ON THE MATHEMATICAL PRINCIPLES UNDERLYING THE THEORY OF THE χ^2 TEST

by

Junjiro Ogawa
University of North Carolina

This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

Institute of Statistics
Mimeograph Series No. 162
January, 1957

ON THE MATHEMATICAL PRINCIPLES UNDERLYING THE THEORY OF THE χ^2 TEST*

by Junjiro Ogawa
Institute of Statistics
University of North Carolina Carolina, Chapel Hill, N. C.

Preface

Many contributions have been made so far to the theory of the χ^2 test; among others the author would like to mention R.A. Fisher [5]**, H.Hotelling [6], H.Cramér [4], G.A.Barnard [1], S.N.Roy and S.K.Mitra [11] and S.K. Mitra [7]. The author should refer to W.G. Cochran [3] also.

The rigorous proof of the theorem which will be stated in exact form later on, that the χ^2 statistic has the limiting χ^2 distribution with degrees of freedom reduced by the number of the independent parameters which were estimated from the sample, was first given by H. Cramér in his famous textbook [4], but some steps of the proof were skipped. Later S. N. Roy and S. K. Mitra [11] and S. K. Mitra [7] reasoned along the same lines and got theorems adjusted to various physical (or statistical) situations.

The purposes of this note are to present a complete and self-contained proof of Cramér's theorem on the one hand, and on the other hand to explain how the proof of the related theorems got by S.N. Roy and S.K.Mitra could be thrown back on that of Cramér's theorem from the mathematical point of view.

1. Cramér's theorem and its proof. We shall first start with the following.

Theorem 1 (Cramér). Suppose that we are given r functions $p_1(\underline{\alpha})$, ..., $p_r(\underline{\alpha})$ of $s < r$ variables $\underline{\alpha}' = (\alpha_1, \ldots, \alpha_s)$ such that for all points of a

---

* Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill.

** The numbers in square brackets refer to the bibliography listed at the end.

non-degenerate interval A in the s-dimensional space of the α_j, 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 $\frac{\partial p_i}{\partial \alpha_j}$ and $\frac{\partial^2 p_i}{\partial \alpha_j \partial \alpha_k}$.

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

Let the possible results of a certain random experiment $\mathscr{E}$ 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 the interval A. Let ν_i denote the number of results belonging to the i-th group, which occur in a sequence of n repetitions of $\mathscr{E}$, so that $\sum_{i=1}^{r} \nu_i = n$.

The equations

$$\sum_{i=1}^{r} \frac{\nu_i - np_i}{p_i} \frac{\partial p_i}{\partial \alpha_j} = 0, \; j = 1, 2, \ldots, s \tag{1.1}$$

of the modified χ^2-minimum method then have exactly one solution $\hat{\underline{\alpha}}' = (\hat{\alpha}_1, \ldots, \hat{\alpha}_s)$ such that $\hat{\underline{\alpha}}$ converges in probability to $\underline{\alpha}_0$ as $n \longrightarrow \infty$. The value of χ^2 obtained by inserting these values of $\underline{\alpha} = \hat{\underline{\alpha}}$ into

$$\chi^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\underline{\alpha}))^2}{np_i(\underline{\alpha})} \tag{1.2}$$

is, in the limit as $n \longrightarrow \infty$, distributed in a χ^2 distribution with r-s-1 degrees of freedom.

Remark: The proof of this theorem is somewhat intricate, and will be divided into two parts. In the first part it will be shown that the equation (1.1) have exactly one solution $\hat{\underline{\alpha}}$ which converges in probability to the true value $\underline{\alpha}_0$ as $n \longrightarrow \infty$. In the second part we shall consider the variables

$$(1.3) \qquad y_i = \frac{\nu_i - np_i(\hat{\underline{\alpha}})}{\sqrt{np_i(\hat{\underline{\alpha}})}} \quad , i = 1, 2, \ldots, r \quad .$$

where $\hat{\underline{\alpha}}$ is the solution of (1.1), the existence of which was already established in the first part. It will be shown here that $\underline{y}' = (y_1, \ldots, y_r)$ tends to a certain singular r-dimensional normal distribution under the assumption that $\underline{\nu}' = (\nu_1, \ldots, \nu_r)$ is the random vector whose probability distribution is given by the multinomial distribution

$$P_r(\underline{\nu}) = \frac{n!}{\nu_1! \ldots \nu_r!} \; p_1^{\nu_1}(\underline{\alpha}_0) \ldots p_r^{\nu_r}(\underline{\alpha}_0) \quad .$$

Proof of Theorem 1

Part I. Put

$$(1.4) \qquad \omega_j(\underline{\alpha}) = \sum_{i=1}^{r} \frac{\nu_i - np_i^0}{n} \left[\frac{1}{p_i}\frac{\partial p_i}{\partial \alpha_j} - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 \right] - \sum_{i=1}^{r} (p_i - p_i^0) \left[\frac{1}{p_i}\frac{\partial p_i}{\partial \alpha_j} - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 \right]$$

$$- \sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 \left[p_i - p_i^0 - \sum_{k=1}^{s} \left(\frac{\partial p_i}{\partial \alpha_k}\right)_0 (\alpha_k - \alpha_k^0) \right], j=1,2,\ldots,s$$

then the equations (1.1) can be rewritten as

$$(1.5) \qquad \sum_{k=1}^{s} (\alpha_k - \alpha_k^0) \sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 \left(\frac{\partial p_i}{\partial \alpha_k}\right)_0 = \sum_{i=1}^{r} \frac{\nu_i - np_i^0}{np_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 + \omega_j(\underline{\alpha}), j=1,\ldots,s \; .$$

Let

$$\underline{x}' = (x_1, \ldots, x_r) ,$$

where

$$x_i = \frac{\nu_i - np_i^0}{\sqrt{np_i^0}}, \quad i = 1, 2, \ldots, r , \tag{1.6}$$

and

$$B = \begin{bmatrix} \frac{1}{\sqrt{p_1^0}}\left(\frac{\partial p_1}{\partial \alpha_1}\right)_0 & \cdots & \frac{1}{\sqrt{p_1^0}}\left(\frac{\partial p_1}{\partial \alpha_s}\right)_0 \\ \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{p_r^0}}\left(\frac{\partial p_r}{\partial \alpha_1}\right)_0 & \cdots & \frac{1}{\sqrt{p_r^0}}\left(\frac{\partial p_r}{\partial \alpha_s}\right)_0 \end{bmatrix} . \tag{1.7}$$

Using the matrix notations, the equations (15) turn out to be

$$B'B(\underline{\alpha}-\alpha_0) = n^{-\frac{1}{2}} B'\underline{x} + \underline{\omega}(\underline{\alpha}) . \tag{1.8}$$

The rank of the matrix B is s, because it is easily seen that

$$B = \begin{bmatrix} \frac{1}{\sqrt{p_1^0}} & 0 \; \cdots & 0 \\ & \frac{1}{\sqrt{p_2^0}} \; \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 \; \cdots & \frac{1}{\sqrt{p_r^0}} \end{bmatrix} \cdot \begin{bmatrix} \left(\frac{\partial p_1}{\partial \alpha_1}\right)_0 & \left(\frac{\partial p_1}{\partial \alpha_2}\right)_0 & \cdots & \left(\frac{\partial p_1}{\partial \alpha_s}\right)_0 \\ \left(\frac{\partial p_2}{\partial \alpha_1}\right)_0 & \left(\frac{\partial p_2}{\partial \alpha_2}\right)_0 & \cdots & \left(\frac{\partial p_2}{\partial \alpha_s}\right)_0 \\ \cdots & \cdots & \cdots & \cdots \\ \left(\frac{\partial p_r}{\partial \alpha_1}\right)_0 & \left(\frac{\partial p_r}{\partial \alpha_2}\right)_0 & \cdots & \left(\frac{\partial p_r}{\partial \alpha_s}\right)_0 \end{bmatrix}$$

and the rank of the second factor matrix which is the matrix D calculated at $\underline{\alpha} = \underline{\alpha}_0$, is s by the assumption d) of the theorem, and the first factor matrix is a non-singular diagonal matrix of degree r by the assumption b). Hence the $s \times s$ square matrix $B'B$ is positive definite and so has its inverse. Therefore we have from (1.8) that

$$(1.9) \qquad \underline{\alpha} - \underline{\alpha}_0 = n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{\omega}(\underline{\alpha}) .$$

Now let the probability space unerlying the present consideration be $(\Omega, \mathcal{B}, P)$, i.e., let Ω be the set of the elementary events. $\mathcal{B}$ is some additive class of subsets of Ω containing Ω itself and P is a non-negative completely additive set function defined on the subsets belonging to $\mathcal{B}$ such that $P(\Omega) = 1$. We shall denote the elementary event by ξ.

By the inequality due to Tchebyschev-Bienaymé, we have, for any $\lambda > 0$,

$$(1.10) \qquad P\left[\left\{ \xi ; \, | \nu_i(\xi) - np_i^0 | \geq \lambda\sqrt{n}\right\}\right] \leq \frac{p_i^0(1-p_i^0)}{\lambda^2} < \frac{p_i^0}{\lambda^2} .$$

If we denote by Q_{in} the set $\left\{\xi \, ; \, |\nu_i(\xi) - np_i^0| < \lambda\sqrt{n}\right\}$ and put

$$(1.11) \qquad Q_n = \bigcap_{i=1}^{r} Q_{in} ,$$

then we can see that

$$(1.12) \qquad P(Q_n) = 1 - P(\overline{Q}_n) \geq 1 - \sum_{i=1}^{r} P(\overline{Q}_{in}) > 1 - \frac{1}{\lambda^2} .$$

Here we shall choose λ such that simultaneously

$$\frac{\lambda}{\sqrt{n}} \longrightarrow 0 \text{ and } \frac{1}{\lambda^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty ;$$

for example, $\lambda = n^q$, $0 < q < 1/4$ satisfies the above requirements. By the assumption b), it follows for any $\xi \varepsilon Q_n$ that

$$|x_i(\xi)| < \frac{\lambda}{c}, \quad i = 1, 2, \ldots, r. \tag{1.13}$$

For two inner points $\underline{\alpha}_1$ and $\underline{\alpha}_2$ of A, we have the difference

$$\begin{aligned}(1.14)\quad \omega_j(\underline{\alpha}_1)-\omega_j(\underline{\alpha}_2) &= \sum_{i=1}^{r} \frac{\nu_i - np_i^0}{n}\left[\frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right] \\ &- \sum_{i=1}^{r} \frac{\nu_i - np_i^0}{n}\left[\frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right] \\ &- \sum_{i=1}^{r}(p_i^1 - p_i^0)\left[\frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right] + \sum_{i=1}^{r}(p_i^2 - p_i^0)\left[\frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right] \\ &- \sum_{i=1}^{r}\frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\left[p_i^1 - p_i^0 - \sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial \alpha_k}\right)_0(\alpha_k^1 - \alpha_k^0)\right] + \sum_{i=1}^{r}\frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\left[p_i^2 - p_i^0 - \sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial \alpha_k}\right)_0(\alpha_k^2 - \alpha_k^0)\right]\end{aligned}$$

and this can be reduced to the following form:

$$\begin{aligned}\omega_j(\underline{\alpha}_1)-\omega_j(\underline{\alpha}_2) &= \sum_{i=1}^{r} \frac{\nu_i - np_i^0}{n}\left[\frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2\right] \\ (1.15)\quad &- \sum_{i=1}^{r}(p_i^1 - p_i^2)\left[\frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right] - \sum_{i=1}^{r}(p_i^2 - p_i^0)\left[\frac{1}{p_2^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2\right] \\ &- \sum_{i=1}^{r}\frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\left[p_i^1 - p_i^2 - \sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial \alpha_k}\right)_0(\alpha_k^1 - \alpha_k^2)\right].\end{aligned}$$

We shall mention here as a lemma some **results** from elementary calculus: If a function $\phi(x_1, \ldots, x_s)$ is defined in some finite closed and convex domain of s-dimensional Euclidean space such that $\frac{\partial \phi}{\partial x_i}$, $i = 1, \ldots, s$, exist and ~~are continuous everywhere~~ in that domain, then for any two points belonging to that domain, we have

$$|\phi(\underline{x}_1) - \phi(\underline{x}_2)| \leq K \cdot |\underline{x}_1 - \underline{x}_2| \quad . \tag{1.16}$$

where K is a constant depending only on the function $\phi(\underline{x})$

Proof of the Lemma. This can be shown as follows: For any two points $\underline{x}_1$ and $\underline{x}_2$ belonging to the domain, put

$$\underline{z} = \underline{x}_2 + t(\underline{x}_1 - \underline{x}_2) \ , \ 0 \leq t \leq 1 \ .$$

