PANELS AND TIME SERIES ANALYSIS: MARKOV CHAINS AND AUTOREGRESSIVE PROCESSES

BY

T. W. ANDERSON

TECHNICAL REPORT NO. 24
JULY 1976

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

PANELS AND TIME SERIES ANALYSIS:
MARKOV CHAINS AND AUTOREGRESSIVE PROCESSES

By

T. W. Anderson

Technical Report No. 24
July 1976

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

A

Panels and Time Series Analysis:
Markov Chains and Autoregressive Processes

by

T. W. Anderson

Stanford University

1. Introduction

In a panel survey a sample of individuals is interviewed at several points in time; the resulting data are a sequence of responses. The techniques and objectives were described by Lazarsfeld, Berelson, and Gaudet in The People's Choice (1944). That study was based on repeated interviews of many voters in Erie County, Ohio, in 1940. Respondents were asked in May, June, July, August, September, and October for which party (or candidate) the respondent intended to vote. For some purposes the responses to this question were coded as Republican, Democrat, and "Don't Know"; that is, each person at each time was put into one of three categories. The records of the 445 persons who responded to all six interviews can be considered as 445 observations (or "realizations") from a probability distribution of such records (that is, a segment of a stochastic process).

A discrete-state, discrete-time Markov chain can serve as a model for panel data. The development of this model, illustrated by the survey of vote intention, was reported by Anderson (1954). The statistical methodology, developed further in collaboration with Leo Goodman, appeared later in Anderson and Goodman (1957).

In some panel surveys the responses may be quantitative, such as answers to the question how many hours did you spend last month reading the newspapers. In economic surveys the questions are more likely to produce numerical answers: how many hours did you work last week and how much money did you spend on groceries last month. Analysis of such data are sometimes called cross-section studies by econometricians.

A possible model for time series consisting of measurements on one or more continuous variables is a univariate or multivariate autoregressive process. The statistical methods for autoregressive processes have been developed mainly for one observed time series, that is, the record of one unit of observation. However, a characteristic feature of panel data is that there are available the records of several units of observation. Of course, such repeated measurements occur in other situations. A psychologist may obtain a test score on several individuals at several points in time; a physician may read blood pressures of several patients daily.

The purpose of the present paper is to review some statistical methods for Markov chains and present similar methods for corresponding problems in autoregressive processes in the case of repeated measurements. The statistical problems treated are those presented by Anderson and Goodman (1957) and suggest that other procedures for Markov chains have their analogs for autoregressive processes. The development of the methods for autoregressive processes and proofs of the mathematical statements will be given in a later paper.

Statistical methods for a single observed series from a Markov chain* have been discussed extensively by Billingsley (1961). The autoregressive process with one observation on the series is treated in depth in Anderson (1971).

Each section of this paper is divided into three subsections, the first dealing with the Markov chain model, the second treating the autoregressive model, and the third displaying correspondences between the two models. Section 2 defines the models and reviews some of their properties. Section 3 discusses summary data and estimates of parameters. Section 4 develops tests of hypotheses.

2. Probability Models for Time Series.

2.1. A Markov Chain Model for Discrete Data. A Markov chain can serve as a model for the probabilities of a sequence of statistical variables that take on a finite set of values. Let the values or states or categories be labelled $1, \ldots, m$, and let x_t be the statistical variable at time t, $t = 1, \ldots, T$. For instance, $x_1 = 2$ might denote an individual holding opinion 2 (Democrat) at the first interview (May). Then a Markov chain model specifies the probability of state j at time t given state i at time $t - 1$

$$\Pr\{x_t = j | x_{t-1} = i\} = p_{ij}(t) , \qquad i,j = 1, \ldots, m . \tag{2.1}$$

* Bartlett (1951) developed some methods in the context of a single observation.

These transition probabilities satisfy the conditions $p_{ij}(t) \geq 0$ and

$$\sum_{j=1}^{m} p_{ij}(t) = 1 , \qquad i = 1, \ldots, m . \tag{2.2}$$

While the period of observation is usually finite $(t=1,\ldots,T)$, the stochastic process may be defined for all integers, $t = \ldots, -1, 0, 1, \ldots$. In any case, there is a marginal distribution of the statistical variable at each time point; the probability of state i at time t is denoted $p_i(t)$ $[p_i(t) \geq 0$, $\Sigma_{i=1}^{m} p_i(t) = 1]$. The joint probability that $x_{t-1} = i$ and $x_t = j$ is $p_i(t-1)p_{ij}(t)$; thus, the marginal probability that $x_t = j$ follows from the marginal distribution at $t - 1$ by

$$\sum_{i=1}^{m} p_i(t-1)\, p_{ij}(t) = p_j(t) , \qquad j = 1, \ldots, m . \tag{2.3}$$

In order to describe the probabilities of the observed random variables $(t = 1, \ldots, T)$, it is only necessary to prescribe the vector $\underset{\sim}{p}(1) = [p_1(1), \ldots, p_m(1)]'$ and the matrices $\underset{\sim}{P}(t) = [p_{ij}(t)]$, $t=2,\ldots,T$. The distinguishing feature of a Markov chain is that the conditional probability of x_t given the entire past $x_{t-1}, x_{t-2}, \ldots$ depends only on the immediately preceding variable x_{t-1} .

In many situations the transition probabilities are homogeneous; that is, $p_{ij}(t) = p_{ij}$ or $\underset{\sim}{P}(t) = \underset{\sim}{P}$, $t = \ldots, -1, 0, 1, \ldots$. Then (under general conditions) the marginal distributions are homogeneous, that is, $\underset{\sim}{p}(t) = \underset{\sim}{p}$, and the process is stationary. In this case (2.3) is

$$\sum_{i=1}^{m} p_i\, p_{ij} = p_j , \qquad j = 1, \ldots, m . \tag{2.4}$$

The matrix form of (2.4)

$$\underset{\sim}{p}'\underset{\sim}{P} = \underset{\sim}{p}' . \tag{2.5}$$

shows that $\underset{\sim}{p}$ is a left-sided characteristic vector of $\underset{\sim}{P}$ corresponding to a characteristic root of 1, that is, a root of

$$|\underset{\sim}{P} - \lambda \underset{\sim}{I}| = 0 \tag{2.6}$$

of 1. The equations (2.2) can be written $\underset{\sim}{P}\underset{\sim}{\varepsilon} = \underset{\sim}{\varepsilon}$, where $\underset{\sim}{\varepsilon} = (1, \ldots, 1)'$; thus $\underset{\sim}{\varepsilon}$ is a right-sided characteristic vector of $\underset{\sim}{P}$ corresponding to a characteristic root of 1. There are m roots $\lambda_1 = 1, \lambda_2, \ldots, \lambda_m$ of (2.6); each root satisfies $|\lambda_i| \leq 1$. The Markov chain is called irreducible if the root of 1 is of multiplicity 1; then there is a positive probability of going from one state to another in some interval of time. In that case $\underset{\sim}{p}$ is determined uniquely by (2.5) and the normalization $\underset{\sim}{p}'\underset{\sim}{\varepsilon} = 1$.

