The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 147–156

A CHI-SQUARE STATISTIC WITH RANDOM CELL BOUNDARIES[1]

By D. S. Moore

Purdue University

In testing goodness of fit to parametric families with unknown parameters, it is often desirable to allow the cell boundaries for a chi-square statistic to be functions of the estimated parameter values. Suppose M cells are used and m parameters are estimated using BAN estimators based on the sample. Then A. R. Roy and G. S. Watson showed that in the univariate case the asymptotic null distribution of the chi-square statistic is that of $\sum_1^{M-m-1} Z_l^2 + \sum_{M-m}^{M-1} \lambda_l Z_l^2$, where Z_l are independent standard normal and the constants λ_l lie between 0 and 1. They further observed that in the location-scale case the λ_l are independent of the parameters if the cell boundaries are chosen in a natural way, and that in any case all λ_l approach 0 as M is appropriately increased. We extend all of these results to the case of rectangular cells in any number of dimensions. Moreover, we give a method for numerical computation of the exact cdf of the asymptotic distribution and provide a short table of critical points for testing goodness-of-fit to the univariate normal family.

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 [2] showed that if MLE's based on the full sample 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 null distribution is that of

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

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

It is desirable in practice to allow the cell boundaries to be functions of the estimated parameter values. For univariate observations A. R. Roy [11] and G. S. Watson [12, 13, 14] showed that the asymptotic null distribution of the chi-square statistic is given by (1.1) in this case also. In Section 2 we extend this result to the case of rectangular cells in any number of dimensions. This extension is made possible by the use of modern random function methods, which also result in a shorter rigorous proof in the one-dimensional case.

Received October 30, 1969.

[1] Research supported in part by the Office of Naval Research Contract NONR 1100(26) and the Aerospace Research Laboratories Contract AF 33(615)C 1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Section 3 gives two auxiliary results. Roy and Watson both showed that in location-scale problems the λ_i are independent of the parameters if the cell boundaries are chosen in a natural way. This result extends immediately to the multivariate case. Watson also observed that in general all λ_i approach 0 as the number of cells is appropriately increased, though he gave no full proof. This result also extends to our case.

Section 4 presents a method of numerically computing the exact cdf of (1.1) in certain cases. This method is used to obtain a short table of critical points of the asymptotic distribution for testing goodness of fit to the univariate normal family. The table assumes equiprobable cells, a common recommendation discussed in this setting by Watson in Section 1 of [12].

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. $\mathscr{L}(X)$ is the law or distribution of the random variable X. $N(\mu, \Sigma)$ denotes the normal law with vector of means μ and covariance matrix $\mathbf{\Sigma}$. 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 $F(x \mid \theta)$ 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) $F(x \mid \theta)$ has density function $f(x \mid \theta)$ which is continuous in $x \times \theta$ and continuously differentiable in θ.

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

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

(A3) The integrals

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

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

$$\mathbf{J} = ||J_{ij}|| \qquad i,j = 1, \cdots, m$$

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

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

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

$$-\infty \equiv \xi_{i0}(\theta) < \xi_{i1}(\theta) < \cdots < \xi_{i,\, v_i - 1}(\theta) < \xi_{iv_i}(\theta) \equiv \infty,$$

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

(A4) $\partial \xi_{ij}/\partial \theta_s$ exist and are continuous in Ω for

$$i = 1, \cdots, k, j = 1, \cdots, v_i \quad \text{and} \quad s = 1, \cdots, m.$$

A partition of R_k into $M = \prod_{i=1}^{k} v_i$ 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 $F(x \mid \theta)$ falls in the σth cell can be expressed by a familiar difference operator. Let us define the operator $\Delta_\sigma{}^\theta H$ by writing

$$p_\sigma(\theta) = \Delta_\sigma{}^\theta F(z \mid \theta).$$

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 $\theta \in \Omega$ and that

(A5) $M > m$ and for any fixed $\theta_0 \in \Omega$ the $m \times M$ matrix W with entries

$$w_{i\sigma} = \frac{\partial}{\partial \theta_i} \Delta_\sigma{}^{\theta_0} F(z \mid \theta)$$

has rank m.