Then, since the domain was convex it is clear that $\underline{z}$ belongs to the domain for $0 \leq t \leq 1$. We shall apply the mean-value theorem of elementary differential calculus to the function

$$\Phi(t) = \phi(\underline{x}_2 + t(\underline{x}_1 - \underline{x}_2)) \quad ,$$

getting

$$\Phi(t) = \Phi(0) + t . \Phi'(\theta t) \ , \ 0 < \theta < 1 \quad .$$

This means that we have for $t = 1$

$$\phi(\underline{x}_1) - \phi(\underline{x}_2) = \sum_{i=1}^{s} \left(\frac{\partial \phi}{\partial x_i} \right)_{\underline{x} = \underline{x}_2 + \theta(\underline{x}_1 - \underline{x}_2)} . (x_i^1 - x_i^2) \ .$$

Hence by applying Cauchy's inequality we obtain

$$|\phi(\underline{x}_1) - \phi(\underline{x}_2)| \leq \sqrt{\sum_{i=1}^{s} \left(\frac{\partial \phi}{\partial x_i}\right)^2 . \sum_{i=1}^{s} (x_i^1 - x_i^2)^2}$$

or

$$|\phi(\underline{x}_1) - \phi(\underline{x}_2)| \leq \sqrt{\sum_{i=1}^{s} \left(\frac{\partial \phi}{\partial x_i}\right)^2} . |\underline{x}_1 - \underline{x}_2| \quad .$$

The function $\sum_{i=1}^{s}\left(\frac{\partial \phi}{\partial x_i}\right)^2$ is continuous in the domain and, since the domain was finite and closed, $\sum_{i=1}^{s}\left(\frac{\partial \phi}{\partial x_i}\right)^2$ attains its maximum in that domain, which we denote by K^2. Consequently we have proved the inequality (1.16).

By virtue of the inequality (1.16), we get the following

$$(1.18)\qquad \left|\frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2\right| \leq k_{1ij}\cdot\left|\underline{\alpha}_1 - \underline{\alpha}_2\right| ,$$

where k_{1ij} is a constant depending only on the function $\frac{1}{p_i}\frac{\partial p_i}{\partial \alpha_j}$. If we put

$$k_{1j} = \sum_{i=1}^{r} k_{1ij} \quad \text{and} \quad k_1 = \max_{1 \leq j \leq s}\left\{ k_{1j}\right\} ,$$

then we obtain

$$(1.19)\qquad \left|\sum_{i=1}^{r} \frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^2}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_2\right| \leq k_1\cdot\left|\underline{\alpha}_1-\underline{\alpha}_2\right| \quad \text{for all } j .$$

In a similar manner we can get the following inequalities with suitable constants k_2, k_3, k_4 and k_5:

$$(1.20)\qquad \left|\sum_{i=1}^{r}(p_i^1 - p_i^2)\right| \leq k_2\left|\underline{\alpha}_1 - \underline{\alpha}_2\right| ,$$

$$(1.21)\qquad \left|\sum_{i=1}^{r}(p_i^2 - p_i^0)\right| \leq k_3\left|\underline{\alpha}_2 - \underline{\alpha}_0\right| ,$$

$$(1.22)\qquad \left|\sum_{i=1}^{r} \frac{1}{p_i^1}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_1 - \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0\right| \leq k_4\left|\underline{\alpha}_1-\underline{\alpha}_0\right| \quad \text{for all } j ,$$

and

(1.23) $$\left|\sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i}{\partial \alpha_j}\right)_0 \left[p_i^1 - p_i^2 - \sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial \alpha_k}\right)_0 (\alpha_k^1 - \alpha_k^2)\right]\right|$$

$$\leq k_5 \left\{ |\underline{\alpha}_2 - \underline{\alpha}_0| \cdot |\underline{\alpha}_1 - \underline{\alpha}_2| + |\underline{\alpha}_1 - \underline{\alpha}_2|^2 \right\}$$

for all j.

Hence from (1.15) we can obtain

$$|\omega_j(\underline{\alpha}_1) - \omega_j(\underline{\alpha}_2)| \leq K_1 \frac{\lambda}{\sqrt{n}} |\underline{\alpha}_1 - \underline{\alpha}_2| + K_2 |\underline{\alpha}_1 - \underline{\alpha}_2| \, |\underline{\alpha}_1 - \underline{\alpha}_0| + K_3 |\underline{\alpha}_1 - \underline{\alpha}_2| \, |\underline{\alpha}_2 - \underline{\alpha}_0|$$

$$+ K_4 |\underline{\alpha}_1 - \underline{\alpha}_2| \, |\underline{\alpha}_2 - \underline{\alpha}_0| + K_5 |\underline{\alpha}_1 - \underline{\alpha}_2|^2,$$

or

(1.24) $$|\omega_j(\underline{\alpha}_1) - \omega_j(\underline{\alpha}_2)| \leq K |\underline{\alpha}_1 - \underline{\alpha}_2| \left\{ \frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_1 - \underline{\alpha}_0| + |\underline{\alpha}_2 - \underline{\alpha}_0| \right\},$$

and consequently we obtain

(1.25) $$|\underline{\omega}(\underline{\alpha}_1) - \underline{\omega}(\underline{\alpha}_2)| = \sqrt{\sum_{j=1}^{s} (\omega_j(\underline{\alpha}_1) - \omega_j(\underline{\alpha}_2))^2}$$

$$\leq K \cdot |\underline{\alpha}_1 - \underline{\alpha}_2| \cdot \left\{ \frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_1 - \underline{\alpha}_0| + |\underline{\alpha}_2 - \underline{\alpha}_0| \right\},$$

where the constants in (1.24) and (1.25), which were denoted by the same symbol K, may differ.

In order to solve the equation (1.19) for $\xi \in Q_n$, we shall apply the method of successive approximation, i.e., we define the vector $\underline{\alpha}_\nu$ by the recurrence relation

(1.26) $$\underline{\alpha}_\nu = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{\omega}(\underline{\alpha}_{\nu-1}).$$

It will easily be seen from the definition (14) of $\underline{\omega}(\underline{\alpha})$ that

$$\underline{\omega}(\underline{\alpha}_0) = \underline{0}.$$

Hence we have

$$\underline{\alpha}_1-\underline{\alpha}_0 = n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} \tag{1.27}$$

and

$$\underline{\alpha}_{\nu+1}-\underline{\alpha}_\nu = (B'B)^{-1}[\underline{\omega}(\underline{\alpha}_\nu)-\underline{\omega}(\underline{\alpha}_{\nu-1})] . \tag{1.28}$$

Let us denote the greatest absolute value of the elements of $(B'B)^{-1}B'$ and $(B'B)^{-1}$ by g_1; then we have from (1.27) that

$$|\alpha_j^1 - \alpha_j^0| \leq \frac{rg}{c}\frac{\lambda}{\sqrt{n}} .$$

Hence it follows that

$$|\underline{\alpha}_1 - \underline{\alpha}_0| \leqq K_2 \frac{\lambda}{\sqrt{n}} , \tag{1.29}$$

where K_2 is a suitable constant. From (1.28) we can obtain the following inequality by (1.26)

$$|\underline{\alpha}_{\nu+1}-\underline{\alpha}_\nu| \leqq K_3 |\underline{\alpha}_\nu-\underline{\alpha}_{\nu-1}| \left\{ |\underline{\alpha}_\nu-\underline{\alpha}_0| + |\underline{\alpha}_{\nu-1}-\underline{\alpha}_0| + \frac{\lambda}{\sqrt{n}} \right\} . \tag{1.30}$$

From this we can show, for sufficiently large values of n, that

$$|\underline{\alpha}_{\nu+1}-\underline{\alpha}_\nu| \leqq K_2 [(4K_2+1)K_3]^\nu \left(\frac{\lambda}{\sqrt{n}}\right)^{\nu+1} . \tag{1.31}$$

We shall prove (1.31) by the mathematical induction: We assume that (1.31) is true for all values of ν from 0 up to $\nu-1$ and for $\nu=0$ the inequality (1.31) reduces to (1.29). From (1.30) we have

$$|\underline{\alpha}_{\nu+1}-\underline{\alpha}_\nu| \leqq K_3K_2 [(4K_2+1)K_3]^{\nu-1} \cdot \left(\frac{\lambda}{\sqrt{n}}\right)^\nu \left\{ |\underline{\alpha}_\nu-\underline{\alpha}_0| + |\underline{\alpha}_{\nu-1}-\underline{\alpha}_0| + \frac{\lambda}{\sqrt{n}} \right\} . \tag{1.32}$$

On the other hand, it can be shown that

$$\frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_\nu - \underline{\alpha}_0| + |\underline{\alpha}_{\nu-1} - \underline{\alpha}_0| \le \frac{\lambda}{\sqrt{n}} + 2|\underline{\alpha}_1 - \underline{\alpha}_0| + 2|\underline{\alpha}_2 - \underline{\alpha}_1| + \cdots + 2|\underline{\alpha}_{\nu-1} - \underline{\alpha}_{\nu-2}| + |\underline{\alpha}_\nu - \underline{\alpha}_{\nu-1}|$$

$$< \frac{\lambda}{\sqrt{n}} + 2(|\underline{\alpha}_1 - \underline{\alpha}_0| + |\underline{\alpha}_2 - \underline{\alpha}_1| + \cdots + |\underline{\alpha}_\nu - \underline{\alpha}_{\nu-1}| + \cdots)$$

$$\le \frac{\lambda}{\sqrt{n}} + 2K_2 \frac{\lambda}{\sqrt{n}} \left\{ 1 + (4K_2+1)K_3 \frac{\lambda}{\sqrt{n}} + \left[(4K_2+1)K_3 \frac{\lambda}{\sqrt{n}} \right]^2 + \cdots \right\} .$$

Hence for sufficiently large values of n such that

$$0 < (4K_2+1)K_3 \frac{\lambda}{\sqrt{n}} < \frac{1}{2} ,$$

we have

$$(1.33) \qquad \frac{\lambda}{\sqrt{n}} + |\underline{\alpha}_2 - \underline{\alpha}_0| + |\underline{\alpha}_{\nu-1} - \underline{\alpha}_0| < \frac{\lambda}{\sqrt{n}} \left\{ 1 + \frac{2K_2}{1-(4K_2+1)K_3 \frac{\lambda}{\sqrt{n}}} \right\} < (4K_2+1) \frac{\lambda}{\sqrt{n}} .$$

Consequently from (1.32) and (1.33) we can obtain (1.31).

The infinite series

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

converges absolutely for any $\xi \in Q_n$ and for sufficiently large values of n; and if we define $\hat{\underline{\alpha}}$ by

$$(1.34) \qquad \hat{\underline{\alpha}} = \underline{\alpha}_0 + (\underline{\alpha}_1 - \underline{\alpha}_0) + (\underline{\alpha}_2 - \underline{\alpha}_1) + \cdots ,$$

then

$$|\hat{\underline{\alpha}} - \underline{\alpha}_0| \le |\underline{\alpha}_1 - \underline{\alpha}_0| + |\underline{\alpha}_2 - \underline{\alpha}_1| + \cdots < \frac{\lambda}{\sqrt{n}} \frac{1}{1-(4K_2+1)K_3 \frac{\lambda}{\sqrt{n}}} \longrightarrow 0 \text{ as } n \longrightarrow \infty ,$$

for any $\xi \in Q_n$. This means that $\hat{\underline{\alpha}}$ converges in probability to $\underline{\alpha}_0$ as $n \longrightarrow \infty$.

(1.34) can be expressed as

(1.35) $$\underline{\hat{\alpha}} = \lim_{\nu \longrightarrow \infty} \underline{\alpha}_\nu \quad ,$$

for any $\xi \varepsilon Q_n$ and for sufficiently large values of n; hence from (1.26) we get

(1.36) $$\underline{\hat{\alpha}} = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{\omega}(\underline{\hat{\alpha}}) \; ,$$

for any $\xi \varepsilon Q_n$ and for sufficiently large values of n.

