# AN EXPERIMENTAL EVALUATION OF QUEUEING THEORY 

by
David Goyder
B. Com., University of British Columbia, 1969

A Thesis Submitted in Partial Fulfilment of The Requirements for the Degree of Master of Business Administration
in the Faculty of

Commerce and Business Administration

We accept this thesis as conforming to the required standard

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce and Business Administration

The University of British Columbia
Vancouver 8, Canada

Date 30 April, 1970

This thesis examines methods for predicting queue length of single server queues in order to evaluate how the practitioner may achieve greatest accuracy. Because accuracy is dependent on the correct estimation of the rate parameters of the population distributions and the choice of the appropriate method of prediction, the effects of errors in both of these are examined.

A computer simulation model written in GPSS/360 is used to create a real world from which data is drawn and where long run performance represents the correct solution. For four values of rho nine simulations are run, each with a unique combination of inter-arrival and service time distributions. In each of the 36 runs 10,000 arrivals are generated from which two samples of size 36 and 100 are taken and from which the generated queue statistics form the standard. A statistical analysis is used to detect samples taken from exponential distributions. The lack of a suitable test for small samples led to the development of a test based on the correlation coefficient of the sample times and pre-computed standard data.

Estimates for queue length are found with classical queueing formulae and solution methods suggested by Marshill These predictions are done without prior knowledge of rate parameters and queue type which are estimated from the samples. Then the estimated solutions are compared to the real world solution derived from the simulation.

Estimation error for each method is measured and
conclusions are drawn as to their accuracy in predicting queue length. It is found that accurate queue length estimation is possible using methods that can be applied without a great deal of prior mathematical knowledge. The classical formulae are accurate only when applied to queues with exponential inter-arrival times and are found to overestimate when applied to other queue types. The Increasing Failure Rate (IFR) bounds on queue length provide a satisfactory method of estimation for the general class of queues.

## TABLE OF CONTENTS

## CHAPTER

PAGE
I INTRODUCTION ..... 1
Statement of contents ..... 1
Purpose of Study ..... 1
The Standard and Data Source ..... 3
Method of Analysis ..... 6
II STATISTICAL ANALYSIS OF DATA ..... 7
Consideration in Choosing the Test ..... 8
Tests based on Life Testing ..... 9
An Analytic Exponential test ..... 11
a. Correlation Coefficient ..... 14
b. A Regression Line ..... 15
c. A t-test for the Ordinate ..... 15
Chi-Squared test for goodness of Fit ..... 16
Results of Applying Tests to the Sample Data ..... 27
Summary ..... 26
III METHODS FOR THE ANALYSIS OF QUEUES ..... 29
Assumptions ..... 29
$M / M / 1$ Queue ..... 29
M/G/1 Queue ..... 30
G/G/1,G/M/1 Queue ..... 30
G/G/1 Bounds ..... 31
TABLE OF CONTENTS (cont'd)
CHAPTER PAGE
III Computation of Lower Bound. ..... 33
IFR/G/I, DFR/G/I Queues ..... 34
Expected Wait and Queue Length ..... 35
IV QUEUE ANAJYSIS ..... 36
Evaluation of Sampling Errors ..... 37
Evaluation of Calculation Methods with Accurately Estimated Parameters ..... 46
Results from Accurate Data. ..... 47
Results from Inaccurate Data ..... 53
Summary ..... 57
$V$ AN APPLICATION AND CONCLUSION. ..... 58
Conclusions ..... 61
BIBLIOGRAPHY ..... 63
APPENDIX I Increasing and Decreasing Failure Rate ..... 64
II The Correlation Coefficient Test ..... 66

## LIST OF TABLES

Type I and II Errors from Testing forExponentials with Correlation Coefficient20III
IVIVType I and II Errors from Testing forExponential with Coefficient of Variation23
V
Ordinate of a Fitted Regression Line Equals Zero ..... 25
VI Type I and II Errors from Chi-Square Test of Rate Distribution of Sample Data ..... 27
VII Percent Error of Rho, Lamda, and Mu (Sample
Size 36) ..... 41
VIII Percent Error of Rho, Lamda, and Mu (Sample Size 100) ..... 42
IX Sampling Errors for Rho, Lamda, and Mu as a Percent of Intended Simulated Values ..... 44
X Summary of Queue Statistics from Accurate Data ..... 48
XI Summary of Queue Statistics from Inaccurate Data (Sample Size 36) ..... 54
TABLE PAGE
XII Sumary of Queue Statistics from InaccurateData (Sample Size l00) ............................ 55XIII Pre-computed Ordinate Values for theCorrelation Coefficient Test.69

## LIST OF FIGURES

## FIGURE

PAGE

1 Distributions used in the Simulation ...... 5
2 Plots of two Distributions................... 13
3 Effective Direction of Change in Rho for Changes in Lamda and Mu................... 39

4 Distribution of Percent Error in Rho by Sample size。....0.0............................. 40


## ACKNOWLEDGEMENT

I would like to express my appreciation to Professor L. L. George for his guidance with the mathematical portions of this thesis.

I am particularly indebted to Professor H. C. Wilkinson for his direction in the preparation of this paper and for conducting the simulation on which the work is based.

## NOTATION

$T_{n}=$ time between $n$-th and $(n+1)$-th arrival, $T_{n} \sim A(t)$,
$E\left[T_{n}\right]=1 / \lambda$
$S_{n}=$ service time of $n$-th customer, $S_{n} \sim G(t), E\left[S_{n}\right]=1 / \mu$
$U_{n}=S_{n}-T_{n} U_{n} \sim K(t)$
$\rho=\lambda / \mu$
( n )
$v_{f}=$ n-th moment about origin of random variable with
distribution $F$.
$\sigma^{2}=$ variance of a random variable with distribution $F$
$c_{f}^{2}=\sigma_{f}^{2} /\left(v_{f}\right)^{2}$, where $c_{f}$ is the coefficient of variation
$F^{C}(t)=1-F(t)$ for any distribution $F$

## CHAPTER I

## INTRODUCTION

Statement of Contents

This thesis will present an analysis of single server queues to establish how the practitioner might best solve his queueing problems.

The approach is to examine raw data of inter-arrival and service times sampled from a simulated real worldp to calculate queue statistics, and to compare these to the long run performance of the simulation model. This simulation model, in effect, provides the real world solution with which we can compare our calculated predictions of queue length and waiting time。

Some statistical procedures intended to aid the practitioner in arriving at his assumption of queue type will be examined for their applicability, ease of use and accuracy. Purpose of study

Queueing theory has developed considerably over the past twenty years through the use of sophisticated mathematics. Perhaps the most widely applied portion of queueing theory is derived from the classical $M / M / 1$ situation where the Poisson assumption makes the derivation of formulae for queue length and waiting times comparatively simple. The solution of
non- Poisson queues is considerably more difficuit and has not produced such easily calculable formulae. Indeed the solution to general single-server queues is not expressed solely in terms of the moments of the arrival and service distribution but require moments of the idle time distribu* tion which are hard to find. 1

Many queueing situations found in industry (the classic being the supermarket check-out problem) are assumed to be of the $M / M /$. type since the poisson process is found so often in nature and the type of facility being analyzed often displays the characteristics on which the Poisson is founded. This assumption may be false. However with the added convenience in applying the poisson queueing formulae and in the absence of a similarly convenient general set of formulae, the poisson assumption is made almost indiscriminately. This is certainly common in business where the potential benefits associated with greater accuracy are often not considered worth the added cost of analysis.

For these reasons it is our intention to find the level of mathematical rigour required to arrive at solutions to various Poisson and non-Poisson queues which are sufficiently accurate for practical purposes, and to measure the amount of error one can expect from a wrong assumption about queue type.

[^0]The assumptions considered here are that a particular distribution of inter-arrival or service times is, or is not distributed expontentially. Thus a single service queue may be any of the following four types: 1) $M / M / 1$
2) $M / G / 1$
3) $\mathrm{G} / \mathrm{M} / \mathrm{l}$
4) $G / G / 1$

A sub-classification of the general [G] distribution will be introduced later in the analysis to differentiate between those having increasing failure rate and decreasing failure rate. These latter terms are defined in Appendix $I$.

Recognizing that a large part of the solution lies in identifying the inter-arrival and service time distributions, we shall examine methods which differentiate between exponential and non-exponential distributions.

The Standard and Data Source

One of the greatest problems in analysing queues empirically is to establish a suitable standard against which to measure the performance of the hypothesized solution.

In this study use of a computer simulation written in GPSS/360 is made as a real world from which data can be drawn and where long run performance represents the correct solution. Single channel queues with traffic intensities ( $\rho$ ) denotes all other distributions of times.
of 06.7 .0 and .9 were simulated. For each value of p nine simulations were run, each with a unique combination of inter-arrival and service time distributions. The basic distributions from which each pair was selected are shown in Figure 1. In each of the thirty-six runs, 10,000 arrivals generated queue statistics to form the standard. Two samples, of 36 and 100 sequential measures of inter-arrival and service times, were taken at arbitrarily chosen points from each run; the two sample sizes were used to reveal how much effect sample size actually had on the accuracy of the solution. The entire simulation was run independently of the writer and only the two samples from each run were released for analysis. The generating distributions were not identified until all calculations were done and each sample was approached with no prior information whatsoever. Since samples of limited size are the only source of knowledge of the situation a user faces, the queueing analysis as presented in this thesis is worked from the samples alone. This procedure is to duplicate the industrial user's position as far as is possible.

The solution of queue length calculated by various methods is later compared to the simulated standard. Conclusions are then drawn as to the effectiveness of queue type identification, the accuracy of each method and the superiority of one method over another.


Each sample run is analyzed in two stages. First. some statistical tests are used to identify the type of distribution from which the sample times were drawn. As an analyst in business would generally need a readily usable procedure, the tests considered are relatively simple to calculate and are elementary to apply. It is assumed an analyst would be limited to a desk calculator and slide rule for his calculations.

The second step is analysis of queue operating characteristics by several methods of calculation. Solutions for queue length are found from the classical $M / M / 1$ queueing formulae and from equations for G/G/l queues given by Marshall. Values for upper and lower bounds on queue length for both M/M/1 and G/G/1 based solutions are also worked out. Calculations for both $M / M / 1$ and $G / G / 1$ are made for all samples. This will permit us to compare computed queue lengths for each sample under correct and incorrect assumptions of queue type. We can then derive the amount of error that can be expected to result from analysis based on erroneous identification of queue type.

[^1]STATISTICAL ANALYSIS OF DATA

The comparative ease with which one can find solutions for queues with exponential inter-arrival times from the M/M/1 or the Pollaczek - Khintchine formulae makes analysis of this distribution imperative. When a successful analysis is possible it can be extended to the service time distribution with very little additional effort and significant benefits in speed of computation. One is not so much concerned with actually identifying a distribution or classifying it as one of the known theoretical distributions, as with ascertaining if it comes from the exponential or not.

Considerations in Choosing the Test

Standard tests such as the chi-squared and KolmogorovSmirnov tests, while very popular, have certain limitations. The Kolmogorov - Smirnov is very powerful with an extensive body of data but will give errors with less than one hundred observations. ${ }^{4}$ There are not many practical situations where it is possible to take as many as 100 samples in steady state. In fact, in these experiments the two samples tested are size thirty-five and one hundred. It is the smaller one which is

[^2]being tested, being of practical size, and the larger is used mainly as a control. The Kolmogorov - Smirnov test is then of little value except for larger bodies of data. It may be noted here as well that Lilliefors observes that critical values determined in this test are not correct when one or more parameters are estimated from the sample. ${ }^{5}$,The computation of this test is a difficult and lengthly procedure, rendering it yet again unsuitable for our purposes. The chi-squared test for goodness of fit is perhaps the classic test for the fit of distributions. Its application too has limitations when the sample is small. The short-comings encountered with this test are discussed in greater detail later in this chapter.

