A NEW PROOF OF THE PEARSON-FISHER THEOREM

By M. W. Birch

Glasgow University

1. Introduction and summary. This paper is concerned with the theorem that the X^2 goodness of fit statistic for a multinomial distribution with r cells and with s parameters fitted by the method of maximum likelihood is distributed as χ^2 with $r - s - 1$ degrees of freedom. Karl Pearson formulated and proved the theorem for the special case $s = 0$. The general theorem was formulated by Fisher [2]. The first attempt at a rigorous proof is due to Cramér [1]. A serious weakness of Cramér's proof is that, in effect, he assumes that the maximum likelihood estimator is consistent. (To be precise, he proves the theorem for the subclass of maximum likelihood estimators that are consistent. But how are we in practice to distinguish between an inconsistent maximum likelihood estimator and a consistent one?) Rao [3] has closed this gap in Cramér's proof by proving the consistency of maximum likelihood for any family of discrete distributions under very general conditions.

In this paper the theorem is proved under more general conditions than the combined conditions of Rao and Cramér. Cramér assumes the existence of continuous second partial derivatives with respect to the "unknown" parameter while here only total differentiability at the "true" parameter values is postulated. There is a radical difference in the method of proof. While Cramér regards the maximum likelihood estimate as being the point where the derivative of the log-likelihood function is zero, here it is regarded as the point at which the likelihood function takes values arbitrarily near to its supremum.

The method of proof consists essentially of showing that the goodness of fit statistic is a quadratic form in the observed proportions when the observed proportions are close to the expected proportions. The known asymptotic properties of the multinomial distribution are then used. The asymptotic efficiency of the maximum likelihood estimator is proved at the same time.

2. Formal statement of the theorem.

THEOREM. *Suppose* $\boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi_i(\boldsymbol{\theta}), \cdots, \pi_r(\boldsymbol{\theta}))$ *is defined for* $\boldsymbol{\theta} \varepsilon \Theta$, *where* Θ *is a subspace of s-dimensional Cartesian Space* R^s. *For each* $\boldsymbol{\theta}$ *and* i, $\pi_i(\boldsymbol{\theta})$ *is a positive or zero real number and, for each* $\boldsymbol{\theta}$, $\sum_{i=1}^r \pi_i(\boldsymbol{\theta}) = 1$. *Suppose* (A) *that* $\boldsymbol{\theta}_0$ *is an interior point of* Θ. *Suppose* (B) *that, given any* $\epsilon > 0$, *there exists a* $\delta > 0$ *such that* $|\boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{\pi}(\boldsymbol{\theta}_0)| > \delta$ *whenever* $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| > \epsilon$. *Suppose* (C) *that* $\pi_i(\boldsymbol{\theta}_0) > 0$ *for each* i. *Suppose* (D) *that, for each* i, *numbers* a_{ij} *exist such that*

$$\pi_i(\boldsymbol{\theta}) = \pi_i(\boldsymbol{\theta}_0) + [\pi_i(\boldsymbol{\theta}_0)]^{\frac{1}{2}} \sum_j a_{ij}(\theta_j - \theta_{0j}) + o|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \text{ as } \boldsymbol{\theta} \to \boldsymbol{\theta}_0,$$

Received 15 August 1963.

i.e. *that* $\pi_i(\boldsymbol{\theta})$ *is totally differentiable at* $\boldsymbol{\theta}_0$ *with partial derivatives*

$$\partial\pi_i(\boldsymbol{\theta}_0)/\partial\theta_j = a_{ij}[\pi_i(\boldsymbol{\theta}_0)]^{\frac{1}{2}}.$$

Suppose (E) *that the matrix* $\mathbf{A}$ *has rank* s.

Let $\{X_k\}k = 1, 2, \cdots$ *be a sequence of independent random variables, each taking the value* i *with probability* $\pi_{0i} = \pi_i(\boldsymbol{\theta}_0)$. *Let* P_{ni} *be the proportion of* X's *in the first* n *trials taking the value* i.