In a more general model the probability of a state at time t may depend on the states at several time points earlier. For example, a second-order stationary chain is defined by the transition probabilities

$$\Pr\{x_t = k \mid x_{t-2} = i, x_{t-1} = j\} = p_{ijk}, \quad i,j,k = 1, \ldots, m, \tag{2.7}$$

where $p_{ijk} \geq 0$ and $\sum_{k=1}^{m} p_{ijk} = 1$, $i,j = 1, \ldots, m$. Higher-order chains and nonstationary chains are defined similarly. For some purposes it is convenient to redefine states so as to construct a Markov (first-order) chain that is mathematically equivalent to this second-order chain. For example, if $m = 2$, let

$$\begin{aligned} \tilde{x}_t &= 1 \quad \text{if} \quad x_{t-1} = 1, x_t = 1, \\ &= 2 \quad \text{if} \quad x_{t-1} = 1, x_t = 2, \\ &= 3 \quad \text{if} \quad x_{t-1} = 2, x_t = 1, \\ &= 4 \quad \text{if} \quad x_{t-1} = 2, x_t = 2. \end{aligned} \tag{2.8}$$

Then the matrix of transition probabilities for $\tilde{x}(t)$ is

$$\tilde{P} = \begin{bmatrix} p_{111} & p_{112} & 0 & 0 \\ 0 & 0 & p_{121} & p_{122} \\ p_{211} & p_{212} & 0 & 0 \\ 0 & 0 & p_{221} & p_{222} \end{bmatrix} . \tag{2.9}$$

The Markov chain model also includes multivariate cases. As an illustration, consider two dichotomous variables $y_t = 1$ or 2 and $z_t = 1$ or 2. Define x_t by

$$\begin{aligned} x_t &= 1 \quad \text{if } y_t = 1,\ z_t = 1 , \\ &= 2 \quad \text{if } y_t = 1,\ z_t = 2 , \\ &= 3 \quad \text{if } y_t = 2,\ z_t = 1 , \\ &= 4 \quad \text{if } y_t = 2,\ z_t = 2 . \end{aligned} \tag{2.10}$$

The model may be further developed to include the effects of other variables by stratification. If there is a discrete conditioning variable, the transition probabilities can depend on it; that is, the (homogeneous) transition probabilities in the h-th stratum may be the set $[p_{ij}^{(h)}]$.

From a statistician's viewpoint the Markov chain model is constructed from a family of elementary multinomial distributions; in the case of a dichotomous item (that is, two states) these are Bernoulli distributions. Each conditional distribution is such a discrete distribution. The appropriate statistical procedures for a Markov chain are similarly developments of methods for multinomial distributions.

2.2. An Autoregressive Process. An autoregressive process can serve as a model for a sequence of continuous random variables. In the simplest case the (univariate) stochastic process $\{y_t\}$ has the property that the conditional distribution of y_t given $y_{t-1}, y_{t-2}, \ldots,$ has expected value $\beta(t)y_{t-1}$ and $y_t - \beta(t)y_{t-1}$ is statistically independent of $y_{t-1}, y_{t-2}, \ldots$. This is often written

$$y_t = \beta(t)y_{t-1} + u_t , \tag{2.11}$$

where $\{u_t\}$ is a sequence of unobservable independent random variables with means 0. If the autoregression coefficients are homogenous, that is $\beta(t) = \beta$ for some β , $|\beta| < 1$, and the u_t are identically distributed, the process is stationary and we can write

$$y_t = \sum_{s=0}^{\infty} \beta^s u_{t-s} . \tag{2.12}$$

If the u_t's have variance σ^2 , the y_t's have variance $\sigma^2/(1-\beta^2)$ and covariances

$$\mathcal{E} y_t y_s = \frac{\beta^{|t-s|}\sigma^2}{1-\beta^2} . \tag{2.13}$$

If the u_t's are normally distributed, any set of y_t's are normal and the covariance structure (2.13) determines the process.

If $\underset{\sim}{y}_t$ is a p-component vector, a Markov (first-order) vector process is defined by

$$\underset{\sim}{y}_t = \underset{\sim}{B}(t)\underset{\sim}{y}_{t-1} + \underset{\sim}{u}_t , \tag{2.14}$$

where $\underset{\sim}{B}(t)$ is a $p \times p$ matrix and $\{\underset{\sim}{u}_t\}$ is a sequence of independent (unobservable) random vectors with expected values $\mathcal{E}\underset{\sim}{u}_t = \underset{\sim}{0}$, covariance matrices $\mathcal{E}\underset{\sim}{u}_t\underset{\sim}{u}_t' = \underset{\sim}{\Sigma}_t$, and $\underset{\sim}{u}_t$ independent of $\underset{\sim}{y}_{t-1}, \underset{\sim}{y}_{t-2}, \ldots$. Let the covariance matrix of $\underset{\sim}{y}_t$ be $\mathcal{E}\underset{\sim}{y}_t\underset{\sim}{y}_t' = \underset{\sim}{F}_t$. Then from (2.14) and the independence of $\underset{\sim}{y}_{t-1}$ and $\underset{\sim}{u}_t$ we deduce

$$(2.15) \qquad \underset{\sim}{F}_t = \underset{\sim}{B}(t)\, \underset{\sim}{F}_{t-1}\, \underset{\sim}{B}(t)' + \underset{\sim}{\Sigma}_t \, .$$

If the observations are made for $t = 1, \ldots, T$, the model may be specified by the marginal distribution of $\underset{\sim}{y}_1$ and the distributions of $\underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_T$. In particular, if $\underset{\sim}{y}_1$ and the $\underset{\sim}{u}_t$'s are normal, the model for the observation period is specified by
$\underset{\sim}{F}_1, \underset{\sim}{B}(2), \ldots, \underset{\sim}{B}(T), \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_T$.

When the autoregression matrices are homogeneous, that is, $\underset{\sim}{B}(t) = \underset{\sim}{B}$, and the $\underset{\sim}{u}_t$'s are identically distributed with mean $\underset{\sim}{0}$ and covariance matrice $\underset{\sim}{\Sigma}$, the process is stationary if the characteristic roots of $\underset{\sim}{B}$ are less than 1 in absolute value and the process is defined for $t = \ldots, -1, 0, 1, \ldots$ or $\underset{\sim}{y}_1$ is assigned the stationary marginal distribution. In this case the covariance matrix of $\underset{\sim}{y}_t$ is

$$(2.16) \qquad \underset{\sim}{F} = \sum_{s=0}^{\infty} \underset{\sim}{B}^s \underset{\sim}{\Sigma}\, \underset{\sim}{B}'^s \, ,$$

and the covariance of $\underset{\sim}{y}_t$ and $\underset{\sim}{y}_s$ is

$$(2.17) \qquad \mathcal{E}\underset{\sim}{y}_t\underset{\sim}{y}_s' = \underset{\sim}{B}^{t-s}\underset{\sim}{F} \, , \qquad s \le t \, .$$

(Note that $\underset{\sim}{F} = \underset{\sim}{F}_t = \underset{\sim}{F}_{t-1}$ satisfies (2.15) for $\underset{\sim}{B}(t) = \underset{\sim}{B}$ and $\underset{\sim}{\Sigma}_t = \underset{\sim}{\Sigma}$.)

A second-order stationary autoregressive vector process may be defined by

$$y_t = B_1 y_{t-1} + B_2 y_{t-2} + u_t \ . \tag{2.18}$$

This model can be written as a first-order process by writing

$$\tilde{y}_t = \tilde{B} \tilde{y}_{t-1} + \tilde{u}_t \ , \tag{2.19}$$

where

$$\tilde{y}_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \quad \tilde{u}_t = \begin{pmatrix} u_t \\ 0 \end{pmatrix} \ , \tag{2.20}$$

$$\tilde{B} = \begin{pmatrix} B_1 & B_2 \\ I & 0 \end{pmatrix} \ . \tag{2.21}$$

The characteristic roots of $\tilde{B}$ are the roots of

$$|-\lambda^2 I + \lambda B_1 + B_2| = 0 \ . \tag{2.22}$$

For a stationary process these roots are to be less than 1 in absolute value.

The autoregressive processes appropriate to several subpopulations (strata) may be different. In the homogeneous first-order case the matrices of coefficients and the covariance matrices may be different. If other influencing variables are continuous, they may be taken account of by adding them to the regression to yield the model.

$$y_t = B y_{t-1} + \gamma' z_t + u_t \ , \tag{2.23}$$

where z_t is a vector of such variables and γ is a vector of parameters. In particular, when $z_t \equiv 1$ and γ is a scalar, the process $\{y_t\}$ may have a mean different from 0.

