The Annals of Statistics
1979, Vol. 7, No. 3, 671–679

NONPARAMETRIC ESTIMATION OF MARKOV TRANSITION FUNCTIONS[1]

BY SIDNEY YAKOWITZ

University of Arizona

Let $\{X_n\}$ be a Markov chain having a stationary transition function and assume that the state set is an arbitrary set in a Euclidean space. The state transition law of the chain is given by a function $F(y|x) = P[X_{n+1} \leqslant y|X_n = x]$, which is assumed defined and continuous for all x. In this paper we give a statistical procedure for determining a function $F_n(y|x)$ on the basis of the sample $\{X_j\}^n{}_{j=1}$, $n = 1, 2, \cdots$, and prove that if the chain is irreducible, aperiodic, and possesses a limiting distribution π, then with probability 1, $\sup_y|F_n(y|x) - F(y|x)| \to_n 0$ for every x such that any open sphere containing x has positive π probability. This result improves upon a study by Roussas which gives only weak convergence. We demonstrate that a certain clustering algorithm is useful for obtaining efficient versions of our estimates. The potential value of our methods is illustrated by computer studies using simulated data.

1. Introduction. Let $\{X_n\}_{n\geqslant 0}$ be a vector-valued, irreducible, aperiodic Markov chain which has a limiting distribution π and a continuous, time-invariant state transition function F. F governs the process in the sense that for every $n > 0$, every vector y, and x

$$P[X_{n+1} \leqslant y|X_n = x] = F(y|x).$$

The text [1] is adequate background for the terminology and ideas used in this paper. Feller [4, page 264–266] gives a useful sufficient condition for a chain to have limiting distribution. According to [1], the distribution function π is a *limiting distribution* for $\{X_n\}$ if for every y,

$$P[X_n \leqslant y] \to \pi[B(y)] \qquad \text{as} \quad n \to \infty,$$

where $B(y) = \{x : x \leqslant y\}$. Let $S(x, e)$ denote the sphere of radius e centered at x. We define

$$\mathfrak{X} = \{x : \pi[S(x, e)] > 0 \text{ for every } e > 0\}.$$

The purpose of the present work is to give the construction of a function $F_n(y|x)$ from the sample $\{X_i\}_{1\leqslant i \leqslant n}$ and subsequently demonstrate that F_n is a consistent estimator for F in the sense that, with probability 1,

$$\sup_y|F_n(y|x) - F(y|x)| \to_n 0$$

for every $x \in \mathfrak{X}$. We then discuss a numerical study as well as efficient constructions of our estimator.

Received June 1977; revised February 1978.

[1] Research supported, in part, by NSF grants ENG-74-20462 and ENG-76-20280.

AMS 1970 *subject classifications.* Primary 62M05; secondary 62G05.

Key words and phrases. Markov-chain, consistent estimator, nonparametric inference, hydrologic time series.

The only related result known to us is a study by Roussas [9] which gives a slightly different construction and proves only weak convergence, and this under somewhat more restrictive conditions (the state space must be a linear set, the transition function must be differentiable, and π must dominate Lebesgue measure over the entire real line). It is not clear that this construction extends to higher dimensional state spaces. Roussas [9] asserts that a proof of strong convergence "would be desirable".

The search for statistical methods for inference of a Markov chain transition function was motivated by our attempts to model and predict daily river flow, and is part of a continuing effort in this direction [10, 11, 12, 16, 17]. Some of the hydrologic implications of the present study were presented in [14], and a companion paper [16] for a hydrologic journal is being readied. A somewhat related nonparametric method for groundwater analysis was described in [13].

2. Construction of a consistent transition function estimator. Let $\{c_i\}$ be a sequence of vectors which are dense in a Euclidean space, which is presumed to be the state space for the Markov chain. $[a]$ will denote the integer part of the real number a. X_i denotes the ith state of a Markov chain $\{X_i\}_{i>0}$. Let n be some fixed positive integer. The construction of the estimate F_n depends on certain objects defined below.

$$S_{j,n} = \left\{X_i : 1 \leqslant i \leqslant n \text{ and } \|X_i - c_j\| \leqslant \|X_i - c_k\|,\ 1 \leqslant k \leqslant n^{\frac{1}{2}}\right\}.$$

$$C_n = \left\{c_j : 1 \leqslant j \leqslant n^{\frac{1}{2}} \text{ such that } S_{j,n} \text{ has } \geqslant \left[n^{\frac{1}{3}}\right] \text{ elements}\right\};$$

$$c_n(x) = c_v \text{ where } c_v \text{ is element of least index in } C_n \text{ such that}$$
$$\|c_v - x\| \leqslant \|c_k - x\|, \text{ all } c_k \in C_n.$$

