The Annals of Statistics
1975, Vol. 3, No. 3, 599-616

UNIFIED LARGE-SAMPLE THEORY OF GENERAL CHI-SQUARED STATISTICS FOR TESTS OF FIT[1]

BY DAVID S. MOORE AND M. C. SPRUILL

Purdue University

We present a unified large-sample theory of general chi-squared tests of fit under composite null hypotheses and Pitman alternatives. The statistics are quadratic forms in the standardized cell frequencies, and we allow random cells, k-variate observations from not necessarily continuous distributions, and quite general estimates of unknown parameters. Generalizations of published results on a number of specific chi-squared tests follow.

1. Introduction. The original Pearson χ^2 test for goodness-of-fit to a fixed distribution is based on observed cell frequencies in a set of fixed cells. In practice we are often interested in testing the composite null hypothesis that the observations come from a parametric family $F(x \mid \theta)$ of distributions. To obtain estimated cell frequencies for a χ^2 test we must then estimate the parameter θ (which is often a vector). If the estimator used is the maximum likelihood estimator (MLE) of θ based on the cell frequencies (or another asymptotically equivalent estimator), the resulting test is the Pearson–Fisher χ^2 (see Cramér (1946), Section 30.3). If instead the MLE based on the original data (or asymptotically equivalent estimator) is used, the resulting Chernoff–Lehmann (1954) χ^2 does not have a limiting χ^2 null distribution. What is worse, the limiting null distribution usually depends on the (unknown) true value of θ.

This unpleasant situation (and attempts to model the procedures often followed by experimenters) leads to the use of cells which are themselves functions of the data. We call these *random cells*. Watson (1959) and Roy (1956) independently studied the random cell version of the Chernoff–Lehmann χ^2 statistic and observed that if $F(x \mid \theta)$ is a location-scale family and the cells are chosen in the proper manner, then the limiting null distribution does not depend on the true θ. Random cells can of course be employed in the Pearson–Fisher statistic as well: after obtaining the cells we "forget" that they are functions of the data and calculate the usual MLE based on cell frequencies. Witting (1959) studied the case in which cell boundaries are sample quantiles and multivariate generalizations. The univariate case of this random-cell χ^2 statistic was further investigated by Bofinger (1973).

Received December 1973; revised July 1974.

[1] Research sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-72-2350. The United States Government is authorzied to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

AMS 1970 *subject classifications*. Primary 62E20; Secondary 62G10.

Key words and phrases. Chi-squared tests, goodness of fit, limiting distributions.

Large sample theory for fixed-cell χ^2 statistics is based on the multinomial distribution and does not depend on the dimension of the observations. Witting, in his particular case, used the theory of statistically equivalent blocks to give essentially dimension-free proofs. But rigorous proofs for other random-cell cases are more difficult. The essential technique is to show that the difference between the random-cell statistic and a fixed-cell statistic of similar form approaches zero in probability as the sample size increases. This was first done by Roy (1956). Chibisov (1971) and Moore (1971) realized independently that this fact becomes routine when the tools of weak convergence of probability measures are applied. Chibisov obtained the limiting distributions of the random-cell versions of the Pearson–Fisher and Chernoff–Lehmann statistics under the null hypothesis and also under sequences of Pitman alternatives. He gave proofs only for univariate observations. Weak convergence results for empiric processes in R^k must be used if the observations are k-dimensional. This was done by Moore (1970, 1971) who found the limiting null distributions of these random-cell tests for rectangular cells in R^k.

Several other authors have proposed "non-standard" χ^2 goodness-of-fit statistics. Kambhampati (1971) gave a quadratic form (not the sum of squares) of observed minus expected cell frequencies having the property that its limiting null distribution is χ^2 when MLE's from the original data are used. Chase (1972) studied χ^2 statistics of Pearson–Fisher and Chernoff–Lehmann type when the estimator of θ comes from a sample independent of the sample being tested for fit. Murty and Gafarian (1970) studied the same tests when the estimator is based on both the sample being tested and an independent sample. These three papers study only the fixed-cell versions of their tests and obtain only the limiting null distributions.

