AD-A020 129

THE ALMOST REGENERATIVE METHOD FOR STOCHASTIC SYSTEM SIMULATIONS

Francis L. Gunther

California University

Prepared for:

Army Research Office - Durham

December 1975

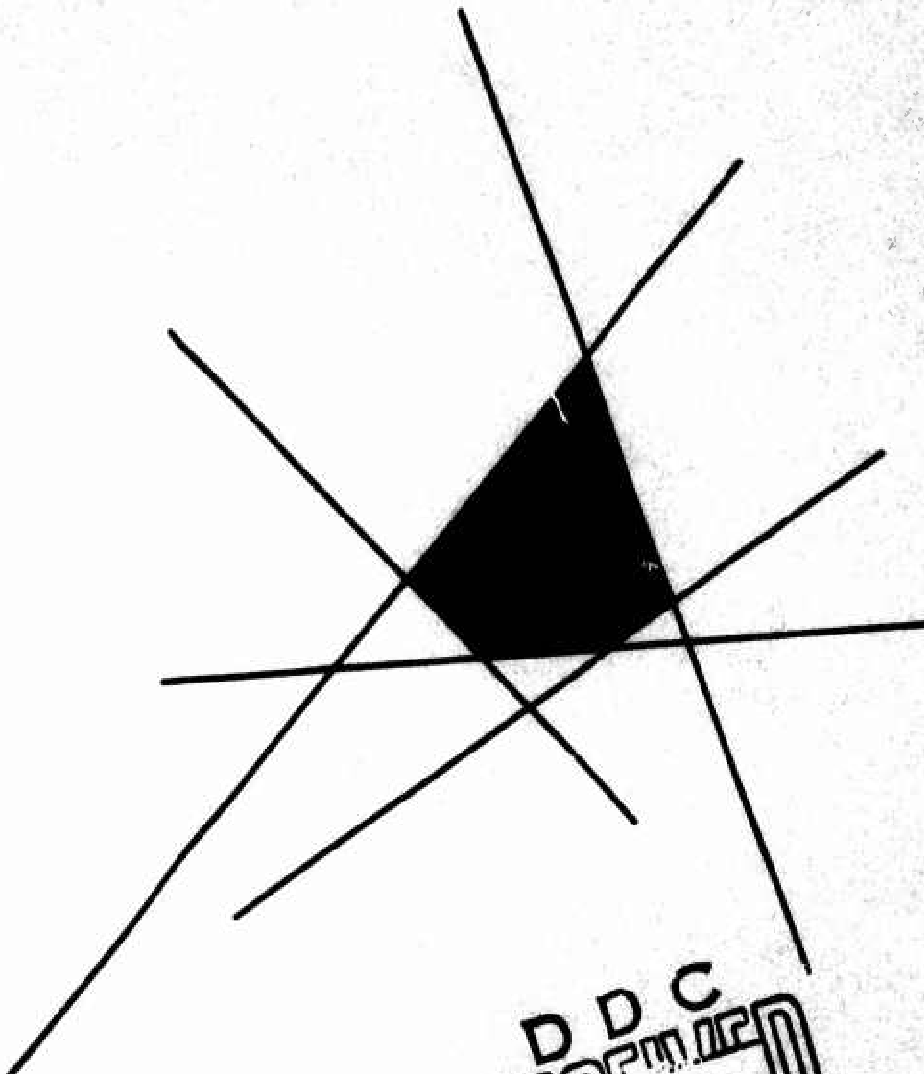
037135

ORC 75-21
DECEMBER 1975

THE ALMOST REGENERATIVE METHOD FOR STOCHASTIC SYSTEM SIMULATIONS

by
FRANCIS L. GUNTHER

A020129



OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

THE ALMOST REGENERATIVE METHOD FOR STOCHASTIC SYSTEM SIMULATIONS

Operations Research Center Research Report No. 75-21

Francis L. Gunther

December 1975

U. S. Army Research Office - Durham

DAHC04-75-G-0163

Department of Industrial Engineering
and Operations Research
University of California, Berkeley

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
ORC 75-21

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
THE ALMOST REGENERATIVE METHOD FOR STOCHASTIC SYSTEM SIMULATIONS

5. TYPE OF REPORT & PERIOD COVERED
Research Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Francis L. Gunther

8. CONTRACT OR GRANT NUMBER(s)
DAHC04-75-G-0163

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Operations Research Center
University of California
Berkeley, California 94720

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
P-12549-M

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office - Durham
Box CM, Duke Station
Durham, North Carolina 27706

12. REPORT DATE
December 1975

13. NUMBER OF PAGES
~~83~~ 84

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Computer Simulation
Regenerative Method
Queues
Renewal Processes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DEDICATION

To Judy, Kerry, and Brendan.

ACKNOWLEDGEMENT

I wish to thank my dissertation advisor, Professor Ronald W. Wolff, for the many enjoyable hours he spent in discussion with me throughout my studies and for the profound influence he had on my completion of this dissertation. I also wish to thank Professors Sheldon Ross and Michael Storebraker for serving on my dissertation committee. Special thanks are due to Sandra Powers who spent many weekends typing the manuscript and handling all the administrative details associated with submitting the dissertation to the university. I wish to express my sincerest appreciation to Herb Kettler, Randy Downing, and George Spiro of Bell Telephone Laboratories for their generous support and their belief in me throughout my Ph.D. studies.

ABSTRACT

The *regenerative method* for stochastic system simulation allows data collection each time the stochastic process enters a specific *single state*, r , called the regeneration state. The generated observations have the desireable property of being independent and identically distributed. Relative to a fixed run length, however, the mean time between entries into r may be excessively long for complicated stochastic systems, thus providing few observations and poor variance estimates.

The *almost regenerative method* is an extension of the regenerative method designed to alleviate this problem for complicated stochastic systems (such as a network of queues). The almost regenerative method allows data collection each time the stochastic process enters a *set of states*. Simulations of simple queueing networks show that the almost regenerative method can provide an order of magnitude improvement over the regenerative method in terms of the mean-square-error of the estimator of total delay in queue, and this relative improvement increases with system complexity. In addition, variance estimates based on observations generated by the almost regenerative method are shown emperically to be superior to variance estimates based on observations generated by the regenerative method.

The almost regenerative method is also shown emperically to consistently provide more efficient estimators, in terms of mean-square-error, than the frequently used fixed time increment method for event-oriented simulations.

The computer program modifications necessary to implement the almost regenerative method are no more complicated than implementing the regenerative method or the fixed time increment method. Also, CPU requirements are virtually the same for each of these methods.

TABLE OF CONTENTS

TABLE OF CONTENTS

List of Tables cont.

List of Figures:

CHAPTER 1

INTRODUCTION AND SUMMARY OF RESULTS

1.1 Introduction

The central problem in simulations of stochastic processes is concerned with the method used to collect observations of the stochastic process under consideration. Consider a stationary stochastic process $\{X(t), t \geq 0\}$, and let f be a real-valued, integrable function defined on $X(t)$. For a fixed simulation run length ℓ, an estimate of $\theta = E\{f(X)\}$ is

$$\hat{\theta}(\ell) = \frac{1}{\ell} \int_0^{\ell} f[X(u)]du ,$$

since it can be shown, under fairly general conditions, that $\hat{\theta}(\ell) \to \theta$ with probability one (w.p.1) as $\ell \to \infty$. Thus, observations may be taken at arbitrary intervals to form an estimate of θ. However, the conditions under which observations are collected is important when estimating the variance of $\hat{\theta}(\ell)$ (which is used, for example, in confidence interval estimation and hypothesis testing), since the observations may be serially correlated. All of the methods considered in this dissertation partition the time series $\{X(t), t \in [0,\ell]\}$ in different ways in an effort to reduce, or eliminate, the dependence between observations. The difference between the methods depends on the use made of knowledge about the process $X(t)$.

The classical *fixed time increment method* (FTM) is the degenerate case, since no use is made of any knowledge of $X(t)$. Observations are collected at fixed interval lengths, say Δ, where $n\Delta = \ell$ and

n is the number of observations. Since these observations may be dependent, the effects of autocorrelation must be allowed for in the estimate of $V\{\hat{\theta}(\ell)\}$, thus creating an estimation problem. Two approaches to this problem have been the method of batch means [Conway (1963)] and the application of time series methods [Conway, Johnson, and Maxwell (1959), and Fishman and Kiviat (1967)].

First suggested, but not developed, by Cox and Smith (1961), the *regenerative method* (RM) for collecting observations of the time series $\{X(t),\ t \in [0,\ell]\}$ has recently been systematically developed by a number of authors [Kabek (1968), Fishman (1973, 1974 abc), Crane and Iglehart (1973, 1974 abc, 1975 ab), and Iglehart (1974, 1975)]. The stochastic process $\{X(t),\ t \geq 0\}$ is called regenerative (a more precise definition is given in Chapter 2) if, every time a certain event occurs, the future of the process from then on is independent of the past and becomes a probabilistic replica of the future after time zero. Such times are called regeneration points. The central idea of the RM is to exploit the fact, when $X(t)$ is regenerative, that certain random variables between successive regeneration times are independent and identically distributed (iid), thus circumventing the autocorrelation problem in estimates of $V\{\hat{\theta}(\ell)\}$.

As a practical issue, the stochastic process $X(t)$ may not have the regenerative property (e.g., a stable multiple channel queue with interarrival and service time distributions such that the system never empties [Whitt (1972)]), or if it does, the expected time between regeneration times may be excessively large (e.g., a network of queues). In either of these cases it may be possible to define an *almost regenerative method* (ARM) which retains the desireable properties of the RM, but which allows more frequent data collection than might be possible if the RM were applied directly.

To see that sample size is important when estimating $V\{\hat{\theta}\}$ consider a sequence of iid random variables $\{Z_i\}_1^n$, each of which has a Normal distribution with mean μ and variance σ^2; i.e., $Z_i \sim N(\mu,\sigma^2)$ for all $i = 1, \ldots, n$. Then $\hat{\theta} = \frac{1}{n}\sum_1^n Z_i$ is an unbiased estimator of μ, and $V\{\hat{\theta}\} = \sigma^2/n$. The usual unbiased estimator of σ^2 is $s^2 = \frac{1}{n-1}\sum_1^n (Z_i - \hat{\theta})^2$. For samples from a Normal distribution, it is well-known that the sample function $(n-1)s^2/\sigma^2$ has the χ^2 distribution with $(n-1)$ degrees of freedom. Thus, for any probability $1-\alpha$, one can select two values of χ^2, depending on n and α, such that

$$(1.1) \qquad P\left\{\frac{\chi_1^2}{n-1} \leq \frac{s^2}{\sigma^2} \leq \frac{\chi_2^2}{n-1}\right\} = 1-\alpha .$$

As a result, this confidence interval can be made as small as desired simply by increasing n.

The purpose of this dissertation has been to investigate and develop the properties of an almost regenerative method suitable for application to variance estimation in complicated, practical queueing systems which do not have frequently occuring regeneration points.

1.2 Summary of Results

The regenerative method (RM) and the almost regenerative method (ARM) are defined and discussed in Chapter 2. The ARM estimators are found to have the same asymptotic properties as the RM estimators when the stochastic process under consideration is regenerative.

When applying the RM to a stochastic process $\{X(t), t \geq 0\}$ with state space E, the main idea is to select a *single state* in E such that the X(t) process regenerates each time it enters this state. The ARM permits one to define a *set of states*, say A, of E such that the X(t) process "almost regenerates," with negligible dependence on the past history of the process, each time X(t) enters A. As a result, the ARM will produce more observations than the RM for a fixed run length. A good choice of A depends on the model parameters being estimated.

The observations generated by the ARM for the queueing networks simulated in Chapters 3 and 4 are found to *act as if they were iid,* thus allowing direct use of the regenerative estimators.

A natural way to compare the relative efficiencies of the ARM, RM, and FTM estimators is to compare their mean-square-errors, where the mean-square-error of the estimator $\hat{\theta}$ of θ is defined by $MSE\{\hat{\theta}\} = E\{(\hat{\theta} - \theta)^2\}$. The ARM estimators for the queueing network models of Chapters 3 and 4 were found emperically to be up to *60 percent more efficient* than either the RM or FTM estimators. This result is attributable to the facts that the ARM can generate observations which act iid, unlike the FTM, and the ARM can generate more observations than the RM for the same run lengths. Examples are given where one might expect the improvement in efficiency of the ARM estimators to be up to an order of magnitude, or more, more efficient than comparable RM or FTM estimators.

When estimating the variance of the delay in queue estimator, the ARM variance estimators were found emperically to be superior to the RM variance estimators; i.e., the ARM variance estimators have a smaller bias and a smaller estimated standard deviation than the RM variance estimators.

Additionally, comparisons of the simple ratio estimator with the jackknife ratio estimator and comparisons of the simple variance estimator with the jackknife variance estimator show emperically that the jackknife estimators are preferable for ARM generated observations.

CHAPTER 2

THE ALMOST REGENERATIVE METHOD

The regenerative method for simulating stochastic systems is reviewed in this chapter, and an extension of this method, the almost regenerative method, is presented which is applicable to simulations of large, complicated stochastic systems. In Chapters 3 and 4, examples of queueing networks are investigated which show emperically that the almost regenerative estimators can be significantly more efficient than the regenerative estimators.

2.1 Regenerative Processes and the Regenerative Method (RM)

A sequence $Y_1, Y_2 \ldots$ of non-negative, independent and identically distributed (iid) random variables is called an *ordinary renewal process*. To avoid trivialities, assume that $P\{Y_n > 0\} > 0$ for all n. The physical principle underlying renewal theory is that of events occuring at arbitrary epochs, where the intervals between these events are the Y_n. Define $S_0 = 0$ and $S_n = Y_n + S_{n-1}$ $n = 1,2, \ldots.$ Then S_n is the epoch of the nth renewal after epoch 0. When Y_1 has a different distribution than the other Y_n, $n > 1$ ($Y_1, Y_2, \ldots$ are still independent), the sequence $Y_1, Y_2 \ldots$ is called a *general (delayed) renewal process*.

A stochastic process $\{X(t), t \geq 0\}$, with state space E, is said to be a *regenerative process* if there exists a sequence of random variables $\{Y_n\}_1^\infty$, with $S_0 = 0$ and $S_n = Y_n + S_{n-1}$, $n = 1,2, \ldots$, such that

(R1) $\{Y_n\}_1^\infty$ is an ordinary renewal process; and,

(R2) for any $n = 0,1, \ldots$, and $m = 1,2, \ldots$,

$$P\{X(S_n + t_i) \in E_i,\ i = 1,2,\ \ldots,\ m \mid X(s),\ s \le S_n\} = P\{X(S_0 + t_i) \in E_i,\ i = 1,2,\ \ldots,\ m\}\ ,$$

where $E_i \subset E$ and $t_i \geq 0$, for $i = 1,2, \ldots, m$. The epochs $S_1, S_2, \ldots$ are called *regeneration points (epochs),* and the $Y_1, Y_2, \ldots$ are the lengths of the regenerative cycles. Thus, a regenerative process is one which "starts over" at each regeneration point in such a way that the future of the process at a regeneration point is independent of the past history of the process, and, furthermore, the future at a regeneration point is a probabilistic replica of the future after epoch S_0. When the $\{Y_n\}_1^\infty$, form a general renewal process, $\{X(t), t \geq 0\}$ is said to be a *general (delayed) regenerative process.* In this case, S_0 is replaced by S_1 in (R2). Unless specified to the contrary, all regenerative processes considered in this chapter will be ordinary regenerative processes.

So far nothing has been said about the state of $X(t)$ at the regeneration points $\{S_n\}_0^\infty$. It is possible that a state change in the $X(t)$ process may not accompany the regeneration points. For example, although the GI/G/1 queue is regenerative ~ those epochs when an arrival finds no customers in the system, the number of customers in queue *does not change* (it remains zero). It is also possible that state changes in $X(t)$ at regeneration points *may not be unique.* Consider the number in system process for the GI/G/1 queue with batches of random size. At the epoch an arriving batch finds the system empty, a regeneration point occurs and the number in system becomes the batch size, which is a random variable. Finally, for a given $X(t)$ process, there may be *many choices* of regeneration points. For example, the number in system process for an M/M/1 queue has regeneration points when an arrival finds n customers in the system, for fixed n, it also has regeneration points when a departure leaves n customers behind in the system, for fixed n. This property is true for all $n = 0,1, \ldots$.

