Some Results about Decomposable (or Markov-type) Models for Multidimensional Contingency Tables: Distribution of Marginals and Partitioning of Tests

Rolf Sundberg

Reprint from

Scandinavian Journal of Statistics

Vol. 2, No. 2, 1975

Scand J Statist 2: 71–79, 1975

Some Results about Decomposable (or Markov-type) Models for Multidimensional Contingency Tables: Distribution of Marginals and Partitioning of Tests

ROLF SUNDBERG

The Royal Institute of Technology, Stockholm

Received April 1974, revised October 1974

ABSTRACT. **For the so-called decomposable or Markov-type models for contingency tables of arbitrary dimension it is shown that the probability of a minimal set of fitted marginals may be expressed in a closed form, analogous to that of the maximum likelihood estimate. As a consequence a closed form expression can be given for the so-called exact test statistic in a test of a decomposable model. Finally a conjecture by A. H. Andersen is proved, stating that when the models are decomposable the exact test statistic and the likelihood ratio may both be completely factorized into exact test statistics and likelihood ratios, respectively, for testing homogeneity in one-dimensional or independence in two-dimensional contingency tables.**

Key words: contingency table, decomposable model, distribution of marginals, exact test, likelihood ratio, test partitioning, homogeneity, independence

1. Notations and background

In this section is given a brief account of those terms and facts about contingency tables which have been judged necessary for the comprehension of the two later sections. For more exhaustive surveys the reader is referred to Andersen (1974) or Haberman (1974).

By an m-dimensional *contingency table* is understood a set

$$x = \{x_{i_1 \dots i_m};\ i_k = 1, \dots, r_k;\ k = 1, \dots, m\} \tag{1.1}$$

of non-negative integers.

The most important statistical models for contingency tables are of multinomial or Poisson type. When discussing the distribution of x we will assume that x may be regarded as a sample of size $N = \Sigma_{i_1, \dots, i_m} x_{i_1 \dots i_m}$ from a multinomial distribution with probabilities $p_{i_1 \dots i_m}$. If instead the $x_{i_1 \dots i_m}$ were to be regarded as independent Poisson variables, conditioning on the total N would give the multinomial distribution, so we might easily go from the one distribution to the other.

Introducing

$$\lambda = \{\log p_{i_1 \dots i_m}\}, \tag{1.2}$$

the probability distribution of x may be written in the following exponential family form,

$$P(x) = \frac{N!}{\prod\limits_{i_1, \dots, i_m} x_{i_1 \dots i_m}!} e^{\lambda \cdot x}, \tag{1.3}$$

where $\lambda \cdot x$ is the ordinary scalar product,

$$\lambda \cdot x = \sum_{i_1, \dots, i_m} \log p_{i_1 \dots i_m} x_{i_1 \dots i_m}. \tag{1.4}$$

We will only consider *hierarchical* models, as defined by Birch (1963). This means that the cell probabilities are permitted to be log-linearly related in such a way that a suitable set of marginals, usually called the minimal set of fitted marginals, is sufficient for the parameters. For a closer discussion we need some convenient notations.

Let $Z = \{1, 2, \dots, m\}$. For a subset Y of Z we define the Y-marginal x^Y as the set of sums obtained from x by summing over all those indices which do not correspond to an element in Y.

The minimal set of fitted marginals is characterized by a class $\{Y_1, \dots, Y_T\}$ of subsets of Z, such that no member Y_t of the class is a subset of any other member of the class. Except for trivial linear relations, the most trivial being that the total sum N is fixed, $\{x^{Y_1}, \dots, x^{Y_T}\}$ shall be a minimal sufficient statistic in the exponential family distribution of x. Let the corresponding set of canonical parameters (corresponding to λ in (1.3); called interactions) be denoted by $\lambda^{(1)}, \dots, \lambda^{(T)}$. The probability of x can now be written

$$P(x) = \frac{N!}{\sum\limits_{i_1, \dots, i_m} x_{i_1 \dots i_m}!} \exp\left[\sum_{t=1}^{T} \lambda^{(t)} \cdot x^{Y_t}\right]. \tag{1.5}$$

The probability of a particular set of marginals

$x^{Y_1}, \ldots, x^{Y_T}$ is obtained by summing (1.5) over all tables with the correct marginals,

$$P(x^{Y_1}, \ldots, x^{Y_T}) = N! \left(\sum_{\substack{x, \text{ given} \\ \text{marginals}}} \frac{1}{\prod_{i_1, \ldots, i_m} x_{i_1 \ldots i_m}!} \right) \exp \left[\sum_{t=1}^{T} \lambda^{(t)} \cdot x^{Y_t} \right]. \quad (1.6)$$

The main result in section 2 is a closed form expression for the combinatorial sum in (1.6) when the model is decomposable.

The notion *decomposable* was introduced by Haberman (1970), see Andersen (1974) or Haberman (1974). His definition is recursive: The model shall be represented by a single set Y ($T=1$) or be the 'union' of two decomposable models. We need not be more detailed here, since we will use another equivalent criterion. Haberman showed that the class $\{Y_t\}$ can be ordered in such a way that for each t,

$$Y_t \cap (\bigcup_{s<t} Y_s) = Y_t \cap Y_\tau \quad (1.7)$$

for some $\tau = \tau(t) < t$, if and only if the model is decomposable. He used this characterization to derive a closed form expression for the maximum likelihood estimates of the cell probabilities $p_{i_1 \ldots i_m}$, see expression (2.5).

