BIOMETRICS 32, 95–108
March, 1976

# ANALOGIES BETWEEN MULTIPLICATIVE MODELS IN CONTINGENCY TABLES AND COVARIANCE SELECTION

NANNY WERMUTH

*Institut für Medizinische Statistik und Dokumentation, Universität Mainz, 65 Mainz, Langenbeckstrasse 1, Federal Republic of Germany*

## SUMMARY

A certain class of patterns of association can be investigated by fitting multiplicative models to a contingency table or by using covariance selection on a covariance matrix. We show that each multiplicative model for a contingency table corresponds to one particular covariance selection model, and we point at the resulting similarities in the interpretation of patterns, in test statistics for each pattern and in implied marginal associations among variable pairs.

## 1. INTRODUCTION

The purpose of analyzing data is to find structures which are complex enough to fit the data but simple enough to facilitate interpretation. Structures describing interrelations among several variables may be called patterns of association. We characterize a certain class of patterns by the concept of zero partial association and show that it is this class of patterns which can be studied by fitting multiplicative models to a contingency table or by fitting certain covariance selection models to a covariance matrix.

The theory of log-linear models (Birch [1963]) as well as the theory of covariance selection (Dempster [1972]) has been developed for many variables. Before these methods can find widespread use as data analytic tools, however, at least three problems have to be resolved: First, and most important, is the question of how to interpret complex models. Second, programmable algorithms should be available to compute estimates and test statistics for a given pattern of association. Third, methods and algorithms have to be developed to guide the choice among several plausible patterns of association for a given set of data. Following is a brief review of the results achieved so far with regard to these aspects:

For *log-linear models,* interpretations of the models and of individual parameters of the models have, for instance, been proposed by Roy and Kastenbaum [1956], Darroch [1962], Bishop [1971] and Goodman [1970, 1973]. The parameters that are known as "high order interactions," however, are difficult to interpret. The computational problems with obtaining cell estimates for a given log-linear model have been solved by Bishop [1967]. A Fortran program to fit log-linear models is now available from Goodman. No programmed methods are yet available for the model search. A Bayesian approach to analyzing contingency tables is due to Lindley [1964].

The parameters in *covariance selection models* have been termed "concentrations," and only their geometrical interpretation has been given (Dempster [1969]). Extensive iterations are required for Dempster's [1972] proposed cyclic-fitting and Newton-type algorithms which can be used to compute the estimates under a given covariance selection model and to select models.

In this paper we make use of analogies between multiplicative models for contingency tables and covariance selection models. In Section 3 we discuss the different possible patterns of association for three variables. We present rules for the interpretation of patterns and for computing test statistics. Furthermore, for each pattern we show the implications for the marginal associations among variable pairs. Sections 4 and 5 give the patterns for four variables and generalizations to many variables.

## 2. NOTATION AND ASSUMPTIONS

Of the many possible interrelations among variables we concentrate on a class of patterns characterized by zero partial associations, that is, by partially independent pairs of variables. This class of patterns is frequently being studied either implicitly or explicitly whenever a covariance selection model is fitted to a covariance matrix, or whenever independence hypotheses are tested in a contingency table.

### *2.1 Covariance Selection.*

More precisely, for covariance selection models it is assumed that $p$ variables follow a joint normal distribution represented by the density

$$f(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{p/2} |\mathbf{\Sigma}|^{-1/2} \exp\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\}, \tag{1}$$

where $\mathbf{\Sigma}$ and its inverse $\mathbf{\Sigma}^{-1}$ are both $p \times p$ positive definite symmetric matrices. The $(i, j)$ element $\sigma_{ij}$ of $\mathbf{\Sigma}$ is the covariance between $x_i$ and $x_j$, and $\sigma^{ij}$, the $(i, j)$ element of $\mathbf{\Sigma}^{-1}$, has been termed "concentration." The $(i, j)$th elements of the $p \times p$ correlation matrix $\mathbf{P}$ and of the inverse correlation matrix $\mathbf{P}^{-1}$ are $\rho_{ij} = \sigma_{ij}/\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}$ and $\rho^{ij} = \sigma_{ii}^{1/2}\sigma_{jj}^{1/2}\sigma^{ij}$, respectively.

If no restrictions are imposed on the parameters $\sigma^{ij}$, the maximum likelihood estimate for the covariance structure is known to be the observed covariance matrix. Dempster [1972] shows the existence of a unique estimate $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ in a covariance selection model with some of the parameters $\sigma^{ij}$ for $i \neq j$ restricted to zero. He proves that this estimate $\hat{\mathbf{\Sigma}}$ is maximum likelihood and that $\hat{\mathbf{\Sigma}}$ is completely defined by the pattern of zeros in the inverse $\mathbf{\Sigma}^{-1}$.

With $n$ observations on each of the $p$ variables, the sample covariance is defined as

$$s_{ij} = \frac{1}{n-1} \sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j), \tag{2}$$

where

$$\bar{x}_i = \frac{1}{n} \sum_{l=1}^{n} x_{il}$$

is the mean observation on the $i$th variable and the $(i, j)$ element of the observed correlation matrix $\mathbf{R}$ is $r_{ij} = s_{ij}/s_{ii}^{1/2}s_{jj}^{1/2}$.

### *2.2 Multiplicative Models for Contingency Tables.*

Birch [1963], on the other hand, used log-linear models to distinguish between different hypotheses in a contingency table. A fixed number of observations, $n$, can be classified by combinations of $p$ variables where variable $j$ has $i_j = 1, \cdots, I_j$ categories, $j = 1, 2, \cdots, p$.

