The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 1792-1807

ORDER OF DEPENDENCE IN A STATIONARY NORMALLY DISTRIBUTED TWO-WAY SERIES

BY P. K. BHATTACHARYA

University of Arizona

Order of dependence is defined in a normally distributed two-way series. Under certain stationarity and symmetry conditions it is shown that when the extents of the series in both directions are large in comparison with the order of dependence, the joint density function reduces to a simple form with a small number of parameters after some adjustments. In this form a central role is played by the Kronecker products of some matrices having common eigenvectors. Maximum likelihood estimates of the parameters and likelihood ratio test criteria for certain hypotheses on order of dependence are derived.

1. Introduction. If, in a normally distributed stationary time series $\{x_1, \cdots, x_T\}$ with zero means, the conditional distribution of x_t given $x_{t-1}, \cdots, x_1$ depends only on $x_{t-1}, \cdots, x_{t-q}$ for $t = q + 1, \cdots, T$, and if T is large in comparison with q, the joint probability density function (pdf) of $x_1, \cdots, x_T$ can be brought to a simple form by some small adjustments. Since these adjustments involve only the end terms of the series, the interdependence in the rest of the series remains practically unaffected and can be studied in a systematic manner in terms of only $q + 1$ parameters under the simplified probability model. The above simplification was adopted by Anderson (1962) whose principal concern was to develop optimum statistical decision procedures for determining the value of q which is called the order of dependence in the time series.

In this paper our aim is to develop a simple model for the interdependence in a normally distributed two-way series $\{x_{st}, s = 1, \cdots, S, t = 1, \cdots, T\}$ satisfying some stationarity and symmetry conditions by limiting the order of dependence in a sense to be defined in the next section. Note that for a time series, i.e., a collection of random variables indexed by the integers, the definition of order of dependence is based of Markov dependence. Since there is no natural ordering among 2-tuples of integers (s, t), the order of dependence in a two-way series can be defined in various ways. The particular definition we have adopted here is motivated by the following considerations, (i) in many physical situations x_{st} denotes some observation on the plane at the point (s, t), e.g., the illumination on a certain grid in a discretized black and white picture, where the dependence of $x_{s_1t_1}$ on $x_{s_2t_2}$ tends to be weak as (s_1, t_1) and (s_2, t_2) are far apart in the sense that $(s_1 - s_2)^2 + (t_1 - t_2)^2$ is large. This has determined the nature of Markov-type dependence in our definition. (ii) This definition lends itself to the methods involving Kronecker products of matrices and (iii) it generalizes (as indicated in Remark III of Section 7) to higher dimensions without much additional difficulty.

Received May 11, 1971; revised May 2, 1972.

Starting from some stationarity and symmetry conditions we then show that when the order of dependence q in a normally distributed two-way series $\{x_{st}, s = 1, \cdots, S, t = 1, \cdots, T\}$ is small in comparison with S and T, its joint pdf can be brought to a simple form by small adjustments. These adjustments involve only those random variables x_{st} for which at least one subscript is either too small or too large and results in a form of density function described in terms of relatively few parameters. Maximum likelihood estimates of these parameters when the order of dependence is known and likelihood ratio test for an unknown order of dependence are derived.

2. Order of dependence. For a two-way series $\{x_{st}\} = \{x_{st}, s = 1, \cdots, S, t = 1, \cdots, T\}$ we write

$$x_t' = (x_{1t}, \cdots, x_{St}), \qquad t = 1, \cdots, T, \tag{1}$$
$$x' = (x_{11}, \cdots, x_{S1}, \cdots, x_{1T}, \cdots, x_{ST}).$$

We assume that x follows an ST-dimensional nonsingular normal distribution. Let $u_{st} = x_{st} - E(x_{st} \mid x_{t-1}, \cdots, x_1)$, $t = 2, \cdots, T$ and by u_t denote the S-dimensional random column vector whose sth coordinate is u_{st}. The following are well-known facts in multivariate normal theory (see, e.g., Anderson (1958)).

LEMMA 1. *$E(x_{st} \mid x_{t-1}, \cdots, x_1)$ is a linear function of $\{x_{s't'}, s' = 1, \cdots, S, t' = 1, \cdots, t-1\}$ and u_t follows an S-dimensional normal distribution with mean vector 0 and covariance matrix Γ_t which does not depend on $x_1, \cdots, x_{t-1}$.*

We now define order of dependence as follows.

DEFINITION 1. The order of dependence in a normally distributed two-way series $\{x_{st}\}$ is defined to be the smallest $q \geqq 0$ so that

(a) For $t = 2, \cdots, T$, the conditional distribution of x_{st} given $x_{t-1}, \cdots, x_1$ depends only on the random variables $x_{s't'}$ for which $(s' - s)^2 + (t' - t)^2 \leqq q^2$, and

(b) For $t = 2, \cdots, T$ and $s = [q] + 1, \cdots, S$, the conditional distribution of u_{st} given $u_{s-1,t}, \cdots, u_{1t}$ depends only on $u_{s-1,t}, \cdots, u_{s-[q],t}$.

Clearly, the ordear of dependence q in a normally distributed two-way series can only be a number of the form $(h^2 + k^2)^{\frac{1}{2}}$ where h and k are integers, viz. $0, 1, 2^{\frac{1}{2}}, 2, 5^{\frac{1}{2}}$ etc.

REMARK 1. In Definition 1, condition (a) is on the nature of dependence between the columns x_t of the two-way series $\{x_{st}\}$ and condition (b) is on the nature of dependence within the columns after eliminating the effects of between column dependence.

REMARK 2. We may replace q by q_1 in condition (a) and by q_2 in condition (b) to obtain a generalization of Definition 1. All the results in this paper can be easily modified for this slightly more general definition of order of dependence. In fact, we can make condition (a) a little more flexible by allowing the conditional distribution of x_{st} given $x_{t-1}, \cdots, x_1$ to depend only on those $x_{s't'}$ for which

$s - a(t-t') \leqq s' \leqq s + a(t-t')$, $t' = t-1, \cdots, t-q_1$ where $a(1) \geqq a(2) \geqq \cdots \geqq a(q_1)$ are given numbers.

REMARK 3. Suppose the dependence in $\{x_{st}\}$ is of order q according to our definition. What can we say about the conditional distribution of x_{st} on $x_{t-1}, \cdots, x_1$ and $x_{s-1,t}, \cdots, x_{1t}$, i.e., the conditional distribution of x_{st} on the random variables preceding itself in the lexicographic ordering according to which the two-way series is arranged to form the vector x in (1)? Let $\xi_s, \eta_t, \varepsilon_{st}$, $s = 1, \cdots, S$, $t = 1, \cdots, T$ be independent normal random variables with mean 0, $E\xi_s^2 = \sigma_1^2 > E\eta_t^2 = \sigma_2^2 > E\varepsilon_{st}^2 = \sigma_3^2$, and let $\alpha = \{1 - \sigma_2^2/\sigma_1^2\}^{\frac{1}{2}}$, $\beta = \{1 - \sigma_3^2/\sigma_2^2\}^{\frac{1}{2}}$. Define $x_{s1} = \xi_s$, $s = 1, \cdots, S$, $x_{1t} = \alpha x_{1,t-1} + \eta_t$, $t = 2, \cdots, T$, and $x_{st} = \alpha x_{s,t-1} + \beta x_{s-1,t} + \varepsilon_{st}$, $s = 2, \cdots, S$, $t = 2, \cdots, T$. Then $\{x_{st}\}$ satisfies the conditions of Definition 1 with $q = 1$, but $E(x_{st} \mid x_{t-1}, \cdots, x_1, x_{s-1,t}, \cdots, x_{1t}) = \alpha x_{s,t-1} + \beta x_{s-1,t} - \alpha\beta x_{s-1,t-1}$. In fact, it can be shown in general, that if the dependence in $\{x_{st}\}$ is of order q according to our definition, then the conditional distribution of x_{st} given the preceding random variables in the lexicographic ordering depends only on $\{x_{s't'} \mid s - [q^2 - (t-t')^2] - [q] \leqq s' \leqq s + [q^2 - (t-t')^2]$, $t' = t-1, \cdots, t - [q]\}$ and on $x_{s-1,t}, \cdots, x_{s-[q],t}$. However, the converse is not true in general.

We shall restrict our attention to series which satisfy the following conditions.

