BOUNDS ON THE DELAY DISTRIBUTION IN GI/G/1 QUEUES

Sheldon M. Ross
California University

Prepared for:
Office of Naval Research
Army Research Office-Durham

January 1973
BOUNDS ON THE DELAY DISTRIBUTION
IN GI/G/1 QUEUES

by
SHELDON M. ROSS

OPERATIONS
RESEARCH
CENTER

COLLEGE OF ENGINEERING
UNIVERSITY OF CALIFORNIA - BERKELEY
BOUNDS ON THE DELAY DISTRIBUTION IN GI/G/1 QUEUES

by

Sheldon M. Ross
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

JANUARY 1973

This research has been partially supported by the U.S. Army Research Office-Durham under Contract DA-31-124-ARO-D-331 and the Office of Naval Research under Contract N00014-69-A-0200-1036 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.
**Title:** BOUNDS ON THE DELAY DISTRIBUTION IN GI/G/1 QUEUES

**Author:** Sheldon M. Ross

**Report Date:** January 1973

**Contract or Grant No.:** DA-31-124-ARO-D-331

**Project No.:** 20014501B14C

**Distribution Statement:** This document has been approved for public release and sale; its distribution is unlimited.

**Supplementary Notes:** Also supported by the Office of Naval Research under Contract NO0014-69-A-0200-1036.

**Abstract:** SEE ABSTRACT.
<table>
<thead>
<tr>
<th>KEY WORDS</th>
<th>LINK A</th>
<th>LINK B</th>
<th>LINK C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delay Distribution</td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
</tr>
<tr>
<td>GI/G/1 Queues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martingale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stopping Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

Bounds are obtained for the limiting distribution of the delay in queue for a GI/G/1 system via Martingale theory. These bounds are somewhat stronger than similar bounds recently obtained by Kingman. Simplifications of the bounds are obtained in the special cases where the service distribution is either IFR, DFR, NBU or NWU.
1. INTRODUCTION

Consider the usual GI/G/1 queue with interarrival times between customers \( X_1, X_2, \ldots \) and service times \( Y_1, Y_2, \ldots \), where \( E(Y_i) < E(X_i) < \infty \). Let \( U_i = Y_i - X_i \) and assume that there exists a nonzero value \( \theta \) such that

\[
E\left[ e^{\theta U_i} \right] = 1.
\]

If such a value of \( \theta \) exists then by Jensen's inequality it must be positive since \( EU_1 < 0 \). Let \( D_n \) denote the delay in queue of the \( n \)th customer, and let

\[
\bar{D}(t) = \lim_{n \to \infty} P(D_n > t).
\]

In [1] and [2] the following inequality was proven by Kingman

\[
ae^{-\theta t} \leq \bar{D}(t) \leq e^{-\theta t}
\]

where

\[
a = \inf_{t>0} \frac{\int_0^\infty dF(y) / \int_0^t e^{\theta(y-t)} dF(y)}{t}
\]

and where \( F \) is the distribution of \( U_1 \). Kingman proved the right side of the above inequality in [1] by using Kolmogorov's inequality for Martingales, and used a different technique in [2] to obtain the complete inequality. Following the Martingale approach of Kingman but using an appropriate stopping time rather than Kolmogorov's inequality a somewhat sharper inequality will now be obtained.
2. THE INEQUALITIES

As was shown by Lindley [3]

\[ D(t) = \Pr \left\{ \sum_{i=1}^{n} U_i > t \text{ for some } n = 1, 2, \ldots \right\}. \]

If we let \( Z_n = e^{\sum_{i=1}^{n} U_i} \), then as \( Z_n \) is the product of independent random variables each with mean 1, it follows that \( \{Z_n \geq 1\} \) is a Martingale. For a fixed positive constant \( A \), define the stopping time \( N_A \) by

\[ N_A = \text{1st } n \text{ such that either } Z_n > e^{\theta t} \text{ or } Z_n < e^{-\theta A}. \]

As is well known the moment generating function of \( N_A \) exists in a region about 0 and thus by Martingale theory

\[ 1 = E(Z_n) = E(Z_{N_A}) \]

\[ \begin{align*}
1 &= E \left[ e^{\sum_{i=1}^{N_A} U_i} \right] P \left( \sum_{i=1}^{N_A} U_i > t \right) + \\
&\quad \quad + E \left[ e^{\sum_{i=1}^{N_A} U_i} \right] P \left( \sum_{i=1}^{N_A} U_i < -A \right) \tag{1}
\end{align*} \]

Now,

\[ E \left[ e^{\sum_{i=1}^{N_A} U_i} \right] = e^{\theta t} E \left[ e^{\sum_{i=1}^{N_A} U_i - \theta t} \right] \tag{2} \]

By conditioning on \( N_A \) and \( \sum_{i=1}^{N_A} U_i \) we obtain that
(3) \[ \inf_{0<r} E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right] \leq E \left[ e^{\theta(U_1)} \mid \sum_{l=1}^{N_A} U_l > t \right] \leq \sup_{0<r} E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right] \]

Also, since

\[ \lim_{A \to \infty} P \left( \sum_{l=1}^{N_A} U_l > t \right) = P \left\{ \sum_{l=1}^{N_A} U_l > t \text{ for some } n < \infty \right\} = \bar{\delta}(t) \]

and

\[ \lim_{A \to \infty} E \left[ e^{\theta \sum_{l=1}^{N_A} U_l - \sum_{l=1}^{N_A} U_l} \mid \sum_{l=1}^{N_A} U_l < -A \right] = 0 \]

we obtain from (1), (2), and (3), by letting \( A \to \infty \) that

(4) \[ \frac{e^{-\theta t}}{\sup_{0<r} E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right]} \leq \bar{\delta}(t) \leq \frac{e^{-\theta t}}{\inf_{0<r} E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right]} \]