To avoid later confusion on the above issues, in this dissertation $X(t)$ will represent a stochastic process such that, if $X(S_n^-) = u$ and $X(S_n) = v$ (i.e., the states of $X(t)$ just before and just after the nth regeneration point) for some fixed n and $u,v \in E$, then $X(S_n^-) = u$ and $X(S_n) = v$ for all $n = 0,1, \ldots$. The state pair (u,v) will be referred to as the *regenerative state transition* for the process $X(t)$, and $X(t)$ will be referred to as a *(u,v) regenerative process* whenever necessary to distinguish the states u and v. This terminology is convenient and compact; it will become clearer when "almost regenerative processes" are introduced in which the state transitions of interest are allowed between sets of states. Further, this terminology will not restrict the generality of our results.

Now consider a regenerative process $\{X(t), t \geq 0\}$ which converges in distribution to a random variable X, denoted by $X(t) \xrightarrow{\mathcal{D}} X$, and suppose a point estimate of the constant $\theta = E\{f(X)\}$ is desired, where f is a real-valued, integrable function defined on X. For example, if $f(x) = x$ then θ is the mean of X. If $f(x) = 1$ when $x \leq a$ and $f(x) = 0$ when $x > a$, then $\theta = P\{X \leq a\}$.

A cumulative process $\{A_n, n = 1,2, \ldots\}$ may be defined on $X(t)$ by

$$(1.1)\qquad A_n = \int_{S_{n-1}}^{S_n} f[X(u)]du$$

where the $\{S_n\}$ are the regeneration points of $X(t)$ [corresponding to a specific choice of regenerative state transition (u,v)]. Since $X(t)$ is regenerative, the $\{A_n\}_1^\infty$ are iid. When $E\int_{S_0}^{S_1} |f[X(u)]|du < \infty$ [Brown and Ross (1970)],

$$(1.2)\qquad \lim_{t\to\infty} \int_0^t f[X(u)]du/t = \theta$$

with probability one (w.p.1). And from Smith (1955), Theorem 7,

$$(1.3)\qquad \lim_{t\to\infty} \int_0^t f[X(u)]du/t = \frac{E\{A_1\}}{E\{Y_1\}} \quad \text{w.p.1} .$$

Relations (1.2) and (1.3) are also true in expectation. From (1.2) and (1.3) it is immediate that

$$(1.4)\qquad \theta = \frac{E\{A_1\}}{E\{Y_1\}} .$$

When $X(t)$ is a general regenerative process, this result can be restated as $\theta = E\{A_2\}/E\{Y_2\}$.

The form of (1.4) suggests a straight-forward method, called the *regenerative method* (RM), for estimating θ . Suppose the regenerative process $\{X(t), t \geq 0\}$ is observed for n regenerative cycles and the pairs (A_i, Y_i) are recorded for cycle i , $i = 1,2, \ldots, n$, where Y_i is the ith cycle length and A_i is defined by (1.1).

Since the pairs $\{(A_i,Y_i)\}_1^n$ are iid, it follows from the Strong Law of Large Numbers (SLLN) and (1.4) that

(1.5) $$\lim_{n\to\infty} \frac{\bar{A}(n)}{\bar{Y}(n)} = \theta \quad \text{w.p.1} ,$$

where $\bar{A}(n) = \sum_1^n A_i/n$ and $\bar{Y}(n) = \sum_1^n Y_i/n$. Thus, a regenerative estimator of θ is

(1.6) $$\hat{\theta}(n) = \frac{\bar{A}(n)}{\bar{Y}(n)} , \; n = 1,2, \ldots .$$

The RM eliminates some of the shortcomings of the fixed time increment method (FTM), which was described in Chapter 1. The observations generated by the FTM may be correlated, thus complicating estimates of estimator variance; there is no autocorrelation between the RM observations $\{(A_i,Y_i)\}_1^n$ which are mutually independent. Also, the FTM observations are influenced by the initial state of the system, thus encouraging one to "throw away" the first few observations to reduce estimator bias [Conway (1963)]; since the RM observations are identically distributed, data can be collected from the beginning of the simulation.

However, the RM does have some drawbacks of its own. First, the ratio estimator (1.6) of θ is biased because, in general, the expected value of a ratio is not the ratio of the expected values. For sufficiently large sample sizes, and under the assumption that the (A_i,Y_i) are iid bivariate Normal, it is known [Tin (1965)] that

(1.7) $$E\{\hat{\theta}(n)\} = \theta[1 - \frac{1}{n}(C_{11} - C_{02})] + O(n^{-1})$$

where

$$(1.8) \qquad C_{ij} = \frac{E\{(A_1 - EA_1)^i (Y_1 - EY_1)^j\}}{E^i\{A_1\}\, E^j\{Y_1\}} \, , \; i,j = 0,1 \ldots$$

is the bivariate coefficient of variation of order (i,j), and $O(n^{-k})$ has the property that $n^k O(n^{-k}) < \infty$ as $n \to \infty$. Note that C_{ij} depends on the choice of regenerative state transition (u,v). Other ratio estimators of θ which have smaller biases and mean-square-errors than those of (1.6) are compared by Tin (1965).

The RM has a more serious defect, however. The expected time between regeneration points, $E\{Y_1\}$, may not be small enough for practical considerations. Of course, it may also be possible that $\{X(t), t \geq 0\}$ may not be regenerative, in which case there are no states (u,v) with the desired property. In spite of this problem, an extension of the RM may still be used to collect data, from simulations of the process, which can be conveniently used to estimate θ and the variance of the estimator of θ.

2.2 The Almost Regenerative Method (ARM)

The two basic approximation strategies one may take when dealing with stochastic processes which do not have convenient regeneration state transitions can be classified as "state space transformation" methods and "almost regenerative" methods.

The central idea of the state space transformation methods [Crane and Iglehart (1975b)] is to redefine the stochastic process in such a way that the new process has convenient regeneration state transitions. The RM may then be applied directly to the new process, from which inferences may be drawn about the process of interest. The danger with these methods is that they may produce estimators which are not consistant in the usual statistical sense. For example, consider a GI/G/1 queue with continuous interarrival and service distributions. By discretizing these two distributions to obtain a regenerative process, certain random variables may be adversely affected. This can be seen by considering the busy cycle length process (the time between successive arrivals finding the system empty). Since there is a positive probability that the only customer in the system will complete his service at the same time an arrival occurs, the definition of what constitutes a busy cycle is a matter of convention. If departures are processed before arrivals, the new arrival begins a new busy cycle, otherwise the new arrival contributes to the continuation of the present busy cycle started by some previous arrival. Thus, in one case the busy cycles are "short" and in the other case they are "long"; neither may provide a good approximation to the actual busy cycle length distribution.

Unlike the state space transformation methods, the almost regenerative methods do not change the process of interest. Rather, they simply modify the conditions under which observations are collected. Recall that observations of the process, when using the RM, are made each time the process makes a (u,v) transition, and the resulting observations are iid. This is a desireable property. As motivation for the concept of an almost regenerative method, consider the following definition due to Smith (1955), which is a weakening of the definition of a regenerative process.

A stochastic process $\{X(t), t \geq 0\}$ with state space E is said to be *loosely regenerative* if, for $n = 0,1, \ldots$ and $m = 1,2, \ldots$ there exists a sequence of epochs $\{Z_n\}$ and a constant $\lambda \equiv \lambda(E_1, \ldots, E_m) \geq 0$ such that

$$\text{(2.1)} \quad P\{X(Z_n + t_i) \in E_i,\ i = 1,2, \ldots, m \mid X(s),\ s \leq Z_n - \lambda\}$$
$$= P\{X(Z_n + \lambda + t_i) \in E_i,\ i = 1,2, \ldots, m\}$$

for $E_i \subset E$ and $t_i \geq 0$, $i = 1,2, \ldots, m$. The difference between a regenerative process and a loosely regenerative process may be explained as follows. In a regenerative process the past history, except in so far as it determines the most recent regeneration point, loses all predictive value whenever a regeneration point occurs. In a loosely regenerative process the past continues to have some influence on the behavior of the process for a further amount of time λ after the epochs $\{Z_n\}$.

As an example of a loosely regenerative process, consider a GI/G/c queue where service time at each channel is a constant, say α , and suppose the process of interest is the number of customers in the system. This process is regenerative at those epochs when an arrival finds the system empty. However, it is not regenerative at those epochs when an arrival finds $n (n > 0)$ customers in the system since the future behavior of the process depends on the remaining service time of those customers in service, as well as n . However, if the history of the process were known for α time units before an arrival epoch, the remaining service times of the customers in service would be known, and thus the future probabilistic behavior of the process would also be known. In this case, the process of interest is loosely regenerative, with $\lambda = \alpha$, when an arrival finds n customers in the system, for fixed $n \geq 0$.

This example suggests an "almost regenerative" method for simulating a queueing system. Let $E\{T\}$ be the expected interarrival time of customers to the above described GI/G/c queue, and suppose that the system is stable, in the sense that $\rho = \alpha/cE\{T\} < 1$. For simulation purposes, one could take observations of the process each time an arrival finds n in the system; these epochs do not constitute regeneration points, but one might expect the influence of the remaining service times on the future to become negligible as the time seperation between two lagged observations increases. To see what is meant by this statement, consider the following definition.

A stationary process $\{X(t), t \geq 0\}$ with state space E is said to be *ϕ-mixing* [cf. Billingsley (1968)] if, for every set $F \subset E$, there exists a monotone non-increasing, real-valued function ϕ such that

$$(2.2) \qquad |P\{X(t + \beta) \in F \mid X(s), s \leq t\} - P\{X(t + \beta) \in F\}| \leq \phi(\beta)$$

for $t,\beta \geq 0$, where only functions ϕ satisfying $\lim_{\beta \to \infty} \phi(\beta) = 0$ are considered. The sequence of observations generated by an arrival finding n customers in the system generally is not stationary, so the ϕ-mixing definition does not apply directly. However, the essense of (2.2) is that separate observations of the process are asymptotically independent.

The data collection method outlined above may be generalized. Consider the stochastic process $\{X(t), t \geq 0\}$ with state space E , and let U and V be two disjoint subsets of E (U and V need not partition E). Also, let $\{Y'_n\}_1^\infty$ be a sequence of non-negative random variables where Y'_n is the time between the $(n - 1)$st and the nth transition of $X(t)$ from any state which is an element of U to any state which is an element of V . Define $S'_0 = 0$ and $S'_n = Y'_n + S'_{n-1}$, $n = 1,2, \ldots$. The sequence $\{Y'_n\}_1^\infty$ will be called the *(U,V) almost renewal process* and $\{S'_n\}_0^\infty$ will be called the *(U,V) almost regeneration points (epochs)* of the $X(t)$ process. Note that no conditions are placed on these definitions other than the transitions of interest be from the set U to the set V .

The following *notational convention* will be observed throughout the remainder of this chapter: all symbols associated with a (U,V) representation of the $X(t)$ process will be so designated by a prime ('). Usually these symbols will have a direct counterpart in a (u,v) regenerative process.

Let $X(t) \overset{\mathcal{D}}{\to} X$ and suppose a point estimate of the constant $\theta = E\{f(X)\}$ is desired, where f is a real-valued, integrable function defined on X . A process $\{A'_n\}_1^\infty$ may be defined by

$$A'_n = \int_{S'_{n-1}}^{S'_n} f[X(u)]du \ , \ n = 1,2, \ldots \tag{2.3}$$

where the $\{S'_n\}_0^\infty$ are the (U,V) almost regenerative points of $X(t)$ corresponding to a particular choice of $U,V \subset E$. The *almost regenerative method* (ARM) for obtaining an estimate of θ is similar to the regenerative method (RM). Suppose the $X(t)$ process is observed for n' (U,V) almost regenerative cycles, and the pairs (A'_i,Y'_i) are recorded for cycle $i = 1,2, \ldots, n'$. Then an estimate of θ is

$$\hat{\theta}'(n') = \frac{\bar{A}'(n')}{\bar{Y}'(n')} \tag{2.4}$$

where $\bar{A}'(n')$ and $\bar{Y}'(n')$ are the sample means.

Conditions exist such that $\hat{\theta}'(n') \underset{n'\to\infty}{\longrightarrow} \theta$ w.p.1 . Although this type of result is felt to hold for fairly general stochastic processes, it will be shown here only when the $X(t)$ process has at least one embedded (u,v) regenerative process; i.e., there exist $u,v \in E$ such that $X(t)$ is (u,v) regenerative. Let $N(n')$ be the number of (u,v) regeneration points during the first n' (U,V) almost regeneration points, and let N_i be the number of (U,V) almost regeneration points during the ith (u,v) cycle. Note that the $\{N_i\}$ are iid. To avoid trivialities, assume for the remainder of this chapter that the sets $U,V \subset E$ are chosen in such a way that $E\{N_1\} < \infty$ and $P\{Y'_n > 0\} > 0$.

The following technical lemma will be needed.

Lemma 2.1:

If X_i is a random variable defined on the ith (u,v) regenerative cycle and $E\{X_i\} < \infty$, then

$$\lim_{n\to\infty} \frac{X_{n+1}}{n} = 0 \text{ w.p.1} . \tag{2.5}$$

Proof:

Since the $\{X_i\}$ are iid, and by the SLLN, (2.6) $$E\{X_1\} = \lim_{n\to\infty} \frac{\sum_1^{n+1} X_i}{n+1} = \lim_{n\to\infty} \frac{\sum_1^{n} X_i}{n} \cdot \frac{n}{n+1} + \lim_{n\to\infty} \frac{X_{n+1}}{n} \cdot \frac{n}{n+1}$$

$$= E\{X_1\} + \lim_{n\to\infty} \frac{X_{n+1}}{n}$$

w.p.1. The result then follows.

Lemma 2.1 can easily be extended to the following lemma which will be stated without proof.

Lemma 2.2:

If X_i is a random variable defined on the ith (u,v) regenerative cycle, $E\{X_i\} < \infty$, and $\{N(n)\}$ is a sequence of random variables that goes to ∞ as $n \to \infty$, then

(2.7) $$\lim_{n\to\infty} \frac{X_{N(n)+1}}{N(n)} = 0 \text{ w.p.1 .}$$

Theorem 2.3:

If $\{X(t), t \geq 0\}$, with state space E, has an embedded (u,v) regenerative process such that $u \in U$ and $v \in V$ for some $U,V \subset E$, and if $E\int_{S_{n-1}}^{S_n} |f[X(u)]|du < \infty$ for all $n = 1,2, \ldots$, then the estimator (2.4) is consistant; i.e., $\lim_{n'\to\infty} \hat{\theta}'(n') = \theta$ w.p.1.

Proof:

Let $K(n') = \sum_1^{N(n')} N_i$. Then

(2.8) $$\lim_{n'\to\infty} \bar{Y}'(n') = \lim_{n'\to\infty} \frac{\sum_1^{n'} Y_i'}{n'} = \lim_{n'\to\infty} \frac{\sum_1^{N(n')} Y_j + \sum_{K(n')+1}^{n'} Y_i'}{K(n') + n' - K(n')} .$$

Since $0 \le \sum_{K(n')+1}^{n'} Y_i' \le Y_{N(n')+1}$, and since $E\{Y_j\} < \infty$ for all j (Y_j is the time between the $(j-1)$st and the jth (u,v) renewal points), it follows from Lemma 2.2 that $\sum_{K(n')+1}^{n'} Y_i'/N(n') \to 0$ w.p.1. Similarily, $[n' - K(n')]/N(n') \to 0$ w.p.1. Hence, dividing both numerator and denominator of (2.8) by $N(n')$ and applying the SLLN yields

$$\text{(2.9)} \qquad \lim_{n'\to\infty} \bar{Y}'(n') = \frac{E\{Y_1\}}{E\{N_1\}} \quad \text{w.p.1.}$$

To obtain a similar result for $\bar{A}'(n')$, let the positive and negative parts of $f(x)$ be defined by $f^+(x) = \max [0,f(x)]$ and $f^-(x) = -\min [0,f(x)]$. Then $f(x) = f^+(x) - f^-(x)$ for all x . Since $E\int_{S_{n-1}}^{S_n} |f[X(u)]|du < \infty$ for all $n = 1,2, \ldots$ (by assumption), and this implies that $\int_{S_{n-1}}^{S_n} |f[X(u)]|du < \infty$ w.p.1 for all $n = 1,2, \ldots$, the same is also true for $f^+(x)$ and $f^-(x)$. Define $A_{n'}'^+$ and $A_{n'}'^-$ to be the positive and negative parts of $A_{n'}'$. These terms are non-negative. Thus

$$\text{(2.10)} \qquad 0 \le \sum_{K(n')+1}^{n'} A_i'^+ \le \int_{S_{N(n')}}^{S_{N(n')+1}} f^+[X(u)]du \ .$$

But the right-hand side of (2.10) has finite expectation, so application of Lemma 2.1 yields $\sum_{K(n')+1}^{n'} A_i'^+ \; N(n') \to 0$ w.p.1 . A similar result is true for the negative parts of A_i' , $i = K(n')+1, \ldots, n'$. Hence the same arguments that led to (2.9) yield

$$\text{(2.11)} \qquad \lim_{n'\to\infty} \bar{A}'(n') = \frac{E\{A_1\}}{E\{N_1\}} \quad \text{w.p.1 .}$$

Then from (2.9), (2.11), and (1.4),

(2.12) $$\lim_{n'\to\infty} \hat{\theta}'(n') = \frac{E\{A_1\}}{E\{Y_1\}} = \theta \quad \text{w.p.1},$$

as was to be shown.

