ON
ALTERNATIVE PARAMETRISATIONS IN STANDARD LOG-LINEAR MODEL

by

Andrzej KUTYLOWSKI*)

Forschungsbericht/
Research Memorandum No. 145

July 1979

*) Institute of Philosophy and Sociology, Polish Academy of Sciences, 72 Nowy Swiat, Warszawa, Poland and Department of Mathematics, Institute for Advanced Studies, Vienna.

I would like to thank Inge Heinrich for helpful comments concerning the style of presentation. I am solely responsible for any remaining errors and would be glad to be notified of their occurence.

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

Content

	Page
Abstract	ii
1. Introduction	1
2. The problem of parametrisation	7
3. Multiplicative interaction components	17
4. Mixed reference parametrisation	23
5. Fixed reference parametrisation	29
6. Conclusions and extensions	33
Notes	39
Appendix	43
Bibliography	47

Abstract

The issue of how to parametrise standard log-linear model has been neglected to some extent in the literature. In this paper general aspects of parametrisation are reviewed, in close connection with linear models theory. Two main alternative parametrisations, which were exclusively used by different authors, are next discussed and compared with reference to theory of multiplicative interactions in contingency tables. Finally, some aspects of implementation of different parametrisations are stressed, with regard to estimation procedures and possible extensions of standard log-linear model.

1. Introduction

In the empirical sciences the cumulation of knowledge proceeds through confrontations of assumed theoretical postulates and empirical evidence. Regardless of many philosophical and methodological issues defining the scenario of such confrontation, data analysis plays in it quite special role. Data analysis links assumptions with their consequences, both in confirmatory manner, when theoretical postulates proceed empirical check, and exploratory way,when data are supposed to contain information which,when extracted, retains some sort of theoretical relevance.

In the same time nondeterministic nature of many phenomena requires the collection of multiple measurements to allow for performing statistical analyses. Consequently, the role of statistical data analysis is crucial, particularly in the social sciences.

Until quite recently, data subject to rigorous statistical methodology were constrained mostly to quantitative rather than qualitative measurements, those latter arising when variables of interest were classifications.

The advent of so-called log-linear models changed this state of affairs quite dramatically. Log-linear models, similarly to their relatives linear models, provide a rich framework in which

i) structural hypotheses about probabilistic relations among variables can be formulated in a parametric way,
ii) the values of unknown structural parameters can be estimated and
iii) the hypotheses can be tested by evaluation of goodness of fit.

Log-linear models differ from earlier developments (of which brief but excellent account is given in LEWIS (1962) and MAXWELL (1961)) by their parametric character, with values of structural parameters informing about the degree of relation. In the other sense log-linear models are also parametric by including assumptions specifying the sample distributions of observed frequencies.[1)]

This first aspect raises the issue of the choice concerning alternative ways in which parametric character of a log-linear model can be realised. In fact, the problem is quite general and concerns the manner in which departures from a given zero structural hypothesis can be formulated. The probabilistic character of log-linear models requires actually consideration of alternative ways in which departures from independence or more general hypothesis can be expressed.

Consequently the details of this problem are comparable to very similar issues arising in traditional methodology of linear modelling, particularly in (so-called) analysis of variance.

Given variety of quite general log-linear models, we principally restrict the discussion (without great loss of generality) to what could be named a <u>standard</u> log-linear model. This model is in fact discussed in most available monographs by BISHOP, FIENBERG and HOLLAND (1975), FIENBERG (1977), PLACKETT (1974) see also HABERMAN (1974a), EVERITT (1977) and UPTON (1978). Among variety of methods which can be used for estimation of this model, maximum likelihood occupies the first place. Alternative approaches to estimation, however, do not interfer with our style of discussion as it pertains to the model construction rather than to the estimation of the values of given parameters.

A standard log-linear model arises when there are J classifications so the value c_j of jth variable is from a finite set $\bar{c}_j$, $c_j \varepsilon \bar{c}_j$, $c_j \leq C_j$. Then, considering the joint cross-classification, we can define its value or category as $u \equiv <c_1, c_2, \ldots, c_J>$, u belonging to a finite set $u \varepsilon \bar{u}$. Consequently, the value of joint variable can be indexed by a positive integer $w \varepsilon \bar{w}$, $w \leq W$ and a (discrete) density can be defined on the set $\bar{u}$ of all u_w . This probabilistic assumption is usually stated by saying that that joint density pertaining to J variables is multinomial with probabilities $\pi_1, \ldots, \pi_w$ and fixed sample size N. It means that the random variable N takes its values n_w with probabilities π_w . Strictly speaking we do not deal with qualitative data, as n_w are quantities, but rather with categorical data containing information about categories u_w and implicitely about relations among J classifications.

In log-linear models the interdependencies among J variates are dealt with by assuming that the unknown probabilities are certain functions of structural parameters, say $\underline{\rho}$, which are supposed to be interpreted in terms of degree of interdependencies among those variates. Further, probabilities are assumed to be linear functions on a logarithmic scale. Writing $\phi_w \equiv \ln \pi_w$, where ln( ) denotes natural logarithm, we can symbolise the model as

$$\phi_w = F_w(\underline{\rho}) \qquad w \leq W \tag{1.1}$$

Usually the specification of $F_w( )$ relies on intuitively appealing way to decompose ϕ_w into intercept term, say, ρ^*, main effects ρ_i^A , interaction ρ_{ij}^{AB} and so on. For illustrative purposes let us take $w=<i,j>$, $i \leq 2$, $j \leq 2$ and write ϕ_y as

$$\phi_y = \rho_*^* + \rho_{i*}^A + \rho_{j*}^B + \rho_{ij*}^{AB} \tag{1.2}$$

Such decomposition naturally generalises to higher dimensions. For J=3 and w=<i,j,K> we get

$$(1.3)\quad \phi_{ijk} = \rho^{*}_{**} + \rho^{A}_{i**} + \rho^{B}_{j**} + \rho^{C}_{k**} + \rho^{AB}_{ij*} + \rho^{BC}_{jk*} + \rho^{AC}_{ik*} + \rho^{ABC}_{ijk*}$$

Two properties of such decomposition should be noted. It always contains an intercept term and it is hierarchical in the sense that if term T is in it, then every term T'ЄT is also included. In other words, we can say that the decomposition contains all <u>marginal</u> terms T' given T terms and that the intercept term is marginal to every other.[2)]

(1.2) and (1.3) specify certain hypothesis about logarithms of probabilities and consequently about probabilities. Alternatively, we could consider the means of multinomial density, $E(N_w) = N\pi_w$ and define a model with respect to their logarithms, as FIENBERG (1977) and BFH (1975) do. Note, however, that because the intercept term in decomposition, the difference in specification is subsumed in redefining only the intercept term.[3)] The hypothesis specified by (1.2) or (1.3) are in a sense the most general ones. One can entertain less general hypothesis by omitting certain terms (f.e. set of parameters λ^{AB}_{ij} in (1.2) and simultaneously preserving the hierarchical character of this less general hypothesis by retaining all necessary marginal terms. Thus, the decomposition

$$\phi_{ij} = \rho^{*}_{**} + \rho^{A}_{i**} + \rho^{AB}_{ij*}$$

is not hierarchical because the parameters ρ^{B}_{j**} , $j \leq J$ are not included. Such nonhierarchical decompositions will not be discussed.

A given decomposition is usually called a model, which does not distinguish between a model and a structural hypothesis. In this paper, and for the reasons of convenience, the term model means the most general decomposition of the type (1.2) or (1.3), so these less general decompositions (hypothesis), are thought as being generated from the model. Thus, it is possible that same sets of structural hypothesis can be generated from two "different" models.

The linearity of hierarchical decomposition is in analogy to linear ("ANOVA") models, where this kind of decomposition appears with respect to the mean and where its advantages concerning interpretation have been recognised.

It should be clear that the suitable specification of $F_w(\ )$ is crucial in model construction, both for generation of structural hypothesis, and interpretation of parameters. Similarly as in linear models, it turns out to be less important for estimation and hypothesis testing. Rather unfortunately, criteria for specifying $F_w(\ )$ and ways of specifying parameters neither seem to be neither sufficiently discussed in theoretical literature nor sufficiently understood in applied literature.
In the 2nd part the general aspects of identifying parameters of log-linear model will be presented, in close analogy with linear model theory. In the log-linear framework, however, the choice among parameterisations depends to some extent on the way in which theory of interactions in contingency tables is formulated. In the 3rd part we present such a formulation in terms of interaction components, which serves as a basis for comparison of two main alternative parametrisations appearing in the literature, discussed respectively in Parts 4 and 5. Conclusions and brief comments concerning extensions are contained in Part 6.

2. The problem of parametrisation

It is easy to notice that the hierarchical decomposition, of which (1.2) and (1.3) are examples, contain parameters which are not identified. Writing for example system of equations (1.2) in a matrix form we obtain

$$(2.1)\qquad \begin{bmatrix}\phi_{11}\\ \phi_{12}\\ \phi_{21}\\ \phi_{22}\end{bmatrix} = \begin{bmatrix}1&1&0&1&0&1&0&0&0\\ 1&0&1&1&0&0&1&0&0\\ 1&1&0&0&1&0&0&1&0\\ 1&0&1&0&1&0&0&0&1\end{bmatrix}\begin{bmatrix}\\ \underset{\sim}{\rho}_{*}\\ \\ \end{bmatrix} \qquad \underset{\sim}{\phi} = \underset{\sim}{R}_{*}\underset{\sim}{\rho}_{*}$$

where $\underset{\sim}{\rho}^{1}_{*} = (\rho^{*}_{1*},\rho^{B}_{1*},\rho^{B}_{2*},\rho^{A}_{1*},\rho^{A}_{2*},\rho^{AB}_{11*},\rho^{AB}_{12*},\rho^{AB}_{21*},\rho^{AB}_{22*})$.