Reversing Haberman's ordering we may formulate the following characterization of decomposability, which will be repeatedly used in the sequel:

Criterion for decomposability. *A model is decomposable if and only if the class $\{Y_t\}$ can be ordered in such a way that each Y_t is composed of one set of elements which are missing in all Y_s for $s>t$ and one set $U_t = Y_t \cap (\bigcup_{s>t} Y_s)$ which is contained in some Y_τ, $\tau = \tau(t) > t$. Furthermore, it is a fact that if such an ordering is possible, a version may be found in which any prescribed set is the last one.*

For a proof, see the references above.

The criterion is used recursively to check decomposability. First find a set in the class $\{Y_t\}$ satisfying the condition on Y_1. Exclude this set Y_1 from the model. It is easily seen that the resulting model is decomposable if and only if the original model was decomposable. Among the remaining sets find a set satisfying the condition on Y_2, eliminate it, and so on as long as possible. The model is decomposable if and only if all sets are eliminated and ordered by this procedure. Furthermore, as mentioned above it should be possible to save any prescribed set to be the last one.

Examples

When $Z = \{1, 2\}$ all models are decomposable. When $Z = \{1, 2, 3\}$ all models are decomposable except the one represented by $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$ (no third order interaction). The sevendimensional model represented by $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{2, 3, 5\}$, $\{1, 3, 6\}$, $\{5, 7\}$ ($T=5$) is decomposable (the criterion works in reverse order, for instance). This latter model has been discussed by Haberman (1970) and Goodman (1971).

From other considerations Goodman (1971) introduced the Markov-type model and gave a recursive elimination procedure which worked and yielded a closed form for the maximum likelihood estimates precisely for these models. Bishop (1971) proposed essentially the same procedure. Goodman's procedure is as follows:

First, eliminate the elements in Z not belonging to any Y_t, i.e. $Z - \bigcup_t Y_t$.

Second, find a set in $\{Y_t\}$, say Y_1, that contains elements which only belong to Y_1, and eliminate these elements. Let U_1 be the set of remaining elements in Y_1, i.e. $U_1 = Y_1 \cap (\bigcup_{s>1} Y_s)$.

Third, define the new model $\{Y_2, \ldots, Y_T\}$ if U_1 is a subset of some Y_τ, $\tau > 1$, $\{U_1, Y_2, \ldots, Y_T\}$ otherwise.

Repeat the second and third steps with the new models generated, as long as possible. If ultimately all elements of Z have been eliminated, Goodman's procedure has worked.

If the model is decomposable our criterion tells that Goodman's procedure can be applied in such a way that each of the sets U_t is a subset of some remaining set Y_τ, hence need not be introduced. It might appear from this as if Goodman's criterion for a Markov-type model is weaker than our criterion for decomposability. They are in fact equivalent, as may be seen in the following way:

If no set Y_1 can be found such that U_1 is a subset of some Y_s, $s>1$, it is easily seen that Goodman's procedure can only be applied to each of those sets, $Y_1, \ldots, Y_t$ say, which contain elements not contained in any other set, and the procedure would fail when the model $\{U_1, \ldots, U_t, Y_{t+1}, \ldots, Y_T\}$ has been arrived at. This ensures the possibility to find the desired set Y_1, and the proof is completed by induction.

Proofs of the equivalence of the two notions decomposable and Markov-type for hierarchical contingency tables may also be found in Andersen (1974) and Haberman (1974).

In sections 2 and 3 so-called exact tests are discussed. By the exact test statistic for testing a hypothesis that specifies a reduction from one model with minimal set $t(x)$ of fitted marginals to another simpler model with minimal set $u(x)$, we mean the conditional probability of t given u under the hypothesis. This conditional probability is equal to the ratio between the respective coefficients in (1.6) for the two models. For testing independence in a 2×2 table, the test is identical with Fisher's exact test in the hyper-geometric distribution. The reduction in this example is from $Y=Z=\{1,2\}$ to $\{Y_1, Y_2\}$, $Y_1=\{1\}$, $Y_2=\{2\}$.

2. A closed form for the distribution of the minimal set of fitted marginals

As described in the preceding section, in order to find a closed form for the distribution of the minimal set of fitted marginals in a hierarchical contingency table we should find a closed form for the sum

$$\sum_{\substack{x,\text{ given}\\ \text{marginals}}} \frac{1}{\prod_{i_1,\dots,i_m} x_{i_1\dots i_m}!} \tag{2.1}$$

where the summation is over all tables x that have a given set of values of the marginals $x^{Y_1}, \dots, x^{Y_T}$. In the theorem below we give such an expression for a decomposable model. The closed form of the distribution of the minimal set of fitted marginals is obtained by inserting this expression in (1.6)

Theorem. *In a decomposable model, represented by* $\{Y_1, \dots, Y_T\}$ *ordered according to the criterion for decomposability, the sum (2.1) may be expressed as*

$$\sum_{\substack{x,\text{ given}\\ \text{marginals}}} \frac{1}{\prod_{i_1,\dots,i_m} x_{i_1\dots i_m}!} = \frac{\prod_{t=1}^{T-1} \prod_{\substack{\text{all } U_t\\ \text{marginals}}} u^{(t)}!}{\prod_{t=1}^{T} \prod_{\substack{\text{all } Y_t\\ \text{marginals}}} y^{(t)}!} \cdot \left(\prod_{k\in Z-\bigcup_t Y_t} r_k\right)^N, \tag{2.2}$$

where $U_t = Y_t \cap (\bigcup_{s>t} Y_s)$, $t=1,\dots,T-1$, *and* $y^{(t)}$ *and* $u^{(t)}$ *represent unspecified* Y_t *and* U_t *marginals, respectively.*

