

2

SECOND-ORDER MOMENTS OF A STATIONARY MARKOV CHAIN AND SOME APPLICATIONS

T. W. Anderson
Stanford University

AD-A206 366

TECHNICAL REPORT NO. 1

February 1989

U. S. Army Research Office

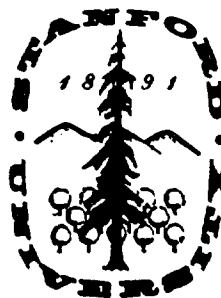
~~Contracts DAAG29-85-K-0230 and DAAL03-89-K-0033~~

Theodore W. Anderson, Project Director

Department of Statistics
Stanford University
Stanford, California

DTIC
ELECTE
S 22 MAR 1989 D
E

Approved for Public Release; Distribution Unlimited.



89 3 21 024

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 22	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) SECOND-ORDER MOMENTS OF A STATIONARY MARKOV CHAIN AND SOME APPLICATIONS		5. TYPE OF REPORT & PERIOD COVERED Technical report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T.W. Anderson		8. CONTRACT OR GRANT NUMBER(s) DAAG29-85-K-0239 DAAL03-89-K-0033
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Department - Sequoia Hall Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE February 1989
		13. NUMBER OF PAGES 13pp.
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) NA		
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Stationary Markov chain, second-order moments, estimation of stationary probabilities, simulation, first-order autoregression		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) See reverse side for Abstract.		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. ABSTRACT.

The i -th state of a finite-state Markov chain can be indicated by a vector with 1 in the i -th position and 0's in the other positions. The vector-valued process defined in this way is autoregressive in the wide sense. The second-order moments, moving average representation, and spectral density of this process are obtained. The numerical-valued chain is a linear function of the vector; a simple condition is derived for it to be first-order autoregressive in the wide sense. The stationary probabilities of the chain can be estimated by the mean of observations on the vector-valued process; this estimator is asymptotically equivalent to the maximum likelihood estimator.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A-1	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

1. Introduction.

A finite-state Markov chain is a stochastic process in which the variable takes on one of a finite number of values. Although the values may be numerical, they need not be; they may be simply states or categories. If they are given numerical values, the values are not necessarily the first so many integers, as in the number of customers waiting in a queue. When the chain is described in terms of states, it is convenient for many purposes to treat the chain as a vector-valued process. The vector has 1 in the position corresponding to the given state and 0 in the other position. Then the vector-valued process is first-order autoregressive in the wide sense when the Markov chain is first-order. Anderson (1979a), (1979b), (1980) pointed out analogies between Gaussian autoregressive processes and Markov chains in terms of moments, sufficient statistics, tests of hypotheses, etc.

In this paper the consequences of the autoregressive structure of the vector-valued process are developed further to yield various second-order moments and the spectral density of the process. It is shown that using the mean of this process as an estimator of the stationary probabilities is asymptotically equivalent to the maximum likelihood estimator and is asymptotically efficient (Section 4). The numerical-valued Markov chain is considered as a linear function of the vector-valued process, and a simple condition is obtained for it to be a wide-sense first-order autoregressive process (Section 5).

2. A Stationary Markov Chain.

A stationary Markov chain $\{x_t\}$ with discrete time parameter and states $1, \dots, m$ is defined by the transition probabilities

$$(1) \quad \Pr\{x_t = j | x_{t-1} = i, x_{t-2} = k, \dots\} = p_{ij}, \quad i, j, k, \dots = 1, \dots, m, \quad t = \dots, -1, 0, 1, \dots,$$

where $p_{ij} \geq 0$ and $\sum_{j=1}^m p_{ij} = 1$. Let $\Pr\{x_t = i\} = p_i$, $i = 1, \dots, m$ ($p_i \geq 0$ and $\sum_{i=1}^m p_i = 1$). Then

$$(2) \quad \sum_{i=1}^m p_i p_{ij} = p_j, \quad j = 1, \dots, m.$$

If $\mathbf{P} = (p_{ij})$ and $\mathbf{p} = (p_i)$, a column vector, then the above properties can be written as

$$(3) \quad \mathbf{p}'\mathbf{P} = \mathbf{p}', \quad \mathbf{P}\boldsymbol{\epsilon} = \boldsymbol{\epsilon}, \quad \mathbf{p}'\boldsymbol{\epsilon} = 1,$$

where $\boldsymbol{\epsilon} = (1, 1, \dots, 1)'$. Let $\boldsymbol{\epsilon}_i$ be the m -component vector with 1 in the i -th position and 0's elsewhere, and let $\{\mathbf{z}_t\}$ be a sequence of m -component random vectors. Then the Markov chain can be written

$$(4) \quad \Pr\{\mathbf{z}_t = \boldsymbol{\epsilon}_j | \mathbf{z}_{t-1} = \boldsymbol{\epsilon}_i, \mathbf{z}_{t-2} = \boldsymbol{\epsilon}_k, \dots\} = p_{ij}, \quad i, j, k, \dots = 1, \dots, m, \quad t = \dots, -1, 0, 1, \dots$$

Note that $\mathbf{z}_t$ has 1 in one position and 0 in the others; hence $\epsilon' \mathbf{z}_t = 1$.