Columns of matrix $\underset{\sim}{R}_{*}$ are linearly dependent, and its rank is 4. This means that although ϕ is uniquely defined as $\phi = R_{*}\rho_{*}$, ρ_{*} is not and in this sense ρ_{*}'s are not identifiable (R^{-1}_{*} does not exist). Matrix $\underset{\sim}{R}_{*}$ can be described as being not of full column rank; its rank is less then the number of columns.

In general the unidentified model can be written as

$$(2.2)\qquad \underset{\sim}{\phi} = \underset{\sim}{R}_{*}\underset{\sim}{\rho}_{*}$$

$\underset{\sim}{\phi}(Wx1)$, $\underset{\sim}{R}_{*}(WxM_{*})$, $\rho_{*}(M_{*}x1)$ with rank of $\underset{\sim}{R}_{*}$, $r(\underset{\sim}{R}_{*}) = M$, $M<W$ $M<M_{*}$. A known matrix $\underset{\sim}{R}_{*}$ can be called a regressor matrix or a design matrix; this latter name, however, suggests a connection with the analysis of experimental data by linear models and is thus slightly misleading.

Because it would be clearly desirable to work with identified parameters, the general problem of parametrisation can be stated as follows:

from unidentifical model (2,2) derive an identified model in which certain parameters $\underset{\sim}{\rho}$ can be uniquely expressed as functions of ϕ's (under given structural hypothesis). The identified model should allow easy specification of structural hypothesis and contain parameters having "reasonable"interpretation.

There are several approaches to this problem in the context of linear model, where $\phi_w$is not a logarithm of probability but a mean. Concise presentation of those can be found in SEBER (1966) or in BOCK (1975). One of approaches particularly relevant in log-linear context, solves the problem of parametrisation of ful rank model through reparametrisation of unidentified model (2.2).

Let us consider M linear functions of unidentified parameters $\underset{\sim}{\rho}_*$

$$\underset{\sim}{\rho} = \underset{\sim}{P}\underset{\sim}{\rho}_* \tag{2.3}$$

such that $\underset{\sim}{\rho}(Mx1), \underset{\sim}{P}(MxM_*)$. It is proved that equation (2.2) can be solved with respect to $\underset{\sim}{\rho}$ only if

$$r\begin{pmatrix} \underset{\sim}{R}_* \\ \underset{\sim}{P} \end{pmatrix} = r(\underset{\sim}{R}_*) = r(\underset{\sim}{P}) = M \tag{2.4}$$

that is, when rows of $\underset{\sim}{P}$ are linearly independent. In this case $\underset{\sim}{R}_*$ can be uniquely factored into

$$\underset{\sim}{R}_* = \underset{\sim}{R}\underset{\sim}{P} \quad , \tag{2.5}$$

matrix $(\underset{\sim}{P}\underset{\sim}{P}')$ is nonsingular and $\underset{\sim}{R}$ is given by

$$\underset{\sim}{\phi} = \underset{\sim}{R}_*\underset{\sim}{\rho}_* = \underset{\sim}{R}\underset{\sim}{P}\underset{\sim}{\rho}_* = \underset{\sim}{R}\underset{\sim}{\rho} \tag{2.6}$$

Furthermore, knowing $\underset{\sim}{P}$ allows to find $\underset{\sim}{R}$ from

$$\underset{\sim}{R} = \underset{\sim}{R}_{*}\underset{\sim}{P}'(\underset{\sim}{P}\underset{\sim}{P}')^{-1} \tag{2.7}$$

and knowing $\underset{\sim}{R}$ allows to find $\underset{\sim}{P}$ from

$$\underset{\sim}{P} = (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}'\underset{\sim}{R}_{*} \quad . \tag{2.8}$$

Matrix $\underset{\sim}{P}$ is called <u>parametrisation</u> matrix, and matrix $\underset{\sim}{R}$ a <u>basis</u> matrix. Algebraic details of this construction are given in BOCK (1975).

From (2.1) we might expect that the knowledge of the type of $\underset{\sim}{P}$ matrix chosen in reparametrisation should be satisfactory to specify the basis matrix. From (2.8) we would expect that the knowledge of the type of $\underset{\sim}{R}$ matrix will be satisfactory to specify the parametrisation matrix. This is actually the case, and BOCK (1975: 300) presents various types of parametrisation matrices and corresponding basis matrices used in linear models, see also FINN (1974).

The typical case when w denotes multiple indices pertaining to variables requires, given particular type of (re)parametrisation, the construction of $\underset{\sim}{P}$ and $\underset{\sim}{R}$ matrices from "elementary" matrices by means of direct (Kronecker) products. Note, that once the correspondence between type of parametric functions and type of basis was established, the interest remains only with construction of the basis matrix.

Let $\underset{\sim}{B}_{Cj}$ be an elementary basis for variable j having C_j categories.

$$\underset{\sim}{B}_{Cj} = [\underset{\sim}{1}_j(*)\underset{\sim}{I}_j(1) \ \ldots \ \underset{\sim}{1}_j(C_j-1)] \quad . \tag{2.9}$$

consisting of C_j column vectors, each of dimension $(C_j x 1)$.

If the same type of "parametrisation" is used for every variable, then in direct products construction only vectors of (2.7) will appear, and typical column vector of $\underset{\sim}{R}$ matrix for $w = \langle c_1, c_2, \ldots, c_J \rangle$ will be

$$\underset{\sim}{r}(c_1,c_2,\ldots,c_J) = \underset{\sim}{b}_1(c_1) \otimes \underset{\sim}{b}_2(c_2) \otimes \ldots \otimes \underset{\sim}{b}_J(c_J) \tag{2.10}$$

giving required decomposition

$$\phi_{c_1c_2\ldots c_J} = \sum_{c_1,c_2,\ldots,c_J} \underset{\sim}{r}(c_1,c_2,\ldots,c_J) \ . \tag{2.11}$$

To illustrate these rather abstract concepts, let us consider a 2x2 case with the unidentifiable model (1.2). Assume that we decided on certain grounds to use orthogonal polynomials for both variables A and B and that, for technical reasons, in the composite subscript $\langle i,j \rangle$ the last index changes fastest, so in our example elements of ϕ have the order like in (2.1). The elementary orthogonal polynomial basis matrix is in our case

$$\underset{\sim}{L}_{2_k} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \left[\underset{\sim}{1}^k(*) \underset{\sim}{1}^k(1) \right] \quad , \quad k = A,B \tag{2.12}$$

In a model for only one dichotomous variable we would have $\underset{\sim}{R} = \underset{\sim}{L}_2$. For two dichotomies the general formula (2.10) specialises to

$$\begin{aligned} \underset{\sim}{\phi} &= \{ \sum_{c_i} \sum_{c_j} (\underset{\sim}{1}^A_i \otimes \underset{\sim}{1}^B_j) \} \underset{\sim}{\lambda} \\ &= \{ \underset{\sim}{1}^A(*) \otimes \underset{\sim}{1}^b(*) \} + \{ (\underset{\sim}{1}^A(*) \otimes \underset{\sim}{1}^B(1) \} + \{ \underset{\sim}{1}^A(1) \otimes \underset{\sim}{1}^B(*) \} + \\ &+ \{ \underset{\sim}{1}^A(1) \otimes \underset{\sim}{1}^B(1) \} \underset{\sim}{\lambda} \\ &= \left\{ \begin{bmatrix} \underset{*}{1}^A(*) \underset{\sim}{1}^B(*) \\ \underset{1}{1}^A(*) \underset{\sim}{1}^B(*) \end{bmatrix} + \begin{bmatrix} \underset{*}{1}^A(*) \underset{\sim}{1}^B(1) \\ \underset{1}{1}^A(*) \underset{\sim}{1}^B(1) \end{bmatrix} + \begin{bmatrix} \underset{*}{1}^A(1) \underset{\sim}{1}^B(*) \\ \underset{1}{1}^A(1) \underset{\sim}{1}^B(*) \end{bmatrix} + \begin{bmatrix} \underset{*}{1}^A(1) \underset{\sim}{1}^B(1) \\ 1^A(1) \underset{\sim}{1}^B(1) \end{bmatrix} \right\} \underset{\sim}{\lambda} \end{aligned} \tag{2.13}$$

$$\begin{bmatrix} \phi_{11} \\ \phi_{12} \\ \phi_{21} \\ \phi_{22} \end{bmatrix} \underset{(*,*)(*,1)(1,*)(1,1)}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}} \begin{bmatrix} \lambda^{*} \\ \lambda_{1}^{B} \\ \lambda_{1}^{A} \\ \lambda_{11}^{AB} \end{bmatrix}$$

In (2.12) subscripts indexing columns of the elementary matrix pertain to the categories of given variable, with categories relabelled from $1,2,3,\ldots C_j$ to $1,2,\ldots,(C-1)_j$. Thus $\underset{\sim}{1}^{k}(*)$ denotes the vector pertaining to the first category of variable k.

