CONTRIBUTIONS TO THE STATISTICAL ANALYSIS OF EXPERIMENTS WITH ONE OR MORE RESPONSES (NOT NECESSARILY NORMAL)

by

Vasant P. Bhapkar

University of North Carolina

This research was supported partly by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 229
July 1959

ACKNOWLEDGMENTS

I am deeply indebted to Professor S. N. Roy for suggesting the problem to me and for his inspiring guidance and constant encouragement. I would also like to express my grateful appreciation to Professor Wassily Hoeffding and Professor W. J. Hall for going through the manuscript and for their suggestions.

I owe my sincere thanks to the U. S. Educational Foundation in India and the Institute of International Education for the Smith-Mundt and Fulbright awards. I am also grateful to the Office of Naval Research for financial assistance.

Finally, thankful appreciation is due to Mrs. Ouida Taylor for the careful work of typing the manuscript, and to the secretarial staff of the Department of Statistics at Chapel Hill and, in particular, to Miss Marianne Byrd for the help received from her at various stages of the work.

TABLE OF CONTENTS

INTRODUCTION

In the general analysis of variance situation, we have observations on a single character of each individual or experimental unit for various factor-combinations. This may be called a multi-factor uni-response situation. Such data are usually analyzed under the assumption of a normal distribution for the response.

More generally, we may have observations on several characters of each individual or experimental unit for various factor-combinations. This may be called a multifactor multi-response situation. Such data also are usually analyzed under the assumption of joint normality of the several responses for each individual.

Similarly when a number of observations on several variables are available, the associations between the variables are generally studied under the assumption of normality.

In Part I of this thesis, we shall be concerned with experimental data given in the form of frequencies in cells determined by a finitely multi-way cross-classification, with predefined categories, finite in number, along each way of classification. We shall pose hypotheses, which might be considered to be generalizations appropriate to this set up of the usual hypotheses (i) in classical "normal" univariate "fixed effects" analysis of variance or ANOVA, (ii) "normal" multivariate "fixed effects" analysis of variance or MANOVA, and (iii) in analysis of various kinds of "normal" independence, and shall offer large sample tests for such hypotheses.

The large sample tests suggested for all these cases in Part I are based on the frequency χ^2-test of Karl Pearson [25]. Analysis of categorical data, thus going back to Karl Pearson, has been developed at subsequent stages, among others, by Fisher [10], Barnard [2], E. S. Pearson [24], Cramer [7] and Neyman [22]. However, Part I of this thesis is along the line, historically going back to Barnard and E. S. Pearson for the simple 2 × 2 table, but developed extensively (and for a long time in ignorance of Barnard and Pearson's prior work) for more general cases by Mitra [19], Roy and Mitra [33], Ogawa [23] and Diamond [8]. So far as the mathematical methods are concerned, Part I of this thesis uses and extends the ones introduced by Cramer [7] and further developed by Mitra [19], Ogawa [23] and Diamond [8].

The general probability model for Part I is that of a product of several multinomial distributions. According as the marginal frequencies along any way or dimension are held fixed or left free, that dimension or way will be said to be associated with a "factor" or a "response."

The model is then $\prod_j \frac{n_{oj}!}{\prod_i n_{ij}!} \prod_i p_{ij}^{n_{ij}}$ such that $\sum_i p_{ij} \equiv p_{oj} = 1$ and $\sum_i n_{ij} \equiv n_{oj}$ is fixed, $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,s$.

Thus here "i" refers to categories of the response while "j" refers to categories of the factor. n_{oj} denotes the pre-assigned sample size for the j-th factor-category, out of which n_{ij} happen to lie in the i-th response-category. It should be further noticed that i may be a multiple subscript, say, $i_1 i_2 \ldots i_k$ with $i_1 = 1,2,\ldots,r_1$; $i_2 = 1,2,\ldots,r_2$; $\ldots$; $i_k = 1,2,\ldots,r_k$; so that, all combinations being supposed to be allowed, $r = r_1 r_2 \ldots r_k$. Likewise, j also might be a multiple subscript, say, $j_1 j_2 \ldots j_\ell$ with $j_1 = 1,2,\ldots,s_1$; $\ldots$; $j_\ell = 1,2,\ldots,s_\ell$; but with this important distinction that all combinations may not be allowed. This will be called a k-response (or k-variate) and ℓ-factor problem, $i_1,i_2,\ldots,i_k$ denoting categories of responses and $j_1,j_2,\ldots,j_\ell$ denoting categories of factors.

According as a set of (real) values is or is not associated with the categories along any way of classification (factor or response), that way of classification will be said to be structured or unstructured. If the categories are class-intervals for a continuous variate (or factor), then the distances from an arbitrary origin of, say, the midpoints of such intervals form a natural set of associated scores. Likewise, if the variate (or factor) is discrete, then these values will be natural scores. If the response (or factor) is not even discrete but is categorical with an implied ranking (like good, fair, bad) then the scores may be assigned to these categories accordingly. It may happen that a system of scores is assigned on some other considerations, even for categories without any implied ranking to start with.

So far as the problems of interest are concerned, the main distinction between the unstructured and the structured case seems to be the following. It is possible and also useful, in the structured case, to define certain over-all aspects of the distribution. Then, while it is still possible to study the same problems as in the unstructured case, such problems are hardly of the same interest for the structured case and the problems involving the newly defined over-all aspects are the ones that become more meaningful.

Part II considers, broadly speaking, the same kinds of problems as Part I, but under the usual probability models of the nonparametric inference. An important distinction between this Part I approach (which might be characterized as one kind of nonparametric set up and hence might be called categorical nonparametric set up) and the nonparametric approach may be pointed out. In the nonparametric set up, while studying some aspect of different populations, it is generally assumed that the population distributions are identical, apart from the aspect that is being studied. In the case of one-way classification, for example, it is assumed that the factor under consideration leaves the entire distribution unaffected except for location. In some situations such an assumption may be quite realistic while in others it might be quite unrealistic. It may also happen that the experimenter is interested in some over-all aspect of the distribution (e.g., average), whatever may be the form of the distribution. In other words, the main point of interest might be just how the different levels of a particular factor are affecting some special feature of the whole response and not the whole response itself. This type of problems, in particular, and some other types, in general, can be tackled more easily (at least for large samples) under the approach of Part I than under the nonparametric approach. Naturally, since the set up is more general with fewer assumptions, we have to be content, at least at this stage, with approximate criteria of asymptotic nature. On the other hand, in the nonparametric set up with some broad restrictive assumptions, it is often possible to have exact criteria for small samples and asymptotic approximations for large samples. Thus the nonparametric approach may be recommended where we have reasons to believe that the restrictive model is not so unrealistic while categorical approach may be recommended for other situations, with the proviso that, as of now, no small sample tests are available. This means that, at the moment, there is nothing we can do in those situations where the categorical approach seems to be more reasonable and where, at the same time, we need exact tests.

In this thesis, the first three chapters (i.e., Part I) deal with the categorical approach, while the last two chapters (i.e., Part II) deal with the nonparametric approach.

In Chapter I, the hypotheses are posed which are considered to be analogues appropriate to this categorical set up of the usual hypotheses in classical normal "fixed effects" ANOVA and MANOVA, and in analysis of various kinds of normal independence. A more detailed version of this material has already appeared elsewhere (see Roy and Bhapkar [31]), but is included here for the sake of completeness.

In Chapter II, two theorems on minimum χ_1^2 are proved. These have been already proved by Neyman [22]. The first one is proved here along Cramer's lines while the second gives the justification for Neyman's linearization procedure.

In Chapter III, some special problems out of those posed in Chapter I are investigated. The univariate two-factor problems are studied in some detail and other problems are considered briefly. It has been shown that for "linear hypotheses," the minimum χ_1^2 is the same as the one obtained by the "general least squares" approach on some asymptotically normal variables.

In Chapter IV, Mood's test [20] for the two-way classification has been extended to cover incomplete block situations. An extension of Hoeffding's theorem [14] on U-statistics is stated and proved and a new test-criterion for the problem of c samples is offered.

In Chapter V, some regression problems and some bivariate problems in the nonparametric set up are studied. Most of the test-criteria developed are asymptotic in nature. The methods employed for the regression problems are extensions of those used by Mood and Brown [20].

NOTATION

As far as possible the following notation will be used, all departures being clearly indicated at the proper places.

Matrices will be denoted by capital letters, small letters underscored will denote column-vectors and row-vectors if they are primed. The transpose of a matrix A is denoted by A'. A matrix M with p rows and q columns will sometimes be written as $M_{p\times q}$ to denote its structure.

d.f. denotes degrees of freedom.

$N(\mu,\sigma)$ denotes a normal variable with mean μ and standard deviation σ.

$N(\underline{\mu},\Sigma)$ denotes a random variable having the multivariate normal distribution with mean-vector $\underline{\mu}$ and variance-covariance matrix Σ.

$\xrightarrow{(p)}$ denotes convergence in probability.

When there are multiple subscripts, as in p_{ijk}, a zero in the place of a subscript indicates the result of summation over that subscript.

A star in the place of a subscript will indicate that the quantity in question is independent of that subscript.

If "i" denotes the categories of a response, we shall, in short, say that "i" is a response.

J_r will denote a matrix $[1]_{r\times r}$.

$\underline{J}$ will denote a column-vector of unities.

$\approx$ will denote 'approximately' (sometimes in probability).

O will denote 'of the order' (sometimes in probability).

$\mathcal{E}$ will denote 'expectation.'

In Chapter IV, sometimes, capital letters denote 'variables' while small letters denote 'fixed quantities.'

In Part I, q with some subscripts will denote a quantity which is not necessarily a probability.

CHAPTER I

SOME ANALOGUES OF THE CUSTOMARY HYPOTHESES IN "NORMAL" ANOVA, MANOVA, AND IN STUDIES OF "NORMAL" ASSOCIATION

1.1 Introduction

Roy and Mitra [33] state: "This is an attempt at a somewhat systematic exposition (i) which is based on a clear distinction between a 'variate' (response) and a 'way of classification' (factor), that stems from differing experimental situations and sampling schemes, (ii) which sets up different probability models for the different situations, and (iii) which poses different types of hypotheses according as it is a 'multivariate analysis' situation or an 'analysis of variance' situation or something of a mixed type."

Accordingly, we shall pose hypotheses, that might be considered to be generalizations appropriate to this categorical set up of the usual hypotheses in the classical 'normal' set up. The problems of interest will naturally depend on the nature of 'responses' and 'factors.' If a set of values is associated with the categories along any way of classification (factor or response), that way of classification will be said to be structured. We shall consider three different types of problems, namely where (i) all responses are unstructured and so also are all factors, (ii) some responses are structured and factors are unstructured,

(iii) responses are structured and so also are some factors. To make everything concrete we shall consider here only some three-way, four-way and five-way tables. This will serve as an illustration and will indicate what happens as we increase the dimensionality of the table.

1.2 A three-way table (ijk) in which "i" is a response and "j" and "k" are factors

Let i denote the categories of the response and j,k denote the categories corresponding to the two factors. Then the probability model is given by the product-multinomial distribution

$$(1.2.1) \qquad \Phi = \prod_{j,k} \left[\frac{n_{ojk}!}{\prod_i n_{ijk}!} \prod_i p_{ijk}^{n_{ijk}} \right] ,$$

where $\sum_i p_{ijk} \equiv p_{ojk} = 1$ and $\sum_i n_{ijk} \equiv n_{ojk}$ (fixed). Suppose $i = 1,2,\ldots,r$; $j = 1,2,\ldots,s$ and $k = 1,\ldots,t$ but with the provision that all combinations (jk) may not be allowed. In other words, given j, k takes a set of values which is a subset of $(1,2,\ldots t)$, depending upon j . We shall refer to these as either a complete design or an incomplete design, as the case may be.

1.2.1 The case where "i", "j" and "k" are all unstructured

The hypothesis of no interaction between "j" and "k" (i.e., between the two factors) means, essentially, that for a given i , there is a lesser number of unknown parameters than would be given by all allowable (jk) combinations. Two specializations are in the same spirit as in ordinary analysis of variance.

$$(1.2.2) \qquad H_{01} : p_{ijk} = q_{i*k}\, q_{ij*} \quad ,$$

and

$$(1.2.3) \qquad H_{01}^{(1)} : p_{ijk} = q_{i*k}^{(1)} + q_{ij*}^{(1)} \quad .$$

As explained in the section on notation, we shall be using q with some subscripts as a general symbol for quantities which are not necessarily probabilities. The physical interpretations are as follows. For (1.2.2), $p_{ijk}/p_{ij'k}$ is independent of "k". This means that, for any two categories of the first factor, the proportions in the i-th category of the response are in the same ratio for any category of the second factor. Similarly, $p_{ijk}/p_{ijk'}$ is independent of "j" with a similar interpretation. Hence we might call (1.2.2) the hypothesis of no interaction (between the two factors) in the multiplicative sense. For (1.2.3), $p_{ijk} - p_{ij'k}$ is independent of "k" and $p_{ijk}-p_{ijk'}$ is independent of "j", with similar interpretations. Hence we might call (1.2.3) the hypothesis of no interaction (between the two factors) in the additive sense.

Now, if the design is complete, then summing both sides of (1.2.2) over "j" and "k" separately and then jointly it is easy to check that (1.2.2) can be rewritten (letting c stand for complete) in the equivalent form

$$(1.2.4) \qquad H_{01}^{(c)} : p_{ijk} = p_{iok}\, p_{ijo}/p_{ioo} \quad .$$

It must be remembered that none of p_{iok}, p_{ijo} or p_{ioo} is a probability, each being based on summation over subscripts belonging to different multinomial distributions. Thus (1.2.4) is only formally similar to (1.3.2) (to be discussed in the next section), but (1.2.4) would be identical with the condition that there is no partial association between "j" and "k", for given "i", if "j" and "k" were variates and not factors as in the present case. Thus the "no interaction" hypothesis in the form (1.2.2) (for a complete design) is related to the hypothesis of "no partial association" in the case where "j" and "k" are variates.

Going back to H_{01} we can test the narrower hypothesis that q_{ij*} is independent of "j", or, in other words, that p_{ijk} is independent of "j", which, without any loss of generality, can be written in the form

(1.2.5) $$H_{02} : p_{ijk} = q_{i*k} \quad .$$

We shall eventually get the same hypothesis if we proceed similarly from $H_{01}^{(1)}$. Now, if we assume for concreteness that "j" stands for treatments and "k" for blocks, (i) H_{01} and (ii) $H_{01}^{(1)}$ state respectively the hypothesis or model of no interaction (i) in the multiplicative sense and (ii) in the additive sense. H_{02} states the hypothesis of no treatment effect. It is open to us (depending upon past knowledge) (a) to start from (1.2.1) as the model and test as a hypothesis either H_{01} or $H_{01}^{(1)}$ or directly even H_{02} , or

(b) to start from a model which is (1.2.1) together with either H_{01} or $H_{01}^{(1)}$, and then to test H_{02} as a hypothesis.

1.2.2 The case when "i" is structured

In this case, the natural analogues of H_{01} and $H_{01}^{(1)}$ seem to be

$$(1.2.6) \qquad H_{03} : \sum_i a_i\, p_{ijk} = q_{**k}\, q_{*j*} \quad ,$$

and

$$(1.2.7) \qquad H_{03}^{(1)} : \sum_i a_i\, p_{ijk} = q_{**k}^{(1)} + q_{*j*}^{(1)} \quad ,$$

where a_i's are the scores associated with the categories of the response. H_{03} and $H_{03}^{(1)}$ are then seen to be hypotheses of no interaction in the multiplicative and the additive sense respectively, appropriate to the case of structured variate where we might be primarily interested in the average response. (1.2.7) seems to be more natural, being in the spirit of the usual hypothesis of no interaction in the analysis of variance.

Remembering that $q_{**k}^{(1)}$ and $q_{*j*}^{(1)}$ are completely unknown, we can rewrite (1.2.7) in the equivalent form

$$(1.2.8) \qquad H_{03}^{(1)} : \sum_i a_i\, p_{ijk} = \sum_i a_i\, q_{i*k}^{(1)} + \sum_i a_i\, q_{ij*}^{(1)} \quad ,$$

or

$$\sum_i a_i[p_{ijk} - q_{i*k}^{(1)} - q_{ij*}^{(1)}] = 0 \quad .$$

It is easy to see that $H_{01}^{(1)} \Longrightarrow H_{03}^{(1)}$ but not conversely, and that (1.2.8) for all sets of a_i's $\Longrightarrow H_{01}^{(1)}$. On the other hand, neither $H_{01} \Longrightarrow H_{03}$ nor does the converse hold (even for all sets of a_i's). Suppose we ask what happens if (1.2.6) is to hold for all sets of a_i's? Rewriting the right side of (1.2.6) as, say, $q_{*j*} \sum_i a_i q_{i*k}$, we can rewrite (1.2.6) as

$$(1.2.9) \qquad H_{03}: \sum_i a_i [p_{ijk} - q_{i*k}\, q_{*j*}] = 0 \quad .$$

This means that if we set up a hypothesis

$$(1.2.10) \qquad H_{04}: \quad p_{ijk} = q_{i*k}\, q_{*j*} \quad ,$$

then $H_{04} \Longrightarrow H_{03}$ but not conversely. However, H_{03} for all a_i's $\Longrightarrow H_{04}$ with a counterpart formed by interchanging "j" and "k" on the right side of (1.2.10). Thus it turns out that H_{03} does not have a natural tie-up with H_{01} in the sense in which $H_{03}^{(1)}$ has a natural tie-up with $H_{01}^{(1)}$. The tie-up of H_{03} (in this sense) is with H_{04} or its counterpart. However, if we sum both sides of (1.2.10) over i , we should have $1 = q_{*j*} \sum_i q_{i*k}$, whence $q_{*j*} = 1 \Big/ \sum_i q_{i*k}$, which really means that both are pure constants. Thus H_{04} is essentially H_{02}, and (1.2.6) for all sets of a_i's is the same as H_{02} and its counterpart.

Going back to H_{02}, we write its analogue in the form

$$(1.2.11) \qquad H_{05}: \sum_i a_i\, p_{ijk} = q_{**k} \quad .$$

Again, as before, $H_{02} \Longrightarrow H_{05}$ but not conversely. But H_{05} (for all sets of a_i's) $\Longrightarrow H_{02}$. The other remarks made after (1.2.5) would also carry over to this case, covering (1.2.6), (1.2.7) and (1.2.11).

Going back to (1.2.6), another hypothesis which has the same relation to H_{01} as $H_{03}^{(1)}$ has to $H_{01}^{(1)}$ is

$$(1.2.12) \qquad H_{06}: \prod_i p_{ijk}^{a_i} = q_{*j*}\, q_{**k} \quad .$$

Rewriting the right side as $\prod_i q_{ij*}^{a_i}\, q_{i*k}^{a_i}$, we can rewrite (1.2.12) as

$$(1.2.13) \qquad H_{06}: \prod_i \left(p_{ijk}\big/ q_{ij*}\, q_{i*k}\right)^{a_i} = 1 \quad .$$

Now $H_{01} \Longrightarrow H_{06}$ but not conversely, and also H_{06} (for all a_i's) $\Longrightarrow H_{01}$. We notice, (1.2.12) implies that the weighted geometric mean of p_{ijk} (over "i") has the 'no interaction property.' How meaningful this interpretation would be is not very clear and thus (1.2.12) is offered quite tentatively.

1.2.3 <u>The case where "i" and "j" are structured</u>

In this case (assuming a given set of weights b_j's to go with "j") the natural analogues of H_{03}, $H_{03}^{(1)}$ and H_{05} seem to be

$$(1.2.14) \qquad H_{07}: \sum_i a_i\, p_{ijk} = q_{**k} \times \text{an assumed function of } b_j\text{'s}$$

$$= q_{**k}(\lambda + \mu b_j)\ , \text{ say,}$$

or

$$= \lambda_k + \mu_k\, b_j \quad \text{(which is more general),}$$

where λ_k and μ_k are unknown functions of k and λ and μ are unknown constants,

$$(1.2.15) \qquad H_{07}^{(1)}: \sum_i a_i\, p_{ijk} = q_{**k}^{(1)} + \mu b_j \quad ,$$

and

$$(1.2.16) \qquad H_{08}: \sum_i a_i\, p_{ijk} = q_{**k} \quad .$$

The remarks made after (1.2.5) would also carry over here. It may be noted that "j" may be a structured factor in a two-dimensional (jk) design or a concomitant variable in a one-way (k) classification. Another meaningful hypothesis parallel to H_{08} is

$$(1.2.17) \qquad H_{09}: \sum_i a_i \, p_{ijk} = \lambda + \mu b_j \quad .$$

1.2.4 The case where "i", "j" and "k" are all structured

In this case, assuming furthermore a given set of weights c_k's to go with "k", it is possible to set up the same H_{07}, $H_{07}^{(1)}$, H_{08} and H_{09} and also similar ones in which the roles of "j" and "k" are interchanged. However, the more interesting hypothesis would seem to be

$$(1.2.18) \qquad H_{010}: \sum_i a_i \, p_{ijk} = \text{an assumed function of } b_j \text{ and } c_k$$
$$= \lambda + \mu b_j + \nu c_k \text{ , say.}$$

(i) Starting from (1.2.1) as a model we can test this hypothesis or directly one in which $\mu = 0$ and/or $\nu = 0$, or (ii) starting from (1.2.1) together with (1.2.18) as a model we can test the hypothesis that $\mu = 0$ and/or $\nu = 0$.

1.3 A three-way table (ijk) in which "i", "j" and "k" are responses

The probability model is given by

$$(1.3.1) \qquad \Phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}} \quad ,$$

where $\sum_{i,j,k} p_{ijk} = 1$ and $\sum_{i,j,k} n_{ijk} = n$ (fixed) .

1.3.1 <u>The case where "i", "j" and "k" are all unstructured</u>

Consider

$$(1.3.2) \qquad H_{011}: \; p_{ijk} = \frac{p_{iok}\, p_{ojk}}{p_{ook}} \quad ,$$

which can be described as the hypothesis of no partial association between "i" and "j", given "k". There are two equivalent ways to get at (1.3.2). One is to notice that the conditional joint distribution of "i" and "j", given "k", is p_{ijk}/p_{ook} and the conditional marginal distribution of "i", given "k", is p_{iok}/p_{ook}, that of "j", given "k", is p_{ojk}/p_{ook}, which leads to (1.3.2). Another way is to start with the condition

(1.3.3) $p_{ijk}/\, p_{ojk}$ is independent of "j" $= q_{i*k}$, say,

which means that the conditional distribution of "i", given "j" and "k", is independent of "j", and then rewrite (1.3.3) in the form $p_{ijk} = p_{ojk}\, q_{i*k}$. Summing over "j" we have that (1.3.3) is equivalent to (1.3.2). This hypothesis and the appropriate χ^2 test have already been discussed in [33].

We next proceed to the hypothesis that "j" and "k" are independent and so also are "i" and "k", i.e.,

$$(1.3.4) \qquad H_{012}: \; p_{ojk} = p_{ojo}\, p_{ook} \;\text{ and }\; p_{iok} = p_{ioo}\, p_{ook} \,.$$

We note that $H_{011} \cap H_{012} \Longrightarrow$

$$(1.3.5) \qquad H_{013}: \; p_{ijk} = p_{ioo}\, p_{ojo}\, p_{ook} \quad ,$$

which is the hypothesis of over-all independence of "i", "j" and "k". It should also be noticed that $H_{012} \not\Longrightarrow$

(1.3.6) $\qquad H_{014}: \; p_{ijk} = p_{ijo}\, p_{ook}$,

which is the hypothesis that there is no multiple association between "ij" and "k". It can be seen that $H_{014} \Longrightarrow H_{012}$, but not conversely (unlike the multivariate normal case). One could also get to (1.3.6) by starting either from the condition that

(1.3.7) $\qquad p_{ijk}/p_{ijo}$ is independent of "ij" ,

or from the one that

(1.3.8) $\qquad p_{ijk}/p_{ook}$ is independent of "k" .

Another hypothesis which might be of interest is that of two by two independence, or in symbols,

(1.3.9) $\qquad H_{015}: \; p_{ijo} = p_{ioo}\, p_{ojo} , \quad p_{iok} = p_{ioo}\, p_{ook}$

$$\text{and} \quad p_{ojk} = p_{ojo}\, p_{ook} \; .$$

It is well known that $H_{013} \Longrightarrow H_{015}$, but not conversely in general. The hypotheses H_{013} and H_{014} have already been discussed in [33].

1.3.2 The case where "i" is structured

We may consider hypotheses analogous to (1.2.6) and (1.2.7), namely

(1.3.10) $\qquad H_{016}: \; \sum_i a_i\, p_{ijk}/p_{ojk} = q_{*j*}\, q_{**k}$,

or

(1.3.11) $\qquad H_{016}^{(1)}: \; \sum_i a_i\, p_{ijk}/p_{ojk} = q_{*j*}^{(1)} + q_{**k}^{(1)}$.

Next let us consider a hypothesis analogous to H_{011}, using the version of H_{011} given by (1.3.3).

(1.3.12) $\quad H_{017}: \sum_i a_i\, p_{ijk}/p_{ojk}$ is independent of "j" .

It is easily seen that (1.3.12) could be written in the equivalent form

(1.3.13) $\quad H_{017}: \sum_i a_i \left[p_{ijk} - \dfrac{p_{iok}\, p_{ojk}}{p_{ook}} \right] = 0$.

Thus $H_{011} \Longrightarrow H_{017}$ but not conversely. However, (1.3.13) (for all a_i's) $\Longrightarrow H_{011}$.

Also analogous to H_{012} we now have

(1.3.14) $\quad H_{018}: \sum_i a_i\, p_{ijo}/p_{ojo}$ is independent of "j" ,

$\sum_i a_i\, p_{iok}/p_{ook}$ is independent of "k" ,

which is equivalent to

(1.3.15) $\quad H_{018}: \sum_i a_i (p_{ijo} - p_{ioo}\, p_{ojo}) = 0$,

$\sum_i a_i (p_{iok} - p_{ioo}\, p_{ook}) = 0$.

The same remarks are applicable as after (1.3.13). Likewise, analogous to H_{014} we have

(1.3.16) $\quad H_{019}: \sum_i a_i\, p_{ijk}/p_{ojk}$ is independent of "jk" ,

which is equivalent to

(1.3.17) $\quad H_{019}: \sum_i a_i (p_{ijk} - p_{ioo}\, p_{ojk}) = 0$.

The same remarks are applicable again as after (1.3.13).

1.3.3 The case where "i" and "j" are structured

If, for concreteness, we fix our attention on "i", then the hypothesis of independence of "i" with respect to "j" and "k" of the types (1.3.12), (1.3.14) and (1.3.16) (as well as the tests) remain as before. However, we have other interesting possibilities. Assuming a set of weights b_j's to go with "j", we can write, for example, a hypothesis related to H_{017} in the form

$$(1.3.18) \qquad H_{020}: \sum_i a_i \, p_{ijk}/p_{ojk} = \text{an assumed function of } b_j \times \text{an assumed function of "k"}$$
$$= (\lambda + \mu b_j) q_{**k} \text{ (say)} ,$$
$$\text{or} \quad = \lambda_k + \mu_k b_j \text{ (which is more general)} .$$

We can test this in the spirit of testing for linearity of regression; or assuming this as a model, we can test the hypothesis that $\mu = 0$ or $\mu_k = 0$.

A hypothesis naturally related to H_{018} would seem to be

$$(1.3.19) \qquad H_{021}: \sum_i a_i \, p_{ijo}/p_{ojo} = \text{an assumed function of } b_j$$
$$= \lambda + \mu b_j \text{ (say)} ,$$
$$\sum_i a_i \, p_{iok}/p_{ook} \text{ is independent of "k"} .$$

1.3.4 The case where "i", "j" and "k" are all structured

Here we have further interesting possibilities. Assuming a set of weights c_k's to go with "k", we have, for example, a hypothesis related to H_{019} in the form

(1.3.20) H_{022}: $\sum_i a_i \, p_{ijk}/p_{ojk}$ = an assumed function of b_j and c_k

$$= \lambda + \mu b_j + \nu c_k \quad \text{(say)} .$$

We can test this hypothesis in the spirit of testing for linearity of regression, treating λ, μ and ν as unknown constants. Or assuming (1.3.20) as a model we can also test the hypothesis that $\mu = 0$ or that $\nu = 0$. Likewise, the one related to (1.3.14) seems to be

(1.3.21) H_{023}: $\sum_i a_i \, p_{ijo}/p_{ojo}$ = an assumed function of b_j

$$= \lambda_1 + \mu b_j \quad \text{(say)} ,$$

$\sum_i a_i \, p_{iok}/p_{ook}$ = an assumed function of c_k

$$= \lambda_2 + \nu c_k \quad \text{(say)} .$$

This also may be tested in the spirit of testing for linearity of regression, or, assuming this as a model we can test for, say, $\mu = 0$ and/or $\nu = 0$.

It will be seen that in this study of association in this categorical set up we have been working in the spirit of "regression" rather than in the spirit of "correlation." We have not been trying to use a single measure for any of the various types of association. Such a single measure seems to have a limited use in this categorical set up. However, such single measures (which come out as the noncentrality parameters in the asymptotic power functions of the respective tests for independence) have already been discussed in [8].

1.4 A four-way ($ijk\ell$) table in which "i", "j", "k" and "ℓ" are all responses

$$\Phi = \frac{n!}{\prod_{i,j,k,\ell} n_{ijk\ell}!} \prod_{i,j,k,\ell} p_{ijk\ell}^{n_{ijk\ell}} , \tag{1.4.1}$$

where $\sum_{i,j,k,\ell} p_{ijk\ell} = 1$ and $\sum_{i,j,k,\ell} n_{ijk\ell} = n$ (fixed).

There is much in common between this four-variate case and the three-variate case discussed in 1.3. However, the four-variate case presents certain new features and we shall state some for purposes of illustration.