3. Second-order Moments.

It follows from the model that

$$(5) \quad \mathcal{E}(z_{jt} | \mathbf{z}_{t-1} = \epsilon_i, \mathbf{z}_{t-2} = \epsilon_k, \dots) = p_{ij},$$

where z_{jt} is the j -th component of $\mathbf{z}_t$. We can write (5) in vector form as

$$(6) \quad \mathcal{E}(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots) = \mathbf{P}' \mathbf{z}_{t-1}.$$

Let

$$(7) \quad \mathbf{v}_t = \mathbf{z}_t - \mathbf{P}' \mathbf{z}_{t-1}$$

be the t -th disturbance. Then

$$(8) \quad \begin{aligned} \mathcal{E} \mathbf{v}_t &= \mathcal{E} \{ \mathcal{E}[(\mathbf{z}_t - \mathbf{P}' \mathbf{z}_{t-1}) | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots] \} \\ &= \mathbf{0}. \end{aligned}$$

In (8) the outer expectation is with respect to $\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots$. Similarly

$$(9) \quad \begin{aligned} \mathcal{E} \mathbf{v}_t \mathbf{z}_{t-s}' &= \mathcal{E} [\mathcal{E}(\mathbf{v}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots) \mathbf{z}_{t-s}'] \\ &= \mathbf{0}, \quad s = 1, 2, \dots \end{aligned}$$

Since $\mathbf{v}_{t-s} = \mathbf{z}_{t-s} - \mathbf{P}' \mathbf{z}_{t-s-1}$,

$$(10) \quad \mathcal{E} \mathbf{v}_t \mathbf{v}_{t-s}' = \mathbf{0}, \quad s = 1, 2, \dots$$

Thus $\{\mathbf{v}_t\}$ is a sequence of uncorrelated random vectors.

We can iterate $\mathbf{z}_t = \mathbf{P}' \mathbf{z}_{t-1} + \mathbf{v}_t$ to obtain

$$(11) \quad \mathbf{z}_t = \mathbf{v}_t + \mathbf{P}' \mathbf{v}_{t-1} + \dots + (\mathbf{P}')^{s-1} \mathbf{v}_{t-s+1} + (\mathbf{P}')^s \mathbf{z}_{t-s}.$$

Then

$$(12) \quad \mathcal{E}(\mathbf{z}_t | \mathbf{z}_{t-s}, \mathbf{z}_{t-s-1}, \dots) = (\mathbf{P}')^s \mathbf{z}_{t-s}.$$

This conditional expected value could alternatively be obtained from the fact that the transition probabilities from $\mathbf{x}_{t-s}$ to $\mathbf{x}_t$ are the elements of $\mathbf{P}^s$.

Since $\{x_t\}$ is stationary, $\{z_t\}$ is stationary and z_t has a marginal multinomial distribution with probabilities $p_1, \dots, p_m$. Hence, the expectation of z_t is

$$(13) \quad \mathcal{E} z_t = p,$$

and the covariance matrix is

$$(14) \quad \text{Var}(z_t) = Dp - pp' = V,$$

say, where Dp is a diagonal matrix with i -th diagonal element p_i . From (11) we also find

$$(15) \quad \mathcal{E} z_t z'_{t-s} = (P')^s \mathcal{E} z_t z' = (P')^s Dp, \quad s = 0, 1, \dots$$

Since $\mathcal{E} z_t = \mathcal{E} z_{t-s} = p$, and $(P')^s p = p$, the covariance matrix between z_t and z_{t-s} is

$$(16) \quad \text{Cov}(z_t, z'_{t-s}) = (P')^s V, \quad s = 0, 1, \dots$$

Thus (14) and (16) determine the second-order moments of $\{z_t\}$.

