A Chi-Square Statistic with Random Cell Boundaries*

by

D. S. Moore

Department of Statistics
Division of Mathematical Sciences

Mimeograph Series No. 206

October 1969

---

*Research supported in part by the Office of Naval Research Contract NONR 1100(26) and the Aerospace Research Laboratories Contract AF 33(615)C1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

A Chi-Square Statistic with Random Cell Boundaries*

by

D. S. Moore

Purdue University

1. Introduction. Standard statistics of chi-square type are defined in terms of cells which are fixed prior to taking observations. Moreover, if parameters are to be estimated from the data they must be estimated by asymptotically good estimators based on the observed cell frequencies. Typically the maximum likelihood estimator (MLE) is used. Chernoff and Lehmann [1] showed that if MLE's based on the raw observations are used, the asymptotic distribution of the statistic need no longer be chi-square. In fact, if M cells are used and m parameters are estimated, the asymptotic distribution is that of

$$(1.1) \qquad \sum_{i=1}^{M-m-1} Z_i^2 + \sum_{i=M-m}^{M-1} \lambda_i Z_i^2 ,$$

where $Z_1, \ldots, Z_{m-1}$ are independent standard normal random variables and the λ's, which may depend on the parameters, lie between 0 and 1.

It would seem desirable in practice to allow the cell boundaries to be functions of the estimators used for the unknown parameters. In section 2 we show that in this case the asymptotic distribution of the chi-square statistic is that of (1.1) for rectangular cells in any number of dimensions. This result

Distribution of this document is unlimited.

* Research supported in part by the Office of Naval Research Contract NONR 1100(26) and the Aerospace Research Laboratories Contract AF 33(615)C1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

generalizes that of A. R. Roy [8] who in his unpublished thesis gave this theorem for one dimension. The extension is made possible by the use of modern random function methods, which also shorten the proof of Roy's result substantially.

The usefulness of such statistics is impaired by the dependence of the λ's on the unknown parameters. In section 3 we give two results intended to lessen this difficulty. First, in location or location-scale problems the λ's are independent of the parameters if the cell boundaries are chosen in a suitably invariant manner (see the statement of Theorem 2). This result was also obtained by Roy in the one-dimensional case. It is then shown that in any case all λ's converge to 0 as the number of cells used is appropriately increased. Thus the standard limiting distribution, chi-square with M-m-1 degrees of freedom, approximates the limiting distribution of the new statistic if many cells are used.

In section 3 several examples are presented. Perhaps the most common use of chi-square statistics with parameters estimated from the data is in testing goodness of fit to the univariate normal family. We therefore give a short table of critical points of the asymptotic distribution in this case. The table assumes equiprobable cells (a common recommendation for standard chi-square) and presents results for M = 5, 7, 9, 11, 15 and 21. Note that for 21 cells the critical points are quite close to those of the chi-square distribution with 18 degrees of freedom.

In the body of the paper we adopt certain conventions of notation. All vectors are column vectors, with prime denoting transpose. Matrices other than vectors are boldface, but vectors are not. If A is a vector, E[A] is the vector of expected values of the components of A. $\mathcal{L}(X)$ is the law or distribution of the random variable X. $N(\mu, \underset{\sim}{\Sigma})$ denotes the normal law with vector of means μ and covariance matrix $\underset{\sim}{\Sigma}$. Finally, $X_n = o_p(1)$ or $X_n \to 0(P)$ denote convergence to 0 in probability.

I am very grateful to Professor Herman Rubin for informing me of Roy's work and suggesting this generalization; and to Professor Carl de Boor for assistance with the numerical analysis required to produce Table 1.

2. Asymptotic distribution. Let $F(x|\theta)$ be a k-variate distribution function depending on an m-dimensional parameter θ which is an element of a parameter space Ω. We will assume that Ω is an open set in Euclidean m-space R_m. F will be called regular if it satisfies the assumptions

(A1) $F(x|\theta)$ has density function $f(x|\theta)$ which is continuous in $x \times \theta$ and continuously differentiable in θ.

(A2) For $i = 1,\ldots,m$

$$\frac{\partial}{\partial\theta_i} \int f(x|\theta)\, dx = \int \frac{\partial}{\partial\theta_i} f(x|\theta)\, dx.$$

(A3) The integrals

$$\int \left(\frac{\partial \log f}{\partial\theta_i} \right)^2 f(x|\theta)\, dx \qquad i=1,\ldots,m$$

are finite for all $\theta \in \Omega$ and the information matrix

$$\underset{\sim}{J} = \| J_{ij} \| \qquad i, j=1,\ldots,m$$

$$J_{ij} = \int \frac{\partial \log f}{\partial\theta_i} \frac{\partial \log f}{\partial\theta_j} f(x|\theta)\, dx$$

is positive definite for all $\theta \in \Omega$.