The autoregressive process is constructed from multivariate regressions. In (2.14), for example, the vector $\underset{\sim}{y}_{t-1}$ constitutes the "independent variables" and the vector $\underset{\sim}{y}_t$ constitutes the "dependent variables" in ordinary regression. To a large extent the statistical methods for autoregressive models are regression or least squares procedures.

[illegible] Correspondence between Markov Chains and Autoregressive Processes.

The discrete variable x_t which takes the values $1, \ldots, m$ can be replaced by the m-component vector $\underset{\sim}{y}_t$ in which the i-th component is 1 if $x_t = i$ and is 0 if $x_t \neq i$. If we define $\underset{\sim}{\varepsilon}_i$ to be the m-component vector with 1 as the i-th component and 0 as the other components, we can define $\underset{\sim}{y}_t$ as $\underset{\sim}{\varepsilon}_i$ when $x_t = i$. Then

$$(2.24) \qquad \Pr\{\underset{\sim}{y}_t = \underset{\sim}{\varepsilon}_j \mid \underset{\sim}{y}_{t-1} = \underset{\sim}{\varepsilon}_i\} = p_{ij}(t), \qquad i,j = 1, \ldots, m .$$

Thus the conditional expectation of $\underset{\sim}{y}_t$ given the past is

$$(2.25) \qquad \mathcal{E}(\underset{\sim}{y}_t \mid \underset{\sim}{y}_{t-1}, \underset{\sim}{y}_{t-2}, \ldots) = \mathcal{E}(\underset{\sim}{y}_t \mid \underset{\sim}{y}_{t-1}) = \underset{\sim}{P}'(t)\underset{\sim}{y}_{t-1} .$$

If we define $\underset{\sim}{u}_t = \underset{\sim}{y}_t - \underset{\sim}{P}(t)'\underset{\sim}{y}_{t-1}$, then (2.25) implies $\mathcal{E}\underset{\sim}{u}_t = \underset{\sim}{0}$ and

$$(2.26) \qquad \mathcal{E}\underset{\sim}{u}_t \underset{\sim}{y}'_{t-s} = \mathcal{E}\mathcal{E}(\underset{\sim}{u}_t \mid \underset{\sim}{y}_{t-1}, \underset{\sim}{y}_{t-2}, \ldots)\underset{\sim}{y}'_{t-s} = \underset{\sim}{0}, \qquad s = 1, 2, \ldots .$$

The latter is equivalent to

$$(2.27) \qquad \mathcal{E}\underset{\sim}{y}_t \underset{\sim}{y}'_{t-1} = \underset{\sim}{P}(t)'\mathcal{E}\underset{\sim}{y}_{t-1} \underset{\sim}{y}'_{t-1} .$$

Note that here $\underset{\sim}{u}_t$ has a singular distribution since $\underset{\sim}{\varepsilon}'\underset{\sim}{u}_t = 0$. Conditional on $\underset{\sim}{y}_{t-1} = \underset{\sim}{\varepsilon}_i$, the variance of the j-th component of $\underset{\sim}{y}_t$ is $p_{ij}(t)[1-p_{ij}(t)]$ and the covariance between the j-th and k-th components is $-p_{ij}(t)p_{ik}(t)$. The unconditional variance is

$\Sigma_{i=1}^{m} p_i(t-1)[p_{ij}(t)\ 1-p_{ij}(t)]$, and the unconditional covariance is $-\Sigma_{i=1}^{m} p_i(t-1)\ p_{ij}(t)\ p_{ik}(t)$.

The properties of a Markov chain are similar to those of an autoregressive model, defined by (2.14), in that the conditional expectation of the observed vector y_t given the past is a linear function of the immediately preceding observed vector y_{t-1} . The autocovariances of y_t can be obtained from (2.27) in a manner analogous to that of the autoregressive model. In particular, in the stationary case (2.17) holds with B replaced by P' and F replaced by $\mathcal{E} y_{t-1} y'_{t-1} - \mathcal{E} y_{t-1} \mathcal{E} y'_{t-1} = \mathcal{E} y_{t-1} y'_{t-1} - p\, p'$; then $f_{jj} = p_j - \Sigma_{i=1}^{m} p_i\, p_{ij}^2$ and $f_{jk} = -\Sigma_{i=1}^{m} p_i\, p_{ij}\, p_{ik}$, $j \neq k$. The representation of the Markov chain differs from the autoregressive process in that the covariance matrix of the residual u_t is singular and depends on y_{t-1} in the conditional distribution; while u_t is uncorrelated with y_{t-1}, it is not statistically independent of y_{t-1} . We also note that the characteristic roots of B are less than 1 in absolute value, but one root of P (when P is irreducible) is exactly 1, corresponding to characteristic vectors of p on the left and ϵ on the right; ϵ is a characteristic vector of F corresponding to a root of 0. In the following pages the parallels between the two processes are used to suggest methods and derivations for one model from the other.

3. Estimation of Parameters.

3.1. Estimation of Transition Probabilities. An observation on an individual consists of the sequence of states for T successive time points. For example, the $T = 6$ successive monthly party preferences of a voter might constitute such a sequence. Let $x_{t\alpha}$ be the state of the α-th individual at the t-th time, $\alpha = 1, \ldots, N$, $t = 1, \ldots, T$. The observed sequence $\underset{\sim}{x}_{1\alpha}, \ldots, \underset{\sim}{x}_{T\alpha}$ is considered an observation from the Markov chain specified by the set of probabilities $[p_i(1)], [p_{ij}(2)], \ldots, [p_{ij}(T)]$. The probability of a given sequence of states $x(1), \ldots, x(T)$ is

$$(3.1) \qquad p_{x(1)}(1)\, p_{x(1),x(2)}(2) \cdots p_{x(T-1)x(T)}(T) .$$

The parameters to be estimated from a sample are the marginal probabilities $[p_i(1)]$ and the transition probabilities $[p_{ij}(2)], \ldots, [p_{ij}(T)]$. The observed sequences are considered as independent observations from the model defined by (3.1).

Let $n_{ij}(t)$ be the number of observed individuals in state i at time $t-1$ and j at time t, and let

$$(3.2) \qquad n_i(t-1) = \sum_{j=1}^{m} n_{ij}(t) = \sum_{k=1}^{m} n_{ki}(t-1) , \quad i = 1,\ldots,m, \quad t = 2,\ldots,T .$$

The set $n_{ij}(t)$, $i,j = 1, \ldots, m$, for each t constitutes the frequencies of individuals in state i at time $t-1$ and state j at time t and would usually be recorded in a two-way table; the row totals are $n_i(t-1)$, $i = 1, \ldots, m$, and the column totals are $n_j(t)$, $j = 1, \ldots, m$.

A sufficient set of statistics for the model (3.1) is $n_{ij}(t)$, $i,j = 1, \ldots, m$, $t = 2, \ldots, T$; statistical inference need only use this information. The maximum likelihood estimates of the parameters are

$$\hat{p}_i(1) = \frac{n_i(1)}{N}, \qquad i = 1, \ldots, m, \tag{3.3}$$

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)}, \qquad i,j = 1, \ldots, m, \quad t = 2, \ldots, T. \tag{3.4}$$

[If $n_i(t-1) = 0$, then $n_{ij}(t) = 0$ and (3.4) is undefined.] The estimates are in effect estimates of multinomial probabilities. The estimates satisfy $\hat{p}_i(1) \geq 0$, $\Sigma_{i=1}^{m} \hat{p}_i(1) = 1$, $\hat{p}_{ij}(t) \geq 0$, and $\Sigma_{j=1}^{m} \hat{p}_{ij}(t) = 1$.

