THE EFFECT OF LONG RANGE
DEPENDENCE IN A SIMPLE
QUEUEING MODEL

by

Sidney Resnick
and Gennady Samorodnitsky

Sidney Resnick and Gennady Samorodnitsky were partially supported by NSF Grant DMS-94-00535 at Cornell University. Sidney Resnick also received partial support from the United States-Israel Binational Science Foundation (BSF) Grant No. 89-00323 and 92-000324 and Gennady Samorodnitsky received some support from the United States-Israel Binational Science Foundation (BSF) Grant No. 92-00074/1.
THE EFFECT OF LONG RANGE DEPENDENCE IN A SIMPLE QUEUEING MODEL

SIDNEY RESNICK AND GENNADY SAMORODNITSKY

Cornell University

ABSTRACT. We offer some specific information about how long range dependence can influence the characteristics of a simple queue. We take the analogue of the G/M/1 queue except that the input stream is altered to exhibit long range dependence. The equilibrium queue size and equilibrium waiting time distributions each have heavy tails. By suitably selecting the parameters of the inputs, the queue size or waiting time can be made to possess infinite variance and even infinite mean. Some simulations dramatically illustrate the potential for undetected long range dependence to significantly alter the queueing behavior compared to what is anticipated with traditional inputs.

1. Introduction.

Long range dependence is a property of stationary time series models whose current state has a strong dependency on the remote past. Definitions vary from author to author but a commonly accepted definition in covariance stationary time series is that a process \( \{X_n\} \) has long range dependence if

\[
\sum_{j=1}^{\infty} |\text{corr}(X_0, X_j)| = \infty
\]

(cf. Brockwell and Davis, 1991). A very useful recent reference on long range dependence is Beran, 1992. In contexts outside traditional times series where correlations may not exist or be difficult to compute or to interpret, other measures of dependence may be more meaningful. In this case, long range dependence may refer to a slowly decreasing dependence between blocks of random variables as the time gap between the blocks grows. In our work, one interpretation of long range dependence in a stationary process \( \{T_n\} \) is that the probability

\[
P[T_i \leq r, i = 1, \ldots, n]
\]

decreases slowly to 0 as \( n \uparrow \infty \) for a suitable choice of constant \( r \).

Statistical evidence is mounting that traffic on certain types of data networks may exhibit long range dependence. For instance, Beran et. al (1992) report on the analysis of several data sets representing the traffic seen as a result of transmitting video conference scenes. The data sets are large, sometimes in excess of 50,000 data points and consequently one may expect estimates of the autocorrelation function to be accurate to very large lags. Beran et. al. conclude that long range dependence is likely to be present in the underlying models which gave rise to many of these data sets. A time series plot of a sample video conference data set of size 48497 is given in Figure 1 and under it is a plot of the sample autocorrelation function to lag 1000.

---


Key words and phrases. long range dependence, heavy tails, Poisson process, G/M/1 queue.

Sidney Resnick and Gennady Samorodnitsky were partially supported by NSF Grant DMS-94-00535 at Cornell University. Sidney Resnick also received partial support from the United States-Israel Binational Science Foundation (BSF) Grant No. 89-00323 and 92-000324 and Gennady Samorodnitsky received some support from the United States-Israel Binational Science Foundation (BSF) Grant No. 92-00074/1.
Notice the sample autocorrelation function does not go to zero rapidly as would be characteristic of short memory models.

The presence of long range dependence is difficult to prove statistically. Indeed, several of the data sets analyzed in Beran et. al. (1992) have been previously analyzed by conventional time series means (cf. Heyman et. al. (1992)) and fitted by such garden variety time series models as a stationary autoregression of order 2. Although the autoregressive polynomial roots of such models are typically near the unit circle, which is necessitated by the requirement of modelling a slowly decreasing sample autocorrelation function, such classical time series models do not exhibit long range dependence. The fact that long range dependence is difficult to distinguish from time series models with autoregressive polynomial roots near the unit circle coupled with persistent engineering skepticism about whether long range dependence would materially effect system performance assuming it was indeed present has led us to inquire if the effects of long range dependence can be quantified at least in simple models.