1.4.1 The case where all are structured

$$H_{024}: \quad p_{ijko} = p_{ioko}\, p_{ojko}/p_{ooko} \quad , \tag{1.4.2}$$

$$p_{ijo\ell} = p_{ioo\ell}\, p_{ojo\ell}/p_{ooo\ell} \quad .$$

1.4.2 The case where "i", "j" and "k" are structured

$$H_{025}: \sum_i a_i\, p_{ijko}/p_{ojko} = \text{an assumed function of } b_j \text{ and } c_k \tag{1.4.3}$$

$$= \lambda + \mu b_j + \nu c_k \quad \text{(say)} ,$$

$$\sum_i a_i\, p_{ijo\ell}/p_{ojo\ell} = q_{***\ell} \times \text{an assumed function of } b_j$$

$$= q_{***\ell}(\lambda_1 + \mu_1 b_j) \quad \text{(say)} .$$

1.5 A five-way $(ijk\ell m)$ table in which "i" and "j" are responses and "k", "ℓ" and "m" are factors

$$(1.5.1) \qquad \Phi = \prod_{k,\ell,m} \left[\frac{n_{ook\ell m}!}{\prod_{i,j} n_{ijk\ell m}!} \prod_{i,j} p_{ijk\ell m}^{n_{ijk\ell m}} \right] ,$$

where $\sum_{i,j} p_{ijk\ell m} \equiv p_{ook\ell m} = 1$ and $\sum_{i,j} n_{ijk\ell m} = n_{ook\ell m}$ (fixed).

There is much in common between this case and the corresponding three-way case discussed in 1.2. We shall discuss some new features.

1.5.1 The case where all are unstructured

We may consider the hypothesis

$$(1.5.2) \qquad H_{026}: \quad p_{iok\ell m} = q_{i**\ell m}\, q_{i*k*m}\, q_{i*k\ell *} ,$$
$$p_{ojk\ell m} = q_{*j*\ell m}\, q_{*jk*m}\, q_{*jk\ell *} ,$$

which may be interpreted as the hypothesis of no three-factor interaction (in the multiplicative sense) or a similar hypothesis $H_{026}^{(1)}$ in the additive set up. Similarly, we may consider

$$(1.5.3) \qquad H_{027}: \quad p_{iok\ell m} = q_{i*k**}\, q_{i**\ell *}\, q_{i***m} ,$$
$$p_{ojk\ell m} = q_{*jk**}\, q_{*j*\ell *}\, q_{*j**m} ,$$

which may be interpreted as the hypothesis of no two-factor interaction and a similar hypothesis $H_{027}^{(1)}$, and finally

(1.5.4) H_{028}: the right side is independent of one or more of the factors "k", "ℓ" and "m" , which may be interpreted as the hypothesis of no corresponding main-effects and a similar hypothesis $H_{028}^{(1)}$.

One may start from (1.5.1) and test H_{026} or $H_{026}^{(1)}$ as a hypothesis, or from H_{026} or $H_{026}^{(1)}$ as a model and test H_{027} or $H_{027}^{(1)}$ as a hypothesis, or from H_{027} or $H_{027}^{(1)}$ as a model and test H_{028} or $H_{028}^{(1)}$ as a hypothesis. There are various intermediate cases.

1.5.2 The case where "i" and "j" are structured

The analogues of the hypotheses in the previous case seem to be as follows:

$$H_{029}: \sum_i a_i \, p_{iok\ell m} = q^{(1)}_{***\ell m} \, q^{(1)}_{**k*m} \, q^{(1)}_{**k\ell *} \quad , \tag{1.5.5}$$

$$\sum_j b_j \, p_{ojk\ell m} = q^{(2)}_{***\ell m} \, q^{(2)}_{**k*m} \, q^{(2)}_{**k\ell *} \quad ,$$

or $H_{029}^{(1)}$ with the additive set up,

$$H_{030}^{(1)}: \sum_i a_i \, p_{iok\ell m} = q^{(1)}_{**k**} + q^{(1)}_{***\ell *} + q^{(1)}_{****m} \quad , \tag{1.5.6}$$

$$\sum_j b_j \, p_{ojk\ell m} = q^{(2)}_{**k**} + q^{(2)}_{***\ell *} + q^{(2)}_{****m} \quad ,$$

or H_{030} with the multiplicative set up, and finally

(1.5.7) H_{031}: the right side is independent of one or more of "k", "ℓ" and "m" ,

or $H_{031}^{(1)}$ with the additive set up.

Finally we consider

1.5.3 The case where "i", "j" and "k" are structured

The hypotheses of interest might be, for example, as follows:

(1.5.8) H_{o32}: $\sum_i a_i \, p_{iok\ell m} = \lambda^{(1)}_{\ell m} + \mu^{(1)}_{\ell m} c_k$ (say) ,

$\sum_j b_j \, p_{ojk\ell m} = \lambda^{(2)}_{\ell m} + \mu^{(2)}_{\ell m} c_k$ (say) ,

or

(1.5.9) H_{o33}: $\sum_i a_i \, p_{iok\ell m} = \lambda^{(1)}_{\ell m} + \mu^{(1)} c_k$ (say) ,

$\sum_j b_j \, p_{ojk\ell m} = \lambda^{(2)}_{\ell m} + \mu^{(2)} c_k$ (say) ,

(1.5.10) H_{o34}: $\sum_i a_i \, p_{iok\ell m} = \lambda^{(1)}_{\ell *} + \lambda^{(1)}_{* m} + \mu^{(1)} c_k$,

$\sum_j b_j \, p_{ojk\ell m} = \lambda^{(2)}_{\ell *} + \lambda^{(2)}_{* m} + \mu^{(2)} c_k$,

and finally

(1.5.11) H_{o35}: the right side is independent of one or more of "ℓ" and "m" with or without μ's being zero.

CHAPTER II

ON SOME BASIC THEOREMS OF NEYMAN ON χ_1^2 AND "LINEARIZATION"

2.1 Introduction

Let

$$\Phi = \prod_j \left[\frac{n_{oj}!}{\prod_i n_{ij}!} \prod_i p_{ij}^{n_{ij}} \right] \tag{2.1.1}$$

denote a product multinomial distribution so that $\sum_i p_{ij} = 1$ and $\sum_i n_{ij} \equiv n_{oj}$ is fixed. If a hypothesis H_0 is given in the form of certain constraints on p_{ij}'s, then the large sample test of H_0 under (2.1.1) for the model is in terms of a statistic given by

$$\sum_{i,j} \frac{(n_{ij} - n_{oj}\,\hat{p}_{ij})^2}{n_{oj}\,\hat{p}_{ij}} \tag{2.1.2}$$

in which $\hat{p}_{ij}$'s are estimates of p_{ij}'s which maximize Φ subject to $\sum_i p_{ij} = 1$ and to the further constraints on p_{ij}'s that define the hypothesis. This statistic, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, is distributed as a χ^2 with degrees of freedom equal to the number of independent constraints on p_{ij}'s that define the hypothesis. It may be observed that instead of the maximum likelihood estimates of p_{ij}'s, one might as well consider any set of estimates belonging to the broader class of BAN estimates, as shown by Neyman [22]. Likewise, instead of the statistic (2.1.2) one might as well consider the slightly different one known as χ_1^2 . If we introduce some additional constraints on p_{ij}'s and thus define a new hypothesis $H_0^* \subset H_0$, then the test of H_0^*, under (2.1.1) for the model, is in terms of a statistic given by

$$(2.1.3) \qquad \sum_{i,j} \frac{(n_{ij} - n_{oj}\, \hat{\hat{p}}_{ij})^2}{n_{oj}\, \hat{\hat{p}}_{ij}}$$

which, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, has the χ^2 distribution with degrees of freedom equal to the number of independent constraints on p_{ij}'s that define H_0^*. However, if we want to test H_0^* under H_0 for the model, then the test will be given in terms of a statistic

$$(2.1.4) \qquad \sum_{i,j} \frac{(n_{oj} - n_{ij}\, \hat{\hat{p}}_{ij})^2}{n_{oj}\, \hat{\hat{p}}_{ij}} - \sum_{i,j} \frac{(n_{oj} - n_{ij}\, \hat{p}_{ij})^2}{n_{oj}\, \hat{p}_{ij}}$$

which, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, has the χ^2 distribution with degrees of freedom equal to the number of additional independent constraints on the p_{ij}'s that define H_0^* under H_0 for the model. As before, any set of BAN estimates may be used, and similarly, the slightly different χ_1^2 statistics may be used.

Neyman shows that the relevant equations (namely maximum likelihood or minimum χ^2 or minimum χ_1^2) have a system of solutions which are BAN estimates of parameters. On the other hand, Cramer [7] shows, in the simplest case, that the maximum likelihood equations (which are the same as modified minimum χ^2 equations) have a unique system of consistent solutions, and then the χ^2 statistic based on this solution has an asymptotic χ^2 distribution. Mitra [19] and Ogawa [23] extend it to more general cases. We shall prove along Cramer's lines the theorem, in the simplest case, for the minimum χ_1^2 estimates and the χ_1^2 statistic. It could then be extended to more general cases. Mitra [19] and Diamond [8] have defined and obtained asymptotic power of the χ^2 tests. The same thing could be done for χ_1^2 tests.

2.2 <u>Theorem 2.2</u>

Suppose that we are given r functions $p_1(\underline{\alpha}), \ldots, p_r(\underline{\alpha})$ of $s < r$ variables $\underline{\alpha}' = (\alpha_1, \ldots, \alpha_s)$ such that, for all points of a nondegenerate interval A in the s-dimensional space of $\underline{\alpha}$, the functions $p_i(\underline{\alpha})$ satisfy the following conditions:

(a) $\sum_{i=1}^{r} p_i(\underline{\alpha}) = 1$.

(b) $p_i(\underline{\alpha}) \geq c^2 > 0$ for all i .

(c) Every $p_i(\underline{\alpha})$ has continuous derivatives $\dfrac{\partial p_i}{\partial \alpha_j}$ and $\dfrac{\partial^2 p_i}{\partial \alpha_j \partial \alpha_k}$.

(d) The matrix $D = \left(\dfrac{\partial p_i}{\partial \alpha_j}\right)_{\substack{i=1,\ldots,r \\ j=1,\ldots,s}}$ is of rank s .

Let the possible results of a certain random experiment E be divided into r mutually exclusive groups and suppose that the probability of obtaining a result belonging to the i-th group is $p_i^0 = p_i(\underline{\alpha}_0)$, where $\underline{\alpha}_0' = (\alpha_1^0, \ldots, \alpha_s^0)$ is an inner point of A. Let v_i denote the number of results belonging to the i-th group, in a sequence of n repetitions of E, so that $\sum_{i=1}^{r} v_i = n$. It is assumed that none of the v_i's is equal to zero. Then, the equations

$$(2.2.1) \qquad \sum_{i=1}^{r} \frac{(v_i - np_i)}{v_i} \frac{\partial p_i}{\partial \alpha_j} = 0 \quad , \quad j = 1,2,\ldots,s \quad ,$$

of the minimum χ_1^2 method, have exactly one solution $\underline{\hat{\alpha}}' = (\hat{\alpha}_1, \ldots, \hat{\alpha}_s)$ such that $\underline{\hat{\alpha}}$ converges in probability to $\underline{\alpha}_0$ as $n \to \infty$, and

$$(2.2.2) \qquad \chi_1^2 = \sum_{i=1}^{r} \frac{[v_i - np_i(\underline{\hat{\alpha}})]^2}{v_i}$$

is, in the limit as $n \to \infty$, distributed as a χ^2 with $r-s-1$ d.f.

Proof: Let $\frac{\partial p_i}{\partial \alpha_j} = p_{ij}$ and $\left(\frac{\partial p_i}{\partial \alpha_j}\right)_{\underline{\alpha}=\underline{\alpha}_0} = p_{ijo}$.

The equations (2.2.1) can then be written as

$$\sum_i \frac{(v_i - np_i^0)}{v_i} (p_{ij} - p_{ijo}) + \sum_i \frac{v_i - np_i^0}{v_i} p_{ijo}$$

$$- n \sum_i \frac{(p_i - p_i^0)}{v_i} (p_{ij} - p_{ijo}) - n \sum_i \frac{(p_i - p_i^0)}{v_i} p_{ijo} = 0 ,$$

$$j = 1,2,\ldots,s .$$

Therefore,

$$(2.2.3)\quad \sum_k (\alpha_k - \alpha_k^0) \sum_i \frac{1}{p_i^0}\, p_{ijo} p_{iko} = \sum_i \frac{v_i - np_i^0}{np_i^0}\, p_{ijo} + w_j(\underline{\alpha}) \;,$$

where

$$(2.2.4)\quad w_j(\underline{\alpha}) = \sum_i \frac{v_i - np_i^0}{v_i}\,(p_{ij} - p_{ijo}) - n \sum_i \frac{p_i - p_i^0}{v_i}(p_{ij} - p_{ijo})$$

$$- \sum_i \frac{p_{ijo}}{p_i^0} \left\{ \frac{np_i^0 - v_i}{v_i}(p_i - p_i^0) + (p_i - p_i^0) - \sum_k (\alpha_k - \alpha_k^0) p_{iko} \right\}$$

$$- \sum_i \frac{(v_i - np_i^0)^2}{nv_i\, p_i^0}\, p_{ijo} \;, \qquad j = 1, 2, \ldots, s \;.$$

Let $$B = \left[\frac{1}{\sqrt{p_i^0}}\, p_{ijo}\right]_{r\times s} \;.$$

Then by condition (d) B is of rank s. Let

$$(2.2.5)\quad x_i = \frac{v_i - np_i^0}{\sqrt{np_i^0}} \;, \quad \underline{x}' = (x_1, x_2, \ldots, x_r) \;,$$

and $$\underline{w}'(\underline{\alpha}) = [w_1(\underline{\alpha}), \ldots, w_s(\underline{\alpha})] \;.$$

Then the equations (2.2.3) can be written as

$$(2.2.6)\quad B'B(\underline{\alpha} - \underline{\alpha}_0) = n^{-\frac{1}{2}} B'\underline{x} + \underline{w}(\underline{\alpha}) \;,$$

so that

$$(2.2.7)\quad \underline{\alpha} = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1} B'\underline{x} + (B'B)^{-1}\,\underline{w}(\underline{\alpha}) \;.$$

Then, following Cramer, we have with a probability greater than $1 - \frac{1}{\lambda^2}$,

$$(2.2.8)\quad |v_i - np_i^0| < \lambda\sqrt{n} \quad \text{for all } i = 1, 2, \ldots, r \;.$$

Until further notice, we shall assume that v_i satisfy (2.2.8).

We shall let λ denote a function of n such that $\lambda \to \infty$ as $n \to \infty$, while $\frac{\lambda^2}{\sqrt{n}} \to 0$ as $n \to \infty$ [e.g. $\lambda = n^q$, $0 < q < \frac{1}{4}$]. All results obtained will then be true with a probability $\geq 1 - \frac{1}{\lambda^2}$, and which $\to 1$ as $n \to \infty$.

By condition (b),

$$(2.2.9) \qquad |x_i| < \frac{\lambda}{c}, \qquad i = 1,2,\ldots,r .$$

Now, for two inner points $\underline{\alpha}_1$ and $\underline{\alpha}_2$ of A, we have

$$w_j(\underline{\alpha}_1) - w_j(\underline{\alpha}_2) = \sum_i \frac{v_i - np_i^0}{v_i}(p_{ij1} - p_{ij2}) - n\sum_i \frac{p_i^1 - p_i^2}{v_i}(p_{ij1} - p_{ij0})$$

$$- n\sum_i \frac{p_i^2 - p_i^0}{v_i}(p_{ij1} - p_{ij2}) - \sum_i \frac{p_{ijo}}{p_i^0}\left\{p_i^1 - p_i^2 - \sum_k p_{iko}(\alpha_k^1 - \alpha_k^2)\right.$$

$$\left.- \frac{v_i - np_i^0}{v_i}(p_i^1 - p_i^2)\right\}, \qquad j = 1, 2, \ldots, s .$$

Now from (2.2.5), (b) and (2.2.9)

$$v_i = np_i^0 + x_i\sqrt{np_i^0} = np_i^0\left[1 + \frac{x_i}{\sqrt{np_i^0}}\right] > np_i^0\left[1 - \frac{\lambda}{c^2\sqrt{n}}\right],$$

so that

$$(2.2.10) \qquad \frac{n}{v_i} < \frac{1}{p_i^0\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)} < \frac{1}{c^2\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)},$$

and

$$\left|\frac{v_i - np_i^0}{v_i}\right| < \frac{\lambda}{c^2\sqrt{n}\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)}, \qquad i = 1,2,\ldots,r .$$

Therefore,

$$c^2\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)\left|w_j(\underline{\alpha}_1) - w_j(\underline{\alpha}_2)\right| \le \frac{\lambda}{\sqrt{n}}\sum_i\left|p_{ij1} - p_{ij2}\right| + \sum_i\left|p_i^1 - p_i^2\right| \times$$

$$\left|p_{ij1} - p_{ijo}\right| + \sum_i\left|p_i^2 - p_i^0\right|\left|p_{ij1} - p_{ij2}\right| + \frac{\lambda}{\sqrt{n}}\sum_i \frac{|p_{ijo}|}{p_i^0}\left|p_i^1 - p_i^2\right| +$$

$$c^2\sum_i \frac{|p_{ijo}|}{p_i^0}\left|p_i^1 - p_i^2 - \sum_k p_{iko}(\alpha_k^1 - \alpha_k^2)\right| , \quad j = 1,2,\ldots,s \ .$$

In view of conditions (b) and (c), $|p_{ij1} - p_{ij2}| \le k_{1ij}|\underline{\alpha}_1 - \underline{\alpha}_2|$, where $|\underline{\alpha}_1 - \underline{\alpha}_2|$ is the distance in the s-space and k_{1ij} is a constant. Let $k_{1j} = \sum_i k_{1ij}$ and $k_1 = \max(k_{1j})$. Then, $\sum_i |p_{ij1} - p_{ij2}| \le k_1|\underline{\alpha}_1 - \underline{\alpha}_2|$ for all j. Similarly,

$$|p_i^1 - p_i^2| \le k_2|\underline{\alpha}_1 - \underline{\alpha}_2| \quad \text{for all } i ,$$

$$\sum_i |p_{ij1} - p_{ijo}| \le k_3|\underline{\alpha}_1 - \underline{\alpha}_0| \quad \text{for all } j ,$$

$$|p_i^2 - p_i^0| \le k_4|\underline{\alpha}_2 - \underline{\alpha}_0| \quad \text{for all } i ,$$

$$\sum_i \frac{1}{p_i^0}|p_{ijo}| < k_5 \quad \text{for all } i,j \text{ , and}$$

$$\left|p_i^1 - p_i^2 - \sum_k p_{iko}(\alpha_k^1 - \alpha_k^2)\right| = \left|\sum_k (p_{ik2} - p_{iko})(\alpha_k^1 - \alpha_k^2) + \frac{1}{2}\sum_k\sum_{k'} \frac{\partial^2 p_i}{\partial\alpha_k\,\partial\alpha_{k'}}(\alpha_k^1 - \alpha_k^2)(\alpha_{k'}^1 - \alpha_{k'}^2)\right|$$

$$\le k_6|\underline{\alpha}_2 - \underline{\alpha}_1||\underline{\alpha}_2 - \underline{\alpha}_0| + k_7|\underline{\alpha}_2 - \underline{\alpha}_1|^2 \ .$$

Also $|\alpha_k^1 - \alpha_k^2| \le |\underline{\alpha}_1 - \underline{\alpha}_2|$ for all k.

Therefore,

$$|w_j(\underline{\alpha}_1) - w_j(\underline{\alpha}_2)| \leq \frac{k_0|\underline{\alpha}_1 - \underline{\alpha}_2|}{1 - \frac{\lambda}{c^2\sqrt{n}}} \left\{\frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_1 - \underline{\alpha}_0| + |\underline{\alpha}_2 - \underline{\alpha}_0|\right\} \quad ,$$

$$j = 1,2,\ldots,s \; .$$

Hence,

$$(2.2.11) \quad |\underline{w}(\underline{\alpha}_1) - \underline{w}(\underline{\alpha}_2)| \leq \frac{k|\underline{\alpha}_1 - \underline{\alpha}_2|}{1 - \frac{\lambda}{c^2\sqrt{n}}} \left\{\frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_1 - \underline{\alpha}_0|\,|\underline{\alpha}_2 - \underline{\alpha}_0|\right\} \quad .$$

We now define a sequence of vectors $\underline{\alpha}_v$ for $v = 1,2,\ldots$ by

$$(2.2.12) \qquad \underline{\alpha}_v = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{w}(\underline{\alpha}_{v-1}) \quad ,$$

and we propose to show that $\{\underline{\alpha}_v\} \to$ a definite limit $\underline{\alpha}$ which is then evidently a solution of (2.2.7). It will be seen from the definition of $\underline{w}(\underline{\alpha})$ that

$$w_j(\underline{\alpha}_0) = -\sum_i \frac{(v_i - np_i^0)^2}{nv_i p_i^0}\, p_{ijo} \; , \quad j = 1,2,\ldots,s. \text{ Also,}$$

$$(2.2.13) \qquad \underline{\alpha}_1 - \underline{\alpha}_0 = n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{w}(\underline{\alpha}_0)$$

$$\text{and} \qquad \underline{\alpha}_{v+1} - \underline{\alpha}_v = (B'B)^{-1}[\underline{w}(\underline{\alpha}_v) - \underline{w}(\underline{\alpha}_{v-1})] \quad .$$

The matrices $(B'B)^{-1}B'$ and $(B'B)^{-1}$ are independent of n. Denoting by g an upper bound of absolute values of the elements, we have from (2.2.8), (2.2.10) and (c)

$$|p_{ijo}|\, \frac{(v_i - np_i^0)^2}{nv_i p_i^0} < k\, \frac{\lambda^2}{n\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)} \quad \text{for all } (i,j),$$

where k is some constant independent of n. Hence, from (2.2.9) and (2.2.13) we have

$$|\alpha_j^1 - \alpha_j^0| \leq \frac{\lambda}{\sqrt{n}}\frac{rg}{c} + g\,\frac{s\ k\ \lambda^2}{n\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)} , \quad j = 1,2,\ldots,s ,$$

so that

(2.2.14) $$|\underline{\alpha}_1 - \underline{\alpha}_0| \leq k' \frac{\lambda}{\sqrt{n}}\left[1 + \frac{\lambda}{\sqrt{n}\left(1 - \frac{\lambda}{c^2\sqrt{n}}\right)}\right]$$, where k' is a suitable constant. Similarly, from (2.2.11) and (2.2.13), we have

(2.2.15) $$|\underline{\alpha}_{v+1} - \underline{\alpha}_v| \leq \frac{k'|\underline{\alpha}_v - \underline{\alpha}_{v-1}|}{1 - \frac{\lambda}{c^2\sqrt{n}}}\left\{\frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_v - \underline{\alpha}_0| + |\underline{\alpha}_{v-1} - \underline{\alpha}_0|\right\},$$

where k' is independent of n and v. Since $\frac{\lambda}{\sqrt{n}} \to 0$ as $n \to \infty$, for sufficiently large values of n, $1 - \frac{\lambda}{c^2\sqrt{n}} > \frac{1}{2}$,

so that $|\underline{\alpha}_1 - \underline{\alpha}_0| \leq k^* \frac{\lambda}{\sqrt{n}}$, and

(2.2.16)
$$|\underline{\alpha}_{v+1} - \underline{\alpha}_v| \leq 2k'|\underline{\alpha}_v - \underline{\alpha}_{v-1}|\left\{\frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_v - \underline{\alpha}_0| + |\underline{\alpha}_{v-1} - \underline{\alpha}_0|\right\} .$$

Then, following Cramer, for large n

(2.2.17) $$|\underline{\alpha}_{v+1} - \underline{\alpha}_v| \leq k^*[2k'(4k^* + 1)]^v\left(\frac{\lambda}{\sqrt{n}}\right)^{v+1} .$$

The infinite series

$$\underline{\alpha}_0 + (\underline{\alpha}_1 - \underline{\alpha}_0) + (\underline{\alpha}_2 - \underline{\alpha}_1) + \ldots$$

converges absolutely for sufficiently large n, and if we define $\underline{\hat{\alpha}}$ by

(2.2.18) $$\underline{\hat{\alpha}} = \underline{\alpha}_0 + (\underline{\alpha}_1 - \underline{\alpha}_0) + (\underline{\alpha}_2 - \underline{\alpha}_1) + \ldots$$

then $\underline{\hat{\alpha}}$ satisfies (2.2.7) and hence (2.2.1). It follows from (2.2.17) that $\underline{\hat{\alpha}} \xrightarrow{(p)} \underline{\alpha}_0$ as $n \to \infty$. Proving uniqueness along Cramer's lines, we thus have a unique solution $\underline{\hat{\alpha}}$ of (2.2.1), which converges in probability to $\underline{\alpha}_0$.

Still assuming (2.2.8), we have, from (2.2.13),

$$(B'B)^{-1}\underline{w}(\underline{\hat{\alpha}}) = \underline{\hat{\alpha}} - \underline{\alpha}_1 + (B'B)^{-1}\underline{w}(\underline{\alpha}_0)$$

$$= (B'B)^{-1}\underline{w}(\underline{\alpha}_0) + (\underline{\alpha}_2 - \underline{\alpha}_1) + (\underline{\alpha}_3 - \underline{\alpha}_2) + \ldots .$$

As seen already, every component of $(B'B)^{-1}\underline{w}(\underline{\alpha}_0)$ is, in absolute value, $< 2k\frac{\lambda^2}{n}$ for large n and therefore, from (2.2.17), we see that every component of $(B'B)^{-1}\underline{w}(\underline{\hat{\alpha}})$ is, in absolute value, $< M\frac{\lambda^2}{n}$, so that (2.2.7) may be written as

(2.2.19) $$\underline{\hat{\alpha}} = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + M'\frac{\lambda^2}{n}\underline{\theta}_1 ,$$

where M' is a constant and $\underline{\theta}_1' = (\theta_1, \ldots, \theta_s)$ such that $|\theta_j| \leq 1$, $j = 1,2,\ldots,s$. Consider now

(2.2.20) $$y_i = \frac{v_i - np_i(\underline{\hat{\alpha}})}{\sqrt{v_i}} , \quad i = 1,2,\ldots,r ,$$

so that $\chi_1^2 = \sum_{i=1}^{r} y_i^2$. Then

$$y_i = \frac{x_i - \sqrt{\frac{n}{p_i^0}}\,[p_i(\underline{\hat{\alpha}}) - p_i^0]}{\sqrt{1 + \frac{x_i}{\sqrt{np_i^0}}}} .$$

Now,

$$\sqrt{n}[p_i(\underline{\hat{\alpha}}) - p_i^0] = \sqrt{n}\sum_j p_{ijo}(\hat{\alpha}_j - \alpha_j^0) + \frac{\sqrt{n}}{2}\sum_j\sum_{j'}\frac{\partial^2 p_i}{\partial\alpha_j\partial\alpha_{j'}}(\hat{\alpha}_j - \alpha_j^0)(\hat{\alpha}_{j'} - \alpha_{j'}^0),$$

and from (2.2.19) $\hat{\alpha}_j - \alpha_j^o = O\left(\frac{\lambda}{\sqrt{n}}\right)$, so that

$$\sqrt{n}[p_i(\hat{\underline{\alpha}}) - p_i^o] = \sqrt{n}\sum_j p_{ijo}(\hat{\alpha}_j - \alpha_j^o) + O\left(\frac{\lambda^2}{n}\right) \quad .$$

Hence

$$y_i = \left\{x_i - \sqrt{\frac{n}{p_i^o}} \sum_j p_{ijo}(\hat{\alpha}_j - \alpha_j^o) + O\left(\frac{\lambda^2}{n}\right)\right\} \left\{1 - \frac{x_i}{2\sqrt{np_i^o}} + O\left(\frac{\lambda^2}{n}\right)\right\}$$

$$= x_i - \sqrt{\frac{n}{p_i^o}} \sum_j p_{ijo}(\hat{\alpha}_j - \alpha_j^o) + O\left(\frac{\lambda^2}{n}\right) \quad , \quad i = 1,2,\ldots,r \; .$$

Thus,

$$\underline{y} = \underline{x} - \sqrt{n}\, B(\hat{\underline{\alpha}} - \underline{\alpha}_0) + k'' \frac{\lambda^2}{\sqrt{n}} \underline{\theta}_2$$

$$= \underline{x} - B(B'B)^{-1}B'\underline{x} + M_0 \frac{\lambda^2}{\sqrt{n}} \underline{\theta}$$

$$(2.2.21) \qquad = [I - B(B'B)^{-1}B']\underline{x} + M_0 \frac{\lambda^2}{\sqrt{n}} \underline{\theta} \quad ,$$

where k's and M's are independent of n and $\underline{\theta}$'s stand for vectors such that $|\theta_i| \leq 1$, $i = 1,2,\ldots,r$. From this point on the rest of the proof goes through along Cramer's lines.

2.3 Linearization

It will be noticed that Cramer's theorem and its generalizations, as well as analogous theorems on χ_1^2, are of the nature of pure existence theorems. They prove the existence of a particular system of solutions for the minimizing equations, for which the theorem stated is true. But neither of them says how to isolate this particular system of solutions. When the equations concerned have just one real solution, there is no problem. However, when there are more than one such system, a theorem due to Wald [36] and Ogawa [23] says that the solution system that maximizes the likelihood in the arithmetical sense is the consistent one.

