VELUT ARBOR ÆVO

A Comparison of Stochastically Ordered Queues

by

Sudha Jain
Department of Statistics
University of Toronto

Technical Report No. 9514, October 16, (1995)

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

A Comparison of Stochastically Ordered Queues

Sudha Jain
Department of Statistics
University of Toronto, Toronto
Ontario, Canada M5S 1A1

ABSTRACT

In this paper, the principle of stochastic ordering is applied to compare performance measures such as waiting times of two queueing systems. The technique of Monte Carlo simulation is employed to study the behaviour of such systems. It is shown that, in practice, to compare two stochastically ordered queues, we require a large number of customers to show significant results. Furthermore, the problem of comparing stochastically ordered queues requires careful interpretation when one is dealing with heavy-traffic queues.

Key Words: Stochastic Ordering, Heavy-traffic, GI/G/1 queues.

1. Introduction

The concept of stochastic orderings can be used to study performance measures of queues under different settings of operation. With any flexible manufacturing system, it is often required to measure the effectiveness of the system and compare it with similar systems. Stochastic comparison of queues requires understanding the phenomenon of stochastic orderings (applicable to queues). For example, certain performance measure such as waiting times of queues can be made stochastically ordered. Stochastic comparisons for different queueing systems based on stochastic ordering of the underlying distributions were made by Yu (1974) and Jacobs and Schach (1972). Stoyan (1983) provides a comprehensive treatment of stochastic comparisons to queues. Whitt (1981) compared waiting time and queue length in multiserver queues based on the principle of stochastic ordering. Gross and Harris (1985) discussed the idea of stochastic dominance in the study of bounds and approximations of queueing systems. It can be easily followed on the principle of stochastic orderings that if the interarrival random variable of one queue is stochastically smaller than that of a second queueing system, and the service time variable of the second system is stochastically smaller than the first system, then it implies that the waiting times of the first queue will be larger than those of the second. This type of ordering was used for finding bounds for a complicated GI/GI/1 queueing systems.

More recently, Harris and Prabhu (1987) discussed stochastic comparisons for single server queues of various quantities of interest such as waiting time, number of customers in the system, and duration of busy periods. However, the practical implications of stochastic orderings in queues are not discussed. The object of this paper is to study the behaviour of stochastically ordered queues. Numerical simulations are performed for a GI/GI/1 queueing system using Monte Carlo technique. This study provides a guideline for under which circumstances the comparison of performance measures would be possible. The problem of heavy-traffic is also investigated when dealing with the problem of stochastic orderings for comparing performance measures of queues.

In section 2, the preliminary results and notations concerning stochastic orderings for queueing systems are presented. Section 3 deals with statistical methods for comparing queues based on the concept of stochastic orderings. In our study, because we are only interested in the alternative hypothesis that the waiting times of one queueing system is stochastically larger or smaller than the second queueing system, the nonparametric Savage test was considered to be appropriate. In Section 4, various GI/GI/1 queueing models are simulated under different conditions for the purpose of comparing performance measures. Monte Carlo simulations are carried out to generate the waiting times distribution with various traffic intensities. Finally, the implications of the results are summarized and their problems are discussed with applications.

2. Stochastic Ordered Queues

Stochastic orderings for the distribution function is defined as follows:

Definition. Let X and Y be random variables and F and G be their respective distribution functions. The random variable X is stochastically smaller than Y and is denoted by $X \underset{s}{\leq} Y$, or equivalently, their respective distribution functions F and G satisfy $F \underset{d}{\leq} G$, if for every t

$$P(X > t) \leq P(Y > t)\ ,$$

and

$$F(t) \geq G(t)$$

The following distribution functions are examined by Stoyan (1983) to compare the parameters of queueing systems.

Exponential distribution. Let

$$F(t) = 1 - exp(-\lambda_1 t) \; ; \; t \geq 0, \; \lambda_1 > 0 \, , \tag{2.1}$$

$$G(t) = 1 - exp(-\lambda_2 t) \; ; \; t \geq 0, \; \lambda_2 > 0 \, . \tag{2.2}$$

Then

$$F \underset{d}{\leq} G \iff \lambda_2 \leq \lambda_1 \, . \tag{2.3}$$

Erlang distribution. Let

$$F(t) = 1 - \sum_{i=0}^{k_1 - 1} \frac{(\lambda_1 t)^i}{i\,!} exp(-\lambda_1 t) \tag{2.4}$$

and

$$G(t) = 1 - \sum_{i=0}^{k_2 - 1} \frac{(\lambda_2 t)^i}{i\,!} exp(-\lambda_2 t) \tag{2.5}$$

Then, it follows that $\lambda_2 \leq \lambda_1$ and $k_1 \leq k_2$ implies $F \underset{d}{\leq} G$.

