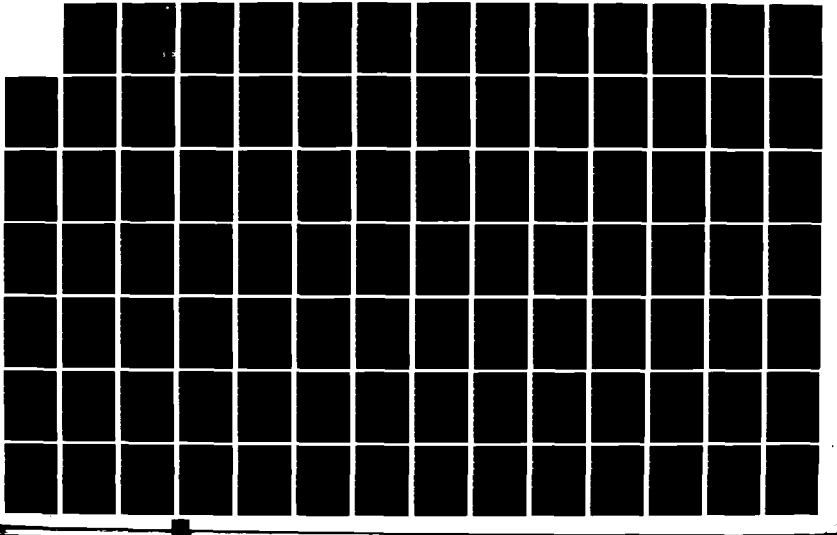
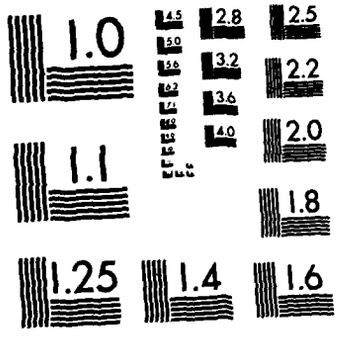
AD-A123 823 AN AUTOREGRESSIVE METHOD FOR SIMULATION OUTPUT ANALYSIS 1/2
(U) STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH
Y L JOW DEC 82 TR-65 N00014-76-C-0578

UNCLASSIFIED F/G 12/1 NL





AN AUTOREGRESSIVE METHOD FOR SIMULATION OUTPUT ANALYSIS

by

Yung-Li Lily Jow

TECHNICAL REPORT NO. 65

December 1982

Prepared under Contract N00014-76-C-0578 (NR 042-343)*

for the

Office of Naval Research

DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

*This research was also partially supported under National Science Foundation Grant MCS79-09139.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In recent years, the use of stochastic systems becomes more extensive in the study of complex phenomena. As the complexity grows, it is more and more difficult in obtaining analytic results. Therefore, the use of computer simulation to study complex stochastic systems has been widely adopted. Typically, we first construct a model which captures the underlying structure of the stochastic system. We then perform sampling experiments on the model and analyze the simulation output sequences to make statistical inferences about the behavior of the system. Since results are based on the observations from our experiments, it is important to develop theoretically sound and computationally efficient methods for simulation output analysis. This is our main concern here.

In general, we want to estimate parameters associated with the steady–state distribution of a stable stochastic process. We use confidence intervals for the quantities of interest to assess the statistical precision of our point estimates. To construct confidence interval for the characteristic of the system under study requires the knowledge of the variance of the estimate. Different methods have been developed to evaluate this quantity. The methods currently being used are the regenerative method, independent replications, batch means, and an autoregressive approach. Except for the

replication method, all methods are based on one output sequence recorded from a single simulation run.

The regenerative method is based on limit theorems developed for regenerative stochastic processes; see Crane and Lemoine (1977) and Iglehart (1978) for an introduction to and a detailed review of the regenerative method. To use it, one has to explore the existence of a regenerative structure and define an appropriate state vector to carry out the simulation. In practice, simulations arise in which simulators may be reluctant to devote the time required to do so or the regenerative property may be absent. In these cases, alternative approaches for simulation output analysis play an important role. The autoregressive method is developed to cope with these limitations of the regenerative method.

The autoregressive approach has been discussed by Fishman (1973, 1978). In this paper we discuss an autoregressive method which is a refinement of the old one and is generalized to multidimensional processes. In contrast to the regenerative method, the autoregressive method, which is based on theorems developed for stationary processes, is a method of approximation. It relies on the assumption that the variance constant required for assessing the precision of point estimates can be approximated arbitrarily closely by the spectral density function at zero of a finite order autoregressive process. Instead of estimating the variance directly we use techniques developed for time series analysis to get an approximation. This is our first goal.

Although simulation is useful, it can be a very expensive tool to use. Since considerable computer time is required for simulation runs, it is therefore important to obtain as precise results as possible from the simulation.

The second goal of this paper is to develop several variance reduction techniques which can be used in conjuction with the autoregressive method for obtaining additional variance reduction for the estimate. Our approach to this objective is to introduce some auxiliary processes, which are correlated with the original process under study; along with the original process and apply the multidimensional version of the autoregressive method.

A natural starting point for achieving our objectives is a review of the theory of stationary processes. Chapter 2 provides a review of the general theory, applicable limit theorems, and some linear models of stationary processes. We also develop some approximation theorems for continuous spectral density function. It is then demonstrated that a continuous spectral density function can be approximated arbitrarily closely by the corresponding spectral density functions of some finite order linear models.

Chapter 3 provides the estimation method for strictly stationary processes. This chapter contains a discussion of order selection criteria along with the establishment of the consistency of estimates. In the last section, we establish some conditions which allow us to apply the autoregressive method to simulation output data and justify the autoregressive method for certain non–stationary processes. Hence, the method is applicable to the simulation of discrete or continuous time Markov processes and semi–Markov processes.

Several variance reduction techniques, which enable us to shorten the confidence interval constructed, are developed in Chapter 4. In the application of these techniques a new point estimate is formed by taking a linear combination of the old point estimate and the point estimates obtained from those auxiliary processes. The coefficients of this linear combination which

minimize the variance of the new point estimate must be estimated.

To see how the method works, Chapter 5 contains several numerical examples. They are (1) the waiting time process in an $M/M/1$ queue, (2) the outflow process of a lake model, (3) the passage time and response time processes in a closed network of queues, (4) the queue length process in a two–station single server cyclic queue. Except for the process in Example 4, which is a semi–Markov process, all processes are Markov processes. All these processes are regenerative processes; hence we include the simulation results obtained by using the regenerative method for comparison. In each example, we are able to provide the theoretical values.

In Chapter 6, we examine the strengths and weaknesses of the autoregressive method.

CHAPTER II

STATIONARY PROCESSES WITH A DISCRETE TIME PARAMETER

2.1. General Definitions and Theorems

Let E be some measurable space, $\mathcal{E}$ a σ–field of subsets of E with a probability measure P. A random variable is a measurable function from E to the real line $\Re$. A stochastic process y with discrete time parameter is a family of random variables $\{ y(n) : n \in T \}$. Here, $y(n)$ is the observation at time n and T is the time range involved, where $T = \{0, 1, 2, \ldots\}$ or $T = \{\ldots, -1, 0, 1, \ldots\}$. And by a d–dimensional random process y we will mean a column vector consisting of d random processes $\{ y_j(n) : n \in T \}$, $j = 1, 2, \ldots, d$,

$$y(n) = \big(y_1(n), y_2(n), \ldots, y_d(n)\big)',$$

where $'$ denotes the transpose of a vector.

(2.1.1) DEFINITION. (a) A stochastic process $y = \{ y(n) : n \in T \}$ is said to be strictly stationary if the joint distribution of $y(n_1 + n)$, $y(n_2 + n)$, $\ldots, y(n_k + n)$ is independent of n for every finite set of integers $\{ n_1, n_2, \ldots, n_k \}$ of T and for every integer n such that $\{ n_1 + n, n_2 + n, \ldots, n_k + n \} \subseteq T$, i.e.,

$$P\{ y_{j_1}(n_1 + n) \in B_1, \ldots, y_{j_k}(n_k + n) \in B_k \} = P\{ y_{j_1}(n_1) \in B_1, \ldots, y_{j_k}(n_k) \in B_k \},$$

for all Borel sets $B_1, \ldots, B_k \in \mathcal{E}$, and every subset $\{ j_1, j_2, \ldots, j_k \}$ of $\{ 1, 2, \ldots, d \}$.

(b) A weakly stationary process y is a stochastic process having finite second moments ($E\{ y_i(n)^2 \} < \infty$), a constant mean $\mu = E\{ y(n) \}$, and its covariance function

$$R(n) = E\{ (y(n+m) - \mu)(y(m) - \mu)^* \}, \tag{2.1.2}$$

exists, is finite, and does not depend on m. Here * denotes the conjugate transpose of a matrix or vector.

Obviously, any process which is stationary in the strict sense and has finite covariance matrix is weakly stationary. Two weakly stationary processes $x = \{ x(n) : n \in T \}$ and $y = \{ y(n) : n \in T \}$ are said to be *stationarily correlated*, if their joint covariance function

$$R_{xy}(m, n) = E\{ (x(m) - \mu_x)(y(n) - \mu_y)^* \},$$

exists and depends only on the difference $m - n$.

(2.1.3) EXAMPLE. White Noise. Let $\ldots, \epsilon(-1), \epsilon(0), \epsilon(1), \ldots$ be a sequence of d–dimensional random vectors with

$$E\{ \epsilon(n) \} = 0,$$
$$E\{ \epsilon(n)^2 \} = G,$$

where 0 is the zero vector and G is non–negative definite[†], and such that any two different vectors are uncorrelated, that is,

$$E\{ \epsilon(n)\epsilon^*(m) \} = 0_d, \qquad \text{for} \quad n \neq m,$$

[†]A matrix G is non-negative definite if $\alpha^* G \alpha \geq 0$ for any complex–valued vector α.

where O_d denotes the $d \times d$ zero matrix. Then the process $\epsilon = \{\epsilon(n) : -\infty < n < \infty\}$ is weakly stationary with covariance function

$$R(k) = \begin{cases} G, & k = 0; \\ O_d, & \text{otherwise.} \end{cases}$$

The time series $\epsilon(n)$ is usually called white noise.

(2.1.4) EXAMPLE. Stationary Markov Chains. Let $\{X(n) : n \geq 0\}$ be a Markov chain for which the initial state $X(0)$ is chosen according to the stationary distribution. Then $\{X(n) : n \geq 0\}$ is strictly stationary.

The class of covariance functions, R, defined by Equation (2.1.2) can be described by the following theorem.

(2.1.5) THEOREM. The covariance function is non–negative definite, that is,

$$\begin{aligned} &R(-n) = R^*(n), \\ &\sum_{m,n=1}^{N} \alpha_n^* R(m-n)\alpha_m \geq 0, \quad N = 1, 2, \ldots \end{aligned} \tag{2.1.6}$$

for every set of complex vectors $\alpha_1, \ldots, \alpha_N$. Conversely, any function R satisfying (2.1.6) is the covariance function of a stationary process.

Proof. See Doob (1953) p. 473. ∎

(2.1.7) THEOREM. If R is the covariance function of a weakly stationary process, then

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} F(d\lambda),$$

where F is a matrix–valued function whose increments, $F(\lambda_1) - F(\lambda_2)$, $\lambda_1 \geq \lambda_2$, are Hermitian non–negative†. The function F is uniquely defined if we require in addition (i) $F(-\pi) = 0$, and (ii) $F(\lambda)$ is right–continuous.

Proof. See Hannan (1970) pp. 34–37. ∎

†A matrix F is Hermitian non–negative if $F = F^*$ and F is non–negative definite.

The function F is called the spectral distribution function and we can write F in the form

$$F = F_{ac} + F_s + F_d,$$

where F_{ac} is the absolutely continuous part of F, F_s is the singular part, and F_d is the discrete part. If F is absolutely continuous, (i.e., $F_s = F_d = O_d$) then

$$F(\lambda) = \int_{-\pi}^{\lambda} f(u)\,du,$$

where f is called the spectral density function. The matrices $f(\lambda)$ are also Hermitian non–negative for $-\pi \leq \lambda \leq \pi$; that is $f(\lambda) = f^*(\lambda)$ and $\alpha^* f(\lambda)\alpha \geq 0$ for every complex–valued vector α. Throughout this paper, we shall only be concerned with *real processes* with absolutely continuous spectral distribution functions. Since we deal only with real processes, $\{y(n) : n \geq 0\}$, $R(n)$ is real, and $R(n) = R'(-n)$. A simple calculation

$$\begin{aligned} R(n) &= \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda)\,d\lambda \\ &= \int_{-\pi}^{\pi} e^{-in\lambda} f(-\lambda)\,d\lambda \\ &= \int_{-\pi}^{\pi} e^{-in\lambda} f'(\lambda)\,d\lambda, \qquad \big(R(n) = R'(-n)\big), \end{aligned}$$

shows that $f(-\lambda) = f'(\lambda)$. If $\sum_{n=-\infty}^{\infty} |R_{ij}(n)|^2 < \infty$, then the numbers $R_{ij}(n)/2\pi$ are simply the Fourier coefficients of the Fourier series expansion of the function $f_{ij}(\lambda)$, thus

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} R(n) e^{-in\lambda}. \tag{2.1.8}$$

Equation (2.1.8) is the Fourier series representation of $f(\lambda)$ and thus f is continuous. Note that Equation (2.1.8) also follows from the condition

$\sum_{n=-\infty}^{\infty}|R_{ij}(n)| < \infty$, since an absolutely summable series is automatically square summable. The next theorem describes the class of spectral distribution functions.

(2.1.9) THEOREM. In order that the matrix function F be the spectral distribution function of some weakly stationary d–dimensional process, it is necessary and sufficient that the matrix $F(\lambda_2) - F(\lambda_1)$ be Hermitian non–negative for $\pi \geq \lambda_2 \geq \lambda_1 \geq -\pi$.

Proof. See Rozanov (1967) pp. 22–23. ∎

Before we state the next theorem, we need the following definitions.

(2.1.10) DEFINITION. A random process $\{ z(t) : -\infty < t < \infty \}$ is said to have orthogonal increments if $E\{ \big(z(\lambda_1) - z(\lambda_2)\big)\big(z(\lambda_3) - z(\lambda_4)\big)^* \} = O_d$, $\lambda_1 > \lambda_2 \geq \lambda_3 > \lambda_4$.

(2.1.11) DEFINITION. Let x_n, $n = 1, 2, \ldots$ be a sequence of random variables for which $E\{ x_n^2 \} < \infty$. Then x_n is said to converge in mean square to a random variable x if

$$\lim_{n \to \infty} E\{ |x_n - x|^2 \} = 0. \tag{2.1.12}$$

A necessary and sufficient condition that x exists satisfying (2.1.12) is the Cauchy condition,

$$\lim_{m,n \to \infty} E\{ |x_m - x_n|^2 \} = 0.$$

We now state the spectral representation theorem for stationary processes.

(2.1.13) THEOREM. Every weakly stationary process $\{ y(n) : -\infty < n < \infty \}$ admits a spectral representation

$$y(n) = \int_{-\pi}^{\pi} e^{in\lambda} z(d\lambda), \tag{2.1.14}$$

where the process $\{ z(\lambda) : -\pi \leq \lambda \leq \pi \}$ has orthogonal increments and

$$E\{ z(d\lambda) z^*(d\lambda) \} = F(d\lambda).$$

Defining $z(\lambda)$ to be right–continuous, it is then uniquely determined neglecting a set in E of probability measure zero.

Proof. See Hannan (1970) p. 41. ∎

The symbol $d\lambda$, which appears in integral (2.1.14), will be thought of as a very small interval containing λ. Also, $d\lambda$, $d\mu, \ldots$ will be small intervals containing λ, $\mu, \ldots$, respectively. Then

$$E\{ z(d\lambda) z^*(d\mu) \} = \begin{cases} F(d\lambda), & \text{if } \lambda = \mu; \\ O_d, & \text{if } \lambda \neq \mu \end{cases}$$

The spectral representation can be modified if $F(\lambda)$ is absolutely continuous(cf. Hannan (1970) pp. 122–123, Rozanov (1967)pp. 39–41). In that case, if there is a measurable matrix $\varphi(\lambda)$ satisfying

$$\varphi(\lambda)\varphi^*(\lambda) = \frac{dF(\lambda)}{d\lambda} = f(\lambda),$$

then there is a process $\{ z_1(\lambda) : -\pi \leq \lambda \leq \pi \}$ with orthogonal increments which satisfies

$$y(n) = \int_{-\pi}^{\pi} e^{in\lambda} \varphi(\lambda) z_1(d\lambda), \quad E\{ z_1(d\lambda) z_1^*(d\lambda) \} = I_d d\lambda, \tag{2.1.15}$$

where I_d denote the $d \times d$ identity matrix. This representation will be used later.

2.2. Linear Transformations of Stationary Processes

Let $y = \{y(n) : -\infty < n < \infty\}$ be some stationary process whose spectral representation is

$$y(n) = \int_{-\pi}^{\pi} e^{in\lambda} z(d\lambda).$$

We will say that the process $x = \{x(n) : -\infty < n < \infty\}$ is obtained from y by a linear transformation, if x admits a spectral representation of the form

$$x(n) = \int_{-\pi}^{\pi} e^{in\lambda} h(\lambda) z(d\lambda),$$

where $h(\lambda)$ satisfies

$$\int_{-\pi}^{\pi} h(\lambda) F(d\lambda) h^*(\lambda) < \infty.$$

We will call the function h the frequency response function of the linear transformation. The joint covariance function of x and y is

$$\begin{aligned} E\{x(m)\, y^*(n)\} &= E\{\,(\int_{-\pi}^{\pi} e^{i'm\lambda}\, h(\lambda) z(d\lambda))(\int_{-\pi}^{\pi} e^{in\mu}\, z(d\mu))^*\,\} \\ &= E\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i'm\lambda - in\mu} h(\lambda) z(d\lambda) z^*(d\mu)\} \\ &= \int_{-\pi}^{\pi} e^{i(m-n)\lambda} h(\lambda) F(d\lambda), \qquad \text{(cf. p. 10)}, \end{aligned}$$

which depends only on $m - n$. Thus the stationary processes x and y are stationarily correlated as defined in Section 2.1. Let us call F_x the spectral distribution function for x, F_y for y, and F_{xy} their joint spectral distribution function. It is not difficult to see

$$\begin{aligned} F_x(\lambda) &= h(\lambda) F_y(\lambda) h^*(\lambda), \\ F_{xy}(\lambda) &= h(\lambda) F_y(\lambda). \end{aligned} \tag{2.2.1}$$

The conditions (2.2.1) are not only necessary, but also sufficient for x to be obtained from y by a linear transformation. This fact is proved in the following theorem.

(2.2.2) THEOREM. Let the stationary processes x and y be stationarily correlated. In order that x be obtainable from y by a linear transformation with frequency response function h, it is necessary and sufficient that the spectral distribution functions F_x, F_y and F_{xy} satisfy conditions (2.2.1).

Proof. See Rozanov (1967) p. 36. ∎

Remark. If the stationary process y has a spectral density f_y then a process x which is obtainable from y by a linear transformation with frequency response function h also has a spectral density, which is given by

$$f_x(\lambda) = h(\lambda) f_y(\lambda) h^*(\lambda). \tag{2.2.3}$$

Let us consider two examples of linear models known as moving average processes and autoregressive processes.

(2.2.4) EXAMPLE. Moving Average Processes. A moving average process $y = \{ y(n) : -\infty < n < \infty \}$ is defined by the following expression

$$y(n) = \sum_{j=-\infty}^{\infty} A(j)\epsilon(n-j), \tag{2.2.5}$$

where $\epsilon = \{ \epsilon(n) : -\infty < n < \infty \}$ is a sequence of uncorrelated random vectors with covariance matrix G and $\{ A(j) : -\infty < j < \infty \}$ are $d \times d$ real matrices. Note that the process y is obtained from ϵ by a linear transformation with $h(\lambda) = \sum_{j=-\infty}^{\infty} A(j) e^{-ij\lambda}$. A necessary and sufficient condition for the series in (2.2.5) to converge in mean square is

$$\sum_{j=-\infty}^{\infty} \|A(j)\|^2 < \infty,$$

where $\|\cdot\|$ stands for any norm of a matrix. It is straightforward to see that the process ϵ has spectral density function $(2\pi)^{-1}G$. From Equation (2.2.3), we easily obtain the spectral density function f for y,

$$f(\lambda) = \frac{1}{2\pi}\Big(\sum_{j=-\infty}^{\infty} A(j)e^{-ij\lambda}\Big) G \Big(\sum_{j=-\infty}^{\infty} A(j)e^{-ij\lambda}\Big)^{*},$$

which is a continuous function. Therefore, the spectral density function of a moving average process is absolutely continuous.

Alternatively, suppose we are given the stationary process y with absolutely continuous spectral distribution function F and spectral density f. Since $f(\lambda)$ is Hermitian non–negative, $f(\lambda)$ can be diagonalized by an unitary matrix $U(\lambda)$[†]:

$$f(\lambda) = U(\lambda)D(\lambda)U^{*}(\lambda),$$

where $D(\lambda)$ is a diagonal matrix and the elements of D are real and non–negative. Define

$$\big(f(\lambda)\big)^{1/2} = U(\lambda)\big(D(\lambda)\big)^{1/2}U^{*}(\lambda),$$

where $\big(D(\lambda)\big)^{1/2}$ is obtained by taking the non–negative square root of all the elements of $D(\lambda)$. Then $\big(f(\lambda)\big)^{1/2}$ is a Hermitian non–negative, measurable function and

$$f(\lambda) = \big(f(\lambda)\big)^{1/2}\big(\big(f(\lambda)\big)^{1/2}\big)^{*}.$$

Then each element of $\big(f(\lambda)\big)^{1/2}$ is square integrable and has a Fourier series expansion, namely,

$$\big(f(\lambda)\big)^{1/2} = \sum_{j=-\infty}^{\infty} A(j)e^{-ij\lambda}, \quad \sum_{j=-\infty}^{\infty} \|A(j)\|^2 < \infty,$$

[†]See, for example, Strang (1976) pp. 212–213.

where

$$A(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \big(f(\lambda)\big)^{1/2}\, e^{ij\lambda} d\lambda.$$

It follows from (2.1.15) that the spectral representation of y can be written as

$$\begin{aligned} y(n) &= \int_{-\pi}^{\pi} e^{in\lambda}\big(f(\lambda)\big)^{1/2} z_1(d\lambda) \\ &= \int_{-\pi}^{\pi} e^{in\lambda} \sum_{j=-\infty}^{\infty} A(j) e^{-ij\lambda} z_1(d\lambda) \\ &= \sum_{j=-\infty}^{\infty} A(j)\epsilon(n-j), \end{aligned}$$

where

$$\epsilon(n) = \int_{-\pi}^{\pi} e^{in\lambda} z_1(d\lambda), \quad E\{\, z_1(d\lambda) z_1^*(d\lambda)\,\} = I_d d\lambda.$$

We see that $\{\, \epsilon(n) : -\infty < n < \infty \,\}$ is a sequence of uncorrelated random vectors. This shows that y is a moving average process. Thus we have proved the following theorem.