The conditional covariance matrix of z_t and v_t is

$$(17) \quad \begin{aligned} \text{Var}(z_t | z_{t-1} = \epsilon_i, z_{t-2}, \dots) &= \text{Var}(v_t | z_{t-1} = \epsilon_i, z_{t-2}, \dots) \\ &= Dp_i - p_i p'_i \\ &= V_i, \end{aligned}$$

say, where Dp_i is a diagonal matrix with p_{ij} as the j -th diagonal element and

$$(18) \quad P = \begin{bmatrix} p'_1 \\ p'_2 \\ \vdots \\ p'_m \end{bmatrix}.$$

From this conditional variance of v_t (which has conditional mean value 0) we find the (marginal) covariance matrix of v_t as

$$(19) \quad \begin{aligned} \mathcal{E} v_t v'_t &= \sum_{i=1}^m p_i V_i \\ &= \sum_{i=1}^m p_i (Dp_i - p_i p'_i) \\ &= Dp - \sum_{i=1}^m p_i p_i p'_i. \end{aligned}$$

by (2). Note that

$$(20) \quad \begin{aligned} P' V P &= P' Dp P - P' p p' P \\ &= \sum_{i=1}^m p_i p_i p'_i - p p'. \end{aligned}$$

Thus

$$(21) \quad V = P'VP + \sum_{i=1}^m p_i V_i.$$

The second-order moments of $\{z_t\}$ are the second-order moments of a first-order autoregressive process with coefficient matrix P' and disturbance covariance matrix $\sum_{i=1}^m p_i V_i$. (See Anderson (1971), Sections 5.2 and 5.3, for example.) However, the conditional covariance matrix of v_t given z_{t-1} depends on z_{t-1} . This fact shows that v_t and z_{t-1} are dependent, though uncorrelated.

Let $\lambda_1 = 1, \lambda_2, \dots, \lambda_m$ the characteristic roots of P , $t_1 = \epsilon, t_2, \dots, t_m$ be the corresponding right-sided (column) characteristic vectors, and $w'_1 = p', w'_2, \dots, w'_m$ be the corresponding left-sided (row) characteristic vectors. It is assumed that there are m linearly independent right- and left-sided vectors (that is, that the elementary divisors of P are simple). Let

$$(22) \quad \Lambda = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \Lambda_2 \end{pmatrix},$$

$$(23) \quad Q = (p, q_2, \dots, q_m) = (p, Q_2),$$

$$(24) \quad T = (\epsilon, t_2, \dots, t_m) = (\epsilon, T_2).$$

Normalize the vectors so $T'Q = I$; that is, $T' = Q^{-1}$ and $Q' = T^{-1}$. Then

$$(25) \quad P = T\Lambda Q' = T\Lambda T^{-1},$$

$$(26) \quad P^s = T\Lambda^s Q' = T\Lambda^s T^{-1}.$$

Since $\epsilon'V = \epsilon'(Dp - pp') = 0$,

$$(27) \quad \begin{aligned} (P')^s V &= Q\Lambda^s T'V \\ &= (p, Q_2) \begin{pmatrix} 1 & 0 \\ 0 & \Lambda_2^s \end{pmatrix} \begin{pmatrix} \epsilon' \\ T_2' \end{pmatrix} (Dp - pp') \\ &= Q_2 \Lambda_2^s T_2' Dp \\ &= (P' - p\epsilon')^s Dp, \quad s = 1, 2, \dots \end{aligned}$$

We shall assume that the chain is irreducible and aperiodic. Then $|\lambda_i| < 1$, $i = 2, \dots, m$. As s increases, the covariance function decreases as a linear combination of $\lambda_2^s, \dots, \lambda_m^s$.

We can write z_t in the moving average representation

$$(28) \quad z_t = \sum_{s=0}^{\infty} (P')^s v_{t-s}.$$

Since $(P')^s v_{t-s} = Q_2 \Lambda_2^s T_2' v_{t-s}$, $s = 1, 2, \dots$, (28) converges. The representation (28) is trivial, however, because

$$(29) \quad \begin{aligned} (P')^s v_{t-s} &= (P')^s (z_{t-s} - P' z_{t-s-1}) \\ &= (P')^s z_{t-s} - (P')^{s+1} z_{t-(s+1)}. \end{aligned}$$

The spectral density of $\{z_t\}$ for $\lambda \neq 0$ is (Hannan (1970), p. 67, for example)

$$(30) \quad \begin{aligned} &(I - P'e^{i\lambda})^{-1} V (I - P'e^{-i\lambda})^{-1} \\ &= [Q(I - \Lambda e^{i\lambda})T']^{-1} V [T(I - \Lambda e^{-i\lambda})Q']^{-1} \\ &= (T')^{-1} (I - \Lambda e^{i\lambda})^{-1} Q^{-1} V (Q')^{-1} (I - \Lambda e^{-i\lambda})^{-1} T^{-1} \\ &= Q(I - \Lambda e^{i\lambda})^{-1} T' V T (I - \Lambda e^{-i\lambda})^{-1} Q' \\ &= Q_2 (I - \Lambda_2 e^{i\lambda})^{-1} T_2' D_p T_2 (I - \Lambda_2 e^{-i\lambda})^{-1} Q_2'. \end{aligned}$$

Since $|\lambda_i| < 1$, $i = 2, \dots, m$, $I - \Lambda_2 e^{i\lambda}$ and $I - \Lambda_2 e^{-i\lambda}$ are nonsingular for all real λ .

