ANTE-DEPENDENCE ANALYSIS OF AN ORDERED SET OF VARIABLES[1]

K. R. Gabriel

Hebrew University, Jerusalem and University of North Carolina

1. Introduction and summary. For a set of variables in a given order, sth ante-dependence will be said to obtain if each one of the variables, given at least s immediate antecedent variables in the order, is independent of all further preceding variables. If the number of variables is p, ante-dependence is of some order between 0 and $p - 1$. 0th ante-dependence and $(p - 1)$st ante-dependence are equivalent to complete independence and to completely arbitrary patterns of dependence, respectively, and are defined irrespective of the ordering of the variables. 1st to $(p - 2)$nd ante-dependence are defined in terms of a specific order only.

If $X_1, X_2, \cdots, X_p$ are multivariate normal, sth ante-dependence is equivalent to each X_i, given $X_{i-1}, X_{i-2}, \cdots, X_{i-s}, \cdots, X_{i-s-z}$, being uncorrelated with $X_{i-s-z-1}, X_{i-s-z-2}, \cdots, X_2, X_1$ for any non-negative z. In other words, the partial correlation of X_i and $X_{i-s-z-k}$, given all the variables $X_{i-1}, X_{i-2}, \cdots, X_{i-s-z}$, is zero for all i, k and z.[2] The hypothesis that the covariance matrix is such that all the above partial correlations vanish will be denoted by D_s $(s = 0, 1, \cdots, p - 1)$, so that, for the multivariate normal distribution, D_s denotes the hypothesis of sth ante-dependence.

It is shown that for any set of ordered variables, normal or otherwise, D_s is equivalent to the following correspondence between the regression equations of X_i on all other variables, and on $X_{i-s}, X_{i-s+1}, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{i+s}$ only: the multiple correlations are equal, and the regression coefficients of $X_{i-s}, X_{i-s+1}, \cdots, X_{i+s}$ are equal in both equations, all other coefficients in the former equation being zero. It is also equivalent to the $(p - s)(p - s - 1)/2$ elements in the upper right (and also lower left) corner of the inverse covariance matrix being zero.[3] Indeed, any null hypothesis on a set of elements of the inverse covariance matrix may be formulated, and tested, as a hypothesis D_s if the variables can be so ordered as to put the zero elements in the upper right and lower left corners of the inverse.

Received November 20, 1959; revised June 12, 1961.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Sets of variables of $\mathbf{X}$ may be defined in a manner as to include formally variables not in $\mathbf{X}$, e.g., the set $X_{i-s}, X_{i-s+1}, \cdots, X_i, \cdots, X_{i-s}$ would have X_{i+s} undefined if $i \leqq s$. In such cases the sets will be understood to include only the variables which are defined in $\mathbf{X}$.

[3] In terms of Greenberg and Sarhan's paper [3] the inverse covariance matrix is diagonal of type $s - 1$.

Maximum likelihood estimates are derived under D_s for the normal case. Likelihood ratio tests of any one D_s against any other follow immediately and may be expressed in terms of the sample partial correlations. Exact distributions are not investigated, but for large samples χ^2 approximations are available. Thus a sequence of tests of D_{p-2} under D_{p-1}, D_{p-3} under D_{p-2}, $\cdots$, D_0 under D_1, is obtained which, in effect, forms a breakdown of the large sample test of independence, D_0, under the general alternative, D_{p-1}.

The assumptions of ante-dependence are clearly analogous to those of Markov processes and autoregressive schemes for time series; the motivation for the study and application of these models is also similar. The present model is more general in that it relaxes the usual autoregression assumptions of equal variances and, more crucial, of equal correlations between all pairs of equidistant variables (distance being meant in terms of the order of the set or time series). This greater generality requires analysis of a sample of observations for the study of ante-dependence, whereas for autoregressive schemes there are methods of analysis based on a single observation of the time series.

The ante-dependence models can be generalized to the case of several variables at each stage of the ordering. This would be analogous to the study of multiple time-series.

s-ante-dependent sets of variables may be generated by s successive summations of independent variables. This may be relevant for some applications of such models.

Ante-dependence models might be applicable to observations ordered in time or otherwise. Observations on growth of organisms up to each of several ages could be analyzed in such a manner. Where growth is recorded on several dimensions, e.g., height and weight, the analysis would proceed in terms of the multidimensional generalization of the model. Other possible fields of application include batteries of psychological tests increasing in complexity, and data on the successive location of travelling objects. A study of some such applications is now under way.[4]