(2.2.6) THEOREM. A weakly stationary process is a moving average process if and only if its spectral distribution function is absolutely continuous.

Proof. See Ronazov (1967) pp. 39–42. ∎

The next two theorems deal with stationary processes with absolutely continuous distribution having spectral densities which are polynomials in $e^{-i\lambda}$. We first state the theorem for scalar case.

(2.2.7) THEOREM. If the spectral density function

$$f(\lambda) = \sum_{j=-q}^{q} \gamma(j) e^{-ij\lambda} \geq 0, \quad \gamma(j) = \gamma(-j), \quad \gamma(q) \neq 0,$$

where $\gamma(j)$ is real, then it may be represented in the form

$$f(\lambda) = \frac{1}{2\pi}|\sum_{j=0}^{q} \alpha(j)e^{-ij\lambda}|^2, \quad \alpha(j) \text{ is real,} \tag{2.2.8}$$

where the polynomial $\sum_{j=0}^{q} \alpha(j)z^j$ has no zeros in the open unit disc.

Proof. We follow the proof given in Hannan (1970) pp. 62–63 . We consider

$$M(z) = \sum_{j=-q}^{q} \gamma(j)z^{-j}.$$

This expression has $2q$ zeros, counting each with its appropriate multiplicity. Let z_k be a zero whose absolute value is not 1, then $\bar{z}_k^{-1}$ is also a zero because $\gamma(j) = \gamma(-j)$. Thus the zeros different from 1 in absolute value can be paired $(z_k, \bar{z}_k^{-1})$. Let there be $r \leq q$ such pairs of zeros taking each pair as often as its multiplicity. Then there are $2s = 2q - 2r$ zeros of absolute value 1, say $e^{i\theta_k}$, $k = 1, \ldots, 2s$. Then

$$\begin{aligned} f(\lambda) &= \gamma(q)e^{-iq\lambda} \prod_{j=1}^{r} (e^{i\lambda} - z_j)(e^{i\lambda} - \bar{z}_j^{-1}) \prod_{k=1}^{2s} (e^{i\lambda} - e^{i\theta_k}) \\ &= \prod_{j=1}^{r} (e^{i\lambda} - z_j)(e^{-i\lambda} - \bar{z}_j)\{\gamma(q)e^{-is\lambda} \prod_{l=1}^{r} (-\bar{z}_l^{-1}) \prod_{k=1}^{2s} (e^{i\lambda} - e^{i\theta_k})\}. \end{aligned}$$

It follows that the bracked factor is real and non–negative. Moreover, the derivative of the bracked factor vanishes at $\lambda = \theta_k$ because it is non–negative and zero at $\lambda = \theta_k$. Thus θ_k occurs in pairs. The $2q$ zeros can be numbered and divide into two sets $(z_1, \ldots, z_q)$ and $(z_{q+1}, \ldots, z_{2q})$ such that if $|z_k| < 1$ then $z_{q+k} = \bar{z}_k^{-1}$ and if $|z_k| = 1$ then $z_{q+k} = z_k$. Then

$$f(\lambda) = \frac{\alpha(q)^2}{2\pi}|\prod_{k=1}^{q} (e^{i\lambda} - z_k)|^2 = \frac{1}{2\pi}|\sum_{j=0}^{q} \alpha(j)e^{-ij\lambda}|^2,$$

where

$$\alpha(q)^2 = 2\pi\gamma(q)\prod_{j=1}^{r}(\bar{z}_k^{-1})\prod_{k=1}^{s}(-e^{i\theta_k}).$$

Since the coefficients in $M(z)$ are real, the zeros are real or occur in conjugate pairs. Then $\alpha(j)$ will be real. ∎

Note that (2.2.8) is the spectral density function of a qth order moving average process (cf. pp. 12–13). There is a similar result for the vector case. Since the proof is rather long, we omit it and refer the interested reader to Hannan (1970) and Rozanov (1967).

(2.2.9) THEOREM. A non–negative matrix function

$$f(\lambda) = \sum_{j=-q}^{q} \Gamma(j)e^{-ij\lambda}, \quad \Gamma(q) \neq O_d, \quad \Gamma(j) = \Gamma'(-j),$$

which has a determinant not identically zero, can be represented in the form

$$f(\lambda) = \frac{1}{2\pi}\Big(\sum_{j=0}^{q} A(j)e^{-ij\lambda}\Big)\Big(\sum_{j=0}^{q} A(j)e^{-ij\lambda}\Big)^{*}, \tag{2.2.10}$$

where $A(0)$ is Hermitian non–negative definite, $A(j)$ are real and all zeros of $\det(\sum_{j=0}^{q} A(j)z^j)$ lie on or outside the unit disc.

Proof. See Hannan (1970) pp. 63–66. ∎

The factorization (2.2.10) leads to a finite order moving average representation (cf. pp. 12–13), namely,

$$y(n) = \sum_{j=0}^{q} A(j)\epsilon(n-j), \quad E\{\epsilon(m)\epsilon^{*}(n)\} = \delta_m^n I_d,$$

where $\delta_m^n = 1$ if $m = n$, $\delta_m^n = 0$ otherwise.

We now study another important linear model which is called the autoregressive process.

(2.2.11) EXAMPLE. Autoregressive Processes. A stationary process $y = \{ y(n) : -\infty < n < \infty \}$ is called a pth order autoregressive process if it satisfies

$$\sum_{j=0}^{p} B(j)y(n-j) = \epsilon(n), \quad B(0) = I_d, \quad B(p) \neq O_d, \tag{2.2.12}$$

where $\epsilon = \{ \epsilon(n) : -\infty < n < \infty \}$ is a process of uncorrelated random vectors with covariance matrix G. If $y(n)$ has constant mean, μ, then we certainly have

$$\sum_{j=0}^{p} B(j)\mu = E\{ \epsilon(n) \},$$

so that (2.2.12) holds for the new process $\{ y(n) - \mu : -\infty < n < \infty \}$ with $\epsilon(n) - E\{ \epsilon(n) \}$ on the right. Therefore, we may assume $E\{ \epsilon(n) \} = 0$ to facilitate our discussion. If all zeros of

$$\det\left(\sum_{j=0}^{p} B(j)z^j \right)$$

lie outside the unit disc, then a solution of (2.2.12) exists and is of the form

$$y(n) = \sum_{j=0}^{\infty} A(j)\epsilon(n-j), \quad A(0) = I_d, \tag{2.2.13}$$

where the $\|A(j)\|$ converges exponentially to zero as j increases (cf. Hannan (1970) p. 326). In order to obtain the spectral density function f of the process y, we observe that ϵ is obtained from y by a linear transformation with $h(\lambda) = \sum_{j=0}^{p} B(j)e^{-ij\lambda}$. From Equation (2.2.3), we obtain $h(\lambda)f(\lambda)h^*(\lambda) = (2\pi)^{-1}G$. Hence

$$f(\lambda) = \frac{1}{2\pi}\left(\sum_{j=0}^{p} B(j)e^{-ij\lambda} \right)^{-1} G \left(\sum_{j=0}^{p} B(j)e^{-ij\lambda} \right)^{*-1}.$$

Multiply (2.2.12) by the transpose of (2.2.13) with n replaced by $n-s$ and take expectation on both sides, we obtain

$$E\{\sum_{j=0}^{p} B(j)y(n-j)y'(n-s)\} = E\{\sum_{k=0}^{\infty} \epsilon(n)\epsilon'(n-s-k)A'(k)\}$$

which leads to

$$\sum_{j=0}^{p} B(j)R(s-j) = \delta_0^s G, \quad s = 0, 1, 2, \ldots \tag{2.2.14}$$

These are often called the Yule-Walker equations.

2.3. The Law of Large Numbers and Central Limit Theorem

The study of strictly stationary processes is usually carried out in the context of measure–preserving transformations, in this section we shall follow the development of Doob (1953), Chapter 10. Let $(E, \mathcal{E}, P)$ be a probability space; that is to say, E is a space of points ω, $\mathcal{E}$ is a σ–field of subsets of E, and P is a probability measure on $\mathcal{E}$.

(2.3.1) DEFINITION. Let T be a transformation of $\mathcal{E}$ onto itself. It is called a measure–preserving set transformation if the following conditions are satisfied :

(C1) T is single valued, modulo set of probability 0: if A_1 is an image of A under T, the class of all images of A is the class of all measurable sets differing from A_1 by sets of probability 0.

(C2) $P(TA) = P(A)$.

(C3) Neglecting ω sets of probability 0,

$$T\Big(A_1 \bigcup A_2\Big) = TA_1 \bigcup TA_2$$

$$T\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \bigcup_{n=1}^{\infty} TA_n$$

$$T(E - A) = E - TA.$$

If every measurable set is the image of some measurable set under T, this transformation must be 1–1 (neglecting sets of probability 0) and the inverse T^{-1} is defined, and is also a measure–preserving set transformation. If T is a measure–preserving set transformation, there is one and only one transformation T_1 defined for every random variable, taking random variables into random variables, and having the following properties:

(P1) T_1 is single-valued modulo the random variables which vanish with probability 1: if x_1 is an image of x under T_1, the class of all images of x is the class of all random variables equal to x_1 with probability 1.

(P2) T_1 is consistent with T: $T_1 1_A$ is 1 almost everywhere on TA and 0 otherwise.

(P3) T_1 is linear: if a, b are constants and if x, y are random variables,

$$T_1(ax + by) \equiv aT_1x + bT_1y \quad \text{a.e.}$$

(P4) T_1 preserves convergence: if

$$\lim_{n\to\infty} x_n = x \quad \text{a.e.},$$

then $\lim_{n\to\infty} T_1 x_n = Tx$ a.e..

We shall use the same notation T for T_1.

(2.3.2) PROPOSITION. If T is a measure–preserving set transformation, and if x is a random variable, the stochastic process

$$\{\, y(n) : n \geq 0 \,\}, \quad y(n) = T^n x, \tag{2.3.3}$$

is strictly stationary, and, if T has an inverse, the stochastic process

$$\{\, y(n) : -\infty < n < \infty \,\}, \quad y(n) = T^n x,$$

is also strictly stationary.

Proof. See Doob (1953) pp. 454–455. ∎

(2.3.4) PROPOSITION. Let $\{y(n) : n \geq 0\}$ be a strictly stationary process, then there is one and only one measure–preserving set transformation T such that $T^n y(0) = y(n)$ *a.e.* for all $n \geq 1$.

Proof. See Doob (1953) pp. 455–456. ∎

The above discussion and results apply equally well to multi–dimensional processes; the interested reader is referred to Rozanov (1967) Chapter 4 for an excellent treatment.

(2.3.5) DEFINITION. A measure–preserving set transformation is called metrically transitive if the only invariant random variables are constant with probability one.

The process defined by (2.3.3) is metrically transitive if T is metrically transitive. We now have the ergodic theorem.

(2.3.6) THEOREM. If $\{y(n) : n \geq 0\}$ is strictly stationary and metrically transitive with $E\{|y_j(n)|\} < \infty$ for $j = 1, \ldots, d$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} y(k) = E\{y(0)\} \quad \text{a.e.}$$

Also if $E\{y_j(n)^2\} < \infty$ for $j = 1, \ldots, d$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} y(m+k)y'(k) = E\{y(m)y'(0)\} \quad \text{a.e.} \qquad (2.3.7)$$

Proof. See Hannan (1970) p. 203. ∎

We shall study a weaker condition than metrically transitive which implies it. Let $\{y(n) : n \geq 0\}$ be a strictly stationary process defined on a probability space $(E, \mathcal{E}, P)$. For integers $0 \leq a \leq b$, let $\mathcal{F}_a^b$ be the σ–field generated by $y_j(a), \ldots, y_j(b)$, $0 \leq j \leq d$ (with the obvious extension for $b = \infty$).

(2.3.8) DEFINITION. We say that the process $\{ y(n) : n \geq 0 \}$ is ϕ–mixing if there exists a non–negative function ϕ of positive integers such that $\lim_{n \to \infty} \phi(n) = 0$ and for each $k \geq 0$, and $n \geq 1$,

$$\sup\{ |P(E_2 \mid E_1) - P(E_2)| : E_1 \in \mathcal{F}_0^k, E_2 \in \mathcal{F}_{k+n}^{\infty} \} \leq \phi(n). \qquad (2.3.9)$$

Condition (2.3.9) simply says that events concerning the "future" of the process become almost independent of events in the past.

Remark 1. Let $\{ X(n) : n \geq 0 \}$ be a stationary Markov chain with finite state space, and let $y(n) = f(X(n))$ where f is some real function on the state space. If $\{ X(n) : n \geq 0 \}$ is irreducible and aperiodic, then $\{ y(n) : n \geq 0 \}$ is ϕ–mixing (cf. Billingsley (1968), pp. 167–168).

Remark 2. If $\{ X(n) : n \geq 0 \}$ is a Markov process with infinite state space, then $\{ y(n) : n \geq 0 \}$ is ϕ–mixing if $\{ X(n) : n \geq 0 \}$ satisfies Doeblin's condition, has one ergodic class, and is aperiodic.†

(2.3.10) THEOREM. Suppose that the process $\{ y(n) : n \geq 0 \}$ is ϕ–mixing with $\sum_n \phi(n)^{1/2} < \infty$, has covariance function R, and has mean zero ($E\{ y \} = 0$. Then

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=0}^{n-1} y(k) \right) \Rightarrow N(0, \Sigma).$$

Here $\Rightarrow$ denotes the convergence in distribution and $N(0, \Sigma)$ is the d–dimensional normal random vector with mean 0 and covariance matrix Σ,

$$\Sigma = (\sigma_{ij}) = \sum_{k=-\infty}^{\infty} R(k). \qquad (2.3.11)$$

†See Doob (1953), p. 190.

The elements of Σ are

$$\sigma_{ij} = E\{ y_i(0)y_j(0) \} + \sum_{k=1}^{\infty} E\{ y_i(0)y_j(k) \} + \sum_{k=1}^{\infty} E\{ y_i(k)y_j(0) \}, \quad (2.3.12)$$

and the series converges absolutely.

Proof. See Billingsley (1968) pp. 174–177. ∎

Let f be the spectral density function of the process y in Theorem (2.3.10). From this theorem we know $\sum_{k=-\infty}^{\infty} |R_{ij}(k)| < \infty$ for $i, j = 1, \ldots, d$, therefore f has Fourier series expansion (cf. Section 2.1),

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} R(j)e^{-ij\lambda}.$$

Observe that $\Sigma = \sum_{j=-\infty}^{\infty} R(j) = 2\pi f(0)$. In the next section we will show that any continuous spectral density function can be approximated arbitrarily closely by the spectral density function at zero of a finite order autoregressive process. Therefore, we can use techniques in the time series literature to obtain an estimate of Σ.

2.4 Approximation of Spectral Density Function

In this section we shall demonstrate that any arbitrary continuous spectral density function f can be approximated by a polynomial in $e^{-i\lambda}$ having the following form

$$g(\lambda) = \sum_{j=-q}^{q} \Gamma(j)e^{-ij\lambda}, \quad \Gamma(j) = \Gamma'(-j), \quad \Gamma(q) \neq O_d. \tag{2.4.1}$$

In Section 2.2, we have seen that g is the spectral density function of a finite order moving average process. We shall also show that f can be approximated arbitrarily closely by the spectral density function of a finite order autoregressive process.

The approximation (2.4.1) is based on the Weierstrass approximation theorem which states:

(2.4.2) THEOREM. If k is a continuous function of period 2π, then corresponding to every positive number ϵ there exists a trigonometric sum

$$S(\lambda) = a_0 + \sum_{j=1}^{n} (a_j \cos \lambda j + b_j \sin \lambda j),$$

such that the inequality

$$|S(\lambda) - k(\lambda)| < \epsilon \tag{2.4.3}$$

is satisfied for all value of λ.

Proof. See Achieser (1956). ∎

The results contained in the following remarks will be used throughout this section.

Remark 1. For any continuous function f of period 2π, let $k(\lambda) = f(\lambda) + \delta_1$ and let $S(\lambda)$ satisfy the Weierstrass approximation theorem (2.4.2) with ϵ

replaced by $\delta_2 > 0$ in (2.4.3). Then

$$\delta_1 - \delta_2 < S(\lambda) - f(\lambda) < \delta_1 + \delta_2.$$

By properly selecting δ_1 and δ_2, we can make the upper bound and lower bound of $S(\lambda) - f(\lambda)$ arbitrarily small.

Remark 2. If $k(\lambda) = k(-\lambda)$ is an even function, let $S_1(\lambda) = \frac{1}{2}(S(\lambda) + S(-\lambda))$, which is also an even function. Then we have

$$\begin{aligned} |S_1(\lambda) - k(\lambda)| &= |\frac{1}{2}(S(\lambda) - k(\lambda)) + \frac{1}{2}(S(-\lambda) - k(-\lambda))| \\ &< \epsilon, \end{aligned}$$

and

$$\begin{aligned} S_1(\lambda) &= \frac{1}{2}(S(\lambda) + S(-\lambda)) \\ &= a_0 + \frac{1}{2}\sum_{j=1}^{n} a_j(\cos \lambda j + \cos(-\lambda j)) + \frac{i}{2}\sum_{j=1}^{n} b_j(\sin \lambda j + \sin(-\lambda j)), \\ &= a_0 + \sum_{j=1}^{n} \frac{1}{2}a_j(e^{ij\lambda} + e^{-ij\lambda}) = \sum_{j=-n}^{n} c_j e^{-ij\lambda} \end{aligned}$$

where $i = \sqrt{-1}$, $c_0 = a_0$, $c_j = c_{-j} = \frac{1}{2}a_j$ are real.

Similarly, if $k(\lambda) = -k(-\lambda)$ is an odd function, we may form an odd function $S_2(\lambda) = \frac{1}{2}(S(\lambda) - S(-\lambda))$. Then

$$|S_2(\lambda) - k(\lambda)| < \epsilon$$

and

$$\begin{aligned} S_2(\lambda) &= \frac{1}{2}\sum_{j=1}^{n} a_j(\cos \lambda j - \cos(-\lambda j)) + \frac{i}{2}\sum_{j=1}^{n} b_j(\sin \lambda j - \sin(-\lambda j)) \\ &= \frac{i}{2}\sum_{j=1}^{n} b_j(e^{ij\lambda} - e^{-ij\lambda}) = i\sum_{j=-n}^{n} d_j e^{-ij\lambda}, \end{aligned}$$

where $d_0 = 0$, $d_j = -d_{-j} = -\frac{1}{2}b_j$ are real.

The next theorem, which can be found in most time series literature (cf. Anderson (1970) pp. 410–411), is an approximation theorem for scalar processes.

(2.4.4) THEOREM. If a spectral density function f is continuous then for any $\epsilon > 0$ there is a spectral density function of the form

$$g(\lambda) = \sum_{j=-m}^{m} \gamma(j) e^{-ij\lambda}, \tag{2.4.5}$$

where $\gamma(j) = \gamma(-j)$ is real, $g(\lambda) > \epsilon/2$ and $|g(\lambda) - f(\lambda)| < \epsilon, -\pi \leq \lambda \leq \pi$.

Proof. Let $\delta_1 = 3\epsilon/4$, $\delta_2 = \epsilon/4$ and take $g(\lambda) = S(\lambda)$ as described in *Remark 1.* Then

$$\epsilon/2 < g(\lambda) - f(\lambda) < \epsilon,$$

and $g(\lambda) > \epsilon/2$ since f is a spectral density function which is non–negative for all λ. Since the spectral density function f is an even function (cf. Section 2.1), we can take $g(\lambda)$ as (2.4.5) according to the argument of *Remark 2.* ∎

(2.4.6) COROLLARY. If a spectral density function f is continuous then for any $\epsilon > 0$ there is a finite order moving average process with positive spectral density, say g, such that

$$|f(\lambda) - g(\lambda)| < \epsilon, \quad -\pi \leq \lambda \leq \pi.$$

Proof. By Theorem (2.4.4), f can be approximated by a spectral density function (2.4.5). Then apply Theorem (2.2.7), it follows that g is the spectral density function of a finite order moving average process. ∎

(2.4.7) COROLLARY. If a spectral density function f is continuous, then for any arbitrary $\epsilon > 0$, there is an autoregressive process with spectral density function, say h, such that $|f(\lambda) - h(\lambda)| < \epsilon, -\pi \leq \lambda \leq \pi$.