4. Estimation of the Stationary Probabilities.

Consider a sequence of observations on the chain, $x_1, \dots, x_N$. These define a sequence $z_1, \dots, z_N$ from the process $\{z_t\}$. Let

$$(31) \quad S = \sum_{t=1}^N z_t.$$

Then $\mathcal{E}S = Np$. The covariance matrix of S is

$$(32) \quad \begin{aligned} &\mathcal{E}(S - Np)(S - Np)' \\ &= \sum_{t,s=1}^N \mathcal{E}(z_t - p)(z_s - p)' \\ &= \sum_{t=1}^N \sum_{s=1}^{t-1} \mathcal{E}(z_t - p)(z_s - p)' + \sum_{s=1}^N \sum_{t=1}^{s-1} \mathcal{E}(z_t - p)(z_s - p)' + \sum_{t=1}^N \mathcal{E}(z_t - p)(z_t - p)' \\ &= \sum_{t=1}^N \sum_{r=1}^{t-1} [\mathcal{E}(z_t - p)(z_{t-r} - p)' + \mathcal{E}(z_{t-r} - p)(z_t - p)'] + NV \\ &= \sum_{t=1}^N \sum_{r=1}^{t-1} [(P')^r V + V P^r] + NV \end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^N \sum_{r=1}^{t-1} [W_2 \Lambda_2^r T_2' D_p + D_p T_2 \Lambda_2^r W_2'] \\
&= \sum_{t=1}^N [W_2 (I - \Lambda_2)^{-1} \Lambda_2 (I - \Lambda_2^{t-1}) T_2' D_p \\
&\quad + D_p T_2 (I - \Lambda_2^{t-1}) \Lambda_2 (I - \Lambda_2)^{-1} W_2'] + NV \\
&= N [W_2 (I - \Lambda_2)^{-1} \Lambda_2 T_2' D_p + D_p T_2 \Lambda_2 (I - \Lambda_2)^{-1} W_2'] \\
&\quad - W_2 (I - \Lambda_2)^{-2} \Lambda_2 (I - \Lambda_2^N) T_2' D_p - D_p T_2 (I - \Lambda_2^N) \Lambda_2 (I - \Lambda_2)^{-1} W_2'] \\
&\quad + NV.
\end{aligned}$$

Then the covariance matrix of

$$(33) \quad \sqrt{N} \bar{z} = \frac{1}{\sqrt{N}} S$$

is

$$\begin{aligned}
(34) \quad &W_2 (I - \Lambda_2)^{-1} T_2' D_p + D_p T_2 (I - \Lambda_2)^{-1} W_2' - D_p + pp' \\
&- \frac{1}{N} [W_2 (I - \Lambda_2)^{-2} \Lambda_2 (I - \Lambda_2^N) T_2' D_p + D_p T_2 (I - \Lambda_2^N) \Lambda_2 (I - \Lambda_2)^{-1} W_2'].
\end{aligned}$$

Theorem.

$$(35) \quad \sqrt{N}(\bar{z} - p) \xrightarrow{L} N[0, W_2 (I - \Lambda_2)^{-1} T_2' D_p + D_p T_2 (I - \Lambda_2)^{-1} W_2' - D_p + pp'].$$

The sum S is the vector of frequencies of the states $1, \dots, m$ being observed, and $\bar{z}$ is the vector of relative frequencies. Since $\mathcal{E} \bar{z} = p$, the vector $\bar{z}$ is an estimator of p . Grenander (1954), Rosenblatt (1956), and Grenander and Rosenblatt (1957) showed that in the case of a scalar process with expected value a linear function of exogenous series and a stationary covariance function the least squares estimator of this linear function is asymptotically efficient among all linear estimators. (See also Anderson (1971), Section 10.2.) The result holds as well for vector processes. In particular, the mean of a set of observations is an asymptotically efficient linear estimator of the constant mean of a wide-sense stationary process, as is the case here.

