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

T. W. Anderson
Stanford University

TECHNICAL REPORT NO. 1
February 1989

U. S. Army Research Office
Contracts DAAG20-85-K-0230 and DAAL03-89-K-0033
Theodore W. Anderson, Project Director

Department of Statistics
Stanford University
Stanford, California

Approved for Public Release; Distribution Unlimited.
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.
<table>
<thead>
<tr>
<th>REPORT NUMBER</th>
<th>GOVT ACCESSION NO.</th>
<th>RECIPENT'S CATALOG NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE (and Subtitle)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SECOND-ORDER MOMENTS OF A STATIONARY MARKOV CHAIN AND SOME APPLICATIONS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AUTHOR(s)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T.W. Anderson</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERFORMING ORGANIZATION NAME AND ADDRESS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics Department - Sequoia Hall Stanford University Stanford, CA 94305-4065</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONTROLLING OFFICE NAME AND ADDRESS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REPORT DATE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>February 1989</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMBER OF PAGES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13 pp.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SECURITY CLASS. (of this Report)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DISTRIBUTION STATEMENT (of this Report)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release; distribution unlimited.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUPPLEMENTARY NOTES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEY WORDS (Continue on reverse side if necessary and identify by block number)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary Markov chain, second-order moments, estimation of stationary probabilities, simulation, first-order autoregression</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ABSTRACT (Continue on reverse side if necessary and identify by block number)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>See reverse side for Abstract.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.
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, ..., \( m \) is defined by the transition probabilities

\[
Pr\{x_t = j | x_{t-1} = i, x_{t-2} = k, \ldots \} = p_{ij}, \quad i, j, k, \ldots = 1, \ldots, m, \quad t = \ldots, -1, 0, 1, \ldots,
\]

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

\[
\sum_{i=1}^{m} p_i p_{ij} = p_j, \quad j = 1, \ldots, m.
\]

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

\[
p'P = p', \quad PE = e, \quad p'e = 1,
\]

where \( e = (1, 1, \ldots, 1)' \). Let \( e_i \) be the \( m \)-component vector with 1 in the \( i \)-th position and 0's elsewhere, and let \( \{x_t\} \) be a sequence of \( m \)-component random vectors. Then the Markov chain can be written

\[
Pr\{x_t = e_j | x_{t-1} = e_i, x_{t-2} = e_k, \ldots \} = p_{ij}, \quad i, j, k, \ldots = 1, \ldots, m, \quad t = \ldots, -1, 0, 1, \ldots
\]
Note that $z_t$ has 1 in one position and 0 in the others; hence $e'z_t = 1$.


It follows from the model that

\[ \mathcal{E}(z_{jt}|z_{t-1} = t, z_{t-2} = \epsilon, \ldots) = p_{ij}, \]

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

\[ \mathcal{E}(z_t|z_{t-1}, z_{t-2}, \ldots) = P'z_{t-1}. \]

Let

\[ v_t = z_t - P'z_{t-1} \]

be the $t$-th disturbance. Then

\[ \mathcal{E}v_t = \mathcal{E}\{\mathcal{E}[(z_t - P'z_{t-1})|z_{t-1}, z_{t-2}, \ldots]\} = 0. \]

In (8) the outer expectation is with respect to $z_{t-1}, z_{t-2}, \ldots$. Similarly

\[ \mathcal{E}v_t z'_{t-s} = \mathcal{E}[\mathcal{E}(v_t|z_{t-1}, z_{t-2}, \ldots)z_{t-s}] = 0, \quad s = 1, 2, \ldots. \]

Since $v_{t-s} = z_{t-s} - P'z_{t-s-1},$

\[ \mathcal{E}v_t v'_{t-s} = 0, \quad s = 1, 2, \ldots. \]

Thus $\{v_t\}$ is a sequence of uncorrelated random vectors.