Let $n_{i_1, i_2, \cdots, i_p}$ be a typical observed cell count in a $\prod_{i=1,\ldots,p} I_i$ contingency table. For log-linear models it is assumed that the cell count $N_{i_1, i_2, \ldots, i_p}$ follows a multinomial distribution,

$$P(N_{i_1, i_2, \cdots, i_p} = n_{i_1, \cdots, i_p}) \propto \prod_{i_1, \cdots, i_p} (m_{i_1, i_2, \cdots, i_p})^{n_{i_1, i_2 \ldots i_1}} \tag{3}$$

with $m_{i_1, i_2, \cdots, i_p}$ as the expected cell value in cell $(i_1, i_2, \cdots, i_p)$ of the $p$-dimensional contingency table. For the log-linear model formulation the logarithm of the expected cell value is regarded as the sum of several parameters, the "interaction terms," denoted here by "$u$". For example, consider the so called saturated model for three variables with $(i_1, i_2, i_3) = (i, j, k)$, $(I_1, I_2, I_3 = I, J, K)$:

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \tag{4}$$

with

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_k u_{3(k)} = 0$$

$$\sum_j u_{12(ij)} = \sum_k u_{13(ik)} = 0 \qquad \forall\ i$$

$$\sum_i u_{12(ij)} = \sum_k u_{23(jk)} = 0 \qquad \forall\ j$$

$$\sum_j u_{23(jk)} = \sum_i u_{13(ik)} = 0 \qquad \forall\ k$$

and

$$\sum_i u_{123(ijk)} = 0 \qquad \forall\ ij$$

$$\sum_j u_{123(ijk)} = 0 \qquad \forall\ ik$$

$$\sum_k u_{123(ijk)} = 0 \qquad \forall\ jk.$$

The different possible unsaturated log-linear models are the result of restricting some of the "interactions," the $u$-parameters, to equal zero. The multiplicative models are unsaturated models for which the expected cell values can be computed multiplicatively from certain marginal tables and for which interpretations in terms of independencies exist. For an extensive discussion see Goodman [1970].

*2.3 Tests for Different Patterns.*

Likelihood ratio-tests, which are known to follow asymptotically a $\chi^2$-distribution, can be used to test whether a hypothesized model or pattern fits the data. The likelihood function evaluated at the maximum likelihood estimates can be written as

$$\ln L \propto -\frac{n}{2} \ln |\hat{\mathbf{P}}|, \tag{5}$$

and

$$\ln L \propto \sum_{i_1, i_2, \cdots, i_p} (n_{i_1, i_2, \cdots, i_p}) \ln \hat{m}_{i_1, i_2, \cdots, i_p} \tag{6}$$

in the multivariate normal and in the multinomial distributions, respectively. Let $L_1$, $L_2$ denote the likelihood for two different patterns, then the likelihood ratio-test statistic,

$-2\ln(L_1/L_2)$, measures the change in fit. These statistics are easy to compute whenever the likelihood can be factorized. We show below how these factorizations can be derived in the contingency table context and for covariance selection models.

## 3. PATTERNS FOR THREE VARIABLES

Three variables give rise to three conceptually distinct patterns characterized by zero partial associations: either one, two or all three variable pairs are conditionally independent.

### *3.1 Multiplicative Models in Contingency Tables.*

Let $m_{ijk}$ denote the expected cell count for the $(i, j, k)$th cell. The requirement of exactly one zero partial association can be expressed either through its effect on the expected cell count $m_{ijk}$, or through its effect on the log-linear model, that is, on the parametrization of $\ln m_{ijk}$. Suppose that variables 2 and 3 are independent given variable 1, then the expected cell count is

$$m_{ijk} = m_{ij.}m_{i.k}/m_{i..} \qquad \forall\ ijk \tag{7}$$

with $m_{ij.} = \sum_k m_{ijk}$, $m_{i.k} = \sum_j m_{ijk}$ as the two-dimensional marginal tables of variables 1, 2 and 1, 3, respectively, and $m_{i..} = \sum_{jk} m_{ijk}$ as the one-dimensional table of variable 1. As Birch has shown

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} \qquad \forall\ ijk \tag{8}$$

is equivalent to (7) where the interaction terms, or $u$-parameters, satisfy the restriction as described for the saturated model in (4). From (3) and (8) it can be shown that the minimal sufficient statistics for the parameters of this pattern are the observed marginal tables of variables 1, 2 and 1, 3. Therefore, the notation 12/13 is being used here.

In analogy with analysis of variance ideas pattern 12/13 has been interpreted as a model with no three-factor-interaction and no two-factor-interaction between variables 2 and 3, that is with $u_{23(jk)} = u_{123(ijk)} = 0$ in the saturated model (3). We emphasize—in analogy to covariance selection models—that it is a pattern with zero partial association of the variable pair (2, 3).

A less complex pattern is defined by

$$m_{ijk} = m_{ij.}m_{..k}/m_{...} \qquad \forall\ ijk, \tag{9}$$

or by

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} \qquad \forall\ ijk.$$

The notation for this model is 12/3. The familiar interpretation is that variable 3 is independent of variables 1 and 2 together, or that it is a pattern with only one two-factor-interaction, $u_{12(ij)}$. We, however, stress that 12/3 is a model with two zero partial associations for (2, 3) and (1, 3). This becomes evident after noticing that the conditions $u_{23(jk)} = u_{123(ijk)} = 0$ and $u_{13(ik)} = u_{123(ijk)} = 0$ are both satisfied.

Finally, for model 1/2/3 the one-dimensional marginal tables of variables 1, 2 and 3 are the minimal sufficient statistics, and

$$m_{ijk} = m_{i..}m_{.j.}m_{..k}/m_{...}^{\ 2} \qquad \forall\ ijk \tag{10}$$

or

$$\ln m_{ijk} = u + u_{1(i)}u_{2(k)} + u_{3(j)}\ .$$

The model can be interpreted either as the pattern with mutually independent variables, as the pattern with no two-factor-interactions or as the model with exactly three zero partial associations.