Figure 1.1

One of the difficulties in studying queueing systems with serially correlated input is that there are no commonly accepted paradigms for arrival processes. Specifying the autocorrelation function of a stationary sequence of interarrival times with given marginals is sufficient to fully describe the model only when this sequence is assumed to be Gaussian. Normal interarrival times are unsuitable because of the theoretical possibility of negative outcomes and also because of lack of skewness. Needless to say, the situation is even more difficult when one wants to model long range dependent interarrival times.
A recent paper of Livny, Melamed and Tsiolis (1993) studying the impact of serial correlation on queues, used two methods to simulate an input process. One is the TES method due to Melamed (1991) and studied by Jagerman and Melamed (1992a, b), while the other is the minification technique due to Lewis and McKenzie (1991) which simulates a min-moving average process of order 1 (see also Davis and Resnick, 1989, 1993 for background on such processes). Both techniques produce a sequence of uniform random variables on (0,1) which can be then transformed to any given marginal distribution by the standard application of the inverse distribution function. The autocorrelation functions in both cases have one free parameter, which happens to be the same in the present context as saying that both stationary streams of uniform random variables can be parametrized by their correlations at lag 1 (say). TES and Minification methods produce very different autocorrelation functions, but both of them decrease to zero exponentially fast with the lag. For the same correlation at lag 1, the decrease is faster for the minification method and slower for the TES method.

In Section 2 we propose a simple modification of the G/M/1 queue. We assume we have a single server who serves according to a homogeneous Poisson process with rate $\mu$. We feed this server an arrival stream modeled by a point process whose interpoint distances $T_n$ is a stationary process. Under a mild condition on the sequence $\{T_n\}$, the variable $X_n$, representing the number in the system seen by the $n$th arriving customer has a limit distribution as $n \to \infty$. We interpret this limit distribution as the equilibrium distribution and write $X_n \Rightarrow X_{\infty}$, where $\Rightarrow$ denotes convergence in distribution and $X_{\infty}$ represents a random variable with the equilibrium distribution.

In Section 3 we construct a class of examples where the interarrival times $\{T_n\}$ have long range dependence and this leads to the property that the distribution of $X_{\infty}$ has a heavy right tail. Contrast this with the classical stable G/M/1 queue where the stationary queue length distribution is geometric (Asmussen, 1987, page 204). So it appears that the presence of long range dependence can indeed have a dramatic effect on system performance.

Section 4 presents a simulation of the $\{T_n\}$ process discussed in Section 3 and shows graphically the effect of the long range dependence by comparing system characteristics of a model with such an input compared with a standard model without the highly dependent input.

2. A reversible process with negative mean drift.

Consider a homogeneous Poisson process $N$ with rate $\mu$. We think of $N$ as a random measure on the Borel subsets of $[0, \infty)$. For an interval $I$, $N(I)$ represents the number of service completions in time interval $I$ by a single server serving at rate $\mu$. Let $\{T_n, n \geq 1\}$ be a stationary sequence of non-negative random variables. Suppose $E T_n = \lambda^{-1}$ and set

$$S_0 = 0, \quad S_n = T_1 + \cdots + T_n, \quad n \geq 1$$

and think of $S_n$ as the time of arrival of the $n$th customer. Setting $S_0 = 0$ is a convenience which could easily be eliminated if desired. We assume $\{S_n\}$ is independent of $N$. For convenience, set $X_0 = 0$ and define the process $\{X_n, n \geq 0\}$ by the Lindley recursion:

(2.1) $$X_{n+1} = (X_n + 1 - N(S_n, S_{n+1})), n \geq 0,$$

where as usual $x^+ = x$, if $x \geq 0$ and $x^+ = 0$ if $x < 0$. We think of $X_n$ as the number in the system seen by the $n$th arriving customer. Finally define

$$K_{n+1} = 1 - N(S_n, S_{n+1}), n \geq 0,$$

so that (2.1) can be rewritten as

(2.1') $$X_{n+1} = (X_n + K_{n+1}), n \geq 0.$$
Note that
\[ EK_n = 1 - \mu ET_n = 1 - \mu \lambda^{-1}, \]
and it is natural to set
\[ \rho = \frac{\lambda}{\mu}, \]
and think of \(\rho\) as the traffic intensity. Then
\[ EK_n = 1 - \rho^{-1} \]
and
\[ EK_n < 0 \text{ iff } \rho < 1. \]
Then paralleling the classical theory we have the following result.