When applying the ARM to a stochastic process $X(t)$, the object is to define the sets $U,V \subset E$ such that the serial correlation between observations and the expected time between successive (U,V) almost regeneration points are sufficiently small to justify use of the regenerative estimators. For example, to estimate the expected delay in queue, d, for an M/M/1 model, Crane and Iglehart (1975b) set

(2.13) $$V = \{x : x \in [d - \varepsilon, d + \varepsilon]\} , \; U = R^+ - V$$

(where d was known) for some $\varepsilon > 0$. They found emperically that values of $\varepsilon < d$ produced little autocorrelation between successive (U,V) almost regenerative cycles. Interestingly, they also found that as ε was chosen to increase the frequency of observation, the accuracy of the estimators *actually improved*. One would expect, however, that this desireable property would not persist as the "trapping interval" width $[d - \varepsilon, d + \varepsilon]$ increases without bound, since successive customer delays are correlated. More will be said about this issue in Chapters 3 and 4.

2.3 The Asymptotic Distribution of the Almost Regenerative Estimator

It is well-known [cf. Law (1974)] for the regenerative estimator $\hat{\theta}(n)$ of θ, given by (1.6), that if $0 < E\{A_1^2\} < \infty$, $0 < E\{Y_1^2\} < \infty$, and $E\int_{S_{n-1}}^{S_n} | f[X(u)] | du < \infty$ for all $n = 1,2, \ldots,$ then

$$\frac{\sqrt{n}\,[\hat{\theta}(n) - \theta]}{\sqrt{M}} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1) \tag{3.1}$$

where

$$M = \frac{V\{A_1 - \theta Y_1\}}{E^2\{Y_1\}} = \theta^2\left[\frac{E\{A_1^2\}}{E^2\{A_1\}} - \frac{2E\{A_1Y_1\}}{E\{A_1\}E\{Y_1\}} + \frac{E\{Y_1^2\}}{E^2\{Y_1\}}\right] \tag{3.2}$$

and $N(0,1)$ is the unit Normal distribution function. Similar results will be shown in this section for the almost regenerative estimator $\hat{\theta}'(n')$ given by (2.4), under the condition that $\{X(t), t \geq 0\}$ is regenerative.

The following obvious lemma will be stated without proof.

Lemma 2.4:

If the $\{X(t), t \geq 0\}$ process, with state space E, has at least one embedded (u,v) regenerative process, for some $U,V \subset E$, then the expected number of (U,V) almost regeneration points, N_i, during the ith (u,v) regenerative cycle is at least one; i.e.,

$$E\{N_1\} \geq 1 . \tag{3.3}$$

Theorem 2.5:

If the $\{X(t), t \geq 0\}$ process is as in Lemma 2.4, and if $0 < E\{A_1^2\} < \infty$, $0 < E\{Y_1^2\} < \infty$, and $E\int_{S_{n-1}}^{S_n} |f[X(u)]|du < \infty$ for all

$n = 1,2, \ldots,$ then

$$(3.4) \qquad \frac{\sqrt{n'}\,[\hat{\theta}'(n') - \theta]}{\sqrt{M'}} \xrightarrow[n'\to\infty]{\mathcal{D}} N(0,1)$$

where

$$(3.5) \qquad M' = E\{N_1\}M = E\{N_1\}\theta^2 \left[\frac{E\{A_1^2\}}{E^2\{A_1\}} - \frac{2E\{A_1Y_1\}}{E\{A_1\}E\{Y_1\}} + \frac{E\{Y_1^2\}}{E^2\{Y_1\}} \right]$$

and $\hat{\theta}'(n')$ is defined by (2.4).

<u>Proof</u>:

Define $Z_j = A_j - \theta Y_j$ and $Z_i' = A_i' - \theta Y_i'$ for $j = 1,2, \ldots, N(n')$, and $i = 1,2, \ldots, n'$. Then

$$(3.6) \qquad \frac{\sqrt{n'}\,[\hat{\theta}'(n') - \theta]}{\sqrt{M'}} = \frac{E\{N_1\} \sum_1^{n'} Z_i'}{E\{Y_1\}\sqrt{n'M'}} \cdot \frac{E\{Y_1\}n'}{E\{N_1\} \sum_1^{n'} Y_i'} ,$$

where $E\{N_1\} \geq 1$ by Lemma 2.4. Also

$$(3.7) \qquad \frac{E\{N_1\} \sum_1^{n'} Z_i'}{E\{Y_1\}\sqrt{n'M'}} = \frac{\sum_1^{N(n')} Z_j}{\left[N(n')V(Z_1)\cdot\frac{1}{E(N_1)}\left[\frac{K(n')}{N(n')} + \frac{n' - K(n')}{N(n')}\right]\right]^{1/2}} + \frac{\sum_{K(n')+1}^{n'} Z_i'}{\left[n'\cdot\frac{V(Z_1)}{E(N_1)}\right]^{1/2}}$$

where $K(n') = \sum_1^{N(n')} N_j$ and $V\{Z_1\} = E\{A_1^2\} - 2\theta E\{A_1Y_1\} + \theta^2 E\{Y_1^2\}$.

Since $0 \le n' - K(n') \le N_{N(n')+1}$, Lemma 2.2 and the SLLN yield

$$\text{(3.8)} \qquad \frac{1}{E\{N_1\}}\left[\frac{K(n')}{N(n')} + \frac{n' - K(n')}{N(n')}\right] \underset{n'\to\infty}{\longrightarrow} 1 \quad \text{w.p.1} .$$

Using a similar argument for the positive and negative parts of Z_j and Z_i' as was used in Theorem 2.3,

$$\text{(3.9)} \qquad \sum_{K(n')+1}^{n'} Z_i' \Big/ \sqrt{n'} \underset{n'\to\infty}{\longrightarrow} 0 \quad \text{w.p.1} .$$

By the central limit theorem for sums of iid random variables,

$$\text{(3.10)} \qquad \frac{\sum_{1}^{N(n')} Z_i}{\sqrt{N(n')V\{Z_1\}}} \underset{n'\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N(0,1) ,$$

since $V\{Z_1\}$ is positive and finite by the assumed conditions on A_1 and Y_1 . Combining (3.7) - (3.10) and applying Theorem 4.4.8 from Chung (1968) $\left[\text{If } K_n \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} K^* \text{ and } L_n \underset{n\to\infty}{\longrightarrow} 1 \text{ w.p.1 , then } K_n L_n \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} K^*\right]$ yields

$$\text{(3.11)} \qquad \frac{E\{N_1\} \sum_{1}^{n'} Z_i'}{E\{Y_1\} \sqrt{n'M'}} \underset{n'\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N(0,1) .$$

Since

$$\text{(3.12)} \qquad \frac{E\{Y_1\}n'}{E\{N_1\} \sum_{1}^{n'} Y_i'} \to 1 \quad \text{w.p.1}$$

from (2.9), a second application of Chung's Theorem 4.4.8 to (3.11) and (3.12) yields the desired result.

Note that (3.2) and (3.5) are the same when $U = \{u\}$ and $V = \{v\}$. Also note that Theorem 2.5 implies Theorem 2.3.

2.4 The Ratio Estimators

Let $\{(X_i, Y_i)\}_1^n$ be n iid observations of the bivariate random variable (X,Y), which has means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and covariance σ_{XY}. Assume that these moments are finite and that $Y_i > 0$ w.p.1. Estimators of $\theta = \mu_X/\mu_Y$ are called *ratio estimators*.

The *simple ratio estimator* of θ [already encountered in (1.6)] is

(4.1) $$\hat{\theta}_c(n) = \frac{\bar{X}(n)}{\bar{Y}(n)},$$

where $\bar{X}(n) = \sum_1^n X_i/n$ and $\bar{Y}(n) = \sum_1^n Y_i/n$ are the sample means. For sufficiently large n, a power series expansion shows that the dominant bias of (4.1) is of order $1/n$,

(4.2) $$E\{\hat{\theta}_c(n)\} = \theta + a/n + b/n^2 + \ldots$$

where a and b are constants of the expression. As a result, an extensive literature has developed in an effort to create efficient, estimators of θ which reduce the bias of (4.2).

One such estimator is the jackknife ratio estimator due originally to Quenouille (1944, 1956) and Durbin (1959) [see Miller (1974) for a comprehensive review of the history and properties of this estimator].

The essence of the jackknife is to divide the n observations into g groups of h observations each $(n = gh)$. Let $\hat{\theta}_c(n)$ be the estimate of θ based on the full sample, and define $\hat{\theta}_{-i}(g,h)$ to be the corresponding estimator based on a sample of size $(g - 1)h$, where the ith group of size h has been deleted. The *generalized jackknife estimator* of θ is

$$\hat{\theta}_J(g,h) = \frac{1}{g} \sum_{i=1}^{g} \hat{\theta}_i(g,h) , \tag{4.3}$$

where

$$\hat{\theta}_i(g,h) = g\hat{\theta}_c(n) - (g - 1)\hat{\theta}_{-i}(g,h) , \quad i = 1,2, \ldots, g . \tag{4.4}$$

are called pseudo-values by Tukey (1958).

Since (4.3) is not restricted to ratio estimation, it is interesting to consider the problem of estimating the mean of a sequence of iid random variables $Z_1, Z_2, \ldots, Z_n$. Then

$$\hat{\theta}_i(g,h) = \frac{1}{h} \sum_{(i-1)h+1}^{ih} Z_k , \quad i = 1,2, \ldots, g \tag{4.5}$$

so that (4.3) is a generalization of a batch means estimator.

When originally proposed, Quenouille considered the "simple" jackknife with $g = 2$ and found that the technique eliminated the $1/n$ term from any bias. Since the (X_i, Y_i) are iid, the $\hat{\theta}_i(g,h)$ are interchangeable random variables. Hence, it is straightforward to show from (4.2) that

$$E\{\hat{\theta}_J(g,h)\} = \theta - b/[g(g - 1)h^2] + \ldots , \tag{4.6}$$

which shows that the bias to order $1/n^2$ is minimized for $g = n$ (hence, $h = 1$). The $1/n^2$ (and higher) bias terms may be eliminated by (repeatedly) jackknifing the pseudo-values (4.4) [Schucany, Gray and Owen (1971)].

The principle attraction of the jackknife estimator is the ease with which estimates of the variance of $\hat{\theta}_J$ can be made. In an abstract, Tukey (1958) proposed that the pseudo-values could be treated as g approximately iid observations from which one could expect

$$(4.7) \qquad \frac{\sqrt{g}\,(\hat{\theta}_J - \theta)}{\left[\frac{1}{g-1}\sum_{1}^{g}(\hat{\theta}_i - \theta)^2\right]^{1/2}}$$

to be approximately distributed as a Student - t random variable with $g - 1$ degrees of freedom, and, in the general setting of U - statistics, Arvesen (1969) proved this conjecture for the limiting case as $h \to \infty$. In the same paper, Arvesen also proved that (4.7), with $g = n(h = 1)$, has a limiting unit Normal distribution as $n \to \infty$.

As a simple example of Arvesen's theorems, consider the real-valued function f defined on (X,Y) such that $E\{f(X,Y)\} = \theta$, and let $\{(X_i,Y_i)\}_1^n$ be a sequence of iid observations. From his Theorem 8, if $g = n(h = 1)$, $\sigma_X^2 < \infty$ and $\sigma_Y^2 < \infty$, then

$$(4.8) \qquad \frac{\sqrt{g}\,(\hat{\theta}_J - \theta)}{\left[\frac{1}{g-1}\sum_{i=1}^{g}(\hat{\theta}_i - \hat{\theta}_J)^2\right]^{1/2}} \xrightarrow[g\to\infty]{D} N\left(0, \theta^2\left[\frac{\sigma_X^2}{\mu_X^2} - \frac{2\sigma_{XY}}{\mu_X\mu_Y} + \frac{\sigma_Y^2}{\mu_Y^2}\right]\right).$$

Also, under the same conditions, a special case of his Theorem 9 yields

$$(4.9) \qquad \frac{1}{g-1}\sum_{i=1}^{g}(\hat{\theta}_i - \hat{\theta}_J)^2 \xrightarrow[g\to\infty]{P} \theta^2\left[\frac{\sigma_X^2}{\mu_X^2} - \frac{2\sigma_{XY}}{\mu_X\mu_Y} + \frac{\sigma_Y^2}{\mu_Y^2}\right],$$

where $\xrightarrow{P}$ denotes convergence in probability. Thus, $\hat{\theta}_J$ has the same asymptotic variance as $\hat{\theta}_c(n)$ [see (3.1) in section 2.3]. The form of (4.9) suggests the *jackknife variance estimator* of $\hat{\theta}_J$,

$$(4.10) \qquad \hat{V}\{\hat{\theta}_J(g,h)\} = \frac{1}{g(g-1)}\sum_{i=1}^{g}(\hat{\theta}_i - \hat{\theta}_J)^2 .$$

The alternative to using the simple or jackknife estimators of θ is to use a ratio estimator such as *Beale's ratio estimator* [Beale (1962)],

$$(4.11) \qquad \hat{\theta}_B(n) = \frac{\bar{X}(n)}{\bar{Y}(n)}\left[\frac{1 + \frac{1}{n}\,\frac{s_{XY}}{\bar{X}(n)\bar{Y}(n)}}{1 + \frac{1}{n}\,\frac{s_{YY}}{\bar{Y}^2(n)}}\right]$$

or Tin's ratio estimator [Tin (1965)]

$$(4.12) \qquad \hat{\theta}_T(n) = \frac{\bar{X}(n)}{\bar{Y}(n)}\left[1 + \frac{1}{n}\left(\frac{s_{XY}}{\bar{X}(n)\bar{Y}(n)} - \frac{s_{YY}}{\bar{Y}^2(n)}\right)\right]$$

where

$$(4.13) \qquad s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}[X_i - \bar{X}(n)][Y_i - \bar{Y}(n)]$$

is the sample covariance. The corresponding estimator of the variance of (4.11) and (4.12) [also (4.1)] is the *simple variance* *estimator*

$$(4.14) \qquad \hat{V}(\hat{\theta}(n)) = \frac{1}{n}\left[\frac{\bar{X}(n)}{\bar{Y}(n)}\right]^2 \left[\frac{s_{XX}}{\bar{X}^2(n)} - \frac{2s_{XY}}{\bar{X}(n)\bar{Y}(n)} + \frac{s_{YY}}{\bar{Y}^2(n)}\right] .$$

Since there is a positive probability that (4.14) can be negative for small samples, the jackknife ratio estimator and its associated variance estimator are usually preferred over (4.1), (4.11), (4.12), and (4.14). This choice is further justified by the growing emperical evidence [Rao and Webster (1966), Hutchison (1971), and Iglehart (1974)] which suggests that the mean-square-error of $\hat{\theta}_J$ for small samples is smaller than that of $\hat{\theta}_c$ and is about the same as those of $\hat{\theta}_B$ and $\hat{\theta}_T$.

The simple ratio estimator (4.1) and the jackknife ratio estimator (4.3), along with their associated variance estimators (4.14) and (4.10), will be used for the emperical results of Chapters 3 and 4. Since the almost regenerative method can provide more observations than the regenerative method, for a fixed run length, one might expect that the estimators (4.1), (4.3), (4.10), and (4.14) would perform better with the almost regenerative data than with the regenerative data. These issues will be discussed in the next two chapters.

CHAPTER 3

APPLICATION OF THE ALMOST REGENERATIVE METHOD TO A TANDEM QUEUEING NETWORK

In this chapter the potential power of the almost regenerative method (ARM) will be demonstrated by simulating a queueing network consisting of two M/M/1 queues in tandem. The ARM is compared with the regenerative method (RM), the fixed time increment method (FTM), and a third method called the random observer method (ROM). This last method consists of taking observations of the process each time a Poisson "observer" arrives. The measure of comparison is the mean-square-error (MSE) of the estimator of total expected delay in queue. For the small sample size considered, the ARM estimates of delay are 20.4, 0.6, and 23.7 percent more efficient (in terms of MSE) than the estimates generated by the RM, FTM, and ROM, respectively, when using the simple ratio estimator. When using the jackknife ratio estimator, the estimated improvements in efficiency are 42.5, 3.0, and 25.4 percent, respectively. In addition, the estimates of the variance of the estimator of delay are superior for the ARM generated data.

