c1

AN ANALYSIS OF MULTIDIMENSIONAL CONTINGENCY TABLES

by

Lilian Mast (née Feuerverger)

B. Sc. Hons., McGill University, 1969

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

APRIL, 1973

Department of Mathematics

The University of British Columbia
Vancouver 8, Canada

Date February 1, 1973.

Abstract

In this thesis we consider the following model for a three-dimensional $r \times s \times t$ contingency table:

$$E(f_{ijk}) = np_{ijk} = n[p_{i..}p_{.j.}p_{..k} + p_{i..}\alpha_{jk} + p_{.j.}\beta_{ik} + p_{..k}\gamma_{ij} + \delta_{ijk}] \geq 0$$

$i = 1,2,..,r;\ j=1,2,..,s;\ k=1,2,...,t$ with

$$\sum_{j=1}^{s} \alpha_{jk} = \sum_{k=1}^{t} \alpha_{jk} = 0,\ \sum_{i=1}^{r} \beta_{ik} = \sum_{k=1}^{t} \beta_{ik} = 0,\ \sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0$$

and $\sum_{i=1}^{r} \delta_{ijk} = \sum_{j=1}^{s} \delta_{ijk} = \sum_{k=1}^{t} \delta_{ijk} = 0$. A dot indicates summation over the replaced subscript. The f_{ijk}'s represent the frequencies and the p_{ijk}'s represent the proportions. The problem we are concerned with is testing the hypothesis H_0: $\delta_{ijk} = 0$ for all i, j, k. i.e. no second order interaction is present. We then seek to extend the model and problem to a w-way table.

We use the method of the likelihood ratio. To assist us in determining the numerator of the likelihood ratio we reformulate a theorem about constrained extrema and Lagrange multipliers and prove this reformulation.

Some general conclusions we draw are: there are two extensions to our 3-way model; results we obtain using our model and methods are in close agreement with results obtained using the models and methods of other statisticians.

TABLE OF CONTENTS

TABLES

ACKNOWLEDGMENT

The author wishes to express her sincere appreciation to Dr. Stanley W. Nash of the Department of Mathematics, University of British Columbia for suggesting the topic of this thesis and for the many hours of guidance and comments during its preparation. She would also like to thank Dr. N. Glick of the Department of Mathematics for reading the final manuscript.

She is also pleased to acknowledge the support of the University of British Columbia for its financial assistance.

Finally, the author wishes to thank Eleanor Lannon and Mary Daisley for their assistance in the typing of this thesis.

CHAPTER I INTRODUCTION

The cross classification of a series of observations according to w characteristics, or attributes, is referred to as a multidimensional contingency table. The entries in the cells of a contingency table are frequencies.

The usual hypotheses investigated in contingency tables are about kinds of stochastic independence or lack of it. The descriptions of stochastic independence for several partitions of the sample space are those given by Kolmogorov (and by Feller for several events). Similar descriptions hold for density functions and cumulative distribution functions of several random variables.

In this thesis we investigate how to handle and analyze multidimensional contingency tables. Quite a bit has been done in recent years, but a lot still remains to be done. We concern ourselves basically with definitions of interaction.

Before we present our model and definition of interaction we will mention, very briefly, the models and definitions of interaction of some of the statisticians who were and may still be concerned with the analysis of contingency tables. H. H. Ku and S. Kullback investigated the problem of interaction in multidimensional contingency tables from the viewpoint of information theory as developed by Kullback. The hypothesis of no rth-order interaction is defined in the sense of an hypothesis of "generalized" independence of classifications with fixed r-th order marginal restraints. For a three-way table, with given cell probabilities p_{ijk}, the minimum discrimination information for a contingency table with marginals $p_{ij.}$, $p_{.jk}$, $p_{i.k}$ is given by the set of cell probabilities $p^*_{ijk} = a_{ij} b_{jk} c_{ik} p_{ijk}$ where a_{ij}, b_{jk}, and c_{ik} are functions of the given marginal probabilities. $\ell n(p^*_{ijk}/p_{ijk}) = \ell n\, a_{ij} + \ell n\, b_{jk} + \ell n\, c_{ik}$, represents no 2nd-order interaction. The minimum discrimination information statistic, asymptotically distributed as χ^2 with appropriate degrees of freedom, is

$$2 \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} \,\ell n\, f_{ijk} - 2 \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} \ln f^*_{ijk} \geq 0,$$ where f_{ijk} are the observed cell frequencies and f^*_{ijk} are the "no interaction" cell frequencies uniquely determined by a simple convergent iteration process of the marginals on p_{ijk}.

Bartlett defined no second order interaction in a 2×2×2 table as $p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}$ where the p's are the cell probabilities. Roy and Kastenbaum extended Bartlett's definition and arrived a set of "no interaction constraints" in an r×s×t table in the form of

$$\frac{p_{rst}p_{ijt}}{p_{ist}p_{rjt}} = \frac{p_{rsk}p_{ijk}}{p_{isk}p_{rjk}} \quad \text{for} \quad \begin{array}{l} i = 1, 2, \ldots, (r-1) \\ j = 1, 2, \ldots, (s-1) \\ k = 1, 2, \ldots, (t-1) \end{array}$$

The "mechanism" used by Roy and Kastenbaum is based on the fact that the two hypotheses

$H_1: p_{i.k} = p_{i..}p_{..k}$

$H_2: p_{ij.} = p_{i..}p_{.j.}$

will not imply $H: p_{ijk} = p_{i..}p_{.jk}$ in a 3-way contingency table. A dot indicates summation over the replaced subscript. The "no interaction" hypothesis is required to generate the set of constraints such that these constraints, when superimposed on $H_1 \cap H_2$ should imply H.

Simpson required the definition of "no-second-order interaction" to be symmetrical with respect to the three attributes of a 2×2×2 table. If some function $\psi(p_{111}, p_{121}, p_{211}, p_{221})$ is chosen to measure the association of classifications A, B and C, then the function must be such that the equation $\psi(p_{111}, p_{121}, p_{211}, p_{221}) = \psi(p_{112}, p_{122}, p_{212}, p_{222})$

$\Leftrightarrow \quad \psi(p_{111}, p_{211}, p_{112}, p_{212}) = \psi(p_{121}, p_{221}, p_{122}, p_{222})$

$\Leftrightarrow \quad \psi(p_{111}, p_{121}, p_{112}, p_{122}) = \psi(p_{211}, p_{221}, p_{212}, p_{222})$. He showed that the function $\psi = \dfrac{p_{121}p_{211}}{p_{111}p_{221}}$ or the cross product ratio used by Bartlett, satisfies this requirement.

Lancaster defined the second order interaction by the partition of the chi-square statistic χ^2, i.e. it is defined as the difference between the total χ^2 for testing complete independence of the three classifications, and the sum of the 3 components corresponding to tests for independence in each of the 3 marginal tables. Lancaster's definition does not always satisfy Simpson's condition of symmetry. In Chapter V of this thesis we will mention the model he gives in his book, which is much like our own, though we have used different approaches. We are in a way defending Lancaster against the criticisms of the preponents of the loglinear model (to be mentioned soon), while what he has done reinforces what we propose.

Darroch made an explicit comparison of the definitions of interaction in multiway contingency tables and in the analysis of variance. The main point he made was that a natural symmetrical definition of "no second-order interaction"

$$p_{ijk} = \frac{p_{.jk}p_{i.k}p_{ij.}}{p_{i..}p_{.j.}p_{..k}} \quad (1)$$

necessarily implies constraints on the marginal probabilities $p_{ij.}$, $p_{.jk}$, $p_{i.k}$,

i.e. $\sum_{k=1}^{t} p_{ijk} = p_{ij.} = \frac{p_{ij.}}{p_{i..}p_{.j.}} \sum_{k=1}^{t} \frac{p_{i.k}p_{.jk}}{p_{..k}}$

or $\sum_{k=1}^{t} \frac{p_{i.k}p_{.jk}}{p_{..k}} = p_{i..}p_{.j.} \;\forall\, i, j$ and the like.

This is undesirable since the condition for "no second-order interaction" should relate p_{ijk} to any given set of marginal probabilities and should not place restrictions on the latter. Consequently Darroch defined a "perfect three-way table" as one for which condition (1) and the resulting constraints on the marginal probabilities are satisfied exactly. He concluded further that "in imperfect tables it is not possible to express p_{ijk} in terms of simple functions of $p_{ij.}$, $p_{i.k}$, $p_{.jk}$ when there is no 2nd order interaction."

The existence and uniqueness of the set p_{ijk} as the solution of

$$\frac{p_{rst}p_{ijt}}{p_{ist}p_{rjt}} = \frac{p_{rsk}p_{ijk}}{p_{isk}p_{rjk}} \quad \text{for} \quad \begin{array}{l} i = 1, 2, \ldots, (r-1) \\ j = 1, 2, \ldots, (s-1) \\ k = 1, 2, \ldots, (t-1) \end{array}$$

for any given set of mutually consistent marginal probabilities was conjectured for $r{\times}s{\times}t$ tables and proved for the $2{\times}2{\times}2$ case. The search for a simple formulation in terms of parameters which are implicitly defined by the marginal probabilities led Darroch to define

$p_{ijk} = \mu A_{jk} B_{ki} C_{ij}$ where

$\sum_{k=1}^{t} A_{jk} = \sum_{i=1}^{r} B_{ki} = \sum_{j=1}^{s} C_{ij} = 1$ and $\mu \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} A_{jk} B_{ki} C_{ij} = 1$; and

to show that $\mu \doteq 1$, $A_{jk} \doteq \dfrac{p_{\cdot jk}}{p_{\cdot j \cdot}}$, $B_{ki} \doteq \dfrac{p_{i \cdot k}}{p_{\cdot \cdot k}}$ and $C_{ij} \doteq \dfrac{p_{ij \cdot}}{p_{i \cdot \cdot}}$.

Since there is no solution in closed form to the maximum likelihood equations for the parameters under hypothesis of no second-order interaction, unless the observed table happens to be perfect, Darroch suggested an interative solution and gave a numerical illustration using the example given by Kastenbaum & Lamphierar, [1959] .

It is of interest to note that Darroch suggested the likelihood ratio test based on $Z_{ABC} = 2 \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} \ell n \dfrac{f_{ijk}}{n\hat{\mu}\hat{A}_{jk}\hat{B}_{ki}\hat{C}_{ij}}$ which is asymptotically distributed as χ^2 with $(r-1) \times (s-1) \times (t-1)$ degrees of freedom.

Good proposed to use the principle of maximum entropy as a heuristic principle for the generation of null hypotheses, with main application to w-dimensional contingency tables. By using his principle, it is shown that for a w-dimensional $2 \times 2 \times \ldots \times 2$ contingency table with $p_i = p_{i_1 i_2 \ldots i_w}$, $i_1, \ldots, i_w = 0, 1$ and with all marginal probabilities down to (w-1)-way assigned, the null hypothesis to be tested is $\prod_i^{|i|\text{ even}} p_i = \prod_i^{|i|\text{ odd}} p_i$

where $|i| = i_1 + i_2 + \ldots + i_w$. This expression reduces to $p_1p_4p_6p_7 = p_2p_3p_5p_8$ ($p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}$) when $m = 3$. Good also generalized the definition to that of no rth-order and all higher-order interactions in a w-dimensional contingency table with a complete set of rth-order restraints by means of discrete Fourier transforms of the logarithms of probabilities. However, the interactions so defined are usually complex valued unless the number of categories within each classification is equal to two.

L. Goodman followed the definition by Good but proposed a test that yields real valued interactions. Goodman has published many papers on contingency tables, some of which are listed in the bibliography of this thesis.

Another method of analysis has implicitly been given by B. Woolf in the case of a 2×2×t table. Let the frequencies in the kth 2×2 table be denoted by f_{1k}, f_{2k} , f_{3k}, f_{4k}, where f_{1k}, f_{2k} occupy the first row and f_{1k}, f_{3k} the first column. Compute $Z_k = \ell n f_{1k} - \ell n f_{2k} - \ell n f_{3k} + \ell n f_{4k}$ and e_k from $\frac{1}{e_k} = \frac{1}{f_{1k}} + \frac{1}{f_{2k}} + \frac{1}{f_{3k}} + \frac{1}{f_{4k}}$.

If there is zero second-order interaction, then $\chi^2 = \sum_{k=1}^{t} e_k Z_k^2 - (\sum_{k=1}^{t} e_k Z_k)^2 / \sum_{k=1}^{t} e_k$ is asymptotically distributed as chi-squared with (t-1) degrees of freedom.

With unrestricted sampling conditions, M.W. Birch states that Roy and Kastenbaum's condition may be rewritten as $\ell n f_{ijk} = \ell n\ f_{rjk} + \ell n\ f_{isk} + \ell n\ f_{ijt} - \ell n\ f_{ist} - \ell n\ f_{rjt} - \ell n\ f_{rsk} + \ell n\ f_{rst}$,

$(1 \le i \le r-1\ ;\ 1 \le j \le s-1\ ;\ 1 \le k \le t-1)$ where $f_{ijk} = np_{ijk}$. This condition is satisfied if, and only if, $\ln f_{ijk}$ can be written in the form $\ln f_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk}$ ($1 \le i \le r$; $1 \le j \le s$; $1 \le k \le t$), where $u_{1.} = u_{2.} = u_{3.} = 0$; $u_{13i.} = u_{12i.} = 0$, each i ; $u_{23j.} = u_{12.j} = 0$, each j ; $u_{23.k} = u_{13.k} = 0$, each k ; (where $u_{1.} = \sum_{i=1}^{r} u_{1i}$) ; but otherwise u , u_{1i}, ..., u_{23jk} are completely arbitrary. In general, we can write $\ln f_{ijk}$ in the form:

$$\ln f_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk} + u_{123ijk}$$

where $u_{123ij.} = 0$, each (i, j) ; $u_{123i.k} = 0$, each (i, k) ; and $u_{123.jk} = 0$, each (j, k) . The u_{123}'s are then the second order interactions.

Y.M.M. Bishop adopts Birch's model. She utilizes his result that appropriate sums of the observed cell frequencies are sufficient statistics for maximum likelihood estimation of the cell frequencies under a specified model. Birch does not give a computing method but refers to iterative computing methods of Norton [1945] , Kastenbaum and Lamphier [1959], and Darroch [1962]. Bishop uses a different computing method, an iterative proportional method which she adapted from Deming and Stephan [1940]. This is illustrated in Chapter VI of this thesis.

The work of S. Fienberg and F. Mosteller is closely allied with that of Bishop and therefore we do not give a separate discussion of their work. Some of their papers are listed in the bibliography.

The above is by no means an exhaustive list of statisticians who have worked with contingency tables. It was our purpose just to present to the reader an idea of some of the alternatives. This being accomplished, we proceed to present our model.

CHAPTER II PRESENTATION OF OUR MODEL

Consider the following model for a three-dimensional $r \times s \times t$ contingency table:

$$E(f_{ijk}) = np_{ijk} = n[p_{i..}p_{.j.}\, p_{..k} + p_{i..}\alpha_{jk} + p_{.j.}\beta_{ik} + p_{..k}\gamma_{ij} + \delta_{ijk}] \geq 0$$

$i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$

with $\sum_{j=1}^{s} \alpha_{jk} = \sum_{k=1}^{t} \alpha_{jk} = 0$, $\sum_{i=1}^{r} \beta_{ik} = \sum_{k=1}^{t} \beta_{ik} = 0$, $\sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0$

and $\sum_{i=1}^{r} \delta_{ijk} = \sum_{j=1}^{s} \delta_{ijk} = \sum_{k=1}^{t} \delta_{ijk} = 0$. A dot indicates as before summation over the replaced subscript. The f_{ijk}'s represent the frequencies and the p_{ijk}'s represent the proportions.

$$\begin{cases} E(f_{.jk}) = np_{.jk} = n[p_{.j.}p_{..k} + \alpha_{jk}] \\ E(f_{i.k}) = np_{i.k} = n[p_{i..}p_{..k} + \beta_{ik}] \\ E(f_{ij.}) = np_{ij.} = n[p_{i..}p_{.j.} + \gamma_{ij}] \end{cases} \Rightarrow$$

$$\begin{cases} \alpha_{jk} = p_{.jk} - p_{.j.}p_{..k} \\ \beta_{ik} = p_{i.k} - p_{i..}p_{..k} \\ \gamma_{ij} = p_{ij.} - p_{i..}p_{.j.} \end{cases} \qquad (1)$$

$$\delta_{ijk} = p_{ijk} - p_{i..}p_{.j.}p_{..k} - p_{i..}\alpha_{jk} - p_{.j.}\beta_{ik} - p_{..k}\gamma_{ij}$$

$= p_{ijk} - p_{i..}p_{.jk} - p_{.j.}p_{i.k} - p_{..k}p_{ij.} + 2p_{i..}p_{.j.}p_{..k}$ after substituting the expressions for α_{jk}, β_{ik}, γ_{ij} .

We explain now why the model presented seems reasonable. Consider first the case of a 2×2×2 contingency table. Let the partitions of the sample space be $\{A, \bar{A}\}$, $\{B, \bar{B}\}$, and $\{C, \bar{C}\}$. Consider the two events A and B . Everyone agrees that $P(AB) - P(A)P(B)$ is a proper way of measuring deviation from pairwise independence. Suppose now that the three events A, B, C are pairwise independent so that

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(AC) &= P(A)P(C) \\ P(BC) &= P(B)P(C) \end{aligned} \tag{2}$$

We show now by an example from Feller that pairwise independence does not necessarily imply that $P(ABC) = P(A)P(B)P(C)$ i.e. that the two events AB and C are independent

($\Leftrightarrow P(ABC) = P(AB)P(C)$) or that BC and A are independent

($\Leftrightarrow P(ABC) = P(A)P(BC)$) or that AC and B are independent

($\Leftrightarrow P(ABC) = P(B)P(AB)$) .

<u>Example:</u> Consider the six permutations of the letters a, b, c as well as the three triples (a, a, a) , (b, b, b) and (c, c,c) . We take these nine triples as points of a sample space and attribute probability 1/9 to each. Denote by A_k the event that the kth place is occupied by the letter a . Obviously each of these three events has probability 1/3 while $P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = 1/9$. The three events are therefore pairwise independent, but they are not mutually independent because also $P(A_1A_2A_3) = 1/9$. (The occurence of A_1 and A_2 implies the occurence of A_3 and so A_3 is not independent of A_1A_2) .

We reserve therefore the term independence for the case where not only (2) holds, but in addition $P(ABC) = P(A)P(B)P(C)$. This equation ensures that A and BC are independent and also that the same is true of B and AC and of C and AB .

Consider a 2×2×2 table. $i = 1, 2; j=1, 2 ; k=1, 2$.

Letting $P_{1..} = P(A)$, $P_{.1.} = P(B)$, $P_{..1} = P(C)$,

$P_{2..} = P(\overline{A})$, $P_{.2.} = P(\overline{B})$, $P_{..2} = P(\overline{C})$,

$P_{11.} = P(AB)$, $P_{1.1} = P(AC)$, $P_{.11} = P(BC)$

$P_{12.} = P(A\overline{B})$, etc.,

we see that the conditions (1) are equivalent to

$$\begin{cases} \alpha_{11} = P(BC) - P(B)P(C) \\ \alpha_{12} = P(B\overline{C}) - P(B)P(\overline{C}) \\ \alpha_{21} = P(\overline{B}C) - P(\overline{B})P(C) \\ \alpha_{22} = P(\overline{B}\overline{C}) - P(\overline{B})P(\overline{C}) \end{cases}$$

as well as four similar equations for the β_{ik}'s and four similar equations for the γ_{ij}'s . Hence the condition that B and C are independent corresponds to $\alpha_{jk} = 0$; the condition that A and C are independent corresponds to $\beta_{ik} = 0$; and the condition that A and B are independent corresponds to $\gamma_{ij} = 0$.

We have seen, by the example considered, that pairwise independence does not necessarily imply mutual independence, hence the presence of the term δ_{ijk} . If δ_{ijk} were not present, then pairwise independence would automatically imply $p_{ijk} = p_{i..}p_{.j.}p_{..k}$ i.e. $P(ABC) = P(A)P(B)P(C)$ (or $P(AB\overline{C}) = P(A)P(B)P(\overline{C})$ etc.) . δ_{ijk}, we have seen, is equal to $p_{ijk} - p_{i..}p_{.j.}p_{..k} - p_{i..}\alpha_{jk} - p_{.j.}\beta_{ik} - p_{..k}\gamma_{ij}$, so that, if there is no first order interaction, δ_{ijk} becomes $p_{ijk} - p_{i..}p_{.j.}p_{..k}$.

In this thesis we are mainly concerned with $r\times s\times t$ contingency tables. Suppose therefore $\{A_1, A_2, \ldots, A_r\}$, $\{B_1, B_2, \ldots, B_s\}$, $\{C_1, C_2, \ldots, C_t\}$ are three partitions of the sample space. The A and B partitions are stochastically independent if $P(A_iB_j) = P(A_i)P(B_j)$, i=1 to r-1 , j=1 to s-1 . These (r-1)(s-1) equations are linearly independent. Similar statements hold for the other two pairs of partitions. There are further conditions for complete independence beyond those for pairwise independence. For three partitions the other conditions are $P(A_iB_jC_k) = P(A_i)P(B_j)P(C_k)$, i = 1 to r-1 ; j=1 to s-1 ; k=1 to t-1 . These (r-1)(s-1)(t-1) equations are linearly independent. A. N. Kolmogorov in his text "Foundations of Probability" and W. Feller in his text "An Introduction to Probability Theory and its Applications" deal with the theory of independence.

All this information is contained in our model $p_{ijk} = p_{i..}p_{.j.}p_{..k} + p_{i..}\alpha_{jk} + p_{.j.}\beta_{ik} + p_{..k}\gamma_{ij} + \delta_{ijk}$.

$i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$.

Let $p_{1..} = P(A_1)$, ..., $p_{r..} = P(A_r)$

$p_{.1.} = P(B_1)$, ..., $p_{.s.} = P(B_s)$

$$p_{..1} = P(C_1), \ldots, p_{..t} = P(C_t)$$

$$p_{11.} = P(A_1 B_1) \text{ etc.}$$

$$\left.\begin{aligned} \alpha_{jk} &= P(B_j C_k) - P(B_j)P(C_k) \\ \beta_{ik} &= P(A_i C_k) - P(A_i)P(C_k) \\ \gamma_{ij} &= P(A_i B_j) - P(A_i)P(B_j) \end{aligned}\right\} \qquad \text{are the}$$

measures of first order interaction or deviations from pairwise independence. In our model, if we have pairwise independence, that is, if the α's, β's, and γ's are all zero we get $p_{ijk} = p_{i..}p_{.j.}p_{..k} + \delta_{ijk}$. In this case $\delta_{ijk} = p_{ijk} - p_{i..}p_{.j.}p_{..k}$ or $\delta_{ijk} = P(A_i B_j C_k) - P(A_i)P(B_j)P(C_k)$. If $P(A_i B_j C_k) = P(A_i)P(B_j)P(C_k)$ for all i, j, k then $\delta_{ijk} = 0$ for all i, j, k and since there is no first and no second order interaction we have complete independence.

CHAPTER III TESTING THE HYPOTHESIS OF NO SECOND ORDER INTERACTION USING THE LIKELIHOOD RATIO TECHNIQUE

A. THE DENOMINATOR OF THE LIKELIHOOD RATIO

In our model second order interaction is present iff $\delta_{ijk} \neq 0$ for some (i, j, k) , i.e. iff $p_{ijk} \neq p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 2p_{i..}p_{.j.}p_{..k}$ for some (i, j, k) . Let $q_{ijk} = p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 2p_{i..}p_{.j.}p_{..k}$ $(p_{ijk} = q_{ijk} + \delta_{ijk}$ whether or not δ_{ijk} is zero) . Considering the given model we may want to test the following hypothesis: H_o: $\delta_{ijk} = 0$ for all (i, j, k) . i.e. no second order interaction is present. Under H_o the model yields:

$$p_{ijk} = p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 2p_{i..}p_{.j.}p_{..k} = q_{ijk}$$

$i = 1$ to r ; $j = 1$ to s ; $k = 1$ to t . We use the method of likelihood ratio.

The most general assumption is that the density or probability function

$$P = n! \prod_{i=1}^{r} \prod_{j=1}^{s} \prod_{k=1}^{t} \frac{p_{ijk}^{f_{ijk}}}{f_{ijk}!} \quad \text{with} \quad 0 < p_{ijk} \leq 1 , \quad \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} p_{ijk} = 1 ,$$

where the f_{ijk} are non-negative integers and $\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} = n$.