2. A sequence of null hypotheses on partial correlations. Consider a vector variable $\mathbf{X} = (X_1, X_2, \cdots, X_p)$ with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Var}(\mathbf{X}) = \Sigma$, where Σ is positive definite. Denote the (i, j)th elements of Σ and Σ^{-1} by σ_{ij} and σ^{ij}, respectively. Define Δ as the $p \times p$ matrix with diagonal elements $\sigma_{11}, \sigma_{22}, \cdots, \sigma_{pp}$ and all other elements zero, and set $P = \Delta^{-\frac{1}{2}}\Sigma\Delta^{-\frac{1}{2}}$ with (i, j)th element $\rho_{ij} = \sigma_{ij}(\sigma_{ii}\sigma_{jj})^{-\frac{1}{2}}$ and P^{-1} with (i, j)th element ρ^{ij}. Next define

$$\rho_{ij\cdot k,l,\cdots,m} = (\rho_{ij\cdot l,\cdots,m} - \rho_{ik\cdot l,\cdots,m}\rho_{jk\cdot l,\cdots,m})(1 - \rho^2_{ik\cdot l,\cdots,m})^{-\frac{1}{2}}(1 - \rho^2_{jk\cdot l,\cdots,m})^{-\frac{1}{2}}$$

and

$$1 - \rho^2_{i(j_1,j_2,\cdots,j_s)} = \prod_{u=1}^{p}\{1 - \rho^2_{ij_u\cdot j_1,j_2,\cdots,j_{u-1}}\}.$$

[4] A generalization of ante-dependence analysis for any linear model, and a specialization of MANOVA under each order of ante-dependence has been presented to the ISI meeting in Paris, August–September, 1961 [9].

These are the usual definitions of the population values of the following parameters:

$$\sigma_{ij} = \text{covariance of } X_i \text{ and } X_j,$$

$$\rho_{ij} = \text{correlation of } X_i \text{ and } X_j,$$

$$\rho_{ij\cdot k,l,\cdots,m} = \text{partial correlation of } X_i \text{ and } X_j \text{ given } X_k, X_l, \cdots, X_m,$$

$$\rho_{i(j_1,j_2,\cdots,j_s)} = \text{multiple correlation of } X_i \text{ on } X_{j_1}, X_{j_2}, \cdots, X_{j_s}.$$

A sequence of hypotheses $G_0, G_1, G_2, \cdots, G_{p-1}$, is defined as

$$G_s : \rho_{i,i+s+1\cdot i+1,i+2,\cdots,i+s} = 0 \qquad \text{for all} \quad i = 1, 2, \cdots, p-s-1$$

together with the sequence of hypotheses

$$D_s = \bigcap_{u=s}^{p-1} G_u \qquad s = 0, 1, 2, \cdots, p-1.$$

Thus, considering D_s under D_{s+1} is equivalent to considering G_s under the assumption that $G_{s+1}, G_{s+2}, \cdots, G_{p-1}$ all hold.

These hypotheses form a chain from completely arbitrary correlations—under D_{p-1} no restrictions except positive definiteness of Σ are put on the correlations—to complete null correlation among all variables—under D_0 (see Theorem 1). As one goes from any D_s to D_{s-1} there are $p - s$ further restrictions on the correlations, amounting in effect to correlation of each X_i with one less of the preceding $X_{i-1}, X_{i-2}, \cdots$, or following $X_{i+1}, X_{i+2}, \cdots$. It will be shown in Theorem 1 that D_s is equivalent to Σ^{-1} having arbitrary elements in the principal diagonal and in the first s off diagonals, and zero elements everywhere else. (The sth off diagonals of Σ^{-1} are defined as the elements σ^{ij} for which $i - j = s$ (upper) or $j - i = s$ (lower)). In other words, the sequence $D_0, D_1, \cdots, D_{p-1}$ is one in which increasingly more off diagonals become arbitrary rather than zero. For D_0 all off diagonals are zero; for D_1 the first off diagonal is arbitrary, all others zero; and so forth until for D_{p-1} all elements of Σ^{-1} are arbitrary.

It should be noted that the hypotheses G_s and D_s are defined in terms of the given order of the variables $X_1, X_2, \cdots$, to X_p. For any other order of the variables the hypotheses would have a different meaning. In fact hypotheses about certain null partial correlations or null elements in the inverse covariance matrix might be formulated as hypotheses D_s by suitable permutation of the order of the variables. Only the two extreme hypotheses, D_0 and D_{p-1}, are defined irrespective of the order.

Lemma 1. *Under D_s, $\rho_{ij\cdot\Phi} = 0$ for all i and j, where Φ denotes any set of s or more successive variables between X_i and X_j.*

Proof. Consider first Φ to be $i + 1, i + 2, \cdots, i + s$, where $i + s + 1 < j$ (for $i + s + 1 = j$ the result is trivial). Then

$$\rho_{ij\cdot\Phi,i+s+1} = (\rho_{ij\cdot\Phi} - \rho_{i,i+s+1\cdot\Phi}\rho_{j,i+s+1\cdot\Phi})(1 - \rho^2_{i,i+s+1\cdot\Phi})^{-\frac{1}{2}}(1 - \rho^2_{j,i+s+1\cdot\Phi})^{-\frac{1}{2}},$$

and since under D_s, $\rho_{i,i+s+1\cdot\Phi} = 0$,

$$\rho_{ij\cdot\Phi,i+s+1} = 0 \leftrightarrow \rho_{ij\cdot\Phi} = 0.$$

Similarly, under D_s,

$$\rho_{ij\cdot\Phi,i+s+1,i+s+2} = 0 \leftrightarrow \rho_{ij\cdot\Phi,i+s+1} = 0,$$

and finally, under D_s,

$$\rho_{ij\cdot\Phi,i+s+1,i+s+2,\cdots,j-1} = 0 \leftrightarrow \rho_{ij\cdot\Phi} = 0.$$

Now the LHS equality holds under D_{j-i-1}, which is implied by D_s as $j - i - 1 > s$. Therefore under D_s the RHS equality also holds.