The algebraic logic of basis construction implicitely specifies the assignment of parameters to columns of $\underset{\sim}{R}$. Here "*" indicates the "independence" of a column (and hence the respective parameter) from particular variable. First column pertains to a parameter "independent" of both A and B, i.e. to the "intercept" parameter λ^{*} . Second column pertains to a parameter "dependent" on the first category of B and "independent" of the first variable; this parameter is denoted by λ_{1}^{B}. Last column pertains to a parameter "dependent" on first category of A and first category of B, i.e. to interaction effect λ_{12}^{AB} . The generations of basis matrices by means of direct products received a very general treatment by HABERMAN (1974a). Rather unfortunately, his very technical presentation has restricted availability and, being constrained to generation of basis matrix, did not show very close analogies between problem of parametrisation in log-linear and linear context. With the exception of BOCK (1975) other authors considered neither the problem of generating an R matrix by direct products technique, nor the construction of the (chosen) R matrix in this way.

Does this consitute an important gap in the theory and for practice of log-linear modelling?

Three aspects should be taken into account in giving answer to this question.

First of all, by continuing the analogy with the linear model, it should be stressed that parametrisation problem in a standard model is equivalent to the problem of parametrisation for a case with one observation per cell w. This fact considerably reduces complexity of the issue, making "usual" parametrisation in linear case an obvious choice. The most "usual" parametrisation discussed for the 'balanced' linear model with one observation per cell states that in two response cases

$$\phi_{ij} = \lambda^{*} + \lambda^{A}_{i} + \lambda^{B}_{j} + \lambda^{AB}_{ij} \tag{2.14}$$

where parameters are constrained to be

$$\sum_i \lambda^{A}_{i} = \sum_j \lambda^{B}_{j} = \sum_i \lambda^{AB}_{ij} = \sum_j \lambda^{AB}_{ij} = 0 \quad .$$

The extension for higher number of responses is obvious; for three response cases we get

$$\phi_{ijk} = \lambda^{*}+\lambda^{A}_{i}+\lambda^{B}_{j}+\lambda^{C}_{k}+\lambda^{AB}_{ij}+\lambda^{AC}_{ik}+\lambda^{BC}_{jk}+\lambda^{ABC}_{ijk} \tag{2.15}$$

with constraints

$$\sum_i \lambda^{A}_{i} = \ldots = \sum_i \lambda^{AB}_{ij} = \sum_j \lambda^{AB}_{ij} = \ldots = \sum_i \lambda^{ABC}_{ijk} = \sum_j \lambda^{ABC}_{ijk} = \sum_k \lambda^{ABC}_{ijk} = 0$$

In general polytomous case this parametrisation seem to be first employed by BIRCH (1963) and subsequently appears in a long series of papers by GOODMAN (see, f.e. 1970, 1971).

It has been exclusively used in monographs by BISHOP, FIENBERG and HOLLAND (1975),FIENBERG (1976), EVERITT (1977) and UPTON (1978).

In (2.15) λ's represent in a certain manner the degree of various interdependencies among A,B,C. By analogy with ANOVA models, λ_i^A can be called a main effect, or 0-th order interaction effect. λ_{ij}^{AB} can be called a (first order) interaction effect, and λ_{ijk}^{ABC} an second order interaction effect. It does not seem to be appriopriate to interprete λ_{ij}^{AB} "as two factor interaction". In (2.15), contrary to ANOVA model, A and B are responses and not factors on which the distribution characterised by π is conditioned on. In these sense, all interactions in standard log-linear model are response interactions.

Another 'main' alternative choice of parametrisation gives in two response cases

$$(2.16) \qquad \phi_{ij} = \gamma^* + \gamma_i^A + \gamma_j^B + \gamma_{ij}^{AB}$$

with constraints $\gamma_a^A = \gamma_b^B = \gamma_{ab}^{AB} = 0$, for certain arbitrarily fixed $a,b, a\varepsilon\overline{I}, b\varepsilon\overline{J}$. Again the extension to higher number of responses is obvious and for three response cases we get

$$(2.17) \qquad \phi_{ijk}=\gamma^*+\gamma_i^A+\gamma_j^B+\gamma_k^C+\gamma_{ij}^{AB}+\gamma_{ik}^{AC}+\gamma_{jk}^{BC}+\gamma_{ijk}^{ABC}$$

with constraints

$$\gamma_a^A = \ldots = \gamma_{ab}^{AB} = \ldots = \gamma_{abc}^{ABC} = 0$$

for certain fixed $a,b,c, a\varepsilon\overline{I}$, $b\varepsilon\overline{J}$, $c\varepsilon\overline{K}$. Typically a,b,c can be taken to be respectively I,J,K, so parameter's subscripts way over $i<I$, $j<J$, $k<K$.

This parametrisation was first employed for general polytomous case by MANTEL (1966) who noted that (2.17) was "stimulated by correspondence with GOODMAN". It was subsequently and exclusively used in a monograph by PLACKETT (1974), THEIL (1970) and in a series of papers by KULLBACK and his associates, see f.e. KU and KULLBACK (1974). Analogously to λ's, γ parameters are called (response) interaction effects of given order.

Two separate lists of references where either γ's or λ's appear suggest that the lack of discussion concerning selection of parametrisations might have been based on continuation of non-overlapping traditions. Such a dominance of historical aspect in evaluation of model construction is clearly undesirable.

Secondly, given the choice from "usual" ANOVA parametrisation in balanced case, is any of (2.15) or (2.17) preferable on grounds of parsimony? There is no definite answer to this question, unless one considers the issue of

(i) estimation of parameters
(ii) testing structural hypothesis and
(iii) interpretation of parameters.

The most general identified hypothesis (model) specifies

$$\underset{\sim}{\phi} = \underset{\sim\sim}{R}\underset{\sim}{\rho} \tag{2.18}$$

with $r(\underset{\sim}{R}) = M = W$ and establishes one-to-one correspondence between $\underset{\sim}{\rho}$'s and ϕ's (this is called a saturated model). In a less general hierarchical hypothesis certain terms (f.e. $\underset{\sim}{\lambda}^{AB}_{ij}$ in 1.12) are omitted, which reduces the dimensions of $\underset{\sim}{\rho}$ and $\underset{\sim}{R}$ and consequently also $r(R)$. An identified parametrisation of saturated model is thus important, allowing obvious generation of structural hypothesis.

Contrary, however, to linear model situation where vector ϕ is unconstrained, probabilities π_w must sum to 1 and the consistency of summations to marginal probabilities should be preserved. This suggests that the evaluation of a given parametrisation requires its comparison with established ways of formulating hypothesis about relationships in contingency tables, i.e. with theories of interactions. Theory of interactions was traditionally formulated directly in terms of probabilities, but not explicit functions of them. For our purposes it needs to be reformulated in a parametric way to provide a basis for comparison between λ and γ parametrisation. We present such reformulation before discussing comparisons with respect to (i),(ii) and (iii).

3. Multiplicative interaction components

In one-way table[4] $\overline{\pi}_i^A \equiv \{\pi_1,\ldots,\pi_I\}$, $i \leq I$, we can define a ratio of π_i to $\pi_{i'}$

$$w_{i'}^{i} \equiv \frac{\pi_i}{\pi_{i'}} \quad . \tag{3.1}$$

This ratio can be interpreted as ratio of odds of occurence of A_i with respect to $A_{i'}$. Without any loss of generality it can be assumed that $i \neq i'$ as $w_{i}^{i'} = 1$. Odds ratios (3.1) and all other odds-ratios we are going to consider, could be named relative odds[5], because in example (3.1) they are defined to relate probability of one category to probability of another. For $I \geq i$ there are $I(I-1)$ such ratios and they will be named 0-th order (relative) interaction components. For I=3 we have

$$w_2^1 = \pi_1/\pi_2 \qquad w_3^1 = \pi_1/\pi_3$$

$$w_1^2 = \pi_2/\pi_2 \qquad \omega_3^2 = \pi_2/\pi_3$$

$$w_1^3 = \pi_3/\pi_1 \qquad w_2^3 = \pi_3/\pi_2 \quad .$$

It is important to note that components $w_{i}^{i'}$ satisfy restriction

$$\prod_{ii'} w_{i}^{i'} = 1 \tag{3.2}$$

(For I=3 we easily check that $w_1^2 w_2^1 w_1^3 w_3^1 w_2^3 w_3^2 = 1$)

In a two-way table $\overline{\Pi}_{ij}^{AB}$ we can define stratified o-th order interaction components by fixing a category of one variable,

say j, to obtain $w^{i}_{i'j}$, $w^{i}_{i'j'}$. The interaction components of the 1st order can be defined as

$$(3.3) \qquad w^{i\,j}_{i'j'} \equiv w^{i}_{i'j}/w^{i}_{i'j'} \qquad\qquad j\neq j'$$

$$= \frac{\pi_{ij}\pi_{i'j'}}{\pi_{i'j}\pi_{ij'}}$$

Similarly to (3.2) $w^{ij}_{ij'}$ satisfy restrictions

$$(3.4) \qquad \prod_{ii'} w^{ij}_{i'j'} = \prod_{jj'} w^{ij}_{i'j'} = 1$$

Generalising this construction to higher dimensions does not present difficulties. In three dimensions for example, we define the 2nd order interaction component to be a ratio of two stratified interaction components of the 1st order, say

$$(3.5) \qquad w^{ijk}_{ij'k'} \equiv w^{ij}_{ijk}/w^{ij}_{ij'k'}$$

$$= \frac{\pi_{ijk}\pi_{i'j'k}\pi_{i'jk}\pi_{ij'k'}}{\pi_{i'jk}\pi_{ij'k}\pi_{ijk'}\pi_{i'j'k'}}$$

Consequently $w^{ijk}_{i'j'k'}$ satisfy restrictions

$$(3.6) \qquad \prod_{ii'} w^{ijk}_{i'j'k'} = \prod_{jj'} w^{ijk}_{i'j'k'} = \prod_{KK'} w^{ijk}_{i'j'k'}$$

Above components can be defined for any choice of stratification variable; the definition (3.3) is equivalent to $w^{ij}_{i'j'} \equiv w^{j}_{ij'}/w^{j}_{ij}$, and the definition (3.5) to that obtained using any of the other two possibilities to choose two variables and stratify with respect to the third. This pro-

perty, sometimes called symmetry, could be described as stratification invariance and is highly desirable for logical reasons, see SIMPSON (1951) for detailed presentation is 2x2x2 case.