CONDITION 1. $$E(x_{st}) = \xi \qquad \text{for all } s, t .$$

CONDITION 2.
$$\operatorname{Cov}(x_{s_1t_1}, x_{s_2t_2}) = \operatorname{Cov}(x_{s_1+h,t_1+k}, x_{s_2+h,t_2+k}) \qquad \text{for all } s_1, t_1, s_2, t_2, h, k.$$

CONDITION 3. $$\operatorname{Cov} x_{st}, x_{s+h,t+k}) = \operatorname{Cov}(x_{st}, x_{s-h,t+k}) \qquad \text{for all } s, t, h, k .$$

Conditions 1 and 2 are stationarity conditions and Condition 3 is a symmetry condition.

Let $y_{st} = x_{st} - \xi$. Then $\{y_{st}\}$ is a normally distributed two-way series satisfying Conditions 1—3 with $E(y_{st}) = 0$. Defining $y_1, \cdots, y_T$ and y from $\{y_{st}\}$ in the same way as $x_1, \cdots, x_T$ and x were defined from $\{x_{st}\}$ in (1), we see that $y_{st} - E(y_{st} \mid y_{t-1}, \cdots, y_1) = x_{st} - E(x_{st} \mid x_{t-1}, \cdots, x_1) = u_{st}$. Thus the order of dependence in $\{y_{st}\}$ is the same as the order of dependence in $\{x_{st}\}$. We shall, therefore, study $\{y_{st}\}$ for a while, remembering that

$$y = x - \xi\mathbf{1} \tag{2}$$

where $\mathbf{1}$ is the ST-dimensional column vector with 1 for each coordinate.

In view of Lemma 1, the following lemma is an immediate consequence of condition (a) in Definition 1 and the stationarity Conditions 1 and 2.

LEMMA 2. *If the dependence in $\{x_{st}\}$ is of order q and if $y_{st} = x_{st} - \xi$, then there exist $S \times S$ matrices $\beta_1, \cdots, \beta_{[q]}$ and Γ so that*

$$E(y_t \mid y_{t-1}, \cdots, y_1) = \textstyle\sum_{k=1}^{[q]} \beta_k y_{t-k}$$
$$E(\{y_t - \textstyle\sum_{k=1}^{[q]} \beta_k y_{t-k}\}\{y_t - \sum_{k=1}^{[q]} \beta_k y_{t-k}\}') = \Gamma .$$

In the next two lemmas we shall draw some conclusions about the structures of the matrices β_k and Γ. Before stating these lemmas let us introduce a few notations and a definition.

For $k = 0, 1, \cdots, [q]$, we denote by $\nu(q, k)$ the largest integer so that $k^2 + \nu^2(q, k) \leqq q^2$. Then $[q] = \nu(q, 0) \geqq \nu(q, 1) \geqq \cdots \geqq \nu(q, [q]) = 0$. If q is an integer then $\nu(q, 0) = q$ and $\nu(q, q) = 0$.

For $r = 1, \cdots, T$, let E_r denote the $T \times T$ matrix with $\frac{1}{2}$'s on the diagonals r elements above and r elements below the main diagonal, and 0's at all other places. For $r = 1, \cdots, S$, let F_r denote the the $S \times S$ matrix with the same property. We shall denote by E_0 the $T \times T$ identity matrix and by F_0 the $S \times S$ identity matrix whenever convenient, but when there is no danger of confusion, we shall use the symbol I for both the $S \times S$ and the $T \times T$ identity matrix.

DEFINITION 2. An $N \times N$ matrix is said to have property π_{m_1,m_2} (π'_{m_1,m_2}) for $m_1, m_2 = 1, \cdots, N$, if it has 0's everywhere except possibly in the first (last) m_2 columns of its first m_1 rows and in the last (first) m_2 columns of its last m_1 rows. Property $\pi_{m,m}$ ($\pi'_{m,m}$) will simply be mentioned as π_m (π_m').

LEMMA 3. *If the dependence in the series $\{x_{st}\}$ is of order q and if $2\nu(q, 1) < S$, then for the matrices $\beta_1, \cdots, \beta_{[q]}$ in Lemma 2, there exist constants $\{\theta_{kl}, k = 1, \cdots, [q], l = 0, 1, \cdots, \nu(q, k)\}$ so that*

$$\beta_k = \sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l + \rho_k$$

where ρ_k has property $\pi_{\nu(q,1),\nu(q,1)+\nu(q,k)}$.

PROOF. Since the order of dependence in $\{x_{st}\}$ is q, the entry on the sth row and s'th column of β_k, being the coefficient of $y_{s',t-k}$ in $E(y_{st} \mid y_{t-1}, \cdots, y_1)$, is 0 for $|s' - s| > \nu(q, k)$. It now follows by definition of the matrices F_l that the entry in the sth row and s'th column of ρ_k is 0 whenever $|s' - s| > \nu(q, k)$. We therefore have (a) the first $\nu(q, 1)$ rows of ρ_k can have nonzeros only in the first $\nu(q, 1) + \nu(q, k)$ columns, (b) the last $\nu(q, 1)$ rows of ρ_k can have nonzeros only in the last $\nu(q, 1) + \nu(q, k)$ columns, and (c) in all other rows of ρ_k all entries that are more than $\nu(q, k)$ elements to the right or to the left of the main diagonal are 0's. To complete the proof it will now be enough to show that except for the first $\nu(q, 1)$ and the last $\nu(q, 1)$ rows in β_k, all entries on the main diagonal are equal and all entries on the diagonals r elements to the right and to the left of the main diagonal are equal for $r = 1, \cdots, \nu(q, k)$. But this follows because for $s = \nu(q, 1) + 1, \cdots, S - \nu(q, 1)$ and $l = 0, 1, \cdots, \nu(q, k)$ the coefficients of $y_{s-l,t-k}$ and $y_{s+l,t-k}$ are equal by the symmetry Condition 3, and these coefficients are the same for all s by the stationarity Condition 2.

LEMMA 4. *If the dependence in the series $\{x_{st}\}$ is of order q and if $\nu(q, 1) + [q] < S$, then for matrix Γ in Lemma 2, there exist constants $\{\gamma_l, l = 0, 1, \cdots, [q]\}$ so that*

$$\Gamma^{-1} = \sum_{l=0}^{[q]} \gamma_l F_l + \rho_0$$

where ρ_0 has property $\pi_{\nu(q,1)+[q]}$.

PROOF. Consider $u_{1t}, \cdots, u_{St}$ for any fixed t between $[q]+1$ and T. By condition (a) in Definition 1,

$$\begin{aligned} u_{st} &= y_{st} - E(y_{st} \mid y_{t-1}, \cdots, y_1) \\ &= y_{st} - E(y_{st} \mid y_{s't'}, t' = t-1, \cdots, t-[q], \\ &\qquad\qquad s = s - \nu(q, t-t'), \cdots, s + \nu(q, t-t')) \end{aligned}$$

for $s = \nu(q, 1)+1, \cdots, S - \nu(q, 1)$. It therefore follows from the stationarity Condition 2 that $\{u_{st}, s = \nu(q, 1)+1, \cdots, S-\nu(q, 1)\}$ is a normally distributed stationary series with mean 0. Now let g and g^* denote the joint pdf of $u_{1t}, \cdots, u_{St}$ and of $u_{1t}, \cdots, u_{\nu(q,1)+[q],t}$ respectively and let $g_s(u_{st} \mid u_{s-1,t}, \cdots, u_{1t})$ denote the conditional pdf of u_{st} given $u_{s-1,t}, \cdots, u_{1t}$.

