A STUDY OF LARGE SAMPLE TEST CRITERIA THROUGH PROPERTIES OF EFFICIENT ESTIMATES

PART I : TESTS FOR GOODNESS OF FIT AND CONTIGENCY TABLES

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. The existence of m.l. and m.l. equation estimates and their efficiency, in the case of sampling from a finite multinomial distribution, have been established under conditions weaker than those assumed by earlier writers. For instance, the existence of the second derivatives of the cell probabilities as functions of the parameters is not necessary. The asymptotic distributions of the chi-square goodness of fit test and test criteria for examining composite hypotheses in contingency tables have been shown to depend mainly on the efficiency (in the new sense) of estimates used in the construction of the test criteria. The m.l. estimate satisfies the efficiency condition and is, in the sense of second order efficiency, more suitable for the construction of test criteria than other efficient estimates. A new test has been proposed for examining the expected frequencies computed under two different hypotheses of a special nature.

1. Introduction

In two previous papers (Rao, 1960a, 1960b), the author gave a new formulation of the concept of asymptotic efficiency of estimates of parameters and established an optimum property (second order efficiency) of the maximum likelihood (m.l.) estimate. This shows that although the class of estimation procedures providing efficient estimates is very wide, the m.l. method is distinguishable from the rest by its maximum second order efficiency. It has been observed that m.l. estimates (which are efficient in the new sense) provide a good summary of data in large samples in the sense that inference based on these estimates is equivalent to that obtained by utilizing the whole data. The approach to the problem of estimation and the specific results obtained have been based on the concepts introduced by Fisher (1921, 1925).

It is also well known that in the construction of large sample test criteria efficient estimates play an important role. For instance, in obtaining a χ^2 goodness of fit, the expected values are calculated by inserting efficient estimates such as those obtained by maximising likelihood or minimising χ^2. The object of the present series of articles is to show how the concept of efficiency in the new sense plays a key role in the construction of large sample criteria and in the derivation of their asymptotic distributions. It is also suggested that preference should be given to m.l. estimates, because of its maximum second order efficiency, among all efficient estimates in the construction of large sample criteria.

Incidentally, the existence of the m.l. estimates, their consistency, efficiency and asymptotic normality of distribution have all been deduced, in the case of the finite multinomial under conditions weaker than those assumed by Cramér (1946);

4

for instance, without assuming second order differentiability of the cell probabilities as functions of unknown parameters. The χ^2 goodness of fit criterion is shown to have the usual asymptotic distribution even under these weaker conditions.

2. Definitions and assumptions

Let $\pi_1, \ldots, \pi_k$, the probabilities in the k cells of a multinomial distribution, depend on a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ of parameters belonging to a q dimensional real space R^q or to a nondegenerate interval of R^q. If $p_1, \ldots, p_k$ are the observed proportions in the k cells for a sample of size n, the log likelihood of $\boldsymbol{\theta}$ is proportional to

$$p_1 \log \pi_1 + \ldots + p_k \log \pi_k.$$

When π_i are differentiable functions of $\boldsymbol{\theta}$, the maximum likelihood (m.l.) equations are defined by

$$z_r = \Sigma \frac{p_i}{\pi_i} \frac{\partial \pi_i}{\partial \theta_r} = 0, \qquad r = 1, \ldots, q. \qquad \ldots (2.1)$$

The information matrix, which is the variance covariance matrix of z_r, is denoted by $\Lambda = BB' = (i_{rs})$, where the matrix B and i_{rs} are defined as follows:

$$B = \left(\frac{1}{\sqrt{\pi_i}} \frac{\partial \pi_i}{\partial \theta_r} \right), \quad i_{rs} = \Sigma \frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \theta_r} \frac{\partial \pi_i}{\partial \theta_s}. \qquad \ldots (2.2)$$

Unless otherwise stated, the variables z_i and the elements i_{rs} denote the values at $\boldsymbol{\theta}_0 = (\theta_1^0, \ldots, \theta_q^0)$, the true value of the parameter supposed to be an interior point of the admissible region of $\boldsymbol{\theta}$, such that $\pi_i(\boldsymbol{\theta}_0) > 0$ for each i.

We introduce the following assumptions.

Assumption 1 : Every π_i has continuous first derivatives $\partial \pi_i / \partial \theta_s$ at the true value $\boldsymbol{\theta}_0$.

Assumption 2a : Given a $\delta > 0$, it is possible to find an $\epsilon > 0$, such that

$$\inf_{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 \geqslant \delta} \Sigma\, \pi_i(\boldsymbol{\theta}_0) \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})} \geqslant \epsilon. \qquad \ldots (2.3)$$

Assumption 2b : $\pi_i(\boldsymbol{\theta}) \neq \pi_i(\boldsymbol{\beta})$ for at least one i, when $\boldsymbol{\theta} \neq \boldsymbol{\beta}$, which is the identifiability condition.

It is easy to see that Assumption 2b $\Longrightarrow$ Assumption 2a, when the admissible interval of $\boldsymbol{\theta}$ is closed and $\pi_i(\boldsymbol{\theta})$ are continuous functions of $\boldsymbol{\theta}$. If the interval is open, it may happen that

$$|\pi_i(\boldsymbol{\theta}_n) - \pi_i(\boldsymbol{\theta}_0)| \to 0, \text{ for each } i \text{ as } \boldsymbol{\theta}_n \to \boldsymbol{\beta} \neq \boldsymbol{\theta}_0$$

Assumption 2a prevents such a thing happening and is in a certain sense a strong identifiability condition.

Assumption 3: There exists an estimate $\dot{\boldsymbol{\theta}} = (\dot{\theta}_1, \ldots, \dot{\theta}_q)$ such that

i) plim $\dot{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$

ii) $\sqrt{n}d_i = \sqrt{n}(\dot{\theta}_i - \theta_i^0)$ has an asymptotic distribution (a.d.).

iii) The estimate $\dot{\boldsymbol{\theta}}$ is efficient in the new sense (Rao, 1960a, 1960b), i.e.,

$$\text{plim } [\sqrt{n}(z_r - i_{r1}d_1 - \ldots - i_{rq}d_q)] = 0, \quad r = 1, \ldots, q \qquad \ldots \quad (2.4)$$

or in matrix notation $$\text{plim } (Z - \Lambda D) = 0$$

where Z and D denote the column vectors of the elements $\sqrt{n}\, z_r$ and $\sqrt{n}\, d_s$ respectively. We may, in theory, allow $\dot{\boldsymbol{\theta}}$, itself to depend on the unknown true value $\boldsymbol{\theta}_0$, but it is necessary for the application of the results derived that $\dot{\pi}_i = \pi_i(\dot{\boldsymbol{\theta}})$, the estimate of $\pi_i(\boldsymbol{\theta})$ should be independent of the assumed true value.