Tests Based on Life Testing
Epstein ${ }^{6}$ wrote a paper concerned with statistical techniques in life testing of exponentials. These are tests of the assumption that any given sample data come from an exponential distribution. Two tests are examined in this

[^3]present thesis; one graphical test which is of some value, and one analytical which is based on a false assumption. For the sake of continuity the latter is presented first.

## An Analytic Exponential Test

This test uses a basic property of the Poisson process which is: ${ }^{7}$
"...if one observes a Poisson process for a fixed length of time $T$ and if $r$ events occur in $[O, T]$ at times, $t_{1} \leqslant t_{2} \leqslant t_{3} \ldots \leqslant t_{r} \leqslant T$, then these times (after being subjected to a random permutation) can be considered as $r$ independent observations of a random variable uniformly distributed over $\left[O_{\rho} T\right]^{n}$. Thus for $r$ even moderately large (assumed to be larger than 10) $\sum_{i=1} t_{i}$ is approximately normally distributed with mean $\frac{r T}{2}$ and variance $\frac{r T^{2}}{12}$ by the Central Limit Theorem. By standardising $\sum_{i=1}^{r} t_{i}$, the resultant statistic should be easily found in a table of areas under the normal curve. At the same time confidence limits around the mean can easily be computed. This is an attractive test for our purpose because of the simple computation and speed in providing a test of the hypothesis.

7 Ibid.. p. 7.

This test however is invalid. Consider the following example:

$$
\therefore \text { Let } t_{2}=t_{2}=\ldots=t_{20}=5 \text { be successive interval times }
$$

from a stochastic process. Let $T=10$. To test that the process is exponential, the 20 times must be tested to see if they are uniformly distributed on $[\mathrm{O}, \mathrm{T}]$. The expectation of


The 95\% confidence acceptance interval for the sample is $100 \pm(1.96)(12.9)=74.7,125.3$. The observed sum

$$
\Sigma t_{i}=100
$$

so, clearly, the hypothesis that the $t_{i}$ are exponential is accepted. But these data came from the degenerate distribution and thus the test is invalidated.

The error lies in the standardization of the random variable. The standardization of a random variable $X$ with $a$ finite mean and variance is defined (denoted by $X *$ ),

$$
X^{*}=\frac{x-E(x)}{\sigma_{X}}:
$$

$X^{*}$ is approximately normally distributed with mean zero and variance one. However any random variable must be standardized with its own mean and variance. When exponential data is tested the standardized variable is $N(O, 1)$. With any other data the distribution of the random variable is not determined. 'In the above example merely juggling $T$ will
cause the test to reject or accept arbitrarily.

A Graphical Exponential Test

In a well known procedure using probability paper, data are plotted to see if they form a straight line. A straight line indicates that the data are distributed with the same function as was used to construct the paper. This procedure is commonly used with normally distributed data and normal probability graph paper. The usual procedure plots the observations on the linear scale and cumulative proportions on the non-linear scale. For small samples one may plot the individual observations against $\frac{i}{n+1}$ for $i=1,2, \ldots, n^{8}$

Epstein shows that a similar plotting procedure can be used for testing exponentials. He suggests that the ordered times statistics $t_{1},<t_{2},<t_{3} \ldots<t_{n}$ be plotted against the quantity.

$$
\begin{equation*}
y=\ln \left(\frac{1}{1-F\left(t_{i}\right)}\right) \quad \text { where } i=1,2, \ldots, n_{0} \tag{1}
\end{equation*}
$$

where $F\left(t_{i}\right)=\frac{i}{n+I}$ and where $y$ is the ordinate.
If the exponential assumption is correct, the plotted points will be fitted well by a straight line passing through the origin with slope equal to the reciprocal mean of the $t_{i}{ }^{\text {' }} \mathrm{s}$,

[^4]but as Cooper ${ }^{9}$ points out:
"One criticism of this procedure is that no
formal test of straightness exists at present".
The expression (l) is developed from the cumulative density function for the exponential $F(t)$ given by
\[

$$
\begin{aligned}
F(t) & =0 \\
& =1-e^{-\lambda t} \quad \text { if } t \geqslant 0
\end{aligned}
$$
\]

and it reduces to,

$$
y=\ln \left(\frac{n+1}{n+1}\right)
$$

for ease of computation.
The data may be plotted in two ways. Plain graph paper may be used to plot the natural logarithms of the ordinate with the ordered times $t_{1}{ }^{\prime} t_{2}, \ldots, t_{n}$. Taking natural logarithms can be avoided by plotting $\frac{n+1}{n+1-1}$ with the times on semi-log graph paper. This latter procedure is recommended for the normal user in business since it considerably reduces the time required to arrive at a satisfactory answer and avoids the conversion of data where mistakes can occur. Figure 2 shows both exponential data and uniformly distributed data plotted on semi-log paper.

It must be further emphasized that this procedure is not a test as such, but a method of providing an indication that the exponential assumption may be justified.

We now extend this graphical procedure by applying some analytical techniques and tests to develop a measure of how
${ }^{9}$ Ibid. . p. 83.

well the sample data is approximated by the exponential distribution.

The theoretical straight line fitted to the plotted points of an exponential has slope equal to the exponential parameter and passes through the origin. There are a number of analytical procedures available for describing plotted data and fitted straight lines. Those considered are:
a. correlation coefficient
b. a regression line $y=a+b x$
c. a t-test for the ordinate

## a) Correlation Coefficient

This is a measure of scatter indicating the degree of relationship between one variable and another. It is comparatively easy to calculate with the aid of an adding machine or mechanical desk calculator and will provide the deg̣ree of relationship between sample times and the statistic $y=\ln \left(\frac{n+1}{n+1-i}\right)$. The user would have pre-computed $y^{\prime} s$ with the sum and sum of squares for several sample sizes. After a sample of a specific size has been gathered, computation of a correlation coefficient would produce a statistic that the analyst could compare to some pre-computed standard confidence intervals. The analysis of the sample generated for this thesis produced a marked difference in correlation coefficient for exponential and non-exponential distributions, indicating that this may be a valid test. Results of application are presented in a later section of this chapter.
b) A Regression line $y=a+b x$

A more precise result is obtained by fitting a regresm sion line to the data. Here the line should be $y=0+b x$ where $b$ is equal to one over the mean of the distribution. We are limited in the analysis of the regression by our lack of knowledge of population parameters. We do know however that the population alpha is zero and a tatest can be used. The $b$ coefficient cannot be tested as the population mean is unknown. A test for linearity is not possible since two or more $y$ values do not occur for each $t_{i}$; there is a one, to one correspondence.
c) A t-test for the Ordinate

If we assume the underlying process to be exponential, the population ordinate will be zero. With the sample standard error of $a, ~ a t-t e s t$ is constructed such that if $|t|$ exceeds the critical value $t \alpha / 2, n-2$ shown in tables, we reject the null hypothesis that the casual distribution was exponential. The following test statistic is computed and the result of application to the sample data follow in a later section of this chapter.

$$
t=\frac{a-0}{S_{a}}=\frac{a}{S_{a}}
$$

The $X^{2}$ test for goodness of fit is probably the best known test. It to has limitations related to the amount of data available. A rule of thumb found in many references requires the number of expected observations in any one class to be at least 5 before the chimsquare test can be considered accurate. Parl gives additional limitations on sample size. ${ }^{10}$ He suggests that the test gives reasonably good results in the case of larger samples:

> -o"defined for present purposes as consisting of at least loo items. Samples of less than 50 items are generally considered unreliable in many applications".

Cochran suggests that the restriction on the expected number in each class can be relaxed to at least one. ${ }^{11} \mathrm{He}$ also suggests that the combining of classes weakens the sensitivity of the chimsquare test.

The chi-square is used in this work to test the hypothesis that the sample data came from the Poisson. Classes were combined to ensure at least one expected observation in each cell. Since the mean is used in generating the Poisson is the sample mean, the number of degrees of

10
Boris Parl, Basic Statistics, New York, Doubleday and Company Inc, 1967, p.199.
$11_{\text {George }}$ W. Snedecor and William G. Cochran, Statistical Methods, Iowa, Iowa State University Press, 1967 , p. 235.
freedom must be reduced by this one estimated parameter. 12 The results of these testsfollow.

Results of Applying Tests to the Sample Data

Each of the procedures or tests described in the previous section was applied to the sample distributions to find out how accurately each test identified those distributions which were generated from the exponential population used in the simulation model. From these results we deduce the relative worth of each test and find that only one the "correlation coefficient test" provides consistent results. In addition the coefficient of variation while not a formal test statistic by itself, is described because it helps to detect those distributions that come from the exponential population. A procedure for testing sample data useful to the practitioner as well as pre-computed data will be found in Appendix II.

## a) Correlation Coefficient

We have previously discussed the correlation coefficient as a measure of scatter. From the computed statistics and the times $t_{i}$ the correlation coefficient was computed. The resulting coefficients for each exponential and nonexponential distribution are collected in Table $I$.

[^5]Frequency Distribution of Correlation Coefficient
for Exponential and General Sample Distribution

|  | Exponential |  | General |  |
| :---: | :---: | :---: | :---: | :---: |
| Sample size | 36 | 100 | 36 | 100 |
| Correlation Coefficient |  |  |  |  |
| . 820-. 829 |  |  | 1 |  |
| .830-.839 |  |  | 22 | 1 |
| .840-.849 |  |  | 0 | 1. |
| .850-.859 |  |  | 4 | 6 |
| .860-.869 |  |  | 4 | 4 |
| .870-. 879 |  |  | 4 | 8 |
| . $880-.889$ |  |  | 6 | 10 |
| . 890-. 899 |  |  | 11 | 11 |
| .900-.909 |  |  | 8 | 2 |
| .910-.919 |  |  | 3 | 5 |
| .920-.929 | 1 |  | 3 |  |
| .930-.939 | 0 |  | 1 |  |
| .940-.949 | 1 |  | 1 |  |
| .950-.959 | 0 |  |  |  |
| .960-.969 | 2 | 1 |  |  |
| .970-. 979 | 3 | 2 |  |  |
| .980-.989 | 8 | 9 |  |  |
| .990-.999 | 9 | 12 |  |  |
| Mean | . 9797 | . 9883 | . 89006 | . 8818 |
| St. Dev. | . 0166 | . 0078 | . 0272 | .0196 |
| 95\% Confidence | 1.0000 | 1.0000 | . 9440 | . 9202 |
| Limits | . 9472 | . 9729 | . 9372 | . 8434 |

An increase in sample size considerably reduces the standard error of the correlation coefficient and may provide a significant test with a yet to be determined relationship between them. This needs to be tested with additional distributions and with differing sample sizes.

It would appear that the significant mark in the current data is 093 , a coefficient greater than this figure indicating exponentiality. Even with a sample size of 36 the type 1 and If errors shown in Table II are not significant in this case and may prove not to be so in general. The 1.96 standard deviation confidence limits around the mean correlation coefficient are shown. It should be noted that a significant difference exists between the lower limit for the non-exponential. This gives further encouragement to the hypothesis that the correlation coefficient is a useful test statistic for detecting exponential distributions and warrants further investigation.

A supplementary indicator of exponentiality is found in the coefficient of variation. This was used and tested in much the same manner as the correlation coefficient and is presented here as an additional statistic to be used with the former recognizing that using it alone it can give very misleading results. This quantity is a relative measure of dispersion where the standard deviation is expressed as a percentage of the mean.

Using the same procedure as before the coefficient of variation for each sample size are grouped in distributions.

TABLE II
Type 1 and II Errors from Testing for Exponentials with Correlation Coefficient

Null Hypothesis $H_{0}:$ Distribution is Exponential if.c.c. > 930

| Sample size | Type 1 Error |  | Type II Error |  |
| :---: | :---: | :---: | :---: | :---: |
|  | No. $/ 24$ * | $\%$ | No. $148 *$ | \% |
| 36 | 1 | 4.2 | 2 | 4.2 |
| 100 | 0 | 0 | 0 | 0 |

*The difference between these quantities is due to the greater number of combinations of queue type with general distributions

These are shown in Table III. Again the exponential can be differentiated. The theoretical coefficient of variation for an exponential distribution is 1 , and all coefficients found for the sample non-exponential distributions were less than .8 . It is suggested that a confidence band of $1 . \pm .3$ could be used to accept random data as being from an exponential. The crucial point to be remembered here is that while a theoretical exponential distribution has a coefficient of variation of one, (from $\left.c_{v}=s_{0} d_{0} / m e a n\right)$ : other non-exponential distribution can also have a coefficient of one Thus only when the correlation coefficient is close to 1.0 can the coefficient of variation be validly used to confirm exponentiality。

This coefficient is used however, as a formal test for the IFR condition in a general distribution as explained in Appendix $I$. This is not to be confused with the use being made of this quantity above.