**Proposition 2.1.** If \(\{T_n\}\) is a reversible, stationary, ergodic process and \(\rho < 1\), we have as \(n \to \infty\)
\[ X_n \Rightarrow X_\infty \overset{d}{=} \left( \sum_{i=1}^{j} K_i \right) \]
where \(\sum_{i=1}^{0} K_i = 0\) and the limit random variable is finite because of the assumption \(EK_i < 0\).

**Proof.** As in the classical case (Asmussen, 1987, page 80; Resnick, 1992, page 514) induction yields
\[ X_n = \max\{X_0 + \sum_{i=1}^{n} K_i, \sum_{i=2}^{n} K_i, \ldots, K_n, 0\}. \]
If the sequence \(\{K_n\}\) is reversible, which means for all integers \(m\)
\[ (K_1, \ldots, K_m) \overset{d}{=} (K_m, \ldots, K_1), \]
then with \(X_0 = 0\)
\[ X_n \overset{d}{=} \max\left\{ \sum_{i=1}^{n} K_i, \sum_{i=1}^{n-1} K_i, \ldots, K_1, 0 \right\} \]
\[ = \max_{j=0}^{n} \left( \sum_{i=1}^{j} K_i \right). \]
If \(EK_i < 0\), then ergodicity of \(\{T_n\}\) and the ergodic theorem implies \(\sum_{i=1}^{n} K_i \sim nEK_1 \to -\infty\) and so as \(n \to \infty\)
\[ \left( \sum_{i=1}^{j} K_i \right) \overset{\text{in probability}}{\to} \left( \sum_{i=1}^{j} K_i \right) < \infty. \]

It only remains to easily verify that \(\{T_n\}\) reversible implies \(\{K_n\}\) is also reversible. We have for any \(m\) and non-negative integers \(k_1, \ldots, k_m\) that
\[ P[K_i = 1 - k_i, i = 1, \ldots, m] \]
\[ = P[N(S_{i-1}, S_i) = k_i, i = 1, \ldots, m] \]
\[ = EP[N(S_{i-1}, S_i) = k_i, i = 1, \ldots, m|T_1, \ldots, T_m] \]
\[ = E \prod_{i=1}^{m} \frac{e^{-\mu T_i (\mu T_i)^{k_i}}}{k_i!} \]
and since \((T_1, \ldots, T_m) \overset{d}{=} (T_m, \ldots, T_1)\) we have this equal to
\[ = E \prod_{i=1}^{m} \frac{e^{-\mu T_{m-i+1}}(\mu T_{m-i+1})^{k_i}}{k_i!} \]
\[ = E \prod_{i=1}^{m} \frac{e^{-\mu T_i}(\mu T_i)^{k_{m-i+1}}}{k_{m-i+1}!} \]
\[ = P[K_i = 1 - k_{m-i+1}, i = 1, \ldots, m] \]
\[ = P[K_m = 1 - k_1, \ldots, K_1 = 1 - k_m]. \quad \Box \]

In Section 3 we will analyze tail behavior of \( X_\infty \) and we will show that for positive constants \( c, r \), it is possible to bound \( P[X_\infty > cn] \) from below by probabilities of the form
\[
\frac{1}{2} P\left[ \frac{S_n}{n} < r \right] \geq \frac{1}{2} P[T_i \leq r, i = 1, \ldots, n].
\]

Thus the right tail probabilities of \( X_\infty \) can be bounded below by left tail probabilities of \( S_n/n \) which may be thought of as the probability that the sample mean is bounded away from the theoretical mean. We will see that such probabilities can decrease to zero very slowly if the \( \{T_n\} \) sequence has a long range dependence property.

3. A family of input processes with long range dependent interarrival times.

In this section we suggest a model for the input process that is parametrized by an infinite sequence of numbers and can have correlations that decrease as slowly or as quickly as one wishes. The model has a dependence structure that is fairly intuitive and is amenable to straightforward computer generation. The basic model generates stationary interarrival times with standard exponential marginals, and these can be transformed to any other desired marginals by the application of the inverse transform method.