Because of the constraints on interaction components, it is reasonable to find such a subset of 0-th, 1st and 2nd order components (or functions of them) which, being satisfactory for formulation of various hypothesis about relationships, allow to represent respectively probabilities from one-way, two-way, three-way distributions in a unique way and are further eventually unconstrained.

Let us consider for heuristic reasons restriction (3.2)

$$\prod_{i\,i'}(\prod w^{i}_{i'}) = \prod_{i'\,i}(\prod w^{i}_{i'}) = 1$$

and note that we can either set here $i' \neq i$ and fix i or set i free for fixed i'. Writing in the first case $w^{i}_{i'} \equiv m^{i}_{a}$ and in the second case $w^{i}_{i'} \equiv f^{i}_{a}$, $a \neq i$ (3.2) becomes

$$\prod_{i\,a}(\prod m^{i}_{a}) = \prod_{a\,i}(\prod f^{i}_{a}) = 1$$

Considering $w^{i}_{i'}$ jointly for i,i' we can call the set $\bar{m}^{i}_{a}$ mixed reference category components (because a varies given i) and the set $\bar{f}^{i}_{a}$ fixed reference category components, (because a is fixed). Note that only in the second expression the product of the term in parenthesis does not depend on i, which suggest that we might be able to dispense with constraint using fixed (reference) components, but not mixed components.

From (3.1), (3.3), (3.5) we now have, given a, $I-1$ fixed 0-th order components f^{i}_{a}, $(I-1)(J-1)$ fixed 1st order components f^{ij}_{ab} and $(I-1)(J-1)(K-1)$ 2nd order components f^{ijk}_{abc}. Those

components are satisfactory to express[5)]

(i) hypothesis of homogeneity in one way table by

(F1) $$\bar{f}_a^i = 1$$

(ii) hypothesis of independence (no 1st order interaction) in two way table by

(F2) $$\bar{f}_{ab}^{ij} = 1$$

(iii) hypothesis of no 2nd order interaction in a three way table by

(F3) $$\bar{f}_{abc}^{ijk} = 1$$

In fact, the lack of 2nd order interaction has been defined "nonparametrically" in terms of probabilities by (F3), see original paper by RØY and KASTENBAUM (1956).

Reconsidering (3.1) in terms of mixed components, we obtain for fixed i (I-1) components m_a^i with restriction $\Pi_i m_\cdot^i = 1$. Hypothesis of homogeneity can now be expressed by

(M1a) $$\bar{m}_a^i = 1 \quad .$$

Note that this is equivalent to (F1) because multiplying m_a^i by π_a/π_i we get $f_i^a = 1$ and changing indices finally gives the equivalence. Furthermore, (M1a) will hold for every i, so (M1a) is equivalent to

(M1b) $$\bar{m}_a^i = 1 \quad , \qquad \text{all } i .$$

Hypothesis of independence can be expressed by

(M2a) $$\bar{m}_{ab}^{ij} = 1$$

which is equivalent to (F2) which we notice by taking reciprocal and changing indices. Consequently, it will hold for every i,j, which causes the equivalence of (M2a) to

(M2b) $$\bar{m}^{ij}_{ab} = 1 \qquad \text{all } i, j.$$

Analogous reasoning shows the equivalence of (F3) to

(M3a) $$\bar{m}^{ijk}_{abc} = 1$$

and final equivalence of (M3a) to

(M3b) $$\bar{m}^{ijk}_{abc} = 1 \qquad \text{all } i, j, k.$$

Finally we come to the conclusion that (M1), (M2) and (M3) are equivalent respectively to $\bar{w}^{i}_{i'} = 1$, $\bar{w}^{ij}_{i'j'} = 1$, $\bar{w}^{ijk}_{i'j'k'} = 1$. This last formulation of independence and no 2nd order interaction was used by DARROCH (1962), who noted its equivalence to original definition by ROY and KASTENBAUM. DARROCH noted also the functional dependence of mixed components, but did not give explicit parametric treatment specifying the type of this functional dependence.

We conclude that whereas using fixed components implicitely satisfies constraints on probabilities, the use of mixed components requires retention of constraints and consequently leads to "redundant" decomposition.

Above presentation generalises an approach by DAVIS (1974) whose presentation of multiplicative components in terms of odds ratios did not distinguish between fixed and mixed components.

This was quite understandable, as DAVIS approach was essentially constrained to multiple dichotomies. Properties of

multiplicative interaction components are presented in more details in KUTYLOWSKI (1979), see also DARROCH (1962, 1974) who discusses other approaches than multiplicative one.

We have shown that formulation of basic hypothesis in contingency tables can be equivalently done by considering two types of functions of probabilities: fixed and mixed interaction components. In the next two parts we compare λ and γ parametrisations by interpreting λ's and γ's in terms of fixed and mixed components.

4. Mixed Reference Parametrisation

We will consider three-way model (2.14), which is satisfactory for our purposes. (2.14) defines a system of IJK linear equations in λ's. Solution of this system gives

$$(4.1)\quad \lambda^{*} = \phi_{+++} ,$$

$$\lambda^{A}_{i} = \phi_{i++} - \phi_{+++} , \ldots ,$$

$$\lambda^{AB}_{ij} = (\phi_{ij+} - \phi_{+j+}) - (\phi_{i++} - \phi_{+++}) = \lambda^{A}_{ij} - \lambda^{A}_{i} , \ldots ,$$

$$\lambda^{ABC}_{ijk} = (\phi_{ijk} - \phi_{+jk} - \phi_{i+k} + \phi_{++k}) - (\phi_{ij+} - \phi_{+j+} - \phi_{i++} + \phi_{+++})$$

$$= \lambda^{AB}_{ijk} - \lambda^{AB}_{ij}$$

λ^{*} turns out to be a general mean of log-probabilities, λ^{A}_{i} an 0-th order interaction effect, λ^{AB}_{ij} a first order interaction effect. In (4.1) λ^{A}_{ij} defines stratified 0-th order effect (for B_j), λ^{AB}_{ijk} defines stratified 1st order effect (for C_k). More detailed interpretation of λ effects is given in many texts treating linear model, see SCHEFFEE (1959) or SEARLE (1971) and will not be presented here.

It is not obvious how do λ parameters relate to the interaction components. GOODMAN (1970, 1971) and BISHOP, FIENBERG and HOLLAND (1975) express λ's for particularly simple models in terms of odds ratios but it seems that general expressions are not available. This prohibits more meaningful interpretation of λ effects, particularly in polytomous cases, making efforts to establish rules of interpretation lacking in generality and constrained to simplest situations (see f.e. PAGE, 1977).

Moreover, showing correspondence between λ's and interaction components would give a firm basis for comparison between λ and ϑ parametrisations.

According to our theorem, proved in Appendix for three-way table, λ's relate to mixed reference components and λ parametrisation is consequently termed mixed reference parametrisation. The theorem states that λ's can be expressed in terms of odds ratios as follows

$$(4.2)\ (i) \qquad \lambda^{A}_{i} = (IJK)^{-1}\{\Sigma\Sigma\Sigma \ln w^{i}_{ajk}\} = (IJK)^{-1}\Sigma\Sigma \ln(m^{i}_{\cdot jk}),\ldots,$$

$$(ii) \qquad \lambda^{AB}_{ij} = (IJK)^{-1}\{\Sigma\Sigma\Sigma \ln w^{ij}_{abk}\} = (IJK)^{-1}\Sigma \ln(m^{ij}_{\cdot\cdot k}),\ldots,$$

$$(iii) \qquad \lambda^{ABC}_{ijk} = (IJK)^{-1}\{\Sigma\Sigma\Sigma \ln w^{ijk}_{abc}\} = (IJK)^{-1}\ln(m^{ijk}_{\cdot\cdot\cdot})$$

λ effects are rather complicated functions of odds-ratios. Every λ depends on mixed interaction components of the same order as the order of given λ. Note that by suitable multiplication we can make each λ effect proportional to the natural logarithm of the geometric mean of respective odds-ratios. F.e. multiplying and dividing λ^{A}_{i} by (I-1)JK we obtain

$$(4.3) \qquad \lambda^{A}_{i} = \frac{(\ln m^{i}_{\cdot\cdot\cdot})/(I-1)JK}{IJK\ /\ \ (I-1)JK}$$

$$= \frac{\ln m^{i}_{\cdot\cdot\cdot}}{(I-1)JK} \times \frac{(I-1)JK}{IJK}$$

$$= \frac{I-1}{I} \ln(m^{i}_{xxx}).$$

By analogous reasoning

$$\lambda_{ij}^{AB} = \frac{(I-1)(J-1)}{IJ} \ln(w_{xxx}^{ij})$$

$$\lambda_{ijk}^{ABC} = \frac{(I-1)(J-1)(K-1)}{IJK} \ln(m_{xxx}^{ijk})$$

Incidentally, results (4.3) were derived for some simple tables and conjectured for more complex tables by DAVIS (1972).