Let $\hat{\boldsymbol{\theta}}_n$ *be any value of* $\boldsymbol{\theta}\ \varepsilon\ \overline{\Theta}$ *for which there exists a sequence* $\{\boldsymbol{\theta}_{nm}\}$, $m = 1, 2, \cdots$ *such that* $\boldsymbol{\theta}_{nm}\ \varepsilon\ \Theta$ *for each* m *and*

$$2n \sum_{i=1}^{r} [P_{ni} \ln \pi_i(\boldsymbol{\theta}_{nm}) - P_{ni} \ln P_{ni}] \rightarrow \sup_{\boldsymbol{\theta}\varepsilon\Theta} 2n \sum_{i=1}^{r} [P_{ni} \ln \pi_i(\boldsymbol{\theta}) - P_{ni} \ln P_{ni}]$$

and $\theta_{nm} \rightarrow \hat{\boldsymbol{\theta}}_n$.

Then, as $n \rightarrow \infty$, *the joint distribution of* $n^{\frac{1}{2}}(\hat{\theta}_{n1} - \theta_{01}), \cdots, n^{\frac{1}{2}}(\hat{\theta}_{ns} - \theta_{0s})$ *and* $n\sum [P_{ni} - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2/\pi_i(\hat{\boldsymbol{\theta}}_n)$ *tends to a distribution in which* $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *and* $n\sum [P_{ni} - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2/\pi_i(\hat{\boldsymbol{\theta}}_n)$ *are independently distributed,* $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *as* $\mathfrak{N}(\mathbf{0}, (\mathbf{A}'\mathbf{A})^{-1})$ *and* $n\sum [P_{ni} - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2/\pi_i(\hat{\boldsymbol{\theta}}_n)$ *as* χ^2_{r-s-1}.

NOTE 1. $n\sum [P_{ni} \ln \pi_i(\boldsymbol{\theta}) - P_{ni} \ln P_{ni}]$ may be interpreted as the likelihood of $\boldsymbol{\theta}$ given the first n "observations" $X_1, \cdots, X_n$. $n\sum [P_{ni} - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2/\pi_i(\hat{\boldsymbol{\theta}}_n)$ is the $\sum (O - E)^2/E$ statistic for testing goodness of fit.

NOTE 2. Regularity Condition B is equivalent to saying that $\boldsymbol{\pi}^{-1}$ is continuous at $\boldsymbol{\pi}_0$.

NOTE 3. $\mathbf{A}'\mathbf{A}$ is the information matrix for $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$.

NOTE 4. Regularity Condition D may be replaced by the stronger but more easily verified condition D$'$ that the partial derivatives $\partial\pi_i/\partial\theta_j$ exist for each i and j in some neighbourhood of $\boldsymbol{\theta}_0$ and are continuous at $\boldsymbol{\theta}_0$.

NOTE 5. Here $\overline{\Theta}$ denotes the closure of Θ in R^s if Θ is bounded and the closure of Θ in R^s plus a point at ∞ if Θ is not bounded. Thus $\hat{\boldsymbol{\theta}}_n$ may be infinite if Θ is not bounded.

NOTE 6. The definition of the maximum likelihood estimate is slightly unconventional. Technical difficulties arise if the maximum likelihood estimate is defined as the point at which the likelihood attains its supremum. For example, no such point may exist. The likelihood function may tend to its supremum at a point of discontinuity or at a point of the boundary of Θ not in Θ or in remote regions of R^s. If the maximum likelihood estimate is defined as above, there is always at least one such estimate, though it may be a point at which $\boldsymbol{\pi}$ has not been defined. Exceptionally there may be several or even a whole interval but, in practice, it is usually found that there is only one.

If Θ is closed and bounded and if $\boldsymbol{\pi}(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta}\ \varepsilon\ \Theta$, then the supremum of the likelihood function is attained at $\boldsymbol{\theta}_n$. If these further conditions are satisfied, $\boldsymbol{\theta}_n$ may be defined simply as a point at which the likelihood attains its supremum and the conclusions of the theorem will follow.