Examples

(*a*) For a two-dimensional contingency table with independence between the classifications the minimal set of fitted marginals is the set of row sums x_{i+} and column sums x_{+j}. Thus $Y_1=\{1\}$, $Y_2=\{2\}$ and U_1 is empty. We obtain the well known result

$$\sum_{\substack{x \text{ given}\\ \{x_{i+},\, x_{+j}\}}} \frac{1}{\prod_{ij} x_{ij}!} = \frac{x_{++}!}{\prod_i x_{i+}! \prod_j x_{+j}!} \tag{2.3}$$

(*b*) As a more complicated example, let us consider the 7-dimensional example given in section 1 and represented by the sets $\{1,2,3\}$, $\{1,2,4\}$, $\{2,3,5\}$, $\{1,3,6\}$, $\{5,7\}$. The reverse ordering satisfies the criterion. We obtain the expression

$$\sum_{\substack{x,\text{ given}\\ \text{marginals}}} \frac{1}{\prod_{i_1,\dots,i_7} x_{i_1\dots i_7}!} = \frac{\prod_{i_5} x_{i_5}^{(5)}!}{\prod_{i_5,i_7} x_{i_5 i_7}^{(5,7)}!} \cdot \frac{\prod_{i_1,i_3} x_{i_1 i_3}^{(1,3)}!}{\prod_{i_1,i_3,i_6} x_{i_1 i_3 i_6}^{(1,3,6)}!}$$
$$\times \frac{\prod_{i_2,i_3} x_{i_2 i_3}^{(2,3)}!}{\prod_{i_2,i_3,i_5} x_{i_2 i_3 i_5}^{(2,3,5)}!} \cdot \frac{\prod_{i_1,i_2} x_{i_1 i_2}^{(1,2)}!}{\prod_{i_1,i_2,i_4} x_{i_1 i_2 i_4}^{(1,2,4)}!} \cdot \frac{1}{\prod_{i_1,i_2,i_3} x_{i_1 i_2 i_3}^{(1,2,3)}!}. \tag{2.4}$$

Comment

The close relations between the expression (2.2) and the closed form for the maximum likelihood estimate of $Np_{i_1\dots i_m}$ under the model should be noted. The latter, see Goodman (1971), Haberman (1970, 1974) or Andersen (1974), is

$$\frac{\prod_{t=1}^{T} y^{(t)}}{\prod_{t=1}^{T-1} u^{(t)}} \cdot \frac{1}{\prod_{k\in Z-\bigcup_t Y_t} r_k}, \tag{2.5}$$

where $y^{(t)}$ and $u^{(t)}$ stand for the marginals consistent with $(i_1, \dots, i_m)$.

Proof. We first explain the rightmost factor in (2.2). The easiest way to understand this factor is to think of the classification of N distinguishable individuals. The number of classifications which yield a particular table x is

$$\frac{N!}{\prod_{i_1,\dots,i_m} x_{i_1\dots i_m}!}, \tag{2.6}$$

and the sum (2.1) is the number of classifications which would give the marginals $x^{Y_1}, \dots, x^{Y_T}$. The number of classifications which would give the marginal table $x^{\bigcup_t Y_t}$ is the rightmost factor in (2.2) multiplied by the expression corresponding to (2.6) for the marginal table $x^{\bigcup_t Y_t}$, since the indices corresponding to $Z-\bigcup_t Y_t$ may be chosen quite arbitrarily for each individual. Thus, after having put aside the rightmost factor in (2.2), we may start anew in the marginal table $x^{\bigcup_t Y_t}$, and we are justified in assuming $\bigcup_t Y_t = Z$ from now on.

The idea of the following proof is to show that the sum (2.1) for varying values of the marginals will appear as the coefficients in an expansion of a particular exponential function, which may be called the generating function, and by expanding this function in a different way to obtain alternative expressions for the coefficients. For this purpose we need to introduce as many variables as there are marginals in the minimal set of fitted marginals. Let

I_t be the set of indices of the Y_t marginals, so that $x^{Y_t} = \{x_j^{Y_t}, j \in I_t\}$. Corresponding to the marginal $x_j^{Y_t}$ we introduce the variable $\alpha_j^{(t)}$. It should always be understood below that the indices from the I_t's are consistent with each other and with the present value of $(i_1, ..., i_m)$. Ordinary expansion in power series gives

$$\exp\left[\sum_{i_1,...,i_m} \prod_{t=1}^{T} \alpha_{j_t}^{(t)}\right]$$
$$= \prod_{i_1,...,i_m} \exp\left[\prod_{t=1}^{T} \alpha_{j_t}^{(t)}\right]$$
$$= \prod_{i_1,...,i_m} \sum_{x_{i_1...i_m} \geq 0} \frac{1}{x_{i_1...i_m}!}\left(\prod_{t=1}^{T} \alpha_{j_t}^{(t)}\right)^{x_{i_1...i_m}}$$
$$= \sum_{x} \prod_{i_1,...,i_m} \left(\frac{1}{x_{i_1...i_m}!} \prod_{t=1}^{T} \alpha_{j_t}^{(t)x_{i_1...i_m}}\right)$$
$$= \sum_{x^{Y_1},...,x^{Y_T}} \prod_{t=1}^{T} \prod_{j_t \in I_t} \alpha_{j_t}^{(t)x_{j_t}^{Y_t}} \sum_{\substack{x, \text{ given} \\ x^{Y_1},...,x^{Y_T}}} \frac{1}{\prod_{i_1,...,i_m} x_{i_1...i_m}!}. \tag{2.7}$$

It is seen that the sum (2.1) appears as coefficient in this power series. We will now expand the exponential in (2.7) in a different way.