The Type I and Type II errors resulting from applying the coefficient to the sample data is shown in Table IV. It is concluded that inferences should be drawn from this table cautiously and the coefficient should be applied with the same degree of caution for the reasons mentioned above.

Frequency Distribution of Coefficient of Variation for Exponential and General Sample Distributions

|  | Exponential |  | General |  |
| :---: | :---: | :---: | :---: | :---: |
| Sample size | 36 | 100 | 36 | 100 |
| Coefficient of Variation |  |  |  |  |
| .1- . 199 |  |  | 2 |  |
| .2-. 299 |  |  | 5 | 6 |
| .3-. 399 | 1 |  | 14 | 12 |
| .4-. 499 | 0 |  | 11 | 13 |
| .5-. 599 | 0 |  | 13 | 16 |
| .6-. 699 | 4 | 2 | 3 | 1 |
| .7-.799 | 3 | 3 |  |  |
| .8-.899 | 5 | 3 |  |  |
| .9-.999 | 3 | 2 |  |  |
| 1.0-1.099 | 3 | 6 |  |  |
| 1.1-1.199 | 1 | 5 |  |  |
| 1.2-1.299 | 3 | 2 |  |  |
| 1.3-1.399 | 1 | 0 |  |  |
| 1.4-1.499 |  | 0 |  |  |
| 1.5-1.599 |  | 1 |  |  |
| Mean | . 9042 | 1.0080 | . 4297 | . 4404 |
| St Dev. | . 2267 | . 2115 | . 1203 | . 1064 |
| 95\% Confidence | 1.3492 | 1.4158 | . 6657 | . 6464 |
| Limits | . 4492 | . 5858 | . 1937 | . 2344 |


b) A Regression line $y=a \neq b x$

No direct use was made of this regression line except in connection with the t-test for the ordinate which is discussed in the next sub-section. Tests of linearity could not be performed for reasons mentioned previously and the lack of knowledge of the population mean prevented an analysis of variance and F-tests.

## c) $t$-Test

This test provided the least interesting results and no conclusions could be drawn except that the test is unsatisfactory. The results are summarized in Table $V$ and are inconclusive since both types I and II errors were plentiful.

## d) Chi-Squared Test

This test did not produce satisfactory results either. It was possible to manipulate the results at will when constructing a Poisson distribution.

Arrival rate distributions were constructed from the sample times using a time unit chosen for its convenience. Initially the time unit was chosen to give the same number of units as events, thereby making the arrival rate equal to 1.0. If this was always done, a user would only need one set of figures for the theoretical distribution thus making his job of testing easier. This test produced such inaccu-

| O | TABLE V <br> Type I and II Errors from t-Test that Ordinate of Fitted Regression line Equals zero* |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Ho: $\mathrm{a}=0$ |  |  |  |  |
| Type I Error Type II Error |  |  |  |  |
| Sample size | No. $/ 24$ | \% | No. $/ 48$ | \% |
| 36 | 14 | 58.0 | 22 | 46.0 |
| $100$ | 22 | 92.0 | 4 | 8.4 |

rate results however that a search was made for a better time unit. No satisfactory time unit was found as can be seen from Table VI which shows the results from chi-square tests on the sample data for eight different time unit values. The eight values in the left-hand column show the number of arrivals per time unit; the actual time unit being a function of the total elapsed time of inter-arrival times over the sample and the number of arrivals. As in the t-test these results are inconclusive.

Summary

There does not appear to be a simple test for exponential distributions that is satisfactory under a variety of conditions especially small sample sizes. For guidance the graphical test is useful. The test based on the correlation coefficient has promise but needs more extensive examination and trial with different sample sizes and a larger selection of distributions. It would be of interest to try some nonexponential distributions constructed to have a coefficient of variation close to one.

The test for the IFR and DFR properties described in Appendix $I$, using the coefficient of variation, appears to be the only reliable test at present. The significance of IFR and DFR properties will become apparent in Chapter IV and therefore deserve reference here due to that importance. A procedure for testing a distribution, easily used by a non-technical person, is shown in Appendix II. This is

## TABLE VI

Type 1 and II Errors from ChimSquare Test* of Rate Distributionsof Sample Data

| Ave rate per time unit** | Sample size |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | 36 |  | 100 |  |
|  | Type 1/24 | Type II/24 | Type 1/48 | Type 1I/48 |
| . 05 | 0 | 48 | 0 | 48 |
| .10 | 0 | 48 | 0 | 48 |
| . 30 | 5 | 33 | 13 | 6 |
| . 50 | 11 | 6 | 17 | 0 |
| . 75 | 9 | 10 | 19 | 0 |
| 1.00 | 13 | 1 | 18 | 0 |
| 1.25 | 16 | 2 | 19 | 1 |
| 1.50 | 14 | 4 | 20 | 0 |

*at . 01\% level of Significance
**see text for explanation of time unit.
the "correlation coefficient test" and uses pre-computed data points for comparison. The method of calculation. with formulae shows a potential user the steps required to arrive at conclusive evidence based on this test as to the nature of the distribution.

## METHODS ROR THE ANALYSIS OF QUEUES

In this chapter we examine the solution methods used to find queue length and waiting times.

Assumptions

A broad and not completely justifiable assumption used through all the queue analysis is that of steady state. This permits the use of equilibrium solutions. However, the majority of situations faced by the small business operator may not be steady state. An example is a spare parts counter where the pace of business increases in the early morning then decreases over the lunch period and may peak again in the afternoon. These problems can be overcome by isolating each period with changed arrival patterns and performing separate analyses. However, this does not fully account for transient behaviour.

M/M/1 Queue

$$
\text { Expected system length, } \begin{align*}
L & =\frac{\rho}{1-\rho}  \tag{1}\\
\because \operatorname{Var}(L) & =\frac{\rho}{(1-\rho)^{2}} \tag{2}
\end{align*}
$$

$$
\begin{align*}
\text { Expected queue length, } L q & =\frac{\rho^{2}}{1-\rho}  \tag{3}\\
& =  \tag{4}\\
\operatorname{Var}(L q) & =\frac{\rho^{2}+\rho^{3}-\rho^{4}}{(1-\rho)^{2}}
\end{align*}
$$

M/G/1 Queue

These are the Pollaczek - Khintchine formulae:

Expected system length, $L=\rho+\frac{\lambda^{2} \sigma_{0}^{2}+\rho^{2}}{2(1-\rho)}$

Expected queue length, $L q=\frac{\lambda^{2} \sigma^{2}+\rho^{2}}{2(1-\rho)}$
G/G/1, G/M/1 Queues

The solution of queue length and waiting time for the general case is complex and the exact formulae depend on the moments of the idle time distribution as shown by Marshall. 13 The expected wait in queue is found to be:

$$
\begin{align*}
W q & =\frac{E\left(U^{2}\right)}{2 E(U)}-\frac{E\left(I^{2}\right)}{2 E(I)} \\
& =\frac{\lambda^{2}\left(\sigma_{a}^{2}+\sigma_{Q}^{2}\right)+(1-\rho)^{2}}{2 \lambda(I-\rho)}-\frac{\nu_{h}(2)}{2 V_{h}} \tag{7}
\end{align*}
$$

where $\mathrm{Un}=\mathrm{Sn}-\mathrm{Tn}$. The idle time distribution is

13 Marshall, p. 4.
found from the fundamental identity,

$$
\begin{aligned}
W_{n+1} & =\max \left[0, W_{n}+U n\right] \\
\text { But letting } X_{n} & =-\min [0, W n+U n] \\
W_{n+1}-X_{n} & =W n+U n
\end{aligned}
$$

From the above where $\mathrm{Xn}>0 \quad \mathrm{Xn}=\mathrm{I}$
The variance of the expected wait in queue is shown by Marshall to be

$$
\begin{align*}
\sigma_{W q}^{2}= & \frac{E(U)^{3}}{-3 E(U)}+\left(\frac{E\left(U^{2}\right)}{-2 E(U)}\right)^{2}+\frac{E\left(I^{3}\right)}{3 E(I)}-\left(\frac{E\left(I^{2}\right)}{2 E(I)}\right)^{2} \\
= & \frac{\lambda\left(v_{G}^{\left.(3)-v_{a}^{(3)}\right)+3\left(\rho v_{G}(2)-v^{(2)}\right)}\right.}{} \begin{aligned}
& 3(1-\rho) \\
&+\left(\frac{\lambda^{2}\left(\sigma_{a}^{2}+\sigma_{g}^{2}\right)+(1-\rho)^{2}}{2 \lambda(1-\rho)}\right)^{2}-\sigma_{e, h}^{2}
\end{aligned}
\end{align*}
$$

where $\sigma_{e, h}^{2}=\frac{v_{h}^{(3)}}{3 v_{h}}-\left(\frac{v_{h}}{2} \psi_{h}^{(2)}\right)^{2}$.
From successive inter-arrival and service times (Tn
and Sn ) one can compute Xn and the moments of the idle time distribution. While a tedious task by hand, they can be derived easily on a computer; such was our case. It is doubtful that the average user under consideration would be a computer or compute these quantities by hand. To obviate this problem, Marshall's bounds on expected wait for the $G / G / 1$ queue are shown.

G/G/1 Bounds
a) Upper bound

$$
\begin{equation*}
W q=\frac{\lambda\left(\sigma_{\mathrm{a}}^{2}+\sigma_{\mathrm{g}}^{2}\right)}{2(1-\rho)} \tag{9}
\end{equation*}
$$

By multiplying both numerator and denominator by $\lambda$ this expression becomes

$$
W q=\frac{\sigma_{a}^{2}+\sigma_{G}^{2}}{2(1 / \lambda-1 / \mu)}
$$

which is Kingman's heavy traffic approximation. The latter states:
${ }^{\circ}$ For this approximation to be valid, the denominator must be small compared to the square root of the numerator". ${ }^{14}$

This bound is expected to be high for rho values of 09 or less, although it will be accurate as an equality for rho greater than . 9 .

This bound is particularly useful because it only depends on the first two moments of the inter-arrival and service time distribution.
b) Lower bound

The lower bound $\ell<E(W)(W q)$ is found as the solution of
$x=\int_{-x}^{\infty} K^{C}(u) d u$, where $U n=(S n-T n) \sim K(u)$.
This bound is shown by Marshall to be unique for rho less than 1.

The computation of $\mathrm{K}^{c}(u)$ requires the convolution of Sn and Tn .

14 J.F.C. Kingman, "The Heavy Traffic Approximation in the Theory of Queues", Proceedings of the Symposium on Conjestion Theory, ed Walter L. Smith and William J. Wilkinson, Chapel Hill, The University of North Carolina Press, 1965, p. 137-157.