Suppose that there exists another solution $\underline{\hat{\hat{\alpha}}}$ which converges in probability to $\underline{\alpha}_0$ as $n \longrightarrow \infty$; then, for sufficiently large values of n,

(1.37) $$\underline{\hat{\hat{\alpha}}} = \underline{\alpha}_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x}(B'B)^{-1}\underline{\omega}(\underline{\hat{\hat{\alpha}}}) \quad ,$$

for any $\xi \varepsilon Q_n^*$ such that $P(Q_n^*) \longrightarrow 1$ as $n \longrightarrow \infty$ and for sufficiently large values of n. It is clear that

$$P(Q_n) = P(Q_n \cap Q_n^*) + P(Q_n \cap \overline{Q}_n^*) \quad ;$$

therefore,

$$\lim_{n \longrightarrow \infty} P(Q_n \cap Q_n^*) = \lim_{n \longrightarrow \infty} P(Q_n) = 1 \quad .$$

For sufficiently large values of n and for any $\xi \varepsilon Q_n \cap Q_n^*$ we have

$$\underline{\hat{\hat{\alpha}}} - \underline{\hat{\alpha}} = (B'B)^{-1}[\underline{\omega}(\underline{\hat{\hat{\alpha}}}) - \underline{\omega}(\underline{\hat{\alpha}})] \; ;$$

and hence it follows by (1.25) that

(1.38) $$|\underline{\hat{\hat{\alpha}}} - \underline{\hat{\alpha}}| \leq K.|\underline{\hat{\hat{\alpha}}} - \underline{\hat{\alpha}}| \left\{ \frac{\lambda}{\sqrt{n}} + |\underline{\hat{\hat{\alpha}}} - \underline{\alpha}_0| + |\underline{\hat{\alpha}} - \underline{\alpha}_0| \right\} \; .$$

If $|\underline{\hat{\hat{\alpha}}} - \underline{\hat{\alpha}}| > 0$ for all n, then (1.38) leads to a contradiction $1 \leq 0$ as $n \longrightarrow \infty$. This means that for sufficiently large n there exists an exactly one solution $\underline{\hat{\alpha}}$ which converges in probability to $\underline{\alpha}_0$ as $n \longrightarrow \infty$.

Part 2. For the solution $\hat{\underline{\alpha}}$ we have

$$(1.39) \qquad (B'B)^{-1}\underline{\omega}(\hat{\underline{\alpha}}) = \hat{\underline{\alpha}} - \underline{\alpha}_1 = (\underline{\alpha}_2-\underline{\alpha}_1)+(\underline{\alpha}_3-\underline{\alpha}_2)+ \cdots ;$$

hence, by using the inequalities (1.31), we can see that every component of $(B'B)^{-1}\underline{\omega}(\hat{\underline{\alpha}})$ is dominated by $K' \frac{\lambda^2}{n}$ in its absolute value, where K' is a certain constant independent of n. Therefore, we can write

$$(1.40) \qquad \hat{\underline{\alpha}}-\underline{\alpha}_0 = n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x}+(B'B)^{-1}\underline{\omega}(\hat{\underline{\alpha}})$$

$$= n^{-\frac{1}{2}}(B'B)B'\underline{x}+ \frac{K'\lambda^2}{n}\underline{\theta}', 0 \leq \theta_j' \leq 1, j=1,\ldots,s .$$

Let $\underline{y}' = (y_1, \ldots, y_r)$, where

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

then it can be rewritten as follows:

$$(1.41) \qquad y_i = \frac{\nu_i-np_i^0}{\sqrt{np_i^0}} - \sqrt{n}\,\frac{p_i(\hat{\underline{\alpha}})-p_i^0}{\sqrt{p_1^0}} + \frac{\nu_i-np_i(\hat{\underline{\alpha}})}{\sqrt{n}}\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}} - \frac{1}{\sqrt{p_i^0}}\right)$$

$$= x_i - \sqrt{\frac{n}{p_i^0}}\,(p_i(\hat{\underline{\alpha}})-p_i^0) + \frac{\nu_i-np_i^0}{\sqrt{n}}\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}} - \frac{1}{\sqrt{p_i^0}}\right)$$

$$-\sqrt{n}(p_i(\hat{\underline{\alpha}})-p_i^0)\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}} - \frac{1}{\sqrt{p_i^0}}\right) .$$

Now

$$\sqrt{\frac{n}{p_i^0}}(p_i(\hat{\underline{\alpha}})-p_i^0) = \sqrt{\frac{n}{p_i^0}}\sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial \alpha_k}\right)_0(\hat{\alpha}_k-\alpha_k^0) + \frac{1}{2}\sqrt{\frac{n}{p_i^0}}\sum_{k,\ell=1}^{s}\left(\frac{\partial^2 p_i}{\partial\alpha_k\partial\alpha_\ell}\right)_*(\hat{\alpha}_k-\alpha_k^0)(\hat{\alpha}_\ell-\alpha_\ell^0).$$

where $\left(\frac{\partial^2 p_i}{\partial\alpha_k\partial\alpha_\ell}\right)_*$ means that the value of $\frac{\partial^2 p_i}{\partial\alpha_k\partial\alpha_\ell}$ is calculated at a certain point on the segment joining $\underline{\alpha}_0$ and $\hat{\underline{\alpha}}$. On the other hand, since

$$\hat{\alpha}_k-\alpha_k^0 = \alpha_k'-\alpha_k^0+\frac{K'\lambda^2}{n}\theta_k',\ \alpha_k'-\alpha_k^0 = O\left(\frac{\lambda}{\sqrt{n}}\right) ,$$

it follows that

$$\hat{\alpha}_k-\alpha_k^0 = O\left(\frac{\lambda}{\sqrt{n}}+\frac{\lambda^2}{n}\right) = O\left(\frac{\lambda}{\sqrt{n}}\right) \quad . \tag{1.42}$$

Consequently we have

$$\sqrt{\frac{n}{p_i^0}}\,(p_2(\hat{\underline{\alpha}})-p_1^0)=\sqrt{\frac{n}{p_i^0}}\sum_{k=1}^{s}\left(\frac{\partial p_i}{\partial\alpha_k}\right)_0(\hat{\alpha}_k-\alpha_k^0)+O\left(\frac{\lambda^2}{\sqrt{n}}\right) \quad . \tag{1.43}$$

In a similar manner we can see that

$$\frac{\nu_i-np_i^0}{\sqrt{n}}\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}}-\frac{1}{\sqrt{p_i^0}}\right)= x_i\sqrt{p_i^0}\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}}-\frac{1}{\sqrt{p_i^0}}\right)=O\left(\frac{\lambda^2}{\sqrt{n}}\right) , \tag{1.44}$$

and

$$\sqrt{n}(p_i(\hat{\underline{\alpha}})-p_i^0)\left(\frac{1}{\sqrt{p_i(\hat{\underline{\alpha}})}}-\frac{1}{\sqrt{p_i^0}}\right)=\sqrt{n}\ O\left(\frac{\lambda^2}{n}\right)=O\left(\frac{\lambda^2}{\sqrt{n}}\right) \quad . \tag{1.45}$$

Hence we obtain the relation

$$\underline{y}=\underline{x}-\sqrt{n}\ B(\hat{\underline{\alpha}}-\underline{\alpha}_0)+\frac{K''\lambda^2}{\sqrt{n}}\underline{\theta}^?\ ,\ 0\leq\ \theta_j^2\ \leq\ ,j = 1,\ \ldots,\ s. \tag{1.46}$$

Substituting (1.40) into (1.46), it follows that

$$\underline{y} = (1-B(B'B)^{-1}B')\underline{x} + \frac{K\lambda^2}{\sqrt{n}}\underline{\theta}\ , \tag{1.47}$$

where $\underline{\theta}$ is a random vector such that $0 < \theta_i \leq 1,\ \ i = 1, \ldots, r,$ and

K is a certain constant independent of n. Since the random vector $\frac{K\lambda^2}{\sqrt{n}}\underline{\theta}$ converges in probability to $\underline{0}$ as $n \longrightarrow \infty$, by using the theorem which was stated in the last paragraph of § 22.6 of Cramér's book [4], we can see that the distribution of $\underline{y}$ is, in the limit as $n \longrightarrow \infty$, the same as that of $(I-B(B'B)^{-1}B')\underline{x}$. The author would like to refer also to Ogawa [8] in this connection.

Now, if we put

$$\underline{p}' = (\sqrt{p_1^0}, \sqrt{p_2^0}, \ldots, \sqrt{p_r^0}) \quad ,$$

the limiting distribution of $\underline{x}$ as $n \longrightarrow \infty$ is the r-dimensional singular normal distribution with mean $\underline{0}$ and the covariance matrix

$$\Lambda = I - \underline{p}\underline{p}' \quad . \tag{1.48}$$

The statistic $\chi^2 = \underline{y}'\underline{y}$, whose distribution is under consideration, has the same limiting distribution as that of the statistic

$$\begin{aligned} \chi_*^2 &= \underline{x}'(I-B(B'B)^{-1}B')'(I-B(B'B)^{-1}B')\,\underline{x} \\ &= \underline{x}'(I-B(B'B)^{-1}B')\,\underline{x} \ . \end{aligned} \tag{1.49}$$

Since the s x s matrix $B'B$ is positive definite (of course symmetric), we can transform it by a suitable orthogonal matrix C into a diagonal form whose diagonal elements are all positive, i.e.,

$$C'(B'B)C = \begin{bmatrix} \mu_1^2 & & & 0 \\ & \mu_2^2 & & \\ & & \ddots & \\ 0 & & & \mu_s^2 \end{bmatrix} \equiv M^2, \text{ say} \ .$$

In other words, it means that

$$(B'B)^{-1} = CM^{-1} \cdot M^{-1}C' ,$$

whence we have the relation

(1.50) $$(BCM^{-1})'(BCM^{-1}) = M^{-1}C'B'BCM^{-1} = I .$$

(1.50) means that s column vectors of the matrix BCM^{-1} form a system of s orthogonal vectors of r-dimensions. If we complete BCM^{-1} to an r-dimensional orthogonal matrix K by adjoining $(r-s)$ column vectors $\underline{b}_{s+1}, \ldots, \underline{b}_r$, i.e., if we set

$$K = (BCM^{-1} \vdots \underline{b}_{s+1} \cdots \underline{b}_r)$$

and

$$K'K = I ,$$

then it will be seen that

(1.51) $$K'B(B'B)^{-1}B'K = (K'BCM^{-1})(K'BCM^{-1})'$$

$$= \begin{bmatrix} I_s \\ \cdots \\ 0 \end{bmatrix} \cdot [\, I_s \vdots 0 \,] = \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix} ,$$

and consequently we obtain

(1.52) $$K'(I - B(B'B)^{-1}B)K = \begin{bmatrix} 0 & 0 \\ 0 & I_{r-s} \end{bmatrix} .$$

Now make the orthogonal transformation

(1.53) $$\underline{\xi} = K'\underline{x} \quad ;$$

then

(1.54) $$\mathcal{E}(\underline{\xi}) = \underline{0} \text{ and } \mathcal{E}(\underline{\xi}\underline{\xi}') = K'(I_r - \underline{p}\underline{p}')K$$

and

(1.55)

$$\chi^2_* = \underline{x}'KK'(I-B(B'B)^{-1})KK'\underline{x} = \underline{\xi}'K'(I-B(B'B)^{-1}B')K\underline{\xi}$$

$$= \underline{\xi}' \begin{pmatrix} 0 & 0 \\ 0 & I_{r-s} \end{pmatrix} \underline{\xi} = \xi_{s+1}^2 + . \quad + \xi_r^2 \quad .$$

Since $\underline{p}'B = \underline{0}'$, we have

$$\underline{p}'K = \underline{p}'(BCM^{-1} \vdots \underline{b}_{s+1} \cdots \underline{b}_r) = (\overbrace{0 \cdots 0}^{s} \vdots e_{s+1} \cdots e_r)$$

where

$$e_g = \underline{p}'\underline{b}_g \,, \; g = s+1, \ldots, r \; .$$

Therefore,

$$K'\underline{pp}'K \quad = \quad \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}, \text{ where}$$

$$Q = (q_{gh}) \; ; \; q_{gh} = e_g e_h \;, \; g, h = s+1, \ldots, r \; .$$

Consequently the covariance matrix of $(\xi_{r+1}, \ldots, \xi_s)$ is

(1.56) $$I_{r-s} - Q \quad .$$

From the relation $\underline{p}'\underline{x} = 0$ we have

(1.57) $$e_{s+1}\,\xi_{s+1} + \cdots + e_r\xi_r = 0 \quad .$$

Consider the following generalized Helmert's orthogonal transformation:

(1.58) $$\underline{\eta}^* = P\underline{\xi}^*$$

where

$$
(1.59)\,P=\begin{pmatrix}
\frac{e_{s+1}}{\sqrt{e_{s+1}^2+\dots+e_r^2}} & \frac{e_{s+2}}{\sqrt{e_{s+1}^2+\dots+e_r^2}} & \cdots & \frac{e_{r-1}}{\sqrt{e_{s+1}^2+\dots+e_r^2}} & \frac{e_r}{\sqrt{e_{s+1}^2+\dots+e_r^2}} \\
\sqrt{\frac{e_{s+2}^2+\dots+e_r^2}{e_{s+1}^2+e_{s+2}^2+\dots+e_r^2}} & \frac{-e_{s+1}e_{s+2}}{\sqrt{(e_{s+1}^2+\dots+e_r^2)(e_{s+2}^2+\dots+e_r^2)}} & \cdots & \frac{-e_{s+1}e_{r-1}}{\sqrt{(e_{s+1}^2+\dots+e_r^2)(e_{s+2}^2+\dots+e_r^2)}} & \frac{-e_{s+1}e_r}{\sqrt{(e_{s+1}^2+\dots+e_2^2)(e_{s+2}^2+\dots+e_r^2)}} \\
0 & \sqrt{\frac{e_{s+3}^2+\dots+e_r^2}{e_{s+2}^2+e_{s+3}^2+\dots+e_r^2}} & \cdots & \frac{-e_{s+2}e_{r-1}}{\sqrt{(e_{s+2}^2+\dots+e_r^2)(e_{s+3}^2+\dots+e_r^2)}} & \frac{-e_{s+2}e_r}{\sqrt{(e_{s+2}^2+\dots+e_r^2)(e_{s+3}^2+\dots e_r^2)}} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & \sqrt{\frac{e_r^2}{e_{r-1}^2+e_r^2}} & \frac{-e_{r-1}e_r}{\sqrt{(e_{r-1}^2+e_r^2)e_r^2}}
\end{pmatrix}
$$

then it is clear from (1.57) that

(1.60) $$\eta_{s+1} = 0$$

and, therefore,

(1.61) $$\chi_*^2 = \eta_{s+2}^2 + \dots + \eta_r^2 \quad .$$

The covariance matrix of $(\eta_{s+1}, \eta_{s+2}, \dots, \eta_r)$ is

$$P(I_{rs} - Q)P' = P(I_{rs} - \underline{e}\underline{e}')P'$$
$$= I_{rs} - P\underline{e} \cdot \underline{e}'P',$$

where

$$\underline{e}' = (e_{s+1}, e_{s+2}, \dots, e_r) \quad ,$$

and

(1.62) $$P\underline{e} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad ,$$

because $\sum_{s+1}^{r} e_g^2 = \text{tr}Q = \text{tr}\ \underline{p}\underline{p}' = \text{tr}\ \underline{p}'p = \sum_1^r p_i^0 = 1$. Whence we have

(1.63) $$I_{r-s} - PQP' = \begin{bmatrix} 0 & \underline{0}' \\ \underline{0} & I_{r-s-1} \end{bmatrix}$$

This means that $\eta_{s+1} = 0$ with probability one, and $\eta_{s+2}, \dots, \eta_r$ are independent normal variates with means $0, \dots, 0$ and covariances I_{r-s-1}. Thus the limiting distribution of χ_*^2 as $n \to \infty$ and consequently the limiting distribution of χ^2, is the χ^2 distribution with degrees of freedom r-1-s. This completes the proof of Theorem 1.

2. Some remarks to Theorem 1 and its proof.

The parameter space of a multinomial probability distribution is, in general, the intersection of the (r-1)-dimensional flat $\sum_1^r p_i = 1$ and the r-dimensional unit interval $\left\{ \underline{p};\ 0 \leq p_i \leq 1,\ i = 1, 2, \ldots, r \right\}$. We shall call the parameter space Π.

Mathematically speaking, the assumptions of Theorem 1 specify the mapping of a non-degenerate s-dimensional interval A to the subset

$$(2.1) \qquad \Pi_0 = \Pi \cap \left\{ \underline{p};p_i > c^2 > 0,\ i=1,2,\ldots,r \right\}$$

of Π. In other words, assuming that the true parameter value $\underline{p}^0$ is an inner point of Π_0, if the subset Π_0 of Π is contained in the image of some non-degenerate interval of Euclidean space mapped by the function $\underline{p}'(\underline{\alpha})=(p_1(\underline{\alpha}), \ldots, p_r(\underline{\alpha}))$ such that $\underline{p}^0$ corresponds to an inner point of that interval and the mapping function $\underline{p}(\underline{\alpha})$ satisfies the condition (c), then the theorem holds true. Condition (d) assures that the dimensionality is preserved by this mapping.

Modified χ^2-minimum equation (1.8) is the same as the usual normal equation of the least-squares method up to the term $\underline{\omega}(\underline{\alpha})$. The fundamental idea of Cramér's proof is that, in the limit as $n \to \infty$, the whole situation can be reduced to that of the usual least-squares method.

In addition to the conditions (a) - (d), if we assume further that $\alpha_1 = \alpha_1^0, \ldots, \alpha_t = \alpha_t^0 (t < s)$ and put

$$(2.2) \qquad B^* = \begin{pmatrix} \frac{1}{\sqrt{p_1^0}}\left(\frac{\partial p_1}{\partial \alpha_{t+1}}\right)_0 & \cdots & \frac{1}{\sqrt{p_1^0}}\left(\frac{\partial p_1}{\partial \alpha_s}\right)_0 \\ \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{p_r^0}}\left(\frac{\partial p_r}{\partial \alpha_{t+1}}\right)_0 & \cdots & \frac{1}{\sqrt{p_r^0}}\left(\frac{\partial p_r}{\partial \alpha_s}\right)_0 \end{pmatrix},$$

$$(2.3) \qquad \underline{\alpha}^{*'} = (\alpha_{t+1}, \ldots, \alpha_s) \quad ,$$

and if $\omega_j^*(\underline{\alpha}^*)$, $j = t+1, \ldots, s$ are defined analogously with $\omega_j(\underline{\alpha})$, then the modified χ^2-minimum equation comes out as follows:

$$(2.4) \qquad B^{*'}B^{*'}(\underline{\alpha}^*-\underline{\alpha}_0^*) = n^{-\frac{1}{2}}B^{*'}\underline{x}+ \underline{\omega}^*(\underline{\alpha}^*);$$

Accordingly we have in this case

$$(2.5) \quad \underline{y}^* = \begin{pmatrix} \dfrac{\nu_1-np_1(\alpha_1^0 \ldots \alpha_t^0 , \underline{\alpha}^*)}{\sqrt{np_1(\alpha_1^0 \ldots \alpha_t^0 , \underline{\alpha}^*)}} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\nu_r-np_r(\alpha_1^0 \ldots \alpha_t^0, \underline{\alpha}^*)}{\sqrt{np_r(\alpha_1^0 \ldots \alpha_t^0, \underline{\alpha}^*)}} \end{pmatrix} =(I-B^*(B^*B^*)^{-1}B^{*'})\underline{x}+ \frac{K^*\lambda^2}{\sqrt{n}}\,\underline{\theta}^* ,$$

and the χ^2 statistic has, in the limit as $n \longrightarrow \infty$, the same distribution as

$$(2.6) \qquad \underline{y}^{*'}\underline{y}^{*'} = \underline{x}'(I-B^*(B^{*'}B^*)^{-1}B^{*'})\,\underline{x} \quad ,$$

which turns out to be the χ^2 distribution with degrees of freedom $r-1(s-t)=$ $r-s+t-1$.

From the usual theory of the least squares [see for example, Ogawa [10]), we can conclude that $\underline{y}^{*'}\underline{y}^*-\underline{y}'\underline{y}$ is independent of $\underline{y}'\underline{y}$ and is distributed as χ^2-distribution with degrees of freedom t in the limit as $n \longrightarrow \infty$.

The author wants to add one more remark: Cramér's theorem assures that there exists one and only one solution which is consistent of the modified χ^2-minimum equation. In our multinomial probability distribution case, it will be easily seen that the modified χ^2-minimum equation coincides with the maximum likelihood equation in the Cramér's sense. A. Wald [12] proved that the statistic which gives the greatest maximum of the likelihood function (if any at all) is consistent under the following 8 assumptions:

Assumption 1: Cumulative distribution function $F(x;\theta)$ is either discrete for all θ or is absolutely continuous for all θ.

Assumption 2: We shall define auxiliary functions as follows:

$$f(x,\theta,\rho) = \sup_{|\theta - \theta'| \leq \rho} f(x;\theta') \text{ and } f^*(x,\theta,\rho) = f(x,\theta,\rho) \text{ if } f(x,\theta,\rho) > 1 \text{ and}$$
$$= 1 \quad \text{otherwise}$$

$$\phi(x,r) = \sup_{|\theta| > r} f(x;\theta) \text{ and } \phi^*(x,r) = \phi(x,r) \text{ if } \phi(x,r) > 1 \text{ and}$$
$$= 1 \quad \text{otherwise .}$$

For sufficiently small ρ and for sufficiently large r the expected values

$$\int_{-\infty}^{\infty} \log f^*(x,\theta,\rho)dF(x;\theta_0) \text{ and } \int_{-\infty}^{\infty} \log \phi^*(x,r)dF(x;\theta_0) \text{ are finite and}$$

θ_0 denotes the true parameter point.

Assumption 3: If $\lim_{i \to \infty} \theta_i = \theta$, then $\lim_{i \to \infty} f(x;\theta_i) = f(x;\theta)$ for all x except perhaps on a set which may depend on the limit point θ (but not on the sequence $\{\theta_i\}$) and whose probability measure is zero under the probability distribution corresponding to the true parameter point θ_0.

Assumption 4: If θ_1 is a parameter point different from the true parameter point θ_0, then $F(x;\theta_1) \neq F(x;\theta_0)$ for at least one value of x.

Assumption 5: If $\lim_{i \to \infty} \theta_i = \infty$, then $\lim_{i \to \infty} f(x;\theta_i) = 0$ for any x except perhaps on a fixed set (independent of the sequence $\{\theta_i\}$ whose probability is zero under the probability distribution corresponding to the true parameter point θ_0.

Assumption 6: For the true parameter point θ_0 we have

$$\int_{-\infty}^{\infty} \left| \log f(x;\theta_0) \right| dF(x;\theta_0) < \infty \quad .$$

Assumption 7: The parameter space Π is a closed subset of the k-dimensional Cartesian space.

Assumption 8: $f(x,\theta,\rho)$ is a measurable function of x for any θ, and .

In our special case we may put

$$(2.7) \qquad \theta = \underline{p} \quad \text{and} \quad \underline{x}_\alpha' = (x_{1\alpha}, x_{2\alpha}, \ldots, x_{r\alpha})$$

$$\nu_i = \sum_{\alpha=1}^{n} x_{i\alpha}, \; i = 1,2, \ldots, r \quad ,$$

$$(2.8) \qquad f(x_\alpha;\theta) = p_1^{x_{1\alpha}} p_2^{x_{2\alpha}} \ldots p_r^{x_{r\alpha}} \; ,$$

where

$$x_{i\alpha} = \begin{cases} 1 & \text{if in the ith repetition of the random experiment the event happens which belongs to ith class} \\ 0 & \text{Otherwise} \end{cases}$$

Assumption 1 is trivial. Assumption 2 is satisfied with

$$f^*(x,\theta,\rho)=1, \quad \phi^*(x,r)= 1 \; .$$

Assumptions 3 and 4 are also trivially satisfied. Assumption 5 is also trivial.

$$\log f(x;\theta_0) \equiv \sum_{i=1}^{r} x_i \log p_i \quad \text{and thus we have}$$

$$\mathcal{E} \left(\left| \log f(x;\theta_0) \right| \right) = - \sum_{i=1}^{r} p_i^0 \log p_i^0 < \infty \quad .$$

Hence Assumption 6 is also satisfied. We shall confine ourselves to the subset $\Pi_0^* = \Pi \cap \{\underline{p}; p_i \geq c^2 > 0,\ i=1, \ldots, r\}$; then Assumption 7 is satisfied. Assumption 8 is trivial because $f(x;\theta)$ is a step function in this case.