Let us partition the x_i-axis by functions of θ,

$$-\infty \equiv \xi_{i0}(\theta) < \xi_{i1}(\theta) < \ldots < \xi_{i,\nu_i-1}(\theta) < \xi_{i\nu_i}(\theta) \equiv \infty ,$$

for each $i = 1,\ldots,k$. We assume

(A4) $\partial\xi_{ij}/\partial\theta_s$ exist and are continuous in Ω for $i = 1,\ldots,k$, $j = 1,\ldots,\nu_i$ and $s = 1,\ldots,m$.

A partition of R_k into $M = \Pi_{i=1}^{k} \nu_i$ cells is formed by the Cartesian products of the cells of the partitions of the coordinate axes. We will index the cells of this partition by σ running from 1 to M (the particular assignment of indices to cells is immaterial). The probability $p_\sigma(\theta)$ that an observation on $F(x|\theta)$ falls in the σth cell can be expressed by a familiar difference operator. Let us define the operator Δ_σ^θ H by writing

$$p_\sigma(\theta) = \Delta_\sigma^\theta F(z|\theta).$$

The superscript specifies the value of θ at which the partitioning functions $\xi_{ij}(\theta)$ are evaluated. We assume that all $p_\sigma(\theta) > 0$ for all $\theta \in \Omega$.

Suppose that $X_1,\ldots,X_n$ is a random sample from $F(x|\theta)$ and that $\hat{\theta}_n = \hat{\theta}_n(X_1,\ldots,X_n)$ is a sequence of estimators of θ. We wish to allow the observations to choose the cells by replacing θ in $\xi_{ij}(\theta)$ by $\hat{\theta}_n$, and to consider the resulting statistic of chi-square type

$$T_n = \sum_\sigma \frac{[N_\sigma - np_\sigma(\hat{\theta}_n)]^2}{np_\sigma(\theta_n)} ,$$

where N_σ is the number of $X_1,\ldots,X_n$ falling in the σth cell.

We must require that $\hat{\theta}_n$ be an asymptotically minimum variance estimator. Let us begin by requiring only asymptotic efficiency in the sense of C. R. Rao ([7] and references therein). Suppose A(x) is the vector of logarithmic derivatives of f,

$$A(x)' = \left(\frac{\partial \log f(x|\theta)}{\partial\theta_1}, \ldots, \frac{\partial \log f(x|\theta)}{\partial\theta_m} \right) .$$

Then Rao's definition of asymptotic efficiency is

(A5) There is a nonsingular m x m matrix of constants $\underset{\sim}{B}(\theta)$, which may depend on θ, such that

$$\sqrt{n}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{B} A(X_i) + o_p(1) .$$

This says that $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically a linear transformation of the vector of derivatives of the log likelihood function. Rao has shown that (A5) implies that the information in $\hat{\theta}_n$ approaches the total information in the sample.

But (A5) does not imply asymptotic efficiency in the usual "minimum variance" sense. For if F is regular, $E[\underset{\sim}{B} A(x)] = 0$ and each component of $\underset{\sim}{B} A(x)$ has finite variance, so that

$$\mathcal{L}_\theta \{ \sqrt{n} (\hat{\theta}_n - \theta) \} \to N(0, \underset{\sim}{B} \underset{\sim}{J} \underset{\sim}{B}').$$

we will therefore require more of $\underset{\sim}{B}$ (see the statement of Theorem 1). Our requirements are satisfied in particular when $\hat{\theta}_n$ is the MLE and the usual conditions for asymptotic efficiency of the MLE hold. In that case $\hat{\theta}_n$ satisfies (A5) with $\underset{\sim}{B} = \underset{\sim}{J}^{-1}$.

THEOREM 1. Suppose (A1) - (A5) are satisfied and that the matrix $\underset{\sim}{B} + \underset{\sim}{B}' - \underset{\sim}{B} \underset{\sim}{J} \underset{\sim}{B}'$ is positive definite for all $\theta \in \Omega$. Then

$$\mathcal{L} \{ T_n \} \to \mathcal{L} \{ \sum_{i=1}^{M-m-1} Z_i^2 + \lambda_1 Z_{M-m}^2 + \cdots + \lambda_m Z_{M-1}^2 \}$$

where $Z_1, \ldots, Z_{M-1}$ are independent N(0,1) r.v.'s and the λ_j, which may depend on θ, satisfy $0 \leq \lambda_j < 1$.