Regardless of the complexity of λ's functions of odds ratios, the derivation of expressions (4.2) or (4.3) might be possibly helpful in interpreting the magnitudes of effects under given structural hypothesis which specifies the absence of certain terms (f.e. that $\bar{\lambda}_{ijk}^{ABC} = 0$). More detailed interpretation of parameters, however, will not be discussed here.

From a practical point of view it is important to know how to obtain mixed reference parametrisation for given model. According to (2.10) and (2.11) the knowledge of elementary basis matrix is satisfactory for this purpose. The identified λ parameters will pretain to respective columns of constructed regressor matrix L, and thus are related to sets of categories $\bar{c}_j$, $c_j \varepsilon \bar{c}_j$, $j \leqslant J$. Before discussing construction of regressor matrix we introduce a general notation for subscripting identified parameters. Similarly as before (see 2.13) let us denote the first category of j-th variable by "$*_j$", the second category by "2_j",..., the last category by $(C_j-1)_j$. Generally, then every effect depends on $c_i,\ldots,c_J$

$$\lambda(c_i,\ldots,c_J) \quad .$$

From the algebraic logic of basis construction it follows that

(i) if $c_j = *$ then λ effect does not depend on variable j.

(ii) if $c \neq *$ then λ effect pertains, not necessary exclusively, to c_j^{-1}-th category of j-th variable. Hence $\lambda(*_1,*_2) \equiv \lambda^*, \ldots, \lambda(1_1,1_1) \equiv \lambda_{11}^{AB}$ as in simple case described in (2.13).

The elementary basis for j-th variable is of dimension $(C_j \times C_j)$ and consists of C_j vectors

$$\underline{L}_{c_j} = \{\underline{l}_j(*_i)\underline{l}_j(1_i)\ldots\underline{l}_j(C_j-1)\}$$

such, that

if $c_j = *$ then all elements of $\underline{l}_j(c_j)$ are ones.

$c_j \neq *$ then
- c_j-th element of $\underline{l}_j(c_j) = 1$
- C_j-th element of $\underline{l}_j(c_j) = -1$
- all other elements equal 0

Mixed reference elementary basis can thus be written as

$$\underline{L}_{c_j} = \begin{bmatrix} \underline{1} & \underline{I} \\ 1 & -\underline{1}' \end{bmatrix}, \tag{4.3}$$

where $\underline{I}$ is $(C_j-1)^x(C_j-1)$ identity matrix. A vector $\underline{l}(c_1,\ldots,c_J)$ of the basis matrix is generated from (4.3) by (2.10). The basis itself is generated by (2.11).

For illustrative purposes we generate mixed reference basis for 2x3 table. Here we have

$\underline{l}_2(*_2) = (1,1,1)'$ $\qquad$ $\underline{l}_1(*_1) = (1,1)'$

$\underline{l}_2(1_2) = (1,0,-1)'$ $\qquad$ $\underline{l}_1(1_1) = (1,-1)'$

$\underline{l}_2(2_2) = (0,1,-1)'$

From (2.10) and rules for subscripting parameters given above we obtain columns of $\underset{\sim}{L}$ matrix (of dimension 6x1) and parameters pertaining to them.

$\underline{1}(*_1,*_2) = (1,1,1,1,1,1)'$ $\qquad \lambda^{*}$

$\underline{1}(*_1,1_2) = (1,0,-1,1,0,1)'$ $\qquad \lambda^{B}_{1}$

$1(*_1,2_2) = (0,1,-1,0,1,-1)'$ $\qquad \lambda^{B}_{2}$

$\underline{1}(1_1,*_2) = (1,1,1,-1,-1,-1)'$ $\qquad \lambda^{A}_{1}$

$\underline{1}(1_1,1_2) = (1,0,-1,-1,0,1)'$ $\qquad \lambda^{AB}_{11}$

$\underline{1}(1_1,2_2) = (0,1,-1,0,-1,1)'$ $\qquad \lambda^{AB}_{12}$

FIENBERG and MASON (1977) derive mixed parametrisation for a particular table with polytomous variables by noting how some vectors in regressor matrix R_{*} of unidentified model are linearly dependent. They obtain required results by suitable manipulation of dependent vectors and also recognition of the character of dependencies. Such approach implicitely uses the parametrization matrix which defines λ's as given functions of ρ_{*}'s and thus defines vectors of $\underset{\sim}{L}$ matrix as functions of vectors of $\underline{R}_{*}$ matrix. In other words, particular version of relation (2.7) is utilised in this approach, but only implicitely. The construction by direct product seems to be not only more general but also simpler and more elegant.

5. Fixed Reference Parametrisation

One of the ways to obtain identified parametrisation from unidentified hierarchical decomposition like (1.2) or (1.3) is to omitt those columns of the matrix $R_{\tilde{*}}$ which are linearly dependent on the previous ones, while starting with intercept column. This leads to γ parametrisation, which is easy to obtain in a 2x2 case by omitting from $R_{\tilde{*}}$ in (2.1) columns corresponding to, say, ρ^{A}_{2*}, ρ^{B}_{2*}, ρ^{AB}_{22}. This procedure is typically done when introducing dummy variables in regression model. Omitted categories are sometimes called reference categories, see NIE et al. (1975:374). The choice of reference category is arbitrary and thus can be made on the grounds of convenience in interpretability. Corresponding to the popularity of this parametrisation in usual regression analysis is its usefulness in linear "ANOVA" models for unbalanced data, see BURKE and SCHUESSLER (1974). Its computational implementation is straightforward, and resulting system of model equations, like (2.16) in a three-way case, is particularly easy to solve.

(5.1)
$$\gamma^{*} = \phi_{abc}$$

$$\gamma^{A}_{i} = \phi_{ibc} - \phi_{abc}, \ldots,$$

$$\gamma^{AB}_{ij} = (\phi_{ijc} - \phi_{ajc}) - (\phi_{ibc} - \phi_{abc}), \ldots,$$

$$\gamma^{ABC}_{ijk} = \{(\phi_{ijk} - \phi_{ajk}) - (\phi_{ibk} - \phi_{abk})\} - \{(\phi_{ijc} - \phi_{ajc}) - (\phi_{ibc} - \phi_{abc})\} \quad .$$

From properties of logarithms we can immediately express the parameters in terms of odds ratios

$$(5.2) \qquad \gamma^{*} = \ln \pi_{abc}$$

$$\gamma^{A}_{i} = \ln w^{i}_{abc} \equiv \ln f^{i}_{abc}$$

$$\gamma^{AB}_{ij} = \ln w^{ij}_{abc} \equiv \ln f^{ij}_{abc}$$

$$\gamma^{ABC}_{ijk} = \ln w^{ijk}_{abc} \equiv \ln f^{ijk}_{abc} \quad .$$

Whereas λ effects are complex functions of mixed components, γ effects are particularly simple functions of fixed components. γ's are just logarithms of appriopriate fixed components of the order equal to the order of given effect.

γ^{A}_{i} is the logarithm of 0-th order odds of A_i with respect to A_a in the stratum A_b, A_c. γ^{AB}_{ij} is logarithm of 1st order odds of A_i, B_j with respect to A_a, B_b in the stratum C_c. γ^{ABC}_{ijk} is the logarithm of 2nd order odds of A_i, B_j, C_k with respect to $A_a, B_b C_c$. Exponentiating (2.17) we obtain π_{ijk} expressed as multiplicative functions of, say, $\zeta \equiv \ln \sigma$, ζ being respective odds ratios.

We will not discuss the interpretation of γ or ζ parameters. It seems, however, that under variety of structural hypothesis, fixed reference parameters will have simpler and more useful interpretation than mixed reference parameters. On the grounds of parameters' interpretability, the fixed parametrisation seems to be preferred to the mixed one.

Elementary basis matrix of fixed parametrisation

$$\underline{G}_{C_j} = \{\underline{g}_j(*_j)\underline{g}_j(1_j)\ldots\underline{g}_j(C_j-1)\}$$

of dimension $C_j \times C_j$ has the following structure

if $c_j =$ then all elements of $\underline{g}_j(c_j)$ are 1

$c_j \neq$ then { c_j-th element of $\underline{g}_j(c_j)$ equals 1
all other elements equal 0.

Hence the elementary basis matrix can be written as

$$\underline{G}_j = \begin{bmatrix} \underline{1} & \underline{I} \\ 1 & \underline{0}' \end{bmatrix}$$

where $\underline{I}$ is $(C_j-1)\times(C_j-1)$ identity matrix.