The same argument is readily extended to cases where Φ does not include X_{i+1}.

LEMMA 2. *Under D_s $\rho_{ij\cdot\Psi} = 0$, for all i and j, where Ψ is any set including at least s successive variables intermediate to X_i and X_j.*

PROOF. Write any s successive variables in Ψ which are intermediate to X_i and X_j as Φ and the rest of Ψ as $X_{k_1}, X_{k_2}, \cdots, X_{k_u}$; X_{k_u} being chosen as the variable from among $\Psi - \Phi$ which is nearest in the ordering either to X_i or to X_j. Now

$$\rho_{ij\cdot\Psi} = \rho_{ij\cdot\Phi,k_1,k_2,\cdots,k_u}$$
$$= \frac{(\rho_{ij\cdot\Phi,k_1,\cdots,k_{u-1}} - \rho_{ik_u\cdot\Phi,k_1,\cdots,k_{u-1}}\rho_{jk_u\cdot\Phi,k_1,\cdots,k_{u-1}})}{(1 - \rho^2_{ik_u\cdot\Phi,k_1,\cdots,k_{u-1}})^{\frac{1}{2}}(1 - \rho^2_{jk_u\cdot\Phi,k_1,\cdots,k_{u-1}})^{\frac{1}{2}}},$$

so that if the Lemma holds for all Ψ with $s + u - 1$ terms or less, $\rho_{ij\cdot\Phi,k_1,\cdots,k_{u-1}} = 0$ and either $\rho_{ik_u\cdot\Phi,k_1,\cdots,k_{u-1}} = 0$ or $\rho_{jk_u\cdot\Phi,k_1,\cdots,k_{u-1}} = 0$, whence the LHS will also be zero and the Lemma hold for all Ψ with $s + u$ terms. But by Lemma 1 it holds for $u = 0$, and therefore it must hold for all $u = 0, 1, 2, \cdots$.

LEMMA 3. *For any non-singular $p \times p$ matrix Σ the following two statements are equivalent for any i:*

(a) $\sigma^{ij} = 0$, *for all j such that $|i - j| > s$;*

(b) *the ith column of the inverse of the principal minor of the $i - s, i - s + 1, \cdots, i, \cdots, i + s$ rows and columns of Σ is made up of the corresponding elements of the inverse of Σ.*

PROOF. By definition, for all $j = 1, 2, \cdots, p$,

$$\sum_{k=1}^{p} \sigma_{jk}\sigma^{ki} = \delta_{ij}$$

(Kroeneker's delta). It can easily be checked that

$$\sum_{k=i-s}^{i+s} \sigma_{jk}\sigma^{ki} = \delta_{ij} \qquad \text{for all } j = 1, 2, \cdots p,$$

if and only if $\sigma^{ki} = 0$ for all $|k - i| > s$.

The second set of equations for $j = i - s, i - s + 1, \cdots, i, \cdots, i + s$ define $(\sigma^{i-s,i}, \sigma^{i-s+1,i}, \cdots, \sigma^{i,i}, \cdots, \sigma^{i+s,i})$ as the ith column of the inverse of the principal minor of the $(i - s, i - s + 1, \cdots, i, \cdots, i + s)$ rows and columns of Σ.

LEMMA 4. *For any non-singular $p \times p$ matrix Σ, the condition*

$$\sigma^{ij} = 0 \qquad \text{for all } i, j \text{ such that } |i - j| > s,$$

implies that in the inverse of any principal minor of consecutive rows and columns of Σ, all elements outside the principal diagonal and the first s off diagonals are zero.

PROOF. Consider any principal minor of the first $u + s$ rows and columns. By definition

$$\sum_{k=1}^{p} \sigma_{jk}\sigma^{ki} = \delta_{ij} \qquad \text{for all } i \text{ and } j,$$

so that if $\sigma^{ij} = 0$ for $|i - j| > s$,

$$\sum_{k=1}^{u+s} \sigma_{jk}\sigma^{ki} = \delta_{ij} \qquad \text{for } i = 1, 2, \cdots, u \quad \text{and} \quad j = 1, 2, \cdots, u + s.$$

But the latter equations define the first u columns of the inverse of the principal minor, which are therefore seen to have the same elements as the corresponding parts of the first u columns of Σ^{-1}. In particular, all elements for which $i > j + s$ must be zero.