Then

$$\begin{aligned} g(u_{1t}, \cdots, u_{St}) &= \text{const.} \exp[-\tfrac{1}{2} u_t' \Gamma^{-1} u_t] \\ g^*(u_{1t}, \cdots, u_{\nu(q,1)+[q],t}) &= \text{const.} \exp[-\tfrac{1}{2}\psi_1(u_{1t}, \cdots, u_{\nu(q,1)+[q],t})] \end{aligned} \tag{3}$$

where ψ_1 is a quadratic form. Also, by condition (b) of Definition 1 and due to the stationarity of $\{u_{st}, s = \nu(q, 1)+1, \cdots, S-\nu(q, 1)\}$ already mentioned,

$$\begin{aligned} &\textstyle\prod_{s=\nu(q,1)+[q]+1}^{S} g_s(u_{st} \mid u_{s-1,t}, \cdots, u_{1t}) \\ &\quad = \text{const.} \exp[-(2\sigma^2)^{-1} \textstyle\sum_{s=\nu(q,1)+[q]+1}^{S-\nu(q,1)} (u_{st} - \alpha_1 u_{s-1,t} - \cdots \\ &\qquad - \alpha_{[q]} u_{s-[q],t})^2 - \tfrac{1}{2}\psi_2(u_{S-\nu(q,1)-[q]+1,t}, \cdots, u_{St})] , \end{aligned} \tag{4}$$

where ψ_2 is a quadratic form, $\sigma^2 = \text{Var}(u_{st})$ and α_l is the coefficient of $u_{s-l,t}$ in $E(u_{st} \mid u_{s-1,t}, \cdots, u_{s-[q],t})$. Now by the same argument that leads to formula (14) of Anderson (1962), we have

$$\begin{aligned} &\sigma^{-2} \textstyle\sum_{s=\nu(q,1)+[q]+1}^{S-\nu(q,1)} (u_{st} - \alpha_1 u_{s-1,t} - \cdots - \alpha_{[q]} u_{s-[q],t})^2 \\ &\quad = \textstyle\sum_{l=0}^{[q]} \gamma_l u_t' F_l u_t + \psi_3(u_{\nu(q,1)+1,t}, \cdots, u_{\nu(q,1)+[q],t}) \\ &\qquad + \psi_4(u_{S-\nu(q,1)-[q]+1,t}, \cdots, u_{S-\nu(q,1),t}) , \end{aligned} \tag{5}$$

where $\gamma_0, \gamma_1, \cdots, \gamma_l$ depend on $\sigma^2, \alpha_1, \cdots, \alpha_{[q]}$, and ψ_3 and ψ_4 are quadratic forms. Using (3), (4) and (5) and the fact that

$$g(u_{1t}, \cdots, u_{St}) = g^*(u_{1t}, \cdots, u_{\nu(q,1)+[q],t}) \cdot \textstyle\prod_{s=\nu(q,1)+[q]+1}^{S} g_s(u_{st} \mid u_{s-1,t}, \cdots, u_{1t})$$

we have,

$$\begin{aligned} u_t' \rho_0 u_t &= u_t' \{\Gamma^{-1} - \textstyle\sum_{l=0}^{[q]} \gamma_l F_l\} u_t \\ &= \psi_1(u_{1t}, \cdots, u_{\nu(q,1)+[q],t}) + \psi_2(u_{S-\nu(q,1)-[q]+1,t}, \cdots, u_{St}) \\ &\quad + \psi_3(u_{\nu(q,1)+1,t}, \cdots, u_{\nu(q,1)+[q],t}) \\ &\quad + \psi_4(u_{S-\nu(q,1)-[q]+1,t}, \cdots, u_{S-\nu(q,1),t}) . \end{aligned}$$

Since $\psi_1 + \psi_3$ is a quadratic form involving only the first $\nu(q, 1)+[q]$ of the random variables $u_{1t}, \cdots, u_{St}$ and $\psi_2 + \psi_4$ is a quadratic form involving only the last $\nu(q, 1)+[q]$ of the random variables $u_{1t}, \cdots, u_{St}$, the matrix ρ_0 has property $\pi_{\nu(q,1)+[q]}$ as was to be proved.

The coefficient θ_{kl} in Lemma 3 is the coefficient of $\frac{1}{2}(y_{s-l,t-k} + y_{s+l,t-k})$ in the regression of y_{st} on $\{y_{s't'}, (s'-s)^2 + (t'-t)^2 \leqq q^2$ and $t' < t\}$ and the coefficients $\gamma_0, \gamma_1, \cdots, \gamma_{[q]}$ of Lemma 4 are related to $\sigma^2, \alpha_1, \cdots, \alpha_{[q]}$ in a way analogous to relations (18) of Anderson (1962).

For convenience, we set $\theta_{00} = 1$, $\theta_{01} = \cdots = \theta_{0,\nu(q,0)} = 0$, and $\beta_0 = I$. Then $\beta_0 = \sum_{l=0}^{\nu(q,0)} \theta_{kl} F_l$ so that we can extend the structural formula for β_k given in Lemma 3 to $k = 0$.

We shall now examine the form of the joint pdf of the series $\{y_{st}\}$ when the order of dependence in the series is q.

It follows from the stationarity Condition 2 that $E(y_t y'_{t+h})$ is the same for all t and depends only on h. We therefore write $E(y_t y'_{t+h}) = \Sigma_h$. By the symmetry Condition 3, $\Sigma_{-h} = \Sigma_h' = \Sigma_h$. Let us denote by $y_{(q)}$ the $[q]S$-dimensional column vector consisting of the first $[q]S$ coordinates of y and let $\Sigma = E(yy')$ and $\Sigma_{(q)} = E(y_{(q)} y'_{(q)})$. We can then express Σ and $\Sigma_{(q)}$ as partitioned matrices involving Σ_h, $h = 0, 1, \cdots, T-1$. The (i, i)th element of $\Sigma_{(q)}$ is $\Sigma_{|i-j|}$, $i, j = 0, 1, \cdots, [q]-1$, and $\Sigma = \Sigma_{(T)}$. If we now denote by $f_{(q)}(y_{(q)})$ the joint pdf of $y_{(q)}$ and by $f_t(y_t \mid y_{t-1}, \cdots, y_1)$ the conditional joint pdf of y_t given $y_{t-1}, \cdots, y_1$, then the joint pdf $f(y)$ of y is,

$$f(y) = f_{(q)}(y_{(q)}) \prod_{t=[q]+1}^{T} f_t(y_t \mid y_{t-1}, \cdots, y_1)\,. \tag{6}$$

Here $f_{(q)}(y_{(q)})$ is the $[q]S$-dimensional normal pdf with mean vector 0 and covariance matrix $\Sigma_{(q)}$ and when the order of dependence in the series is q, it follows from Lemmas 1 and 2 that $f_t(y_t \mid y_{t-1}, \cdots, y_1)$ is the S-dimensional normal pdf with mean vector $\sum_{k=1}^{[q]} \beta_k y_{t-k}$ and covariance matrix Γ. Incorporating these facts in (6), we have

$$\begin{aligned} f(y) = (2\pi)^{-\frac{1}{2}ST}|\Sigma_{(q)}|^{-\frac{1}{2}}|\Gamma|^{-\frac{1}{2}(T-[q])} \cdot \exp[-\tfrac{1}{2}y'_{(q)}\Sigma_{(q)}^{-1} y_{(q)} \\ - \tfrac{1}{2} \textstyle\sum_{t=[q]+1}^{T} (y_t - \sum_{k=1}^{[q]} \beta_k y_{t-k})'\Gamma^{-1}(y_t - \sum_{k=1}^{[q]} \beta_k y_{t-k})]\,. \end{aligned} \tag{7}$$

Now the exponent of (7) is easily seen to be $-\frac{1}{2}$ times

$$\begin{aligned} &\textstyle\sum_{t=1}^{T} y_t'(\sum_{k=0}^{[q]} \beta_k'\Gamma^{-1}\beta_k)y_t \\ &\quad + 2 \textstyle\sum_{t=[q]+1}^{T-[q]} \sum_{j=1}^{[q]} y_t'(\sum_{k=0}^{[q]-j} \beta_k'\Gamma^{-1}\beta_{k+j} - 2\Gamma^{-1}\beta_j)y_{t-j} \\ &\qquad + \psi_5(y_1, \cdots, y_{[q]}) + \psi_6(y_{T-[q]+1}, \cdots, y_T)\,, \end{aligned} \tag{8}$$

where ψ_5 and ψ_6 are quadratic forms. So far we have proceeded exactly as Anderson (1962), but at this point we shall find it profitable to express the sums in (8) as quadratic forms in y. This is done by writing the sums in (8) in terms of Kronecker products of matrices. We now have the following lemma which summarizes the development in this section.