## Computation of lower bound

As an alternative to computing the convolution of ( $\mathrm{Sn}-\mathrm{Tn}$ ) the distribution of $\mathrm{K}(\mathrm{u})$ is found from each sample using empirical density functions.

$$
\text { Let } T \sim \hat{F}_{T}(x) \text { and } S \sim \hat{F}_{S}(x)
$$

If
then

$$
\hat{F}_{\mathrm{T}}(x)=\left\{\begin{array}{cc}
0 & x<x_{1} \\
1 / T & x_{1}<x<x_{2} \\
\vdots & \\
1.0 & x_{T T}<x
\end{array}\right.
$$

$$
\hat{F}_{T}(x)=\frac{1}{E} \sum_{i=1}^{t} U\left(x_{i}-X\right) \text { where } U \text { is the unit step fan. }
$$

Then

$$
f_{x}^{*}(s)=\frac{1}{t} \sum_{i=1}^{t} e^{-S x_{i}}
$$

The Laplace transform of $U_{n}=S_{n}-T_{n}$
is

$$
f_{s}^{*}(s) \cdot f_{T}^{*}(-s)
$$

which equals

$$
\begin{aligned}
& =\frac{1}{n} \sum_{i=1}^{n} e^{s(S i)} \cdot \frac{1}{n} \sum_{j=1}^{n} e^{-s(T j)} \\
& =\frac{1}{n^{2}} \sum_{i, j=1}^{n^{2}} e^{s(S i)}-s(T j) \\
& =\frac{1}{n^{2}} \sum_{k=1}^{n^{2}} e^{-s(U k)} \quad k=0,1, \cdots, n^{2} \\
& =f_{u}^{*}(s) \\
& =\hat{F}_{n}(U)
\end{aligned}
$$

We now solve for x the lower bound in

$$
x=\int_{-x}^{\infty}\left(1-\hat{F}_{n}(u)\right) d u
$$

which is solved computationally:

$$
x=\sum_{m=n^{2}}^{x} U_{m}
$$

where

$$
U_{n}=S_{1}-T_{1} \text { and } U_{n}{ }^{2}=S_{n}-T_{n}
$$

Thus the bound can be found very easily from the sample data.

One great asset in this lower bound is the selfgeneration of data points for the $K^{C}(u)$ distribution. For each $n$ arrivals $n^{2}$ points are used in the distribution ${ }^{2}$ thus for a very small amount of data quite a precise distribution can be expected.

IFR/G/1, DFR/G/1 Queues

Tighter bounds on the expected wait in queue can be found for these two classes. (See Appendix I)

The bounds on $\operatorname{IFR}(D F R) / G / 1$ queues are generally stronger than those mentioned previously, bounding the expected number in the queue to at most, one customer.
a) $I F R / G / 1$ Queue

$$
\begin{gathered}
J-\frac{\left(\rho+C_{a}^{2}\right)}{2 \lambda}<W_{q}<J \\
\lambda J-\frac{\left(\rho+C_{a}^{2}\right)}{2}<L_{q}<\lambda J
\end{gathered}
$$

where

$$
J=\frac{C_{a}^{2}+\lambda^{2} \sigma_{g}^{2}}{2 \lambda(1-\rho)}
$$

b) $\mathrm{DFR} / \mathrm{G} / 1$ Queue.

$$
\ell<W q<J-\frac{\rho+C_{a}^{2}}{2 \lambda}
$$

and

$$
\lambda \ell<L_{q}<J-\left(\rho+C_{a}^{2}\right) / 2
$$

where $J$ is as above and $\ell$ is the solution to the lower bounds in the G/G/1 case.

Expected Wait and Queue Length

Comparative queue characteristics can easily be computed from the following relationships:

$$
\begin{aligned}
L & =\lambda W^{l 5} \\
L q & =\lambda W q \\
W & =W q+\frac{1}{\mu} \\
L q & =L-\rho
\end{aligned}
$$

## Computation

All formulae shown in this section are coded for solution on the computer. This is a straight forward task except for the solution to the lower bound equation (10) which was found, as explained, through the use of the empirical density function.

[^6]
## CHAPTER IV

QUEUE ANALYSIS

In this chapter we apply the theoretical methods to the sample data generated from the simulation model in order that we may estimate queue length. These estimators will then be compared to the simulated real world solution and their accuracy evaluated.

There are three steps in this process. The firstwis to establish the accuracy of the sample parameters as estimates of the population parameters. This is important since many of the classical queueing calculations depend solely on the expeated values of the population distributions. Only the calculations for the general lower bound (10) ${ }^{16}$ and the moments of the idle time distribution on the other hand, rely directly on the interaction between specific sequential inter-arrival and service times, as can be appreciated from $U_{n}=S_{n}-T_{n}$. The importance of the first moment of the sample distributions is emphasized by its use in the "correlation coefficient test" and in the computation of rho which, of course, subsequentlyaffects queue length estimates.

The second step is an evaluation of the queueing calculation methods themselves. Since some serious errors do result when estimating the population distribution parameters only those queue with a calculated rho within $5 \%$ of the simulated

[^7]rho will be used in this step. This will give us a measure of the ability of the various methods chosen for estimating queue length to perform under the same parameters that the simulation used. One would not expect to find disagreement between the simulated and calculated queue length for these samples.

The third step is to evaluate the overall result. Having established some knowledge of the behaviour of the methods used in queue length estimation under optimum conditions we will better understand the errors in estimating queue length resulting from errors in estimating distribution parameters. Runs which have an estimated rho considerably in error from the simulated rho will be examined to find the best method of approach to minimize overall error in estimating queue length.

## Step One: Evaluation of Sampling Errors

Since the majority of queue operating characteristics depend on the expected inter-arrival and service times through their composition of rho, and as the Pollaczek-Khintchine formulae depend on the coefficient of variation squared as well, it is most desireable to accurately predict these expected values if one anticipates their successful application in queueing analysis.

In some exploratory simulation runs, to test the operation of the experiment, large discrepancies between the simulation parameter mean and the sampled mean occurred. It was thought that this was due to the integer form of data with a low mean of
ten. To prevent distortion or bias from the simulation, the mean was raised to one thousand, in the case of inter-arrival times, to produce integers with three significant digits. This new data was then read by the queue programs with the decimal point displaced two positions left. The new data then was accurate to one one hundredth of a time unit. The blame for "bad" data could not now be laid with the simulation We should note at this time that the random number generator is considered to be sufficiently random that results will not be biased by cycling of the random number string. The IBM standard random number generator is tested to generate two to the twenty-ninth digits before repeating.

Before we look at the specific figures for lamda, mup and rho, consider the effects on rho of a change in lamda and mu. As would be expected from the well known relationship $\rho=\lambda / \mu$, a fractional change in lamda will cause a proportional change in rho. Thus if the arrival rate is increased by ten percent, the traffic intensity (rho) will increase by the same ten percent. This is not true with mu. A fractional change in mu, denoted $b$, will cause a change in rho equal to $-\frac{\mathrm{b}}{\mathrm{b}+1}$. This means that a $20 \%$ increase in the service rate will improve the traffic intensity by only $16.7 \%$. The combined effect of change in lamda and mu, denoted by $a$ and $b$ respectively, on rho can be expressed as $c=\frac{a-b}{b+1}$, where $c$ is the fractional change in rho. This relationship is very pertinent to the application

17 Subroutine RANDU, IBM System/360 Scientific Subroutine Package Version III, White Plains, N. Y., IBM Corp.
of queueing results when considering the cost of altering or service rates. It is mentioned here solely to aid the reader in understanding how errors in the estimation of population parameters can effect the subsequent queue analysis. Essentially errors in mu will have less effect on rho although the diminishing effect is only obvious when errors in mu are larger than approximately ten percent.

An additional problem concerns the direction of change in rho caused by the combined errors in lamda and mu. The following Figure 3 shows the directional change in lamda and mu.

| Sign of $a$ | Sign of $b$ | Change in <br> rho | Condition |
| :---: | :---: | :---: | :---: |
| + | + | + | $a>b$ |
| + | + | - |  |
| + | + | - | $a>b$ |

Fig. 3 Effective direction of change in rho, for changes of lamda and mu.

In the following presentation of the actual data, all errors are shown as a percent error from the intended value from the simulation model. This gives us a unit of comparison that is common to all quantities.

Since each sample of data was generated independently, each estimated rho is considered as a separate entity and while the mean percent error overall in the values of rho may
be close to zero, this is not an indicator of accuracy as one can see from the data presented. Tables VII and VIII show sample values as well as percent error of lama, mu, and rho for the two sample sizes.

To help the reader in interpreting these tables, errors in the estimation of rho, shown as a percent of the intended values of rho in the simulation model, are grouped by their deviation from zero in five percent steps ignoring the direction of error. Figure 4 below shows this grouping for the two sample sizes.

| \% Error <br> Class <br> Boundries | Frequency by 36 | Sample 100 | Size |
| :---: | :---: | :---: | :---: |
| 0-4.9 | 7 | 15 |  |
| $5-9.9$ | 12 | 7 |  |
| $10-14.9$ | 8 | 6 |  |
| 15-19.9 | 2 | 3 | - |
| 20-24.9 | 3 | 2 |  |
| 25-29.9 | 1 | 2 |  |
| 30-34.9 | 2 | 1 |  |
| 35-39.9 | 1 |  |  |
|  | 36 | 36 |  |

Fig. 4 Distribution of Percent Error in Rho by Sample Size.

It should be emphasized here that the error in rho is calculated on the intended value of rho and not the actual simulated result. As the largest deviation of the simulated rho from the intended value was less than one percent in all but two cases, both of whose deviations were within $1.5 \%$, the intended values are considered sufficiently accurate. Use of the intended values also aides in the presentation of data where, for example, 9 is

## Key to Tables VII and VIIr

## Column

1. Run Number

2 Intended Rho
3 Estimated Rho
4 Percent Error in Estimating Rho
5 Intended Lamda
6 Percent Error in Estimating Lamda
7 Intended Mu
8 Percent Error in Estimating Mu

Percent Error of Rho Lamda, and Mu (Sample Size 36)

| $(1)$ | $(2)$ | $(3)$ | $(4)$ | $(5)$ | $(6)$ | $(7)$ | $(8)$ |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 1 | 0.9 | 1.130 | 25.555 | 0.1 | 20.000 | 0.111 | -3.700 |
| 2 | 0.8 | 0.745 | -6.875 | 0.1 | 7.000 | 0.125 | 14.400 |
| 3 | 0.7 | 0.855 | 22.143 | 0.1 | 19.000 | 0.143 | -2.700 |
| 4 | 0.6 | 0.590 | -1.667 | 0.1 | -1.000 | 0.167 | 0.800 |
| 5 | 0.9 | 0.979 | -8.778 | 0.1 | 7.000 | 0.111 | -1.900 |
| 6 | 0.8 | 0.741 | -7.375 | 0.1 | -8.000 | 0.125 | -0.800 |
| 7 | 0.7 | 0.779 | 11.286 | 0.1 | 11.000 | 0.143 | -0.600 |
| 8 | 0.6 | 0.624 | 4.000 | 0.1 | 1.000 | 0.167 | -2.800 |
| 9 | 0.9 | 0.921 | 2.333 | 0.1 | -12.000 | 0.111 | -13.600 |
| 10 | 0.8 | 0.846 | 5.750 | 0.1 | 6.000 | 0.125 | 0.800 |
| 11 | 0.7 | 0.472 | -32.571 | 0.1 | -22.000 | 0.143 | 15.500 |
| 12 | 0.6 | 0.532 | -11.333 | 0.1 | 11.000 | 0.167 | 24.800 |
| 13 | 0.9 | 1.060 | 17.77 .8 | 0.1 | 1.000 | 0.111 | -14.500 |
| 14 | 0.8 | 0.850 | 6.250 | 0.1 | 0.0 | 0.125 | -5.600 |
| 15 | 0.7 | 0.624 | -10.857 | 0.1 | -2.000 | 0.143 | 9.900 |
| 16 | 0.6 | 0.553 | -7.833 | 0.1 | -7.000 | 0.167 | 1.400 |
| 17 | 0.9 | 0.980 | 8.889 | 0.1 | 5.000 | 0.111 | -2.800 |
| 18 | 0.8 | 0.856 | 7.000 | 0.1 | 0.0 | 0.125 | -6.400 |
| 19 | 0.7 | 0.700 | -0.000 | 0.1 | 2.000 | 0.143 | 1.500 |
| 20 | 0.6 | 0.601 | 0.167 | 0.1 | -5.000 | 0.167 | -5.200 |
| 21 | 0.9 | 0.694 | -22.889 | 0.1 | -18.000 | 0.111 | 6.200 |
| 22 | 0.8 | 0.703 | -12.125 | 0.1 | 2.000 | 0.125 | 16.000 |
| 23 | 0.7 | 0.781 | 11.571 | 0.1 | 9.000 | 0.143 | -2.000 |
| 24 | 0.6 | 0.542 | -9.667 | 0.1 | -6.000 | 0.167 | 3.800 |
| 25 | 0.9 | 0.889 | -1.222 | 0.1 | 9.000 | 0.111 | 3.800 |
| 26 | 0.8 | 1.070 | 33.750 | 0.1 | 16.000 | 0.125 | -12.800 |
| 27 | 0.7 | 0.844 | 20.571 | 0.1 | 19.000 | 0.143 | -1.300 |
| 28 | 0.6 | 0.505 | -15.833 | 0.1 | -7.000 | 0.167 | 9.800 |
| 29 | 0.9 | 0.811 | -9.889 | 0.1 | -5.000 | 0.111 | 6.200 |
| 30 | 0.8 | 0.693 | -13.375 | 0.1 | -17.000 | 0.125 | -4.800 |
| 31 | 0.7 | 0.957 | 36.714 | 0.1 | 26.000 | 0.143 | -7.600 |
| 32 | 0.6 | 0.519 | -13.500 | 0.1 | -7.000 | 0.167 | 7.400 |
| 33 | 0.9 | 1.020 | 13.333 | 0.1 | 6.000 | 0.111 | -6.600 |
| 34 | 0.8 | 0.765 | -4.375 | 0.1 | -6.000 | 0.125 | -1.600 |
| 35 | 0.7 | 0.640 | -8.571 | 0.1 | -1.000 | 0.143 | 8.500 |
| 36 | 0.6 | 0.545 | -9.167 | 0.1 | -12.000 | 0.167 | -2.800 |