For reasons which will become clearer when we describe the efficient application of the algorithm to follow, the c_j's will sometimes be referred to as *representative states*. We will let $G_{n,c_v}(y)$ be the empirical distribution function constructed from the successor elements to those in $S_{v,n}$. Thus

$$G_{n,c_v}(y) = A(v,y)/B(v) \tag{2.1}$$

where $A(v,y)$ is the number of X_i's in $S_{v,n}$ such that $X_{i+1} \leqslant y$ and $B(v)$ is the total number of X_i's in $S_{v,n}$. We say that $X_{i+1} \leqslant y$ if each coordinate of the vector X_{i+1} is less than or equal to the corresponding coordinate of y. We will have occasion to use formula (2.1) only in cases in which $S_{v,n}$ is not empty. For arbitrary vectors x and y, we define F_n by

$$F_n(y|x) = G_{n,c_n(x)}(y). \tag{2.2}$$

THEOREM 1. *Let $\{X_i\}_{i\geqslant 0}$ be an irreducible, aperiodic Markov chain with stationary transition function $F(y|x)$ which for every y is continuous in x. Suppose further that $\{X_i\}$ has a limiting distribution, π, which for each x, dominates the* *measure associated with* $F(y|x)$. *Then, with probability* 1,

$$\sup_y |F_n(y|x) - F(y|x)| \to 0.$$

PROOF. From the analysis to follow, it will be seen that for any fixed y, $F_n(y|x) \to F(y|x)$, with probability 1. But this implies the asserted uniform convergence, as one may confirm (as noted in [9]) by looking at the details of the proof of the Glivenko-Cantelli theorem.

Under the assumption that the initial state X_1 is distributed as the limiting distribution π, $\{X_i\}$ is a stationary stochastic process and is readily seen to be ergodic (see [1] for definitions and discussion) with respect to the time shift operation. For that reason, the ergodic theorem implies that for any Borel set A, with probability 1,

$$1/n\Sigma_{1\leqslant i\leqslant n}1_A(X_i) \to \pi(A), \tag{2.3}$$

where $1_A(\cdot)$ denotes the indicator function for the event A. Our first step in the proof will be to use this fact to show that for any $x \in \mathfrak{X}$, with probability 1,

$$c_n(x) \to x. \tag{2.4}$$

Let A denote any sphere centered at x. We will affirm (2.4) by demonstrating that, with probability 1, there is some random time N such that $c_n(x) \in A$ for every $n > N$. Let M be an integer large enough that c_j is in A' for some $j < M$. A' denotes the sphere centered at x which has half the radius of A. We will assume henceforth that n is larger than M^2. Since $\pi(A') > 0$, by (2.3), with probability 1, for some random N_1 and for some $\alpha > 0$,

$$\Sigma_{i\leqslant n}1_A(X_i) > \alpha n, \qquad \text{all} \quad n > N_1,$$

and of course, eventually, $n^{\frac{1}{3}} < \alpha n$. These facts imply that for some N_2, $C_n \cap A$ is not empty for $n > N_2$. This fact establishes (2.4), inasmuch as A was an arbitrary sphere centered at x.

We next establish that for any fixed vector y, with probability 1, as $n \to \infty$,

$$\cdot|F_n(y|x) - F(y|c_n(x))| \to 0, \tag{2.5}$$

which, when we recall the assumed continuity of F in x, in conjunction with equation (2.4), implies the theorem.

Let ε be an arbitrary positive number and for each $n \geqslant 1$, E_n is defined to be the event that

$$|F_n(y|x) - F(y|c_n(x))| \geqslant \varepsilon.$$

Let A_1 be a sphere centered at x such that if $x', x'' \in A_1$, then $|F(y|x') - F(y|x'')| < \varepsilon/2$, and let N_1 be the least (random) time such that $c_n(x) \in A_1$ for all $n > N_1$.