For illustrative purposes we will construct basis matrix $\underline{G}$ for a 2x3 table. From "elementary" vectors

$\underline{g}_2(*_2) = (1, 1, 1)'$ $\qquad$ $\underline{g}_1(*_1) = (1, 1)'$

$\underline{g}_2(1_2) = (1, 0, 0)'$ $\qquad$ $\underline{g}_1(1_1) = (1, 0)'$

$\underline{g}_2(2_2) = (0, 1, 0)'$

we obtain columns of $\underline{G}$ by using (2.10) and specify corresponding parameters by rules given in Part 4.

$\underline{g}(*_1,*_2) = (1,1,1,1,1,1)'$ $\qquad$ γ^*

$\underline{g}(*_1,1_2) = (1,0,0,1,0,0)'$ $\qquad$ γ^B_1

$\underline{g}(*_1,2_2) = (0,1,0,0,1,0)'$ $\qquad$ γ^B_2

$\underline{g}(1_1,*_2) = (1,1,1,0,0,0)'$ $\qquad$ γ^A_1

$\underline{g}(1_1,1_2) = (1,0,0,0,0,0)'$ $\qquad$ γ^{AB}_{11}

$\underline{g}(1_1,2_2) = (0,1,0,0,0,0)$ $\qquad$ γ^{AB}_{12}

It should be noted that elementary basis of fixed parametrisation is not included in repertoire of most popular basis matrices for linear models, see Bock (1975:300).

6. Conclusions and extensions

In Parts 4 and 5 two main parametrisations were presented and characterised with respect to how they involve interaction components. A comparison of fixed and mixed parametrisations should also take into account issues of specifying structural hypotheses and estimating parameters.

In the context of standard linear model[7)] the maximum likelihood method is usually used for estimation. Its use is justified not only on theoretical grounds of optimality properties of estimators but also on practical grounds as many easily available and efficient computational algorithms give maximum likelihood estimates.(So called)Iterative Proportional Fitting Algorithm (IPF) is probably simplest and most popular among those algorithms. It is extensively discussed in most texts (f.e. Bishop et al.,1975).

In case of λ or γ parametrisations there is a very close connection between maximum likelihood estimation and generation of structural hypotheses, because every hierarchical hypothesis can be equivalently discribed either by a set of terms present or a set of estimated (sub)margins of the table equal to empirical (sub)margins.

This equivalence is based on the fact that every structural hypothesis which is hierarchical has specific probabilistic interpretation (for example mutual independence when margins A,B, C are fitted), which can be expressed either by λ functions of mixed components or γ functions of fixed components.
In Part 3 it was shown that fixed and mixed components are equivalent with respect to hypotheses of homogeneity, independence and 2-nd order interaction. Similar results hold for more complicated hypotheses in more dimensions. We will not show that, in fact, probabilistic interpretations of structural hypotheses are the same whether one uses λ or γ parametrisation. This more general result is, however, intuitive in the light of equivalencies noted in Part 3.

The explicit treatment of hypotheses for three-dimensional case can be found in Bishop et al.(1975) for λ parameters and in Plackett(1974) for γ parameters. In three dimensions there is 8 different classes of hypotheses, containing typically the following fitted margins: AB,BC,AC; AC,AB;AB,C;AB;A,B,C; A,B, N. Note that there are f.e. three hypotheses of the second kind, but only one hypothesis of the first kind.

The 7 classes of hypotheses have thus the same probabilistic interpretation regardless of the use of λ's or γ's. If a hypothesis contins ABC as a fitted margin (thus, if it contains $\bar{\lambda}^{ABC}_{ijk}$ or $\bar{\gamma}^{ABC}_{ijk}$) a perfect fit will be obtained in a three-dimensional case, which corresponds to a saturated model.

The invariance of hypothesis interpretation with respect to the choice between γ's or λ's comes from the fact that, as long as hierarchical hypothesis are concerned for fixed or mixed parametrisation, it is satisfactory and necessary to specify them in terms of fitted margins in maximum likelihood estimation. This fact is consequently displayed when using Iterative Proportional Fitting procedure, as well as when more general algorithms are used, see f.e. NELDER's discussion of Iterative Weighted Least Squares, (NELDER, 1974).

It should be noted that first estimated $\hat{\phi}$'s are obtained from IPF and then estimated parameters $\underset{\sim}{\hat{\rho}}$ can be calculated from

$$\underset{\sim}{R}^{-1}\underset{\sim}{\hat{\phi}} = \underset{\sim}{\hat{\rho}} \quad . \tag{6.1}$$

For $\underset{\sim}{R} = \underset{\sim}{L}$ or $\underset{\sim}{R} = \underset{\sim}{G}$ estimated $\hat{\phi}$'s will implicitely satisfy the conditions specifying given hypothesis in parametric way,i.e. that certain terms are zero.

Concerning the comparison of fixed and mixed parametrisations we conclude that the interpretation of parameters is the only criterion which matters, and that according to this criterion the fixed parametrisation will often be preferred on the grounds of simplicity. It would be desirable that existing computer programs performing IPF algorithm gave the user the opportunity to choose between these two parametrisations.

It should be also noted that because of the correspondence between fixing terms at 0 and fitting the margins fixed and mixed parametrisations are in a sense "natural".

The "reverse" way of obtaining parameter estimates does not prohibit wide choice of alternative "non-natural" parametrisations, both for saturated and unsaturated models. IPF procedure gives the possibility first to specify the hypothesis, obtain ϕ's and then solve (6.1) for appropriate parameters given choice of parametrisation. However, it seems that it is not necessary that certain other functions than λ or γ will equal 0 under given structural hypothesis. F.e. in a 2x2 case $\lambda_{11}^{AB}=0 \equiv \phi_{11}+\phi_{22}-\phi_{12}-\phi_{21} = 0$, but $2\phi_{11}+\phi_{22}-\phi_{12}-\phi_{21} = 0$ only if $\phi_{11}+\phi_{22}-\phi_{12}-\phi_{21} = -\phi_{11}$. Then, if such "non-natural" functions appear in R matrix, it is necessary to omitt certain terms from $\underset{\sim}{R}$ in (6.1) which reduces its dimensionality and makes $\underset{\sim}{R}$ automatically noninvertible.Thus, (6.1) would apply only to saturated model.

Fortunately the possibility of parametrisation choice can be discussed in slightly more general way.

From (2.2) we know, that $\underset{\sim}{\phi}=\underset{\sim}{R}_{*}\underset{\sim}{\rho}_{*}$ and from (2.8) that $\underset{\sim}{P} = (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}^{1}\underset{\sim}{R}_{*}$, so finally

$$\begin{aligned}\underset{\sim}{\rho} &= (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}'\underset{\sim}{R}_{*}\underset{\sim}{\rho}_{*} \\ &= (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}'\underset{\sim}{\phi} \\ &\equiv \underset{\sim}{T}\underset{\sim}{\phi}\end{aligned}$$

The inverse $(\underset{\sim}{R}'\underset{\sim}{R})^{-1}$ always exist because every $\underset{\sim}{R}$ is of full column rank. Consequently, for given $\underset{\sim}{\phi}$, there exists a vector of parameters estimates ρ, which can be calculated on the basis of (6.2) from $\underset{\sim}{\phi}$ obtained by IPF under a structural hypothesis. Thus, after estimation of $\underset{\sim}{\phi}$, the choice of parametrisation can be made and ρ can be obtained. It is not necessary that M=W, because (6.2) holds for any $M \leq W$ giving $\underset{\sim}{T}(MxW)$. The rank of $\underset{\sim}{R}$ must be, however, equal to the rank of structural hypothesis. Moreover, $\underset{\sim}{T}$ matrix should be obtained from regressor matrix properly constructed according to principles presented in Part 2.[8)]

Any of the typical parametrisations used in linear model context can be potentially useful in expressing hierarchical structural hypothesis. Elementary basis matrices for typical parametrisations are presented and discussed, for example, in BOCK (1975, Ch. 5) or FINN (1974).

It should be noted, that it is possible to consider separate "parametrisations" for different terms or subsets of terms of the model which would lead in a two-way case to, say, λ parametrisation of main effects of A, γ parametrisation of main effects of B and "mixed" parametrisation of 1st order interaction effects.

There is, however, not much evidence concerning exploration of possibilities alternative parametrisations offer, both in the theoretical and applied literature. One of the reasons could be a relative lack of attention given to the interpretation of parameters, even in case of two "natural" parametrisations discussed above. Moreover, the unavailability of closed-form expressions for other parameters in terms of meaningful components (like interaction components) adds to difficulties in interpreting parameters or justifying "non-natural" parametrisations.

Within standard log-linear model more general maximum likelihood estimation procedures allow any parametrisation which leads to hypothesis specified by a regressor matrix of full rank. In special case nonhierarchical hypothesis can be considered, however, they present particular difficulties in interpretation (HABERMAN, 1974a).

Hierarchical hypothesis based on orthogonal (unnormalized) or orthonormal (normalized) polynomial parametrisation can be considered as another special case.[9] Orthogonal polynomials parametrisation seems to be particularly suitable when categories of variables are ordered and it is meaningful to assign to them integer values assuming usually equal (or possibly nonequal) spacing. This topic has been treated in some detail by HABERMAN (1974b), see also GOODMAN (1978).

More general procedures allow also ML estimation under finer constraints on parameters, when the dimension of regressor matrix is not reduced. KULLBACK and FISHER (1973) give interesting illustration of such constraints.