Proof. Let $f_\epsilon(\lambda) = f(\lambda) + \frac{\epsilon}{2}$, $-\pi \leq \lambda \leq \pi$. Obviously $f_\epsilon(\lambda)$ is continuous and positive; therefore, the reciprocal of $f_\epsilon(\lambda)$, $f_\epsilon^{-1}(\lambda)$, exists and is also

continuous and positive. An application of Theorem (2.1.9) shows that $f_\epsilon^{-1}(\lambda)$ is the spectral density function of a stationary process. By Corollary (2.4.6) there is a positive spectral density function of a moving average process, say g, such that $|f_\epsilon^{-1}(\lambda) - g(\lambda)| < \epsilon'$, $-\pi \leq \lambda \leq \pi$, where

$$\epsilon' = \frac{1}{2}\epsilon \cdot \frac{\min\, g(\lambda)}{\max\, f_\epsilon(\lambda)}.$$

Then

$$\begin{aligned} |f(\lambda) - g^{-1}(\lambda)| &\leq \frac{\epsilon}{2} + |f_\epsilon(\lambda) - g^{-1}(\lambda)| \\ &\leq \frac{\epsilon}{2} + |f_\epsilon(\lambda)| \cdot |g^{-1}(\lambda)| \cdot |f_\epsilon^{-1}(\lambda) - g(\lambda)| \\ &< \epsilon. \end{aligned}$$

This proves the theorem. ∎

To prove similar theorems for vector processes, we shall begin with 2-dimensional case ($d = 2$), then extend the results to $d > 2$. Let f be a continuous 2×2 spectral density matrix. Since $f(\lambda) = f^*(\lambda)$ (cf. Section 2.1), we can take

$$f(\lambda) = \begin{pmatrix} f_1(\lambda) & p_1(\lambda) + ip_2(\lambda) \\ p_1(\lambda) - ip_2(\lambda) & f_2(\lambda) \end{pmatrix}$$

where f_k and p_k are real continuous functions. In Section 2.1, we have shown that $f(\lambda) = f'(-\lambda)$, therefore f_k and p_1 are even functions and p_2 is an odd function. Then we can properly select δ_1, δ_2 and the approximation functions g_k, q_k, for f_k and p_k, $k = 1, 2$ respectively, such that for $\epsilon > 0$

$$g_k(\lambda) = \sum_{j=-n_k}^{n_k} \gamma_{kj} e^{-ij\lambda}, \quad \gamma_{kj} = \gamma_{k,-j} \quad \text{is real},$$

$$\epsilon < \Delta_k(\lambda) = g_k(\lambda) - f_k(\lambda) < 2\epsilon,$$

and

$$q_1(\lambda) = \sum_{j=-m_1}^{m_1} d_{1j} e^{-ij\lambda}, \quad d_{1j} = d_{1,-j} \quad \text{is real},$$

$$\epsilon/4 < U(\lambda) = q_1(\lambda) - p_1(\lambda) < \epsilon/2,$$

and

$$q_2(\lambda) = i \sum_{j=-m_2}^{m_2} d_{2j} e^{-ij\lambda}, \quad d_{2j} = -d_{2,-j} \quad \text{is real},$$

$$\epsilon/4 < V(\lambda) = q_2(\lambda) - p_2(\lambda) < \epsilon/2.$$

Put

$$g(\lambda) = \begin{pmatrix} g_1(\lambda) & q_1(\lambda) + iq_2(\lambda) \\ q_1(\lambda) - iq_2(\lambda) & g_2(\lambda) \end{pmatrix}$$

$$= \sum_{j=-m}^{m} \Gamma(j) e^{-ij\lambda} = f(\lambda) + \Delta(\lambda)$$

where

$$\Delta(\lambda) = \begin{pmatrix} \Delta_1(\lambda) & U(\lambda) + iV(\lambda) \\ U(\lambda) - iV(\lambda) & \Delta_2(\lambda) \end{pmatrix}$$

$$m = \max(n_k, m_k),$$

$$\Gamma_{kk}(j) = \gamma_{kj}, \quad \gamma_{kj} = 0, \quad \text{if } |j| > n_k, \quad k = 1, 2,$$

$$\Gamma_{12}(j) = d_{1j} - d_{2j}, \quad d_{kj} = 0, \quad \text{if } |j| > m_k, \quad , k = 1, 2,$$

$$\Gamma_{21}(j) = d_{1j} + d_{2j},$$

and $\|\Delta(\lambda)\| < 3\epsilon$, if we define the norm to be maximum row sum. Clearly, $\Gamma(m) \neq O_d$, $\Gamma(j)$ is real and $\Gamma'(j) = \Gamma(-j)$. To show the non-negativity of $g(\lambda)$, we observe that for any complex-valued vector α

$$\alpha' = a' + ib' = (a_1, a_2) + i(b_1, b_2), \quad a_i, b_i \in \Re, \quad i = 1, 2,$$

$$\alpha^* g(\lambda)\alpha = \alpha^* f(\lambda)\alpha + \alpha^* \Delta(\lambda)\alpha,$$

$$\geq \alpha^* \Delta(\lambda)\alpha \quad \text{(by the non-negativity of } f(\lambda))$$

$$= a_1^2\Delta_1(\lambda) + a_2^2\Delta_2(\lambda) + 2a_1a_2U(\lambda) + b_1^2\Delta_1(\lambda)$$

$$+ b_2^2\Delta_2(\lambda) + 2b_1b_2U(\lambda) + 2(a_2b_1 - a_1b_2)V(\lambda).$$

Since Δ_k,U and V are positive, the worst case occurs when $a_1a_2 < 0$, $b_1b_2 < 0$ and $(a_2b_1 - a_1b_2) < 0$. In this case, we use the following inequalities:

$$\Delta_k(\lambda) > \epsilon,$$

$$U(\lambda), V(\lambda) < \epsilon/2.$$

We then have

$$\alpha^* g(\lambda)\alpha \geq \frac{\epsilon}{2}((a_1+a_2)^2+(b_1+b_2)^2+(a_2+b_1)^2+(a_1-b_2)^2) \geq 0.$$

Thus, we have shown that we can find an approximation g which is also a spectral density function when $d=2$. Now, assume $d>2$ and f is a $d\times d$ spectral density matrix having continuous components. We may obtain an approximation function (2.4.1), say g, by applying the following algorithm.

(2.4.8) ALGORITHM. (Approximation of Spectral Density Function)

A1. [Initialization.] Let $g^{(k)}(\lambda)=f(\lambda)$ for $k=0$. Set $k=0$, $i=1$, and $j=2$.

A2. [Polynomial approximation.] Let

$$w(\lambda)=\begin{pmatrix} g_{ii}^{(k)}(\lambda) & g_{ij}^{(k)}(\lambda) \\ g_{ji}^{(k)}(\lambda) & g_{jj}^{(k)}(\lambda)\end{pmatrix}.$$

If all components of $w(\lambda)$ are polynomials in $e^{-i\lambda}$ then go to A4. Otherwise, find a 2×2 matrix function, denoted by h, which approximates $w(\lambda)$ using the procedure described in pp. 27–28.

A3. [Update.] Form $g^{(k+1)}$ by substituting h for w in $g^{(k)}$ and update $k=k+1$.

A4. [Done?] Set $j=j+1$, if $j\leq d$ then go to A2. Otherwise, set $i=i+1$. If $i<d$ then put $j=i+1$ and go to A2.

A5. [Obtain g.] Put $g(\lambda)=g^{(k)}(\lambda)$. ∎

Note that all $g^{(k)}$ are spectral density functions, since they are Hermitian non-negative (cf. Section 2.1). Note also that $\|g^{(k+1)}(\lambda)-g^{(k)}(\lambda)\|<3\epsilon$ for

all k. This algorithm will terminate in a finite number of steps, M ($M \leq d(d-1)/2$). Step **A5.** puts $g = g^{(M)}$ which is a spectral density function having the form (2.4.1), and

$$\|g(\lambda) - f(\lambda)\| < C \cdot \epsilon,$$

for some finite constant C ($C \leq 3M$). Thus we have established the following result.

(2.4.9) THEOREM. If f is a $d \times d$ spectral density matrix with continuous components, for any $\epsilon > 0$ there is a spectral density matrix of the form

$$g(\lambda) = \sum_{j=-m}^{m} \Gamma(j) e^{-ij\lambda}, \quad \Gamma(m) \neq O_d, \quad \Gamma(j) = \Gamma'(-j),$$

and $\|g(\lambda) - f(\lambda)\| < \epsilon, \quad -\pi \leq \lambda \leq \pi.$

(2.4.10) COROLLARY. Let f be as described in Theorem (2.4.9), then for any $\epsilon, \epsilon' > 0$, there is a finite order moving average process with spectral density function g, such that $\|f(\lambda) - g(\lambda)\| < \epsilon, -\pi \leq \lambda \leq \pi$. Also, there is a finite order autoregressive process with spectral density function h, such that $\|f(\lambda) - h(\lambda)\| < \epsilon', -\pi \leq \lambda \leq \pi$.

Proof. The first part follows directly from Theorem (2.4.9) and Theorem (2.2.7). For the second part let

$$f_\epsilon(\lambda) = f(\lambda) + \epsilon \cdot I_d,$$

then every entry of $f_\epsilon(\lambda)$ is continuous and $f_\epsilon(\lambda)$ is positive definite for all λ. Thus the reciprocal of $f_\epsilon(\lambda)$, $k(\lambda) = f_\epsilon^{-1}(\lambda)$, exists for all λ and every

component of $k(\lambda)$ is continuous. Let

$$S(\lambda) = \left(\sum_{j=0}^{q} A(j)e^{-ij\lambda}\right) G \left(\sum_{j=0}^{q} A(j)e^{-ij\lambda}\right)^*, \quad A(0) = I_d, \quad A(j) \text{ is real,}$$

such that $\|S(\lambda) - k(\lambda)\| < \epsilon_1$, then

$$\begin{aligned} \|f(\lambda) - S^{-1}(\lambda)\| &\leq \|f(\lambda) - f_\epsilon(\lambda)\| + \|f_\epsilon(\lambda) - S^{-1}(\lambda)\| \\ &\leq \epsilon + \|f_\epsilon(\lambda)\| \cdot \|S^{-1}(\lambda)\| \cdot \|g(\lambda) - k(\lambda)\|. \end{aligned}$$

And the reciprocal of $S(\lambda)$, $S^{-1}(\lambda)$, exists except perhaps for finitely many λ and is the spectral density function of an autoregressive process. By properly choosing ϵ_1, and letting $h(\lambda) = S^{-1}(\lambda)$, we obtain $\|f(\lambda) - h(\lambda)\| < \epsilon'$. ∎

CHAPTER III

THE AUTOREGRESSIVE METHOD AND ITS APPLICATIONS

Based on the results of Chapter 2, it is clear that a continuous spectral density function can be approximated arbitrarily closely by the spectral density function of a finite order autoregressive process. We shall see that it is possible to use this result to obtain consistent point estimates and asymptotically valid confidence intervals for quantities of interest.

Let $\{ y(n) : n \geq 0 \}$ be a d–dimensional strictly stationary process which is ϕ–mixing with $\sum_{n=1}^{\infty} \phi(n)^{1/2} < \infty$. We wish to estimate the quantity

$$r = E\{ y(n) \}.$$

From Theorem (2.3.6) and Theorem (2.3.10) we know

$$\begin{aligned} \hat{r}_n &= \frac{1}{n} \sum_{i=1}^{n} y(i) \to r \quad \text{a.s.} \\ \sqrt{n}\,(\hat{r}_n - r) &\Rightarrow N(0, \Sigma), \end{aligned} \tag{3.0.1}$$

where $\Sigma = \sum_{k=-\infty}^{\infty} R(k)$ and the series converges absolutely. Corollary (2.4.10) shows that for any arbitrarily small $\epsilon > 0$ there is a Σ_ϵ such that $\|\Sigma - \Sigma_\epsilon\| < \epsilon$, and this Σ_ϵ is 2π times the spectral density function at zero of a finite order autoregressive process. Instead of estimating Σ directly, the autoregressive method is designed to estimate the approximation Σ_ϵ.

3.1. The Autoregressive Method

We consider the kth order vector autoregressive process $\{x(n) : n \geq 0\}$ which satisfies

$$\sum_{j=0}^{k} B(j)x(n-j) = \epsilon(n), \quad B(0) = I_d, \quad B(k) \neq O_d, \tag{3.1.1}$$

where the $\epsilon(n)$ are i.i.d. (identically and independently distributed) with mean zero and covariance matrix G and the $B(j)$ are square matrices. The mean zero assumption of $\epsilon(n)$ implies that $x(n)$ has mean zero. This is not always the case. If the mean μ of $x(n)$ is not zero, we make a mean correction (i.e., we replace $x(n)$ by $x(n) - \mu$). If all zeros of the

$$\det\Big(\sum_{j=0}^{k} B(j)z^j\Big)$$

lie outside the unit disc, then a solution of (3.1.1) exists (cf. Section 2.2) and is of the form

$$x(n) = \sum_{j=0}^{\infty} A(j)\epsilon(n-j), \quad A(0) = I_d.$$

Recall that the spectral density function g of $\{x(n) : n \geq 0\}$ is

$$g(\lambda) = \Big(\sum_{j=0}^{k} B(j)e^{-ij\lambda}\Big)^{-1} G \Big(\sum_{j=0}^{k} B(j)e^{-ij\lambda}\Big)^{*-1}.$$

To estimate the parameters $B(j)$ and G, it is natural to use the Yule-Walker equations, which are

$$\sum_{j=0}^{k} B(j)R(s-j) = \delta_0^s G, \quad s = 0, 1, 2, \ldots \tag{3.1.2}$$

Since there are $k+1$ unknown matrices, namely $B(1), \ldots, B(k)$, and G, we use the first $k+1$ equations of (3.1.2). By taking transpose on both sides and

using $R'(n) = R(-n)$ (cf. Section 2.1), we obtain

$$\sum_{j=0}^{k} B(j)R(j) = G,$$

$$\sum_{j=1}^{k} R(j-s)B'(j) = -R(-s), \quad s = 1, 2, \ldots, k.$$

Rewrite the equations in matrix form, we obtain

$$\begin{pmatrix} R(0) & R(1) & \ldots R(k-1) \\ R(-1) & R(0) & \ldots R(k-2) \\ \vdots & \vdots & \vdots \\ R(1-k) & R(2-k) & \ldots \quad R(0) \end{pmatrix} \begin{pmatrix} B'(1) \\ B'(2) \\ \vdots \\ B'(k) \end{pmatrix} = -\begin{pmatrix} R(-1) \\ R(-2) \\ \vdots \\ R(-k) \end{pmatrix}, \qquad (3.1.3)$$

denoted by $\mathcal{R}_k B_k = -r_k$.

Given a sample of size n, say $\{ x(i) : i = 0, \ldots, n-1 \}$, our estimation equations become

$$C_k \hat{B}_k = -c_k$$

where

$$C(m) = (n-m)^{-1} \sum_{j=1}^{n-m} x(m+j)x'(j),$$

replaces $R(m)$ to obtain C_k from $\mathcal{R}_k$, c_k from r_k, and $\hat{B}_k = (\hat{B}_k(1), \ldots, \hat{B}_k(k))'$ is an estimate of B_k. Since

$$G = \sum_{j=0}^{k} B(j)R(-j),$$

we estimate G by

$$\hat{G}_k = \sum_{j=0}^{k} \hat{B}_k(j)C(-j),$$

and $g(\lambda)$ by

$$\hat{g}_k(\lambda) = \left(\sum_{j=0}^{k} \hat{B}_k(j)e^{-ij\lambda} \right)^{-1} \hat{G}_k \left(\sum_{j=0}^{k} \hat{B}_k(j)e^{-ij\lambda} \right)^{*-1}.$$

The consistency of $\hat{B}_k(j)$ and $\hat{G}_k$ are justified by the following theorem:

(3.1.4) THEOREM. If $\{ x(n) : n = 0, 1, \ldots, n-1 \}$ is generated by (3.1.1), where $\epsilon(n)$ and $B(j)$ are as stated below the equation, then $\hat{B}_k$, C_k and $\hat{G}_k$ converge almost surely to B_k, $\mathcal{R}_k$, and G respectively as $n \to \infty$.

Proof. See Hannan (1970) pp. 329–332. ∎

An alternative approach for the estimation of parameters is to assume that the $\epsilon(n)$ are normally distributed and to use the method of maximum likelihood. This does not lead to the estimates just discussed. If n is relatively large, there will be little difference between the maximum likelihood estimates and those derived from Equation (3.1.3) (cf. Anderson (1971) pp. 183–186).

We assume that $\{ y(n) : n \geq 0 \}$ is generated by a k_0th order autoregressive process with parameters $B_{k_0} = (B_{k_0}(1), \ldots, B_{k_0}(k_0))'$ and G_{k_0}. If k_0 is known, we can obtain consistent estimates $\hat{B}_{k_0}$ and $\hat{G}_{k_0}$ from a realization. Unfortunately, the value k_0 is usually not known, therefore we must estimate the true order according to certain order selection rules. We shall study this topic in detail later. Now let us assume we have an order selection criterion and assume that there is a finite constant K, known a priori such that $k_0 \leq K < \infty$. From Theorem (3.1.4) above, it follows directly that $\hat{B}_k$ and $\hat{G}_k$ are consistent estimates of B_{k_0} and G_{k_0} respectively for all $k \geq k_0$. So, in order to estimate Σ which appeared in (3.0.1) we propose the following autoregressive method.

(3.1.4) ALGORITHM (The Autoregressive Method)

A1. [Which criterion?] Select a constant K which serves as the maximum order of the autoregressive model, and choose a criterion for order determination.

A2. [Parameters estimation.] Obtain the estimates $\hat{B}_k$ and $\hat{G}_k$ by fitting the observations $\{y(0), \ldots, y(n-1)\}$ to a kth order autoregressive process for $k = 0, 1, \ldots, K$.

A3. [Select order.] Determine the order according to the selected criterion. We denote this selection by $\hat{k}$.

A4. [Obtain $\hat{\Sigma}$.] Estimate Σ by the quantity

$$\hat{\Sigma} = \left(\sum_{j=0}^{\hat{k}} \hat{B}_{\hat{k}}(j)\right)^{-1} \hat{G}_{\hat{k}} \left(\sum_{j=0}^{\hat{k}} \hat{B}_{\hat{k}}(j)\right)'^{-1}. \quad \blacksquare$$

Justification for the autoregressive method appears in the following sections.

3.2 The Univariate Autoregressive Method

In this section we shall study the univariate case in detail. The variance constant appearing in (3.0.1) is simply a scalar, which we denote by σ^2. The autoregressive method yields an estimate, say s^2, which is an approximation of σ^2. We will examine several order selection criteria for the autoregressive process and derive asymptotic properties for the estimates.

Recall that a kth order autoregressive process $x = \{x(n) : n \geq 0\}$ satisfies

$$\sum_{j=0}^{k} \beta(j)x(n-j) = \epsilon(n), \quad \beta(0) = 1, \quad \beta(k) \neq 0 \tag{3.2.1}$$

where $\beta(j)$'s are constants and the $\epsilon(n)$ are uncorrelated random variables with $E\{\epsilon(n)\} = 0$ and variance $E\{\epsilon(n)^2\} = \sigma^2(\epsilon)$. The spectral density function f_x of x is given by

$$f_x(\lambda) = \frac{\sigma^2(\epsilon)}{2\pi|\sum_{j=0}^{k}\beta(j)e^{-ij\lambda}|^2}. \tag{3.2.2}$$

Let $y = \{y(n) : n \geq 0\}$ be a strictly stationary process which is ϕ–mixing with $\sum_{n=1}^{\infty}\phi(n)^{1/2} < \infty$. Let the spectral representation of the process y be

$$y(n) = \int_{-\pi}^{\pi} e^{in\lambda}z(d\lambda), \quad E\{|z(d\lambda)|^2\} = F(d\lambda).$$

First, we will show that we can find a finite order autoregressive process x which is close to the process y in the sense of mean square. From the discussion in Section 2.1, Equation (2.1.15), we know that the representation can be modified if F is absolutely continuous. In that case, if h is a Borel measurable function satisfying

$$|h(\lambda)|^2 = \frac{dF(\lambda)}{d\lambda} = f(\lambda), \tag{3.2.3}$$

then there is a z_1 process with orthogonal increments which satisfies

$$y(n) = \int_{-\pi}^{\pi} e^{in\lambda} h(\lambda) z_1(d\lambda), \quad E\{|z_1(d\lambda)|^2\} = d\lambda.$$

Since y is a real process and the series $\sum_{k=-\infty}^{\infty} R(k)$ converges absolutely (cf. Theorem (2.3.10)), f is real, non–negative, and continuous (cf. p. 23). Then we may take $h(\lambda) = \sqrt{f(\lambda)}$ which is the positive square root of $f(\lambda)$. From Corollary (2.4.7), we know that for given $\delta > 0$ there is a spectral density function (3.1.3), say $g(\lambda) = 1/|\sum_{j=0}^{k} \beta(j) e^{-ij\lambda}|^2$, such that $|f(\lambda) - g(\lambda)| < \delta$, $-\pi \leq \lambda \leq \pi$. Let $x = \{x(n) : n \geq 0\}$ be defined as follows:

$$x(n) = \int_{-\pi}^{\pi} e^{in\lambda} \sqrt{g(\lambda)} z_1(d\lambda).$$

Since g is real and non–negative for all λ, the positive square root of $g(\lambda)$, $\sqrt{g(\lambda)}$, exists thus $x(n)$ is well–defined. The covariance function of x is

$$\begin{aligned} E\{x(n+m)x^*(n)\} &= \int_{-\pi}^{\pi} e^{i(n+m)\lambda} \sqrt{g(\lambda)}\, e^{-in\lambda} \sqrt{g(\lambda)} d\lambda \\ &= \int_{-\pi}^{\pi} e^{im\lambda} g(\lambda) d\lambda, \end{aligned}$$

which depends only on m. This shows that x is a weakly stationary process. The processes x and y are close in the sense of mean square, since

$$\begin{aligned} E\{|x(n) - y(n)|^2\} &= E\{|\int_{-\pi}^{\pi} e^{in\lambda}(\sqrt{f(\lambda)} - \sqrt{g(\lambda)}) z_1(d\lambda)|^2\} \\ &= \int_{-\pi}^{\pi} |\sqrt{f(\lambda)} - \sqrt{g(\lambda)}|^2 d\lambda \\ &\leq \int_{-\pi}^{\pi} |\sqrt{f(\lambda)} - \sqrt{g(\lambda)}| \cdot |\sqrt{f(\lambda)} + \sqrt{g(\lambda)}| d\lambda \\ &\leq \int_{-\pi}^{\pi} |f(\lambda) - g(\lambda)| d\lambda \\ &\leq 2\pi\delta. \end{aligned}$$

Since $|1/\sum_{j=0}^{k}\beta(j)e^{-ij\lambda}|$ also satisfies (3.2.3), the spectral representation of x can also be written as

$$x(n) = \int_{-\pi}^{\pi} e^{in\lambda} \frac{1}{(\sum_{j=0}^{k}\beta(j)e^{-ij\lambda})} z_2(d\lambda), \quad E\{\,|z_2(d\lambda)|^2\,\} = d\lambda.$$

Define

$$\epsilon(n) = \sum_{j=0}^{k}\beta(j)x(n-j)/\beta(0).$$

Then

$$\epsilon(n) = \int_{-\pi}^{\pi} \frac{1}{\beta(0)} e^{in\lambda} z_2(d\lambda),$$

and

$$E\{\,\epsilon(n+m)\epsilon(n)\,\} = \int_{-\pi}^{\pi} \frac{1}{\beta^2(0)} e^{im\lambda} d\lambda = \begin{cases} 2\pi/\beta^2(0), & \text{if } m = 0; \\ 0, & \text{if } m \neq 0. \end{cases}$$

This shows that the sequence $\epsilon(n)$ are uncorrelated. Thus, x is a kth order autoregressive process.