If $\boldsymbol{\pi}(\boldsymbol{\theta})$ is continuous but Θ is not necessarily closed and bounded, the logarithm of the likelihood is maximised at $\hat{\boldsymbol{\theta}}_n$ except when $\hat{\boldsymbol{\theta}}_n \varepsilon \overline{\Theta} - \Theta$. The conclusions of the theorem show that the probability of the latter happening tends to zero as $n \to \infty$.

We deduce, therefore, that, if the further regularity Condition (F), that $\boldsymbol{\pi}(\boldsymbol{\theta})$ is continuous at all points $\boldsymbol{\theta}$ of Θ, is satisfied, then $\hat{\boldsymbol{\theta}}_n$ may be defined as a point at which the likelihood attains its supremum and the conclusions of the theorem will still apply.

NOTE 7. Regularity Condition B is due to Rao [3], regularity Condition E, to Cramér [1]. Regularity Conditions D and D′ are believed to be new. Regularity Condition B is sufficient for the consistency of the maximum likelihood estimators.

3. Proof of the theorem. This will be broken up into a series of lemmas.

LEMMA 1. *Let $p_1, \cdots, p_r$ be any numbers $\geqq 0$ such that $\sum_{i=1}^r p_i = 1$. Then*

$$-2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] \geqq |\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta})|^2.$$

PROOF. If we interpret $p_i \ln \pi_i(\boldsymbol{\theta})$ as $-\infty$ if $\pi_i(\boldsymbol{\theta}) = 0$ and $p_i > 0$ and as 0 if $\pi_i(\boldsymbol{\theta}) = p_i = 0$ and if we interpret $p_i \ln p_i$ as 0 if $p_i = 0$, then it is easily verified that $-2[p_i \ln \pi_i - p_i \ln p_i] - 2(p_i - \pi_i) \geqq (p_i - \pi_i)^2$ if p_i or $\pi_i = 0$. When p_i and π_i are both >0 we find by application of Taylor's theorem that

$$p_i \ln p_i = \pi_i \ln \pi_i + (1 + \ln \pi_i)(p_i - \pi_i) + \tfrac{1}{2} w_i^{-1}(p_i - \pi_i)^2$$

where w_i is between p_i and π_i and therefore $\leqq 1$. Therefore

$$-2[p_i \ln \pi_i - p_i \ln p_i] - 2(p_i - \pi_i) = w_i^{-1}(p_i - \pi_i)^2 \geqq (p_i - \pi_i)^2.$$

Summation over i gives the required result.

LEMMA 2. *Let $p_1, \cdots, p_r$ be as in Lemma 1. Put*

$$y_i = (p_i - \pi_{0i})/\pi_{0i}^{\frac{1}{2}} \qquad i = 1, \cdots, r.$$

Then, as $\mathbf{p} \to \boldsymbol{\pi}_0$ and $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$

$$-2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i]$$
$$= [y - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]'[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] + o[|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2].$$

PROOF. If we apply Taylor's theorem in the same way as in Lemma 1 we get

$$-2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] = \sum_{i=1}^{r} w_i^{-1}[\pi_i(\boldsymbol{\theta}) - p_i]^2,$$

where, for each i, w_i is between $\pi_i(\boldsymbol{\theta})$ and p_i. (Because of regularity Conditions C and D, $\pi_i(\boldsymbol{\theta})$ and p_i are >0 when $|\boldsymbol{\theta} - \boldsymbol{\theta}_0|$ and $|\mathbf{p} - \boldsymbol{\pi}_0|$ are small enough.)