The given constraints are $\sum_{i=1}^{r} \delta_{ijk} = \sum_{j=1}^{s} \delta_{ijk} = \sum_{k=1}^{t} \delta_{ijk} = 0$ as well as

$$\sum_{j=1}^{s} \alpha_{jk} = \sum_{k=1}^{t} \alpha_{jk} = 0 , \quad \sum_{i=1}^{r} \beta_{ik} = \sum_{k=1}^{t} \beta_{ik} = 0 , \quad \sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0 .$$

These constraints are equivalent to $\sum_{i=1}^{r} p_{ijk} = p_{.jk}$, $\sum_{j=1}^{s} p_{ijk} = p_{i.k}$, $\sum_{k=1}^{t} p_{ijk} = p_{ij}$. To maximize P under the most general set of assumptions we use the constraint $\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} p_{ijk} = 1$ explicitly. The α_{jk} , β_{ik} , γ_{ij}, δ_{ijk} do not have to brought in explicitly in dealing with the general case. The constraints we placed on the α's , β's , γ's and δ's were chosen so that the contingency table would add up. We just keep in mind what $p_{i..}$, $p_{.j.}$, $p_{..k}$ and $p_{ij.}$, $p_{i.k}$, $p_{.jk}$ stand for in terms of the p_{ijk}'s

$$p_{i..} = \sum_{j=1}^{s} \sum_{k=1}^{t} p_{ijk} \qquad p_{.jk} = \sum_{i=1}^{r} p_{ijk}$$

$$p_{.j.} = \sum_{i=1}^{r} \sum_{k=1}^{t} p_{ijk} \qquad p_{i.k} = \sum_{j=1}^{s} p_{ijk}$$

$$p_{..k} = \sum_{i=1}^{r} \sum_{j=1}^{s} p_{ijk} \qquad p_{ij.} = \sum_{k=1}^{t} p_{ijk}$$

Let $\hat{p}_{ijk}$ denote the maximum likelihood estimator (MLE) of p_{ijk} in the general case. It is well known that $\hat{p}_{ijk} = \frac{f_{ijk}}{n}$. Hence the denominator of our likelihood ratio is $n! \prod_{i=1}^{r} \prod_{j=1}^{s} \prod_{k=1}^{t} (\frac{f_{ijk}}{n})^{f_{ijk}} \Big/ f_{ijk}!$

B. A THEOREM ABOUT LAGRANGE MULTIPLIERS

The usual Lagrange multiplier procedure is to have linearly independent constraints on the variables, and the number of such constraints must be fewer than the number of variables. For reasons of symmetry it is useful to have a form of the procedure in which the constraints no longer need be linearly independent and may even exceed in number the number of variables, though the number of linearly independent constraints must still be fewer than the number of variables. The following is a theorem about the extended procedure. The notation of this theorem and its proof is independent of that in the rest of the thesis. The proof of this theorem closely parallels Apostol's proof of the original Lagrange multiplier theorem. (See Tom M. Apostol, "Mathematical Analysis, A Modern Approach to Advanced Calculus". Reading, Mass.,1957. pp. 152-156).

Theorem: Let f be a real valued function having continuous first-order partial derivatives on an open set S in n-dimensional Euclidean space E_n. Let $g_1, \ldots, g_p$ be p real-valued functions also having continuous first-order partial derivatives on S, and assume that the number of functionally independent members of $\{g_1, \ldots, g_p\}$ on S is r, with $r < n$. Let X_o be that subset of S on which $g_1, \ldots, g_p$ all vanish, that is, in set, vector notation, $X_o = \{\underline{x} | \underline{x} \varepsilon S, \underline{g}(\underline{x}) = \underline{0}\}$. Assume that $x_o \varepsilon X_o$ and assume that there exists a neighborhood $\eta(\underline{x}_o)$ such that $f(\underline{x}) \leq f(x_o)$ for all $\underline{x}$ in $X_o \cap \eta(\underline{x}_o)$ or that $f(\underline{x}) \geq f(\underline{x}_o)$ for all $\underline{x}$ in $X_o \cap \eta(\underline{x}_o)$. Assume also that the matrix $[D_i g_j(\underline{x}_o)] = [\frac{\partial}{\partial x_i} g_j(\underline{x}) \Big|_{\underline{x}=\underline{x}_o}]$ has rank r, with $r < n$. Then there exist p real numbers $\lambda_1, \ldots, \lambda_p$ such that the following n equations are satisfied:

$D_j f(\underline{x}_o) + \sum_{i=1}^{p} \lambda_i D_j g_i(\underline{x}_o) = 0 \quad (j = 1, 2, \ldots, n)$.[1] If $p > r$, then then $(p-r)$ of the multipliers $\lambda_1, \ldots, \lambda_p$ can be assigned arbitrary values, provided only that the r constraints associated with the other r multipliers form a set of r functionally independent constraints.

We now give an adaptation of the argument in Apostol:

Proof: Assume, without loss of generality, that the $(r\times r)$ minor in the upper left-hand corner of the $(n\times p)$ matrix $[D_i g_j(\underline{x}_o)]$ has rank r. Consider the following system of r linear equations in the p unknowns

$$\lambda_1, \ldots, \lambda_p: \quad \sum_{k=1}^{p} \lambda_k D_h g_k(\underline{x}_o) = -D_h f(\underline{x}_o) \quad (h = 1, 2, \ldots, r) .$$

Once $\lambda_{r+1}, \ldots, \lambda_p$ have been assigned arbitrary values, there is a unique solution for $\lambda_1, \ldots, \lambda_r$. It remains to show that the other $(n-r)$ linear equations are also satisfied for this set of values for $\lambda_1, \ldots \lambda_p$.

To do this we apply the implicit function theorem.(See Apostol (1957), pp. 147-148). Since $r < n$, every point $\underline{x}$ in S can be written in the form $\underline{x} = (\underline{u} ; \underline{v})$, say, where $\underline{u} \in E_r$ and $\underline{v} \in E_{n-r}$. In the remainder of this proof we will write $\underline{u}$ for $(x_1, \ldots, x_r)$ and $\underline{v}$ for $(x_{r+1}, \ldots, x_n)$, so that $v_\ell = x_{r+\ell}$. In terms of the vector valued function $\underline{g} = (g_1, \ldots, g_p)$, we can now write $\underline{g}(\underline{u} ; \underline{v}) = \underline{0}$ whenever $\underline{x} = (\underline{u} ; \underline{v})$ belongs to X_o. Since $\underline{g}$ has continuous first-order partial derivatives on S, and since the $(r\times r)$ minor $[D_i g_j(\underline{x}_o)]$ mentioned above has rank r, all the conditions of the implicit function theorem are satisfied. Therefore, there exists an $(n-r)$- dimensional neighborhood V_o of $\underline{v}_o$ and a unique vector-valued function $\underline{H} = (H_1, \ldots, H_r)$, defined on V_o and having values in E_r, such that $\underline{H}$ has continuous first-order partial derivatives on V_o, $\underline{H}(v_o) = \underline{u}_o$, and for every $\underline{v}$ in V_o , we have $\underline{g}(\underline{H}(\underline{v})\ ;\ \underline{v}) = \underline{0}$. This amounts to saying that the system of r equations $g_1(x_1, \ldots, x_n) = 0$, $\ldots$, $g_r(x_1, \ldots, x_n) = 0$ can be "solved" for $x_1, \ldots, x_r$ in terms of $x_{r+1}, \ldots, x_n$, giving the solutions in the form $x_\ell = H_\ell(x_{r+1}, \ldots, x_n)$, $\ell = 1, 2, \ldots, r$. We shall now "substitute" these expressions for $x_1, \ldots, x_r$ into the expression $f(x_1, \ldots, x_n)$ and also into each expression $g_k(x_1, \ldots, x_n)$. That is to say, we define a new function F as follows:

$$F(x_{r+1}, \ldots, x_n) = f[H_1(x_{r+1}, \ldots, x_n), \ldots, H_r(x_{r+1}, \ldots, x_n], x_{r+1}, \ldots, x_n]$$

and we define r new functions $G_1, \ldots, G_r$ as follows:

$$G_k(x_{r+1}, \ldots, x_n) = g_k[H_1(x_{r+1}, \ldots, x_n), \ldots, H_r(x_{r+1}, \ldots, x_n), x_{r+1}, \ldots x_n]$$

$k = 1, 2, \ldots, r$. More briefly, we can write $F(\underline{v}) = f(\underline{R}(\underline{v}))$ and $G_k(\underline{v}) = g_k(\underline{R}(\underline{v}))$, where $\underline{R}(\underline{v}) = (\underline{H}(\underline{v}), \underline{v})$. Here $\underline{v}$ is restricted to lie in the set V_o .

Each function G_k so defined is identically zero on the set V_o by the implicit function theorem. Therefore, each derivative $D_h G_k$ is also identically zero on V_o , and, in particular, $D_h G_k(\underline{v}_o) = 0$. But by the chain rule (Apostol (1957) , pp. 112-114) $D_\ell G_k(\underline{v}_o) =$ $\sum_{h=1}^{n} D_h g_k(\underline{x}_o) D_\ell R_h(\underline{v}_o) = 0$ $(\ell = 1, 2, \ldots, (n-r))$. But $R_k(\underline{v}) =$ $H_k(\underline{v})$ if $1 \leqq k \leqq r$ and $R_k(\underline{v}) = x_k$ if $(r+1) \leqq k \leqq n$. Therefore, when

$(r+1) \leqq k \leqq n$, we have $D_\ell R_k(\underline{v}) = \begin{cases} 1 & \text{if } r+\ell = k \\ 0 & \text{if } r+\ell \neq k \end{cases}$.

Hence the above set of equations becomes

$$\sum_{h=1}^{r} D_h g_k(\underline{x}_o) D_\ell H_h(\underline{v}_o) + D_{r+\ell} g_k(\underline{x}_o) = 0 \qquad \begin{array}{l} k = 1, 2, \ldots, r \\ \ell = 1, 2, \ldots, (n-r) . \end{array} \tag{2}$$

By continuity of $\underline{H}$, there will be a neighborhood $\eta(\underline{v}_o) \subset V_o$ such that $\underline{v} \in \eta(\underline{v}_o)$ implies $(\underline{H}(\underline{v}) ; \underline{v}) \in \eta(\underline{x}_o)$, where $\eta(\underline{x}_o)$ is the neighborhood in Apostol's statement of the theorem. Hence $\underline{v} \in \eta(\underline{v}_o)$ implies $(\underline{H}(\underline{v}) ; \underline{v}) \in \{X_o \cap \eta(\underline{x}_o)\}$ and therefore, by hypothesis, we have either $F(\underline{v}) \leqq F(\underline{v}_o)$ for all $\underline{v}$ in $\eta(\underline{v}_o)$ or else we have $F(\underline{v}) \geqq F(\underline{v}_o)$ for all $\underline{v}$ in $\eta(\underline{v}_o)$. That is, F has a local maximum or a local minimum at the interior point $\underline{v}_o$. Each partial derivative $D_\ell F(\underline{v}_o)$ must therefore be zero. If we use the chain rule to compute these derivatives,

we find $D_\ell F(\underline{v}_o) = \sum_{h=1}^{n} D_h f(\underline{x}_o) D_\ell R_h(\underline{v}_o) = 0$

$(\ell = 1, 2, \ldots, (n-r))$, and hence we can write

$$\sum_{h=1}^{r} D_h f(\underline{x}_o) D_\ell H_h(\underline{v}_o) + D_{r+\ell} f(\underline{x}_o) = 0 \quad (\ell = 1, 2, \ldots, (n-r)) \tag{3}$$

If we now multiply (2) by λ_k , sum on k , and add the result to (3) , we find

$$\left[\sum_{h=1}^{r} D_h f(\underline{x}_o) D_\ell H_h(\underline{v}_o) + D_{r+\ell} f(\underline{x}_o)\right] + \sum_{k=1}^{r} \lambda_k \left[\sum_{h=1}^{r} D_h g_k(\underline{x}_o) D_\ell H_h(\underline{x}_o) + D_{r+\ell} g_k(\underline{x}_o)\right]$$

$$= \sum_{h=1}^{r} [D_h f(\underline{x}_o) + \sum_{k=1}^{r} \lambda_k D_h g_k(\underline{x}_o)] D_\ell H_h(\underline{v}_o) + D_{r+\ell} f(\underline{x}_o) + \sum_{k=1}^{r} \lambda_k D_{r+\ell} g_k(\underline{x}_o) = 0$$

for $\ell = 1, 2, \ldots, (n-r)$. In the sum over h , the expression in the square brackets becomes $[-\sum_{k=r+1}^{p} \lambda_k D_h g_k(\underline{x}_o)]$ because of the way $\lambda_1, \ldots, \lambda_r$ were defined. Thus we are left with $[D_{r+\ell} f(\underline{x}_o) + \sum_{k=1}^{r} \lambda_k D_{r+\ell} g_k(\underline{x}_o)] + [-\sum_{k=r+1}^{p} \lambda_k \sum_{h=1}^{r} D_h g_k(\underline{x}_o) D_\ell H_h(\underline{v}_o)] = 0$ for $\ell =$ (4) $1, 2, \ldots, (n-r)$. But, referring back to equation (2), the second square bracket of the last equation (4) becomes $\sum_{k=r+1}^{p} \lambda_k D_{r+\ell} g_k(\underline{x}_o)$ $(\ell = 1, 2, \ldots, (n-r))$. Substituting this into (4) one gets $D_{r+\ell} f(\underline{x}_o) + \sum_{k=1}^{p} \lambda_k g_k(\underline{x}_o) = 0$ $(\ell = 1, 2, \ldots, (n-r))$, which is equivalent to equation (1), and these are exactly the equations required to complete the proof.

C. PREVIEW OF THE ARGUMENT TO COME

We will be considering the maximum likelihood estimation of the q_{ijk} , the hypothetical p_{ijk} with δ_{ijk} set equal to zero. Consider the equations $\frac{\partial q}{\partial p_{abc}} = 0$ where $q = \log(n!) - \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \log(f_{ijk}!) + \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} \log p_{ijk} - \theta [\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} p_{ijk} - 1] + \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{ijk} [p_{ijk} - q_{ijk}]$,

a = 1 to r ; b = 1 to s ;
c = 1 to t .

$$\frac{\partial q}{\partial p_{abc}} = \frac{f_{abc}}{p_{abc}} - \{\theta + \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{ijk} \frac{\partial g_{ijk}}{\partial p_{abc}}\}$$

a = 1 to r ; b = 1 to s ;

c = 1 to t ,

where $g_{ijk} = p_{ijk} - q_{ijk}$.

These give us rst non-homogeneous linear equations in the unknowns θ and the η_{ijk} . The matrix of coefficients of the unknowns is

$[\frac{\partial}{\partial p_{abc}} g_h(\underline{p})]$, h = 0, 1, ..., rst; a = 1 to r ; b = 1 to s ; c = 1 to t .

The first step in applying the Lagrange multiplier theorem is to find the rank of this matrix.

It is shown that $[\begin{smallmatrix} 1 & \underline{0}' \\ \underline{0} & X \otimes Y \otimes Z \end{smallmatrix}][\frac{\partial g_{h(p)}}{\partial p_{ijk}}]$,and hence $[\frac{\partial}{\partial p_{ijk}} g_h(\underline{p})]$ has rank $(r-1)(s-1)(t-1) + 1$, where $[\begin{smallmatrix} 1 & \underline{0}' \\ \underline{0} & X \otimes Y \otimes Z \end{smallmatrix}]$ is an $((rst + 1) \times (rst + 1))$ nonsingular matrix to be described later. Upon eliminating θ from the equations $\frac{\partial q}{\partial p_{abc}} = 0$, the resulting matrix of coefficients of the η_{ijk} has rank $(r-1)(s-1)(t-1)$.

Ordinarily the next step would be to solve for the Lagrange multipliers η_{ijk} and , having found them, to substitute these into the equations $\frac{\partial q}{\partial p_{abc}} = 0$ and to solve these equations for $\hat{q}_{ijk}$, the restricted maximum likelihood estimators of the p_{ijk} , given that the

$\delta_{ijk} = 0$. In the present case this is not necessary. It will be shown that the equations are consistent and have a solution (actually systems of solutions) for the p_{ijk} when one sets p_{ijk} equal to

$$\hat{q}_{ijk} = \frac{f_{i..}f_{.jk} + f_{.j.}f_{i.k} + f_{..k}f_{ij.}}{n^2} - \frac{f_{i..}f_{.j.}f_{..k}}{n^3} .$$

Thus the $\hat{q}_{ijk}$ are the restricted maximum likelihood estimators of the p_{ijk} when $\delta_{ijk} = 0$. Generally the equations are inconsistent for other sets of values of the p_{ijk} .

D. SATISFYING THE CONDITIONS OF THE THEOREM

Define $f(p_{111}, p_{112}, \ldots, p_{11t}, \ldots, p_{rst}) = \log n! - \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \log (f_{ijk}!)$

$$+ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{ijk} \log p_{ijk} , \; p_{ijk} > 0.$$

Let $S = \{(x_1, \ldots, x_{rst}) \mid \; x_\ell > 0 \;\text{ for }\; \ell = 1 \;\text{ to }\; rst\}. \dfrac{\partial f}{\partial p_{ijk}} = \dfrac{f_{ijk}}{p_{ijk}}$,

which is continuous on S , $i = 1, \ldots, r$
$j = 1, \ldots, s$
$k = 1, \ldots, t$.

Consider the $rst + 1$ constraints $g_h(p_{111}, \ldots, p_{rst}) = 0$;

$h = 0, 1, \ldots, rst$:

$$g_o(\underline{p}) = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} p_{ijk} - 1$$

$$g_{ijk}(\underline{p}) = p_{ijk} - q_{ijk} ,$$

$i = 1$ to r ; $j = 1$ to s ; $k = 1$ to t , where $\underline{p} = [p_{111}, \ldots, p_{rst}]$.

$\frac{\partial g_o}{\partial p_{ijk}} = 1$ and, since constant functions are continuous, we have that g_o has continuous first order partial derivatives on S. Essentially, the $g_{ijk}(\underline{p})$ are polynomials in the p_{abc}, hence all of the partial derivatives are continuous.

Now $\Omega = \{(x_1,\ldots,x_{rst}) \mid x_1+\ldots+x_{rst} = 1 \, , \, x_\ell > 0 \text{ for } \ell = 1 \text{ to } rst\}$ is the subset of S where $g_o(\underline{x}) = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} x_{ijk} - 1 = 0$. Ω is the X_o to use in the earlier case for the general multinomial case. When second order interactions are zero, then $X_o = \{(x_1,\ldots,x_{rst}) \mid g_h(\underline{x}) = 0 \, ,$

$x_\ell > 0$ for $\ell = 1$ to rst
$h = 0,1,\ldots,rst\}$

This $X_o \subset \Omega \subset S$ is what we need here.

Let us now find how many $g_h(\underline{p})$ are functionally independent on the set S of the Lagrange multiplier theorem. The number of $g_h(\underline{p})$ that are functionally independent on the set S equals the rank of the matrix $[\frac{\partial}{\partial p_{abc}} g_h(\underline{p})]$, where $\underline{p} \in S$. If we multiply the above matrix $[\frac{\partial \underline{g}}{\partial \underline{p}}]$ by a non-singular matrix, the rank remains unchanged. Let us use $\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & X \otimes Y \otimes Z \end{bmatrix}$, which we describe below. Note that $\otimes$ stands for the direct product of the matrices.

Let $\{X^{(u)}\}$, $u = 0,\ldots,(r-1)$, denote a set of functions orthonormal with respect to $\{p_{1..},\ldots,p_{r..}\}$; $\{Y^{(v)}\}$, $v = 0,\ldots,(s-1)$, a set orthonormal with respect to $\{p_{.1.},\ldots,p_{.s.}\}$; $\{Z^{(w)}\}$, $w = 0,\ldots,(t-1)$, a set orthonormal with respect to $\{p_{..1},\ldots,p_{..t}\}$; with $X^{(0)} \equiv 1$, $Y^{(0)} \equiv 1$, $Z^{(0)} \equiv 1$. Let $X = [X_i^{(u)}]$, $i = 1$ to r ; $Y = [Y_j^{(w)}]$, $j = 1$ to s ; $Z = [Z_k^{(w)}]$, $k = 1$ to t . The use of $X \otimes Y \otimes Z$, where the superscript indexes the rows, and the subscript indexes the columns, and $X_i^{(u)} Y_j^{(v)} Z_k^{(w)}$ is the element of $X \otimes Y \otimes Z$ in the (u,v,w)th row and (i,j,k)th column should reduce $[\frac{\partial}{\partial p_{ijk}} g_h(\underline{p})]$ to much simpler form.

$$\frac{\partial g_o}{\partial p_{abc}} = 1$$

$$\frac{\partial g_{abc}}{\partial p_{abc}} = 1 - [p_{.bc} + p_{a.c} + p_{ab.} + p_{a..} + p_{.b.} + p_{..c} - 2(p_{.b.}p_{..c} + p_{a..}p_{..c} + p_{a..}p_{.b.}]$$

$$(k \neq c) \quad \frac{\partial g_{abk}}{\partial p_{abc}} = -[p_{.bk} + p_{a.k} + p_{..k} - 2(p_{.b.}p_{..k} + p_{a..}p_{..k})]$$

$$(j \neq b) \quad \frac{\partial g_{ajc}}{\partial p_{abc}} = -[p_{.jc} + p_{aj.} + p_{.j.} - 2(p_{.j.}p_{..c} + p_{a..}p_{.j.})]$$

$$(i \neq a) \quad \frac{\partial g_{ibc}}{\partial p_{abc}} = -[p_{i.c} + p_{ib.} + p_{i..} - 2(p_{i..}p_{..c} + p_{i..}p_{.b.})]$$

$$\begin{pmatrix} j \neq b \\ k \neq c \end{pmatrix} \quad \frac{\partial g_{ajk}}{\partial p_{abc}} = -[p_{.jk} - 2(p_{.j.}p_{..k})]$$

$$\begin{pmatrix} i \neq a \\ k \neq c \end{pmatrix} \quad \frac{\partial g_{ibk}}{\partial p_{abc}} = -[p_{i.k} - 2(p_{i..}p_{..k})]$$

$$\begin{pmatrix} i \neq a \\ j \neq b \end{pmatrix} \quad \frac{\partial g_{ijc}}{\partial p_{abc}} = -[p_{ij.} - 2(p_{i..}p_{.j.})]$$

$$\begin{pmatrix} i\neq a \\ j\neq b \\ k\neq c \end{pmatrix} \frac{\partial g_{ijk}}{\partial p_{abc}} = 0 \; .$$

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} Y_b^{(v)} Z_c^{(w)} -$$

$$\Big\{ X_a^{(u)} \Big[\sum_{j=1}^{s} \sum_{k=1}^{t} Y_j^{(v)} Z_k^{(w)} p_{.jk} \Big] + Y_b^{(v)} \Big[\sum_{i=1}^{r} \sum_{k=1}^{t} X_i^{(u)} Z_k^{(w)} p_{i.k} \Big] +$$

$$Z_c^{(w)} \Big[\sum_{i=1}^{r} \sum_{j=1}^{s} X_i^{(u)} Y_j^{(v)} p_{ij.} \Big] + Y_b^{(v)} Z_c^{(w)} \Big[\sum_{i=1}^{r} X_i^{(u)} p_{i..} \Big]$$

$$+ X_a^{(u)} Z_c^{(w)} \Big[\sum_{j=1}^{s} Y_j^{(v)} p_{.j.} \Big] + X_a^{(u)} Y_b^{(v)} \Big[\sum_{k=1}^{t} Z_k^{(w)} p_{..k} \Big]$$

$$+ 2 \Big[X_a^{(u)} \Big[\sum_{j=1}^{s} Y_j^{(v)} p_{.j.} \Big] \Big[\sum_{k=1}^{t} Z_k^{(w)} p_{..k} \Big] + Y_b^{(v)} \Big[\sum_{i=1}^{r} X_i^{(u)} p_{i..} \Big] \Big[\sum_{k=1}^{t} Z_k^{(w)} p_{..k} \Big]$$

$$+ Z_c^{(w)} \Big[\sum_{i=1}^{r} X_i^{(u)} p_{i..} \Big] \Big[\sum_{j=1}^{s} Y_j^{(v)} p_{.j.} \Big] \Big] \Big\} \; .$$

We now make use of the orthogonality relations and also introduce Lancaster's correlation notation:

$\rho_{uvw} = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} p_{ijk}$. Since $X^{(0)} \equiv 1$, $Y^{(0)} \equiv 1$, $Z^{(0)} \equiv 1$, this gives $\rho_{ovw} = \sum_{j=1}^{s} \sum_{k=1}^{t} Y_j^{(v)} Z_k^{(w)} p_{.jk}$, $\rho_{uvo} = \sum_{i=1}^{r} \sum_{j=1}^{s} X_i^{(u)} Y_j^{(v)} p_{ij.}$ and $\rho_{uow} = \sum_{i=1}^{r} \sum_{k=1}^{t} X_i^{(u)} Z_k^{(w)} p_{i.k}$.

$$\rho_{oow} = \sum_{k=1}^{t} Z_k^{(w)} p_{..k} = \sum_{k=1}^{t} Z_k^{(w)} Z_k^{(0)} p_{..k} = \begin{cases} 1 & \text{if } w = 0 \\ 0 & \text{if } w \neq 0 \; . \end{cases}$$

Similarly $\rho_{uoo} = \begin{cases} 1 & \text{if } u = 0 \\ 0 & \text{if } u \neq 0 \end{cases}$ and $\rho_{ovo} = \begin{cases} 1 & \text{if } v = 0 \\ 0 & \text{if } v \neq 0 \end{cases}$.

Note that there is no second order interaction in our model i.e. $\delta_{ijk} = 0 \;\forall$ (ijk) if and only if $\sum_{u=1}^{r-1} \sum_{v=1}^{s-1} \sum_{w=1}^{t-1} \rho_{uvw} = 0$.

(Refer to Lancaster's "The Chi-Squared Distribution" for proof).