PROOF. Denote the σth cell of the partition generated by $\xi_{ij}(\theta)$ by $I_\sigma(\theta)$. N_σ is the number of $X_1, \ldots, X_n$ falling in $I_\sigma(\hat{\theta}_n)$ and we let n_σ be the number of $X_1, \ldots, X_n$ falling in $I_\sigma(\theta_0)$, where θ_0 is the true parameter value. Then if $F_n(x)$ is the empiric cdf,

$$N_\sigma - np_\sigma(\hat{\theta}_n) = n [\Delta_\sigma^{\hat{\theta}} F_n(x) - \Delta_\sigma^{\hat{\theta}} F(x|\hat{\theta})]$$

$$n_\sigma - np_\sigma(\theta_0) = n [\Delta_\sigma^{\theta_0} F_n(x) - \Delta_\sigma^{\theta_0} F(x|\theta_0)] .$$

Defining the empiric cdf process $W_n(x) = \sqrt{n} [F_n(x) - F(x|\theta_0)]$, we have

$$(2.1) \quad n^{-\frac{1}{2}} [N_\sigma - np_\sigma(\hat{\theta}_n)] = n^{-\frac{1}{2}} [n_\sigma - np_\sigma(\theta_0)]$$

$$+ [\Delta_\sigma^{\hat{\theta}} W_n(x) - \Delta_\sigma^{\theta_0} W_n(x)] - \sqrt{n}\, \Delta_\sigma^{\hat{\theta}} [F(x|\hat{\theta}) - F(x|\theta_0)] .$$

Now $W_n(x)$ is a weakly convergent process. When k = 1 the limit is the familiar Brownian bridge; for k > 1 the limit may depend on F, but exists nonetheless (see Dudley [3]). Since ξ_{ij} are continuous and $\hat{\theta}_n \to \theta_o$ (P) we therefore conclude that

$$\Delta_\sigma^{\hat{\theta}} W_n(x) - \Delta_\sigma^{\theta_o} W_n(x) = o_p(1) . \tag{2.2}$$

Define the vector ∂F by

$$\partial F' = \left(\frac{\partial F(x|\theta)}{\partial \theta_1} , \dots , \frac{\partial F(x|\theta)}{\partial \theta_m} \right) \Big|_{\theta=\theta_o}$$

and agree that $\Delta_\sigma^\theta \partial F$ will mean the vector whose components are Δ_σ^θ applied to the components of ∂F. Then by Taylor's theorem, continuity of ∂F in θ, and (2.2), (2.1) becomes

$$n^{-\frac{1}{2}} [N_\sigma - np_\sigma(\hat{\theta}_n)] = n^{-\frac{1}{2}} [n_\sigma - np_\sigma(\theta_o)]$$

$$- (\Delta_\sigma^{\hat{\theta}} \partial F)' \sqrt{n} (\hat{\theta}_n - \theta_o) + o_p (1) .$$

Since $\sqrt{n} (\hat{\theta}_n - \theta)$ is $O_p(1)$ and ∂F is continuous,

$$(\Delta_\sigma^{\hat{\theta}} \partial F)' \sqrt{n} (\hat{\theta}_n - \theta_o) - (\Delta_\sigma^{\theta_o} \partial F)' \sqrt{n} (\hat{\theta}_n - \theta_o) = o_p (1). \tag{2.3}$$

Furthermore, assumption (A2) implies that

$$\Delta_\sigma^{\theta_o} [\partial F(x|\theta)/\partial \theta_s] = \frac{\partial}{\partial \theta_s} \Delta_\sigma^{\theta_o} F(x|\theta) \tag{2.4}$$

$$= \frac{\partial}{\partial \sigma_s} \int_{I_\sigma (\theta_o)} f(x|\theta) \, dx$$

$$= \int_{I_\sigma(\theta_o)} \frac{\partial f(x|\theta)}{\partial \theta_s} \, dx .$$

Define the vector $w_\sigma(\theta)$ by

$$w_\sigma(\theta)' = \left(\int_{I_\sigma(\theta_o)} \frac{\partial f(x|\theta)}{\partial\theta_1}\, dx, \ldots, \int_{I_\sigma(\theta_o)} \frac{\partial f(x|\theta)}{\partial\theta_m}\, dx \right).$$