We proceed by induction with respect to the dimension m. For $m = 0$ (one single element x) the theorem is trivially true, since in that case the minimal set of fitted marginals is characterized by the single set $Y = \{\ \}$. So we may proceed to the induction step, and assume that the theorem holds for tables of dimension less than m.

By assumption Y_1 contains elements missing in all other Y_t's (the elements $1, ..., k$, say, $k \geq 1$) and the set U_1 of remaining indices is a subset of some Y_t, $t \geq 2$. Let $U_1 = \{k+1, ..., l\}$, $l \geq k$, that is $Y_1 = \{1, ..., l\}$. For further simplifications of the notations, let $U_1 \subset Y_2$ (this is no restriction; only Y_1 must satisfy the criterion at this stage).

Performing the summation over $i_1, ..., i_k$ in the generating function to the left in (2.7) we obtain

$$\exp\left[\sum_{i_1,...,i_m} \prod_{t=1}^{T} \alpha_{j_t}^{(t)}\right]$$
$$= \exp\left[\sum_{i_{k+1},...,i_m} \left(\sum_{i_1,...,i_k} \alpha_{j_1}^{(1)}\right) \prod_{t=2}^{T} \alpha_{j_t}^{(t)}\right]$$
$$= \exp\left[\sum_{i_{k+1},...,i_m} \beta_{j_2} \prod_{t=3}^{T} \alpha_{j_t}^{(t)}\right], \tag{2.8}$$

where we have introduced new variables β_{j_2} by

$$\beta_{j_2} = \alpha_{j_2}^{(2)} \sum_{i_1,...,i_k} \alpha_{j_1}^{(1)}. \tag{2.9}$$

Since $U_1 \subset Y_2$ the new variables β_{j_2} will depend on the same indices as $\alpha_{j_2}^{(2)}$. This means that the right hand side of (2.8) is the generating function of an $(m-k)$-dimensional decomposable table with the minimal set $x^{Y_2}, ..., x^{Y_T}$ of fitted marginals. Hence, expanding this generating function in analogy to (2.7) the general coefficient will be a sum of the same type as (2.1), but for the $(m-k)$-dimensional marginal table. Since the dimension is less than m we know that this sum may be expressed according to the theorem. We obtain the desired product (2.2) with the exception of the factor

$$\prod_{\substack{\text{all } U_1 \\ \text{marginals}}} u^{(t)}! \Big/ \prod_{\substack{\text{all } Y_1 \\ \text{marginals}}} y^{(t)}!$$
$$= \prod_{i_{k+1},...,i_l} x_{i_{k+1}...i_l}^{U_1}! \Big/ \prod_{i_1,...,i_l} x_{i_1...i_l}^{Y_1}! \tag{2.10}$$

But this factor is obtained when we go back from β to the original variables $\alpha^{(1)}$ and $\alpha^{(2)}$, since first

$$\prod_{j_2 \in I_2} \beta_{j_2}^{x_{j_2}^{Y_2}} \prod_{j_2 \in I_2} \Big(\sum_{i_1,...,i_k} \alpha_{j_1}^{(1)}\Big)^{x_{j_2}^{Y_2}} \prod_{j_2 \in I_2} \alpha_{j_2}^{(2)x_{j_2}^{Y_2}}, \tag{2.11}$$

and further, multiplying over the indices corresponding to $Y_2 - U_1$, we can write

$$\prod_{j_2 \in I_2} \Big(\sum_{i_1,...,i_k} \alpha_{j_1}^{(1)}\Big)^{x_{j_2}^{Y_2}} = \prod_{i_{k+1},...,i_l} \Big(\sum_{i_1,...,i_k} \alpha_{j_1}^{(1)}\Big)^{x_{i_{k+1}...i_l}^{U_1}}$$
$$= \prod_{i_{k+1},...,i_l} \sum_{\substack{x^{Y_1} \\ \text{given} \\ x^{U_1}}} \frac{x_{i_{k+1}...i_l}^{U_1}!}{\prod_{i_1,...,i_k} x_{i_1...i_l}^{Y_1}!} \prod_{i_1,...,i_k} \alpha_{j_1}^{(1)x_{i_1...i_l}^{Y_1}}. \tag{2.12}$$

The summation sign here may be removed to the left of all multiplications and brought together with the summation over the marginals $x^{Y_2}, ..., x^{Y_T}$, after which an identification of coefficients yields the desired result. *end of proof*