Now $w_i^{-1} = \pi_{0i}^{-1} + o(1)$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ and $\mathbf{p} \to \boldsymbol{\pi}_0$. Moreover, for each i, $(\pi_i - p_i)^2 \leqq 2(\pi_i - \pi_{0i})^2 + 2(p_i - \pi_{0i})^2$. Because of regularity Condition D, $\pi_i - \pi_{0i}$ is

$O|\boldsymbol{\theta} - \boldsymbol{\theta}_0|$. Therefore, $(\pi_i - p_i)^2$ is $O[|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 + |\mathbf{y}|^2]$ for each i and $\sum_{i=1}^r w_i^{-1}(\pi_i - p_i)^2 = \sum_{i=1}^r \pi_{0i}^{-1}(\pi_i - p_i)^2 + o[|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 + |\mathbf{y}|^2]$.

Now, for each i,

$$\pi_i - p_i = (\pi_{0i})^{\frac{1}{2}}\sum_{j=1}^r [a_{ij}(\theta_j - \theta_{0j})] - y_i(\pi_{0i})^{\frac{1}{2}} + o|\boldsymbol{\theta} - \boldsymbol{\theta}_0|.$$

And so

$$\begin{aligned}\sum_{j=1}^r \pi_{0i}^{-1} (\pi_i - p_i)^2 &= \sum_{i=1}^r [y_i - \sum_j a_{ij}(\theta_j - \theta_{0j}) + o\,|\boldsymbol{\theta} - \boldsymbol{\theta}_0|]^2 \\ &= \sum_{i=1}^r [y_i - \sum_j a_{ij}(\theta_j - \theta_{0j})]^2 + o(|\mathbf{y}| + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|)\,|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \\ &= \sum_{i=1}^r [y_i - \sum_j a_{ij}(\theta_j - \theta_{0j})]^2 + o\,[|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2]\end{aligned}$$

since

$$[|\mathbf{y}| + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|]|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = \tfrac{1}{2}|\mathbf{y}|^2 + \tfrac{3}{2}|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 - \tfrac{1}{2}(|\mathbf{y}| - |\boldsymbol{\theta} - \boldsymbol{\theta}_0|)^2 \leqq \tfrac{3}{2}(|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2).$$

This gives the required result.

LEMMA 3. *Let* $\boldsymbol{\theta}^*(\mathbf{p}) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} + \boldsymbol{\theta}_0$. (*Because of regularity Condition* E, $\mathbf{A}'\mathbf{A}$ *is non-singular and* $\boldsymbol{\theta}^*$ *is well-defined.*) *Then*

$$-2\sum_{i=1}^r [p_i \ln \pi_i(\theta) - p_i \ln p_i] = R + (\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p}))'\mathbf{A}'\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p})) + o\,(|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p})|^2) \text{ as } \mathbf{p} \to \boldsymbol{\pi}_0 \text{ and } \boldsymbol{\theta} \to \boldsymbol{\theta}_0,$$

where

$$R = [\mathbf{y} - \mathbf{A}(\boldsymbol{\theta}^*(\mathbf{p}) - \boldsymbol{\theta}_0)]'[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta}^*(\mathbf{p}) - \boldsymbol{\theta}_0)] = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}.$$

PROOF. The quadratic form $[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]'[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]$, obtained in Lemma 2, is at a minimum when $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ and it may easily be verified that

$$[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]'[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] = R + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{A}'\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*).$$

It follows from Lemma 2 that

$$-2\sum_{i=1}^r [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] = R + (\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p}))'\mathbf{A}'\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p})) + o(|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2).$$

But $|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 \leqq 2[|\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2 + |\boldsymbol{\theta}^* - \boldsymbol{\theta}_0|^2]$ and $\boldsymbol{\theta}^* - \boldsymbol{\theta}_0$ is $O|\mathbf{y}|$. The required result follows.