A trade-off should be noted between the simplicity of structural hypothesis and complexity of a parametrisation. With IPF the choice of parametrisation is constrained to a set of hypothesis having definite probabilistic interpretations. In this case the choice of parametrisation pertains to representation of given hypothesis, f.e. to representation of 1st order interaction. With more general procedures the choice of parametrisation is not constrained to such well defined hypothesis (like in case of nonhierarchical hypothesis) and interpretation of parameters becomes more difficult but also more important. It seems that traditional techniques of displaying results of estimation for linear (ANOVA) models can be helpful here, see their brief review in FINN(1974) and an example of application to a log-linear model by LITTLE (1978).

There is a great variety of models which could be described as non-standard log-linear models in the sense of

i) specifying more complex functions $\underset{\sim}{\phi}$ of log-probabilities (or log-means) of multivariate distributions

ii) considering disjoint variables, with some (usually one) of them being responses conditional on values of certain factors

iii) specifying more complex regressor matrices than $\underset{\sim}{R}_*$.

The issues of basis matrix construction and comparison between fixed and mixed parametrisations discussed in this paper in some detail are also fundamental in more complex context of various non-standard models. For example analysing factorial structures of the type (2.1) for ϕ's being more complex functions of probabilities also leads to the choice of parametrisation. Consequently the interpretation of λ or γ parameters of an non-standard model will be only a modification of results derived with respect to $\phi = \ln \pi$. In many cases the construction of regressor matrix from elementary basis matrices is required, see f.e. LANDIS et al. (1976), which makes the details of such construction of practical value to the analyst.

Some expecially important references to (i) models are LANDIS et al. (1976) and to (ii) and (iii) models BOCK (1975) and NERLOVE and PRESS (1976)*).

---

*) The recent volume of Sociological Methodology 1979 (San Francisco: Jossey-Bass) contains two references relevant to the topic of this paper (by FIENBERG, EVERS and NAMBOODIRI). I have been made aware of their existence after completion of this paper.

NOTES:

1 There seem to be some confusion concerning this second aspect. Davis, for example, stressing the importance of parametric character in the first sense, praises log-linear models for being "new lines in nonparametric methodology" (DAVIS,1972:37).

2 "Term" is understood to be a set of parameters with common configuration of variables, f.e. λ_i^A, $i \leqslant I$, or λ_{ij}^{AB}, $i \leqslant I$, $j \leqslant J$. Such sets are denoted respectively by $\bar{\lambda}_{ij}^{AB}$ and $\bar{\lambda}_{ij}^{AB}$. Thus the term $\bar{\lambda}_i^A$ is marginal to $\bar{\lambda}_{ij}^{AB}$ because configuration A is a subset of configuration AB. Detailed explanation of configurations appear in BISHOP et al. (1975) and of marginality in NELDER (1974).

3 To make this statement more explicit note that for, say, $w=\langle i,j \rangle$ we can write

$$\varphi_{ij}^* = \rho_*^* + F_w(\underline{\rho}_*^*)$$

where elements of $\underline{\rho}_*^*$ contain all other parameters than intercept ρ^*. Taking logarithm of $N\pi_{ij}$ we get

$$\begin{aligned} \varphi^* &\equiv \ln N\pi_{ij} \\ &= \ln N + \varphi_{ij} \\ &= \ln N + \rho_*^* + F_w(\underline{\rho}_*^*) \end{aligned}$$

which does not affect any parameter from (1.2) except the intercept.

4 The set of probabilities π_i^A, $i \leqslant I$ is denoted by $\bar{\pi}_I^A$, where superscipt A over i indicates that i index varies. This notation is applied to more general situations in an obvious way.

5 This is DARROCH's terminology, see DARROCH(1962).

6 A set of relations will be symbolised by using set operator "-" ; thus $\bar{\omega}_i^A = 1$ means $\omega_i^A = 1$ for all i. We also accept a convention that an arithmetic mean over a subscript is denoted by "+" and a geometric mean by "x". Further, a sum is concisely symbolised by " 0 " and a product by ".".

7 Model discussed with reference to Iterative Proportional Fitting assumes that $\phi_w = \ln N\pi_w$ which, as we shown before, does not affect interpretation of any parameters except intercept.

8 ECTA, one of the most popular programs for standard log-linear modelling, allows some reparametrizations based on (6.2). According to program description, see Goodman and Fay(1973), a situation analysed by ECTA (card 52) occurs when elementary basis matrix for variable $j \leq J$, say $\underline{R}_j$ has orthogonal columns. Each column of $\underline{R}_j = (\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_M)$ is further divided by its squared norm (length) $\underline{r}_m'\underline{r}_m$; in this sense $\underline{R}_j$ is normalized. It follows from column - wise orthogonality that the typical element of $(\underline{R}_j'\underline{R}_j)$ is 0 for off-diagonal elements and $\underline{r}_m'\underline{r}_m$ for diagonal elements. Thus $(\underline{R}_j'\underline{R}_j)$ is a diagonal matrix, its inverse is also diagonal with typical element $(1/\underline{r}_m'\underline{r}_m)$.

Let $\underline{D}_j^{-1}$ be diagonal $\{1/\underline{r}_1'\underline{r}_1, \ldots, 1/\underline{r}_M'\underline{r}_M\}$, and note that dividing columns of $\underline{R}_j$ by $\underline{r}_m'\underline{r}_m$ is equivalent to $\underline{R}_j\underline{D}_j^{-1}$, thus $(\underline{R}_j\underline{D}_j^{-1}) = \underline{D}_j^{-1}\underline{R}_j' \equiv \tilde{\underline{R}}_j'$, where $\tilde{\underline{R}}_j$ denotes normalized matrix. Consequently $(\underline{R}_j\underline{R}_j')^{-1} = \underline{D}_j^{-1}$ and we notice that due to column-wise orthogonality assumption ECTA avoids computing the inverse of nondiagonal $(\underline{R}_j'\underline{R}_j)$ matrix. However, this feature prohibits the input of elementary basis matrices for fixed parametrisation. To demonstrate this we consider simplest situation with one trichotomous variable so regressor matrix $\underline{R}$ reduces to elementary basis matrix, say, $\underline{G}_3 = (\underline{g}_1, \underline{g}_2, \underline{g}_3)$

$$\underline{G}_3 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

We immediately note that columns $\underline{g}_1, \underline{g}_2$ and $\underline{g}_1, \underline{g}_3$ are not orthogonal. ECTA does not check the orthogonality of columns which can lead to obtaining results not expected by the user. When $\underline{G}_3$ was inputed, program generated $\underline{T}$ matrix equal to

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

which is different from the correct $\underline{T}$ matrix (equal to the inverse of $\underline{G}_3$ because our model is saturated)

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} .$$

Thus, given that model $\underline{T}$ matrix is correctly generated from "elementary" $\underline{T}_j$ matrices, it should be still possible to obtain fixed parametrisation. Unfortunately, ECTA does not allow the input of full $\underline{T}_j$ matrix but only the input of $\underline{T}_j$ with first row omitted (card 51). Noting different rows in two matrices above we see that in one-variable case only the interpretation of intercept term would be affected. If construction of T matrix has to give correct results, and thus equivalent to construction through elementary basis matrices described in this paper, we would expect that in multivariable case incorrect first rows of $\underline{T}_j$ will as well affect the interpretation of other parameters. Regardless of those limitations, reparametrisation options of ECTA allow orthogonal and orthonormal polynomials with equal and non-equal spacing.

9 Orthogonal polynomials parametrisation can of course be employed to some extent also with IPF and using (6.2). IPF demands, however, that for C_j category variable all (C_j-1) polynomial effects have to be included. In three category case the hypothesis of no quadratic (main) effect can not be directly formulated, which is possible with more general procedures like Iterative Weighted Least Squares.

Appendix: Proof of theorem from § 4

Proof is very easy and based on a simple fact that $w^{i'}_{i}$, a ratio of two identical probabilities, is equal to 1. Consider first (i): from (4.1) we have

$$\begin{aligned}
\lambda^A_i &= \phi_{i++} - \phi_{+++} \\
&= (KJ)^{-1}\phi_{i00} - (IJK)^{-1}\phi_{00} \\
&= I(IJK)^{-1}\phi_{i00} - (IJK)^{-1}\phi_{00} \\
&= (IJK)^{-1}\{I\phi_{i00} - \phi_{000}\} \\
&= (IJK)^{-1}\{\phi_{i00} + (I-1)\phi_{i00} - (\phi_{a_1 00} + \sum_{r=2}^{I}\phi_{r00})\}
\end{aligned}$$

Index r is such that $r \neq a_i, i \leq I-1$ so we can successively eliminate $a \neq i$ until the "last" index a_I which equals i and thus $\phi_{i00} - \phi_{a_I 00} = 0$.