In a non-decomposable model the recursive procedure cannot be completely carried through. However, as far as Goodman's less restrictive procedure may be carried through, the summation problem in (2.1) may be reduced to that in a lower-dimensional table (the reduction may be performed to a model which might be called the purely nondecomposable kernel). This is seen from the induction step in the proof above. If U_1 is a subset of some Y_t, $t \geq 2$, the induction step works without change. If U_1 is not contained in any Y_t, $t \geq 2$, we introduce a new variable for the sum $\Sigma_{i_1,...,i_k} \alpha_{j_1}^{(1)}$. Continuing analogously we obtain the relation

$$\sum_{\substack{x,\text{ given}\\ x^{Y_1},\dots,x^{Y_T}}} \frac{1}{\prod_{i_1,\dots,i_m} x_{i_1\dots i_m}!} = \frac{\prod_{\substack{\text{all } U_1\\ \text{marginals}}} u^{(1)}!}{\prod_{\substack{\text{all } Y_1\\ \text{marginals}}} y^{(1)}!} \times \sum_{\substack{x^*,\text{ given}\\ x^{U_1},x^{Y_2},\dots,x^{Y_T}}} \frac{1}{\prod_{i_{k+1},\dots,i_m} x_{i_{k+1}\dots i_m}!}, \tag{2.13}$$

where $x^* = \{x_{i_{k+1}\dots i_m}\}$ is an $(m-k)$-dimensional table. The only difference from before is that the right hand sum is conditioned on x^{U_1}, too. The next step, if possible, is applied to the sum in the right member of (2.13), to reduce the dimension further.

As remarked in the introduction the so called exact test statistic for testing if a model may be reduced to a less general model is the ratio between the respective quantities (2.1). The theorem above gave the closed form formula (2.2) of (2.1) for a decomposable model. Thus we have the following corollary.

Corollary. *The exact test statistic for testing a hypothesis that specifies a reduction from one decomposable model to another, less general decomposable model may be expressed in a closed form.*

In the same situation the likelihood ratio may be expressed in closed form, as is well-known. The likelihood ratio is the quotient of two expressions of form (see (2.5))

$$\prod_{i_1,\dots,i_m} \hat{p}_{i_1\dots i_m}^{x_{i_1\dots i_m}} = \frac{\prod_{t=1}^{T} \prod_{\substack{\text{all } Y_t\\ \text{marginals}}} (y^{(t)})^{y^{(t)}}}{\prod_{t=1}^{T-1} \prod_{\substack{\text{all } U_t\\ \text{marginals}}} (u^{(t)})^{u^{(t)}}} \cdot \frac{N^{-N}}{\left(\prod_{k\in Z-\cup_t Y_t} r_k\right)^{N}}, \tag{2.14}$$

one for each model. Comparison of (2.14) with (2.2) reveals that the likelihood ratio may be regarded as approximating the exact test statistic in the same way as n^n approximates $n!$

These results about tests may be somewhat generalized. It is sufficient that the two models have the same purely non-decomposable kernel remaining when Goodman's procedure has been applied as far as possible. The reason is that this non-decomposable kernel will then only influence a factor common to numerator and denominator of the test statistic, and hence this factor may be crossed out. For the exact test statistic this is seen from the discussion made in connexion with equation (2.13). For the likelihood ratio it follows directly from the results by Goodman (1971) for instance, in an analogous way. We return to this generalization in section 3.

3. Partitioning of test statistics

The exact test statistic (see section 1) and the likelihood ratio may both be used to test models for a contingency table. More precisely we consider hypotheses specifying a reduction from one hierarchical model to a less general one.

Andersen (1973) formulated a conjecture stating that when both models are decomposable, each of the two test statistics may be factorized into a product of test statistics of the respective type, all factors corresponding to tests of homogeneity (equiprobability) in one-dimensional and tests of independence in two-dimensional tables. We will prove this conjecture here. A partitioning may be used in an analysis of the causes of a disagreement between data and a particular model. From some aspects, the possibility of partitioning hypotheses and likelihood ratios has been considered by Goodman (1970, 1971), cf. remarks to Lemma 3 below.

Theorem. *The exact test statistic and the likelihood ratio for testing a hypothesis specifying a reduction from one decomposable model to a less general decomposable model may both be factorized into a product of exact test statistics and likelihood ratios, respectively, each factor corresponding to a test of homogeneity in a one-dimensional table or a test of independence in a two-dimensional table.*

Remarks. The proof will be made independent of the precise closed form expressions possible for the test statistics. The reasons are two: A use of these expressions would only slightly have simplified and shortened the proof, and it would have obscured the possibility of generalizations. It may probably be asserted that the most interesting and essential case is expressed in the given formulation of the theorem, but as an inspection of the proof will reveal, extensions are possible in several directions:

(i) The statistical distribution of the table may be permitted to be of another exponential type. For instance, an analogous theorem could be formulated if the cell contents were independently normally distributed with a known and common variance. (Note that the model representations and the concept of decomposability in many respects may be regarded from a purely set-theoretic point of view, without any relation to the statistical distributions.)

(ii) It is not necessary that the models be decomposable. It is sufficient that the two models have the same purely non-decomposable kernel. For as described in the last part of section 2 the factor corresponding to this non-decomposable kernel may be crossed out of the test statistic by division. If we now instead insert a factor corresponding to the single set W = 'the union of sets in the kernel', the test statistic will not be changed but the two models will be changed to decomposable models, and the theorem will be applicable.

(iii) The two test statistics considered are not unique in their satisfying the theorem (but they are the most important ones).