Thus the solution $\hat{\underline{\alpha}}$ should be the one which gives the greatest maximum of the likelihood function.

3. On some related theorems. S. N. Roy and S. K. Mitra [11] proposed the following modification of the theorem of Cramér.

Theorem 2. (Roy and Mitra) In addition to the statements of Theorem 1, let now $f_k(\underline{\alpha})$, $k = 1, \ldots, t < s$ be t real valued-functions of $\underline{\alpha}$ such that

(e) at every point $\underline{\alpha}$ in A, the functions $f_k(\underline{\alpha})$ are continuous and possess continuous partial derivatives with respect to α_j up to the second order.

(f) the matrix $\left(\frac{\partial f_k}{\partial \alpha_j}\right)$ is of rank t for every $\underline{\alpha}$ in A.

For $k = 1,2,\ldots,t$, let f_k^0 be a member in the range of f_k in A. Denote by H_0 the hypothesis

$$H_0 \colon \underline{f}(\underline{\alpha}) = \underline{f}^0 \quad ,$$

then if H_0 is true, the equations

$$\sum_{i=1}^{r} \frac{\nu_i - np_i}{p_i} \frac{\partial p_i}{\partial \alpha_j} + \sum_{k=1}^{t} \lambda_k \frac{\partial f_k}{\partial \alpha_j} = 0, \quad j = 1,2,\ldots,s \tag{3.1}$$
$$\underline{f}(\underline{\alpha}) = \underline{f}^0$$

for minimizing χ^2 in the modified sense subject to H_0, have exactly one system of solutions $\hat{\hat{\underline{\alpha}}}' = (\hat{\hat{\alpha}}_1 \ldots \hat{\hat{\alpha}}_s)$, $\hat{\hat{\underline{\lambda}}}' = (\hat{\hat{\lambda}}_1 \ldots \hat{\hat{\lambda}}_t)$, such that $\hat{\hat{\underline{\alpha}}} \longrightarrow \underline{\alpha}_0$ in probability as $n \longrightarrow \infty$ and the statistic

(3.2) $$\chi_1^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\hat{\hat{\underline{\alpha}}}))^2}{np_i(\hat{\hat{\underline{\alpha}}})}$$

is, in the limit as $n \longrightarrow \infty$, distributed as χ^2 distribution with degrees of freedom r-s+t-1; also $\chi^2_{H0} = \chi_1^2 - \chi^2$ under H_0 is, in the limit as $n \longrightarrow \infty$, distributed independently of χ^2 as a χ^2 distribution with degrees of freedom t.

This can be proved easily from the remarks in § 2. We shall consider the transformation

(3.3) $$f_1(\underline{\alpha}) = \xi_1, \ldots, f_t(\underline{\alpha}) = \xi_t, \alpha_{t+1} = \xi_{t+1}, \ldots, \alpha_s = \xi_s,$$

here we have assumed, without any loss of generality, that the principal minor does not vanish identically, i.e., that

$$\begin{vmatrix} \frac{\partial f_1}{\partial \alpha_1} & \cdots & \frac{\partial f_1}{\partial \alpha_t} \\ \vdots & & \vdots \\ \frac{\partial f_t}{\partial \alpha_1} & \cdots & \frac{\partial f_t}{\partial \alpha_t} \end{vmatrix} \neq 0 .$$

This means that $\alpha_1, \ldots, \alpha_t$ can be expressed as functions of $\xi_1 \ldots \xi_t$ and $\alpha_{t+1}, \ldots, \alpha_s$.

More precisely, in some neighborhood of the point

$$\xi_1 = f_1^0, \ldots, \xi_t = f_t^0, \xi_{t+1} = \alpha_{t+1}^0, \ldots, \xi_s = \alpha_s^0 \quad ,$$

there exists implicit functions

(3.4) $$\alpha_i = \phi_i(\xi_1 \ldots \xi_t \xi_{t+1} \ldots \xi_s), \; i = 1, \ldots, t \quad ,$$

such that

$$(3.5) \qquad \alpha_i^0 = \phi_i(f_1^0 \ldots f_t^0\, \alpha_{t+1}^0 \ldots \alpha_s^0), \; i=1, \ldots, t \quad ,$$

and they are continuous and possess continuous partial derivatives up to the second order.

If we denote by S the image in ξ-space of the interval A mapped under the transformation (3.3), the assumptions of the theorem assure that the neighborhood contains some non-degenerate interval of s-dimensional Cartesian space.

Hence under the hypothesis H_0, the functions

$$(3.6) \qquad \alpha_i = \phi_i(f_1^0 \ldots f_t^0 \xi_{t+1} \ldots \xi_s), \; i = 1, \ldots, s$$

are defined in a non-degenerate (s-t)-dimensional interval S_0 and possess continuous partial derivatives with respect to $\xi_{t+1}, \ldots, \xi_s$ up to the 2nd order. Consequently the r functions p_i as functions of $\xi_{t+1}, \ldots, \xi_s$ possess continuous derivatives up to the second order. In fact, for $j = t+1, \ldots, s$

$$(3.7) \qquad \frac{\partial p_i}{\partial \xi_j} = \frac{\partial p_i}{\partial \alpha_1}\frac{\partial \alpha_1}{\partial \xi_j} + \ldots + \frac{\partial p_i}{\partial \alpha_t}\frac{\partial \alpha_t}{\partial \xi_j} + \frac{\partial p_i}{\partial \alpha_j}$$

and

$$(3.8) \qquad \frac{\partial^2 p_i}{\partial \xi_j \partial \xi_k} = \sum_{g=1}^{t} \left\{ \sum_{h=1}^{t} \frac{\partial^2 p_i}{\partial \alpha_g \partial \alpha_h}\frac{\partial \alpha_g}{\partial \xi_j}\frac{\partial \alpha_k}{\partial \xi_k} + \frac{\partial^2 p_i}{\partial \alpha_g \partial \alpha_k}\frac{\partial \alpha_g}{\partial \xi_j} \right\}$$

$$+ \sum_{g=1}^{t} \frac{\partial p_i}{\partial \alpha_g}\frac{\partial^2 \alpha_g}{\partial \xi_j \partial \xi_k} + \sum_{h=1}^{t} \frac{\partial^2 p_i}{\partial \alpha_j \partial \alpha_g}\frac{\partial \alpha_g}{\partial \xi_k} + \frac{\partial^2 p_i}{\partial \alpha_j^2} \quad ,$$

and these are all continuous in S_0. Furthermore, if for some point in S_0

there exist linear relations

$$(3.9) \qquad a_{t+1}(\underline{\xi}^*)\frac{\partial p_i}{\partial \xi_{t+1}} + \dots + a_s(\underline{\xi}^*)\frac{\partial p_i}{\partial \xi_s} = 0, \; i=1,\dots,r \quad ,$$

then this means, from (3.7), that

$$\frac{\partial p_i}{\partial \alpha_1}\cdot\left(\sum_{t=1}^{s} a_g(\underline{\xi}^*)\frac{\partial \alpha_1}{\partial \xi_g}\right)+\dots+\frac{\partial p_i}{\partial \alpha_t}\left(\sum_{t+1}^{s} a_g(\underline{\xi}^*)\frac{\partial \alpha_t}{\partial \xi_g}\right)+a_{t+1}(\underline{\xi}^*)\frac{\partial p_i}{\partial \xi_{t+1}}+\dots+a_s(\underline{\xi}^*)\frac{\partial p_i}{\partial \xi_s}=0;$$

$$i=1, \dots, r \; ,$$

thus we get $a_{t+1}(\underline{\xi}^*) = \dots = a_s(\underline{\xi}^*) = 0$, i.e., the rank of the matrix

$\left(\frac{\partial p_i}{\partial \xi_j}\right) \quad \begin{matrix} i = 1, \dots, r \\ j = t+1,\dots,s \end{matrix}$ is s-t .

Thus Theorem 2 is reduced to Theorem 1 with $\xi_{t+1}, \dots, \xi_s$ playing the role of $\alpha_1, \dots, \alpha_s$.

The second part of Theorem 2 is clear from the remark in § 2.

Next we shall consider the following

Theorem 3: (Roy and Mitra) Let the possible results of a certain random experiment $\mathcal{E}$ be divided into r mutually exclusive groups, and suppose that the probability of obtaining a result belonging to the ith group is $p_i^0 \geq c^2 > 0$. Let ν_i denote the number of results belonging to the ith group, which occur in a sequence of n repetitions of $\mathcal{E}$ so that $\sum_1^r \nu_i = n$.

Let us consider the hypothesis

$$H_0: f_j(p_1,\dots,p_r)=0, \; j=1, \dots, r-s-1$$

where f_j possesses continuous partial derivatives up to the second order and the matrix $\left(\frac{\partial f_i}{\partial p_t}\right) \begin{matrix} j =1,\dots,r-s-1 \\ t =1,\dots,r-1 \end{matrix}$ is of rank r-s-1

Then the modified χ^2 minimum equation subject to H_0

$$(3.10)\quad \frac{\nu_i - np_i}{p_i} - \frac{\nu_r - np_r}{p_r} + \sum_{k=1}^{r-s-1} \lambda_k\left(\frac{\partial f_k}{\partial p_i} - \frac{\partial f_k}{\partial p_r}\right) = 0, \quad i=1, \ldots, r-1 \quad .$$

$$f_k(\underline{p}) = 0, \quad k = 1, \ldots, r-s-1$$

$$\Sigma p_i = 1$$

have one and only one solution $\hat{\underline{p}}$ and $\hat{\underline{\lambda}}$ such that $\hat{\underline{p}} \longrightarrow \underline{p}^0$ in probability as $n \longrightarrow \infty$ and

$$(3.11)\qquad \chi^2 = \sum_{i=1}^{r} \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

is, in the limit as $n \longrightarrow \infty$, distributed as a χ^2 with degrees of freedom $r-s-1$.

This theorem can also be reduced to Theorem 1. Among $r-1$ independent parameters $p_1, \ldots, p_{r-1}$, there are $r-s-1$ relations (functionally independent) and the assumptions of the theorem assure us that $r-s-1$ of them can be expressed as function of the other s parameters and these functions satisfy the conditions of Theorem 1.

We now state and sketch proofs of the analogous theorems in the analysis of variance situation.

<u>Theorem 1.*</u> Suppose that rs real-valued and one-valued functions $p_{ij}(\underline{\theta})$, $i=1,\ldots,r$; $j=1,\ldots,s$ of $\underline{\theta}$ are defined on a certain non-degenerate interval A of t-dimensional Euclidean space such that

(a) $\sum_{i=1}^{r} p_{ij}(\underline{\theta}) = p_{0j}(\underline{\theta}) = 1$ for $j=1,\ldots,s$ and all $\underline{\theta}$ in A,

(b) $p_{ij}(\underline{\theta}) \geq c^2 > 0$ for all i,j and $\underline{\theta}$ in A,

(c) every $p_{ij}(\underline{\theta})$ has continuous partial derivatives with respect to the component of $\underline{\theta}$ up to the second order, and

(d) the matrix $D(\underline{\theta})=(\frac{\partial p_{ij}}{\partial \theta_k})$ is of rank t for all $\underline{\theta}$ in A.

We shall consider s mutually independent random experiments $\mathcal{E}_1, \ldots, \mathcal{E}_s$, and in each experiment the possible outcomes are divided into r groups. The probability of getting a result which belongs to the ith group in the jth experiment $\mathcal{E}_j$ is $p^0_{ij} = p_{ij}(\underline{\theta}_0)$, where $\underline{\theta}_0$ is an inner point of A. Let the frequency of the result belonging to the ith group in n_{0j} repetitions of $\mathcal{E}_j$ be n_{ij}, so that

$$\sum_{i=1}^{r} n_{ij} = n_{0j}, \; j=1,\ldots,s \; . \quad \sum_{j=1}^{s} n_{0j} = n \; .$$

Then the modified χ^2-minimum equations

$$(3.12) \qquad \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{n_{ij}-n_{0j}p_{ij}(\underline{\theta})}{p_{ij}(\underline{\theta})} \frac{\partial \, p_{ij}(\underline{\theta})}{\partial \; \theta_k} = 0, k=1,2,\ldots,t \quad ,$$

have one and only one solution $\hat{\underline{\theta}}' = (\hat{\theta}_1 \ldots \hat{\theta}_t)$ which is consistent, and the statistic

$$(3.13) \qquad \chi^2 = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{(n_{ij}-n_{0j}p_{ij}(\hat{\underline{\theta}}))^2}{n_{0j}p_{ij}(\hat{\underline{\theta}})}$$

is, in the limit as $n \longrightarrow \infty$ subject to the ratios $r_j = n_{0j}/n$ being held fixed, distributed as a χ^2 with degrees of freedom $rs-s-t$.