Similar emperical results are shown for simulations of an $H_4/H_4/1$ tandem queue network, where H_4 represents a hyperexponential distribution with squared coefficient of variation of 4.

3.1 The Almost Regenerative State Space Approach

Consider a system which consists of two M/M/1 queues in tandem where the departure stream of the first queue is the arrival stream for the second queue. Assume that arrivals to the first queue occur at Poisson rate λ and service at each queue is exponential at rate μ. Assume further that each queue is stable in the sense that $\rho \equiv \lambda/\mu < 1$. Let D_k be the steady state delay (not including service) of an arbitrary customer in queue k, $k = 1, 2$, and let his total delay in the network be $D_T = D_1 + D_2$. Then it is well-known that

$$(1.1) \qquad d \equiv E\{D_T\} = E\{D_1\} + E\{D_2\} ,$$

where

$$(1.2) \qquad E\{D_k\} = \frac{\lambda E^2\{S_k\}}{1 - \rho} , \; k = 1,2$$

and $E\{S_k\} = 1/\mu$ is the expected service time in system k.

Suppose an arrival finds the network in state (i,j), where i is the number of customers in system 1 and j is the number in system 2. From a workload point of view, a job is merely a package of "work" for the server. When $i,j > 0$, the total work, V, in the network found by an arrival consists of $2i - 1$ service times for those customers in the first system, $j - 1$ additional service times for those customers in the second system, plus the remaining service times of the two customers in service. From the "memoryless" property of the exponential distribution, the expected remaining service time of a customer in service is the same as the expected service time of an arbitrary customer. Thus, the expected work in the network found by an arrival who finds state (i,j) is $E\{V \mid (i,j)\} = (2i + j)E\{S\}$.

Since service and interarrival times in the M/M/1 tandem network are exponentially distributed, it is well-known that the epochs at which arrivals find state (i,j), for fixed i and j, form a regenerative process. Let p_{ij} be the steady state probability that the network is in state (i,j). Since Poisson arrivals "see" time averages [cf. Wolff (1970)], p_{ij} is also the proportion of time arrivals find the network in state (i,j). Hence, the arrival rate of those arrivals who find state (i,j) is λp_{ij}, and the mean time between such arrivals is

$$E\{T_{ij}\} = \frac{1}{\lambda p_{ij}} \quad . \tag{1.3}$$

It is also well-known [Reich (1957)] that, for the M/M/1 tandem network,

$$p_{ij} = (1 - \rho)^2 \rho^{i+j} \quad , \tag{1.4}$$

where $\rho \equiv \lambda/\mu \in (0,1)$, so that (1.3) has a unique minimum at $i = j = 0$. Thus

$$E\{T_{oo}\} = \frac{1}{\lambda(1 - \rho)^2} \tag{1.5}$$

is the smallest expected length of a regenerative cycle, over all possible choices of (i,j), for the M/M/1 tandem network. Since $E\{T_{oo}\}$ increases and is unbounded as ρ increases toward 1, the regenerative method might require excessively long simulation runs to obtain a specified degree of accuracy.

The almost regenerative method, on the other hand, can lessen the severity of this problem. Define the set of states

(1.6) $A(n) = \{(i,j) : 2i + j = n ; i,j = 0,1, \ldots \}$

for $n = 0,1, \ldots.$ Then each time an arrival to the tandem network finds any state $(i,j) \varepsilon A(n)$, for fixed n , he finds n equivalent services ahead of him. The expected work in the network found by an arrival to the set $A(n)$ is thus

(1.7) $$E\{V \mid A(n)\} = nE\{S\} .$$

Arrivals to the set $A(n_o)$, for fixed n_o , form an "almost" regenerative process. Note that this process is the superposition of a number of regenerative processes [namely those generated by the elements of $A(n_o)$] , and thus the asymptotic results of Theorems 2.2 and 2.4 will hold.

Analogous to the regenerative case, define T_n' to be the time between successive arrivals who find n equivalent services in the network and p_n' to be the steady state probability that there are n equivalent services in the network. The distribution $\{p_n'\}_0^\infty$ is then given by

(1.8) $$p_n' = \sum_{A(n)} (1 - \rho)^2 \rho^{i+j}, \; n = 0,1, \ldots$$

which simplifies to

(1.9) $$p_{2n}' = (1 - \rho)\rho^n(1 - \rho^{n+1}), \; p_{2n+1}' = \rho p_{2n}', \; n = 0,1, \ldots.$$

As in (1.3), the expected time between successive arrivals finding n equivalent services in the network is

$$E\{T'_n\} = \frac{1}{\lambda p'_n} . \tag{1.10}$$

Since one purpose of the ARM is to increase the number of observations, for fixed run length, over that possible from the RM, it is useful to know the conditions under which (1.10) is less than (1.3). Straightforward algebra reveals that (1.10) achieves its minimum value at n_o , where n_o is the nearest non-negative integer to

$$\gamma = - \frac{\ln 2\rho}{\ln \rho} \tag{1.11}$$

for $\rho \in (0,1)$. From (1.11) n_o is positive if, and only if, ρ is greater than $(1/2)^{2/3}$, or about .630. As a consequence, $E\{T'_{n_o}\} < E\{T_{oo}\}$ if, and only if, the same condition is satisfied. Otherwise $E\{T'_{n_o}\} = E\{T_{oo}\}$. For this reason a value for ρ which is greater than .630 was used for the simulation experiment described below.

3.2 Simulation of the M/M/1 Tandem Network

A simulation of the network consisting of two M/M/1 queues in tandem was made to compare the relative efficiencies of four data collection methods for estimating the total expected delay in queue, d . The first method is the regenerative method, where regeneration points occur when an arrival finds the system in state (i_o,j_o) . This method will be referred to as the RM(i_o,j_o) . The second method is the almost regenerative method, ARM$[A(n_o)]$, where n_o is picked according to (1.11), and the third method is the fixed time increment method, FTM, where the constant sampling interval size is $\Delta = E\{T'_{n_o}\}$. The fourth method consists of taking observations at random intervals according to a Poisson process with mean $E\{T'_{n_o}\}$. The purpose of this "random observer" method, denoted by ROM, is to demonstrate the obvious fact that observations of the cumulative delay process may not be independent, even when the times between observations are. In contrast, however, observations of the cumulative process for the ARM $[A(n_o)]$ will be shown to act as if they were iid.

The system utilization factor, ρ, was set at .8 for each M/M/1 queue, with $\lambda = 1$ and $\mu = 1.25$. From (1.1) and (1.2)

$$E\{D_T\} = 6.400 . \tag{2.1}$$

And from (1.3), (1.10) and (1.11),

$$\begin{aligned} i_o = j_o = 0 \ , \ E\{T_{oo}\} = 25 , \\ n_o = 4 \ , \ E\{T'_{n_o}\} = 16 . \end{aligned} \tag{2.2}$$

One hundred independent replications, of m sample observations each, were made for each of the four methods: the RM(0,0), the ARM[A(4)], the FTM(Δ = 16), and the ROM with mean inter-observation time $E\{T'_4\} = 16$. The parameter m was set equal to 100 for each method, with the exception of the RM(0,0) for which m was set according to

$$m = 100\ E\{T'_{n_o}\}/E\{T_{oo}\} , \tag{2.3}$$

so that $m = 64$ for the RM(0,0). The expected number of arrivals

(160,000) to the system was then the same for each method. The interarrival and service times were generated using the CDC 6400 RANF pseudo-random number generator. The initial conditions for each replication were the same; i.e., no customers in the system and all servers idle.

The following notation will be helpful in understanding the simulation results. For each simulation run, r replication estimates of d, $\{\hat{d}_j\}_1^r$, and r replication estimates of $V(\hat{d}_j)$, $\{\hat{V}(\hat{d}_j)\}_1^r$, were collected. The estimators for the jth replication were based on m observations of the cumulative delay process, $\{A_{ij}\}_{i=1}^r$, and m observations of the cumulative arrival process, $\{N_{ij}\}_{i=1}^r$ [see Law (1974)]. The estimates of interest are the estimate of d,

(2.4)
$$\hat{d} = \frac{1}{r} \sum_{j=1}^{r} \hat{d}_j ,$$

the estimated variance of $\hat{d}_j$,

(2.5)
$$\hat{\sigma}_d^2(j) = \frac{1}{r-1} \sum_{j=1}^{r} (\hat{d}_j - \hat{d})^2 ,$$

the estimate of the standard deviation of $\hat{d}$,

(2.6)
$$s_d = \frac{\hat{\sigma}_d(j)}{\sqrt{r}} ,$$

the estimated MSE of $\hat{d}_j$,

(2.7)
$$\hat{MSE}(\hat{d}_j) = \frac{1}{r} \sum_{j=1}^{r} (\hat{d}_j - d)^2 ,$$

and the estimated standard deviation of $\hat{MSE}(\hat{d}_j)$,

(2.8) $$s_{MSE} = \left[\frac{1}{r(r-1)} \sum_{j=1}^{r} [(\hat{d}_j - d)^2 - \hat{MSE}(\hat{d})]^2\right]^{1/2}$$

The two estimates of d are the simple ratio estimator and the jackknife ratio estimator (see Chapter 2, section 2.4). The estimated variance of $\hat{d}_j$ is an *external* estimate since it is formed at the end of the simulation run. An *internal* estimate of the variance of $\hat{d}_j$ may be obtained from the m observations $\{(A_{ij},N_{ij})\}_{i=1}^{m}$ using the simple variance estimator and the jackknife variance estimator described in Chapter 2, section 2.4. These internal estimates will be referred to with tildes ($\sim$), $\tilde{\sigma}_d(j)$, with corresponding estimated standard deviation given by

(2.9) $$\tilde{s}_d = \frac{\tilde{\sigma}_d(j)}{\sqrt{r}} .$$

The simulation results are summarized in Tables 3.1 and 3.2. When using the simple ratio estimator, the ARM estimates of the $MSE(\hat{d})$ are 20.4, 0.6, and 23.7 percent less than the estimated $MSE(\hat{d})$ generated by the RM, FTM, and ROM, respectively. These improvements in efficiency are 42.5, 3.0, and 25.4 percent, respectively, when using the jackknife ratio estimator. The standard deviation of the ARM estimates of d and the MSE are generally the smallest among the four methods, thus lending additional confidence to the results.

The jackknife ratio estimator appears to be less biased than the simple ratio estimator, as expected. Also, the jackknife estimates

M/M/1 Tandem Queue: Comparison of MSE Estimates

Table 3.1

(ρ = .8, d = 6.4, 100 independent replications of m observations each)

Method	CPU Time (sec)	Arrivals (10^3)	m	Simple Ratio Estimator					Jackknife Ratio Estimator				
				$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\hat{MSE}(d_j)$	s_{MSE}	$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\hat{MSE}(d_j)$	s_{MSE}
RM(0,0)	230.9	159.656	64	6.117	.138	1.910	1.991	.386	6.394	.164	2.691	2.691	.723
ARM[A(4)]	238.1	159.701	100	6.149	.123	1.521	1.584	.196	6.402	.124	1.546	1.546	.263
FTM(Δ=16)	238.9	158.834	100	6.237	.125	1.567	1.593	.254	6.237	.125	1.567	1.593	.254
ROM[E{T}=16]	244.6	161.851	100	6.202	.143	2.038	2.077	.552	6.201	.143	2.032	2.072	.550

M/M/1 Tandem Queue: Comparison of Variance Estimates

Table 3.2

(ρ = .8, 100 independent replications of m observations each)

Method	m	Simple Variance Estimator				Jackknife Variance Estimator			
		External Estimate		Internal Estimate		External Estimate		Internal Estimate	
		$\hat{\sigma}_d^2(j)$	$\hat{s}_d$	$\tilde{\sigma}_d^2(j)$	$\tilde{s}_d$	$\hat{\sigma}_d^2(j)$	$\hat{s}_d$	$\tilde{\sigma}_d^2(j)$	$\tilde{s}_d$
RM(0,0)	64	1.910	.386	1.585	.298	2.691	.723	1.723	.655
ARM[A(4)]	100	1.521	.196	1.569	.166	1.546	.263	1.546	.388

of d from the RM and ARM data are nearly unbiased, while the jackknife estimates of d from the FTM and ROM data show no bias improvement over the simple ratio estimates of the same data. This is attributable to the fact that the FTM and ROM observations are not iid, and thus the bias of each observation is not the same; i.e., the bias for the initial observations is less than the bias for the final observations. Note that the jackknife ratio estimator is designed for equal bias in all segments of the data. On the other hand, the RM observations are iid, and the ARM observations "act" as if they were iid. This property of the ARM data will be discussed in the next section.

Assuming that the external estimates of the variance of $\hat{d}_j$ are unbiased (this is not unreasonable since this estimate is based on r independent replications of the experiment), it is interesting to note from Table 3.2 that the internal estimates of $V(\hat{d}_j)$, both for the simple variance estimator and for the jackknife variance estimator, are biased considerably more for the RM data than they are for the ARM data. In fact, the bias reduction is about 80 percent when using the ARM data. The fact that the ARM generates more observations, which act as if they were iid, than the RM, for a fixed run length, is primarily responsible for this bias reduction. The FTM and ROM internal estimates of $V(\hat{d}_j)$ are not shown since the FTM and ROM generated data is highly correlated and produces variance estimates of poor quality.

To give additional confidence to the M/M/1 tandem network results, the same experiment was also conducted for the $H_4/H_4/1$ tandem network. The symbol H_k represents a hyperexponential distribution function with squared coefficient of variation k . The squared coefficient of variation, $CV^2(X)$, for the random variable X , with non-zero mean, is $CV^2(X) = V(X)/E^2(X)$. The distribution H_k for the experiment was taken to be a mixture of the unit step function at zero and an exponential distribution; i.e.,

$$P\{X \leq x\} \equiv H_k(x) = bU_o(x) + (1 - b)e^{-\beta x} \tag{2.10}$$

where $U_o(x) = 1$ if $x = 0$, and $U_o(x) = 0$ if $x > 0$. When $E(X)$ and k are specified, $b = (k - 1)/(k + 1)$ and $\beta = 2[(k + 1)E(X)]^{-1}$. In the case of an arrival process, this distribution represents batch Poisson arrivals, where the batches occur at rate β and the size of the batch is geometrically distributed. In the case of service, H_k represents batch service, where the service rate for a batch is β .

As in the M/M/1 tandem network simulation, ρ was set at .8 for each queue. Since the theoretical steady state distribution of the total number of customers in the network is not known, it was estimated by making a pilot run of 1500 arrivals using the FTM. The mode of the estimated distribution, after smoothing, was near 8, and thus the almost regenerative data collection set [defined by (1.6)] was chosen to be A(8).