Assumption 3 may appear as a blanket assumption but is deliberately introduced to demonstrate the key role played by the efficiency condition (2.4). Whatever may be the method by which $\dot{\boldsymbol{\theta}}$ is obtained the propositions considered in the paper are valid provided only it satisfies the Assumption 3.

We will, in fact, show in Section 3.5 that Assumptions 1 and 2b imply Assumption 3 when the rank of Λ is q, and Assumptions 1 and 2a imply the stronger result that m.l. estimates exist and satisfy the conditions of Assumption 3 when the rank of Λ is q. The conditions under which the results relating to existence of m.l. estimates, their consistency, efficiency (implying asymptotic normality of distribution) are obtained are much weaker than those considered by earlier writers. It can be proved that under the same or similar less restrictive conditions other methods of estimation such as minimum χ^2 (Neyman, 1949; Rao, 1955a), modified minimum χ^2 (Neyman, 1949), minimum discrepancy (Haldane, 1951), minimum Hellinger distance (Rao, 1960b), also provide estimates satisfying the Assumption 3.

Situations, however, exist where the Assumption 3 is satisfied without even the rank of Λ being q, so that the treatment of the paper is, in some respects, more general.

We have not made any assumption about the rank of Λ at the true value $\boldsymbol{\theta}_0$ and the proofs adopted are valid whatever may be the rank of Λ. If the rank is full, i.e., equal to q, then the Assumption 3 is equivalent to

$$\underset{n\to\infty}{\text{plim}} [\sqrt{\ } (d_r - i^{r1} z_1 - \ldots - i^{rq} z_q)] = 0 \qquad \ldots \quad (2.5)$$

and the proofs of the propositions considered are extremely simple when (2.5) holds.

It has been shown in (Rao, 1960b) that the rate of convergence in (2.4) is highest when $\dot{\boldsymbol{\theta}}$ is an m.l. estimate, which indicates some merit in using an m.l. estimate in preference to any other efficient estimate.

3. Chi-square goodness of fit and associated problems

3.1. *Preliminary lemmas.* We shall assemble, to begin with, the key notations and their relationships to be used in the lemmas. The transpose of a column vector is indicated by a prime. The probabilities π_i and their derivatives are taken at the true value θ_0 of the parameter unless stated otherwise.

$$V' = \left(\sqrt{n}\frac{p_1-\pi_1}{\sqrt{\pi_1}}, \ldots, \sqrt{n}\frac{p_k-\pi_k}{\sqrt{\pi_k}} \right)$$

$$U' = \left(\sqrt{n}\frac{\dot{\pi}_1-\pi_1}{\sqrt{\pi_1}}, \ldots, \sqrt{n}\frac{\dot{\pi}_k-\pi_k}{\sqrt{\pi_k}} \right)$$

$$Z' = (\sqrt{n}\, z_1, \ldots, \sqrt{n}\, z_q), \quad z_r = \Sigma \frac{p_i}{\pi_i} \frac{\partial \pi_i}{\partial \theta_r}$$

$$D' = (\sqrt{n}\, d_1, \ldots, \sqrt{n}\, d_q), \quad d_i = \dot{\theta}_i - \theta_i^0$$

$B(q \times k) = \left(\frac{1}{\sqrt{\pi_i}} \frac{\partial \pi_i}{\partial \theta_r} \right)$ so that $BB' = \Lambda$, the dispersion matrix of Z and $Z = BV$.

Observe that $U \sim B'D$, since

$$\sqrt{n}\, \frac{\dot{\pi}_i-\pi_i}{\sqrt{\pi_i}} \sim \Sigma\, \frac{\sqrt{n}}{\sqrt{\pi_i}} \frac{\partial \pi_i}{\partial \theta_r}\, d_r$$

which follows from the continuity of the derivatives of $\pi_i(\theta)$, where $\sim$ stands for equivalence of asymptotic distributions.

By cov(X, Y), where X and Y are vectors, is meant the matrix of covariances

$$\begin{pmatrix} \text{cov}\,(x_1, y_1) & \ldots & \text{cov}\,(x_1, y_k) \\ \ldots & \ldots & \ldots \\ \text{cov}\,(x_m, y_1) & \ldots & \text{cov}\,(x_m, y_k) \end{pmatrix}$$

For instance, by straight computation, we find

$$\text{cov}\,(V, Z) = B'.$$

Lemma 1: *The a.d. (asymptotic distribution) of Z is q-variate normal with dispersion matrix Λ.*

This result is a consequence of every linear function $Z'L$, (L being a non-random vector) having an asymptotic normal distribution (Wald and Wolfowitz, 1944). The distribution is, however, singular when the rank of Λ is not full.

Lemma 2a: *Under the Assumptions 1 and 3, the a.d. of $D' \Lambda D$ is $\chi^2(q)$ when the rank of Λ is full.*

If the rank of Λ is q, then by (2.5)

$$D \sim \Lambda^{-1} Z, \quad D' \Lambda D \sim Z' \Lambda^{-1} Z.$$

The a.d. of $Z' \Lambda^{-1} Z$ is $\chi^2(q)$ in virtue of the result of Lemma 1 and the same is then true of $D' \Lambda D$.

Lemma 2b : *Uunder the Assumptions 1 and 3, the a.d. of $D' \Lambda D$ is $\chi^2(t)$, where t is the rank of Λ.*

There exists a matrix C, non-singular such that $C \Lambda C' = \Delta$, where Δ is diagonal and of order q, but with only t elements > 0 and the rest equal to zero. Denoting by ∇ the matrix obtained by replacing the non-zero elements of Δ by their reciprocals we find[1]

$$\begin{aligned} \Lambda &= \Lambda C' \nabla C \Lambda \\ D' \Lambda D &= D' \Lambda C' \nabla^{\frac{1}{2}} \nabla^{\frac{1}{2}} C \Lambda D \\ &\sim Z' C' \nabla^{\frac{1}{2}} \nabla^{\frac{1}{2}} C Z. \end{aligned} \tag{3.1.1}$$

The a.d. of $D' \Lambda D$ is same as that of sum of squares of linear functions $Z' C' \nabla^{\frac{1}{2}}$ of Z. The dispersion matrix of $Z' C' \nabla^{\frac{1}{2}}$ is

$$\nabla^{\frac{1}{2}} C \Lambda C' \nabla^{\frac{1}{2}} = \nabla^{\frac{1}{2}} \Delta \nabla^{\frac{1}{2}} = I_t$$

where I_t is a diagonal matrix with t elements equal to unity and the rest to zero. Hence the a.d. of (3.1.1) is $\chi^2(t)$ and so is that of $D' \Lambda D$.

