It is clear that the path is not allowed to cross the line $y=R$. The problem of determining the probability of such a path is now a bit more complicated than in Theorem 1, since there are paths which cross $y=N$ from above in non lattice points. So paths, which cross the line $y=N$ for a first time in a non-lattice point are inadmissable and have to be excluded from further consideration. Consider now the segment of an admissable path from height $m+a$ to the point, where the last busy period begins. Its probability is conveniently derived by means of the basic process $X_n^\star$, in particular we find that this probability equals (see the remark on the subscript n in the proof of Theorem 1)

$$P(X_n^\star = N, X_i^\star > N, 0 \le i < n | X_0^\star = m + a),$$

which in turn equals

$$\begin{aligned}
&\gamma P(X^\star_{n-1} = N + R, X^\star_i > N, 0 \le i < n | X^\star_0 = m + a) = \\
&= \gamma P(X^\star_{n-1} = R, X^\star_i > 0, 0 \le i < n | X^\star_0 = m + a - N) \\
&= \gamma s^\star_{n-1}(m + a - N, R) \\
&= \gamma \bar{s}_{n-1}(R, m + a - N).
\end{aligned}$$

The tail segment due to the last unfinished busy period has probability

$$\begin{aligned}
P(X^\star_n = k, X^\star_i > R - 1, 0 \le i \le n | X^\star_0 = N) &= s^\star_n(N - R + 1, k - R + 1) \\
&= \bar{s}_n(k - R + 1, N - R + 1).
\end{aligned}$$

Taking into account that there have been $a > 0$ arrivals during idle periods, the result follows.

To prove (21) we note that the first idle period requires $N - m$ arrivals to terminate. Thus the required probability is

$$g_n(N - m) * P(Q_n = k, Z_n = \sigma | Q_0 = N, Z_0 = \sigma),$$

and the result follows by using (20). To prove (22) suppose, the last idle period, which is still in progress at time n was initiated by $i, 0 \le i < R$ customers. We observe that after cutting out segments due to idle periods we are left with a path segment having the same distribution as $T_{m+a,i}$. The total number of arrivals during all completed idle periods has probability $g_n(a)$, the arrivals during the last idle period have probability $h_n(k - i)$ and since

$$g_n(a) * h_n(k - i) = h_n(a + k - i),$$

the result follows after summation on i.

In a similar way (23) is proved, observing that the initial idle period requires $N - m$ arrivals to terminate.

In order to complete the proof, we have to show, that the mapping described above is a bijection. The transformed path consists of two parts: (i) from height m to $m + a$ without any down step, which represents all idle periods, and (ii) from $m + a$ to the end, which represents all busy periods. In the second part (see Figure 4) consider the segment starting from $m+a$ and reaching the first lattice point after crossing downward the line $y = m+a-R$ for the first time. This corresponds to the first busy period in Figure 3. Let the lattice point marking the end of the first busy period have height $N+a-a_1, (N-R \le a_1 \le N)$. Then the number of arrivals in the first idle period is a_1. Take a segment from (i) of height a_1 which corresponds to the first idle period and join this segment after the segment corresponding to the first busy period. For the next busy period take the segment in (ii) from height $N + a - a_1$ to the first lattice point after crossing downward the line $y = R + a - a_1$ for the first time. Let this point have height $N + a - a_1 - a_2, (N - R \leq a_2 \leq N)$. This segment is nothing but the second busy period and can be joined after the first idle period. Now take from (i) a segment from $m + a_1$ to $m + a_1 + a_2$ (which is the second idle period) and join at the end of the second busy period. This procedure of concatenation will finally restore the original path. Thus the mapping is bijective.

■

4 Waiting times

In this section we discuss briefly the distribution of the waiting time W_n of a customer arriving at time n. The derivation of the distribution of W_n poses no particular problem, once the queue length distributions are known.