Now consider the case of homogeneous transition probabilities, but with the initial probabilities $p_i(1)$ arbitrary. Then a sufficient set of statistics is $n_i(1)$, $i = 1, \ldots, m$, and

$$n_{ij} = \sum_{t=2}^{T} n_{ij}(t), \qquad i,j = 1, \ldots, m. \tag{3.5}$$

The two-way table of frequencies (3.5) is the sum of the $T - 1$ two-way tables with entries $n_{ij}(t)$. [However, $n_i(1)$ is not a marginal total of this table.] The maximum likelihood estimates are (3.3) and

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i^*}, \qquad i,j = 1, \ldots, m, \tag{3.6}$$

where $n_i^* = \Sigma_{j=1}^{m} n_{ij}$.

In an alternative model the states at the initial time $t = 1$ are considered as given; that is, $x_{1\alpha}$, $\alpha = 1, \ldots, N$, are treated as nonstochastic. Then $n_i(1)$, $i = 1, \ldots, m$, are considered as given, that is, are parameters, not statistics. The sufficient set of statistics is the set $n_{ij}(t)$, $t = 2, \ldots, T$, or n_{ij}, $i,j = 1, \ldots, m$ as the case may be.

In the case of homogeneous transition probabilities, it may be desired to treat the process as stationary; the marginal probability distribution $[p_i]$ is determined by $[p_{ij}]$ as the solution to (2.4); in particular, the initial distribution $[p_i(1)]$ must be $[p_i]$. A sufficient set of statistics is $n_i(1)$, $i = 1, \ldots, m$, and n_{ij}, $i,j = 1, \ldots, m$. While the parameter set can be reduced to $[p_{ij}]$, the maximum likelihood estimates are not (3.6) in this case because the likelihood function depends on $n_i(1)$ and $p_i(1) = p_i$, $i = 1, \ldots, m$, the latter being functions of $[p_{ij}]$. The estimates are too complicated to give explicitly.

To assess sampling variability and to evaluate confidence in inferences it is desirable to know the distributions of the estimates. Since the exact distributions for given sample sizes are too complicated to be useful we consider "large-sample theory". Anderson and Goodman (1957) developed asymptotic theory for the number of observations N getting large; this large-sample theory is appropriate for panel studies where the number of time points is small (sometimes $T = 2$) and the number of respondents is large. When the transition probabilities are homogeneous, the parameters do not depend on the time t except the initial probabilities $[p_i(1)]$; then it is meaningful to consider asymptotic distributions as $T \to \infty$. This theory is appropriate when the data consist of one (or several) long series; the measurements are not necessarily repeated. Bartlett (1951) gave some of this theory. In general, when a limiting distribution hold for $T \to \infty$ it will hold for arbitrary N; in fact, with proper normalization the same limiting distribution holds for $N \to \infty$ and fixed T. In such a case we will say the limit holds as N and/or $T \to \infty$. [In mathematical terms the error is arbitrarily small if N is sufficiently large or T is sufficiently large, or both.]

The asymptotic theory for $N \to \infty$ is the usual multinomial theory, for the N observations on the chain are independent. Then by the usual multinomial theory $\sqrt{N}[\hat{p}_i(1)-p_i(1)]$, $i = 1, \ldots, m$, have a limiting normal distribution with means 0 and covariance matrix $[p_i(1)\,\delta_{ij} - p_i(1)\,p_j(1)]$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$. If $p_i(t-1) > 0$, the set $\sqrt{n_i(t-1)}[\hat{p}_{ij}(t)-p_{ij}(t)]$, $j = 1, \ldots, m$, have a limiting normal distribution with means 0 and covariance matrix $[p_{ij}(t)\delta_{jk} \; p_{ij}(t)p_{ik}(t)]$; the sets for different values of i and/or different values of t are independent in the limiting distribution. The limiting distribution of the estimates of the rows of the transition probability matrices is the same as that of estimates of independent multinomial distributions.

If the transition probabilities are homogeneous and the chain is irreducible, then for each i the set $\sqrt{n_i^*}(\hat{p}_{ij}-p_{ij})$, $j = 1, \ldots, m$, has a limiting normal distribution with means 0, variances $p_{ij}(1-p_{ij})$ and covariances $-p_{ij}p_{ik}$, $j \neq k$, and the sets for different values of i are independent in the limiting distribution. The limits in the homogeneous case are valid as $N \to \infty$ and/or $T \to \infty$.

3.2. Estimation of Autoregressive Coefficients. Let $\underset{\sim}{y}_{t\alpha}$ be the p-component vector of measurements of the α-th individual at the t-th time point, $\alpha = 1, \ldots, N$, $t = 1, \ldots, T$. The model is a first-order autoregressive model (2.14) with $\underset{\sim}{u}_t$ having the normal distribution $N(\underset{\sim}{0}, \underset{\sim}{\Sigma}_t)$ and $\underset{\sim}{y}_1$ having the normal distribution $N(\underset{\sim}{0}, \underset{\sim}{F}_1)$. The probability density of the sequence $\underset{\sim}{y}_{1\alpha}, \ldots, \underset{\sim}{y}_{T\alpha}$ for a given α is

$$(3.7) \quad \frac{1}{(2\pi)^{\frac{1}{2}Tp} |\underset{\sim}{F}_1|^{\frac{1}{2}} \prod_{t=2}^{T} |\underset{\sim}{\Sigma}_t|^{\frac{1}{2}}}$$

$$\times \exp\left\{-\frac{1}{2}\left[\underset{\sim}{y}_{1\alpha}' \underset{\sim}{F}_1^{-1} \underset{\sim}{y}_{1\alpha} + \sum_{t=2}^{T} (\underset{\sim}{y}_{t\alpha} - \underset{\sim}{B}(t)\underset{\sim}{y}_{t-1,\alpha})' \underset{\sim}{\Sigma}_t^{-1} (\underset{\sim}{y}_t - \underset{\sim}{B}(t)\underset{\sim}{y}_{t-1,\alpha})\right]\right\} .$$

Then a sufficient set of statistics for $\underset{\sim}{F}_1, \underset{\sim}{B}(1), \ldots, \underset{\sim}{B}(T), \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_T$ is $\Sigma_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t\alpha}'$, $t = 1, \ldots, T$, and $\Sigma_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}_{t\alpha}'$, $t = 2, \ldots, T$. The maximum likelihood estimates of the parameter matrices are

$$(3.8) \qquad \hat{\underset{\sim}{F}}_1 = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{1\alpha} \underset{\sim}{y}_{1\alpha}' ,$$

$$(3.9) \qquad \hat{\underset{\sim}{B}}(t) = \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t-1,\alpha}' \left(\sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}_{t-1,\alpha}'\right)^{-1} , \qquad t = 2, \ldots, T ,$$

$$(3.10) \quad \hat{\underset{\sim}{\Sigma}}_t = \frac{1}{N} \sum_{\alpha=1}^{N} \left(\underset{\sim}{y}_{t\alpha} - \hat{\underset{\sim}{B}}(t)\underset{\sim}{y}_{t-1,\alpha}\right)\left(\underset{\sim}{y}_{t\alpha} - \hat{\underset{\sim}{B}}(t)\underset{\sim}{y}_{t-1,\alpha}\right)'$$

$$= \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t\alpha}' - \hat{\underset{\sim}{B}}(t) \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}_{t-1,\alpha}' \hat{\underset{\sim}{B}}(t)' , \quad t = 2, \ldots, T .$$

The components of $\hat{\underset{\sim}{B}}(t)$ are least squares estimates. [See Anderson (1958), Chapter 8, and Anderson (1971), Chapter 5, for example.]