The results for the $H_4/H_4/1$ tandem model are summarized and compared in Tables 3.3 and 3.4. To facilitate comparisons, the assumption was made that the regenerative estimates were unbiased. This assumption is actually not true since the jackknife estimators are biased even for iid observations. However, it provides a good approximation since the bias contribution is small in the MSE

$H_4/H_4/1$ Tandem Queue: Comparison of MSE Estimates

Table 3.3

(ρ = .8, d = 7.0, 100 independent replications of m observations each)

Method	CPU Time (sec)	Arrivals (10^3)	m	Simple Ratio Estimator					Jackknife Ratio Estimator				
				$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\widehat{MSE}(d_j)$	s_{MSE}	$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\widehat{MSE}(d_j)$	s_{MSE}
RM(0,0)	455.0	377.413	90	6.744	.142	2.009	2.077	.272	7.000	.156	2.430	2.430	.345
ARM[A(8)]	461.2	366.204	100	6.969	.135	1.819	1.819	.283	7.225	.147	2.175	2.226	.371
FTM(Δ=8.54)	482.9	366.344	100	7.186	.146	2.124	2.162	.310	7.186	.151	2.280	2.314	.309
ROM[E{T}=8.54]	477.4	372.295	100	7.122	.151	2.277	2.291	.268	7.127	.163	2.650	2.667	.273

$H_4/H_4/1$ Tandem Queue: Comparison of Variance Estimates

Table 3.4

(ρ = .8, 100 independent replications of m observations each)

Method	m	Simple Variance Estimator				Jackknife Variance Estimator			
		External Estimate		Internal Estimate		External Estimate		Internal Estimate	
		$\hat{\sigma}_d^2(j)$	$\hat{s}_d$	$\tilde{\sigma}_d^2(j)$	$\tilde{s}_d$	$\hat{\sigma}_d^2(j)$	$\hat{s}_d$	$\tilde{\sigma}_d^2(j)$	$\tilde{s}_d$
RM(0,0)	90	2.009	.272	1.347	.628	2.430	.345	2.151	2.195
ARM[A(8)]	100	1.819	.283	1.639	.882	2.175	.371	2.416	2.779

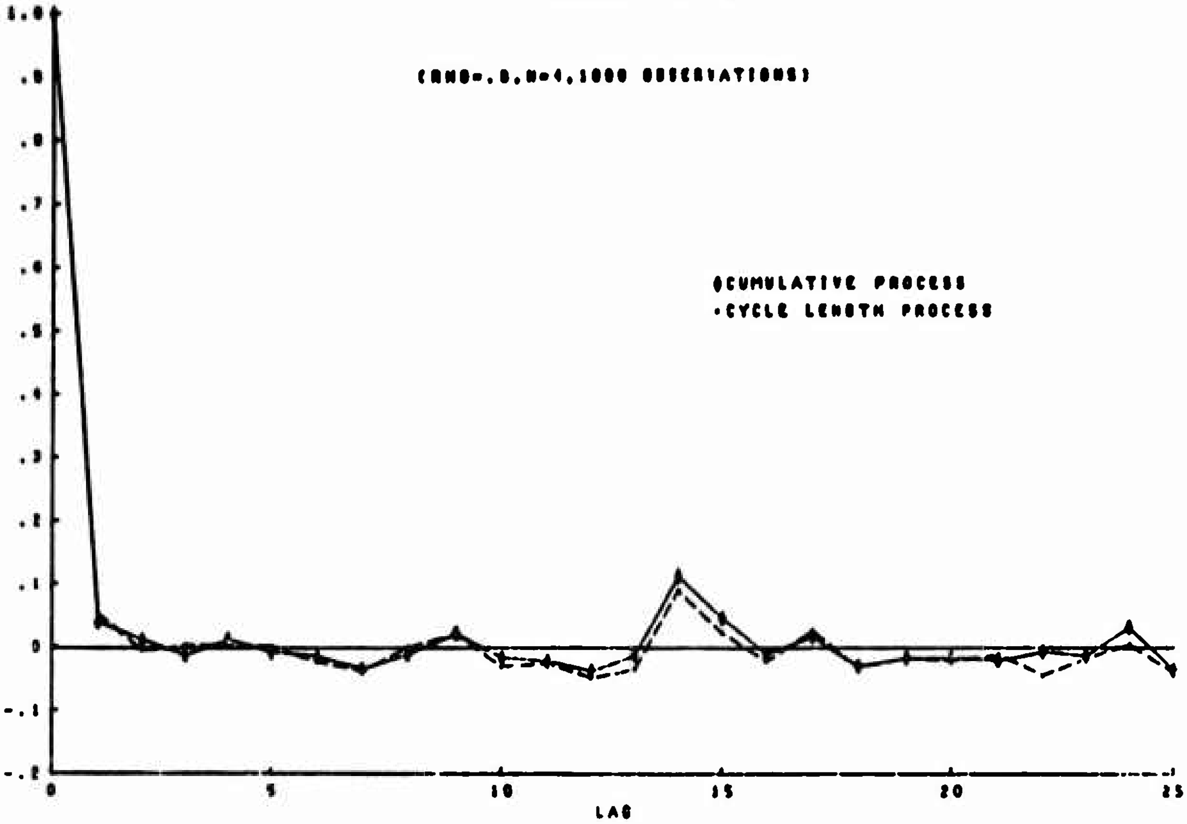
estimates (see Table 3.3). As in the M/M/1 case, the ARM dominates the other methods for estimating d and $V(\hat{d}_j)$.

3.3 Tests of the Renewal Hypothesis

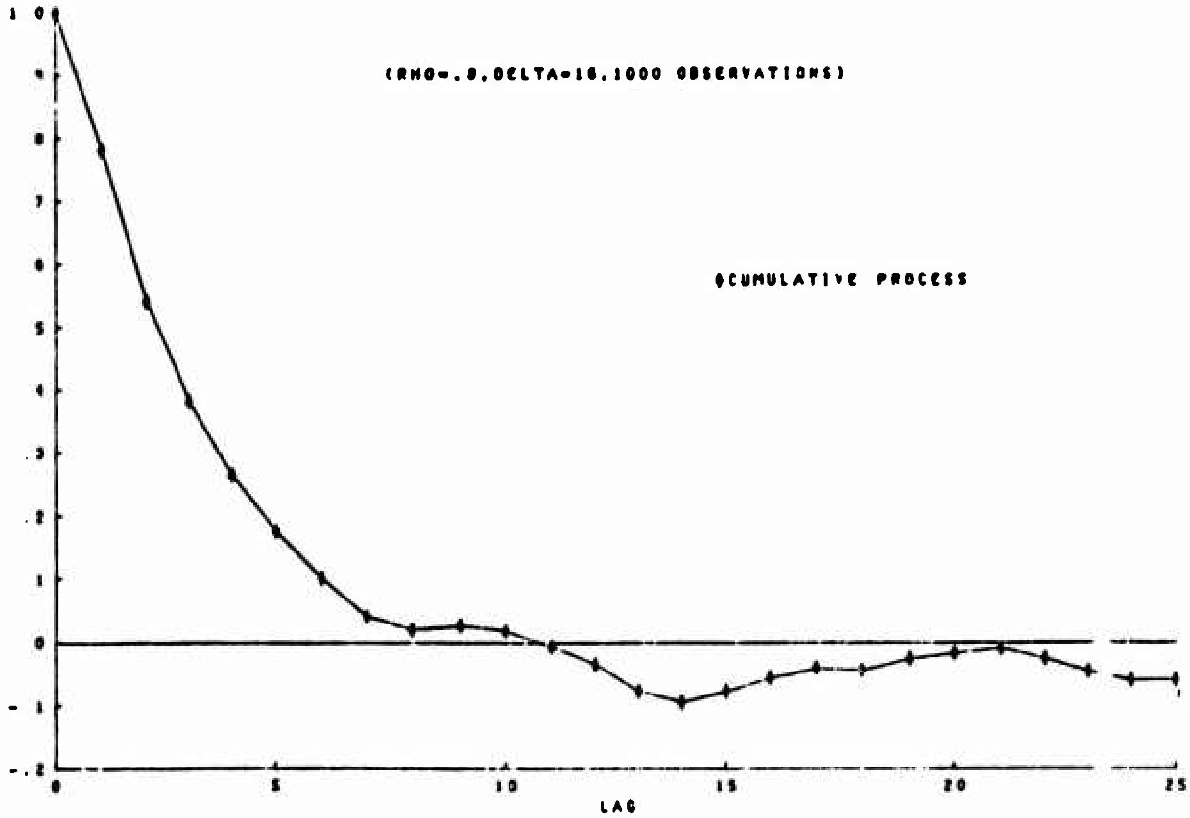
The tests of Appendix I were applied to the almost regenerative cycle length process $\{Y_i'\}_1^m$, where Y_i' is the time between the (i - 1)st and the ith arrivals to find the system in state $A(n_o)$, and the corresponding cumulative delay process $\{A_i'\}_1^m$, where A_i' is the cumulative network delay during Y_i' , for the M/M/1 and the $H_4/H_4/1$ tandem queue models. The purpose of the tests was to determine whether the observed processes *act* like those of a regenerative process.

Figure 3.1 contains a plot of the estimated autocorrelation functions, based on 1000 observations, for the M/M/1 cycle length process and the M/M/1 cumulative delay process generated by the ARM[A(4)]. With the exception of the spike at lag 14, there appears to be no significant autoregressive behavior in either process. The correlation at lag 14 is difficult to explain; possibly it is due to slight nonrandom behavior in the CDC RANF pseudo-random number generator which is known to have short periods in its low order bits [Hutchinson (1966)]. Tests of RANF, using the same initial seeds as for the simulation experiment, confirmed that some periodicities do exist in its estimated spectrum. It is interesting to note that the estimated autocorrelation functions for the RM(0,0), applied to the M/M/1 tandem network, also showed what appeared to be significant autoregressive behavior at lags 9 and 34. The unexplained autoregressive behavior persisted, but at *different* lags,

AUTOCORRELATION FUNCTION FOR 25 LAGS
M/M/1 TANDEM ARM WITH A(N) STATE SPACE
FIGURE 3.1



AUTOCORRELATION FUNCTION FOR 25 LAGS
M/M/1 TANDEM FIXED TIME INCREMENT PROCESS
FIGURE 3.2



when new random number seeds were used for RANF. It appears that this spurious correlation is not significant for the purposes of the present investigation.

The estimated autocorrelation function for the M/M/1 FTM cumulative process, shown in Figure 3.2, has significant correlation for at least five lags. The corresponding processes for the $H_4/H_4/1$ tandem model, shown in Figures 3.3 and 3.4, lead to similar conclusions.

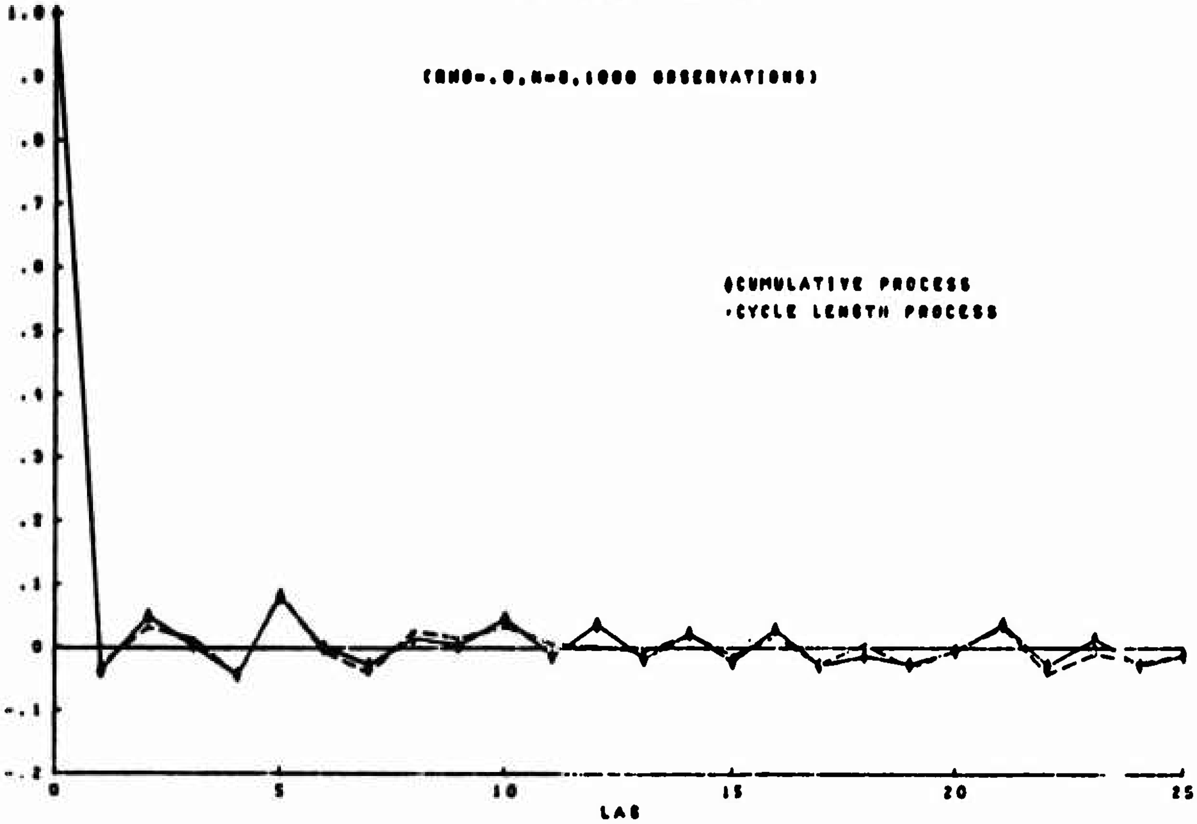
Table 3.5 contains a summary of the renewal test statistics for each of the observed processes. It is interesting that the processes generated by the ARM comfortably pass all tests, at the 95 per cent significance level, for both the M/M/1 *and* the $H_4/H_4/1$ tandem models. Further, none of these processes exhibits any discernable periodicites, at the 95 per cent significance level, in its estimated spectrum (not shown). Thus, the ARM, when applied to the M/M/1 and the $H_4/H_4/1$ tandem network models, generates observations which *act* as though they were iid. As expected, the FTM processes do not pass any of the renewal tests at the 95 per cent level.

3.4 Distribution of the Times Between Transitions

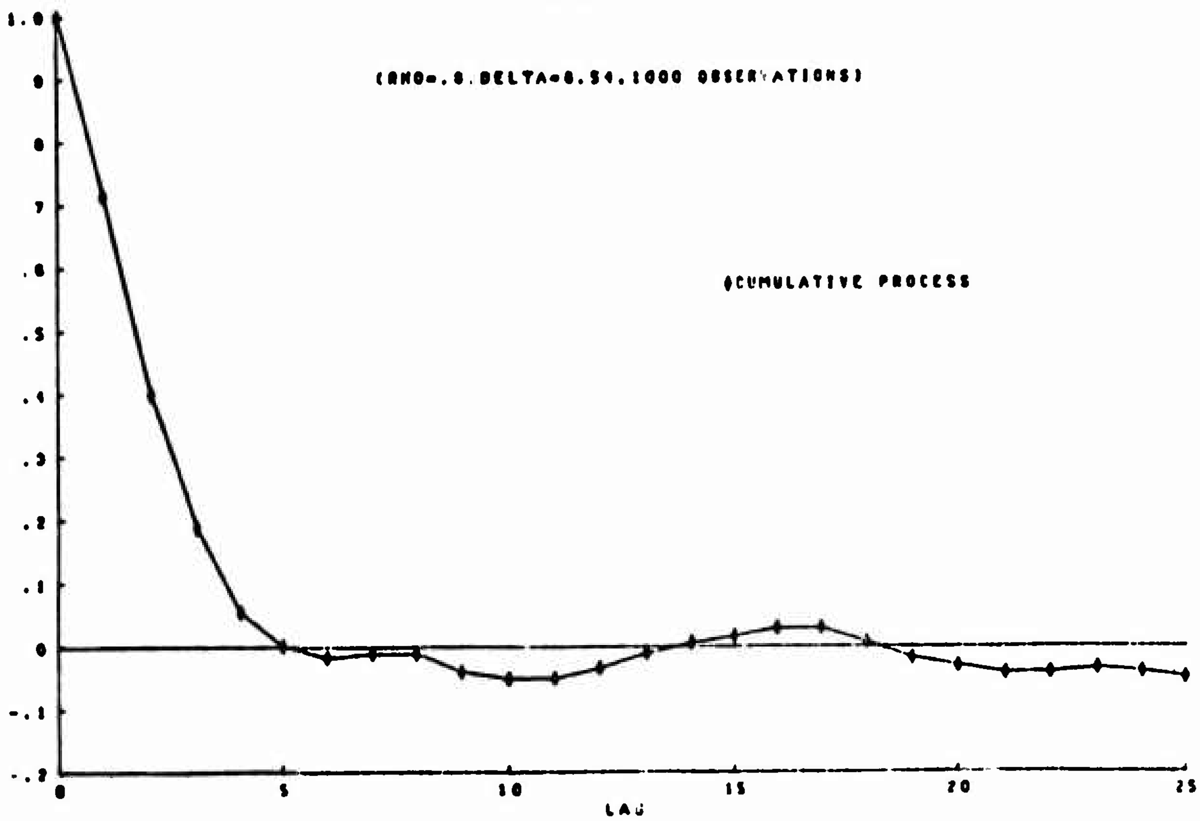
In this section expressions will be derived for the times between state transitions, conditioned on the number of equivalent services in the system, for the M/M/1 tandem network, and it will be shown that these times are distributed "almost" exponentially for certain values of $\rho = \lambda/\mu$.