Suppose that $X_1, \cdots, X_n$ is a random sample from $F(x \mid \theta)$ and that $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$ 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

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

where N_σ is the number of $X_1, \cdots, X_n$ 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 ([10] and references therein). Suppose $A(x)$ is the vector of logarithmic derivatives of f,

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

Then Rao's definition of asymptotic efficiency is

(A6) There is a nonsingular $m \times m$ matrix of constants $\mathbf{B}(\theta)$, which may depend on θ, such that

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

This says that $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ is asymptotically a linear transformation of the vector of derivatives of the log likelihood function. Rao has shown that (A6) implies that the information in $\hat{\theta}_n$ approaches the total information in the sample. But (A6) does not imply asymptotic efficiency in the usual "minimum variance" sense. For if F is regular, $E[\mathbf{B}A(x)] = 0$ and each component of $\mathbf{B}A(x)$ has finite variance, so that

$$\mathscr{L}_\theta\{n^{\frac{1}{2}}(\hat{\theta}_n-\theta)\} \to N(0, \mathbf{BJB}'),$$

we will therefore require more of $\mathbf{B}$ (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 (A6) with $\mathbf{B} = \mathbf{J}^{-1}$.

THEOREM 1. *Suppose* (A1)–(A6) *are satisfied and that the matrix* $\mathbf{B}+\mathbf{B}'-\mathbf{BJB}'$ *is positive definite for all* $\theta \in \Omega$. *Then*

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

where $Z_1, \cdots, Z_{M-1}$ *are independent* $N(0, 1)$ rv'*s and the* λ_j, *which may depend on* θ, *satisfy* $0 \leqq \lambda_j < 1$.

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

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

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

Defining the empiric cdf process $W_n(x) = n^{\frac{1}{2}}[F_n(x) - F(x \mid \theta_0)]$, we have

$$n^{-\frac{1}{2}}[N_\sigma - np_\sigma(\hat{\theta}_n)] = n^{-\frac{1}{2}}[n_\sigma - np_\sigma(\theta_0)] + [\Delta_\sigma{}^{\hat{\theta}} W_n(x) - \Delta_\sigma{}^{\theta_0} W_n(x)] - n^{\frac{1}{2}}\Delta_\sigma{}^{\hat{\theta}}[F(x \mid \hat{\theta}) - F(x \mid \theta_0)]. \quad (2.1)$$

We wish to show that

$$\Delta_\sigma{}^{\hat{\theta}} W_n(x) - \Delta_\sigma{}^{\theta_0} W_n(x) = o_p(1). \quad (2.2)$$

Since ξ_{ij} are continuous and $\hat{\theta}_n \to \theta_0(P)$, this will follow if $\eta_n \to c(P)$ implies $W_n(\eta_n) - W_n(c) = o_p(1)$. Define a continuous function H mapping R_k onto the unit cube in such a way that the rv's $Y_i = H(X_i)$ have uniform marginal distributions in each direction. If $U(u)$ is the common cdf of $Y_1, \cdots, Y_n$ and $U_n(u)$ is the empiric cdf, then the empiric process $W_n^*(u) = n^{\frac{1}{2}}(U_n(u) - U(u))$ satisfies $W_n(x) = W_n^*(H(x))$. The path functions of W_n^* fall in the space D_k of functions on the unit cube having only jump discontinuities. It is possible to define on D_k a metric of Skorohod type such that D_k is complete and separable, convergence to a continuous limiting function is uniform, and W_n^* converges weakly to a Gaussian process W^* such that $W^*(u)$ is continuous w.p. 1. These results are contained in Neuhaus [8] and the weak convergence result is in fact a consequence of the stronger results of Dudley [4].