We can iterate $z_t = P'z_{t-1} + v_t$ to obtain

\[ z_t = v_t + P'v_{t-1} + \cdots + (P')^{s-1}v_{t-s+1} + (P')^s z_{t-s}. \]

Then

\[ \mathcal{E}(z_t|z_{t-s}, z_{t-s-1}, \ldots) = (P')^s z_{t-s}. \]

This conditional expected value could alternatively be obtained from the fact that the transition probabilities from $x_{t-s}$ to $x_t$ are the elements of $P^s$. 2
Since \( \{x_t\} \) is stationary, \( \{z_t\} \) is stationary and \( z_t \) has a marginal multinomial distribution with probabilities \( p_1, \ldots, p_m \). Hence, the expectation of \( z_t \) is

\[
\mathbb{E} z_t = p,
\]

and the covariance matrix is

\[
\text{Var}(z_t) = Dp - pp' = V,
\]
say, where \( Dp \) is a diagonal matrix with \( i \)-th diagonal element \( p_i \). From (11) we also find

\[
\mathbb{E} z_t z'_{t-s} = (P')^s \mathbb{E} z_t z'_t = (P')^s Dp, \quad s = 0, 1, \ldots.
\]

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

\[
\text{Cov}(z_t, z'_{t-s}) = (P')^s V, \quad s = 0, 1, \ldots.
\]

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

The conditional covariance matrix of \( z_t \) and \( v_t \) is

\[
\text{Var}(z_t | z_{t-1} = \epsilon_i, z_{t-2}, \ldots) = \text{Var}(v_t | z_{t-1} = \epsilon_i, z_{t-2}, \ldots)
\]

\[
= Dp_i - p_i p'_i
\]

\[
= V_i,
\]
say, where \( Dp_i \) is a diagonal matrix with \( p_{ij} \) as the \( j \)-th diagonal element and

\[
P = \begin{bmatrix}
p'_1 \\
p'_2 \\
\vdots \\
p'_m
\end{bmatrix}.
\]

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

\[
\mathbb{E} v_t v'_t = \sum_{i=1}^m p_i V_i
\]

\[
= \sum_{i=1}^m p_i (Dp_i - p_i p'_i)
\]

\[
= Dp - \sum_{i=1}^p p_i p_i p'_i.
\]

by (2). Note that

\[
P' V P = P' Dp P - P' pp' P
\]

\[
= \sum_{i=1}^p p_i p_i p'_i - pp'.
\]
Thus

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

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 \( \lambda_1 = 1, \lambda_2, \ldots, \lambda_m \) the characteristic roots of \( P, \) \( t_1 = e, t_2, \ldots, t_m \) be the corresponding right-sided (column) characteristic vectors, and \( w'_1 = p', w'_2, \ldots, w'_m \) 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

\[
A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & A_2 \end{pmatrix},
\]

\[
Q = (p, q_2, \ldots, q_m) = (p, Q_2),
\]

\[
T = (\epsilon, t_2, \ldots, t_m) = (\epsilon, T_2).
\]

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

\[
P = T\Lambda Q' = T\Lambda T^{-1},
\]

\[
P' = T\Lambda'Q' = T\Lambda'T^{-1}.
\]

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

\[
(P')'V = QA'T'V = (p, Q_2) \begin{pmatrix} 1 & 0 \\ 0 & A_2' \end{pmatrix} (\begin{pmatrix} \epsilon' \end{pmatrix} (Dp - pp')) = Q_2 \Lambda'_2 T'_{s} Dp = (P' - pe')'Dp, \quad s = 1, 2, \ldots.
\]

We shall assume that the chain is irreducible and aperiodic. Then \(|\lambda_i| < 1, i = 2, \ldots, m. \) As \( s \) increases, the covariance function decreases as a linear combination of \( \lambda_2', \ldots, \lambda_m'. \)
We can write $z_t$ in the moving average representation