By the same reasoning about the first u rows of the inverse of that principal minor, all elements for which $j > i + s$ must be zero.

The same argument holds for principal minors of the last $u + s$ consecutive rows. For minors of s rows and columns the argument is trivial.

Now, any principal minor of consecutive rows and columns of Σ can be obtained from Σ by first taking the principal minor whose last rows and columns are the ones concerned, and then taking the required minor as the last so and so many rows and columns of that. For example, in a matrix of four rows and columns one would obtain the principal minor of the second and third rows and columns by first striking out the fourth, and then from the remainder striking out the first. As has been shown, the property that all elements outside the first s off diagonals vanish is preserved when taking minors in this manner. Therefore that property holds for any principal minor of consecutive rows and columns of Σ if it holds for Σ.

THEOREM 1. *For a vector variable* $\mathbf{X}$ *with variance* Σ *and correlation* P, *the following three statements are equivalent*:

D_s : $\rho_{i,i+u+1 \cdot i+1, \cdots, i+u} = 0$ *for all* i *and all* $u \geqq s$;

D'_s : $\sigma^{ij} = 0$ *for all* $|i - j| > s$;

D''_s : *For all* i, *the regressions of* X_i

(1) *on all other variables in* $\mathbf{X}$, *and*

(2) *on all other* X_j *such that* $|i - j| \leqq s$, *have equal multiple correlations and regression co-efficients of* X_j *for* $|i - j| \leqq s$, *and all other regression co-efficients in* (1) *are null.*

PROOF. By Lemmas 1 and 2, D_s implies that $\rho_{ij \cdot \text{all others in } \mathbf{X}} = 0$ if $|i - j| > s$. But

$$\rho_{ij \cdot \text{all others in } \mathbf{X}} = -\sigma^{ij}(\sigma^{ii}\sigma^{jj})^{-\frac{1}{2}}$$

([2]—23.4.2—gives this in terms of co-factors). Thus $D_s \rightarrow D'_s$.

By Lemma 4, D_s' implies that, for any i, in the inverse of the principal minor of the $(i, i + 1, \cdots i + u + 1)$th rows and columns of Σ, the $(i, i + u + 1)$th element is zero if $u \geqq s$. But $\rho_{i,i+u+1\cdot i+1,\cdots,i+u}$ is a multiple of that element (by the same expression quoted above), so it also is zero under D_s'. Thus $D_s' \rightarrow D_s$.

The regression of any X_i on all the other variables in $\mathbf{X}$ has (regression co-efficient of X_j) $= -\sigma^{ii}/\sigma^{ij}$, ([2]—23.2.4—) and (multiple correlation) $= (1 - 1/\sigma_{ii}\sigma^{ii})^{\frac{1}{2}}$, ([2]—23.5.2—).

Under D_s' clearly the regression co-efficients of X_j for $|i - j| > s$ are zero. By Lemma 3, σ^{ij} for $|i - j| \leqq s$ are equal to the corresponding elements in the ith column of the principal minor of the $(i - s, i - s + 1, \cdots, i, \cdots, i + s)$th rows and columns of Σ. But the regression of X_i on $X_{i-s}, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{i+s}$ involves the elements of the ith column of that minor in the same formulae. As the elements involved are equal for both regressions, the regression co-efficients and multiple correlations are necessarily also equal.

Conversely, if the regressions are the same, application of the formulae above shows the non-null elements of the ith columns of Σ and the minor to be the same. By the converse of Lemma 3, $\sigma^{ji} = 0$ for j such that $|i - j| > s$. Thus $D_s' \leftrightarrow D_s''$.

This completes the proof of the theorem.

It should be noted that D_s could equivalently have been stated in terms of $\sigma_{i,i+u+1\cdot i+1,\cdots,i+u}$, and D_s' in terms of ρ^{ij}.

The above are hypotheses of null correlation or covariance, not of independence. For a multivariate normally distributed $\mathbf{X}$, however, they are equivalent to independence [7], and D_s would correspond to the definition of sth ante-dependence given in the introduction.

The hypotheses, or models $D_0, D_1, \cdots, D_{p-1}$ may be considered in terms of regression on preceding variables in the sequence. Each variable X_i may be partitioned into its regression on $X_{i-1}, X_{i-2}, \cdots, X_{i-s}$, plus a residual which, under D_s, is uncorrelated with $X_1, X_2, \cdots, X_{i-s-1}$. This model is a generalization of an autoregressive scheme of order s [6], though the latter usually involves the additional assumption of equal regression functions at each stage. Moreover, the statistical treatment of autoregression commonly is in terms of a single observation on the whole series, whereas the present discussion is in terms of repeated samples of the sequence.