LEMMA 5. *If the dependence in series $\{x_{st}\}$ is of order q and if $y_{st} = x_{st} - E(x_{st})$, then the joint* pdf *of y is of the form*

$$\begin{aligned} f(y) = \text{const.} \exp[-\tfrac{1}{2}y'\{E_0 \otimes \textstyle\sum_{k=0}^{[q]} \beta_k'\Gamma^{-1}\beta_k \\ + 2 \textstyle\sum_{j=1}^{[q]} E_j \otimes (\sum_{k=0}^{[q]-j} \beta_k'\Gamma^{-1}\beta_{k+j} - 2\Gamma^{-1}\beta_j)\}y + \psi_5 + \psi_6] \end{aligned}$$

where the structures of β_k *and* Γ^{-1} *are as given in Lemmas* 3 *and* 4, *and* ψ_5 *is a quadratic form in* $y_1, \cdots, y_{[q]}$ *and* ψ_6 *is a quadratic form in* $y_{T-[q]+1}, \cdots, y_T$.

3. A modified density function. The form of the density function f arrived at in Lemma 5 is such that relatively few parameters θ_{kl} and γ_l determine the interdependence between most of the random variables in the collection $\{y_{st}\}$ whereas the rest of the parameters (which are not mentioned explicitly) are required only to complete the description of the relatively few remaining random variables. To see this, we note that

(a) If U is an $S \times S$ matrix with property π_u and V is an $S \times S$ matrix with property π_v, then UV has property $\pi_{\max(u,v)}$, and

(b) If U is an $S \times S$ matrix with property π_u, then both $U(\sum_{l=0}^{[q]} \gamma_l F_l)$ and $(\sum_{l=0}^{[q]} \gamma_l F_l)U$ have property $\pi_{u+[q]+1}$.

Using these two facts we now conclude from Lemmas 3 and 4 that

$$\sum_{k=0}^{[q]} \beta_k'\Gamma^{-1}\beta_k = \sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l) + \rho_0^*$$

and for $j = 1, \cdots, [q]$,

$$\begin{aligned}\sum_{k=0}^{[q]} \beta_k'\Gamma^{-1}\beta_{k+j} - 2\Gamma^{-1}\beta_j &= \sum_{k=0}^{[q]-j} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} F_l) \\ &\quad - 2(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,j)} \theta_{jl} F_l) + \rho_j^*\end{aligned}$$

where the matrices $\rho_0^*, \rho_1^*, \cdots, \rho_{[q]}^*$ all have property $\pi_{3[q]+1}$. If we now examine the quadratic form

$$y'(E_0 \otimes \rho_0^* + 2 \sum_{j=1}^{[q]} E_j \otimes \rho_j^*)y ,$$

we notice that it involves only those terms $y_{st}\, y_{s't'}$ for which s and s' are either both among the first $3[q]+1$ or both among the last $3[q]+1$ of the integers $1, \cdots, S$. Incorporating these facts in Lemma 5, we have the following lemma which justifies the remark made at the beginning of this section.

LEMMA 6. *If the dependence in the series* $\{x_{st}\}$ *is of order* q *and if* $y_{st} = x_{st} - E(x_{st})$, *then the joint* pdf *of* y *is of the form*

$$\begin{aligned}f(y) = \text{const.} \exp[-\tfrac{1}{2}y'\{E_0 \otimes \textstyle\sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l) \\ + 2 \textstyle\sum_{j=1}^{[q]} E_j \otimes \sum_{k=0}^{[q]-j} ((\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} F_l) \\ - 2(\textstyle\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,j)} \theta_{jl} F_l))\}y + \psi_7(y)]\end{aligned} \tag{9}$$

where $\psi_7(y)$ *is a quadratic form involving only those* y_{st} *for which at least one of the subscripts is either among the first* $3[q]+1$ *or among the last* $3[q]+1$ *integers in its domain.*

When S and T are large compared to q, the quadratic form

$$\begin{aligned}y'\{E_0 \otimes \textstyle\sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l) \\ + 2 \textstyle\sum_{j=1}^{[q]} E_j \otimes {}_{k=0}^{[q]-j} ((\sum_{l=0}^{\nu(q,k)} \theta_{kl} F_l)'(\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} F_l) \\ - 2(\textstyle\sum_{l=0}^{[q]} \gamma_l F_l)(\sum_{l=0}^{\nu(q,j)} \theta_{jl} F_l))\}y\end{aligned} \tag{10}$$

will tend to dominate $\psi_7(y)$ which suggestes a modification of $f(y)$ by dropping $\psi_7(y)$ from the exponent of (9). The probability model that arises in this way involves only a small number of parameters but the manner in which they enter the form of the density function still makes it unsuitable for analysis. We shall now investigate another modification of $f(y)$ by replacing the matrices $E_1, \cdots, E_{[q]}$ by $A_1, \cdots, A_{[q]}$ where the matrices $A_1, \cdots, A_{[q]}$ are nonsingular, all having the same eigenvectors and $E_r - A_r$ having property π_r or π_r' (see Definition 2), and, similarly replacing the matrices $F_1, \cdots, F_{[q]}$ by $B_1, \cdots, B_{[q]}$ where the matrices $B_1, \cdots, B_{[q]}$ are nonsingular, all having the same eigenvectors and $F_r - B_r$ having property π_r or π_r'. Several such systems of matrices are known. One such system (in which $E_r - A_r$ and $F_r - B_r$ have property π_r') is obtained in the following way. Let C_m be the $m \times m$ circulant whose ith row, for $i = 1, \cdots, m - 1$, is the $(i + 1)$th row of the $m \times m$ identity matrix and whose mth row is the 1st row of the $m \times m$ identity matrix. Then for $q < \frac{1}{2}\min(S, T)$, $A_r = \frac{1}{2}[C_T{}^r + (C_T{}')^r]$, $r = 1, \cdots, [q]$ and $B_r = \frac{1}{2}[C_S{}^r + (C_S{}')^r]$, $r = 1, \cdots, [q]$ have the desired properties. These systems of matrices not only have common eigenvectors, but their eigenvectors and eigenvalues are also expressed by very simple formulas. For more information on these systems we refer to Anderson (1962).

We now note that the quadratic form obtained by replacing the matrices E_r by A_r and the matrices F_r by B_r in (10) differs from (10) by a quadratic form $\psi_8(y)$ which like $\psi_7(y)$ involves only those y_{st} for which at least one of the subscripts is either among the first $3[q] + 1$ or among the last $3[q] + 1$ integers in its domain. We thus arrive at the following theorem.

THEOREM 1. *If the dependence in the series* $\{x_{st}\}$ *is of order* q, *then the joint* pdf *of* x *is of the form*

$$g(x) = \text{const.} \exp[-\tfrac{1}{2}(x - \xi\mathbf{1})'R(\theta, \gamma)(x - \xi\mathbf{1}) + \psi(x)] \qquad \textit{where}$$

$$\begin{aligned} R(\theta, \gamma) = A_0 \otimes \textstyle\sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} B_l)'(\sum_{l=0}^{[q]} \gamma_l B_l)(_{l=0}^{\nu(q,k)} \theta_{kl} B_l) \\ + 2 \textstyle\sum_{j=1}^{[q]} A_j \otimes \sum_{k=0}^{[q]-j} \{(\sum_{l=0}^{\nu(q,k)} \theta_{kl} B_l)' \\ \times (\textstyle\sum_{l=0}^{[q]} \gamma_l B_l)(\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} B_l) \\ - 2(\textstyle\sum_{l=0}^{[q]} \gamma_l B_l)(\sum_{l=0}^{\nu(q,j)} \theta_{jl} B_l)\} \end{aligned} \tag{11}$$

with $A_0 = E_0 = I$, $B_0 = F_0 = I$, $E_r - A_r$ *and* $F_r - B_r$ *having property* π_r *or* π_r', $r = 1, \cdots, [q]$, *and* ψ *is a quadratic form involving only those* x_{st} *for which at least one subscript is either among the first* $3[q] + 1$ *or among the last* $3[q] + 1$ *integers in its domain. Furthermore, the matrices* $A_1, \cdots, A_{[q]}$ *can be so chosen as to have the same eigenvectors and the matrices* $B_1, \cdots, B_{[q]}$ *can be so chosen as to have the same eigenvectors.*

In view of Theorem 1, we propose to examine the interdependence in a normally distributed two-way series $\{x_{st}, s = 1, \cdots, S, t = 1, \cdots, T\}$ with S and T large in comparison with its order of dependence q, under the model

$$p(x \mid \xi, \theta, \gamma) = (2\pi)^{-\frac{1}{2}ST}\{\det R(\theta, \gamma)\}^{\frac{1}{2}} \exp[-\tfrac{1}{2}(x - \xi\mathbf{1})'R(\theta, \gamma)(x - \xi\mathbf{1})], \tag{12}$$

where $\theta = \theta_{kl}, k = 1, \cdots, [q], l = 0, 1, \cdots, \nu(q, k)\}$ and $\gamma = \{\gamma_l, l = 0, 1, \cdots, [q]\}$ are such that the matrix $R(\theta, \gamma)$ given by (11) is positive definite with $\theta_{00} = 1$, $\theta_{01} = \cdots \theta_{0,\nu(q,0)} = 0$.