TABLE VIII
$(1)$
1
2
3
4
5
6
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
(2)
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6
0.9
0.8
0.7
0.6

Percent Error of Rho, Lamda, and Mu (Sample Size 100)
(3)
0.894 1.050
0.631
0.923
0.711
0.608
0.754
0.842
0.689
0.579
0.878
0.714
-0.667
31.
-5.
5
2.
-6
1.
1.

$$
\begin{array}{r}
5 \\
-1 \\
-3 \\
-2 \\
-10
\end{array}
$$

(5) (6)
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.1
0.567
0.589
-11.833
13.333
-4.250
2.571
-10.167
8.333
21.125
-19.000
-1.833
.
1.
8.0
18.0
-5.0
-5.0
-1.0
2.0
10.0
10.0
-10.0

| (7) | (8) |
| :---: | :---: |
| 0.111 | 8.900 |
| 0.125 | -9.600 |
| 0.143 | 5.700 |
| 0.167 | -9.400 |
| 0.111 | -2.800 |
| 0.125 | 8.800 |
| 0.143 | 8.500 |
| 0.167 | 8.000 |
| 0.111 | 7.100 |
| 0.125 | 1.600 |
| 0.143 | 0.800 |
| 0.167 | -1.000 |
| 0.111 | 4.400 |
| 0.125 | 16.800 |
| 0.143 | -7.600 |
| 0.167 | 3.200 |
| 0.111 | 0.800 |
| 0.125 | 0.800 |
| 0.143 | -4.100 |
| 0.167 | 4.400 |
| 0.111 | 5.300 |
| 0.125 | -5.400 |
| 0.143 | 21.800 |
| 0.167 | 16.400 |
| 0.111 | -20.800 |
| 0.125 | 1.600 |
| 0.1 .43 | -21.600 |
| 0.167 | 8.000 |
| 0.111 | -1.900 |
| 0.125 | -3.200 |
| 0.143 | -7.600 |
| 0.167 | 8.000 |
| 0.111 | -1.000 |
| 0.125 | -4.000 |
| 0.143 | 14.800 |
| 0.157 | 6.800 |

$\omega$
sufficiently accurate for .893 and the extra decimals do not provide any significant information.

Estimates of rho from the larger samples are not significantly more accurate than those from the smaller sample. The notable difference lies in the larger number of estimates with errors of five percent or less resulting from the larger sample size. Even so, approximately half of the estimates are still greater than ten percent in error from the real world rho as determined in the simulation model. This leads us to con= clude that no significant benefit can be gained from the increased sample size. It also follows that the large, samples do not improve our knowledge of the accuracy of the smaller samples in their ability to satisfactorily estimate rho. There does not appear to be any evidence here to justify increasing the sample size from 36 to 100 .

Furthering our analysis of the error in estimation of rho, the data from the two sample sizes is grouped in Table IX by intended rho value. We notice that the mean percent error is small for all parameters with no significant differences between the small and large sample results. This proximity to zero for all parameters is expected since group means calculated from both positive and negative values would tend toward zero unless an unusual bias was present at the time of selection of specific inter-arrival and service times. There is no reason to believe that there was.

There appears to be a slight difference between samples in their respective maximum negative errors. Eight of the twelve values in the larger sample are smaller than the

Sampling Errors for Lamda, Mu, and Rho
as Percentage of Intended Simulated Values

| Sample Size: |  |  | 36 |  |  | $100$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |
| Variable |  | $\begin{aligned} & \text { Mean } \\ & \text { \% Error } \end{aligned}$ | $\begin{aligned} & \mathrm{Stn} \\ & \text { Dev. } \end{aligned}$ | $\begin{aligned} & \text { Max }(\phi) \\ & \text { \& Error } \end{aligned}$ | $\begin{aligned} & \text { Max }(-) \\ & \% \text { Error } \end{aligned}$ | $\begin{aligned} & \text { Mean } \\ & \text { \% Error } \end{aligned}$ | $\begin{aligned} & \text { Stn } \\ & \text { Dev. } \end{aligned}$ | Max (+) | Max(-) |
|  |  | \% Error |  |  |  |  |  | \% Error |
| . 9 | $\lambda$ |  | 1.44 | 11.54 | 20.00 | 18.00 | 0.11 | 8.57 | 12.00 | 12.00 |
|  | $\mu$ | -2.29 | 8.56 | 9.80 | 14.50 | 0.0 . | 8.82 | 8.90 | 20.80 |
|  | p | 4.74 | 14.70 | 25.56 | 22.89 | 0.46 | 9.45 | 13.33 | 16.22 |
| . 8 | $\lambda$ | 0.0 | 9.57 | 16.00 | 17.00 | 3.99 | 8.47 | 18.00 | 8.00 |
|  | $\mu$ | -0.08 | 9.52 | 16.00 | 12.80 | . 71 | 8.06 | 16.80 | 9.60 |
|  | p | . 95 | 14.50 | 33.75 | 13.38 | 4.13 | 13.73 | 31.25 | 10.75 |
| . 7 | $\lambda$ | 6.77 | 14.48 | 26.00 | 22.00 | -0.33 | 8.27 | 14.00 | 12.00 |
|  | $\mu$ | 2.35 | 7.36 | 15.50 | 7.60 | 1.18 | 13.16 | 21.80 | 21.60 |
|  | $\rho$ | 5.58 | 20.89 | 36.71 | 32.57 | 0.38 | 17.48 | 27.14 | 27.71 |
| . 6 | $\lambda$ | -3.66 | 6.65 | 11.00 | 12.00 | -0.77 | 5.73 | 10.00 | 8.00 |
|  | $\mu$ | 4.13 | 9.16 | 24.80 | 5.20 | 4.93 | 7.13 | 16.40 | 9.40 |
|  | $\rho$ | 7.20 | 6.63 | 4.00 | 15.83 | -5.15 | 7.90 | 5.17 | 21.00 |

corresponding values in the sample of 36 observations. This trend is reversed in the maximum positive errors where only five of the twelve are smaller. This tends to confirm the lack of any objective evidence of benefits in accuracy in estimating population parameters by increasing sample size. Considering each of the 72 values of lamda, mu, and rho individually, the following Figure 5 verifies that little can be gained from the larger sample in estimating the three parameters.

| Error from <br> 100 size was | Lamda | Mu | Rho |
| :--- | ---: | ---: | ---: |
| Larger | 17 | 19 | 14 |
| Same | 0 | $\frac{1}{2}$ | 0 |
| Smaller | $\frac{19}{36}$ | $\frac{16}{36}$ | $\frac{22}{36}$ |

Fig. 5 Comparison of percent error between sample sizes.

These errors are of course further compounded by the different effects that errors in lamda and mu have on rho, especially with respect to sign. This is noticed most of all in the calculations for $M / M / l$ queues (see Tables VII and VIII) where both lamda and mu have greater error in the larger sample size, yet rho has less error overall, as indicated in the first run of each sample size.

One remembers from Chapter II that the chance of an incorrect identification of a distribution was reduced by using the larger sample size. It will be shown later in this chapter
that identification of the shape of the distribution is important in arriving at accurate estimates of queue length. For this purpose alone the user is justified in taking a larger sample.

## Step 2: Evaluation of Calculation Methods With Accurately Estimated Parameters.

It is important to decide which of the methods shown in the previous chapter we will consider within the grasp of the average user Having settled this, we can then test to see if any are satisfactory in consistently and accurately estimating queue length.

We will assume that the analyst is able to find a best estimate for lamda and mu from the data available. From this the calculation of rho follows. With this the user has access to solutions to all the classical formulae for $M / M / 1$ queues. From the variance of the service time distribution he can compute the coefficient of variation and is subsequently able to solve for $M / G / 1$ queues with the Pollaczek-Khintchine formulae. The computation required for application of the $G / G / 1$ equation for $W q(7)$ is considered too complex to be used in everyday analysis. The procedure for finding the general lower bound is of this class too. The upper $G / G / 1$ bound is dependent on the first two moments of the inter-arrival and service distributions only and is thus usable by the analyst. The bounds on queues that meet the IFR condition (see Appendix I and Chapter II (12)) rely on the same bounds and are also
included. Only the upper bound (13) for the DFR condition is included since the lower DFR bound is the same as the G/G/1 lower bound.

In Step 1 we saw that errors in tho are caused by errors in predicting the estimated arrival and service rates. Before the true values of rho and queue length were revealed to this writer, the estimation and calculation of queue lengths for each of the 72 runs had been completed using the most accurate estimates of the population parameters that could be derived from the sample data. In this section we will evaluate the ability of the methods described in Chapter III to consistently find accurate solutions. This can only be accomplished if one has accurate input data. Otherwise the errors arising from inaccurate estimation of the parameters will interfere with the estimation of queue length.