*3.2 Covariance Selection Models.*

We shall now identify the different patterns in a correlation matrix. Let $\mathbf{R}$ be the symmetric observed correlation matrix $[r_{ij}]$ with determinant $D_{123}$ and inverse

$$\mathbf{R}^{-1} = \frac{1}{D_{123}} \begin{bmatrix} 1 - r_{23}^2 & -(r_{12} - r_{23}r_{13}) & -(r_{13} - r_{12}r_{23}) \\ & 1 - r_{13}^2 & -(r_{23} - r_{12}r_{13}) \\ & & 1 - r_{12}^2 \end{bmatrix}. \tag{11}$$

The inverse can be expressed in terms of partial correlation coefficients $r_{ij.k}$ and determinants of submatrices $D_{ij}$ where

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{(1 - r_{ik}^2)^{1/2}(1 - r_{jk}^2)^{1/2}} \tag{12}$$

and

$$D_{ij} = 1 - r_{ij}^2.$$

Thus,

$$\mathbf{R}^{-1} = \frac{1}{D_{123}} \begin{bmatrix} D_{23} & -r_{12.3}(D_{13}D_{23})^{1/2} & -r_{13.2}(D_{12}D_{23})^{1/2} \\ & D_{13} & -r_{23.1}(D_{12}D_{13})^{1/2} \\ & & D_{12} \end{bmatrix}. \tag{13}$$

From (1), (13) and an analogous expression for the expected inverse correlation matrix $\mathbf{P}^{-1}$ it follows that the "concentration" $\rho^{ij}$ is simply a multiple of the partial correlation coefficient $\rho_{ij.k}$. Thus, with the assumption of a positive definite covariance matrix, zero concentrations are equivalent to zero partial correlations, and therefore distinct covariance selection models can be interpreted as patterns requiring zero partial associations. For example, consider again model 12/13, the case of zero partial association between variables 2 and 3, that is the pattern with $\rho^{23} = 0$. This implies $\rho_{23.1} = 0$, or $\rho_{23} = \rho_{12}\rho_{13}$, and as estimated correlations

$$\hat{\mathbf{P}} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ & 1 & r_{12}r_{13} \\ & & 1 \end{bmatrix}. \tag{14}$$

Furthermore, it can be seen either directly or from (15),

$$D_{123} = D_{ij}D_{ik}(1 - r_{jk.i}^2); \qquad i \neq j \neq k, \tag{15}$$

that the determinant of $\hat{\mathbf{P}}$ in (14) is

$$|\hat{\mathbf{P}}| = D_{12}D_{13}. \tag{16}$$

When we require $\rho_{13.2} = \rho_{23.1} = 0$, we have model 12/3 and both $\rho_{13} = \rho_{12}\rho_{23}$ and $\rho_{23} = \rho_{13}\rho_{12}$ have to hold. This can be satisfied only if $\rho_{13} = \rho_{23} = 0$; therefore

$$\hat{\mathbf{P}} = \begin{bmatrix} 1 & r_{12} & 0 \\ & 1 & 0 \\ & & 1 \end{bmatrix}; \qquad |\hat{\mathbf{P}}| = D_{12} . \tag{17}$$

Similarly, for model 1/2/3 we require $\rho_{12.3} = \rho_{13.2} = \rho_{23.1} = 0$ and, as the estimated correlation matrix, simply the identity matrix.

We have shown how patterns with zero partial associations can be regarded as multiplicative models for contingency tables and as covariance selection models as well. Thus, similar interpretations are possible whether qualitative or quantitative variables are being studied. Similarities in the likelihood ratio tests based on (5) and (6) also become obvious.

### *3.3 Analogies in Test Statistics.*

Consider, for instance, the test against the saturated model for pattern 12/13 or, equivalently, the test for zero partial association of variable pair (2, 3). From (3), (6) and (7) it follows that the likelihood ratio-test on $I \times (J - 1) \times (K - 1)$ degrees of freedom is

$$2\{(\sum_{ijk} n_{ijk} \ln n_{ijk}) - [(\sum_{ij} n_{ij.} \ln m_{ij.}) + (\sum_{ik} n_{i.k} \ln n_{i.k}) - (\sum_{i} n_{i..} \ln n_{i..})]\}. \tag{18}$$

On the other hand, for a correlation matrix it follows from (1), (5) and (14) that the likelihood ratio test on one degree of freedom is

$$-n\{\ln D_{123} - [\ln D_{12} + \ln D_{13} - \ln D_1]\} \tag{19}$$

with $D_1$ as the determinant of the correlation "matrix" with one element, that is with $\ln D_1 = 0$.

The analogies for three variables are summarized in Table 1.

### *3.4 Analogies in Implied Marginal Associations.*

In the case of a correlation matrix, we obtained for each pattern explicit expressions of the implied marginal correlations. It can be expected that corresponding patterns in a

TABLE 1

SUMMARY TABLE FOR THREE VARIABLES

| Type of pattern | | | Examples | | | |
|---|---|---|---|---|---|---|
| Case | Number of patterns | Number of conditionally independent variable pairs | Conditionally independent variable pairs | Notation for the pattern | Log-likelihood evaluated at the max. likelihood estimates | Implied marginal associations |
| a | $\binom{3}{1}$ | 1 | (2,3) | 12/13 | $T^*_{12}+T_{13}-T_1$ | $\rho_{23} = \rho_{12}\rho_{13}$ |
| b | $\binom{3}{2}$ | 2 | (2,3)(1,3) | 12/3 | $T_{12}+T_3-T_0$ | $\rho_{23} = \rho_{13} = 0$ |
| c | $\binom{3}{3}$ | 3 | (2,3)(1,3)(1,2) | 1/2/3 | $T_1+T_2+T_3-2T_0$ | $\rho_{23} = \rho_{13} = \rho_{12} = 0$ |