In many situations, the hypothesis can be equivalently expressed in terms of the constraints on p's, say, for example

$$F_t(\underline{p}) = 0 \quad , \quad t = 1,2,\ldots,\mu \quad . \tag{2.3.1}$$

In the particular case, when these constraints are linear in p's, the method of minimum χ_1^2 reduces the problem to the solution of a system of linear equations and hence is more convenient. When these constraints are not linear, Neyman [22] has suggested the following "linearization" technique. Taylor's formula is applied to obtain the expansion of $F_t(\underline{p})$ about the point $\underline{p} = \underline{q}$, where q's are sample proportions. Thus,

$$F_t(\underline{p}) = F_t^*(\underline{q},\underline{p}) + \frac{1}{2}\sum_i\sum_j c_{tij}(p_i - q_i)(p_j - q_j) \quad , \tag{2.3.2}$$

where

$$F_t^*(\underline{q},\underline{p}) = F_t(\underline{q}) + \sum_i b_{ti}(p_i - q_i) \quad . \tag{2.3.3}$$

Here b_{ti} represents the partial derivative of $F_t(p)$ with respect to p_i taken at $\underline{p} = \underline{q}$. Thus b_{ti} does not depend upon the p's, so that $F_t^*(\underline{q},\underline{p})$ is a linear function of the p's. On the other hand, the coefficients c_{tij} are functions of both the p's and the q's. Neyman considers two models for minimizing the "generalized distance" between $\underline{p}$ and $\underline{q}$. (i) Under the first model, the minimization is effected with respect to such variation of the p's as is consistent with (2.3.1) and (ii) under the second model, the minimization is effected with respect to such variation of the p's as is consistent with

$$F_t^*(\underline{p}) = 0 \quad , \quad t = 1,2,\ldots,\mu \, . \tag{2.3.4}$$

He then proves that if the generalized distance is such that its minimization under the restrictions (2.3.1) leads to BAN estimate of $\underline{p}^0$, then the minimization of the same distance under the conditions (2.3.4) also leads to the BAN estimate of $\underline{p}^0$. In particular, the minimization of χ_1^2 under the conditions (2.3.4) leads to the BAN estimate of the $\underline{p}^0$ and thus the problem is reduced to the solution of a system of linear equations.

Some steps in Neyman's proof are not very clear. He says, for example, "In fact, the numbers p_k^0 satisfy restrictions (2.3.1) and (2.3.4). . . ." We shall give an independent direct proof for the simple case and this could be easily generalized to the product multinomial situation.

2.4 Theorem 2.4

Suppose that we are given $\underline{p}' = (p_1, p_2, \ldots, p_r)$ such that, for all points of a nondegenerate interval A in the r-dimensional space of $\underline{p}$, p_i's satisfy the following conditions:

(a) $\sum_{i=1}^{r} p_i = 1$.

(2.4.1) (b) In addition, $F_t(\underline{p}) = 0$, $t = 1,2,\ldots,\mu$, where F's are independent functions of $\underline{p}$.

(c) $p_i \geq c^2 > 0$ for all i .

(d) Every $F_t(\underline{p})$ has continuous derivatives

$$\frac{\partial F_t}{\partial p_i} \quad \text{and} \quad \frac{\partial^2 F_t}{\partial p_i \partial p_j} \quad .$$

Let the possible results of a certain random experiment E be divided into r mutually exclusive groups and suppose that the probability of obtaining a result belonging to the i-th group is p_i^0 , where $\underline{p}_0' = (p_1^0, \ldots, p_r^0)$ is an inner point of A such that $F_t(\underline{p}_0) = 0$. Let v_i denote the number of results belonging to the i-th group, which occur in a sequence of n repetitions of E , so that $\sum_{i=1}^{r} v_i = n$.

Then, (i) the equations minimizing χ_1^2 with respect to such variation of the p's as is consistent with (a), (b*), (c), and (d), where

$$(2.4.2) \quad (b^*) \quad F_t^*(\underline{p}) = F_t(\underline{q}) + \sum_i b_{ti}(p_i - q_i) = 0 \quad ,$$

$$b_{ti} = \left.\frac{\partial F_t}{\partial p_i}\right|_{\underline{p}=\underline{q}} \qquad \text{and} \qquad q_i = \frac{v_i}{n}$$

have a unique solution $\hat{\hat{\underline{p}}}$, (ii) $\hat{\hat{\underline{p}}}$ is a consistent estimate of $\underline{p}_0$, and (iii) $\sum_{i=1}^{r} \frac{(v_i - n\hat{\hat{p}}_i)^2}{v_i}$ is, in the limit as $n \to \infty$, distributed as a χ^2 with μ degrees of freedom.

Remark: What this theorem does, in effect, is to offer a test for the original hypothesis $F_t(\underline{p}) = 0$ $(t = 1,2,\ldots,\mu)$ in terms of a χ_1^2 in which the estimates for the p's are obtained not from the minimization of χ_1^2 under the original constraints but under the constraints (2.4.2).

Proof: Under constraints (2.4.1) on $\underline{p}$, eliminating μ p's we have $r - 1 - \mu = s$ independent p's, say p_1, p_2, ..., p_s , so that

(2.4.3) $$p_i = f_i(\underline{p}^*) \qquad i = s + 1, \ldots, r \ ,$$

where $\underline{p}^{*\prime}_{1\times s} = (p_1, p_2, \ldots, p_s)$.

Then, from (2.2.19), we have a minimum χ_1^2 estimate, subject to (2.4.1), given by

(2.4.4) $$\hat{\underline{p}}^*_{s\times 1} = \underline{p}^*_{0\,s\times 1} \pm n^{-\frac{1}{2}} (B'B)^{-1}_{s\times s} \, B'_{s\times r} \, \underline{x}_{r\times 1} + k \frac{\lambda^2}{n} \underline{\theta}_{1\,s\times 1} \ ,$$

where

(2.4.5) $$B = \left[\frac{1}{\sqrt{p_i^0}} \left(\frac{\partial p_i}{\partial p_j} \right)_0 \right] \qquad \begin{array}{l} i = 1,2,\ldots,r \\ j = 1,2,\ldots,s \end{array} \ .$$

Under constraints (2.4.2), we have to minimize

$$f = \sum_i \frac{(v_i - np_i)^2}{v_i} - 2n \sum_{t=1}^{\mu+1} \lambda_t [F_t(\underline{q}) + \sum_i b_{ti}(p_i - q_i)] \ ,$$

assuming that the $(\mu + 1)$-th equation is $\sum_{i=1}^{r} p_i = 1$. The equations are

$$(2.4.6) \qquad \frac{v_i - n\hat{\hat{p}}_i}{v_i} + \sum_t \hat{\lambda}_t b_{ti} = 0 \quad , \; i = 1,2,\ldots,r \; ,$$

and

$$(2.4.7) \qquad F_t(\underline{q}) + \sum_i b_{ti}(\hat{\hat{p}}_i - q_i) = 0 \quad , \quad t = 1,2,\ldots,(\mu+1) = \mu' \; ,$$

say.

Let

$$(2.4.8) \qquad A = (b_{ti})_{\mu' \times r} \; , \qquad Q_{r\times r} = \text{diagonal } (q_1,\ldots,q_r)$$

and $$\underline{f}'_{1\times\mu'} = [F_1(\underline{q}), \; \ldots, \; F_{\mu'}(\underline{q})] \; .$$

Then, (2.4.6) and (2.4.7) can be written as

$$\underline{q}_{r\times 1} - \hat{\underline{\hat{p}}}_{r\times 1} + \underset{r\times r}{Q} \; \underset{r\times\mu'}{A'} \; \underset{\mu'\times 1}{\hat{\lambda}} = 0 \quad \text{and}$$

$$\underline{f}_{\mu'\times 1} + A_{\mu'\times r}(\hat{\underline{\hat{p}}} - \underline{q})_{r\times 1} = 0 \quad .$$

Therefore, $$\underline{f} + (AQA')\underset{\mu'\times\mu'}{\hat{\underline{\lambda}}} = 0 \quad .$$

Since the restrictions (2.4.1) are independent, A is of rank μ' so that AQA' is nonsingular. Hence

$$\hat{\underline{\lambda}} = -\,(AQA')^{-1}\underline{f} \quad , \quad \text{so that}$$

$$(2.4.9) \qquad \hat{\underline{\hat{p}}}_{r\times 1} = \underline{q}_{r\times 1} - QA'(AQA')^{-1}\,\underline{f} \quad .$$

Thus $\hat{\underline{\hat{p}}}$ is unique.

Now, in the notation of Theorem 2.2, with probability $> 1 - \frac{1}{\lambda^2}$, $|x_i| < \frac{\lambda}{c}$, $i = 1,2,\ldots,r$. Also, (2.4.10) $\quad q_i = p_i^o + x_i\sqrt{p_i^o/n}$, and hence

$$b_{ti} = \left.\frac{\partial F_t}{\partial p_i}\right|_{\underline{p}=\underline{q}} = \left.\frac{\partial F_t}{\partial p_i}\right|_{\underline{p}=\underline{p}_o} + \sum_k \frac{\partial^2 F_t}{\partial p_i\,\partial p_k}(q_k - p_k^o)$$

$$= b_{tio} + O\left(\frac{\lambda}{\sqrt{n}}\right) , \quad \text{where} \quad b_{tio} = \left.\frac{\partial F_t}{\partial p_i}\right|_{\underline{p}=\underline{p}_o} .$$

Similarly,

$$F_t(\underline{q}) = F_t(\underline{p}_o) + \sum_i b_{tio}(q_i - p_i^o) + \frac{1}{2}\sum_i\sum_j \frac{\partial^2 F_t}{\partial p_i \partial p_j}(q_i - p_i^o)(q_j - p_j^o)$$

$$= \sum_i b_{tio}\, x_i \sqrt{p_i^o/n} + O\left(\frac{\lambda^2}{n}\right) .$$

Now,

$$(AQA')_{ij} = \sum_k b_{ik}\, q_k\, b_{jk}$$

$$= \sum_k \left[b_{iko} + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]\left[p_k^o + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]\left[b_{jko} + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]$$

$$= \sum_k b_{iko}\, p_k^o\, b_{jko} + O\left(\frac{\lambda}{\sqrt{n}}\right) , \quad \text{so that}$$

$$(AQA') = A_o D A_o' + \Theta , \quad \text{where} \quad \Theta = (\theta_{ij}) ,$$

$$D = \text{diagonal } (p_1^o, \ldots, p_r^o) \quad \text{and} \quad \theta_{ij} = O\left(\frac{\lambda}{\sqrt{n}}\right) .$$

Suppose $\underset{\mu' \times \mu'}{\Theta} = \underset{\mu' \times \ell}{P}\ \underset{\ell \times \mu'}{R}$, where ℓ = Rank P = Rank R

and $P = O\left(\frac{\lambda}{\sqrt{n}}\right)$. Then,

$$(AQA')^{-1} = (A_o D A_o')^{-1} + (A_o D A_o')^{-1} P\, \Lambda\, R (A_o D A_o')^{-1} ,$$

where Λ is a suitable matrix. Hence,

$$(2.4.11) \qquad (AQA')^{-1} = (A_0DA_0')^{-1} + \underset{\mu'\times\mu'}{O}\left(\frac{\lambda}{\sqrt{n}}\right) .$$

Similarly,

$$(QA')_{ij} = q_i b_{ji} = \left[p_i^0 + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]\left[b_{jio} + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]$$

$$= p_i^0 b_{jio} + O\left(\frac{\lambda}{\sqrt{n}}\right) , \quad \text{so that}$$

$$QA' = DA_0' + O\left(\frac{\lambda}{\sqrt{n}}\right) .$$

Thus, $QA'(AQA')^{-1} = DA_0'(A_0DA_0')^{-1} + O\left(\frac{\lambda}{\sqrt{n}}\right)$, and

$$QA'(AQA')^{-1}\underline{f} = \left[DA_0'(A_0DA_0')^{-1} + O\left(\frac{\lambda}{\sqrt{n}}\right)\right]\left[n^{-\frac{1}{2}}A_0D^{\frac{1}{2}}\underline{x} + O\left(\frac{\lambda^2}{n}\right)\right]$$

$$= n^{-\frac{1}{2}}DA_0'(A_0DA_0')^{-1}A_0D^{\frac{1}{2}}\underline{x} + O\left(\frac{\lambda^2}{n}\right) ,$$

where $D^{\frac{1}{2}} = \text{diagonal}(\sqrt{p_1^0}, \ldots, \sqrt{p_r^0})$.

Therefore, from (2.4.9) ,

$$(2.4.12) \qquad \underset{r\times1}{\underline{\hat{p}}} = \underset{r\times1}{\underline{p}_0} + n^{-\frac{1}{2}}D^{\frac{1}{2}}\underline{x} - n^{-\frac{1}{2}}DA_0'(A_0DA_0')^{-1}A_0D^{\frac{1}{2}}\underline{x} + O\left(\frac{\lambda^2}{n}\right)$$

$$= \underset{r\times1}{\underline{p}_0} + n^{-\frac{1}{2}}[D^{\frac{1}{2}} - DA_0'(A_0DA_0')^{-1}A_0D^{\frac{1}{2}}]\underline{x} + O\left(\frac{\lambda^2}{n}\right) ,$$

which shows that $\underline{\hat{p}} \to \underline{p}_0$ in probability. If we consider, as before, only the independent estimates, say $\hat{p}_1, \ldots, \hat{p}_s$, then we have, from (2.4.12),

$$(2.4.13) \qquad \underset{s\times1}{\underline{\hat{p}}^*} = \underset{s\times1}{\underline{p}_0^*} + n^{-\frac{1}{2}}[D_1^{\frac{1}{2}} \vdots O][I - D^{\frac{1}{2}}A_0'(A_0DA_0')^{-1}A_0D^{\frac{1}{2}}]\underline{x} + O\left(\frac{\lambda^2}{n}\right) ,$$

$$\text{where} \quad D = \begin{array}{cc} \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} & \begin{array}{c} s \\ r-s \end{array} \\ \begin{array}{cc} s & \quad r-s \end{array} & \end{array} \quad \text{say} .$$

We shall show that

$$(2.4.14) \qquad (B'B)^{-1}B' = [D_1^{\frac{1}{2}} \mid 0][I - D^{\frac{1}{2}}A_0'(A_0DA_0')^{-1}A_0D^{\frac{1}{2}}] .$$

Now $\left[\dfrac{\partial p_i}{\partial p_j}\right]_{0\ r\times s} = \begin{bmatrix} I \\ G \end{bmatrix}\begin{matrix} s \\ \mu' \end{matrix}$ where $G = \left[\dfrac{\partial f_i}{\partial p_j}\right]_0$, so that

$$(2.4.15) \qquad B = D^{-\frac{1}{2}}\begin{bmatrix} I \\ G \end{bmatrix}\begin{matrix} s \\ \mu' \end{matrix} ; \; D^{-\frac{1}{2}} = \text{diagonal}\left(\frac{1}{\sqrt{p_i^0}} , i=1,\ldots,r\right) .$$

Since $F_t(\underline{p}) = 0$, $t = 1,2,\ldots,\mu'$, we have

$$\frac{\partial F_t}{\partial p_j} + \sum_{i=s+1}^{r} \frac{\partial F_t}{\partial p_i}\frac{\partial p_i}{\partial p_j} = 0 , \qquad j = 1,2,\ldots,s .$$

Then

$$(2.4.16) \qquad \left[\frac{\partial p_i}{\partial p_j}\right]_{\mu'\times s} = -\left[\frac{\partial F_t}{\partial p_i}\right]^{-1}_{\mu'\times\mu'}\left[\frac{\partial F_t}{\partial p_j}\right]_{\mu'\times s} .$$

We notice that $\left(\dfrac{\partial F_t}{\partial p_i}\right)$ is nonsingular in view of independence of F_t , $t = 1,2,\ldots,\mu'$. Thus $G = -A_2^{-1}A_1$, where $A = (\underset{s}{A_1} \mid \underset{\mu'}{A_2})\mu'$, from (2.4.8). (For convenience, we shall drop the suffix 0.) Hence, $B = D^{-\frac{1}{2}}\begin{bmatrix} I \\ -A_2^{-1}A_1 \end{bmatrix}$, so that

$$\begin{aligned} B'B &= D_1^{-1} + A_1'A_2'^{-1}D_2^{-1}A_2^{-1}A_1 \\ &= D_1^{-1} + A_1'A_2'^{-1}D_3^{-1}D_3^{-1}A_2^{-1}A_1 , \text{ where } D_3 = D_2^{\frac{1}{2}} . \end{aligned}$$

Let $(B'B)^{-1} = D_1 + D_1A_1'A_2'^{-1}D_3^{-1}\varphi D_3^{-1}A_2^{-1}A_1D_1$.

Then,

$$(2.4.17) \qquad [I + D_3^{-1}A_2^{-1}A_1D_1A_1'A_2'^{-1}D_3^{-1}]\varphi = -I \ .$$

Thus, $(B'B)^{-1}B' = [D_1 + D_1A_1'A_2'^{-1}D_3^{-1}\varphi D_3^{-1}A_2^{-1}A_1D_1][I \vdots -A_1'A_2'^{-1}]D^{-\frac{1}{2}}$

$$= [D_1 + D_1A_1'A_2'^{-1}D_3^{-1}\varphi D_3^{-1}A_2^{-1}A_1D_1 \vdots D_1A_1'A_2'^{-1}D_3^{-1}\varphi D_3]D^{-\frac{1}{2}}$$

in view of (2.4.17). Now

$$ADA' = A_1D_1A_1' + A_1D_1A_1' + A_2D_2A_2' \quad .$$

Hence, $(ADA')^{-1} = A_2'^{-1}D_2^{-1}A_2^{-1} + A_2'^{-1}D_2^{-1}A_2^{-1}A_1D_4\varphi_1D_4A_1'A_2'^{-1}D_2^{-1}A_2^{-1}$,

where $D_4 = D_1^{\frac{1}{2}}$. Then,

$$(2.4.18) \quad [I + D_4A_1'A_2'^{-1}D_2^{-1}A_2^{-1}A_1D_4]\varphi_1 = -I \quad .$$

From (2.4.17) and (2.4.18), we can see that

$$(2.4.19) \qquad D_4A_1'A_2'^{-1}D_3^{-1}\varphi = \varphi_1D_4A_1'A_2'^{-1}D_3^{-1} \quad .$$

Now

$$\begin{aligned}
\text{right side of (2.4.14)} &= D_1^{\frac{1}{2}}[(I|O) - (D_1^{\frac{1}{2}}|O)A'(ADA')^{-1}AD^{\frac{1}{2}}] \\
&= D_1^{\frac{1}{2}}[(I|O) - D_1^{\frac{1}{2}}A_1'(ADA')^{-1}AD^{\frac{1}{2}}] \\
&= D_1^{\frac{1}{2}}[(I|O) - D_1^{\frac{1}{2}}A_1'(A_2'^{-1}D_2^{-1}A_2^{-1} + A_2'^{-1}D_2^{-1}A_2^{-1}A_1D_4D_4A_1'A_2'^{-1}D_3^{-1} \times \\
&\qquad \varphi D_3^{-1}A_2^{-1})AD^{\frac{1}{2}}] \quad \{\text{from (2.4.19)}\} \\
&= D_1^{\frac{1}{2}}[I - D_1^{\frac{1}{2}}A_1'(A_2'^{-1}D_2^{-1}A_2^{-1} + A_2'^{-1}D_2^{-1}A_2^{-1}A_1D_1A_1'A_2'^{-1}D_3^{-1}\varphi D_3^{-1}A_2^{-1})A_1D_1^{\frac{1}{2}} \vdots \\
&\qquad - D_1^{\frac{1}{2}}A_1'A_2'^{-1}D_2^{-\frac{1}{2}} - D_1^{\frac{1}{2}}A_1'A_2'^{-1}D_2^{-1}A_2^{-1}A_1D_1A_1'A_2'^{-1}D_3^{-1}\varphi] \\
&= D_1^{\frac{1}{2}}[I - D_1^{\frac{1}{2}}A_1'\{A_2'^{-1}D_2^{-1}A_2^{-1} - A_2'^{-1}D_3^{-1}(I+\varphi)D_3^{-1}A_2^{-1}\}A_1D_1^{\frac{1}{2}} \vdots \\
&\qquad D_1^{\frac{1}{2}}A_1'A_2'^{-1}D_3^{-1}\varphi] \quad \text{from (2.4.17)'} \\
&= (B'B)^{-1}B' = \text{left side of (2.4.14)} \quad .
\end{aligned}$$

Thus we have

$$(2.4.20) \qquad \underset{s\times 1}{\hat{\hat{\underline{p}}}^*} = \underset{s\times 1}{\underline{p}_0^*} + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + O\left(\frac{\lambda^2}{n}\right) .$$

Now, from Theorem 2,2,

$$\chi_1^2 = \sum_{i=1}^{r} \frac{(v_i - n\hat{p}_i)^2}{v_i}$$

is, in the limit as $n \to \infty$, distributed as a χ^2 with μ d.f. Let

$$(2.4.21) \qquad \chi_1^{*2} = \sum_{i=1}^{r} \frac{(v_i - n\hat{\hat{p}}_i)^2}{v_i} .$$

Since the first two terms in (2.4.4) and (2.4.20) are identical, using the same argument as for the minimum χ_1^2 under the original constraints, we have the limiting χ^2 distribution with μ d.f. for χ_1^{*2} .

CHAPTER III

SOME SPECIAL PROBLEMS POSED IN CHAPTER I

3.1 Introduction

Reiersol [27] considers binomial experiments and makes use of results of Neyman [22] to determine χ^2 tests for the hypotheses appropriate to factorial experiments. Mitra [19] not only generalizes Reiersol's theorems to multinomial experiments, but also avoids his restriction that the parameter sets in the different linear forms occurring in the hypothesis be non-overlapping. We shall prove theorems, analogous to Mitra's theorem, to cover the cases that cannot be treated by his theorem. In particular, when the hypothesis H_0 specifies linear functions of p's as known linear functions of some unknown parameters, the minimum χ_1^2 to test H_0 is exactly the same as the minimum sum of squares of residuals obtained by the general least squares technique on the linear functions of q's to estimate the unknown parameters. We shall then treat the various problems posed in Chapter I as direct applications of either these theorems or Cramer's general theorems as extended by Mitra [19] and Ogawa [23].

3.2 On the test of linear hypotheses on the responses by the use of χ_1^2 statistic and the χ_1^2 minimization method of estimation

Let us consider a product-multinomial distribution in the usual notation, so that "i" refers to categories of the response and "j" refers to s different multinomial distributions. Also $\sum_{i=1}^{r} p_{ij} = 1$ and $\sum_{i=1}^{r} n_{ij} = n_{oj}$ (fixed), $j = 1,2,\ldots,s$. Let $p_{ij} > 0$ and also $n_{ij} > 0$ for all (i,j) . We shall consider the hypothesis H_0 defined by m linearly independent constraints on p_{ij}'s [independent of $\sum_{i=1}^{r} p_{ij} = 1$], say,

$$\text{(3.2.1)} \quad H_0: \; F_t(\underline{p}) = \sum_i \sum_j f_{tij} p_{ij} + h_t = 0 \;, \quad t = 1,2,\ldots,m \;,$$

where f_{tij} and h_t are known constants such that (i) $\text{Rank}(f_{tij})_{m\times(rs)} = m \leq (r-1)s$, (ii) the above equations, together with $\sum_{i=1}^{r} p_{ij} = 1$ $(j = 1,2,\ldots,s)$ have at least one set of solutions $\{p_{ij}\}$ for which $p_{ij} > 0$ for all (ij). Then, $\chi_1^2 = \sum_j n_{oj} \sum_i \frac{(q_{ij} - p_{ij})^2}{q_{ij}}$, where $q_{ij} = \frac{n_{ij}}{n_{oj}}$.

Let $b_{tj} = \sum_i f_{tij}\, q_{ij}$, $\quad b_t = \sum_j b_{tj}$, $\quad c_t = b_t + h_t$,

$$\text{(3.2.2)} \quad e_{tt'j} = \sum_i (f_{tij} - b_{tj})(f_{t'ij} - b_{t'j}) q_{ij}$$

and

$$g_{tt'} = \sum_j \frac{1}{n_{oj}} e_{tt'j} \;, \qquad t,t' = 1,2,\ldots,m \;.$$

We notice that b_{tj} is in the nature of a "sample mean" of "F_t" for j-th sample, while $e_{tt'j}$ is in the nature of a sample covariance of "F_t" and "$F_{t'}$" for the j-th sample. Since F_t's are linearly independent, it follows that $G = (g_{tt'})_{m\times m}$ is positive-definite. We shall prove

Theorem 3.2.1

$$\chi_1^{*2} = \underset{\text{subject to } H_0}{\text{Min}}\ \chi_1^2 = \underline{c}'\, G^{-1}\, \underline{c} \quad .$$

Proof: To minimize χ_1^2 subject to the constraints we introduce Lagrangian multipliers and write

$$f = \sum_j n_{oj} \sum_i \frac{(p_{ij} - q_{ij})^2}{q_{ij}} - 2\sum_j \lambda_j\Big(\sum_i p_{ij} - 1\Big) - 2\sum_t \mu_t F_t(\underline{p}) \quad .$$

Differentiating with respect to p_{ij} and equating this to zero, we get the minimizing equations

$$n_{oj}\,\frac{(p_{ij} - q_{ij})}{q_{ij}} - \lambda_j - \sum_t \mu_t\, f_{tij} = 0 \ , \qquad \begin{array}{l} i = 1,2,\ldots,r \ , \\ j = 1,2,\ldots,s \ . \end{array}$$

Multiplying by q_{ij} and summing over i, we get

$$-\lambda_j - \sum_t \mu_t\, b_{tj} = 0 \quad .$$

Eliminating the λ's we get

$$(3.2.3) \qquad p_{ij} = q_{ij}\left[1 + \frac{\sum_t \mu_t (f_{tij} - b_{tj})}{n_{oj}}\right] \ ,$$

where μ's are to be determined from (3.2.1). Hence,

$$\sum_i \sum_j f_{tij} q_{ij} [1 + \frac{1}{n_{oj}} \sum_{t'} \mu_{t'} (f_{t'ij} - b_{t'j})] + h_t = 0 \quad , \qquad t = 1, 2, \ldots, m \; .$$

These may be written as

$$G\, \underline{\mu} + \underline{c} = \underline{0} \quad , \quad \text{where} \quad \underline{c}' = (c_1, c_2, \ldots, c_m)$$

$$\text{and} \quad \underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_m) \quad .$$

$$\text{Hence} \quad \underline{\mu} = - G^{-1} \underline{c} \quad .$$

$$\text{Then} \quad \chi_1^{*2} = \text{Minimum } \chi_1^2$$

$$= \sum_j n_{oj} \sum_i q_{ij} \left\{ \frac{1}{n_{oj}} \sum_t \mu_t (f_{tij} - b_{tj}) \right\}^2$$

$$= \sum_j \frac{1}{n_{oj}} \sum_t \sum_{t'} \mu_t \, \mu_{t'} \, e_{tt'j}$$

$$= \underline{\mu}' \, G \, \underline{\mu}$$

$$(3.2.4) \qquad = \underline{c}' \, G^{-1} \, \underline{c} \quad .$$

Then by Neyman's theorem, if H_0 is true, χ_1^{*2} is distributed in the limit as a χ^2 with m d.f.

The form of (3.2.4) suggests that χ_1^{*2} may be the same as the one we would obtain if we test the hypothesis (3.2.1) by considering b_t's, the unbiased estimates of $\sum_i \sum_j f_{tij}\, p_{ij}$ and using asymptotic normality. We have

$$b_t = \sum_i \sum_j f_{tij} \, q_{ij} \; , \text{ so that}$$

$$\mathcal{E}(b_t) = \sum_i \sum_j f_{tij} \, p_{ij} = - h_t \quad \text{if } H_0 \text{ is true,}$$

and

$$\operatorname{cov}(b_t, b_{t'}) = \sum_j \sum_i f_{tij} f_{t'ij} \frac{p_{ij}(1 - p_{ij})}{n_{oj}} - \sum_j \sum_{i \neq i'} f_{tij} f_{t'i'j} \frac{p_{ij} p_{i'j}}{n_{oj}}$$

$$= \sum_j \sum_i \frac{f_{tij} f_{t'ij} p_{ij}}{n_{oj}} - \sum_j \frac{1}{n_{oj}} \left(\sum_i f_{tij} p_{ij}\right)\left(\sum_i f_{t'ij} p_{ij}\right)$$

$$= \varphi_{tt'} \quad \text{say} .$$

Hence, in the limit, when H_0 is true, $\underline{c} \approx N(\underline{0}, \varphi)$,so that $\underline{c}' \varphi^{-1} \underline{c}$ is asymptotically distributed as a χ^2 with m d.f. If we replace p_{ij}'s in φ by q_{ij}'s we get G . Hence G may be considered as an estimate of φ . Thus we have proved

Theorem 3.2.2: The minimum χ_1^2 method to test the linear hypothesis of the type (3.2.1) is exactly equivalent to the "large sample test" based on the asymptotic normality of the unbiased estimates of $F_t(\underline{p})$, whose variance-covariance matrix is estimated by the "sample variance-covariance matrix."

Invariance

We then expect the χ_1^2 statistic to be invariant under the choice of linearly independent constraints (on p's) defining the hypothesis (3.2.1). This can be easily proved.

Suppose we start from

$$\sum_i \sum_j f^*_{tij} \, p_{ij} + h^*_t = 0 \quad , \quad \text{where}$$

$$f^*_{tij} = \sum_u k_{tu} \, f_{uij} \quad \text{and} \quad h^*_t = \sum_u k_{tu} \, h_u$$

where $K = (k_{tu})$ is nonsingular.

Then $\chi_1^2 = \underline{c}'\, G^{-1} \underline{c}$, from (3.2.4) .

Similarly $\chi_1^{*2} = \underline{c}^{*'} G^{*-1} \underline{c}^*$.