The sample runs analyzed below all have a rho that was confirmed to be within plus or minus five percent of the true value in the parent simulation run, when the true values were revealed. These twenty two runs of the seventy-two will be considered the "accurate data". Because six estimated rho values exceeded 1.0 they were discarded and the remaining 44
$=$ runs comprise the data which will be examined in Step 3 . Results from Accurate Data

The operating characteristics of these runs are shown in Table $X$. In most cases there is at least one method for which the error in estimation is small while the most consistent

```
Key to Tables X, XI, and XII
```

Column
1 Run Number
2 Queue Type
3. Intended Rho

4 Average Queve Length from Simulation
5 Estimated Queue Length by General Method
6 Percent Error in (5)
7. Estimated Queue Length from Classical Formulae

8: Percent Error in (7)
9 IFR Lower Bound, DFR bounds marked \#
10 . IFR Upper Bound, DFR bounds marised \#
11 Percent Error in the Mean of (9) and (10) compared to (4)

## Sample Size 36

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 4 | M/ M/I | 0.6 | 0.82 | 0.82 | 0.61 | 0.85 | 3.92 | 0.69 | 1.39 | 27.23 |
| 8 | M/G2/1 | 0.6 | 0.64 | 0.64 | 1.10 | 0.72 | 13.34 | 0.38 | 1.01 | 9.73 |
| 9 | M/G3/1 | 0.9 | 5.37 | 7.50 | 39.65 | 7.56 | 40.86 | 7.10 | 8.03 | 40.91 |
| 19 | G2/G2/1 | 0.7 | 0.50 | 0.66 | 31.60 | 1.63 | 225.80 | 0.39 | 0.91 | 30.40 |
| 20 | G2/G2/1 | 0.6 | 0.28 | 0.22 | -20.85 | 0.90 | 224.32 | 0.02 | 0.48 | -10.43 |
| 25 | C3/ M/1 | 0.9 | 5.17 | 6.77 | 31.00 | 7.12 | 37.79 | 6.65 | 7.40 | 35.93 |
| 34 | G3/G3/1 | 0.8 | 1.85 | 1.21 | -34.84 | 2.49 | 33.98 | 0.97 | 1.59 | -31.02 |
| Sample Size 100 |  |  |  |  |  |  |  |  |  |  |
| 1 | $\mathrm{M} / \mathrm{M} / 1$ | 0.9 | 7.74 | 8.90 | 15.10 | 7.50 | -3.02 | 0.32 | 4.63 | -68.02 |
| 5 | $\mathrm{M} / \mathrm{C} 2 / 1$ | 0.9 | 4.10 | 7.83 | 91.05 | 7.24 | 76.57 | 0.25 | 1.06 | -84.10 |
| 7 | M/G2/1 | 0.7 | 1.07 | 1.12 | 4.10 | 1.29 | 20.50 | 1.17 | 1.99 | 47.06 |
| 8 | M/G2/1 | 0.6 | 0.64 | 0.61 | -3.77 | 0.65 | 1.73 | 0.13 | 0.47 | -52.83 |
| 11 | $\mathrm{M} / \mathrm{G} / 1$ | 0.7 | 1.18 | 0.90 | -23.79 | 1.17 | -0.59 | 0.17 | 1.15 | -43.86 |
| 12 | M/G3/1 | 0.6 | 0.67 | 0.58 | $-13.25$ | 0.60 | -10.13 | 0.43 | 1.15 | 17.36 |
| 13 | G2/G3/1 | 0.9 | 3.08 | 2.85 | -7.50 | 6.30 | 104.64 | 2.54 | 3.16 | -7.49 |
| 16 | G2/G3/1 | 0.6 | 0.34 | 0.41 | 21.89 | 0.82 | 143.79 | 0.20 | 0.70 | 34.32 |
| 18 | G2/G2/1 | 0.8 | 0.97 | 0.92 | -5.25 | 2.71 | 179.30 | 0.62 | 1.18 | -7.26 |
| 20 | G2/G2 / 1 | 0.6 | 0.28 | 0.27 | -4.32 | 0.84 | 201.08 | 0.05 | 0.51 | 0.72 |
| 21 | G2/ M/1 | 0.9 | 5.48 | 4.58 | -16.55 | 5.89 | 7.38 | 4.16 | 4.81 | -18.30 |
| 26 | G3/ M/1 | 0.8 | 2.49 | 2.18 | -12.15 | 2.92 | 17.38 | 2.05 | 2.69 | -4.61 |
| 30 | G3/G2/1 | 0.8 | 1.36 | 0.94 | -30.98 | 2.51 | 84.21 | 0.68 | 1.29 | -27.64 |
| 31 | G3/G2/1 | 0.7 | 0.69 | 0.67 | -2.62 | 1.82 | 165.65 | 0.44 | 1.04 | 8.08 |
| 36 | 63/G3/1 | 0.6 | 0.43 | 0.48 | 12.18 | 0.84 | 97.42 | 0.27 | 0.84 | 29.39 |

method overall is the G/G/I solution. This, unfortunately, is one method assumed too complex for the average user. The predictions shown in Table $X$ will be evaluated in two groups; those with exponential inter-arrivals first, followed by the general arrival queues. To make reference to the various solution methods in the following discussion easier, the solution method for the $M / M / 1$ or $M / G / 1$ queue type will be referred to as the exponential solution similarly the solution method for the $G / G / 1$ queue will be called the general solution. Reference to the bounds used in estimating queue length will be made in the normal manner, $i_{\text {. }} e_{\text {. the }}$ IFR bounds.

## a) $M / M / 1$ and $M / G / 1$ Queues

Nine of the twenty-two sample queues are discussed here and will be referred to by their unique run number. Runs 4 and 1 are $M / M / 1$ and the exponential solution estimates queue length accurately in both instances. Even in the relatively congested queue, run 1 , the solution is within $3 \%$ of the simulated result. This degree of accuracy is expected and helps to confirm the satisfactory operation of the entire experiment. The general method estimates queue length with an acceptable degree of error although these solutions are less consistent than the exponential ones.

The exponential solutions to the $M / G / 1$ queues are less satisfactory although only in runs 9 and 5 are the errors excessive. These latter runs are marked by very consistent solutions from the general method as well. The consistency
between the general and exponential solutions in these runs suggests that either these methods both overestimate queue length or the simulated value is not a steady state value. In the following evaluation of the $G / G / 1$ queues, we will see that the exponential method considerably overestimates queue length and it is suspected that this may be the case in these $M / G / 1$ queues. In view of this it is hard to reconcile the consistency of the general and exponential estimations in these two runs, particularly when the error in the general estimation is distinctly higher than in all other general estimations in the twenty-two runs. Except for runs 9 and 5 the estimation of queue length for $M / G / 1$ queues is considered acceptable by both methods although the only method used by an analyst will be the exponential.

The IFR bounds span the simulated solution in only three of the nine runs under discussion. In two cases the coefficient of variation of the interoarrival time distributions exceeds 1.0 and DFR bounds are computed. These substantially underestimate the simulated solution. The bounds that fail to include the simulated solution within their span are both high and low with no apparent trend, and are not considered a reliable estimator of queue length in the exponential arrival time case.

One may conclude from this first set of runs that the classical queueing formulae will generally be reliable in estimating queue length in those situations for which they are intended, and will be of practical use to the average user
in the business community. Errors in estimation may occur when service times are not distributed exponentially, but the magnitude of these errors and any method of their prediction is not known.

## b) $G / G / 1$ Queues

The estimated solutions to these queues are more consistent with each other than in the previous section. These solutions are estimated within $30 \%$ of the true figure by the general method while the errors in the IFR bound estimated solution tend to be compatable in size and in direction.

Table 10 shows that the general solution and the IFR bound solution estimated more accurately in the larger sample queue than in the smaller. We have already suggested that little benefit is gained from the larger sample size in estimating queue length since population parameters are not estimated with noticeable improvement in accuracy. The evidence in this table tends to refute this arguement and support the idea that the larger sample does permit better estimates for queue length for the general inter-arrival time queue. Equation (7) from Chapter III for the expected wait in the queue depends on the moments of the idle time distribution. The same number of observations that are in the sample go into computing the shape of this distribution so that a more accurate estimation of the iale time distribution parameters will result from the larger sample. This would account for greater accuracy in the general solution method.

The estimated solutions to queue length by the IFR bounds seem more reliable in these general queues than was found in the exponential examples. Nine of the thirteen bounds span the simulated queue results, and the remaining four are not grossly inaccurate. The average of these two bounds is compared to the simulated figure and the percent differences are shown in Table $X$. This is a crude and unscientific method of estimation as we have no knowledge of where specifically any solution should lie within the span of these bounds. It is done here to measure how large estimation error will be with a makeshift method based on these bounds. Even so, errors in estimation of queue length are not significantly worse than from the general solution. This means that the IFR bounds can be used to give as reliable, and in some cases more reliable, solutions than the general case. Reliable in this context is within 20\%.

In the absence of a more accurate alternate method this amount of error must be accepted, even though for most purposes it is not considered to be excessively incorrect. This is particularly so when we are discussing fractional expected queue lengths. A $20 \%$ error in an expected queue length of say .6 is only . 12 and estimation in this case is not seriously in error.

The exponential solution method overestimates in all cases in this second group of queues. The resultant solutions do not contribute at all to our finding of actual queue lengths but certainly do point out the necessity of accurate distribution identification. In all but three cases, at least 100\%
error occurred. By assuming that a queueing problem is $M / M / 1$ when in fact it is not, the analyst is likely to considerably overestimate the real queue length.

From experimental runs analyzed one may conclude that provided correct identification of the inter-arrival and service distributions is made, and it appears from Chapter II that this can be done, an analyst can find estimates of queue length (Lq) with a high probability of less than $20 \%$ error. This, of course, assumes that population parameters, lamda and mu, are accurately estimated. The assumption that a queue is of the M/M/ 1 type should be made cautiously. Overestimation errors exceed $100 \%$ if the assumption is false. On the other hand if all queues are assumed to be $G / G / 1$, errors from applying the IFR solution method will not be as correspondingly excessive. It appears safer to assume that all queues are of the general type unless some positive identification of the inter-arrival pattern is made. Only then is it advisable to apply the classical queueing formulae for exponential arrivals.

Step 3: Results from Inaccurate Samples

Calculation of estimated queue length from this body of data by any of the methods, did not provide satisfactory answers. All errors in estimation were large, indicating a relationship between errors in estimating rho and queue length. The results are shown in Tables XI and XII。
a) $M / M / 1$ Queues

The classical formulae produce exceptionally high

TABLE XI
Summary of Queue Statistics From Inaccurate Data (Sample Size 36)

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2 | M/ M/I | 0.8 | 2.96 | 2.05 | -30.80 | 2.18 | -25.27 | 2.00 | 2.77 | -19.44 |
| 3 | M/ M/1 | 0.7 | 1.64 | 4.73 | 188.71 | 5.02 | 206.22 | 0.25 | 2.08 | -28.86 |
| 5 | M/G2/1 | 0.9 | 4.10 | 32.31 | 687.76 | 30.55 | 644.84 | 0.28 | 1.41 | -79.31 |
| 6 | M/G2/1 | 0.8 | 1.92 | 2.01 | 4.52 | 1.47 | $-23.45$ | 0.19 | 0.96 | -70.15 |
| 7 | M/G2/1 | 0.7 | 1.07 | 1.44 | 33.83 | 1.88 | 75.68 | 1.43 | 2.21 | 69.66 |
| 10 | H/03/1 | 0.8 | 2.43 | 4.06 | 66.94 | 3.55 | 45.97 | 0.24 | 1.55 | -61.25 |
| 11 | M/G3/1 | 0.7 | 1.18 | 0.34 | -71.21 | 0.32 | -72.48 | 0.18 | 0.78 | -59.57 |
| 12 | M/G3/1 | 0.6 | 0.67 | 0.31 | -53.50 | 0.49 | -27.12 | 0.09 | 0.55 | -51.79 |
| 14 | G2/03/1 | 0.8 | 1.32 | 1.68 | 27.42 | 4.82 | 265.00 | 1.43 | 2.01 | 30.34 |
| 15 | G2/G3/1 | 0.7 | 0.61 | 0.28 | -54.41 | 1.03 | 68.95 | 0.08 | 0.49 | -53.35 |
| 16 | G2/G3/1 | 0.6 | 0.34 | 0.29 | -12.72 | 0.68 | 102.37 | 0.03 | 0.47 | -26.92 |
| 17 | c2/c2/1 | 0.9 | 2.36 | 15.05 | 536.90 | 46.84 | 1882.27 | 14.72 | 15.38 | 536.86 |
| 18 | G2/G2/1 | 0.8 | 0.97 | 1.78 | 83.63 | 5.09 | 423.79 | 1.39 | 2.00 | 74.41 |
| 21 | G2/ M/1 | 0.9 | 5.48 | 0.79 | -35.52 | 1.57 | -71.30 | 0.54 | 0.98 | -86.13 |
| 22 | G2/ $11 / 1$ | 0.8 | 1.85 | 0.93 | -50.24 | 1.66 | -10.83 | 0.69 | 1.22 | -48.87 |
| 23 | G2/ M/I | 0.7 | 0.97 | 1.98 | 102.87 | 2.78 | 185.54 | 1.75 | 2.40 | 113.28 |
| 24 | G2/ M/1 | 0.6 | 0.52 | 0.20 | -61.27 | 0.64 | 23.51 | 0.05 | 0.47 | -50.19 |
| 27 | G3/ $1 / 1$ | 0.7 | 1.17 | 2.77 | 137.50 | 4.58 | 292.29 | 2.63 | 3.33 | 155.18 |
| 28 | G3/ M/I | 0.6 | 0.65 | 0.35 | -47.18 | 0.51 | -21.37 | 0.25 | 0.74 | -24.66 |
| 29 | G3/G2/I | 0.9 | 3.62 | 1.42 | -60.89 | 3.48 | $-4.00$ | 1.17 | 1.81 | -58.88 |
| 30 | G3/G2/1 | 0.8 | 1.36 | 0.68 | -50.29 | 1.56 | 14.54 | 0.36 | 0.91 | -53.52 |
| 31 | G3/G2/1 | 0.7 | 0.69 | 10.11 | 1371.03 | 21.24 | 2931.41 | 0.52 | 10.35 | 1346.29 |
| 32 | G3/G2/1 | 0.6 | 0.37 | 0.31 | -16.26 | 0.56 | 51.76 | 0.06 | 0.56 | -16.80 |
| 35 | G3/G3/1 | 0.7 | 0.80 | 0.66 | -17.46 | 1.14 | 43.34 | 0.40 | 0.99 | -12.19 |
| 36 | G3/G3/1 | 0.6 | 0.43 | 0.32 | -25.76 | 0.65 | 52.93 | 0.13 | 0.62 | -12.88 |