<u>Proof of Theorem 1*</u>

Put

$$q_{ij}(\underline{\theta}) = r_j p_{ij}(\underline{\theta})$$

so that $\sum_{i=1}^{r} q_{ij}(\underline{\theta}) = r_j$ for all j; then the equation (3.12) can be written as

$$(3.12^*) \qquad \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{n_{ij} - nq_{ij}(\underline{\theta})}{q_{ij}(\underline{\theta})} \frac{\partial q_{ij}(\underline{\theta})}{\partial \theta_k} = 0. \; k = 1,2,\ldots,t \quad .$$

Put

$$\omega_k(\underline{\theta}) = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{n_{ij} - nq_{ij}^0}{n} \left[\frac{1}{q_{ij}} \frac{\partial q_{ij}}{\partial \theta_k} - \frac{1}{q_{ij}^0} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 \right]$$

$$- \sum_{j=1}^{s} \sum_{i=1}^{r} (q_{ij} - q_{ij}^0) \left[\frac{1}{q_{ij}} \frac{\partial q_{ij}}{\partial \theta_k} - \frac{1}{q_{ij}^0} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 \right]$$

$$- \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{1}{q_{ij}^0} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 \left[q_{ij} - q_{ij}^0 - \sum_{\ell=1}^{t} \left(\frac{\partial q_{ij}}{\partial \theta_\ell}\right)(\theta_\ell - \theta_\ell^0) \right]$$

then (3.12^*) can be rewritten as follows:

$$(3.15) \qquad \sum_{\ell=1}^{t} (\theta_\ell - \theta_\ell^0) \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{1}{q_{ij}^0} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 \left(\frac{\partial q_{ij}}{\partial \theta_\ell}\right)_0 = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{n_{ij} - n}{nq_{ij}^0} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 + \omega_k(\underline{\theta}) ,$$

$$k = 1,2,\ldots,t \; .$$

Let

$$\underline{x} = (x_{11} \ldots x_{r1} \ldots x_{1s} \ldots x_{rs}),$$

where

$$(3.16) \qquad x_{ij} = \frac{n_{ij} - n_{0j} p_{ij}^0}{\sqrt{n_{0j} p_{ij}^0}} = \frac{n_{ij} - nq_{ij}^0}{\sqrt{nq_{ij}^0}} \quad ,$$

and, furthermore, let

$$\underline{\omega}(\underline{\theta}) = (\omega_1(\underline{\theta}), \ldots, \omega_R(\underline{\theta}))$$

and

(3.17) $$B = \left(\frac{1}{\sqrt{q^0_{ij}}} \left(\frac{\partial q_{ij}}{\partial \theta_k}\right)_0 \right) \quad .$$

Using matrix notation, (3.15) can be expressed as

(3.18) $$B'B(\underline{\theta}-\underline{\theta}_0) = n^{-\frac{1}{2}} B'\underline{x} + \underline{\omega}(\underline{\theta}) \;;$$

hence we get

(3.19) $$\underline{\theta} - \underline{\theta}_0 = n^{-\frac{1}{2}}(B'B)^{-1}B'\underline{x} + (B'B)^{-1}\underline{\omega}(\underline{\theta}) \; .$$

For each experiment $\mathcal{E}_j$, Tchebyschev-Bienaymé inequality becomes

$$P\left[\left\{\omega ;\; |n_{ij}(\omega)-n_{0j}p^0_{ij}| \geq \lambda\sqrt{n_{0j}}\right\}\right] \leq \frac{p^0_{ij}(1-p^0_{ij})}{\lambda^2} < \frac{p^0_{ij}}{\lambda^2}$$

or

$$P\left[\left\{\omega ;\; |n_{ij}(\omega)-nq^0_{ij}| \geq \lambda\sqrt{n}\right\}\right] \leq \frac{q^0_{ij}}{\lambda^2 r_j} \quad .$$

Let $r = \min_j . \; r_j$; then

(3.20) $$P\left[\left\{\omega ;\; |n_{ij}(\omega)-nq^0_{ij}| \geq \lambda\sqrt{r_j}\sqrt{n}\right\}\right] < \frac{q^0_{ij}}{\lambda^2 r} \;, \text{ for all } j \quad .$$

Put $Q^{(0)}_{in} = \left\{\omega ;\; n_{ij}(\omega)-n_{0j}p^0_{ij}| < \lambda\sqrt{n_{0j}}\right\}$, $Q^{(j)}_n = \bigcap_i Q^{(j)}_{in}$, $Q_n = \bigcap_j Q^{(j)}_n$;

then we have

(3.21) $$P(Q_n) > 1 - \frac{s}{\lambda^2 r} \quad ,$$

and, for any $\quad \varepsilon\; Q_n$, we have

$$(3.22) \qquad |x_{ij}| = \frac{|n_{ij}-nq^0_{ij}|}{\sqrt{nq^0_{ij}}} = \frac{|n_{ij}-n_{0j}p^0_{ij}|}{\sqrt{nr_jp^0_{ij}}} < \frac{\lambda}{c} \quad .$$

As before, we can show that for any two points $\underline{\theta}_1$ and $\underline{\theta}_2$ in A

$$(3.23) \qquad |\underline{\omega}(\underline{\theta}_1) - \underline{\omega}(\underline{\theta}_2)| \leq K|\underline{\theta}_1-\underline{\theta}_2| \left\{ \frac{\lambda}{\sqrt{n}} + |\underline{\theta}_1-\underline{\theta}_0| + |\underline{\theta}_2-\underline{\theta}_0| \right\},$$

where K is a constant depending only on the functional form of p_{ij} and the closure of the interval A. Thus we can show that the equation (3.19) has one and only one consistent solution $\hat{\underline{\theta}}$.

Let $\underline{y} = (y_{11} \ldots y_{r1} \ldots y_{1s} \ldots y_{rs})$, where

$$(3.24) \qquad y_{ij} = \frac{n_{ij}-n_{0j}p_{ij}(\hat{\underline{\theta}})}{\sqrt{n_{0j}p_{ij}(\hat{\underline{\theta}})}}$$

$$= \frac{n_{ij}-nq_{ij}(\hat{\underline{\theta}})}{\sqrt{nq_{ij}(\hat{\underline{\theta}})}} \quad .$$

Then we obtain in exactly the same manner as before

$$(3.25) \qquad \underline{y} = \underline{x} - \sqrt{n}\, B(\hat{\underline{\theta}} - \underline{\theta}_0) + \frac{K'\lambda^2}{\sqrt{n}}\underline{(1)}$$

$$= (I-B(B'B)^{-1}B')\underline{x} + \frac{K\lambda^2}{\sqrt{n}}\underline{(1)} \quad .$$

The limiting distribution of $\chi^2 = \underline{y}'\underline{y}$, as $n \longrightarrow \infty$, is the same as that of

$$(3.26) \qquad \chi^2_* = \underline{x}'(I-B(B'B)^{-1}B')\,\underline{x} \quad .$$

The limiting distribution of $\underline{x}$ as $n \longrightarrow \infty$ is a rs-dimensional normal distribution with mean $\underline{0}$ and the covariance matrix

$$(3.27)\quad \Lambda = \begin{pmatrix} I - \frac{1}{r_1}\underline{q}_1\underline{q}_1' & & 0 \\ & I - \frac{1}{r_2}\underline{q}_2\underline{q}_2' & \\ 0 & & I - \frac{1}{r_s}\underline{q}_s\underline{q}_s' \end{pmatrix} ,$$

where

$$\underline{q}_j' = (q_{1j} q_{2j} \ldots q_{rj}),\ j = 1,2,\ldots,s \ .$$

We shall apply the orthogonal transformation

$$(3.28)\qquad \underline{\xi} = K'\underline{x} \quad ,$$

where K, defined analogously to K in the proof of Theorem 1, is given by

$$(3.29)\qquad K = (BCM^{-1}\underline{b}_{t+1} \ldots \underline{b}_{rs}) \ .$$

Then

$$(3.30)\qquad \chi^2_{*} = \xi^2_{t+1} + \ldots + \xi^2_{rs} \ .$$

The covariance matrix of $\underline{\xi}$ is

$$(3.31)\qquad K'\Lambda K = I - \sum_{j=1}^{s} \frac{1}{r_j} K'\underline{q}_j^{*}\underline{q}_j^{*'}K \quad ,$$

where $\underline{q}_j^{*'} = (0 \ldots 0 \sqrt{q^0_{1j}} \ \ldots \sqrt{q^0_{rj}} \ 0 \ldots 0)$, $j = 1, \ldots, s$.

Since

$$\underline{q}_j^{*'}K = (\overbrace{0 \ldots 0}^{t}\ e^{(j)}_{t+1} \ldots e^{(j)}_{rs}),$$

where

$$(3.32)\qquad e^{(j)}_h = \underline{q}_j^{*'}\underline{b}_h \ ,\ h = t+1, \ldots, rs \quad ,$$

we have

$$(3.33) \qquad 1-K'\Lambda K = \begin{bmatrix} I(t) & 0 \\ 0 & I(rs-t)-Q \end{bmatrix} ,$$

where

$$Q = (q_{gh}),\ q_{gh} = \sum_{j=1}^{s} q_{gh}^{(j)}\ ,\ q_{gh}^{(j)} = \frac{1}{r_j}\, e_g^{(j)} e_h^{(j)} .$$

On the other hand, we have mutually orthogonal s relations, i.e.,

$$\underline{q}_j^{*\prime}\,\underline{x} = 0\ ,\quad j = 1, \ldots, s \quad ,$$

or

$$(3.34) \qquad e_{t+1}^{(j)}\,\xi_{t+1} + \ldots + e_{rs}^{(j)}\,\xi_{rs} = 0,\ j=1,\ldots,s .$$

Let

$$\underline{e}^{(j)'} = (e_{t+1}^{(j)} \ldots e_{rs}^{(j)}) ;$$

then

$$(3.35) \qquad Q = \frac{1}{r_1}\,\underline{e}^{(1)}\underline{e}^{(1)'} + \ldots + \frac{1}{r_s}\,\underline{e}^{(s)}\underline{e}^{(s)'} .$$

We shall apply the orthogonal linear transformation

$$(3.36) \qquad \underline{\eta}^{*} = P\,\underline{\xi}^{*} \quad ,$$

where the first s rows of P are

$$\underline{e}^{(1)'}\Big/\left|\underline{e}^{(1)}\right| \quad , \ldots ,\ \underline{e}^{(s)'}\Big/\left|\underline{e}^{(s)}\right| \ ;$$

then we have

$$(3.37) \qquad PQP' = \sum_{j=1}^{s} \frac{1}{r_j}\, P\underline{e}^{(j)}\,.\,\underline{e}^{(j)'}P' = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} .$$

Thus we have shown that

$$(3.38) \qquad \chi_*^2 = \eta_{t+s+1}^2 + \ldots + \eta_{rs}^2 ,$$

and the covariance matrix of $\eta_{t+1}, \ldots, \eta_{rs}$ is

$$(3.39) \qquad I(rs-t)-PQP' = \begin{bmatrix} 0 & 0 \\ 0 & I(rs-t-s) \end{bmatrix} .$$

<u>Theorem 2*</u>. (S.N.Roy) Under the same conditions as Theorem 1*, suppose we have u $(< t)$ relations:

$$(3.40) \qquad H_0: \; f_k(\underline{\theta}) = f_k^0, \quad k = 1, 2, \ldots, u ,$$

such that

(e) t functions $f_k(\underline{\theta})$, $k = 1, \ldots, u$ possess continuous partial derivatives up to the second order, and

(f) the matrix $\left(\dfrac{\partial f_i}{\partial \theta_k}\right)$ is of rank u for all $\underline{\theta}$ in A.