Lemma 3 : *Under the Assumptions 1 and 3 the a.d. of $U'U = n\Sigma(\dot{\pi}_i - \pi_i)^2/\pi_i$ is $\chi^2(t)$ where $t = \text{rank } \Lambda$.*

Since $U'U \sim D'BB'D = D' \Lambda D$, the result follows from Lemma 2a or 2b.

Lemma 4: *The Assumptions 1 and 3 imply plim $(V-U)'U = 0$ or writing out in full*

$$\text{plim } n \, \Sigma \, \frac{(p_r - \dot{\pi}_r)(\dot{\pi}_r - \pi_r)}{\pi_r} = 0.$$

This follows from

$$(V-U)'U = V'B'D - D' \Lambda D = (Z' - D' \Lambda) D \to 0 \text{ in probability by (2.4).}$$

Lemma 5: *Under the same conditions as in Lemma 4, a.c.$(V-U, U) = 0$.*

Since $BB' = \Lambda$, there exists a matrix G such that $B = \Lambda G$, and therefore, $U \sim B'D = G' \Lambda D \sim G'Z$. Consider

$$\begin{aligned} \text{a.c. } (V - G'Z, G'Z) &= \text{a.c.}(V, G'Z) - \text{a.c. } (G'Z, G'Z) \\ &= \text{c.}(V, Z)G - G' \Lambda G \\ &= B'G - G' \Lambda G = (B' - G' \Lambda) G' = 0. \end{aligned}$$

[1] The matrix $C' \nabla C$ is defined to be a pseudo-inverse of Λ. The properties of such inverses and their use in statistics are discussed by the author (Rao, 1955b). The relationship $\Lambda = \Lambda C' \nabla C \Lambda$ is established by post and pre-multiplying with C' and C and using the relation $C \Lambda C' = \Delta$.

Lemma 6* : *Let X be a p-variate normal variable with mean zero and dispersion matrix H of rank $f \leqslant p$. If $Q_1(X)$ and $Q_2(X)$ are two quadratic forms such that $[Q_1(X)+Q_2(X)]$ is $\chi^2(c)$, $Q_1(X)$ is $\chi^2(b)$ and $Q_2(X)$ is non-negative, then $Q_2(X)$ is $\chi^2(c-b)$.*

Transform X to W such that $w_1, \ldots, w_f$ are each independently $N(0, 1)$ and $w_{f+1}, \ldots, w_p$ are all $N(0, 0)$. The transformed quadratic forms $Q_1(W)$, $Q_2(W)$ are, therefore, equivalent, with probability one, to those obtained by omitting $w_{f+1}, \ldots, w_p$. The problem reduces to the case of $G = I$ and $p = f$.

Consider an orthogonal transformation from X to Y such that

$$X'X = Y'Y \text{ and } Q_1(X)+Q_2(X) = \Sigma\lambda_i y_i^2.$$

Since $Q_1(X)+Q_2(X)$ is $\chi^2(c)$, it follows that $\lambda_i = 1$, $i \leqslant c$ and zero otherwise. Hence

$$y_1^2+\ldots+y_c^2 = Q_1(Y)+Q_2(Y). \qquad \ldots \quad (3.1.2)$$

Since $Q_2(Y)$ is non-negative and $Q_1(Y)$ is non-negative being distributed as χ^2, $Q_1(Y)$ and $Q_2(Y)$ contain only the variables $y_1, \ldots, y_c$. The problem is thus reduced to the case of c independent normal variables, each with zero mean and unit variance, together with the condition (3.1.2). Under such conditions, given that $Q_1(Y)$ is $\chi^2(b)$, it follows that $Q_2(Y)$ is $\chi^2(c-b)$, as may be proved by considering an orthogonal transformation from Y to Z such that

$$\Sigma y_i^2 = \Sigma z_i^2 \quad \text{and} \quad Q_1(Y) = z_1^2+\ldots+z_b^2.$$

3.2. *The goodness of fit test.*

Theorem 1: *Under the Assumptions* 1 *and* 3, *the a.d. of*

$$\Sigma \frac{n(p_r-\dot{\pi}_r)^2}{\dot{\pi}_r}$$

is $\chi^2(k-t-1)$, where t = rank Λ.

Consider

$$\Sigma \frac{n(p_r-\pi_r)^2}{\pi_r} = \Sigma \frac{n(p_r-\dot{\pi}_r)^2}{\pi_r} + \Sigma \frac{n(\dot{\pi}_r-\pi_r)}{\pi_r} + \text{product terms}$$

or in matrix notation,

$$V'V \sim (V-U)'(V-U)+U'U \qquad \ldots \quad (3.2.1)$$

since the product term$\rightarrow 0$ in probability by Lemma 4. The a.d. of $V'V$ is $\chi^2(k-1)$ and that of $U'U$ is $\chi^2(t)$ by Lemma 3. Hence by an application of Lemma 6, $(V-U)'(V-U)$ is $\chi^2(k-1-t)$. Alternatively, since a.c. $(V-U, U) = 0$ by Lemma 5, $(V-U)'(V-U)$ and $U'U$ are asymptotically independently distributed. Hence the a.d. of $(V-U)'(V-U)$ is $\chi^2(k-1-t)$. In order to prove the result of Theorem 1 we observe that

$$\Sigma \frac{n(p_r-\dot{\pi}_r)}{\dot{\pi}_r} \sim \Sigma \frac{n(p_r-\dot{\pi}_r)}{\pi_r}$$

because of continuity of $\pi_r(\theta)$.

* This lemma was proved and included at the suggestion of my colleague, Dr. S. K. Mitra. For proving Theorem 1, we need either Lemma 5 or Lemma 6.

General remarks on the result of Theorem 1. The decomposition of the total χ^2 in (3.2.1) deserves some comments. If the observed frequency in any cell is indicated by $O = np$, the hypothetical frequency by $H = n\pi$ and the estimated (or expected) by $E = n\dot{\pi}$, the equation (3.2.1) may be written

$$\Sigma \frac{(O-H)^2}{H} \sim \Sigma \frac{(O-E)^2}{E} + \Sigma \frac{(E-H)^2}{H} \qquad \ldots \quad (3.2.2)$$

$$\chi^2(k-1) = \chi^2(k-t-1) + \chi^2(t).$$

For an application of the χ^2 goodness of fit test, it is necessary to enquire whether the rank of Λ is constant over the admissible set of parameters, since the true value of the parameter $\boldsymbol{\theta}$ is unknown. What is asserted by Theorem 1 is that the degrees of freedom of a.d. is dependent only on the rank of Λ at the true value, although the rank of Λ may not be same at all points.