An alternative method of estimating p from the data $z_1, \dots, z_N$ is to estimate P and find its left-sided characteristic vector corresponding to the characteristic root 1. If z_1 is given (that is, fixed), the maximum likelihood estimator of P is

$$(36) \quad \hat{P} = \left(\sum_{t=2}^N z_{t-1} z_{t-1}' \right)^{-1} \sum_{t=2}^N z_{t-1} z_t'.$$

(See Anderson and Goodman (1957).) Note that

$$(37) \quad \sum_{t=2}^N \mathbf{z}_{t-1} \mathbf{z}'_{t-1} = \begin{pmatrix} \sum_{t=1}^{N-1} z_{1t} & 0 & \dots & 0 \\ 0 & \sum_{t=1}^{N-1} z_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_{t=1}^{N-1} z_{mt} \end{pmatrix}$$

is diagonal. In fact, the diagonal elements of (37) are the components of $\mathbf{S} - \mathbf{z}_N$. If there is no absorbing state, every component of $\mathbf{p}$ is positive and the probability that (37) is nonsingular approaches 1 as $N \rightarrow \infty$ and $\hat{\mathbf{P}}$ is well-defined. Since $\mathbf{P}$ is a consistent estimator of $\mathbf{P}$, as $N \rightarrow \infty$ the probability approaches 1 that $\epsilon' \hat{\mathbf{p}} = 1$ and

$$(38) \quad \hat{\mathbf{p}}' \hat{\mathbf{P}} = \hat{\mathbf{p}}$$

have a unique solution for $\hat{\mathbf{p}}$.

We observe that

$$(39) \quad \begin{aligned} (\bar{\mathbf{z}} - \frac{1}{N} \mathbf{z}_N)' \hat{\mathbf{P}} &= \frac{1}{N} \epsilon' \sum_{t=2}^N \mathbf{z}_{t-1} \mathbf{z}'_t \\ &= \frac{1}{N} \sum_{t=2}^N \mathbf{z}'_t \\ &= (\bar{\mathbf{z}} - \frac{1}{N} \mathbf{z}_1)'. \end{aligned}$$

From (38) and (39) we obtain

$$(40) \quad \begin{aligned} N^{a-1}(\mathbf{z}'_N \hat{\mathbf{P}} - \mathbf{z}'_1) &= N^a(\bar{\mathbf{z}} - \hat{\mathbf{p}})'(\hat{\mathbf{P}} - \mathbf{I}) \\ &= N^a(\bar{\mathbf{z}} - \hat{\mathbf{p}})' \hat{\mathbf{T}}_2(\hat{\mathbf{A}}_2 - \mathbf{I}) \hat{\mathbf{Q}}'_2, \end{aligned}$$

where

$$(41) \quad \hat{\mathbf{P}} = (\epsilon, \hat{\mathbf{T}}_2) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{A}}_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{p}}' \\ \hat{\mathbf{W}}'_2 \end{pmatrix}.$$

Since the left-hand side of (40) approaches $\mathbf{0}$ as $N \rightarrow \infty$, we deduce that

$$(42) \quad N^a(\bar{\mathbf{z}} - \hat{\mathbf{p}}) \xrightarrow{\mathbf{P}} \mathbf{0}.$$

The two estimators of $\mathbf{p}$ are asymptotically equivalent. See, also, Henry (1970).

In many situations a sample $x_1, \dots, x_N$ (or equivalently, $\mathbf{z}_1, \dots, \mathbf{z}_N$) is not drawn from a stationary process, but from a process starting with some given state. In that case one would might discard enough initial observations to ensure that over the remaining period of observation

the process is stationary or almost stationary. Thus the use of $\bar{z}$ calculated from the remaining observations as an estimator of p would waste the initial observations. (If one wanted to run a simulation and insisted on *independent* observations, one would start the process with a possibly random state, let it run until stationarity is achieved, and make one observation. Then one would repeat the procedure. Clearly this is expensive in computational resources and is unnecessary.)

To estimate P and then p does not require stationarity; one can start from any state. The estimator of P is a maximum likelihood estimator, and hence the estimator of p (and Λ_2 , W_2 , and T_2) is maximum likelihood and asymptotically efficient. (Several different sequences from the same chain can be aggregated; see Anderson and Goodman (1957).)

The investigator will typically want to estimate the asymptotic covariance matrix of the estimator, given in (35), which involves

$$(43) \quad \begin{aligned} D_p T_2 (I - \Lambda_2)^{-1} Q'_2 &= D_p \sum_{j=2}^m t_j (1 - \lambda_j)^{-1} q'_j \\ &= D_p \sum_{j=2}^m \sum_{s=0}^{\infty} \lambda_j^s t_j q'_j. \end{aligned}$$

The rate of convergence is governed by the root that is largest in absolute value. The asymptotic covariance matrix can also be written as

$$(44) \quad D_p \sum_{s=0}^{\infty} (P' - \epsilon p')^s + [D_p \sum_{s=0}^{\infty} (P' - \epsilon p')^s]' - D_p + p p'.$$

Since

$$(45) \quad (P - \epsilon p')^s = P^s - \epsilon p', \quad s = 0, 1, \dots,$$

the covariance matrix (44) is

$$(46) \quad D_p \sum_{s=0}^{\infty} P^s + D_p \left(\sum_{s=0}^{\infty} P^s \right)' - D_p - p p'.$$

The infinite sum can be approximated by a finite sum since the sum is convergent. For an estimator P and p are replaced by $\hat{P}$ and $\hat{p}$, respectively.