Proof. Let the two models be represented by $\{X_1, ..., X_S\}$ and $\{Y_1, ..., Y_T\}$, with the former being the more general one, that is each Y_t being a subset of some X_s. The main part of the proof will be devoted to a demonstration of the possibility of a suitable partitioning of the model reduction $\{X_1, ..., X_S\} \to \{Y_1, ..., Y_T\}$ into a chain of successive reductions. Each such reduction will be of one of two types, which we call one-factor homogeneity and two-factor independence hypotheses, respectively.

By a *one-factor homogeneity hypothesis* we mean a reduction of type

$$\{Y, Y_1, ..., Y_T\} \to \{Y^*, Y_1, ..., Y_T\}, \tag{3.1}$$

where Y is a set containing some element not belonging to any of the Y_t's, and Y^* is obtained from Y by excluding this element. Y^* is permitted to be a subset of some Y_t, in which case Y^* is automatically excluded as being superfluous.

By a *general independence hypothesis* regarding a decomposable model $\{Y, Y_1, ..., Y_T\}$ we mean a reduction of type

$$\{Y, Y_1, ..., Y_T\} \to \{Y^*, Y^{**}, Y_1, ..., Y_T\}, \tag{3.2}$$

where Y^* and Y^{**} are proper subsets of Y with union $Y^* \cup Y^{**} = Y$ and are such that each Y_t satisfies either $Y_t \cap Y = Y_t \cap Y^*$ or $Y_t \cap Y = Y_t \cap Y^{**}$. Any of the sets Y^* or Y^{**} may be a subset of some Y_t, in which case it is automatically excluded as being superfluous. The hypothesis specifies independence in the Y marginal table between the two subsets $Y - Y^*$ and $Y - Y^{**}$. By a *two-factor independence hypothesis* we mean the particular case when these two latter subsets consist of one element each. We define the general independence hypothesis only for decomposable models. Note that the model $\{Y^*, Y^{**}, Y_1, ..., Y_T\}$ is decomposable in consequence. For if we choose an ordering of the more general model with Y as the last set according to the criterion for decomposability, this ordering will work also in the other model, with Y^*, Y^{**} as the last two sets.

Lemma 1. *Given two decomposable models, one more general than the other. The reduction from the more general model to the less general one may be performed in steps, each step implying a test of a hypothesis of one-factor homogeneity or two-factor independence.*

We will carry through the remainder of the proof of the theorem before we give a proof of Lemma 1, since the latter requires a good deal of work and further lemmas.

From the definitions of the exact test statistic and the likelihood ratio as quotients it is evident that both these statistics may be given a factorization corresponding to the partitioning of the model reduction asserted by the lemma. This factorization is not enough, however. It remains to make sure that each of these factors may be further factorized, into a product of the statistics for testing homogeneity in one-dimensional and independence in two-dimensional tables.

This may be checked in the closed form expressions of the exact test statistic and the likelihood ratio, which are obtainable from the formulas (2.2) and (2.14), respectively. But we will give a proof which is more general and does not explicitly use any results from section 2. We start this proof with a lemma, which will also be used in the proof of Lemma 1.

Lemma 2. *Consider a reduction from one decomposable model $\{X_1, ..., X_S\}$ to a less general decomposable model $\{Y_1, ..., Y_t\}$. If an element is not appearing at all in the two models, or is only contained in such sets X_s which remain unchanged under the reduction, that is $X_s = Y_t$ for some t, this element may be omitted from the two models before the tests are calculated (the test statistics are the same in the marginal table obtained by summing over the omitted index as in the original table).*

The proof of Lemma 2 is found at the end of this section.

In particular, the general independence hypothesis (3.2) may be considered within the marginal table $\boldsymbol{x}^Y$, where it assumes the simple form $\{Y\} \to \{Y^*, Y^{**}\}$. If the hypothesis is that of two-factor independence between two indices, it is now obvious that it means independence between these two indices for each set of fixed values of the other indices in the marginal table $\boldsymbol{x}^Y$. But with the other indices fixed, this is nothing but independence in a two-dimensional table. In this way a hypothesis of two-factor independence is made up from a number of hypotheses of independence in disjoint two-dimensional tables. These two-dimensional tables are obviously independently distributed with no parameters in common, whether the hypothesis of independence is true or not. Hence the two test statistics may both be correspondingly factorized into mutually independent test statistics for test of independence in these two-dimensional tables.

This is the desired factorization of the two-factor independence test. The proof for the one-factor homogeneity test is analogous but simpler, and is therefore omitted.

end of proof of theorem

We now start proving Lemma 1. First, in Lemma 3 below, the partitioning asserted in Lemma 1 will be established in the particular but most important case when the more general model is maximally general, the 'all-embracing' model $S=1$, $X_1=Z$. A proof in this case could alternatively be extracted from the two papers by Goodman (1970, 1971), but it is not explicitly stated there. For this reason and for completeness we give a proof of Lemma 3.

Lemma 3. *In the particular case when the more general model is maximally general, $X_1=Z$, Lemma 1 holds, that is the reduction to a less general model may be performed in steps corresponding to hypotheses of one-factor homogeneity and two-factor independence.*

Example. Consider tests of the decomposable model $\{Y_1, Y_2\}$, $Y_1=\{1, 3\}$, $Y_2=\{2, 3\}$ for a three-dimensional contingency table, that is tests of a reduction from $Z=\{1, 2, 3\}$ to $\{Y_1, Y_2\}$. The tests may be calculated as the product over the third index of the tests of independence between the two indices in the two-dimensional subtable obtained when the third index is fixed.