Toward bounding $P[E_n]$, $n \geqslant N_1$, we employ a large deviation result (associated with the Laplace-DeMoivre theorem) given in [5]. Let $S(n)$ denote $S_{c_n(x), n}$. $B(n)$ is the number of elements in $S(n)$. Assume, for the moment, that $F(y|X_j) = F(y|x) = p$ for every $X_j \in S(n)$. We put superscript p on F to remind ourselves of this assumption and for later convenience of notation. $q \equiv 1 - p$. Let $\mathbf{B}$ denote the sequence of numbers $\{B(n)\}$. Then

$$\begin{aligned} P[E_n|\mathbf{B}, N_1] &= P\left[(B(n)/pq)^{\frac{1}{2}}|F_n^p(y|x) - p| > \varepsilon(B(n)/pq)^{\frac{1}{2}}|\mathbf{B}, N_1\right] \\ &< P\left[(B(n)/pq)^{\frac{1}{2}}|F_n^p(y|x) - p| > \varepsilon B(n)^{\frac{1}{10}}/ (pq)^{\frac{1}{2}}|\mathbf{B}, N_1\right]. \end{aligned}$$

One may verify that $B(n)F_n^p(y|x)$ is the sum of $B(n)$ independent Bernoulli trials having parameter p and consequently $(B(n)/pq)^{\frac{1}{2}}(F_n^p(y|x) - p)$ is identically the variable $S^*_{B(n)}$ defined in connection with the theorem on page 193 in Feller [5]. Further, we may identify $\varepsilon B(n)^{\frac{1}{10}}/(pq)^{\frac{1}{2}}$ with the quantity $x_{B(n)}$ of that theorem. It may be seen that the condition that $(x_{B(n)})^3/B(n)^{\frac{1}{2}}$ converge to 0 with increasing $B(n)$ holds, and consequently, by virtue of equation (6.7) of that theorem, for $B(n)$ large enough,

$$P[E_n|\mathbf{B}, N_1] < \exp\left[-B(n)^{\frac{1}{5}}\varepsilon^2/pq\right].$$

By construction, $B(n) \geqslant n^{\frac{1}{3}}$, and hence

$$\Sigma_{n>N_1}P[E_n|\mathbf{B}, N_1] < \textstyle\int_0^\infty \exp\left(-\varepsilon^2 y^{\frac{1}{15}}\right)dy < \infty.$$

Consequently, the Borel-Cantelli lemma implies

$$P\left[|F_n^p(y|x) - F^p(y|c_n(x))| \to 0|\mathbf{B}, N_1\right] = 1. \tag{2.6}$$

The result (2.6) holds regardless of the values of the conditioning variables $\mathbf{B}$ and N_1 and so, by taking the marginal distribution with respect to these quantities, we may conclude that (2.5) holds for the case in which $F(y|x) = p$.

To account for the variability of $F(y|X_j)$ with $X_j \in S(n)$, we note that if x is an interior point of $\mathcal{X}$, then it is evident that, with probability 1, the diameter $S(n)$ goes to 0 with increasing n. In any event, if we define $\hat{S}(n)$ to be that $(1 - n^{-\frac{1}{2}})$ proportion of the points in $S(n)$ which lie closest to $c_n(x)$ and if $\hat{G}_n$ is then defined to be the empirical distribution function constructed from the successors to the points in $\hat{S}(n)$, then, with probability 1, the diameter of $\hat{S}(n)$ converges to 0 and also $\hat{G}_n(y) - F_n(y|x) \to 0$. For these reasons, our conclusions to follow will remain valid if we make the (technically faulty) assumption that for A_1 and N_1 as used earlier in this proof,

$$|F(y|X_j) - F(y|X_k)| < \varepsilon/2$$

for every $X_j, X_k \in S(n)$, $n \geqslant N_1$. Thus for $n > N_1$, if $p_1 = \min_{x\in A_1}F(y|x), p_2 = \max_{x\in A_1}F(y|x)$, then by our earlier argument

$$\begin{aligned} P[F_n(y|x) - F(y|c_n(x)) > \varepsilon|\mathbf{B}, N_1] &\leqslant P[F_n^{p_2}(y|x) - p_1 > \varepsilon|\mathbf{B}, N_1] \\ &= P[F_n^{p_2}(y|x) - p_2 > \varepsilon + p_1 - p_2|\mathbf{B}, N_1] \\ &\leqslant P[F_n^{p_2}(y|x) - p_2 > \varepsilon/2|\mathbf{B}, N_1] \\ &\leqslant \exp\left[-1/2(\varepsilon^2/4)B(n)^{\frac{1}{5}}\right]. \end{aligned}$$

The same bound pertains to $P[F_n(y|x) - F(y|c_n(x)) < -\varepsilon|\mathbf{B}, N_1]$ and these bounds allow us to apply an earlier argument to conclude that (2.5) holds in the general case.