The statistic $\Sigma(E-H)^2/H$ may be used to test the hypothesis that a particular set of probabilities is true given that the admissible set has the representation $\pi_i(\boldsymbol{\theta})$.

3.3. *Test for deviations in any particular set of cells.* Cochran (1954), and Rao and Chakravarty (1956) considered problems where attention is concentrated on the deviation of the observed from the expected frequency in a particular cell or the deviations in a particular set of cells. Tests for examining the singificance of such deviations are extremely useful as they may lead to a suitable explanation of the departure from hypothesis when indicated by a large value of the χ^2 goodness of fit test. The result of the following lemma will be useful in computing the variances and covariances of the deviations.

From Lemma 5 we have

$$\text{a.c. } (V-U, U) = 0 \Longrightarrow \text{a.c.}(V, U) = \text{a.c. } (U, U).$$

Hence

$$\left.\begin{aligned} \text{a.c. } (V-U, V-U) &= \text{c. } (V, V) - \text{a.c. } (U, U) \\ &= \text{c.}(V, V) - G' \Lambda G \text{ where } G \text{ is as defined in Lemma 4} \\ &= \text{c.}(V, V) - B' \Lambda^{-1} B \text{ when rank } \Lambda = q. \end{aligned}\right\} \qquad \ldots \quad (3.3.1)$$

We have, thus, very simple formulae for finding the asymptotic variances and covariances of the deviations $(p_r - \dot{\pi}_r)$. For instance, to test the departure in the r-th cell we need

$$\text{a.v.}[\sqrt{n}(p_r - \dot{\pi}_r)] = \text{v.}[\sqrt{n}(p_r - \pi_r)] - \text{a.v.}[\sqrt{n}(\dot{\pi}_r - \pi_r)]$$
$$= \pi_r(1-\pi_r) - \text{a.v.}[\sqrt{n}(\dot{\pi}_r - \pi_r)].$$

When the rank of Λ is q, the last expression can be evaluated in terms of $\Lambda^{-1} = (i^{rs})$ which represents the asymptotic variances and covariances of the estimates $\dot{\theta}_r$

$$\text{a.v.}[\sqrt{n}(\dot{\pi}_r - \pi_r)] = \Sigma\Sigma \frac{\partial \pi_r}{\partial \theta_u} \frac{\partial \pi_r}{\partial \theta_v} i^{uv}$$

The deviation $p_r - \dot{\pi}_r$ can be tested using its standard error.

If the deviations in a number of cells have to be tested simultaneously we have to compute the variance-covariance matrix of the deviations by using the result (3.3.1). With the inverse of the matrix so computed we can construct a quadratic form in the deviations, which has a χ^2 distribution with d.f. equal to the number of deviations to be examined. It may, however, happen that the rank of the variance-covariance matrix of the deviations is not full, in which case the significance of all the deviations cannot be examined simultaneously. We consider only those deviations for which the dispersion matrix remains non-singular.

3.4. *Test for the hypothesis that the parameters belong to a subset.* We need some further assumptions to consider the problem of testing whether the parameter $\boldsymbol{\theta}$ belongs to a subset assuming the specification $\pi_i(\boldsymbol{\theta})$ to be true.

Assumption 4 : The locus of $\boldsymbol{\theta}$ in the subset is represented by

$$\theta_i = g_i(\alpha_1, \ldots, \alpha_r), \quad i = 1, \ldots, q,$$

where g_i admit continuous first derivatives and $(\alpha_1, \ldots, \alpha_r)$ is confined to R^r or to some non-degenerate interval in R^r.

Assumption 5: There exist consistent estimates α_i^*, having asymptotic distributions such that

$$\text{plim} \quad \sqrt{n}\, [y_u - j_{u1}(\alpha_1^* - \alpha_1^0) - \ldots - j_{ur}(\alpha_r^* - \alpha_r^0)] = 0, \; u = 1, \ldots, r$$

where $y_u = \dfrac{\partial}{\partial \alpha_u} \Sigma\, p_i \log \pi_i$, and (j_{mp}) is the information matrix for the parameters $\alpha_1, \ldots, \alpha_r$.

Assumption 5 is same as Assumption 3 in terms of the new parameters.

Lemma 7 : *The Assumptions* 1, 3, 4 *and* 5 *imply*

$$plim \; n\, \Sigma \, \frac{(\dot{\pi}_u - \pi_u^*)(\pi_u^* - \pi_u)}{\pi_u} = 0.$$

By the same argument as in Lemma 4 we have

$$\text{plim} \; n\, \Sigma \, \frac{(p_u - \pi_u^*)(\pi_u - \pi_u)}{\pi_u} = 0$$

and, therefore, Lemma 8 is true if

$$\text{plim} \; n\, \Sigma \, \frac{(p_u - \dot{\pi}_u)(\pi_u^* - \pi_u)}{\pi_u} = 0. \qquad \ldots \quad (3.4.1)$$

Since $\sqrt{n}(\pi_u^* - \pi_u) \sim \sqrt{n}\, \Sigma \, \dfrac{\partial \pi_u}{\partial \alpha_i} (\alpha_i^* - \alpha_i^0)$, to prove (3.4.1) it is enough to show

$$\text{plim} \quad \Sigma \, \frac{\sqrt{n}}{\pi_u} \, \frac{\partial \pi_u}{\partial \alpha_i} (p_u - \dot{\pi}_u) = 0. \qquad \ldots \quad (3.4.2)$$

The left hand side of (3.4.2) is equal to

$$\text{plim} \sum_t \frac{\partial g_t}{\partial \alpha_i} \sum_u \frac{\sqrt{n}}{\pi_u} \frac{\partial \pi_u}{\partial \theta_t}(p_u - \dot{\pi}_u).$$

$$= \sum \frac{\partial g_t}{\partial \alpha_i} \text{ plim } \sum \frac{\sqrt{n}}{\pi_u} \frac{d\pi_u}{\partial \theta_t} (p_u - \dot{\pi}_u)$$

$$= \sum \frac{\partial g_t}{d\alpha_i} \text{. plim } \sqrt{n}(z_t - i_{t1}d_1 - \ldots - i_{tq}d_q) = 0 \text{ by (2.4).}$$

Theorem 2: *Under the Assumptions* 1, 3, 4 *and* 5, *the a.d. of*

$$n \sum \frac{(\dot{\pi}_m - \pi^*_m)^2}{\pi^*_m}$$

is $\chi^2(t-a)$, *where* $t = rank\ \Lambda$ *and* $a = rank\,(j_{rs})$.