To describe the model we need to introduce some terminology and provide some background information. Let \( \mathcal{B} \) be the Borel \( \sigma \)-field on \( \mathbb{R} \), and let \( M \) be an independently scattered infinitely divisible random measure on \( \mathcal{B} \) with control measure
\[
(3.1) \quad \lambda(ds) = ds, \quad s \in \mathbb{R},
\]
and instantaneous Lévy measures
\[
(3.2) \quad q(dx, s) = \frac{e^{-x}}{x} 1(x > 0)dx, \quad x \in \mathbb{R}, \quad s \in \mathbb{R}.
\]

That is, \( M \) is just a stochastic process indexed by Borel sets of finite Lebesgue measure such that if \( A_1, A_2, \ldots \) are pairwise disjoint, then \( M(A_1), M(A_2), \ldots \) are independent (that is what independently scattered means), and if furthermore the union of the \( A_n \)'s above has a finite Lebesgue measure as well, then \( M(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} M(A_n) \) with probability 1. The fact that control measure \( \lambda \) is Lebesgue measure and that the instantaneous Lévy measure \( q \) does not depend on \( s \) implies that the distribution of the random variable \( M(A) \) depends only on the Lebesgue measure of the set \( A \). This is easily seen from the fact that each \( M(A) \) is a non-negative infinitely divisible random variable with Laplace transform
\[
(3.3) \quad E e^{-\alpha M(A)} = \exp\{-\int_0^\infty (1 - e^{-\alpha x}) \nu_{M(A)}(dx)\}, \quad \alpha \geq 0
\]
and Lévy measure \( \nu_{M(A)} \) given by
\[
(3.4) \quad \nu_{M(A)}(dx) = \int_A q(dx, s) \lambda(ds) = |A| \frac{e^{-x}}{x} 1(x > 0)dx.
\]
In particular, $M(A)$ is a Gamma random variable with scale parameter 1, and shape parameter equal to $\lambda(A)$ (cf. Feller, 1971, page 451 for example). If $\lambda(A) = 1$, then $M(A)$ is the standard exponential random variable. Note that $\{M((0, t], t \geq 0)\}$ is a Gamma process and has stationary independent increments. We refer the reader to Rajput and Rosinski (1989) for more information on infinitely divisible random measures and stochastic integrals with respect to these measures.

Another description of the process $M$ is as follows: Let

$$\sum_{t_k \in A} \epsilon(t_k, j_k)$$

be a Poisson random measure (PRM) on $[0, \infty) \times (0, \infty]$ with mean measure $\lambda \times x^{-1}e^{-x}dx$. (Cf. Resnick, 1987.) Then we may define for any $A \in \mathcal{B}$ such that $\lambda(A) < \infty$

$$M(A) = \sum_{t_k \in A} j_k.$$

Our basic input process is a moving average with respect to the random measure $M$, which formally is a stochastic integral of the form

$$\xi_n = \int_{-\infty}^{\infty} f(n+s)M(ds), \ n = 0, 1, 2, \ldots.$$  

(3.5)

Here $f \geq 0$ is a kernel function, which is our degree of freedom in specifying the dependence between the interarrival times. The moving average (3.5) is well defined if and only if

$$\int_{[0, \infty) \times (0, \infty]} (f(t)x \wedge 1) dt \frac{e^{-x}}{x} dx < \infty$$

(3.6)

(cf. Rajput and Rosinski (1989)). In this case $\{\xi_n, n \geq 0\}$ is a (strictly) stationary process (infinitely divisible, in fact), such that each $\xi_n$ is a non-negative infinitely divisible random variable with Laplace transform

$$Ee^{-\alpha \xi_n} = \exp\{-\int_0^{\infty} (1-e^{-\alpha x})\nu_\xi(dx)\}, \ \alpha \geq 0$$

(3.7)

and Lévy measure $\nu_\xi$

$$\nu_\xi(B) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1(xf(s) \in B)q(dx, s)\lambda(ds).$$

(3.8)

Our specific choice of kernel $f$ is now given. Let $\{a_n, n \geq 1\}$ be a nonincreasing sequence of nonnegative numbers such that

$$\sum_{n=1}^{\infty} a_n = 1.$$  

(3.9)

Let $I_j$ denote the interval $(j-1, j-1+a_j)$, $j = 1, 2, \ldots$ and define

$$f(u) = 1(u \in \cup_{j=1}^{\infty} I_j).$$

(3.10)
One readily checks that (3.6) holds, and therefore our input process (3.5) is a stationary process, and we see from (3.8) that each $\xi_n$ has the standard exponential distribution. Note also that we may write

$$\xi_n = \sum_{j=1}^{\infty} M((j-1-n, j-1-n + a_j)),$$

so that $\xi_n$ is a sum of Gamma distributed random variables; recall that

$$M((j-1-n, j-1-n + a_j))$$

is Gamma distributed with shape parameter $a_j$.