Let $\Delta_{\ell m} = \begin{cases} 1 & \text{if } \ell = m \\ 0 & \text{if } \ell \neq m \end{cases}$

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} Y_b^{(v)} Z_c^{(w)} - \{X_a^{(u)} \rho_{ovw} + Y_b^{(v)} \rho_{uow} +$$

$$+ Z_c^{(w)} \rho_{uvo} + Y_b^{(v)} Z_c^{(w)} \Delta_{uo} + X_a^{(u)} Z_c^{(w)} \Delta_{vo} + X_a^{(u)} Y_b^{(v)} \Delta_{wo}\} + 2\{X_a^{(u)} \Delta_{vo} \Delta_{wo} +$$

$$Y_b^{(v)} \Delta_{uo} \Delta_{wo} + Z_c^{(w)} \Delta_{uo} \Delta_{vo}\} .$$

Let us now consider the following cases:

$(u = 0 , v = 0 , w = 0)$. Recall $\rho_{ooo} = 1$.

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \frac{\partial g_{ijk}}{\partial p_{abc}} = 1 - 6 + 2(3) = 1 \qquad \text{(1 equation)}$$

$(u = 0, v = 0, w \neq 0)$, $\rho_{oow} = 0$, $\rho_{ooo} = 1$.

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = Z_c^{(w)} - 3Z_c^{(w)} + 2Z_c^{(w)} = 0 \qquad \text{((t-1)eq'ns)}$$

$(u = 0, v \neq 0, w = 0)$, $\rho_{ovo} = 0$, $\rho_{ooo} = 1$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} Y_j^{(v)} \frac{\partial g_{ijk}}{\partial p_{abc}} = Y_b^{(v)} - 3Y_b^{(v)} + 2Y_b^{(v)} = 0 \quad ((s-1)\text{eq'ns})$$

$(u \neq 0,\ v = 0,\ w = 0)$, $\rho_{uoo} = 0$, $\rho_{ooo} = 1$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} X_i^{(u)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} - 3X_a^{(u)} + 2X_a^{(u)} = 0 \quad ((r-1)\text{eq'ns})$$

$(u = 0,\ v \neq 0,\ w \neq 0)$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = Y_b^{(v)} Z_c^{(w)} - \{\rho_{ovw} + Y_b^{(v)} Z_c^{(w)}\}$$

$$= -\rho_{ovw} \quad ((s-1)(t-1)\text{eq'ns})$$

$(u \neq 0,\ v = 0,\ w \neq 0)$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} X_i^{(u)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} Z_c^{(w)} - \{\rho_{uow} + X_a^{(u)} Z_c^{(w)}\}$$

$$= -\rho_{uov} \quad ((r-1)(t-1)\text{eq'ns})$$

$(u \neq 0,\ v \neq 0,\ w = 0)$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} Y_b^{(v)} - \{\rho_{uvo} + X_a^{(u)} Y_b^{(v)}\}$$

$$= -\rho_{uvo} \quad ((r-1)(s-1)\text{eq'ns})$$

$(u \neq 0,\ v \neq 0,\ w \neq 0)$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = X_a^{(u)} Y_b^{(v)} Z_c^{(w)} - \{X_a^{(u)} \rho_{ouw} + Y_b^{(v)} \rho_{uow} + Z_c^{(w)} \rho_{uvo}\}$$

$$((r-1)(s-1)(t-1)\text{eq'ns})$$

Suppose we let h index the rows of $[\frac{\partial}{\partial p_{abc}} g_h(\underline{p})]$ and (a, b, c) the columns. We can put $h = 0$ first and then $h = (i, j, k)$ afterwards in some convenient order, say lexicographic order. Similarly we order the (a, b, c). The matrix is $(rst + 1) \times rst$.

Let $X = [X_i^{(u)}]$, $(r \times r)$, with u indexing rows and i columns. Similarly for $Y = [Y_j^{(v)}]$, $(s \times s)$ and $Z = [Z_k^{(w)}]$, $(t \times t)$. Then $X \otimes Y \otimes Z$, $(rst \times rst)$, has rows indexed by (u, v, w) lexicographically and columns by (i, j, k) lexicographically.

Consider $\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & X \otimes Y \otimes Z \end{bmatrix} [\frac{\partial}{\partial p_{ijk}} g_h(\underline{p})]$ $(rst + 1) \times rst$. The first row of the product matrix consists of 1's. The row corresponding to $u=0$, $v=0$, $w=0$, consists of 1's. The $(r-1)$ rows corresponding to $u \neq 0$, $v=0$, $w=0$ consist of 0's. The $(s-1)$ rows corresponding to $u = 0$, $v \neq 0$, $w = 0$ consist of 0's. The $(t-1)$ rows corresponding to $u = 0$, $v = 0$, $w \neq 0$ consist of 0's. The $(s-1)(t-1)$ rows corresponding to $u = 0$, $v \neq 0$, $w \neq 0$ consist of $-\rho_{ovw}$'s.

The $(r-1)(t-1)$ rows corresponding to $u \neq 0$, $v = 0$, $w \neq 0$ consist of $-\rho_{uow}$'s.

The $(r-1)(s-1)$ rows corresponding to $u \neq 0$, $v \neq 0$, $w = 0$ consist of $-\rho_{uvo}$'s.

Obviously the rank of the submatrix consisting of the

$$1 + 1 + (r-1) + (s-1) + (t-1) + (s-1)(t-1) + (r-1)(t-1) + (r-1)(s-1)$$

rows listed above is one. In general the submatrix consisting of the remaining (r-1)(s-1)(t-1) rows will have rank (r-1)(s-1)(t-1) because of the linear independence of the orthonormal functions. We now prove this statement.

The matrix $V((r-1)(s-1)(t-1) \times rst)$ with elements $X_a^{(u)}Y_b^{(v)}Z_c^{(w)}$, $u \neq 0$, $v \neq 0$, $w \neq 0$, (u,v,w) indexing rows, (a,b,c) indexing columns, has rank (r-1)(s-1)(t-1), since its rows are among those of the non-singular matrix $X \otimes Y \otimes Z$. The last (r-1)(s-1)(t-1) rows of

$$\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & X \otimes Y \otimes Z \end{bmatrix} \begin{bmatrix} \dfrac{\partial g_o}{\partial \underline{p}} \\ \dfrac{\partial \underline{g}}{\partial \underline{p}} \end{bmatrix}$$ have elements

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}}$$

$$= X_a^{(u)}Y_b^{(v)}Z_c^{(w)} - \{X_a^{(u)}\rho_{ovw} + Y_b^{(v)}\rho_{uow} + Z_c^{(w)}\rho_{uvo}\} .$$

(u,v,w) indexes rows $(u = 1$ to $(r-1)$, $v = 1$ to $(s-1)$, $w = 1$ to $(t-1))$. (a,b,c) indexes columns $(a = 1$ to r, $b = 1$ to s, $c = 1$ to $t)$. Let us denote this $((r-1)(s-1)(t-1) \times rst)$ matrix by Ω. Let π be an $(rst \times rst)$ diagonal matrix with diagonal elements $p_{a..}p_{.b.}p_{..c} > 0$. The rank of $\Omega\pi V'$ is $\leq$ rank of Ω. If we can show that $\Omega\pi V'$ has full rank (r-1)(s-1)(t-1), then Ω must have rank $\geq$ (r-1)(s-1)(t-1). But it cannot have rank > (r-1)(s-1)(t-1). The element in the (u,v,w)th row and (u', v', w')th column of $\Omega\pi V'$ is :

$$\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} p_{a..}p_{.b.}p_{..c}[X_a^{(u)}Y_b^{(v)}Z_c^{(w)} - \{X_a^{(u)}\rho_{ovw} + Y_b^{(v)}\rho_{uow} + Z_c^{(w)}\rho_{uvo}\}]$$

$$\times [X_a^{(u')}Y_b^{(v')}Z_c^{(w')}] .$$

$$= \sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} p_{a..}p_{.b.}p_{..c}\, X_a^{(u)}Y_b^{(v)}Z_c^{(w)}X_a^{(u')}Y_b^{(v')}Z_c^{(w')}$$

$$- \sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} p_{a..}p_{.b.}p_{..c}\{X_a^{(u)}\rho_{ovw} + Y_b^{(v)}\rho_{uow} + Z_c^{(w)}\rho_{uvo}\}X_a^{(u')}Y_b^{(v')}Z_c^{(w')}$$

The first term on the right equals

$$(\sum_{a=1}^{r} p_{a..}X_a^{(u)}X_a^{(u')})(\sum_{b=1}^{s} p_{.b.}Y_b^{(v)}Y_b^{(v')})(\sum_{c=1}^{t} p_{..c}Z_c^{(w)}Z_c^{(w')}) = \begin{cases} 1 \text{ if } (u,v,w)=(u',v',w') \\ 0 \text{ if } (u,v,w)\neq(u',v',w') \end{cases}$$

In the second term

$$\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} p_{a..}p_{.b.}p_{..c}X_a^{(u)}\rho_{ovw}X_a^{(u')}Y_b^{(v')}Z_c^{(w')}$$

$$= \rho_{ovw}(\sum_{a=1}^{r} p_{a..}X_a^{(u)}X_a^{(u')})(\sum_{b=1}^{s} p_{.b.}Y_b^{(v')})(\sum_{c=1}^{t} p_{..c}Z_c^{(w')}) = 0$$

Since $\sum_{b=1}^{s} p_{.b.}Y_b^{(v')} = 0$ for $v' \neq 0$ and $\sum_{c=1}^{t} p_{..c}Z_c^{(w')} = 0$ for $w' = 0$, similar calculations for the other two parts of the second term show that each part equals 0, hence the whole second term equals 0. The diagonal elements of $\Omega\pi v'$ are 1's and the off diagonal elements are 0's. Hence $\Omega\pi v'$ has full rank and hence Ω has rank $(r-1)(s-1)(t-1)$. This means that $(rs + rt + st - r - s - t + 1)$ of the Lagrange multipliers can be assigned convenient arbitrary values, and then we must solve for the other $1 + (r-1)(s-1)(t-1)$.

E. SOLVING FOR ONE OF THE LAGRANGE MULTIPLIERS

Consider $q = \log(n!) - \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} \log(f_{ijk}!) + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} f_{ijk} \log p_{ijk}$

$$-\{\theta[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} p_{ijk} - 1] + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ijk}[p_{ijk} - q_{ijk}].$$

$$\frac{\partial q}{\partial p_{abc}} = \frac{f_{abc}}{p_{abc}} - \{\theta + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ijk} \frac{\partial g_{ijk}}{\partial p_{abc}}\} \quad a = 1 \text{ to } r\ ;\ b = 1 \text{ to } s\ ;\ c = 1 \text{ to } t\ .$$

What we have actually found is the rank of the coefficient matrix of the homogeneous part of these rst equations.

$$\frac{\partial q}{\partial p_{abc}} = \frac{f_{abc}}{p_{abc}} - \{\theta + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ijk} \frac{\partial g_{ijk}}{\partial p_{abc}}\}$$

$$= \frac{f_{abc}}{p_{abc}} - \theta - \eta_{abc} + \sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ajk}\, p_{.jk} + \sum_{i=1}^{r}\sum_{k=1}^{t} \eta_{ibk}\, p_{i.k}$$

$$+ \sum_{i=1}^{r}\sum_{j=1}^{s} \eta_{ijc}\, p_{ij.} + \sum_{i=1}^{r} \eta_{ibc}\, p_{i..} + \sum_{j=1}^{s} \eta_{ajc} p_{.j.} + \sum_{k=1}^{t} \eta_{abk}\, p_{..k}$$

$$- 2\{\sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ajk}\, p_{.j.}\, p_{..k} + \sum_{i=1}^{r}\sum_{k=1}^{t} \eta_{ibk}\, p_{i..}\, p_{..k} + \sum_{i=1}^{r}\sum_{j=1}^{s} \eta_{ijc}\, p_{i..}\, p_{.j.}\}$$

Since $p_{abc} = (p_{a..}p_{.bc} + p_{.b.}p_{a.c} + p_{..c}p_{ab.} - 2p_{a..}p_{.b.}p_{..c})$ by hypothesis,

$$\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} p_{abc} \frac{\partial q}{\partial p_{abc}} = n-\theta - \sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} \eta_{abc}\, q_{abc} + \sum_{a=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} \eta_{ajk}\, p_{a..}\, p_{.jk}$$

$$+ \sum_{i=1}^{r}\sum_{b=1}^{s}\sum_{k=1}^{t} \eta_{ibk}\, p_{.b.} p_{i.k} + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{c=1}^{t} \eta_{ijc}\, p_{..c}\, p_{ij.} + \sum_{i=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} \eta_{ibc}\, p_{i..}\, p_{.bc}$$

$$+ \sum_{a=1}^{r}\sum_{j=1}^{s}\sum_{c=1}^{t} \eta_{ajc}\, p_{.j.}\, p_{a.c} + \sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{k=1}^{t} \eta_{abk}\, p_{..k}\, p_{ab.}$$

$$- 2 \{\sum_{a=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{ajk}\, p_{a..}p_{.j.}p_{..k} + \sum_{i=1}^{r} \sum_{b=1}^{s} \sum_{k=1}^{t} \eta_{ibk}p_{i..}p_{.b.}p_{..k}$$

$$+ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{c=1}^{t} \eta_{ijc}p_{i..}p_{.j.}p_{..c}\}$$

$$\sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} p_{abc} \frac{\partial q}{\partial p_{abc}} = n-\theta + \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} \eta_{abc}[p_{a..}p_{.bc} + p_{.b.}p_{a.c} + p_{..c}p_{ab.} - 4p_{a..}p_{.b.}p_{..c}] \quad (5)$$

Set this equation equal to 0 .

Then $\theta = n + \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{ijk}[p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 4p_{i..}p_{.j.}p_{..k}]$

The equations $\frac{\partial q}{\partial p_{abc}} = 0$ $a = 1$ to r ; $b = 1$ to s ; $c = 1$ to t represent rst equations in $rst + 1$ unknowns; the η's and θ . We found that there are $(r-1)(s-1)(t-1) + 1$ linearly independent equations among these rst equations. To arrive at the equation (5) in θ we multiply the first of the $(r-1)(s-1)(t-1) + 1$ independent equations by a suitable nonzero constant and add to this multiples of the other remaining independent equations. We can then eliminate θ from the $(r-1)(s-1)(t-1)$ equations. It is clear that these $(r-1)(s-1)(t-1)$ equations are linearly independent.

F. THE SYSTEM OF EQUATIONS HAVE A FINITE SOLUTION

Now let us go back and look at the original equations in the form $p_{abc} \frac{\partial q}{\partial p_{abc}} = 0$. If θ has been replaced by the expression in the η_{abc}'s, then there are rst non-homogeneous linear equations in rst unknowns. The rank of the coefficient matrix now is $(r-1)(s-1)(t-1)$. In order for the system of equations to have a finite solution, the rank of the augmented matrix must be the same as the rank of the coefficient matrix.

The equations $p_{abc} \frac{\partial q}{\partial p_{abc}} = 0$ after the expression for θ has been substituted become

$$p_{abc} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{ijk}\left[\frac{g_{ijk}}{p_{abc}} + p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 4p_{i..}p_{.j.}p_{..k}\right]$$

$= (f_{abc} - np_{abc})$ (6) . Now the term of $p_{abc} \frac{\partial q}{\partial p_{abc}} = 0$ not involving the η_{abc} , after θ has been eliminated, is $(f_{abc} - np_{abc})$. In order for the augmented matrix to have rank $(r-1)(s-1)(t-1)$, it is necessary that

$\sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} (f_{abc} - np_{abc}) = 0$ whenever one or more of the indices u,v,w are zero. The above requirement for the equations $p_{abc} \frac{\partial q}{\partial p_{abc}} = 0$ to be consistent automatically rules out many otherwise possible sets of values for the p_{abc} .

Let $X^{(u)}$, $Y^{(v)}$, $Z^{(w)}$ represent orthonormal functions with respect to the distributions $\{p_{i..}\}$, $\{p_{.j.}\}$, $\{p_{..k}\}$ respectively, as before. Let $\hat{X}^{(u)}$, $\hat{Y}^{(v)}$, $\hat{Z}^{(w)}$ represent orthonormal functions with respect to the empirical distributions $\{\frac{f_{i..}}{n}\}$, $\{\frac{f_{.j.}}{n}\}$, $\{\frac{f_{..k}}{n}\}$ respectively.

Let $\rho_{uvw} = \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} p_{abc}$

$$\hat{\rho}_{uvw} = \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} \hat{X}_a^{(u)} \hat{Y}_b^{(v)} \hat{Z}_c^{(v)} \frac{f_{abc}}{n}$$

$$\tilde{\rho}_{uvw} = \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} \frac{f_{abc}}{n}$$

Then $\rho_{ooo} = 1$, $\rho_{uoo} = \rho_{ovo} = \rho_{oow} = 0$ for $u \neq 0$, $v \neq 0, w \neq 0$.

Similarly for $\hat{\rho}$ but not for $\tilde{\rho}$, though $\tilde{\rho}_{ooo} = 1$.

Write $\hat{q}_{abc} = \left(\frac{f_{a..} f_{.bc} + f_{.b.} f_{a.c} + f_{..c} f_{ab.}}{n^2} - \frac{2 f_{a..} f_{.b.} f_{..c}}{n^3}\right)$

$$\sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} \hat{X}_a^{(u)} \hat{Y}_b^{(v)} \hat{Z}_c^{(w)} \hat{q}_{abc} = \Delta_{uo} \hat{\rho}_{ovw} + \Delta_{vo} \hat{\rho}_{uow} + \Delta_{wo} \hat{\rho}_{uvo} - 2 \Delta_{uo} \Delta_{vo} \Delta_{wo}$$

where $\Delta_{gh} = \begin{cases} 1 \text{ for } g = h \\ 0 \text{ for } g \neq h \end{cases}$, since $\sum_{a=1}^{r} \hat{X}_a^{(u)} \hat{X}_a^{(0)} \frac{f_{a..}}{n} = \begin{cases} 1 \text{ for } u = 0 \\ 0 \text{ for } u \neq 0 \end{cases}$

Thus $$\sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} \hat{X}_a^{(u)} \hat{Y}_b^{(v)} \hat{Z}_c^{(w)} \hat{q}_{abc} = \begin{cases} 1 \text{ for } u = 0,\ v = 0,\ w = 0 \\ 0 \text{ for any two of } u, v, w = 0 \\ \hat{\rho}_{ovw} \text{ for } u = 0, \text{ but } v \neq 0,\ w \neq 0 \\ \hat{\rho}_{uow} \text{ for } v = 0, \text{ but } u \neq 0,\ w \neq 0 \\ \hat{\rho}_{uvo} \text{ for } w = 0, \text{ but } u \neq 0,\ v \neq 0 \\ 0 \text{ for } u \neq 0,\ v \neq 0,\ w \neq 0 \end{cases}$$

Similarly for $\sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} p_{abc}$, where the p_{abc} represent their hypothetical values, except that p is without $\wedge$.

$$\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} \hat{X}_a^{(u)}\hat{Y}_b^{(v)}\hat{Z}_c^{(w)} \frac{f_{abc}}{n} = \hat{\rho}_{uvw} = \begin{cases} 1 \text{ for } u = 0,\ v = 0,\ w = 0 \\ 0 \text{ for any two of } u,\ v,\ w = 0 \\ \hat{\rho}_{ovw} \text{ for } u = 0, \text{ but } v \neq 0,\ w \neq 0 \\ \hat{\rho}_{uow} \text{ for } v = 0, \text{ but } u \neq 0,\ w \neq 0 \\ \hat{\rho}_{uvo} \text{ for } w = 0, \text{ but } u \neq 0,\ v \neq 0 \\ \hat{\rho}_{uvw} \text{ for } u \neq 0,\ v \neq 0,\ w \neq 0 \end{cases}$$

Thus $$\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} \hat{X}_a^{(u)}\hat{Y}_b^{(v)}\hat{Z}_c^{(w)} (f_{abc} - n\hat{q}_{abc}) = \begin{cases} 0 & \text{when at least one of the } u,\ v,\ w \text{ is zero} \\ n\hat{\rho}_{uvw} & \text{when all indices are not zeros} \end{cases}$$

On the other hand $\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} X_a^{(u)}Y_b^{(v)}Z_c^{(w)} \frac{f_{abc}}{n} = \tilde{\rho}_{uvw}$, which usually $\neq \rho_{uvw}$ where $p_{a..} \neq \frac{f_{a..}}{n}$, etc. Also $\tilde{\rho}_{uoo}$, $\tilde{\rho}_{ovo}$, $\tilde{\rho}_{oow}$ can not be expected to be zero when $p_{a..} \neq \frac{f_{a..}}{n}$ etc. Thus $\sum_{a=1}^{r}\sum_{b=1}^{s}\sum_{c=1}^{t} X_a^{(u)}Y_b^{(v)}Z_c^{(w)}(f_{abc} - np_{abc}) = n(\tilde{\rho}_{uvw} - \rho_{uvw})$, where p_{abc} has its hypothetical value. For $u = 0$, $v = 0$, $w = 0$, one has $n(\tilde{\rho}_{ooo} - \rho_{ooo}) = 0$. In the other cases, one has $n(\tilde{\rho}_{uvw} - \rho_{uvw})$.

We conclude that if the equations (6) are multiplied by $X_a^{(u)}Y_b^{(v)}Z_c^{(w)}$ and then summed with respect to a, b, c, then, in general, the resulting rst equations will have on the right side $n(\tilde{\rho}_{uvw} - \rho_{uvw})$, of which (rst - 1) may be expected to be different from zero. However, the rank of the matrix of coefficients of η_{abc} on the left side is $(r-1)(s-1)(t-1)$, so that the equations will be inconsistent except for special sets

$\{p_{a..}\}$, $\{p_{.b.}\}$, $\{p_{..c}\}$ which leave an appropriate set of not more than $(r-1)(s-1)(t-1)$ terms on the right of the equations non-zero. We want to verify that $\{p_{a..}\} = \{\frac{f_{a..}}{n}\}$, $\{p_{.b.}\} = \{\frac{f_{.b.}}{n}\}$, $\{p_{..c}\} = \{\frac{f_{..c}}{n}\}$ will accomplish this.

A set of non-homogeneous linear equations are consistent, that is, have a solution $\Leftrightarrow$ rank of coefficient matrix = rank of augmented matrix $\Leftrightarrow$ column space of coefficient matrix = column space of the augmented matrix.

We have shown that the vector $\hat{\underline{q}}$ lies in the column space of the vectors $\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)}$, where $u \neq 0$, $v \neq 0$, $w \neq 0$, but that, in general, the same can not be said of other possible $\underline{p}$. If we can show that the column space of

$$\left[\frac{\partial g_{ijk}}{\partial p_{abc}} + p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 4p_{i..}p_{.j.}p_{..k}\right]$$ equals the

column space of $\{\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)} \mid u \neq 0,\ v \neq 0,\ w \neq 0\}$

then the set of equations will be consistent when $\underline{p} = \hat{\underline{q}}$. Note that here (a, b, c) indexes rows, (i, j, k) and (u, v, w) columns.

The column space of the coefficient matrix above is a subspace of rst-dimensional Euclidean space, E_{rst} . The matrix was shown to have rank $(r-1)(s-1)(t-1)$, hence its column space is an $(r-1)(s-1)(t-1)$ subspace of E_{rst} .

Now the set of vectors $\left\{\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)} \;\middle|\; \begin{array}{l} u = 0 \text{ to } (r-1) \\ v = 0 \text{ to } (s-1) \\ w = 0 \text{ to } (t-1) \end{array}\right\}$

forms a basis for E_{rst} . We will show that the column space of the coefficient matrix is orthogonal to the subspace spanned by the set of vectors $\{\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)} \mid$ one or more of u, v, w is zero$\}$. This is a $1 + (r-1) + (s-1) + (t-1) + (s-1)(t-1) + (r-1)(t-1) + (r-1)(s-1) = rst - (r-1)(s-1)(t-1)$ dimensional subspace. Hence its complement, the $(r-1)(s-1)(t-1)$-dimensional subspace spanned by the set of vectors $\{\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)} \mid u \neq 0, v \neq 0, w \neq 0\}$, is the same as the column space of the coefficient matrix. We have shown that the vector $(\underline{f} - n\hat{\underline{q}})$ lies in this subspace. Hence the set of equations would be consistent when $(\underline{f} - n\hat{\underline{q}})$ is the non-homogeneous part of the equations.

To show that the column space of the coefficient matrix is orthogonal to the subspace spanned by the set of vectors $\{\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)} \mid$ one or more of u, v, w is zero$\}$, recall

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{ijk}}{\partial p_{abc}} = \begin{cases} 1 \text{ for } u = v = w = 0 \\ 0 \text{ if two of } u, v, w \text{ are } 0, \text{ one not } 0 \\ -\rho_{uvw} \text{ if one of } u, v, w \text{ is } 0, \text{ two not } 0 \\ X_a^{(u)} Y_b^{(v)} Z_c^{(w)} - \{X_a^{(u)} \rho_{ovw} + Y_b^{(v)} \rho_{uow} + \\ \qquad Z_c^{(w)} \rho_{uvo}\} \text{if none of } u, v, w \text{ is } 0 . \end{cases}$$

Recall also that $\sum_{i=1}^{r} X_i^{(u)} p_{i..} = 0$ for $u \neq 0$

$$\sum_{j=1}^{s} \sum_{k=1}^{t} Y_j^{(v)} Z_k^{(w)} p_{.jk} = \rho_{ovw} \qquad , \text{etc.}$$

Hence $$\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \left[\frac{\partial g_{ijk}}{\partial p_{abc}} + p_{i..} p_{.jk} + p_{.j.} p_{i.k} + p_{..k} p_{ij.} - 4 p_{i..} p_{.j.} p_{..k}\right]$$

$$= \begin{cases} 0 \text{ if at least one of } u, v, w \text{ is } 0 \\ X_a^{(u)} Y_b^{(v)} Z_c^{(w)} - \{X_a^{(u)} \rho_{ovw} + Y_b^{(v)} \rho_{uow} + Z_c^{(w)} \rho_{uvo}\} \text{ if none of } u, v, w \text{ is } 0 \,. \end{cases}$$

This shows that the column space of the matrix

$[\frac{\partial g_{ijk}}{\partial p_{abc}} + p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 4p_{i..}p_{.j.}p_{..k}]$ is orthogonal to

all the vectors $\underline{X}^{(u)} \otimes \underline{Y}^{(v)} \otimes \underline{Z}^{(w)}$ for which at least one of u, v, w is 0 .