The remainder of this paper will be devoted to the derivation of the maximum likelihood estimates of ξ, θ and γ on the basis of independent realizations $x_i = \{x_{sti}\}$ of the process $\{x_{st}\}$ when the order of dependence q is known, and to the derivation of the likelihood ratio test of the null hypothesis H_0: order of dependence is q_0 against the alternative hypotheses H_1: order of dependence is q_1, where $q_0 < q_1$ are two numbers of the form $(h^2 + k^2)^{\frac{1}{2}}$, h, k integers. To facilitate these derivations, we shall first transform the observed random variables to independent random variables. This is done in the next section by taking advantage of the fact that the A_r matrices have the same eigenvectors and the B_r matrices have the same eigenvectors.

4. Transformation of $\{x_{st}\}$ to independent random variables. Let $a_1, \cdots, a_T$ denote the normalized eigenvectors of each of the matrices $A_1, \cdots, A_{[q]}$ and let λ_{jt} denote the eigenvalue of A_j corresponding to the eigenvector a_t. Similarly, let $b_1, \cdots, b_S$ denote the normalized eigenvectors of each of the matrices $B_1, \cdots, B_{[q]}$ and let μ_{js} denote the eigenvalue of B_j corresponding to the eigenvector b_s. Define matrices P with columns $a_1, \cdots, a_T$ and Q with columns $b_1, \cdots, b_S$, i.e., $P = (a_1, \cdots, a_T)$, $Q = (b_1, \cdots, b_S)$, and let $\Lambda_j = \text{diag}(\lambda_{jt})$, $j = 0, 1, \cdots, [q]$, $M_j = \text{diag}(\mu_{js})$, $j = 0, 1, \cdots, [q]$ where $\lambda_{0t} = \mu_{0s} = 1$ for convenience.

The following properties of these matrices are easy to verify and we omit their proofs. Note that in all the expressions of Lemma 7, the right-hand side is a diagonal matrix.

LEMMA 7. (a) *P and Q are orthonormal, i.e., $P'P = I$, $Q'Q = I$.*

(b) *For any real $c_0, c_1, \cdots, c_{[q]}$, $P'(\sum_{j=0}^{[q]} c_j A_j)P = \sum_{j=0}^{[q]} c_j \Lambda_j$, and*

$$Q'(\textstyle\sum_{j=0}^{[q]} c_j B_j)Q = \sum_{j=0}^{[q]} c_j M_j \,.$$

(c) *For any j, k, A_j and A_k commute and B_j and B_k commute.*

(d) *For any j, k, $P'(A_j A_k)P = \Lambda_j \Lambda_k$, $Q'(B_j B_k)Q = M_j M_k$.*

(e) *$P \otimes Q$ is orthonormal.*

(f) *For any j, k, $(P \otimes Q)'(A_j \otimes B_k)(P \otimes Q) = (P'A_j P) \otimes (Q'B_k Q) = \Lambda_j \otimes M_k$.*

Using these properties of P and Q, we note that the matrix $R(\theta, \gamma)$ given by (11) is diagonalized when pre-multiplied by $(P \otimes Q)'$ and post-multiplied by $P \otimes Q$, and the determinant of $R(\theta, \gamma)$ is also easily obtained. This is shown in the following lemma.

LEMMA 8. *$(P \otimes Q)'R(\theta, \gamma)(P \otimes Q)$ is diagonal, and*

$$\det R(\theta, \gamma) = \textstyle\prod_{s=1}^{S} \prod_{t=1}^{T} H_{st}(\theta)\{\prod_{s=1}^{S} K_s(\gamma)\}^T \quad \textit{where} \tag{13}$$

$$H_{st}(\theta) = \lambda_{0t} \textstyle\sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} \mu_{ls})^2 + 2 \sum_{j=1}^{[q]} \lambda_{jt}\{\sum_{k=0}^{[q]-j} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} \mu_{ls}) \times (\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} \mu_{ls}) - 2 \sum_{l=0}^{\nu(q,j)} \theta_{jl} \mu_{ls}\} \tag{14}$$

and

$$K_s(\gamma) = \sum_{l=0}^{[q]} \gamma_l \mu_{ls} \,. \tag{15}$$

PROOF. We first note that since the matrices B_j are symmetric and since their linear combinations commute, $R(\theta, \gamma)$ can be written as

$$\begin{aligned} R(\theta, \gamma) = {} & A_0 \otimes [(\textstyle\sum_{l=0}^{[q]} \gamma_l B_l) \sum_{k=0}^{[q]} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} B_l)^2] \\ & + 2 \textstyle\sum_{j=1}^{[q]} A_j \otimes [(\sum_{l=0}^{[q]} \gamma_l B_l)\{\sum_{k=0}^{[q]-j} (\sum_{l=0}^{\nu(q,k)} \theta_{kl} B_l) \\ & \times (\textstyle\sum_{l=0}^{\nu(q,k+j)} \theta_{k+j,l} B_l) - 2 \sum_{l=0}^{\nu(q,j)} \theta_{jl} B_l\}] \,. \end{aligned} \tag{16}$$

Using the properties of the matrices P and Q given in Lemma 7 we now see that $D(\theta, \gamma) = (P \otimes Q)' R(\theta, \gamma)(P \otimes Q)$ is obtained by replacing the matrices A_j by Λ_j and the matrices B_l by M_l on the right-hand side of (16). D is diagonal because Λ_j and M_l are diagonal. It is easy to see that sth element in the tth block of D is $d_{st}(\theta, \gamma) = H_{st}(\theta) K_s(\gamma)$ where H_{st} and K_s are as given in (14) and (15). Furthermore, since $P \otimes Q$ is orthonormal,

$$\det R(\theta, \gamma) = \det D(\theta, \gamma) = \textstyle\prod_{s=1}^{S} \prod_{t=1}^{T} d_{st}(\theta, \gamma)$$

from which (13) follows.

THEOREM 2. *If x is a normally distributed two-way series with joint* pdf $p(x \mid \xi, \theta, \gamma)$ *given by* (12) *and if $z = (P \otimes Q)'x$, then the joint* pdf *of z is given by*

$$\begin{aligned} p^*(z \mid \xi, \theta, \gamma) = {} & [(2\pi)^{-ST} \textstyle\prod_{s=1}^{S} \prod_{t=1}^{T} H_{st}(\theta)\{\prod_{s=1}^{S} K_s(\gamma)\}^T]^{\frac{1}{2}} \\ & \times \exp[-\tfrac{1}{2} \textstyle\sum_{s=1}^{S} K_s(\gamma) \sum_{t=1}^{T} H_{st}(\theta)(z_{st} - c_{st}\xi)^2] \end{aligned} \tag{17}$$

where z_{st} and c_{st} are the $(t-1)S + s$th coordinates of z and $(P \otimes Q)'\mathbf{1}$ respectively.

PROOF. The theorem follows immediately from Lemma 8.