* Definition of symbols in Table 1:

| symbol | contingency table | covariance matrix | symbol | contingency table | covariance matrix |
|---|---|---|---|---|---|
| $T_{12}$ | $\sum_{ij} n_{ij.} \ln n_{ij.}$ | $\ln D_{12}$ | $T_2$ | $\sum_j n_{.j.} \ln n_{.j.}$ | $\ln D_2$ |
| $T_{13}$ | $\sum_{ik} n_{i.k} \ln n_{i.k}$ | $\ln D_{13}$ | $T_1$ | $\sum_i n_{i..} \ln n_{i..}$ | $\ln D_1$ |
| $T_3$ | $\sum_k n_{..k} \ln n_{..k}$ | $\ln D_3$ | $T_0$ | $n_{...} \ln n_{...}$ | $\ln 1$ |

contingency table have similar implications for the marginal associations. For attribute data a "correlation coefficient" has been defined (see Kendall and Stuart, [1961]) as

$$g_{23} = \left(\frac{\chi^2}{n}\right)^{1/2} = \left(\sum_{jk} \frac{(n_{jk}n_{..} - n_{j.}n_{.k})^2}{n_{j.}n_{.k}n_{..}^2}\right)^{1/2}. \tag{20}$$

This coefficient is computed from an observed two-dimensional contingency table. A coefficient $\hat{\gamma}$ as implied by the maximum likelihood estimates $\hat{m}_{ijk}$ can be defined similarly. Suppose, for instance, that pattern 12/13 describes a $2^3$ table, then $\hat{m}_{.jk} = \sum_i n_{ij.}n_{i.k}/n_{i..}$, and it can be shown that

$$\hat{\gamma}_{23} = \left(\sum_{jk} \frac{(\hat{m}_{.jk}n_{..} - n_{.j.}n_{..k})^2}{n_{.j.}n_{..k}n_{..}^2}\right)^{1/2} = g_{12}g_{13}\,. \tag{21}$$

Furthermore, pattern 12/3 implies $\hat{\gamma}_{13} = \hat{\gamma}_{23} = 0$, and pattern 1/2/3 implies $\hat{\gamma}_{12} = \hat{\gamma}_{13} = \hat{\gamma}_{23} = 0$. These results are all analogous to those obtained for the correlation matrix. Whenever the variables in a contingency table are not dichotomous, we conjecture that a relation similar to (21) holds, that is $\hat{\gamma}_{23} \leq g_{12}g_{13}$.

The data analyst can use this information about implied coefficients of association to decide for a given constellation of observed marginal associations, whether a hypothesized pattern is likely to be consistent with the data or not. If, for instance, the sample size is large and we observe $r_{12} = r_{13} = .8$ and $r_{23} = .63$ (or $g_{12} = g_{13} = .8$ and $g_{23} = .63$), then only pattern 12/13 will fit the data. All other patterns can be ruled out because the implied correlations do not agree well with the observed correlations. Pattern 12/23, for instance, implies $\hat{\gamma}_{13} = r_{12}r_{23} = .5$ while the observed correlation is .8. Similarly, each of the patterns 12/3, 13/2, 23/1 or 1/2/3 implies zero marginal associations while all observed correlations are large.

The main difference between patterns in a covariance matrix and patterns in a contingency table is that the (partial) correlations—by assumption—have to be the same for all values of the other variables while in a contingency table partial associations may differ within different layers of the other variables. Once the data are reduced to a correlation matrix, changes in partial association with different values of a third variable can no longer be detected. On the other hand, in a contingency table tests are available (e.g. Bishop [1971]) to determine whether partial associations are the same or not.

One effect of this difference is that in a contingency table a given constellation of marginal associations of variable pairs can look as if it were produced by a simple pattern although it is the result of changing partial associations. It is only in a correlation matrix that there exists a one-to-one relationship between implied coefficients of association and the described patterns of association.

## 4. PATTERNS FOR FOUR VARIABLES

With four variables, there are $\binom{4}{2} = 6$ variable pairs that can have zero partial association but the number of conceptually distinct patterns is ten. The patterns, or models, differ not only by the number of zero partial associations but also by how often one given variable is partially unrelated to the other variables.

### *4.1 Ten Patterns with Different Interpretations.*

We describe one version for each of the ten distinct patterns in some detail:

a) There are $\binom{6}{1}$ possible versions with exactly one zero partial association. If, for example, variables 3 and 4 are independent given variables 1 and 2 jointly, then the model is denoted as 123/124.

b) If two variable pairs are to have zero partial association, $\binom{6}{2}$ possible versions exist, and either one variable is involved twice or not.

*Case* $b_1$ :
If, for instance, the variable pairs (3, 4) and (2, 4) are conditionally unrelated, then model 123/14 results. The familiar interpretation for this pattern is that variable 4 is independent of variables 23 conditional on variable 1.

*Case* $b_2$ :
If, instead, the variable pairs (1, 2) and (3, 4) are partially unrelated, the model is denoted as 13/14/23/24. The two variable pairs with zero partial association have no variable in common and the pattern is not a multiplicative one.

c) With three zero partial associations there are $\binom{6}{3}$ versions and three distinct patterns: one variable is involved three times in the zero partial associations (e.g. model 123/4), two variables appear twice among the partially unrelated pairs (e.g. 13/14/23) or three variables are involved twice. The last requirement is equivalent to having one variable with no zero partial association (e.g. model 12/13/14).

*Case* $c_1$ :
For pattern 123/4, the partially unrelated variable pairs are (1, 4), (2, 4) and (3, 4), and variable 4 is completely independent of variables 1, 2, and 4 together, as well.

*Case* $c_2$ :
Let the partially unrelated variable pairs be (1, 2), (2, 4) and (3, 4), then the pattern is denoted as 13/14/23 and variable 4 is independent of 2 and 3 conditional on variable 1 while variable 2 is independent of 1 and 4 conditional on variable 3.