Consider the case of batch arrivals first and assume that a batch arriving at time n finds the server busy with $k > 0$ customers waiting including the one being served. For simplicity we confine ourselves to the case $Z_0 = \sigma$. Then we have in virtue of Theorem 1:

$$P(W_n = w | Z_n = \sigma, Z_0 = \sigma, Q_0 = m) = \tag{24}$$

$$= \sum_{k>0} P(Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \sigma) b_w(k),$$

where

$$b_n(a) = \binom{n-1}{a-1} \gamma^a (1-\gamma)^{n-a}.$$

If the server is idle at time n, then we get:

$$P(W_n = w | Z_n = \omega, Z_0 = \sigma, Q_0 = m) = \tag{25}$$

$$= \sum_{0 \leq k < c} P(Q_n = kR, Z_n = \omega | Q_0 = m, Z_0 = \sigma) \left[g_w(c-k-1) * b_w(kR) \right].$$

Similarly through Theorem 2 the case where customers are served in batches yields:

$$P(W_n = w | Z_n = \sigma, Z_0 = \sigma, Q_0 = m) = \tag{26}$$

$$= \sum_{k>0} \sum_{i=0}^{R-1} P(Q_n = kR + i, Z_n = \sigma | Q_0 = m, Z_0 = \sigma) b_w(k),$$

and

$$P(W_n = n | Z_n = \omega, Z_0 = \sigma, Q_0 = m) = \tag{27}$$
$$= \sum_{A} P(Q_n = kR + i, Z_n = \omega | Q_0 = m, Z_0 = \sigma) \left[g_w(N - kR - i - 1) * b_w(k) \right],$$

where in the summation above the index set $A = \{(i,k) : 0 \leq i < R, 0 \leq kR + i < N\}$.

5 Steady state results

If customers arrive in batches, then the steady state distribution may be found by renewal theory.

Let B denote the length of a busy cycle. Its pgf. is given by

$$\sum_{n \geq 0} P(B = n) t^n = W^c(t) \xi^N(t),$$

where $W(t)$, the pgf. of the interarrival time distribution, is given by

$$W(t) = \frac{\alpha t}{1 - t(1 - \alpha)},$$

and $\xi(t)$ is the root (8). If $t \uparrow 1$, then $W(t) \to 1$ and $\frac{d}{dt} W(t) \to 1/\alpha$. Furthermore

$$\frac{d}{dt} \xi(t) = \frac{\gamma + \beta \xi(t) + \alpha \xi^{R+1}(t)}{1 - t(\beta + (R+1)\xi^R(t))} \to \frac{1}{\gamma - R\alpha},$$

and as a result $\xi(t) \to 1$ if $R\alpha \leq \gamma$. Hence the expected length of a busy cycle is found to be

$$\begin{aligned} E(B) &= \left[N \xi'(t) \xi^{N-1}(t) W^c(t) + c \xi^N(t) W^{c-1}(t) W'(t) \right]_{t=1} \\ &= \frac{N}{R} \frac{\rho^{-1}}{\gamma - R\alpha}, \qquad \rho = \alpha / \gamma. \end{aligned}$$

Thus the event $\{Q_n = k, Z_n = \sigma\}$ is recurrent if $\rho < 1/R$, and we find by the Erdös-Feller-Pollard theorem (Feller (1968), p. 254):

$$\lim_{n\to\infty} P(Q_n = k, Z_n = \sigma) \;=\; \frac{1}{E(B)} \sum_{n\geq 0} s_n(N,k).$$

It remains to determine $\sum s_n(N,k)$. The pgf. of $s_n(N,k)$ is given by

$$\sum_{n\geq 0} t^n s_n(N,k) = ([z^{k-N}] - \xi^N(t)[z^k])\frac{z}{z - t(\gamma + \beta z + \alpha z^{R+1})}.$$

If we let $t \uparrow 1$, we find

$$\begin{aligned}\sum_{n\geq 0} s_n(N,k) &= ([z^{k-N}] - [z^k])\frac{z}{z - (\gamma + \beta z + \alpha z^{R+1})} \\ &= \sum_{\nu\geq 0}\left[\binom{\nu;\alpha,\beta,\gamma}{\nu + k - N}_R - \binom{\nu;\alpha,\beta,\gamma}{\nu + k}_R\right].\end{aligned}$$

Thus if $\rho < 1/R$,

$$\lim_{n\to\infty} P(Q_n = k, Z_n = \sigma) = \frac{R}{N}\alpha(1 - \rho R)\sum_{\nu\geq 0}\left[\binom{\nu;\alpha,\beta,\gamma}{\nu + k - N}_R - \binom{\nu;\alpha,\beta,\gamma}{\nu + k}_R\right]. \qquad (28)$$

Similary we determine

$$\begin{aligned}\lim_{n\to\infty} P(Q_n = kR, Z_n = \omega) &= \frac{1}{E(B)}\sum_{n\geq 0} h_n(k) \qquad (29)\\ &= \frac{R}{N}(1 - \rho R),\end{aligned}$$

which is surprisingly independent of k.