G. THE NUMERATOR OF THE LIKELIHOOD RATIO

The numerator of our likelihood ratio we now see to be

$$n! \prod_{i=1}^{r} \prod_{j=1}^{s} \prod_{k=1}^{t} \frac{\left(\dfrac{f_{i..}f_{.jk} + f_{.j.}f_{i.k} + f_{..k}f_{ij.}}{n^2} - \dfrac{2f_{i..}f_{.j.}f_{..k}}{n^3}\right)^{f_{ijk}}}{f_{ijk}!}$$

H. SUMMARY OF CHAPTER III

In this chapter we discussed testing the hypothesis of no second order interaction. We chose the method of the likelihood ratio. There are no difficulties in determining the denominator of the ratio. To assist us in determining the numerator of the likelihood ratio we introduced a theorem about Lagrange multipliers, modified it, and proved the modified version. We then proceeded to show that our problem does satisfy the conditions of the modified theorem and hence our use of the modified theorem was valid. We finally arrived at an expression for the numerator of the likelihood ratio.

One could alternately think of this chapter as finding the maximum likelihood estimators of the p_{ijk}, first under the most general assumptions, then under the hypothesis of no second order interaction, that is, $\delta_{ijk} = 0$ for every i, j, k. We could then decide to test the hypothesis of no second order interaction using the likelihood ratio criterion or $-2 \log$ (likelihood ratio), or else we could use

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \left(\frac{f_{ijk} - n\hat{q}_{ijk}}{n\hat{q}_{ijk}}\right)^2,$$ where this $\hat{q}_{ijk}$ is the restricted maximum likelihood estimator of p_{ijk} under the hypothesis that second order interaction is zero.

CHAPTER IV CONTINGENCY TABLES FOR WHICH THE HYPOTHESIS OF NO SECOND ORDER INTERACTION IS NOT MEANINGFUL

Ordinarily the p_{ijk} can vary over a finite region. When the marginal totals are specified, that leaves an $(r-1)(s-1)(t-1)$ subspace for the p_{ijk} to vary over. We expect that the maximum likelihood estimators of the q_{ijk} will usually lie in the interior of this region. Occasionally they may have to lie on the boundary, on one of the faces of the region or where two or more of the faces intersect.

Now $P(\text{neither } A \text{ nor } B, \text{ but } C) = P(C) - P\{(A \cup B)C\} = P(C) - P\{(AC) \cup (BC)\}$ (1)
$= P(C) - P(AC) - P(BC) + P(ABC) \geq 0$. Hence $P(ABC) \geq P(AC) + P(BC) - P(C)$.
Thus it follows that

$$\max\{0, p_{i.k} + p_{.jk} - p_{..k}, p_{ij.} + p_{i.k} - p_{i..}, p_{ij.} + p_{.jk} - p_{.j.}\}$$

$$\leq p_{ijk} \leq \min\{p_{.jk}, p_{i.k}, p_{ij.}\}$$. We have not really used these restrictions in deriving the maximum likelihood solution.

In a three-way contingency table it is sometimes possible for one or more q_{ijk} to be

$$< \max\{0, p_{i.k} + p_{ij.} - p_{i..}, p_{ij.} + p_{.jk} - p_{.j.}, p_{.jk} + p_{i.k} - p_{..k}\}$$

or $> \min\{p_{.jk}, p_{i.k}, p_{ij.}\}$, which it is impossible for p_{ijk} to satisfy. Similarly for $\hat{q}_{ijk}$ and the corresponding relative frequencies.

Let Ω' denote the set of real non-negative numbers $\{p_{111}, \ldots, p_{rst}\}$ which are possible probabilities in an $r \times s \times t$ contingency table with given marginal probabilities. In the cases indicated above with the p_{ijk} set equal to q_{ijk} or $\hat{p}_{ijk}$ set equal to $\hat{q}_{ijk}$, the point does not lie in Ω'. There is no solution $(\hat{p}_{111},\ldots,\hat{p}_{rst})$ lying within Ω' under the conditions imposed. We do not want to tamper with the marginal totals or the two-way table interactions, so what can we do?

Since $p_{ijk} = p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 2p_{i..}p_{.j.}p_{..k} + \delta_{ijk}$, expression $q_{ijk} < \max\{0, p_{i.k} + p_{ij.} - p_{i..}, p_{ij.} + p_{.jk} - p_{.j.}, p_{.jk} + p_{i.k} - p_{..k}$ implies that $\delta_{ijk} > 0$, while expression $q_{ijk} > \min\{p_{.jk}, p_{i.k}, p_{ij.}\}$ implies that $\delta_{ijk} < 0$. This suggests that when $(\hat{p}_{111},\ldots,\hat{p}_{rst})$ lies outside of Ω' we not hypothesize that all of the δ_{ijk} are zero, but rather allow certain contrasts of the δ_{ijk} to be non-zero while hypothesizing that all linear combinations of the δ_{ijk} orthogonal to the given contrasts be zero. Possibly one could estimate the direction of this vector (the vector of coefficients of the δ_{ijk} in the non-zero contrast), but it is much simpler to prescribe the direction to begin with and just worry about the length of the vector.

Consider the following marginal totals of a $3\times2\times2$ contingency table:

$C_1 \cup C_2$	B_1	B_2	Sum
A_1	1/12	3/12	1/3
A_2	2/12	2/12	1/3
A_3	3/12	1/12	1/3
Sum	1/2	1/2	1

$B_1 \cup B_2$	C_1	C_2	Sum
A_1	1/24	7/24	1/3
A_2	2/24	6/24	1/3
A_3	3/24	5/24	1/3
Sum	1/4	3/4	1

$A_1 \cup A_2 \cup A_3$	C_1	C_2	Sum
B_1	1/12	5/12	1/2
B_2	2/12	4/12	1/2
Sum	1/4	3/4	1

Note that $p_{1..} = p_{2..} = p_{3..} = 1/3$

$p_{.1.} = p_{.2.} = 1/2$

$p_{..1} = 1/4$, $p_{..2} = 3/4$

Consider the following table:

(ijk)	$p_{ij.}+p_{i.k}-p_{i..}$	$p_{ij.}+p_{.jk}-p_{.j.}$	$p_{i.k}+p_{.jk}-p_{..k}$	possible values of p_{ijk}
111	$\frac{1}{12}+\frac{1}{24}-\frac{8}{24}=\frac{-5}{24}$	$\frac{1}{12}+\frac{1}{12}-\frac{6}{12}=\frac{-8}{24}$	$\frac{1}{24}+\frac{1}{12}-\frac{6}{24}=\frac{-3}{24}$	0 to $\frac{1}{24}$
112	$\frac{1}{12}+\frac{7}{24}-\frac{8}{24}=\frac{1}{24}$	$\frac{1}{12}+\frac{5}{12}-\frac{6}{12}=0$	$\frac{7}{24}+\frac{5}{12}-\frac{18}{24}=\frac{-1}{24}$	$\frac{1}{24}$ to $\frac{2}{24}$
121	$\frac{3}{12}+\frac{1}{24}-\frac{8}{24}=\frac{-1}{24}$	$\frac{3}{12}+\frac{2}{12}-\frac{6}{12}=\frac{-2}{24}$	$\frac{1}{24}+\frac{2}{12}-\frac{6}{24}=\frac{-1}{24}$	0 to $\frac{1}{24}$
122	$\frac{3}{12}+\frac{7}{24}-\frac{8}{24}=\frac{5}{24}$	$\frac{3}{12}+\frac{4}{12}-\frac{6}{12}=\frac{2}{24}$	$\frac{7}{24}+\frac{4}{12}-\frac{18}{24}=\frac{-3}{24}$	$\frac{5}{24}$ to $\frac{6}{24}$
211	$\frac{2}{12}+\frac{2}{24}-\frac{8}{24}=\frac{-2}{24}$	$\frac{2}{12}+\frac{1}{12}-\frac{6}{12}=\frac{-6}{24}$	$\frac{2}{24}+\frac{1}{12}-\frac{6}{24}=\frac{-2}{24}$	0 to $\frac{2}{24}$
212	$\frac{2}{12}+\frac{6}{24}-\frac{8}{24}=\frac{2}{24}$	$\frac{2}{12}+\frac{5}{12}-\frac{6}{12}=\frac{2}{24}$	$\frac{6}{24}+\frac{5}{12}-\frac{18}{24}=\frac{-2}{24}$	$\frac{2}{24}$ to $\frac{4}{24}$
221	$\frac{2}{12}+\frac{2}{24}-\frac{8}{24}=\frac{-2}{24}$	$\frac{2}{12}+\frac{2}{12}-\frac{6}{12}=\frac{-4}{24}$	$\frac{2}{24}+\frac{2}{12}-\frac{6}{24}=0$	0 to $\frac{2}{24}$
222	$\frac{2}{12}+\frac{6}{24}-\frac{8}{24}=\frac{2}{24}$	$\frac{2}{12}+\frac{4}{12}-\frac{6}{12}=0$	$\frac{6}{24}+\frac{4}{12}-\frac{18}{24}=\frac{-4}{24}$	$\frac{2}{24}$ to $\frac{4}{24}$
311	$\frac{3}{12}+\frac{3}{24}-\frac{8}{24}=\frac{1}{24}$	$\frac{3}{12}+\frac{1}{12}-\frac{6}{12}=\frac{-4}{24}$	$\frac{3}{24}+\frac{1}{12}-\frac{6}{24}=\frac{-1}{24}$	$\frac{1}{24}$ to $\frac{2}{24}$
312	$\frac{3}{12}+\frac{5}{24}-\frac{8}{24}=\frac{3}{24}$	$\frac{3}{12}+\frac{5}{12}-\frac{6}{12}=\frac{4}{24}$	$\frac{5}{24}+\frac{5}{12}-\frac{18}{24}=\frac{-3}{24}$	$\frac{4}{24}$ to $\frac{5}{24}$
321	$\frac{1}{12}+\frac{3}{24}-\frac{8}{24}=\frac{-3}{24}$	$\frac{1}{12}+\frac{2}{12}-\frac{6}{12}=\frac{-6}{24}$	$\frac{3}{24}+\frac{2}{12}-\frac{6}{24}=\frac{1}{24}$	$\frac{1}{24}$ to $\frac{2}{24}$
322	$\frac{1}{12}+\frac{5}{24}-\frac{8}{24}=\frac{-1}{24}$	$\frac{1}{12}+\frac{4}{12}-\frac{6}{12}=\frac{-2}{24}$	$\frac{5}{24}+\frac{4}{12}-\frac{18}{24}=\frac{-5}{24}$	0 to $\frac{2}{24}$

TABLE XIV UPPER AND LOWER BOUNDS OF p_{ijk} .

Let us evaluate $q_{111} \cdot q_{111} = p_{1..}p_{.11} + p_{.1.}p_{1.1} + p_{..1}p_{11.} - 2p_{1..}p_{.1.}p_{..1}$ $= (\frac{1}{3} \times \frac{1}{12}) + (\frac{1}{2} \times \frac{1}{24}) + (\frac{1}{4} \times \frac{1}{12}) - 2(\frac{1}{3} \times \frac{1}{2} \times \frac{1}{4}) = \frac{-1}{72}$. Similarly $q_{211} = \frac{2}{72}$. We now enter the q_{ijk} in the cells of the table. For (i, j, k) = (1, 1, 1) the entry is -1/72 ; for (i, j, k) =(2, 1, 1) it is 2/72.

C_1	B_1	B_2	Sum
A_1	$\frac{-1}{72}$	$\frac{4}{72}$	$\frac{3}{72}$
A_2	$\frac{2}{72}$	$\frac{4}{72}$	$\frac{6}{72}$
A_3	$\frac{5}{72}$	$\frac{4}{72}$	$\frac{9}{72}$
Sum	$\frac{6}{72}$	$\frac{12}{72}$	$\frac{18}{72}$

C_2	B_1	B_2	Sum
A_1	$\frac{7}{72}$	$\frac{14}{72}$	$\frac{21}{72}$
A_2	$\frac{10}{72}$	$\frac{8}{72}$	$\frac{18}{72}$
A_3	$\frac{13}{72}$	$\frac{2}{72}$	$\frac{15}{72}$
Sum	$\frac{30}{72}$	$\frac{24}{72}$	$\frac{54}{72}$

$C_1 \cup C_2$	B_1	B_2	Sum
A_1	$\frac{6}{72}$	$\frac{18}{72}$	$\frac{24}{72}$
A_2	$\frac{12}{72}$	$\frac{12}{72}$	$\frac{24}{72}$
A_3	$\frac{18}{72}$	$\frac{6}{72}$	$\frac{24}{72}$
Sum	$\frac{36}{72}$	$\frac{36}{72}$	1

$A_1 \cup A_2 \cup A_3$	C_1	C_2	Sum
B_1	$\frac{6}{72}$	$\frac{30}{72}$	$\frac{36}{72}$
B_2	$\frac{12}{72}$	$\frac{24}{72}$	$\frac{36}{72}$
Sum	$\frac{18}{72}$	$\frac{54}{72}$	1

$B_1 \cup B_2$	C_1	C_2	Sum
A_1	$\frac{3}{72}$	$\frac{21}{72}$	$\frac{24}{72}$
A_2	$\frac{6}{72}$	$\frac{18}{72}$	$\frac{24}{72}$
A_3	$\frac{9}{72}$	$\frac{15}{72}$	$\frac{24}{72}$
Sum	$\frac{18}{72}$	$\frac{54}{72}$	1

The circled entries are impossible.

The intuitive way to apportion the δ contributions of the layers for A_2 and A_3 is to make the δ contributions proportional to $p_{2..}$ and $p_{3..}$ (or to $f_{2..}$ and $f_{3..}$) respectively. In the example $p_{2..} = p_{3..} = 1/3$ (or at least $f_{2..} = f_{3..} = 24$) .

$$\text{Let } \tilde{q}_{ijk} = \frac{f_{i..}f_{.jk} + f_{.j.}f_{i.k} + f_{..k}f_{ij.}}{n^2} - \frac{2f_{i..}f_{.j.}f_{..k}}{n^3}$$

In our example the $\tilde{q}_{ijk}$ give us impossible estimates:

$\tilde{q}_{111} = -1/72 < 0$, etc. Suppose that

$$\delta_{111} = \frac{-(f_{2..} + f_{3..})\lambda}{n} = -(1 - \frac{f_{1..}}{n})\lambda$$

$$\delta_{211} = \frac{f_{2..}}{n}\lambda$$

$$\delta_{311} = \frac{f_{3..}}{n}\lambda$$

Then $\delta_{111} + \delta_{211} + \delta_{311} = 0$. The δ's in the other columns are automatically specified, since $\delta_{ij1} + \delta_{ij2} = 0$, $\delta_{i1k} + \delta_{i2k} = 0$.

The problem then is to estimate λ , which is proportional to the length of the coefficient vector of the δ contrast.

$$p_{111} = \tilde{q}_{111} - \frac{(f_{2..} + f_{3..})}{n}\lambda \qquad p_{121} = \tilde{q}_{121} + \frac{(f_{2..} + f_{3..})}{n}\lambda$$

$$p_{211} = \tilde{q}_{211} + \frac{f_{2..}}{n}\lambda \qquad p_{221} = \tilde{q}_{221} - \frac{f_{2..}}{n}\lambda$$

$$p_{311} = \tilde{p}_{311} + \frac{f_{3..}}{n}\lambda \qquad p_{321} = \tilde{q}_{321} - \frac{f_{3..}}{n}\lambda$$

$$p_{212} = \tilde{q}_{212} \frac{-f_{2..}}{n}\lambda \qquad p_{222} = \tilde{q}_{222} + \frac{f_{2..}}{n}\lambda$$

$$p_{312} = \tilde{q}_{312} \frac{-f_{3..}}{n}\lambda \qquad p_{322} = \tilde{q}_{322} + \frac{f_{3..}}{n}\lambda$$

$$p_{112} = \tilde{q}_{112} + \frac{(f_{2..} + f_{3..})\lambda}{n} \qquad p_{122} = \tilde{q}_{122} - \frac{(f_{2..} + f_{3..})\lambda}{n}$$

$$p = n! \prod_{i=1}^{3} \prod_{j=1}^{2} \prod_{k=1}^{2} \frac{p_{ijk}^{f_{ijk}}}{f_{ijk}!}$$

$$Q = \log n! + \sum_{i=1}^{3} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log p_{ijk} - \sum_{i=1}^{3} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk}!$$

$$\frac{\partial Q}{\partial \lambda} = \frac{-f_{111}(\frac{f_{2..}+f_{3..}}{n})}{\tilde{q}_{111} - (\frac{f_{2..}+f_{3..}}{n})\lambda} + \frac{f_{211}\frac{f_{2..}}{n}}{\tilde{q}_{211} + \frac{f_{2..}}{n}\lambda} + \frac{f_{311}\frac{f_{3..}}{n}}{\tilde{q}_{311} + \frac{f_{3..}}{n}\lambda}$$

$$+ \frac{f_{121}(\frac{f_{2..}+f_{3..}}{n})}{\tilde{q}_{121} + (\frac{f_{2..}+f_{3..}}{n})\lambda} - \frac{f_{221}\frac{f_{2..}}{n}}{\tilde{q}_{221} - \frac{f_{2..}}{n}\lambda} - \frac{f_{321}\frac{f_{3..}}{n}}{\tilde{q}_{321} - \frac{f_{3..}}{n}\lambda}$$

$$+ \frac{f_{112}(\frac{f_{2..}+f_{3..}}{n})}{\tilde{q}_{112} + (\frac{f_{2..}+f_{3..}}{n})\lambda} - \frac{f_{212}\frac{f_{2..}}{n}}{\tilde{q}_{212} - \frac{f_{2..}}{n}\lambda} - \frac{f_{312}\frac{f_{3..}}{n}}{\tilde{q}_{312} - \frac{f_{3..}}{n}\lambda}$$

$$- \frac{f_{112}(\frac{f_{2..}+f_{3..}}{n})}{\tilde{q}_{112} - (\frac{f_{2..}+f_{3..}}{n})\lambda} + \frac{f_{222}\frac{f_{2..}}{n}}{\tilde{q}_{222} + \frac{f_{2..}}{n}\lambda} + \frac{f_{322}\frac{f_{3..}}{n}}{\tilde{q}_{322} + \frac{f_{3..}}{n}\lambda}$$

In C.R. Rao's text "Advanced Statistical Methods in Biometric Research" we find the method of scoring for the estimation of parameters. This method works well with maximum likelihood equations which are too complicated to solve directly. The above equation is such a one. The general method in such cases is to assume a trial solution and derive linear equations for small additive corrections. The process can be repeated till the corrections become negligible.

The quantity $\frac{\partial Q}{\partial \lambda}$ is defined as the efficient score for λ. The maximum likelihood estimate of λ is that value of λ for which the efficient score vanishes. If λ_o is the trial value of the estimate, then expanding $\frac{\partial Q}{\partial \lambda}$ and retaining only the first power of $d\theta = \theta - \theta_o$,

$$\frac{\partial Q}{\partial \lambda} \simeq \frac{\partial Q}{\partial \lambda_o} + d\lambda \frac{\partial^2 Q}{\partial \lambda_o^2}$$

$$\simeq \frac{\partial Q}{\partial \lambda_o} - d\lambda\, I(\lambda_o)$$ where $I(\lambda_o)$, the information at the value $\lambda = \lambda_o$, is the expected value of $\frac{\partial^2 Q}{d\lambda^2}$. In large samples the difference between $-I(\lambda_o)$ and $\frac{\partial^2 Q}{\partial \lambda_o^2}$ will be of $0(1/n)$, where n is the number of observations, so that the above approximation holds to the first order of small quantities. The correction $d\lambda$ is obtained from the equation

$$d\lambda I(\lambda_o) = \frac{\partial Q}{\partial \lambda_o}$$

$$d\lambda = \frac{\partial Q}{\partial \lambda_o} \div I(\lambda_o)$$

The first approximation is $(\lambda_o + d\lambda)$, and the above process can be repeated with this as the new trial value.

This chapter has by no means exhausted the possible methods of treating three-way contingency tables for which the hypothesis of no second order interaction is not meaningful. It does however propose an idea of how these situations can be dealt with.

CHAPTER V EXTENSION TO HIGHER DIMENSIONAL CONTINGENCY TABLES.

A. TWO EXTENSIONS OF THE THREE DIMENSIONAL MODEL TO THE FOUR DIMENSIONAL CASE.

Consider a $q \times r \times s \times t$ contingency table. Consider as well the following two possible extensions of our model:

I. $$E(f_{hijk}) = np_{hijk} = n[p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k} + p_{h...}\, p_{.i..}\, \alpha^{(34)}_{jk} + p_{h...}\, p_{..j.}\, \alpha^{(24)}_{ik} + p_{h...}\, p_{...k}\, \alpha^{(23)}_{ij} + p_{.i..}\, p_{..j.}\, \alpha^{(14)}_{hk} + p_{.i..}\, p_{...k}\, \alpha^{(13)}_{hj}$$

$$+ p_{..j.}\, p_{...k}\, \alpha^{(12)}_{hi} + p_{h...}\, \delta^{(234)}_{ijk} + p_{.i..}\, \delta^{(134)}_{hjk} + p_{..j.}\, \delta^{(124)}_{hik} + p_{...k}\, \delta^{(123)}_{hij}$$

$+ \theta_{hijk}]$, $h = 1$ to q, $i = 1$ to r, $j = 1$ to s, $k = 1$ to t with

$\alpha^{(34)}_{.k} = \alpha^{(34)}_{j.} = 0$; $\alpha^{(24)}_{.k} = \alpha^{(24)}_{i.} = 0$; $\alpha^{(23)}_{.j} = \alpha^{(23)}_{i.} = 0$; $\alpha^{(14)}_{.k} = \alpha^{(14)}_{h.} = 0$;

$\alpha^{(13)}_{.j} = \alpha^{(13)}_{h.} = 0$; $\alpha^{(12)}_{.i} = \alpha^{(12)}_{h.} = 0$; $\delta^{(234)}_{.jk} = \delta^{(234)}_{i.k} = \delta^{(234)}_{ij.} = 0$;

$\delta^{(134)}_{.jk} = \delta^{(134)}_{h.k} = \delta^{(134)}_{hj.} = 0$; $\delta^{(124)}_{.ik} = \delta^{(124)}_{h.k} = \delta^{(124)}_{hi.} = 0$; $\delta^{(123)}_{.ij} = \delta^{(123)}_{h.j}$

$= \delta^{(123)}_{hi.} = 0$; $\theta_{.ijk} = \theta_{h.jk} = \theta_{hi.k} = \theta_{hij.} = 0$ where the dot indicates summation over the replaced subscript.

II. $$E(f_{hijk}) = np_{hijk} = n[p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k} + p_{hi..}\, \alpha'^{(34)}_{jk}$$

$$+ p_{h.j.}\, \alpha'^{(14)}_{ik} + p_{h..k}\, \alpha'^{(23)}_{ij} + p_{.ij.}\, \alpha'^{(14)}_{hk} + p_{.i.k}\, \alpha'^{(13)}_{hj} + p_{..jk}\, \alpha'^{(12)}_{hi}$$

$$+ p_{h...}\, \delta'^{(234)}_{ijk} + p_{.i..}\, \delta'^{(134)}_{hjk} + p_{..j.}\, \delta'^{(124)}_{hik} + p_{...k}\, \delta'^{(123)}_{hij} + \theta'_{hijk}],$$

$h = 1$ to q, $i = 1$ to r, $j = 1$ to s, $k = 1$ to t with

$$\alpha'^{(34)}_{.k} = \alpha'^{(34)}_{j.} = 0;\ \alpha'^{(24)}_{.k} = \alpha'^{(24)}_{i.} = 0;\ \alpha'^{(23)}_{.j} = \alpha'^{(23)}_{i.} = 0;\ \alpha'^{(14)}_{.k} = \alpha'^{(14)}_{h.} = 0;$$

$$\alpha'^{(13)}_{.j} = \alpha'^{(13)}_{h.} = 0;\ \alpha'^{(12)}_{.i} = \alpha'^{(12)}_{h.} = 0;\ \delta'^{(234)}_{.jk} = \delta'^{(234)}_{i.k} = \delta'^{(234)}_{ij.} = 0;$$

$$\delta'^{(134)}_{.jk} = \delta'^{(134)}_{h.k} = \delta'^{(134)}_{hj.} = 0;\ \delta'^{(124)}_{.ik} = \delta'^{(124)}_{h.k} = \delta'^{(124)}_{..k} = 0;\ \delta'^{(123)}_{.ij} = \delta'^{(123)}_{h.j}$$

$= \delta'^{(123)}_{hi..} = 0;\ \theta'_{.ijk} = \theta'_{h.jk} = \theta'_{hi.k} = \theta'_{hij.} = 0$ where the dot indicates summation over the replaced subscript.