The assumption that the autoregression matrices are homogeneous and the disturbances identically distributed leads to considerable simplification. The sufficient set of statistics for $\underset{\sim}{F}_1$, $\underset{\sim}{B}$, and $\underset{\sim}{\Sigma}$ is $\Sigma_{\alpha=1}^{N} \underset{\sim}{y}_{1\alpha} \underset{\sim}{y}_{1\alpha}'$, $\Sigma_{t=2}^{T} \Sigma_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t\alpha}'$, and $\Sigma_{t=2}^{T} \Sigma_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t-1,\alpha}'$. The maximum likelihood estimate of $\underset{\sim}{F}_1$ is (3.8) and the maximum likelihood estimates of the other matrices are

$$(3.11) \qquad \hat{\underset{\sim}{B}} = \sum_{t=2}^{T} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t-1,\alpha}' \left(\sum_{t=2}^{T} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}_{t-1,\alpha}' \right)^{-1},$$

$$(3.12) \qquad \hat{\underset{\sim}{\Sigma}} = \frac{1}{N(T-1)} \sum_{t=2}^{T} \sum_{\alpha=1}^{N} (\underset{\sim}{y}_{t\alpha} - \hat{\underset{\sim}{B}} \underset{\sim}{y}_{t-1,\alpha})(\underset{\sim}{y}_{t\alpha} - \hat{\underset{\sim}{B}} \underset{\sim}{y}_{t-1,\alpha})'$$

$$= \frac{1}{N(T-1)} \left[\sum_{t=2}^{T} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t\alpha} \underset{\sim}{y}_{t\alpha}' - \hat{\underset{\sim}{B}} \sum_{t=2}^{T} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}_{t-1,\alpha}' \hat{\underset{\sim}{B}}' \right].$$

An alternative model is to consider $\underset{\sim}{y}_{1\alpha}$, $\alpha = 1, \ldots, N$, as nonstochastic or fixed and treat $\underset{\sim}{y}_{t\alpha}$, $t = 2, \ldots, T$, $\alpha = 1, \ldots, N$, conditionally. Then the maximum likelihood estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$ are (3.11) and (3.12).

When $\underset{\sim}{y}_{1\alpha}$ is considered to have the marginal normal distribution determined by the stationary process, the covariance matrix $\underset{\sim}{F}_1$ is a function of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$ as given by (2.16). Then the maximum likelihood estimates are much more complicated. (As $T \to \infty$, (3.11) and (3.12) are asymptotically equivalent to the maximum likelihood estimates, but not as $N \to \infty$.)

We now consider the asymptotic properties of the estimates as $N \to \infty$ and/or $T \to \infty$. As $N \to \infty$, $\hat{F}_1$, $\hat{B}(t)$, and $\hat{\Sigma}_t$ are consistent estimates of F_1, $B(t)$, and Σ_t, $t = 2, \ldots, T$, respectively. The elements of $\sqrt{N}[\hat{B}(t) - B(t)]$ have a limiting normal distribution with means 0 and covariances constituting the Kronecker product $\Sigma_t \otimes F_{t-1}^{-1}$, $t = 2, \ldots, T$. The matrix F_{t-1} is estimated consistently by $(1/N)\Sigma_{\alpha=1}^{N} y_{t-1,\alpha} y_{t-1,\alpha}'$.

In the case of homogeneous autoregressive coefficients, regardless of the distribution of $y_{1\alpha}$, $\alpha = 1, \ldots, N$, and of the value of N, as $T \to \infty$ $\hat{B}$ and $\hat{\Sigma}$ are consistent estimates of B and Σ, respectively, and $\sqrt{T}[\hat{B} - B]$ has a limiting normal distribution with means 0 and covariances constituting $(1/N)\,\Sigma \otimes F^{-1}$. The matrix F is consistently estimated by $[1/N(T-1)]\,\Sigma_{t=2}^{T} \Sigma_{\alpha=1}^{N} y_{t-1,\alpha} y_{t-1,\alpha}'$.

If T is fixed and $N \to \infty$, $\sqrt{N}(\hat{B} - B)$ has a limiting normal distribution with means 0 and covariances constituting

$$\Sigma \otimes \left(\sum_{t=2}^{T} F_t \right)^{-1}, \tag{3.13}$$

3.3. Correspondence of Sufficient Statistics and Estimates in Markov Chains and Autoregressive Processes. In Section 2.3 a Markov chain was represented as a vector process with $y_t = \varepsilon_j$ with conditional probability $p_{ij}(t)$ given $y_{t-1} = \varepsilon_i$. From this definition we can write the second-order moment matrices for the Markov chain as

$$\sum_{\alpha=1}^{N} y_{t\alpha} y_{t\alpha}' = \sum_{i=1}^{m} n_i(t) \varepsilon_i \varepsilon_i', \tag{3.14}$$

which is a diagonal matrix with $n_i(t)$ as the i-th diagonal element, and

$$(3.15) \qquad \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}'_{t\alpha} = \sum_{i,j=1}^{m} n_{ij}(t) \underset{\sim}{\varepsilon}_i \underset{\sim}{\varepsilon}'_j ,$$

which is an $m \times m$ matrix with $n_{ij}(t)$ as the i,j-th element. Note that the elements of (3.14) can be derived from the elements of (3.15). Here (3.15) for $t = 2, \ldots, T$ constitute a sufficient set of statistics for the nonhomogeneous Markov chain. The estimates (3.4) constitute elements of the matrix estimates (3.9) under this correspondence.

If the transition probabilities are homogeneous with arbitrary initial probabilities, a sufficient set of statistics is (3.14) for $t = 1$ and

$$(3.16) \qquad \sum_{t=2}^{T} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{t-1,\alpha} \underset{\sim}{y}'_{t\alpha} = \sum_{i,j=1}^{m} n_{ij} \underset{\sim}{\varepsilon}_i \underset{\sim}{\varepsilon}'_j .$$

The estimates (3.6) are elements of (3.11).

4. Tests of Hypotheses.

4.1. Tests for Markov Chains. The tests for Markov chains presented in this section were developed by Anderson and Goodman (1957); they can be applied for any value of T (≥ 2) and for large N. Bartlett (1951) developed some of the tests as valid for one observed sequence of states from a homogeneous chain when $T \to \infty$. For convenience, we shall assume $p_{ij}(t) > 0$ and $p_{ij} > 0$ as the case may be. Some of the procedures were illustrated in Anderson (1954). In Section 4.2 test criteria for the corresponding hypotheses for autoregressive models are given (in the same sequence) and in Section 4.3 the correspondences are discussed.

Specified Probabilities. In the homogeneous chain to test the null hypothesis that $p_{ij} = p^o_{ij}$, $j = 1, \ldots, m$, where the set p^o_{ij} are specified, for a given i one can use the criterion

$$n_i^* \sum_{j=1}^{m} \frac{(\hat{p}_{ij} - p^o_{ij})^2}{p^o_{ij}} , \tag{4.1}$$

which under the null hypothesis has a limiting χ^2-distribution with $m - 1$ degrees of freedom as $N \to \infty$ and/or as $T \to \infty$. The criteria for different i are asymptotically independent. To test the null hypothesis $p_{ij} = p^o_{ij}$, $i,j = 1, \ldots, m$, the sum of (4.1) over i can be used; it has a limiting χ^2-distribution with $m(m-1)$ degrees of freedom when the null hypothesis is true. If the transition probabilities are not necessarily homogeneous one can test the null hypothesis $p_{ij}(t) = p^o_{ij}(t)$, $j = 1, \ldots, m$, for given i and t by use of the criterion

$$n_i(t-1) \sum_{j=1}^{m} \frac{[\hat{p}_{ij}(t) - p^o_{ij}(t)]^2}{p^o_{ij}(t)} , \tag{4.2}$$

which under the null hypothesis has a limiting χ^2-distribution as $N \to \infty$. The criteria for different i and t are asymptotically independent; they can be summed over i and/or t to form χ^2-criteria for combined hypotheses. These criteria are analogs of the χ^2 goodness-of-fit criterion for multinomial distributions. The test procedures can be inverted in the usual fashion to obtain confidence regions for the transition probabilities.