A possible computational method is to power P until the rows of P^s are similar enough, that is, until P^s is similar to $\epsilon p'$. Then (46) follows. Note that at each step only P^s and $\sum_{s=0}^r P^s$ need be held in memory.

As a way of simulating a probability distribution of a finite set of outcomes, Persi Diaconis (personal communication) has suggested setting up a Markov chain with this probability distribution as the stationary probability distribution. An example of particular interest is *simulating the*

uniform distribution of matrices with nonnegative integer entries and fixed column and row sums (contingency tables). Diaconis and Efron (1986) have discussed the analysis of two-way tables based on this model.

Diaconis sets up the following procedure to generate a Markov chain in which all possible tables (with assigned marginals) are the states of the chain. At each step select a pair of rows (g, h) and columns (i, j) at random; with probability $\frac{1}{2}$ add 1 to the g, i -th cell and to the h, j -th cell and subtract 1 from the g, j -th cell and from the h, i -th cell. If a step would lead to a negative entry, cancel it. Each step leaves the row and column sums as given and so determines the transition probabilities of a Markov chain with the possible tables as states. Since the matrix of transition probabilities is doubly stochastic (the probability of going from one state to another is the same as the probability of the reverse), the stationary probabilities are uniform; that is, p is proportional to ϵ .

Diaconis and Efron were interested in the distribution of the χ^2 goodness-of-fit statistic, say $T(\mathbf{z})$, where $\mathbf{z}$ is the vector representation of the two-way tables; that is, they want $\Pr\{T(\mathbf{Z}) \leq \tau\}$ for arbitrary $\tau \in [0, \infty)$. Given $p_i = \Pr\{\mathbf{Z} = \epsilon_i\}$ the probability can be calculated

$$(47) \quad \Pr\{T(\mathbf{Z}) \leq \tau\} = \sum_{i=1}^m p_i I[T(\epsilon_i) \leq \tau],$$

where $I(\cdot)$ is the indicator function; that is, $I[T(\epsilon_i) \leq \tau] = 1$ if $T(\epsilon_i) \leq \tau$, and $= 0$ if $T(\epsilon_i) > \tau$. Given a sample $\mathbf{z}_1, \dots, \mathbf{z}_N$, the probability (47) is estimated by

$$(48) \quad \frac{1}{N} \sum_{t=1}^N I[T(\mathbf{z}_t) \leq \tau].$$

The estimator of the cdf of $T(\mathbf{Z})$ is the empirical cdf of $T(\mathbf{Z})$ although $\mathbf{z}_1, \dots, \mathbf{z}_N$ are not independent. However, the sample variance of $I[T(\mathbf{z}_t) \leq \tau]$ divided by N is not an estimator of the variance of (48). If a subset of $\mathbf{z}_1, \dots, \mathbf{z}_N$ is taken, say, $\mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_n}$, so that $\min |t_i - t_j|$ is great enough that $\mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_n}$ can be considered independent, the sample variance of $I[T(\mathbf{z}_{t_i}) \leq \tau]$ provides an estimator of a lower bound to the variance of (48).

If the specified row and column totals are such that each possible value of T can come from only one table, there is a 1 - 1 correspondence between the statistic and the table. That is, the value of the statistic is simply another label for the state. Then the generation of the statistic is given by the Markov chain, and its empirical cdf is an asymptotically efficient estimator of the distribution of T . These facts suggest that even if there is not a 1 - 1 correspondence between the statistic and the table, the empirical cdf is an asymptotically efficient estimator.

5. Scoring a Markov chain.

Suppose each state is assigned a numerical value α_i , $i = 1, \dots, m$. Let $y_t = \alpha_i$ if and only if x_t is in state i . For example, the state of a queue may be indicated by the number of customers

waiting. In that case it is convenient to label the states $0, 1, \dots, N$ and set $\alpha_i = i$; here N is the maximum number of customers who can be waiting. (In many other cases the index of the state may have no numerical meaning.)