In this paper we follow the approach of Moore (1971) to give a large-sample theory for a general class of χ^2 statistics (*n*nd quadratic forms in the standardized cell frequencies) which includes all statistics mentioned above and many other variations as well. The theory allows multivariate observations and quite general estimators of θ, and considers sequences of local alternatives as well as the composite null hypothesis. We further include some cases in which $F(x \mid \theta)$ is not continuous in x, in particular the case in which the set of possible discontinuities is known and fixed. This covers many problems of interest, such as testing the fit of integer-valued observations to a specific family such as the Poisson. Example 4.1 shows that some other discontinuous $F(x \mid \theta)$ are also included.

The set of χ^2 statistics studied here is limited by the method of proof, which establishes that the difference between a random-cell statistic and a corresponding fixed-cell statistic converges to zero in probability. If the number of cells grows with the number n of observations at a rate faster than $O(n^{\frac{1}{2}})$, this method fails and the large-sample theory takes a different form. Thus, for example, Kempthorne's (1968) χ^2 statistic cannot be studied in our framework, and has a normal limiting null distribution rather than the linear combination of χ^2 variates which is the limiting law of all statistics considered here.

Section 2 introduces necessary notation and assumptions. Section 3 presents a useful association between an arbitrary distribution on R^k and a continuous distribution on the unit cube with all univariate marginals uniform. Section 4 presents the general theory, while Section 5 obtains specific results for one-sample tests (Pearson–Fisher, Chernoff–Lehmann and Kambhampati) and Section 6 treats two-sample tests (Chase and Murty–Gafarian). Knowing the limiting distributions of these statistics under Pitman alternatives makes possible some power comparisons. These are discussed in Section 7.

2. Notation, definitions and assumptions. We observe $Y_1, Y_2, \cdots$ independent R^k-valued random variables with df $F(x \mid \theta, \eta)$. The parameter θ ranges over an open set Ω_1 in R^m, while η ranges over a neighborhood of a point η_0 in R^p. We write

$$F(x \mid \theta, \eta_0) = F(x \mid \theta)$$

so that the composite null hypothesis that the Y_i have a df in the family $F(x \mid \theta)$ becomes

$$H_0 : \eta = \eta_0 .$$

We will explore the large-sample behavior of tests for H_0 under the sequence of parameter values (θ_0, η_n) where $\theta_0 \in \Omega_1$ and $\eta_n = \eta_0 + n^{-\frac{1}{2}}\gamma$ for fixed $\gamma \in R^p$. H_0 is the special case $\gamma = 0$. (For sufficiently large n, η_n is in the neighborhood of η_0 for which $F(x \mid \theta, \eta)$ is defined. We will not constantly repeat the qualification "for sufficiently large n.") Many common alternatives are covered by this model, for example the "contamination alternative" under which

$$F(x \mid \theta, \eta) = (1 - \eta)F(x \mid \theta) + \eta H(x)$$

for $0 \leqq \eta \leqq 1$ and H a fixed df. Our model is that used by Chibisov (1971) in his study of the univariate Pearson–Fisher and Chernoff–Lehmann statistics and by Durbin (1973) in his study of the empiric process.

The cells for our χ^2 tests are rectangles in R^k with edges parallel to the coordinate axes. They are functions of a variable φ defined on an open set Ω_2 in R^r. The resulting cells are denoted by $I_\sigma(\varphi)$, $\sigma = 1, 2, \cdots, M$ and are understood to be closed to the "north and east." (Usually $r = m$ and in the actual test statistics φ is replaced by an estimator of θ. Thus a common choice of cells in testing fit to the univariate normal family uses cell boundaries of the form $\bar{Y} + a_i s_Y$ for constants a_i. In this case $r = m = 2$ and φ takes values in $\Omega_2 = \{(x, y) : -\infty < x < \infty, y > 0\}$. Our formulation allows in addition boundaries $\bar{Y} + a_i$ with $r = 1$ or cells bounded by sample α_i-quantiles, in which case $r = M - 1$.)

In forming χ^2 statistics the unknown parameter θ is estimated by $\theta_n = \theta_n(Y_1, \cdots, Y_n)$ and the cells are chosen by $\varphi_n = \varphi_n(Y_1, \cdots, Y_n)$. We will assume that under (θ_0, η_n), $\varphi_n - \varphi_0 = o_p(1)$ for some $\varphi_0 \in \Omega_2$ and $\theta_n - \theta_0 = o_p(1)$. *We* *will suppress arguments θ, φ, η whenever they take the values θ_0, φ_0, η_0 respectively.* In particular, expected values and derivatives not otherwise identified are computed under (θ_0, η_0).