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

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

$$ (P')^s v_{t-s} = (P')^s (z_{t-s} - P' z_{t-s-1})$$

$$= (P')^s z_{t-s} - (P')^{s+1} z_{t-(s+1)}. \quad (29)$$

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

$$ (I - P' e^{i\lambda})^{-1} V (I - Pe^{-i\lambda})^{-1}$$

$$= [Q(I - \Lambda e^{i\lambda}) T']^{-1} V [T(I - \Lambda e^{-i\lambda}) Q']^{-1}$$

$$= (T')^{-1} (I - \Lambda e^{i\lambda})^{-1} Q^{-1} V (Q')^{-1} (I - \Lambda e^{-i\lambda})^{-1} T^{-1}$$

$$= Q(I - \Lambda e^{i\lambda})^{-1} T' V T (I - \Lambda e^{-i\lambda})^{-1} Q'$$

$$= Q_2 (I - \Lambda_2 e^{i\lambda})^{-1} T_2^2 D p T_2 (I - \Lambda_2 e^{-i\lambda})^{-1} Q_2'. \quad (30)$$

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


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

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

Then $ES = Np$. The covariance matrix of $S$ is

$$\mathcal{E}(S - Np)(S - Np)'$$

$$= \sum_{t=1}^{N} \mathcal{E}(z_t - p)(z_t - p)'$$

$$= \sum_{t=1}^{N} \sum_{s=1}^{t-1} \mathcal{E}(z_t - p)(z_s - p)' + \sum_{s=1}^{N} \sum_{t=1}^{s-1} \mathcal{E}(z_t - p)(z_s - p)' + \sum_{t=1}^{N} \mathcal{E}(z_t - p)(z_t - p)'$$

$$= \sum_{t=1}^{N} \sum_{r=1}^{t-1} [\mathcal{E}(z_t - p)(z_{t-r} - p)' + \mathcal{E}(z_{t-r} - p)(z_t - p)'] + NV$$

$$= \sum_{t=1}^{N} \sum_{r=1}^{t-1} [(P')^r V + V P^r'] + NV$$

5
\[
\begin{align*}
N^{-1} &= \sum_{i=1}^{N} \sum_{t=1}^{N-1} \left[ W_2 A_2' T_2'D_p + D_p T_2 A_2' W_2' \right] \\
&= \sum_{t=1}^{N} \left[ W_2(I - A_2)^{-1} A_2(I - A_2^{-1}) T_2'D_p + D_p T_2(I - A_2^{-1}) A_2(I - A_2)^{-1} W_2' \right] + NV \\
&= N \left[ W_2(I - A_2)^{-1} A_2 T_2'D_p + D_p T_2 A_2(I - A_2)^{-1} W_2' \right] \\
&\quad - W_2(I - A_2)^{-2} A_2(I - A_2^N) T_2'D_p - D_p T_2(I - A_2^N) A_2(I - A_2)^{-2} W_2' \\
&\quad + NV.
\end{align*}
\]

Then the covariance matrix of

\[(33)\]
\[\sqrt{N} \tilde{z} = \frac{1}{\sqrt{N}} S\]

is

\[(34)\]
\[W_2(I - A_2)^{-1} T_2'D_p + D_p T_2(I - A_2)^{-1} W_2' - D_p + pp'\]
\[\quad - \frac{1}{N} \left[ W_2(I - A_2)^{-2} A_2(I - A_2^N) T_2'D_p + D_p T_2(I - A_2^N) A_2(I - A_2)^{-2} W_2' \right].\]

**Theorem.**

\[(35)\]
\[\sqrt{N}(\tilde{z} - \mu) \overset{d}{\to} N[0, W_2(I - A_2)^{-1} T_2'D_p + D_p T_2(I - A_2)^{-1} W_2' - D_p + pp'].\]

The sum \(S\) is the vector of frequencies of the states \(1, \ldots, m\) being observed, and \(\tilde{z}\) is the vector of relative frequencies. Since \(E \tilde{z} = \mu\), the vector \(\tilde{z}\) is an estimator of \(\mu\). 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 \(\mu\) from the data \(z_1, \ldots, z_N\) 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

\[(36)\]
\[\hat{P} = \left( \sum_{t=2}^{N} z_{t-1}' z_{t-1}' \right)^{-1} \sum_{t=2}^{N} z_{t-1}' z_t'.\]
(See Anderson and Goodman (1957).) Note that

\begin{equation}
\sum_{t=2}^{N} z_{t-1} z'_{t-1} = \begin{pmatrix}
\sum_{t=1}^{N-1} z_{1t} & 0 & \ldots & 0 \\
0 & \sum_{t=1}^{N-1} z_{2t} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sum_{t=1}^{N-1} z_{mt} \\
\end{pmatrix}
\end{equation}