Then by (A5), (2.3) and (2.4),

$$n^{-\frac{1}{2}} [N_\sigma - np_\sigma(\hat{\theta}_n)] = n^{-\frac{1}{2}} [n_\sigma - np_\sigma(\theta_o)]$$

$$- w_\sigma(\theta_o)' \underset{\sim}{B}\, n^{-\frac{1}{2}} \sum_{i=1}^{n} A(X_i) + o_p(1)$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} [C_\sigma(X_i) - w_\sigma' \underset{\sim}{B}\, A(X_i)] + o_p(1)$$

where the argument θ_o is assumed whenever θ is suppressed and

$$C_\sigma(x) = 1 - p_\sigma(\theta_o) \qquad x \in I_\sigma(\theta_o)$$

$$= -p_\sigma(\theta_o) \qquad x \notin I_\sigma(\theta_o) .$$

It follows by the multivariate central limit theorem that

$$\mathcal{L}_{\theta_o} \{ n^{-\frac{1}{2}} [N_\sigma - np_\sigma(\hat{\theta}_n)] : \sigma = 1,\ldots,M \} \to N(0, \underset{\sim}{\Sigma}(\theta_o))$$

where $\underset{\sim}{\Sigma}(\theta_o)$ is $M \times M$ with entries

(2.5) $$\Sigma_{\sigma\tau} = E_{\theta_o} [(C_\sigma(X) - w_\sigma \underset{\sim}{B}\, A(X)) \cdot (C_\tau(X) - w_\tau \underset{\sim}{B}\, A(X))]$$

Finally $p_\sigma(\hat{\theta}_n) / p_\sigma(\theta_o) \to 1\ (P)$, so that

$$\mathcal{L}_{\theta_o} \left\{ \frac{N_\sigma - np_\sigma(\hat{\theta}_n)}{(np_\sigma(\theta_n))^{\frac{1}{2}}} : \sigma = 1,\ldots,M \right\} \to N(0, \underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{\Sigma}\, \underset{\sim}{P}^{-\frac{1}{2}}) ,$$

where $\underset{\sim}{P}$ is the $M \times M$ matrix with entries

$$P_{\sigma\sigma} = p_\sigma(\theta_o) , \qquad \sigma = 1,\ldots,M$$

$$P_{\sigma\tau} = 0 , \qquad \sigma \neq \tau .$$

It is well known that if an M x 1 vector U satisfies $\mathcal{L}\{U\} = N(0, \underset{\sim}{C})$ and $\underset{\sim}{C}$ has characteristic roots $\lambda_1, \ldots, \lambda_M$, then $\mathcal{L}\{U'U\} = \mathcal{L}\{\Sigma_{i=1}^M \lambda_i Z_i^2\}$ where the Z_i are independent N(0,1). Suppose that $\lambda_1, \ldots, \lambda_M$ are the characteristic roots of $\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{\Sigma} \underset{\sim}{P}^{-\frac{1}{2}}$. Then

$$\mathcal{L}_{\theta_o} \{T_n\} \to \mathcal{L}\{ \Sigma_{i=1}^M \lambda_i Z_i^2 \} .$$

The remainder of the proof consists of an investigation of the λ_i, and is a straightforward generalization of Roy's work for k = 1.

Let $\underset{\sim}{W}$ be the m x M matrix with columns w_σ for $\sigma = 1, \ldots, M$. Let also p be the M x 1 vector with entries $p_\sigma(\theta_o)$. Then straightforward computation from (2.5) yields

$$\underset{\sim}{\Sigma} = \underset{\sim}{P} - pp' - \underset{\sim}{W}' \underset{\sim}{B} \underset{\sim}{W} - \underset{\sim}{W}'\underset{\sim}{B}'\underset{\sim}{W} + \underset{\sim}{W}' \underset{\sim}{B} \underset{\sim}{J} \underset{\sim}{B}' \underset{\sim}{W} = \underset{\sim}{P} - \underset{\sim}{C} ,$$

where

$$\underset{\sim}{C} = pp' + \underset{\sim}{W}' (\underset{\sim}{B} + \underset{\sim}{B}' - \underset{\sim}{B} \underset{\sim}{J} \underset{\sim}{B}') \underset{\sim}{W}.$$

It is easily seen that the λ_i are also the characteristic roots of $\underset{\sim}{P}^{-1} \underset{\sim}{\Sigma}$. All $\lambda_i \geq 0$ since $\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{\Sigma} \underset{\sim}{P}^{-\frac{1}{2}}$ is a covariance matrix. We observe that

$$\Sigma_\sigma p_\sigma = 1; \quad \Sigma_\sigma (w_\sigma)_s = 0, \quad s = 1, \ldots, m.$$

The sum of the columns of $\underset{\sim}{\Sigma}$ is therefore 0, so that at least one $\lambda_i = 0$. Denote by r $(\underset{\sim}{D})$ the rank of any matrix $\underset{\sim}{D}$. Set $q = r(\underset{\sim}{P}^{-1} \underset{\sim}{C}) = r(\underset{\sim}{C})$. Then it follows from

$$(2.6) \qquad \det [\lambda \underset{\sim}{I} - \underset{\sim}{P}^{-1} \underset{\sim}{\Sigma}] = \pm \det [(1-\lambda) \underset{\sim}{I} - \underset{\sim}{P}^{-1} \underset{\sim}{C}]$$

that exactly M-q of the $\lambda_i = 1$.