If $\{X_n\}$ is aperiodic and for each x, π dominates the measure associated with $F(y|x)$, then the conclusion of the theorem is valid for any initial probability for X_1. For the proof of the theorem depends on $X_1 \sim \pi$ only in order to conclude that, with probability 1, equation (2.3) holds. For any event B in the field of events of the process $\{X_i\}$, let $P_\pi(B)$ and $P_x(B)$ denote, respectively, the probability of B under the assumption that $X_1 \sim \pi$ and under the assumption that $X_1 = x$. Under the above assumptions, Theorem 7.18 and Proposition 7.12 of [1] imply that for any tail event B, and for any x,

$$P_\pi(B) = P_x(B),$$

and thus if G is any initial distribution function for X_1, since the event (2.3) is a tail event, we have $1 = P_\pi(\lim_n 1/n\Sigma_{i=1}^n 1_A(X_i) = \pi(A)) = \int P_x(\lim 1/n\Sigma 1_A(X_i) = \pi(A))dG(x)$. Although the results of [1] which we have just employed are cited for one-dimension, they are based on developments in [3] which hold for an arbitrary state space.

3. Numerical applications. Our first study consisted of analysis of the Markov chain in which $X_{n+1}|X_n = x$ is normally distributed, with mean and variance depending on x. Specifically,

$$X_{n+1}|(X_n = x) \sim N(\operatorname{sign}(x)|x|^{0.8}, 0.25/(1 + \exp(x))). \tag{3.1}$$

In (3.1), sign(x) is 1 or -1 according to whether x is positive or negative. A chain segment of 10,000 samples was simulated according to the transition function determined by (3.1) and used as the observed sample input to the transition function estimation algorithm. The normal observations were simulated by the Box-Muller algorithm (described, for example, in [15]). The set of representative states c_j, $1 \leqslant j \leqslant 100$, was the evenly spaced grid given by

$$c_j = X_{\min} + j/100(X_{\max} - X_{\min}), \qquad 1 \leqslant j \leqslant 100, \tag{3.2}$$

where $X_{\max}$ and $X_{\min}$ are respectively the maximum and the minimum of the input sample. The distributions $G_{10^4,j}$ in (2.2) were taken to be normal (rather than the empirical distribution function as per the formal description of the algorithm) with mean and variance being the sample mean and variance of $S_{j,10^4}$. For those sets $S_{j,10^4}$ having fewer than 10 members, we arbitrarily set $G_{10^4,j}$ to be normal with mean c_j and a variance of $1/2$. In our experimental studies assessing the effectiveness of various statistical procedures for Markov chains, we have found the pragmatic expedient of just comparing computer plots of simulated time series and their approximations more satisfying than various conceivable measures of "goodness of fit" for time series. Thus in Figure 1, we have presented for the reader's comparison plots of the simulated process associated with (3.1) and a simulation of its approximation obtained by using F_n constructed as above as the Markov

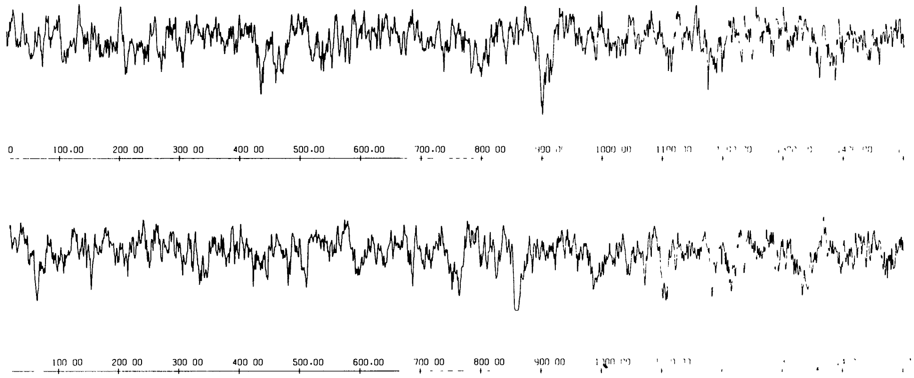

FIG. 1. *Comparison of a Markov chain* (top) *and its approximation* (*bottom*).

transition function. As an additional comparison, Table 1 gives the sample means and variances of successors to $S_{j,\,10^4}$, $50 \leqslant j \leqslant 100$ and j a multiple of 5, and compares these with their theoretical values (namely $\text{sign}(c_j)|c_j|^{0.8}$ and $0.25/(1 + \exp(c_j))$, respectively). Also in this table, we have compared the number of X_i observations in the simulated realizations of the actual and approximated Markov chains which lie closest to our tabulated values of c_j.