Note. If S and T are even and if A_j and B_j are obtained from the S- and T-dimensional circulants C_S and C_T in the way mentioned in Section 3, then the common normalized eigenvectors of A_j are

$$\begin{aligned} a_t' &= (2/T)^{\frac{1}{2}}(\cos 2\pi t/T, \cos 4\pi t/T, \cdots, \cos (T-1)2\pi t/T, \cos 2\pi t) \,, \\ &\qquad\qquad t = 1, \cdots, T/2 - 1 \\ &= (2/T)^{\frac{1}{2}}(\sin 2\pi t/T, \sin 4\pi t/T, \cdots, \sin (T-1)2\pi t/T, \sin 2\pi t) \,, \\ &\qquad\qquad t = T/2 + 1, \cdots, T - 1 \\ a'_{T/2} &= T^{-\frac{1}{2}}(-1, 1, \cdots, -1, 1) \,, \qquad a_T' = T^{-\frac{1}{2}}(1, 1, \cdots, 1, 1) \end{aligned}$$

and the common normalized eigenvectors of B_j are

$$\begin{aligned} b_s' &= (2/S)^{\frac{1}{2}}(\cos 2\pi s/S, \cos 4\pi s/S, \cdots, \cos (S-1)2\pi s/S, \cos 2\pi s) \,, \\ &\qquad\qquad s = 1, \cdots, S/2 - 1 \\ &= (2/S)^{\frac{1}{2}}(\sin 2\pi s/S, \sin 4\pi s/S, \cdots, \sin (S-1)2\pi s/S, \sin 2\pi s) \,, \\ &\qquad\qquad s = S/2 + 1, \cdots, S - 1 \\ b'_{S/2} &= S^{-\frac{1}{2}}(-1. 1, \cdots, -1, 1) \,, \qquad b_S' = S^{-\frac{1}{2}}(1, 1, \cdots, 1, 1) \,. \end{aligned}$$

Hence the Tth row sum of P' is $T^{\frac{1}{2}}$, the Sth row sum of Q' is $S^{\frac{1}{2}}$ and the sums of all other rows of P' and Q' are 0. Since c_{st} is the product of the sth row sum of Q' and the tth row sum of P', it follows that for the above choice of the matrices A_j and B_j, $c_{ST} = (ST)^{\frac{1}{2}}$ and for all other (s, t), $c_{st} = 0$.

5. Maximum likelihood estimation of the ξ, θ, γ. Let $\{x_{sti}, s = 1, \cdots, S, t = 1, \cdots, T\}$, $i = 1, \cdots, N$ be N independent realizations of a normally distributed two-way series whose joint pdf $p(x \mid \xi, \theta, \gamma)$ is given by (12). As in (1), let x_i denote the ST-dimensional column vector whose $(t-1)S + s$th coordinate is x_{sti}. If we now transform $z_i = (P \otimes Q)'x_i$, $i = 1, \cdots, N$, and let z_{sti} denote the $(t-1)S + s$th coordinate of z_i, then by Theorem 2, the log likelihood function of ξ, θ, γ given $z_1, \cdots, z_N$ is found to be

$$\begin{aligned} L(\xi, \theta, \gamma) &= \textstyle\sum_{i=1}^{N} \log p^*(z_i \mid \xi, \theta, \gamma) \\ &= -\tfrac{1}{2}NST \log(2\pi) + \tfrac{1}{2}N \textstyle\sum_{s=1}^{S} \sum_{t=1}^{T} \log H_{st}(\theta) \\ &\quad + \tfrac{1}{2}NT \textstyle\sum_{s=1}^{S} \log K_s(\gamma) \\ &\quad - \tfrac{1}{2} \textstyle\sum_{s=1}^{S} K_s(\gamma) \sum_{t=1}^{T} H_{st}(\theta) \sum_{i=1}^{N} (z_{sti} - c_{st}\xi)^2 . \end{aligned} \tag{18}$$

Let us reindex the parameters $\{\theta_{kl}\}$ as $\theta_{10} = \theta_1, \cdots, \theta_{1,\nu(q,1)} = \theta_{\nu(q,1)+1}$, $\theta_{20} = \theta_{\nu(q,1)+2}, \cdots, \theta_{[q],0} = \theta_{\alpha(q)}$ where

$$\alpha(q) = \textstyle\sum_{k=1}^{[q]} \nu(q, k) + [q] .$$

Then the likelihood equations become

$$\frac{1}{N} \cdot \frac{\partial L}{\partial \xi} = \textstyle\sum_{s=1}^{S} K_s(\gamma) \sum_{t=1}^{T} H_{st}(\theta) c_{st}(\bar{z}_{st} - c_{st}\xi) = 0$$

$$\begin{aligned} \frac{1}{N} \cdot \frac{\partial L}{\partial \theta_j} &= \tfrac{1}{2} \textstyle\sum_{s=1}^{S} \sum_{t=1}^{T} \{H_{st}(\theta)\}^{-1} \dfrac{\partial H_{st}(\theta)}{\partial \theta_j} \\ &\quad - \tfrac{1}{2} \textstyle\sum_{s=1}^{S} K_s(\gamma) \sum_{t=1}^{T} \dfrac{\partial H_{st}(\theta)}{\partial \theta_j} V_{st}(\xi) = 0 , \qquad j = 1, \cdots, \alpha(q) \end{aligned}$$

$$\begin{aligned} \frac{1}{N} \cdot \frac{\partial L}{\partial \gamma_j} &= \tfrac{1}{2}T \textstyle\sum_{s=1}^{S} \{K_s(\gamma)\}^{-1} \dfrac{\partial K_s(\gamma)}{\partial \gamma_j} - \tfrac{1}{2} \textstyle\sum_{s=1}^{S} \dfrac{\partial K_s(\gamma)}{\partial \gamma_j} \sum_{t=1}^{T} H_{st}(\theta) V_{st}(\xi) \\ &= 0 , \qquad\qquad j = 0, 1, \cdots, [q] \end{aligned}$$

where

$$\bar{z}_{st} = \textstyle\sum_{i=1}^{N} z_{sti}/N \qquad \text{and} \qquad V_{st}(\xi) = \textstyle\sum_{i=1}^{N} (z_{sti} - c_{st}\xi)^2/N .$$

These equations cannot be solved exactly. We therefore use an iterative method commonly known as the "method of scoring" (see, e.g., Rao (1965)). In order to apply this method, let us examine the first partial derivatives with respect to ξ, θ_j and γ_j of the log likelihood function given a single realization z_i. Since the z_i consists of independent normally distributed random variables $\{z_{sti}, s = 1, \cdots, S, t = 1, \cdots, T\}$ with

$$E(z_{sti}) = c_{st}\xi \qquad \text{and} \qquad \operatorname{Var}(z_{sti}) = 1/K_s(\gamma)H_{st}(\theta) ,$$

let us first compute

$$\frac{\partial \log p^*_{st}(z_{sti} \mid \xi, \theta, \gamma)}{\partial \xi} = K_s(\gamma) H_{st}(\theta) c_{st}(z_{sti} - c_{st}\xi)$$

$$\frac{\partial \log p^*_{st}(z_{sti} \mid \xi, \theta, \gamma)}{\partial \theta_j} = \frac{1}{2} \frac{\partial H_{st}(\theta)}{\partial \theta_j} \{H_{st}(\theta)\}^{-1} - \tfrac{1}{2} K_s(\gamma) \frac{\partial H_{st}(\theta)}{\partial \theta_j} (z_{sti} - c_{st}\xi)^2$$

$$\frac{\partial \log p^*_{st}(z_{sti} \mid \xi, \theta, \gamma)}{\partial \gamma_j} = \frac{1}{2} \frac{\partial K_s(\gamma)}{\partial \gamma_j} \{K_s(\gamma)\}^{-1} - \frac{1}{2} \frac{\partial K_s(\gamma)}{\partial \gamma_j} H_{st}(\theta)(z_{sti} - c_{st}\xi)^2 .$$

Hence

$$J_{st}(\xi, \xi) = E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \xi}\right)^2\right] = c^2_{st} H_{st}(\theta) K_s(\gamma)$$

$$J_{st}(\xi, \theta_j) = E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \xi}\right)\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_j}\right)\right] = 0$$

$$J_{st}(\xi, \gamma_j) = E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \xi}\right)\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_j}\right)\right] = 0$$

$$\begin{aligned} J_{st}(\theta_j, \theta_k) &= E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_j}\right)\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_k}\right)\right] \\ &= \frac{1}{2}\left\{\frac{1}{H_{st}(\theta)} \cdot \frac{\partial H_{st}(\theta)}{\partial \theta_j}\right\}\left\{\frac{1}{H_{st}(\theta)} \cdot \frac{\partial H_{st}(\theta)}{\partial \theta_k}\right\} \end{aligned}$$

$$\begin{aligned} J_{st}(\theta_j, \gamma_k) &= E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_j}\right)\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_k}\right)\right] \\ &= \frac{1}{2}\left\{\frac{1}{H_{st}(\theta)} \cdot \frac{\partial H_{st}(\theta)}{\partial \theta_j}\right\}\left\{\frac{1}{K_s(\gamma)} \cdot \frac{\partial K_s(\gamma)}{\partial \gamma_k}\right\} \end{aligned}$$

$$\begin{aligned} J_{st}(\gamma_j, \gamma_k) &= E_{\xi,\theta,\gamma}\left[\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_j}\right)\left(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_k}\right)\right] \\ &= \frac{1}{2}\left\{\frac{1}{K_s(\gamma)} \cdot \frac{\partial K_s(\gamma)}{\partial \gamma_j}\right\}\left\{\frac{1}{K_s(\gamma)} \cdot \frac{\partial K_s(\gamma)}{\partial \gamma_k}\right\} . \end{aligned}$$