First, consider a non-negative random variable $X \sim F(t)$; i.e., $P\{X \leq t\} = F(t)$. The *Laplace-Stieltjes transform* (LS-transform) of X is defined by

AUTOCORRELATION FUNCTION FOR 25 LAGS
H/H/1 TANDEM ARM WITH A(N) STATE SPACE
FIGURE 3.3



AUTOCORRELATION FUNCTION FOR 25 LAGS
H/H/1 TANDEM FIXED TIME INCREMENT PROCESS
FIGURE 3.4



SUMMARY OF TEST STATISTICS FOR THE TANDEM QUEUE MODEL

TABLE 3.5

Model	Data Collection	Process+	Lag 1 Autocorrelation	Kolmogorov-Smirnov	Anderson-Darling	Wald-Wolfowitz
M/M/1	ARM	$\{A_i'\}$	.042	.660	.099	1.557
M/M/1	ARM	$\{Y_i'\}$	.055	.727	.164	.975
M/M/1	FTM	$\{A_i\}$	.781	8.978	35.832	21.532
H/H/1	ARM	$\{A_i'\}$	-.037	.885	.126	.570
H/H/1	ARM	$\{Y_i'\}$	-.028	.787	.083	.520
H/H/1	FTM	$\{A_i\}$	.715	7.943	28.179	19.676
Upper 5% significance point			.062	1.358	.461	1.960
Upper 1% significance point			.081	1.628	.743	2.575

+1000 observations

(4.1) $$\tilde{F}(s) = \int_0^\infty e^{-st} dF(t)$$

for $s \geq 0$. The LS-transform can be interpreted as an expectation in the following sense

(4.2) $$\tilde{F}(s) = E\{e^{-sX}\} .$$

This definition will be useful in the following discussion.

Let T be the time between state transitions in the M/M/1 tandem network model. In section 3.1 the steady state probability there are n equivalent services in the network was found to be

(4.3) $$p'_{2n} = (1 - \rho)\rho^n(1 - \rho^{n+1}),\ p'_{2n+1} = \rho p'_{2n},\ n = 0,1, \ldots .$$

Since the LS-transform of an exponential distribution with parameter γ is $\gamma/(\gamma + s)$, the conditional distribution of T, given the number of equivalent services in the network, has the LS-transform

(4.4) $$E\{e^{-sT}|2n\} = \begin{cases} \dfrac{\lambda}{\lambda + s}, & n = 0 \\[2ex] \dfrac{\lambda + \mu}{\lambda + \mu + s} a_{2n} + \dfrac{\lambda + 2\mu}{\lambda + 2\mu + s}(1 - a_{2n}), & n = 1,2, \ldots \end{cases}$$

and

(4.5) $$E\{e^{-sT}|2n + 1\} = \frac{\lambda + \mu}{\lambda + \mu + s} b_{2n+1} + \frac{\lambda + 2\mu}{\lambda + 2\mu + s}(1 - b_{2n+1}),$$
$$n = 0,1, \ldots$$

where

$$(4.6)\qquad \begin{aligned} a_{2n} &= (p_{o,2n} + p_{n,o}) \left[\sum_{(i,j)\varepsilon A(2n)} p_{ij}\right]^{-1}, \quad n = 1,2,\ldots \\ b_{2n+1} &= p_{o,2n+1} \left[\sum_{(i,j)\varepsilon A(2n+1)} p_{ij}\right]^{-1}, \quad n = 0,1,\ldots \end{aligned}$$

and p_{ij} is the asymptotic probability that there are i customers in system 1 and j customers in system 2. From (4.4) and (4.5) it is easy to see that, w.p.1,

$$(4.7)\qquad E\{e^{-sT}|2n\} \to (1-\rho)\frac{\lambda+\mu}{\lambda+\mu+s} + \rho\,\frac{\lambda+2\mu}{\lambda+2\mu+s}, \text{ as } n \to \infty$$

and

$$(4.8)\qquad E\{e^{-sT}|2n+1\} \to \frac{\lambda+2\mu}{\lambda+2\ \ +s}, \text{ as } n \to \infty .$$

Thus, $\{T|2n\}$ converges to a hyperexponential distribution while $\{T \mid 2n+1\}$ converges to an exponential distribution as n becomes large. And, as $\rho \to 1$ and $n \to \infty$, the steady state distribution of T is an exponential distribution.

The squared coefficient of variation of T given n is defined by

$$(4.9)\qquad CV^2\{T|n\} = \frac{V\{T|n\}}{E^2\{T|n\}} = \frac{E\{T^2|n\}}{E^2\{T|n\}} - 1, \quad n = 0,1,\ldots .$$

From (4.7) and (4.8)

$$(4.10)\qquad CV^2\{T|2n\} \to 1 + \frac{\rho}{2}(1-\rho), \text{ as } n \to \infty$$

and

$$(4.11)\qquad CV^2\{T|2n+1\} \to 1, \text{ as } n \to \infty .$$

More interesting, however, is the behavior of T, given n, for small n. From (4.4) and (4.5)

(4.12)

$$CV^2\{T|2n\} = \begin{cases} 1 , & n = 0 \\ 2(1-\rho^{n+1}) \dfrac{[(1-\rho)(2+\rho)^2(1+\rho^n)+\rho(1+\rho)^2(1-\rho^{n-1})]}{[(1-\rho)(2+\rho)(1+\rho^n) + \rho(1+\rho)(1-\rho^{n-1})]^2} - 1 , & n = 1,2, \ldots \end{cases}$$

and

$$(4.13) \quad CV^2\{T|2n+1\} = 2(1 - \rho^{n+1}) \frac{[(1-\rho)\rho^n(2+\rho)^2 + (1-\rho^n)(1+\rho)^2]}{[(1-\rho)\rho^n(2+\rho) + (1-\rho^n)(1+\rho)]^2} - 1 , \quad n = 0,1, \ldots .$$

Table 3.6 shows selected values of (4.12) and (4.13) as functions of ρ and n. For all values of ρ and n, $\{T|2n\}$ behaves very much like an exponential random variable (which has a coefficient of variation of 1). However, $\{T|2n + 1\}$ has a relatively high coefficient of variation for small $n > 0$.

Unconditioning on n in (4.4) and (4.5) yields

$$(4.14) \qquad E\{e^{-sT}\} = (1-\rho)^2 \frac{\lambda}{\lambda+s} + 2\rho(1-\rho) \frac{\lambda+\mu}{\lambda+\mu+s} + \rho^2 \frac{\lambda+2\mu}{\lambda+2\mu+s}$$

which has squared coefficient of variation

$$(4.15) \qquad CV^2\{T\} = \frac{5\rho^4 + 2\rho^3 - 19\rho^2 + 12\rho + 4}{(\rho^2 - \rho + 2)^2}$$

Table 3.6 contains values of $CV^2\{T\}$, as a function of ρ, from which one can argue that T is distributed "almost" exponentially for large values of ρ. It is interesting to note, as shown earlier, that the almost regenerative method is most powerful for values of $\rho > (1/2)^{2/3}$.

Selected Values for $CV^2\{T|2n\}$, $CV^2\{T|2n+1\}$, and $CV^2\{T\}$

Table 3.6

n \ ρ	$CV^2\{T\|2n\}$ .1	.3	.5	.7	.8	.9	$CV^2\{T\|2n+1\}$ .1	.3	.5	.7	.8	.9
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	1.000	1.000	1.000	1.000	1.000	1.000	1.117	1.152	1.132	1.109	1.098	1.088
2	1.041	1.078	1.083	1.077	1.072	1.067	1.015	1.065	1.091	1.094	1.091	1.087
3	1.045	1.097	1.109	1.100	1.093	1.087	1.001	1.021	1.051	1.070	1.074	1.075
4	1.045	1.103	1.118	1.107	1.099	1.091	1.000	1.007	1.027	1.050	1.058	1.063
5	1.045	1.104	1.122	1.109	1.097	1.089	1.000	1.002	1.014	1.035	1.045	1.054
6	1.045	1.105	1.123	1.109	1.097	1.086	1.000	1.001	1.007	1.024	1.036	1.046
7	1.045	1.105	1.124	1.108	1.094	1.082	1.000	1.000	1.003	1.017	1.028	1.039
8	1.045	1.105	1.125	1.107	1.092	1.079	1.000	1.000	1.002	1.012	1.022	1.034
9	1.045	1.105	1.125	1.107	1.090	1.075	1.000	1.000	1.001	1.008	1.017	1.029
10	1.045	1.105	1.125	1.107	1.088	1.072	1.000	1.000	1.000	1.006	1.014	1.025
∞	1.045	1.105	1.125	1.107	1.080	1.045	1.000	1.000	1.000	1.000	1.000	1.000

ρ	.0	.1	.2	.3	.4	.411	.5	.6	.7	.8	.9	1.0
$CV^2\{T\}$	1.0	1.374	1.673	1.868	1.942	1.943	1.898	1.756	1.553	1.333	1.137	1.0

CHAPTER 4

USING THE ALMOST REGENERATIVE METHOD TO OBTAIN ORDER OF MAGNITUDE IMPROVEMENTS IN MSE EFFICIENCY

A simple queueing network model is presented in this chapter for which the ARM can provide a 60 percent improvement in MSE efficiency over the RM, and a 50 percent improvement over the FTM, for estimating the expected total delay in queue. Equally important, the ARM generated data provides better estimates of estimator variance than the RM. These emperical findings have strong implications for the potential power of the ARM when applied to general queueing networks, since it appears that the relative efficiency of the ARM increases with system complexity.

In addition, comparisons of the simple ratio estimator with the jackknife ratio estimator and the simple variance estimator with the jackknife variance estimator show that the jackknife estimators are preferable when making estimates from ARM generated data.

4.1 The M/M/c Tandem Queue Network

Let p_n be the steady state probability that there are n customers in a single M/M/c queue, where arrivals occur at Poisson rate λ and the service time at each channel is exponentially distributed at rate μ. If $\rho \equiv \lambda/c\mu < 1$, it is well-known that

(1.1) $$p_n = \begin{cases} p_o(\lambda/\mu)^n/n! & , n = 0,1, \ldots, c-1 \\ p_c(\lambda/c\mu)^{n-c} & , n = c, c+1, \ldots \end{cases}$$

where

(1.2) $$p_o = \left[\sum_{n=0}^{c} \frac{(\lambda/\mu)^n}{n!} + \frac{\mu p_o(\lambda/\mu)^c}{(c-1)!(c\mu - \lambda)} \right]^{-1}$$

It is also well-known that the expected delay in queue of a typical customer, given delay is positive, is distributed exponentially with parameter $(c\mu - \lambda)$. Hence

$$(1.3) \qquad E\{D\} = \frac{1}{c\mu - \lambda} \sum_{n \geq c} p_n = \frac{\mu p_o (\lambda/\mu)^c}{(c-1)!(c\mu - \lambda)^2} .$$

Now consider a network consisting of two M/M/c queues in tandem, so that departures from the first system are arrivals at the second system. Again, assume that arrivals to the first system are Poisson at rate λ and that the service time at each channel, for each queue, is exponentially distributed at rate μ . Let p_{ij} represent the steady state probability that there are i customers in the first system and j in the second system; i.e., the network is in state (i,j). From Reich (1963), the departure stream from the first system is Poisson at rate λ , and $p_{ij} = p_i p_j$. Also, the expected total delay in system is $d \equiv E\{D_T\} = 2E\{D\}$.

One possible choice of regenerative state transition for this tandem network occurs when an arrival to the first system finds the network in state (i,j), for fixed i and j . Using the same arguments as in Chapter 3, it is well-known that the mean time between such regeneration points is

$$(1.4) \qquad E\{T_{ij}\} = \frac{1}{\lambda p_{ij}} .$$

This expression is minimized for the state (i_o,j_o) that satisfies $i_o = j_o$ and $i_o < \lambda/\mu < i_o + 1$. As in Chapter 3, the regenerative method with regenerative state (i_o,j_o) will be referred to as the RM(i_o,j_o) .

Consider the set of states

(1.5) $A(n_1,n_2) = \{(i,j) : i + j \in [n_1,n_2] \ , \ i,j = 0,1, \ldots\}$

for n_1, $n_2 = 0,1, \ldots$. An almost regeneration point will be said to occur whenever an arrival to the first system finds any one of the states $(i,j) \in A(n_1,n_2)$, for fixed n_1,n_2 . The probability of such an event is

$$(1.6) \qquad p'_{n_1,n_2} = \sum_{A(n_1,n_2)} p_{ij} = \sum_{n=n_1}^{n_2} \sum_{i=0}^{n} p_{i,n-i} \ ,$$

and the expected time between these events is

$$(1.7) \qquad E\{T'_{n_1,n_2}\} = \frac{1}{\lambda p'_{n_1,n_2}} \ .$$

The almost regenerative method with almost regenerative state $A(n_1,n_2)$ will be referred to as the ARM[$A(n_1,n_2)$] .

4.2 Simulation of the M/M/2 Tandem Network Using ARM [$A(n_1,n_2)$]

As shown emperically in Chapter 3, the ARM can generate more observations, which act as if they were iid, than the RM for the same simulation run length. As a result, the ARM can be more efficient, in terms of the MSE of the estimator, than the RM or the FTM when estimating the expected delay, d , for simple queueing networks. One might expect that the MSE efficiency of the ARM could be further improved by increasing the number of states in $A(n_1,n_2)$. But increasing the size of $A(n_1,n_2)$ involves a balance between increased observation frequency and more pronounced serial correlation in the observations. In addition, computer CPU time increases as $A(n_1,n_2)$ becomes larger since more observations are processed for the same run length. In this section the $ARM[A(n_1,n_2)]$ will be compared with the $RM(i,j)$ and the $FTM(\Delta=E\{T'_{n_1,n_2}\})$ for various values of n_1 and n_2.

For $c = 2$, $\lambda = 1$, and $\mu = .625$ (so that $\rho = .8$), (1.4) is minimized when $i_o = j_o = 1$. From (1.1) - (1.7) it is straightforward to compute

$$
(2.1)\qquad
\begin{aligned}
&d = 5.689, && E\{T_{oo}\} = 81.0\ , && E\{T_{11}\} = 31.641,\\
&E\{T'_{44}\} = 12.361, && E\{T'_{45}\} = 6.180, && E\{T'_{46}\} = 4.175,\\
&E\{T'_{36}\} = 3.171, && E\{T'_{37}\} = 2.578, && E\{T'_{38}\} = 2.202,\\
&E\{T'_{28}\} = 1.933, && E\{T'_{29}\} = 1.733.
\end{aligned}
$$

One hundred independent simulation replications, [illegible] m observations each, were made for each data collection method. The parameter m was set to ensure the same expected number of arrivals to the system (123,000) for each method. The interarrival and service times (for each system) were generated with seperate random number seeds, which were the same for each method, using the CDC 6400 RANF pseudo-random number generator. The estimates of d were based on the cumulative delay process, $\{A_j\}_1^m$, and the cumulative arrival process, $\{N_j\}_1^m$. The parameters of interest were computed using ralations (2.4) - (2.8) of Chapter 3, section 3.2. Each replication began with identical initial conditions (i.e., no customers in queue and all servers idle).

The simulation results are summarized in Tables 4.1, 4.2, 4.3, and 4.4. Referring to Tables 4.1 and 4.2, the simple ratio estimates of

$MSE(\hat{d}_j)$ for the ARM are up to 60 percent less than the RM simple ratio estimates of $MSE(\hat{d}_j)$ and up to 53 percent less than the corresponding FTM estimates. This improvement reaches a peak for the ARM[A(3,8)]. The ARM jackknife estimates of $MSE(\hat{d}_j)$ are nearly 45 percent less than the RM jackknife estimates of $MSE(\hat{d}_j)$ and up to 43 percent less than the corresponding FTM estimates. This improvement also reaches a maximum for the ARM[A(3,8)]. The standard deviation of the ARM estimates of $MSE(\hat{d}_j)$ is generally less than the standard deviation of the RM and FTM estimates, thus increasing confidence in these results.

As observed for the M/M/1 tandem network of Chapter 3, the ARM and RM jackknife estimates of d for the M/M/2 network have smaller biases than the corresponding simple ratio estimates of d . However, the FTM jackknife estimates of d have about the same bias as the corresponding simple ratio estimates. This result is directly attributable to the superior iid qualities of the ARM generated data when compared to the FTM data.

The internal and external estimates of $V(\hat{d}_j)$ are shown in Table 4.3 for the RM and ARM generated data. The variance estimates using the FTM data are not shown since the serial correlation between the FTM observations degrades the estimator accuracy and increases the CPU time necessary to form the estimator (since lagged autocorrelations must be included in the computation).