The result follows from the decomposition

$$n \sum \frac{(\dot{\pi}_m - \pi_m)^2}{\pi_m} \sim n \sum \frac{(\dot{\pi}_m - \pi^*_m)^2}{\pi^*_m} + n \sum \frac{(\pi^*_m - \pi_m)^2}{\pi_m} \qquad \ldots \quad (3.4.4)$$

where the left hand side is asymptotically $\chi^2(t)$ by Theorem 1, and by Lemma 3, the second term on the right hand side is asymptotically $\chi^2(a)$. Lemma 7 justifies the expansion (3.4.4) and finally an application of Lemma 6 gives the desired result.

General comments on the result of Theorem 2. If we represent by $E_2 = n\pi^*$, expected or estimated frequency under the second hypothesis (that the parameter $\boldsymbol{\theta}$ is confined to a subset) the test criterion is

$$\sum \frac{(E_1 - E_2)^2}{E_2} \qquad \ldots \quad (3.4.5)$$

where it may be noted that E_1 is written for E, the expected frequency under the first specification. Combining the decompositions (3.2.2) and (3.4.4) we have

$$\sum \frac{(O-H)^2}{H} \sim \sum \frac{(O-E_1)^2}{E_1} + \sum \frac{(E_1 - E_2)^2}{E_2} + \sum \frac{(E_2 - H)^2}{H}. \qquad \ldots \quad (3.4.6)$$

Some years ago, the author (Rao, 1948) suggested a general criterion from which several large sample tests having asymptotic χ^2 distribution were deduced as special cases. The test (3.4.5) is, however, different from the previous test. Under the null hypothesis both are asymptotically equivalent and so also other tests proposed for the purpose (Mitra, 1955; Rao and Chakravarty, 1956). But differences may exist in their relative efficiencies in finite samples.

3.5. *Sufficient conditions for the validity of the Assumption* 3. The Assumption 3 or 5 as stated may be difficult to verify and it may, therefore, be useful to have some simple conditions under which they are true. The following lemmas provide some answers.

Lemma 8 : *The Assumptions* 1 *and* 2a, *together with the non-singularity of* Λ *imply the Assumption* 3.

We first consider the set S of $\boldsymbol{\theta}$, such that

$$\Sigma\, [\pi_i(\boldsymbol{\theta}_0)+\epsilon_1] \log \pi_i(\boldsymbol{\theta}_0) > \Sigma\, [\pi_i(\boldsymbol{\theta}_0)-\epsilon_1] \log \pi_i(\boldsymbol{\theta})$$

where $\epsilon_1 > 0$ is fixed to make $\pi_i(\boldsymbol{\theta}_0)-\epsilon_1 > 0$. For p_i such that $\pi_i(\boldsymbol{\theta}_0)-\epsilon_2 \leqslant p_i \leqslant \pi_i\,(\boldsymbol{\theta}_0)+\epsilon_2$, where $\epsilon_2 < \epsilon_1$,

$$\Sigma\, p_i \log \pi_i(\boldsymbol{\theta}_0) \geqslant \Sigma\, [\pi_i(\boldsymbol{\theta}_0)+\epsilon_1] \log \pi_i(\boldsymbol{\theta}_0) > \Sigma\, [\pi_i(\boldsymbol{\theta}_0)-\epsilon_1] \log \pi_i(\boldsymbol{\theta}) \geqslant \Sigma\, p_i \log \pi_i(\boldsymbol{\theta}).$$

For $\boldsymbol{\theta}$ outside S, and $(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 \geqslant \delta$, $\log \{\pi_i(\boldsymbol{\theta}_0)/\pi_i(\boldsymbol{\theta})\}$ are bounded in which case

$$\Sigma\, p_i \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})}$$

can be made uniformly close (within ϵ difference) to

$$\Sigma\, \pi_i(\boldsymbol{\theta}_0) \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})}$$

by choosing p_i sufficiently close to $\pi(\boldsymbol{\theta}_0)$, say $\pi_i(\boldsymbol{\theta}_0)-\epsilon_3 \leqslant p_i \leqslant \pi_i(\boldsymbol{\theta}_0)+\epsilon_3$. Since by Assumption 2

$$\inf_{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 \geqslant \delta} \Sigma\, \pi_i(\boldsymbol{\theta}_0) \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})} \geqslant \epsilon > 0$$

$$\Sigma\, p_i \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})} > 0 \text{ for all } \boldsymbol{\theta} \text{ such that } (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 \geqslant \delta \qquad (3.5.1)$$

when $\pi_i(\boldsymbol{\theta}_0)-\epsilon_4 \leqslant p_i \leqslant \pi_i(\boldsymbol{\theta}_0)+\epsilon_4$, where ϵ_4 is smaller of ϵ_2 and ϵ_3.

Since $\pi_i(\boldsymbol{\theta}_0) > 0$, and $\pi_i(\boldsymbol{\theta})$ are continuous, the result (3.5.1) shows that the supremum of $\Sigma\, p_i \log \pi_i(\boldsymbol{\theta})$ is attained in the open interval $(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 < \delta$. As δ can be chosen arbitrarily small, and p_i is close to $\pi_i(\boldsymbol{\theta}_0)$ with probability one, the value $\dot{\boldsymbol{\theta}}$ at which the supremum is attained provides a consistent estimate of $\boldsymbol{\theta}$. So far we have used only Assumption 2 and continuity of $\pi_i(\boldsymbol{\theta})$.

If $\pi_i(\boldsymbol{\theta})$ are differentiable, the derivative of $\Sigma\, p_i \log \pi_i(\boldsymbol{\theta})$ vanishes at $\dot{\boldsymbol{\theta}}$. This shows that the m.l. equation has at least one root which maximises the likelihood and which provides a consistent estimate.

We shall use the condition that rank Λ is q, and the continuity of the derivatives, to prove Assumption 3. The rank of the matrix Λ, however, does not play a significant role in the proof. It is felt that a weaker condition than this may be sufficient for this purpose.