*Case* $c_3$ :
If only variable 1 is partially related to all of the other three variables, then the pairs (2, 3), (2, 4) and (3, 4) are the three pairs with zero partial association. The resulting model is denoted as 12/13/14. It implies independence of variables 2, 3 and 4 conditional on variable 1.

d) With exactly four zero partial associations there are (because of symmetry reasons, as under b) two distinct types of patterns within the $\binom{6}{4}$ versions.

*Case* $d_1$ :
For the zero partial associations of pairs (1, 2), (1, 3), (2, 4) and (3, 4) we get variables 1 and 4 being independent of 2 and 3 jointly. The pattern is denoted as 14/23.

*Case* $d_2$ :
If the pairs (1, 4), (2, 3), (2, 4) and (3, 4) are partially unrelated, the resulting pattern is 12/13/4. Here, variable 4 is independent of 1, 2 and 3; furthermore, variables 2 and 3 are conditionally independent given variable 1.

e) With five variable pairs having zero partial associations, there are $\binom{6}{5}$ versions. If only variables 1 and 2 are partially and marginally related, the pattern is 12/3/4.

f) Finally, there is only one pattern with exactly six partially unrelated variable pairs, the pattern with four mutually independent variables: 1/2/3/4.
The ten different patterns and their implied marginal associations are listed in Table 2, together with their factorization of the likelihood for the multivariate normal and for the multinomial distribution.

### *4.2 Pattern 123/4 in a Contingency Table.*

Take as an example Case $b_1$, that is pattern 123/14, with exactly two zero partial associations, namely for variable pairs (3, 4) and (2, 4). In the contingency table context the zero partial association is achieved if in the corresponding log-linear model the two-factor-interactions and all higher order interactions involving the variable pair are zero. More precisely, compared with the saturated model with $(i_1, i_2, i_3, i_4) = (i, j, k, l)$ we require for pattern 123/14 that

$$\begin{aligned} u_{34(kl)} &= u_{134(ikl)} = u_{234(jkl)} = u_{1234(ijkl)} = 0 \\ &\text{and} \\ u_{24(jl)} &= u_{124(ijl)} = u_{234(jkl)} = u_{1234(ijkl)} = 0. \end{aligned} \tag{22}$$

Under those assumptions the log-linear model is

$$\ln m_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} + u_{23(jk)} + u_{123(ijk)}\,. \tag{23}$$

An equivalent expression in terms of the expected cell counts is

TABLE 2

SUMMARY TABLE FOR FOUR VARIABLES

| Type of pattern | | Examples | | | |
|---|---|---|---|---|---|
| Case | No. of pattern | Conditionally independent variable pairs | Notation for the pattern | Log-likelihood evaluated at the max. likelihood estimates | Implied marginal associations in a correlation matrix |
| a | 6 | (3,4) | 123/124 | $T^*_{123}+T_{124}-T_{12}$ | $\varrho_{34}=\dfrac{(\varrho_{13}\varrho_{14}+\varrho_{23}\varrho_{24})-\varrho_{12}(\varrho_{13}\varrho_{24}+\varrho_{14}\varrho_{23})}{1-\varrho^2_{12}}$ |
| $b_1$ | 12 | (3,4)(2,4) | 123/14 | $T_{123}+T_{14}-T_1$ | $\varrho_{34}=\varrho_{13}\varrho_{14}$ ; $\varrho_{24}=\varrho_{12}\varrho_{14}$ |
| $b_2$ | 3 | (3,4)(1,2) | 13/14/23/24 | iterative solution needed | |
| $c_1$ | 4 | (3,4)(2,4)(1,4) | 123/4 | $T_{123}+T_4-T_0$ | $\varrho_{34}=\varrho_{24}=\varrho_{14}=0$ |
| $c_2$ | 12 | (3,4)(2,4)(1,2) | 13/14/23 | $T_{13}+T_{14}+T_{23}-(T_1+T_3)$ | $\varrho_{34}=\varrho_{13}\varrho_{14}$; $\varrho_{24}=\varrho_{13}\varrho_{14}\varrho_{23}$; $\varrho_{12}=\varrho_{13}\varrho_{23}$ |
| $c_3$ | 4 | (3,4)(2,4)(2,3) | 12/13/14 | $T_{12}+T_{13}+T_{14}-2T_1$ | $\varrho_{34}=\varrho_{13}\varrho_{14}$; $\varrho_{24}=\varrho_{12}\varrho_{14}$; $\varrho_{23}=\varrho_{12}\varrho_{13}$ |
| $d_1$ | 3 | (3,4)(2,4)(1,2)(1,3) | 14/23 | $T_{14}+T_{23}-T_0$ | $\varrho_{34}=\varrho_{24}=\varrho_{12}=\varrho_{13}=0$ |
| $d_2$ | 12 | (3,4)(2,4)(1,4)(2,3) | 12/13/4 | $T_{12}+T_{13}+T_4-(T_1+T_0)$ | $\varrho_{34}=\varrho_{24}=\varrho_{14}=0$; $\varrho_{23}=\varrho_{12}\varrho_{13}$ |
| e | 6 | all except (1,2) | 12/3/4 | $T_{12}+T_3+T_4-2T_0$ | $\varrho_{34}=\varrho_{24}=\varrho_{23}=\varrho_{14}=\varrho_{13}=0$ |
| f | 1 | all | 1/2/3/4 | $T_1+T_2+T_3+T_4-3T_0$ | $\varrho_{34}=\varrho_{24}=\varrho_{23}=\varrho_{14}=\varrho_{13}=\varrho_{12}=0$ |