The number of $Y_1, \cdots, Y_n$ falling in the cell $I_\sigma(\varphi)$ will be denoted by $N_{n\sigma}(\varphi)$. The cell probability for this cell under (θ, η) is

$$p_\sigma(\theta, \eta, \varphi) = \int_{I_\sigma(\varphi)} dF(x \mid \theta, \eta) .$$

Thus the "estimated cell probability" used in χ^2 statistics is $p_\sigma(\theta_n, \varphi_n)$. The standardized cell frequencies are

$$v_{n\sigma}(\theta, \eta, \varphi) = \frac{N_{n\sigma}(\varphi) - np_\sigma(\theta, \eta, \varphi)}{[np_\sigma(\theta, \eta, \varphi)]^{\frac{1}{2}}}$$

which is the σth component of an M-vector $V_n(\theta, \eta, \varphi)$.

If $K(\theta, \varphi)$ is a *n*nd symmetric $M \times M$ matrix for each (θ, φ) in $\Omega_1 \times \Omega_2$, a general χ^2 statistic has the form

$$T_n = V_n'(\theta_n, \varphi_n)K(\theta_n, \varphi_n)V_n(\theta_n, \varphi_n) . \tag{2.1}$$

Standard χ^2 statistics have this form with $K(\theta_n, \varphi_n) \equiv I_M$ (the identity matrix) so that the statistic is $||V_n(\theta_n, \varphi_n||^2$. The Pearson–Fisher statistic uses $\theta_n = \bar{\theta}_n$ where $\bar{\theta}_n$ maximizes

$$\sum_{\sigma=1}^{M} N_{n\sigma}(\varphi_n) \log p_\sigma(\theta, \varphi_n) .$$

The Chernoff–Lehmann statistic uses $\theta_n = \hat{\theta}_n$ where $\hat{\theta}_n$ maximizes

$$\sum_{i=1}^{n} \log f(Y_i \mid \theta)$$

and $f(x \mid \theta)$ is the pdf of the df $F(x \mid \theta)$ with respect to a σ-finite dominating measure which does not depend on θ. We will see that the choice of φ_n affects the limiting distributions only through its limit in probability φ_0, as was observed by Watson (1959) and Chibisov (1971) for the cases they studied.

Our general assumptions—not all of which will always be invoked—are as follows. Recall that $\eta_n = \eta_0 + n^{-\frac{1}{2}}\gamma$.

A1. Under (θ_0, η_n), $\theta_n - \theta_0 = O_p(n^{-\frac{1}{2}})$ and $\varphi_n - \varphi_0 = o_p(1)$. Every vertex $x(\varphi)$ of every cell $I_\sigma(\varphi)$ is a continuous R^k-valued function of φ in a neighborhood of φ_0.

A2. For each σ, $p_\sigma(\theta, \eta, \varphi)$ is continuous in (θ, η, φ) and continuously differentiable in (θ, η) in a neighborhood of $(\theta_0, \eta_0, \varphi_0)$. Moreover, $\sum_1^M p_\sigma = 1$ and $p_\sigma > 0$ for each σ.

A3. $F(x) = F(x \mid \theta_0, \eta_0)$ is continuous at every vertex $x(\varphi_0)$ of every cell $I_\sigma(\varphi_0)$. As $n \to \infty$, $\sup_x |F(x \mid \eta_n) - F(x)| \to 0$.

A4. $K(\theta, \varphi) = S(\theta, \varphi)S(\theta, \varphi)'$ for an $M \times M$ matrix $S(\theta, \varphi)$ with entries continuous in (θ, φ) at (θ_0, φ_0)

A5. Under (θ_0, η_n)

$$n^{\frac{1}{2}}(\theta_n - \theta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} h(Y_i, \eta_n) + A\gamma + o_p(1)$$

for some $m \times p$ matrix A and measurable function $h(x, \eta)$ from $R^k \times R^p$ to R^m

satisfying

$$E[h(Y, \eta_n) \mid (\theta_0, \eta_n)] = 0$$
$$E[h(Y, \eta_n)h(Y, \eta_n)' \mid (\theta_0, \eta_n)] = L(\eta_n)$$

where $L(\eta_n)$ is a nnd $m \times m$ matrix converging to the finite nnd matrix $L = E[h(Y)h(Y)']$ as $n \to \infty$.

A6. The df's $F(x \mid \eta)$ possess pdf's $f(x \mid \eta)$ with respect to a σ-finite dominating measure ν. As $n \to \infty$, $f(x \mid \eta_n) \to f(x \mid \eta_0)$ and $h(y, \eta_n) \to h(y)$ a.e. (ν).