Homogeneity. In treating panel data the investigator may question whether conditions change enough over the time of observation to require the use of an inhomogeneous chain. To test the null hypothesis that $p_{ij}(t) = p_{ij}$, $t = 2, \ldots, T$, for some p_{ij}, $i,j = 1, \ldots, m$, one may use the criterion

$$\sum_{t=2}^{T} \sum_{i,j=1}^{m} n_i(t-1) \frac{[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2}{\hat{p}_{ij}} \quad . \tag{4.3}$$

Under the null hypothesis this criterion has a limiting χ^2-distribution with $m(m-1)(T-2)$ degrees of freedom as $N \to \infty$. If one thinks of the set of probabilities $p_{ij}(t)$ and the set of frequencies $n_{ij}(t)$ in (three-way) $m \times m \times (T-1)$ arrays, the criterion (4.3) is the usual χ^2-criterion for testing the independence of the categorization (i,j) and the classification t.

Independence. If $p_{ij} = p_j$, $i = 1, \ldots, m$, for some p_j, $j = 1, \ldots, m$, the random variable x_t is independent of x_{t-1} in a homogeneous Markov chain. To test the null hypothesis of independence one may use the criterion

$$\sum_{i,j=1}^{m} n_i^* \frac{(\hat{p}_{ij} - \hat{p}_j)^2}{\hat{p}_j} \quad , \tag{4.4}$$

where

$$\hat{p}_j = \frac{\Sigma_{i=1}^{m} n_{ij}}{N(T-1)} \quad , \qquad j = 1, \ldots, m \quad . \tag{4.5}$$

Under the null hypothesis the criterion has a limiting χ^2-distribution with $(m-1)^2$ degrees of freedom as $N \to \infty$ and/or $T \to \infty$. The criterion is the χ^2-criterion for independence in the two-way table $\{n_{ij}\}$.

Given Order. An investigator may consider a more elaborate model in which the probability of a state observed at time t depends on the state observed in the last r time points. He may question whether it would be appropriate to use a model of lower order. Independence is order $r = 0$; in the previous section this hypothesis was tested against the alternative that $r = 1$. As another example, we consider testing the null hypothesis that a homogeneous second-order chain is first-order, that is, that $p_{ijk} = p_{jk}$ for some suitable $p_{jk}(p_{jk} \geq 0, \Sigma_{k=1}^{m} p_{jk} = 1)$.

In a second-order homogeneous chain with $n_i(1)$ and $n_{ij}(2)$ as given [or $p_i(1)$ and $p_{ij}(2)$ as arbitrary] the maximum likelihood estimates of p_{ijk} are

(4.6) $$p_{ijk} = \frac{n_{ijk}}{n^*_{ij}} , \qquad i,j,k = 1, \ldots, m ,$$

where $n_{ijk} = \Sigma_{t=3}^{T} n_{ijk}(t)$, $n^*_{ij} = \Sigma_{k=1}^{m} n_{ijk}$, and $n_{ijk}(t)$ is the number of observations of state i at $t - 2$, j at $t - 1$, and k at t . Then a criterion for testing that an assumed second-order chain is actually a first-order chain is

(4.7) $$\sum_{i,j,k=1}^{m} n^*_{ij} \frac{(\hat{p}_{ijk} - \hat{p}^*_{jk})^2}{\hat{p}^*_{jk}} ,$$

where

(4.8) $$\hat{p}^*_{jk} = \frac{\Sigma_{i=1}^{m} n_{ijk}}{\Sigma_{i=1}^{m} n^*_{ij}} .$$

When the null hypothesis is true, (4.7) has a limiting χ^2-distribution with $m(m-1)^2$ degrees of freedom as $N \to \infty$ and/or $T \to \infty$.

Several Chains Identical. A population may be stratified into several subpopulations and the transition probabilities may be different. Suppose we have samples from s Markov chains with transition probabilities $p_{ij}^{(h)}$, $h = 1, \ldots, s$, and we wish to test the null hypothesis that the chains are identical, that is, that $p_{ij}^{(h)} = p_{ij}$, $h = 1, \ldots, s$, for some p_{ij}. Let $\hat{p}_{ij}^{(h)}$ be the maximum likelihood estimate of the transition probability from the h-th sample, and let $\hat{p}_{ij}^{(\cdot)}$ be the estimate based on all s samples under the assumption of the null hypothesis. The criterion is

$$(4.9) \qquad \sum_{h=1}^{s} \sum_{i,j=1}^{m} n_i^{*(h)} \frac{[\hat{p}_{ij}^{(h)} - \hat{p}_{ij}^{(\cdot)}]^2}{\hat{p}_{ij}^{(\cdot)}} ,$$

which has the χ^2-distribution with $(s-1)m(m-1)$ degrees of freedom under the null hypothesis as $N \to \infty$ and/or $T \to \infty$.

Independence of Two Sets of States. Suppose the state of a Markov chain is determined by a pair of responses (that is, answers to two questions). Denote the state as (α,β), $\alpha = 1, \ldots, A$, and $\beta = 1, \ldots, B$, where α denotes the first answer or class and β the second, and the transition probabilities as $p_{\alpha\beta,\mu\nu}$. Is the sequence of changes in one classification independent of that in the second? The null hypothesis is

$$(4.10) \qquad p_{\alpha\beta,\mu\nu} = q_{\alpha\mu} \, r_{\beta\nu} , \qquad \alpha,\mu = 1, \ldots, A, \quad \beta,\nu = 1, \ldots, B ,$$

where $q_{\alpha\mu}$ is a transition probability for the first classification and $r_{\beta\nu}$ is for the second. Let $n_{\alpha\beta,\mu\nu}(t)$ be the number of individuals in state (α,β) at $t - 1$ and (μ,ν) at t, and let

$n_{\alpha\beta,\mu\nu} = \Sigma_{t=2}^{T} n_{\alpha\beta,\mu\nu}(t)$. The maximum likelihood estimate of $p_{\alpha\beta,\mu\nu}$ is

$$(4.11) \qquad \hat{p}_{\alpha\beta,\mu\nu} = \frac{n_{\alpha\beta,\mu\nu}}{\Sigma_{\gamma=1}^{A} \Sigma_{\delta=1}^{B} n_{\alpha\beta,\gamma\delta}} , \quad \alpha,\mu = 1,\ldots,A, \quad \beta,\gamma = 1,\ldots,B,$$

when the null hypothesis is not assumed and is $\hat{q}_{\alpha\mu}\ \hat{r}_{\beta\nu}$, where

$$(4.12) \qquad \hat{q}_{\alpha\mu} = \frac{\Sigma_{\beta,\nu=1}^{B} n_{\alpha\beta,\mu\nu}}{\Sigma_{\gamma=1}^{A} \Sigma_{\beta,\nu=1}^{B} n_{\alpha\beta,\gamma\nu}} , \qquad \hat{r}_{\beta\nu} = \frac{\Sigma_{\alpha,\mu=1}^{A} n_{\alpha\beta,\mu\nu}}{\Sigma_{\alpha,\mu=1}^{A} \Sigma_{\delta=1}^{B} n_{\alpha\beta,\mu\delta}}$$

when the null hypothesis is assumed. The χ^2-criterion for testing the null hypothesis of independence is

$$(4.13) \qquad \sum_{\alpha,\mu=1}^{A} \sum_{\beta,\nu=1}^{B} n_{\alpha\beta}^{*} \frac{(\hat{p}_{\alpha\beta,\mu\nu} - \hat{q}_{\alpha\mu}\ \hat{r}_{\beta\nu})^2}{\hat{q}_{\alpha\mu}\ \hat{r}_{\beta\nu}} ,$$

where $n_{\alpha\beta}^{*} = \Sigma_{\gamma=1}^{A} \Sigma_{\delta=1}^{B} n_{\alpha\beta,\gamma\delta}$. When the null hypothesis is true, the criterion has a limiting χ^2-distribution with $AB(AB-1) - A(A-1) - B(B-1) = (A-1)(B-1)\ (AB+A+B)$ degrees of freedom as $N \to \infty$ and/or $T \to \infty$.