*Definition of selected symbols in Table 2:

| symbol | contingency table | covariance matrix | symbol | contingency table | covariance matrix |
|---|---|---|---|---|---|
| $T_{123}$ | $\sum_{ijk} n_{ijk.} \ln n_{ijk.}$ | $\ln D_{123}$ | $T_{14}$ | $\sum_{il} n_{i..l} \ln n_{i..l}$ | $\ln D_{14}$ |
| | $\sum_{i} n_{i...} \ln n_{i...}$ | $\ln D_1$ | $T_0$ | $n_{....} \ln n_{....}$ | $\ln 1$ |

$$m_{ijkl} = \frac{m_{ijk.}m_{i..l}}{m_{i...}}. \tag{24}$$

The minimal sufficient statistics for the $u$-parameters in (23) are the marginal tables of variables 1, 2, 3 and 14, that is, $n_{ijk}$ and $n_{i..l}$ , respectively. This explains the notation 123/14. The pattern can be interpreted in either of three ways: If from (23) it is a model with four two-factor interactions and one three factor interaction, if from (24) it is the model with variables 2 and 3 being independent of variable 4 within each layer of variable 1, if from (22) it is a pattern with two zero partial associations.

The likelihood can be factorized into the marginal tables of variables 1, 2, 3, of variables 1, 4 and of variable 1. From (6) and (24)

$$\ln L \propto (\sum_{ijk} n_{ijk.} \ln n_{ijk.}) + (\sum_{il} n_{i..l} \ln n_{i..l}) - (\sum_{i} n_{i...} \ln n_{i...}). \tag{25}$$

Further, it can be shown that in a $2^4$-table the marginal associations implied by pattern 123/14 are

$$\hat{\gamma}_{24} = g_{14}g_{24} \quad \text{and} \quad \hat{\gamma}_{34} = g_{13}g_{14} . \tag{26}$$

*4.3 Pattern 123/14 in a Correlation Matrix.*

On the other hand, in a correlation matrix a variable pair has zero partial association if its partial correlation coefficient is equal to zero. And, since an element of an inverse covariance matrix, the "concentration" is a multiple of a partial correlation coefficient, each pattern with zero partial correlations is a covariance selection model as well.

In general, with four variables we have

$$r^{ij} = \frac{-r_{ij.kl}(D_{ikl}D_{jkl})^{1/2}}{D_{ijkl}} \tag{27}$$

and

$$r_{ij.kl} = \frac{r_{ij.k} - r_{il.k}r_{jl.k}}{(1 - r_{il.k}^2)^{1/2}(1 - r_{jl.k}^2)^{1/2}} = \frac{r_{ij.k} - r_{il.k}r_{jl.k}}{(D_{ikl}/D_{kl}D_{ik})^{1/2}(D_{jkl}/D_{kl}D_{jk})^{1/2}} \tag{28}$$

and the determinant $D_{1234}$ can be expressed in six different ways as

$$D_{1234} = \frac{D_{134}D_{234}}{D_{34}}(1 - r_{12.34}^2), = \frac{D_{124}D_{234}}{D_{24}}(1 - r_{13.24}^2), \cdots ,$$

$$= \frac{D_{123}D_{124}}{D_{12}}(1 - r_{34.12}^2). \tag{29}$$

When we require $\rho^{34} = \rho^{24} = 0$ or, equivalently, $\rho_{34.12} = \rho_{24.13} = 0$, the resulting pattern is (analogy to the contingency table context) denoted as 123/14. The two zero partial correlations imply that $\rho_{34.1} = \rho_{23.1}\rho_{24.1}$ and $\rho_{24.1} = \rho_{23.1}\rho_{34.1}$ , which in turn can only be satisfied if $\rho_{34.1} = \rho_{24.1} = 0$, that is if $\rho_{34} = \rho_{13}\rho_{14}$ and $\rho_{24} = \rho_{12}\rho_{14}$. Therefore, the maximum likelihood estimates of the marginal correlations assume a rather simple form:

$$\hat{\rho}_{34} = r_{13}r_{14} , \qquad \hat{\rho}_{24} = r_{12}r_{14}$$

and (30)

$$\hat{\rho}_{ij} = r_{ij} \quad \text{if} \quad \begin{cases} ij \neq 34 \\ ij \neq 24 \end{cases}.$$

The determinant of this estimated correlation matrix can be computed from submatrices involving variables 1, 2, 3 and variables 1, 4 as

$$|\hat{\mathbf{P}}| = D_{123}D_{14}\,. \tag{31}$$

Consequently, the likelihood function evaluated at the maximum-likelihood estimates is from (1) and (31):

$$\ln L \propto -\frac{n}{2}(\ln D_{123} + \ln D_{14} - \ln D_1), \tag{32}$$

and the likelihood ratio test—as in the case of the multiplicative model in contingency tables—can be computed without evaluating the maximum likelihood estimates explicitly. The advantage of this fact increases with the number of variables being studied.

## 5. GENERALIZATIONS TO $p$ VARIABLES

We now indicate how the analogies in the interpretation of patterns, test statistics and implied marginal associations look like for $p$ variables.

### *5.1 The Pattern With One Zero Partial Association.*

To simplify notation, let $K$ denote the set with $p - 2$ indices from 1 to $p$ except for $i$ and $j$,

$$K = \{n \mid n = 1, \cdots, p, n \neq i, n \neq j\}. \tag{33}$$

Similarly, let $iK = \{n \mid n = 1, \cdots, p; n \neq j\}$, $ijK = \{n \mid n = 1, \cdots, p\}$. Then, the inverse elements $r^{ij}$, $r^{ii}$, $r^{jj}$ of the correlation matrix $\mathbf{R}$ can be expressed in terms of the partial correlation coefficient $r_{ij.K}$, and in terms of determinants $D_{iK}$, $D_{jK}$, $D_{ijK}$:

$$r^{ij} = \frac{-r_{ij.K}(D_{iK}D_{jK})^{1/2}}{D_{ijK}}; \quad r^{ii} = \frac{D_{jK}}{D_{ijK}}; \quad r^{jj} = \frac{D_{iK}}{D_{ijK}}. \tag{34}$$