An interesting way to generate sth ante-dependent variables is by repeated (s times) summation of uncorrelated variables. If $X_i(s)$ be the ith value of the sth successive sum of $Z_1, Z_2, \cdots$, then the standard expression for an nth difference is $Z_i = \sum_{k=0}^{s} (-1)^k \binom{s}{k} X_{i-k}^{(s)}$. This can be rewritten

$$X_i^{(s)} = \sum_{k=1}^{s} (-1)^{k-1} \binom{s}{k} X_{i-k}^{(s)} + Z_i$$

showing that $X_i^{(s)}$, given $X_{i-1}^{(s)}, X_{i-2}^{(s)}, \cdots, X_{i-k}^{(s)}$ is uncorrelated with $X_{i-k-1}, \cdots, X_2, X_1$, i.e., that D_s holds for $\mathbf{X}^{(s)}$.

The case $s = 1$ is known as Guttman's Perfect Simplex [4, 5]. It has been proposed as a model for batteries of psychological tests increasing in complexity by successive addition of independent increments. For $s > 1$ the relative weight of the earlier increments becomes greater—no simple application is known.

3. Likelihood ratio tests of partial independence. For $\mathbf{X}$ multivariate normal, consider a sample of N, i.e., $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$ where $\mathbf{x} = (x_{1\nu}, x_{2\nu}, \cdots, x_{p\nu})$. Define $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)$ where $\bar{x}_i = N^{-1}\sum_{\nu=1}^{N} x_{i\nu}$, also define $s_{ij} = N^{-1}\sum_{\nu=1}^{N} (x_{i\nu} - \bar{x}_i)(x_{j\nu} - \bar{x}_j)$ for all $i, j = 1, 2, \cdots, p$ and let s_{ij} be the (i, j)th element of S. Similarly to the population parameters, define

S^{-1} with element s^{ij},
D with diagonal $s_{11}, s_{22}, \cdots, s_{pp}$ and other elements zero,
$R = D^{-\frac{1}{2}}SD^{-\frac{1}{2}}$ with element r_{ij},
R^{-1} with element r^{ij},
$r_{ij\cdot k,l,\cdots m} = (r_{ij\cdot l,\cdots,m} - r_{ik\cdot l,\cdots,m}\, r_{jk\cdot l,\cdots,m})(1 - r^2_{ik\cdot l,\cdots,m})^{-\frac{1}{2}}(1 - r^2_{jk\cdot l,\cdots,m})^{-\frac{1}{2}}$,

and

$$1 - r^2_{i(j_1, j_2, \cdots, j_s)} = \prod_{u=1}^{s} \{1 - r^2_{ij_u\cdot j_1, j_2, \cdots, j_{u-1}}\}.$$

THEOREM 2. *Let* $\mathbf{X}$ *have a multivariate normal distribution with expectation* $\boldsymbol{\mu}$ *and variance* Σ, *and* $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$ *be a sample of* N *from it. Let* Γ *be a set of elements of unordered pairs of numbers* (i, j), $0 \leqq i, j \leqq p$, *including* $(1, 1)$, $(2, 2), \cdots, (p, p)$, *and let* $\sum_\gamma$ *denote summation over all pairs* $(i, j) \in \Gamma$. *Consider the hypothesis*

$$D_\Gamma : \sigma^{ij} = 0 \qquad \text{if } (i, j) \notin \Gamma.$$

Then under D_Γ *the maximum likelihood is*

$$(2\pi)^{-Np/2}\, |\hat{\Sigma}_{(\Gamma)}|^{-N/2} \exp[-Np/2],$$

where $\hat{\Sigma}_{(\Gamma)}$ *has elements* $\hat{\sigma}_{ij} = s_{ij}$ *for* $(i, j) \in \Gamma$, *and* $\hat{\Sigma}^{-1}_{(\Gamma)}$ *satisfies* $\hat{\sigma}^{ij} = 0$ *for* $(i, j) \notin \Gamma$.

PROOF. Under D_Γ the parameters $\boldsymbol{\mu}$ and σ_{ij} for $(i, j) \in \Gamma$ specify the distribution completely. The likelihood is maximized by putting

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$$

and choosing $\hat{\sigma}_{ij}$ to satisfy

$$\sum_{v=1}^{p} \sum_{z=1}^{p} (\sigma_{vz} - s_{vz}) \frac{\partial \sigma^{vz}}{\partial \sigma^{ij}} = 0$$

for all $(i, j) \in \Gamma$. (This is a simple extension of the well known theory. See [1], Sec. 3.2). The latter is satisfied by putting

$$\hat{\sigma}_{ij} = s_{ij} \qquad \text{for all } (i, j) \in \Gamma.$$

The maximum likelihood can be written

$$(2\pi)^{-Np/2}\, |\hat{\Sigma}_{(\Gamma)}|^{-N/2} \exp[-\tfrac{1}{2}N \operatorname{Tr} \hat{\Sigma}^{-1}_{(\Gamma)}(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})' - \tfrac{1}{2}N \operatorname{Tr} \hat{\Sigma}^{-1}_{(\Gamma)}S],$$

where $\hat{\mathbf{y}}$, as above, and $\hat{\Sigma}_{(\Gamma)}$ are the maximum likelihood estimates under D_Γ. Clearly the first term in the exponent vanishes. The second term is