Now let $J_{st}(\theta, \theta)$ be the $\alpha(q) \times \alpha(q)$ matrix with $J_{st}(\theta_j, \theta_k)$ in the jth row and kth column, $J_{st}(\theta, \gamma)$ the $\alpha(q) \times \{[q] + 1\}$ matrix with $J_{st}(\theta_j, \gamma_k)$ in the jth row and kth column and $J_{st}(\gamma, \gamma)$ the $\{[q] + 1\} \times \{[q] + 1\}$ matrix with $J_{st}(\gamma_j, \gamma_k)$ in the jth row and kth column. Finally, let $J_{st}(\xi, \theta, \gamma)$ denote the partitioned matrix

$$J_{st}(\xi, \theta, \gamma) = \begin{bmatrix} J_{st}(\xi, \xi) & 0 & 0 \\ 0 & J_{st}(\theta, \theta) & J_{st}(\theta, \gamma) \\ 0 & J_{st}(\theta, \gamma)' & J_{st}(\gamma, \gamma) \end{bmatrix}$$

and let

$$J(\xi, \theta, \gamma) = \textstyle\sum_{s=1}^{S} \sum_{t=1}^{T} J_{st}(\xi, \theta, \gamma) .$$

Then $J(\xi, \theta, \gamma)$ is the information matrix from one realization z_i. In other words, letting $\varphi_{sti}(\xi, \theta, \gamma)'$ denote the $1 + \alpha(q) + \{[q] + 1)\}$-dimensional row vector

$$\varphi_{sti}(\xi, \theta, \gamma)' = \Big(\frac{\partial \log p^*_{st}(z_{sti})}{\partial \xi}, \frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_1}, \cdots, \frac{\partial \log p^*_{st}(z_{sti})}{\partial \theta_{\alpha(q)}}, \frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_0}, \cdots, \frac{\partial \log p^*_{st}(z_{sti})}{\partial \gamma_{[q]}}\Big)$$

and $\varphi_i(\xi, \theta, \gamma) = \sum_{s=1}^S \sum_{t=1}^T \varphi_{sti}(\xi, \theta, \gamma)$, we can easily see that

$$J_{st}(\xi, \theta, \gamma) = E_{\xi,\theta,\gamma}[\varphi_{sti}(\xi, \theta, \gamma)\varphi_{sti}(\xi, \theta, \gamma)']$$

and

$$J(\xi, \theta, \gamma) = E_{\xi,\theta,\gamma}[\varphi_i(\xi, \theta, \gamma)\varphi_i(\xi, \theta, \gamma)'] .$$

Now if $\tilde{\xi}, \tilde{\theta}, \tilde{\gamma}$ are first approximations to the maximum likelihood estimates, then the method of scoring gives the next approximations by the formula

$$(\hat{\xi}, \hat{\theta}, \hat{\gamma}) = (\tilde{\xi}, \tilde{\theta}, \tilde{\gamma}) + \tilde{\varphi}'\tilde{J}^{-1} \tag{19}$$

where $\tilde{J} = J(\tilde{\xi}, \tilde{\theta}, \tilde{\gamma})$ and $\tilde{\varphi}_i = \sum_{i=1}^N \varphi_i(\tilde{\xi}, \tilde{\theta}, \tilde{\gamma})/N$. From the structure of J it is clear that the adjustment of $\tilde{\xi}$ to $\hat{\xi}$ can be carried out separately from the adjustment of the estimates of the other parameters. Furthermore, if the first approximations $\tilde{\xi}, \tilde{\theta}, \tilde{\gamma}$ are such that $N^{\frac{1}{2}}(\tilde{\xi} - \xi)$, $N^{\frac{1}{2}}(\tilde{\theta} - \theta)$ and $N^{\frac{1}{2}}(\tilde{\gamma} - \gamma)$ are $O_p(1)$, i.e., if for any given $\varepsilon > 0$ there exist $\eta > 0$ and N_0 so that $P[|N^{\frac{1}{2}}(\tilde{\xi} - \xi)| < \eta] > 1 - \varepsilon$, $P[|N^{\frac{1}{2}}(\tilde{\theta} - \theta)| < \eta] > 1 - \varepsilon$ and $P[|N^{\frac{1}{2}}(\tilde{\gamma} - \gamma)| < \eta] > 1 - \varepsilon$, for $N > N_0$, then the estimates $\hat{\xi}. \hat{\theta}. \hat{\gamma}$ obtained by only one iteration of the formula (19) has the same asymptotic distribution of the maximum likelihood estimates. In other words, in such a case $N^{\frac{1}{2}}(\hat{\xi} - \xi, \hat{\theta} - \theta, \hat{\gamma} - \gamma)$ is asymptotically normally distributed with mean vector 0 and covariance matrix J^{-1}.

The first approximations $\tilde{\xi}, \tilde{\theta}, \tilde{\gamma}$ can be obtained in the following way.

In the special case when $A_1, \cdots, A_{[q]}$ and $B_1, \cdots, B_{[q]}$ are derived from the circulants, we take

$$\tilde{\xi} = \bar{z}_{ST}/(ST)^{\frac{1}{2}}$$

which is exactly the maximum likelihood estimate of ξ and $\hat{\xi} = \tilde{\xi}$, i.e., no adjustment is needed. Otherwise, we take

$$\tilde{\xi} = \sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^N x_{sti}/NST .$$

To obtain $\tilde{\theta}$ and $\tilde{\gamma}$, let $\tilde{x}_{sti} = x_{sti} - \tilde{\xi}$ and let $\tilde{x}_{st}$ be the N-dimensional column vector whose ith coordinate is $\tilde{x}_{sti}$. In view of the remark immediately following Lemma 3, if the joint pdf of x_i were the function $g(x)$ of Theorem 1, $\frac{1}{2}\theta_{kl}$ would be the common coefficient of $x_{s-l,t-k} - \xi$ and $x_{s+l,t-k} - \xi$ in the regression of x_{st} on $\{x_{s't'}, (s' - s)^2 + (t' - t)^2 \leqq q^2\}$. However, we are now considering $p(x)$ to be the joint pdf of x_i, but for large S, T, $p(x)$ differs from $g(x)$ only slightly and the actual role of θ_{kl} in the regression of x_{st} on $\{x_{s't'}, (s' - s)^2 + (t' - t)^2 \leqq q^2\}$ would still tend to be nearly the same as above. For this reason, we construct for each $s = \nu(q, 1) + 1, \cdots, S - \nu(q, 1)$, and $t = [q] + 1, \cdots, T$, an estimate

$$\tilde{\theta}_{st} = (\tilde{X}'_{st}\tilde{X}_{st})^{-1}\tilde{X}'_{st}\tilde{x}_{st}$$

of θ, where $\tilde{X}_{st}$ is an $N \times \alpha(q)$ matrix whose $\{\nu(q, k-1) + k - 1 + l\}$th column is $\frac{1}{2}(\tilde{x}_{s-l,t-k} + \tilde{x}_{s+l,t-k})$, and finally take the average of all these estimates to obtain

$$\tilde{\theta} = \sum_{s=\nu(q,1)+1}^{S-\nu(q,1)} \sum_{t=[q]+1}^{T} \tilde{\theta}_{st}/(S - 2\nu(q, 1))(T - [q]) .$$

To obtain $\tilde{\gamma}$ we now recall that $V_{st}(\xi) = \sum_{i=1}^{N} (z_{sti} - c_{st}\xi)^2/N$, has mean $1/K_s(\gamma)H_{st}(\theta)$ which indicates that for each s, t, $1/V_{st}(\tilde{\xi})H_{st}(\tilde{\theta})$ is a good estimate of $K_s(\gamma) = \sum_{l=0}^{[q]} \mu_{ls}\gamma_l$. Averaging these estimates over t, we obtain

$$\tilde{w}_s = T/\sum_{t=1}^{T} V_{st}(\tilde{\xi})H_{st}(\tilde{\theta})$$

as an estimate of $\sum_{l=0}^{[q]} \mu_{ls}\gamma_l$. For this reason we take

$$\tilde{\gamma} = (M'M)^{-1}M'\tilde{w}_s$$

where M is a $S \times \{[q] + 1\}$ matrix with μ_{ls} in its sth row and lth column.