The dependence structure of the input process is obviously determined by the numbers $\{a_n, n \geq 1\}$. Since the kernel $f(n+\cdot)$ in (3.5) is defined by the shifting the set $\bigcup_{j=1}^{\infty} I_j$, and since the random measure $M$ assigns independent random weights to disjoint parts of the real line, it follows that the larger the intersection of the set $\bigcup_{j=1}^{\infty} I_j$ with the shifted copies of itself, the stronger the dependence the interarrival time process $\{T_n, n \geq 0\}$ exhibits. In particular, in the extreme case $a_1 = 1, a_n = 0$ for every $n \geq 2$ this intersection is empty, and the interarrival times are i.i.d. On the other hand, we can make the dependence stronger by making the sequence $\{a_n, n \geq 1\}$ decrease in a regular and slow way.

We now discuss second order properties of $\{\xi_n\}$ and start by explaining the distributional structure of $\xi_0, \xi_1, \ldots, \xi_{n-1}$, which will be needed in the sequel. For a subset $A \subset \{0, 1, \ldots, n-1\}$ let

$$I_A = \{x \in \mathbb{R} : f(i+x) = 1 \text{ for all } i \in A, f(i+x) = 0 \text{ for all } i \notin A\}.$$ 

Then, obviously,

$$\xi_i = \sum_{A \ni i} M(I_A), \ i = 0, 1, \ldots, n-1.$$ 

Since for different $A$'s the sets $I_A$'s are pairwise disjoint, the corresponding terms $M(I_A)$'s are independent. Moreover, each $M(I_A)$ has Gamma distribution with scale parameter 1 and shape parameter $|I_A|$.

It is easy to understand the structure of the numbers $|I_A|$'s. First of all,

$$|I_A| = 0 \text{ if for some } i < j < k, i \in A, k \in A, j \notin A,$$

and so we consider now sets $A$ of the form

$$A(i, j) = \{i, i+1, \ldots, j\}, \ i \leq j.$$ 

Let $\gamma(i, j) = |I_{A(i, j)}|$. It is simple to see that

$$\gamma(i, j) = \begin{cases} a_{j+1}, & \text{if } i = 0, j = 0, 1, \ldots, n-2, \\ \sum_{k=n}^{\infty} a_k, & \text{if } i = 0, j = n-1, \\ a_{n-i}, & \text{if } i = 1, \ldots, n-1, j = n-1, \\ a_{j-i} - a_{j-i+1}, & \text{if } 0 < i < j < n-1. \end{cases}$$

Therefore, denoting $\Gamma(i, j) = M(I_{A(i, j)}), 0 \leq i \leq j \leq n-1$, we conclude that

$$\xi_k = \sum_{i=0}^{k} \sum_{j=k}^{n-1} \Gamma(i, j), \ k = 0, \ldots, n-1,$$

where $\Gamma(i, j), 1 \leq i \leq j \leq n$ are independent Gamma random variables with scale parameter 1, and shape parameters given by (3.12).

Now focus on the second order properties of the process. Even though second order properties in nonnormal models are not as critical as in the normal case, it is still somewhat illuminating to see how variance-covariance relations behave here. Let $\rho_n = \text{corr}(\xi_0, \xi_n), n \geq 0$ and $S_n = \xi_0 + \ldots + \xi_{n-1}, n \geq 1$. Note that part (i) of Proposition 3.1 shows, in particular, that $\rho_n$ can decrease to zero as slowly as we wish, if $a_n$'s are chosen to go to zero slowly enough.
Proposition 3.1. (i) For the correlations \( \{ \rho_n \} \) of the \( \{ \xi_n \} \) process we have

\[
\rho_n = \sum_{i=n+1}^{\infty} a_i,
\]

In particular, if

\[
a_n \sim cn^{-(1+\theta)}, \quad n \to \infty
\]

for some \( \theta > 0 \) and \( c > 0 \), then

\[
\rho_n \sim c\theta^{-1}n^{-\theta}, \quad n \to \infty.
\]

(ii) For the variance of \( S_n = \xi_0 + \ldots + \xi_{n-1} \), \( n \geq 1 \) we have

\[
\text{Var}(S_n) = n^2 \sum_{i=n}^{\infty} a_i + \sum_{i=1}^{n-1} (n-i)^2 \left( (i+1)a_{n-i} - (i-1)a_{n-i+1} \right)
\]

\[
= n \left( 1 + 2 \sum_{h=1}^{n-1} \left( 1 - \frac{h}{n} \right) \sum_{i=h+1}^{\infty} a_i \right).
\]