$$\begin{aligned}
-\tfrac{1}{2}N \operatorname{Tr} \hat{\Sigma}_{(\Gamma)}^{-1} S &= -\tfrac{1}{2}N \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\sigma}^{ij} s_{ij} \\
&= -\tfrac{1}{2}N \sum_{\gamma} \hat{\sigma}^{ij} s_{ij} \\
&= -\tfrac{1}{2}N \sum_{\gamma} \hat{\sigma}^{ij} \hat{\sigma}_{ij} \\
&= -\tfrac{1}{2}N \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\sigma}^{ij} \hat{\sigma}_{ij} \\
&= -\tfrac{1}{2}N \operatorname{Tr} \hat{\Sigma}_{(\Gamma)}^{-1} \hat{\Sigma}_{(\Gamma)} \\
&= -\tfrac{1}{2}Np,
\end{aligned}$$

for as $\hat{\sigma}^{ij} = 0$ for $(i, j) \not\in \Gamma$, $\sum_\gamma \hat{\sigma}^{ij} a_{ij} = \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\sigma}^{ij} a_{ij}$ for any a_{ij}.

LEMMA 5. *For any correlation matrix P*

$$|P| = \prod_{i=1}^{p-1} (1 - \rho_{i,i+1}^2) \prod_{i=1}^{p-2} (1 - \rho_{i,i+2\cdot i+1}^2) \cdots \prod_{i=1}^{p-s} (1 - \rho_{i,i+s\cdot i+1,i+2,\cdots,i+s-1}^2) \cdots (1 - \rho_{1p\cdot 2,\cdots,p-1}^2).$$

PROOF. It is well known ([2]—23.5.2—) that

$$\frac{|P|}{|P_{11}|} = 1 - \rho_{1(23\cdots p)}^2,$$

whence

$$\frac{|P|}{|P_{11}|} = (1 - \rho_{12}^2)(1 - \rho_{13.2}^2) \cdots (1 - \rho_{1p\cdot 23\cdots p-1}^2).$$

Similarly

$$\frac{|P_{11}|}{|P_{1122}|} = (1 - \rho_{23}^2)(1 - \rho_{24\cdot 3}^2) \cdots (1 - \rho_{2p\cdot 34\cdots p-1}^2),$$

and so on for further principal minors of P. Multiplying these results together, the above result follows.

COROLLARY 2.1. *Under D_s the maximum likelihood is*

$$(2\pi)^{-Np/2} |\hat{\Sigma}_{(s)}|^{-N/2} \exp [-Np/2]$$

where

$$|\hat{\Sigma}_{(s)}| = \prod_{i=1}^{p} s_{ii} \prod_{i=1}^{p-1} (1 - r_{i,i+1}^2) \prod_{i=1}^{p-2} (1 - r_{i,i+2\cdot i+1}^2) \cdots \prod_{i=1}^{p-s} (1 - r_{i,i+s\cdot i+1,i+2,\cdots i+s-1}^2).$$

PROOF. It was shown in Theorem 1 that D_s is a hypothesis of the type considered in Theorem 2. Hence by Theorem 2 the above expression for the likelihood follows.

Now under D_s the maximum likelihood estimators of σ_{ii} are s_{ii}, and those of the non-null partial correlation co-efficients are the sample partial correlations. This follows from the fact that the set of variances and correlations ρ_{ij} where $|i - j| \leqq s$ are a one-to-one transformation of the set of variances and non-null partial correlations under D_s. (For similar considerations see [1] Corollary 3.2.1 and section 4.3.1).

Next $\Sigma = \Delta^{\frac{1}{2}} P \Delta^{\frac{1}{2}}$, whence $|\Sigma| = |\Delta|\, |P|$. Now $|\Delta| = \prod_{i=1}^{p} \sigma_{ii}$, and under D_s

$$|P| = \prod_{i=1}^{p-1} (1 - \rho_{i,i+1}^2) \prod_{i=1}^{p-2} (1 - \rho_{i,i+2\cdot i+1}^2) \cdots \prod_{i=1}^{p-s} (1 - \rho_{i,i+s\cdot i+1,\cdots i+s-1}^2).$$

Introducing the maximum likelihood estimates one obtains the expression for $|\hat{\Sigma}_{(s)}|$.

COROLLARY 2.2. *The α-level likelihood ratio test of D_s under D_{s+1} is,*

$$\textit{accept } D_s \textit{ if } -N \sum_{p=1}^{p-s-1} \log_e(1 - r_{i,i+s+1\cdot i+1,\cdots i+s}^2) \leqq \chi_{1-\alpha}^2,$$

reject D_s otherwise,

where $\chi_{1-\alpha}^2$ is the $100(1 - \alpha)$ percentage point of the χ^2 distribution with $(p - s - 1)$ degrees of freedom. This test is asymptotically valid.