Now $\eta_n \to c(P)$ implies $H(\eta_n) \to H(c)(P)$. This with $W_n^* \to W^*$ weakly implies by Theorem 4.4 of [1] that the joint distribution of $(W_n^*, H(\eta_n))$ converges weakly on $D_k \times R_k$ to that of $(W^*, H(c))$. Since $W^*(u)$ is continuous w.p. 1 and convergence to a continuous limit is uniform, the function $\varphi: D_k \times R_k \to R_1$ defined by $\varphi(f, a) = f(a) - f(H(c))$ is continuous w.p. 1 with respect to the distribution of $(W^*, H(c))$. Thus by Theorem 5.1 of [1], $\varphi(W_n^*, H(\eta_n)) = W_n^*(H(\eta_n)) - W_n^*(H(c))$ converges in law to $\varphi(W^*, H(c))$. Since $\varphi(W^*, H(c)) \equiv 0$ this is just the statement $W_n(\eta_n) - W_n(c) = W_n^*(H(\eta_n)) - W_n^*(H(c)) \to O(P)$.

Define the vector ∂F by

$$\partial F' = \left(\frac{\partial F(x\,|\,\theta)}{\partial \theta_1}, \cdots, \frac{\partial F(x\,|\,\theta)}{\partial \theta_m}\right)\Bigg|_{\theta=\theta_0}$$

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

$$n^{-\frac{1}{2}}[N_\sigma - np_\sigma(\hat{\theta}_n)] = n^{-\frac{1}{2}}[n_\sigma - np_\sigma(\theta_0)] - (\Delta_\sigma^\theta \partial F)' n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) + o_p(1).$$

Since $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ is $O_p(1)$ and ∂F is continuous,

$$(2.3) \qquad (\Delta_\sigma^\theta \partial F)' n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) - (\Delta_\sigma^{\theta_0} \partial F)' n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = o_p(1).$$

Furthermore, assumption (A2) implies that

$$\begin{aligned}
\Delta_\sigma^{\theta_0}[\partial F(x\,|\,\theta)/\partial\theta_s] &= \frac{\partial}{\partial\theta_s}\Delta_\sigma^{\theta_0} F(x\,|\,\theta) \\
(2.4) \qquad &= \frac{\partial}{\partial\theta_s}\int_{I_\sigma(\theta_0)} f(x\,|\,\theta)\,dx \\
&= \int_{I_\sigma(\theta_0)} \frac{\partial f(x\,|\,\theta)}{\partial\theta_s}\,dx.
\end{aligned}$$

Define the vector $w_\sigma(\theta)$ as the σth column of the matrix W in (A5):

$$w_\sigma(\theta)' = \left(\int_{I_\sigma(\theta_0)} \frac{\partial f(x\,|\,\theta)}{\partial\theta_1}\,dx, \cdots, \int_{I_\sigma(\theta_0)} \frac{\partial f(x\,|\,\theta)}{\partial\theta_m}\,dx\right).$$

Then by (A6), (2.3) and (2.4),

$$\begin{aligned}
n^{-\frac{1}{2}}[N_\sigma - np_\sigma(\hat{\theta}_n)] &= n^{-\frac{1}{2}}[n_\sigma - np_\sigma(\theta_0)] - w_\sigma(\theta_0)'\mathbf{B}n^{-\frac{1}{2}}\textstyle\sum_{i=1}^n A(X_i) + o_p(1) \\
&= n^{-\frac{1}{2}}\textstyle\sum_{i=1}^n [C_\sigma(X_i) - w_\sigma'\mathbf{B}A(X_i)] + o_p(1)
\end{aligned}$$

where the argument θ_0 is assumed whenever θ is suppressed and

$$\begin{aligned}
C_\sigma(x) &= 1 - p_\sigma(\theta_0) \qquad & x \in I_\sigma(\theta_0) \\
&= -p_\sigma(\theta_0) \qquad & x \notin I_\sigma(\theta_0).
\end{aligned}$$