The m.l. equation is

$$\Sigma\, \frac{p_r}{\dot{\pi}_r} \frac{\partial \dot{\pi}_r}{\partial \dot{\theta}_s} = 0, \quad s = 1, \ldots, q$$

$$\text{or } \Sigma\, \frac{p-\dot{\pi}_r}{\dot{\pi}_r} \frac{\partial \dot{\pi}_r}{\partial \dot{\theta}_s} = 0 = \Sigma\, \frac{\sqrt{n}(p-\pi_r)}{\dot{\pi}_r} \frac{\partial \dot{\pi}_r}{\partial \dot{\theta}_s} - \Sigma\, \frac{\sqrt{n}(\dot{\pi}_r-\pi_r)}{\dot{\pi}_r} \frac{\partial \dot{\pi}_r}{\partial \dot{\theta}_s} \quad \ldots \quad (3.5.2)$$

The first term on the right hand side of (3.5.2) is $\sim \sqrt{n}\, z_s$. Using the expansion

$$\sqrt{n}(\dot{\pi}_r - \pi_r) = \Sigma\left(\frac{\partial \pi_r}{\partial \theta_t} + e_{rt}\right)\sqrt{n}\, d_t, \text{ where } e_{rt} \to 0 \text{ as } \dot{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$$

the second term on the right hand side of (3.5.2) becomes

$$\sqrt{n}\, \Sigma\, j_{ts}\, d_t \qquad (3.5.3)$$

where
$$j_{ts} = \Sigma\, \frac{1}{\pi_r} \frac{\partial \dot{\pi}_r}{\partial \theta_s}\left(\frac{\partial \pi_r}{\partial \theta_t} + e_{rt}\right)$$

$$= \Sigma\left(\frac{1}{\pi_r}\frac{\partial \pi_r}{\partial \theta_s} + e'_{rs}\right)\left(\frac{\partial \pi_r}{\partial \theta_t} + e_{rt}\right),\ e'_{rs} \to 0 \text{ as } \dot{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$$

$$= i_{ts} + e''_{ts},\quad e''_{ts} \to 0 \quad \text{as} \quad \dot{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0 \qquad \ldots \quad (3.5.4)$$

$$\to i_{ts} \text{ as } \dot{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0.$$

We may write the equation (3.5.2) as the equivalence

$$\sqrt{n}\, z_s \sim \Sigma\, j_{ts} \sqrt{n}\, d_t. \qquad (3.5.5)$$

By the assumption that $|i_{ts}| \neq 0$, when $\dot{\boldsymbol{\theta}}$ is sufficiently close to $\boldsymbol{\theta}_0$, $|j_{rs}| \neq 0$ in virtue of (3.5.4). Therefore, the equivalence (3.5.5) may be written

$$\sqrt{n}\, d_s \sim \Sigma\, j^{ts} \sqrt{n}\, z_t,\ (j^{ts}) = (j_{ts})^{-1} \qquad \ldots \quad (3.5.6)$$

$$\sim \Sigma i^{ts} \sqrt{n}\, z_t \text{ as } j^{ts} \to i^{ts} \text{ in probability} \qquad \ldots \quad (3.5.7)$$

which shows that the a.d. of $\sqrt{n}\, d_1, \ldots, \sqrt{n}\, d_q$ is multivariate normal with dispersion matrix Λ^{-1}. Inverting the relation (3.5.7) we have

$$\sqrt{n}\, z_s \sim \Sigma\, i_{ts} \sqrt{n}\, d_t = \sqrt{n}\, \Sigma\, i_{ts}\, d_t$$

which is (iii) of the Assumption 3.

The Assumption 2 used in Lemma 8 is important as it specifies the condition under which we can assert that the m.l. estimate has the properties mentioned in Assumption 3. But if merely the existence of estimates satisfying Assumption 3 has to be established, Assumption 2 may be replaced by the weaker identifiability condition.

Lemma 9 : *Assumptions* 1 *and* 2b *imply Assumption* 3.

We consider a sphere of radius δ round $\boldsymbol{\theta}_0$. Since

$$\inf\ \Sigma\ \pi_i(\boldsymbol{\theta}_0) \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})} \qquad \ldots \quad (3.5.8)$$

over the sphere is attained, because of continuity of $\pi(\boldsymbol{\theta})$, at some point on the sphere the identifiability Assumption 2b ensures that the expression (3.5.8) is greater than $\epsilon > 0$. The argument of Lemma 8 applied to points over the chosen sphere shows that the likelihood at θ_0 exceeds the likelihood of any point on the sphere with probability unity as $n \to \infty$. This shows that a local maximum of the likelihood is attained at some point inside the sphere. The first derivative of log likelihood vanishes at that point. Hence there exists a root of the m.l. equation which is consistent. The rest of the argument is as in Lemma 8.

3.6. *Goodness of fit tests when the Assumption* 3 *is not satisfied.* If the estimated parameters do not satisfy the Assumption 3 the statistic $\Sigma(O-E)^2/E$ need not have a χ^2 distribution. But in some cases, as when the method of moments is followed one may construct an alternative statistic with its a.d. as χ^2 on the same d.f. as the goodness of fit χ^2.

Let the parameters be estimated by the method of moments i.e., by equating the observed moments of the grouped distribution to the hypothetical values for the grouped distribution. As the estimating equations are linear in the frequencies, the estimated deviations

$$p_i-\dot{\pi}_1, \ \ldots, \ p_k-\dot{\pi}_k \quad \ldots \quad (3.6.1)$$

are subject to as many linear restrictions as there are independent parameters, besides their sum being zero. If the matrix of derivatives of π_i has rank q, the asymptotic dispersion matrix of the deviations will have rank $(k-q-1)$. We need only choose $(k-q-1)$ independent deviations and using the reciprocal of their dispersion matrix construct the quadratic form. This has χ^2 distribution with $(k-q-1)$ d.f. The asymptotic dispersion matrix of the deviations can be easily computed by the usual methods.

We can also test the significance of deviations in any particular cell or deviations in a particular set of cells as in Section 3.3 using the asymptotic variances and covariances of the deviations (3.6.1).

There is another situation where the parameters are estimated by an efficient method utilizing the individual observations and not simply the observed frequencies in certain class intervals. In such a case the statistic $\Sigma(O-E)^2/E$, when used as $\chi^2(k-q-1)$ over estimates significance. The extent to which this happens has been studied by Chernoff and Lehmann (1954). An alternative expression is given for the excess in terms of the difference between estimates obtained in two different ways.

Let us indicate by θ^*, the estimate of θ, and by π^* that of π, obtained from the original observations by an efficient method of estimation such as the m.l. The information matrix for a single observation is denoted by (j_{rs}), reserving (i_{rs}) for the grouped distribution. Let $y_s = n^{-1}(\partial \log L/\partial\theta_s)$, where $L = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$ and $f(x, \theta)$ is differentiable. We make the following assumption regarding the estimate θ^*.