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

\begin{equation}
\hat{\mathbf{p}}' \hat{\mathbf{P}} = \mathbf{p}
\end{equation}

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

We observe that

\begin{equation}
(z - \frac{1}{N} \mathbf{z}_N)' \hat{\mathbf{P}} = \frac{1}{N} \mathbf{e}' \sum_{t=2}^{N} z_{t-1} z'_t
= \frac{1}{N} \sum_{t=2}^{N} z'_t
= (z - \frac{1}{N} \mathbf{z}_1)'.
\end{equation}

From (38) and (39) we obtain

\begin{equation}
N^{a-1}(z'_N \hat{\mathbf{P}} - z'_1) = N^a (z - \hat{\mathbf{p}})' (\hat{\mathbf{P}} - I)
= N^a (z - \hat{\mathbf{p}})' \hat{\mathbf{T}}_2 (\hat{\mathbf{A}}_2 - I) \hat{\mathbf{Q}}'_2,
\end{equation}

where

\begin{equation}
\hat{\mathbf{P}} = (\mathbf{e}, \hat{\mathbf{T}}_2) \begin{pmatrix} 1 & 0 \\ 0 & \hat{\mathbf{A}}_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{p}}' \\ \mathbf{W}'_2 \end{pmatrix}.
\end{equation}

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

\begin{equation}
N^a (z - \hat{\mathbf{p}}) \overset{p}{\to} \mathbf{0}.
\end{equation}

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

In many situations a sample \( x_1, \ldots, x_N \) (or equivalently, \( \mathbf{z}_1, \ldots, \mathbf{z}_N \)) 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 \( \hat{z} \) calculated from the remaining observations as an estimator of \( \mathbf{p} \) 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 \( \mathbf{P} \) and then \( \mathbf{p} \) does not require stationarity; one can start from any state. The estimator of \( \mathbf{P} \) is a maximum likelihood estimator, and hence the estimator of \( \mathbf{p} \) (and \( \Lambda_2, W_2, \) 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

\[
D_\mathbf{p}T_2(I - \Lambda_2)^{-1}Q_2' = D_\mathbf{p} \sum_{j=2}^{m} t_j(1 - \lambda_j)^{-1} q_j'
= D_\mathbf{p} \sum_{j=2}^{m} \sum_{s=0}^{\infty} \lambda_j^s t_j q_j'.
\]

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

\[
D_\mathbf{p} \sum_{s=0}^{\infty} (\mathbf{P}' - \epsilon \mathbf{p}')^s + [D_\mathbf{p} \sum_{s=0}^{\infty} (\mathbf{P}' - \epsilon \mathbf{p}')^s]' - D_\mathbf{p} + \mathbf{p}'p'.
\]

Since

\[
(P - \epsilon \mathbf{p}')^s = P^s - \epsilon \mathbf{p}', \quad s = 0, 1, \ldots,
\]

the covariance matrix (44) is

\[
D_\mathbf{p} \sum_{s=0}^{\infty} P^s + \mathbf{D}_\mathbf{p}(\sum_{s=0}^{\infty} P^s)' - D_\mathbf{p} - \mathbf{p}'p'.
\]

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

A possible computational method is to power \( \mathbf{P} \) until the rows of \( \mathbf{P}^s \) are similar enough, that is, until \( \mathbf{P}^s \) is similar to \( \epsilon \mathbf{p}' \). Then (46) follows. Note that at each step only \( \mathbf{P}^s \) and \( \sum_{t=0}^{s} \mathbf{P}^t \) 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 \(e\).