To determine q we use the assumption that $\underset{\sim}{D} = \underset{\sim}{B} + \underset{\sim}{B}' - \underset{\sim}{B} \underset{\sim}{J} \underset{\sim}{B}'$ is positive definite. Then $\underset{\sim}{W}' \underset{\sim}{D} \underset{\sim}{W}$ has rank m since W does, and since the vectors p and w_σ

are linearly independent, $q = r(\underset{\sim}{C}) = 1 + m$. Thus exactly M-m-1 $\lambda_i = 1$. $\underset{\sim}{C}$ is non-negative definite, since pp' and $\underset{\sim}{W}' \underset{\sim}{D} \underset{\sim}{W}$ are. It therefore follows from (2.6) that all $\lambda_i \leq 1$. This completes the proof.

3. Application of the statistic. The most useful case of Theorem 1 is of course that for estimators asymptotically equivalent to the MLE. In the remainder of this paper we will therefore assume that $\underset{\sim}{B} = \underset{\sim}{D} = \underset{\sim}{J}^{-1}$. The applicability of Theorem 1 is restricted by the dependence of the λ_i on θ. We first remark that this dependence vanishes in the location-scale case.

THEOREM 2. *Suppose that* (A1) - (A5) *with* $\underset{\sim}{B} = \underset{\sim}{J}^{-1}$ *hold and*

(A) $F(x|\theta) = F(x-\theta)$ *and* $\xi_{ij}(\theta) = \theta_i + a_{ij}$, a_{ij} *constants;*

or

(B) $F(x|\theta,\varphi) = F\left(\frac{x_1-\theta_1}{\varphi_1}, \ldots, \frac{x_m-\theta_m}{\varphi_m}\right)$ *and*

$\xi_{ij}(\theta,\varphi) = \theta_i + a_{ij}\varphi_i$, a_{ij} *constants.*

In either case the λ_i *do not depend on the true values of the parameters.*

PROOF. We give only a sketch of the proof, which is straightforward. In the location parameter case (A) it is easy to see that $\underset{\sim}{J}$ and $p_\sigma(\theta)$ are independent of θ and that $w_\sigma(\theta_o)$ is independent of θ_o. Since the λ_i are characteristic roots of $\underset{\sim}{P} - pp' - \underset{\sim}{W}' \underset{\sim}{J}^{-1} \underset{\sim}{W}$ they are also independent of θ_o.

In the location-scale case (B) the matrices $\underset{\sim}{J}$ and $\underset{\sim}{W}$ depend on the scale parameters $\varphi_1,\ldots,\varphi_m$. But it is easy to see that $\underset{\sim}{W}' \underset{\sim}{J}^{-1} \underset{\sim}{W}$ is independent of (θ,φ) and hence that the λ_i are also parameter-free.

We next make the important observation that as we increase the number of cells used (precisely, as we refine the partition of R_m generated by the $\xi_{ij}(\theta)$) all λ_i converge to zero. The chi-square distribution with M-m-1 degrees of freedom therefore approximates the asymptotic distribution of T_n for large M.

THEOREM 3. Suppose that (A1) - (A5) hold with $\underset{\sim}{B} = \underset{\sim}{J}^{-1}$. Suppose that $M \to \infty$ and that the $\xi_{ij}(\theta)$ are chosen so that $\xi_{i1}(\theta) \to -\infty$ and $\xi_{i,\nu_i -1} \to \infty$ for all i and all $\theta \in \Omega$, and so that $\sup_j | \xi_{ij}(\theta) - \xi_{i,j-1}(\theta) | \to 0$ for all i and all $\theta \in \Omega$. Then $\lambda_i \to 0$ for $i = 1,\dots,m$ and all $\theta \in \Omega$.

PROOF. We can write

$$\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{\Sigma} \underset{\sim}{P}^{-\frac{1}{2}} = \underset{\sim}{I} - \underset{\sim}{P}^{-\frac{1}{2}} pp' \underset{\sim}{P}^{-\frac{1}{2}} - \underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{W}' \underset{\sim}{J}^{-1} \underset{\sim}{W} \underset{\sim}{P}^{-\frac{1}{2}}$$

$$= \underset{\sim}{I} - (\underset{\sim}{P}^{-\frac{1}{2}} p)(\underset{\sim}{P}^{-\frac{1}{2}} p)' - (\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}})(\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}})'$$

$$= \underset{\sim}{I} - \underset{\sim}{A} \underset{\sim}{A}'$$

where $\underset{\sim}{A}$ is the $M \times (m+1)$ matrix

$$\underset{\sim}{A} = (\underset{\sim}{P}^{-\frac{1}{2}} p \quad \underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}}) .$$