Assumption 6:

$$\text{plim} \quad \sqrt{n}[y_s - j_{s1}(\theta_1^*-\theta_1^0) - \ldots - j_{sq}(\theta_q^*-\theta_q^0)] = 0, \ s = 1, \ldots, q.$$

Lemma 11 : *Under the Assumptions* 3 *and* 6

i) a.v. $\sqrt{n}(p_r-\pi_r^*) = $ v.$[\sqrt{n}(p_r-\pi_r)]-$a.v.$[\sqrt{n}(\pi_r^*-\pi_r)]$

ii) a.c.$[\sqrt{n}(p_i-\pi_i^*),\ \sqrt{n}(p_j-\pi_j^*)]$

$=$ c.$[\sqrt{n}(p_i-\pi_i), \sqrt{n}(p_j-\pi_j)]-$a.c.$[\sqrt{n}(\pi_i^*-\pi_i),\ \sqrt{n}(\pi_j^*-\pi_j)]$

iii) $\text{plim}\ n\ \Sigma\ \dfrac{(p_r-\dot{\pi}_r)(\dot{\pi}_r-\pi_r^*)}{\pi_r} = 0.$

The results (i) and (ii) are immediate consequences of the equations

$$\text{a.c. } [\sqrt{n}(p_r-\pi_r^*), \sqrt{n}(\pi_s^*-\pi_s)] = 0 \text{ for all } r \text{ and } s \qquad \ldots \quad (3.6.2)$$

which are true if

$$\text{a.c. } [\sqrt{n}(p_r-\pi_r^*), \sqrt{n}\, y_s] = 0 \quad \text{for all } r \text{ and } s. \qquad \ldots \quad (3.6.3)$$

Let us assume for simplicity that the ranks of (j_{rs}) and (i_{rs}) are both equal to q, the number of parameters. To prove (3.6.3) we have to show that

$$\text{a.c. } [\sqrt{n}(p_r-\pi_r), \sqrt{n}\, y_s] = \text{a.c. } [\sqrt{n}(\pi_r^*-\pi_r),\ \sqrt{n}\, y_s].$$

$$= \frac{\partial \pi_r}{\partial \theta_s} \text{ (obtained by expanding } \pi_r^* \text{ in terms of } y_i) \qquad \ldots \quad (3.6.4)$$

Since $E(p_r) = \pi_r$, i.e.,

$$\int p_r f(x_1) \ldots f(x_n) dx_1 \ldots dx_n = \pi_r$$

differentiating with respect to θ_s, we obtain

$$\int p_r L \frac{\partial \log L}{\partial \theta_s} dx_1 \ldots dx_n = \int (\sqrt{n}\, p_r)\, (\sqrt{n}\, y_s) L\ dx_1 \ldots dx_n = \frac{\partial \pi_r}{\partial \theta_s} \qquad \ldots \quad (3.6.5)$$

Comparing (3.6.4) and (3.6.5) we find (3.6.3) is true. Further $\sqrt{n}(\dot{\pi}_r-\pi_r^*) \sim \sqrt{n}$ $\Sigma\ \dfrac{\partial \pi_r}{\partial \theta_s}(\dot{\theta}_s-\theta_s^*)$. Hence result (iii) of the lemma follows if

$$\text{plim}\ \sqrt{n}\ \Sigma\ \frac{\partial \pi_r}{\partial \theta_s}\ \frac{(p_r-\dot{\pi}_r)}{\pi_r} = 0$$

which is true in virtue of (3.4.3).

The results of Lemma 11 are important in many ways. For instance the significance of the deviations in any given set of cells can be tested, although the estimates are not obtained from the grouped distribution, since the variances and covariances of the deviations can be easily computed by using the results (i) and (ii). Thus

$$\text{a.v. } [\sqrt{n}(p_r-\pi_r^*)] = \pi_r(1-\pi_r) - \Sigma\ \Sigma\ \frac{\partial \pi_r}{\partial \theta_s}\ \frac{\partial \pi_r}{\partial \theta_t} j^{st}$$

$$\text{a.c. } [\sqrt{n}(p_r-\pi_r^*)\ ,\ \sqrt{n}(p_u-\pi_u^*)] = -\pi_r\pi_u - \Sigma\ \Sigma\ \frac{\partial \pi_r}{\partial \theta_s}\ \frac{\partial \pi_u}{\partial \theta_t} j^{st}.$$

It may be observed that the dispersion matrix of all the deviations $(p_i-\pi_i^*)$ may have the maximum rank $(k-1)$, so that a χ^2 based on $(k-1)$ degrees of freedom may be constructed to test goodness of fit. But this is not likely to be an efficient test. On the other hand the statistic

$$n \Sigma \frac{(p_i-\pi_i^*)^2}{\pi_i^*}$$

need not have a χ^2 distribution asymptotically as shown below. We use the result (iii) of Lemma 11 to obtain the decomposition

$$n \Sigma \frac{(p_i-\pi_i^*)^2}{\pi_i^*} \sim n \Sigma \frac{(p_i-\dot{\pi}_i)^2}{\dot{\pi}_i} + n \Sigma \frac{(\dot{\pi}_i-\pi_i^*)^2}{\pi_i^*} .$$

The first term on the right hand side is asymptotically $\chi^2(k-q-1)$ as shown in Theorem 1. The second term which is asymptotically independent of the first provides the excess and depends on the numerical difference between estimates obtained in two ways.

4. Tests in contingency tables

Problems associated with a single multinomial distribution have been fully discussed in the earlier sections. Another group of problems, which is important in practice, is related to contingency tables, or independent samples from a number of multinomial distributions. Let us suppose that we have samples of sizes $n_1, \ldots, n_m$ $(n_1+\ldots+n_m = n)$ from m finite multinomial populations, all not necessarily having an equal number of cells. One of the problems is to test the hypothesis that the cell probabilities could be represented in terms of q parameters, $\theta_1, \ldots, \theta_q$. The likelihood of the parameters is the product of the likelihoods

$$L_1 L_2 \ldots L_m$$

corresponding to the m samples. Let us define

$$z_{ij} = \frac{1}{n_i} \frac{\partial \log L_i}{\partial \theta_j}$$

$$n_i = n\lambda$$

and (i^t_{rs}) is the information matrix per single observation from the t-th multinomial distribution. We make the following assumptions.

Assumption 7 : The rank of (g_{rs}) is q, where $g_{rs} = \Sigma \lambda_t i^t_{rs}$.

Assumption 8 : Every cell probability π_{ij} as a function of the parameters admits continuous first order differential coefficients.

Assumption 9 : If $\pi_{i1}, \ldots, \pi_{ik_i}$ are the probabilities for the i-th multinomial then

$$\sum_{j=1}^{k_i} \pi_{ij}(\boldsymbol{\theta}_0) \log \frac{\pi_{ij}(\boldsymbol{\theta}_0)}{\pi_{ij}(\boldsymbol{\theta})}$$

is bounded away from zero when $(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2 \geqslant \delta$, however small, for each i.