Consider the following queueing systems S_1 and S_2.

$S_1 : GI_1/GI_1/1$. Let interarrival time distribution be $A_1 \approx$ Erlang (λ_1, l_1) and let service time distribution be $B_1 \approx$ Erlang (μ_1, k_1)

$S_2 : GI_2/GI_2/1$. Let interarrival time distribution be $A_2 \approx$ Erlang (λ_2, l_2) and let service time distribution be $B_2 \approx$ Erlang (μ_2, k_2).

Where $l_1, \; k_1, \; l_2, \; k_2,$ are shape parameters; and $\lambda_1, \; \lambda_2, \; \mu_1, \; \mu_2$ are the scale parameters. It can be easily seen that queueing systems S_1 and S_2 satisfy the following conditions:

a) let $\lambda_1 \leq \lambda_2$ and $l_2 \leq l_1$

then

$$A_1(t) \geq A_2(t) \tag{2.6}$$

b) let $\mu_2 \leq \mu_1$ and $k_1 \leq k_2$

then

$$B_1(t) \leq B_2(t) \tag{2.7}$$

Under the above conditions, Jacob and Schach (1972) have shown that system 2 has a stochastically shorter waiting time and queue size than system 1.

Assuming that the service time parameters μ_1 and μ_2 for the two Erlang queueing systems are equal (i.e. $\mu_1 = \mu_2 = \mu$) and that they have the same number of phases for the service time distribution (i.e. $k_1 = k_2 = k$), but different scale parameters λ_1 and λ_2, then the traffic intensities are given by

$$\rho_1 = \frac{\lambda_1 k}{\mu} = \lambda_1 C \tag{2.8}$$

and

$$\rho_2 = \frac{\lambda_2 k}{\mu} = \lambda_2 C \tag{2.9}$$

where

$$C = \frac{k}{\mu}.$$

Lehmann (1983) stated that a test for location provides a good method for testing the equivalence of two distributions, but it is rarely realistic even when it is reasonable to assume that the density function is known. It is desirable to allow the model to contain an unknown scale parameter and justify the equivalence of the scale parameter. Since the traffic intensity parameter, ρ, is an important system parameter, then for design and control of queues, it would be desirable to test the equality of ρ_1 and ρ_2. The problem of testing the equality of ρ_1 and ρ_2 is equivalent to that of testing the following null hypothesis:

$$H_o : \lambda_1 = \lambda_2 \tag{2.10}$$

In our statistical analysis, we treat the special and important queueing quantity, the waiting time of the two queueing systems, by constructing a null hypothesis for the scale parameters. For M/E_k/1 queue, the mean waiting times W_{q_1} and W_{q_2} for systems S_1 and S_2 are given by [Gross and Harris (1988), page 178]:

$$W_{q_1} = \frac{(k+1)\lambda_1}{(2k)\mu(\mu - \lambda_1)}, \tag{2.11}$$

and

$$W_{q_2} = \frac{(k+1)\lambda_2}{(2k)\mu(\mu - \lambda_2)}, \tag{2.12}$$

respectively.

From (2.11) and (2.12), it can be easily seen that the problem of comparing the means of the waiting times of the queueing systems is equivalent to one of testing the equality of the scale parameters of the service time distribution.

3. Statistical Methods

In order to make comparisons for waiting times of systems 1 and 2, we need statistical methods for testing the hypotheses which are applicable to stochastically ordered queues such as the one proposed in (2.10). There are various nonparametric methods which are useful in comparing such ordered observations. In this paper, the Savage test statistic is applied because it is the most powerful nonparametric test for comparing waiting times of stochastically ordered queues. This is simulated based on exponentially distributed random variables of arrival and service pattern. A more detailed discussion of generating the waiting time distribution will be given in the next section.

Let $x_1, x_2, ..., x_m$ be m independent observations on a random variable X with cumulative distribution function (CDF) F(x), and $x_{m+1}, x_{m+2}, ..., x_{m+n}$ be n independent observations on a random variable Y with CDF G(y). Consider testing the following null hypothesis:

$$H_o : p(x_1, x_2, ..., x_N) = \prod_{i=1}^{N} f(x_i)$$

against an alternative

$$H_1 : q(x_1, x_2, ..., x_N) = \prod_{i=1}^{m} f(x_i - d_i) \prod_{j=1}^{n} f(x_{m+j} - d_{m+j})$$

where $d = (d_1, d_2, ..., d_N)$ is an arbitrary vector and N=m+n.