PROOF. The ratio of the likelihood under D_s to that under D_{s+1} is, from Corollary 2.1,

$$\lambda_s = \{|\hat{\Sigma}_{(s)}|/|\hat{\Sigma}_{(s+1)}|\}^{-N/2} = \left\{ \prod_{i=1}^{p-s-1} (1 - r_{i,i+s+1.i+1,\cdots,i+s}^2) \right\}^{N/2},$$

so that

$$-2 \log_e \lambda_s = -N \sum_{i=1}^{p-s-1} \log_e(1 - r_{i,i+s+1\cdot i+1,\cdots,i+s}^2).$$

Now D_s specifies $(p - s)(p - s - 1)/2$ zero elements in Σ^{-1} and D_{s+1} specifies only $(p - s - 1)(p - s - 2)/2$ of them as zero. Thus by the well known results on the asymptotic distribution of the likelihood criterion [8], $-2 \log_e \lambda_s$ is asymptotically distributed as χ^2 with $(p - s - 1)$ degrees of freedom; nothing definite is known about the closeness of the asymptotic approximation for any finite sample size.

A closer approximation to the asymptotic distribution of the likelihood criterion results from the following considerations. For zero partial correlation $\rho_{i,i+s+1\cdot i+1,\cdots,i+s}$ one uses the even moments of the partial correlation [1] to calculate

$$E(1 - r_{i,i+s+1\cdot i+1,\cdots,i+s}^2) = 1 - 1/(N - s - 1)$$

and

$$\mathrm{Var}\,(1 - r_{i,i+s+1\cdot i+1,\cdots,i+s}^2) = 2(N - s - 2)/(N - s + 1)(N - s - 1)^2.$$

Now taking the logarithm of $1 - r^2$—omitting the subscripts—and expanding it as a Taylor's series, one obtains for the first two terms,

$$\begin{aligned} E \log_e(1 - r^2) &= \log_e E(1 - r^2) - \operatorname{Var}(1 - r^2)/2[E(1 - r^2)]^2 \\ &= \log_e[1 - 1/(N - s - 1)] - 1/(N - s + 1)(N - s - 2) \\ &= -1/(N - s - 1) - 1/2(N - s - 1)^2 - 1/3(N - s - 1)^3 \\ &\quad - \cdots - 1/(N - s + 1)(N - s - 2) \\ &= -1/(N - s - 1) - 0[1/(N - s - 1)^2]. \end{aligned}$$

Hence

$$\begin{aligned} E(-2 \log_e \lambda_s) &= -N \sum_{i=1}^{p-s-1} \{-1/(N - s - 1) - 0[1/(N - s - 1)^2]\} \\ &= (p - s - 1)N/(N - s - 1) + 0[1/(N - s - 1)]. \end{aligned}$$

The distribution of $-2 \log_e \lambda_s$ is asymptotically χ^2 with $p - s - 1$ degrees of freedom, so its expectation should be $p - s - 1$. This expectation will be approximated more rapidly by using the criterion

$$-2 \log_e \lambda_s' = -(N - s - 1) \sum_{i=1}^{p-s-1} \log_e(1 - r^2_{i,i+s+1\cdot i+1,\cdots,i+s}).$$

Tests for each D_s under D_{s+1}, $s = 0, 1, 2, \cdots, p - 2$, are given in Corollary 2.2. As these are likelihood ratios, successive tests can be combined by adding the statistics and the degrees of freedom of χ^2, e.g., to test D_1 under D_3, the test statistic

$$-2 \log_e \lambda_1 - 2 \log_e \lambda_2 = -N\left\{\sum_{i=1}^{p-2} \log_e(1 - r^2_{i,i+2.i+1}) + \sum_{i=1}^{p-3} \log_e(1 - r^2_{i,i+3.i+1,1+2})\right\}$$

is distributed asymptotically as χ^2 with $(p - 2) + (p - 3) = 2p - 5$ degrees of freedom. Another example is the test of D_0 under D_{p-1}, i.e., of independence under the most general alternative, whose statistic is

$$\sum_{i=0}^{p-2} -2 \log_e \lambda_s = -2 \log_e \prod_{i=0}^{p-2} \lambda_s = N \log_e\{|\hat{\Sigma}_{(0)}|/|\hat{\Sigma}_{(p-1)}|\} = -N \log_e |\hat{P}|$$

distributed as χ^2 with $p(p - 1)/2$ degrees of freedom, giving the usual likelihood ratio test for independence [1].

Thus the present sequence of tests breaks up the general test of independence into a sequence of intermediate steps of increasing degrees of ante-dependence. Procedures for inferring the degree of ante-dependence may be adapted from these tests, for instance in analogy to the procedures for inferring the degree of polynomial regression by the use of tests of successive regression co-efficients.

All tests, except that of D_0 under D_{p-1}, depend on the ordering of the variables.

This is assumed given; procedures for inferring possible orderings of the variables have not been investigated.