Following the arguments of Lemma 8 it can be shown that there exist consistent estimates $\dot{\theta}_1, \ldots, \dot{\theta}_q$ having asymptotic distributions, such that

$$\text{plim} \sqrt{n}[\Sigma \lambda_i z_{ij} - g_{j1}(\theta_1 - \theta_1^0) - \ldots - g_{jq}(\theta_q - \theta_q^0)] = 0$$

for $j = 1, \ldots, m$ and fixed $\lambda_1, \ldots, \lambda_m$. We may now prove the following in the same way as Theorems 1 and 2.

Theorem 3: *Under the Assumptions 7, 8 and 9 the asymptotic distribution of the statistic*

$$\sum_{i=1}^{m} \sum_{j=1}^{k_i} \frac{n_i(p_{ij} - \dot{\pi}_{ij})^2}{\dot{\pi}_{ij}} = \Sigma \, \Sigma \frac{(O-E)^2}{E}$$

is $\chi^2(\Sigma k_i - m - q)$.

We use the decomposition

$$\Sigma \, \Sigma \frac{n_i(p_{ij} - \pi_{ij})^2}{\pi_{ij}} \sim \Sigma \, \Sigma \frac{n_i(p_{ij} - \dot{\pi}_{ij})^2}{\dot{\pi}_{ij}} + \Sigma \, \Sigma \frac{n_i(\dot{\pi}_{ij} - \pi_{ij})^2}{\pi_{ij}} \quad \ldots \quad (4.1)$$

where the left hand side is asymptotically χ^2 $(\Sigma k_i - m)$ and the last term in (4.1) is asymptotically $\chi^2(q)$.

Theorem 4: *Under the Assumptions 7, 8 and 9 and the further assumptions that* (i) $\theta_1, \ldots, \theta_q$ *can be represented as functions of* $t < q$ *parameters* $\alpha_1, \ldots, \alpha_t$, (ii) *every* θ_i *admits first partial derivatives in* α_j, *which are continuous and* (iii) *a condition similar to Asumption 9 in terms of* α_i *is satisfied, the a.d. of*

$$\sum_i \sum_j \frac{n_i(\dot{\pi}_{ij} - \pi_{ij}^*)^2}{\pi_i^*} = \Sigma \, \Sigma \frac{(E_1 - E_2)^2}{E_2}$$

is $\chi^2(q-t)$, *where* π_{ij}^* *denotes the estimate of* π_{ij} *as a function of parameters* α_i, *and* E_1, E_2 *are the expectations under the original and new specifications.*

To prove the result we first obtain the decomposition

$$\sum_i \sum_j \frac{n_i(\dot{\pi}_{ij} - \pi_{ij})^2}{\pi_{ij}} \sim \Sigma \, \Sigma \frac{n_i(\dot{\pi}_{ij} - \pi_{ij}^*)^2}{\pi_{ij}^*} + \Sigma \, \Sigma \frac{n_i(\pi_{ij}^* - \pi_{ij})^2}{\pi_{ij}}$$

or

$$\Sigma \, \Sigma \frac{(E_1 - H)^2}{H} \sim \Sigma \, \Sigma \frac{(E_1 - E_2)^2}{E_2} + \Sigma \, \Sigma \frac{(E_2 - H)^2}{H}$$

and proceed as in Theorem 2.

Theorems 3 and 4 concerning several multinomial distributions cover the multidimensional contingency tables in so far as hypotheses specifying the cell probabilities are concerned.

As an application let us consider the phenotypic frequencies of O, A, B, AB, blood groups in two samples of individuals from two communities. The first hypothesis is that for each community the frequencies are consistent with Bernstein's theory. There are then four parameters, p_1, q_1, representing the A and B gene frequencies in one community and p_2, q_2, in the second. The second hypothesis specifies further that the gene frequencies in the two communities are same, so that all the cell probabilities involve only two parameters p, q. For estimating these parameters by the m.l. method reference may be made to Rao (1952).

phenotype	community 1			community 2		
	O	E_1	E_2	O	E_1	E_2
O	121	118.89	117.94	118	119.03	121.62
A	120	122.44	105.52	95	92.68	108.81
B	70	81.54	98.14	121	118.74	101.10
AB	33	30.13	31.40	30	32.65	32.37

The test for the first specification (consistency with Bernstein's theory) is

$$\Sigma \frac{(O-E_1)^2}{E_1} = 0.44 \quad \text{for community 1,} \quad \text{d.f.} = 1$$

$$= 0.35 \quad \text{for} \quad ,, \quad 2, \quad \text{d.f.} = 1$$

or a total of 0.79, which is small for $\chi^2(2)$. To test the equality of gene frequencies the statistic is

$$\Sigma\ \Sigma\ \frac{(E_1-E_2)^2}{E_2} = 11.04, \text{ with } 4-2 = 2 \text{ d.f.}$$

This is significant at 1% level, indicating differences in the gene frequencies. In a previous paper (Rao, 1948), the author examined the second hypothesis by using a different large sample test. The present test based on the difference between expected values under the two hypotheses seems to be more attractive in practice.

REFERENCES

CHERNOFF, H. and LEHMANN, E. L. (1954): The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Stat.*, **25**, 579–586.

COCHRAN, W. G. (1954): Some methods of strengthening χ^2 tests. *Biometrics*, 10, 417.

CRAMÉR, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press.

FISHER, R. A. (1921): On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, A, **222**, 309.

——— (1925) Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700.

HALDANE, J. B. S. (1951): A class of efficient estimates of a parameter. *Bull. Int. Stat. Inst.*, **33**, 231.

MITRA, S. K. (1955): Statistical analysis of categorical data. No. 142, Mimeographed Series, Institute of Statistics, North Carolina.

NEYMAN, J. (1949): Contributions to the theory of the χ^2 test. First Berkeley Symposium on Mathematical Statistics and Probability, 239.

RAO, C. R. (1948): Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proc. Camb. Phil. Soc.*, **44**, 50.

——— (1955a): Theory of the method of estimation by minimum chisquare. Proc. 29th Session of the International Statistical Institute, 25-32.

——— (1955b): Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā*, **15**, 253.

——— (1958): Maximum likelihood estimation for the multinomial distribution. *Sankhyā*, **18**, 139.

——— (1960a): Apparent anomalies and irregularities in maximum likelihood estimation. 32nd session of the International Statistical Institute, Tokyo, 1960.

——— (1960b): Asymptotic efficiency and limiting information (to appear in the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability).

RAO, C. R. and CHAKRAVARTY, I. M. (1956): Some small sample tests of significance for a Poisson distribution. *Biometrics*, **12**, 264-282.

WALD, A. and WOLFOWITZ, J. (1944): Statistical tests based on permutations of the observations. *Ann. Math. Stat.*, **15**, 358-372.

Paper received : November, 1960.