It follows by the multivariate central limit theorem that

$$\mathscr{L}_{\theta_0}\{n^{-\frac{1}{2}}[N_\sigma - np_\sigma(\hat{\theta}_n)]:\sigma = 1,\cdots,M\} \to N(0,\mathbf{\Sigma}(\theta_0))$$

where $\mathbf{\Sigma}(\theta_0)$ is $M \times M$ with entries

$$\Sigma_{\sigma\tau} = E_{\theta_0}[(C_\sigma(X) - w_\sigma \mathbf{B}A(X))\cdot(C_\tau(X) - w_\tau \mathbf{B}A(X))]. \tag{2.5}$$

Finally $p_\sigma(\hat{\theta}_n)/p_\sigma(\theta_0) \to 1(P)$, so that

$$\mathscr{L}_{\theta_0}\left\{\frac{N_\sigma - np_\sigma(\hat{\theta}_n)}{(np_\sigma(\theta_n))^{\frac{1}{2}}}:\sigma = 1,\cdots,M\right\} \to N(0,\mathbf{P}^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{P}^{-\frac{1}{2}}),$$

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

$$P_{\sigma\sigma} = p_\sigma(\theta_0), \qquad \sigma = 1,\cdots,M$$

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

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

$$\mathscr{L}_{\theta_0}\{T_n\} \to \mathscr{L}\{\textstyle\sum_{i=1}^M \lambda_i Z_i^2\}.$$

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 $\mathbf{W}$ be the $m \times M$ matrix with columns w_σ for $\sigma = 1, \cdots, M$. Let also p be the $M \times 1$ vector with entries $p_\sigma(\theta_0)$. Then straightforward computation from (2.5) yields

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

where

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

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

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

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

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

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

To determine q we use the assumption that $\mathbf{D} = \mathbf{B} + \mathbf{B}' - \mathbf{B}\mathbf{J}\mathbf{B}'$ is positive definite. Then $\mathbf{W}'\mathbf{D}\mathbf{W}$ has rank m since $\mathbf{W}$ does, and since the vectors p and w_σ are linearly independent, $q = r(\mathbf{C}) = 1 + m$. Thus exactly $M - m - 1$ $\lambda_i = 1$. $\mathbf{C}$ is nonnegative definite, since pp' and $\mathbf{W}'\mathbf{D}\mathbf{W}$ are. It therefore follows from (2.6) that all $\lambda_i \leqq 1$. 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 $\mathbf{B} = \mathbf{D} = \mathbf{J}^{-1}$. 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)–(A6) *with* $\mathbf{B} = \mathbf{J}^{-1}$ *hold and*

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

or

(B) $F(x \mid \theta, \varphi) = F\left(\dfrac{x_1-\theta_1}{\varphi_1}, \cdots, \dfrac{x_m-\theta_m}{\varphi_m}\right)$ *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 $\mathbf{J}$ and $p_\sigma(\theta)$ are independent of θ and that $w_\sigma(\theta_0)$ is independent of θ_0. Since the λ_i are characteristic roots of $\mathbf{P} - pp' - \mathbf{W}'\mathbf{J}^{-1}\mathbf{W}$ they are also independent of θ_0.

In the location-scale case (B) the matrices $\mathbf{J}$ and $\mathbf{W}$ depend on the scale parameters $\varphi_1, \cdots, \varphi_m$. But it is easy to see that $\mathbf{W}'\mathbf{J}^{-1}\mathbf{W}$ 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)–(A6) *hold with* $\mathbf{B} = \mathbf{J}^{-1}$. *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* $i = 1, \cdots, m$ *and all* $\theta \in \Omega$.