6. Likelihood ratio test for the order of dependence. Let $q_0 < q_1$ denote two numbers of the form $(h^2 + k^2)^{\frac{1}{2}}$, h, k integers. In this section we consider the problem of testing the null hypothesis H_0: "order of dependence in $\{x_{st}\}$ is q_0" against the alternative hypothesis H_1: "order of dependence in $\{x_{st}\}$ is q_1" on the basis of N independent realizations of $\{x_{st}\}$. Again we note that if the joint pdf of $\{x_{st}\}$ were $g(x)$ with $q = q_1$, then the hypothesis of the smaller order of dependence q_0 would be equivalent to the hypothesis, $\theta_{kl} = 0$ for all k, l so that $q_0^2 < k^2 + l^2 \leqq q_1^2$ and $\gamma_l = 0$ for $l = [q_0] + 1, \cdots, [q_1]$. Hence when S and T are large and the joint pdf $p(x)$ of x with $q = q_1$ differs slightly from $g(x)$, the null hypothesis

$$\begin{aligned} H_0^*: \quad \theta_{kl} &= 0 \qquad \text{for all} \quad k, l \quad \text{so that} \quad q_0^2 < k^2 + l^2 \leqq q_1^2 \qquad \text{and} \\ \gamma_l &= 0 \qquad \text{for} \quad l = [q_0] + 1, \cdots, [q_1] \end{aligned}$$

and the alternative hypothesis, H_1^*: at least one of the parameters listed in H_0^* is nonzero, have practically the same meaning of H_0 and H_1 respectively. We now obtain the likelihood ratio λ for H_0^* against H_1^*. From (18) we immediately obtain

$$\begin{aligned} -2 \log \lambda = N \sum_{s=1}^{S} \sum_{t=1}^{T} \Big[&\log \frac{H_{st}(\hat{\theta})}{H_{st}(\theta^*)} + \log \frac{K_s(\hat{\gamma})}{K_s(\gamma^*)} \\ &+ K_s(\gamma^*)H_{st}(\theta^*)V_{st}(\xi^*) - K_s(\hat{\gamma})H_{st}(\hat{\theta})V_{st}(\hat{\xi})\Big] \end{aligned}$$

where $(\hat{\xi}, \hat{\theta}, \hat{\gamma})$ and $(\xi^*, \theta^*, \gamma^*)$ are maximum likelihood estimates of the parameters for $q = q_1$ and $q = q_0$ respectivly. The null hypothesis H_0^* is rejected when the test statistic $-2 \log \lambda$ is too large. Asymptotically, $-2 \log \lambda$ follows a χ^2-distribution with $\alpha(q_1) + [q_1] - \alpha(q_0) - [q_0]$ degrees of freedom under H_0^* as $N \to \infty$.

7. Miscellaneous remarks.

REMARK I. *Effect of large S and T on the convergence of the maximum likelihood estimates.* In Sections 5 and 6 we have discussed the asymptotic properties of the maximum likelihood estimates of the parameters ξ, θ, γ for a given order of dependence and of the likelihood ratio test for the order of dependence as $N \to \infty$. In this paragraph we shall give some heuristic arguments which seem to indicate that if S and T are moderately large in comparison with q, these asymptotic results will become applicable with only moderately large N. Our argument is based on the fact that the NST observations $\{x_{sti}\}$ are transformed into NST independent though not identically distributed random variables $\{z_{sti}\}$ from which the estimates and the tests are obtained. For example, if the order of dependence $q = 2$ and if $S = T = 30$, then with $N = 20$, we have 18,000 independent $\{z_{sti}\}$ to estimate only 7 parameters from. What we need are (i) the convergence of $\bar{J}_{ST} = \sum_{s=1}^{S} \sum_{t=1}^{T} J_{st}/ST$ to a nonsingular matrix J^* as $S \to \infty$, $T \to \infty$, and (ii) the convergence of the joint distribution of $(NST)^{-\frac{1}{2}}$ times the first partial derivatives of $\sum_{i=1}^{N} \log p^*(z_i \mid \xi, \theta, \gamma)$ with respect to $\xi, \theta_1, \cdots, \theta_{\alpha(q)}, \gamma_0, \gamma_1, \cdots, \gamma_{[q]}$, to a multivariate normal distribution with mean 0 and covariance matrix J^*. Routine computations are needed to check convergence (i) and we have done so for $q = 2$ when $A_1, \cdots, A_{[q]}$ and $B_1, \cdots, B_{[q]}$ are obtained from the circulants, but not attempted to do this in any generality. In order to check convergence (ii) one would have to verify conditions that will ensure that a multivariate central limit theorem for sums of independent but nonidentically distributed random vectors holds here (see Bergström (1949)).

REMARK II. *Discrepancy between the density functions g and p when S and T are large.* In Section 3 the density function $p(x)$ was adopted because (a) it differs from $g(x)$ of Theorem 1 only by the absence of a quadratic form $\psi(x)$ in the exponent which involves relatively few of the random variables $\{x_{st}\}$ when S and T are large, and (b) it is much easier than $g(x)$ to work with. The same arguments led Anderson (1962) in his study of order of dependence in one-way series. A question that we have not attempted to look into is the following. If we write g_{ST} for the function g in Theorem 1 and p_{ST} for the function p in (12), then does the discrepancy between g_{ST} and p_{ST} tend to disappear in some sense as $S, T \to \infty$? The remark (a) above is only a heuristic argument for expecting some such convergence. To pose the question formally one may consider $g_{mn,ST}$ and $p_{mn,ST}$ to be the marginal distributions of $\{x_{st}, s = 3[q] + 2, \cdots, m, t = 3[q] + 2, \cdots, n\}$ obtained from g_{ST} and p_{ST} respectively for $S = m + 1, m + 2, \cdots$, and $T = n + 1, n + 2, \cdots$, and ask whether for every fixed m, n these two sequences converge to a common function as $S, T \to \infty$.

REMARK III. *Order of dependence in normlly distributed r-way series for $r \geqq 3$.* Let $\{x_{t_1 \cdots t_r}, t_1 = 1, \cdots, T_1, \cdots, t_r = 1, \cdots, T_r\}$ be a normally distributed r-way series. We can extend Definition 1 to define the order of dependence for such a r-way series inductively as follows. Let $u_{t_1 \cdots t_r} = x_{t_1 \cdots t_r} - E(x_{t_1 \cdots t_r} \mid x_{t_1' \cdots t_r'}, t_1' = 1, \cdots, T_1, \cdots, t_{r-1}' = 1, \cdots, T_{r-1}, t_r' = 1, \cdots, t_r - 1)$ for $t_r = 2, \cdots, T_r$. We then define

DEFINITION 1′. The order of dependence in a normally distributed r-way series $\{x_{t_1 \cdots t_r}\}$ is defined to be the smallest $q \geqq 0$ so that

(a) For $t_r = 2, \cdots, T_r$, the conditional distribution of $x_{t_1 \cdots t_r}$ given $\{x_{t_1' \cdots t_r'}, t_1' = 1, \cdots, T_1, \cdots, t'_{r-1} = 1, \cdots, T_{r-1}, t_r' = 1, \cdots, t_r - 1\}$ depends only on the random variables $x_{t_1' \cdots t_r'}$ for which $\sum_{j=1}^{r} (t_j' - t_j)^2 \leqq q^2$, and

(b) For each $t_r = 2, \cdots, T_r$, the order of dependence in the normally distributed $(r-1)$-way series $\{u_{t_1, \cdots, t_r}, t_1 = 1, \cdots, T_1, \cdots, t_{r-1} = 1, \cdots, T_{r-1}\}$ is q.

Now Definition 1′ along with Anderson's (1962) definition for one-way series defines the order of dependence of r-way series for all r. Definition 1 of Section 2 can now be seen as a special case of Definition 1′ with $r = 2$.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] ANDERSON, T. W. (1962). Determination of the order of dependence in normally distributed time series in *Proceedings of the Symposium on Time Series Analysis.* Wiley, New York.

[3] BERGSTÖM, H. (1949). On the central limit theorem in the case of not equally distributed random variables. *Skand. Aktuarietidskr.* **33** 37–62.

[4] RAO, C. RADHAKRISHNA. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721