In particular, if (3.15) holds, then as \( n \to \infty \),

\[
\text{Var}(S_n) \sim \begin{cases} 
2cn^{2-\theta}/(1-\theta)(2-\theta), & \text{if } 0 < \theta < 1, \\
2cn \log n, & \text{if } \theta = 1, \\
2cn^{\infty} \sum_{i=1}^{\infty} (2i-1)a_i, & \text{if } \theta > 1.
\end{cases}
\]

Proof. (i) One can proceed directly from (3.11) or by noting that

\[
(\xi_0, \xi_n) \overset{d}{=} (G_1 + G_2, G_1 + G_3),
\]

where \( G_1, G_2 \) and \( G_3 \) are independent Gamma random variables with scale parameters equal to 1, and shape parameters being \( \sum_{i=n}^{\infty} a_i, \sum_{i=1}^{n-1} a_i \) and \( \sum_{i=1}^{n-1} a_i \) respectively, from which (3.14) follows. Under the assumption (3.15) the estimate (3.16) is an easy consequence of (3.14).

(ii) It is a standard fact (for example see Brockwell and Davis, 1991, page 219) that

\[
\text{Var}(S_n) = n + 2 \sum_{i=1}^{n-1} (n-i)\rho_i.
\]

Then (3.17) follows from (3.14) and (3.18) is a straightforward consequence of (3.14) and (3.15). \( \square \)

Observe that by (3.18) the variance of the partial sums \( S_n \) increase faster than linearly whenever, e.g., the sequence \( \{ a_n, \ n \geq 1 \} \) satisfies (3.15) with \( 0 < \theta \leq 1 \).

The process \( \{ \xi_n, \ n \geq 0 \} \) defined by (3.5) with \( f \) as in (3.11) has standard exponential marginals. We can transform these marginals to any desired marginal distribution by the inverse transform method. Given a distribution function \( F \) we define

\[
Y_{n}^{(F,1)} = F^{-1}(1 - e^{-\xi_n}), \quad n \geq 0
\]
and

\[ Y_n^{(F,2)} = F^{-\xi_n}, \quad n \geq 0. \]

Since \(1 - e^{-\xi_n}\) has uniform distribution on \((0,1)\), both \(Y_n^{(F,1)}\) and \(Y_n^{(F,2)}\) have the prescribed distribution \(F\). Furthermore, as the processes \(\{Y_n^{(F,1)}, n \geq 0\}\) and \(\{Y_n^{(F,2)}, n \geq 0\}\) are each a pointwise monotone transformation of the process \(\{\xi_n, n \geq 0\}\), they naturally have a dependence structure similar to that of \(\{\xi_n\}\). It is also clear that these two processes have, in general, different multivariate distributions.

Recall from Section 2 that the tail properties of the stationary queue length distribution depend on the distribution of \(S_n/n\) and we now discuss how dependence properties of \(\{T_n\}\) influence the distribution of \(X_\infty\). It turns out that at least one of the processes defined above is long-range dependent in the appropriate sense, as long as the sequence \(\{a_n, n \geq 1\}\) decreases to zero slowly enough. This is presented in the following proposition.

**Proposition 3.2.** For every \(r \in R\) such that \(F(r) > 0\) there is a finite positive constant \(c(r, F)\) such that for every \(n \geq 1\)

\[
P(Y_i^{(F,2)} \leq r, i = 0, 1, \ldots, n-1) \geq c(r, F) \left( \Gamma \left( \sum_{j=n}^\infty a_j \right) \right)^{-1}
\]

\[
\sim c(r, F) \sum_{j=n}^\infty a_j = c(r, F) \rho_{n-1}, \quad \text{as } n \to \infty.
\]

**Proof.** Obviously, for any \(r\) with \(F(r) > 0\) we have

\[
P(Y_i^{(F,2)} \leq r, i = 0, 1, \ldots, n-1) = P(T_i \geq \log \left( \frac{1}{F(r)} \right), i = 0, 1, \ldots, n-1),
\]

\(n = 1, 2, \ldots\). It follows from (3.13) that

\[
T_i \geq \Gamma(0, n-1) = \sum_{j=0}^\infty M((j, j + a_j + 1)), \quad i = 0, \ldots, n-1.
\]