Then the modified χ^2-minimum equations subject to H_0 has one and only one consistent solution $\hat{\underline{\theta}}$, and the statistic

$$(3.41) \qquad \chi_0^2 = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{(n_{ij}-n_{ij}p_{ij}(\hat{\underline{\theta}}))^2}{n_{0j}p_{ij}(\hat{\underline{\theta}})}$$

is, in the limit as $n \longrightarrow \infty$ subject to the ratios $r_j = n_{0j}/n$ being held fixed, distributed as a χ^2 with degrees of freedom rs-s-t+u.

And if the hypothesis H_0 is true, then

$$\chi_0^2 - \chi^2$$

is, in the limit as $n \longrightarrow \infty$ subject to $r_j = n_{0j}/n$ being held fixed, independent of χ^2 and is distributed as a χ^2 with degrees of freedom u.

<u>Proof of Theorem 2*</u>

Without any loss of generality, we can assume that

$$(3.42) \qquad \left|\begin{matrix} \frac{\partial f_1}{\partial \theta_1} & \cdots & \frac{\partial f_1}{\partial \theta_u} \\ \cdot & \cdots & \cdot \\ \frac{\partial f_u}{\partial \theta_1} & \cdots & \frac{\partial f_u}{\partial \theta_u} \end{matrix}\right|_{\underline{\theta} = \underline{\theta}_0} \neq 0 \quad ,$$

and the determinant of the left-hand side of (3.42) is a continuous function of $\underline{\theta}$; therefore (3.42) holds true in some neighborhood of $\underline{\theta}_0$. Let a non-degenerate interval A_0 of which θ_0 is an inner point be contained in that neighborhood of θ_0. We can proceed by taking A_0 as a domain of $\underline{\theta}$, instead of A. Thus we may assume that (3.42) holds for all $\underline{\theta}$ in A.

Consider the transformation

$$(3.43) \qquad f_1(\underline{\theta}) = \xi_1, \ldots, f_u(\underline{\theta}) = \xi_u, \theta_{u+1} = \xi_{u+1} \cdots \theta_t = \xi_t;$$

then by a well-known theorem on implicit functions, we have

$$(3.44) \qquad \theta_1 = \phi_1(\underline{\xi}), \ldots, \theta_u = \phi_u(\underline{\xi}), \theta_{u+1} = \xi_{u+1}, \ldots, \theta_t = \xi_t,$$

where the functions $\phi(\underline{\xi})$'s are continuous and possess continuous partial derivatives up to the second order such that

$$(3.45) \qquad \theta_i^0 = \phi_i(f_1^0 \ldots f_u^0, \theta_{u+1}^0 \ldots \theta_t^0), i=1,2,\ldots,t \quad .$$

Substituting (3.44) into $p_{ij}(\underline{\theta})$, we get

$$(3.46) \qquad p_{ij}(\underline{\theta}) = q_{ij}(\underline{\xi})$$

such that functions $q_{ij}(\underline{\xi})$ are continuous and possess continuous partial derivatives up to the second order.

Hence under the null-hypothesis H_0, rs functions $p_{ij}(\underline{\theta})$ becomes functions $q_{ij}(\underline{\xi}^*)$, where $\underline{\xi}^{*\prime} = (\xi_{u+1}, \dots, \xi_t)$; and these functions $q_{ij}(\underline{\xi}^*)$ satisfy the conditions (a), (b), (c) and (d). We shall show that only (d) is satisfied; indeed, since,

$$(3.47) \qquad \frac{\partial q_{ij}}{\partial \xi_h} = \frac{\partial p_{ij}}{\partial \theta_1}\frac{\partial \theta_1}{\partial \xi_h} + \dots + \frac{\partial p_{ij}}{\partial \theta_u}\frac{\partial \theta_u}{\partial \xi_h} + \frac{\partial p_{ij}}{\partial \theta_h}, \; h = u+1, \dots, t \; ,$$

if the rank of the matrix $\left(\frac{\partial q_{ij}}{\partial \xi_h}\right)$ is less than t-u at some point $\underline{\xi}^*$ in A^* (which is the projection of A into the sub-space spanned by the last t-u coordinate axes), then we can find

$$\lambda_{u+1}(\underline{\xi}^*), \dots, \lambda_t(\underline{\xi}^*)$$

such that

$$0 = \sum_{h=u+1}^{t} \lambda_h(\underline{\xi}^*)\frac{\partial q_{ij}}{\partial \xi_u} = \frac{\partial p_{ij}}{\partial \theta_1}\sum_{n} \lambda_h(\underline{\xi}^*)\frac{\partial \theta_1}{\partial \xi_h} + \dots + \frac{\partial p_{ij}}{\partial \theta_u}\sum_{h} \lambda_h(\underline{\xi}^*)\frac{\partial \theta_u}{\partial \xi_h}$$

$$+ \sum_{h} \lambda_h(\underline{\xi}^*)\frac{\partial p_{ij}}{\partial \theta_h} \quad .$$

On the other hand, we have assumed that the matrix $\left(\frac{\partial p_{ij}}{\partial \theta_k}\right)$ is of rank t for all $\underline{\theta}$ in 4. Thus we get

$$\lambda_h(\underline{\xi}^*) = 0, \; h = u+1, \dots, t \quad .$$

Consequently the problems are reduced to the case of Theorem 1^* with independent t-u parameters $\underline{\xi}^*$.

Theorem 3^* (S.N.Roy). Suppose that we have s mutually independent random experiments $\mathcal{E}_1, \dots, \mathcal{E}_s$, and the possible outcomes in each experiment are divided into r exclusive groups. Let the probability of getting a result belonging to the ith class in $\mathcal{E}_j$ be

$$p^0_{ij} \geq c^2 > 0 \quad .$$

Let n_{ij} denote the number of results belong to the ith class in n_{0j} repetitions of $\mathcal{E}_j$, so that $\sum_{i=1}^{r} n_{ij}=n_{0j}$, and let

$$\sum_{j=1} n_{0j}=n \quad .$$

Let us consider the hypothesis

$$H_0: \quad f_k(\underline{p}) = 0, \; k=1, \ldots, rs-s-t$$

where f_k are continuous and possess continuous partial derivatives up to the second order in the parameter space

$$\pi = \left\{ \underline{p};\; 0 < c^2 \leq p_{ij} \leq 1, \; \sum_i p_{ij}=p_{0j}=1, j=1,\ldots,s \right\}$$

of which $\underline{p}^0$ is an inner point and the matrix $\left(\frac{\partial f_k}{\partial p_{ij}}\right)$ is of rank rs-s-t.

Then the modified χ^2 minimum equations subject to H_0 have one and only one consistent solution $\hat{\underline{p}}$, and the statistic

$$\chi^2 = \sum_j \sum_i \frac{(n_{ij}-n_{0j}\hat{p}_{ij})^2}{n_{0j}\hat{p}_{ij}}$$

is, in the limit as $n \longrightarrow \infty$ subject to all ratios r_j being held fixed, distributed as a χ^2 with degrees of freedom

$$rs-s-t \quad .$$

This theorem can also be reduced to Theorem 1*.

4. <u>On the distribution of a modified χ^2 statistic</u>.

In testing goodness of fit by the χ^2 method, there are many cases in which we already have sample values (not the frequencies). For simplicity of discussion, we shall mention the simplest situation, because in this way the statistical feature of the problem will become clearer. For the multiparameter case the author would like to refer to Herman Chernoff and E. L. Lehmann [1].

Suppose that we have random samples $x_1, x_2, \ldots, x_n$ from a certain population having an absolutely continuous distribution function $F(x;\alpha)$ and a probability density function $f(x;\alpha)$ which satisfies the following regularity conditions: (see Cramér [4])

1) For almost all x, the derivatives $\frac{\partial \log f}{\partial \alpha}$, $\frac{\partial^2 \log f}{\partial \alpha^2}$ and $\frac{\partial^3 \log f}{\partial \alpha^3}$ exist for every α belonging to a non-degenerate interval A.

2) For every α in A, we have $\frac{\partial f}{\partial \alpha} < F_1(x)$, $\frac{\partial^2 f}{\partial \alpha^2} < F_2(x)$ and $\frac{\partial^3 \log f}{\partial \alpha^3} < H(x)$, the functions $F_1(x)$ and $F_2(x)$ being integrable over $(-\infty, +\infty)$, while

$$\int_{-\infty}^{\infty} H(x) f(x;\alpha)\, dx < M,$$

where M is independent of α.

3) For every α in A, the integral $\int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \alpha}\right)^2 f dx$ is finite and positive. Furthermore, we assume that the true value of the unknown parameter is α_0.

Let the whole interval of x be divided into r classes as follows:

$$a_0 = -\infty < a_1 < a_2 < \ldots < a_{r-1} < a_r = +\infty \;;$$

and let the class frequencies be $\nu_1, \ldots, \nu_r$, so that $\sum_1^r \nu_i = n$. In other words, if we define auxiliary random variables as follows:

(4.1) $$x'_{ij} = \begin{cases} 1 & \text{if } x_j \leq a_i \\ 0 & \text{otherwise} \end{cases}$$

and

(4.2) $$\nu(a_i) = \sum_{j=1}^{n} x'_{ij}, \quad i = 1, 2, \ldots, r \quad ,$$

then we get

(4.3) $$\nu_i = \nu(a_i) - \nu(a_{i-1}), \; i=1,2,\ldots,r \; .$$

Further let

(4.4) $$\pi_i^0 = F(a_i;\alpha_0), \; i = 1, 2, \ldots, r \quad ,$$

where, of course, we have put $\pi_0 = F(-\infty;\ \alpha_0)=0$, $\pi_r=F(\infty;\alpha_0)=1$, and

(4.5) $$p_i^0 = \pi_i^0 - \pi_{i-1}^0, \; i = 1,2,\ldots,r \quad .$$

Then it is a well-known fact that the statistic

(4.6) $$\sum_{i=1}^{r} \frac{(\nu_i - np_i^0)^2}{np_i^0}$$

is, in the limit as $n \to \infty$, distributed as χ^2 with degrees of freedom r-1. If $\hat{\alpha}$ is that modified χ^2-minimum estimate of α, in Cramér's sense, from the frequency data which satisfies the condition of consistency, then Theorem 1 in § 1 tells us that the statistic

$$\sum_{i=1}^{r} \frac{(\nu_i - np_i(\hat{\alpha}))^2}{np_i(\hat{\alpha})}$$

is, in the limit as $n \to \infty$, distributed as χ^2 with degrees of freedom r-2.

Let α^* be the maximum likelihood estimate from the original sample in Cramér's sense, i.e., a solution of the maximum likelihood equation

$$\frac{\partial \log L}{\partial \alpha} = 0$$

which is a consistent estimator of α, then it was shown by Cramér [4] that

$$\sqrt{n}(\alpha^*-\alpha_0)=\frac{1}{k^2\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial \log f(x;\alpha)}{\partial \alpha}\right)_{\alpha_0} + O(1) \text{ in probability ,} \tag{4.7}$$

where

$$k^2 = \mathscr{E}_{\alpha_0}\left\{ \left(\frac{\partial \log f(x;\alpha)}{\partial \alpha}\right)^2_{\alpha_0}\right\} . \tag{4.8}$$

Now let us consider the limiting distribution of the statistic

$$\chi^{*2} = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\alpha^*))^2}{np_i(\alpha^*)} , \tag{4.9}$$

in the limit as $n \to \infty$.

Since

$$\pi_i(\alpha^*)=F(a_i;\alpha^*)=F(a_i;\alpha_0)+(\alpha^*-\alpha_0)\frac{\partial F}{\partial \alpha}(a_i;\alpha_0)+O(\frac{1}{\sqrt{n}}) \text{ in probability} \tag{4.10}$$

$$= \pi_i^0 + (\alpha^*-\alpha_0)\left(\frac{\partial \pi_i}{\partial \alpha_0}\right) + O(\frac{1}{\sqrt{n}}) \text{ in probability ,}$$

it follows that

$$\sqrt{n}\left(\frac{\nu(a_i)}{n}-F(a_i;\alpha^*)\right)=\sqrt{n}\left(\frac{\nu(a_i)}{n} - \pi_i^0 - (\alpha^*-\alpha_0)\left(\frac{\partial \pi_i}{\partial \alpha}\right)_0\right)+ O(1) \text{ in prob.}$$

$$=\sqrt{n}\left(\frac{\nu(a_i)}{n} - \pi_i^0\right) - \frac{\left(\frac{\partial \pi_i}{\partial \alpha}\right)_0}{k^2\sqrt{n}} \sum_{j=1}^{n} \frac{\partial \log f(x_j;\alpha_0)}{\partial \alpha} + O(1)$$

in probability.