Denoting $\phi_{i00} - \phi_{a00}$ by ϕ^i_{a00} we finally get

$$\begin{aligned}
\lambda^A_i &= (IJK)^{-1}\{\phi^i_{a_1 00} + \phi^i_{a_2 00} + \ldots + \phi^i_{a_{I-1} 00} + \phi^i_{a_I 00}\} \\
&= (IJK)^{-1}\{\sum_{ajk}\sum\sum\phi^i_{ajk}\} \\
&= (IJK)^{-1}\sum_{ajk}\sum\sum \ln \pi_{ijk} - \ln \pi_{ajk} \\
&= (IJK)^{-1}\sum_{ajk}\sum\sum \ln w^i_{ajk} \\
&= (IJK)^{-1}\sum_{jk}\sum \ln \prod_{a} w^i_{ajk} \\
&\equiv (IJK)^{-1}\sum_{jk}\sum \ln m^i_{.jk}
\end{aligned}$$

It is easier to consider (ii) by using the equation $\lambda^{AB}_{ij} = \lambda^{A}_{ij} - \lambda^{A}_{i}$. We know by now that $\lambda^{A}_{ij} = (I\ K)^{-1}\Sigma\Sigma\phi^{i}_{ajk}$. Consequently

$$\lambda^{AB}_{ij} = (IJK)^{-1}\{J\Sigma_a\ \phi^{i}_{ajO} - \Sigma\Sigma_{aj}\ \phi^{i}_{ajO}\}$$

$$= (IJK)^{-1}\ \{\Sigma_a\phi^{i}_{ajO}+(J-1)\Sigma_a\phi^{i}_{ajO}-(\Sigma_a\phi^{i}_{ab_1O}+\Sigma\Sigma_{sa}\phi^{i}_{asO})\}$$

Similarly as before let index s be $s\neq b_j$ $j\leq J-1$ so we successively eliminate $b_j\neq j$ until the "last" giving

$$\phi^{i}_{ajO}-\phi^{i}_{abjO} \equiv \phi^{ij}_{ajO} = 0 \ :$$

$$\lambda^{AB}_{ij} = (IJK)^{-1}\ [\Sigma_a\{(\phi^{i}_{ajO}-\phi^{i}_{ab_1O})+(\phi^{i}_{ajO}-\phi^{i}_{ab_2O})+\ldots+(\phi^{I}_{ajO}-\phi^{i}_{ab_{J-1}O})\}]$$

$$= (IJK)^{-1}\ \Sigma_a\ \Sigma_{b\neq j}\phi^{i}_{ajO}-\phi^{i}_{abO}$$

$$= (IJK)^{-1}\ \Sigma\Sigma_{ab}(\ln w^{i}_{aj.}-\ln w^{i}_{ab.})$$

$$= (IJK)^{-1}\Sigma\Sigma_{ab}\ \ln(w^{ij}_{ab.})$$

$$\equiv (IJK)^{-1}\ln(m^{ij}_{..})$$

Finally, we will use the equation $\lambda^{ABC}_{ijk}=\lambda^{AB}_{ijk}-\lambda^{AB}_{ij}$. We know from (ii) that $\lambda^{AB}_{ijk}=(IJ\ )^{-1}\Sigma\Sigma_{ab}\phi^{ij}_{abk}$ so

$$\lambda^{ABC}_{ijk} = (IJK)^{-1}\{K\Sigma\Sigma_{ab}\phi^{ij}_{abk}-\Sigma\Sigma\Sigma_{abk}\phi^{ij}_{abk}\}$$

$$= (IJK)^{-1}[\Sigma\Sigma_{ab}\phi^{ij}_{abk}+(K-1)\Sigma\Sigma_{ab}\phi^{ij}_{abk}-\Sigma\Sigma_{ab}\phi^{ij}_{abc_1}+\Sigma\Sigma\Sigma_{abt}\phi^{ij}_{abt}]$$

$$= (IJK)^{-1}[\Sigma\Sigma_{ab}\{(\phi^{ij}_{abk}-\phi^{ij}_{abc_1})+(\phi^{ib}_{ajk}-\phi^{ij}_{abc_2})+\ldots+(\phi^{ib}_{ajk}-\phi^{ij}_{abc_{K-1}})\}]$$

$$= (IJK)^{-1}\Sigma\Sigma_{ab}\Sigma_{c\neq k}(\phi^{ij}_{abk}-\phi^{ij}_{abc})$$

$$= (IJK)^{-1}\sum_{abc}\sum\sum(\ln w^{ij}_{abk} - \ln w^{ij}_{abc})$$

$$= (IJK)^{-1}\sum_{abc}\sum\sum\ln(w^{ijk}_{abc})$$

$$\equiv (IJK)^{-1}\ln(m^{ijk}_{\cdot\cdot\cdot})$$

Again, $t \neq c_k$, $k<K-1$ and the "last" expression is eliminated because $c_K = k$. The generalization of theorem to arbitrary n-way tables could be done by induction and would not involve any new ideas.

BIBLIOGRAPHY:

Birch,M.W. (1963) Maximum likelihood in three-way contingency tables. J. Roy. Statist. Soc. B 25, 22o-33.

Bock, R.D. (1975) Multivariate Statistical Methods in Behavioral Research. New York:MIT

Bishop, Y.M., Fienberg, S.E., Holland, P.W. (1975) Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass.:M I T

Burke, P.J., Schuessler, K.(1974) Alternative approaches to analysis of variance tables, pp. 145-88 in: H.L. Costner (Ed.) Sociological Methodology 1973-1974. San Francisco: Jossey-Bass

Darroch, J.N. (1962) Interactions in multi-factor contingency tables.J.Roy.Statist. Soc.B 24, 251-63

————(1974) Multiplicative and additive interaction in contingency tables. Biometrika 61, 2o7-14

Davis, J.A. (1972) The Goodman log-linear system for assessing effects in multivariate contingency tables. University of Chicago: National Opinion Research Center

————(1974) Hierarchical models for significance tests in multivariate contingency tables: an exegesis of Goodman`s recent papers; pp. 189-231 in: H.L.Costner (Ed.) Sociological Methodology 1973-1974. San Francisco:Jossey-Bass

Everitt, B. (1977) The Analysis of Contingency Tables. London: Chapman and Hall

Finn, J. (1974) A General Model for Multivariate Analysis. New York: Holt, Rinehard and Winston

Fienberg, S.E. (1977) The Analysis of Cross-Classified Categorical Data. Cambridge, Mass.:M I T

————, Mason, W.M. (1977) Identification and estimation of age-period-cohort models in the analysis of discrete archival data. University of Minnesota, School of Statistisc: Technical Report No. 286

Goodman, L.A. (1970) The multivariate analysis of qualitative data: interactions among multiple classofications. J. Amer. Stat. Assoc.,65, 226-56

———— (1971) The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics 13, 33-61

Haberman, S.J. (1974a) The Analysis of Frequency Data. Chicago: University of Chicago Press

————(1974b) Log-linear models for contingency tables with ordered classifications. Biometrics 30, 589-600

Ku, H.H., Kullback,S. (1974) Log-linear models in contingency table analysis. The American Statistician 28,115-22

Kutylowski, A. (1979) Log-linear probability models. Ch. 4 in: Structural models of relationships among qualitative variables and their sociological applications. In preparation.

Lewis,B.N. (1962) On the analysis of interaction in multidimensional contingency tables. J.Roy. Statist. Soc., A 125, 88-117

Mantel, N.(1966) Models for complex contingency tables and polychotomous dosage response curves. Biometrics 22,83-95

Maxwell, A.E.(1961) Analyzing Qualitative Data. London: Methuen

Nelder, J.A. (1974) Analysis of variance pragrams, least squa-

res and square tables (letter to the Editor). Applied Statistics 23,232

———— (197) General Linear Interactive Modelling: Glim manual. Oxford: Numarical Algorithm Group

Nie,N.H., Hull,C.H., Jenkins, J.G., Steinbrenner, K., Bent, D.H. (1975) Statistical Package for the Social Sciences. New York: McGraw-Hill

Nerlove, M., Press, J.S. (1976) Multivariate log-linear probability models for the analysis of qualitative data. Discussion Paper No.1. Northwestern University, Center for Statistics and Probability

Page, W.F. (1977) Interpretation of Goodman's log-linear model effects. An odds ratio approach. Sociological Methods and Research., 5, 419-35

Plackett, R.L. (1974) Analysis of Categorical Data. New York: Hafner

Roy, S.N., Kastenbaum, M.A. (1956) On the hypothesis of no interaction in a multiway contingency table. Ann. Math. Statist.,27,749-57

Scheffè ,H. (1959) The Analysis of Variance. New York: Wiley

Searle, S.R. (1971) Linear Models. New York:Wiley

Simpson, E.H. (1951) The interpretation of interaction in contingency tables. J.Roy. Statist. Soc. B 13, 238-41

Theil,H.(1979) On the estimation of relationships involving qualitative variables. Amer. J. Sociol. 76, 103-54

Upton, G.J.G. (1978) The Analysis of Cross-tabulated Data. Chichester: Wiley

Nelder, J. (1974) Log-linear models for contingency tables: generalisation of classical least squares. Applied Statistics 23, 323-29

Goodman, L.A.(1978) Simple models for the analysis of association in cross-classifications having ordered row and column categories. Technical Report No 61. University of Chicago: Department of Statistics

Kullback, S., Fisher, M. (1973) Partitioning second order interaction in three-way contingency tables. Applied Statistics 22,172-84

Little, R.J.A. (1978)Generalised linear models for cross-classified data from World Fertility Survey. World Fertility Survey Technical Report

Landis, R.J., Stanish, W.M., Freeman,J.L., Koch, G.G. (1976) A computer program for the generalised chi-square analysis of categorical data using weighted least squares (GENCAT). Computer Programs in Biomedicine 6, 196-231

Goodman, L.A. Fay, R.(1973) ECTA Program. Description for Users (December). University of Chicago: Department of Statistics