An autoregressive (AR) process is a very specialized Markov chain having the structure

$$X_{n+1} = \Sigma_{i=0}^{N} a_i X_{n-i} + N_n, \tag{3.3}$$

where the $\{N_n\}$ sequence is taken to be some noise process (usually uncorrelated) which is assumed independent of the X_i's, $i \leqslant n$. For example, the autoregressive moving average time series (analyzed at length in [2]) is, subsequent to its exploitation in [6], the most popular model for streamflow, although there seems to be scant physical motivation for the linear difference equation part of (3.3). (In fact, the dynamic equations of channel flow are nonlinear [8]).

Our opinion is that there are many other instances where the ARMA model is applied to patently nonlinear phenomena, such application stemming from the lack of availability of alternative methods. We are hopeful that techniques such as given in the present paper will lead to more appropriate statistical methodology for nonlinear stochastic systems.

In order to display the potential advantage of our nonparametric method over a misapplied parametric technique, we compared the predictive powers of our model against autoregressive models of various orders. Specifically, we sought to compare the performance of a minimum square error predictor for an AR model (discussed in [2]) with the least square predictor associated with our nonparametric method. One may readily see that our least square predictor of $X_{n+1}|X_n = x$ is given by the sample mean of the elements in $S_{j,\,n}$, where j is the index such that c_j is the closest

TABLE 1

A Comparison of Parameters and Statistics of Actual and Approximating Chains

j	c_j representative state	No. of x_j's nearest c_j in actual chain	No. of x_j's nearest c_j in approx. chain	Sample mean of successors to states in S_j, n	Theoretic mean of successor to c_j (Sign $(c_j)(c_j)^{0.8}$)	Sample var. successors to states in S_j, n	Theoretical Var. of successor to c_j $(0.25/(H \exp(c_j))$
50	.642	8	12	.734	.701	.092	.086
55	.718	37	28	.767	.767	.088	.081
60	.793	125	97	.826	.831	.076	.077
65	.869	225	218	.887	.894	.081	.073
70	.945	471	423	.960	.956	.070	.069
75	1.021	556	584	1.016	1.016	.064	.066
80	1.096	406	431	1.077	1.076	.061	.062
85	1.172	172	162	1.143	1.135	.057	.059
90	1.248	17	18	1.158	1.194	.047	.055
95	1.324	2	1	1.259	1.251	.037	.052

representative state to x. In this study, we chose as the transition function for the underlying Markov chain the (highly nonlinear) rule

$$X_{n+1}|(X_n = x) = \sin(\pi x) + N_n, \tag{3.4}$$

where the N_n's are independent and identically distributed.

With probability 0.8, $N_n = 0$ and with probability 0.2, N_n is chosen uniformly from the interval $[-1/2, 1/2]$. We note that with this process, nothing is to be gained by using an autoregressive moving average approximation in place of the purely AR process, since the noise actually is uncorrelated. The AR parameters and the approximating transition function F_n of our procedure were both inferred from the same 20,000 points of a simulated chain. An additional 1,500 observations of the underlying chain were simulated and for each time n and for each predictor, the quantity $(X_{n+1} - \hat{X}_{n+1})^2$ was calculated, and the root mean square (rms) error $(\Sigma_n(X_{n+1} - \hat{X}_{n+1})^2/1500)^{\frac{1}{2}}$ tabulated. X_{n+1} was the observed value of the n + 1st Markov state, and $\hat{X}_{n+1}$ was the predicted value based on observations X_n, or, in the case of an rth order AR process, $\hat{X}_{n+1}$ depended on the r-tuple $(X_n, \cdots, X_{n-r+1})$ of observed states. The representative vectors c_j, $1 \leqslant j \leqslant 100$, for our transition function approximation; were again chosen by rule (3.2).

The optimum one step predictor for this process is, of course,

$$\hat{X}_{n+1} = \sin(\pi X_n).$$

For this predictor, the square root of the expected error is $(E[N_n^2])^{\frac{1}{2}} = (0.2/12)^{\frac{1}{2}} = 0.129$. The observed rms error for this optimum predictor calculated from the 1500 observation periods of the chain was 0.132. For the predictor based on our nonparametric transition function estimate, the rms error was 0.135. The AR predictor of order $N = 1, 2$ and 3 (calibrated according to the standard methods described, for example, in [2]) gave rms errors of, respectively, 0.534, 0.526 and 0.522.