We now turn to the problem of parameter estimation. The autoregressive method requires us to estimate $\beta_k(j)$ and $\sigma_k^2(\epsilon)$ for different values of k. As indicated in Section 3.1, we solve Equation (3.1.3) to obtain the parameters. Since we deal with scalar process, $R(-j) = R(j)$. Therefore, Equation (3.1.3) becomes

$$\begin{pmatrix} R(0) & R(1) & \dots & R(k-1) \\ R(1) & R(0) & \dots & R(k-2) \\ \vdots & \vdots & & \vdots \\ R(k-1) & R(k-2) & \dots & R(0) \end{pmatrix} \begin{pmatrix} \beta_k(1) \\ \beta_k(2) \\ \vdots \\ \beta_k(k) \end{pmatrix} = -\begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(k) \end{pmatrix}, \tag{3.2.4}$$

denoted by $\mathcal{R}_k\beta_k = -r_k$. The solution for (3.2.4) is $\beta_k = -\mathcal{R}_k^{-1}r_k$ and $\sigma_k^2(\epsilon) = \sum_{j=0}^{k}\beta_k(j)R(j)$. The equations for the coefficients of the $(k+1)$th

order autoregressive model are

$$\begin{pmatrix} \mathcal{R}_k & q_k \\ q_k' & R(0) \end{pmatrix} \begin{pmatrix} \beta_{k+1}^{(1)} \\ \beta_{k+1}(k+1) \end{pmatrix} = -\begin{pmatrix} r_k \\ R(k+1) \end{pmatrix}, \tag{3.2.5}$$

where $q_k = \big(R(k), \ldots, R(1)\big)'$ and $\beta_{k+1}^{(1)} = \big(\beta_{k+1}(1), \ldots, \beta_{k+1}(k)\big)'$. The partitioned equations of (3.2.5) are

$$\mathcal{R}_k \beta_{k+1}^{(1)} + q_k \beta_{k+1}(k+1) = -r_k, \tag{3.2.6}$$

$$q_k' \beta_{k+1}^{(1)} + R(0)\beta_{k+1}(k+1) = -R(k+1). \tag{3.2.7}$$

Elimination of $\beta_{k+1}^{(1)}$ from (3.2.6) and (3.2.7) yields

$$\beta_{k+1}(k+1) = \frac{q_k' \mathcal{R}_k^{-1} r_k - R(k+1)}{R(0) - q_k' \mathcal{R}_k^{-1} q_k}.$$

We observe that $\mathcal{R}_k^{-1} q_k = -\big(\beta_k(k), \beta_k(k-1), \ldots, \beta_k(1)\big)'$, therefore

$$R(0) - q_k' \mathcal{R}_k^{-1} q_k = \sum_{j=0}^{k} \beta_k(j) R(j) = \sigma_k^2(\epsilon).$$

Thus

$$\beta_{k+1}(k+1) = -\frac{\sum_{j=0}^{k} R(k+1-j)\beta_k(j)}{\sigma_k^2(\epsilon)}.$$

Substitution of $\beta_{k+1}(k+1)$ into (3.2.6) gives

$$\beta_{k+1}(j) = \beta_k(j) + \beta_k(k+1-j)\beta_{k+1}(k+1), \quad j = 1, 2, \ldots, k.$$

And

$$\begin{aligned} \sigma_{k+1}^2(\epsilon) &= \sum_{j=0}^{k+1} \beta_{k+1}(j) R(j) \\ &= R(0) + \sum_{j=1}^{k} \beta_k(j) R(j) + \beta_{k+1}(k+1) \Big(\sum_{j=1}^{k+1} \beta_k(k+1-j) R(j) \Big) \\ &= \sigma_k^2(\epsilon) + \beta_{k+1}(k+1) \Big(\sum_{j=0}^{k} R(k+1-j)\beta_k(j) \Big) \\ &= \sigma_k^2(\epsilon)\big(1 - \beta_{k+1}^2(k+1)\big). \end{aligned}$$

To summarize this method, we have the following recursive formulae: for each $k \geq 0$

$$\begin{aligned}
&\beta_{k+1}(k+1) = -\frac{\sum_{j=0}^{k} R(k+1-j)\beta_k(j)}{\sigma_k^2(\sigma)}, \\
&\beta_{k+1}(j) = \beta_k(j) + \beta_k(k+1-j)\beta_{k+1}(k+1), \quad j = 1, \ldots, k, \\
&\beta_{k+1}(0) = 1, \\
&\sigma_{k+1}^2(\epsilon) = \sigma_k^2(\epsilon) \cdot \left(1 - \beta_{k+1}^2(k+1)\right),
\end{aligned} \tag{3.2.8}$$

with initial conditions

$$\sigma_0^2(\epsilon) = R(0), \quad \beta_0(0) = 1. \tag{3.2.9}$$

This is a simple and efficient method to compute parameters $\beta_k(j)$ and $\sigma_k^2(\epsilon)$ (cf. Durbin (1960) pp. 139–153). Given observations $\{y(0), \ldots, y(n-1)\}$, we replace $R(n)$ by $C(n)$ in (3.2.8) and (3.2.9) to obtain estimates $\hat{\beta}_k$ and $\hat{\sigma}_k^2(\epsilon)$ for β_k and $\sigma_k^2(\epsilon)$ respectively.

The problem of determining the order for an autoregressive process has gained much attention during the past forty years. We could trace interest in this area to as early as the mid–1940's. Quenouille (1947) considered a method for determining the goodness–of–fit for autoregressive model. Later, Akaike (1969, 1970) introduced a method by using the concept of final prediction error (FPE). A few years later, Akaike (1974) proposed a new criterion called An Information Criterion (AIC) which has become very popular. Other criteria have been introduced since then are, for example, the BIC suggested by Akaike (1977), and Schwarz (1978), and the h criteria by Hannan and Quinn[†] (1979). In this paper, we shall consider the AIC, BIC, and h–criterion. We consider a stationary process, $\{x(n) : n \geq 0\}$, generated by

[†]In Hannan and Quinn's paper they used ϕ instead. We change it to h to avoid confusion with the ϕ–mixing condition for stationary processes.

(3.2.1). The estimate of the true order, which we shall call k_0, is obtained by minimizing one of the following quantities :

$$AIC(k) = n \log \hat{\sigma}_k^2(\epsilon) + 2k,$$
$$BIC(k) = n \log \hat{\sigma}_k^2(\epsilon) + k \log n,$$
$$h(k) = n \log \hat{\sigma}_k^2(\epsilon) + 2kc \log\log n, \quad c > 1,$$

where $\hat{\sigma}_k^2(\epsilon)$ are the estimates of $\sigma^2(\epsilon)$ obtained from the kth order autoregressive model based on a sample of size n[†]. We denote the selected order by $\hat{k}_n$; we will use $\hat{k}$ whenever there is no confusion.

We now state without proof certain theorems on the asymptotic properties of these estimates. The proofs may be found in Shibata (1976) and Hannan and Quinn (1979). We assume that

(C1) $\sum_{j=0}^{k} \beta(j) x^j \neq 0, \quad |x| \leq 1; \quad E\{\epsilon(m)\epsilon(n)\} = \delta_m^n \sigma^2(\epsilon)$.

(C2) $\{\epsilon(n) : n \geq 0\}$ *consists of independent random variables with the same normal distribution* $N(0, \sigma^2(\epsilon))$.

(C3) *The true order* k_0 *is bounded above by some constant* $K < \infty$ *which is known a priori.*

(3.2.10) THEOREM. Under the conditions described above the asymptotic distribution of $\hat{k}_n$ selected by AIC is given by

$$\lim_{n \to \infty} P(\hat{k}_n = k) = \begin{cases} p_{k-k_0} \cdot q_{K-k}, & k_0 \leq k \leq K; \\ 0, & \text{otherwise,} \end{cases}$$

where

$$p_n = \sum_{\substack{r_1 + 2r_2 + \cdots + nr_n = n \\ r_i \geq 0, \text{ integers}}} \left(\prod_{i=1}^{n} \frac{1}{r_i!} \left(\frac{\alpha_i}{i} \right)^{r_i} \right),$$

$$q_n = \sum_{\substack{r_1 + 2r_2 + \cdots + nr_n = n \\ r_i \geq 0, \text{ integers}}} \left(\prod_{i=1}^{n} \frac{1}{r_i!} \left(\frac{1 - \alpha_i}{i} \right)^{r_i} \right),$$

[†]Unless it is important to have the subscripts n, we shall drop them for convenience.

and $\alpha_i = P(\chi^2(i) > 2i)$, $i = 1, \ldots, K$. Here, $\chi^2(i)$ is a random variable having the Chi–square distribution with i degree of freedom.

Proof. See Shibata (1976) pp. 119–120. ∎

The next theorem holds under a weaker condition : we may replace (C2) by

(D2) $E\{\epsilon(n) \mid \mathcal{F}_{n-1}\} = 0$, $E\{\epsilon(n)^2 \mid \mathcal{F}_{n-1}\} = \sigma^2$, $E\{\epsilon(n)^4\} < \infty$, *where* $\mathcal{F}_n$ *is the* σ*–field generated by* $y(m)$, $m \leq n$.

(3.2.11) THEOREM. Under the conditions (C1), (D2), (C3) and $E\{|\epsilon(n)|^r\} < \infty$, for some $r > 4$, the estimates $\hat{k}_n$ obtained via $BIC(k)$ or $h(k)$ are strongly consistent.

Proof. See Hannan and Quinn (1979) pp. 192–193. ∎

Besides the order selection criteria discussed above, one may resort to a statistical test(cf. Anderson (1971) p. 215, Fishman (1978) p. 251). We test the null hypothesis H_0: autoregressive processes of order $k < K$ against the alternative H_1: autoregressive process of order K. The statistic for testing is

$$T_{K-k} = n\left(1 - \frac{\hat{\sigma}_K^2(\epsilon)}{\hat{\sigma}_k^2(\epsilon)}\right)$$

which has a limiting χ^2–distribution as $n \to \infty$ with $K-k$ degrees of freedom when the null hypothesis is true. We reject the null hypothesis if

$$T_{K-k} > \chi^2_{1-\alpha}(K-k)$$

where $\chi^2_{1-\alpha}(K-k)$ is the $(1-\alpha)$th quantile of the χ^2–distribution with $K-k$ degrees of freedom. We select the order $\hat{k}$ to be the first k, $k = 0, 1, \ldots, K$,

such that H_0 is accepted. For sufficiently large n, this test has approximate significance level α. Other types of statistical test will not be considered in this paper; the interested reader is referred to Anderson (1971), Quenouille (1947), Grenander and Rosenblatt (1957).

The variance constant obtained by the autoregressive method is

$$\hat{s}_{\hat{k}}^2 = \frac{1}{2\pi} \frac{1}{|\sum_{j=0}^{\hat{k}} \hat{\beta}_{\hat{k}}(j)|^2}.$$

To prove the consistency of $\hat{s}_{\hat{k}}^2$, let us define the K–dimensional vectors β, $\hat{\beta}_k$ as follows:

$$\begin{aligned} \beta' &= (\beta(1), \beta(2), \ldots, \beta(k_0), 0, \ldots, 0), \\ \hat{\beta}_k' &= (\hat{\beta}_k(1), \hat{\beta}_k(2), \ldots, \hat{\beta}_k(k), 0, \ldots, 0), \quad k = 1, \ldots, K, \end{aligned}$$

where $\hat{\beta}_k(j)'s$ are the estimates of the coefficients of the kth order autoregressive process. It should be noted that $\hat{\beta}_k$ and $\hat{\sigma}_k^2(\epsilon)$ are consistent for $k \geq k_0$ (cf. Anderson (1970) pp. 188–200). Let

$$S(x, y) = \frac{y}{(1 + x'e)^2}, \quad e' = (1, \ldots, 1) \in \Re^K, x \in \Re^K, y \in \Re.$$

We observe that

$$s^2 = \frac{\sigma^2(\epsilon)}{2\pi |\sum_{j=0}^{k_0} \beta(j)|^2} = \frac{1}{2\pi} S(\hat{\beta}, \sigma^2(\epsilon)),$$

and $\hat{s}_k^2 = S(\hat{\beta}_k, \hat{\sigma}_k^2(\epsilon))$ is an estimate of s^2 obtained from the kth order autoregressive model. Since S is a continuous function, by using a continuous mapping argument (cf. Billingsley (1968) pp. 30–31), we know that $\hat{s}_k^2$ is a consistent estimate of s^2 for $k \geq k_0$. If $\hat{k}$ is obtained by minimizing $BIC(k)$ or $h(k)$, the consistency of $\hat{s}_{\hat{k}}^2$ follows immediately from the strong consistency

of the estimate $\hat{k}$. If $\hat{k}$ is obtained via minimizing $AIC(k)$, then for any $\delta > 0$, $\epsilon > 0$

$$\begin{aligned} P\{|\hat{s}_k^2 - s^2| > \delta\} &= \sum_{k=0}^{K} P\{|\hat{s}_k^2 - s^2| > \delta, \hat{k} = k\} \\ &= \sum_{k=0}^{K} P\{|\hat{s}_k^2 - s^2| > \delta \mid \hat{k} = k\} \cdot P\{\hat{k} = k\} \\ &= \sum_{k=1}^{K} P\{|\hat{s}_k^2 - s^2| > \delta\} \cdot P\{\hat{k} = k\}. \end{aligned}$$

From Theorem (3.2.10) and the consistency of $\hat{s}_k^2$ for $k \geq k_0$, for sufficiently large sample size M we have

$$\sum_{k<k_0} P\{\hat{k} = k\} < \frac{\epsilon}{2}, \quad \text{and}$$

$$P\{|\hat{s}_k^2 - s^2| > \delta\} < \frac{\epsilon}{2}, \quad k_0 \leq k \leq K.$$

Thus

$$\begin{aligned} P\{|s_k^2 - s^2| > \delta\} &\leq \frac{\epsilon}{2} \sum_{k \geq k_0}^{K} P\{\hat{k} = k\} + \frac{\epsilon}{2} \\ &< \epsilon \end{aligned}$$

This shows that $\hat{s}_k^2$ is a consistent estimate of s^2.

3.3. The Multidimensional Autoregressive Method

For the vector case, we first derive recursive formulae for obtaining parameters of the kth order autoregressive model, B_k and G_k (cf. Section 3.1). To solve

$$\mathcal{R}_k B_k = -r_k \tag{3.3.1}$$

recursively requires an auxiliary equation (cf. Lee (1980))

$$\begin{pmatrix} R(0) & R(1) & \dots R(k-1) \\ R(-1) & R(0) & \dots R(k-2) \\ \vdots & \vdots & \vdots \\ R(1-k) & R(2-k) & \dots \quad R(0) \end{pmatrix} \begin{pmatrix} A_k'(k) \\ A_k'(k-1) \\ \vdots \\ A_k'(1) \end{pmatrix} = -\begin{pmatrix} R(k) \\ R(k-1) \\ \vdots \\ R(1) \end{pmatrix}, \tag{3.3.2}$$

denoted by $\mathcal{R}_k A_k = -q_k$. Note that Equation (3.3.2) is equivalent to

$$\sum_{j=0}^{k} A_k(j) R(s-k+j) = O_d, \quad s = 0, 1, \dots, k-1, \tag{3.3.3}$$

with $A_k(0) = I_d$. If we define

$$\begin{aligned} \xi_k(n) &= A_k(k)x(n) + \cdots + A_k(1)x(n-k+1) + A_k(0)x(n-k) \\ &= \sum_{j=0}^{k} A_k(j)x(n-k+j), \end{aligned}$$

then Equation (3.3.3) implies

$$E\{\, \xi_k(n)x'(n-s)\,\} = O_d, \quad s = 0, 1, \dots, k-1. \tag{3.3.4}$$

Let H_k denote the covariance matrix of $\xi_k(n)$, then

$$\begin{aligned} H_k &= E\{\, \xi_k(n)\xi_k'(n)\,\} = E\{ \sum_{j=0}^{k} A_k(j)x(n-k+j)\xi_k'(n)\,\} \\ &= E\{\, x(n-k)\xi_k'(n)\,\} \qquad \text{(by (3.3.4))} \\ &= \sum_{j=0}^{k} R(-j)A_k'(j). \end{aligned}$$

Equations (3.3.1) and (3.3.2) can be solved in an efficient recursive manner (cf. Kailath (1974), Wiggins and Robinson (1965)). The equations for the coefficients of autoregressive model of order $(k+1)$ are

$$\begin{pmatrix} \mathcal{R}_k & q_k \\ q_k' & R(0) \end{pmatrix} \begin{pmatrix} B_{k+1}^{(1)} \\ B_{k+1}'(k+1) \end{pmatrix} = -\begin{pmatrix} r_k \\ R(-k-1) \end{pmatrix} \tag{3.3.5}$$

and

$$\begin{pmatrix} R(0) & r_k' \\ r_k & \mathcal{R}_k \end{pmatrix} \begin{pmatrix} A_{k+1}'(k+1) \\ A_{k+1}^{(1)} \end{pmatrix} = -\begin{pmatrix} R(k+1) \\ q_k \end{pmatrix} \tag{3.3.6}$$

where

$$B_{k+1}^{(1)} = \big(B_{k+1}(1), B_{k+1}(2), \ldots, B_{k+1}(k)\big)',$$
$$A_{k+1}^{(1)} = \big(A_{k+1}(k), A_{k+1}(k-1), \ldots, A_{k+1}(1)\big)'.$$

The partitioned equations for (3.3.5) and (3.3.6) are

$$\mathcal{R}_k B_{k+1}^{(1)} + q_k B_{k+1}'(k+1) = -r_k, \tag{3.3.7}$$
$$q_k' B_{k+1}^{(1)} + R(0) B_{k+1}'(k+1) = -R(-k-1), \tag{3.3.8}$$

and

$$R(0) A_{k+1}'(k+1) + r_k' A_{k+1}^{(1)} = -R(k+1), \tag{3.3.9}$$
$$r_k A_{k+1}'(k+1) + \mathcal{R}_k A_{k+1}^{(1)} = -q_k. \tag{3.3.10}$$

The solutions for (3.3.1) and (3.3.2) are $B_k = -\mathcal{R}_k^{-1} r_k$ and $A_k = -\mathcal{R}_k^{-1} q_k$ respectively. Substituting them into Equations (3.3.7)—(3.3.10), we obtain

$$B_{k+1}^{(1)} = B_k + A_k B_{k+1}'(k+1),$$
$$B_{k+1}'(k+1) = -\big(q_k' A_k + R(0)\big)^{-1}\big(R(-k-1) + q_k' B_k\big),$$
$$A_{k+1}^{(1)} = A_k + B_k A_{k+1}'(k+1),$$
$$A_{k+1}'(k+1) = -\big(R(0) + r_k' B_k\big)^{-1}\big(R(k+1) + r_k' A_k\big),$$

and $$G_k = \sum_{j=0}^{k} B_k(j) R(-j) = R(0) + r_k' B_k,$$

$$H_k = \sum_{j=0}^{k} A_k(j) R(j) = R(0) + q_k' A_k.$$

Note that H_k is the covariance matrix of $\xi_k(n)$, hence is symmetric and positive definite.

We recursively define

$$\begin{aligned} \Delta_{k+1} &= \sum_{j=0}^{k} B_k(j) R(k+1-j) \\ \rho_{k+1} &= G_k^{-1/2} \Delta_{k+1} \big(H_k^{-1/2}\big)', \end{aligned} \tag{3.3.11}$$

with initial conditions

$$G_0 = H_0 = R(0), \qquad A_k(0) = B_k(0) = I_d. \tag{3.3.12}$$

Then

$$\begin{aligned} B_{k+1} &= \begin{pmatrix} B_k \\ 0_d \end{pmatrix} - \begin{pmatrix} A_k \\ I \end{pmatrix} H_k^{-1} \Delta_{k+1}', \\ G_{k+1} &= G_k + \Delta_{k+1} B_{k+1}'(k+1) \\ &= G_k \big(I - \rho_{k+1} \rho_{k+1}'\big), \end{aligned} \tag{3.3.13}$$

$$\begin{aligned} A_{k+1} &= \begin{pmatrix} 0_d \\ A_k \end{pmatrix} - \begin{pmatrix} I \\ B_k \end{pmatrix} G_k^{-1} \Delta_{k+1}, \\ H_{k+1} &= H_k + A_{k+1}(k+1) \Delta_{k+1} \\ &= H_k \big(I - \rho_{k+1}' \rho_{k+1}\big), \end{aligned} \tag{3.3.14}$$

Equations (3.3.11)—(3.3.14) give a simple way to compute the parameters of a given order autoregressive process.