Hájek and Šidak (1967) defined the following statistic S_c for testing the hypothesis (3.1):

$$S_c = \sum_{i=1}^{N} (C_i - \overline{c}) a(R_i) \tag{3.2}$$

where $\overline{c} = \frac{1}{N} \sum_{i=1}^{N} c_i$, and $a(R_i)$ is the value of the "score function" a(·) at R_i.

Various statistics can be derived from (3.2) such as the Wilcoxon, Van der Waerden (or Normal Score) and Savage tests. The Savage statistic is described as follows:

Savage Test.

The Savage test statistic is defined by

$$S = \sum_{i=1}^{m} a(R_i) \tag{3.3}$$

where

$$a(i) = \sum_{j=N-i+1}^{N} (\frac{1}{j})$$

It can be shown that under H_o, the expectation and variance of (3.3) is given by

$$E(S) = m \tag{3.4}$$

$$Var(S) = \frac{mn}{N-1}\left(1 - \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{j}\right)\right) \tag{3.5}$$

The distribution of S is approximately normal for min(m,n) large.

Thus, the test statistic (3.3) is easy to use in our application to compare the waiting time distribution of the two stochastically ordered queues.

4. Simulations

Monte Carlo simulations were carried out to generate the waiting time distribution of $GI_1/GI_1/1$ and $GI_2/GI_2/1$ queueing systems under conditions (a) and (b). For simulation purposes, we consider a special case of the Erlang distribution for systems S_1 and S_2 when $k_1 = k_2 = l_1 = l_2 = 1$, which implies the exponential arrival rate and service time distribution. For various combination of parameters, queueing systems S_1 and S_2 are generated. Tables 4.1 to 4.4 provides a brief summary of the simulated results for systems S_1 and S_2.

Simulations were run of sizes 100, 200, 500, 1000, 2000, and 5000, and were conducted using various traffic intensities in systems S_1 and S_2 with the different scale parameters. The Savage test was applied to the simulated waiting times for each of the simulated data. We are interested in rejecting the null hypothesis (2.10). The percentage of rejections of the null hypothesis are given in tables 4.1 to 4.4.

Tables 4.1 and 4.2 present results with moderate traffic intensities, while table 4.2 considers traffic intensities getting closer for systems S_1 and S_2. Similarly, tables 4.3 and 4.4 are based on queues generated with heavy traffic intensities. Obviously, this requires a larger number of observations in order to reject the null hypothesis. Furthermore, when traffic intensities are getting closer under heavy traffic conditions, we need a substantially larger sample size. These results are given in table 4.4.

Table 4.1. System S_1 waiting times are generated with $\lambda_1 = 0.4$, $\mu_1 = 0.5$ and $\rho_1 = 0.8$. System S_2 waiting times are generated with $\lambda_2 = 0.4$, $\mu_2 = 1.0$ and $\rho_2 = 0.4$.

m=n	Number of Simulations	Rejection of H_o in %
10	100	56.00
	500	53.90
	1000	53.80
	2000	53.80
	5000	54.02
20	100	74.00
	500	74.25
	1000	72.00
	2000	70.90
	5000	70.86
30	100	85.00
	500	80.50
	1000	79.50
	2000	80.00
	5000	79.70
40	100	79.50
	500	85.60
	1000	85.10
	2000	83.20
	5000	83.40
50	100	87.00
	500	87.25
	1000	87.90
	2000	87.10
	5000	88.74
100	100	96.00
	500	97.00
	1000	96.00
	2000	97.25
	5000	97.50

Table 4.2. System S_1 waiting times are generated with $\lambda_1 = 0.8$, $\mu_1 = 1.0$ and $\rho_1 = 0.8$. System S_2 waiting times are generated with $\lambda_2 = 0.6$, $\mu_2 = 1.0$ and $\rho_2 = 0.6$.

m=n	Number of Simulations	Rejection of H_o in %
20	200	35.50
	500	38.20
	1000	38.50
	2000	39.90
	5000	39.70
30	200	44.50
	500	47.60
	1000	46.20
	2000	47.50
	5000	46.80
40	200	48.50
	500	51.60
	1000	51.20
	2000	50.85
	5000	51.50
50	200	50.50
	500	54.00
	1000	54.20
	2000	55.25
	5000	55.70
200	200	79.00
	500	79.50
	1000	81.00
	2000	81.20
	5000	81.50
500	200	94.50
	500	93.20
	1000	94.80
	2000	95.50
	5000	95.00