Note that if we sum model I over h we obtain

$$E(f_{.ijk}) = np_{.ijk} = n[p_{.i..}\ p_{..j.}\ p_{...k} + p_{.i..}\ \alpha^{(34)}_{jk} + p_{..j.}\ \alpha^{(24)}_{ik}$$

$$+ p_{...k}\ \alpha^{(23)}_{ij} + \delta^{(234)}_{ijk}\ ]$$ which we note is the model for the three dimensional case. The same result holds true if we sum over h in the second extension of our model.

Let us now determine θ_{hijk} . $E(f_{..jk}) = np_{..jk}$

$$= n[p_{..j.}p_{...k} + \alpha^{(34)}_{jk}\ ]$$

$$\Rightarrow \alpha^{(34)}_{jk} = p_{..jk} - p_{..j.}\ p_{...k}$$

Similarly for $\alpha^{(24)}_{ik}$, $\alpha^{(23)}_{ij}$, $\alpha^{(14)}_{hk}$, $\alpha^{(13)}_{hj}$, $\alpha^{(12)}_{hi}$

$$E(f_{.ijk}) = np_{.ijk} = n[p_{.i..}p_{..j.}p_{...k} + p_{.i..}\ \alpha^{(34)}_{jk} + p_{..j.}\ \alpha^{(24)}_{ik} + p_{...k}\alpha^{(23)}_{ij} + \delta^{(234)}_{ijk}\ ]$$

$$\Rightarrow \delta_{ijk}^{(234)} = p_{.ijk} - p_{.i..}p_{..j.}p_{...k} - p_{.i..}\alpha_{jk}^{(34)} - p_{..j.}\alpha_{ik}^{(24)} - p_{...k}\alpha_{ij}^{(23)}$$

$$= p_{.ijk} + 2p_{.i..}p_{..j.}p_{...k} - p_{.i..}p_{..jk} - p_{..j.}p_{.i.k} - p_{...k}p_{.ij.}$$

Similarly for $\delta_{hjk}^{(134)}$, $\delta_{hik}^{(124)}$, $\delta_{hij}^{(123)}$.

$$\theta_{hijk} = p_{hijk} - p_{h...}p_{.i..}p_{..j.}p_{...k} - p_{h...}p_{.i..}\alpha_{jk}^{(34)}$$

$$- p_{h...}p_{..j.}\alpha_{ik}^{(24)} - p_{h...}p_{...k}\alpha_{ij}^{(23)} - p_{.i..}p_{..j.}\alpha_{hk}^{(14)} - p_{.i..}p_{...k}\alpha_{hj}^{(13)}$$

$$- p_{..j.}p_{...k}\alpha_{hi}^{(12)} - p_{h...}\delta_{ijk}^{(234)} - p_{.i..}\delta_{hjk}^{(134)} - p_{..j.}\delta_{hik}^{(124)}$$

$- p_{...k}\delta_{hij}^{(123)}$. After substituting the expressions just derived for $\alpha_{jk}^{(34)}$ and the other α's and for $\delta_{ijk}^{(234)}$ and the other δ's into the formula for θ_{hijk}, and combining terms we obtain:

$$\theta_{hijk} = p_{hijk} - 3p_{h...}p_{.i..}p_{..j.}p_{...k} - p_{h...}p_{.ijk} - p_{.i..}p_{h.jk}$$

$$- p_{..j.}p_{hi.k} - p_{...k}p_{hij.} + p_{h...}p_{.i..}p_{..jk} + p_{h...}p_{..j.}p_{.i.k}$$

$$+ p_{h...}p_{...k}p_{.ij.} + p_{.i..}p_{..j.}p_{h..k} + p_{.i..}p_{...k}p_{h.j.}$$

$$+ p_{..j.}p_{...k}p_{hi..}.$$

Let us now determine θ'_{hijk}. From the second extension of our model $E(f_{..jk}) = np_{..jk} = n[p_{..j.}p_{...k} + \alpha'^{(34)}_{jk}] \Rightarrow \alpha'^{(34)}_{jk} = p_{..jk} - p_{..j.}p_{...k}$.

Similarly for $\alpha'^{(24)}_{ik}$, $\alpha'^{(23)}_{ij}$, $\alpha'^{(14)}_{hk}$, $\alpha'^{(13)}_{hj}$, $\alpha'^{(12)}_{hi}$. $\delta'^{(234)}_{ijk} = p_{.ijk}$

$+ 2p_{.i..}\, p_{..j.}\, p_{...k} - p_{.i..}\, p_{..jk} - p_{..j.}\, p_{.i.k} - p_{...k}\, p_{.ij.}$

Similarly for $\delta'^{(134)}_{hjk}$, $\delta'^{(124)}_{hik}$, $\delta'^{(123)}_{hij}$. $\theta'_{hijk} = p_{hijk} -$

$p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k} - p_{hi..}\, \alpha'^{(34)}_{jk} - p_{h.j.}\, \alpha'^{(24)}_{ik} - p_{h..k}\, \alpha'^{(23)}_{ij}$

$- p_{.ij.}\, \alpha'^{(14)}_{hk} - p_{.i.k}\, \alpha'^{(13)}_{hj} - p_{..jk}\, \alpha'^{(12)}_{hi} - p_{h...}\, \delta'^{(234)}_{ijk} - p_{.i..}\, \delta'^{(134)}_{hjk}$

$- p_{..j.}\, \delta'^{(124)}_{hik} - p_{...k}\, \delta'^{(123)}_{hij}$. After substituting the expressions just derived for $\alpha'^{(34)}_{jk}$ and the other α''s and for $\delta'^{(234)}_{ijk}$ and the other δ''s into the formula for θ'_{hijk}, and combining terms we obtain:

$$\theta'_{hijk} = p_{hijk} - 9p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k} + 3p_{hi..}\, p_{.j..}\, p_{...k} + 3p_{h.j.}\, p_{.i..}\, p_{...k} + 3p_{h..k}\, p_{.i..}\, p_{..j.} + 3p_{.ij.}\, p_{h...}\, p_{...k} + 3p_{.i.k}\, p_{h...}\, p_{..j.}$$

$$+ 3p_{..jk}\, p_{h...}\, p_{.i..} - 2p_{hi..}\, p_{..jk} - 2p_{h.j.}\, p_{.i.k} - 2p_{.ij.}\, p_{h..k}$$

$$- p_{h...}\, p_{.ijk} - p_{.i..}\, p_{h.jk} - p_{..j.}\, p_{hi.k} - p_{...k}\, p_{hij.}.$$

B. LANCASTER'S DEFINITION OF INTERACTIONS COMPARED TO OURS

Consider for a moment the following definition of interactions given by Lancaster in his text "The Chi-Squared Distribution" (1969) pp. 254-256: Let F_{i_1}, $F_{i_1 i_2}$, $F_{i_1 i_2 i_3}, \ldots$ denote the one-, two-, three-, ... dimensional distribution functions, where $i_1 < i_2 < i_3 < \ldots\ldots$.

Let F' denote the (n-1)-dimensional distribution function of the set complementary to the particular random variable chosen, F'' as the (n-2) dimensional distribution function of the (n-2) variables complementary to the particular pair (i_1, i_2) chosen, and so on. With this convention, interactions of the $(n-1)^{th}$ order are said to exist if, and only if, F cannot be displayed as a sum

$$F = \Sigma F_{i_1} F' - \Sigma F_{i_1} F_{i_2} F'' + \Sigma F_{i_1} F_{i_2} F_{i_3} F''' - \ldots (-1)^n \Sigma F_{i_1} F_{i_2} \ldots F_{i_{(n-1)}} F^{(n-1)}$$

$$+ (-1)^{n-1} F_{i_1} F_{i_2} \ldots F_{i_n}$$

where the summation is over all combinations of indices $(i_1, i_2, \ldots, i_k)$, $k = 1, 2, \ldots, (n-1)$. We have written $F^{(n-1)}$ to mean F with (n-1) primes as superscripts. The last two terms are of the same form and may be consolidated as $(-1)^n (n-1) F_{i_1} F_{i_2} \ldots F_{i_n}$, $n > 2$.

For three variables, X, Y, Z second order interactions exist if F cannot be written as the following sum:

$F = F_1(X) F_{23}(YZ) + F_2(Y) F_{13}(XZ) + F_3(Z) F_{12}(XY) - 2F_1(X) F_2(Y) F_3(Z)$.

Now let X, Y, Z be code random variables: $X = i$ when an observation falls in the A_i category of the first classification, $i = 1$ to r, and so on. The distributions are then discrete. Then $p_{ijk} = \Delta^3 F_{XYZ}(i,j,k)$ or $\Delta^3 F_{123}(i,j,k)$, where Δ^3 denotes a first difference with respect to three variables. (See Samuel S. Wilks (1962) Mathematical Statistics pp 39-41, 49-50 about differences.)

If one has $F(x_1,\ldots,x_m)\ g(x_{m+1},\ldots,x_n)$, then $\Delta^n(Fg) = (\Delta^m F)(\Delta^{n-m} g)$. Thus Lancaster's expression in terms of cumulative distributions is equivalent to ours. Our expression is the three variable first difference of Lancaster's expression. This would be so even if the distributions were not discrete.

Consider now the case of four variables W,X,Y,Z. Third order interactions exist if F cannot be written as the following sum:

$$F = F_1(W)\ F_{234}(X,Y,Z) + F_2(X)\ F_{134}(W,Y,Z) + F_3(Y)\ F_{124}(W,X,Z) +$$
$$F_4(Z)\ F_{123}(X,Y,Z) - F_1(W)\ F_2(X)\ F_{34}(YZ) - F_1(W)\ F_3(Y)\ F_{24}(X,Z) -$$
$$F_1(W)\ F_4(Z)\ F_{23}(X,Y) - F_2(X)\ F_3(Y)\ F_{14}(W,Z) - F_2(X)\ F_4(Z)\ F_{13}(W,Y) -$$
$$F_3(Y)\ F_4(Z)\ F_{12}(X,Y) + 3F_1(W)\ F_2(X)\ F_3(Y)\ F_4(Z)\ .$$

Similar to the case of three variables, let W, X, Y, Z be code random variables which take on the values 1 to q, 1 to r, 1 to s, 1 to t respectively. It is easy to see that third order interaction is present iff

$$p_{hijk} \neq p_{h...}\ p_{.ijk} + p_{.i..}\ p_{h.jk} + p_{..j.}\ p_{hi.k} + p_{...k}\ p_{hij.}$$
$$- p_{h...}\ p_{.i..}\ p_{..jk} - p_{h...}\ p_{..j.}\ p_{.i.k} - p_{h...}\ p_{...k}\ p_{.ij.} - p_{.i..}$$
$$p_{..j.}\ p_{h..k} - p_{.i..}\ p_{...k}\ p_{h.j.} - p_{..j.}\ p_{...k}\ p_{hi..} + 3p_{h...}\ p_{.i..}$$
$$p_{..j.}\ p_{...k}$$ for some h, i, j, k. (This is the four variable first difference of Lancaster's expression.) i.e. iff θ_{hijk} of our first extension of our 3-way model $\neq 0$ for some h, i, j, k. Hence for the

4-dimensional case, only our definition of the presence of 3rd order interaction in our first extension of our original model coincides with that of Lancaster's.

C. THE GENERAL CASE FOR THE FIRST EXTENSION OF OUR 3-WAY MODEL.

There are w partitions or classifications of the sample space. We divide the set $\{1, 2, \ldots, w\}$ into two disjoint parts, $\{r_1, r_2, \ldots, r_u\}$ and $\{s_1, s_2, \ldots, s_v\}$, with $u + v = w$. For definiteness we take $r_1 < r_2 < \ldots < r_u$ and $s_1 < s_2 < \ldots < s_v$. Let T_u denote the ordered u-tuple $(r_1, r_2, \ldots, r_u)$ and ψ_v the ordered v-tuple $(s_1, s_2, \ldots, s_v)$. Let $i_h = 1$ to m_{r_h} be the subscript for the r_h-th partition and $j_k = 1$ to m_{s_k} the subscript for the s_k-th partition. Let α_u denote the ordered u-tuple $(i_1, i_2, \ldots, i_u)$. Let σ_v denote the ordered v-tuple $(j_1, j_2, \ldots, j_v)$. We will write $p(r_h; i_h)$ to denote $p_{\underbrace{\ldots.}_{(r_{h-1})\text{dots}} i_h \underbrace{\ldots.}_{(w-r_h)\text{dots}}}$.

The v-factor interactions involving the partitions $s_1, s_2, \ldots, s_v$ can be written $A(\psi_v; \sigma_v)$ or sometimes $A(s_1 s_2, \ldots, s_v; j_1, j_2, \ldots j_v)$. A constraint on such a v-factor interaction is

$$\sum_{j_k=1}^{m_{s_k}} A(s_1, s_2, \ldots, s_v; j_1, j_2, \ldots, j_v) = 0 \quad \text{where } k = 1, 2, \ldots, \text{ or } v \text{ where}$$

$v \geq 2$.

The general case for the first extension of our 3-way model. can be written

$$E(f_{\ell_1\ell_2\ldots\ell_w}) = np_{\ell_1\ell_2\ldots\ell_w} = n\{\prod_{h=1}^{w} p(h,\ell_h) + \sum_{v=2}^{w} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})]\ A(\psi_v;\ \sigma_v)\}$$

where it is understood that $u = w-v$, and that $i_h = \ell_{r_h}$. The summation $\sum_{\psi_v}$ is a summation over all possible v-tuples with $1 \leq s_1 < s_2 < \ldots < s_v \leq w$. It is a v-fold summation, and there are $\binom{w}{v}$ terms altogether. (When $u = 0$, $\prod_{h=1}^{0} p(r_h;\ \ell_{r_h}) = 1$, since empty products are conventionally always set = 1.)

The terms of the sum for $p_{\ell_1\ell_2\ldots\ell_w}$ can be separated into two classes: (1) those in which the wth subscript appears in the $\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})$ part as the factor $p(w;\ \ell_w)$, and (2) those in which the wth subscript appears in the $A(\psi_v;\ \sigma_v)$ factor, which is then $A(s_1,\ldots,s_{v-1},\ w;\ \ell_{s_1},\ldots,\ell_{s_{v-1}},\ \ell_w)$. The two classes of terms are mutually exclusive and take in all possible cases. Now we sum the $p_{\ell_1\ell_2\ldots\ell_w}$ term by term with respect to ℓ_w (the other ℓ's held fixed). The terms of the first class have a factor $\sum_{\ell_w=1}^{m_w} p(w;\ \ell_w) = 1$ and those of the second class a factor

$\sum_{\ell_w=1}^{m_w} A(s_1,\ldots,s_{v-1},\ w;\ \ell_{s_1},\ldots,\ell_{s_{v-1}},\ \ell_w) = 0$. What is left is the marginal model for $p_{\ell_1 \ldots \ell_{w-1}}$. We could have done the same sort of thing for any other subscript and its partition. One can repeat the argument and get the marginal model for (w-2) partitions, etc. To avoid ambiguity with lots of dots instead of subscripts, one might write $p(1,2,..,w;\ell_1,\ell_2,..\ell_w)$ for $p_{\ell_1\ell_2.....\ell_w}$, etc. The first set of letters or numbers within the parenthesis indicates the partitions involved, the second set indicates the indices or categories considered of the corresponding partitions in the first set. E.g. Suppose there are 5 partitions. Then $p(1,2,4;\ h,i,k) = p_{hi.k.}$. Here the old notation is good enough, but, when there are, say u subscripts and v dots, things can become pretty confusing.

Using the information of the previous paragraph we can arrive at an expression for $A(1,2,..,w;\ \ell_1,\ell_2,..\ell_w)$ or (w-1)th order interaction. We prove the following by induction:

$$A(1,2,..,w;\ \ell_1,\ell_2,..,\ell_w) = p_{\ell_1\ell_2\ldots\ell_w} +$$

$$\sum_{u=1}^{w-2} (-1)^u \sum_{\psi_v} \left[\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})\right] p(\psi_v;\ \sigma_v) +$$

$$(-1)^{w-1} (w-1) \prod_{k=1}^{w} p(k;\ \ell_k).$$ Here

$p(\psi_v;\ \sigma_\sigma) = p(s_1,s_2,..,s_v\ ;\ \ell_{s_1},\ell_{s_2},\ldots,\ell_{s_v})$, and $u = 1$ to $(w-2)$ is equivalent to $v = (w-1)$ to 2 or, turned around, $v = 2$ to $(w-1)$, since

$u + v = w$.

Thus the (w-1)th order interaction of our first model is the same as that of Lancaster's model.

Proof:

Lancaster's model can be written as

$$p_{\ell_1\ell_2\ldots\ell_w} = \sum_{u=1}^{w-2} \sum_{\psi_v} (-1)^{u-1} \left[\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})\right] p(\psi_v;\ \sigma_v) + (-1)^w (w-1) \prod_{h=1}^{w} p(h;\ \ell_h)$$

$$+\ a(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)$$

$$a(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w) = p_{\ell_1\ell_2\ldots\ell_w} + \sum_{u=1}^{w-2} \sum_{\psi_v} (-1)^u \left[\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})\right] \times$$

$$p(\psi_v;\ \sigma_v) + (-1)^{w-1} (w-1) \prod_{h=1}^{w} p(h;\ \ell_h)$$

From our first extension of our three-way model

$$p_{\ell_1\ell_2\ldots\ell_w} = \prod_{h=1}^{w} p(h;\ \ell_h) + \sum_{u=1}^{w-2} \sum_{\psi_v} \left[\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})\right] A(\psi_v;\ \sigma_v)$$

$$+\ A(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)$$

$$A(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots\ell_w) = p_{\ell_1\ell_2\ldots\ell_w} - \prod_{h=1}^{w} p(h;\ \ell_h) - \sum_{u=1}^{w-2} \sum_{\psi_v} \left[\prod_{h=1}^{u} p(r_h;\ell r_h)\right] \times$$

$$A(\psi_v,\sigma_v)$$

It has been shown that the two models are equivalent for $w = 2, 3$. To show by induction that they are equivalent for any whole number $w \geq 2$, suppose our $A(\psi_w;\sigma_w)$ equals Lancaster's $a(\psi_w;\sigma_w)$ for $w = 2, 3, \ldots, k$. We now want to show that our $A(1,2,\ldots, k+1, \ell_1,\ell_2,\ldots,\ell_{k+1})$ = Lancaster's $a(1,2,\ldots, k+1; \ell_1,\ell_2\ldots\ell_{k+1})$.

$$\text{Our } A(1,2,\ldots, k+1; \ell_1,\ell_2,\ldots\ell_{k+1}) = p_{\ell_1\ell_2,\ldots\ell_{k+1}}$$

$$- \prod_{h=1}^{k+1} p(h; \ell_h) - \sum_{u=1}^{k-1} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h; \ell_{r_h})] \{p(\psi_v; \sigma_v) + \sum_{m=1}^{v-2} \sum_{\psi_n} (-1)^m \times$$

$$[\prod_{h=1}^{m} p(r_h; \ell_{r_h})]\, p(\psi_n; \sigma_n) + (-1)^{v-1} (v-1) \prod_{h=1}^{v} p(h; \ell_{s_h})\} \quad (*)$$

$$\text{Here} \quad u + v = w = k + 1 \qquad \sum_{u=1}^{w-2} \iff \sum_{v=2}^{w-1} \quad \text{and} \quad \sum_{u=1}^{k-1} \iff \sum_{v=2}^{k}$$

$$m + n = v \qquad \sum_{m=1}^{v-2} \iff \sum_{n=2}^{v-1}$$

The last term on the right of equation (*) is

$$\sum_{u=1}^{k-1} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h; \ell_{r_h})] (-1)^v (v-1) \prod_{h=1}^{v} p(h; \ell_{s_h})$$

$$= \sum_{v=2}^{k} \sum_{\psi_v} (-1)^v (v-1) [\prod_{h=1}^{k+1} p(h; \ell_h)] = [\prod_{h=1}^{k+1} p(h; \ell_h)] \sum_{v=2}^{k} (-1)^v (v-1) \sum_{\psi_v} 1$$

$$\text{But } \sum_{\psi_v} 1 = \binom{w}{v} = \binom{k+1}{v} = \frac{(k+1)!}{v!\,(k-v+1)!}$$

$$(v-1) \sum_{\psi_v} 1 = \frac{(k+1)!}{(v-1)!\,(k-v+1)!} - \frac{(k+1)!}{v!\,(k-v+1)!} = (k+1) \binom{k}{v-1} - \binom{k+1}{v}$$

$$\sum_{v=2}^{k} (-1)^v (v-1) \sum_{\psi_v} 1 = -(k+1) \sum_{v=2}^{k} (-1)^{v-1} \binom{k}{v-1} - \sum_{v=2}^{k} (-1)^v \binom{k+1}{v}$$

$$= -\{(k+1) \sum_{u=1}^{k-1} (-1)^u \binom{k}{u} + \sum_{v=2}^{k} (-1)^v \binom{k+1}{v}\}$$

Now $(1-1)^s = \sum_{r=0}^{s} (-1)^r \binom{s}{r} = 0$ for s a whole number.

Thus $\sum_{u=1}^{k-1} (-1)^u \binom{k}{u} = -\{\binom{k}{0} + (-1)^k \binom{k}{k}\} = \begin{cases} 0 & \text{if } k \text{ is odd} \\ -2 & \text{if } k \text{ is even} \end{cases}$,

and $\sum_{v=2}^{k} (-1)^v \binom{k+1}{v} = -\{\binom{k+1}{0} - \binom{k+1}{1} + (-1)^{k+1} \binom{k+1}{k+1}\} = k + (-1)^k$

$$= \begin{cases} (k-1) & \text{if } k \text{ is odd} \\ (k+1) & \text{if } k \text{ is even} \end{cases}.$$

$$\sum_{v=2}^{k} (-1)^v (v-1) \sum_{\psi_v} 1 = \begin{cases} -\{(k+1)\,0 + (k-1)\} = -(k-1) & \text{if } k \text{ is odd} \\ -\{(k+1)\,(-2) + (k+1)\} = (k+1) & \text{if } k \text{ is even} \end{cases}$$

$$= (-1)^k k + 1 .$$

Thus $\sum_{u=1}^{k-1} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})]\ (-1)^v (v-1) \prod_{h=1}^{v} p(h,\ \ell_{s_h})$

$$= \{(-1)^k\ k+1\} \prod_{h=1}^{k+1} p(h;\ \ell_h)\ .$$

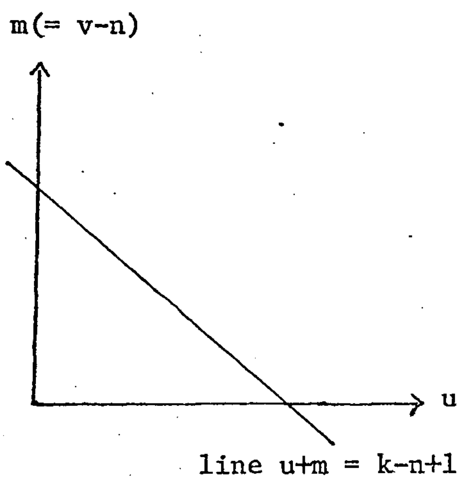
The next to last term on the right of equation (*) is

$$- \sum_{u=1}^{k-1} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})] \sum_{m=1}^{v-2} \sum_{\psi_n} (-1)^m [\prod_{h=1}^{m} p(r_h;\ \ell_{r_h})]\ p(\psi_n;\ \sigma_n). \qquad (**)$$

First the $w = k + 1$ letters are separated into two disjoint sets $T_u = \{r_1, r_2, \ldots, r_u\}$ and $\psi_v = \{s_1, s_2, \ldots, s_v\}$, then the latter is separated into two disjoint sets T_m and ψ_n. Let $T_{u+m} = T_u \cup T_n$. It is the complement of ψ_n. Let us look at the terms of the sum above involving a particular ψ_n. Here n is some whole number such that $2 \leq n \leq v-1$. The partial sum involving a particular ψ_n consists of several terms, each of the form

$$(-1)^{m-1} [\prod_{h=1}^{u+m} p(r_h;\ \ell_{r_h})]\ p(\psi_n;\ \sigma_n) \quad (u+m = w-n = k-n+1).$$

The problem is to find how many such terms there are.



u	m	v = k-u+1
1	k-n	k
2	k-n-1	k-1
3	k-n-2	k-2
.	.	.
.	.	.
.	.	.
.	.	.
k-n-1	2	n+2
k-n	1	n+1

Out of the $k-n+1$ letters not in ψ_n, there are $\binom{k-n+1}{u} = \binom{k-n+1}{m}$ ways to assign u letters to T_u and m letters to T_m.

The sum of the coefficients of terms involving

$[\prod_{h=1}^{k-n+1} p(r_h;\ \ell_{r_h})]\ p(\psi_n;\ \sigma_n)$, including the factor $(-1)^{m-1}$, is

$$\sum_{m=1}^{k-n} (-1)^{m-1} \binom{k-n+1}{m} = \binom{k-n+1}{0} + (-1)^{k-n+1}\binom{k-n+1}{k-n+1} = 1 + (-1)^{k-n+1}$$

$$= \begin{cases} 0 & \text{if } (k-n+1) \text{ is odd, } (k-n) \text{ is even} \\ 2 & \text{if } (k-n+1) \text{ is even, } (k-n) \text{ is odd} \end{cases} .$$

Hence the sum (**) is

$\sum_{n=2}^{k-1} \sum_{\psi_n} \{1 + (-1)^{k-n+1}\} [\prod_{h=1}^{k-n+1} p(r_h;\ \ell_{r_h})]\ p(\psi_n;\ \sigma_n)$. One can replace the dummy variable n by v.