TABLE XII

## Summary of Queue Statistics From Inaccurate Data (Sample Size 100)

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 3 | M/ M/1 | 0.7 | 1.64 | 1.09 | -33.37 | 1.09 | -33.43 | 0.88 | 1.52 | -26.63 |
| 4 | M/ M/I | 0.6 | 0.82 | 0.79 | -2.82 | 1.08 | 32.31 | 0.83 | 1.52 | 43.88 |
| 6 | M/G2/1 | 0.8 | 1.92 | 1.42 | -25.90 | 1.53 | -20.49 | 0.18 | 0.80 | -74.54 |
| 9 | 1:/G3/1 | 0.9 | 5.37 | 1.85 | -65.46 | 1.80 | -66.42 | 0.20 | 1.63 | -82.94 |
| 10 | 1:/G3/1 | 0.8 | 2.43 | 4.64 | 90.71 | 3.50 | 44.12 | 0.26 | 1.78 | -58.20 |
| 14 | G2/G3/1 | 0.8 | 1.32 | 0.85 | -35.38 | 1.78 | 35.15 | 0.60 | 1.15 | -33.83 |
| 15 | G2/G3/1 | 0.7 | 0.61 | 1.11 | 81.86 | 2.73 | 346.24 | 0.86 | 1.44 | 87.99 |
| 17 | $\mathrm{G} 2 / \mathrm{G} 2 / 1$ | 0.9 | 2.36 | 1.01 | -57.17 | 3.94 | 66.57 | 0.76 | 1.29 | -56.58 |
| 19 | $\mathrm{G} 2 / \mathrm{G} 2 / 1$ | 0.7 | 0.50 | 1.56 | 212.60 | 4.11 | 722.40 | 1.28 | 1.93 | 221.10 |
| 22 | G2/ M/1 | 0.8 | 1.86 | 2.09 | 11.85 | 4.52 | 142.31 | 1.90 | 2.47 | 17.32 |
| 23 | G2/ M/1 | 0.7 | 0.97 | 0.26 | -72.82 | 0.52 | -46.77 | 0.14 | 0.53 | -65.49 |
| 24 | G2/ M/1 | 0.6 | 0.52 | 0.23 | -55.88 | 0.43 | -17.73 | 0.08 | 0.46 | -48.07 |
| 25 | G3/ M/1 | 0.9 | 5.17 | 232.39 | 4396.79 | 286.16 | 5437.11 | 231.62 | 232.34 | 4388.81 |
| 27 | c3/ M/l | 0.7 | 1.17 | 4.53 | 287.76 | 7.18 | 514.38 | 4.29 | 4.99 | 297.39 |
| 28 | G3/ $1 / 1$ | 0.6 | 0.65 | 0.31 | -52.37 | 0.59 | -9.4.7 | 0.17 | 0.66 | -36.49 |
| 32 | G3/G2/1 | 0.6 | 0.37 | 0.32 | -12.20 | 0.63 | 70.73 | 0.06 | 0.58 | -12.47 |
| 33 | 63/63/1 | 0.9 | 3.67 | 21.26 | 479.92 | 38.73 | 956.36 | 20.35 | 21.72 | 482.00 |
| 34 | G3/G3/1 | 0.8 | 1.86 | 16.88 | 808.89 | 30.76 | 1556.38 | 16.72 | 17.51 | 821.54 |
| 35 | G3/G3/1 | 0.7 | 0.80 | 0.39 | -50.88 | 0.74 | -6.41 | 0.20 | 0.74 | -40.45 |

queue lengths when rho approaches one. As a result, overestimation of rho at high values provides theoretically correct but results impossible in practical situations. Yet underestimation of rho at lower values may provide almost no effect at all on queue length. It is virtually impossible to draw consistent conclusions as to the size of error from the $M / M / l$ queue results, in the absence of any apparent trend, other than to say that all estimates were increasingly inaccurate as rho was estimated with greater error.

## b) $G / G / 1$ Queues

The need for an accurate estimation of rho is indicated in these runs too. The relationship between errors has a more consistent trend in the general arrival queues. In each run shown in Tables XI and XII for this class of queues, the sign of the estimation error in rho and in queue length is the same. Little use can be made of this fact in reducing prediction errors when rho error is unknown. It is interesting to note that even in terms of errors, the $G / G / 1$ queue tends to behave with greater stability than the $M / M / 1$ queue.

In none of the runs with inaccurately found rho values did errors compensate for themselves to produce good estimates for queue length from poor data. The occurence of accurate queue prediction in this case would, of course, be by accident. One cannot conclude that this form of accident would never occur; one can suggest that it is very unlikely, based on the runs generated in this experiment.

In cases where a poor prediction of rho is recognized
by an and because of his experience and his feel for the system some compensation for expected error can be made. This may be in the form of additional sampling to provide some addim tional comparative estimates of rho or it may be merely an adjustment based on intuition. The analyst is advised in this situation to compute solutions for a range of rho values to determine the behaviour of the specific queue in question. He then can determine the effect of erroneous rho values.

## Summary

Essential to successful analysis of any queueing problem are accurate predictions of population parameters. Easily applied methods exist for finding of adequate solutions to queueing problems provided the parameters used are accurate to approximately $5 \%$ of the true value. If these parameters are not good predictions even the expert will fail to arrive at satism factory answers.

The method of solution for $M / M / 1$ queues, the classical queueing formulae, are shown to provide accurate predictions of queue length provided the exponential assumption is correct. Misapplication of these formulae will seriously overestimate queue length. The IFR bounds while not particularly satisfactory for the $M / M / 1$ situation do give as accurate a solution as any of the other methods considered when applied to $G / G / 1$ queues.

The actual process that an analyst could use is exemplified here. This is followed by a recapitualation of our findings.

An Application

Let us assume that a service station operator also offers a car washing service on his lot. Because of the nature of the extra services he offers in the washing and cleaning of cars, such as vacuuming, tar removal, and engine cleaning, cars to not move through his plant deterministically as is usual. Let us assume that he thinks he will lose customers if they are forced to wait in a queue of three cars or longer on the average. He needs to know how long the average queue of cars will be and how long a customer will spend on the average getting his car washed.

The service station operator has special cards that are stamped by a time clock when a car enters the wash and are stamped again when it leaves. To collect the data he needs for his queueing analysis he records the time of arrival on the card when a car joins the queue for washing. From the three times recorded for each car he computes the time each car spent in the queue and in the wash. These are the interarrival and service times.

The mean inter-arrival and service times are easy to
compute. By squaring each of the inter-arrival and service times, summing them, and dividing by the number of cars observed the variance is found. Let us suppose that the values. found in this example are:
a) average inter-arrival time 10 minutes, with variance 54.4 minutes and,
b) average service time 6.45 minutes, with variance 23.7 minutes.

The parameters lamda and mu are the reciprocals of the averages shown above.

$$
\begin{aligned}
& \lambda=.1 \text { arrivals per minute。 } \\
& \mu=.155 \text { services per minute. }
\end{aligned}
$$

Rho is then computed $\rho=\frac{\lambda}{\mu}=\frac{.1}{.155}=.64$;
this value being the intensity of use of the facility expressed as a fraction of one.

Identification of the type of each distribution is the next step in this analysis. Reference to Appendix II shows precomputed data with which to compute the test statistic and the method of computation. Assume that the following correlation coefficients were computed:
a) inter-arrival distribution86
b) service distribution . 87

This would indicate that both distributions are not exponentially distributed and the IFR bound method will be used. To verify that both distributions do have the IFR property, the test mentioned in Appendix I will be used. The coefficients of
variation are computed with,

$$
c^{2}=\frac{\text { variance }}{\text { mean }^{2}}
$$

The example coefficients are:
a) inter-arrival distribution $\quad \frac{54 \& 4}{10}=.544$
b) service distribution $\quad \frac{23.7}{6.45}=.570$;
therefore both distributions have the IFR property. Using equation (12) of Chapter III the IFR bounds are found:

$$
\begin{aligned}
J=\frac{C a^{2}+\lambda^{2} \sigma_{\mathrm{F}}^{2}}{2(1-\rho)} & =\frac{(.541)+(.1)(.1)(23.7)}{(2)(.1)(.36)} \\
& =\frac{.781}{.072} \\
& =10.82
\end{aligned}
$$

Lower bound on queue length:

$$
\begin{aligned}
L B=\lambda J-\frac{\left(\rho^{2}+C a^{2}\right)}{(2)} & =1.082-\frac{.64+0.54 .4}{(2)} \\
& =1.082-.592 \\
& =.590
\end{aligned}
$$

Upper bound on queue length:

$$
U B=\lambda J=1.082
$$

The average of these bounds is . 84 and this is assumed to be the estimate of queue length. Using the relationships between queue length and time spent queueing, the service station operator can compute the total time that a customer will spend having his car washed on the average.

From Wq $=\frac{\mathrm{Lq}}{\lambda}=\frac{.84}{.1}=8.4$ we find the average wait in queue to be 8.4 minutes. The total time spent in the car wash is then $W=W q+\frac{1}{\mu}=8.4+6.45=14.85$ minutes. It takes approximately 15 minutes on the average to be processed by this facility.

Conclusions

Successful application of queue length estimation methods is dependent upon the identification of exponentially distributed intermarrival and service time distributions. Identification is made difficult by the lack of a consistent test which detects exponentiality in small samples. A test is developed in this thesis from a well known graphical procedure. This test consistently detects samples that are drawn from exponential distributions and rejects those that are not. Samples of 36 and 100 observations were tested. An attractive feature of this test is its ease in application.

When identification is possible, appropriate methods for estimating queue length are accurate for most practical applications. Classical queueing methods when applied to $M / M / 1$ and M/G/l queues provide satisfactory estimates but serious overestimation results from their application to $G / G / l$ and G/M/l queues. Thus the assumption that a queue is exponential must be made with care. The method for successful queue length prediction in general is considered too complex for practical application and $I F R$ and $D F R$ bounds are a satisfactory
replacement. The average of the upper and lower bounds in the IFR case provides a prediction that is as accurate as the solution from the complex general calculation.

Regardless of which method is chosen for estimation,
inaccurate prediction of queue length will result from inaccurate estimation of rate parameters of the sampled population distributions.

## BIBLIOGRAPHY

Barlow, E. Richard and Frank Proschan, Mathematical Theory of Reliability, New York, John Wiley \& Sons, Inc., 1967. p.12-35.

Cooper, B.E. Statistics for Experimentalists. Braunschweig, Germany, Pergamon Press, 1969, pp. 366 .

Cox, D.R., and Walter L. Smith, Queues, Methuen's Monograohs on Applied Probability and Statistics, London, Metheun \& Co. Ltd. 1967, pp.180.

Goddard, L.S. Mathematical Techniques of Operational Research, New York, Pergamon Press, 1963, p.85-140.