4.2. Tests for Autoregressive Processes. The development and application of procedures is based on asymptotic theory as $N \to \infty$ and/or $T \to \infty$. In this section we study methods of testing hypotheses about the matrices of autoregressive coefficients and, in some cases, hypotheses about covariance matrices. The hypotheses correspond to those concerning Markov chains presented in Section 4.1. The procedures are analogs of procedures in multivariate regression. [See Anderson (1958), Chapter 8, for example.] For many hypotheses there are choices of test criteria that have the same limiting distributions under the respective null hypotheses. We have usually chosen a trace criterion (the Lawley-Hotelling criterion in the case of a linear hypothesis) as the criterion most similar to the criterion for the corresponding hypothesis about Markov chains.

Specified Autoregression Matrices. If the autoregression matrices are the same, that is, $B(1) = \dots = B(T) = B$, one can test the null hypothesis that B is a specified $p \times p$ matrix B^0 by use of the criterion

$$(4.14) \qquad \operatorname{tr} (\hat{B} - B^0) \sum_{t=2}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} (\hat{B} - B^0)' \hat{\Sigma}^{-1} .$$

As $N \to \infty$ and/or $T \to \infty$, this criterion has a limiting χ^2-distribution with p^2 degrees of freedom. Other criteria, such as the likelihood ratio criterion, could be used. A set of test procedures can be inverted to obtain confidence regions; a confidence region for B consists of all matrices B^0 such that (4.14) is less than a suitable number from the χ^2-tables.

Equality of Autoregression Matrices given Equality of Covariance Matrices. Suppose $\Sigma_t = \Sigma$, $t = 2, \dots, T$, for some covariance matrix Σ; [illegible] consider testing the null hypothesis $B(t) = B$ for some matrix B. The criterion

$$(4.15) \quad \operatorname{tr} \sum_{t=2}^{T} [\hat{B}(t) - \hat{B}] \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} [\hat{B}(t) - \hat{B}]' \hat{\Sigma}^{-1} =$$

$$= \operatorname{tr} \left[\sum_{t=2}^{T} \hat{B}(t) \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} \hat{B}(t)' - \hat{B} \sum_{t=2}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} \hat{B}' \right] \hat{\Sigma}^{-1}$$

has a limiting χ^2-distribution with $(T-2)p^2$ degrees of freedom as

$N \to \infty$ when the null hypothesis is true.

Independence. In the first-order autoregressive model independence at different time points is identical to $B(2) = \ldots = B(T) = 0$, given $\Sigma_2 = \ldots = \Sigma_T$, or in the homogeneous case $B = 0$. In the latter case the null hypothesis of independence is that the autoregression matrix is the zero matrix and the criterion is (4.14) with $B^0 = 0$, that is,

$$\operatorname{tr} \hat{B} \sum_{t=2}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} \hat{B}' \hat{\Sigma}^{-1} . \tag{4.16}$$

Given Order. The second-order homogeneous process is defined by (2.19). It is a first-order process if $B_2 = 0$. To test this null hypothesis we need estimates of B_1 and B_2

$$[\hat{B}_1 \ \hat{B}_2] = \left[\sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t\alpha} y'_{t-1,\alpha} \quad \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t\alpha} y'_{t-2,\alpha} \right] \begin{bmatrix} \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} & \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-2,\alpha} \\ \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-2,\alpha} y'_{t-1,\alpha} & \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-2,\alpha} y'_{t-2,\alpha} \end{bmatrix}^{-1} . \tag{4.17}$$

These are maximum likelihood estimates if $y_{1\alpha}, y_{2\alpha}, \alpha = 1, \ldots, N$, are considered as fixed. The covariance matrix Σ is estimated by

$$\hat{\Sigma} = \frac{1}{N(T-2)} \sum_{t=3}^{T} \sum_{\alpha=1}^{N} [y_{t\alpha} - \hat{B}_1 y_{t-1,\alpha} - \hat{B}_2 y_{t-2,\alpha}] [y_{t\alpha} - \hat{B}_1 y_{t-1,\alpha} - \hat{B}_2 y_{t-2,\alpha}]' . \tag{4.18}$$

A criterion for testing the null hypothesis is

$$\operatorname{tr} \hat{B}_2 \left[\sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-2,\alpha} y'_{t-2,\alpha} - \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-2,\alpha} y'_{t-1,\alpha} \left(\sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-1,\alpha} \right)^{-1} \right.$$

$$\left. \sum_{t=3}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha} y'_{t-2,\alpha} \right] \hat{B}'_2 \hat{\Sigma}^{-1} . \tag{4.19}$$

Under the null hypothesis the criterion has a limiting χ^2-distribution with p^2 degrees of freedom as $N \to \infty$ and/or $T \to \infty$.

Several Processes Identical. Consider testing the null hypothesis that the matrices of autoregressive coefficients of s first-order homogeneous autoregressive processes with identical covariance matrices Σ are equal on the basis of N_h observed time series of length T from the h-th process, $h = 1, \ldots, s$. If $\hat{B}^{(h)}$ is the estimate of the matrix for the h-th process, $\hat{B}$ is the pooled estimate of the hypothetically equal matrices, and $\hat{\Sigma}$ is the pooled estimate of Σ, a criterion for testing the null hypothesis is

$$\operatorname{tr} \sum_{h=1}^{s} (\hat{B}^{(h)} - \hat{B}) \sum_{t=2}^{T} \sum_{\alpha=1}^{N_h} y^{(h)}_{t-1,\alpha} y^{(h)'}_{t-1,\alpha} (\hat{B}^{(h)} - \hat{B})' \hat{\Sigma}^{-1} \tag{4.20}$$

$$= \operatorname{tr} \left[\sum_{h=1}^{s} \hat{B}^{(h)} \sum_{t=2}^{T} \sum_{\alpha=1}^{N_h} y^{(h)}_{t-1,\alpha} y^{(h)'}_{t-1,\alpha} \hat{B}^{(h)'} \right.$$

$$\left. - \hat{B} \sum_{h=1}^{s} \sum_{t=2}^{T} \sum_{\alpha=1}^{N_h} y^{(h)}_{t-1,\alpha} y^{(h)'}_{t-1,\alpha} \hat{B}' \right] \hat{\Sigma}^{-1} .$$

Under the null hypothesis this has a χ^2-distribution with $(s - 1)p^2$ degrees of freedom.