Unfortunately we are not able to give an expression for the steady state solution of the queueing model with batch service, unless $N = R$. In this case, however, the solution is already known (Jackson and Nickols (1956)).

6 Continuous time results

Discrete time queueing models may also serve as a basis for the discussion of queueing processes in continuous time, where discrete and continuous time models are linked by an appropriate passage to the limit. This limiting procedure will be discussed in more detail in this section, where we confine ourselves to Markovian queueing processes. The reason for this restriction is, that the analysis of convergence properties of non Markovian processes is considerably more involved than it is in the Markovian case. On the other hand this restriction is not too severe, since it is known, that in many non Markovian queueing models the Markov property may be retained, either by introducing supplementary variables or by the construction of Markov processes by embedding at renewal points.

So far we have considered queueing processes Q_n at discrete time marks $0, 1, \ldots, n$. Now let us change the unit in which time is measured to $\Delta = t/n$. Then the new time marks $0, \Delta, 2\Delta, \ldots$ form a partition of the interval $[0, t]$ into n slots of equal length Δ.

Consider now an aperiodic Markov chain Q_n with denumerable state space and one step transition matrix $\mathbf{P}$, which has entries p_{ij}, and define

$$\begin{aligned} p_{ij} &= \frac{t}{n}\lambda_{ij} \qquad (i \neq j) \\ p_{ii} &= 1 - \frac{t}{n}\sum_{j\neq i}\lambda_{ij} \end{aligned}$$

where the constants $\lambda_{ij} \geq 0$. Then the transition matrix $\mathbf{P} = \mathbf{P}(t/n)$ may be written as

$$\mathbf{P}(t/n) = \mathbf{I} + \frac{t}{n}\mathbf{W},$$

where $(\mathbf{W})_{ij} = \lambda_{ij}$ if $i \neq j$ and $(\mathbf{W})_{ii} = -\sum_{j\neq i}\lambda_{ij}$.

Now let $\mathcal{L}$ denote the set of all numerical sequences $\{u_n\}$ and introduce the norm $\|u\| = \sup|u_n|$, thus $\mathcal{L}$ is a Banach space and the matrices $\mathbf{P}$ and $\mathbf{W}$, which are in the general case of infinite dimension, act as linear operators on $\mathcal{L}$. Let us assume further, that the operator $\mathbf{W}$ is bounded. It follows from a standard result in functional analysis (Yosida (1968), p. 269) that

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}^n(t/n) &= \lim_{n\to\infty}(\mathbf{I} + \frac{t}{n}\mathbf{W})^n \qquad (30) \\ &= \exp[\mathbf{W}t], \end{aligned}$$

which is the transition semi group of a continuous time Markov process having infinitesimal generator **W**. The result given above in turn implies weak convergence of the process $Q_n = Q_n(t/n)$ to a Markov jump process $Q(t)$ (for a rigorous discussion see Ethier and Kurtz, (1984), pp. 167-169). So not only the finite dimensional distributions of Q_n have a well defined limit, even the sample paths of $Q_n(t/n)$ converge with respect to the Skorohod metric to the sample paths of $Q(t)$.

In the most general case it may be quite difficult to find the limiting expression (30). However, in some queueing situations the structure of the one-step transition matrix **P** is relatively simple, so that the transient solutions in discrete time can be given. In this case the passage to the limit will usually pose no particular problem.

Let us now return to discrete time queues with batch arrivals we have discussed in section 2. We set $\alpha = \lambda t/n$ and $\gamma = \mu t/n$. It then follows from the discussion above that the vector process $(Q_n(t/n), Z_n)$ converges weakly to the Markov process $(Q(t), Z(t))$, which can be identified as the N-policy $M/E_R/1$ queueing model. The basic result is given in the following

Lemma 2. *If $\alpha = \lambda t/n, \gamma = \mu t/n$ and λ, μ and t are fixed, then*

$$\lim_{n\to\infty} \binom{n; \alpha, \beta, \gamma}{n+k}_R = e^{-(\lambda+\mu)t} \rho^{k/(1+R)} I^R_{-k}(2wt), \tag{31}$$

where $w = \lambda^{\frac{1}{1+R}} \mu^{\frac{R}{1+R}}$ and $I^R_j(x)$ denotes the Luchak function (see for instance Prabhu (1965), p.98):

$$I^R_j(x) = \sum_{\nu\geq 0} \frac{(x/2)^{j+\nu(R+1)}}{\nu!(j+R\nu)!}.$$

Proof. We recall that the generalized R-trinomial coefficients are the coefficients of the powers of z in polynomial expansion of $(\gamma + \beta z + \alpha z^{R+1})^n$. Thus

$$\begin{aligned} \binom{n; \alpha, \beta, \gamma}{n+k}_R &= [z^k](\gamma/z + \beta + \alpha z^R)^n \\ &= [z^k] \left(\frac{\mu t}{nz} + 1 - \frac{t}{n}(\lambda + \mu) = \frac{\lambda t z^R}{m} \right)^n. \end{aligned}$$