From Table 4.3, the internal ARM estimates of $V(\hat{d}_j)$ are less biased and more stable than the RM estimates, where the measure of stability is the estimated standard deviation of the estimator. It is interesting to observe that the RM(0,0) internal variance statistic

M/M/2 Tandem Queue: Comparison of MSE Estimates

Table 4.1

($\lambda = 1$, $\mu = .625$, $d = 5.689$, 100 independent replications of m observations each)

Method	CPU Time	Arrivals	m	Simple Ratio Estimator					Jackknife Ratio Estimator				
	(sec)	(10^3)		$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\hat{MSE}(\hat{d}_j)$	s_{MSE}	$\hat{d}$	s_d	$\hat{\sigma}_d^2(j)$	$\hat{MSE}(\hat{d}_j)$	s_{MSE}
RM(0,0)	225.7	131.821	16	5.409	.197	3.896	3.974	.412	5.719	.212	4.514	4.517	.716
RM(1,1)	218.6	126.166	40	5.434	.151	2.239	2.354	.374	5.748	.188	3.516	3.519	1.094
ARM[A(4,4)]	219.3	123.270	100	5.360	.140	1.968	2.076	.304	5.602	.159	2.523	2.531	.459
ARM[A(4,5)]	228.6	126.523	200	5.432	.135	1.815	1.881	.264	5.642	.149	2.211	2.213	.355
ARM[A(4,6)]	229.0	124.169	296	5.424	.132	1.751	1.821	.299	5.637	.149	2.213	2.216	.433
ARM[A(3,6)]	232.3	123.940	389	5.471	.135	1.833	1.881	.285	5.691	.153	2.334	2.334	.456
ARM[A(3,7)]	236.0	123.763	479	5.429	.131	1.717	1.785	.276	5.630	.148	2.185	2.188	.409
ARM[A(3,8)]	239.7	123.880	561	5.507	.125	1.554	1.587	.239	5.697	.139	1.937	1.937	.324
ARM[A(2,8)]	238.9	121.532	639	5.396	.142	2.020	2.106	.287	5.584	.157	2.473	2.484	.395
ARM[A(2,9)]	245.9	123.681	713	5.447	.133	1.775	1.834	.240	5.612	.147	2.160	2.166	.322
FTM(Δ=12.361)	215.7	123.750	100	5.530	.170	2.889	2.914	.289	5.529	.170	2.887	2.912	.288
FTM(Δ=6.180)	219.5	123.744	200	5.520	.171	2.937	2.965	.300	5.517	.171	2.935	2.964	.299
FTM(Δ=4.175)	232.2	123.721	296	5.584	.172	2.968	2.979	.319	5.580	.172	2.965	2.977	.318
FTM(Δ=3.171)	226.4	123.472	339	5.496	.183	3.331	3.369	.436	5.492	.182	3.328	3.367	.435
FTM(Δ=2.578)	230.0	123.625	479	5.524	.172	2.945	2.972	.355	5.520	.172	2.943	2.971	.354
FTM(Δ=2.202)	233.6	123.663	561	5.543	.184	3.391	3.413	.471	5.540	.184	3.388	3.411	.469
FTM(Δ=1.933)	236.4	123.654	639	5.536	.181	3.260	3.283	.407	5.532	.180	3.257	3.281	.406
FTM(Δ=1.733)	239.4	123.702	713	5.556	.174	3.031	3.049	.345	5.552	.174	3.028	3.047	.344

M/M/2 Tandem Network: Comparison of MSE Efficiencies[†]

Table 4.2

	RM(0,0)	RM(1,1)	Values of Δ for the FTM							
			12.361	6.180	4.175	3.171	2.578	2.202	1.933	1.733
ARM[A(4,4)]	49.5/44.1[†]	14.0/28.2	31.9/17.0							
ARM[A(4,5)]	53.4/51.0	20.7/37.1		36.6/25.3						
ARM[A(4,6)]	55.1/51.0	23.5/37.1			38.6/25.2					
ARM[A(3,6)]	53.0/48.3	19.9/33.6				44.2/30.7				
ARM[A(3,7)]	55.9/51.6	25.0/37.9					47.0/35.0			
ARM[A(3,8)]	60.1/57.1	32.1/44.9						53.5/43/2		
ARM[A(2,8)]	48.2/45.2	11.8/29.7							38.3/27.2	
ARM[A(2,9)]	54.4/52.1	22.5/38.6								39.8/28.9

[†] Entry a_{ij} has two components: the first component refers to the percent improvement in MSE efficiency of method i over method j when using the simple ratio estimator and the second entry refers to a similar quantity when using the jackknife ratio estimator. Each entry is computed by

$$a_{ij} = \frac{\hat{MSE}_j - \hat{MSE}_i}{\hat{MSE}_j} \times 100\%$$

For example, the ARM[A(3,8)] jackknife estimate of $MSE(\hat{d})$ is 57.1 percent less than the RM(0,0) jackknife estimate of $MSE(\hat{d})$.

M/M/2 Tandem Queue: Comparison of Variance Estimators

Table 4.3

(λ = 1, μ = .625, 100 independent replications of m observations each)

Method	m	Simple Variance Estimator				Jackknife Variance Estimator			
		External Estimate		Internal Estimate		External Estimate		Internal Estimate	
		$\hat{\sigma}^2_d(j)$	$\hat{s}_d$	$\tilde{\sigma}^2_d(j)$	$\tilde{s}_d$	$\hat{\sigma}^2_d(j)$	$\hat{s}_d$	$\tilde{\sigma}^2_d(j)$	$\tilde{s}_d$
RM(0,0)	16	3.896	.412	1.569	.215	4.514	.716	2.879	.505
RM(1,1)	40	2.289	.374	1.388	.186	3.516	1.094	2.370	.446
ARM[A(4,4)]	100	1.968	.304	1.311	.148	2.523	.459	2.027	.340
ARM[A(4,5)]	200	1.815	.264	1.352	.144	2.211	.355	1.912	.265
ARM[A(4,6)]	296	1.751	.299	1.344	.138	2.213	.433	1.937	.280
ARM[A(3,6)]	389	1.833	.285	1.386	.158	2.334	.456	2.026	.365
ARM[A(3,7)]	479	1.717	.276	1.324	.143	2.185	.409	1.904	.294
ARM[A(3,8)]	561	1.554	.239	1.437	.164	1.937	.324	1.980	.276
ARM[A(2,8)]	639	2.020	.287	1.414	.168	2.473	.395	1.969	.297
ARM[A(2,9)]	713	1.775	.240	1.292	.158	2.160	.322	1.758	.264

M/M/2 Tandem Queue Renewal Test Statistics for ARM[A(n_1,n_2)]

Table 4.4

(λ = 1, μ = .625, 1000 observations of each process)

n_1	n_2	Lag 1 Autocorrelation				Kolmogorov-Smirnov				Anderson-Darling				Wald-Wolfowitz			
		$\{A_i\}$	$\{Y_i\}$	$\{N_i\}$	$\{Z_i\}$	$\{A_i\}$	$\{Y_i\}$	$\{N_i\}$	$\{Z_i\}$	$\{A_i\}$	$\{Y_i\}$	$\{N_i\}$	$\{Z_i\}$	$\{A_i\}$	$\{Y_i\}$	$\{N_i\}$	$\{Z_i\}$
4	4	.00	.02	.02	.00	.29	.45	.44	.33	.01	.03	.03	.02	1.19	1.48	1.63	1.60
4	5	.01	.03	.01	.00	.27	.38	.47	.51	.01	.05	.08	.03	4.53	7.29	2.01	1.94
4	6	.06	.04	.04	.07	.74	.56	.58	.80	.19	.08	.11	.22	.01	3.08	2.44	2.09
3	6	.05	.03	.04	.07	.68	.56	.57	.73	.16	.06	.08	.21	6.01	1.06	10.7	4.11
3	7	.01	.02	.02	.00	1.11	1.10	1.17	.94	.27	.25	.28	.20	5.42	1.21	10.8	5.40
3	8	.00	.02	.01	.00	.47	.60	.55	.35	.04	.05	.05	.03	6.73	1.23	13.7	9.48
2	8	.00	.01	.00	.00	.41	.52	.51	.32	.50	.06	.06	.03	11.6	1.11	15.5	11.4
2	9	.10	.09	.08	.10	1.12	1.18	.97	1.12	.49	.46	.32	.52	11.9	.26	19.8	12.4
		Upper 5% Significance Point															
		.062				1.356				.461				1.960			
		Upper 1% Significance Point															
		.081				1.628				.743				2.575			

$\{A_i\}$ - cumulative delay process $\{N_i\}$ - cumulative arrival process

$\{Y_i\}$ - cycle length process $\{Z_i\}$ - jackknife transformation of $\{A_i\}$ and $\{N_i\}$

underestimates significantly, but the corresponding standard deviation estimate does not provide an indication of this fact; i.e., the RM(0,0) internal variance estimate has a large bias while its standard deviation implies that the variance estimate is relatively good. This problem is also characteristic of the ARM variance estimates, but the severity of the problem is reduced. The importance of this improvement cannot be overly emphasized, since in most simulations the variance of the estimator of interest is usually an internal estimate based on a single simulation run.

When comparing the simple variance estimator and the jackknife variance estimator for the different data collection methods (see Table 4.3), the internal jackknife variance estimator is consistently within one standard deviation of the external estimate. This is not always the case for the simple variance estimator.

The renewal test statistics for the ARM[$A(n_1,n_2)$] are shown in Table 4.4. The lack of any discernable periodicities in the associated spectrums (not shown), at the 95 percent level, supports the conclusion from Table 4.4 that the jackknifed transformations of the ARM[A(4,4)] and the ARM[A(4,5)] generated observations act as if they were iid. The fact that data generated by larger $A(n_1,n_2)$ sets do not pass the renewal tests does not seem to degrade the corresponding estimates of d and $V(\hat{d}_J)$. It seems clear, however, that the jackknife transformation improves the iid properties of the data.

There are two reasons why the ARM estimates of d are more efficient than the FTM estimates of d. First, the ARM generated data has better iid properties than the FTM data, and, secondly, the jackknife transformation of the data further improves its iid qualities. Also, it appears that the jackknife estimators do not reduce estimator bias when the data has significant serial correlation.

Increasing the size of $A(n_1,n_2)$ does not appear to cause severe degradation of these results when n_1 and n_2 are within a "reasonable" range of each other. However, as the separation between n_1 and n_2 increases, the number of observations, for a fixed run length, also increases, thereby increasing the CPU time to process the additional observations.

4.3 Comments on the Implied Power of the Almost Regenerative Method

When comparing the emperical results of this chapter with those of Chapter 3, it appears that the MSE efficiency of the ARM estimators, relative to the RM and the FTM, increases with system complexity. In addition, the internal estimates of estimator variance are improved when the estimates are based on the ARM generated data rather than the RM generated data. These observations have far-reaching consequences, since it is not difficult to imagine queueing networks for which the ARM could provide an order of magnitude improvement in MSE efficiency.

These properties of the ARM are due primarily to the fact that the ARM can generate observations which act as if they were iid, unlike the FTM, and that the ARM can generate more observations, for a fixed run length, than the RM. To illustrate this second point, the theoretical expected time between some M/M/c tandem network regeneration and almost regeneration points are shown in Table 4.5 for selected values of λ , μ , and c .

Two M/M/c Queues in Tandem: Comparison of $E\{T\}$ for the ARM and the RM

Table 4.5

ρ	c	RM $E\{T_{oo}\}^{(1)}$	RM $E\{T_{i_o i_o}\}^{(2)}$	ARM $E\{T'_{n_o,n_o}\}^{(3)}$	ARM $E\{T'_{n_o,n_o+1}\}^{(4)}$
.8	2	81.0	31.6	12.4	6.2
.8	3	316.8	38.2	12.5	6.2
.8	4	1345.5	45.0	12.7	6.3
.8	5	5929.0	52.1	13.5	6.7
.9	2	361.0	111.4	25.6	13.0
.9	3	1612.0	121.3	25.9	13.0
.9	4	7892.5	130.5	26.0	13.0
.9	5	40671.6	139.3	26.2	13.1
.95	2	1521.0	421.3	53.0	26.5
.95	3	7237.7	438.8	53.1	26.6
.95	4	37996.4	454.3	53.1	26.6
.95	5	210786.8	468.5	53.2	26.7

(1) Expected time between arrivals who find state $(i,j) = (0,0)$.

(2) Expected time between arrivals who find state (i_o,j_o), where $i_o = j_o$ and $i_o < \lambda/\mu < i_o + 1$.

(3) Expected time between arrivals who find some state $(i,j) \in A(n_o,n_o)$, where n_o minimizes $E\{T'_{n_o,n_o}\}$.

(4) Expected time between arrivals who find some state $(i,j) \in A(n_o,n_o+1)$ where n_o minimizes $E\{T'_{n_o,n_o+1}\}$.

CHAPTER 5

THE ALMOST REGENERATIVE METHOD IN PERSPECTIVE

There are two related, but distinct, issues involved when conducting a simulation experiment. The *internal problem* is concerned with the conditions under which data is collected from the simulation, and the *external problem* is concerned with the use made of the data to form reliable estimates of the parameters of interest.

The regenerative method (RM) solves the internal problem for simulations of regenerative processes by generating observations which are iid. As a result, the external problem is eased since autocovariances need not be computed when estimating the variance of the estimators. However, the RM does have at least one serious failing. It cannot be applied to simulations of stochastic processes which are not regenerative, nor can it be applied effectively to regenerative processes which have excessively long expected times between regeneration times. A second, less serious, defect concerns the external problem of forming ratio estimators of the data. All known ratio estimators are biased, and the bias depends on the regenerative state chosen.

The alternative methods for collecting simulation data are the standard fixed time increment method (FTM), which makes no use of any prior knowledge of the process under consideration, and approximate regeneration methods. The approximate methods can be classified as either state space transformation methods, which distort the original process by inducing a regenerative approximation to the process of interest, and the almost regenerative methods.

The purpose of this dissertation has been to develop an almost regenerative method which alleviates the internal data collection problem created by the RM (i.e., possibly infrequent observations) and yet exploits the desireable property of the RM for producing iid observations.

When compared to the RM, the ARM has been shown emperically to have the potential for an order of magnitude reduction in the MSE of the total delay in queue estimator for simple queueing networks. In fact, this relative improvement increases with system complexity. Similar results were shown when comparing the ARM with the FTM. Moreover, the ARM variance estimates have a smaller bias and a smaller estimated standard deviation than the RM variance estimators. The programming considerations necessary to implement the ARM are no more complicated than those required to implement the RM or the FTM for most event-oriented simulations.

However, the ARM is not without fault. The choice of the almost regenerative pair of sets (U,V), which trigger data collection, remains largely an art, since an intelligent choice of (U,V) depends on the characteristics of the stochastic process being simulated, the parameters being measured, and the desired estimator accuracy. Increases in the sizes of U and V provide more observations for fixed run length, but only at the possible expense of variance estimator accuracy. That is, once the data loses its similarity to iid data, the autocorrelation between observations must be included in the variance estimates and this requires an additional pass through the data, and probably a degradation of the variance estimator.

As an external issue, the jackknife transformation seems to improve the "iid-like" properties of the data (see Table 4.3, Chapter 4), thus improving variance estimates. This fact needs to be balanced with the larger storage requirements of the jackknife estimator, when compared to the storage requirements of the other estimators.

APPENDIX I

TESTING FOR RENEWAL PROCESSES

The tests in this Appendix can be found in Cox and Lewis (1966).

Let $\{X_i\}_1^n$ be a random sample for a stationary random variable X with unknown continuous distribution $F_X(t)$, and let $F_o(t)$ be a completely specified distribution. The tests of this Appendix are designed to test the null hypothesis

$$H_o: \; F_X = F_o$$

against the two-sided alternative

$$H_1: \; F_X \neq F_o \; .$$

The first two tests which will be used later are distribution-free in the sense that the null hypothesis does not depend on the specific form of F_X and F_o . This property is a consequence of the probability integral transformation which states that if the random variable X has the distribution F , then $U = F(X)$ is uniformly distributed over $[0,1]$. Thus, the probability integral transformation reduces the problem to one of testing whether the observations $\{F_o(X_i)\}_1^n$ are uniformly distributed.