The process $\{y_t\}$ can also be defined by

$$(49) \quad y_t = \alpha' z_t,$$

where $\alpha' = (\alpha_1, \dots, \alpha_m)$. The second-order moments of $\{y_t\}$ can be found from those of $\{z_t\}$. The mean is

$$(50) \quad \mathcal{E} y_t = \alpha' p = \mu;$$

say; the variance is

$$(51) \quad \begin{aligned} \text{Var}(y_t) &= \alpha' V \alpha = \alpha' D_p \alpha - (\alpha' p)^2 \\ &= \alpha' Q T' D_p \alpha - (\alpha' p)^2 \\ &= \sum_{i=2}^m (\alpha' q_i)(\alpha' D_p t_i); \end{aligned}$$

and the covariances are given by

$$(52) \quad \begin{aligned} \text{Cov}(y_t, y_{t-s}) &= \alpha' \text{Cov}(z_t, z_{t-s}) \alpha \\ &= \alpha' (P')^s V \alpha \\ &= \alpha' Q_2 \Lambda_2^s T_2' D_p \alpha \\ &= \sum_{i=2}^m (\alpha' q_i)(\alpha' D_p t_i) \lambda_i^s. \end{aligned}$$

A similar result was derived by Reynolds (1972) in a different way.*

The nature of the process $\{y_t\}$ can be found from the nature of the $\{z_t\}$ process. Let $\mathcal{L}$ be the lag operator; that is, $\mathcal{L} z_t = z_{t-1}$. Then the $\{z_t\}$ process can be written in autoregressive form as

$$(53) \quad (I - P' \mathcal{L}) z_t = v_t.$$

Since $P = T \Lambda Q'$ and $T' Q = I$, multiplication of (53) on the left by T' gives

$$(54) \quad (I - \Lambda \mathcal{L}) w_t = u_t,$$

where

$$(55) \quad w_t = T' z_t = \begin{pmatrix} \epsilon' \\ T_2' \end{pmatrix} z_t = \begin{pmatrix} 1 \\ w_t^{(2)} \end{pmatrix}$$

* I am indebted to Jayaram Muthuswamy for calling my attention to this problem and associated literature.

and

$$(56) \quad u_t = T'v_t = \begin{pmatrix} \epsilon' \\ T_2' \end{pmatrix} v_t = \begin{pmatrix} 0 \\ u_t^{(2)} \end{pmatrix}.$$

We can re-write the last $m - 1$ components of (54) as

$$(57) \quad (I - \Lambda_2 \mathcal{L})w_t^{(2)} = u_t^{(2)}.$$

The first component of (54) is $(1 - \mathcal{L})1 = 0$. Then

$$(58) \quad w_t^{(2)} = (I - \Lambda_2 \mathcal{L})^{-1} u_t^{(2)}.$$

Multiplication of (54) on the left by Q yields

$$(59) \quad y_t = \alpha' z_t = \alpha'(p, Q_2) \begin{pmatrix} 1 \\ w_t^{(2)} \end{pmatrix} = \alpha' p + \alpha' Q_2 (I - \Lambda_2 \mathcal{L})^{-1} u_t^{(2)} \\ = \alpha' p + \alpha' Q_2 (I - \Lambda_2 \mathcal{L})^{-1} T_2' v_t.$$

Multiplication of (59) by $|I - \Lambda_2 \mathcal{L}|$ yields

$$(60) \quad \left[\prod_{j=2}^m (1 - \lambda_j \mathcal{L}) \right] (y_t - \mu) = \sum_{j=2}^m (\alpha' q_j) \prod_{h \neq 1, j} (1 - \lambda_h \mathcal{L}) t_j' v_t.$$

This equation defines an autoregressive moving average process (in the wide sense) with an autoregressive part of order $m - 1$ and a moving average part of order at most $m - 2$. Note that if $\alpha_j = 1$ for some j and $\alpha_i = 0$ for $i \neq j$, then (60) defines the marginal process of z_{it} .

The above development is in terms of real roots and vectors. If a pair of roots are complex conjugate, the corresponding pairs of vectors are complex conjugate and the analysis goes through as before.

The covariance function (52) will be the covariance function of a first-order autoregression if there is only one term in the sum on the right-hand side. Let $Q_2 = (q_2, Q_3)$, $T_2 = (t_2, T_3)$, and

$$(61) \quad \Lambda_2 = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \Lambda_3 \end{pmatrix}$$

for a real root λ_2 . Then

$$(62) \quad \text{Cov}(y_t, y_{t-s}) = (\alpha' q_2)(\alpha' D_p t_2) \lambda_2^s + \alpha' Q_3 \Lambda_3^s T_3' D_p \alpha.$$

This is the covariance function of a first-order autoregressive process if λ_2 is real and either $\alpha' Q_3 = 0$ or if $\alpha' D_p T_3 = 0$. This fact was given by Lai (1978) in his Theorem 2.3 under the assumption that all of the characteristic roots are real and distinct.