Independence of Two Subprocesses. Suppose $\underset{\sim}{y}_t = (\underset{\sim}{y}_t^{(1)\prime}\ \underset{\sim}{y}_t^{(2)\prime})'$ and we ask the question whether the first-order autoregressive process is such that the two subprocesses $\{\underset{\sim}{y}_t^{(1)}\}$ and $\{\underset{\sim}{y}_t^{(2)}\}$ are independent. Let

$$\underset{\sim}{B} = \begin{pmatrix} \underset{\sim}{B}_{11} & \underset{\sim}{B}_{12} \\ \underset{\sim}{B}_{21} & \underset{\sim}{B}_{22} \end{pmatrix}, \quad \underset{\sim}{\Sigma} = \begin{pmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} \\ \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{pmatrix} . \tag{4.21}$$

The two subprocesses (in the Gaussian case) are independent if and only if

$$\underset{\sim}{B}_{12} = \underset{\sim}{0},\ \underset{\sim}{B}_{21} = \underset{\sim}{0} , \tag{4.22}$$

$$\underset{\sim}{\Sigma}_{12} = \underset{\sim}{\Sigma}_{21}' = \underset{\sim}{0} . \tag{4.23}$$

The estimates (3.11) and (3.12) are partitioned similarly. A test of the hypothesis (4.23) can be based on the criterion

$$N(T-1)\ \mathrm{tr}\ \hat{\underset{\sim}{\Sigma}}_{11}^{-1}\ \hat{\underset{\sim}{\Sigma}}_{12}\ \hat{\underset{\sim}{\Sigma}}_{22}^{-1}\ \hat{\underset{\sim}{\Sigma}}_{21} , \tag{4.24}$$

which has a limiting χ^2-distribution with p_1p_2 degrees of freedom, (as $N \to \infty$ and/or $T \to \infty$) , where p_1 is the number of coordinates in $\underset{\sim}{y}_t^{(1)}$ and p_2 is the number in $\underset{\sim}{y}_t^{(2)}$

When $\underset{\sim}{\Sigma}_{12} = \underset{\sim}{\Sigma}_{21}' = \underset{\sim}{0}$, a criterion to test the null hypothesis $\underset{\sim}{B}_{12} = \underset{\sim}{0}$ is

$$(4.25)\qquad \operatorname{tr} \hat{B}_{12}\Bigg[\sum_{t=2}^{T}\sum_{\alpha=1}^{N} y^{(2)}_{t-1,\alpha}\, y^{(2)'}_{t-1,\alpha} - \sum_{t=2}^{T}\sum_{\alpha=1}^{N} y^{(2)}_{t-1,\alpha}\, y^{(1)'}_{t-1,\alpha} \left(\sum_{t=2}^{T}\sum_{\alpha=1}^{N} y^{(1)}_{t-1,\alpha}\, y^{(1)'}_{t-1,\alpha}\right)^{-1} \sum_{t=2}^{T}\sum_{\alpha=1}^{N} y^{(1)}_{t-1,\alpha}\, y^{(2)'}_{t-1,\alpha}\Bigg] \hat{\Sigma}'_{12}\,\hat{\Sigma}^{-1}_{11} ;$$

its limiting distribution as $N \to \infty$ and/or $T \to \infty$ is a χ^2-distribution with $p_1 p_2$ degrees of freedom. A criterion to test the null hypothesis $B_{21} = 0$ is (4.25) with 1 and 2 interchanged; its limiting distribution is also a χ^2-distribution with $p_1 p_2$ degrees of freedom. The three criteria are asymptotically independent.

4.3. Correspondence of Tests. The matrix P has $m(m-1)$ elements to specify because the sum of elements in each row is 1, while the matrix B has p^2. In most cases, if a χ^2-test for a hypothesis about P or $P(t)$ has as the number of degrees of freedom a multiple of $m(m-1)$, the corresponding test for B or $B(t)$ has the same multiple of p^2 as the number of degrees of freedom. According to the correspondence set up in Section 3.3, n_i^* is the i-th diagonal element of the diagonal matrix $\sum_{t=2}^{T} \sum_{\alpha=1}^{N} y_{t-1,\alpha}\, y'_{t-1,\alpha}$ and the diagonal matrix with diagonal elements $1/p_{ij}$ is a generalized inverse of the limiting covariance matrix of $\sqrt{n_i^*}\,(\hat{p}_{i1} - p_{i1}), \ldots, \sqrt{n_i^*}\,(\hat{p}_{im} - p_{im})$. Then the sum of (4.1) over i and (4.14) correspond as criteria for specified matrices.

Since $n_i(t-1)$ is the i-th diagonal element of $\sum_{\alpha=1}^{N} y_{t-1,\alpha}\, y'_{t-1,\alpha}$ of the representation for the Markov chain, (4.3) and (4.15) correspond as tests of homogeneity.

For independence (4.4) and (4.16) correspond; in the discrete case the covariance matrix of $\hat{p}_{ij}$ does not depend on i . The correspondence between degrees of freedom, however, is $(m - 1)^2$ and p^2 because independence in the discrete case is not defined by setting $\underset{\sim}{P} = \underset{\sim}{0}$. (Independence is $\underset{\sim}{P} = \underset{\sim}{\varepsilon}\,\underset{\sim}{p}'$.)

For testing whether a second-order model is actually first-order criteria (4.7) and (4.19) correspond. In the latter $\hat{\underset{\sim}{\Sigma}}$ could be replaced by (3.12), which is a consistent estimate of $\underset{\sim}{\Sigma}$ when the null hypothesis is true, or $\hat{\underset{\sim}{\Sigma}}$ could be replaced by the matrix in brackets divided by $N(T - 2)$. Note the degrees of freedom are $m(m - 1)^2$ and p^2 , respectively.

To test equality of matrices criteria (4.9) and (4.20) correspond. The tests of independence, however, do not have similar structures.

Acknowledgments. The author is indebted to Persi Diaconis, Neil Henry, and Paul Shaman for reading earlier versions of this paper and making useful suggestions.

REFERENCES

Anderson, T. W. (1954), "Probability Models for Analyzing Time Changes in Attitudes", Mathematical Thinking in the Social Sciences, Paul F. Lazarsfeld, ed., The Free Press, Glencoe, Illinois, pp. 17-66.

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York.

Anderson, T. W. (1971), The Statistical Analysis of Time Series, John Wiley and Sons, Inc., New York.

Anderson, T. W., and Goodman, Leo A. (1957), "Statistical Inference about Markov Chains", Annals of Mathematical Statistics, Vol. 28, pp. 89-110.

Bartlett, M. S. (1951), "The Frequency Goodness of Fit Test for Probability Chains", Proceedings of the Cambridge Philosophical Society, Vol. 47, pp. 86-95.

Billingsley, Patrick (1961), "Statistical Methods in Markov Chains", Annals of Statistics, Vol. 32, pp. 12-40.

Lazarsfeld, Paul F., Berelson, Bernard, and Gaudet, Hazel (1948), The People's Choice, Second Edition, Columbia University Press, New York.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time Series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations," Steffen L. Lauritzen, May, 1974.

18. "On the Interrelationships among Sufficiency, Total Sufficiency and Some Related Concepts," Steffen L. Lauritzen, June, 1974.

19. "Statistical Inference for Multiply Truncated Power Series Distributions," T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains," T. W. Anderson, June 1975.

21. "Asymptotic Properties of Some Estimators in Moving Average Models," Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting," Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear Discriminant Function," Chien-Pai Han, April 1976.

24. "Panels and Time Series Analysis: Markov Chains and Autoregressive Processes," T. W. Anderson, July 1976.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR-24	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): PANELS AND TIME SERIES ANALYSIS: MARKOV CHAINS AND AUTOREGRESSIVE PROCESSES	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Theodore W. Anderson	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Stanford University, Stanford, California 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics & Probability Program Code 436, Arlington, Virginia 22217	12. REPORT DATE: July 1976
	13. NUMBER OF PAGES: 31
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)

Markov chain, autoregressive process, panel survey, time series analysis, maximum likelihood estimates, chi-square tests.

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

SEE REVERSE SIDE

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)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT

Statistical inference in two time series models is developed for cases where there are several observations on the entire time series. Emphasis is on first-order processes: the Markov chain for discrete data and the first-order autoregressive process for vectors of continuous variables. The models are not necessarily homogeneous (or stationary) in time. Sufficient statistics and maximum likelihood estimates are presented. (Continuous variables are assumed normally distributed.) Test criteria for various hypotheses are developed; on a large-sample basis these criteria have χ^2-distributions. The close correspondence between the properties and statistical methods for the two models is pointed out.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)