Table 4.3. System S_1 waiting times are generated with $\lambda_1 = 0.95$, $\mu_1 = 1.0$ and $\rho_1 = 0.95$. System S_2 waiting times are generated with $\lambda_2 = 0.8$, $\mu_2 = 1.0$ and $\rho_2 = 0.8$.

m=n	Number of Simulations	Rejection of H_o in %
50	200	49.50
	500	49.20
	1000	50.10
	2000	50.50
	5000	50.20
100	200	64.00
	500	65.20
	1000	65.10
	2000	67.20
	5000	67.80
200	200	73.00
	500	72.50
	1000	74.25
	2000	75.50
	5000	75.60
400	200	84.00
	500	83.25
	1000	84.50
	2000	84.25
	5000	84.80
500	200	89.50
	500	89.60
	1000	89.50
	2000	89.90
	5000	89.60
1000	200	96.00
	500	96.20
	1000	96.50
	2000	96.20
	5000	96.70

Table 4.4. System S_1 waiting times are generated with $\lambda_1 = 0.95$, $\mu_1 = 1.0$ and $\rho_1 = 0.95$. System S_2 waiting times are generated with $\lambda_2 = 0.9$, $\mu_2 = 1.0$ and $\rho_2 = 0.9$.

m=n	Number of Simulations	Rejection of H_o in %
100	200	56.00
	500	55.10
	1000	55.50
	2000	56.50
	5000	56.10
200	200	60.50
	500	59.10
	1000	60.60
	2000	60.50
	5000	60.10
500	200	62.00
	500	61.00
	1000	62.50
	2000	61.10
	5000	62.10
1000	200	65.60
	500	65.75
	1000	65.25
	2000	66.10
	5000	66.00
2000	200	81.00
	500	80.50
	1000	81.25
	2000	80.70
	5000	81.20
2500	200	85.00
	500	85.25
	1000	85.60
	2000	85.00
	5000	85.90

Concluding Remarks

It is clear from tables 4.1 to 4.4 that increasing the number of simulations from 200 to 5000 does not significantly affect the percentage of rejecting the null hypothsis. As expected, the implication of rejecting the null hypothesis is that system 2 has a stochastically shorter waiting time than system 1. From tables 4.1 and 4.3, it can be easily seen that if there is a significant difference in traffic intensities of systems S_1 and S_2, then a smaller sample size (or smaller number of customers) would be sufficient to give statistically significant waiting times of the two systems. Furthermore, if the traffic intensities are closer, then it requires a larger number of customers in order to show a significant difference in waiting times.

However, under heavy traffic conditions (i.e. $\rho \rightarrow 1$), it seems that we need a very large number of customers in order to make a meaningful comparison. Tables 4.3 and 4.4 present results under heavy traffic conditions. Obviously, we expect the sample size to be large, in order to build-up the queue, which would be reflected in increased waiting times. Smaller sample sizes under heavy traffic conditions will not reflect the true picture of what is happening to the performance measure (i.e. waiting times of the customers). In this paper, the practical utility of stochastic comparisons is demonstrated under both moderate and heavy traffic conditions.

Acknowledgements. The author is grateful to NSERC for supporting this research project. The author is also grateful to Professor J. G. C. Templeton for his comments.

References

Buzacott, J. A. and Shanthikumar, J. G. *Stochastic Models of Manufacturing Systems. Prentice Hall, 1993.*

Gross, D. and Harris, C. M. *Fundamentals of Queueing Theory. Second Edition, Wiley, 1985.*

Hâjek, J. *A Course in Nonparametric Statistics. Holden-Day. 1969.*

Harris, C. M. and Prabhu, N. U. Stochastic comparisons for single server queues. *Naval Res. Logist. Quart., Vol. 34, 555-567, 1987.*

Jacobs Jr., D. R. and Schach, S. Stochastic order relationships between GI/G/k systems. *Ann. Math. statist., Vol. 43, 1623-1633, 1972.*

Lehmann, E. L. *Theory of Point Estimation. Wiley. 1983.*

Shaked, M. and Shanthikumar, J. G. *Stochastic Orders and Their Applications. Academic Press. 1992.*

Whitt, W. (1981). Comparing point processes and queues. *Adv. Probab., 13, 207-220.*

Yu, O. S. (1974). Stochastic bounds for heterogeneous-server queues with Erlang service times. *J. Appl. Probab., 11, 785-796.*