LEMMA 4. *Let* $\hat{\boldsymbol{\theta}}(\mathbf{p})$ *be any value of* θ *for which there exists a sequence* $\{\boldsymbol{\theta}_m(\mathbf{p})\}$, $m = 1, 2, \cdots$ *such that* $\boldsymbol{\theta}_m(\mathbf{p}) \to \hat{\boldsymbol{\theta}}(\mathbf{p})$ *and*

$$n \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}_m(\mathbf{p})) - p_i \ln p_i] \to \sup_{\boldsymbol{\theta} \varepsilon \Theta} n \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i].$$

Then

$$\hat{\boldsymbol{\theta}}(\mathbf{p}) - \boldsymbol{\theta}^*(\mathbf{p}) = o|\mathbf{y}| \text{ as } \mathbf{p} \to \boldsymbol{\pi}_0,$$

i.e. $\hat{\boldsymbol{\theta}}(\mathbf{p}) = \boldsymbol{\theta}_0 + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} + o|\mathbf{y}|$ as $\mathbf{p} \to \boldsymbol{\pi}_0$.

PROOF. It is sufficient to prove that, given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^*|>\epsilon|\mathbf{y}|} -2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] > -2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i]$$
$$\left(\geqq \inf_{\boldsymbol{\theta} \varepsilon \Theta} -2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i]\right)$$

whenever $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta$. For then $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*| \leqq \epsilon|\mathbf{y}|$ whenever $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta$.

Now let λ be the smallest eigenvalue of $\mathbf{A}'\mathbf{A}$. Because of assumption C, $\mathbf{A}'\mathbf{A}$ is non-singular and $\lambda > 0$.

Choose $\delta_1 > 0$ and $\eta > 0$ such that, whenever $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_1$ and $|\boldsymbol{\theta} - \boldsymbol{\theta}^*(\mathbf{p})| < \eta$ and $\boldsymbol{\theta} \varepsilon \Theta$,

$$R + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{A}'\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - [\epsilon^2 \lambda/(\epsilon^2 + 2)] (|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2)$$
$$< -2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] < R + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{A}'\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$
$$+ [\epsilon^2\lambda/(\epsilon^2 + 2)] (|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2)$$

(δ_1 and η can be so chosen because of Lemma 3).
Now, for $\epsilon|\mathbf{y}| < |\boldsymbol{\theta} - \boldsymbol{\theta}^*| < \eta$, $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_1$,

$$-2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] > R + \lambda |\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2 - \frac{\epsilon^2 \lambda}{\epsilon^2 + 2} (|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2)$$
$$> R + \left(\lambda - \frac{\epsilon^2 \lambda}{\epsilon^2 + 2}\right) \epsilon^2|\mathbf{y}|^2 - \frac{\epsilon^2 \lambda}{\epsilon^2 + 2} |\mathbf{y}|^2 = R + \frac{\epsilon^2 \lambda}{\epsilon^2 + 2} |\mathbf{y}|^2$$

while

$$-2 \sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i] < R + \frac{\epsilon^2 \lambda}{\epsilon^2 + 2} |\mathbf{y}|^2.$$

Therefore, when $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_1$, $\inf_{\eta > |\boldsymbol{\theta}-\boldsymbol{\theta}^*|>\epsilon|\mathbf{y}|} -2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] > -2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i]$.

Because of regularity condition B there exists an η' such that

$$|\boldsymbol{\pi} - \boldsymbol{\pi}_0| > \eta' \quad \text{whenever} \quad |\boldsymbol{\theta} - \boldsymbol{\theta}_0| > \tfrac{1}{2}\eta. \quad \text{Now, as} \quad \mathbf{p} \to \boldsymbol{\pi}_0, \mathbf{y} \to \mathbf{0}.$$

Therefore, by Lemma 3, $\boldsymbol{\theta}^* \rightarrow \boldsymbol{\theta}_0$ and

$$-2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i] \rightarrow 0$$

as $\mathbf{p} \rightarrow \boldsymbol{\pi}_0$.