These relations, in turn, allow the partial correlation coefficient in terms of inverse elements to be written as

$$r_{ij.K} = \frac{-r^{ij}}{(r^{ii})^{1/2}(r^{jj})^{1/2}}. \tag{35}$$

Furthermore, the determinant of the $p$-dimensional correlation matrix can be computed as

$$D_{ijK} = \frac{D_{iK}D_{jK}}{D_K}(1 - r_{ij.K}^2). \tag{36}$$

Thus, it follows that the likelihood $|\hat{\mathbf{P}}|$ for a covariance selection model with exactly one zero concentration ($\sigma^{ij} = 0$) or, equivalently, with one zero partial correlation coefficient ($\rho_{ij.K} = 0$) is:

$$L \propto |\hat{\mathbf{P}}|^{-n/2} = \left(\frac{D_{iK}D_{jK}}{D_K}\right)^{-n/2}, \tag{37}$$

that is, it may be computed from determinants of submatrices without evaluating the maximum likelihood estimate $\hat{\rho}_{ij}$.

If there is exactly one zero partial association for variable pair $(i, j)$, a simple matrix formula for $\hat{\rho}_{ij}$ can nevertheless be given in terms of the observed correlations. Let the $p \times p$ matrix $\mathbf{R}$ be partitioned as

$$\mathbf{R} = \left[\begin{array}{cc|c} 1 & r_{ij} & \mathbf{R}_{iK} \\ r_{ji} & 1 & \mathbf{R}_{jK} \\ \hline \mathbf{R}_{iK}^T & \mathbf{R}_{jK}^T & \mathbf{R}_{KK} \end{array}\right] \begin{array}{l} \}\, 2 \\ \\ \}\, p-2 \end{array}, \qquad \underbrace{\quad}_{2} \quad \underbrace{\quad}_{p-2} \tag{38}$$

then the observed partial correlation for $(i, j)$ is

$$r_{ij.K} = \frac{r_{ij} - \mathbf{R}_{iK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{jK}^T}{(1 - \mathbf{R}_{iK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{iK}^T)^{1/2}(1 - \mathbf{R}_{jK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{jK}^T)^{1/2}}. \tag{39}$$

*Proof*: From

$$\begin{bmatrix} 1 - \mathbf{R}_{iK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{iK}^T & r_{ij} - \mathbf{R}_{iK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{jK}^T \\ & 1 - \mathbf{R}_{jK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{jK}^T \end{bmatrix} = \begin{bmatrix} r^{ii} & r^{ij} \\ & r^{jj} \end{bmatrix}^{-1}$$

and from (34) and (36) we get

$$\begin{bmatrix} r^{ii} & r^{ij} \\ & r^{jj} \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{D_{iK}}{D_K} & \dfrac{r_{ij.K}(D_{iK}D_{jK})^{1/2}}{D_K} \\ & \dfrac{D_{jK}}{D_K} \end{bmatrix},$$

then relation (39) follows. A few additional arguments lead from (39) to the maximum-likelihood estimate as

$$\hat{\rho}_{ij} = \mathbf{R}_{iK}\mathbf{R}_{KK}^{-1}\mathbf{R}_{jK}^T, \quad \text{or to}$$
$$\hat{\rho}_{ij} = r_{ij} - \frac{r^{ij}}{r^{ii}r^{jj} - r^{ij}r^{ij}}. \tag{40}$$

In the contingency table context the most complex of the multiplicative models with $(I_i - 1)(I_j - 1)\prod_{l \epsilon K} I_l$ degrees of freedom is the pattern with exactly one zero partial association: For all layers of the variables in $K$, variables $i$ and $j$ are independent; the expected and estimated cell counts satisfy:

$$m_{ijK} = \frac{m_{.jK}m_{i.K}}{m_{..K}}; \qquad \hat{m}_{ijK} = \frac{n_{.jK}n_{i.K}}{n_{..K}}. \tag{41}$$

The log-likelihood can then be computed from observed marginal tables,

$$\ln L \propto (\sum_{jK} n_{.jK} \ln n_{.jK}) + (\sum_{iK} n_{i.K} \ln n_{i.K}) - (\sum_{K} n_{..K} \ln n_{..K}) \tag{42}$$

just as in a correlation matrix where the log-likelihood could be computed from submatrices (see (37)):

$$\ln L \propto -\frac{n}{2}[\ln D_{jK} + \ln D_{iK} - \ln D_K]. \tag{43}$$

Furthermore, in a $2^p$ contingency table the implied marginal association

$$\hat{\gamma}_{ij} = \left(\sum \frac{(\hat{m}_{ij.}n_{...} - n_{i..}n_{j..})^2}{n_{i..}n_{.j.}n_{..}^2}\right)^{1/2}$$

will be similar to the one implied in a correlation matrix (similar to (40)).

*5.2 A Pattern With Several Zero Partial Associations.*

All multiplicative models can be derived and interpreted as patterns with several zero partial associations, be it in the contingency table context or in the correlation matrix context. For instance, suppose we want to test the following hypothesis in a correlation matrix: Variables 1, 2, and 3 are independent after eliminating the influence of variables 4 to $p$. This hypothesis corresponds to pattern $145, \cdots, p/245, \cdots, p/345, \cdots, p$, or to zero partial associations of the pairs (1, 2), (1, 3) and (2, 3). The likelihood for this pattern and thus a likelihood ratio test statistic may be found by a stepwise elimination of partial associations, that is by repeatedly applying a rule like (37)

$$\begin{aligned}
\rho^{12} &= 0 \rightarrow \frac{D_{1345,\cdots,p}D_{2345,\cdots,p}}{D_{345,\cdots,p}} = |\hat{\mathbf{P}}| \\
\rho^{12} = \rho^{13} &= 0 \rightarrow \frac{D_{145,\cdots,p}D_{2345,\cdots,p}}{D_{45,\cdots,p}} = |\hat{\mathbf{P}}| \qquad (44) \\
\rho^{12} = \rho^{13} = \rho^{23} &= 0 \rightarrow \frac{D_{145,\cdots,p}D_{245,\cdots,p}D_{345,\cdots,p}}{(D_{45,\cdots,p})^2} = |\hat{\mathbf{P}}|.
\end{aligned}$$