Proof of Lemma 3. First, if Z contains some elements not contained in any Y_t, we may reduce the model $\{Z\}$ to the model $\{\bigcup_t Y_t\}$ by a sequence of one-factor homogeneity hypotheses, each such hypothesis eliminating one of these elements.

Let $\{Y_t\}$ be ordered according to the criterion of decomposability. Primarily we partition the reduction from $\{\bigcup_t Y_t\}$ to $\{Y_1, ..., Y_T\}$ in the following way:

$$\left\{\bigcup_{t=1}^{T} Y_t\right\} \to \left\{Y_1, \bigcup_{t=2}^{T} Y_t\right\} \to \left\{Y_1, Y_2, \bigcup_{t=3}^{T} T_t\right\} \to ... \to \{Y_1, ..., Y_T\}. \quad (3.6)$$

Each of these reductions represents a general independence hypothesis, as this was defined. This is verified from the fact that for each τ, $Y_\tau \cap \bigcup_{t>\tau} Y_t = Y_\tau \cap Y_t$ for some $t>\tau$. Evidently, all the models are decomposable.

We must also show that a general independence hypothesis may be partitioned into a sequence of hypotheses of two-factor independence. According to Lemma 2 we may disregard those sets which are left unchanged by the reduction and write a general independence hypothesis as $\{Y^* \cup Y^{**}\} \to \{Y^*, Y^{**}\}$. Let say that $Y^*=\{1, ..., k, k+1, ..., l\}$ and $Y^{**}=\{1, ..., k, l+1, ..., m\}$, $0 \leq k < l < m$ (test of independence between $\{k+1, ..., l\}$ and $\{l+1, ..., m\}$ in an m-dimensional model). The desired partitioning of the general independence hypothesis into a sequence of two-factor independence hypotheses may be performed by first considering independence between the $(k+1)$st index and one at a time of the $m-l$ last indices, next between the $(k+2)$nd and the $m-l$ last, and so on. That this works is formally seen in the following way.

$$\begin{aligned}
&Y^* \cup Y^{**} = \{1, ..., m\} \to \{1, ..., k, k+2, ..., m\}, \\
&\{1, ..., m-1\} \to \{1, ..., k, k+2, ..., m\}, \{1, ..., k, \\
&k+2, ..., m-1\}, \{1, ..., m-2\} = [\text{exclusion of sub-} \\
&\text{set}] = \{1, ..., k, k+2, ..., m\}, \{1, ..., m-2\} \to ... \to \\
&\{1, ..., k, k+2, ..., m\}, \{1, ..., l\} \to \{1, ..., k, k+3, ..., \\
&m\}, \{1, ..., k, k+2, ..., m-1\}, \{1, ..., l\} \to ... [\text{as} \\
&\text{above}] \to \{1, ..., k, k+3, ..., m\}, \{1, ..., k, k+2, ..., l\}, \\
&\{1, ..., l\} = \{1, ..., k, k+3, ..., m\}, \{1, ..., l\} \to ... \\
&\to \{1, ..., k, l+1, ..., m\}, \{1, ..., l\} = Y^{**}, Y^* \qquad (3.7)
\end{aligned}$$

end of proof of Lemma 3

Proof of Lemma 1. We prove the existence of a partitioning of desired type by induction over the number of elements (the number of indices). We will show that it is possible to partition the reduction into one reduction step involving at least one element less and a number of reduction steps corresponding to one-factor homogeneity or two-factor independence hypotheses.

In the more general of the two models, $\{X_1, ..., X_S\}$, choose an element that belongs to exactly one of the sets, which is possible since the model is decomposable. Let this set be X_1, say, and let the element be the integer 1. Let X_1^* denote the set obtained from X_1 by excluding the element 1, that is $X_1^*=X_1-\{1\}$. The element 1 may or may not belong to $\cup Y_t$. The latter case is the simplest. In this case we make the partitioning

$$\{X_1, X_2, ..., X_S\} \to \{X_1^*, X_2, ..., X_S\} \to \{Y_1, ..., Y_T\}. \quad (3.8)$$

The first reduction corresponds to a one-factor homogeneity hypothesis and the remaining reduction does not involve the element 1.

Now suppose that the element 1 belongs to at least one of the Y_t's. Then we will make use of the following lemma, stating useful properties of decomposable models.

Lemma 4. *In a decomposable model represented by* $\{Y_1, ..., Y_T\}$, *consider an arbitrary element belonging to* $\cup_t Y_t$, *the element 1 say, and define W to be the union of all the sets* Y_t *which contain this element, the sets* $Y_{\tau+1}, ..., Y_T$ *say,* $\tau < T$. *If the single set W is substituted for this collection of sets the resulting, less restrictive model* $\{Y_1, ..., Y_\tau, W\}$ *is also decomposable (and no set in this representation is superfluous). In fact, in both models the sets may be ordered according to the criterion of decomposability in such a way that the collection of sets* $\{Y_1, ..., Y_\tau\}$ *comes first (a fact that justifies our numbering of the sets).*

The proof of this lemma will be given after the present proof.