Now $b_{tj}^* = \sum_i f_{tij}^*\, q_{ij} = \sum_u k_{tu}\, b_{uj}$, so that

$$b_t^* = \sum_j b_{tj}^* = \sum_u k_{tu}\, b_u \quad .$$

Thus $c_t^* = b_t^* + h_t^* = \sum_u k_{tu}\, c_u$, so that

$$\underline{c}^* = K\, \underline{c} \quad .$$

Also, $e_{tt'j}^* = \sum_i (f_{tij}^* - b_{tj}^*)(f_{t'ij}^* - b_{t'j}^*) q_{ij}$

$$= \sum_u \sum_v k_{tu}\, k_{t'v}\, e_{tt'j} \quad ,$$ so that

$$g_{tt'}^* = \sum_j \frac{e_{tt'j}^*}{n_{oj}} = \sum_u \sum_v k_{tu}\, k_{t'v}\, g_{tt'}$$ and hence

$$G^* = K\, G\, K' \quad .$$

Then $\chi_1^{*2} = \underline{c}^{*'}\, G^{*-1} \underline{c}^* = \underline{c}' K' K'^{-1} G^{-1} K^{-1} K\, \underline{c} = \chi_1^2$.

3.2.1 Structured response

Sometimes a linear hypothesis is defined by linear functions of unknown parameters. Theoretically, of course, this can be reduced to the case, already considered, where the hypothesis is defined by linear constraints on p's. But in many cases this equivalent expression in terms of linearly independent constraints on p's may be tedious. We shall prove a theorem, which might be considered as another version of theorems 3.2.1 and 3.2.2, which reduces the problem to that of least squares.

Theorem 3.2.3: With the same product-multinomial distribution for the model that was used in theorem 3.2.1, let the linear hypothesis H_0 be defined by

(3.2.5) $$H_0: \sum_i a_i p_{ij} = d_{j1}\theta_1 + d_{j2}\theta_2 + \dots + d_{jt}\theta_t , \quad j = 1,2,\dots,s ,$$

where d's are known constants and θ's are unknown parameters. Then the minimum χ_1^2 to test H_0 is the same as the minimum sum of squares of residuals obtained by the general least squares technique on $\sum_i a_i q_{ij}$, with the variances estimated by "sample-variances."

Proof: Let $D = (d_{jk})_{s \times t}$ and Rank $D = u \le s$.

Now $H_0 \Longrightarrow \sum_j \ell_j \sum_i a_i p_{ij} = \sum_j \ell_j \sum_k d_{jk}\theta_k = \sum_k \theta_k \sum_j \ell_j d_{jk}$.

Hence, if $\sum_j \ell_j d_{jk} = 0$, $(k = 1,2,\dots,t)$, then $\sum_j \sum_i \ell_j a_i p_{ij} = 0$. On the other hand, if $\sum_j \sum_i \ell_j a_i p_{ij} = 0$, whatever may be θ's, then $\sum_j \ell_j d_{jk} = 0$, $k = 1,2,\dots,t$.

Such a linear function $\sum_j \ell_j \sum_i a_i p_{ij}$ may be called a "hypothesis constraint."

Since Rank $D = u$, the number of linearly independent s-vectors $(\ell_{v1}, \ell_{v2}, \dots, \ell_{vs})$ satisfying $\sum_j \ell_{vj} d_{jk} = 0$ $(k = 1,2,\dots,t)$ is $s - u$.

Let $L = (\ell_{vj})_{(s-u)\times s}$ be a matrix whose row-vectors satisfy the above conditions. Then Rank $L = (s-u)$ and $LD = 0$. Now, $H_0 \Longrightarrow \sum_i \sum_j \ell_{vj}\, a_i\, p_{ij} = 0$, $(v = 1,2,\ldots, s-u)$. On the other hand, $\sum_i \sum_j \ell_{vj}\, a_i\, p_j = 0$, $(v=1,2,\ldots,s-u)$ $\Longrightarrow \sum_j \ell_{vj}\, m_j = 0$ $(v=1,2,\ldots,s-u)$, where $m_j = \sum_i a_i\, p_{ij}$. Thus $(m_1, m_2, \ldots, m_s)$ is orthogonal to row-vectors of L. But the row-vectors of L form a basis of the vector-space orthogonal to column-vectors of D. Hence $(m_1, \ldots, m_s)$ belongs to the vector-space generated by the column-vectors of D. Thus, there are θ's, not all zero, such that

$$m_j = \sum_k \theta_k\, d_{jk} \quad , \quad j = 1,2,\ldots,s \quad .$$

Therefore, $\sum_i a_i\, p_{ij} = \sum_k d_{jk}\, \theta_k$ $(j = 1,2,\ldots,s)$. Thus $H_0 \Longleftrightarrow \sum_i \sum_j \ell_{vj}\, a_i\, p_{ij} = 0$ $(v = 1,2,\ldots,s-u)$, where $LD = 0$ and L is of rank $(s-u)$. Then by theorem 3.2.1, the minimum χ_1^2 is given by (3.2.4). Here

$$f_{vij} = \ell_{vj}\, a_i \quad , \quad \text{so that}$$

$$b_{vj} = \sum_i f_{vij}\, q_{ij} = \ell_{vj} \sum_i a_i\, q_{ij} = \ell_{vj}\, \alpha_j \quad , \quad \text{where}$$

$$\alpha_j = \sum_i a_i\, q_{ij} \quad ,$$

$$b_v = \sum_j b_{vj} = \sum_j \ell_{vj}\, \alpha_j \quad .$$

Also $h_v = 0$ and hence $c_v = b_v$. Thus,

$$\underline{c} = L\,\underline{\alpha} \quad .$$

Also, $e_{vv'j} = \sum_i (f_{vij} - b_{vj})(f_{v'ij} - b_{v'j})q_{ij}$

$$= \ell_{vj}\,\ell_{v'j}\,\beta_j \quad , \quad \text{where}$$

$$\beta_j = \sum_i (a_i - \alpha_j)^2\, q_{ij} \quad .$$

Then, $g_{vv'} = \sum_j \frac{1}{n_{oj}}\, e_{vv'j} = \sum_j \frac{\beta_j}{n_{oj}}\, \ell_{vj}\,\ell_{v'j}$,

so that $G = L\,\Lambda\,L'$, where

$\lambda_j = \frac{\beta_j}{n_{oj}}$ and $\Lambda = \text{diagonal}\ (\lambda_j\ ,\ j = 1,2,\ldots,s)$. Hence,

(3.2.6) $$\chi_1^2 = \underline{\alpha}'L'(L\,\Lambda\,L')^{-1}\,L\,\underline{\alpha} \quad .$$

Note that α_j's are independent and

$$\text{var}(\alpha_j) = \sum_i a_i^2\,\frac{p_{ij}(1 - p_{ij})}{n_{oj}} - \sum_{i \neq i'}\sum a_i\, a_{i'}\,\frac{p_{ij}\,p_{i'j}}{n_{oj}}$$

$$= \frac{1}{n_{oj}}\left[\sum_i a_i^2\, p_{ij} - \left\{\sum_i a_i\, p_{ij}\right\}^2\right] \quad .$$

Hence "sample var" $(\alpha_j) = \frac{1}{n_{oj}}\left[\sum_i a_i^2\, q_{ij} - \left\{\sum_i a_i\, q_{ij}\right\}^2\right]$

$$= \lambda_j \quad .$$

If we use the least squares technique on α_j's (using λ_j's for the variance), then the sum of squares to be minimized with respect to the parameters is

$$S^2 = \sum_j (\alpha_j - d_{j1}\theta_1 - d_{j2}\theta_2 - \cdots - d_{jt}\theta_t)^2 \frac{1}{\lambda_j}$$

$$= \sum_j (\gamma_j - \delta_{j1}\theta_1 - \cdots - \delta_{jt}\theta_t)^2 \quad , \quad \text{where}$$

$$\gamma_j = \alpha_j \lambda_j^{-\frac{1}{2}} \quad \text{and} \quad \delta_{jk} = d_{jk}\lambda_j^{-\frac{1}{2}} \quad ,$$

so that $\underline{\gamma} = \Lambda^{-\frac{1}{2}}\underline{\alpha}$ and $\Delta = (\delta_{jk}) = \Lambda^{-\frac{1}{2}}D$,

where $\Lambda^{-\frac{1}{2}} = \text{diagonal}(\lambda_j^{-\frac{1}{2}} , \; j = 1,2,\ldots,s)$. Then it is well known that (for example, [29])

$$\operatorname*{Min}_{\theta} S^2 = \underline{\gamma}'\underline{\gamma} - \underline{\gamma}'\Delta_1(\Delta_1' \Delta_1)^{-1}\Delta_1'\underline{\gamma} \quad , \quad \text{where}$$

$$\underset{s\times u}{\Delta_1} = \text{Basis of } \underset{s\times t}{\Delta} \quad .$$

We note that $\Delta = \Lambda^{-\frac{1}{2}}D \Longrightarrow \text{Rank } \Delta = \text{Rank } D = u$ and $\Delta_1 = \Lambda^{-\frac{1}{2}}D_1$, where D_1 is a basis of D . Hence

$$\text{(3.2.7)} \qquad \text{Min } S^2 = \underline{\alpha}'\Lambda^{-1}\underline{\alpha} - \underline{\alpha}'\Delta^{-1}D_1(D_1'\Lambda^{-1}D_1)^{-1}D_1'\Lambda^{-1}\underline{\alpha} \quad .$$

It remains to show that (3.2.6) = (3.2.7). Now

$$LD = 0 \iff LD_1 = 0 \quad .$$

Let us choose L , subject to $LD_1 = 0$, so that $L \Lambda L' = I$. Then, (3.2.6) $= \underline{\alpha}'L'L\underline{\alpha}$. Let $M = L\,\Lambda^{\frac{1}{2}}$ so that $L = M\,\Lambda^{-\frac{1}{2}}$ and hence $MM' = I$ and $M\Lambda^{-\frac{1}{2}}D_1 = 0$. Also $D_1'\Lambda^{-1}D_1$ is positive-definite, so that there is a nonsingular X such that

$$X'D_1'\Lambda^{-1}D_1X = I \quad , \quad \text{that is,} \quad D_1'\Lambda^{-1}D_1 = (XX')^{-1} \quad .$$

Now
$$\begin{array}{c} s-u \\ u \end{array}\begin{bmatrix} M \\ \hdashline X'D_1'\Lambda^{-\frac{1}{2}} \end{bmatrix}\begin{bmatrix} M' & \vdots & \Lambda^{-\frac{1}{2}}D_1X \end{bmatrix} s = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} .$$
$$\begin{array}{ccc} s & s-u & u \end{array}$$

Hence, $\begin{bmatrix} M' & \vdots & \Lambda^{-\frac{1}{2}}D_1X \end{bmatrix}\begin{bmatrix} M \\ X'D_1'\Lambda^{-\frac{1}{2}} \end{bmatrix} = I$. Thus,

$$M'M + \Lambda^{-\frac{1}{2}}D_1XX'D_1'\Lambda^{-\frac{1}{2}} = I \quad , \quad \text{so that}$$

$$\Lambda^{\frac{1}{2}}L'L\Lambda^{\frac{1}{2}} + \Lambda^{-\frac{1}{2}}D_1(D_1'\Lambda^{-1}D_1)^{-1}D_1'\Lambda^{-\frac{1}{2}} = I \quad .$$

Hence, $\quad L'L = \Lambda^{-1} - \Lambda^{-1}D_1(D_1'\Lambda^{-1}D_1)^{-1}D_1'\Lambda^{-1}$.

Then (3.2.6) = $\underline{\alpha}'L'L\underline{\alpha}$ = (3.2.7) , which completes the proof of the theorem.

It may be noted that the minimum χ_1^2 is, in the limit, distributed as a χ^2 with $s-u$ d.f., where $u = \text{Rank } D$.

3.2.2 Application of 3.2.1 to linear hypotheses on structured uniresponse

In what follows, "i" will denote the structured variate. We shall consider some simple cases.

(i) One dimensional design ("j" $\rightarrow$ factor) .

(3.2.8) $H_0: \sum_i a_i p_{ij}$ is independent of j , $j = 1,2,\ldots,s$.

(3.2.9) $$\chi_1^2 = \sum_{j=1}^{s} \frac{n_{oj}\alpha_j^2}{\beta_j} - \frac{\left[\sum_{j=1}^{s} \frac{n_{oj}\alpha_j}{\beta_j}\right]^2}{\sum_{j=1}^{s} \frac{n_{oj}}{\beta_j}} \quad , \quad \text{d.f.} = s-1 \; ,$$

where $\alpha_j = \sum_i a_i q_{ij}$ and $\beta_j = \sum_i (a_i - \alpha_j)^2 q_{ij}$.

(ii) Two-dimensional design $\left\{ \begin{array}{l} (\text{"j"} \to \text{"Treatments"}) \\ (\text{"k"} \to \text{"Blocks"}) \end{array} \right.$

(a) Hypothesis of no treatment effects on the basic model

(3.2.10) $H_0: \sum_i a_i p_{ijk} = q_{**k}$, $[H_{05}$, (1.2.11)]

$$\begin{array}{l} j = 1, 2, \ldots, s \\ k = 1, 2, \ldots, t \end{array} \quad \{\text{Design may be incomplete}\} .$$

(3.2.11) $$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \sum_{k=1}^{t} [\sum_j h_{jk}\alpha_{jk}]^2 \Big/ \sum_j h_{jk} ,$$

d.f. = M - t ,

where $\alpha_{jk} = \sum_i a_i q_{ijk}$, $\beta_{jk} = \sum_i (a_i - \alpha_{jk})^2 q_{ijk}$

and $h_{jk} = n_{ojk} / \beta_{jk}$,

the summation is over allowable (jk) combinations and M is the number of (jk) combinations. When the design is complete, M = st , so that d.f. = (s - 1)t .

(b) Hypothesis of no interaction (in the additive set up)

(3.2.12) $H_0: \sum_i a_i p_{ijk} = q_{*j*} + q_{**k}$, $[H_{03}^{(1)}$, (1.2.7)] .

(3.2.13) $$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \sum_{k=1}^{t} \frac{B_k^2}{h_{ok}} - \sum_{j=1}^{s} Q_j t_j ,$$

d.f. = M - (s + t - 1) ,

where t's satisfy

$$Q_j = \sum_{j'=1}^{s} c_{jj'} t_{j'} , \quad j = 1,2,\ldots,s , \tag{3.2.14}$$

and

$$B_k = \sum_j \alpha_{jk} h_{jk} , \quad T_j = \sum_k \alpha_{jk} h_{jk} ,$$

$$h_{ok} = \sum_j h_{jk} , \quad h_{jo} = \sum_k h_{jk} ,$$

$$Q_j = T_j - \sum_k \frac{B_k}{h_{ok}} h_{jk} , \quad c_{jj} = h_{jo} - \sum_k \frac{h_{jk}^2}{h_{ok}}$$

$$\text{and} \quad c_{jj'} = -\sum_k \frac{h_{jk}\, h_{j'k}}{h_{ok}} .$$

Here M is, as before, the number of (jk) combinations and the summations are over allowable combinations only.

It may be noted that (3.2.13) and (3.2.14) are similar to "error sum of squares" and "normal equations," respectively, in analysis of variance, T_j and B_k playing the roles of "treatment total" and "block total," respectively. The fundamental difference, however, is that $c_{jj'}$'s here depend not only on the design but also on the observed proportions. In normal ANOVA, The designs can be chosen suitably so that the normal equations have neat closed solutions. This approach fails here for the corresponding equations (3.2.14). For example, even for a complete design (which may be called "randomized block design"), there is no essential simplification in the equations (3.2.14). [The degrees of freedom for χ_1^2 in that case are = $(s-1)(t-1)$.]

(c) Hypothesis of no treatment effects on the no interaction model

$$(3.2.16) \qquad \chi_1^2 = (3.2.11) - (3.2.13)$$

$$= \sum_{j=1}^{s} Q_j t_j \quad , \quad \text{d.f.} = s - 1 \quad ,$$

where Q's and t's are defined as before.

(d) $H_0: \sum_i a_i p_{ijk} = \lambda + \mu b_j$ $[H_{09}$, (1.2.17)] .

$$(3.2.17) \quad \chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \frac{1}{h\ell - m^2} [G^2\ell - 2G\gamma m + \gamma^2 h] ,$$

$$\text{d.f.} = M - 2 ,$$

where $h = \sum_j \sum_k h_{jk}$, $m = \sum_{j=1}^{s} b_j h_{jo}$,

$\ell = \sum_{j=1}^{s} b_j^2 h_{jo}$, $G = \sum_{j=1}^{s} T_j = \sum_{k=1}^{t} B_k$ and

$\gamma = \sum_{j=1}^{s} b_j T_j$ (other quantities as before).

(e) $H_0: \sum_i a_i p_{ijk} = \lambda_k + \mu b_j$ $[H_{07}^{(1)}$, (1.2.15)] .

$$(3.2.18) \quad \chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \sum_{k=1}^{t} \frac{B_k^2}{h_{ok}} - \frac{\left[\gamma - \sum_{k=1}^{t} \frac{B_k m_k}{h_{ok}}\right]^2}{\ell - \sum_{k=1}^{t} \frac{m_k^2}{h_{ok}}} ,$$

$$\text{d.f.} = M - t - 1 ,$$

where $m_k = \sum_j b_j h_{jk}$ (other quantities defined as before).

(f) $\underline{H_0: \sum_i a_i\ p_{ijk} = \lambda_k + \mu_k b_j}$ $[H_{07}$, (1.2.14)]

(3.2.19) $$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2\ h_{jk} - \sum_{k=1}^{t} \frac{B_k^2}{h_{ok}} - \sum_{k=1}^{t} \frac{\left[\gamma_k - \frac{B_k m_k}{h_{ok}}\right]^2}{\ell_k - \frac{m_k^2}{h_{ok}}} ,$$

d.f. = M - 2t ,

where $\gamma_k = \sum_j \alpha_{jk}\ h_{jk}\ b_j$ and $\ell_k = \sum_j b_j^2\ h_{jk}$

(other quantities defined as before).

(iii) Two-dimensional design ("j" → factor)
("k" → another factor)

$H_0: \sum_i a_i\ p_{ijk} = \lambda + \mu b_j + \nu c_k$ $[H_{010}$, (1.2.18)] .

(3.2.20) $$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2\ h_{jk} - \hat{\lambda}G - \hat{\mu}\gamma - \hat{\nu}\delta ,$$

d.f. = M - 3 ,

where $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\nu}$ satisfy the equations

$$G = \hat{\lambda}h + \hat{\mu}m + \hat{\nu}w$$

$$\gamma = \hat{\lambda}m + \hat{\mu}\ell + \hat{\nu}x$$

$$\delta = \hat{\lambda}w + \hat{\mu}x + \hat{\nu}y ,$$ where

$$\delta = \sum_{k=1}^{t} c_k B_k , \quad w = \sum_{k=1}^{t} c_k\ h_{ok} ,$$

$$x = \sum_j \sum_k b_j\ c_k\ h_{jk} \quad \text{and} \quad y = \sum_{k=1}^{t} c_k^2\ h_{ok}$$

(other quantities defined as before).

3.2.3 Application of 3.2.1 to linear hypotheses on structured multiresponse

Let us now suppose that $i = (i_1, i_2, \ldots, i_p)$, that is, there are p responses indicated by $i_1, i_2, \ldots, i_p$ such that $i_1 = 1,2,\ldots,r_1, \ldots, i_p = 1,2,\ldots,r_p$. If these responses are structured, the linear hypotheses, in general, will be of type

$$\text{(3.2.21)}\quad H_0: \sum_{i_1=1}^{r_1} \sum_j f_{ti_1**\ldots*j}\, p_{i_1oo\ldots oj} + h_t^{(1)} = 0$$

$$\cdots\cdots$$

$$\sum_{i_p=1}^{r_p} \sum_j f_{t**\ldots*i_pj}\, p_{oo\ldots oi_pj} + h_t^{(p)} = 0 \ ,$$

$$t = 1,2,\ldots,m \ ,$$

where the linear functions are linearly independent. We can write these as

$$\sum_i \sum_j f_{t*\ldots*i_k*\ldots*j}\, p_{ij} + h_t^{(k)} = 0 \ , \quad k = 1,2,\ldots,p \ , \ t = 1,2,\ldots,m \ ,$$

and hence (3.2.21) is a particular case of (3.2.1). Let

$$b_{tj}^{(k)} = \sum_i f_{t*\ldots*i_k*\ldots*j}\, q_{ij} = \sum_{i_k=1}^{r_k} f_{t*\ldots*i_k*\ldots*j}\, q_{o\ldots oi_ko\ldots oj},$$

$$b_t^{(k)} = \sum_j b_{tj}^{(k)} \ , \qquad c_t^{(k)} = b_t^{(k)} + h_t^{(k)} \quad \text{and}$$

$$\underline{c}'_{1\times mt} = \left(c_1^{(1)}, \ldots, c_m^{(1)};\ c_1^{(2)}, \ldots, c_m^{(2)};\ \ldots\ ,\ c_m^{(p)}\right) \ .$$

Similarly, let

$$g_{tt'}^{(kk')} = \sum_j \frac{1}{n_{oj}}\, e_{tt'j}^{(kk')} \ ,$$

where

$$e_{tt'j}^{(kk')} = \sum_i \left\{f_{t*\ldots*i_k*\ldots*j} - b_{tj}^{(k)}\right\}\left\{f_{t'*\ldots*i_{k'}*j} - b_{t'j}^{(k')}\right\} q_{ij}$$

$$= \sum_{i_k=1}^{r_k} \sum_{i_{k'}=1}^{r_{k'}} f_{t*\ldots*i_k*\ldots*j}\, f_{t'*\ldots*i_{k'}*\ldots*j} \times$$

$$q_{o\ldots oi_k o\ldots oi_{k'} o\ldots oj} - b_{tj}^{(k)}\, b_{t'j}^{(k')} \quad .$$

Thus,

$$\underset{pm\times pm}{G} = \begin{bmatrix} G^{11} & G^{12} & \ldots & G^{1p} \\ G^{21} & G^{22} & \ldots & G^{2p} \\ \cdot & \cdot & \cdot & \cdot \\ G^{p1} & \ldots & \ldots & G^{pp} \end{bmatrix} \text{, where}$$

$$\underset{m\times m}{G^{kk'}} = \left(g_{ij}^{(kk')}\right) \qquad \begin{matrix} i = 1,2,\ldots,m \\ j = 1,2,\ldots,m \end{matrix} \quad .$$

Then $$X_1^2 = \underline{c}'\, G^{-1}\, \underline{c} \quad , \quad \text{d.f.} = pm \quad .$$

We shall not go into the details of various special cases, for example, $H_{029}^{(1)}$, $H_{030}^{(1)}$ (1.5.6), $H_{031}^{(1)}$ (1.5.7), H_{032} (1.5.8), H_{033} (1.5.9), H_{034} (1.5.10) and H_{035} (1.5.11). By considering "hypothesis constraints," one can reduce the above hypotheses (expressed in terms of parameters) to equivalent forms like (3.2.21).

3.2.4 Unstructured response

In theorem 3.2.1, we considered the test criterion appropriate to a linear hypothesis. Its equivalence to a certain least squares technique, for the linear hypotheses in structured cases, was shown in theorem 3.2.3. We shall establish a similar equivalence for the linear hypotheses in unstructured cases in theorem 3.2.4.

Theorem 3.2.4

Under the product-multinomial set up, as in 3.2, let the linear hypothesis be defined by

$$(3.2.22)\quad H_0:\ p_{ij} = d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + \dots + d_{jt}\theta_{it}\ , \qquad \begin{array}{l} i = 1,2,\dots,r \\ j = 1,2,\dots,s\ , \end{array}$$

where d's are known constants and θ's are unknown parameters. Then the minimum χ_1^2 to test H_0 is the same as the minimum "generalized sum of squares" of residuals obtained by the "generalized least squares technique" on q_{ij}, with the covariance-matrix estimated by the "sample covariance-matrix."

Proof: Let $D = (d_{jk})_{s\times t}$ and Rank $D = u \leq s$.

Then as in theorem 3.2.3, we can show that

$$H_0 \Longleftrightarrow \sum_{j=1}^{s} \ell_{vj}\, p_{ij} = 0 \qquad \begin{array}{l} v = 1,2,\dots,s-u \\ i = 1,2,\dots,r\ , \end{array}$$

where $LD = 0$ and L is of rank $(s-u)$. Hence we can apply theorem 3.2.1. Here $f_{(vi)i'j} = \delta_{ii'}\,\ell_{vj}$ $\begin{cases} \delta_{ii'} = 1 \text{ if } i = i' \\ \quad\ = 0 \text{ otherwise}\end{cases}$,

so that

$$b_{(vi)j} = \sum_{i'} f_{(vi)i'j}\, q_{i'j} = \ell_{vj}\, q_{ij}$$

and, hence,

$$h_{(vi)} = \sum_j b_{(vi)j} = \sum_j \ell_{vj} q_{ij} \quad .$$

Also $h_{(vi)} = 0$, so that $c_{(vi)} = \sum_j \ell_{vj}\, q_{ij}$. Thus,

$$\underline{c} = L^* \underline{q} \quad ,$$

where $\underset{1\times r(s-u)}{\underline{c}'} = (c_{11},\ldots,c_{1r}, \ldots, c_{v1},\ldots,c_{vr},\ldots,c_{(s-u)})$,

$$\underset{r(s-u)\times rs}{L^*} = \underset{(s-u)\times s}{L\ X} \cdot I_r = \begin{bmatrix} \ell_{11}I_r & \cdots & \ell_{1s}I_r \\ \cdot\ \cdot\ \cdot & \cdots & \cdot\ \cdot\ \cdot \\ \ell_{s-u,1}I_r & \cdots & \ell_{s-u,s}I_r \end{bmatrix}$$

and $\underset{1\times rs}{\underline{q}'} = (q_{11}, q_{21},\ldots, q_{r1}, \ldots, q_{ij},\ldots,q_{rj},\ldots,q_{rs})$.

Similarly,

$$\begin{aligned} e_{(vi)}e_{(ui')j} &= \sum_{i''} [f_{(vi)i''j} - b_{(vi)j}][f_{(ui')i''j} - b_{(ui')j}]q_{i''j} \\ &= \sum_{i''} (\delta_{ii''} - q_{ij})\ell_{vj}(\delta_{i'i''} - q_{i'j})\ell_{uj}q_{i''j} \\ &= \ell_{vj}\ell_{uj}[\delta_{ii'}q_{ij} - q_{ij}q_{i'j}] \\ &= \ell_{vj}\ell_{uj}y^{j}_{ii'} \quad , \text{ say.} \end{aligned}$$

Then, $g_{(vi)(ui')} = \sum_j \ell_{vj}\ell_{uj} \frac{1}{n_{oj}}\, y^{j}_{ii'}$.

Hence $\underset{r(s-u)\times r(s-u)}{G} = L^* \underset{rs\times rs}{Y} L^{*'}$, where

$$Y = \begin{bmatrix} Y_1/n_{o1} & 0 & \cdots & 0 \\ 0 & Y_2/n_{o2} & \cdots & 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & & \cdots & \\ 0 & & & Y_s/n_{os} \end{bmatrix} \quad \text{and } Y_j = \left(y^{j}_{ii'}\right)_{r\times r} .$$

Then, from (3.2.4),

(3.2.23) $$\chi_1^2 = \underline{q}'L^{*'}[L^*\,Y\,L^{*'}]^{-1}L^*\underline{q} \quad .$$

[This has already been shown by Mitra [19] in a slightly different form.]

On the other hand, if we consider asymptotically normal variables q_{ij}'s , then

$$\mathrm{cov}(q_{ij}, q_{i'j'}) = 0 \quad \text{when} \quad j \neq j' ,$$

and
$$\mathrm{cov}(\underline{q}_j') = \mathrm{cov}(q_{1j}, \ldots, q_{rj}) = \frac{1}{n_{oj}} Y_j \quad .$$

Let
$$S^2 = \sum_j n_{oj}(\underline{q}_j - d_{j1}\underline{\theta}_1 - \cdots - d_{jt}\underline{\theta}_t)' Y_j^{-1} (\underline{q}_j - d_{j1}\underline{\theta}_1 - \cdots - d_{jt}\underline{\theta}_t) ,$$

where $\underline{\theta}_k' = (\theta_{1k}, \theta_{2k}, \ldots, \theta_{rk})$. S^2 may be called the "generalized sum of squares of residuals," and S^2 is to be minimized with respect to $\underline{\theta}$'s.

Since Y_j is positive-definite, there is $P_{j(r\times r)}$ (nonsingular) such that

$$n_{oj} Y_j^{-1} = P_j' P_j \quad , \quad j = 1,2,\ldots,s .$$

Let
$$\underline{q}_j^* = P_j \underline{q}_j \quad , \quad \text{so that}$$

$$S^2 = \sum_j (\underline{q}_j^* - d_{j1}P_j\underline{\theta}_1 - \cdots - d_{jt}P_j\underline{\theta}_t)'(\underline{q}_j^* - d_{j1}P_j\underline{\theta}_1 - \cdots - d_{jt}P_j\underline{\theta}_t)$$
$$= (\underline{q}^* - C\underline{\theta})'(\underline{q}^* - C\underline{\theta}) \quad , \quad \text{where}$$

$$\underset{1\times rs}{\underline{q}^{*\prime}} = (\underline{q}_1^{*\prime}, \ldots, \underline{q}_s^{*\prime}) \ , \quad \underset{1\times rt}{\underline{\theta}'} = (\underline{\theta}_1', \ldots, \underline{\theta}_t')$$

and
$$C_{rs\times rt} = \begin{bmatrix} P_1 & 0 & 0 & \ldots & 0 \\ 0 & P_2 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \ldots & P_s \end{bmatrix}_{rs\times rs} \left(\underset{s\times t}{D} \times I_r\right)_{rs\times rt} \quad .$$

Then, it is well known that

(3.2.24) $\underset{\theta}{\text{Min}}\ S^2 = \underset{\sim}{g}^{*'}[I_{rs} - C_1(C_1'C_1)^{-1}C_1']\underset{\sim}{g}^*$,

where $\underset{rs\times ru}{C_1}$ is a basis of C . We notice that if D_1 is the corresponding basis of D , then

$$C_1 = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & & P_s \end{bmatrix} D_1^* \text{ , where } D_1^* = D_1 X \cdot I_r \quad .$$

Then $C_1'C_1 = D_1^{*'}\ Y^{-1}D_1^*$.