The middle term (third term) on the right of equation (*) is

$$-\sum_{u=1}^{k-1} \sum_{\psi_v} [\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})]\ p(\psi_v;\ \sigma_v) = -\sum_{v=2}^{k} \sum_{\psi_v} [\prod_{h=1}^{k-v+1} p(r_h;\ \ell_{r_h})]\ p(\psi_v;\ \sigma_v).$$

Now we substitute the above results into (*). Our

$$A(1,2,\ldots,\ k+1;\ \ell_1,\ell_2,\ldots,\ell_{k+1}) = p_{\ell_1\ell_2\ldots\ell_{k+1}} - \prod_{h=1}^{k+1} p(h;\ \ell_h)$$

$$-\sum_{v=2}^{k} \sum_{\psi_v} [\prod_{h=1}^{k-v+1} p(r_h;\ \ell_{r_h})]\ p(\psi_v;\ \sigma_v) +$$

$$\sum_{v=2}^{k-1} \sum_{\psi_v} \{1 + (-1)^{k-v+1}\} \left[\prod_{h=1}^{k-v+1} p(r_h;\ \ell_{r_h})\right] p(\psi_v;\ \sigma_v) + \{1+(-1)^k\ k\} \times \prod_{h=1}^{k+1} p(h;\ \ell_h).$$

Our $A(1,2,\ldots,\ k+1;\ \ell_1,\ell_2,\ldots,\ell_{k+1}) = p_{\ell_1\ell_2\ldots\ell_{k+1}} + (-1)^k\ k \prod_{h=1}^{k+1} p(h,\ell_k)$

$$+ \sum_{v=2}^{k} \sum_{\psi_v} (-1)^{k-v+1} \left[\prod_{h=1}^{k-v+1} p(r_h;\ \ell_{r_h})\right] p(\psi_v;\ \sigma_v)$$

$$= p_{\ell_1\ell_2\ldots\ell_{k+1}} + \sum_{u=1}^{k-1} \sum_{\psi_v} (-1)^u \left[\prod_{h=1}^{u} p(r_h;\ \ell_{r_h})\right] p(\psi_v;\ \sigma_v)$$

$$+ (-1)^k\ k \prod_{h=1}^{k+1} p(h;\ \ell_h)$$

= Lancaster's $a(1,2,\ldots,\ k+1;\ \ell_1,\ell_2,\ldots,\ell_{k+1})$ with $w = k + 1$.

Thus the induction is complete.

In testing the hypothesis $A(1,2,\ldots,w;\ \ell_1,\ \ell_2,\ldots,\ell_w) = 0$ for all $(\ell_1,\ell_2,\ldots,\ell_w)$, using a similar method as was used to test the hypothesis $\delta_{ijk} = 0$ for all (i,j,k), our first concern is whether any problems arise in determining the rank of $\left[\frac{\partial\ g_H\ (\underline{p})}{\partial p(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots\ell_w)}\right]$

where H ranges through $0,\ 111\ldots1,\ 111\ldots2,\ldots,\ m_1m_2\times\ldots\times m_w$ and

$$g_0\ (\underline{p}) = \sum_{\ell_1=1}^{m_1} \cdots\cdots \sum_{\ell_w=1}^{m_w} p(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w) - 1$$ and $g_{\ell_1\ell_2\ldots\ell_w}\ (\underline{p})$

$= A(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)$. Let $\Delta_{ij} = \Delta(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$.

$$\frac{\partial g_{\ell_1\ell_2\ldots\ell_w}}{\partial p(1,2,\ldots,w;\ a_1,a_2\ldots,a_w)} = \prod_{k=1}^{w} \Delta(\ell_k,a_k) + \sum_{u=1}^{w-2} (-1)^u \sum_{\psi_v}$$

$$\{ \prod_{h=1}^{u} p(r_h;\ \ell_{r_h}) \prod_{k=1}^{v} \Delta(\ell_{s_k},\ a_{s_k}) + \sum_{y=1}^{u} \Delta(\ell_{r_y},a_{r_y}) [\prod_{\substack{h=1\\h\neq y}}^{u} p(r_h;\ell_{r_h})]\ p(\psi_v;\ \sigma_v)\}$$

$$- (-1)^w\ (w-1) \sum_{z=1}^{w} \Delta(\ell_z,a_z)\ [\prod_{\substack{k=1\\k\neq z}}^{w} p(k,\ell_k)]$$

We multiply the matrix $[\frac{\partial g_H\ (\psi)}{\partial p(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)}]$ on the left by $\left[\begin{matrix} 1 & \underline{0}' \\ \underline{0} & X_1 \otimes X_2 \otimes \cdots \otimes X_w \end{matrix}\right]$ where $\{X_i^{(u_i)}\}$, $u_i = 0,\ldots,(m_i-1)$, denote a set of functions orthonormal with respect to $\{p(i;1),\ p(i;2),\ldots\ p(i,m_i)\}$ and $X_i = [X_{i,j}^{(u)}]$ $j = 1$ to m_i.

Then $$\sum_{\ell_1=1}^{m_1} \sum_{\ell_2=1}^{m_2} \cdots\cdots \sum_{\ell_w=1}^{m_w} X_{1\ell_1}^{(u_1)} X_{2\ell_2}^{(u_2)} \cdots X_{w\ell_w}^{(u_w)} \frac{\partial g_{\ell_1\ell_2\ldots\ell_w}}{\partial p(1,2,\ldots,w;\ a_1,a_2,\ldots,a_w)}$$

$$= X_{1a_1}^{(u_1)} X_{2a_2}^{(u_2)} \cdots X_{wa_w}^{(u_w)} + \sum_{u=1}^{w-2} (-1)^u \sum_{\psi_v} \{ \prod_{h=1}^{u} [\sum_{\ell_{r_h}=1}^{m_{r_h}} X(u_{r_h},\ell_{r_h}) \times$$

$$p(r_h;\ \ell_{r_h})]\ \prod_{k=1}^{v} X(u_{s_k}, a_{s_k})$$

$$+ \sum_{y=1}^{u} X(u_{r_y}, a_{r_y}) \prod_{\substack{h=1\\h\neq y}}^{u} [\sum_{\ell_{r_h}=1}^{m_{r_h}} X(u_{r_h}, \ell_{r_h})\ p(r_h;\ \ell_{r_h})] \ \times$$

$$\sum_{\ell_{s_1}=1}^{m_{s_1}} \sum_{\ell_{s_2}=1}^{m_{s_2}} \sum_{\ell_{s_3}=1}^{m_{s_3}} \cdots \sum_{\ell_{s_v}=1}^{m_{s_v}} [\prod_{k=1}^{v} X(u_{s_k}, \ell_{s_k})]\ p(\psi_v;\ \sigma_v)\}$$

$$- (-1)^w\ (w-1) \sum_{z=1}^{w} X(u_z, a_z) \prod_{\substack{k=1\\k\neq z}}^{w} [\sum_{\ell_k=1}^{m_k} X(u_k;\ \ell_k)\ p(k;\ \ell_k)]$$

where $X(a,b)$ stands for X with a as superscript and b as subscript.

Thus $$\sum_{\ell_1=1}^{m_1} \sum_{\ell_2=1}^{m_2} \cdots \sum_{\ell_w=1}^{m_w} X_{1\ell_1}^{(u_1)} X_{2\ell_2}^{(u_2)} \cdots X_{w\ell_w}^{(u_w)} \frac{\partial g_{\ell_1\ell_2\ldots\ell_w}}{\partial p(1,2,\ldots,w;\ a_1,a_2,\ldots,a_w)}$$

$$= \prod_{k=1}^{w} X_{ka_k}^{(u_k)} + \sum_{u=1}^{w-2} (-1)^u \sum_{\psi_v} \{[\prod_{h=1}^{u} \Delta(u_{r_h}, 0)] \prod_{k=1}^{v} X(u_{s_k}, a_{s_k})$$

$$+ \sum_{y=1}^{u} X(u_{r_y}, a_{r_y}) \ [\prod_{\substack{h=1\\h\neq y}}^{u} \Delta(u_{r_h}, 0)]\ \rho\ (s_1,s_2,\ldots,s_v;\ u_{s_1},u_{s_2},\ldots,u_{s_v})\}$$

$$- (-1)^w\ (w-1) \sum_{z=1}^{w} X(u_z, a_z)\ [\prod_{\substack{k=1\\k\neq z}}^{w} \Delta(u_k, 0)],\quad \text{since}$$

$$\sum_{\ell_k=1}^{m_k} X(u_k, \ell_k)\, p(k,\ell_k) = \Delta(u_k,0) = \begin{cases} 1 & \text{if } u_k = 0 \\ 0 & \text{if } u_k \neq 0 \end{cases} .$$

$\rho(s_1,s_2,\ldots,s_v;\ u_{s_1},u_{s_2},\ldots,u_{s_v})$ is the v-factor correlation with subscript u_{s_k} at the s_k-th subscript position. E.g. When $w = 3$,

$\rho_{012} = \rho(2,3;\ 1,2) = \sum_{j=1}^{s} \sum_{k=1}^{t} Y_j^{(1)}\ Z_k^{(2)}\ p_{.jk}$. By examining the above equation carefully it becomes evident that the rank of the submatrix of

$$\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & X_1 \otimes X_2 \otimes \ldots \otimes X_w \end{bmatrix} \left[\frac{\partial g_H\ (p)}{\partial p(1,2,..,w;\ \ell_1\ell_2\ldots\ell_w)} \right]$$

consisting of the first row and the rows where one or more of $u_1,..,u_w$ is zero, is one. That the submatrix consisting of the remaining rows has rank $(m_1-1)(m_2-1)\ldots(m_w-1)$ follows as a direct extension for the proof where $w = 3$, which we have already done.

Consider $Q = \log(n!) - \sum_{\ell_1=1}^{m_1} \sum_{\ell_2=1}^{m_2} \cdots \sum_{\ell_w=1}^{m_w} \log\ (f_{\ell_1\ldots\ell_w}!)$

$$+ \sum_{\ell_1=1}^{m_1} \sum_{\ell_2=1}^{m_2} \cdots\cdot \sum_{\ell_w=1}^{m_w} f_{\ell_1\ldots\ell_w} \log p_{\ell_1\ldots\ell_w} - \{\theta[\sum_{\ell_1=1}^{m_1} \cdots \sum_{\ell_w=1}^{m_w} p_{\ell_1\ldots\ell_w} - 1]$$

$$+ \sum_{\ell_1=1}^{m_1} \cdots\cdot \sum_{\ell_w=1}^{m_w} \eta_{\ell_1\ldots\ell_w}\, g_{\ell_1\ldots\ell_w} \}$$

$$\frac{\partial Q}{\partial p_{a_1\ldots a_w}} = \frac{f_{a_1\ldots\cdot a_w}}{p_{a_1\ldots\cdot a_w}} - \{ \theta + \sum_{\ell_1=1}^{m_1} \cdots\cdot \sum_{\ell_w=1}^{m_w} \eta_{\ell_1\ldots\ell_w} \frac{\partial g_{\ell_1\ldots\ell_w}}{\partial p_{a_1a_2\ldots a_w}} \}$$

$a_i = 1$ to m_i ; $i = 1$ to w.

The rank of the homogeneneous part of these $m_1 m_2 \ldots m_w$ equations is $(m_1-1)(m_2-1)\ldots(m_w-1) + 1$. If we multiplied the above equations by hypothetical $p_{a_1\ldots a_w}$ and then summed with respect to $a_1,\ldots,a_w$ we would obtain an equation of the following form:

$$\sum_{a_1=1}^{m_1} \cdots\cdots \sum_{a_w=1}^{m_w} p_{a_1\ldots a_w} \frac{\partial Q}{\partial p_{a_1\ldots a_w}} = n - \theta + \sum_{a_1=1}^{m_1} \cdots \sum_{a_w=1}^{m_w} \eta_{a_1\ldots a_w} h_{a_1\ldots a_w}$$

from which we could solve for θ in terms of the η's.

The equations $\dfrac{\partial Q}{\partial p_{a_1\ldots a_w}} = 0$, $a_i = 1$ to m_i, $i = 1,\ldots,w$

represent $m_1\ldots m_w$ equations in $m_1\ldots m_w + 1$ unknowns, the η's and θ. We found that there are $(m_1-1)\ldots(m_w-1) + 1$ linearly independent equations among these $m_1\ldots m_w$ equations. After we eliminate θ we have $(m_1-1)\ldots(m_w-1)$ linearly independent equations. This follows by similar reasoning to that used in the case of three variables.

The term of $p_{a_1\ldots a_w} \dfrac{\partial Q}{\partial p_{a_1\ldots a_w}} = 0$ not involving the

$\eta_{a_1\ldots a_w}$ after θ has been eliminated is $(f_{a_1\ldots a_w} - np_{a_1\ldots a_w})$. In order for the augmented matrix to have rank $(m_1-1)\ldots(m_w-1)$ it is necessary that $\sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} \cdots \sum_{a_w=1}^{m_w} X_{1a_1}^{(u_1)} \cdots X_{wa_w}^{(u_1)} (f_{a_1 \ldots a_w} - np_{a_1 \ldots a_w})$

$= 0$ whenever one or more of the indices $u_1, \ldots, u_w$ are zero. In this case let $\hat{X}_1^{(u_1)}, \hat{X}_2^{(u_2)}, \ldots, \hat{X}_w^{(u_w)}$ represent orthonormal functions with respect to $\{\frac{f_{\ell_1 \ldots}}{n}\}, \{\frac{f_{.\ell_2 \ldots}}{n}\}, \ldots, \{\frac{f_{\ldots \ell_w}}{n}\}$ respectively.

Write $\hat{q}_{a_1 \ldots a_w} = \sum_{u=1}^{w-2} \frac{(-1)^{u+1}}{n^{u+1}} \sum_{\psi_v} [\prod_{h=1}^{u} f(r_h, \ell_{r_h})] f(\psi_v; \sigma_v) +$

$\frac{(-1)^w}{n^w} (w-1) [\prod_{h=1}^{u} f(r_h; \ell_{r_h})]$, where $f(r_h, \ell_{r_h}) = f\underbrace{.. \ \ ..}_{(r_h-1)\text{dots}} \ell_{r_h} \underbrace{.. \ \ ..}_{(w-r_h)\text{dots}}$

and $f(\psi_v; \sigma_v) = f(s_1, s_2, \ldots, s_v; j_1, j_2, \ldots, j_v)$

$$\sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} \cdots \sum_{a_w=1}^{m_w} \hat{X}_{1a_1} \cdots \hat{X}_{wa_w}^{(u_w)} q_{a_1 \ldots a_w} = \sum_{u=1}^{w-2} (-1)^{u+1} \sum_{\psi v} \prod_{h=1}^{u} \Delta(u_{r_h}, 0) \times$$

$$\rho(s_1, s_2, \ldots, s_v; u_{s_1}, u_{s_2}, \ldots, u_{s_v}) + (-1)^w (w-1) \prod_{h=1}^{w} \Delta(u_h, 0) .$$

Thus

$$\sum_{a_1=1}^{m_1} \cdots \sum_{a_w=1}^{m_w} \hat{X}_{1a_1}^{(u_1)} \cdots \hat{X}_{wa_w}^{(u_w)} \hat{q}_{a_1 \ldots a_w} = \begin{cases} 1 \text{ for } u_i = 0, i = 1 \text{ to } w & (1) \\ 0 \text{ for any } (w-1) \text{ of the } u_i = 0 & (2) \\ \hat{\rho}(s_{t_1}, \ldots, s_{t_p}; u_{t_1}, u_{t_2}, \ldots, u_{t_p}) & \\ \text{for } u_{t_1}, \ldots, u_{t_p} \neq 0,\ p \geq 2 & (3) \\ \hat{\rho}(1, 2, \ldots, i-1, i+1, \ldots, w; u_1, u_2, \ldots, u_{i-1}, & \\ \qquad u_{i+1}, \ldots, u_w) & \\ u_i = 0, \text{ all others nonzero} & (4) \\ 0, \text{ all } u_i = 0 & (5) \end{cases}$$

(1) follows from the equality $\sum_{k=1}^{w} (-1)^{k+1} \binom{w}{k} = 1$.

$$\left[\sum_{k=1}^{w} (-1)^{k+1} \binom{w}{k} = - \sum_{k=0}^{w} (-1)^{k} \binom{w}{k} X^{k} \Big|_{X=1} + 1 = -((1-X)^{w}-1) \Big|_{X=1} = -(-1) = 1 \right]$$

(2), (3), (4), (5) follow by close examination of above equation.

Similarly for

$$\sum_{a_1=1}^{m_1} \cdots \sum_{a_w=1}^{m_w} X_{1a_1}^{(u_1)} \cdots X_{wa_w}^{(u_w)} \; p_{a_1 \ldots a_w}, \quad \text{where the} \quad p_{a_1 \ldots a_w}$$

represent their hypothetical values, except that ρ is without ^ .

$$\sum_{a_1=1}^{m_1} \cdots \cdots \sum_{a_w=1}^{m_w} \hat{X}_{1a_1}^{(u_1)} \cdots \hat{X}_{wa_w}^{(u_w)} \frac{f_{a_1 \ldots a_w}}{n} = \hat{\rho}_{u_1 \ldots u_w} =$$

$$\begin{cases} 1 \text{ for } u_i=0, i=1 \text{ to } w \\ 0 \text{ for any } (w-1) \text{ of the } u_i=0 \\ \hat{\rho}\,(s_{t_1}, s_{t_2}, \ldots, s_{t_p};\ u_{t_1}, u_{t_2}, \ldots, u_{t_p}) \text{ for } u_{t_1}, u_{t_2}, \ldots, u_{t_p} \neq 0, p \geq 2 \\ \hat{\rho}(1,2,..,i-1,i+1,..,w;\ u_1,u_2,..u_{i-1}u_{i+1},\ldots,u_w), u_i=0, \text{all others nonzero} \\ \hat{\rho}(1,2,..,w;\ u_1,u_2,..,u_w) \text{ all } u_i \neq 0. \end{cases}$$

Thus $$\sum_{a_1=1}^{m_1} \cdots \sum_{a_w=1}^{m_w} \hat{X}_{1a_1}^{(u_1)} \cdots \hat{X}_{wa_w}^{(u_w)} (f_{a_1 \ldots a_w} - n\, \hat{p}_{a_1 \ldots a_w}) =$$

$$\begin{cases} 0 \text{ when at least one of the } u_i=0 \\ n\, \hat{\rho}(1,2,..,w;\ u_1,u_2,\ldots,u_w) \text{ when all indices are not zero's .} \end{cases}$$

On the other hand $\sum_{a_1=1}^{m_1} \ldots \sum_{a_w=1}^{m_w} X_{1a_1}^{(u_1)} \ldots X_{wa_w}^{(u_w)} \dfrac{f_{a_1 \ldots a_w}}{n} = \tilde{\rho}_{u_1 \ldots u_w}$,

which usually $\neq \rho_{u_1, \ldots u_w}$ when $p(1,a_1) \neq \dfrac{f(1,a_1)}{n}$, etc. Also

$\tilde{\rho}(1;u_1)$, $\tilde{\rho}(2,u_2), \ldots, \tilde{\rho}(w,u_w)$ cannot be expected to be zero

when $p(1;a_1) \neq \dfrac{f(1,a_1)}{n}$ etc. Thus $\sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} \ldots \sum_{a_w=1}^{m_w} X_{1a_1}^{(a_1)} \ldots X_{wa_w}^{(a_w)} \times$

$(f_{a_1 \ldots a_w} - np_{a_1 \ldots a_w}) = n(\tilde{\rho}_{u_1 \ldots u_w} - \rho_{u_1 \ldots u_w})$. For $u_i=0$, $i=1$ to w one has

$n(\tilde{\rho}_{00 \ldots 0} - \rho_{00 \ldots 0}) = 0$. In the other cases, one has $n(\tilde{\rho}_{u_1 \ldots u_w} - \rho_{u_1 \ldots u_w})$.

Consider the equations $p_{a_1 \ldots a_w} \dfrac{\partial Q}{\partial p_{a_1 \ldots a_w}} = 0$ after θ has been replaced by its expression in the η's. Suppose now that these equations are multiplied by $X_{1a_1}^{(a_1)} \ldots X_{wa_w}^{(u_w)}$ and then summed with respect to $a_1, \ldots, a_w$. We conclude that, in general, the resulting $m_1 \ldots m_w$ equations will have on the right side $n(\tilde{\rho}_{u_1 \ldots u_w} - \rho_{u_1 \ldots u_w})$, of which $(m_1 \ldots m_w - 1)$ may be expected to be different from zero depending on the values the u's take on. However, the rank of the matrix of coefficients of the $\eta_{a_1 \ldots a_w}$

on the left side is $(m_1-1)\ldots(m_w-1)$ so that the equations will be inconsistent except for special sets $\{p(1,a_1)\}$, $\{p(2,a_2)\},\ldots,\{p(w,a_w)\}$ which leave an appropriate set of not more than $(m_1-1)(m_2-1)\ldots(m_w-1)$ terms on the right of the equations non-zero. We want to verify that

$$\{p(1,a_1)\} = \{\frac{f(1,a_1)}{n}\},\ \{p(2,a_2)\} = \{\frac{f(2,a_2)}{n}\},\ldots,\ \{p(w,a_w)\} = \{\frac{f(w;a_w)}{n}\}$$

will accomplish this.

We have shown that the vector $\hat{\underline{q}}$ lies in the column space of the vectors $\underline{X}_1^{(u_1)}\otimes\ldots\otimes\underline{X}_w^{(u_w)}$, where $u_i \neq 0$; $i = 1$ to w, but that, in general the same can not be said of the other possible $\underline{p}$. If we can show that the column space of $[\frac{\partial g_{\ell_1\ell_2..\ell_w}}{\partial p_{a_1a_2..a_w}} + h_{\ell_1\ell_2\ldots\ell_w}]$ equals the column

space of $\{\underline{X}_1^{(u_1)} \times \underline{X}_2^{(u_2)} \otimes \ldots \otimes \underline{X}_w^{(u_w)} \mid u_i \neq 0;\ i = 1 \text{ to } w\}$ then the set of equations will be consistent when $\underline{p} = \hat{\underline{q}}$. Note that here $(a_1,..,a_w)$ indexes rows, $(\ell_1,..,\ell_w)$ and $(u_1,..,u_w)$ columns.

The column space of the coefficient matrix above is a subspace of $m_1m_2\ldots.m_w$ - dimensional Euclidean space. The matrix was shown to have rank $(m_1-1)\ (m_2-1)\ldots(m_w-1)$, hence its column space is an $(m_1-1)(m_2-1)\ldots(m_w-1)$ subspace of $E_{m_1m_2\ldots m_w}$.

Now the set of vectors $\{\underline{X}_1^{(u_1)} \otimes \underline{X}_2^{(u_2)} \otimes \ldots \otimes \underline{X}_w^{(u_w)} \left| \begin{array}{l} u_1=0 \text{ to } (m-1) \\ \vdots \\ u_w=0 \text{ to } (m_w-1) \end{array} \right.\}$

forms a basis for $E_{m_1 m_2 \cdots m_w}$. We will show that the column space of the coefficient matrix is orthogonal to the subspace spanned by the set of vectors $\{\underline{X}_1^{(u_1)} \otimes \underline{X}_2^{(u_2)} \otimes \ldots \otimes \underline{X}_w^{(u_w)} \mid$ one or more of the u_i is zero$\}$. This is a $m_1 m_2 \ldots m_w - (m_1-1)(m_2-1) \ldots (m_w-1)$ dimensional subspace. Hence its complement, the $(m_1-1)(m_2-1)\ldots(m_w-1)$ dimensional subspace spanned by the set of vectors $\{\underline{X}_1^{(u_1)} \otimes \underline{X}_2^{(u_2)} \otimes \ldots \otimes \underline{X}_w^{(u_w)} \mid u_i \neq 0;\ i=1 \text{ to } w\}$ is the same as the column space of the coefficient matrix. We have shown that $(\underline{f}-n\hat{\underline{q}})$ lies in this subspace. Hence the set of equations are consistent when $(\underline{f}-n\hat{\underline{q}})$ is the non-homogeneous part of the equations.

That the column space of the coefficient matrix is orthogonal to the subspace spanned by the set of vectors $\{\underline{X}_1^{(u_1)} \otimes \underline{X}_2^{(u_2)} \otimes \ldots \otimes \underline{X}_w^{(u_w)}$ $\mid$ one or more of $u_1, u_2, \ldots, u_w$ is zero $\}$ follows by direct extension from the proof for $w=3$ and so is left to the reader.

D. THE GENERAL CASE FOR THE SECOND EXTENSION OF OUR 3-WAY MODEL.

The general case for the second extension of our 3-way model can be written

$$E(f_{\ell_1 \ell_2 \ldots \ell_w}) = np_{\ell_1 \ell_2 \ldots \ell_w} = n \prod_{i=1}^{w} p(i,\ell_i) + n \sum_{v=2}^{w} \sum_{\psi_v} p(T_u;\alpha_u)\ B(\psi_v;\sigma_v)$$

where it is understood that $u = w-v$. (When $u = 0$, $p(T_u;\alpha_u) = 1$).