By using (4.2) this can be rewritten as follows:

$$(4.11)\quad \sqrt{n}\left(\frac{\nu(a_i)}{n}-F(a_i;\alpha^*)\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(x'_{ij}-\pi_i^0-\frac{\left(\frac{\partial \pi_i}{\partial \alpha}\right)_0}{k^2}\frac{\partial \log f(x_j;\alpha_0)}{\partial \alpha}\right)+o(1)$$

in probability.

Now it can easily be shown under the regularity conditions 1) - 3) that

$$(4.12)\qquad \mathcal{E}_{\alpha_0}x'_{ij}=\int_{-\infty}^{a_i}dF(x;\alpha_0)=\pi_i^0,\quad \mathcal{E}_{\alpha_0}\left(\frac{\partial \log f(x;\alpha_0)}{\partial \alpha}\right)=0 \quad ,$$

and after some calculations the covariance σ_{ij} between two variables

$$x'_{i\ell}-\pi_i^0-\frac{1}{k^2}\frac{\partial \pi_i^0}{\partial \alpha}\frac{\partial \log f(x_\ell;\alpha_0)}{\partial \alpha} \quad \text{and} \quad \pi'_{j\ell'}-\pi_j^0-\frac{1}{k^2}\frac{\partial \pi_j^0}{\partial \alpha}\frac{\partial \log f(x_{\ell'};\alpha_0)}{\partial \alpha}$$

comes out to be

$$(4.13)\qquad \sigma_{ij}=\delta_{\ell\ell'}\left(\pi_i^0(1-\pi_j^0)-\frac{1}{k^2}\frac{\partial \pi_i^0}{\partial \alpha}\frac{\partial \pi_j^0}{\partial \alpha}\right) \quad ,$$

where $\delta_{\ell\ell'}=1$ if $\ell=\ell'$, and $=0$ otherwise.

By using the Central Limit Theorem, we can see that the limiting distribution of

$$(4.14)\quad \left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(x'_{1j}-\pi_1^0-\frac{1}{k^2}\frac{\partial \pi_1^0}{\partial \alpha}\frac{\partial \log f(x_j;\alpha_0)}{\partial \alpha}\right),\ldots,\right.$$

$$\left.\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(x'_{rj}-\pi_r^0-\frac{1}{k^2}\frac{\partial \pi_r^0}{\partial \alpha}\frac{\partial \log f(x_j;\alpha_0)}{\partial \alpha}\right)\right)$$

is the r-dimensional normal distribution with mean $\underline{0}'$ and the covariance matrix

$$(4.15)\qquad \Lambda=\left(\pi_i^0(1-\pi_j^0)-\frac{1}{k^2}\frac{\partial \pi_i^0}{\partial \alpha}\frac{\partial \pi_j^0}{\partial \alpha}\right)_{i,j=1,\ldots,r} \quad .$$

Therefore, the limiting distribution of the random vector

(4.16) $$\left(\sqrt{n}\left(\frac{\nu(a_1)}{n} - F(a_1;\alpha^*)\right), \ldots, \sqrt{n}\left(\frac{\nu(a_r)}{n} - F(a_r;\alpha^*)\right) \right),$$

as $n \longrightarrow \infty$, is also the same r-dimensional normal distribution.

Now, since

(4.17) $$\sqrt{n}\left(\frac{\nu_i}{n} - p_i(\alpha^*)\right) = \sqrt{n}\left(\frac{\nu(a_i)}{n} - F(a_i;\alpha^*)\right) - \sqrt{n}\left(\frac{\nu(a_{i-1})}{n} - F(a_{i-1};\alpha^*)\right),$$

$$i = 1,2,\ldots,r$$

we can conclude that they are distributed, in the limit as $n \longrightarrow \infty$, as the r-dimensional normal distribution with mean $\underline{0}'$ and the covariance matrix

(4.18) $$\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} \Lambda \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

$$= \left(p_i^0\delta_{ij} - p_i^0 p_j^0 - \frac{1}{k^2}\frac{\partial p_i^0}{\partial \alpha}\frac{\partial p_j^0}{\partial \alpha}\right)_{ij=1,2,\ldots,r}$$

On the other hand, we get

$$\frac{\nu_i - np_i(\alpha^*)}{\sqrt{np_i(\alpha^*)}} = \frac{\nu_i - np_i(\alpha^*)}{\sqrt{np_i^0}} + o(1) \text{ in prob, as } n \longrightarrow \infty;$$

whence we see that the vector

$$\left(\frac{\nu_1 - np_1(\alpha^*)}{\sqrt{np_1(\alpha^*)}}, \ldots, \frac{\nu_r - np_r(\alpha^*)}{\sqrt{np_r(\alpha^*)}} \right)$$

is, in the limit as $n \longrightarrow \infty$, distributed as the r-dimensional normal distribution with the mean $\underline{0}'$ and the covariance matrix

$$\left(\delta_{ij}-\sqrt{p_i^0 p_j^0}-\frac{1}{k^2\sqrt{p_i^0 p_j^0}}\frac{\partial p_i^0}{\partial \alpha}\frac{\partial p_j^0}{\partial \alpha}\right)ij = 1,\ldots,r$$

or

(4.19) $$= I - \underline{pp}' - \underline{hh}' \quad ,$$

where

(4.20) $$\underline{p} = \begin{bmatrix} \sqrt{p_1^0} \\ \cdot \\ \cdot \\ \cdot \\ \sqrt{p_r^0} \end{bmatrix} \quad \text{and } \underline{h} = \frac{1}{k}\begin{bmatrix} \frac{1}{\sqrt{p_1^0}}\frac{\partial p_1^0}{\partial \alpha} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{\sqrt{p_r^0}}\frac{\partial p_r^0}{\partial \alpha} \end{bmatrix} \quad .$$

Now

$$\underline{h}'\underline{p} = \underline{p}'\underline{h} = \frac{1}{k}\sum_{i=1}^{r}\frac{\partial p_i^0}{\partial \alpha} = 0$$

(4.21) $$\underline{h}'\underline{h} = \frac{1}{k^2}\sum_{i=1}^{r}\frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 \equiv c^2, \text{ say} \quad .$$

It will easily be seen by Schwarz's inequality that

$$\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 = \left(\int_{a_{i-1}}^{a_i}\frac{\partial f(x;\alpha_0)}{\partial \alpha}dx\right)^2 \leq \int_{a_{i-1}}^{a_i}\frac{1}{f(x;\alpha_0)}\left(\frac{\partial f(x;\alpha_0)}{\partial \alpha}\right)^2 dx \cdot \int_{a_{i-1}}^{a_i} f(x;\alpha_0)dx,$$

i.e.

(4.22) $$\frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 \leq \int_{a_{i-1}}^{a_i}\frac{1}{f(x;\alpha_0)}\left(\frac{\partial f(x;\alpha_0)}{\partial \alpha}\right)^2 dx \quad .$$

Summing up (4.22) with respect to i, we get

$$\sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 \leq \int_{-\infty}^{\infty} \frac{1}{f(x;\alpha_0)}\left(\frac{\partial f(x;\alpha_0)}{\partial \alpha}\right)^2 dx = k^2 ,$$

and consequently we obtain

$$(4.23) \qquad 0 \leq c^2 \leq \sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 \Big/ k^2 \leq 1 .$$

This quantity c^2, which had already appeared in author's 1952 paper [9], should be called the relative efficiency of the grouped data. In fact, the amount of information of the original data at $\alpha = \alpha_0$ is given by

$$(4.24) \qquad \mathcal{E}_{\alpha_0}\left(\frac{\partial \log L}{\partial \alpha}\right)^2_{\alpha_0} = n\, \mathcal{E}_{\alpha_0}\left(\frac{\partial \log f(x;\alpha_0)}{\partial \alpha}\right)^2 = nk^2 ,$$

where $L = \prod_{i=1}^{n} f(x_i;\alpha)$, while the amount of information of the grouped data at $\alpha = \alpha_0$ is given by

$$(4.25) \qquad \mathcal{E}_{\alpha_0}\left(\frac{\partial \log L^{(f)}}{\partial \alpha}\right)^2_{\alpha_0} = \mathcal{E}_{\alpha_0}\left(\sum_{i=1}^{r} \nu_i \frac{\partial \log p_i^0}{\partial \alpha}\right)^2 = n\sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 ,$$

where

$$L^{(f)} = \frac{n!}{\prod_{i=1}^{r} \nu_i!} \prod_{i=1}^{r} p_i^{\nu_i} .$$

Therefore the relative efficiency of the grouped data compared to the original date is given by

$$(4.26) \qquad \mathcal{E}_{\alpha_0}\left(\frac{\partial \log L^{(f)}}{\partial \alpha}\right)^2_{\alpha_0} \Big/ \mathcal{E}_{\alpha}\left(\frac{\partial \log L}{\partial \alpha}\right)^2_{\alpha_0} = \frac{1}{k^2}\sum_{i=1}^{r} \frac{1}{p_i^0}\left(\frac{\partial p_i^0}{\partial \alpha}\right)^2 = c^2 .$$

If we choose r-2 columns appropriately such that the matrix

$$K = (\overbrace{\cdots}^{r-2} \ \frac{\underline{h}}{c} , \underline{p})$$

is an orthogonal matrix, then it is seen that

$$(4.27) \quad K'(I-\underline{pp}'-\underline{hh}')K = I-K'\underline{pp}'K-K'\underline{hh}'K$$

$$= \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 1 \end{bmatrix} - \begin{bmatrix} 0 & & & 0 \\ & \ddots & & \\ & & 0 & \\ 0 & & & 1 \end{bmatrix} - \begin{bmatrix} 0 & & & 0 \\ & \ddots & & \\ & & c^2 & \\ 0 & & & 0 \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ & & & 1-c^2 \\ 0 & & & 0 \end{bmatrix} .$$

This means that the statistic χ^{*2} is, in the limit as $n \longrightarrow \infty$, distributed as

$$\chi^2_{r-2} + (1-c^2)\chi^2_1 .$$

where χ^2_{r-2} and χ^2_1 are mutually independent χ^2-variates with degrees of freedom r-2 and 1 respectively. This result shows, so to speak, that, if we use the estimator α^* instead of $\hat{\alpha}$, then only the c^2 fraction of one degree of freedom is reduced.

REFERENCES

[1] Bernard, G.A.,"Significance tests for 2 x 2 tables",Biom.,Vol. 34(1947), pp.123-138.

[2] Chernoff, H. and Lehmann, E.L., "The use of the maximum likelihood estimates in χ^2 tests for goodness of fit", Ann. Math. Stat., Vol. 25(1954),pp.579-586.

[3] Cochran,W.G., "The χ^2 tests of goodness of fit", Ann.Math.Stat., Vol. 23(1952), pp. 314-345.

[4] Cramér, H., Mathematical Methods of Statistics. Princeton Univ. Press, 1946.

[5] Fisher, R.A., "On the interpretation of chi-square from contingency tables and the calculation of p", Journ. of R.S S., Vol. 455 (1922), pp. 87-94.

[6] Hotelling, H., "The consistency and ultimate distribution of optimum statistics", Trans. Amer. Math. Soc.,Vol 32(1930), pp. 847-859.

[7] Mitra, S.K., "Contribution to the statistical analysis of categorical data", North Carolina Institute of Stat. Memeo. Series 142, 1955.

[8] Ogawa, J., "A limit theorem of Cramér, its generalizations and an application". Unpublished.

[9] Ogawa, J., "Contributions to the theory of systematic statistics I", Osaka Math. Journ.,Vol. 3(1952).

[10] Ogawa, J., "On the sampling distributions of classical statistics in multivariate analysis", Osaka Math. Journ., Vol 5(1954).

[11] Roy, S.N. and Mitra, S.K., "An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis", Biom.,Vol.43(1956),pp. 361-376.

[12] Wald,A., "Note on the consistency of the maximum likelihood estimate", Ann. Math. Stat., Vol. 20(1949), pp. 595-601.

[13] Pearson, E.S., "The choice of statistical tests illustrated on the interpretation of data classed in a 2 x 2 table", Biometrika, Vol.34(1947), pp. 139-167.