The characteristic roots of $\underset{\sim}{P}^{-\frac{1}{2}} \underset{\sim}{\Sigma} \underset{\sim}{P}^{-\frac{1}{2}}$ are therefore one minus the characteristic roots of $\underset{\sim}{A} \underset{\sim}{A}'$. The nonzero characteristic roots of $\underset{\sim}{A} \underset{\sim}{A}'$ are the same as the nonzero characteristic roots of

$$\underset{\sim}{A}' \underset{\sim}{A} = \begin{bmatrix} p' \underset{\sim}{P}^{-1} p & p' \underset{\sim}{P}^{-1} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}} \\ \underset{\sim}{J}^{-\frac{1}{2}} \underset{\sim}{W} \underset{\sim}{P}^{-1} p & \underset{\sim}{J}^{-\frac{1}{2}} \underset{\sim}{W} \underset{\sim}{P}^{-1} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}} \end{bmatrix} .$$

But

$$p' \underset{\sim}{P}^{-1} p = \Sigma_\sigma p_\sigma = 1$$

and

$$p' \underset{\sim}{P}^{-1} \underset{\sim}{W}' = \left(\int \frac{\partial}{\partial\theta_1} f(x|\theta)dx, \dots, \int \frac{\partial}{\partial\theta_m} f(x|\theta)\, dx \right)$$

$$= (0, \dots, 0)$$

by (A2). Hence

$$(3.1) \qquad \underset{\sim}{A}' \underset{\sim}{A} = \begin{bmatrix} 1 & \underset{\sim}{0} \\ \underset{\sim}{0}' & \underset{\sim}{J}^{-\frac{1}{2}} \underset{\sim}{W} \underset{\sim}{P}^{-1} \underset{\sim}{W}' \underset{\sim}{J}^{-\frac{1}{2}} \end{bmatrix} .$$

But computation shows that the (i,j)th entry of $\underset{\sim}{W}\,\underset{\sim}{P}^{-1}\,\underset{\sim}{W}'$ is

$$(3.2) \qquad \sum_{\sigma=1}^{M} \left(\int_{I_\sigma} \frac{\partial f}{\partial \theta_i}\, dx \right) \left(\int_{I_\sigma} \frac{\partial f}{\partial \theta_j}\, dx \right) p_\sigma^{-1} .$$

Use of the mean value theorem for integrals shows that (3.2) is approximately a Riemann sum for the information integral J_{ij}. It is routine to show that as the partition is refined as in the statement of the theorem, each entry of $\underset{\sim}{W}\,\underset{\sim}{P}^{-1}\,\underset{\sim}{W}'$ converges to the corresponding entry of $\underset{\sim}{J}$. (Details of a very similar argument can be found in the proof of Theorem 2 in [5].) Thus $\underset{\sim}{A}'\,\underset{\sim}{A}$ converges to the identity matrix and $\lambda_1, \ldots, \lambda_m$ therefore converge to zero.

4. An example: the univariate normal family. The manipulations used in the proof of Theorem 3 also facilitate the computation of the λ_i, although the integrals appearing in 3.2 are often quite awkward for $k > 1$. In this section we discuss in detail the common case in which the underlying distribution is $N(\theta_1, \theta_2)$ with both parameters unknown. The discussion concludes with a short table of upper critical points of the asymptotic distribution. The methods used in calculating this table are applicable to the problem of calculating the distribution of (1.1) in general.

We will of course define our cells in this example by $\xi_i(\theta) = \theta_1 + a_i \theta_2$ and estimate the parameters by $\hat{\theta}_1 = \bar{x}$ and $\hat{\theta}_2 = s$, the sample mean and standard deviation, respectively. Then we have

$$\int_{I_\sigma} \frac{\partial f}{\partial \theta_1}\, dx = \frac{1}{\theta_2} \int_{a_{\sigma-1}}^{a_\sigma} \frac{1}{\sqrt{2\pi}}\, x\, e^{-\frac{1}{2} x^2}\, dx = w_\sigma / \theta_2$$

and

$$\int_{I_\sigma} \frac{\partial f}{\partial \theta_2}\, dx = \frac{1}{\theta_2} \int_{a_{\sigma-1}}^{a_\sigma} \frac{1}{\sqrt{2\pi}} (x^2 - 1)\, e^{-\frac{1}{2}x^2}\, dx = t_\sigma/\theta_2 ,$$

where

$$w_\sigma = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{1}{2}a_{\sigma-1}^2} - e^{-\frac{1}{2}a_\sigma^2} \right)$$

$$t_\sigma = \frac{1}{\sqrt{2\pi}} \left(a_{\sigma-1}\, e^{-\frac{1}{2}a_{\sigma-1}^2} - a_\sigma\, e^{-\frac{1}{2}a_\sigma^2} \right) .$$