Some comments on these assumptions are in order. A2 is a familiar assumption slightly generalized to include η. Since θ_0 and φ_0 are unknown, in practice this and other regularity assumptions must usually be assumed to hold for all θ in Ω_1 and φ in Ω_2. An exception appears in Example 4.1. A3 is used to handle the alternative case. In the null case it requires only that each vertex $x(\varphi_0)$ be a continuity point of $F(x \mid \theta_0)$. The continuity of $F(x \mid \eta)$ assumed in A3 allows $F(x \mid \theta, \eta)$ which are continuous in x or which have mass points fixed for all (θ, η), and some other cases as well.

A5 specifies the asymptotic behavior of the estimator θ_n in a form used by Durbin (Assumption A1 in [12]), who explains its motivation. In particular, the MLE's $\hat{\theta}_n$ and $\bar{\theta}_n$ both satisfy A5 in most cases. In the case of $\hat{\theta}_n$, arguments of Davidson and Lever (1970) can be used to show that when strong regularity conditions hold, we have under (θ_0, η_n)

$$(2.2) \qquad n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = n^{-\frac{1}{2}} \sum_{i=1}^n J^{-1} \frac{\partial \log f(Y_i \mid \eta_n)}{\partial \theta} + J^{-1}J_{12}\gamma + o_p(1) .$$

Here J is the information matrix for $F(x \mid \theta)$ at θ_0,

$$J = E\left[\left(\frac{\partial \log f}{\partial \theta}\right)\left(\frac{\partial \log f}{\partial \theta}\right)'\right],$$

J_{12} is the $m \times p$ matrix

$$J_{12} = E\left[\left(\frac{\partial \log f}{\partial \theta}\right)\left(\frac{\partial \log f}{\partial \eta}\right)'\right],$$

and we have used the convention that $\partial g/\partial \theta$ is the m-vector of derivatives with respect to the components of θ.

The representation (2.2) for $\hat{\theta}_n$ holds in many less regular cases as well. For example, if $F(x \mid \theta, \eta)$ has pdf

$$f(x \mid \theta, \eta) = (1 - \eta)\tfrac{1}{2}e^{-|x-\theta|} + \eta k(x) \qquad \eta_0 = 0$$

for a pdf $k(x)$, the MLE $\hat{\theta}_n$ for the double exponential distribution is the median. Even though $\partial \log f/\partial \theta$ is a step function, calculation shows that (2.2) holds.

In the case of the MLE $\bar{\theta}_n$ based on cell frequencies, Watson (1959) observed that under suitable regularity conditions these estimators in the random-cell case have the same limiting behavior as in the fixed-cell case under the null hypothesis, namely

$$(2.3) \qquad n^{\frac{1}{2}}(\bar{\theta}_n - \theta_0) = (B'B)^{-1}B'V_n + o_p(1)$$

where the $M \times m$ matrix B has (i, j)th entry

$$p_i^{-\frac{1}{2}} \frac{\partial p_i}{\partial \theta_j} .$$

There is little difficulty in going on to show that in the presence of sufficiently strong regularity conditions we have a form for $\bar{\theta}_n$ which satisfies A5, namely that under (θ_0, η_n)

$$n^{\frac{1}{2}}(\bar{\theta}_n - \theta_0) = (B'B)^{-1}B'V_n(\eta_n) + (B'B)^{-1}B'B_{12}\gamma + o_p(1) , \quad (2.4)$$

where the $M \times p$ matrix B_{12} has (i, j)th entry

$$p_i^{-\frac{1}{2}} \frac{\partial p_i}{\partial \eta_j} .$$

The representation (2.4) also holds in many cases in which the regularity conditions needed to derive it do not hold.

Assumption A6 is used to obtain the specific behavior under (θ_0, η_n) of statistics T_n with estimators θ_n satisfying A5. It is not needed when limiting null distributions are being studied. Although A6 is considerably more restrictive than our other assumptions, it is less restrictive that the conditions on $f(x \mid \eta)$ used by Chibisov, the only other author to obtain limiting alternative distributions for general random-cell tests.

In general, the assumptions given become less restrictive if only the null case is of interest. We have usually not commented separately on the null case, but expect readers to set $F(x \mid \theta, \eta) \equiv F(x \mid \theta)$ and $\gamma = 0$ to obtain null case results.