Note that the left side inequality is just the left side of Kingman's inequality while the right sided inequality is stronger than Kingman's.\(^\dagger\)

A somewhat weaker though probably more useful inequality based on the service distribution can be obtained from the above as follows:

\[ E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right] = E \left[ e^{\theta(Y_1-(X_1+r))} \mid Y_1 > X_1 + r \right]. \]

Hence, by conditioning on \( X_1 \) we obtain that

\[ \inf_{0<s} E \left[ e^{\theta(Y_1-s)} \mid Y_1 > s \right] \leq E \left[ e^{\theta(U_1-r)} \mid U_1 > r \right] \leq \sup_{0<s} E \left[ e^{\theta(Y_1-s)} \mid Y_1 > s \right] \]

\(^\dagger\)The sup and inf are taken over all nonnegative values of \( r \) for which \( P(U_1 \mid Y_1 = s) > 0 \).
As this is true for all \( r \), we obtain that

\[
\sup_{0 < r} E \left[ e^{\theta (U_1 - r)} \mid U_1 > r \right] \leq \sup_{0 < s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right]
\]

\[
\inf_{0 < r} E \left[ e^{\theta (U_1 - r)} \mid U_1 > r \right] \geq \inf_{0 < s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right].
\]

Thus, from (4) we obtain

\[
\frac{e^{-\theta t}}{\sup_{0 < s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right]} \leq \overline{D}(t) \leq \frac{e^{-\theta t}}{\inf_{0 < s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right]}.
\]

A very important special case occurs when the service times are exponentially distributed with mean \( 1/\mu \). That is, when the system is a G/M/1 queue. In this case, using the lack of memory of the exponential distribution we have that the conditional distribution of \( Y_1 - s \) given that \( Y_1 > s \) is just exponential with mean \( 1/\mu \). Hence,

\[
\sup_{s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right] = \inf_{s} E \left[ e^{\theta (Y_1 - s)} \mid Y_1 > s \right] = \frac{\mu}{\mu - \theta}
\]

and thus in the G/M/1 case

\[
\overline{D}(t) = \frac{\theta}{\mu - \theta} e^{-\theta t}.
\]

In certain special cases the sup and inf in equation (5) can be more easily expressed. We say that the service distribution \( G \) is NBU if

\[
\overline{G}(s + t) \leq \overline{G}(s) \overline{G}(t) \quad \text{for all } s, t > 0
\]

and it is said to be NWU if

\[
\overline{G}(s + t) = \overline{G}(s) \overline{G}(t) \quad \text{for all } s, t > 0.
\]

\[\dagger\] The sup and inf in equation (5) are taken over all nonnegative values of \( s \) for which \( P(Y_1 > s) > 0 \).
\[ \tilde{G}(s + t) \geq \tilde{G}(s)\tilde{G}(t) \quad \text{for all } s, t \geq 0 \]

where \( \tilde{G}(t) = 1 - G(t) \). Thus \( G \) is NBU (NWU) means that the remaining service time of a customer who has already been in service for some fixed time is always stochastically smaller (larger) than the service time of a customer just entering service. 

Since \( V \) being stochastically larger than \( W \) implies that 
\[ E[f(V)] > E[f(W)] \] 
for all increasing functions \( f \), we obtain that if \( G \) is NBU with \( G(0) = 0 \) then 
\[
\sup_{0 < s} \left[ e^{\theta(Y_1 - s)} \mid Y_1 > s \right] = E\left[ e^{\theta Y_1} \right]
\]
while if \( G \) is NWU with \( G(0) = 0 \) then 
\[
\inf_{0 < s} \left[ e^{\theta(Y_1 - s)} \mid Y_1 > s \right] = E\left[ e^{\theta Y_1} \right].
\]

If we make the stronger (than NBU) assumption that \( G \) is IFR (that is, that \( \frac{\tilde{G}(s + t)}{\tilde{G}(t)} \) decreases in \( t \) for all \( s \)) then it easily follows when \( G(0) = 0 \) that 
\[
\sup_{0 < s} \left[ e^{\theta(Y_1 - s)} \mid Y_1 > s \right] = E\left[ e^{\theta Y_1} \right]
\]
and 
\[
\inf_{0 < s} \left[ e^{\theta(Y_1 - s)} \mid Y_1 > s \right] = \lim_{s \rightarrow \infty} \inf_{0 < s} \left[ e^{\theta(Y_1 - s)} \mid Y_1 > s \right]
\]

where \( \tilde{M} = \sup\{t : \tilde{G}(t) > 0\} \).

The terminology NBU (new better than used) originated in reliability literature where it means that if \( G \) is the distribution of the lifetime of a component then the remaining life of any \( s \) year old (that is, any used) item is stochastically smaller than the lifetime of a new item.
Similarly if $G$ is assumed to be DFR (that is, if it is assumed that \( \frac{G(s+t)}{G(t)} \) increases in $t$ for all $s$) then it follows when $G(0) = 0$ that

\[
\sup_{0 \leq s} E\left[ e^{\theta(Y_1-s)} \mid Y_1 > s \right] = \lim_{s \to \infty} E\left[ e^{\theta(Y_1-s)} \mid Y_1 > s \right]
\]

\[
\inf_{0 \leq s} E\left[ e^{\theta(Y_1-s)} \mid Y_1 > s \right] = E\left[ e^{\theta Y_1} \right].
\]

The importance of DFR service time distributions partly derives from the fact that if the actual service distribution is a mixture of other distributions each of which is DFR (for instance, each may be exponential) then the service distribution is also DFR.
REFERENCES