Setting $p_\sigma = \Phi(a_\sigma) - \Phi(a_{\sigma-1})$ (here Φ is the standard normal cdf), it follows from (3.2) that

$$\underset{\sim}{W}\, \underset{\sim}{P}^{-1}\, \underset{\sim}{W}' = \theta_2^{-2} \begin{bmatrix} \Sigma_{\sigma=1}^{M}\, w_\sigma^2/p_\sigma & \Sigma_{\sigma=1}^{M}\, w_\sigma t_\sigma/p_\sigma \\ \Sigma_{\sigma=1}^{M}\, w_\sigma t_\sigma/p_\sigma & \Sigma_{\sigma=1}^{M}\, t_\sigma^2/p_\sigma \end{bmatrix} .$$

Since J is diagonal, it follows trivially from the proof of Theorem 3 that λ_1 and λ_2 are the characteristic roots of the matrix

$$(4.1) \qquad \begin{bmatrix} 1 - \Sigma\, w_\sigma^2/p_\sigma & 2^{-\frac{1}{2}}\, \Sigma\, w_\sigma t_\sigma/p_\sigma \\ 2^{-\frac{1}{2}}\, \Sigma\, w_\sigma t_\sigma/p_\sigma & 1 - \frac{1}{2}\, \Sigma\, t_\sigma^2/p_\sigma \end{bmatrix} .$$

The use of equiprobable cells is commonly recommended in the standard chi-square test of goodness of fit. Let us adopt that recommendation here, and agree to use an odd number of cells for computational reasons which will become apparent. The M-1 constants a_σ determining the cell boundaries can now be obtained by inverse interpolation in tables ([6], for example) of the standard normal cdf. Since the set of a_σ is symmetric about the origin, it is easy to see that

$\Sigma w_\sigma t_\sigma / p_\sigma = 0$. Thus the λ_i are simply the diagonal terms of the matrix (4.1). We will investigate the distribution of

$$(4.2) \qquad \sum_{i=1}^{M-3} Z_i^2 + \lambda_1 Z_{M-2}^2 + \lambda_2 Z_{M-1}^2 , \quad 0 < \lambda_1 < \lambda_2 < 1 ,$$

for M = 5, 7, 9, 11, 15 and 21. The values of λ_1 and λ_2 obtained for each M are given in Table 1. If $L = (M-3)/2$, the characteristic function of the random variable (4.2) is

$$\varphi(u) = (1-2iu)^{-L} [(1-2 \lambda_1 iu) (1-2 \lambda_2 iu)]^{-\frac{1}{2}} .$$

We choose M odd to obtain an integral power of $(1-2iu)^{-1}$ here. The function $\varphi(u)$ has a pole of order L at $u = -i/2$ and branch points at $u = -i/2\lambda_1$ and $u = -i/2\lambda_2$. If F(x) is the cdf of (4.2), then we have

$$(4.3) \qquad 1 - F(x) = \frac{1}{2\pi} \int \frac{e^{-iux}}{iu} \varphi(u) \, du$$

where the integral is along a line $u = t - iA$, $0 < A < \frac{1}{2}$, in the lower half-plane. This result is an easy consequence of an inversion formula of Gurland (formula (2) of [4]) which was pointed out to me by Professor Rubin.

Standard use of Cauchy's theorem shows that the right side of (4.3) is the sum of an integral around the pole and an integral around the branch points (avoiding the cut along the imaginary axis between the branch points). Both countours are described clockwise. The integral about the pole is $-2\pi i$ times the residue of the integrand at $u = -i/2$. The residue is computed in the usual way by multiplying the integrand by $(u+i/2)^L$ and differentiating L-1 times.

In computing the integral about the branch points, use is made of the fact that the radical portion of $\varphi(u)$ changes sign in crossing the cut. Standard manipulation reduces this integral to the real integral

$$\pi^{-1} \int_{1/2\lambda_2}^{1/2\lambda_1} \frac{e^{-xt}}{t} (1-2t)^{-L} [(1-2\lambda_1 t)(2\lambda_2 t-1)]^{-\frac{1}{2}} \, dt.$$