If the implied marginal associations, the maximum likelihood estimates $\hat{\rho}_{12}$, $\hat{\rho}_{13}$, $\hat{\rho}_{23}$, are of interest, they can be computed as

$$\hat{\rho}_{ij} = \mathbf{R}_{i45,\cdots,p}\mathbf{R}_{45,\cdots,p}^{-1}\mathbf{R}_{j45,\cdots,p}^{T} \quad \text{for} \quad (i, j) = (1, 2); (1, 3); (2, 3). \qquad (45)$$

This follows from (40) and from arguments similar to those described previously for pattern 123/14, that is from

$$\rho^{12} = \rho^{13} = 0 \left\{ \Rightarrow \begin{matrix} \rho_{12.345,\cdots,p} = 0 \\ \rho_{13.245,\cdots,p} = 0 \end{matrix} \right\} \Rightarrow \begin{cases} \rho_{12.45,\cdots,p} = 0 \\ \rho_{13.45,\cdots,p} = 0 \\ \rho_{23.145,\cdots,p} = \rho_{23.45,\cdots,p} . \end{cases} \qquad (46)$$

In a contingency table the estimated cell counts and the test statistic for pattern $145, \cdots, p/245, \cdots, p/345, \cdots, p$ can be obtained stepwise—similarly as in (44)—to be

$$\hat{m}_{i_1i_2,\cdots,i_p} = \frac{(\sum_{i_2,i_3} n_{i_1,\cdots,i_p})(\sum_{i_1,i_3} n_{i_1,\cdots,i_p})(\sum_{i_1,i_2} n_{i_1,\cdots,i_p})}{(\sum_{i_1,i_2,i_3} n_{i_1,\cdots,i_p})^2}. \qquad (47)$$

Furthermore, the marginal associations $\hat{\gamma}_{12}$, $\hat{\gamma}_{13}$, $\hat{\gamma}_{23}$ in a $2^p$ table will be similar to those in (45).

*5.3 Applications.*

Patterns with zero partial associations may be of interest either within the context of confirmative or of explorative types of analysis. Examples for the first case are hypotheses on conditional independence that occur in observational studies if the influence of several confounding variables is suspected. For instance, in studies on changes in blood circulation potential confounding variables are age, weight and length; in studies on prenatal malformations birth order or sex may be important background variables. Explorative model search procedures, on the other hand, may be used in all those situations where a factor analysis or a cluster analysis could be considered; for instance, to study the interrelations of personality traits or of the symptoms of a disease. In a forthcoming paper we shall propose a model search procedure among multiplicative models for quantitative or qualitative variables.

## ACKNOWLEDGMENT

I wish to thank the referees for their very constructive criticism on an earlier version of this paper.

## ANALOGIES ENTRE MODELES MULTIPLICATIFS DANS LES TABLES DE CONTINGENCE ET SELECTION DE COVARIANCE

### RÉSUMÉ

On peut étudier une certaine classe de modèles d'association en ajustant des modèles multiplicatifs à une table de contingence ou en utilisant la sélection de covariance sur une matrice de covariance.

Nous montrons que chaque modèle multiplicatif pour une table de contingence correspond à un modèle particulier de sélection de covariance et nous mettons en évidence les ressemblances qui en résultent pour l'interprétation des modèles, pour les tests statistiques utilisés pour chaque modèle, et pour les associations marginales implicites entre paires de variables.

## REFERENCES

Birch, M. W. [1963]. Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc. 25*, 220–33.

Bishop, Y. M. M. [1967]. *Multi-dimensional contingency tables: cell estimators*. Ph.D. thesis, Dept. of Statist., Harvard Univ.

Bishop, Y. M. M. [1969]. Full contingency tables, logits and split contingency tables. *Biometrics 25*, 383–400.

Bishop, Y. M. M. [1971]. Effects of collapsing multi-dimensional contingency tables. *Biometrics 27*, 545–62.

Darroch, J. N. [1962]. Interactions in multifactor contingency tables. *J. R. Statist. Soc. 24*, 251–63.

Dempster, A. P. D. [1969]. *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Massachusetts.

Dempster, A. P. D. [1972]. Covariance selection. *Biometrics 28*, 157–75.

Goodman, L. A. [1970]. The multivariate analysis of qualitative data: interactions among multiple classifications. *J. Amer. Statist. Assoc. 65*, 226–56.

Goodman, L. A. [1973]. The analysis of multi-dimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika 60*, 179–92.

Kendall, M. G. and Stuart, A. [1961]. *The Advanced Theory of Statistics*. C. Griffin, London.

Lindley, D. V. [1964]. The Bayesian analysis of contingency tables. *Ann. Math. Statist. 35*, 1622–43.

Roy, S. N. and Kastenbaum, M. A. [1956]. On the hypothesis of no interaction in a multiway contingency table. *Ann. Math. Statist. 27*, 749–57.

Wermuth, N. [1976]. Model search among multiplicative models. *Biometrics 32*, (in press).

Wermuth, N, Hodapp, V. and Weyer, G. [1976]. Die Methode der Kovarianzselektion als Alternative zur Faktorenanalyse untersucht an persönlickkeitsmerkmalen. *Zeitschriff für experimentelle und angewandte Psychologie* (in press).

*Received February* 1974, *Revised August* 1974

*Key Words:* Log-linear models, Covariance selection, Patterns of association.