PROOF. By matrix manipulations parallel to those performed by Watson ([13] pages 53–54) we obtain that the nonzero λ_i are one minus the nonzero characteristic roots of

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0}' & \mathbf{J}^{-\frac{1}{2}}\mathbf{W}\mathbf{P}^{-1}\mathbf{W}'\mathbf{J}^{-\frac{1}{2}} \end{bmatrix}. \tag{3.1}$$

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

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

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 $\mathbf{WP}^{-1}\mathbf{W}'$ converges to the corresponding entry of $\mathbf{J}$. (Details of a very similar argument can be found in the proof of Theorem 2 in [6].) Thus (3.1) converges to the identity matrix and $\lambda_1, \cdots, \lambda_m$ therefore converge to zero.

4. Asymptotic cdf in the univariate normal case. It is desirable in practice to know the critical points of the distribution of (1.1), especially in the common case in which the underlying distribution is $N(\theta_1, \theta_2)$ with both parameters unknown. Watson ([12] Section 5) gives an approximate method of calculation. We show here how to effectively compute values of the cdf of (1.1) and use this method to compile a short table of exact upper critical points.

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 the sample mean and sample standard deviation. We adopt the common recommendation of equally probable cells, and choose 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 ([7], for example) of the standard normal cdf. We then compute λ_1 and λ_2 from the simple expressions given on page 345 of [12].

We will investigate the distribution of

$$\sum_{i=1}^{M-3} Z_i^2 + \lambda_1 Z_{M-2}^2 + \lambda_2 Z_{M-1}^2, \qquad 0 < \lambda_1 < \lambda_2 < 1 \tag{4.1}$$

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 (4.1) 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 (4.1), then we have

$$1 - F(x) = \frac{1}{2\pi}\int \frac{e^{-iux}}{iu}\varphi(u)\,du \tag{4.2}$$

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 [5]) which was pointed out to me by Professor Rubin.

Standard use of Cauchy's theorem shows that the right side of (4.2) 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 contours 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

$$\pi^{-1}(\lambda_1\lambda_2)^{-\frac{1}{2}}\int_{-1}^{1}H(s)[1-s^2]^{-\frac{1}{2}}\,ds \tag{4.3}$$

where

$$H(s)=\exp\left[-\tfrac{1}{2}x(As+B)\right](As+B)^{-1}(1-As-B)^{-L}$$

$$A=\frac{\lambda_2-\lambda_1}{2\lambda_1\lambda_2},\qquad B=\frac{\lambda_1+\lambda_2}{2\lambda_1\lambda_2}.$$

The integral (4.3) is easily evaluated by use of the Gaussian quadrature formula ([3], page 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 (4.1) 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.

TABLE 1

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

			p						
M	λ_1	λ_2	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] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[2] 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.

[3] DAVIS, P. J. and RABINOWITZ, P. (1967). *Numerical Integration*. Blaisdell. Waltham, Mass.

[4] DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

[5] GURLAND, JOHN (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19** 228–237.

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

[7] NATIONAL BUREAU OF STANDARDS (1953). *Tables of Normal Probability Functions.* Applied Mathematics Series 23.

[8] NEUHAUS, GEORG. (1969). On weak convergence of stochastic processes with multi-dimensional time parameter. Technical report, Institut für Mathematische Statistik, University of Münster.

[9] PYKE, RONALD (1968). Applications of almost surely convergent constructions of weakly convergent processes. *International Symposium on Probability Theory and Information Theory at McMaster University.* Springer-Verlag, New York.

[10] RAO, C. R. (1963). Criteria of estimation in large samples. *Sankhyā* **25** 189–206.

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

[12] WATSON, G. S. (1957). The χ^2 goodness-of-fit test for normal distributions. *Biometrika* **44** 336–348.

[13] WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc. Ser. B* **20** 44–61.

[14] WATSON, G. S. (1959). Some recent results in χ^2 goodness-of-fit tests. *Biometrics* **15** 440–468.