Also $\underset{\sim}{g}^{*'}\underset{\sim}{g}^* = \sum_j \underset{\sim}{g}_j^{*'}\underset{\sim}{g}_j^* = \sum_j \underset{\sim}{g}_j'\ n_{oj}\ Y_j^{-1}\underset{\sim}{g}_j = \underset{\sim}{g}'Y^{-1}\underset{\sim}{g}$

and $\underset{\sim}{g}^{*'}C_1 = \underset{\sim}{g}'Y^{-1}D_1^*$. Hence, from (3.2.24),

(3.2.25) $\underset{\theta}{\text{Min}}\ S^2 = \underset{\sim}{g}'Y^{-1}\underset{\sim}{g} - \underset{\sim}{g}'Y^{-1}D_1^*[D_1^{*'}Y^{-1}D_1^*]^{-1}D_1^{*'}Y^{-1}\underset{\sim}{g}$.

Also $LD = 0 \iff LD_1 = 0 \iff L^*D_1^* = 0$. (3.2.23) and (3.2.25) can, then, be seen to be identical by exactly the same argument as the one used for (3.2.6) and (3.2.7).

3.2.5 <u>Application of 3.2.4 to linear hypotheses on unstructured uniresponse</u>

Two-dimensional design $\begin{cases} \text{"j"} \rightarrow \text{"Treatments"} \\ \text{"k"} \rightarrow \text{"Blocks"} \end{cases}$

(i) <u>Hypothesis of no treatment effects, on the basic model</u>

$$H_0:\ p_{ijk} = q_{i*k} \qquad [H_{02},\ (1.2.5)] \ .$$

This has already been considered by Roy and Mitra [33]. In this case, maximum likelihood equations give a unique solution and using this in a χ^2 statistic we have

$$(3.2.26) \qquad \chi^2 = \sum_{i,j,k} \frac{\left(n_{ijk} - \frac{n_{ojk}\, n_{iok}}{n_{ook}}\right)^2}{\frac{n_{ojk}\, n_{iok}}{n_{ook}}} ,$$

d.f. $= (r-1)(M-t)$ in the usual notation

$= t(r-1)(s-1)$ for a complete design .

(ii) Hypothesis of no interaction (additive set-up)

$$H_0: p_{ijk} = q_{i*k} + q_{ij*} \quad [H_{01}^{(1)}, \quad (1.2.3)] .$$

We shall not consider the details of the computation from (3.2.23).

d.f. $= (r-1)(M-s-t+1) = (r-1)(s-1)(t-1)$
for a complete design.

(iii) Hypothesis of no treatment effects, on the model of no interaction (additive set-up)

$\chi^2 = (3.2.26) - \chi_1^2$ for (ii) above.

d.f. $= (r-1)(s-1)$.

3.2.6 Linear hypotheses on unstructured multiresponse

As in 3.2.3, let us now suppose that $i = (i_1, i_2, \ldots, i_p)$, that is, there are p responses indicated by $i_1, i_2, \ldots, i_p$, such that $i_1 = 1,2,\ldots,r_1$, ..., $i_p = 1,2,\ldots,r_p$. If these are unstructured, the linear hypotheses, in general, will be of the type

$$H_0: \sum_j f_{tj}\, p_{i_1 oo\ldots oj} + h_t = 0$$

$$\cdots\cdots\cdots\cdot$$

$$\sum_j f_{tj}\, p_{oo\ldots oi_p j} + h_t = 0\,, \quad t = 1,2,\ldots,m\,.$$

It may be seen that it is not a particular case of (3.2.22) or of its equivalent form. But it is a particular case of (3.2.1) (by suitable definition of f_{tij}'s). Hence the χ_1^2 statistic may be worked out. We shall not go into these details. This will cover, for example, the cases $H_{026}^{(1)}$, $H_{027}^{(1)}$ and $H_{028}^{(1)}$.

3.3 On the test of nonlinear hypotheses on p's in ANOVA and MANOVA situations

As already mentioned in Chapter II, we can adopt Neyman's technique of linearization, so that the problem is reduced to one of the previous cases. On the other hand, it may happen in some cases that the maximum likelihood equations are fairly simple so that the χ^2 statistic (using maximum likelihood estimates) may be used.

3.3.1 Minimum χ_1^2 by "linearization"

Suppose the hypothesis is defined by $F_t(\underline{p}) = 0$, so that "linearization" gives

$$F_t^*(\underline{q}, \underline{p}) = F_t(\underline{q}) + \sum_i \sum_j \left[\frac{\partial F_t(\underline{p})}{\partial p_{ij}}\right]_{\underline{p}=\underline{q}} (p_{ij} - q_{ij}) = 0$$

$$t = 1,2,\ldots,m\,.$$

Let $$\left[\frac{\partial F_t(\underline{p})}{\partial p_{ij}}\right]_{\underline{p}=\underline{q}} = f_{tij}$$

and $$h_t = F_t(\underline{q}) - \sum_i \sum_j f_{tij}\, q_{ij} \quad .$$

Then by theorem 3.2.1, we have

$$\chi_1^2 = \underline{f}'G^{-1}\underline{f} \quad , \quad \text{d.f.} = m \quad , \quad \text{where}$$

$$\underline{f}' = [F_1(\underline{q}), \ldots, F_m(\underline{q})] \quad .$$

We shall not go into the details of special cases, except for the case discussed in the next subsection. The cases that may be worked out in this way are, for example, H_{01} (1.2.2), H_{03} (1.2.6), H_{06}(1.2.12), H_{026} (1.5.2), H_{027} (1.5.3), H_{028} (1.5.4), H_{029} (1.5.5), H_{030} and H_{031} .

3.3.2 The hypothesis of no interaction (multiplicative set-up) in the two-dimensional design

$$H_0: p_{ijk} = q_{ij*}\, q_{i*k} \qquad [H_{01}, \ (1.2.2)] \ .$$

This may be tested by the "linearization" technique, as mentioned in the last paragraph. On the other hand, in the case of a complete design, the maximum likelihood equations appear to be fairly simple and may admit an iterative solution. It can be shown that $H_0 \Longleftrightarrow$

(3.3.1) $$p_{ijk}\, p_{ist} = p_{isk}\, p_{ijt} \ , \quad \begin{array}{l} i = 1,2,\ldots,r \\ j = 1,2,\ldots,s-1 \\ k = 1,2,\ldots,t-1 \end{array} \ .$$

Then the maximum likelihood equations, subject to (3.3.1) and $\sum_i p_{ijk} = 1$, can be obtained by differentiating

$$f = \sum_{i,j,k} n_{ijk} \log p_{ijk} - \sum_{j=1}^{s} \sum_{k=1}^{t} \lambda_{jk}[\sum_{i} p_{ijk} - 1]$$

$$- \sum_{i=1}^{r} \sum_{j=1}^{s-1} \sum_{k=1}^{t-1} \mu_{ijk}[\log p_{ijk} + \log p_{ist} - \log p_{isk} - \log p_{ijt}]$$

with respect to the p's, where λ's and μ's are Lagrange multipliers. The final equations are

$$(3.3.2) \quad \frac{(n_{ijk} - \mu_{ijk})(n_{ist} - \mu_{ioo})}{(n_{isk} + \mu_{iok})(n_{ijt} + \mu_{ijo})} = \frac{(n_{ojk} - \mu_{ojk})(n_{ost} - \mu_{ooo})}{(n_{osk} + \mu_{ook})(n_{ojt} + \mu_{ojo})} ,$$

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

$$\mu_{iok} = \sum_{j=1}^{s-1} \mu_{ijk} , \quad \mu_{ijo} = \sum_{k=1}^{t-1} \mu_{ijk} ,$$

$$\mu_{ojk} = \sum_{i=1}^{r} \mu_{ijk} , \quad \text{and so on.}$$

In particular, when $r = s = t = 2$, we have just two equations (linear) and these can be explicitly solved. In this special case, Bartlett [3] posed another hypothesis of no interaction, but the solution of the maximum likelihood equation comes out as a root of a certain cubic equation. Mitra [19] shows that it is the numerically smallest real root that gives consistent solution. The equations, in the present case, thus seem to be simpler. Roy and Kastenbaum [32] extended Bartlett's hypothesis to more general cases where "i", "j" and "k" are variables, and they get equations similar to (3.3.2).

3.4 On the hypotheses about association for a single multinomial

Most of the hypotheses, in this case, are nonlinear. Hence, we can always apply Neyman's linearization technique. On the other hand, in some cases, the maximum likelihood equations are very simple, so that χ^2 statistic may be used. This is so, for example, in the cases H_{011} (1.3.2), H_{013} (1.3.5) and H_{014} (1.3.6), which have been considered by Roy and Mitra [33].

There is another possibility, namely, the conditional probability approach, which we shall illustrate by considering the case of three variables (i,j,k). Here

$$\Phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}} \quad , \; [\sum_{i,j,k} p_{ijk} = 1] \; ,$$

which can be written as

$$= \left[\prod_{j,k} \frac{n_{ojk}!}{\prod_i n_{ijk}!} \prod_i \left(\frac{p_{ijk}}{p_{ojk}}\right)^{n_{ijk}}\right] \left[\frac{n!}{\prod_{j,k} n_{ojk}!} \prod_{j,k} p_{ojk}^{n_{ojk}}\right]$$

$$= \Phi_1 \times \Phi_2 \; , \text{ say} \; .$$

Let $\frac{p_{ijk}}{p_{ojk}} = p^*_{ijk}$, so that $\sum_i p^*_{ijk} = 1$. Then Φ_1 denotes the conditional probability density of n_{ijk}'s, given n_{ojk}'s, while Φ_2 denotes the probability density of n_{ojk}'s. Also, the number of independent parameters in Φ, Φ_1 and Φ_2 is $rst - 1$, $st(r-1)$ and $st-1$, respectively. We may consider

p^*_{ijk}'s and p_{ojk}'s instead of p_{ijk}'s. Then it is logical to require that the hypotheses, which are expressed in terms of p^*_{ijk}'s only, should be tested by criteria based on Φ_1 only. Hence the test on $p^{*'}$s only will be the same as that on p's if "j" and "k" were factors.

This approach seems to be similar to the "step-down procedure" in normal multivariate analysis [e.g. [28], [30]]. This procedure reduces the problems of association to problems of analysis of variance. Exactly the same thing seems to be happening here by the conditional probability approach. We shall illustrate by considering three simple examples.

(i) <u>Hypothesis of independence in two-dimensional table</u>

$$H_0: \quad p_{ij} = p_{io}\, p_{oj} \quad .$$

Hence $p^*_{ij} = \dfrac{p_{ij}}{p_{oj}}$ is independent of j . Thus H_0 , by the conditional probability approach, is equivalent to

$$H^*_0: \quad p^*_{ij} = q_{i*} \quad ,$$

the hypothesis of homogeneity when "j" is a factor. It is well known that

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n_{io}n_{oj}/n)^2}{n_{io}n_{oj}/n} \quad , \quad d.f. = (r-1)(s-1)$$

may be used to test both H_0 and H^*_0 .

(ii) <u>Hypothesis of "no partial association" in a three-dimensional table</u>

$$H_0: \quad p_{ijk} = \frac{p_{iok}\, p_{ojk}}{p_{ook}} \quad .$$

Then $p^*_{ijk} = \frac{p_{ijk}}{p_{ojk}}$ is independent of j. Thus H_0, by the conditional probability approach, is equivalent to

$$H_0^*: \quad p^*_{ijk} = q_{i*k} \quad ,$$

the hypothesis of "no treatment effects" when "j" and "k" are factors. It has already been noticed [19] that

$$\chi^2 = \sum_i \sum_j \sum_k \frac{\left(n_{ijk} - \frac{n_{ojk} n_{iok}}{n_{ook}}\right)^2}{n_{ojk} \frac{n_{iok}}{n_{ook}}} \text{, d.f.} = t(r-1)(s-1)$$

may be used to test both H_0 and H_0^*.

(iii) <u>Hypothesis of "multiple independence" in a three-dimensional table</u>

$$H_0: \; p_{ijk} = p_{ioo} \, p_{ojk} \quad .$$

As noted earlier

$p^*_{ijk} = \frac{p_{ijk}}{p_{ojk}}$ is independent of j and k. Thus H_0, by the conditional probability approach, is equivalent to

$$H_0^*: \quad p^*_{ijk} = q_{i**} \quad ,$$

when "j" and "k" are factors. It may be seen that

$$\chi^2 = \sum_i \sum_j \sum_k \frac{\left(n_{ijk} - \frac{n_{ioo} \, n_{ojk}}{n}\right)^2}{\frac{n_{ioo} \, n_{ojk}}{n}} \text{, d.f.} = (r-1)(st-1) \; ,$$

may be used to test both H_0 and H_0^*.

Applications

(i) $H_0 : \sum_i a_i p_{ijk} \Big/ p_{ojk}$ is independent of j

[H_{017} , (1.3.12)]. It is equivalent to

$$H_0^* : \sum_i a_i p_{ijk}^* = q_{**k} \quad ,$$

when "j" and "k" are factors. H_0^* has already been considered in 3.2 and the appropriate χ_1^2 is given by (3.2.11). Hence, the same statistic may be used for H_0 .

(ii) $H_0 : \sum_i a_i \; p_{ijk} \big/ p_{ojk}$ is independent of (jk)

[H_{019}, (1.3.16)]

It is equivalent to

$$H_0^* : \sum_i a_i \; p_{ijk}^* \text{ is independent of (jk)} \quad ,$$

when "j" and "k" are factors. The criterion for H_0^* can be easily derived and the same may be used for H_0.

Similarly $H_{016}^{(1)}$ (1.3.11), H_{020} (1.3.18), H_{021} (1.3.19), H_{022} (1.3.20) will follow from corresponding cases when "j" and "k" are factors.

P A R T I I

NONPARAMETRIC SET-UP

CHAPTER IV

SOME UNIVARIATE PROBLEMS

4.1 Introduction

Much of the usual analysis of variance rests on the assumption of normality and homoscedasticity. When these assumptions are not realistic, two different kinds of approaches have been made so far. One is the transformation of variates and the other is a nonparametric development of the whole problem.

Transformation of variates with a view to "normalizing" and "stabilizing the variance" has been in vogue for a long time and has, by and large, served a useful purpose. But, at the same time, one feels that there are some dangers in the indiscriminate use of such procedures. We should try to pose physically meaningful models and then state hypotheses or, in general, decision problems in terms of the original variates themselves. If we make a transformation even before we have posed any model and hypothesis, then proceed with the usual analysis of variance and reach some conclusions on that basis, such conclusions may not have much physical meaning in terms of the original variates and, in any case, such conclusions should be interpreted in terms of the original variates.

Under the non-parametric approach, various methods have been suggested so far to avoid the assumption of normality implicit in the analysis of variance. Friedman [13] made use of ranks in the problem of m rankings. His χ_r^2 statistic, to test the hypothesis H_0 that the rankings by m "observers" of n "objects" are independent, essentially offers a rank test for two-way classification with one observation per cell. For large m, when H_0 is true, χ_r^2 is distributed asymptotically as a χ^2 with $n-1$ d.f. Durbin [9] has given a generalization for the balanced incomplete block design. Benard and Van Elteren [4] have generalized it still further.

Fisher and Yates [12] proposed that each observation be replaced not by its rank but by its normalized rank, defined as the average value of an observation having the corresponding rank in samples of the same size from $N(0,1)$. They proposed that ordinary one-way analysis of variance be applied to these normalized ranks. The argument seems that χ^2 approximation might be approached more rapidly with normalized ranks or some other set of numbers which resemble normal form more than do ranks.

Another technique that has been suggested to get around the assumption of normality, is the use of tests based on permutations. Permutation tests, which seem to have been first proposed by Fisher [11], accept or reject the null hypothesis according to the probability of a test statistic among all relevant permutations of the observed numbers. Application of the permutation method to important cases may be found in articles by Pitman [26] and by Welch [37]. Kruskal and Wallis [16] have proposed an analogue (of one-way F-test) based on ranks and called the H test, to decide whether c samples come from the same population (assuming that the populations are approximately of the same form, in the sense that if they differ it is by a shift or translation). If the samples come from identical populations and the sample sizes are not too small, H is distributed as a χ^2 with $c-1$ d.f. When $c=2$, the H test is essentially the same as Wilcoxon's test [39].

Mood and Brown [20] generalize to c samples the test proposed for two samples by Westenberg [38], utilizing the number of observations above the median of the combined sample. Massey [18] extends Mood's technique to use other order statistics, besides the median, from combined samples. Mood and Brown consider also the two-way classification with one observation per cell or the same number of observations per cell.

Mosteller [21] proposed a multi-decision procedure for accepting either H_0 (that c samples come from the same population) or one of c alternatives that the i-th population is translated to the right. His criterion is in terms of the number of observations in the sample containing the largest observation that exceed all observations in the samples.

Terpstra [34] gives a statistic (for the problem of c samples) for testing against trend. His statistic is based on Wilcoxon's statistic for all pairs of samples and is asymptotically normal. He also [35] gives a test based on a statistic Q, again based on Wilcoxon's statistics for all pairs of samples and shows that Q is asymptotically χ^2 with $\binom{c}{2}$ d.f.

In this chapter, we shall first extend Mood's test for two-way classification to cover incomplete block situations. Then, a new test for the problem of c samples is offered. For this purpose, Hoeffding's theorem on U-statistics [14] extended by Lehmann [17] to the case of two samples, has been extended in a straightforward manner to the case of c samples. This is, then, applied to derive a new test-criterion for c samples. The statistic derived may be considered as an extension of Wilcoxon's statistic.

4.2 Extension of Mood's test for two-way classification to cover "incomplete designs"

Mood and Brown [20] have considered a test for equality of "row" effects in the usual set up with one observation per cell or h observations per cell, with r rows and c columns. The distributions of x_{ij} have medians $v_{ij} = \alpha_i + \beta_j + v$ where the median of the numbers

α_i is zero as is the median of β_j. The distributions of the x_{ij}'s are assumed to be continuous and identical, except for location.

Under the null hypothesis that the row effects α_i are equal (i.e., zero), all the observations in a given column have the same distribution. Let $\tilde{x}_j$ be the median of observations in the j-th column, and in the two-way table let the observation x_{ij} be replaced by a plus sign if it exceeds $\tilde{x}_j$, or by a minus sign if it does not. Let m_i be the number of plus signs in the i-th row. The test criterion used is

$$\chi^2 = \frac{r(r-1)}{ca(r-a)} \sum_{i=1}^{r} \left(m_i - \frac{ca}{r}\right)^2 \quad , \tag{4.2.1}$$

where $a = \frac{r}{2}$ if r is even or $\frac{r-1}{2}$ if r is odd. Unless c is small, the χ^2 approximation with $(r-1)$ d.f. is used. For practical purposes the large sample distribution is satisfactory if $c \geq 10$ or even if $c = 5$ provided $rc \geq 20$. For smaller c, exact probabilities could be computed. We shall consider the generalization to incomplete blocks.

4.2.1 Let us write $n_{ij} = 1$ if (ij) combination is allowed or zero otherwise. Let the number of observations in the i-th row be c_i $(i = 1, 2, \ldots, r)$ and in the j-th column be r_j $(j = 1, 2, \ldots, c)$. Let $a_j = r_j/2$ if r_j is even or $\frac{r_j - 1}{2}$ if r_j is odd. Then there are a_j plus signs in the j-th column. Let m_i's be defined as before. Then we expect (under H_0) m_i to be approximately equal to $c_i/2$.

Following Mood, let us derive the generating function to exhibit the distribution of the m_i's. Suppose t_i is associated with a plus sign in the i-th row $(i=1,2,\ldots,r)$. Let $\Phi_{a_j}(t_1, \ldots, t_{r_j})$ consist of the sum of all terms that can be formed by multiplying t's together a_j at a time. Each term of Φ describes a possible arrangement of signs in a given column. Furthermore, each arrangement of signs is equally likely; hence the probability of a particular arrangement is $1\Big/\binom{r_j}{a_j}$.

Suppose the j-th column contains observations in the $j_1, j_2, \ldots, j_{r_j}$-th rows. Then consider the function

$$\Phi = \prod_{j=1}^{c} \frac{\Phi_{a_j}(t_{j_1}, \ldots, t_{j_{r_j}})}{\binom{r_j}{a_j}} \quad . \tag{4.2.2}$$

There is a one-to-one correspondence between ways of getting terms $t_1^{m_1}\ t_2^{m_2}\ \ldots\ t_r^{m_r}$ in the numerator of Φ and arrangements of signs in the $r \times c$ table which gives rise to m_i plus signs in the i-th row $(i = 1,2,\ldots,r)$. Hence

$$\Phi = \sum_{m_1} \sum_{m_2} \cdots \sum_{m_r} g(m_1, \ldots, m_r) t_1^{m_1}\ t_2^{m_2}\ \ldots\ t_r^{m_r} \ , \tag{4.2.3}$$

where g is the density function for the m_i's.

Note that $\Phi_{a_j}(1,1,\ldots,1) = \binom{r_j}{a_j}$. Φ is thus a factorial-moment generating function for the m_i's. Then $\mathcal{E}(m_i) = \frac{\partial\Phi}{\partial t_i}$ with all t_i's = 1 . We note that

$$\frac{\partial\Phi_{a_j}}{\partial t_i}(t_{j_1},\ldots,t_{j_{r_j}}) = 0 \quad \text{if} \quad n_{ij} = 0$$

$$= \Phi_{a_j-1}(t_{j_1},\ldots,t_{j_{r_j}}) \quad \text{if} \quad n_{ij} = 1 \, ,$$

where one of the t's from the previous bracket is missing.

Hence,

$$(4.2.4) \quad \frac{\partial\Phi}{\partial t_i} = \frac{1}{\prod_{j=1}^{c}\binom{r_j}{a_j}}\left[\sum_{j=1}^{c}\left\{\prod_{j'\neq j}\Phi_{a_{j'}}(t_{j'_1},\ldots,t_{j'_{r_{j'}}})\right\} \times n_{ij}\Phi_{a_j-1}(t_{j_1}\ldots)\right] .$$

Then,

$$(4.2.5) \quad \mathcal{E}(m_i) = \frac{1}{\prod_{j=1}^{c}\binom{r_j}{a_j}}\left[\sum_{j=1}^{c}\left\{\prod_{j'\neq j}\binom{r_{j'}}{a_{j'}}\right\}n_{ij}\binom{r_j-1}{a_j-1}\right] = \sum_{j=1}^{c} n_{ij}\,\frac{a_j}{r_j} .$$

Similarly,

$$\sigma_{ii} = \mathrm{var}(m_i) = \left[\frac{\partial^2\Phi}{\partial t_i^2}\right]_{t=1} + \mathcal{E}(m_i) - [\,\mathcal{E}(m_i)]^2 .$$

From (4.2.4), we have

$$\frac{\partial^2\Phi}{\partial t_i^2} = \frac{1}{\prod_{j=1}^{c}\binom{r_j}{a_j}}\left[\sum_{j=1}^{c} n_{ij}\Phi_{a_j-1}(t_{j_1},\ldots,t_{j_{r_j}})\sum_{j'\neq j}\frac{\partial}{\partial t_i}\Phi_{a_{j'}}(t_{j'_1},\ldots,t_{j'_{r_{j'}}}) \times \prod_{k\neq j\neq j'}\Phi_{a_k}(t_{k_1},\ldots,t_{k_{r_k}})\right] .$$

Hence,

$$\left[\frac{\partial^2\Phi}{\partial t_i^2}\right]_{t=1} = \frac{1}{\prod_{j=1}^{c}\binom{r_j}{a_j}}\left[\sum_{j=1}^{c} n_{ij}\binom{r_j-1}{a_j-1}\sum_{j'\neq j} n_{ij'}\binom{r_{j'}-1}{a_{j'}-1}\prod_{k\neq j\neq j'}\binom{r_k}{a_k}\right]$$

$$= \sum_{j=1}^{c}\sum_{j'\neq j} n_{ij}n_{ij'}\frac{a_j}{r_j}\frac{a_{j'}}{r_{j'}} = \left[\sum_{j=1}^{c} n_{ij}\frac{a_j}{r_j}\right]^2 - \sum_{j=1}^{c} n_{ij}^2\frac{a_j^2}{r_j^2}$$

$$= [\,\mathcal{E}(m_i)]^2 - \sum_{j=1}^{c} n_{ij}\frac{a_j^2}{r_j^2} \quad ,$$

so that

$$\text{(4.2.6)} \qquad \sigma_{ii} = \sum_{j=1}^{c} n_{ij}\left(\frac{a_j}{r_j} - \frac{a_j^2}{r_j^2}\right) \quad .$$

Similarly,

$$\sigma_{ij} = \operatorname{cov}(m_i, m_j) = \left[\frac{\partial^2\Phi}{\partial t_i \partial t_j}\right]_{t=1} - \mathcal{E}(m_i)\,\mathcal{E}(m_j) \quad .$$

From (4.2.4), we have

$$\frac{\partial^2\Phi}{\partial t_i \partial t_j} = \frac{1}{\prod_{k=1}^{c}\binom{r_k}{a_k}}\left[\sum_{k=1}^{c} n_{ik}\left\{\prod_{j'\neq k}\Phi_{a_{j'}} n_{jk}\Phi_{a_k-2} + \Phi_{a_k-1} \times \sum_{j'\neq k} n_{jj'}\Phi_{a_{j'}-1}\prod_{\ell\neq k\neq j'}\Phi_{a_\ell}\right\}\right] \quad .$$

Hence,

$$\left(\frac{\partial^2\Phi}{\partial t_i \partial t_j}\right)_{t=1} = \sum_{k=1}^{c} n_{ik}\left\{n_{jk}\frac{\binom{r_k-2}{a_k-2}}{\binom{r_k}{a_k}} + \frac{\binom{r_k-1}{a_k-1}}{\binom{r_k}{a_k}}\sum_{j'\neq k} n_{jj'}\frac{\binom{r_{j'}-1}{a_{j'}-1}}{\binom{r_{j'}}{a_{j'}}}\right\}$$

$$= \sum_{k=1}^{c} n_{ik}n_{jk}\frac{a_k(a_k-1)}{r_k(r_k-1)} + \sum_{k=1}^{c} n_{ik}\frac{a_k}{r_k}\sum_{j'\neq k} n_{jj'}\frac{a_{j'}}{r_{j'}} \; ,$$

so that

(4.2.7) $$\sigma_{ij} = -\sum_{k=1}^{c} n_{ik}n_{jk}\,\frac{a_k}{r_k^2}\,\frac{r_k - a_k}{r_k - 1} \quad .$$

4.2.2 Asymptotic normality

We have

$$\Phi(\underline{t}) = \frac{1}{\prod_{j=1}^{c}\binom{r_j}{a_j}} \prod_{j=1}^{c} \Phi_{a_j}(t_{j_1},\ldots,t_{j_{r_j}}) = \sum_{m_1}\cdots\sum_{m_r} g(m_1,\ldots,m_r) \times t_1^{m_1} \cdots t_r^{m_r} \; .$$

Replacing t_i in Φ by $e^{s_i/\sqrt{c_i}}$, we have

$$\Phi'(\underline{s}) = \sum_{m_1}\cdots\sum_{m_r} g(m_1,\ldots,m_r)\, e^{\sum_i \frac{m_i s_i}{\sqrt{c_i}}}$$

$$= \text{moment generating function of } \frac{m_i}{\sqrt{c_i}}\text{'s} \; .$$

Let us consider $\log \Phi'$ for large c_i's. We have

(4.2.8) $$\log \Phi' = \sum_{j=1}^{c} \log \Phi'_{a_j}\Big/\binom{r_j}{a_j} \quad .$$

Now $$\frac{1}{\binom{r_j}{a_j}}\Phi'_{a_j} = \frac{1}{\binom{r_j}{a_j}}\sum e^{\sum_{i=1}^{a_j} \frac{s_{j_{k_i}}}{\sqrt{c_{j_{k_i}}}}} \; ,$$

where the summation is over $\binom{r_j}{a_j}$ combinations of type $k_1, k_2, \ldots, k_{a_j}$ out of $(1, 2, \ldots, r_j)$. Hence,

$$\frac{1}{\binom{r_j}{a_j}}\,\Phi'_{a_j} = \frac{1}{\binom{r_j}{a_j}}\sum\left\{1+\sum_{i=1}^{a_j}\frac{s_{j_{k_i}}}{\sqrt{c_{j_{k_i}}}}+\frac{1}{2}\left[\sum_{i=1}^{a_j}\frac{s_{j_{k_i}}}{\sqrt{c_{j_{k_i}}}}\right]^2+O(c^{-3/2})\right\}$$

$$= \frac{1}{\binom{r_j}{a_j}}\left[\binom{r_j}{a_j}+\sum_{i=1}^{r} n_{ij}\binom{r_j-1}{a_j-1}\frac{s_i}{\sqrt{c_i}}+\frac{1}{2}\sum_{i=1}^{r} n_{ij}\binom{r_j-1}{a_j-1}\frac{s_i^2}{c_i}\right.$$

$$\left.+\frac{1}{2}\sum_{i'\neq i}\sum n_{ij}n_{i'j}\binom{r_j-2}{a_j-2}\frac{s_i s_{i'}}{\sqrt{c_i c_{i'}}}+O(c^{-3/2})\right]$$

$$= 1+\sum_{i=1}^{r} n_{ij}\frac{a_j}{r_j}\frac{s_i}{\sqrt{c_i}}+\frac{1}{2}\sum_{i=1}^{r} n_{ij}\frac{a_j}{r_j}\frac{s_i^2}{c_i}+\frac{1}{2}\sum_{i'\neq i}\sum n_{ij}n_{i'j}\times$$

$$\frac{a_j(a_j-1)}{r_j(r_j-1)}\frac{s_i s_{i'}}{\sqrt{c_i c_{i'}}}+O(c^{-3/2})\ ,$$

so that

$$\log\frac{1}{\binom{r_j}{a_j}}\,\Phi'_{a_j} = \sum_{i=1}^{r} n_{ij}\frac{a_j}{r_j}\frac{s_i}{\sqrt{c_i}}+\frac{1}{2}\sum_{i=1}^{r} n_{ij}\frac{a_j}{r_j}\frac{s_i^2}{c_i}+\frac{1}{2}\sum_{i'\neq i}\sum n_{ij}n_{i'j}\times$$

$$\frac{a_j(a_j-1)}{r_j(r_j-1)}\frac{s_i s_{i'}}{\sqrt{c_i c_{i'}}}-\frac{1}{2}\left[\sum_{i=1}^{r} n_{ij}\frac{a_j}{r_j}\frac{s_i}{\sqrt{c_i}}\right]^2+O(c^{-3/2})\ .$$

Then, from (4.2.8), we have

$$\log\Phi' = \sum_{i=1}^{r}\mathcal{E}(m_i)\frac{s_i}{\sqrt{c_i}}+\frac{1}{2}\sum_{i=1}^{r}\sigma_{ii}\frac{s_i^2}{c_i}+\frac{1}{2}\sum_{j\neq i}\sum\sigma_{ij}\frac{s_i s_j}{\sqrt{c_i c_j}}+O(c^{-3/2})\ .$$

Thus for large c_i's, we have the distribution of $\frac{m_i}{\sqrt{c_i}}$'s

approximated by the multivariate normal distribution. Hence,

$$(4.2.9) \qquad \underline{m} \sim N(\mathcal{E}(\underline{m}),\ [\sigma_{ij}]) \ .$$

Since $\sum_{i=1}^{r} m_i = \sum_{j=1}^{c} a_j$, it follows that m_i's are linearly dependent. Hence the above distribution is singular. Considering only $m_1, m_2, \ldots, m_{r-1}$, which have an asymptotic nonsingular normal distribution, we shall have a chi-square criterion with $r-1$ d.f., given by

$$(4.2.10) \qquad \chi^2 = \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} [m_i - \mathcal{E}(m_i)][m_j - \mathcal{E}(m_j)]\sigma^{ij}_{(rr)} \ ,$$

where $(\sigma^{ij}_{(rr)}) = \Sigma^{-1}_{(rr)}$, $\Sigma_{(rr)}$ being the cofactor of σ_{rr} in (σ_{ij}) .