By the continuity theorem for Laplace-Stieltjes transforms we have

$$\begin{aligned}
\lim_{n\to\infty} \binom{n;\alpha,\beta,\gamma}{n+k}_R &= [z^k]e^{-(\lambda+\mu)t}\exp[t(\lambda z^R+\mu/z)] \\
&= e^{-(\lambda+\mu)t}\sum_{\nu\geq 0}\frac{t^{\nu(R+1)-k}\lambda^{\nu}\mu^{R\nu-k}}{\nu!(R\nu-k)!} \\
&= e^{-(\lambda+\mu)t}\rho^{k/(1+R)}I^R_{-k}(2t\lambda^{1/(1+r)}\mu^{R/(1+R)}).
\end{aligned}$$

■

It will be convenient to denote the limit (31) of $p_n(m,k)$ by $p_t(m,k)$. As an immediate consequence of Lemma 2 and (13) we have

$$\begin{aligned}
\lim_{n\to\infty} q_n(m)\frac{n}{t} &= q_t(m) \\
&= \lim_{n\to\infty} p_n(0,-m) \\
&= \frac{m}{t}p_t(0,-m) \\
&= \frac{m}{t}e^{-(\lambda+\mu)t}\rho^{-m/(1+R)}I^R_m(2wt), \qquad (32)
\end{aligned}$$

yielding the density of the duration of a busy period in continuous time. Because of the factor n/t on the left hand side of (32), this limit has to be understood in the sense $q_n(m)\to q_t(m)\,dt$.

Furthermore we have

$$\begin{aligned}
\lim_{n\to\infty} s_n(m,k) &= s_t(m,k) \\
&= p_t(m,k) - q_t(m) * p_t(0,k) \\
&= e^{-(\lambda+\mu)t}\rho^{\frac{k-m}{1+R}}\left[I^R_{m-k}(2wt) - m\int_0^t I^R_m(2ws)I^R_{-k}(2w(t-s))\frac{ds}{s}\right]. \qquad (33)
\end{aligned}$$

Finally we note that

$$\lim_{n\to\infty} g_n(a) = g_t(a)\,dt = \frac{\lambda^a t^{a-1}e^{-\lambda t}}{(a-1)!}\,dt,$$

and

$$\lim_{n\to\infty} h_n(a) = \frac{e^{-\lambda t}(\lambda t)^a}{a!}.$$

Lemma 2 together with (32) and (33) provide us with the basic tools to give the continuous time counterpart of Theorem 1:

Theorem 3. *For $m, k > 0$:*

$$\begin{aligned}
&\lim_{n\to\infty} P(Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \sigma) = \qquad (34)\\
&= \; P(Q(t) = k, Z(t) = \sigma | Q(0) = m, Z(0) = \sigma)\\
&= \; s_t(m,k) + \sum_{i\geq 1}[g_t(ci)\,dt] * [s_t(m+Ni,k) - s_t(m+N(i-1),k-N)]
\end{aligned}$$

In the case of batch service we proceed in an analogous manner. The bivariate Markov chain $(Q_n(t/n), Z_n)$ converges weakly to the continuous time process $(Q(t), Z(t))$, which may be identified as the queue length process of $E_R/M/1$ with N-policy.

We recall that the transient solution has been formulated in terms of the probabilities $\bar{s}_n, \bar{p}_n$ and $\bar{q}_n$, which we derived from s_n, p_n and q_n by interchanging the probabilities α and γ. Similary we interchange λ and μ to get $\bar{s}_t, \bar{p}_t$ and $\bar{q}_t$ from s_t, p_t and q_t.