We may choose a number $\delta_2 < \frac{1}{2}\eta'$ such that

$$|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0| < \tfrac{1}{2}\eta \quad \text{and} \quad -2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i] < \tfrac{1}{4}\eta'^2$$

whenever $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_2$. Now, when $|\boldsymbol{\theta} - \boldsymbol{\theta}^*| \geqq \eta$ and $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_2$, $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \geqq |\boldsymbol{\theta} - \boldsymbol{\theta}^*| - |\boldsymbol{\theta}^* - \boldsymbol{\theta}_0| > \frac{1}{2}\eta$ and therefore $|\boldsymbol{\pi} - \boldsymbol{\pi}_0| > \eta'$:

$$-2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] \geqq |\mathbf{p} - \boldsymbol{\pi}|^2 \text{ (by Lemma 1)}$$

$$\geqq [|\boldsymbol{\pi} - \boldsymbol{\pi}_0| - |\mathbf{p} - \boldsymbol{\pi}_0|]^2 > (\eta' - \delta_2)^2 > \tfrac{1}{4}\eta'^2.$$

Therefore, when $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta_2$,

$$\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^*| \geqq \eta} - 2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] > -2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i].$$

Finally, take $\delta = \min(\delta_1, \delta_2)$. For $|\mathbf{p} - \boldsymbol{\pi}_0| < \delta$,

$$\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^*| > \epsilon|\mathbf{y}|} - 2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}) - p_i \ln p_i] > -2\sum_{i=1}^{r} [p_i \ln \pi_i(\boldsymbol{\theta}^*) - p_i \ln p_i].$$

LEMMA 5.

$$\sum_{i=1}^{r} (1/\pi_i(\boldsymbol{\theta}))(p_i - \pi_i(\boldsymbol{\theta}))^2 = [\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]'[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] + o[|\mathbf{y}|^2 + |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2]$$

as $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}_0$ *and* $\mathbf{p} \rightarrow \boldsymbol{\pi}_0$.

PROOF. Similar to the proof of Lemma 2. The only difference is that we do not need to apply Taylor's Theorem at the beginning.

LEMMA 6.

$$\sum_{i=1}^{r} (1/\pi_i(\boldsymbol{\theta}))(p_i - \pi_i(\boldsymbol{\theta}))^2 = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} + o|\mathbf{y}|^2$$

as $\mathbf{p} \rightarrow \boldsymbol{\pi}_0$.

PROOF. By combining Lemma 4 with Lemma 5.

LEMMA 7. *Define* $\mathbf{Y}_n$ by $Y_{ni} = (P_{ni} - \pi_{0i})/(\pi_{0i})^{\frac{1}{2}}$. *Then the joint distribution of* $n^{\frac{1}{2}}Y_{n1}, \cdots, n^{\frac{1}{2}}Y_{nr}$ *tends as* $n \rightarrow \infty$ *to a (singular) multivariate normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{I}_r - \mathbf{v}\mathbf{v}'$ *where* $\mathbf{v}$ *is the column vector with elements* $(\pi_{01})^{\frac{1}{2}}, \cdots, (\pi_{0r})^{\frac{1}{2}}$.

PROOF. See Cramér [1] (method of characteristic functions).

LEMMA 8. *Let* $\mathbf{Z}$ *be a random variable having a multivariate normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{I}_r - \mathbf{v}\mathbf{v}'$. *Then* $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z}$ *and* $\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z}$ *are independently distributed, the former as* $\mathfrak{N}(\mathbf{0}, (\mathbf{A}'\mathbf{A})^{-1})$ *and the latter as* χ^2_{r-s-1}.

PROOF. (Essentially due to Cramér [1].) Let $\mathbf{U}$ be an orthogonal matrix with $\mathbf{v}$ as its last column. Let it be partitioned as $(\mathbf{U}_1 \vdots \mathbf{v})$. Let

$$\mathbf{W} = \mathbf{U}'\mathbf{Z}, \qquad \mathbf{W}_1 = \mathbf{U}'\mathbf{Z}_1 \quad \text{and} \quad W_r = \mathbf{v}'\mathbf{Z}$$

so that

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \cdots \\ W_r \end{bmatrix}.$$