4.2.3 Special case

Suppose $c_1 = c_2 = \ldots = c_r = c_0$ say

and $r_1 = r_2 = \ldots = r_c = r_0$ say .

Then $a_1 = a_2 = \ldots = a_c = a_0$ say, where

$$a_0 = \frac{r_0}{2} \text{ if } r_0 \text{ is even}$$

$$= \frac{r_0 - 1}{2} \text{ otherwise.}$$ Also $rc_0 = cr_0$.

Then, from (4.2.5), $\mathcal{E}(m_i) = \frac{a_0}{r_0} c_0$, $i = 1,2,\ldots,r$, and

$$\sigma_{ii} = \frac{c_0 a_0}{r_0^2}(r_0 - a_0), \quad i=1,2,\ldots,r, \text{ from } (4.2.6),$$

$$(4.2.11) \qquad \sigma_{ij} = -\frac{a_0(r_0 - a_0)}{r_0^2(r_0 - 1)} \lambda_{ij} \quad j \neq i \quad , \text{ from } (4.2.7) ,$$

where $\lambda_{ij} = \sum_k n_{ik}\, n_{jk}$.

(i) Balanced Incomplete Block Designs:

Let $\lambda_{ij} = \lambda$ for all (ij), ($i \neq j$). Then we have $c_0(r_0 - 1) = \lambda(r - 1)$. Also,

$$\sigma_{ii} = \frac{c_0 a_0}{r_0^2}(r_0 - a_0) = \alpha \quad \text{say , and}$$

$$\sigma_{ij} = \frac{a_0(r_0 - a_0)}{r_0^2(r_0 - 1)}\lambda = \beta \quad \text{say .}$$

Thus, $\Sigma_{(rr)} = (\alpha - \beta)I_{r-1} + \beta J_{r-1}$.

Then $$\Sigma^{-1}_{(rr)} = \frac{1}{\alpha - \beta} I_{r-1} - \frac{\beta}{(\alpha - \beta)[\alpha + (r-2)\beta]} J_{r-1}$$

$$= \gamma I_{r-1} + \delta J_{r-1} \quad \text{say .}$$ Let

$\underline{z}'_{1\times(r-1)} = [(m_i - \frac{a_0 c_0}{r_0}),\ i = 1,2,\ldots,r-1]$. Then, from (4.2.10),

$$\chi^2 = \underline{z}'\Sigma^{-1}_{(rr)}\underline{z} = \gamma\, \underline{z}'\underline{z} + \delta\, \underline{z}'\, J_{r-1}\, \underline{z} \quad .$$

Now $$\underline{z}'J_{r-1}\,\underline{z} = \left(\sum_{i=1}^{r-1} z_i\right)^2 = \left[\sum_{i=1}^{r-1}\left(m_i - \frac{c_0 a_0}{r_0}\right)\right]^2$$

$$= \left[a_0 c_0 - m_r - \frac{(r-1)c_0 a_0}{r_0}\right]^2 = \left(m_r - \frac{c_0 a_0}{r_0}\right)^2 \quad .$$

It can also be seen that $\gamma = \delta$, so that

$$\text{(4.2.12)} \qquad \chi^2 = \frac{r_0^2(r_0 - 1)}{a_0(r_0 - a_0)\lambda r} \sum_{i=1}^{r}\left(m_i - \frac{c_0 a_0}{r_0}\right)^2 \quad .$$

If we put $\lambda = c$ and hence $r_0 = r$, $c_0 = c$ and $a_0 = a$, we get back to (4.2.1).

(ii) Partially Balanced Incomplete Block Designs:

Let us consider rows as treatments, so that $\lambda_{ij} = \lambda_\ell$ if i and j are ℓ-th associates. Then $\Sigma = \alpha I_r + \sum_{\ell=1}^{m} \beta_\ell B_\ell$, where m is the number of associate classes, α is defined as before, B's are association matrices [6] and

$$\beta_\ell = -\frac{a_0(r_0 - a_0)}{r_0^2(r_0 - 1)} \lambda_\ell \quad .$$

Let $\beta_0 = \frac{a_0(r_0 - a_0)c_0}{r_0^2}$ and $q = \frac{a_0(r_0 - a_0)}{r_0(r_0 - 1)}$. Then

$$\Sigma = \sum_{\ell=0}^{m} \beta_\ell B_\ell \text{ , where } B_0 = I \text{ ,}$$

$= q\ \alpha$ in the notation of [5].

Using the results derived in [5] and simplifying, we have

$$\chi^2 = \frac{r_0(r_0 - 1)}{a_0(r_0 - a_0)} \sum_{i=1}^{r} \sum_{j=1}^{r} c_{ij}\left(m_i - \frac{c_0 a_0}{r_0}\right)\left(m_j - \frac{c_0 a_0}{r_0}\right) ,$$

where $C = (c_{ij})$ is such that the solution of "normal equations" for $\underline{t}$ in the analysis of variance for PBIBD is given by $\underline{t} = C\underline{Q}$, $\underline{Q}$ being defined in the usual notation.

4.3 Generalization of Hoeffding's theorem on U-statistics to c samples

Let $x_1, x_2, \ldots, x_{n_1}$ be independent and identically distributed (real or vector-valued) random variables with distribution function F. Similarly let $y_1, y_2, \ldots, y_{n_2}$ be with distribution function G and so on ... and

$z_1, z_2, \ldots, z_{n_c}$ be with distribution function H. Consider

$$(4.3.1) \quad u = u_{n_1, n_2, \ldots, n_c} = \frac{1}{\prod_{i=1}^{c} n_i(n_i-1)\ldots(n_i - m_i + 1)} \times$$

$$\sum \Phi(x_{\alpha_1}, \ldots, x_{\alpha_{m_1}}; y_{\beta_1}, \ldots, y_{\beta_{m_2}}; \ldots; z_{\gamma_1}, z_{\gamma_2}, \ldots, z_{\gamma_{m_c}}) ,$$

where $\sum$ denotes the sum over all permutations of $(\alpha_1, \alpha_2, \ldots, \alpha_{m_1})$ such that $1 \leq \alpha_i \leq n_i$ and are all different, and similarly for β's, ..., and γ's. Let

$$(4.3.2) \quad \Phi^*(x_1, \ldots, x_{m_1}; y_1, \ldots, y_{m_2}; \ldots; z_1, z_2, \ldots z_{m_c})$$

$$= \frac{1}{\prod_{i=1}^{c}(m_i)!} \sum \Phi(x_{\alpha_1}, \ldots, x_{\alpha_{m_1}}; y_{\beta_1}, \ldots, y_{\beta_{m_2}}; \ldots; z_{\gamma_1}, \ldots, z_{\gamma_{m_c}}) ,$$

where $\sum$ here denotes the sum over all permutations $(\alpha_1, \ldots, \alpha_{m_1})$ of the m_1 integers $(1, 2, \ldots, m_1)$ and so on for β's, ..., and for γ's. Then Φ^* is symmetric in x's, symmetric in y's, and so on. Then

$$(4.3.3) \quad u = \frac{1}{\prod_{i=1}^{c}\binom{n_i}{m_i}} \sum \Phi^*(x_{\alpha_1}, \ldots, x_{\alpha_{m_1}}; \ldots; z_{\gamma_1}, \ldots, z_{\gamma_{m_c}}) ,$$

where $\sum$ denotes the sum over all combinations $(\alpha_1, \ldots, \alpha_{m_1})$ of m_1 integers chosen from $(1, 2, \ldots, n_1)$, and so on for β's, ..., γ's.

We shall now assume that

$$(4.3.4) \qquad \mathcal{E}(\Phi) = \eta \quad \text{and} \quad \mathcal{E}(\Phi^2) < \infty .$$

Then,

$$(4.3.5) \qquad \mathcal{E}(\Phi^*) = \eta \quad \text{and} \quad \mathcal{E}(\Phi^{*2}) < \infty ,$$

from Schwarz's inequality. Then,

$$\mathcal{E}(U) = \eta \quad . \tag{4.3.6}$$

Also,

$$\operatorname{var}(U) = \mathcal{E}(U-\eta)^2 = \mathcal{E}\left\{\frac{1}{\prod_{i=1}^{c}\binom{n_i}{m_i}} \sum \Psi^*(X_{\alpha_1},\ldots,X_{\alpha_{m_1}};\ \ldots;\ Z_{\gamma_1},\ldots,Z_{\gamma_{m_c}})\right\}^2,$$

where $\Psi^* = \Phi^* - \eta$. Thus,

$$\left[\prod_{i=1}^{c}\binom{n_i}{m_i}\right]^2 \operatorname{var}(U) = \mathcal{E}\sum\sum \Psi^*(X_{\alpha_1},\ldots,X_{\alpha_{m_1}};\ldots; Z_{\gamma_1},\ldots,Z_{\gamma_{m_c}})\Psi^*(X_{\alpha_1'},\ldots,X_{\alpha_{m_1}'};\ldots;Z_{\gamma_1'},\ldots,Z_{\gamma_{m_c}'}) \quad . \tag{4.3.7}$$

Let

$$\Psi^*_{d_1,d_2,\ldots,d_c}(x_1,\ldots,x_{d_1};y_1,\ldots,y_{d_2};\ldots;z_1,z_2,\ldots,z_{d_c}) = \mathcal{E}\Psi^*(x_1,\ldots,x_{d_1},X_{d_1+1},\ldots,X_{m_1};\ldots;z_1,z_2,\ldots,z_{d_c}, Z_{d_c+1},\ldots,Z_{m_c}) \tag{4.3.8}$$

and

$$\xi_{d_1,d_2,\ldots,d_c} = \operatorname{var} \Psi^*_{d_1,d_2,\ldots,d_c}(X_1,\ldots,X_{d_1};\ldots; Z_1,\ldots,Z_{d_c}) = \mathcal{E}\,\Psi^{*2}_{d_1,d_2,\ldots,d_c} \quad , \tag{4.3.9}$$

which exists in view of Schwarz's inequality and (4.3.5), and

$$\xi_{0,0,\ldots,0} = 0 \ .$$

Then, from (4.3.7),

$$\prod_{i=1}^{c}\binom{n_i}{m_i}^2 \operatorname{var}(U) = \sum_{d_1=0}^{m_1}\cdots\sum_{d_c=0}^{m_c} N_{d_1,d_2,\ldots,d_c}\ \xi_{d_1,d_2,\ldots,d_c} \quad ,$$

where $N_{d_1,d_2,\ldots,d_c}$ = number of combinations that can be formed by choosing m_1 distinct integers α's and m_1

distinct $\alpha''s$ out of $(1,2,\ldots,n_1)$ such that exactly d_1 are common between α's and $\alpha''s$ and similarly for β's ... and γ's. Then we can see that

$$N_{d_1,d_2,\ldots,d_c} = \prod_{i=1}^{c} \binom{n_i}{m_i}\binom{m_i}{d_i}\binom{n_i - m_i}{m_i - d_i} .$$

Thus

$$\text{(4.3.10)} \quad \operatorname{var}(U) = \frac{1}{\prod_{i=1}^{c}\binom{n_i}{m_i}} \sum_{d_1=0}^{m_1} \cdots \sum_{d_c=0}^{m_c} \prod_{i=1}^{c} \binom{m_i}{d_i}\binom{n_i - m_i}{m_i - d_i} \xi_{d_1,d_2,\ldots,d_c} .$$

4.3.1 <u>Asymptotic behavior of var(U)</u>

Let us study var(U) as n_i's $\to \infty$ in such a way that

$$\frac{n_i}{\sum_{i=1}^{c} n_i} \to p_i > 0 , \quad i = 1,2,\ldots,c . \quad \text{Then,}$$

$$\frac{\binom{m_i}{d_i}\binom{n_i - m_i}{m_i - d_i}}{\binom{n_i}{m_i}} \approx \frac{\binom{m_i}{d_i}(n_i - m_i)^{m_i - d_i}}{(m_i - d_i)!} \; \frac{m_i!}{n_i^{m_i}}$$

$$\approx \frac{m_i!}{(m_i - d_i)!} \binom{m_i}{d_i} \frac{1}{n_i^{d_i}} .$$

Therefore,

$$\operatorname{var}(U) \approx \sum_{d_1=0}^{m_1} \cdots \sum_{d_c=0}^{m_c} \prod_{i=1}^{c} \frac{m_i!}{(m_i - d_i)!} \binom{m_i}{d_i} \frac{1}{n_i^{d_i}} \xi_{d_1,d_2,\ldots,d_c} .$$

Hence,

$$\text{(4.3.11)} \qquad \operatorname{var}(U) = \sum_{i=1}^{c} \frac{m_i^2}{n_i} \xi_{0,0,\ldots,0,1,0,\ldots,0} + O(N^{-2}) ,$$

$$\text{where } N = \sum_{i=1}^{c} n_i .$$

4.3.2 Asymptotic normality

We shall now show that $\sqrt{N}(U_{n_1,n_2,\ldots,n_c} - \eta) \approx N(0,\sigma^2)$, where

$$(4.3.12) \qquad \sigma^2 = \sum_{i=1}^{c} \frac{m_i^2}{p_i}\, \xi_{0,\ldots,0,1,0,\ldots,0} \quad ,$$

and $\frac{n_i}{N} \to p_i > 0$.

<u>Proof</u>: Let $V_N = \sqrt{N}(U_{n_1,\ldots,n_c} - \eta)$ and

$$W_N = \frac{m_1}{\sqrt{n_1 p_1}} \sum_{\alpha=1}^{n_1} \Psi^*_{1,0,\ldots,0}(X_\alpha) + \cdots + \frac{m_c}{\sqrt{n_c p_c}} \sum_{\gamma=1}^{n_c} \Psi^*_{0,0,\ldots,0,1}(Z_\gamma)$$

$$= W_{1,n_1} + \cdots + W_{c,n_c} \qquad \text{say,}$$

where $W_{1,n_1} = \frac{m_1}{\sqrt{n_1 p_1}} \sum_{\alpha=1}^{n_1} \Psi^*_{1,0,\ldots,0}(X_\alpha)$ and so on. Then W_{1,n_1} is a sum of n_1 independent and identical random variables with zero mean and variance $\frac{m_1^2}{n_1 p_1}\xi_{1,0,\ldots,0}$ and, hence, by the central limit theorem W_{1,n_1} has a normal limiting distribution with mean zero and variance $\frac{m_1^2}{p_1}\xi_{1,0,\ldots,0}$. Similarly for other W_{i,n_i}. Moreover, $W_{1,n_1}, W_{2,n_2}, \ldots W_{c,n_c}$ are independent. Hence W_N has a normal limiting distribution with mean zero and variance σ^2, given by (4.3.12).

We shall now show that $V_N - W_N$ converges to zero in probability. It is sufficient to prove that

$$\mathcal{E}(V_N - W_N)^2 \to 0 \ .$$

Now

$$\mathcal{E}(V_N - W_N)^2 = \mathcal{E}V_N^2 + \mathcal{E}W_N^2 - 2\,\mathcal{E}V_N W_N$$

$$(4.3.13) \qquad = N\Big[\sum_{i=1}^{c} \frac{m_i^2}{n_i}\,\xi_{0,\ldots,0,1,0,\ldots,0} + O(N^{-2})\Big] + \sum_{i=1}^{c} \frac{m_i^2}{p_i}\,\xi_{0,\ldots,0,1,0,\ldots,0} - 2\,\mathcal{E}V_N W_N\,, \text{ and}$$

$$\mathcal{E}\,V_N W_{1,n_1} = \frac{\sqrt{N}}{\prod_{i=1}^{c}\binom{n_i}{m_i}}\,\mathcal{E}\Big[\sum \Psi^*(X_{\alpha_1},\ldots,X_{\alpha_{m_1}};\ldots;Z_{\gamma_1},\ldots,Z_{\gamma_{m_c}})\Big] \times \Big[\frac{m_1}{\sqrt{n_1 p_1}} \sum_{\alpha=1}^{n_1} \Psi^*_{1,0,\ldots,0}(X_\alpha)\Big]$$

$$(4.3.14) \qquad = \frac{\sqrt{N}\;m_1}{\prod_{i=1}^{c}\binom{n_i}{m_i}\sqrt{n_1 p_1}} \sum \sum_{\alpha=1}^{n_1} \mathcal{E}\,\Psi^*(X_{\alpha_1},\ldots,X_{\alpha_{m_1}};\ldots,Z_{\gamma_1},\ldots,Z_{\gamma_{m_c}})\,\Psi^*_{1,0,\ldots,0}(X_\alpha)\;.$$

Expectation of the product term is zero unless α is one of the integers $\alpha_1,\ldots,\alpha_{m_1}$, in which case, expectation is $\xi_{1,0,\ldots,0}$. Hence, from (4.3.14)

$$\mathcal{E}(V_N W_{1,n_1}) = \frac{\sqrt{N}\;m_1}{\sqrt{n_1 p_1}\,\prod_{i=1}^{c}\binom{n_i}{m_i}}\,M_1\,\xi_{1,0,\ldots,0}\quad, \text{ where}$$

M_1 is the number of terms in the summation such that α is one of the integers $\alpha_1,\ldots,\alpha_{m_1}$ for $1 \leq \alpha \leq n_1$ and all possible combinations of α's, β's, ..., and γ's. Thus

$$M_1 = n_1\binom{n_1-1}{m_1-1}\prod_{i=2}^{c}\binom{n_i}{m_i}\quad.$$

Hence,

$$\mathcal{E}(V_N W_{1,n_1}) = \frac{\sqrt{N}\, m_1 n_1}{\sqrt{n_1 p_1}\binom{n_1}{m_1}} \binom{n_1 - 1}{m_1 - 1} \xi_{1,0,\ldots,0}$$

$$= m_1^2 \frac{\sqrt{N}}{\sqrt{n_1 p_1}} \xi_{1,0,\ldots,0} \quad .$$

Similarly for other terms, so that

$$\mathcal{E}(V_N W_N) = \sum_{i=1}^{c} \mathcal{E}(V_N W_{i,n_i}) = \sum_{i=1}^{c} m_i^2 \sqrt{\frac{N}{n_i p_i}}\, \xi_{0,\ldots,0,1,0,\ldots,0} \; .$$

Therefore,

$$\mathcal{E}(V_N - W_N)^2 = \sum_{i=1}^{c} m_i^2 \left\{\frac{N}{n_i} + \frac{1}{p_i} - 2\sqrt{\frac{N}{n_i p_i}}\right\} \xi_{0,\ldots,0,1,0,\ldots,0} + O(N^{-1}) \text{ , and}$$

hence $\to 0$ as $N \to \infty$, since $\frac{n_i}{N} \to p_i$. Since $V_N = W_N + (V_N - W_N)$ and $(V_N - W_N) \to 0$ in probability, V_N is also $\approx N(0, \sigma^2)$.

4.3.3 Extension to a vector U-statistic

Suppose we consider $\underline{\Phi}' = (\Phi^{(1)}, \Phi^{(2)}, \ldots, \Phi^{(g)})$ and $\underline{U}' = (U^{(1)}, U^{(2)}, \ldots, U^{(g)})$, where $U^{(i)}$ is a U-statistic corresponding to $\Phi^{(i)}$. Then under the assumption of existence of second moments of Φ's, we shall have extensions of the previous results.

Let $\mathcal{E}\Phi^{(i)} = \eta^{(i)}$ and $\underline{\eta}' = (\eta^{(1)}, \ldots, \eta^{(g)})$. Also let

$$\xi^{(i,j)}_{d_1, d_2, \ldots, d_c} = \mathcal{E}[\Psi^{*(i)}(X_1, \ldots, X_{d_1}, X_{d_1+1}, \ldots, X_{m_1^{(i)}}, \ldots, Z_1, \ldots, Z_{d_c}, Z_{d_c+1}, \ldots, Z_{m_c^{(i)}}) \times \Psi^{*(j)}(X_1, \ldots, X_{d_1}, X_{m_1^{(i)}+1}, \ldots, X_{m_1^{(j)}+m_1^{(i)}-d_1}; \ldots; Z_1, \ldots, Z_{d_c}, Z_{m_c^{(i)}+1}, \ldots, Z_{m_c^{(j)}+m_c^{(i)}-d_c})]$$

and $\xi^{(i,j)}_{0,0,\ldots,0} = 0$. Then we shall have in a manner as before

$$\prod_{k=1}^{c} \binom{n_k}{m_k^{(i)}} \binom{n_k}{m_k^{(j)}} \operatorname{cov}(U^{(i)}, U^{(j)}) = \sum_{d_1=0}^{m_1^{(ij)}} \cdots \sum_{d_c=0}^{m_c^{(ij)}} N^{(ij)}_{d_1,\ldots,d_c} \times \xi^{(i,j)}_{d_1,\ldots,d_c}$$

where $m_1^{(ij)} = \min(m_1^{(i)}, m_1^{(j)})$, and $N^{(ij)}_{d_1,d_2,\ldots,d_c}$ = number of combinations that can be formed by choosing $m_1^{(i)}$ distinct integers α's and similarly $m_1^{(j)}$ distinct integers α''s out of $(1,2,\ldots,n_1)$ such that exactly d_1 are common between α's and α''s and similarly for β's, ... and γ's. Then,

$$N^{(ij)}_{d_1,d_2,\ldots,d_c} = \prod_{k=1}^{c} \binom{n_k}{m_k^{(i)}} \binom{m_k^{(i)}}{d_k} \binom{n_k - m_k^{(i)}}{m_k^{(j)} - d_k} .$$ Therefore,

$$(4.3.15) \quad \operatorname{cov}(U^{(i)}, U^{(j)}) = \frac{1}{\prod_{k=1}^{c} \binom{n_k}{m_k^{(j)}}} \sum_{d_1=0}^{m_1^{(ij)}} \cdots \sum_{d_c=0}^{m_c^{(ij)}} \prod_{k=1}^{c} \binom{m_k^{(i)}}{d_k} \times \binom{n_k - m_k^{(i)}}{m_k^{(j)} - d_k} \xi^{(i,j)}_{d_1,d_2,\ldots,d_c} .$$

<u>Asymptotic behavior</u>. Let us study the behavior of $\operatorname{cov}(U^{(i)}, U^{(j)})$ for large n_k's. Now

$$\frac{\binom{m_k^{(i)}}{d_k}\binom{n_k - m_k^{(i)}}{m_k^{(j)} - d_k}}{\binom{n_k}{m_k^{(j)}}} \approx \binom{m_k^{(i)}}{d_k} \frac{m_k^{(j)}!}{(m_k^{(j)} - d_k)!} \frac{1}{n_k^{d_k}} .$$

Hence, from (4.3.15), we have

$$\text{(4.3.16)} \quad \mathrm{cov}(U^{(i)},U^{(j)}) = \sum_{k=1}^{c} \frac{m_k^{(i)}\, m_k^{(j)}}{n_k}\, \underset{\substack{\uparrow \\ (k\text{-th place})}}{\xi(i,j)_{0,\ldots,0,1,0,\ldots,0}} + O(N^{-2}) \ .$$

Asymptotic normality. We shall now show that

$$\sqrt{N}(\underline{U}_{n_1,n_2,\ldots,n_c} - \underline{\eta}) \approx N(\underline{O}, \Sigma) \ ,$$

where

$$\text{(4.3.17)} \qquad \sigma_{ij} = \sum_{k=1}^{c} \frac{m_k^{(i)}\, m_k^{(j)}}{p_k}\, \xi(i,j)_{0,\ldots,0,1,0,\ldots,0} \ ,$$

$\Sigma = (\sigma_{ij})$ and $\frac{n_i}{N} \to p_i > 0$.

Proof: Let $\underline{V}_N = \sqrt{N}(\underline{U}_{n_1,\ldots,n_c} - \underline{\eta})$ and

$$W_N^{(i)} = \frac{m_1^{(i)}}{\sqrt{n_1 p_1}} \sum_{\alpha=1}^{n_1} \Psi^{*(i)}_{1,0,\ldots,0}(X_\alpha) + \ldots + \frac{m_c^{(i)}}{\sqrt{n_c p_c}} \sum_{\gamma=1}^{n_c} \times$$

$$\Psi^{*(i)}_{0,0,\ldots,0,1}(Z_\gamma)$$

$$= \sum_{k=1}^{c} W_{k,n_k}^{(i)} \ .$$

Then, as before,

$$\mathrm{var}(W_N^{(i)}) = \sum_{k=1}^{c} \frac{m_k^{2(i)}}{p_k}\, \xi_{0,\ldots,0,1,0,\ldots,0}(i,i) = \sigma_{ii}$$

and, similarly,

$$\mathrm{cov}(W_N^{(i)},W_N^{(j)}) = \sum_{k=1}^{c} \frac{m_k^{(i)} m_k^{(j)}}{p_k}\, \xi(i,j)_{0,\ldots,0,1,0,\ldots,0} = \sigma_{ij} \ .$$

By the central-limit theorem we have asymptotic normality of $\underline{W}_{k,n_k}$ and, since these are independent for $k=1,2,\ldots,c$, we have asymptotic normality of $\underline{W}_N$ with mean-vector zero and covariance matrix Σ. Now

$$\underline{V}_N = \underline{W}_N + (\underline{V}_N - \underline{W}_N) \quad \text{and as before,}$$

$$\mathcal{E}(V_N^{(i)} - W_N^{(i)})^2 \to 0 \quad (i = 1,2,\ldots,g) \quad \text{so that}$$

$$V_N^{(i)} - W_N^{(i)} \to 0 \text{ in probability, that is,}$$

$$\underline{V}_N - \underline{W}_N \to 0 \text{ in probability.}$$

Hence the assertion follows.

4.4 An application to a certain nonparametric test for c samples

Let $x_1, x_2, \ldots, x_{n_1}$ be independent (real-valued) observations from a population with distribution function F. Similarly let $y_1, \ldots, y_{n_2}$ be independent (real-valued) observations from G, ..., and $z_1, \ldots, z_{n_c}$ be from H. We shall assume that the distributions are continuous. We shall consider a certain nonparametric test for the hypothesis

$$H_0: \ F = G = \ldots = H \quad . \tag{4.4.1}$$

If we assume that the populations are approximately of the same form, in the sense that if they differ it is by a shift or translation, then we may say that we are testing for the equality of location parameters. Let

$$\Phi^{(1)}(x_\alpha, y_\beta, \ldots, z_\gamma) = \begin{cases} 1 \text{ if } & x_\alpha < y_\beta, \ldots, x_\alpha < z_\gamma \\ 0 \text{ otherwise} & , \end{cases}$$

and

$$v^{(1)} = \sum_{\alpha=1}^{n_1} \sum_{\beta=1}^{n_2} \cdots \sum_{\gamma=1}^{n_c} \Phi^{(1)}(x_\alpha, y_\beta, \ldots, z_\gamma)$$

= number of c-plets that can be formed by choosing some $x_\alpha, y_\beta, \ldots, z_\gamma$ such that x_α is the smallest.