A linear change of variables transforms this into

$$(4.4) \qquad \pi^{-1} (\lambda_1\lambda_2)^{-\frac{1}{2}} \int_{-1}^{1} H(s) \, [1-s^2]^{-\frac{1}{2}} \, ds$$

where

$$H(s) = e^{-\frac{1}{2}x(As+B)} (As+B)^{-1} (1-As-B)^{-L}$$

$$A = \frac{\lambda_2 - \lambda_1}{2\lambda_1\lambda_2}$$

$$B = \frac{\lambda_1 + \lambda_2}{2\lambda_1 \lambda_2} \quad .$$

The integral (4.4) is easily evaluated by use of the Gaussian quadrature formula ([2], p. 75)

$$\int_{-1}^{1} H(s) \, [1-s^2]^{-\frac{1}{2}} \, ds = \frac{\pi}{n} \sum_{k=1}^{n} H(s_k) + \frac{\pi}{(2n)!2^{2n-1}} H^{(2n)}(\xi)$$

for some $-1 < \xi < 1$. Here $s_k = \cos((2k-1)\pi/2n)$ are the zeros of Chebyshev polynomials of the second kind. A table of any distribution of the form (4.2) can now be produced very rapidly. We present only Table 1 of upper critical points. Computation of this table required less than 4 seconds of central processor time on Purdue University's CDC 6500.

An approximation to upper critical points of $F(x)$ which is adequate for many practical purposes is obtained as follows. Neglect the contribution of the branch points, which is $\leq 10^{-4}$ for all x if $M \geq 15$ and $\leq 10^{-4}$ at the seventy-fifth percentile if $M \geq 7$. Expand the radical portion of $\varphi(u)$ about $u = -i/2$ and keep only the first term in the integral about the pole. If $F_*(x)$ is the cdf of the chi-square distribution with M-3 degrees of freedom, the resulting approximation is

$$1-F(x) = [(1-\lambda_1)\ (1-\lambda_2)\ ]^{-\frac{1}{2}}\ [1 - F_*(x)\ ]\ .$$

TABLE 1

Upper critical points x_p such that $F(x_p) = p$

M	λ_1	λ_2	p						
			0.75	0.80	0.90	0.95	0.99	0.995	0.999
5	.1030	.5317	3.559	4.023	5.442	6.844	10.077	11.464	14.683
7	.0655	.4037	5.908	6.518	8.322	10.038	13.837	15.423	19.034
9	.0470	.3259	8.241	8.961	11.055	13.007	17.234	18.971	22.885
11	.0361	.2737	10.544	11.358	13.694	15.843	20.430	22.296	26.468
15	.0242	.2077	15.084	16.052	18.792	21.270	26.463	28.547	33.158
21	.0156	.1530	21.777	22.932	26.163	29.043	34.981	37.332	42.489

REFERENCES

[1] Chernoff, Herman and Lehmann, E.L. (1954). The use of maximum likelihood estimates in χ^2 test for goodness of fit. *Ann. Math. Statist.* **25** 579-586.

[2] Davis, P. J. and Rabinowitz, P. (1967). *Numerical Integration*. Blaisdell. Waltham, Mass.

[3] Dudley, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on euclidean spaces. *Illinois Jour. Math.* **10** 109-126.

[4] Gurland, John (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19** 228-237.

[5] Moore, D. S. (1969). Asymptotically nearly efficient estimators of multivariate location parameters. *Ann. Math. Statist.* **40** 1809-1823.

[6] National Bureau of Standards (1953). *Tables of Normal Probability Functions.* Applied Mathematics Series 23.

[7] Rao, C. R. (1963). Criteria of estimation in large samples. *Sankhyā* **25** 189-206.

[8] Roy, A. R. (1956). On χ^2 statistics with variable intervals. Technical Report No. 1, Department of Statistics, Stanford University.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Purdue University	Unclassified
	2b. GROUP

3. REPORT TITLE

A Chi-Square Statistic with Random Cell Boundaries

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, October 1969

5. AUTHOR(S) (Last name, first name, initial)

Moore, David S.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October, 1969	19	8

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
NONR 1100 (26) and AF 33(615)C1244	Mimeo Series # 206
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Also supported by the Aerospace Research Laboratories, Wright Patterson AFB, Ohio	Office of Naval Research Washington, D.C.

13. ABSTRACT

Given a family $F(x|\theta)$ of sufficiently smooth distributions in k-dimensional space depending on an m-dimensional parameter θ, estimate θ by an asymptotically efficient estimator $\hat{\theta}$. Define a chi-square statistic on random cells whose boundaries depend on $\hat{\theta}$. The asymptotic distribution of such statistics is obtained, generalizing work of A. R. Roy for k=1. The limiting distribution depends on θ in general, but is independent of θ for location and location-scale families. Moreover, the distribution approaches χ^2 (M-m-1) as the number of cells $M\to\infty$. A table of upper critical points is given for several values of M for the case where $F(x|\theta)$ is univariate normal. This should be useful in testing goodness of fit to the normal family.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
chi-square statistic						
asymptotic distribution						
goodness of fit						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)   D 23551

Unclassified

Security Classification