4. Generalization to several dimensions. A multidimensional generalization of the foregoing will be to consider p sets of variables $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_p$ with k_i variables in set $\mathbf{X}_i = (X_{i_1}, X_{i_2}, \cdots, X_{i_{k_i}})$. The corresponding hypotheses can be written

$$\mathbf{G}_s : \rho_{i_\alpha,(i+s+1)_\beta\cdot i+1,i+2,\cdots,i+s} = 0$$

for all $i = 1, 2, \cdots p - s - 1$ and all $\alpha = 1, 2, \cdots k_i$ and $\beta = 1,2, \cdots k_{i+s+1}$, the subscript $i + u$ denoting all the variables of $\mathbf{X}_{i+u}$. As before,

$$\mathbf{D}_s : \bigcap_{u=s}^{p-1} \mathbf{G}_u .$$

This generalized model is analogous (for constant k_i) to the extension of autoregressive schemes for multiple time series. The present hypotheses relate only to the order between the sets, and are invariant under permutations of the variables within any or all sets (unlike multiple time series).

The results of the previous section can be extended to this more general case by partitioning the matrix Σ of $\sum_{i=1}^{p} k_i$ rows and columns, into a $p \times p$ supermatrix of minors with $k_1, k_2, \cdots, k_p$ rows and columns, respectively. The main theorem now becomes

THEOREM 1′. *For a vector variable* $\mathbf{X} = (X_{1_1}, \cdots, X_{1_{k_1}}, X_{2_1}, \cdots, X_{p_{k_p}})$ *with variance* Σ *and correlation* P, *the following three statements are equivalent*:

$\mathbf{D}_s : \rho_{i_\alpha(i+u+1)_\beta\cdot i+1,i+2,\cdots,i+u} = 0$ *for all* $u = s, s+1, \cdots, p-1$ *and all* α, β;

$\mathbf{D}_s' : \sigma^{i_\alpha j_\beta} = 0$ *for all* α, β *and all* i, j *such that* $|i - j| > s$;

$\mathbf{D}_s''$: *for any* i *and any* α *in the regressions of* X_{i_α} *on*:

(1) *all other variables in* $\mathbf{X}$ *apart from those in* $\mathbf{X}_i$,

(2) *all other* X_{j_β} *with any* β *and* j *such that* $0 < |i - j| \leqq s$, *the multiple correlations and the regression co-efficients of* X_{j_β} *are equal, and all other regression co-efficients in* (1) *are zero.*

Theorem 2 stands as it is also for the generalized case.

Denote the partial correlation of X_{i_α} and X_{j_β}, given $X_{i_{\alpha+1}}, X_{i_{\alpha+2}}, \cdots, X_{j_{\beta-1}}$, as $\rho_{i_\alpha j_\beta\cdot\text{int}}$ (if $j \leqq i$ and/or $\beta \leqq \alpha$, the definition would be adjusted suitably). Define next, Γ_s' the class of all pairs (i_α, j_β) such that $j - i = s$. Then Lemma 5 and Corollary 2.2 can be reformulated as

LEMMA 5′.

$$|P| = \prod_{u=0}^{p-1} \prod_{(i_\alpha,j_\beta)\varepsilon\Gamma_s'} (1 - \rho^2_{i_\alpha j_\beta\cdot\text{int}});$$

and

COROLLARY 2.2′. *The* α*-level likelihood ratio test of* $\mathbf{D}_s$ *under* $\mathbf{D}_{s+1}$ *is*:

accept $\mathbf{D}_s$ *if* $-N \sum_{(i_\alpha,j_\beta)\varepsilon\Gamma_s'} \log_e(1 - r^2_{i_\alpha j_\beta\cdot\text{int}}) \leqq \chi^2_{1-\alpha}$,

reject $\mathbf{D}_s$ *otherwise*,

where $\chi^2_{1-\alpha}$ *is the* $100(1 - \alpha)$ *percentage point of the* χ^2 *distribution with* $\sum_{i=1}^{p-s-1} k_i k_{i+s+1}$ *degrees of freedom.*

This provides a sequence of tests corresponding to those in Section 3.

Acknowledgment. I am greatly indebted to Professor S. N. Roy for his stimulating lectures and discussions and for his suggestions on the presentation of this work. I am also indebted to Professor L. Guttman who first set me thinking along the lines developed in this paper.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[2] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] B. G. Greenberg and A. E. Sarhan, "Matrix inversion, its interest and application in analysis of data," *J. Amer. Stat. Assn.*, Vol. 45 (1959), pp. 755–766.

[4] L. Guttman, "A new approach to factor analysis: the radex," In *Mathematical Thinking in the Social Sciences*, ed. P. F. Lazarsfeld, The Free Press, Glencoe, 1954.

[5] L. Guttman, "A generalized simplex for factor analysis," *Psychometrika*, Vol. 20 (1955), pp. 173–192.

[6] Maurice G. Kendall, *The Advanced Theory of Statistics*, Vol. 2, Charles Griffin and Co., London, 1948.

[7] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[8] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[9] K. R. Gabriel, "The model of ante-dependence for data of biological growth," *Bull. Inst. Internat. Stat.*, 33rd Session, Paris, 1961.