We assume that $\{ y(n) : n \geq 0 \}$ is generated by a k_0th order autoregressive process with parameters $B_{k_0} = \big(B_{k_0}(1), \ldots, B_{k_0}(k_0)\big)$ and G_{k_0}. Given a sample of size n, we use Equations (3.3.11)–(3.3.14) to compute the parameters for different orders and we adopt Akaike's criterion (cf. Akaike (1974)) to select the order. As in the scalar case, we assume $k_0 \leq K < \infty$, where K is known a priori. We select the order $\hat{k}_n$ which minimizes

$$AIC(k) = n \log|\hat{G}_k| + 2kd^2,$$

where $|\cdot|$ denotes the determinant of a matrix and $\hat{G}_k$ is the estimate of G_{k_0} obtained from the kth order autoregressive process. We would like to establish some asymptotic properties for $\hat{k}_n$. Before we can proceed we need to prove the following lemma.

(3.3.15) LEMMA. If the matrix $I - AA'$ is positive definite, then

$$0 < |I - AA'| < 1 \quad \text{if } A \neq O_d.$$

Proof. Let λ be an eigenvalue of $I - AA'$ and x be the corresponding eigenvector, then

$$(I - AA')x = \lambda x.$$

Multiply both sides by x' to obtain

$$(1 - \lambda)x'x = x'AA'x \geq 0.$$

Because x is an eigenvector, x is not the zero vector. This implies $\lambda \leq 1$, and since every eigenvalue of a positive definite matrix is positive we have $0 < \lambda \leq 1$. It is well known that

$$|I - AA'| = \prod_i \lambda_i,$$

where λ_i are the eigenvalues of $I - AA'$. If $|I - AA'| = 1$ then we must have $\lambda_i = 1$ for all i, and this means $I - AA' = I$ or $A = O_d$. ∎

(3.3.16) THEOREM. The asymptotic distribution of $\hat{k}_n$ has the following property:

$$\lim_{n \to \infty} P\{\hat{k}_n = k\} = 0, \qquad \text{if } k < k_0.$$

Proof. Using Equation (3.3.13) we obtain

$$\begin{aligned}|\hat{G}_{k+1}| &= |\hat{G}_k| \cdot |I - \hat{\rho}_{k+1}\hat{\rho}'_{k+1}| \\ &= |\hat{G}_0| \cdot |(I - \hat{\rho}_1\hat{\rho}'_1)| \cdots |(I - \hat{\rho}_{k+1}\hat{\rho}'_{k+1})|,\end{aligned}$$

and from the consistency of the estimates for parameters of the autoregressive process of order k_{k_0}, we have in probability

$$\lim_{n\to\infty} \hat{\rho}_{k_0}\hat{\rho}'_{k_0} = \rho_{k_0}\rho'_{k_0} = G_{k_0-1}^{-1/2}\Delta_{k_0}B'_{k_0}(k_0)G_{k_0-1}^{-1/2}.$$

Obviously, $G_{k_0-1} \neq O_d$ and since the true order is k_0, we know that $B_{k_0}(k_0) \neq O_d$. Also note that $B_{k_0}(k_0) = -H_{k_0-1}^{-1}\Delta_{k_0}$, this implies $\Delta_{k_0} \neq O_d$. So, $\lim_{n\to\infty} \hat{\rho}_{k_0}\hat{\rho}'_{k_0} \neq O_d$ in probability.

For $0 \leq k < k_0$, we have

$$\frac{|\hat{G}_k|}{|\hat{G}_{k_0}|} = \Big(\prod_{i=k+1}^{k_0} |I - \hat{\rho}_i\hat{\rho}'_i|\Big)^{-1} \geq \frac{1}{|I - \hat{\rho}_{k_0}\hat{\rho}'_{k_0}|} > 1.$$

Since every $I - \hat{\rho}_i\hat{\rho}'_i$ is positive definite, by Lemma (3.3.15) we know $0 < |I - \hat{\rho}_i\hat{\rho}'_i| < 1$. Therefore, for any $\epsilon > 0$, there exists a $\delta > 0$ and an integer $M > 0$, such that for any $n \geq M$

$$\exp\big(2(k_0 - k)d^2/n\big) < 1 + \delta,$$

$$P\{\,|\hat{G}_k|/|\hat{G}_{k_0}| \leq 1 + \delta\,\} < \epsilon.$$

From the definition of $\hat{k}$, we have

$$\begin{aligned}P\{\,\hat{k} = k\,\} &= P\{\,AIC(k) \leq AIC(m), 0 \leq m \leq K\,\} \\ &\leq P\{\,AIC(k) \leq AIC(k_0)\,\} \\ &= P\{\,|\hat{G}_k|/|\hat{G}_{k_0}| \leq \exp\big(2(k_0 - k)d^2/n\big)\,\} \\ &< \epsilon,\end{aligned}$$

this proves the theorem. ∎

The estimate of the covariance matrix Σ appearing in (3.0.1) is given by

$$\hat{\Sigma} = \Big(\sum_{j=0}^{\hat{k}} \hat{B}_{\hat{k}}(j)\Big)^{-1} \hat{G}_{\hat{k}} \Big(\sum_{j=0}^{\hat{k}} \hat{B}_{\hat{k}}(j)\Big)^{*-1}.$$

The consistency of $\hat{\Sigma}$ can be shown in a manner analogous to that used in the last paragraph of Section 3.2 by properly modifying the function S.

3.4. Applications to Markov Processes

We now apply the results derived in previous sections to Markov processes. Let $(E, \mathcal{E})$ be a measurable space.

(3.4.1) DEFINITION. A function $P : (E, \mathcal{E}) \to [0, 1]$ is said to be a probability transition function if :

(a) for each $x \in E, P(x, \cdot)$ is a probability measure on $(E, \mathcal{E})$,

(b) for each $B \in \mathcal{E}, P(\cdot, B)$ is a measurable function with respect to $\mathcal{E}$.

The n–step probability transition functions are defined by setting $P^1(x, B) = P(x, B)$ and

$$P^{n+1}(x, B) = \int_E P^n(y, B) P(x, dy).$$

Let $E^\infty = E \times E \times \cdots$ and $\mathcal{E}^\infty = \mathcal{E} \times \mathcal{E} \times \cdots$. For any $\omega = (\omega_0, \omega_1, \ldots) \in E^\infty$, let $X(i)(\omega) = \omega_i$. Then given P and an initial probability distribution μ, there is a probability measure P_μ on $(E^\infty, \mathcal{E}^\infty)$ such that for all $n > 0$ and $B_0, B_1, \ldots, B_n \in \mathcal{E}$

$$P_\mu\{X(0) \in B_0, X(1) \in B_1, \ldots, X(n) \in B_n\} = \int_{B_0} \mu(dx_0) \int_{B_1} P(x_0, dx_1) \cdots \int_{B_n} P(x_{n-1}, dx_n).$$

It can be shown that

$$P_\mu\{X(n+1) \in B \mid X(0), \ldots, X(n)\} = P_\mu\{X(n+1) \in B \mid X(n)\}. \quad (3.4.2)$$

Equation (3.4.2) is called the Markov property and $\{X(n) : n \geq 0\}$ is said to be a Markov process with state space E, initial distribution μ and stationary probability transition function P.

For our simulation studies we assume that $X(n) \Rightarrow X$ as $n \to \infty$ where X has stationary distribution π. It is known that if we take the initial

distribution to be π then the process $\{X(n) : n \geq 0\}$ is a strictly stationary process. If we make the further assumption that $\{X(n) : n \geq 0\}$ is ϕ–mixing with $\sum_n \phi(n)^{1/2} < \infty$, then Theorem (2.3.10) holds. Therefore, we can apply the autoregressive method to this stationary process and obtain reasonable estimates about the steady state quantities. But usually we do not know the stationary distribution π in real situation, otherwise we could compute the results analytically and would have no need to simulate. So we choose $X(0)$ according to an initial distribution μ, and simulate the process. This procedure yields a non–stationary process. Since this non–stationary process converges in distribution to a stationary random variable, the process is asymptotically stationary. In order to apply the autoregressive method and obtain consistent estimates, we need to justify the method for non–stationary processes.

Let $\{X(n) : n \geq 0\}$ be a Markov process with initial distribution π which satisfies the regularity conditions for Theorem (2.3.10). Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n-1}(X(i)-\mu)\right) \Rightarrow N(0,\Sigma).$$

The autoregressive method says that for $\epsilon > 0$, there exists a Σ_ϵ such that $\|\Sigma - \Sigma_\epsilon\| < \epsilon$ and that from a sample of size n we can obtain a consistent estimate $\hat{\Sigma}_n$ for Σ_ϵ under the stationary distribution π. Now, let μ be any initial distribution which is absolutely continuous with respect to π, then the Radon–Nikodym derivative $\frac{d\mu}{d\pi}$ exists (cf. Halmos 1950). We also assume that there is a constant C, such that $\frac{d\mu}{d\pi} \leq C < \infty$. For any $\delta > 0$, let

$$A_n(\delta) = \{\omega \in E : \|\hat{\Sigma}_n(\omega) - \Sigma_\epsilon\| > \delta\},$$

then $\lim_{n\to\infty} P_\pi\{A_n(\delta)\} = 0$. Thus

$$\begin{aligned}\lim_{n\to\infty} P_\mu\{A_n(\delta)\} &= \lim_{n\to\infty}\int_E P_x\{A_n(\delta)\}\mu(dx)\\ &= \lim_{n\to\infty}\int_E P_x\{A_n(\delta)\}\frac{d\mu}{d\pi}\pi(dx)\\ &\le \lim_{n\to\infty} C\cdot P_\pi\{A_n(\delta)\} = 0.\end{aligned}$$

So, $\hat{\Sigma}_n$ also converges in probability to Σ_ϵ under P_μ. Therefore, we have found some conditions which allow us to apply the autoregressive method to simulation output data.

Remark 1. If $\{X(n) : n \ge 0\}$ is a finite state space irreducible Markov chain then $\frac{d\mu}{d\pi} \le C$ always holds for any μ. For countable state space, $\frac{d\mu}{d\pi} \le C$ holds if μ has finite support.

Remark 2. If π has an atom x, then $\mu = \delta_x$ satisfies $\frac{d\mu}{d\pi} \le C$.

It is often the case that we need to study the steady-state behavior of a continuous time Markov chain. Since the process is simulated in continuous time, we need a technique which converts the continuous time process to a discrete time series in order to apply the autoregressive method. This can be done by either sampling the continuous time Markov chain (then we need to consider the problem of what can be inferred about the full process from the sample) or by using the discrete time method (cf. Hordijk, Iglehart, and Schassberger (1976)). We shall briefly review this method below.

Let $\{X(t) : t \ge 0\}$ be a continuous time Markov chain with countable state space E. Assume the Markov chain is irreducible and positive recurrent. Then $X(t) \Rightarrow X$ as $t \to \infty$. Let the probability transition function be

$$p_{ij}(t) = P\{X(t) = j \mid X(0) = i\},$$

and let

$$q_{ij} = \frac{d}{dt} p_{ij}(t) \mid_{t=0},$$

$Q = \{ q_{ij} : i, j \in E \}$, and $q_i = -q_{ii}$ (assume $0 < q_i < \infty$). The quantities q_{ij}, $i, j \in E$, are called the infinitesimal parameters of the process. Let $\pi = \{ \pi_i : i \in E \}$ be the stationary distribution of the process. Then $\pi' Q = 0$, or equivalently

$$\sum_{i \in E} \pi_i q_{ij} = 0, \quad j \in E.$$

Define the matrix $R = \{ r_{ij} : i, j \in E \}$ by

$$r_{ij} = \begin{cases} 0, & \text{if } i = j; \\ q_{ij}/q_i, & \text{if } i \neq j. \end{cases} \tag{3.4.4}$$

Let $f : E \to \Re$. We are interested in estimating

$$r = E\{ f(X) \} = \sum_{i \in E} f(i) \pi_i.$$

We use the embedded jump chain $\{ X(n) : n \geq 0 \}$ which is a Markov chain with probability transition function R. Let $X(0)$ be choosen according to the stationary distribution p of the embedded jump chain. We define a new function g for $\{ X(n) : n \geq 0 \}$ as follows :

$$g(i) = f(i) q_i^{-1}, \qquad i \in E.$$

Define

$$\begin{aligned} u(n) &= g(X(n)), \\ v(n) &= q_{X(n)}^{-1}, \\ w(n) &= u(n) - r \cdot v(n). \end{aligned} \tag{3.4.5}$$

Since $\{ X(n) : n \geq 0 \}$ is strictly stationary, so are the processes $u = \{ u(n) : n \geq 0 \}$, $v = \{ v(n) : n \geq 0 \}$, and $w = \{ w(n) : n \geq 0 \}$. Note that

$p'R = p'$, or we may write

$$\sum_{i \in E} p_i r_{ij} = p_j, \quad \text{for all } j \in E,$$

which implies

$$\sum_{i \in E} p_i q_i^{-1} q_{ij} = 0.$$

Since the stationary distribution is unique we must have $p_i q_i^{-1} = c\pi_i$ for some constant c. Therefore,

$$\begin{aligned} \frac{E\{u(0)\}}{E\{v(0)\}} &= \frac{E\{f(X(0))q_{X(0)}^{-1}\}}{E\{q_{X(0)}^{-1}\}} \\ &= \frac{\sum_{i \in E} f(i) q_i^{-1} p_i}{\sum_{j \in E} q_j^{-1} p_j} \\ &= \sum_{i \in E} f(i)\pi_i = r. \end{aligned} \tag{3.4.6}$$

Thus $w(n)$ has mean zero. Applying Theorem (2.3.10) to process w yields, as $n \to \infty$,

$$\frac{\sqrt{n}\bar{w}}{\sigma} = \frac{\sqrt{n}\big((\bar{u}/\bar{v}) - r\big)}{\sigma/\bar{v}} \Rightarrow N(0,1). \tag{3.4.7}$$

By using a continuous mapping argument we may replace $\bar{v}$ by $E\{v(0)\}$ and Equation (3.4.7) becomes

$$\frac{\sqrt{n}(\bar{u}/\bar{v} - r)}{\sigma/E\{v(0)\}} \Rightarrow N(0,1).$$

The variance constant σ^2 is

$$\sigma^2 = \sum_{k=-\infty}^{\infty} R_w(k), \tag{3.4.8}$$

and by the definition of w, we have

$$R_w(k) = R_u(k) - 2rR_{uv}(k) + r^2 R_v(k).$$

Therefore

$$\sigma^2 = \sum_{k=-\infty}^{\infty} R_u(k) - 2r \sum_{k=-\infty}^{\infty} R_{uv}(k) + r^2 \sum_{k=-\infty}^{\infty} R_v(k). \qquad (3.4.9)$$

Here, $R_w(k)$, $R_u(k)$, and $R_v(k)$ are the covariance functions of the processes w, u, and v respectively, and $R_{uv}(k)$ is the cross–covariance function of the processes u and v. Thus, we have two alternatives to implement the discrete time method. The first alternative is observe the processes u and v, form $\{w(n) : n \geq 0\}$ by replacing $\hat{r}_n$ for r then apply the univariate autoregressive method and use Equation (3.4.8). The other alternative is to use the vector process $y = \{y(n) : n \geq 0\}$ defined by $y(n) = \big(u(n), v(n)\big)'$ directly and obtain an estimate of σ^2 by using Equation (3.4.9) through the spectral density function of y.

To apply the autoregressive method to semi–Markov processes simulated in continuous time, we first apply the discrete time method (cf. Hordijk, Iglehart, and Schassberger (1976)). Then we either use Equation (3.4.7) or (3.4.8) to calculate σ^2.

CHAPTER IV

VARIANCE REDUCTION TECHNIQUES

Although simulation is an important tool for analyzing stochastic systems, in many practical applications considerable computer time is required for simulation runs. Therefore, it is desirable to develop methods that allow us to obtain, based on the same realizations, improved statistical accuracy. Such methods are called variance reduction techniques. We shall develop several variance reduction techniques and incorporate them into our autoregressive method.

Throughout this chapter, we let $\{ X(n) : n \geq 0 \}$ be a strictly stationary process. Since $X(n)$ has the same distribution for all n, it holds trivially that $X(n) \Rightarrow X$ as $n \to \infty$. The quantity of interest is

$$r = E\{ f(X) \}, \tag{4.0.1}$$

where $f : E \to \Re$ is a real-valued measurable function. We make the further assumption that $\{ X(n) : n \geq 0 \}$ is ϕ-mixing with $\sum_n \phi(n)^{1/2} < \infty$. Then

$$\hat{r}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(X(i))$$

is a strongly consistent unbiased estimate for r and the following central limit theorem holds for $\hat{r}(n)$ (cf. Section 2.33) :

$$\frac{\sqrt{n}\,(\hat{r}_n - r)}{\sigma} \Rightarrow N(0,1), \tag{4.0.2}$$

where

$$\sigma^2 = Var\{ f(X(0)) \} + 2 \sum_{i=1}^{\infty} Cov\{ f(X(i)), f(X(0)) \}.$$

Our goal is to find another strongly consistent estimate for r and a central limit theorem analogous to (4.0.2) with a smaller variance constant σ^2. The motivation for doing so is to be able to form a shorter confidence interval for r.

4.1. Control Variables Method

In this section we will discuss the use of control variables to achieve variance reduction in the simulation and hence obtain a shorter confidence interval. A good introduction to this technique is given in the book by Gaver and Thompson (1973) pp. 582–591. Detailed accounts of various kinds of control variables applications can be found in Iglehart and Lewis (1979), Lavenberg, Moeller, and Welch (1977), and Gaver and Shedler (1971). A control variable is a random variable whose expectation is known and which is correlated with the process under study.

Let a sequence of processes $\{ C_j(n) : n \geq 0 \}$, $j = 1, 2, \ldots, k$, which will serve as the control variables, having the following properties:

(P1) They are fairly easy to obtain; i.e., we do not spend too much time generating them.

(P2) They are correlated with the original process $\{ X(n) : n \geq 0 \}$.

(P3) The mean $E\{ C_j(n) \} = \mu_j$ is known or can be calculated analytically.

Define a $(k+1)$ dimensional column vector

$$\begin{aligned} y(n) &= (y_0(n), y_1(n), \ldots, y_k(n))' \\ &= (f(X(n)), C_1(n), \ldots, C_k(n))', \quad n \geq 0, \end{aligned}$$

and assume that the process $\{y(n) : n \geq 0\}$ is strictly stationary and ϕ–mixing with $\sum_n \phi(n)^{1/2} < \infty$. Then we have the following results (cf. Section 2.3):

$$\hat{y}_n = \frac{1}{n}\sum_{i=0}^{n-1} y(i) \to \mu \qquad \text{a.s.} \qquad \text{and}$$

$$\sqrt{n}(\hat{y}_n - \mu) \Rightarrow N(0, \Sigma), \tag{4.1.1}$$

where

$$\mu = (r, \mu_1, \ldots, \mu_k)',$$
$$\Sigma = (\sigma_{ij}), \quad i, j = 0, 1, \ldots, k,$$

and σ_{ij} is defined by Equation (2.3.12). Now let β be a $(k+1)$ dimensional column vector of real numbers. If we take $\beta = (1, \beta_1, \ldots, \beta_k)'$ and form $y_\beta(n) = f\big((X(n))\big) + \sum_{j=1}^{k} \beta_j\big(C_j(n) - \mu_j\big)$, then

$$\hat{r}_\beta(n) = \frac{1}{n}\sum_{i=0}^{n-1}\left(f\big(X(i)\big) + \sum_{j=1}^{k}\beta_j\big(C_j(i) - \mu_j\big)\right)$$

is an estimate of r. We have $\hat{r}_\beta(n) \to r$ a.s., and a simple application of the continuous mapping theorem (cf. Billingsley (1968)) yields:

$$\frac{\sqrt{n}(\beta'\hat{y}_n - \beta'\mu)}{\sigma_\beta} \Rightarrow N(0, 1), \tag{4.1.2}$$

where $\sigma_\beta^2 = \beta'\Sigma\beta$. Note that Equation (4.1.2) can be written as

$$\frac{\sqrt{n}\,\big(\hat{r}_\beta(n) - r\big)}{\sigma_\beta} \Rightarrow N(0, 1). \tag{4.1.3}$$

Since there is no restriction in selecting the β_j's, $j = 1, \ldots, k$, we pick $\beta = \beta^* = (1, \beta_1^*, \ldots, \beta_k^*)'$ where β^* minimizes σ_β^2. This will produce the smallest possible confidence interval for r. To minimize σ_β^2 we need to solve

the following non–linear programming problem:

$$\begin{aligned} &\text{minimize} && \sigma_\beta^2 = \beta' \Sigma \beta \\ &\text{subject to} && \beta_0 = 1. \end{aligned} \tag{4.1.4}$$

Let $b = (\beta_1, \ldots, \beta_k)'$, $A = \{\sigma_{ij}\}$, $i, j = 1, \ldots, k$ and $a = (\sigma_{01}, \ldots, \sigma_{0k})'$, then

$$\begin{aligned} \sigma_\beta^2 &= (1, b') \begin{pmatrix} \sigma_{10} & a' \\ a & A \end{pmatrix} \begin{pmatrix} 1 \\ b \end{pmatrix} \\ &= \sigma_{00} + 2a'b + b'Ab. \end{aligned}$$

Therefore, the problem (4.1.4) becomes

$$\text{minimize} \qquad b'Ab + 2a'b. \tag{4.1.5}$$

It is easily seen that the optimal solution for (4.1.5) is $b^* = -A^{-1}a$. We denote the corresponding β by β^*, then

$$\begin{aligned} \sigma_{\beta^*}^2 &= \sigma_{00} - 2a'A^{-1}a + (A^{-1}a)'A(A^{-1}a) \\ &= \sigma_{00} - a'A^{-1}a. \end{aligned}$$

Since A is the covariance matrix of the control variables, it is positive definite and thus so is A^{-1}. Hence, we reduce the variance by a positive amount $a'A^{-1}a$. Since the covariance matrix is not known, it becomes necessary to estimate Σ. If $\hat{\Sigma}_n$ is a strong estimate of Σ, then $\hat{a}_n \to a$ a.s. and $\hat{A}_n \to A$ a.s. as $n \to \infty$. Hence, $\hat{A}_n^{-1} \to A^{-1}$ a.s.. Let $\hat{b}_n^* = -\hat{A}_n^{-1}\hat{a}_n'$, it is clear that $\hat{b}_n^* \to b$ a.s. as $n \to \infty$. Then $\hat{r}_{\hat{\beta}^*}(n)$ and $\hat{\sigma}_{\hat{\beta}^*}$ are strongly consistent estimates of r and σ_{β^*} respectively.