Denoting \(z_n = \sum_{k=n}^\infty a_k, \lambda = \log(1/F(r))\), and using (3.12) we immediately conclude that

\[
P(Y_i^{(F,2)} \leq r, i = 0, 1, \ldots, n-1) \geq P(\Gamma(0, n-1) \geq \lambda)
\]

\[
= \int_\lambda^\infty e^{-t \Gamma(z_n^{-1})} dt
\]

\[
\geq c(r, F) \left( \Gamma(z_n) \right)^{-1}
\]

where we set

\[
c(r, F) := e^{-2\lambda} (1 \wedge 2\lambda)/2,
\]

from which the first part of (3.22) follows. The second part of (3.22) follows from \(\Gamma(z) \sim z^{-1}, z \to 0\). This completes the proof. \(\Box\)

We conclude from (3.22) that the probability on the left side of (3.22) can be made to decrease arbitrarily slowly to zero, provided \(a_j\) decreases to zero at the proper rate.
Now we recall the framework of Section 2. Let $F$ be the distribution function of a random variable on $(0,\infty)$ with expected value $\lambda^{-1}$. Let

$$T_n = Y_n^{(F,2)}, \ n \geq 1,$$

so that $\{T_n, n \geq 1\}$ is a stationary sequence of random variables with common distribution $F$. From the construction of the sequence $\{T_n\}$, it is clear that $\{T_n\}$ is a reversible sequence. It is also ergodic (cf. Cambanis et. al. (1991)). If we suppose that $\rho = \lambda/\mu < 1$, then Proposition 2.1 is applicable.

Assume that there is a $r < 1/\mu$ such that $F(r) > 0$ and set

$$S_0 = 0, \ S_n = T_1 + \ldots + T_n, \ n \geq 1.$$

As promised at the end of Section 2, we now bound the tail of the stationary distribution of the number $X_\infty$ of customers in the queue seen by an arriving customer as follows:

$$P(X_\infty > (1 - r\mu)k) = P\left[\bigvee_{j=0}^{\infty} (j - N(0, S_j)) > (1 - r\mu)k\right]$$

$$\geq P[N(0, rk) \leq rk\mu, S_k < rk]$$

$$\geq P\left[\frac{N(0, rk) - rk\mu}{\sqrt{rk\mu}} \leq 0\right] P[S_k < rk]$$

$$\sim \frac{1}{2} P(S_k < rk)$$

$$\geq \frac{1}{2} P[T_i \leq r, i = 1, \ldots, k]$$

$$\geq c \sum_{j=k}^{\infty} a_j,$$

where $c$ is a finite positive constant, and we have applied (3.22). Thus

$$P(X_\infty > k) \geq c \sum_{j=\lceil k/(1-r\mu) \rceil}^{\infty} a_j,$$

for all $k$ large enough.

One immediate conclusion is that, if the $a_n$'s satisfy (3.15) with $\theta \leq 2$, then $X_\infty$ has infinite second moment, whereas if $\theta \leq 1$, then even the mean of $X_\infty$ is infinite! This is, of course, in sharp contrast to what happens in the classical $G/M/1$ model.

**Remarks:** (i) The circumstance where the assumption that $F(r) > 0$ for some $r < 1/\mu$ fails is somewhat degenerate since failure of the assumption means $F(\mu^{-1}) = 0$, in which case $P[T_1 > \mu^{-1}] = 1$ and *every* single interarrival time will be longer than the mean service time. The queue length will be bounded above by that of a $D/M/1$ queue and for such a queue, the stationary queue length distribution has tails which decay exponentially fast.

(ii) The more customers in the system, the longer a new arrival will spend in the system. So a lower bound on the tail of $X_\infty$ as in (3.24) should imply a corresponding lower bound on the tail of the time in the system of a typical customer. Indeed, let $W_n$ be the waiting time in the system of the $n$th customer arriving to the queue so that $W_n$ satisfies

$$W_n \overset{d}{=} \sum_{i=1}^{X_{n+1}} Q_i, \ n \geq 1,$$
where \( Q_1, Q_2, \ldots \) are iid exponentially distributed random variables with common parameter \( \mu \) and independent of \( X_n \). It follows from Proposition 2.1 that

\[
W_n \Rightarrow W_\infty \overset{d}{=} \sum_{i=1}^{X_\infty+1} Q_i,
\]