Here $m_1^{(1)} = m_2^{(1)} = \ldots = m_c^{(1)} = 1$, so that

$$\Phi^{(1)} = \Phi^{*(1)} \quad \text{and} \quad \Psi^{(1)} = \Psi^{*(1)} .$$

Also,

(4.4.2) $$u^{(1)} = \frac{1}{n_1 n_2 \ldots n_c} v^{(1)} \qquad \text{and}$$

(4.4.3) $$\eta^{(1)} = \mathcal{E}[\Phi^{(1)}(X,Y,\ldots,Z)] = \Pr[X < Y,\ldots,X < Z] .$$

If H_0 is true, all orderings of $X,Y,\ldots,Z$ are equally probable, and hence

(4.4.4) $$\eta = \frac{(c-1)!}{c!} = \frac{1}{c} .$$

Also

$$\xi^{(1,1)}_{1,0,\ldots,0} = \mathcal{E}[\Psi^{(1)^2}_{1,0,\ldots,0}(X)] ,$$

where

$$\begin{aligned}\Psi^{(1)}_{1,0,\ldots,0}(x) &= \mathcal{E}\,\Psi^{(1)}(x,Y,\ldots,Z) = \mathcal{E}[\Phi^{(1)}(x,Y,\ldots,Z) - \tfrac{1}{c}] \\ &= [1 - F(x)]^{c-1} - \frac{1}{c} ,\end{aligned}$$

so that

$$\begin{aligned}\xi^{(1,1)}_{1,0,\ldots,0} &= \mathcal{E}[1 - F(X)]^{2c-2} - \frac{1}{c^2} \\ &= \int_{-\infty}^{\infty} [1 - F(x)]^{2c-2}\, dF(x) - \frac{1}{c^2} \\ &= \int_0^1 (1 - t)^{2c-2}\, dt - \frac{1}{c^2} \\ &= \frac{(c-1)^2}{c^2(2c-1)} .\end{aligned}$$

Similarly,

$$\Psi^{(1)}_{0,1,0,\ldots,0}(y) = \mathcal{E}\,\Phi^{(1)}(X,y,\ldots,Z) - \frac{1}{c}$$

$$= \int_{-\infty}^{y}[1 - F(x)]^{c-2}dF(x) - \frac{1}{c}$$

$$= \int_{0}^{F(y)} (1 - t)^{c-2}dt - \frac{1}{c}$$

$$= \frac{1 - [1 - F(y)]^{c-1}}{c - 1} - \frac{1}{c} \quad ,$$

so that

$$\xi^{(1,1)}_{0,1,0,\ldots,0} = \mathcal{E}\left\{\Psi^{(1)^2}_{0,1,0,\ldots,0}(Y)\right\}$$

$$= \mathcal{E}\left[\frac{1 - 2\{1 - F(Y)\}^{c-1} + \{1 - F(Y)\}^{2c-2}}{(c - 1)^2}\right] - \frac{1}{c^2}$$

$$= \frac{1}{c^2(2c - 1)} \quad .$$

Similarly,

$$\xi^{(1,1)}_{0,\ldots,0,1,0,\ldots,0} = \frac{1}{c^2(2c - 1)} \quad .$$

Hence, from (4.3.17), we have

$$(4.4.5) \qquad \sigma_{11} = \frac{1}{c^2(2c - 1)}\left[\frac{(c - 1)^2}{p_1} + \sum_{i=2}^{c}\frac{1}{p_i}\right] \quad ,$$

where $p_i = \dfrac{n_i}{N}$.

In general, if we define

$$\Phi^{(1)}(x_\alpha, y_\beta, \ldots, z_\gamma) = \begin{cases} 1 & \text{if } x_\alpha < y_\beta, \ldots, x_\alpha < z_\gamma \\ 0 & \text{otherwise} \end{cases}$$

$$\Phi^{(2)}(x_\alpha, y_\beta, \ldots, z_\gamma) = \begin{cases} 1 & \text{if } y_\beta < x_\alpha, \ldots, y_\beta < z_\gamma \\ 0 & \text{otherwise} \end{cases}$$

$$\cdots$$

$$\Phi^{(c)}(x_\alpha, y_\beta, \ldots, z_\gamma) = \begin{cases} 1 & \text{if } z_\gamma < x_\alpha,\ z_\gamma < y_\beta, \ldots \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$v^{(k)} = \sum_{\alpha=1}^{n_1} \cdots \sum_{\gamma=1}^{n_c} \Phi^{(k)}(x_\alpha, y_\beta, \ldots, z_\gamma) , \quad (k = 1, 2, \ldots, c)$$

= number of c-plets that can be formed by choosing one observation from each sample such that the observation from the k-th sample is the least.

Then $m_1^{(k)} = m_2^{(k)} = \cdots = m_c^{(k)} = 1$, $u^{(k)} = \frac{1}{n_1 n_2 \cdots n_c} v^{(k)}$, so that $\underline{u} = \frac{1}{\prod_{i=1}^{c} n_i} \underline{v}$. If H_0 is true, $\eta^{(k)} = \frac{1}{c}$, so that $\underline{\eta} = \frac{1}{c} \underline{J}$. Then, similar to (4.4.5), we shall have

$$(4.4.6) \quad \sigma_{ii} = \frac{1}{c^2(2c-1)} \left[\frac{(c-1)^2}{p_i} + \sum_{k \neq i} \frac{1}{p_k} \right] .$$

Now

$$\begin{aligned} \xi_{1,0,\ldots,0}(1,2) &= \mathcal{E}[\Psi^{(1)}(X, Y_1, \ldots, Z_1)\Psi^{(2)}(X, Y_2, \ldots, Z_2)] \\ &= \mathcal{E}[\Phi^{(1)}(X, Y_1, \ldots, Z_1)\Phi^{(2)}(X, Y_2, \ldots, Z_2)] - \frac{1}{c^2} \\ &= \Pr[X < Y_1, \ldots, X < Z_1; Y_2 < X, \ldots, Y_2 < Z_2] - \frac{1}{c^2} \\ &= \mathcal{E}\left\{[1 - F(x)]^{c-1} \int_{-\infty}^{X} [1 - F(y_2)]^{c-2} dF(y_2)\right\} - \frac{1}{c^2} \end{aligned}$$

$$= \mathcal{E}\left\{[1 - F(X)]^{c-1} \int_0^{F(X)} (1 - t)^{c-2} dt\right\} - \frac{1}{c^2}$$

$$= \frac{-(c - 1)}{c^2(2c - 1)} \quad .$$

In general, $\xi(i,j)_{0,\ldots,0,1,0,\ldots,0} = \dfrac{-(c - 1)}{c^2(2c - 1)}$.

(at the i-th or j-th place)

Similarly,

$$\xi(2,c)_{1,0,\ldots,0} = \mathcal{E}[\Psi^{(2)}(X,Y_1,\ldots,Z_1)\Psi^{(c)}(X,Y_2,\ldots,Z_2)]$$

$$= \mathcal{E}[\Phi^{(2)}(X,Y_1,\ldots,Z_1)\Phi^{(c)}(X,Y_2,\ldots,Z_2)] - \frac{1}{c^2}$$

$$= \Pr[Y_1 < X,\ldots,Y_1 < Z_1; Z_2 < X, Z_2 < Y_2,\ldots] - \frac{1}{c^2}$$

$$= \mathcal{E}\left[\int_{-\infty}^{X}[1 - F(y_1)]^{c-2} dF(y_1) \int_{-\infty}^{X}[1 - F(z_2)]^{c-2} dF(z_2)\right] - \frac{1}{c^2}$$

$$= \mathcal{E}\left[\int_0^{F(X)} (1 - t)^{c-2} dt\right]^2 - \frac{1}{c^2}$$

$$= \frac{1}{c^2(2c - 1)} \quad .$$

Hence, from (4.3.17), we have

$$(4.4.7) \quad \sigma_{ij} = \sum_{k=1}^{c} \frac{1}{p_k}\, \xi(i,j)_{0,\ldots,0,1,0,\ldots,0}$$

$$= \frac{1}{c^2(2c - 1)}\left[- \frac{(c - 1)}{p_i} - \frac{(c - 1)}{p_j} + \sum_{\substack{k \neq i \\ \neq j}} \frac{1}{p_k}\right]$$

$$= \frac{1}{c^2(2c - 1)}\left[\sum_{k=1}^{c} \frac{1}{p_k} - \frac{c}{p_i} - \frac{c}{p_j}\right] \quad .$$

Thus from (4.4.6) and (4.4.7), we have

$$(4.4.8) \quad c^2(2c-1)\Sigma = \left(\sum_{k=1}^{c} \frac{1}{p_k}\right) J + c^2 D - c\underline{q}\underline{J}' - c\underline{J}\underline{q}' \quad ,$$

where $D = \text{diagonal } (\frac{1}{p_k}\ , \ k = 1,2,\ldots,c)$ and

$\underline{q}' = \left(\frac{1}{p_1}, \frac{1}{p_2}, \ldots, \frac{1}{p_c}\right)$.

Now $\sum_{k=1}^{c} v^{(k)}$ = number of possible c-plets

$= n_1 n_2 \ldots n_c$.

Hence $u^{(1)},\ldots,u^{(c)}$ are subject to one linear relation $\sum_{k=1}^{c} u^{(k)} = 1$. Hence the distribution of u's is singular and hence the asymptotic distribution should also be singular. Thus, Σ should be singular. In fact we expect $\sum_{i=1}^{c} \sigma_{ii} + \sum_{i\neq j} \sigma_{ij}$ to be zero. Now

$$c^2(2c-1)\underline{J}'\Sigma = \left(\sum_{k=1}^{c} \frac{1}{p_k}\right) \underline{J}'J + c^2\underline{J}'D - c\underline{J}'\underline{q}\underline{J}' - c\underline{J}'\underline{J}\underline{q}' \quad .$$

But $\underline{J'q} = \sum_{k=1}^{c} \frac{1}{p_k}$, $\underline{J}'J = c\underline{J}'$, $\underline{J}'\underline{J} = c$ and $\underline{J}'D = \underline{q}'$,

so that $\underline{J}'\Sigma = 0$. Hence Σ is singular.

Let us consider only $u^{(1)}, u^{(2)}, \ldots, u^{(c-1)}$ and their asymptotic normal distribution. If Σ_0 denotes the covariance-matrix of the asymptotic normal distribution of $u^{(1)}, \ldots, u^{(c-1)}$, then from (4.4.8) we have

$$(4.4.9) \quad c^2(2c-1)\Sigma_0 = \left(\sum_{k=1}^{c} \frac{1}{p_k}\right) J_0 + c^2 D_0 - c\underline{q}_0\underline{J}_0' - c\underline{J}_0\underline{q}_0' \quad ,$$

where $J_0 = (1)_{(c-1)\times(c-1)}$, $\underline{J}_0' = (1)_{1\times(c-1)}$,

$$D_0 = \text{diagonal}\left(\frac{1}{p_1},\ldots,\frac{1}{p_{c-1}}\right) \text{ and } \underline{g}_0' = \left(\frac{1}{p_1},\ldots,\frac{1}{p_{c-1}}\right) .$$

4.4.1 Special case

Let us first consider the special case when $n_1 = n_2 = \ldots = n_c$. Then $p_i = \frac{n_i}{N} = \frac{1}{c}$, so that $D_0 = cI$, $\underline{g}_0' = c\underline{J}_0'$ and hence $(2c-1)\Sigma_0 = cI - J_0$. Therefore, $\frac{1}{2c-1}\Sigma_0^{-1} = \frac{1}{c}[I + J_0]$. If we denote $\underline{b}' = \sqrt{N}(\underline{U}' - \frac{1}{c}J') = (b_1,\ldots,b_c)$ and $\underline{b}_0' = (b_1,b_2,\ldots,b_{c-1})$, then from the asymptotic normality of $\underline{b}_0'$, we have $\underline{b}_0'\Sigma_0^{-1}\underline{b}_0$ distributed, in the limit, as a χ^2 with $c-1$ d.f. Hence, here,

$$\chi^2 = \underline{b}_0'\Sigma_0^{-1}\underline{b}_0 = \underline{b}_0' \frac{2c-1}{c}[I + J_0]\underline{b}_0$$

$$= \frac{N(2c-1)}{c}\left\{\sum_{i=1}^{c-1}(u^{(i)} - \tfrac{1}{c})^2 + \left[\sum_{1}^{c-1}(u^{(i)} - \tfrac{1}{c})\right]^2\right\}$$

$$(4.4.10) \qquad = \frac{N(2c-1)}{c}\sum_{i=1}^{c}(u^{(i)} - \tfrac{1}{c})^2 .$$

4.4.2 General case

Let us now suppose that not all n's are equal. Then $\underline{g}_0$ and $\underline{J}_0$ are linearly independent. From (4.4.9) we have

$$c^2(2c-1)\Sigma_0 = a\,J_0 + c^2D_0 - c\underline{g}_0\underline{J}_0' - c\underline{J}_0\underline{g}_0' ,$$

when $a = \sum_{k=1}^{c}\frac{1}{p_k}$. Therefore,

$$c^2(2c-1)\Sigma_0 = c^2D_0 - \left(c\underline{g}_0 \vdots \underline{J}_0\right)\begin{pmatrix}\underline{J}_0' \\ \cdots\cdots \\ c\underline{g}_0' - a\underline{J}_0'\end{pmatrix}$$

$$= c^2D_0 - EF \quad \text{say} .$$

Then $\frac{1}{c^2(2c-1)} \Sigma_0^{-1} = \frac{1}{c^2} D_0^{-1} - D_0^{-1} E \Lambda F D_0^{-1}$, where Λ is given by

$$[-c^2 I_2 + F D_0^{-1} E]\Lambda = \frac{1}{c^2} I_2 \quad . \tag{4.4.11}$$

Simplifying and using the relation $\sum_{i=1}^{c} u^{(i)} = 1$, we finally have, by similar considerations, the limiting χ^2 distribution of

$$N(2c-1)\left[\sum_{i=1}^{c} p_i (u^{(i)} - \tfrac{1}{c})^2 - \left\{\sum_{i=1}^{c} p_i (u^{(i)} - \tfrac{1}{c})\right\}^2\right]$$

with $c-1$ d.f. When $p_1 = p_2 = \dots = p_c = \frac{1}{c}$, this reduces to the earlier expression.

Remarks. It will be interesting to investigate the asymptotic power of the test against some specific alternatives. The general alternative in mind behind the test is $F_i(x) = F(x - \theta_i)$ where θ's are not all equal $(F \to F_1, G \to F_2, \dots, H \to F_c)$. In this respect, it is similar to Kruskal's test or Mood's test for c samples. In a way, it is similar to Mosteller's test, but his test is against alternatives where one population is shifted to the right and correspondingly his test-statistic is also with reference to one particular sample. On the other hand, our statistic is symmetric with respect to all samples and hence covers much more general alternatives.

Kruskal [16] says, "Unfortunately, for the H test as for many nonparametric tests the power is difficult to investigate and little is yet known about it." Recently, Andrews [1] investigated the power of Kruskal's test and Mood's test for c samples and concluded that the asymptotic efficiency of Kruskal's test relative to Mood's test for c samples is $\lesseqgtr 1$ depending on the alternatives. It will be interesting to carry out similar studies on this test with respect to these two tests. It is expected that the same type of conclusion will be reached in view of the very nature of such nonparametric problems.

CHAPTER V

SOME REGRESSION AND BIVARIATE PROBLEMS

5.1 Introduction

Mood and Brown [20] have considered some simple regression problems. On the basis of a sample of n observations (x_i, y_i), where x is in the nature of a concomitant variable and y, given x, is a continuous variate whose median is of the form $\alpha + \beta x$ where α and β are unknown parameters, they consider the problem of estimating α and β and testing hypotheses about them. They also discuss briefly the general linear regression under this nonparametric set up.

In this chapter we shall extend their methods to discuss some additional regression problems. Next we shall consider some bivariate analysis of variance problems. We shall use the "step-down procedure" [28,30] to reduce the problem to univariate cases with the other variate as a concomitant variate. The regression methods developed earlier will be used here in these bivariate problems. The method seems to be perfectly general and could be extended to three or more variates--that is, to the general multivariate situation. Most of the tests are offered on heuristic considerations. They are expected to be "good" for large samples.

5.2 Some regression problems

We shall first state a lemma [20] which will be useful for later applications.

Lemma 5.2.1 Let

$$(5.2.1) \qquad g(m_1, m_2, \ldots, m_k) = \prod_{i=1}^{k} \binom{n_i}{m_i} \Big/ \binom{n}{m} ,$$

where $n = \sum_{i=1}^{k} n_i$ and $m = \sum_{i=1}^{k} m_i$, denote the density function for the m_i's. Then

$$(5.2.2) \qquad \chi^2 = \frac{n(n-1)}{m(n-m)} \sum_{i=1}^{k} \frac{1}{n_i} \left(m_i - \frac{n_i m}{n}\right)^2$$

has an asymptotic χ^2 distribution with $k-1$ d.f. for large n.

Mood [20] says, "The expression (5.2.1) has a distribution very accurately approximated by the chi-square distribution with $k-1$ d.f. even if n is only of the order of twenty provided all the n_i are at least five."

5.2.1 One sample

Let $(x_1, y_1), \ldots, (x_n, y_n)$ denote a sample of n observations. We shall assume that

(a) the distribution of y for any x is continuous and identical apart from a shift or translation, and

(b) the regression is linear, that is, the location parameter (usually the median), given x, is $\alpha + \beta x$, where α and β are unknown parameters.

To estimate α and β, Mood and Brown [20] suggest that the estimates $\hat{\alpha}$ and $\hat{\beta}$ should be determined by

(5.2.3) Median of $(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$ for $x_i \leq \tilde{x}$

and

(5.3.4) Median of $(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$ for $x_i > \tilde{x}$,

where $\tilde{x}$ is the median of x_i's. If it happens that several x values fall at $\tilde{x}$, then the sign $\leq$ in (5.2.3) and $>$ sign in (5.2.4) may be replaced by $<$ and $\geq$ if such a replacement would more nearly divide the points into groups of equal size. They also give an iteration procedure to determine $\hat{\alpha}$ and $\hat{\beta}$.

We shall find it convenient to speak of $x_i \leq \tilde{x}$ as group one and of $x_i > \tilde{x}$ as group two. Then (5.2.3) and (5.2.4) may be equivalently written as

(5.2.5) $$\hat{\alpha} = \text{Median } (y_i - \hat{\beta}x_i)$$

and

(5.2.6) $$\underset{\text{I}}{\text{Median}} (y_i - \hat{\beta}x_i) = \underset{\text{II}}{\text{Median}} (y_i - \hat{\beta}x_i) ,$$

when I and II stand for groups one and two, respectively.

<u>Test for</u> $\beta = \beta_0$. If $\beta = \beta_0$, α is estimated by $\hat{\alpha} = \text{Median } (y_i - \beta_0 x_i)$. Mood considers the number of points, say m_1 and m_2, above the line $y = \hat{\alpha} + \beta_0 x$ in each group. Let us, for convenience, assume that n is even. Then the probability density of m_1 and m_2 is given by

(5.2.7) $$p(m_1,m_2) = \frac{\binom{n/2}{m_1}\binom{n/2}{m_2}}{\binom{n}{n/2}} ,$$

so that, by lemma 5.2.1, Mood obtains

$$(5.2.8) \qquad \chi^2 \approx \frac{16}{n}\left(m_1 - \frac{n}{4}\right)^2 \quad , \quad \text{d.f.} = 1 \quad ,$$

as the test-statistic. It may be seen that the supposition that n be even may be relaxed.

We may arrive at (5.2.8) on some heuristic considerations. Assuming n is even, as before, we have $\frac{n}{2}$ points in each group and we note that $m_1 + m_2 = \frac{n}{2}$. If the hypothesis is true, we expect m_1 and m_2 to be approximately $= \frac{n}{4}$. Now m_1 is equal to the number of positive $y_i - \hat{\alpha} - \beta_0 x_i$'s from the first group and, similarly, for m_2. Now, $y_i - \alpha - \beta_0 x_i$'s have identical distribution and, also, $\hat{\alpha} - \alpha \xrightarrow{(p)} 0$ as $n \rightarrow \infty$, so that, on heuristic considerations

$$(5.2.9) \qquad p(m_1, m_2) \approx \frac{\binom{n/2}{m_1}\binom{n/2}{m_2}}{\binom{n}{n/2}} \qquad \text{for large } n$$

and, by lemma 5.2.1, we again have asymptotic χ^2 statistic, given by (5.2.8)

If we are willing to assume, in addition, that

(c) the mean and variance of y exist for any x, then taking the mean as a location parameter given by $\alpha + \beta x$, α and β can be immediately estimated by the usual least squares estimators. In the above case, $\hat{\alpha} = \bar{y} - \beta_0 \bar{x}$, where $\bar{y}$ is the mean of y's and similarly for $\bar{x}$. Then also $\hat{\alpha} - \alpha \xrightarrow{(p)} 0$. In this case, if b denotes the number of points above the regression line, we have by a similar heuristic argument

$$p(m_1, m_2) \approx \frac{\binom{n/2}{m_1}\binom{n/2}{m_2}}{\binom{n}{b}} \qquad \text{for large } n ,$$

where m_1 and m_2 are defined as before. Hence, by lemma 5.2.1, we have an alternate test-statistic

(5.2.10) $$\chi^2 \approx \frac{4n}{b(n-b)}\left(m_1 - \frac{b}{2}\right)^2 , \quad \text{d.f.} = 1 .$$

Consistency of $\hat{\alpha}$ and $\hat{\beta}$ determined by (5.2.5) and (5.2.6)

Let $z_i = y_i - \alpha - \beta x_i$. Then z_i's have identical distribution with median zero. Now (5.2.6) may be written as

(5.2.11) $$\underset{I}{\text{Median}}\,[z_i + (\beta - \hat{\beta})x_i] = \underset{II}{\text{Median}}\,[z_i + (\beta - \hat{\beta})x_i] .$$

Now as $n \longrightarrow \infty$, $\left|\underset{I}{\text{Median}}(z_i) - \underset{II}{\text{Median}}(z_i)\right| \xrightarrow{(p)} 0$, so that intuitively it seems that $\hat{\beta} \approx \beta$ will satisfy (5.2.11), that is, $|\hat{\beta} - \beta| \xrightarrow{(p)} 0$. It has not been possible yet to give a general formal proof. We shall give a formal proof for a special case, when x's are constants at our disposal, so that they can be chosen suitably.

Proof for the special case. Let $\tilde{x}_n$, θ_{1n}, θ_{2n} and $\hat{\beta}_n$ denote the median of x's, $\underset{I}{\text{Median}}(z_i)$, $\underset{II}{\text{Median}}(z_i)$ and $\hat{\beta}$ respectively when the sample size is $2n$. Let us suppose that for $n \geq n_0$, (i) x's are chosen alternately from group I $(x \leq \tilde{x}_{n_0})$ and from group II $(x > \tilde{x}_{n_0})$, so that for $n \geq n_0$, $\tilde{x}_n = \tilde{x}_{n_0}$, and (ii) all x's in group II are greater than or equal to $\tilde{x}_{n_0} + \delta$, where δ is a fixed positive number, however small. [For example, δ may be in the nature of a unit of measurement.]

Since $\theta_{1n} \xrightarrow{(p)} 0$ and $\theta_{2n} \xrightarrow{(p)} 0$, given $\eta, \varepsilon > 0$, there is n_1 such that

(5.2.12) $\quad |\theta_{2n}| < \varepsilon$ and $|\theta_{1n}| < \varepsilon$ for $n \geq n_1$,

with probability greater than $1 - \eta$. Consider n greater than $\max(n_0, n_1)$. Let $\beta - \hat{\beta}_n = \theta_n$.

Case (1): Suppose $\theta_n \geq 0$. Then

$$\underset{I}{\text{med}}.[z_i + (\beta - \hat{\beta}_n)x_i] \leq \theta_{1n} + \theta_n \tilde{x}_{n_0} \quad ,$$

and

$$\underset{II}{\text{med}}.[z_i + (\beta - \hat{\beta}_n)x_i] \geq \theta_{2n} + \theta_n(\tilde{x}_{n_0} + \delta) \quad .$$

Then (5.2.11) $\Longrightarrow \theta_{2n} + \theta_n(\tilde{x}_{n_0} + \delta) \leq \theta_{1n} + \theta_n \tilde{x}_{n_0}$,

so that $\quad \theta_n \delta \leq \theta_{1n} - \theta_{2n} < 2\varepsilon$ from (5.2.12).

Hence $\quad \theta_n = |\theta_n| < \frac{2\varepsilon}{\delta} = \varepsilon'$ say.

Case (2): Suppose $\theta_n \leq 0$. Then

$$\underset{I}{\text{med}}.[z_i + (\beta - \hat{\beta}_n)x_i] \geq \theta_{1n} - |\theta_n| \tilde{x}_{n_0} \quad ,$$

and

$$\underset{II}{\text{med}}.[z_i + (\beta - \hat{\beta}_n)x_i] \leq \theta_{2n} - |\theta_n|(\tilde{x}_{n_0} + \delta) \quad .$$

Again, (5.2.11) $\Longrightarrow \theta_{1n} - |\theta_n| \tilde{x}_{n_0} \leq \theta_{2n} - |\theta_n|(\tilde{x}_{n_0} + \delta)$,

so that $\quad |\theta_n| \delta \leq \theta_{2n} - \theta_{1n} < 2\varepsilon$ from (5.2.12).

Hence, again, $\quad |\theta_n| < \frac{2\varepsilon}{\delta} = \varepsilon'$.

Thus, given η and $\varepsilon' > 0$, there is $n^* = \max(n_0, n_1)$, such that $|\theta_n| < \varepsilon'$ with probability $> 1 - \eta$ for $n \geq n^*$.

It may be seen that for large n (5.2.8) $\approx$ the corresponding expression obtained by lemma 5.2.1, if (5.2.7) were replaced by $\dfrac{\begin{pmatrix}\frac{n-1}{2}\\ m_1\end{pmatrix}\begin{pmatrix}\frac{n+1}{2}\\ m_2\end{pmatrix}}{\begin{pmatrix}n\\ \frac{n-1}{2}\end{pmatrix}}$ where n, the sample size, is odd. Hence, in the previous argument, for $n \geq n_0$ and sample size $2n + 1$, we may take group I as $x \leq \tilde{x}_{n_0}$, as before, and group II as $x > \tilde{x}_{n_0}$ and the previous argument goes through. Thus

$\theta_n \xrightarrow{(p)} 0$, that is, $\hat{\beta}_n \xrightarrow{(p)} \beta$, and the proof is complete for the special case mentioned above.

Remark: It may be seen that this proof hinges on the existence of δ. We may note that if we decide $x \leq x_0$ as group I and $x > x_0$ as group II (even though x_0 is not the median of x's), then the test statistic (5.2.8) can be modified suitably and the above proof does not require condition (1).

Consistency of $\hat{\alpha}$: Let us assume that $\hat{\beta} \xrightarrow{(p)} \beta$. Now $\hat{\alpha} = \text{med.}(y_i - \hat{\beta}x_i) = \alpha + \text{med.}[z_i + (\beta - \hat{\beta})x_i]$. We shall assume that x's are bounded (at least bounded with probability $\geq 1 - \eta_1$). Suppose $|x_i| < M$ for all i. Also $\hat{\beta} \xrightarrow{(p)} \beta \Longrightarrow$ given $\varepsilon, \eta > 0$, there is n^* such that $|\beta - \hat{\beta}| < \frac{\varepsilon}{M}$ for all $n \geq n^*$, with probability $> 1 - \eta$. Then

$$\text{med}(z_i) - \varepsilon \leq \text{med}[z_i + (\beta - \hat{\beta})x_i] \leq \text{med}(z_i) + \varepsilon \quad ,$$

for $n \geq n^*$ with probability $> 1 - \eta$. Also $\text{med}(z_i) \xrightarrow{(p)} 0$, so that $\text{med}[z_i + (\beta - \hat{\beta})x_i] \xrightarrow{(p)} 0$. Thus, $\hat{\alpha} \xrightarrow{(p)} \alpha$.

Henceforth we shall assume that the condition (2) is obeyed, so that $\hat{\alpha}$ and $\hat{\beta}$ are consistent.

5.2.2 c samples

Let us suppose that we have n_i independent observations (x_{ij}, y_{ij}) $j = 1,2,\ldots,n_i$ from the i-th population, $i = 1,2,\ldots,c$. We shall assume (a) as before and (b) that the regression is linear, that is, the location parameter (usually the median) of y_{ij}, given x_{ij}, is $\alpha_i + \beta_i x_{ij}$.

(i) To test $\beta_i = \beta_{io}$, $i = 1,2,\ldots,c$.

We shall have c independent χ^2 statistics with 1 d.f. each, giving a χ^2 statistic with c d.f. No new problem is presented here.

(ii) To test $\beta_1 = \beta_2 = \ldots = \beta_c$.

On this hypothesis, y_{ij}'s have medians $\alpha_i + \beta x_{ij}$. We may estimate α_i's and β by

$$\hat{\alpha}_i = \underset{j=1,2,\ldots,n_i}{\text{median}} (y_{ij} - \hat{\beta}x_{ij})$$

and

$$\underset{\text{I}}{\text{median}}(y_{ij} - \hat{\alpha}_i - \hat{\beta}x_{ij}) = \underset{\text{II}}{\text{median}}(y_{ij} - \hat{\alpha}_i - \hat{\beta}x_{ij}) \quad .$$

For convenience, we shall take group I as $x \leq \tilde{x}$ (median of all x's) and group II as $x > \tilde{x}$, though the test-statistics can be modified to suit other cases.

Let $\sum_1^c n_i = N$. For simplicity, let us take n_i to be even. Let m_i be the number of points from the i-th sample belonging to the second group and ℓ_i be the number of points out of these m_i that lie above $y = \hat{\alpha}_i + \hat{\beta}x$. Then $\sum_1^c m_i = \frac{N}{2}$ and $\sum_1^c \ell_i \approx \frac{N}{4}$. If the hypothesis is true, we expect ℓ_i to be $\approx \frac{m_i}{2}$. Let ℓ_i' be the number of observations from the m_i in the second group of the i-th sample, such that $z_{ij} = y_{ij} - \alpha_i - \beta x_{ij}$ is > 0 . Since $\hat{\alpha}_i - \alpha_i \xrightarrow{(p)} 0$ and $\hat{\beta} - \beta \xrightarrow{(p)} 0$, $\ell_i - \ell_i' \xrightarrow{(p)} 0$ as n_i's $\longrightarrow \infty$. Therefore, heuristically, ℓ_i's have the same distribution for large n_i's as ℓ_i''s subject to $\sum_1^c \ell_i' \approx \frac{N}{4}$. Since z_{ij}'s have identical distribution,

$$p(\ell_1', \ldots, \ell_c') = \frac{\prod_1^c \binom{m_i}{\ell_i'}}{\binom{N/2}{N/4}} , \quad \text{so that}$$

$$p(\ell_1, \ell_2, \ldots, \ell_c) \approx \frac{\prod_1^c \binom{m_i}{\ell_i}}{\binom{N/2}{N/4}} \quad \text{for large } n_i\text{'s} .$$

Hence, by lemma 5.2.1, we have

$$\chi^2 \approx 4 \sum \frac{1}{m_i} \left(\ell_i - \frac{m_i}{2}\right)^2 , \quad \text{d.f.} = c - 1 . \tag{5.2.13}$$

If some $m_i = 0$, the corresponding term will be absent and d.f. will be reduced by one. We could have considered group I instead of group II. It may be seen now that the condition n_i be even may be relaxed.