As before the terms of the sum $p_{\ell_1 \ell_2 \ldots \ell_w}$ can be separated into two classes: (1) those in which the wth subscript appears in the

$p(T_u;\alpha_u)$ part, and (2) those in which the wth subscript appears in the $B(\psi_v;\sigma_v)$ factor. The two classes of terms are mutually exclusive and take in all possible cases. Now we sum the $p_{\ell_1\ell_2\ldots\ell_w}$ term by term with respect to ℓ_w (the other ℓ's held fixed). The terms of the first class have a factor $\sum_{\ell_w=1}^{m_w} p(T_u;\Omega_u)$ and those of the second class have

a factor $\sum_{\ell_w=1}^{m_w} B(s_1,\ldots,s_{w-1},w;\ \ell_{s_1},\ldots,\ell_{s_{v-1}},\ell_w) = 0$. What is left is

the marginal model for $p_{\ell_1\ldots\ell_{w-1}}$. We could have done the same sort of thing for any other subscript and its partition. One can repeat the argument and get the marginal model for (w-2) partitions, etc.

We would like now to establish that the sum of the coefficients of the terms of $B(1,2,..,w;\ \ell_1,\ell_2,\ldots,\ell_w)$ is zero. For the 3-way case we got

$$\delta_{ijk} = p_{ijk} - p_{i..}\, p_{.jk} - p_{.j.}\, p_{i.k} - p_{..k}\, p_{ij.} + 2p_{i..}\, p_{.j.}\, p_{..k}\ .$$

Note that the sum of the coefficients is zero. For the four variable case, recall we obtained

$$\begin{aligned}\theta'_{hijk} = p_{hijk} &- p_{h...}p_{.ijk} - p_{.i..}\, p_{h.jk} - p_{..j.}\, p_{hi.k} - p_{...k}\, p_{hij.} \\ &- 2p_{hi..}\, p_{..jk} - 2p_{h.j.}\, p_{.i.k} - 2p_{.ij.}\, p_{h..k} \\ &+ 3p_{hi..}\, p_{..j.}\, p_{...k} + 3p_{h.j.}\, p_{.i..}\, p_{...k} + 3p_{h..k}\, p_{.i..}\, p_{..j.} \\ &+ 3p_{.ij.}\, p_{h...}p_{...k} + 3p_{.i.k}\, p_{h...}\, p_{..j.} + 3p_{..jk}\, p_{h...}\, p_{.i..} \\ &- 9p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k}\end{aligned}$$

Again, note the sum of the coefficients is zero. Let us examine the 4-way case somewhat more carefully by writing the terms of θ'_{hijk} as follows:

$$\theta'_{hijk} = p_{hijk} - p_{h...}\, p_{.i..}\, p_{..j.}\, p_{...k} - p_{hi..}\, \alpha'^{(34)}_{jk} - p_{h.j.}\, \alpha'^{(24)}_{ik}$$

$$- p_{h..k}\, \alpha'^{(23)}_{ij} - p_{.ij.}\, \alpha'^{(14)}_{hk} - p_{.i.k}\, \alpha'^{(13)}_{hj} - p_{..jk}\, \alpha'^{(12)}_{hi} - p_{h...}\, \delta'^{(234)}_{ijk}$$

$$- p_{.i..}\, \delta'^{(134)}_{hjk} - p_{..j.}\, \delta'^{(124)}_{hik} - p_{...k}\, \delta'^{(123)}_{hij} .$$

Note that the coefficients of the first two terms cancel out. Since each α consists of two terms, one with positive 1 coefficient and one with negative 1 coefficient, the coefficients of all the terms involving an α cancel out. Since the sum of the coefficients of the terms that make up any δ is zero the coefficients of the terms involving the δ's cancel out. Hence the sum of the coefficients of the terms of θ'_{hijk} is zero.

We prove by induction that the result holds true for $B(1,2,..,w;\ \ell_1,\ell_2,..,\ell_w)$. Suppose now the result holds true for

$$B(s_1,s_2,..,s_k;\ \ell_{s_1}, \ell_{s_2},....\ell_{s_k}) ,...,\ B(s_1,s_2;\ \ell_{s_1},\ell_{s_2}) .$$

Consider $B(1,2,...,k,\ k+1;\ \ell_1,\ell_2,...,\ell_k,\ell_{k+1}) = p_{\ell_1\ell_2...\ell_{k+1}}$

$$- \prod_{i=1}^{k+1} p(i,\ell_i) - \sum_{v=2}^{k} \sum_{\psi_v} p(T_u,\Omega_u)\, B(\psi_v;\sigma_v) .$$ The coefficients of the first two terms cancel out. By our inductive hypothesis the coefficients of the remaining terms cancel out. Hence the sum of the coefficients

of the terms of $B(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)$ is zero.

In testing the hypothesis $B(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w) = 0$ for all $(\ell_1,\ell_2,\ldots,\ell_w)$ using a similar method as was used to test the hypothesis $\delta_{ijk} = 0$ for all (i,j,k), again our first concern is whether any problems arise in determining the rank of $\left[\dfrac{\partial g_H'(\underline{p})}{\partial p(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)}\right]$

where H ranges through θ, $111\ldots1$, $111\ldots2$, ..., $m_1m_2\ldots m_w$ and

$$g_0'(\underline{p}) = \sum_{\ell_1=1}^{m_1} \cdots \sum_{\ell_w=1}^{m_w} p(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w) - 1 \quad \text{and}$$

$$g'_{\ell_1\ell_2\ldots\ell_w}(\underline{p}) = B(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)\ .$$

It is not immediately obvious what $g'_{\ell_1\ell_2\ldots\ell_w}$ is.

$$g'_{\ell_1\ell_2\ldots\ell_w} = p_{\ell_1\ell_2\ldots\ell_w} - \prod_{i=1}^{w} p(i,\ell_i) - \sum_{v=2}^{w-1} \sum_{\psi_v} p(T_u;\Omega_u)\, B(\psi_v;\sigma_v)$$

Consider the following table:

2-way table

	#	coef.	x
(2)	1	1	1
(1^2)	1	-1	-1
	2		0

3-way table

	#	coef.	x
(3)	1	1	1
$(1,2)$	3	-1	-3
(1^3)	1	2	2
	5		0

4-way table

	#	coef.	x
(4)	1	1	1
$(1,3)$	4	-1	-4
$(2,2)$	3	-2	-6
$(1^2,2)$	6	3	18
(1^4)	1	-9	-9
	15		0

5-way table

	#	coef.	x
(5)	1	1	1
$(1,4)$	5	-1	-5
$(2,3)$	10	-2	-20
$(1^2,3)$	10	3	30
$(1,2^2)$	15	4	60
$(1^3,2)$	10	-11	-110
(1^5)	1	44	44
	52		0

6-way table

	#	coef.	x
(6)	1	1	1
$(1,5)$	6	-1	-6
$(2,4)$	15	-2	-30
(3^2)	10	-2	-20
$(1^2,4)$	15	3	45
$(1,2,3)$	60	4	240
(2^3)	15	6	90
$(1^3,3)$	20	-11	-220
$(1^2,2^2)$	45	-14	-630
$(1^4,2)$	15	53	795
(1^6)	1	-265	-265
	203		0

TABLE I COEFFICIENTS IN THE EXPRESSION FOR THE HIGHEST ORDER INTERACTION

The first column in the table represents the p's, for example, $(1,2^2)$ stands for any product of three p's, one of which has one subscript, the other two each two subscripts, the subscripts h,i,j,k,ℓ all being represented. The second column gives the number of different terms of the form $(1,2^2)$. There are $\frac{\binom{5}{2}\binom{3}{2}}{2} = 15$ terms of the form $(1,2^2)$. The third column gives the coefficient of the term in the first column and the entries in the last column are the products of the entries in the second and third column.

Suppose one is dealing with the w-way case. The subscripts are $i_1,i_2,\ldots,i_w$. Consider the following grouping of these subscripts into say n groups $(c_1^{d_1}, c_2^{d_2},\ldots,c_m^{d_m})$. Here $d_1+d_2+\ldots+d_m = n$ and the c_j, d_j are all positive integers. Also $c_1d_1+c_2d_2+\ldots+c_md_m = w$, and $m \leq n$. The number of partitions of this type are

$$\frac{w!}{d_1!\ldots d_m!\,(c_1!)^{c_1}\ldots(c_m!)^{c_m}}$$

. For example, consider $w=6$ and the type of partition $(1^2,2^2)$. Here $h=2, k=4$. The number of distinct partitions of this type $(1^2,2^2)$ is $\frac{6!}{2!2!\,(1!)^2\,(2!)^2} = \frac{720}{16} = 45$.

One can also write the partition as $(1^{e_1},2^{e_2},\ldots,w^{e_w})$, where e_j may be 0. If e_j is 0, then there are no groupings of j letters. Then $e_1+e_2+\ldots+e_w = n$, the number of groups. $1e_1+2e_2+\ldots+we_w = w$.

The number of distinct partitions of this type is

$\dfrac{w!}{e_1!e_2!\ldots e_w!\ (1!)^{e_1}(2!)^{e_2}\ldots(w!)^{e_w}}$. This is a type of symmetric function and can be found in tables of symmetric functions. For example,

$$(1^2,2^2) = (1^2,2^2,3^0,4^0,5^0,6^0).\ \frac{6!}{2!2!0!0!0!0!(1!)^2(2!)^2(3!)^0(4!)^0(5!)^0(6!)^0} = 45.$$

While it is not immediately obvious what the remaining entries for the w-way table would be, we can however still make a few observations. If we denote by c_k, $k = 2,3,\ldots$, the coefficient of the last term for each table, that is, $c_2 = -1$, $c_3 = 2$, $c_4 = -9$ etc. we note the following relationship: $c_{k+1} = -(k+1)\ c_k - 1$ (and $c_1 = 0$). We have proved before that the sum of the coefficients of the terms of $B(1,2,\ldots,w;\ \ell_1,\ell_2,\ldots,\ell_w)$ is zero. Hence the sum of the entries under the heading X must always be zero. Note also that the sign of a coefficient = $(-1)^{\text{no. of groups } -1}$.

Note that a coefficient, say, in the 6-way table can be derived from the coefficients of the previous tables of lower dimension. Consider the term (1,2,3). We obtain from $p(1;g)\ B(2,3,4,5,6;\ h,i,j,k,\ell)$ and the table of coefficients for the five-way table a coefficient of 2, the negative of the coefficient entry in the (2,3) row. From $p(2,3;h,i)\cdot B(1,4,5,6;\ g,j,k,\ell)$ we get a coefficient of 1, the negative of the coefficient entry in the (1,3) row of the table for the four-way case. From $p(4,5,6;\ j,k,\ell)\ B(1,2,3;\ g,h,i)$ we get a coefficient of 1, the negative of the coefficient entry in the (1,2) row of the table for the three-way case. Since (1,2,3) has three groups the coefficient of

(1,2,3) is + (2+1+1) or 4.

Finally note that we can obtain the sum of the coefficients of, say, $\frac{\partial g_{ijk\ell}}{\partial p_{ijk\ell}}$ by referring to the four-way table, multiplying each number in the X column by the number of groups in the entry in the first column. To illustrate, the sum of the coefficients of the terms in $\frac{\partial g_{ijk\ell}}{\partial p_{ijk\ell}}$ is

$1 \times 1 + 2 \times (-4) + 2 \times (-6) + 3 \times 18 + 4 \times (-9) = -1.$

We complete this section by referring back to a four-way table. The work for the w-way table will simply involve more terms. We want to find the rank of $[\frac{\partial}{\partial p_{hijk}} g_H'(\underline{p})]$ where H takes on the values 0, 1111, 1112,...,qrst.

Again we consider the product $\begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & W \otimes X \otimes Y \otimes Z \end{bmatrix} [\frac{\partial}{\partial p_{hijk}} g_H'(\underline{p})]$.

After calculating all the derivatives $\frac{\partial}{\partial p_{dabc}} g_{hijk}'$, multiplying by $W_h^{(m)} X_i^{(u)} Y_j^{(v)} Z_k^{(w)}$ and summing we arrive at

$$\sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} W_h^{(m)} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g_{hijk}'}{\partial p_{dabc}} = W_d^{(m)} X_a^{(u)} Y_b^{(v)} Z_c^{(w)}$$

$$- Z_c^{(w)} \rho_{muvo} - Y_b^{(v)} \rho_{muow} - X_a^{(u)} \rho_{movw} - W_d^{(m)} \rho_{ouvw} - W_d^{(m)} X_a^{(u)} Y_b^{(v)} \Delta_{wo}$$

$$- W_d^{(m)} X_a^{(u)} Z_c^{(w)} \Delta_{vo} - W_d^{(m)} Y_b^{(v)} Z_c^{(w)} \Delta_{uo} - X_a^{(u)} Y_b^{(v)} Z_c^{(w)} \Delta_{mo}$$

$$+ 3Z_c^{(w)} \Delta_{mo} \rho_{ouvo} + 3Y_b^{(v)} \Delta_{mo} \rho_{ouow} + 3Z_c^{(w)} \Delta_{uo} \rho_{movo} + 3Z_c^{(w)} \Delta_{vo} \rho_{muoo}$$

$$+ 3Y_b^{(v)} \Delta_{uo} \rho_{moow} + 3Y_b^{(v)} \Delta_{wo} \rho_{muoo} + 3X_a^{(u)} \Delta_{mo} \rho_{oovw} + 3X_a^{(u)} \Delta_{vo} \rho_{moow}$$

$$+ 3X_a^{(u)} \Delta_{wo} \rho_{movo} + 3W_d^{(m)} \Delta_{uo} \rho_{oovw} + 3W_d^{(m)} \Delta_{vo} \rho_{ouow} + 3W_d^{(m)} \Delta_{wo} \rho_{ouvo}$$

$$+ 3Y_b^{(v)} Z_c^{(w)} \Delta_{mo} \Delta_{uo} + 3X_a^{(u)} Z_c^{(w)} \Delta_{mo} \Delta_{vo} + 3W_d^{(m)} Z_c^{(w)} \Delta_{uo} \Delta_{vo}$$

$$+ 3X_a^{(u)} Y_b^{(v)} \Delta_{mo} \Delta_{wo} + 3W_d^{(m)} Y_b^{(v)} \Delta_{uo} \Delta_{wo} + 3W_d^{(m)} X_a^{(u)} \Delta_{vo} \Delta_{wo} - 9W_d^{(m)} \Delta_{uo}\Delta_{vo}\Delta_{wo}$$

$$- 9X_a^{(u)} \Delta_{mo} \Delta_{vo} \Delta_{wo} - 9Y_b^{(v)} \Delta_{mo} \Delta_{uo} \Delta_{wo} - 9Z_c^{(w)} \Delta_{mo} \Delta_{uo} \Delta_{vo}$$

$$- 2Y_b^{(v)} Z_c^{(w)} \rho_{muoo} - 2X_a^{(u)} Z_c^{(w)} \rho_{movo} - 2W_d^{(m)} Z_c^{(w)} \rho_{ouvo}$$

$$- 2W_d^{(m)} Y_b^{(v)} \rho_{ouow} - 2W_d^{(m)} X_a^{(u)} \rho_{oovw} - 2X_a^{(u)} Y_b^{(v)} \rho_{moow} .$$

$$\sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} W_h^{(m)} X_i^{(u)} Y_j^{(v)} Z_k^{(w)} \frac{\partial g'_{hijk}}{\partial p_{dabc}} =$$

$$\begin{cases} -1 \text{ for } m=0,\ u=0,\ v=0,\ w=0 \text{ (1 equation)} \\ 0 \text{ for } m=0,\ u=0,\ v=0,\ w\neq 0 \text{ ((t-1) equations)} \\ 0 \text{ for } m=0,\ u=0,\ v\neq 0,\ w=0 \text{ ((s-1) equations)} \\ 0 \text{ for } m=0,\ u\neq 0,\ v=0,\ w=0 \text{ ((r-1) equations)} \\ 0 \text{ for } m\neq 0,\ u=0,\ v=0,\ w=0 \text{ ((q-1) equations)} \end{cases}$$

$$
\begin{cases}
2\rho_{oovw} & \text{for } m=0,\ u=0,\ v\neq 0,\ w\neq 0 \ ((s-1)(t-1) \text{ equations}) \\
2\rho_{ouow} & \text{for } m=0,\ u\neq 0,\ v=0,\ w\neq 0 \ ((r-1)(t-1) \text{ equations}) \\
2\rho_{moow} & \text{for } m\neq 0,\ u=0,\ v=0,\ w\neq 0 \ ((q-1)(t-1) \text{ equations}) \\
2\rho_{ouvo} & \text{for } m=0,\ u\neq 0,\ v\neq 0,\ w=0 \ ((r-1)(s-1) \text{ equations}) \\
2\rho_{movo} & \text{for } m\neq 0,\ u=0,\ v\neq 0,\ w=0 \ ((q-1)(s-1) \text{ equations}) \\
2\rho_{muoo} & \text{for } m\neq 0,\ u\neq 0,\ v=0,\ w=0 \ ((q-1)(r-1) \text{ equations}) \\
-\rho_{muvo} & \text{for } m\neq 0,\ u\neq 0,\ v\neq 0,\ w=0 \ ((q-1)(r-1)(s-1) \text{ equations}) \\
-\rho_{muow} & \text{for } m\neq 0,\ u\neq 0,\ v=0,\ w\neq 0 \ ((q-1)(r-1)(t-1) \text{ equations}) \\
-\rho_{movw} & \text{for } m\neq 0,\ u=0,\ v\neq 0,\ w\neq 0 \ ((q-1)(s-1)(t-1) \text{ equations}) \\
-\rho_{ouvw} & \text{for } m=0,\ u\neq 0,\ v\neq 0,\ w\neq 0 \ ((r-1)(s-1)(t-1) \text{ equations}) \\
W_d^{(m)} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} - Z_c^{(w)} \rho_{muvo} - Y_b^{(v)} \rho_{muow} - X_a^{(u)} \rho_{movw} & \\
-W_d^{(m)} \rho_{ouvw} - 2Y_b^{(v)} Z_c^{(w)} \rho_{muoo} - 2X_a^{(u)} Z_c^{(w)} \rho_{movo} & \\
-2W_d^{(m)} Z_c^{(w)} \rho_{ouvo} - 2W_d^{(m)} Y_b^{(v)} \rho_{ouow} - 2W_d^{(m)} X_a^{(u)} \rho_{oovw} & \\
-2X_a^{(u)} Y_b^{(v)} \rho_{moow} & \text{for } m\neq 0,\ u\neq 0,\ v\neq 0,\ w\neq 0, ((q-1)(r-1)(s-1)(t-1) \text{ equations}
\end{cases}
$$

$$
\text{Consider} \begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & W \otimes X \otimes Y \otimes Z \end{bmatrix} \begin{bmatrix} \dfrac{\partial}{\partial p_{hijk}} g_H'(\underline{p}) \end{bmatrix} \quad (qrst + 1) \times qrst
$$

The submatrix consisting of the first row of the product matrix and the qrst − (q−1)(r−1)(s−1)(t−1) rows where one or more of the m,u,v,w is zero has rank one. The submatrix consisting of the rows where $m \neq 0$, $u \neq 0$, $v \neq 0$, $w \neq 0$ has rank (q−1)(r−1)(s−1)(t−1) because of the linear independence of the orthonormal functions. The proof of this, though not a direct extension of the proof for the 3-way case, is very similar. Note that terms of the form $\sum_{d=1}^{q} \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} p_{d\ldots} p_{.a..} p_{..b.} p_{...c} \times$

$$Y_b^{(v)} Z_c^{(w)} \rho_{muoo} W_d^{(m')} X_a^{(u')} Y_b^{(v')} Z_c^{(w')} = \rho_{muoo} \sum_{d=1}^{q} p_{d\ldots} W_d^{(m')} \times$$

$$\sum_{a=1}^{r} p_{.a..} X_a^{(u')} \sum_{b=1}^{s} p_{..b.} Y_b^{(v)} Y_b^{(v')} \sum_{c=1}^{t} p_{...c} Z_c^{(w)} Z_c^{(w')} = 0 \text{ since } m' \neq 0.$$

Hence we conclude that $\left[\dfrac{\partial g'_H(\underline{p})}{\partial p_{dabc}}\right]$ has rank (q−1)(r−1)(s−1)(t−1) + 1.

$$\text{Consider } q_4 = \log(n!) - \sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \log(f_{hijk}!)$$

$$+ \sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} f_{hijk} \log p_{hijk} - \{\theta[\sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} p_{hijk} - 1] +$$

$$\sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{hijk} g'_{hijk}\} .$$

We found that the rank of the coefficient matrix of the homogeneous part of these qrst equations is (q−1)(r−1)(s−1)(t−1) + 1 .
If we obtain the derivative $\dfrac{\partial q_4}{\partial p_{dabc}}$, multiply it by hypothetical p_{dabc}

and then sum with respect to d, a, b, c we obtain the following equation:

$$\sum_{d=1}^{q} \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} p_{dabc} \frac{\partial q_4}{\partial p_{dabc}} = n - \theta + \sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \eta_{hijk} [p_{h...} p_{.ijk}$$

$$+ p_{.i..} p_{h.jk} + p_{..j.} p_{hi.k} + p_{...k} p_{hij.} - 6p_{h...} p_{.ij.} p_{...k}$$

$$- 6p_{.i..} p_{h.j.} p_{...k} - 6p_{..j.} p_{hi..} p_{...k} - 6p_{h...} p_{.i.k} p_{..j.}$$

$$- 6p_{.i..} p_{h..k} p_{..j.} - 6p_{h...} p_{..jk} p_{.i..} + 27p_{h...} p_{.i..} p_{..j.} p_{...k}$$

$$+ 2p_{.i.k} p_{h.j.} + 2p_{..jk} p_{hi..} + 2p_{h..k} p_{.ij.}]$$

If we equate this equation to zero we obtain an expression for θ in terms of the η's.

There are $(q-1)(r-1)(s-1)(t-1)$ linearly independent equations among

$\frac{\partial q_4}{\partial p_{dabc}} = 0$ d=1 to q, a = 1 to r, b=1 to s, c=1 to t after θ has been eliminated.

Let us now look at the original equations in the form $p_{dabc} \times \frac{\partial q_4}{\partial p_{dabc}} = 0$. If θ has been replaced by the expression in the η_{dabc}'s, then there are qrst non-homogeneous linear equations in qrst unknowns. The rank of the coefficient matrix now is $(q-1)(r-1)(s-1)(t-1)$. In order for the system of equations to have a finite solution, the rank of the augmented matrix must be the same as the rank of the coefficient matrix.

Now the term of $p_{dabc} \frac{\partial q_4}{\partial p_{dabc}} = 0$ not involving the η_{dabc}, after θ has been eliminated, is $(f_{dabc} - np_{dabc})$. In order for the augmented matrix to have rank $(q-1)(r-1)(s-1)(t-1)$ it is necessary that $\sum_{d=1}^{q} \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} W_d^{(m)} X_a^{(u)} Y_b^{(v)} Z_c^{(w)} (f_{dabc} - np_{dabc}) = 0$

whenever one or more of the indices m, u, v, w are zero. Let $\hat{W}^{(m)}$, $\hat{X}^{(u)}$, $\hat{Y}^{(v)}$, $\hat{Z}^{(w)}$ be as before.

$$\text{Let } \hat{q}'_{dabc} = \Big[\frac{f_{d...}f_{.abc} + f_{.a..}f_{d.bc} + f_{..b.}f_{da.c} + f_{...c}f_{dab.}}{n^2}$$

$$+ \frac{2f_{da..}f_{..bc} + 2f_{d.b.}f_{.a.c} + 2f_{.ab.}f_{d..c}}{n^2}$$

$$- \frac{3f_{da..}f_{..b.}f_{...c} + 3f_{d.b.}f_{.a..}f_{...c} + 3f_{d..c}f_{.a..}f_{..b.} + 3f_{.ab.}f_{d...}f_{...c} + 3f_{.a.c}f_{d...}f_{..b.} + 3f_{..bc}f_{d...}f_{.a..}}{n^3}$$

$$+ \frac{9f_{d...}f_{.a..}f_{..b.}f_{...c}}{n^4}\Big]$$

By similar method as before we arrive at

$$\sum_{d=1}^{q} \sum_{a=1}^{r} \sum_{b=1}^{s} \sum_{c=1}^{t} \hat{W}_d^{(m)} \hat{X}_a^{(u)} \hat{Y}_b^{(v)} \hat{Z}_c^{(w)} (f_{dabc} - n\hat{q}'_{dabc}) = \begin{cases} 0 & \text{when at least one of } m,u,v,w = 0 \\ n\,\hat{\rho}_{muvw} & \text{when all indices are not zero.} \end{cases}$$

From here the remainder follows similarly to the three-way case and so is left to the reader.

CHAPTER VI NUMERICAL ILLUSTRATION USING OUR MODEL AND METHODS AND OTHER MODELS AND METHODS.

A. THE LIKELIHOOD RATIO WITH OUR MODEL

We will now consider a numerical example to illustrate our three-way model. Consider the data in table II below. This data was taken from Kullback's text "Information Theory and Statistics" (p. 180). It represents the number of items passing, P, or failing, F, two tests T_1, T_2 on certain manufactured products from manufacturers A, B, C, D.

T_1	P	F	Total
A	112	32	144
B	76	20	96
C	87	9	96
D	41	7	48
Total	316	68	384

T_2	P	F	Total
A	84	24	108
B	86	10	96
C	58	14	72
D	40	8	48
Total	268	56	324

TABLE II: DATA REPRESENTING NUMBER OF ITEMS PASSING OR FAILING TWO TESTS ON CERTAIN MANUFACTURED PRODUCTS.