where \( \{Q_i\} \) are independent of \( X_\infty \). From (3.26) we get the tail behavior of \( W_\infty \):

\[
P[W_\infty > \lambda] \geq P[X_\infty \geq [\mu \lambda]] P\left[ \sum_{i=1}^{[\mu \lambda] + 1} Q_i > \lambda \right] \\
\geq P[X_\infty \geq [\mu \lambda]] P\left[ \sum_{i=1}^{[\mu \lambda] + 1} Q_i - \left( \frac{[\mu \lambda] + 1}{\mu} \right) > \lambda - \left( \frac{[\mu \lambda] + 1}{\mu} \right) \right] \\
\geq P[X_\infty \geq [\mu \lambda]] P\left[ \sum_{i=1}^{[\mu \lambda] + 1} Q_i - \left( \frac{[\mu \lambda] + 1}{\mu} \right) > 0 \right]
\]

and by applying the central limit theorem we get

\[
\sim \frac{1}{2} P[X_\infty \geq [\mu \lambda]] \\
\geq \frac{1}{2} c \sum_{j=|[\mu \lambda]/(1-\mu)|}^{\infty} a_j,
\]

as \( \lambda \to \infty \).

As with the number in the system, we see that it is quite possible that \( W_\infty \) might have infinite variance or even infinite mean if \( a_n \downarrow 0 \) slowly enough.

4. Simulation results.

We simulated the behavior of the simple queue described in the previous sections when the input process of interarrival times \( \{T_n, n \geq 1\} \) is given by (3.21). We have selected \( F(x) = 1 - e^{-x} \), \( x \geq 0 \), so that the interarrival times have marginally the standard exponential distribution. The service rate has been chosen to be \( \mu = 10/7 \), so that the traffic intensity in the system is

\[
\rho = .7.
\]

We have run the simulation with two choices of \( a_n \)'s satisfying (3.15): first with

\[
a_n = c(1 + n)^{-\theta}, \quad n \geq 1,
\]

(here \( \theta = 1 \)) and then with

\[
a_n = c(1 + n)^{-1.5}, \quad n \geq 1,
\]

where \( \theta = .5 \). The constants \( c \) in both cases are chosen in such a way that (3.9) holds. Note that it follows from (3.25) that \( X_\infty \) has in both cases infinite mean, and that the lower bound on the probability tail of \( X_\infty \) is asymptotically bigger under (4.3) than under (4.2). For the purpose of comparison we present also simulations of the same simple queue when the input process is actually given by (3.20) with the same choice of the marginal distribution function \( F \), and the same two choices of the interval lengths given by (4.2) and
Accordingly. All the queues were simulated up to the time of the arrival of the 10,000th customer. The results are presented below. The long range dependent input process \( \{T_n, n \geq 1\} \) has been simulated on a Spark 10 station using the representation (3.13). The results were then fed into the package \textit{Sigma} (cf. Schruben, 1994) to produce the queueing characteristics.

**Figure 4.1:** What 10,000 arriving customers see.
Dependent interarrival times

For comparison we simulated the classical \( G/M/1 \) (actually, \( M/M/1 \)) queue, with the same arrival and service rates.

**Figure 4.2:** What 10,000 arriving customers see, \( G/M/1 \) case

Clearly, the customers in the queues with interarrival times generated by our procedure tend to see much longer lines, and the lines get longer as \( \theta \) decreases (as expected from our discussion in the previous section). Further the long range dependent processes given by (3.21) tend to create the longest lines.
These conclusions are further strengthened by looking at the histograms of the marginal distributions of the numbers of customers seen by new arrivals in the five cases considered above. The distribution tails seem to be much fatter for the long range dependent input (Figure 4.1) than for the renewal input (Figure 4.3); in the latter case the actual marginal distribution is well known to be geometric, with parameter $1 - \rho = .3$.

Figure 4.3: Histograms for dependent input

Figure 4.4: Histogram for independent input
Acknowledgement: Ranjini Natarajan generously shared with us her knowledge of computing and statistical simulation.

REFERENCES


SIDNEY RESNICK, CORNELL UNIVERSITY, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, ETC BUILDING, ITHACA, NY 14853 USA
E-mail: sid@orie.cornell.edu

GENNADY SAMORODNITSKY, CORNELL UNIVERSITY, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, ETC BUILDING, ITHACA, NY 14853 USA
E-mail: gennady@orie.cornell.edu