If we are willing to assume, in addition, (c) as before, then we may take least squares estimates

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}\bar{x}_i \quad \text{where} \quad \hat{\beta} = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_i)x_{ij}}{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2} ,$$

so that $\hat{\alpha}_i \xrightarrow{(p)} \alpha_i$ and $\hat{\beta} \xrightarrow{(p)} \beta$. If ℓ_i denotes the number of points from the i-th sample above the corresponding regression line and $\sum_1^c \ell_i = \ell$, then by a similar heuristic argument,

$$p(\ell_1,\ldots,\ell_c) \approx \frac{\prod_1^c \binom{n_i}{\ell_i}}{\binom{N}{\ell}} \qquad \text{for large } n_i\text{'s} ,$$

so that by lemma 5.2.1 ,

$$\chi^2 \approx \frac{N^2}{\ell(N-\ell)} \sum_1^c \frac{1}{n_i} [\ell_i - \frac{n_i}{N}\ell]^2 \text{ , d.f.} = c-1 . \tag{5.2.14}$$

(iii) <u>To test</u> $\alpha_1 = \alpha_2 = \cdots = \alpha_c$, <u>when</u> $\beta_1 = \beta_2 = \cdots = \beta_c$.

On this hypothesis, y_{ij}'s have medians $\alpha + \beta x_{ij}$. α and β may be estimated by

$$\hat{\alpha} = \text{med}(y_{ij} - \hat{\beta}x_{ij})$$

and

$$\underset{\text{I}}{\text{med}}[y_{ij} - \hat{\beta}x_{ij}] = \underset{\text{II}}{\text{med}}[y_{ij} - \hat{\beta}x_{ij}] \quad ,$$

where, for convenience, we take groups I and II as $x \leq \tilde{x}$ (median of all x's) and $x > \tilde{x}$ respectively. Let N be even and ℓ_i be the number of points in the i-th sample above the regression line $y = \hat{\alpha} + \hat{\beta}x$. If the hypothesis is true, we expect $\ell_i \approx \frac{n_i}{2}$. We note that $\sum_1^c \ell_i = \frac{N}{2}$. Let ℓ_i' denote the number of positive terms in $z_{ij} = y_{ij} - \alpha - \beta x_{ij}$ $(j = 1,2,\dots,n_i)$. Since $\hat{\alpha} \xrightarrow{(p)} \alpha$ and $\hat{\beta} \xrightarrow{(p)} \beta$, $\ell_i - \ell_i' \xrightarrow{(p)} 0$ as $N \longrightarrow \infty$. Hence by similar heuristic arguments, the distribution of ℓ_i's for large N is approximately the same as that of ℓ_i''s subject to $\sum_1^c \ell_i' = \frac{N}{2}$. Hence,

$$(5.2.15) \qquad p(\ell_1,\ell_2,\dots,\ell_c) \approx \frac{\prod_1^c \binom{n_i}{\ell_i}}{\binom{N}{N/2}} \quad \text{for large } N ,$$

so that by lemma 5.2.1

$$(5.2.16) \qquad \chi^2 = 4 \sum_1^c \frac{1}{n_i} \left(\ell_i - \frac{n_i}{2}\right)^2 , \quad \text{d.f.} = c - 1 .$$

If we are willing to assume, in addition, (c), that is, the existence of mean and variance, then we can have least-square estimates $\hat{\alpha}$ and $\hat{\beta}$, such that $\hat{\alpha} \xrightarrow{(p)} \alpha$ and $\hat{\beta} \xrightarrow{(p)} \beta$. If we denote $\sum_1^c \ell_i$ by d , then by the same heuristic argument

$$p(\ell_1,\dots,\ell_c) \approx \prod_1^c \binom{n_i}{{}_i} \Big/ \binom{N}{d} \quad \text{for large } N ,$$

so that

$$\chi^2 \approx \frac{N^2}{d(N-d)} \sum_1^c \frac{1}{n_i} \left(\ell_i - \frac{n_i}{N} d\right)^2 , \quad \text{d.f.} = c - 1 ,$$

We shall indicate here briefly a formal proof for (5.2.15), which was first derived on heuristic considerations.

Let $u_{ij} = y_{ij} - \hat{\beta}x_{ij}$. Then

ℓ_i = number of positive $y_{ij} - \hat{\alpha} - \hat{\beta}x_{ij}$ $(j = 1,2,\ldots,n_i)$
= number of u_{ij}'s $> \hat{\alpha} = \underset{i,j}{\text{median}}\,(u_{ij})$.
$(j = 1,2,\ldots,n_i)$

Also $\sum_1^c \ell_i = \frac{N}{2}$. Let z_a be the a-th $(a = \frac{N}{2})$ u_{ij} in magnitude. Then the joint density function of $\ell_1,\ldots,\ell_c$ and z_a, under the hypothesis, is

$$\text{(5.2.17)}\quad \sum_{i=1}^{c}\sum F_{11_1}(z_a)\ldots F_{11_{n_1-\ell_1}}(z_a)[1 - F_{11_{n_1-\ell_1+1}}(z_a)]\ldots$$
$$[1 - F_{11_{n_1}}(z_a)]\ldots F_{ii_1}(z_a)\ldots F_{ii_{n_i-\ell_i-1}}(z_a)\times$$
$$[1 - F_{ii_{n_i-\ell_i+1}}(z_a)]\ldots[1 - F_{ii_{n_i}}(z_a)]$$
$$dF_{ii_{n_i-\ell_i}}(z_a)\ldots F_{cc_1}(z_a)\ldots F_{cc_{n_c-\ell_c}}(z_a)\times$$
$$[1 - F_{cc_{n_c-\ell_c+1}}(z_a)]\ldots[1 - F_{cc_{n_c}}(z_a)]$$

where $F_{ij}(z_a) = \Pr[u_{ij} \le z_a]$, the i-th term indicates that z_a is from the i-th sample and $\sum$ denotes the sum over all possible combinations.

Since $\hat{\beta} \xrightarrow{(p)} \beta$, given $\varepsilon,\eta > 0$, there is N_0 such that, for $N \ge N_0$, $|\hat{\beta} - \beta| < \varepsilon$ with probability

$> 1 - \eta$. Then for $N \geq N_0$, with probability $> 1 - \eta$, we have

$Pr[Y_{ij} - \beta x_{ij} \leq z_a - \varepsilon x_{ij}] \leq F_{ij}(z_a) \leq Pr[Y_{ij} - \beta x_{ij} \leq z_a + \varepsilon x_{ij}]$,

that is

$$F(z_a - \varepsilon x_{ij}) \leq F_{ij}(z_a) \leq F(z_a + \varepsilon x_{ij}) \quad ,$$

where F denotes the distribution function of all $Y_{ij} - \beta x_{ij}$. In view of the continuity of F ,

$$F_{ij}(z_a) = F(z_a) + \delta_{ij} \quad ,$$

where δ's are arbitrarily small and tend to zero as $N \to \infty$. Then (5.2.17) becomes

$$= \sum_{i=1}^{c} \sum F^{n_1-\ell_1}(z_a)[1 - F(z_a)]^{\ell_1} \cdots F^{n_i-\ell_i-1}(z_a)[1 - F(z_a)]^{\ell_i} dF(z_a)$$
$$\cdots F^{n_c-\ell_c}(z_a)[1 - F(z_a)]^{\ell_c} + O(\delta)$$

$$= \sum_{i=1}^{c} \sum F^{\frac{N}{2}-1}(z_a)[1 - F(z_a)]^{\frac{N}{2}} dF(z_a) + O(\delta)$$

$$= \sum_{i=1}^{c} \binom{n_1}{\ell_1} \cdots \binom{n_{i-1}}{\ell_{i-1}} \frac{n_i!}{\ell_i!(n_i - \ell_i - 1)!} \binom{n_{i+1}}{\ell_{i+1}} \cdots \binom{n_c}{\ell_c} \times$$
$$F^{\frac{N}{2}-1}(z_a)[1 - F(z_a)]^{N/2} dF(z_a) + O(\delta) \quad .$$

On integrating out z_a we have the joint density of $\ell_1, \ell_2, \ldots, \ell_c$

$$= \sum_{i=1}^{c} \binom{n_1}{\ell_1} \cdots \frac{n_i!}{\ell_i!(n_i - \ell_i - 1)!} \cdots \binom{n_c}{\ell_c} \int_0^1 x^{\frac{N}{2}-1}(1 - x)^{\frac{N}{2}} dx + O(\delta)$$

$$= \prod_{1}^{c} \binom{n_i}{\ell_i} B\left(\frac{N}{2}, \frac{N}{2} + 1\right) \sum_{i=1}^{c} (n_i - \ell_i) + O(\delta)$$

$$= \frac{\prod_{1}^{c} \binom{n_i}{\ell_i}}{\binom{N}{N/2}} + O(\delta) ,$$

which is the same as (5.2.15).

(iv) To test $\beta = 0$, when $\beta_1 = \beta_2 = \ldots = \beta_c = \beta$ say.

On this hypothesis, y_{ij}'s have medians α_i's. We may take

$$\hat{\alpha}_i = \underset{j=1,2,\ldots,n_i}{\text{median}} (y_{ij}) \quad .$$

For simplicity let n_i be even. Then $\frac{n_i}{2}$ points from the i-th sample are above the corresponding line. Also $\frac{N}{2}$ points are to the right of $\tilde{x}$, the median of all the x's. Let ℓ_i be the number of points from the i-th sample to the right of $\tilde{x}$ and above the corresponding line and let $\ell = \sum_{1}^{c} \ell_i$. We expect, then, ℓ to be $\approx \frac{N}{4}$. Let m_i's and m be defined similarly for $x \leq \tilde{x}$. Then, by the same heuristic argument, for which a formal proof could be given as in (iii), we have

$$p(\ell, m) \approx \frac{\binom{N/2}{\ell}\binom{N/2}{m}}{\binom{N}{N/2}} \quad \text{for large } N ,$$

and, hence, by lemma 5.2.1, we have

$$(5.2.18) \qquad \chi^2 \approx \frac{16}{N}\left(\ell - \frac{N}{4}\right)^2 , \quad \text{d.f.} = 1 \quad .$$

The condition that n_i be even, then, may be relaxed.

5.2.3 Testing linearity of regression

As in the normal analysis, it is necessary that we have a number of observations for each x_i. Let the observations be (x_i, y_{ij}), $j = 1,2,\ldots,n_i$, $i = 1,2,\ldots,k$. We shall assume that the distribution of y, given x, is continuous and the same apart from location, say $h(x)$, which may depend on x. We want to test the hypothesis that the "regression" is linear, that is,

$$h(x) = \alpha + \beta x \quad .$$

Let $\sum_1^k n_i = N$ and these N observations be divided into two groups, say $x \leq x_{k_1}$ forming the first group and $x > x_{k_1}$ forming the second group, as evenly as possible. Let us suppose that observations corresponding to x_i $(i = 1,2,\ldots,k_1)$ belong to the first group and the rest to the second. Let the groups contain a and $N - a$ observations respectively. We may then estimate α and β by

$$\operatorname{med}(y_{ij} - \hat{\alpha} - \hat{\beta}x_i) = 0 \quad ,$$

and

$$\operatorname*{med}_{I}(y_{ij} - \hat{\beta}x_i) = \operatorname*{med}_{II}(y_{ij} - \hat{\beta}x_i) \quad .$$

Consider the n_i observations corresponding to x_i. If the regression is linear, we expect these n_i to be split evenly by the regression line $y = \hat{\alpha} + \hat{\beta}x$. Let ℓ_i, out of these n_i, be above the line. We expect $\ell_i \approx \frac{n_i}{2}$.

Then $\sum_{i=1}^{k_1} \ell_i = \frac{a}{2}$ and $\sum_{i=k_1+1}^{k} \ell_i = \frac{N-a}{2}$, assuming for convenience that a and $N-a$ are even.

Let $z_{ij} = y_{ij} - \alpha - \beta x_i$. Then on the null hypothesis, z_{ij}'s have identical distribution. Let ℓ_i' be the number of positive terms in z_{ij} $(j = 1, 2, \ldots, n_i)$.

Since $\hat{\alpha} \xrightarrow{(p)} \alpha$ and $\hat{\beta} \xrightarrow{(p)} \beta$, $\ell_i - \ell_i' \xrightarrow{(p)} 0$. Hence on heuristic considerations as before, the distribution of ℓ_i's is the same (asymptotically) as that of ℓ_i''s subject to $\sum_{i=1}^{k_1} \ell_i' = \frac{a}{2}$ and $\sum_{i=k_1+1}^{k} \ell_i' = \frac{N-a}{2}$. Thus,

$$p(\ell_1, \ell_2, \ldots, \ell_k) \approx \frac{\prod_{1}^{k_1} \binom{n_i}{\ell_i}}{\binom{a}{\frac{a}{2}}} \quad \frac{\prod_{k_1+1}^{k} \binom{n_i}{\ell_i}}{\binom{N-a}{\frac{N-a}{2}}} ,$$

so that by lemma 5.2.1

$$\chi_I^2 \approx 4 \sum_{1}^{k_1} \frac{1}{n_i} \left(\ell_i - \frac{n_i}{2}\right)^2 , \quad \text{d.f.} = k - 1 ,$$

and

$$\chi_{II}^2 \approx 4 \sum_{k_1+1}^{k} \frac{1}{n_i} \left(\ell_i - \frac{n_i}{2}\right)^2 , \quad \text{d.f.} = k - k_1 - 1 ,$$

so that

$$\chi^2 \approx 4 \sum_{1}^{k} \frac{1}{n_i} \left(\ell_i - \frac{n_i}{2}\right)^2 , \quad \text{d.f.} = k - 2 .$$

5.3 Some bivariate problems

5.3.1 One-way classification

Let there be n_i independent observations (x_{ij}, y_{ij}) $j = 1,2,\ldots,n_i$, from the i-th population, $i = 1,2,\ldots,k$, and let $\sum_1^k n_i = N$.

Suppose $F_i(x,y)$ denotes the distribution function of (X,Y) for the i-th population. We shall assume that

(i) F's are continuous,

(ii) the distributions are identical except for location, and

(iii) the median of the conditional distribution of Y, given X, is a linear function of X. We note that

(i) $\Longrightarrow$ the conditional probability, given X, is also a probability measure. Let $f_i(x,y)$, $f_i(x)$ and $f_i(y/x)$ denote the densities of (X,Y), X and Y/X respectively.

Also, (ii) $\Longrightarrow$

(5.3.1) $$F_i(x,y) = F(x-\zeta_i,\ y-\eta_i) \quad .$$

We want to test whether the populations are identical. Thus

(5.3.2) $$H_0:\ \zeta_1 = \zeta_2 = \ldots = \zeta_k \quad \text{and} \quad \eta_1 = \eta_2 = \ldots = \eta_k \quad .$$

(5.3.1) $\Longrightarrow$ $f_i(x,y) = f(x-\zeta_i,\ y-\eta_i)$,

so that $f_i(x) = f_1(x-\zeta_i)$ say .

(iii) $\Longrightarrow$ $f(x,y) = f_1(x)f_2(y-\alpha-\beta x)$, so that $f(x-\zeta_i, y-\eta_i) = f_1(x-\zeta_i)f_2[y-\eta_i-\alpha-\beta(x-\zeta_i)]$

$$= f_1(x-\zeta_i)f_2(y-\alpha_i-\beta x) \quad \text{say} .$$

Thus, we see that

$$H_0 \Longleftrightarrow \zeta_1 = \zeta_2 = \dots = \zeta_k$$
$$\text{and} \quad \alpha_1 = \alpha_2 = \dots = \alpha_k \quad .$$

It may be noted that we have relaxed just the normality of the distribution, but retained other features from the classical set up.

We shall use a step-down procedure to test H_0. A step-down procedure for H_0 with a level γ will be a test for

(5.3.3) $\qquad H_{ox}: \ \zeta_1 = \zeta_2 = \dots = \zeta_k \quad ,$

with a level γ_1, and if it is not rejected, a further test for

(5.3.4) $\qquad H_{oy/x}: \ \alpha_1 = \alpha_2 = \dots = \alpha_k$

with a level γ_2, where γ_1 and γ_2 are chosen suitably so that

$$(1-\gamma) = (1-\gamma_1)(1-\gamma_2) \quad .$$

The test for $H_{oy/x}$ will be derived from the conditional distribution of Y, given x, so that the x's then can be regarded as fixed.

For H_{ox}, we consider only x's. Let us consider the test given by Mood [20]. [We could have used either Kruskal's test or the test derived in Chapter IV.] Let m_i

denote the number of observations in the i-th sample greater than the median of all x's. Mood shows that the density function, if H_{ox} is true, is

$$(5.3.5) \qquad p(m_1, m_2, \ldots, m_k) = \frac{\prod_1^k \binom{n_i}{m_i}}{\binom{N}{a}} ,$$

where $a = \frac{N}{2}$ if N is even or $\frac{N-1}{2}$ if N is odd. The test-statistic proposed by him for large N is

$$(5.3.6) \qquad \chi^2 = \frac{N(N-1)}{a(N-a)} \sum_1^k \frac{1}{n_i} \left(m_i - \frac{n_i a}{N}\right)^2 , \quad \text{d.f.} = k-1 .$$

For small n's, the probability is computed from the exact distribution (5.3.5).

The test for $H_{oy/x}$ is seen to be precisely the same as that considered in 5.2. Hence we may take (5.2.16) (in its modified form) as a test-statistic, if the condition (ii) mentioned on page 102 holds good. As already stated, it may be possible to prove that $\hat{\alpha} \xrightarrow{(p)} \alpha$ without using condition (ii), in which case (5.2.16) may be used for large samples in general.

5.3.2 Two-way classification

For simplicity, we shall consider only the case of one observation per cell, when the design is complete. Let "i" denote "treatments" and "j" denote "blocks." Suppose

$$i = 1, 2, \ldots, t$$
$$j = 1, 2, \ldots, b \quad \text{and} \quad N = bt .$$

Let $F_{ij}(x,y)$ denote the distribution function of (X,Y) for the (ij)-th cell.

We shall assume that

(i) $F_{ij}(x,y)$ is continuous,

(ii) the distributions are identical except for location, that is,

$$F_{ij}(x,y) = F(x - \alpha_{ij}, y - \beta_{ij}) \quad ,$$

(iii) the model is additive, that is

$$\alpha_{ij} = \zeta_i + \eta_j \quad \text{and} \quad \beta_{ij} = \gamma_i + \delta_j \text{ , and}$$

(iv) the "regression" of Y on X is linear.

As before, we notice that we have relaxed just the normality of the distribution while retaining other features of the classical set up.

Let $f_{ij}(x,y)$, $f_{ij}(x)$ and $f_{ij}(y/x)$ denote the densities of (X,Y), (X) and (Y/X) respectively.

$$\text{(ii)} \Longrightarrow \quad f_{ij}(x,y) = f(x - \alpha_{ij}, y - \beta_{ij}) \ ,$$

$$\text{and} \qquad f_{ij}(x) = f_1(x - \alpha_{ij}) \quad \text{say} \ .$$

$$\text{(iv)} \Longrightarrow \quad f(x,y) = f_1(x) f_2(y - \alpha - \beta x) \ ,$$

so that

$$f_{ij}(x,y) = f_1(x - \alpha_{ij}) f_2[y - \beta_{ij} - \alpha - \beta(x - \alpha_{ij})]$$

$$= f_1(x - \zeta_i - \eta_j) f_2[y - \alpha - \gamma_i + \beta\zeta_i - \delta_j + \beta\eta_j - \beta x] \ . \tag{5.3.7}$$

We will be interested in the usual hypothesis

$$H_0: \ \zeta_1 = \zeta_2 = \dots = \zeta_t$$

$$\text{and } \gamma_1 = \gamma_2 = \dots = \gamma_t \quad .$$

We shall consider a step-down procedure to test H_0. Considering x's separately, we can test, at a level α_1,

$$H_{ox}: \ \zeta_1 = \zeta_2 = \dots = \zeta_t$$

by the criterion, given by Mood [20],

$$\chi^2 = \frac{t(t-1)}{ba(t-a)} \sum_{i=1}^{t} \left(m_i - \frac{ba}{t}\right)^2 , \quad \text{d.f.} = t-1 , \tag{5.3.8}$$

where $a = \frac{t}{2}$ if t is even or $\frac{t-1}{2}$ otherwise, and m_i = the number of x_{ij}'s $(j = 1,2,\dots,b)$ greater than $\tilde{x}_j$, the median of the j-th column. Then, considering the conditional distribution of y_{ij}'s, given x_{ij}'s, we have to test

(5.3.9) $H_{oy/x}$: y_{ij}'s have medians $\lambda_j + \beta x_{ij}$,

at a level α_2, so that

$$(1-\alpha) = (1-\alpha_1)(1-\alpha_2) \quad .$$

We may estimate λ_j and β by

$$\hat{\lambda}_j = \underset{i=1,2,\dots,t}{\text{median}} (y_{ij} - \hat{\beta}x_{ij}) \quad , \quad \text{and}$$

$$\underset{I}{\text{med}}(y_{ij} - \hat{\lambda}_j - \hat{\beta}x_{ij}) = \underset{II}{\text{med}}(y_{ij} - \hat{\lambda}_j - \hat{\beta}x_{ij}) \quad ,$$

where the groups are with respect to x's as usual. We note that a, defined as above, out of t $y_{ij} - \hat{\lambda}_j - \hat{\beta}x_{ij}$'s for each j, are positive and hence in all ab out of bt $y_{ij} - \hat{\lambda}_j - \hat{\beta}x_{ij}$'s are positive. Let ℓ_i denote the number of positive terms out of b $y_{ij} - \hat{\lambda}_j - \hat{\beta}x_{ij}$, for given i.

Then we expect $\ell_i \approx \frac{b}{2}$ if (5.3.9) is true. Also $\sum_{i=1}^{t} \ell_i = ab$.

Let ℓ_i' denote the number of positive terms out of b $y_{ij} - \lambda_j - \beta x_{ij}$, for given i. On heuristic considerations, for large samples $\lambda_j \approx \hat{\lambda}_j$ and $\beta \approx \hat{\beta}$, so that the distribution of ℓ's is asymptotically the same as that of ℓ's subject to $\sum_1^t \ell_i' = ab$. Hence,

$$p(\ell_1, \ell_2, \ldots, \ell_t) \approx \frac{\prod_1^t \binom{b}{\ell_i}}{\binom{N}{ab}} \quad \text{for large } N, \ (N = bt),$$

so that by lemma 5.2.1,

$$\chi^2 \approx \frac{N^2}{ab(N-ab)} \sum_1^t \frac{1}{b}\left(\ell_i - \frac{b}{N}\,ab\right)^2$$

$$\text{(5.3.10)} \qquad \approx \frac{t^2}{ba(t-a)} \sum_1^t \left(\ell_i - \frac{ba}{t}\right)^2 , \quad \text{d.f.} = t-1 .$$

The same remark as that at the end of 5.3.1 will hold good here. Also, it may seem that we require t large (since we require $\lambda_j \approx \hat{\lambda}_j$ in the above argument), but if we give a formal proof, similar to that given in 5.2.2 (iii), we shall note that $\beta \approx \hat{\beta}$ is sufficient to reduce the proof to the one given by Mood. This does not require large t but only large bt. Hence (5.3.10) gives a test-criterion for large b.

BIBLIOGRAPHY

[1] Andrews, F. C., "Asymptotic Behavior of Some Rank Tests for Analysis of Variance," Annals of Mathematical Statistics, XXV (1954), 724-736.

[2] Barnard, G. A., "Significance Tests for 2 × 2 Tables," Biometrika, XXXIV (1947), 123-138.

[3] Bartlett, M. S., "Contingency Table Interactions," Supplement to the Journal of the Royal Statistical Society, II (1935), 248-252.

[4] Benard, A., and Ph. van Elteren, "A Generalization of the Method of m Rankings," Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, LVI (1953), 358-369.

[5] Bhapkar, V. P., "A Note on Confidence Bounds connected with ANOVA and MANOVA for Balanced and Partially Balanced Incomplete Block Designs," North Carolina Institute of Statistics Mimeograph Series, No. 216, January 1959.

[6] Bose, R. C., and D. M. Mesner, "On Linear Associative Algebras corresponding to Association Schemes of Partially Balanced Designs," Annals of Mathematical Statistics, XXX (1959), 21-38.

[7] Cramer, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

[8] Diamond, E. L., "Asymptotic Power and Independence of Certain Classes of Tests on Categorical Data," North Carolina Institute of Statistics Mimeograph Series, No. 196, April 1958.

[9] Durbin, J., "Incomplete Blocks in Ranking Experiments," British Journal of Psychology, IV (1951), 85-90.

[10] Fisher, R. A., "On the Interpretation of Chi-square from Contingency Tables and the Calculation of p," Journal of the Royal Statistical Society, LXXXV (1922), 87-94.

[11] --------The Design of Experiments, Edinburgh, Oliver and Boyd Ltd., 1935.

[12] Fisher, R. A., and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, Edinburgh, Oliver and Boyd Ltd., 1938.

[13] Friedman, M., "The Use of Ranks to avoid the Assumption of Normality implicit in the Analysis of Variance, Journal of the American Statistical Association, XXXII (1937), 675-701.

[14] Hoeffding, W., "A Class of Statistics with Asymptotically Normal Distributions," Annals of Mathematical Statistics, XIX (1948), 293-325.

[15] Kruskal, W. H., "A Nonparametric Test for the Several Sample Problem," Annals of Mathematical Statistics, XXIII (1952), 525-540.

[16] --------and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," Journal of the American Statistical Association, XLVII (1952), 583-621.

[17] Lehmann, E. L., "Consistency and Unbiasedness of Certain Non-parametric Tests," Annals of Mathematical Statistics, XXII (1951), 165-179.

[18] Massey, F. J., "A Note on a Two-sample Test," Annals of Mathematical Statistics, XXII (1951), 304-306.

[19] Mitra, S. K., "Contributions to the Statistical Analysis of Categorical Data," North Carolina Institute of Statistics Mimeograph Series, No. 142, December 1955.

[20] Mood, A. M., Introduction to the Theory of Statistics, New York, McGraw-Hill Book Co., 1950.

[21] Mosteller, F., "A k-sample Slippage Test for an Extreme Population," Annals of Mathematical Statistics, XIX (1948), 58-65.

[22] Neyman, J., "Contributions to the Theory of χ^2 Test," Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, (1949), 239-273.

[23] Ogawa, J., "On the Mathematical Principles underlying the Theory of the χ^2 Test," North Carolina Institute of Statistics Mimeograph Series, No. 162, January 1957.

[24] Pearson, E. S., "The Choice of Statistical Tests illustrated on the Interpretation of Data Classed in a 2 × 2 Table," Biometrika, XXXIV (1947), 139-167.

[25] Pearson, K., "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling," Philosophical Magazine, Series 5, L (1900), 157-172.

[26] Pitman , E.J.G., "Significance Tests which may be applied to Samples from any Populations. III. The Analysis of Variance Test," Biometrika, XXIX (1937), 322-335.

[27] Reiersol, O., "Tests of Linear Hypotheses concerning Binomial Experiments," Skandinavisk Aktuarietidskrift, XXXVII (1954), 38-59.

[28] Roy, J., "Step-down Procedure in Multivariate Analysis," Annals of Mathematical Statistics, XXIX (1958), 1177-1187.

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

[30] --------and Bargmann, R. E., "Tests of Multiple Independence and the Associated Confidence Bounds," Annals of Mathematical Statistics , XXIX (1958), 491-503.

[31] Roy, S. N., and V. P. Bhapkar,"Some Nonparametric Analogues of 'Normal' ANOVA, MANOVA, and of Studies in 'Normal' Association," North Carolina Institute of Statistics Mimeograph Series, No. 215, January 1959.

[32] Roy, S. N., and M. A. Kastenbaum, "On the Hypothesis of No Interaction in a Multiway Contingency Table," Annals of Mathematical Statistics , XXVII (1956), 749-757.

[33] Roy, S. N., and S. K. Mitra, "An Introduction to Some Nonparametric Generalizations of Analysis of Variance and Multivariate Analysis," Biometrika, XLIV (1956), 361-376.

[34] Terpstra, T. J., "The Exact Probability Distribution of the T Statistic for Testing against Trend and its Normal Approximation," Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, LVI (1953), 433-437.

[35] --------"A Nonparametric Test for the Problem of k Samples," Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, LVII (1954), 505-512.

[36] Wald, A., "Note on the Consistency of the Maximum Likelihood Estimate," Annals of Mathematical Statistics, XX (1949), 595-601.

[37] Welch, B. L., "On the z-test in Randomized Blocks and Latin Squares," Biometrika, XXIX (1937), 21-52.

[38] Westenberg, J., "Significance Test for Median and Interquartile Range in Samples from Continuous Populations of any Form," Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, LI (1948), 252-261.

[39] Wilcoxon, F., "Individual Comparisons by Ranking Methods," Biometrics Bulletin (now Biometrics), I (1945), 80-83.