Now

$$E[\mathbf{W}_1\mathbf{W}_1'] = E[\mathbf{U}_1'\mathbf{Z}\mathbf{Z}'\mathbf{U}_1] = \mathbf{U}_1'[\mathbf{I}_r - \mathbf{v}\mathbf{v}']\mathbf{U}_1 = \mathbf{I}_{r-1}$$

and

$$E[W_r^2] = E[\mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v}] = \mathbf{v}'[\mathbf{I}_r - \mathbf{v}\mathbf{v}']\mathbf{v} = 1 - 1 = 0.$$

Thus $W_r = 0$, and $\mathbf{W}_1, \cdots, W_{r-1}$ (the components of $\mathbf{W}_1$) are normally and independently distributed, each with mean 0 and variance 1.

$\mathbf{Z} = \mathbf{U}\mathbf{W} = \mathbf{U}_1\mathbf{W}_1$. Put $\mathbf{B} = \mathbf{U}_1\mathbf{A}$. Now

$$(\mathbf{v}'\mathbf{A})_j = \sum_{i=1}^{r} a_{ij}(\pi_{0i})^{\frac{1}{2}} = \sum_{i=1}^{r} \left\{ \left[\frac{\partial \pi_i}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\} = \left[\frac{\partial}{\partial \theta_j} \left\{ \sum_{i=1}^{r} \pi_i \right\} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$$

since $\sum_{i=1}^{r} \pi_i = 1$.

Thus $\mathbf{v}'\mathbf{A} = 0$. It follows that $\mathbf{B} = \mathbf{U}'\mathbf{A}$. $\mathbf{U}$ is non-singular and $\mathbf{A}$ has rank s (by regularity condition E) and therefore $\mathbf{B}$ has rank s.

$$(\mathbf{A}'\mathbf{A})^{-1} = (\mathbf{B}'\mathbf{B})^{-1}, \quad (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{W}_1$$

and

$$\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z} = \mathbf{W}_1'\mathbf{W}_1 - \mathbf{W}_1'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{W}_1.$$

By normal least squares theory $(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{W}_1$ and $\mathbf{W}_1'\mathbf{W}_1 - \mathbf{W}_1'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{W}_1$ are independently distributed, the former as $\mathfrak{N}(\mathbf{0}, (\mathbf{B}'\mathbf{B})^{-1})$ and the latter as χ^2_{r-s-1}. This completes the proof.

COMPLETION OF PROOF OF THEOREM. Lemma 4 gives

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}(\mathbf{P}_n) - \boldsymbol{\theta}_0) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'n^{\frac{1}{2}}\mathbf{Y}_n + o(n^{\frac{1}{2}}|\mathbf{Y}_n|)$$

while Lemma 6 gives

$$n\sum_{i=1}^{r} (1/\pi_i(\hat{\boldsymbol{\theta}}_n))[P_i - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2 = n[\mathbf{Y}_n'\mathbf{Y}_n - \mathbf{Y}_n'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}_n] + o(n|\mathbf{Y}_n|^2).$$

In the limit, therefore, using Lemma 7, $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and

$$n \sum_{i=1}^{r} (\pi_i(\hat{\boldsymbol{\theta}}_n))^{-1}[P_i - \pi_i(\hat{\boldsymbol{\theta}}_n)]^2$$

are distributed as $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z}$ and $\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z}$ respectively, where $\mathbf{Z}$ is $\mathfrak{N}(\mathbf{0}, \mathbf{I}_r - \mathbf{v}\mathbf{v}')$.

Lemma 8 completes the proof of the theorem.

REFERENCES

[1] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton.

[2] FISHER, R. A. (1928). On a property connecting the χ^2 measure of discrepancy with the method of maximum likelihood. *Atti Congresso Internat. Mat., Bologna*, **6** 95–100. Reprinted in *Contributions to Mathematical Statistics* by R. A. Fisher (1950). Wiley, New York.

[3] RAO, C. R. (1958). Maximum likelihood estimation for the multinomial distribution with infinite number of cells. *Sankhyā* **20** 211–218.