It follows that for $0 < \gamma < 1/2$, the $100(1-2\gamma)\%$ confidence interval for r is

$$\hat{I}(n) = \left[\hat{r}_{\hat{\beta}^*}(n) - z_{1-\gamma}\hat{\sigma}_{\hat{\beta}^*}/\sqrt{n}, \hat{r}_{\hat{\beta}^*}(n) + z_{1-\gamma}\hat{\sigma}_{\hat{\beta}^*}/\sqrt{n}\right],$$

where $z_{1-\gamma} = \Phi^{-1}(1-\gamma)$ and $\Phi(\cdot)$ is the distribution function of the standard normal random variable.

This variance reduction technique is also applicable to continuous time Markov chains and semi–Markov processes. We first apply the discrete time method discussed in Section 3.4 and then the control variables method.

4.2. Multiple Estimates Method

The multiple estimates method of variance reduction was introduced by Heidelberger (1977) for regenerative Markov processes. In this section we shall slightly modify the method to adapt it to our situation. Here $\{X(n): n \geq 0\}$ is a strictly stationary ϕ–mixing process.

The multiple estimates for r are formed by choosing new measurable functions $f_j : E \to \Re$ such that $E\{f_j(X)\} = r$ for $1 \leq j \leq k$. Assuming $E\{|f_j(X(0))|\} < \infty$, then

$$\hat{r}_j(n) = \frac{1}{n}\sum_{i=0}^{n-1} f_j(X(i)) \to r \qquad \text{a.s.}$$

Each $\hat{r}_j(n)$ is a strongly consistent estimate for r. Define a $(k+1)$–dimensional vector y as

$$\begin{aligned} y(n) &= (y_0(n), y_1(n), \ldots, y_k(n)) \\ &= (f(X(n)), f_1(X(n)), \ldots, f_k(X(n)))', \qquad n \geq 0. \end{aligned}$$

It is not difficult to see that the process $y = \{y(n) : n \geq 0\}$ is strictly stationary. If $\mathcal{N}_a^b$ is the σ–field generated by $X(a), \ldots, X(b)$ and if $\mathcal{M}_a^b$ is the σ–field generated by $y_j(a), \ldots, y_j(b)$, $0 \leq j \leq k$, then $\mathcal{M}_a^b \subseteq \mathcal{N}_a^b$. Since $\{X(n) : n \geq 0\}$ is ϕ–mixing it follows that y is ϕ–mixing. Thus Equation (4.1.1) holds with $\mu = (r, r, \ldots, r)'$. Let $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ so that $\sum_{j=0}^{k} \beta_j = 1$ and form

$$\hat{r}_\beta(n) = \sum_{j=0}^{k} \beta_j \hat{r}_j(n).$$

We have $\hat{r}_\beta(n) \to r$ a.s. and

$$\frac{\sqrt{n}\,(\hat{r}_\beta(n) - r)}{\sigma_\beta} \Rightarrow N(0,1), \tag{4.2.1}$$

where $\sigma_\beta^2 = \beta' \Sigma \beta = \sum_{i,j=0}^{k} \beta_i \sigma_{ij} \beta_j$.

To maximize the variance reduction we solve the non–linear programming problem:

$$\begin{aligned} &\text{minimize} && \sigma_\beta^2 = \beta' \Sigma \beta \\ &\text{subject to} && \beta' e = 1, \end{aligned}$$

where $e = (1, 1, \ldots, 1)'$. The optimal solution is

$$\begin{aligned} \beta^* &= \Sigma^{-1} / e' \Sigma^{-1} e, \\ \sigma_{\beta^*}^2 &= 1 / e' \Sigma^{-1} e. \end{aligned} \tag{4.2.2}$$

To apply the multiple estimates method, we need to find the proper functions f_j's. We shall discuss the choices for Markov processes. Let $\{X(n) : n \geq 0\}$ be a discrete time Markov process with probability transition function P. One choice suggested by Heidelberger (1977) is setting

$$\begin{aligned} f_0 &= f, \\ f_j &= P^j f, \qquad j \geq 1, \end{aligned}$$

which is defined by

$$f_j(x) = \int_E P^j(x, dy) f(y).$$

Obviously, $f_j = P f_{j-1}$, $j \geq 1$. It can be shown that $\pi f = \pi(Pf)$. By induction it follows that

$$\begin{aligned} r_j &= \pi f_j = \pi(P f_{j-1}) \\ &= \pi f_{j-1} = \cdots = r. \end{aligned}$$

This idea is also applicable to continuous time Markov chain $\{X(t) : t \geq 0\}$ with discrete state space. We first apply the discrete time method discussed in Section 3.4. Let $\{X(n) : n \geq 0\}$ be the stationary embedded jump chain. For $j = 1, \ldots, k$ we define

$$\begin{aligned} g_j(i) &= f_j(i) q_i^{-1}, \quad i \in E \\ u_j(n) &= g_j\big(X(n)\big) \\ v(n) &= q_{X(n)}^{-1} \\ w_j(n) &= u_j(n) - r v(n) \end{aligned}$$

and

$$w(n) = \big(w_0(n), \ldots, w_k(n)\big)'$$
$$u(n) = \big(u_0(n), \ldots, u_k(n)\big)'.$$

$\{ w(n) : n \geq 0 \}$ is strictly stationary and ϕ–mixing. For $0 \leq j \leq k$, we have $r_j = E\{ u_j(n) \}/E\{ v(n) \}$ (cf. Section 3.4). Thus $w(n)$ has mean zero,

$$\hat{r}_j(n) = \frac{\bar{u}_j}{\bar{v}} \rightarrow r \qquad \text{a.s.} \qquad \text{and,}$$

$$\sqrt{n}\,\bar{w} = \frac{\sqrt{n}\,(\bar{u}/\bar{v} - \mu)}{1/\bar{v}} \Rightarrow N(0, \Sigma_w),$$

where $\Sigma_w = \sum_{j=-\infty}^{\infty} R_w(j)$. By using a continuous mapping argument we may replace $\bar{v}$ by $E\{ v(0) \}$,

$$\frac{\sqrt{n}\,(\hat{r}(n) - \mu)}{1/E\{ v(0) \}} \Rightarrow N(0, \Sigma_w), \tag{4.2.3}$$

where $\hat{r}(n) = \big(\hat{r}_0(n), \ldots, \hat{r}_k(n)$. By applying the continuous mapping theorem again, we immediately obtain $\Sigma = \Sigma_w / E\{ v(0) \}^2$.

To select f_j we recall that (cf. Section 3.4)

$$\pi' R = \pi',$$

where π is the stationary distribution of the process $\{ X(t) : t \geq 0 \}$ and R is defined by Equation (3.4.4). Therefore, $\pi' f = \pi' R f$. This suggests we let $f_0 = f$ and choose

$$f_j = R^j f \qquad j \geq 1.$$

Again $\pi' f_j = r$ for all j.

4.3. Innovation Control Method

The third variance reduction technique we will discuss is the innovation control. This method utilizes the so–called Wold decomposition of the stationary process. For the purpose of variance reduction we need only discuss the scalar case.

Following Anderson (1971), we use the double–infinite sequence of random variables $\{ x(n) : -\infty < n < \infty \}$ to generate the relevent Hilbert space. Let M_n be the closed subspace spanned by $x(m)$, $m \leq n$. Thus M_n contains all finite linear combinations, $\sum_{j \in S} a(j)x(j)$, where S is a finite set of integers which are less than or equal to n, as well as their limits in mean square. If x and y are two elements of this Hilbert space, then $E\{ xy \}$ is called the inner product. Clearly, $M_m \subseteq M_n$, $m \leq n$. We put $M_{-\infty} = \bigcap_{n=-\infty}^{\infty} M_n$ and $M_\infty = M$. The best linear prediction of $x(n)$ by $x(n-1), x(n-2), \ldots$ is the projection of $x(n)$ on M_{n-1}, denoted by $\hat{x}(n)$. Put $\epsilon(n) = x(n) - \hat{x}(n)$, then $\epsilon(n) \perp M_{n-1}$, i.e., $\epsilon(n)$ is orthogonal to every element in M_{n-1}. The random sequence $\epsilon(n)$ is usually called the *innovation*. If $E\{ \epsilon(n)^2 \} = \sigma^2 = 0$, the process is said to be purely deterministic. If $E\{ \epsilon(n)^2 \} = \sigma^2 > 0$, the process is called *regular*. The Wold decomposition, Wold (1954), clarifies the structure of a stationary process.

(4.3.1) WOLD DECOMPOSITION THEOREM. If $\{ x(n) : -\infty < n < \infty \}$ is a regular stationary stochastic process with $E\{ x(n) \} = 0$, it can be written as

$$x(n) = \sum_{j=0}^{\infty} a(j)\epsilon(n-j) + v(n) = u(n) + v(n) \qquad (4.3.2)$$

where $\sum_{j=0}^{\infty} a(j)^2 < \infty$, $a(0) = 1$, $E\{\epsilon(n)\} = E\{v(n)\} = 0$, $\epsilon(n) \in M_n$ and $v(n) \in M_{-\infty}$. The sequences $\{\epsilon(n) : -\infty < n < \infty\}$ and $\{v(n) : -\infty < n < \infty\}$ are unique.

Proof. See Anderson (1971) pp. 420--421. ∎

The next theorem states the spectral functions of $u(n)$ and $v(n)$

(4.3.3) THEOREM. If $x(n)$ is regular with spectral distribution function $F = F_{ac} + F_s + F_d$, and f is the derivative of F_{ac}, then f is the spectral density function of $u(n)$ and $F_s + F_d$ is the spectral distribution function of $v(n)$. Furthermore,

$$f(\lambda) = \frac{\sigma^2}{2\pi} |\sum_{j=0}^{\infty} a(j) e^{-ij\lambda}|^2.$$

Proof. See Hannan (1970)pp. 140–141. ∎

Now we apply these theorems to obtain some variance reduction. Let $\{X(n) : n \geq 0\}$ be a strictly stationary and ϕ–mixing process. We wish to estimate $r = E\{g(X)\}$[†] for a given function g. Let $x(n) = g(X(n))$ and assume the process $\{x(n) : n \geq 0\}$ is regular. Applying Theorem (4.3.1) yields

$$x(n) = \sum_{j=0}^{\infty} a(j)\epsilon(n-j) + v(n).$$

Let

$$\hat{x}(n) = \sum_{j=1}^{\infty} a(j)\epsilon(n-j) + v(n),$$
$$\epsilon(n) = x(n) - \hat{x}(n), \quad \text{and} \quad \sigma^2 = E\{\epsilon(n)^2\}.$$

Define

$$y(n) = \big(x(n), \epsilon(n)\big)'.$$

[†]We reserve f for the spectral density function.

Then $\{ y(n) : n \geq 0 \}$ is strictly stationary and ϕ–mixing, therefore

$$\frac{1}{n} \sum_{i=0}^{n-1} y(i) \rightarrow \mu \quad \text{a.s., and}$$

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=0}^{n-1} y(i) - \mu \right) \Rightarrow N(O, \Sigma),$$

where

$$\mu = (r, 0)'$$

$$\Sigma = \sum_{n=-\infty}^{\infty} R(n) = \{ \sigma_{ij} \}.$$

Here σ_{ij}, $i, j = 1, 2$ is defined by Equation (2.3.13). Let $F(\lambda)$ be the spectral distribution matrix of $y(n)$ and $f(\lambda) = dF_{ac}(\lambda)/d\lambda$. Then $R(n) = \int_{-\pi}^{\pi} e^{in\lambda} F(d\lambda)$. Note that

$$\sigma_{11} = \sum_{n=-\infty}^{\infty} R_{11}(n) = \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} e^{in\lambda} F_{11}(d\lambda)$$

$$= 2\pi f_{11}(0) + c = \sigma^2 | \sum_{j=0}^{\infty} a(j)|^2 + c,$$

where c is the infinite sum of the integral from the discrete and singular parts of $F^{11}(\lambda)$. And

$$\sigma_{12} = \sum_{n=-\infty}^{\infty} Cov\{ x(n), \epsilon(0) \}$$

$$= \sum_{n=-\infty}^{\infty} Cov\{ \sum_{j=0}^{\infty} a(j)\epsilon(n-j) + v(n), \epsilon(0) \}$$

$$= \sigma^2 \sum_{n=0}^{\infty} a(n),$$

$$\sigma_{22} = \sum_{n=-\infty}^{\infty} E\{ \epsilon(n)\epsilon(0) \} = \sigma^2.$$

As we have done in Section 4.1, we take $\beta = (1, b)'$ and form $\beta' y(n)$. Then

$$\hat{r}_\beta(n) = \frac{1}{n} \sum_{i=0}^{n-1} (f(X(i)) + b\epsilon(i))$$

is an estimate of r and $\hat{r}_\beta(n) \to r$ a.s. as $n \to \infty$. Also the central limit theorem holds:

$$\sqrt{n}(\hat{r}_\beta(n) - r)/\sigma_\beta \Rightarrow N(0,1) \quad \text{as} \quad n \to \infty$$

where $\sigma_\beta^2 = \sigma_{11} + 2b\sigma_{12} + b^2\sigma_{22}$. Again, we may choose β^* to minimize σ_β^2 which is

$$\beta^* = (1, -\sigma_{12}/\sigma_{22})' = (1, -\sum_{n=0}^{\infty} a(n))'$$

$$\text{and} \quad \sigma_{\beta^*}^2 = \sigma_{11} - \sigma_{12}^2/\sigma_{22} = c.$$

We notice that if the process $\{x(n) : n \geq 0\}$ has absolutely continuous spectral distribution then $\sigma_{\beta^*}^2 = c = 0$ that means $\hat{r}_{\beta^*}(n)$ is a constant random variable r.

Although this method seems extremely good, we encounter difficulties in application. In general, we are not able to obtain the innovation $\epsilon(n)$, unless we know the true values of the parameters (e.g., $a(j)$'s, ...) for $\{x(n) : n \geq 0\}$.

CHAPTER V

NUMERICAL EXAMPLES

In this chapter, we present four examples to demonstrate how well our method performs. One of the examples, the outflow process in a lake model, is actually a first order autoregressive process. All other examples come from the area of queueing theory. They are the waiting time process in an $M/M/1$ queue, the passage time and response time processes in a closed network of queues, and the queue length process in a cyclic queue. These processes are regenerative processes for which we can calculate the theoretical values of the parameters being estimated. Therefore we are able to make the comparison between theoretical values and simulation estimates. For all examples, the results obtained from the autoregressive method are quite satisfactory.

To determine the order of autoregressive model, we use AIC, BIC, h–criterion, and the statistical test discussed in Section 3.2 for the univariate case, and AIC for multivariate autoregressive method. Among different criteria, the results obtained by using the AIC are usually the closest ones to true values. We may conclude that AIC is the best one for many applications.

All problems were run on a DEC–20 computer and we used the build–in uniform random number generator to simulate all processes.

5.1. EXAMPLE. Waiting Time Process in an $M/M/1$ Queue

Suppose customers arrive at a single service facility. If a customer finds the server idle, then he receives service immediately, otherwise, he waits his turn to be served. Assume the zeroth customer arrives at time $t_0 = 0$, finds a free server, and experiences a service time $v(0)$. The nth customer arrives at time t_n and experiences a service time $v(n)$. Let the interarrival times $u(n) = t_n - t_{n-1}$, where $n \geq 1$. Also assume the two sequences $\{ u(n) : n \geq 1 \}$ and $\{ v(n) : n \geq 0 \}$ each consists of i.i.d. random variables and are themselves independent. Let $E\{ v(0) \} = \mu^{-1}$, $E\{ u(1) \} = \lambda^{-1}$, and the traffic intensity $\rho = \lambda/\mu$. Let $W(n)$ be the waiting time of the nth customer. Then $W(n)$ can be defined recursively by

$$\begin{aligned} W(0) &= 0, \\ W(n) &= [\, W(n-1) + X(n) \,]^+ \\ &= \max\{ 0, W(n-1) + X(n) \}, \quad n \geq 1, \end{aligned} \tag{5.1.1}$$

where $X(n) = v(n-1) - u(n)$. It is known that if $\rho < 1$ then there exists a random variable W such that $W(n) \Rightarrow W$ as $n \to \infty$. This model is commonly called the $GI/G/1$ queue. If the arrivals form a Poission process with rate λ and service times are exponentially distributed with rate μ, then the queue is called an $M/M/1$ queue. We are interested in estimating $E(W)$, which is finite if $E\{ v(n)^2 \} < \infty$.

We observe the $\{ W(n) : n \geq 0 \}$ process for the univariate autoregressive method. For the variance reduction techniques, we apply the control variables and multiple estimates methods. It is natural to use the service time and interarrival time as control variables; i.e., we use $(W(n), u(n), v(n-1))$ for

the control variables method. We use the column vector

$$y(n) = \big(f_0(W(n)), f_1(W(n)), f_2(W(n))\big)', \tag{5.1.2}$$

for the multiple estimates method. Here, the f function is $f(x) = x$, $f_j(x) = P^j f$, $j \geq 0$. In order to calculate f_j, we need to find the probability transition function for the $\{W(n)\}$ process. For the $M/M/1$ queue it is easy to show that

$$P\{X(n) \leq x\} = \begin{cases} \frac{\mu}{\lambda+\mu} e^{\lambda x}, & \text{for } x < 0; \\ 1 - \frac{\lambda}{\lambda+\mu} e^{-\mu x}, & \text{for } x \geq 0. \end{cases}$$

Thus $g(x) = \frac{d}{dx} P\{X(n) \leq x\}$ exists for all x and we write $P\{X(n) \in dx\} = g(x)dx$. Now to evaluate $f_1(x)$ we have

$$\begin{aligned} f_1(x) &= \int_0^\infty P(x, dy) f(y) = \int_0^\infty y P\{W(n+1) \in dy \mid W(n) = x\} \\ &= \int_0^\infty y P\{x + X(n+1) \in dy\} = \int_0^\infty y g(y-x) dy. \end{aligned}$$

We find

$$f_1(x) = x + \frac{\lambda - \mu}{\lambda\mu} + \frac{\mu}{\lambda(\lambda+\mu)} e^{-\lambda x}.$$

To evaluate $f_2(x)$ we compute the integral

$$f_2(x) = \int_0^\infty P(x, dy) f_1(y).$$

After this computation we find

$$f_2(x) = f_1(x) + \frac{\lambda - \mu}{\lambda\mu} + \left(\frac{\mu}{\lambda+\mu}\right)^2 e^{-\lambda x} \left(\frac{1}{\lambda} + \frac{1}{\lambda+\mu} + x\right).$$

It is possible to calculate exactly the covariance matrix Σ for $y(n)$ in (5.1.2), so that we can compute the theoretical values of variance reductions for the multiple estimates method. The idea is to use the stationary process

$\{W(n) : n \geq 0\}$, i.e., $W(0)$ is distributed according to its stationary distribution. Define the generating function of the joint Laplace transform of the stationary waiting time process by

$$C(z_1, z_2, s) = \sum_{n=0}^{\infty} E\{ e^{-z_1 W(0) - z_2 W(n)} \} \cdot s^n,$$

where $|s| < 1$. $C(z_1, z_2, s)$ has been calculated by Blomqvist (1967), he also gave the exact form for the covariance function of $\{W(n) : n \geq 0\}$ which is

$$\begin{aligned} R(k) &= Cov\{W(k), W(0)\} \\ &= \frac{1}{\mu^2} \cdot \frac{1-\rho^2}{\rho^2} \cdot \sum_{i=k+3}^{\infty} \left(\frac{\rho}{(1+\rho)^2} \right)^i \cdot \frac{(2i-3)!}{i!(i-2)!} \cdot (i-k-1) \cdot (i-k-2). \end{aligned}$$

We can obtain σ_{ij} from $C(z_1, z_2, s)$, e.g.

$$\sigma_{11} = \sigma^2\{W(0)\} + 2 \sum_{n=1}^{\infty} Cov\{W(n), W(0)\},$$

$$\begin{aligned} \sigma_{12} = \sigma^2\{W(0)\} &+ 2 \sum_{n=1}^{\infty} Cov\{W(0), W(n)\} \\ &+ \frac{\mu}{\lambda(\lambda+\mu)} \Big(\sum_{n=1}^{\infty} Cov\{W(0), e^{-\lambda W(n)}\} + \sum_{n=1}^{\infty} Cov\{W(n), e^{-\lambda W(0)}\}, \end{aligned}$$

and

$$\sum_{n=0}^{\infty} E\{W(0)W(n)\} = \lim_{s \to 1} \frac{\partial^2 C(z_1, z_2, s)}{\partial z_1 \partial z_2} \Big|_{z_1=0, z_2=0},$$

$$\sum_{n=0}^{\infty} Cov\{W(0)e^{-\lambda W(n)}\} = -\lim_{s \to 1} \frac{\partial C(z_1, z_2, s)}{\partial z_1} \Big|_{z_1=0, z_2=\lambda}.$$

After we have calculated σ_{ij}, we can obtain σ_{β}^2. by using Equation (4.2.2).

The exact form of the covariance function R allows us to demonstrate the autoregressive approximation. Recall that the autoregressive method enables us to approximate σ^2 (cf. Equation (4.0.2)) by 2π times the spectral density

function at 0 of some finite order autoregressive process (cf. Section 3.1). If we assume that $\{W(n) : n \geq 0\}$ is a pth order order autoregressive process, then $\{W(n) : n \geq 0\}$ satisfies

$$\sum_{j=0}^{p} \beta_p(j) W(n-j) = \epsilon(n), \quad \beta_p(0) = 1, \quad \beta_p(p) \neq 0,$$

where $\epsilon(n)$ are i.i.d. random variables with variance $\sigma_p^2(\epsilon)$. We solve the following Yule–Walker equations (cf. Section 2.2), namely

$$\sum_{j=0}^{p} \beta_p(j) R(s-j) = 0, \quad s = 1, \ldots, p,$$

to obtain $\beta_p(j)$'s and $\sigma_p^2(\epsilon) = \sum_{j=0}^{p} \beta_p(j) R(-j)$. The corresponding approximation for σ^2 is $2\pi f_k(0) = \sigma_p^2(\epsilon)/|\sum_{j=0}^{p} \beta_p(j)|^2$. Table 1 contains the values of $R(k)$ for $k \leq 10$, it also gives $2\pi f_k(0)$ and simulation results for $2\pi f_k(0)$ for $k \leq 10$. From this table we can see that $2\pi f_k(0)$ converges fairly fast to $\sigma^2 = 2\pi f_\infty(0)$ (the stationary waiting time process is an infinite order autoregressive process). Observe that $2\pi f_1(0)$ is 23.99 which is 83% of σ^2 ($2\pi f_\infty(0) = 29$). If we want an estimate with accuracy 90% of the true value, we may pick the order of the autoregressive model to be as low as 3.