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

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

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 \(z_1, \ldots, z_N\), the probability (47) is estimated by

\[
\frac{1}{N} \sum_{i=1}^{N} I[T(z_i) \leq \tau].
\]

The estimator of the cdf of \(T(Z)\) is the empirical cdf of \(T(Z)\) although \(z_1, \ldots, z_N\) are not independent. However, the sample variance of \(I[T(z_i) \leq \tau]\) divided by \(N\) is not an estimator of the variance of (48). If a subset of \(z_1, \ldots, z_N\) is taken, say, \(z_{t_1}, \ldots, z_{t_n}\), so that \(\min|t_i - t_j|\) is great enough that \(z_{t_1}, \ldots, z_{t_n}\) can be considered independent, the sample variance of \(I[T(z_i) \leq \tau]\) 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 \(\alpha_i\), \(i = 1, \ldots, m\). 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, . . . , 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

\[
y_t = \alpha' z_t,
\]

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

\[
\mathbb{E}y_t = \alpha' p = \mu;
\]
say; the variance is

\[
\text{Var}(y_t) = \alpha' V \alpha = \alpha' D p \alpha - (\alpha' p)^2
\]

\[
= \alpha' Q T' D p \alpha - (\alpha' p)^2
\]

\[
= \sum_{i=2}^{m} (\alpha' q_i)(\alpha' D p t_i);
\]

and the covariances are given by

\[
\text{Cov}(y_t, y_{t-s}) = \alpha' \text{Cov}(z_t, z_{t-s}) \alpha
\]

\[
= \alpha' (P')^s V \alpha
\]

\[
= \alpha' Q_2 A_2^s T_2' D p \alpha
\]

\[
= \sum_{i=2}^{m} (\alpha' q_i)(\alpha' D p t_i) \lambda_i^s.
\]

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

\[
(I - P' \mathcal{L}) z_t = v_t.
\]

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

\[
(I - \Lambda \mathcal{L}) w_t = u_t,
\]

where

\[
w_t = T' z_t = \begin{pmatrix} \varepsilon' \\ T' \end{pmatrix} z_t = \begin{pmatrix} 1 \\ w_i^{(2)} \end{pmatrix}
\]

---

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

\[ u_t = T'v_t = \begin{pmatrix} e' \\ T'_2 \end{pmatrix} v_t = \begin{pmatrix} 0 \\ u_t^{(2)} \end{pmatrix}. \]

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

\[ (I - A_2 \mathcal{L})w_t^{(2)} = u_t^{(2)}. \]

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

\[ w_t^{(2)} = (I - A_2 \mathcal{L})^{-1} u_t^{(2)}. \]

Multiplication of (54) on the left by \( Q \) yields

\[ y_t = \alpha' z_t = \alpha'(p, Q_2) \begin{pmatrix} 1 \\ u_t^{(2)} \end{pmatrix} = \alpha' p + \alpha' Q_2 (I - A_2 \mathcal{L})^{-1} u_t^{(2)} \]

\[ = \alpha' p + \alpha' Q_2 (I - A_2 \mathcal{L})^{-1} T'_2 v_t. \]

Multiplication of (59) by \(|I - A_2 \mathcal{L}|\) yields

\[ \sum_{j=2}^{m} (1 - \lambda_j \mathcal{L}) \begin{pmatrix} \alpha' q_j \\ 0 \end{pmatrix} = \sum_{j=2}^{m} (1 - \lambda_j \mathcal{L})^t v_t. \]

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 \( i \neq j,\) than (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

\[ \Lambda_2 = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \Lambda_3 \end{pmatrix} \]

for a real root \( \lambda_2.\) Then

\[ \text{Cov}(y_t, y_{t-s}) = (\alpha' q_2)(\alpha' Dp t_2) \lambda_2^s + \alpha' Q_3 \Lambda_3^s T_3^t Dp \alpha. \]

This is the covariance function of a first-order autoregressive process if \( \lambda_2 \) is real and either \( \alpha' Q_3 = 0 \) or if \( \alpha' Dp T_3 = 0.\) 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) \[ \alpha'Q = (\mu, \alpha'q_2, 0) \]

and

(64) \[ \alpha'D_pT = (\mu, \alpha'D_p t_2, 0). \]

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

(65) \[ \alpha' = \mu \varepsilon' + (\alpha'q_2)t'_2 \]

and

(66) \[ \alpha' = \mu \varepsilon' + (\alpha'D_p t_2)q'_2 D_p^{-1}. \]

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 \( \alpha \) is a linear combination of \( \varepsilon \) 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