According to Lemma 4 we may without restriction assume that precisely for $t > \tau$ the element 1 belongs to Y_t, $0 \leq \tau < T$, with $\{Y_t\}$ ordered according to the criterion. Since the model $\{X_1, ..., X_S\}$ is more general than $\{Y_1, ..., Y_T\}$ we must have $Y_t \subset X_1$ for each $t > \tau$, and consequently $W \subset X_1$ (with possible equality). We make the partitioning

$$\{X_1, ..., X_S\} \to \{W, X_1^*, X_2, ..., X_S\} \to \{W, Y_1, ..., Y_\tau\}$$
$$\to \{Y_{\tau+1}, ..., Y_T, Y_1, ..., Y_\tau\} = \{Y_1, ..., Y_T\}. \quad (3.9)$$

In the second model the set X_1^* may happen to be superfluous. One or several of these reductions may be empty, for instance if $W = X_1$ or if $\tau = T-1$ and $W = Y_T$, but otherwise these three reduction steps may be characterized in the following way.

The first step corresponds to a general independence hypothesis (note that for each $t > 1$, $X_t \cap X_1 \subset X_1^*$), and may be simplified to $\{X_1\} \to \{W, X_1^*\}$. It follows from Lemma 3 that this hypothesis can be partitioned into hypotheses of two-factor independence. It also follows that the second model is decomposable.

In the two intermediate models in (3.9) the element 1 belongs only to W, which is common for the two models. Both models are decomposable, so according to Lemma 2 we may simplify the second reduction step by changing W to $W^* = W - \{1\}$. Observing that $W^* \subset X_1^*$ we obtain the equivalent reduction

$$\{X_1^*, X_2, ..., X_S\} \to \{W^*, Y_1, ..., Y_\tau\}. \quad (3.10)$$

The first of these models is obviously decomposable, and the second model is decomposable according to Lemma 4. Thus, we have a reduction from one decomposable model to another, none of the models involving the element 1.

The third reduction step in (3.9) is (according to Lemma 2) equivalent to the simplified reduction

$$\{W\} \to \{Y_{\tau+1}, ..., Y_T\} \quad (3.11)$$

(Note that for each $t \leq \tau$, $Y_t \cap W$ is a subset of some of the sets $Y_{\tau+1}, ..., Y_T$, according to Lemma 4.) The reduction (3.11) can be partitioned into one-dimensional homogeneity and two-dimensional independence hypotheses by means of Lemma 3.

end of proof of Lemma 1

Proof of Lemma 4. The crucial point is that a recursive ordering and elimination in the original model may be performed according to our criterion for decomposability in such a way that $Y_1, ..., Y_\tau$ are eliminated first. Assume that we try this but fail at some step.

First, some remaining set among $Y_1, ..., Y_\tau$ must have a non-void intersection with W, since otherwise the class of these remaining sets must have represented a decomposable model in itself, a fact contradicting the failure.

Since the model is decomposable, a continuation of the recursive elimination must be possible. The elimination of sets among $Y_1, ..., Y_\tau$ can only be restarted by the elimination of one of the sets with non-void intersection with W. Let Y_t be any such set. Any element from W contained in Y_t also belongs to some set containing the element 1, which must be eliminated before Y_t. It follows that all sets containing 1 must be eliminated before Y_t. Thus, irrespective of the set Y_t to be next eliminated among $Y_1, ..., Y_\tau$, all sets containing the element 1 must be eliminated first. But this contradicts the fact that any set may be saved to be the last one, as remarked after the criterion in section 1. Hence it must be possible to eliminate $Y_1, ..., Y_\tau$ before the other sets.

The other statements of Lemma 4 now follow immediately.

end of proof of Lemma 4

We will refer to Lemma 4 in the proof of Lemma 2, too.

Proof of Lemma 2. We use the constructions of closed forms for the test statistics described in the last part of section 2. We prove the lemma for the exact test statistic. The proof for the likelihood ratio is quite analogous.

Let say that the element 1 does not occur in any set X_s or Y_t. Then it will affect the exact test statistic only by a factor r_1^N in both numerator and denominator. Crossing out this factor is equivalent to calculating the test statistic within the marginal table obtained by summing over the first index.

Next, suppose the element 1 occurs in the sets of some common subclass of $\{X_s\}$ and $\{Y_t\}$ which is not affected by the hypothesis. Applying Lemma 4 we may choose an ordering satisfying the criterion for decomposability and being such that the sets containing 1 come last. From the explicit form of the test statistic is now seen that all factors involving the first index occur in both numerator and denominator. So omitting the first index in all of them does not change the test statistic, and the resulting appearance of the statistic is identical to the test statistic obtained when the first index is omitted everywhere already before the test is constructued.

end of proof of Lemma 2

Hereby all the lemmas have been proved and the proof of the theorem is complete.

Acknowledgement

The author is grateful to the referee for valuable comments.

References

Andersen, A. H. (1973). Kontingenstabeller (in Danish). Read at the 5th Nordic Conf. on Math. Stat. Unpublished version of Andersen (1974).

Andersen, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1**, 115–127.

Birch, N. W. (1963). Maximum likelihood in three-way contingency tables. *J. Royal Statist. Soc. B.* **25**, 220–233.

Bishop, Y. M. M. (1971). Effect of collapsing multidimensional contingency tables. *Biometrics* **27**, 545–562.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. *J. Amer. Statist. Ass.* **65**, 226–256.

Goodman, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *J. Amer. Statist. Ass.* **66**, 339–344.

Haberman, S. J. (1970). The general log-linear model. Ph.D. thesis. University of Chicago.

Haberman, S. J. (1974). *The analysis of frequency data.* Univ. Chicago Press, Chicago.

Dr Rolf Sundberg
Department of Mathematics
The Royal Institute of Technology
S-100 44 Stockholm 70
Sweden