To see how well the autoregressive method performs in an actual simulation, we have taken $\mu = 1$, $\rho = 0.5$. To obtain a $100(1-2\gamma)\%$ confidence interval with half length $100\delta\%$ of the true value of $E\{W\}$ requires $z_{1-\gamma}\sigma/\sqrt{n} = \delta \cdot E\{W\}$ for some γ. If we take $\gamma = 0.05$, $\delta = 0.1$ then the number of customers, n, that need to be simulated is roughly 7850 or 3900 regenerative cycles. For the purpose of comparing the autoregressive method and the regenerative method, we simulated the waiting time process in cycles for a total run length of 4000 cycles (the expected total number of customers is

8000). We consider 500, 1000, 2000, 3000, 4000 cycles, the longer runs being continuations of the shorter runs.

All runs were replicated 30 times. For each replication we form point estimates and confidence intervals for various parameters of interest. We then average over the 30 independent replications and form 90% confidence intervals for each parameter; this is done by using the central limit theorem for i.i.d. random variables. The upper bound provided for order selection was 10 for all criteria. Table 2 shows the point estimate for $E\{W\}$ either with or without a variance reduction technique. Table 3 contains the estimate for σ^2 from the regenerative method as well as the approximation of σ^2 from the autoregressive method for various kinds of order selection criteria. By inspecting the results, we conclude that among the order selection criteria, AIC seems to be the most satisfactory one for $\{W(n) : n \geq 0\}$. Table 5 gives the average order of autoregressive model determined by various order selection criteria. We notice that every criterion yielded a low order autoregressive model, this indicates the choice of the upper bound $K = 10$ is quite sufficient. In Tables 4 and 7 we report the coverage probability defined by

$$\text{c.p.} = \frac{\text{number of 90\% confidence intervals covering } E\{W\}}{\text{total number of confidence intervals formed}}$$

which has expected value 0.9.

Table 6 contains the results of $\sigma^2_{\beta^*}$ for different variance reduction techniques. Here we estimated the optimal multipliers β^* by using equations derived in Chapter 4. From Table 6 we see that by using 3 different f functions for the multiple estimates, we get a substantial amount of variance reduction. In order to judge a variance reduction technique, we must make a

fair comparison. Suppose we simulate n_1 customers using no variance reduction technique (method 1) and n_2 customers using multiple estimates with 3 f functions (method 2). In order to obtain the same statistical accuracy (i.e. same half lengths of confidence intervals) we have

$$z_{1-\alpha}\frac{\sigma_1}{\sqrt{n_1}} = z_{1-\alpha}\frac{\sigma_2}{\sqrt{n_2}},$$

that is equivalent to saying

$$\frac{\sigma_1^2}{\sigma_2^2} = \frac{n_1}{n_2}.$$

From Table 6 we find that for method 2 we can cut the run length to 1/30 of the run length of method 1 and still obtain the same statistical accuracy. Of course, we recognize that by using method 2 we do a certain amount of extra work. Since the computing time depends on the run length as well as the upper bound selected for the autoregressive method, it is hard to find the relation of times between these two methods. However, based on our results, the half length of confidence interval constructed by method 2 is 30 $(= 29/.948)$ times smaller than that of method 1 but requires 6 times (for example, take the last row of Table 8, $6 \approx (110.96 + 372.16)/(43.61 + 39.74)$) as much CPU time.

TABLE 1

Point Estimates and 90% Confidence Intervals for $2\pi f(0)$

Using Various Orders of k in the Autoregressive Method for

the Waiting Time Process in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$

Order	True Values		Simulation Results for different number of cycles				
k	$\rho(k)$	$2\pi f_k(0)$	500	1000	2000	3000	4000
1	0.778	23.99	27.90 ± 6.62	23.92 ± 3.66	24.47 ± 3.03	24.48 ± 2.77	24.20 ± 1.95
2	0.617	25.56	30.06 ±7.69	25.37 ±3.99	26.27 ±3.44	26.20 ±3.16	25.82 ±2.28
3	0.497	26.50	30.76 ±7.91	26.05 ±4.15	26.90 ±3.71	26.94 ±3.46	26.64 ±2.51
4	0.403	27.10	31.19 ±8.49	26.79 ±4.53	27.70 ±4.11	27.69 ±3.61	27.32 ±2.65
5	0.330	27.57	31.35 ±8.56	27.11 ±4.75	28.06 ±4.32	28.26 ±3.81	27.83 ±2.81
6	0.272	27.86	31.05 ±8.66	27.30 ±5.09	28.16 ±4.55	28.66 ±4.04	28.13 ±2.95
7	0.225	28.12	31.87 ±8.99	28.05 ±5.39	28.39 ±4.56	29.02 ±4.19	28.48 ±3.08
8	0.187	28.28	33.13 ±9.77	28.88 ±5.83	28.85 ±4.72	29.45 ±4.39	28.78 ±3.19
9	0.157	28.46	33.20 ±9.90	28.92 ±5.84	29.22 ±4.70	29.59 ±4.41	28.94 ±3.23
10	0.131	28.51	32.93 ±9.78	29.07 ±6.02	29.06 ±4.66	29.86 ±4.60	29.13 ±3.31
∞	0	29.00	—	—	—	—	—

• Results are based on 30 independent replications, $\rho(k) = R(k)/R(0)$, $R(0) = 3$.

TABLE 2

Simulation Results for $E\{W\} = 1.0$ in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$, Point Estimates and 90% Confidence Intervals

Number of Cycles	No Variance Reduction	Control Variables	Multiple Estimates	
			2 functions	3 functions
500	1.0039 ±0.0610	0.9868 ± 0.0379	0.9823 ± 0.0160	0.9821 ± 0.0082
1000	0.9901 ±0.0428	0.9812 ± 0.0227	0.9861 ± 0.0130	0.9928 ± 0.0067
2000	0.9942 ±0.0316	0.9953 ± 0.0171	0.9964 ± 0.0101	0.9962 ± 0.0050
3000	0.9978 ±0.0250	0.9938 ± 0.0160	0.9953 ± 0.0080	0.9951 ± 0.0040
4000	0.9949 ±0.0189	0.9980 ± 0.0118	1.0003 ± 0.0069	0.9982 ± 0.0032

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction techniques.
- The AIC criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 10$.

TABLE 3

Simulation Results for $\sigma^2 = 29.0$ in an $M/M/1$ Queue

with $\mu = 1.0$, $\rho = 0.5$, Point Estimates and 90% Confidence Intervals

Number of Cycles	Regenerative Method	Univariate Autoregressive Method			
		AIC	*BIC*	h	Stat. Test
500	29.33 ± 7.83	29.99 ± 7.94	28.67 ± 7.11	29.75 ± 7.80	28.58 ± 6.69
1000	27.36 ± 5.07	28.54 ± 5.75	24.57 ± 3.92	26.47 ± 4.82	26.04 ± 4.79
2000	29.40 ± 5.56	28.15 ± 4.51	26.02 ± 3.84	26.59 ± 3.83	26.39 ± 3.86
3000	30.12 ± 5.07	29.21 ± 4.47	26.07 ± 3.41	27.21 ± 3.61	26.86 ± 3.59
4000	29.30 ± 3.75	28.25 ± 3.12	25.64 ± 2.49	26.83 ± 3.06	26.04 ± 2.75

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Use the scalar process $\{W(n) : n \geq 0\}$.
- The maximum order for the autoregressive model is $K = 10$.

TABLE 4

Simulation Results for Coverage Probability (= 0.9)

in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$,

Point Estimates and 90% Confidence Intervals

Number of Cycles	Regenerative Method	Univariate Autoregressive Method			
		AIC	*BIC*	h	Stat. Test
500	0.70 ±0.14	0.67 ± 0.14	0.67 ± 0.14	0.67 ± 0.14	0.67 ± 0.14
1000	0.77 ±0.13	0.73 ± 0.14	0.73 ± 0.14	0.73 ± 0.14	0.73 ± 0.14
2000	0.80 ±0.12	0.80 ± 0.12	0.80 ± 0.12	0.80 ± 0.12	0.80 ± 0.12
3000	0.80 ±0.12	0.80 ± 0.12	0.77 ± 0.13	0.80 ± 0.12	0.77 ± 0.13
4000	0.90 ±0.09	0.90 ± 0.09	0.83 ± 0.11	0.90 ± 0.09	0.90 ± 0.09

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Use the scalar process $\{W(n) : n \geq 0\}$.
- The maximum order for the autoregressive model is $K = 10$.

TABLE 5

Order of Autoregressive Model Selected by Different Criteria

in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$,

Point Estimates and 90% Confidence Intervals

Number of Cycles	Univariate Autoregressive Method			
	AIC	*BIC*	h	Stat. Test
500	2.37 ± 0.56	1.10 ± 0.09	1.43 ± 0.23	1.13 ± 0.13
1000	3.30 ± 0.83	1.10 ± 0.09	1.70 ± 0.35	1.47 ± 0.44
2000	2.80 ± 0.60	1.33 ± 0.20	1.67 ± 0.41	1.53 ± 0.42
3000	3.83 ± 0.81	1.43 ± 0.22	2.23 ± 0.51	2.00 ± 0.53
4000	3.90 ± 0.75	1.47 ± 0.22	2.14 ± 0.47	1.73 ± 0.41

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Use the scalar process $\{W(n) : n \geq 0\}$.
- The maximum order for the autoregressive model is $K = 10$.

TABLE 6

Point Estimates and 90% Confidence Intervals for σ^2 by Applying Variance Reduction Technique in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$

Multivariate Method	True Value	Simulation results for different number of cycles				
		500	1000	2000	3000	4000
Control Variables	—	13.75 ± 3.45	12.57 ± 2.12	13.36 ± 11.89	13.92 ± 2.14	13.85 ± 1.53
Multiple Estimates(2 f's)	4.23	3.67 ±1.65	3.40 ±1.11	3.92 ±0.98	4.78 ±1.77	4.59 ±1.37
Multiple Estimates(3 f's)	0.948	0.589 ±0.373	0.554 ±0.230	0.679 ±0.220	0.930 ±0.441	0.918 ±0.367

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction
- The AIC criterion was used for order selection with maximum order for the autoregressive model of $K = 10$.

TABLE 7

Point Estimates and 90% Confidence Intervals

for Coverage Probability by Applying Variance Reduction Technique

in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$

Multivariate Method	True Value	Simulation results for different number of cycles				
		500	1000	2000	3000	4000
Control Variables	0.9	0.80 ± .12	0.80 ± .12	0.93 ± .08	0.83 ± .11	0.90 ± .09
Multiple Estimates(2 f's)	0.9	0.73 ±.14	0.77 ±.13	0.83 ±.11	0.80 ±.12	0.90 ±.09
Multiple Estimates(3 f's)	0.9	0.67 ±.14	0.67 ±.14	0.77 ±.13	0.77 ±.13	0.90 ±.09

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction
- The AIC criterion was used for order selection with Maximum order for the autoregressive model of $K = 10$.

TABLE 8

Comparison of Total CPU Time (in seconds) for Different Methods for Estimating $E\{W\}$ in an $M/M/1$ Queue with $\mu = 1.0$, $\rho = 0.5$

Number of Cycles	Generate sample path		Autoregressive method	
	No v.r.	Multiple 2 f's	Univariate	Multivariate
500	5.85	13.69	5.95	54.26
1000	11.51	26.92	11.02	98.13
2000	22.39	55.20	20.76	187.81
3000	33.85	84.69	31.06	288.63
4000	43.61	110.96	39.74	372.16

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- v.r. is the abbreviation for variance reduction.
- f_1 and f_2 were calculated for the variance reduction method.
- Multiple estimate f_1, f_2.

5.2. EXAMPLE. The Outflow Process in a Lake

Let $S(n)$, $Z(n)$, and $X(n)$ denote the volume of water stored in a lake, the inflow of water, and the outflow of water respectively at time n. Then the amount of water stored at time $n+1$ is defined by the relation

$$S(n+1) = S(n) + Z(n+1) - X(n+1). \tag{5.2.1}$$

If one assumes that output increases with storage through a linear relation, namely

$$X(n) = a \cdot S(n), \qquad 0 < a < 1,$$

then Equation (5.2.1) has the form

$$X(n+1) = \rho X(n) + \epsilon(n+1) \tag{5.2.2}$$

where $\rho = 1/(1+a)$ and $\epsilon(n) = aZ(n)/(1+a)$. We also assume that $\{\epsilon(n) : n \geq 0\}$ is a sequence of i.i.d. random variables with

$$P\{\epsilon(n) \in B\} = p \cdot \delta_0(B) + (1-p)\int_B f(y)dy.$$

This yields a Markov process for the outflow process $\{X(n) : n \geq 0\}$ with state space $E = [0, \infty)$ and probability transition function

$$P(x, B) = p \cdot \delta_{\rho x}(B) + (1-p)\int_B f(y - \rho x)dy.$$

We notice that $\{X(n) : n \geq 0\}$ is actually a first order autoregressive processes. It can be shown that $X(n) \Rightarrow X$ as $n \to \infty$ with

$$E\{X\} = E\{\epsilon(1)\}/(1-\rho) = E\{Z(1)\}.$$

A simulation was carried out to estimate $E\{X\}$. We have taken $\rho = 0.75$ and the $\epsilon(n)$ to have distribution

$$P\{\epsilon(n) \leq x\} = 1 - (1-\rho)e^{-x}, \qquad x \geq 0,$$

which results in X being exponentially distributed with parameter 1.

We observe $\{X(n) : n \geq 0\}$ and use the univariate autoregressive method. It is also possible to observe the variables $\epsilon(n)$ during the simulation, since we know ρ. Thus we were able to apply the innovation control method for variance reduction. We use the vector process $\{y(n) : n \geq 0\}$, where $y(n) = (X(n), \epsilon(n))'$, for the innovation control method. We consider 1000, 2000, 5000, 10000 observations, the longer runs being continuations of the shorter runs. All runs were replicated 30 times. Notice that the process $\{X(n) : n \geq 0\}$ is indeed a first order autoregressive process and the process $\{y(n) : n \geq 0\}$ can be written as

$$\begin{pmatrix} X(n+1) \\ \epsilon(n+1) \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X(n) \\ 0 \end{pmatrix} + \begin{pmatrix} \epsilon(n+1) \\ \epsilon(n+1) \end{pmatrix}$$

which results in $y(n)$ being a first order autoregressive process. Therefore, the upper bound provided for order selection was 5 for all criterions. Table 9–12 summarize the simulation results for the lake model. We estimated the optimal β^* by Equation (4.3.4) and we observe that $\sigma^2_{\beta^*}$ was actually reduced to zero as expected. The fact that $\{X(n) : n \geq 0\}$ and $\{y(n) : n \geq 0\}$ are finite order autoregressive processes offers us a chance to validate the asymptotic property of $\hat{k}$ (cf. Section 3.3 and 3.4) obtained by various order selection criteria. Table 12 contains the average order determined by each criterion. The table shows what is to be expected, namely overestimation of the order by AIC.

TABLE 9

Simulation Results for $E\{X\}$ (the outflow) in the Lake Model with $\rho = 0.75$, Point Estimates and 90% Confidence Intervals

Number of Events	True Value	No Variance Reduction	Innovation Control
1000	1.0	1.0179 ±0.0239	0.9912 ±0.0010
2000	1.0	1.0229 ±0.0173	0.9984 ±0.0006
5000	1.0	1.0151 ±0.0092	0.9994 ±0.0001
10000	1.0	1.0069 ±0.0070	0.9997 ±0.0001

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 5$.

TABLE 10

Simulation Results of σ^2 for $\{X(n)\}$ in the Lake Model with $\rho = 0.75$, Point Estimates and 90% Confidence Intervals

Number of Events	Univariate Autoregressive Method					Innovation Control	
	True V.	*AIC*	*BIC*	h	Stat. T.	True V.	Simulation
1000	7.0	7.2582 ± .5358	7.2172 ± .4759	7.1686 ± .4940	7.2172 ± .4759	0.0	0.0238 ± .0149
2000	7.0	7.1428 ± .3473	7.1924 ± .3189	7.1402 ± .3068	7.1924 ±.3189	0.0	0.0149 ± .0117
5000	7.0	7.0935 ± .2349	7.1171 ± .2356	7.0945 ± .2305	7.1171 ±.2356	0.0	0.0030 ± .0006
10000	7.0	6.9887 ± .1616	7.0245 ± .1536	7.0118 ± .1463	7.0245 ±.1536	0.0	0.0022 ± .0011

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressiv method.
- The maximum order for the autoregressive model is $K = 5$.
- True V. is the abbreviation for True Value.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 11

Simulation Results for Coverage Probability in the Lake Model with $\rho = 0.75$, Point Estimates and 90% Confidence Intervals

Number of Events	Univariate Autoregressive Method					Innovation Control	
	True V.	*AIC*	*BIC*	h	Stat. T.	True V.	Simulation
1000	0.9	0.90 ± .09	0.90 ± .09	0.90 ± .09	0.90 ± .09	0.9	1.0 ± .0
2000	0.9	0.90 ± .09	0.90 ± .09	0.90 ± .09	0.90 ± .09	0.9	1.0 ± .0
5000	0.9	0.97 ± .05	0.97 ± .05	0.97 ± .05	0.97 ± .05	0.9	1.0 ± .0
10000	0.9	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.9	1.0 ± .0

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 5$.
- True V. is the abbreviation for True Value.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 12

Order of Autoregressive Model Selected by Different Criteria in Lake Model, with $\rho = 0.75$, Point Estimates and 90% Confidence Intervals

Number of Events	Univariate Autoregressive Method				MAR Method
	AIC	*BIC*	h	Stat. Test	
1000	1.3333 ± .2410	1.0 ± .0	1.0667 ± .0762	1.0 ± .0	1.0000 ± .0000
2000	1.6333 ± .3391	1.0 ± .0	1.0667 ± .0762	1.0 ± .0	1.1333 ± .1282
5000	1.3333 ± .2410	1.0 ± .0	1.0667 ± .0762	1.0 ± .0	1.1667 ± .1778
10000	1.5000 ± .2586	1.0 ± .0	1.0333 ± .0548	1.0 ± .0	1.1000 ± .1908

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction techniques.
- *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 5$.
- MAR Method is abbreviation for multivariate autoregressive method.

5.3. EXAMPLE. Closed Network of Queues

The closed queueing network has been used in computer studies to model multiprogrammed computer system (cf. Gaver and Shedler (1973)). Iglehart and Shedler (1980) have written a monograph dealing with mathematical and statistical methods for discrete event simulation of network of queues. Here we consider the simple closed network of queues with feedback shown in Figure 1.

There are two servers and a fixed number of jobs N circulating in the network. When a job completes service at center 1, in accordance with a binary–valued random variable ψ, the job joins the end of the queue at center 1 (when ψ=1) or joins the end of the queue at center 2 (when $\psi = 0$). When the job finishes service at center 2, the job rejoins the end of the queue at center 1. We assume that the service discipline is first–come–first–serve. In this example we are interested in estimating the limiting passage time (denoted by P) and the limiting response time (denoted by R). Informally, a passage time of a job is the time for a job to traverse a portion of a network and a response time is the time for a job to complete one cycle of a network.

For the simulation of this queuing network, we make the following probabilistic assumptions:

(1) the two sequences of service times at both centers each consists of i.i.d. random variables, exponentially distributed with rate λ_i;

(2) ψ is a Bernoulli random variable with $P\{\psi = 1\} = p$ and values of ψ form a sequence of i.i.d. random variables;

(3) the sequences in (1) and (2) are mutually independent.

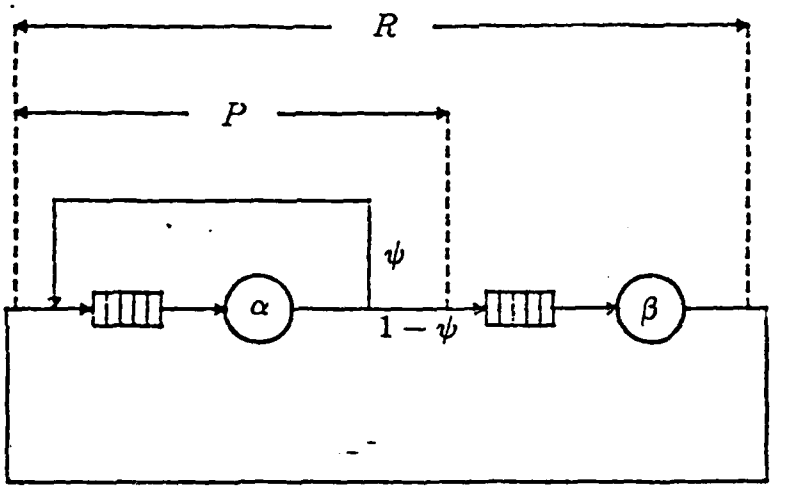


Fig. 1. Closed network of queues.

Let $\{P(n) : n \geq 1\}$ (respectively $\{R(n) : n \geq 1\}$) be the sequence of passage times (respectively response times) enumerated in order of passage time starts and let $\{S(n) : n \geq 1\}$ be the service time experienced at center 1 associated with $P(n)$. We notice that $\{S(n)\}$ is a sequence of i.i.d. random variables. Iglehart and Shedler (1980) have shown that $P(n) \Rightarrow P$ and $R(n) \Rightarrow R$ as $n \to \infty$.