The first test is the well-known Komogorov-Smirnov test. Define the statistic

$$C_n = \sup_{-\infty<t<\infty} |F_n(t) - F_o(t)| = \max\,(D_n^+, D_n^-)$$

where

$$D_n^+ = \sup_{-\infty<t<\infty} \{F_n(t) - F_o(t)\} = \max_{i=1,\ldots,n} \left\{ \frac{i}{n} - F_o(X_{(i)}) \right\}$$

and

$$D_n^- = \sup_{-\infty<t<\infty} \{F_o(t) - F_n(t)\} = \max_{i=1,\dots,n-1} \left\{ F_o(X_{(i)}) - \frac{i-1}{n} \right\} .$$

In these expressions, $F_n(t)$ is the emperical distribution obtained by

$$F_n(t) = \frac{\text{number of } X_i\text{'s} \leq t}{n}$$

and $X_{(i)}$ is the ith order statistic for the sequence $\{X_i\}_1^n$. If $d_{n,\alpha}$ is a value such that

$$P\{D_n > d_{n,\alpha}\} = \alpha ,$$

then the null hypothesis is rejected at level α if the observed value of D_n is greater than $d_{n,\alpha}$. It is well-known [cf. Darling (1957)] that D_n^+ and D_n^- have the asymptotic distribution

(A.1)
$$\lim_{n\to\infty} P\{D_n^+\sqrt{n} > d\} = e^{-2d^2}$$

for $d \in [0,\infty)$; (A.1) is generally valid for $n > 20$.

The second type of test is based on the Cramér-von Mises statistic

$$W_n^2 = \int_{-\infty}^{\infty} [F_n(t) - F_o(t)]^2 dF_o(t)$$

which has the computational form

$$W_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left[F_o(X_{(i)}) - \frac{2i-1}{2n} \right]^2 .$$

Again the null hypothesis is rejected at level α if the observed value of ω_n^2 is greater than $w_{n,\alpha}$ where

$$P\{\omega_n^2 > w_{n,\alpha}\} = \alpha .$$

Both of these tests are measures of the distance between the two distributions. It is interesting to note that these two tests are distribution free, consistant, and do not require arbitrary groupings of the data. Other better known tests, for example the chi-squared goodness of fit test, do not have all these properties. Many other less powerful tests for uniformity are in Knuth (1969).

A.1 Serial Correlation Coefficient Tests

There are essentially two types of tests for renewal processes: those based on the serial correlation coefficients, and those based on the spectrum of intervals. The serial correlation coefficient tests will be described first.

As before, let $\{X_i\}_1^n$ be a sequence of observations $X_i \overset{\mathcal{D}}{\to} X$ from some stationary stochastic process $\{X_i\}_0^\infty$ where $X_i \overset{\infty}{\to} X$. The usual unbiased estimator for the serial correlation coefficient of lag j , r_j , is given by

$$\hat{r}_j = \frac{R_j}{R_o} \tag{A.3}$$

where

$$R_j = \frac{1}{n} \sum_{i=1}^{n-j} [X_i - \bar{X}(n)][X_{i+j} - \bar{X}(n)] , \quad j = 0,1, \ldots, n-1$$

is an estimate of the covariance at lag j ,

$$C_j = E\{[X_i - EX][X_{i+j} - EX]\} ,$$

for all $i,j = 0,1, \ldots$. Since

$$\frac{\hat{r}_1 - \hat{r}}{\sqrt{n - 1}} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{w.p.1} .$$

[Bartlett (1955)], a test at the α level that $r_1 = 0$ is to reject independence if

$$|\hat{r}_1| > \frac{Q_{\alpha/2}}{\sqrt{n - 1}}$$

where $Q_{\alpha/2}$ is the upper $\alpha/2$ point of the unit Normal distribution. Although this test is simple, it has a number of shortcomings. First, it is difficult to extend this test to sample serial correlation coefficients of lag greater than one. Secondly, the small sample distribution of $\hat{r}_1$ under the independence hypothesis is not known; clearly, it depends on the distribution of times between events.

A test designed to alleviate this second difficulty is the Wald-Wolfowitz rank product-moment test for lag 1 [Wald and Wolfowitz (1943)]. The central idea is to replace the X_i's in (A.3) by their ranks, where the rank of X_i is the argument of its corresponding order statistic. Evidently, this procedure has the advantage of controlling the influence of spurious observations. Wald and Wolfowitz showed that the resulting rank product-moment statistic, $\tilde{r}_1$, is asymptotically normally distributed with mean and variance given by

$$E(\tilde{r}_1) = \frac{1}{12}(n - 1)(n + 1)(3n + 2) ,$$

and

$$V(\tilde{r}_1) = \frac{5n^6 + 16n^5 - 14n^4 - 80n^3 - 35n^2 + 64n + 44}{720(n - 1)}$$

A.2 Tests Based on the Spectrum of Intervals

More complicated, but more powerful, tests for renewal processes are based in the spectrum of intervals. Consider the stationary sequence $\{X_i\}_1^\infty$ which has a *spectral density* defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} r_j \cos(j\omega) \ , \ -\pi \le \omega \le \pi \tag{A.4}$$

and a *power spectrum* defined by

$$\sigma^2 f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} C_j \cos(j\omega) \ , \ -\pi \le \omega \le \pi \tag{A.5}$$

where σ^2 is the variance of the r.v. X . If the $\{X_i\}$ are iid, then, from (A.4),

$$f(\omega) = \frac{1}{2\pi} \ , \ -\pi \le \omega \le \pi \tag{A.6}$$

so that a test for renewal process is to compute the spectral density function and check to see if it is of the form (A.6). Define $\omega_p = 2\pi p/n$, where $p = 1,2, \ldots, [\frac{1}{2} n]$. It can be shown that the *periodogram*, defined by

$$I_n(\omega_p) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} X_j e^{-i\omega_p j} \right| \ , \ \omega_p \ne 0 \ (\mathrm{mod}\ 2\pi)$$

where $i = \sqrt{-1}$, is an estimator of the power spectrum [cf. Brillinger (1975)]. Bartlett (1955) and Hannan (1960) showed, among other things, that when the X_i are iid the periodogram asymptotically has an exponential distribution with mean

$$\text{(A.8)} \qquad E\{I_n(\omega_p)\} = \frac{\sigma^2}{2\pi} = \sigma^2 f(\omega_p) \ , \ \omega_p \neq 0 \pmod{2\pi} \ .$$

Cox and Lewis (1966) suggest testing the quantities

$$\text{(A.9)} \qquad U_{(i)} = \frac{\sum_{p=1}^{i} I_n(\omega_p)/f(\omega_p)}{\sum_{p=1}^{n} I_n(\omega_p)/f(\omega_p)}$$

as samples from a uniform distribution to test the hypothesis that $I_n(\omega_p)$ is exponentially distributed. The Kolmogorov-Smirov and Cramér-von Mises statistics are particularly well-suited for this kind of test.

A.3 A Test of the Tests

The following linear model was simulated, for different values of a and ℓ,

$$\text{(A.10)} \qquad X_t = aX_{t-\ell} + \varepsilon_t \ , \ \varepsilon_t \sim N(0,1) \ , \ t = 0, \pm 1, \pm 2, \ldots,$$

to test the sensitivity of the tests in this section for accepting the renewal hypothesis. The results of these tests are summarized in Table A.1. In all those cases where the renewal hypothesis was accepted on the basis of the tests, strong periodicies appeared in the estimated spectrum of $\{X_t\}$ at the 95 per cent significance level. Thus, the renewal hypothesis was rejected in each case.

Let $\bar{X}(n) = \sum X_i/n$ be the usual estimator of the mean of the process defined by (A.10). Since this process is stationary, one may easily show that the variance of $\bar{X}(n)$ is given by [cf. Fishman (1973b), pg. 280)

TESTS OF THE RENEWAL PROCESS TEST STATISTICS

TABLE A.1

(Results are based on 1000 observations of each model)

Model	%[†]	Lag 1 Autocorrelation	Kolmogorov-Smirnov	Anderson-Darling	Wald-Wolfowitz	Passes Tests*
$X_t = .1X_{t-1} + \varepsilon_t$	16.7	.054	.756	.162	2.031	no
$X_t = .2X_{t-1} + \varepsilon_t$	28.6	.155	1.778	1.230	5.052	no
$X_t = .3X_{t-1} + \varepsilon_t$	37.5	.256	2.818	3.390	8.040	no
$X_t = .1X_{t-3} + \varepsilon_t$	16.6	-.044	.878	.153	-1.030	yes
$X_t = .2X_{t-3} + \varepsilon_t$	28.5	-.423	1.188	.318	-1.096	no
$X_t = .3X_{t-3} + \varepsilon_t$	37.4	-.041	1.504	.616	-1.235	no
$X_t = .1X_{t-5} + \varepsilon_t$	16.6	-.040	.612	.105	-.889	yes
$X_t = .2X_{t-5} + \varepsilon_t$	28.5	-.034	.754	.136	-.730	yes
$X_t = .3X_{t-5} + \varepsilon_t$	37.4	-.275	.930	.213	-.494	no
Upper 5% significance point		.062	1.358	.461	1.960	
Upper 1% significance point		.081	1.628	.743	2.575	

[†] Percent of variance lost by assuming data is i.i.d.

*Indicates whether or not the model passed the tests at the upper 5% level.

(A.11) $$V\{\bar{X}(n)\} = \frac{R}{n}\left[1 + 2a\left(1 - \frac{\ell}{n}\right)\right]$$

where

(A.12) $$R = E[(X_t - EX_t)^2] \ .$$

If one were to assume that no serial dependence existed in n successive observations of the process defined by (A.10), then the percent of the true variance given by (A.11) that would be lost can be expressed as

(A.13) $$\%\text{ variance lost} = 1 - \frac{1}{1 + 2a(1 - \ell/n)}$$

$$= \frac{2a(1 - \ell/n)}{1 + 2a(1 - \ell/n)} \ .$$

Values for (A.13) are also shown in Table A.1.

REFERENCES

[1] Arvesen, J. N., "Jackknifing U-Statistics," Annals of Mathematical Statistics, Vol. 40, pp. 2076-100, (1969).

[2] Bartlett, M. S., AN INTRODUCTION TO STOCHASTIC PROCESSES, Cambridge University Press, Cambridge, (1955).

[3] Beale, E. M. L., "Some Uses of Computers in Operational Research," Industrielle Organisation, Vol. 31, pp. 27-28, (1962).

[4] Billingsley, P., CONVERGENCE OF PROBABILITY MEASURES, John Wiley & Sons, New York, (1968).

[5] Brillinger, D. R., TIME SERIES: DATA ANALYSIS AND THEORY, Holt, Rinehart, and Winston, New York, (1975).

[6] Brown, M. and S. M. Ross, "Asymptotic Properties of Cumulative Processes," Technical Report 120, Dept. of Operations Research, Cornell University, Ithaca, New York, (1970).

[7] Chung, K. L., A COURSE IN PROBABILITY THEORY, Harcourt, Brace, and World, New York, (1968).

[8] Conway, R. W., "Some Tactical Problems in Digital Simulation," Management Science, Vol. 10, pp. 47-61, (1963).

[9] Conway, R. W., B. M. Johnson, and W. L. Maxwell, "Some Problems in Digital Systems Simulation," Management Science, Vol. 6, pp. 72-110, (1959).

[10] Cox, D. R. and P. A. W. Lewis, THE STATISTICAL ANALYSIS OF SERIES OF EVENTS, Methuen, London, (1966).

[11] Crane, M. A. and D. L. Iglehart, "A New Approach to Simulating Stable Stochastic Systems," in PROCEEDINGS OF THE 1973 WINTER SIMULATION CONFERENCE, San Francisco, pp. 264-272, (1973).

[12] Crane, M. A. and D. L. Iglehart, "Statistical Analysis of Discrete Event Simulations," PROCEEDINGS OF THE 1974 WINTER SIMULATION CONFERENCE, Washington, D. C., pp. 513-521, (1974a).

[13] Crane, M. A. and D. L. Iglehart, "Simulating Stable Stochastic Systems, I: General Multiserver Queues," Journal of the Association for Computing Machinery, Vol. 21, pp. 103-113, (1974b).

[14] Crane, M. A. and D. L. Iglehart, "Simulating Stable Stochastic Systems, II: Markov Chains," Journal of the Association for Computing Machinery, Vol. 21, pp. 114-123, (1974c).

[15] Crane, M. A. and D. L. Iglehart, "Simulating Stable Stochastic Systems, III: Regenerative Processes and Discrete-Event Simulations," Operations Research, Vol. 23, pp. 33-45, (1975a).

[16] Crane, M. A. and D. L. Iglehart, "Simulating Stable Stochastic Systems, IV: Approximation Techniques," Management Science, Vol. 21, pp. 1215-1224, (1975b).

[17] Durbin, J., "A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios," Biometrika, Vol. 46, pp. 477-480, (1959).

[18] Darling, D. A., "The Kolmogorov-Smirnov, Cramèr-von Mises Tests," Annals of Mathematical Statistics, Vol. 28, pp. 823-838, (1957).

[19] Fishman, G. S., "Statistical Analysis for Queueing Simulations," Management Science, Vol. 20, pp. 363-369, (1973a).

[20] Fishman, G. S., CONCEPTS AND METHODS IN DISCRETE EVENT DIGITAL SIMULATION, John Wiley & Sons, New York, (1973b).

[21] Fishman, G. S., "Simulation Data Analysis Using Continuous Time Markov Processes," Technical Report 74-1, Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, (1974a).

[22] Fishman, G. S., "Estimation in Multiserver Queueing Simulations," Operations Research, Vol. 20, pp. 72-78, (1974b).

[23] Fishman, G. S., "Achieving Specified Accuracy in Simulation Output Analysis," Technical Report 74-4, Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, (1974c).

[24] Fishman, G. S. and P. J. Kiviat, "The Analysis of Simulation Generated Time Series," Management Science, Vol. 13, pp. 525-557. (1967).

[25] Hannan, E. J., TIME SERIES ANALYSIS, Methuen, London, (1960).

[26] Hutchinson, D. W., "A New Uniform Pseudorandom Number Generator," Communications of the Association for Computing Machinery, Vol. 9, pp. 432-40, (1966).

[27] Hutchison, M. C., "A Monte Carlo Comparison of Some Ratio Estimators," Biometrika, Vol. 58, pp. 313-321, (1971).

[28] Iglehart, D. L., "Simulating Stable Stochastic Systems, V: Comparison of Ratio Estimators," Technical Report 86-14, Control Analysis Corp., Palo Alto, Calif., (1974).

[29] Iglehart, D. L., Simulating Stable Stochastic Systems, VI: Quantile Estimation," Technical Report 86-15, Control Analysis Corp., Palo Alto, Calif., (1975).

[30] Kabek, I. W., "Stopping Rules for Queueing Systems," Operations Research, Vol. 16, pp. 431-437, (1968).

[31] Knuth, D. E., THE ART OF COMPUTER PROGRAMMING, Vol. II, Addison-Wesley, Reading, Mass., (1969).

[32] Law, A. M., "Efficient Estimators for Simulated Queueing Systems," ORC Report 74-7, University of California, Berkeley, (1974).

[33] Miller, R. G., "The Jackknife - A review," Biometrika, Vol. 61, pp. 1-15, (1974).

[34] Quenouille, M. H., "Approximate Tests of Correlation in Time Series," Journal of the Royal Statistical Society, Series B, Vol. 11, pp. 68-84, (1949).

[35] Quenouille, M. H., "Notes on Bias in Estimation," Biometrika, Vol. 43, pp. 353-360, (1956).

[36] Rao, J. N. K. and J. Webster, "On Two Methods of Bias Reduction in the Estimation of Ratios," Biometrika, Vol. 53, pp. 571-577, (1966).

[37] Reich, E., "Waiting Times When Queues are in Tandem," Annals of Mathematical Statistics, Vol. 28, pp. 768-773, (1957).

[38] Reich, E., "Note on Queues in Tandem," Annals of Mathematical Statistics, Vol. 34, pp. 338-341, (1963).

[39] Schucany, W. R., H. L. Gray, and D. B. Owen, "On Bias Reduction in Estimation," Journal of the American Statistical Association, Vol. 66, pp. 524-533, (1971).

[40] Smith, W. L., "Regenerative Stochastic Processes," Proceedings of the Royal Society of London, Series A, Vol. 232, pp. 6-31, (1955).

[41] Tin, M., "Comparison of Some Ratio Estimators," Journal of the American Statistical Association, Vol. 60, pp. 294-307, (1965).

[42] Tukey, J. W., "Bias and Confidence in Not-Quite Large Samples," (Abstract), Annals of Mathematical Statistics, Vol. 29, p. 614, (1958).

[43] Wald, A., and J. Wolfowitz, "An Exact Test of Randomness in the Nonparametric Case Based on Serial Correlation," Annals of Mathematical Statistics, Vol. 14, pp. 378-388, (1943).

[44] Whitt, W., "Embedded Renewal Processes in the GI/G/s Queue," Journal of Applied Probability, Vol. 9, pp. 650-58, (1972).

[45] Wolff, R. W., "Work-Conserving Priorities," Journal of Applied Probability, Vol. 7, pp. 327-37, (1970).