This is a 4×2×2 table. The denominator of the likelihood ratio is

$n! \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \left(\frac{f_{ijk}}{n}\right)^{f_{ijk}} \Big/ f_{ijk}!$. The numerator of the likelihood ratio is

$$n! \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \frac{\left(\dfrac{f_{i..}f_{.jk} + f_{.j.}f_{i.k} + f_{..k}f_{ij.}}{(n)^2} - \dfrac{2f_{i..}f_{.j.}f_{..k}}{(n)^3}\right)^{f_{ijk}}}{f_{ijk}!}$$

The likelihood ratio becomes

$$L = \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \left(\frac{n(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.}) - 2f_{i..}f_{.j.}f_{..k}}{(n)^2 \, f_{ijk}} \right)^{f_{ijk}}$$

$$\text{Log}_e L = \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log \{n(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.}) - 2f_{i..}f_{.j.}f_{..k}\}$$

$$- 2 \times n \log n - \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log f_{ijk}$$

112 log 112	= 528.4718735951	84 log 84	= 372.1886111028
32 log 32	= 110.9035488896	24 log 24	= 76.2732919284
76 log 76	= 329.1357338618	86 log 86	= 383.0738674778
20 log 20	= 59.9146454711	10 log 10	= 23.025809299
87 log 87	= 388.5340063229	58 log 58	= 235.5056946117
9 log 9	= 19.7750211960	14 log 14	= 36.9468026146
41 log 41	= 152.2564547349	40 log 40	= 147.5551781646
7 log 7	= 13.6213710434	8 log 8	= 16.6355323334
	1602.6126551048		1291.2048291632

TABLE III VALUES OF $f_{ijk} \log f_{ijk}$

$2 \times 708 \log 708 = 9292.4208366704$

The last two terms of $\text{Log}_e L \doteq - 12186.2383$

Now $f_{1..} = 252$ $\quad f_{.1.} = 584$ $\quad f_{..1} = 384$ $\quad f_{.11} = 316$

$f_{2..} = 192$ $\quad f_{.2.} = 124$ $\quad f_{..2} = 324$ $\quad f_{.21} = 68$

$f_{3..} = 168$ $\quad f_{.12} = 268$

$f_{4..} = 96$ $f_{.22} = 56$

$f_{11.} = 196$ $f_{12.} = 56$ $f_{1.1} = 144$ $f_{1.2} = 108$

$f_{21.} = 162$ $f_{22.} = 30$ $f_{2.1} = 96$ $f_{2.2} = 96$

$f_{31.} = 145$ $f_{32.} = 23$ $f_{3.1} = 96$ $f_{3.2} = 72$

$f_{41.} = 81$ $f_{42.} = 15$ $f_{4.1} = 48$ $f_{4.2} = 48$

Let $R_{ijk} = [n(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.}) - 2f_{i..}f_{.j.}f_{..k}]$

Consider the following table:

(ijk)	R_{ijk}	log R_{ijk}	f_{ijk} log R_{ijk}
111	56181312	17.8441	1998.5392
211	40578048	17.5188	1331.4288
311	41351040	17.5376	1525.7712
411	20289024	16.8256	689.8496
121	16000704	16.5881	530.8192
221	7543296	15.8361	316.7220
321	6770304	15.7280	141.5520
421	3771648	15.1431	106.0017
112	42066432	17.5548	1474.6032
212	40626720	17.5200	1506.7200
312	31332240	17.2601	1001.0858
412	20313360	16.8266	673.0640
122	12070080	16.3060	391.3440
222	7494624	15.8298	158.2980
322	4758768	15.3755	215.2570
422	3747312	15.1365	121.0920
			12182.1477

TABLE IV

VALUES OF f_{ijk} log R_{ijk} where $R_{ijk} = [n(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.}) -2f_{i..}f_{.j.}f_{.k}]$

$$\sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} f_{ijk} \log R_{ijk} = \sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} f_{ijk} \log[n(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.}) - 2f_{i..}f_{.j.}f_{..k}]$$

$$= 12182.1477$$

Hence $\mathrm{Log}_e L = -4.0906$ and $U = -2\,\mathrm{Log}_e L = 8.1812$. The dimension of Ω, the parameter space, is $rst-1$. The dimension of ω, the subset of the parameter space such that 2nd order interaction is zero, is $(r-1) + (s-1) + (t-1) + (r-1)(s-1) + (r-1)(t-1) + (s-1)(t-1)$. In this case dimension of Ω is 15 and dimension of ω is 12. U is asymptotically distributed as a χ^2 with 3 (= 15-12) degrees of freedom. Since $\chi^{2^3}_{.95} = 7.82$ and $\chi^{2^3}_{.90} = 6.25$ and since $8.1812 > 7.82$ we reject H_0: $\delta_{ijk} = 0$ for all i,j,k at the five percent level.

B. THE CHI-SQUARED STATISTIC WITH OUR MODEL

We now calculate the chi-square statistic $\chi^2 = \Sigma \frac{(O-E)^2}{E}$ where the summation extends over all cells. Consider table V :

	O_{ijk}	$E_{ijk} = {}^{R_{ijk}}/_{(708)^2}$
111	112	112.079287
211	76	80.951450
311	87	82.493536
411	41	40.475725
121	32	31.920712
221	20	15.048549
321	9	13.506463
421	7	7.524274
112	84	83.920712
212	86	81.048549
312	58	62.506463
412	40	40.524274
122	24	24.079287
222	10	14.951450
322	14	9.493536
422	8	7.475725
Sum	708	707.999992

	$\lvert(O_{ijk}-E_{ijk})\rvert$	$(O_{ijk}-E_{ijk})^2$	$\frac{(O_{ijk}-E_{ijk})^2}{E_{ijk}}$
111	.079287	.006286	.000056
211	4.951450	24.516857	.302858
311	4.506464	20.308217	.246179
411	.524275	.274864	.003331
121	.079288	.006286	.000196
221	4.951451	24.516867	1.629184
321	4.506463	20.308208	1.503591
421	.524274	.274863	.036530
112	.079288	.006286	.000074
212	4.951451	24.516867	.302496
312	4.506463	20.308208	.324897
412	.524274	.274863	.006782
122	.079287	.006286	.000261
222	4.951450	24.516857	1.639764
322	4.506464	20.308217	2.139162
422	.524275	.274864	.036767
Sum			8.172128

TABLE V CALCULATION OF χ^2 USING EXPECTED VALUES = ${}^{R_{ijk}}/_{(708)^2}$

$\chi^2 = 8.172128$. As before we reject $H_0 : \delta_{ijk} = 0$ for all i, j, k at the five percent level.

C. BISHOP'S MODEL AND METHODS

We will now examine the contingency table using Yvonne M.M. Bishop's model and methods and compare the results we obtain to those from our model. She defines $np_{ijk} = E(f_{ijk})$ and expresses this expected value in the logarithmic scale as

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

where u is the overall mean value and the subscripted u-terms are the main and multiple-factor effects. The numerical subscripts denote the variables involved and the alphabetic subscripts the categories for these variables in the same order. Thus $u_{12(ij)}$ is the two-factor effect between variables 1 and 2 at levels i and j, respectively.

The subscripted terms are deviations, as in the linear models familiar in analysis of variance of quantitative data. Thus, $\sum_{i=1}^{r} u_{1(i)} = \sum_{i=1}^{r} u_{12(ij)} = \sum_{j=1}^{s} u_{12(ij)} = 0$ and, in general, each u-term sums to zero over any of its variables.

Using this notation, Birch has shown that models corresponding to different hypotheses are defined by omitting one or more terms from expression (1) in order of descending hierarchy. For instance, if we wish to postulate that there is no three-factor effect then $u_{123(ijk)} = 0$

for all i,j,k, or more briefly, $u_{123} = 0$ and the last term of expression (1) disappears.

It is not possible to write down the estimates for the elementary cells as direct products of the configuration cells for all models. When it is not possible, the estimates can be obtained iteratively. In three dimensions the only model that requires iteration is the one mentioned above, that of no three factor effect.

We now describe the iterative procedure. Preliminary values $Y_{ijk}^{(0)}$ are put in every elementary cell of the matrix; in practice we use the value 1 for every cell and it is apparent that as $\log 1 = 0$ we have not introduced unwanted multiple-factor effects. Any constant could be used, or any set of numbers that do not exhibit higher-order effects than those we wish to estimate. The preliminary values, $Y_{ijk}^{(0)}$, are then adjusted to yield estimates $Y_{ijk}^{(1)}$ in each cell where

$Y_{ijk}^{(1)} = \dfrac{Y_{ijk}^{(0)} f_{ij.}}{Y_{ij.}^{(0)}}$. The new estimates are again adjusted to yield

$Y_{ijk}^{(2)} = \dfrac{Y_{ijk}^{(1)} f_{i.k}}{Y_{i.k}^{(1)}}$. The cycle is completed when these values are

adjusted to yield $Y_{ijk}^{(3)} = \dfrac{Y_{ijk}^{(2)} f_{.jk}}{Y_{.jk}^{(2)}}$. The cycle is repeated until no difference is discernible between $Y_{ijk}^{(3r)}$ and $Y_{ijk}^{(3r-1)}$. In practice we proceed until no cell estimate differs from the preceding estimate for this cell by more than 0.01.

Applying the iterative procedure to our considered contingency table we obtain the following sequence of tables:

$Y_{ijk}^{(0)}$			Sum
	1	1	2
	1	1	2
	1	1	2
	1	1	2
Sum	4	4	8

			Sum
	1	1	2
	1	1	2
	1	1	2
	1	1	2
Sum	4	4	8

TABLE VI

$Y_{ijk}^{(1)}$			Sum
	98.0	28.0	126.0
	81.0	15.0	96.0
	72.5	11.5	84.0
	40.5	7.5	48.0
Sum	292.0	62.0	354.0

			Sum
	98.0	28.0	126.0
	81.0	15.0	96.0
	72.5	11.5	84.0
	40.5	7.5	48.0
Sum	292.0	62.0	354.0

TABLE VII

$Y_{ijk}^{(2)}$			Sum
	112.0000 0000	32.0000 0000	144.0000 0000
	81.0000 0000	15.0000 0000	96.0000 0000
	82.8571 4286	13.1428 5714	96.0000 0000
	40.5000 0000	7.5000 0000	48.0000 0000
Sum	316.3571 4286	67.6428 5714	384.0000 0000

$Y_{ijk}^{(2)}$			Sum
	84.0000 0000	24.0000 0000	108.0000 0000
	81.0000 0000	15.0000 0000	96.0000 0000
	62.1428 5714	9.8571 4286	72.0000 0000
	40.5000 0000	7.5000 0000	48.0000 0000
Sum	267.6428 5714	56.3571 4286	324.0000 0000

TABLE VIII

$Y_{ijk}^{(3)}$			Sum
	111.8735 6061	32.1689 5460	144.0425 1521
	80.9085 5724	15.0791 9747	95.9877 5471
	82.7636 0354	13.2122 4920	95.9758 5274
	40.4542 7860	7.5395 9874	47.9938 7734
Sum	315.9999 9999	68.0000 0001	384.0000 0000

$Y_{ijk}^{(3)}$			Sum
	84.1120 8968	23.8479 0874	107.9599 9842
	81.1080 8647	14.9049 4296	96.0130 2943
	62.2257 8060	9.7946 7681	72.0204 5741
	40.5540 4325	7.4524 7148	48.0065 1473
	268.0000 0000	55.9999 9999	323.9999 9999

TABLE IX

We continue this cycle three more times and arrive at the following table after rounding off to the fifth decimal place:

$Y_{ijk}^{(12)}$			Sum
	111.84777	32.15223	144.00000
	80.90964	15.09036	96.00000
	82.78777	13.21223	96.00000
	40.45482	7.54518	48.00000
Sum	316.00000	68.00000	384.00000

$Y_{ijk}^{(12)}$			Sum
	84.15223	23.84777	108.00000
	81.09036	14.90964	96.00000
	62.21223	9.78777	72.00000
	40.54518	7.45482	48.00000
Sum	268.00000	56.00000	324.00000

TABLE X

TABLES VI- X CALCULATION OF EXPECTED VALUES USING ITERATIVE METHOD

We now calculate $\chi^2 = \Sigma \frac{(O-E)^2}{E}$ where the summation extends over all the cells. We use the values in the last table as the expected cell values.

Consider the following table:

(ijk)	0_{ijk}	E_{ijk}	$\lvert 0_{ijk}-E_{ijk}\rvert$	$(0_{ijk}-E_{ijk})^2$	$\frac{(0_{ijk}-E_{ijk})^2}{E_{ijk}}$
111	112	111.84777	.15223	.02317	.00021
211	76	80.90964	4.90964	24.10456	.29792
311	87	82.78777	4.21223	17.74288	.21432
411	41	40.45482	.54518	.29722	.00735
121	32	32.15223	.15223	.02317	.00072
221	20	15.09036	4.90964	24.10456	1.59735
321	9	13.21223	4.21223	17.74288	1.34291
421	7	7.54518	.54518	.29722	.03939
112	84	84.15223	.15223	.02317	.00028
212	86	81.09036	4.90964	24.10456	.29726
312	58	62.21223	4.21223	17.74288	.28520
412	40	40.54518	.54518	.29722	.00733
122	24	23.84777	.15223	.02317	.00097
222	10	14.90964	4.90964	24.10456	1.61671
322	14	9.78777	4.21223	17.74288	1.81276
422	8	7.45482	.54518	.29722	.03987
					7.56055

TABLE XI CALCULATION OF χ^2 USING EXPECTED VALUES DERIVED FROM ITERATIVE METHOD

We note that the expected values in Table V and in Table XI differ by very little. $\chi^2 = 7.56055$. Hence we would reject $H_0: u_{123(ijk)} = 0$ for all i, j, k at the ten percent level but not at the five percent level.

Finally let us calculate $U = -2 \log_e L$ using Bishop's expected values.

$$L = \frac{n! \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \dfrac{\left(\dfrac{\hat{f}_{ijk}}{n}\right)^{f_{ijk}}}{f_{ijk}!}}{n! \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \dfrac{\left(\dfrac{f_{ijk}}{n}\right)^{f_{ijk}}}{f_{ijk}!}}$$

$$= \prod_{i=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} \left(\frac{\hat{f}_{ijk}}{f_{ijk}}\right)^{f_{ijk}}$$

, where $\hat{f}_{ijk}$ are Bishop's expected values.

$$\operatorname{Log}_e L = \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log \hat{f}_{ijk} - \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log f_{ijk}$$

$$= \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} f_{ijk} \log \hat{f}_{ijk} - 2893.81748 .$$

(five digits)

$\hat{f}_{ijk}$	$\log \hat{f}_{ijk}$	f_{ijk}	$f_{ijk} \log \hat{f}_{ijk}$
111.85	4.71716	112	528.32192
80.910	4.39333	76	333.89308
82.788	4.41628	87	384.21636
40.455	3.70019	41	151.70779
32.152	3.47047	32	111.05504
15.090	2.71403	20	54.28060
13.212	2.58112	9	23.23008
7.5452	2.02091	7	14.14637
84.152	4.43262	84	372.34008
81.090	4.39555	86	378.01730
62.212	4.13054	58	239.57132
40.545	3.702410	40	148.09640
23.848	3.17170	24	76.12080
14.910	2.702030	10	27.02030
9.7878	2.28114	14	31.93596
7.4548	2.00886	8	16.07088
			2890.02428

TABLE XI VALUES OF $f_{ijk} \log \hat{f}_{ijk}$ WHERE

$\hat{f}_{ijk}$ ARE BISHOP'S EXPECTED VALUES

$\log_e L = 2890.02428 - 2893.81748 = -3.79320$. $U = -2\log_e L = 7.58640$, hence we reject H_0: $u_{123(ijk)} = 0$, for all i, j, k at the ten percent level but not at the 5% level. These results are in close agreement with our previous results.

D. COMPARISON OF ADDITIVE AND MULTIPLICATIVE (LOG LINEAR) MODELS FOR A THREE-WAY CONTINGENCY TABLE.

For the additive model of an $r \times s \times t$ contingency table:

$$p_{ijk} = p_{i..}p_{.j.}p_{..k} + p_{i..}\alpha_{jk} + p_{.j.}\beta_{ik} + p_{..k}\gamma_{ij} + \delta_{ijk},$$

where $\alpha_{jk} = p_{.jk} - p_{.j.}\,p_{..k}$,

$\beta_{ik} = p_{i.k} - p_{i..}p_{..k}$,

$\gamma_{ij} = p_{ij.} - p_{i..}p_{.j.}$, and

$$\delta_{ijk} = p_{ijk} - p_{i..}p_{.jk} - p_{.j.}p_{i.k} - p_{..k}p_{ij.} + 2p_{i..}p_{.j.}p_{..k} .$$

The most interesting hypotheses for this model are

I $\delta_{ijk} = 0$, for all i,j,k, $p_{ijk} = p_{i..}p_{.jk} + p_{.j.}p_{i.k} + p_{..k}p_{ij.} - 2p_{i..}p_{.j.}p_{..k}$

II $\delta_{ijk} = 0$ and $\gamma_{ij} = 0$ for all i,j,k, $p_{ijk} = p_{i..}p_{.jk} + p_{.j.}p_{i.k} - p_{i..}p_{.j.}p_{..k}$

III $\delta_{ijk} = 0$, $\gamma_{ij} = 0$, $\beta_{ik} = 0$ for all i,j,k, $p_{ijk} = p_{i..}p_{.jk}$

IV $\delta_{ijk} = 0$, $\gamma_{ij} = 0$, $\beta_{ik} = 0$, $\alpha_{jk} = 0$ for all i,j,k, $p_{ijk} = p_{i..}p_{.j.}p_{..k}$

The maximum likelihood estimators for these hypotheses can be obtained by Lagrangian methods simlar to those presented earlier in this thesis.

Parameter Constraints		$\hat{P}_{ijk}$ (MLE) for Likelihood Ratio or χ^2 Tests: Additive Model	Log-Linear Model
	No Constraints on model parameters	$\frac{f_{ijk}}{n}$	$\frac{f_{ijk}}{n}$ Observed cell proportions.
I	Zero 2nd-order interactions; no constraints on 1st order (AB, AC,BC) interaction.	$\frac{1}{n^2}(f_{i..}f_{.jk}+f_{.j.}f_{i.k}+f_{..k}f_{ij.})$ $-\frac{2}{n^3}(f_{i..}f_{.j.}f_{..k})$	no closed-form expression; requires iterative numerical estimation procedure.
II	Zero 2nd-order interactions and zero AB 1st order interaction .	$\frac{1}{n^2}(f_{i..}f_{.jk}+f_{.j.}f_{i.k})$ $-\frac{1}{n^3}(f_{i..}f_{.j.}f_{..k})$ $= [\frac{f_{.jk}}{f_{..k}}\frac{f_{i..}}{n} + \frac{f_{i.k}}{f_{..k}}\frac{f_{.j.}}{n}$ $-\frac{f_{i..}}{n}\frac{f_{..j}}{n}]\frac{f_{..k}}{n}$	$\frac{f_{i.k}f_{.jk}}{n f_{..k}}$ $=\frac{f_{i.k}}{f_{..k}}\frac{f_{.jk}}{f_{..k}}\frac{f_{..k}}{n}$ variables A and B conditionally independent given level of C.
III	Zero 2nd-order, AB, AC interaction.	$\frac{f_{.jk}f_{i..}}{n^2}$	$\frac{f_{.jk}f_{i..}}{n^2}$ A independent of B and C jointly .
IV	Independence zero 2nd order, AB, BC, and AC interaction.	$\frac{f_{i..}f_{.j.}f_{..k}}{n^3}$	$\frac{f_{i..}f_{.j.}f_{..k}}{n^3}$

TABLE XIII ADDITIVE MODEL V.S. MULTIPLICATIVE (LOG LINEAR) MODEL.

The two models are equivalent in cases of full independence or one set of first-order interactions. In case II, with two sets of first-order interactions, the log-linear model has a concrete, useful interpretation: if this hypothesis is accepted, then the overall 2-way table of A cross classified with B is regarded as a pooling (over levels of C) of several independent B × C tables. Case II for the additive model does not have an obvious interpretation.

One virtue of the additive model is that maximum likelihood estimates under all its natural hypotheses are intuitive, are easy to verify theoretically, and can be calculated in closed form, whereas the log-linear model's hypothesis of zero second order interactions requires an iterative numeric procedure to compute maximum likelihood estimates.

B I B L I O G R A P H Y

[1] Apostol M.T. (1957) Mathematical Analysis, A Modern Approach to Advanced Calculus. Reading, Mass.: Addison-Wesley Co., Inc. pp.152-156.

[2] Bartlett M.S. (1935) "Contingency Table Interactions." J. Roy. Statist. Soc., Suppl. 2. pp.248-252.

[3] Birch M.W. (1963) "Maximum Likelihood in Three-way Contingency Tables." J. Roy. Statist. Soc., Ser. B, Vol. 25. pp.220-233.

[4] Bishop Y.M.M. (1969) "Full Contingency Tables, Logits, and Split Contingency Tables." Biometrics, Vol. 25. pp. 383-399.

[5] Bishop Y.M.M. and Fienberg S.E. (1969) "Incomplete Two- Dimensional Contingency Tables." Biometrics, Vol. 25. pp. 119-128.

[6] Bishop Y.M.M. and Mosteller F. (1969) "Smoothed Contingency Table Analysis." Chapt. IV - 3 of the National Halothane Study, U.S. Gov. Printing Office. pp. 238-272.

[7] Darroch J.N. (1962) "Interactions in Multifactor Contingency Tables." J. Roy. Statist. Soc., Ser B, Vol 25. pp251-263.

[8] Deming E. W. and Stephan F.F. (1940) "On a Least Squares Adjustment of a Sampled Frequency Table When the Marginal Totals Are Known." Ann. Math. Statist., Vol. 11. pp. 427-444.

[9] Feller W. (1968) An Introduction to Probability Theory and Its Applications. Vol. I, 3rd. Edition. New York: John Wiley & Sons, Inc. pp.125-128.

[10] Fienberg S.E. (1970) "The Analysis of Multidimensional Contingency Tables." Ecology, Vol. 51, No. 3. pp.419-433.

[11] Fienberg S.E. (1970) "An Iterative Procedure for Estimation in Contingency Tables." Ann. Math. Statist., Vol. 41. pp.907-917.

[12] Good I.J. (1963) "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables." Ann. Math. Statist., Vol. 34. pp.911-934.

[13] Good I.J. (1965) "The Estimation of Probabilities: An Essay on Modern Bayesian Methods." Research Monographs, No. 30. Cambridge, Mass.: The M.I.T. Press.

[14] Good I.J. (1966) "How to Estimate Probabilities." J. Inst. Math. Appl., Vol. 2. pp.364-383.

[15] Goodman L.A. (1964a) "Interactions in Multidimensional Contingency Tables." Ann. Math. Statist., Vol. 35. pp.632-646.

[16] Goodman L.A. (1964b) "Simple Method for Analyzing Three-Factor Interaction in Contingency Tables." J. Amer. Statist. Assoc., Vol. 59. pp.319-352.

[17] Goodman L.A. (1968) "The Analysis of Cross-Classified Data." J. Amer. Statist. Assoc., Vol. 63 pp.1091-1131.

[18] Goodman L.A. (1969) "On Partitioning Chi-Square and Detecting Partial Association in Three-Way Contingency Tables." J. Roy. Statist. Soc., Ser B, Vol 31. pp.486-498.

[19] Kastenbaum M.A. and Lamphier D.E. (1959) "Calculation of Chi-Square to Test the No Three-Factor Interaction Hypothesis." Biometrics, Vol. 15. pp.107-115.

[20] Kolmogorov A.N. (1956) Foundations of the Theory of Probability. New York: Chelesa Pub. Co.

[21] Ku H.H. and Kullback S. (1968) "Interaction in Multidimensional Contingency Tables: An Information Theoretic Approach." J. Res. Nat. Bur. Standards, Sect. B. Vol. 72, No. 3. pp.159-199.

[22] Kullback S. (1959) Information Theory and Statistics. New York: John Wiley & Sons, Inc. pp.155-188.

[23] Lancaster H.O. (1951) "Complex Contingency Tables Treated by the Partition of χ^2." J. Roy. Statist. Soc., Ser. B, Vol. 13. pp.242-249.

[24] Lancaster H.O. (1969) The Chi-Squared Distribution. New York: John Wiley & Sons, Inc.

[25] Mardia K.V. (1970) Families of Bivariate Distributions . London: Charles Griffin & Co., Ltd. pp.30-33.

[26] Maxwell A.E. (1961) Analyzing Qualitative Data. London: Methuen & Co., Ltd.

[27] Mosteller F. (1968) "Association and Estimation in Contingency Tables" J. Amer. Statist. Assoc., Vol. 63. pp.1-28.

[28] Norton H.W. (1945) "Calculation of Chi-Square for Complex Contingency Tables" J. Amer. Statist. Assoc., Vol. 40. pp.251-258.

[29] Rao C.R. (1952) Advanced Statistical Methods in Biometric Research. New York: John Wiley & Sons, Inc. pp.165-166.

[30] Roy S.N. and Kastenbaum M.A. (1956) "On the Hypothesis of No Interaction in a Multiway Contingency Table." Ann. Math. Statist., Vol. 27. pp.749-757.

[31] Simpson E.H. (1951) "The Interpretation of Interaction in Contingency Tables." J. Roy. Statist. Soc., Ser. B, Vol. 13. pp.238-241.

[32] Sisam C.H. (1946) College Mathematics. New York: Henry Holt & Co. pp.382-383.

[33] Wilks S.S. (1962). Mathematical Statistics. New York: John Wiley & Sons, Inc. pp.39-41, 49-50.

[34] Woolf B. (1955) "On Estimating the Relation Between Blood Group and Disease." Ann. Hum. Gen., Vol. 19. pp.251-253.