The simulation was carried out to estimate both $E\{P\}$ and $E\{R\}$ and we have taken $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$ and $p = 0.75$. We observe the $\{P(n) : n \geq 0\}$, $\{R(n) : n \geq 0\}$, and $\{S(n) : n \geq 0\}$ processes and use $S(n)$ as the control variable when we estimate $E(P)$ and $E(R)$. We consider 1000, 2000, 3000, 5000, and 10000 observations, the longer runs being continuations of the shorter runs. All runs were replicated 30 times.

Table 13-19 summarize the simulation results for estimating $E\{P\}$ and

$E\{R\}$. The calculation of $E\{P\}$, $E\{R\}$ and the variance constant corresponding to passage times can be found in Section 9 of Iglehart and Shedler (1980). To evaluate the variance reduction for the control variables method, we carried out a similar computation discussed in Iglehart and Shedler (1980). Although we included the simulation results for the variance constant of response times, we can not provide the theoretical value.

TABLE 13

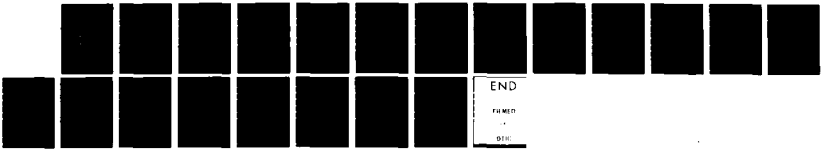
Simulation Results for E{ P } (Passage Time) and E{ R } (Response Time) in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	P (Passage Time)			R (Response Time)		
	True Value	No v.r.	With v.r.	True Value	No v.r.	With v.r.
500	6.667	6.713 ± .120	6.684 ± .019	9.333	9.363 ± .128	9.334 ± .044
1000	6.667	6.616 ± .079	6.667 ± .014	9.333	9.293 ± .091	9.342 ± .035
2000	6.667	6.638 ± .053	6.661 ± .009	9.333	9.309 ± .060	9.331 ± .023
5000	6.667	6.660 ± .039	6.666 ± .005	9.333	9.327 ± .034	9.332 ± .014
10000	6.667	6 668 ± .027	6.664 ± .003	9.333	9.337 ± .024	9.333 ± .009

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- v.r. is the abbreviation for variance reduction.

AD-A123 823 AN AUTOREGRESSIVE METHOD FOR SIMULATION OUTPUT ANALYSIS 2/2
(U) STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH
Y L JOW DEC 82 TR-65 N00014-76-C-0578

UNCLASSIFIED F/G 12/1 NL



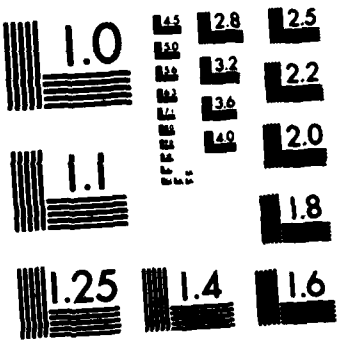


TABLE 14

Simulation Results for σ^2 of Passage Time (P)

in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method					Control Variables	
	True V.	*AIC*	*BIC*	h	Stat. T.	True V.	Simulation
500	58.667	59.887 ± 4.511	57.081 ± 3.392	58.164 ± 3.201	48.922 ± 3.744	1.581	2.300 ± .132
1000	58.667	58.794 ± 2.768	55.910 ± 2.399	56.220 ± 2.349	53.305 ± 2.887	1.581	2.054 ± .119
2000	58.667	58.128 ± 2.168	55.607 ± 1.510	56.203 ± 1.533	54.830 ± 1.721	1.581	1.941 ± .070
5000	58.667	59.746 ± 1.112	57.262 ± 0.798	58.565 ± 1.025	56.989 ± 0.782	1.581	1.801 ± .027
10000	58.667	58.737 ± 0.995	57.051 ± 0.677	58.218 ± 0.807	56.474 ± 0.575	1.581	1.763 ± .028

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.
- True V. is the abbreviation for True Value.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 15

Simulation Results for σ^2 of Response Time (R)

in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method				Control Variables Method
	AIC	*BIC*	h	Stat. T.	
500	60.274 ± 4.840	62.094 ± 4.119	61.891 ± 3.834	58.065 ± 4.468	9.025 ± 0.582
1000	59.989 ± 3.471	60.527 ± 2.752	61.039 ± 2.745	60.216 ± 3.088	8.879 ± 0.450
2000	59.602 ± 1.755	62.133 ± 1.600	60.621 ± 1.643	61.923 ± 1.885	8.650 ± 0.323
5000	60.909 ± 1.067	62.493 ± 0.928	61.152 ± 1.113	63.041 ± 0.834	8.305 ± 0.277
10000	59.160 ± 0.863	60.182 ± 0.832	60.006 ± 0.806	60.985 ± 0.861	8.354 ± 0.187

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 16

Simulation Results for Coverage Probability(=0.9) of Passage Times (P) in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method				Control Variables Method
	AIC	*BIC*	*h*	Stat. Test	
500	0.90 ± .09	0.90 ± .09	0.90 ± .09	0.83 ± .11	0.93 ±0.08
1000	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.87 ± .10	0.87 ±0.10
2000	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.90 ±0.09
5000	0.80 ± .12	0.80 ± .12	0.80 ± .12	0.80 ± .12	0.90 ±0.09
10000	0.83 ± .11	0.83 ± .11	0.83 ± .11	0.83 ± .11	0.93 ±0.08

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 17

Simulation Results for Coverage Probability(=0.9) of Response Times (R) in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method				Control Variables Method
	AIC	*BIC*	*h*	Stat. T.	
500	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.80 ±0.12
1000	0.80 ± .12	0.80 ± .12	0.80 ± .12	0.83 ± .11	0.87 ±0.12
2000	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.87 ±0.10
5000	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.87 ±0.10
10000	0.87 ± .10	0.87 ± .10	0.87 ± .10	0.86 ± .10	0.93 ±0.08

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.
- Stat. T. is the abbreviation for Statistical Test.

TABLE 18

Order of Autoregressive Model Selected by Different Criteria

for the Passage Times (P)

in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method			
	AIC	*BIC*	h	Stat. Test
500	2.07 ± .62	0.73 ± .14	0.97 ± .17	0.27 ± .16
1000	1.83 ± .40	0.93 ± .11	1.03 ± .15	0.70 ± .14
2000	1.97 ± .33	1.03 ± .06	1.20 ± .15	0.93 ± .08
5000	1.93 ± .31	1.10 ± .09	1.43 ± .22	1.07 ± .08
10000	2.43 ± .41	1.27 ± .13	1.67 ± .16	1.13 ± .10

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- The *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.

TABLE 19

Order of Autoregressive Model Selected by Different Criteria

for the Response Times (R)

in a Closed Network of Queues with $N = 2$, $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $p = 0.75$

Number of Events	Univariate Autoregressive Method			
	AIC	*BIC*	*h*	Stat. Test
500	1.90 ± .53	0.83 ± .21	1.27 ± .25	0.47 ± .17
1000	2.10 ± .54	0.97 ± .15	1.20 ± .20	0.73 ± .14
2000	1.97 ± .39	1.03 ± .55	1.40 ± .19	0.93 ± .08
5000	2.30 ± .43	1.20 ± .15	1.50 ± .17	1.10 ± .12
10000	2.77 ± .42	1.50 ± .15	1.70 ± .18	1.30 ± .14

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- Point estimates are obtained by estimating β^* for the variance reduction techniques.
- *AIC* criterion was used to select the order for the multivariate autoregressive method.
- The maximum order for the autoregressive model is $K = 25$.

5.4. EXAMPLE. Queue Length Process in a Single Server Cyclic Queue

We consider a closed system consisting of two service stations shown in Figure 2. There are a fixed number of jobs N circulating in the system. The departure process of each queue is the arrival process of the other. We assume that the service times at both stations are mutually independent and have general distribution functions.

For $t \geq 0$, let $X(t)$ be the number of jobs both waiting and being served at station A at time t. The state space E of the process $\{X(t) : t \geq 0\}$ is $E = \{0, 1, \ldots, N\}$. Then the process $\{X(t) : t \geq 0\}$ is a generalized semi-Markov process (GSMP) (cf. Whitt (1980)). Let $\{(X_n, C_n) : n \geq 0\}$ be the embedded jump process for the GSMP, then we can reconstruct $\{X(t) : t \geq 0\}$ from $\{(X_n, C_n) : n \geq 0\}$. First let

$$\tau_n = \sum_{m=0}^{n-1} \min_{C_{m,i}>0} C_{m,i},$$

where $C_{m,i}$ is the value of the ith clock reading at the mth jump of $X(t)$. Then

$$X(t) = \sum_{k=0}^{\infty} \delta_t([\tau_k, \tau_{k+1}))\, X_k.$$

The only events that can occur are a service completion by A or by B. Therefore, the clock vector c is a pair recording the service times left at stations A and B respectively.

We are interested in estimating the expected number of jobs at station A when the state is in equilibrium. We have taken $N = 3$, and the service times to be gamma (2,1) for server A and gamma (3,1) for server B. We consider

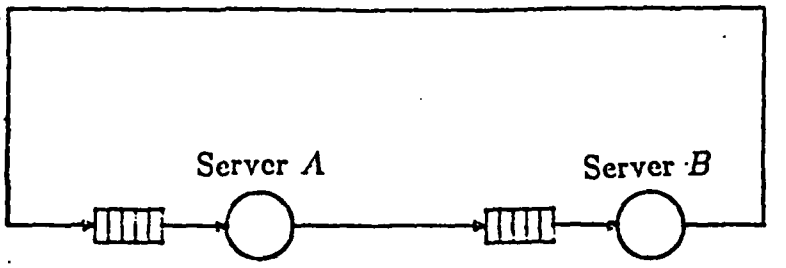


Fig. 2. Single server cyclic queue.

400, 800, 1200, 1600, 2000 cycles, the longer runs being continuations of the shorter runs. For all our runs the cycles were based on returns to the state 0. We use the discrete time method discussed in Section 3.4. All runs were replicated 30 times. The simulation results for this model are displayed in Tables 20–23.

TABLE 20

Simulation Results for Queue Length Process at Station A
in a Single Server Cyclic Queue, 3–jobs, $\Gamma(2,1)$, $\Gamma(3,1)$ servers.
Point Estimates and 90% Confidence Intervals

Number of Cycles	True Value	Simulation
100	.9002	.8691 ± .0259
300	.9002	.8971 ±.0140
500	.9002	.9028 ±.0096
800	.9002	.9042 ±.0066
1000	.9002	.9075 ±.0071

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- The maximum order for the autoregressive model is $K = 25$.

TABLE 21

Simulation Results for $\sigma^2 = 4.0771$ in a Single Server

Cyclic Queue, 3–jobs, $\Gamma(2,1)$, $\Gamma(3,1)$ Servers.

Point Estimates and 90% Confidence Intervals

Number of Cycles	Regenerative Method	Univariate Autoregressive Method			
		AIC	*BIC*	h	Stat. Test
100	3.3900 ± .3211	3.9525 ± .4073	3.8601 ± .4598	3.8960 ± .3724	3.88703 ± .4364
300	3.9643 ±.2077	4.3605 ± .3279	4.2183 ± .2059	4.2844 ± .2443	4.22976 ± .2989
500	3.9909 ±.1502	4.2345 ± .3187	4.2534 ± .1689	4.3404 ± .2345	4.1984 ± .3059
800	4.1250 ±.1265	4.2213 ± .2274	4.4250 ± .1621	4.3522 ± .2131	4.3413 ± .2092
1000	4.1558 ±.1178	4.2460 ± .1948	4.4268 ± .1506	4.3391 ± .1634	4.3069 ± .2063

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- The maximum order for the autoregressive model is $K = 25$.

TABLE 22

Simulation Results for Coverage Probability (0.90) in a Single Server Cyclic Queue, 3–jobs, $\Gamma(2,1), \Gamma(3,1)$ Servers.

Point Estimates and 90% Confidence Intervals

Number of Cycles	Regenerative Method	Univariate Autoregressive Method			
		AIC	*BIC*	h	Stat. Test
100	0.90 ±0.09	0.93 ± .08	0.93 ± .08	0.93 ± .08	0.93 ± .08
300	0.93 ±0.08	0.97 ± .06	0.97 ± .06	1.00 ± .00	0.97 ± .06
500	0.97 ±0.06	1.00 ± .00	1.00 ± .00	1.00 ± .00	1.00 ± .00
800	1.00 ±0.00	1.00 ± .00	1.00 ± .00	1.00 ± .00	1.00 ± .06
1000	0.90 ±0.09	0.97 ± .06	0.97 ± .06	0.97 ± .06	0.97 ± .06

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- The maximum order for the autoregressive model is $K = 25$.

TABLE 23

Order of Autoregressive Model selected by Different Criteria for Queue Length Process at Station A in a Single Cycle Queue, 3–jobs, $\Gamma(2,1)$, $\Gamma(3,1)$ servers.

Number of Cycles	Univariate Autoregressive Method			
	AIC	*BIC*	h	Stat. Test
100	6.10 ± 1.53	2.27 ± .33	3.20 ± .58	4.33 ± 1.69
300	12.57 ± 1.78	2.70 ± .32	5.43 ± .89	8.53 ± 1.79
500	18.10 ± 1.65	3.70 ± .58	7.70 ± 1.01	14.20 ± 1.75
800	21.93 ± 1.03	5.73 ± .76	11.90 ± 1.14	18.9 ± 1.43
1000	23.57 ± .55	6.33 ± .81	13.47 ± 1.15	21.40 ± 1.04

- Results are based on 30 independent replications; the central limit theorem for i.i.d. random variables was used to form confidence intervals.
- The maximum order for the autoregressive model is $K = 25$.

CHAPTER VI

CONCLUSIONS

As the use of computer simulation becomes more important in the study of complex phenomena, the need to develop theoretically sound and computationally efficient methods for simulation output analysis becomes more pressing. The autoregressive method proposed in this paper uses techniques developed for time series analysis to provide both point and interval estimates for parameters associated with the steady-state distribution. In this chapter we shall examine the advantages and disadvantages of the autoregressive method.

The major advantage of the autoregressive method is obvious. It serves as a black box; users provide the simulation output sequence, the black box will produce results automatically. Users need not devote time in analyzing the system as they must for the regenerative method. Furthermore, it seems that the autoregressive method applies to a much broader class of stochastic processes than the regenerative method does. With the generalization to multidimensional processes, the method enables us to apply variance reduction techniques to get more accurate point estimates along with more precise interval estimates.

The disadvantages of the autoregressive method are ~~also clear~~. First, the covariance matrix obtained by the autoregressive method is just an ap-

proximation for the covariance matrix present in the central limit theorem used to construct confidence intervals. Second, the assumptions we put on the system under study are stricter than we would like. The most important of these are the requirement that the process be ϕ–mixing, the initial distribution μ be absolutely continuous with respect to the stationary distribution π, and the Radon–Nikodym derivative $\frac{d\mu}{d\pi}$ be bounded above.

We would like to point out some areas that present potential for research and development. First, try to relax the assumptions we have made about the system under study. Second, to justiy as well as find some conditions which allow us to apply the autoregressive method for processes other than Markovian or Semi–Markovian. Finally, to design order selection criteria especially for the multidimensional autoregressive method.

BIBLIOGRAPHY

1. Achieser, N.I. (1956). *Theory of Approximation.* Frederick Ungar Publishing Co., New York.

2. Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21**, 243–247.

3. Akaike, H. (1970). Statistical prediction identification. *Ann. Inst. Statist. Math.* **22**, 203–217.

4. Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control,* **AC–19**, 716–723.

5. Akaike, H. (1977). On entropy maximization principle. *In application of statistics* (P.R. Krishnaian, ed.), 27–41. Amsterdam: North Holland.

6. Anderson, T.W. (1971). *The Statistical Analysis of Time Series,* John Wiley, New York.

7. Billingsley, P. (1968). *Convergence of Probability Measure,* John Wiley, New York.

8. Blomqvist, N.(1967). The covariance function of the $M/G/1$ queuing system. *Scand. Actuar. J.* **50**, 157–174.

9. Çinlar, E. (1975). *Introduction to Stochastic Processes.* Prentice–Hall, Englewood Cliffs, New Jersey.

10. Chung, K.L. (1974). *A Course in Probability Theory.* Academic Press, New York.

11. Crane, M.A. and Iglehart, D.L. (1975). Simulating stable stochastic systems: III, Regenerative processes and discrete event simulation. *Operations Res.* **23**, 33–45.

12. Crane, M.A. and Lemoine, A.J. (1977). *An Introduction to the Regenerative Method for Simulation Analysis.* Lecture Notes in Control and Information Sciences, **4**, Springer–Verlag, Berlin.

13. Doob, J.L. (1953). *Stochastic Processes*, John Wiley, New York.

14. Durbin, J. (1960). Estimation of parameters in time–series regression models. *J. Roy. Statist. Soc. Ser. B.*, **22**, 139–153.

15. Fishman, G.S. (1973). *Concepts and Methods in Discrete Event Digital Simulation*, John Wiley, New York.

16. Fishman, G.S. (1978). *Principles of Discrete Event Simulation*, John Wiley, New York.

17. Gaver, D.P. and Shedler, G.S. (1971). Control Variable Methods in the Simulation of a Model of a Multiprogrammed Computer System. *Nav. Res. Logist. Q.*, **18**, 435–450.

18. Gaver, D.P. and Thompson, G.L. (1973). *Programming and Probability Models in Operations Research*, Brooks/Cole Pub. Co., Montery, Calif.

19. Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, John Wiley, New York.

20. Hannan, E.J. (1970). *Multiple Time Series*, John Wiley, New York.

21. Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *J. R. Statist. Soc. B* **41**, 190–195.

22. Heidelberger, P. (1977). Variance reduction techniques for the simulation of Markov processes, I: Multiple estimates. Technical Report No. 42. Dept. of Operations Research. Stanford University. Stanford, California.

23. Hordijk, A., Iglehart, D.L. and Schassberger, R. (1976). Discrete time methods for simulating continuous time Markov chains. *Adv. Appl. Prob.* 8, 772–788.

24. Iglehart, D.L. (1978). The regenerative methol for simulation analysis. *Current Trends in Programming Methodology Vol. III: Software Engineering.* K.M. Chandy and R.T. Yeh (eds), 52–71. Prentice–Hall. Englewood Cliffs, New Jersey.

25. Iglehart, D.L. and Lewis, P.A.W. (1979). Regenerative simulation with internal controls, *J. Ass. Coput. Mach.,* **26**, 271–282.

26. Iglehart, D.L. and Shedler, G.S. (1980). *Regenerative Simulation of Response Times in Network of Queues. Lecture Notes in Control and Information Sciences,* **26**. Springer–Verlan, Berlin.

27. Kleinrock, L. (1975). *Queueing Systems Volumn I: Theory.* John Wiley and Sons, New york.

28. Kailath, T. (1974). A view of three decades of linear filtering theory, *IEEE Trans. Inform. Theory,* **IT–20**, 145–181.

29. Lavenberg, S.S., Moeller, T.L. and Welch, P.D. (1977). Control variables applied to the simulation of queueing models of computer systems. *IBM Res. Rep. RC6630,* Yorktown Heights, New York.

30. Lee, D.T.L. (1980). Canonical ladder form realizations and fast estimation algorithms, Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University. Stanford, California.

31. Quenouille, M.H. (1947). A large–sample test for the goodness of fit of autoregressive schemes, *J. R. Statist. Soc., Ser. A,* **110**, 123–129.

32. Rozanov, Y.A. (1967). *Stationary Random Processes,* Holden–Day, San Francisco.

33. Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.,* **6**, 461–464.

34. Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika,* **63**, 117–126.

35. Strang, G. (1976). *Linear Algebra and Its applications,* Academic Press, New York.

36. Whitt, W. (1980). Continuity of generalized Semi–Markov processes. *Math. Operat. Res.* **5**, 494–501.

37. Wiggins, R.A. and Robinson, E.A. (1965). recursive solution to the multichannel filtering problem. *J. Geophys. Res.,* **70**, 1885–1891.

38. Wold, H. (1954). *A Study in the Analysis of Stationary Time Series,* 2nd Edition, Almqvist and Wiksell Book Co., Uppsala.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 65	2. GOVT ACCESSION NO. AD-A123823	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN AUTOREGRESSIVE METHOD FOR SIMULATION OUTPUT ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) YUNG-LI LILY JOW		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0578
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPARTMENT OF OPERATIONS RESEARCH STANFORD UNIVERSITY STANFORD, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR 042-343)
11. CONTROLLING OFFICE NAME AND ADDRESS STATISTICS AND PROBABILITY PROGRAM OFFICE OF NAVAL RESEARCH (Code 436) ARLINGTON, VA 20360		12. REPORT DATE DECEMBER 1982 13. NUMBER OF PAGES 111
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)

Simulation, Output Analysis, Confidence Interval, Time Series Analysis, Stationary Processes, Autoregressive Processes, Spectral Density Function, Variance Reduction.

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

SEE ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Let $\{ y(n) : n \geq 0 \}$ be a d–dimensional strictly stationary process which satisfies some regularity conditions so that the central limit theorem

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n-1} y(i) - \mu\right) \Rightarrow N(0, \Sigma) \quad \text{as} \quad n \to \infty$$

holds, where $\Rightarrow$ denotes weak convergence and $N(0, \Sigma)$ is a d–dimensional normal vector with mean 0 and covariance matrix Σ. It is demonstrated that Σ can be approximated arbitrarily closely by the spectral density function $f(\lambda)$ at zero of an autoregressive process. Based on this approximation, we propose an autoregressive method to estimate the covariance matrix Σ from an observed sequence. We derive some recursive formulae to estimate the parameters for a given order autoregressive process and discuss several criteria used to select the order. Under suitable assumptions the consistency of the estimate for Σ can be established. The method is applicable to the simulation of some discrete or continuous time Markov chains and semi–Markov processes. Here the central limit theorem above can be used to construct confidence intervals for parameters associated with the steady–state behavior of these processes. The method can also be justified for certain non–stationary processes. Several variance reduction techniques are developed which enable us to shorten the confidence intervals constructed.


3-8
DT