4. Efficient multivariate estimation. Let us suppose that we have decided to use M representative vectors $\{c_1, \cdots, c_M\}$ for inferring F_n from the observed multivariate Markov chain segment $\{X_1, \cdots, X_n\}$. We describe in this section a largely heuristic principle which we have found useful in our computer studies. To motivate this principle, suppose that for some fixed y we are trying to minimize $F(y|x) - F_n(y|x)$. We may intuitively anticipate that the smaller the magnitude of $\|x - c_n(x)\|$, the closer will be our approximation. Thus in keeping with this principle, ideally one would like to choose $\{c_1, \cdots, c_M\}$ to minimize $E[\|X - c_n(X)\|^2]$, where X is the π-distributed random vector. By the ergodicity of the chain, π may be approximated by H_n, where H_n is the empirical distribution function constructed from the chain segment $\{X_i\}_{1\leqslant i \leqslant n}$. A reasonable procedure is, therefore, to minimize the expectation $E[\|X - c_n(X)\|^2]$ with the expectation associated with H_n instead of π. This procedure is equivalent to minimizing expression

$$J(c_i, \cdots, c_M) = \Sigma_{i=1}^n(X_i - c_n(X_i))^2. \tag{4.1}$$

It turns out that for even moderately large values of M and n, it is not computationally feasible to find the minimizing set $\{c_1, \cdots, c_M\}$. However, K-means algorithm, which is described in Chapter 4 of Hartigan's book [7], is designed to provide a computationally feasible approximation to the set $\{c_1, c_2, \cdots, c_M\}$ of representative states which minimize J. In [14, 16] we found the K-means algorithm to be useful for computing representative vectors in our analysis of daily flow of the Cheyenne River. We refer the reader to [16] for a discussion of the many details of this application.

Acknowledgment. The author gratefully acknowledges helpful conversations pertaining to this work with Professors J. Denny, J. Krimmel and F. Szidarovszky. Richard Creps ably accomplished the programming tasks associated with this project.

REFERENCES

[1] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading Mass.

[2] Box, G. and Jenkins, G. (1970). *Time Series Analysis, Forecasting, and Control*. Holden-Day, San Francisco.

[3] Doob, J. (1948). Asymptotic properties of Markov transition probabilities. *Trans. Amer. Math. Soc.* **63** 393–421.

[4] Feller, W. (1966). *An Introduction to Probability Theory and Its Application, Volume II*. Wiley, New York.

[5] Feller, W. (1968). *An Introduction to Probability Theory and Its Application, Volume I, 3rd edition*. Wiley, New York.

[6] Fiering, M. (1967). *Streamflow Synthesis*. Harvard Univ. Press, Cambridge, Mass.

[7] Hartigan, J. (1975). *Clustering Algorithms*. Wiley, New York.

[8] Henderson, F. (1966). *Open Channel Flow*. The Macmillan Company, New York.

[9] Roussas, G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40** 1386–1400.

[10] Yakowitz, S. (1972). A statistical model for daily streamflow records with application to the Rillito River. *Proceedings International Symposium on Uncertainties in Hydrologic and Water Resource Systems*, 273–283 Univ. Arizona, Tucson.

[11] Yakowitz, S. (1973). A stochastic model for daily river flows in an arid region. *Water Resources Research* **9** 1271–1285.

[12] Yakowitz, S. (1976^a). Small sample hypothesis tests of Markov order, with application to simulated and hydrologic chains. *J. Amer. Statist. Assoc.* **71** 132–136.

[13] Yakowitz, S. (1976^b). Model-free statistical methods for water table prediction. *Water Resources Research* **12** 836–844.

[14] Yakowitz, S. (1977^a). Statistical models and methods for rivers in the southwest. *Proceedings, 21st Annual Meeting Arizona Academy of Sciences,* Las Vegas, Nevada.

[15] Yakowitz, S. (1977^b). *Computational Probability and Simulation*. Addison-Wesley, Reading, Mass.

[16] Yakowitz, S. (1977^c). A nonparametric Markov model for daily river flow. *Water Resources Research*. To appear.

[17] Yakowitz, S. and Denny, J. (1973). On the statistics of hydrologic time series. *Proceedings, 17th Annual Meeting Arizona Academy of Sciences* **3** 146–163. Tucson, Arizona.

Department of Systems and Industrial Engineering
University of Arizona
Tucson, Arizona 85721