Theorem 4. *For $m, k \geq R$:*

$$\begin{aligned}
&\lim_{n\to\infty} P(Q_n = k, Z_n = \sigma | Q_0 = m, Z_0 = \sigma) = \qquad (35)\\
&= \; P(Q(t) = k, Z(t) = \sigma | Q(0) = m, Z(0) = \sigma)\\
&= \; \bar{s}_t(k-R+1, m-R+1) +\\
&= \; \mu \sum_{a>0}[g_t(a)\,dt] * \bar{s}_t(k-R+1, N-R+1) * [\bar{s}_t(R, m+a-N)\,dt]
\end{aligned}$$

7 References

[1] Bhat N.U. (1964) On single server bulk queueing processes with binomial input. *Oper. Res.*, **12**, 527-533.

[2] Bhat U.N. (1968) *A Study of the Queueing Systems M/G/1 and GI/M/1*, Lecture Notes in Operations Research and Mathematical Economics. Springer, Berlin.

[3] Böhm W., Mohanty S.G. (1990) The transient solution of M/M/1 queues with (M,N)-policy. Paper presented at the 2nd Conference on Lattice Path Combinatorics and Applications, McMaster University 1990.

[4] Chaudhry M.L., Templeton J.G.C. (1983) *A First Course in Bulk Queues*, John Wiley and Sons, New York.

[5] Dafermos S., Neuts M.F. (1973) A single server in discrete time. *Cahiers du Centre de Recherche Opérationelle*, **13**, 23-40.

[6] Ethier S.N., Kurtz T.G. (1986) *Markov Processes. Characterization and Convergence.* John Wiley and Sons, New York.

[7] Falin G. I. (1989) Approximation of synchronous data transmission systems with a small time cycle. *Automation and Remote Control* **4**, 523-528.

[8] Feller W. (1968), *An Introduction to Probability Theory and Its Applications*, 3nd. ed., John Wiley and Sons, New York.

[9] Hayman D.P. (1968), Optimal operating policies for M/G/1 queueing systems. *Oper. Res.*, **16**, 362-382.

[10] Hunter J.J. (1983) *Mathematical Techniques of Applied Probability, Vol. 2; Discrete Time Models: Techniques and Applications*, Academic Press, New York.

[11] Jackson R.R.P. and Nickols D.G. (1956) Some equilibrium results for the queueing process $E_k/M/1$. *J. Roy. Soc.*, **B18**, 275-279.

[12] Kanwar Sen, Jain J.L. (1990) Combinatorial approach to Markovian queueing models. Paper presented at the 2nd Conference on Lattice Path Combinatorics and Applications, McMaster University 1990.

[13] Kanwar Sen, Jain J.L., Gupta J.M. (1990) Lattice path approach to transient solution of M/M/1 queues with (0,K) control policy. Paper presented at the 2nd Conference on Lattice Path Combinatorics and Applications, McMaster University 1990.

[14] Keilson J. (1964) Some comments on single server queueing models and some new results. *Proc. Camb. Phil. Soc.*, **60**, 237-251.

[15] Kobayashi H. (1983) Stochastic modeling: queueing models; in G. Louchard and G. Latouche (eds.): *Probability Theory and Computer Science*, Academic Press, New York.

[16] Meilsing T. (1958) Discrete-time queueing theory. *J. Opns. Res. Soc. Am.* , **6**, 96-105.

[17] Mohanty S.G. (1979) *Lattice Path Counting and Applications.* Academic Press, New York.

[18] Mohanty S.G., Panny W. (1990a) A discrete time analogue of the M/M/1 queue and the transient solution: a geometric approach. *Sankya, Ser. A.*, 364-370.

[19] Mohanty S.G., Panny W. (1990b) A discrete time analogue of the M/M/1 queue and the transient solution: an analytic approach. *Proceedings of the 3rd. Hungarian Colloquium on Limit Theorems in Probability and Statistics*, P. Révész ed., Amsterdam, North Holland.

[20] Neuts M.F. (1973) The single server queue in discrete time-numerical analysis I. *Nav. Res. Log. Quart.*, **20**, 297-304.

[21] Neuts M.F. (1989) *Structured Stochastic Matrices of M/G/1 Type and Their Application*, Marcel Dekker, New York.

[22] Panny W. (1984) *On the Maximum Deviation of Lattice Paths.* Hain/Hanstein/ Königstein.

[23] Prabhu N.U. (1965) *Queues and Inventories*, John Wiley and Sons, New York.

[24] Takács L. (1962) *Introduction to the Theory of Queues.* New York, Oxford University Press.

[25] Takács L. (1975) Combinatorial and analytic methods in the theory of queues. *Adv. Appl. Prob.*, **7**, 607-635.

[26] Yosida K. (1968) *Functional Analysis.* 2nd ed. Springer, New York.