The alternative conditions may be written as

$$(63) \quad \alpha' Q = (\mu, \alpha' q_2, 0)$$

and

$$(64) \quad \alpha' D_P T = (\mu, \alpha' D_P t_2, 0).$$

Multiply (63) on the right by $Q^{-1} = T'$ and (64) by $T^{-1} D_P^{-1} = Q' D_P^{-1}$ to obtain

$$(65) \quad \alpha' = \mu \epsilon' + (\alpha' q_2) t_2'$$

and

$$(66) \quad \alpha' = \mu \epsilon' + (\alpha' D_P t_2) q_2' D_P^{-1}.$$

The conclusion is that $\{y_t\}$ is a first-order autoregressive process (in the wide sense) if the Markov chain is irreducible and aperiodic, if there are m linearly independent characteristic vectors, and if the vector α is a linear combination of ϵ and of either a right-sided characteristic vector of P corresponding to a real root (other than 1) or a left-sided vector corresponding to a real root multiplied by D_P^{-1} .

References

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York.
- Anderson, T. W. (1979a). Panels and time series analysis: Markov chains and autoregressive processes, *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld* (Robert K. Merton, James S. Coleman, and Peter H. Rossi, eds.), The Free Press, New York, 82-97.
- Anderson, T. W. (1979b). Some relations between Markov chains and vector autoregressive processes, *International Statistical Institute: Contributed Papers, 42nd Session, December 4-14, 1979*, Manila, 25-28.
- Anderson, T. W. (1980). Finite-state Markov chains and vector autoregressive processes, *Proceedings of the Conference on Recent Developments in Statistical Methods and Applications*, Directorate-General of Budget, Accounting and Statistics, Executive Yuan, Taipei, Taiwan, Republic of China, 1-12.
- Anderson, T. W., and Leo A. Goodman (1957), Statistical inference about Markov chains. *Annals of Mathematical Statistics*, **28**, 89-110.
- Diaconis, Persi, and Bradley Efron (1985). Testing for independence in a two-way table: New interpretations of the chi-square statistic. *Annals of Statistics*, **13**, 845-913.
- Grenander, Ulf (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics*, **25**, 252-272.
- Grenander, Ulf, and Murray Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. John Wiley and Sons, Inc., New York.
- Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons, Inc., New York.
- Henry, Neil Wylie (1970). *Problems in the Statistical Analysis of Markov Chains*. Doctoral Dissertation, Columbia University Libraries.
- Lai, C. D. (1978). First order autoregressive Markov processes. *Stochastic Processes and their Applications*, **7**, 65-72.
- Reynolds, John F. (1972). Some theorems on the transient covariance of Markov chains. *Journal of Applied Probability*, **9**, 214-218.
- Rosenblatt, Murray (1956). Some regression problems in time series analysis. *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, **1**, 165-186.

Technical Reports
U.S. Army Research Office
Contracts DAAG29-82-K-0156, DAAG29-85-K-0239, and DAAL03-89-K-0033

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T. W. Anderson and Kai-Tai Fang, September 1982.
2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.
3. "Some Further Applications of Finite Difference Operators," Kai-Tai Fang, September 1982.
4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.
5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai-Tai Fang and Shu-Dong He, October 1982.
6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.
7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.
8. "Orthogonal Expansion of Quantile Function and Components of the Shapiro-Francia Statistic," Akimichi Takemura, January 1983.
9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.
10. "Relationships Among Classes of Spherical Matrix Distributions," Kai-Tai Fang and Han-Feng Chen, April 1984.
11. "A Generalization of Autocorrelation and Partial Autocorrelation Functions Useful for Identification of ARMA(p,q) Processes," Akimichi Takemura, May 1984.
12. "Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models," T. W. Anderson and N. D. Singpurwalla, October 1984.
13. "Why Do Noninvertible Estimated Moving Averages Occur?" T. W. Anderson and Akimichi Takemura, November 1984.
14. "Invariant Tests and Likelihood Ratio Tests for Multivariate Elliptically Contoured Distributions," Huang Hsu, May 1985.
15. "Statistical Inferences in Cross-lagged Panel Studies," Lawrence S. Mayer, November 1985.
16. "Notes on the Extended Class of Stein Estimators," Suk-ki Hahn, July 1986.
17. "The Stationary Autoregressive Model," T. W. Anderson, July 1986.
18. "Bayesian Analyses of Nonhomogeneous Autoregressive Processes," T. W. Anderson, Nozer D. Singpurwalla, and Refik Soyer, September 1986.
19. "Estimation of a Multivariate Continuous Variable Panel Model," Lawrence S. Mayer and Kun Mao Chen, June 1987.
20. "Consistency of Invariant Tests for the Multivariate Analysis of Variance," T. W. Anderson and Michael D. Perlman, October 1987.
21. "Likelihood Inference for Linear Regression Models," T. J. DiCiccio, November 1987.
22. "Second-order Moments of a Stationary Markov Chain and Some Applications," T. W. Anderson, February 1989.