Hillier, Frederick S., and Gerald J. Lieberman, Introduction to Operation Research, San Francisco, Calif.. Holden-Day, Inc., 1968, p.285-306.

Marshall, Kneale Thomas, Some Inequalities for Single Server Queues, Berkeley, Operation Research Center, College of Engineering, University of California, 1966. p.1-29.

Parl, Boris, Basic Statistics, Garden City, New York, Doubleday \& Company Inc., 1967. p.223-244.

Parzen, Emanuel, Modern Probability Theory and its Applications, New York, John Wiley \& Sons, Inc., 1965, D.371-374.

Smith, L. Walter and Williem E. Wilkenson, eds. Proceedings of the Symposium on Conjestion Theory, Held at the University of North Carolina at Chapel Hill, Aug.24-26, 1964. Chapel Hill, The University of North Carolina Press, 1965, p.137-157.

Snedecor, George and William G. Cochran, Statistical Methods, Ames, Iowa, The Iowa State University Press, Sixth ed. 1967, p.231-238.

Wagner, Harvey M. Principles of Operations Research, with Applications Eo Managerial Decisions, Englewood Cliffs, New Jersey, Prentice Hall Inc., 1969, p.8-64.

APPENDICES

## APPENDIX I

## INCREASING AND DECREASING FAILURE RATE

The information contained in this appendix is taken from Barlow and Proschan: Mathematical Theory of Reliability. (John Wiley \& Sons, 1967) which may be used as a more extensive reference.

It has been found that certain objects tend to fail. as time progresses, with increasing or decreasing probability. Phenomena with increasing failure rate (IFR) fail in the period [t, $t+d t$ ] over time ( $t$ ) with increasing probability. Those with decreasing failure rate (DFR) fail with less probability over time.

Applications are found, in the case of IFR in the computation of mortality tables and empirical evidence suggests DFR is found in most solid state electronic components.

When applied to queueing, IFR is used to mean that the probability of an arrival during [ $t, t+d t$ ] increases over time。 Conversely DFR implies a decreasing probability. Hence reference will be made to $I F R / G / 1$ and $D F R / G / 1$ queues.

A Statistical Test for $\operatorname{IFR}$ and DFR

The coefficient of variation (s.d./mean) for IFR distributions is less than one and for $D F R$ is greater than one. A distribution with a coefficient of variation equal to one is exponentially distributed and displays both IFR and DFR properties. Identification thus displays both IFR and DFR properties. Identification of these properties can be made by
comparing a distribution"s coefficient of variation to one and drawing the appropriate conclusion.

## APPENDIX II

In this appendix we will test two samples of data to present in detail the test for exponentiality using the correlation coefficient. This procedure uses pre-computed data for comparison with the observed data. The precomputed figures are shown in Table XIII and are arranged for use with sample sizes of $25,35,50,75$, and 100 . The sum, sum of squares, mean and variance are included. These figures are natural logarithms of the statistic for the correlation coefficient test shown in Chapter II and represent the ordinate values. For examole Size $n$, the numbers in Table XIII are computed from

$$
y=\ln \left(\frac{n+1}{n+1-i}\right)
$$

The following times are the inter-arrival times recorded for 35 successive arrivals of customers in a queue. They should be read by column for the correct sequence or occurrence.

| .60 | 25.77 | 9.43 | 34.40 | 2.75 |
| ---: | ---: | ---: | ---: | ---: |
| 1.88 | .74 | 4.10 | 3.73 | 29.71 |
| 65 | 1.52 | 26.64 | .46 | 4.92 |
| 15.59 | 7.07 | 10.57 | 3.09 | 27.08 |
| 17.37 | 4.04 | 5.36 | 20.61 | 3.87 |
| 5.83 | 3.16 | 1.05 | 14.72 | .99 |
| 4.13 | 7.41 | 17.37 | 11.65 | 5.94 |

Before actually calculating the correlation coefficent, several summations are to be found that make it easier.

Representing the specific interarrival times by $x$ we compute,

$$
\Sigma x=333.24, \Sigma x^{2}=6310.0, \Sigma x y=594.8, \frac{\Sigma x}{n}=9.52
$$

where $y$ is defined above.
Calculating the correlation coefficient using the $y$ values for sample size 35 taken from Table XIII we find,

$$
\begin{aligned}
r & =\frac{\sum x y-\bar{x} \sum y}{\sqrt{\left(\Sigma x^{2}-\bar{x} \Sigma x\right)\left(\Sigma y^{2}-\bar{y} \sum y\right)}} \\
& =\frac{594.8-(9.52)(33.28)}{\sqrt{(6310.0-(9.52)(333.24))(56.98-(.95)(33.28))}} \\
& =\frac{277.968}{281.924} \\
& =.986
\end{aligned}
$$

Since this figure exceeds 093 we assume that the interarrival time distribution is exponential.

The service times corresponding to the interarrival times are listed by column in order of occurrence.

| 5.04 | 4.62 | 12.58 | 13.01 | 2.93 |
| ---: | ---: | ---: | ---: | ---: |
| .28 | 17.44 | 9.41 | 2.46 | 10.24 |
| 11.12 | 15.15 | 4.08 | 15.69 | 16.90 |
| .01 | 12.54 | 5.04 | 14.92 | 3.94 |
| .77 | 4.42 | 16.27 | 13.42 | 10.47 |
| 17.36 | 5.27 | 7.90 | 14.63 | 14.43 |
| 7.07 | 1.71 | 11.66 | 3.78 | 14.27 |

As above we compute,

$$
x=321.7, x=4033.1, \quad x y=452.8, \frac{x}{n}=9.19
$$

and using the same $y$ values as above find for the correlation coefficient,

$$
r=\frac{56.98-(9.19)(33.28)}{\sqrt{(4033.1-(9.19)(321.7))(56.98-(.95)(33.28))}}
$$

$=\frac{146.82}{165.06}$
$=.889$

This $r$ is less than .93 and we conclude that the service times are not exponentially distributed. This queue is identified as $M / G / 1$.

## TABLE XIII

## Pre-computed Ordinate Values for the Correlation Coefficient Test

Sample Size 25
0.039
0.080
0.123
0.167
0.214
0.262
0.314
0.368
0.425
0.486
0.550
0.619
0.693
0.773
0.860
0.956
1.061
1.179
1.312
1.466
1.649
1.872
2
0.080
0.123
0.167
0.214
0.262
0.314
0.368
0.425
0.486
0.550
0.619
0.693
0.773
0.860
0.956
1.061

1. 179
1.312
1.466
1.649
1.872
2.565
3.258

$$
\begin{aligned}
\Sigma y & =23.449 \\
\Sigma y^{2} & =38.698 \\
\bar{y} & =.938
\end{aligned}
$$

Sample Size 35
0.028
0.057
0.087
0.118
0.150
0.182
0.216
0.251
0.288
0.325
0.365
0.105
0.448
0.492
0.539
0.588
0.639
0.693
0.750
0.811
0.875
0.344
1.019
1.099
1.186
1.281
1.386
1.504

1. 638
1.792
1.974
2.197
2.485
2.890
3.584

$$
\begin{aligned}
\Sigma y & =33.287 \\
\Sigma y^{2} & =56.987 \\
\bar{y} & =.951
\end{aligned}
$$

## TABLE XIII (cont'd)

## Sample Size 50

| 0.020 | 0.713 |
| :---: | :---: |
| 0.040 | 0.754 |
| 0.061 | 0.796 |
| 0.082 | 0.841 |
| 0.103 | 0.887 |
| 0.125 | 0.936 |
| 0.148 | 0.987 |
| 0.171 | 1.041 |
| 0.194 | 1.099 |
| 0.218 | 1.159 |
| 0.243 | 1.224 |
| 0.268 | 1.293 |
| 0.294 | 1.367 |
| 0.321 | 1.447 |
| 0.348 | 1.534 |
| 0.376 | 1.629 |
| 0.405 | 1.735 |
| 0.435 | 1.852 |
| 0.466 | 1.986 |
| 0.498 | 2.140 |
| 0.531 | 2.322 |
| 0.565 | 2.546 |
| 0.600 | 2.833 |
| 0.636 | 3.239 |
| 0.674 | 3.932 |

$$
\begin{aligned}
\Sigma Y & =48.113 \\
\Sigma Y^{2} & =85.038 \\
\bar{Y} & =.962
\end{aligned}
$$

## Sample Size 75



$$
\begin{aligned}
\Sigma y & =72.914 \\
\Sigma y^{2} & =132.656 \\
\bar{y} & =.972
\end{aligned}
$$

TABLE XIII (cont'd)

Sample Size 100

| 0.010 | 0. 425 | 1.119 |
| :---: | :---: | :---: |
| 0.020 | 0.441 | 1.149 |
| 0.030 | 0.456 | 1.181 |
| 0.040 | 0.1472 | 1.21 .4 |
| 0.051 | 0.488 | 1.248 |
| 0.061 | 0.504 | 1.283 |
| 0.072 | 0.521 | 1.319 |
| 0.083 | 0.538 | 1.357 |
| 0.093 | 0.555 | 1.396 |
| 0.104 | 0.572 | 1.437 |
| 0.115 | 0.590 | 1.480 |
| 0.126 | 0.608 | 1.524 |
| 0.138 | 0.626 | 1.571 |
| 0.149 | 0.645 | 1.619 |
| 0.161 | 0.664 | 1.671 |
| 0.172 | 0.683 | 1.725 |
| 0.184 | 0.703 | 1.782 |
| 0.196 | 0.723 | 1.843 |
| 0.208 | 0.744 | 1.907 |
| 0.221 | 0.765 | 1.976 |
| 0.233 | 0.786 | 2.050 |
| 0.246 | 0.808 | 2.130 |
| 0.258 | 0.831 | 2.217 |
| 0.271 | 0.854 | 2.31 .3 |
| 0.2814 | 0.877 | 2.418 |
| 0.298 | 0.902 | 2.536 |
| 0.311 | 0.926 | 2.669 |
| 0.325 | 0.952 | 2.823 |
| 0.338 | 0.978 | 3.006 |
| 0.352 | 1.004 | 3.229 |
| 0.367 | 1.032 | 3.517 |
| 0.381 | 1.060 | 3.922 |
| 0.396 | 1.089 | 4.615 |

$$
\begin{aligned}
\Sigma y & =97.772 \\
\Sigma y^{2} & =180.862 \\
\bar{Y} & =.978
\end{aligned}
$$


[^0]:    $1_{\text {K. T. Marshall, Some }}$ Inequalities for Single Service Queues, Berkeley, OperaEions Research Center, University of California, 1966, p. 6.

[^1]:    $3^{3}$ Ibid. ${ }^{\text {p. } 4 .}$

[^2]:    ${ }^{4}$ W. Feller, "On the Kolmogorov - Smirnov Limit Theorems for Empirical Distribution", Annals of Math. Stat., 19, pp. 177-189, cited in IBM SysEem/360 Scientific Subroutine Package, Version III, White Plains, N.Y., IBM Corporation.

[^3]:    $5^{5}$.W. Lilliefors, "On the Kolmogrov - Smirnov Test for Normality with Mean and Variance Unknown", J.A.S.A. 62, (1967) pp.399-402, cited in IBM System/360, Scientific Subroutine Package, Version III, White Plains, N.Y. IBM Corporation.
    ${ }^{6}$ B. Epstein, Test for the Validity of the Assumption that the underlying Distribution of Life is Exponential, Washington, D.C., U.S. Dept. of Commerce, office of Technical Services, 1959.

[^4]:    8B.E. Cooper, Statistics for Experimentalists, Braunschweig, Germany, 1969, p. 82.

[^5]:    12d.f. $=$ (No. of classes) - (No. of estimated parameters) - 1 .

[^6]:    $15 \mathrm{~J} . \mathrm{D}$. Little, "A Proof of the Queueing Formula: $L=\lambda W "$, Operations Research, Vol. 9, (1961), pp. 383-387.

[^7]:    16 Numbers in brackets in this section refer to equation numbers in Chapter III.