3. Preliminary results. The basic tool used to relate random-cell chi-squared tests to fixed-cell tests is weak convergence of empiric df processes on the unit cube E^k of R^k. These processes are considered as probability distributions on the space D_k of functions on E^k whose only discontinuities are "jumps". The set C_k of continuous functions on E^k is a subset of D_k. The theory of weak convergence of measures on D_k equipped with a Skorohod-type topology $\mathscr{D}_k$ is given by Bickel and Wichura (1971) and Neuhaus (1971). Suppose $G(x \mid \theta, \eta)$ is a family of continuous df's on E^k having all univariate marginal df's uniform on [0, 1] and such that $G(x \mid \theta_0, \eta)$ is continuous in η near η_0. Let G_n be the empiric df after n observations from G and define the process

$$y_n(x) = n^{\frac{1}{2}}\{G_n(x) - G(x \mid \theta_0, \eta_n)\}$$

where η_n is as in Section 2. Neuhaus has given a result (Satz 3.1 of Neuhaus (1973)) of which the following lemma is a special case.

LEMMA 3.1 (Neuhaus). *Under the assumptions stated, y_n under (θ_0, η_n) converges weakly on $(D_k, \mathscr{D}_k)$ to a Gaussian process y_0 such that $P(y_0 \in C_k) = 1$.*

Since Lemma 3.1 (and most other weak convergence results for processes with k-dimensional time) refers to continuous df's on E^k having uniform univariate marginal distributions, we require an association between distributions on R^k and such distributions. The following lemma provides such an association.

LEMMA 3.2. *Let F be an arbitrary* df *on* R^k *with univariate marginal* df's $F_1, \cdots, F_k$. *Define* $M: R^k \rightarrow E^k$ *by*

$$M(t_1, \cdots, t_k) = (F_1(t_1), \cdots, F_k(t_k)) .$$

Then there exists a continuous df *G on* E^k *with all univariate marginals uniform on* [0, 1] *such that for all t in* R^k

$$F(t) = G(M(t)) .$$

PROOF. Suppose $X = (X_1, \cdots, X_k)$ is a random variable with df F. Suppose that $V_1, V_2, \cdots, V_k$ are independent random variables uniform on [0, 1] and independent of X. Define for $i = 1, \cdots, k$

$$U_i = (1 - V_i)F_i(X_i-) + V_i F_i(X_i)$$

and let G be the joint df of $U_1, \cdots, U_k$. It is clear that each U_i is uniformly distributed on [0, 1]. G thus has all univariate marginals uniform, and must be continuous since all its univariate marginals are. It is not difficult to check that for any $t = (t_1, \cdots, t_k)$ in R^k

$$P[U_1 \leqq F_1(t_1), \cdots, U_k \leqq F_k(t_k)] = P[X_1 \leqq t_1, \cdots, X_k \leqq t_k] .$$

The G whose existence is asserted by Lemma 3.1 need not be unique. The particular G constructed in the proof above has an additional property which is sometimes useful but not used here: if F is replaced by a family $F(t \mid \xi)$ satisfying $\sup_t |F(t \mid \xi) - F(t \mid \xi_0)| \rightarrow 0$ as $\xi \rightarrow \xi_0$, then the corresponding family $G(x \mid \xi)$ on E^k also satisfies $\sup_x |G(x \mid \xi) - G(x \mid \xi_0)| \rightarrow 0$ as $\xi \rightarrow \xi_0$.

4. General χ^2 statistics. Both cell frequencies and cell probabilities for the rectangular cell $I_\sigma(\varphi)$ can be expressed in terms of the difference operator $\Delta_\sigma(\varphi)$ defined by

$$p_\sigma(\varphi) = \int_{I_\sigma(\varphi)} dF(x) = \Delta_\sigma(\varphi)F .$$

$\Delta_\sigma(\varphi)$ can be expressed explicitly as a linear combination of $F(x(\varphi))$ for vertices $x(\varphi)$ of $I_\sigma(\varphi)$. Define the empiric process

$$W_n(x) = n^{\frac{1}{2}}\{F_n(x) - F(x \mid \eta_n)\} .$$

The troublesome error terms in assessing large sample behavior of χ^2 statistics arise from the difference between the random cells $I_\sigma(\varphi_n)$ actually used and the fixed cells $I_\sigma(\varphi_0)$ which they approach. The following lemma disposes of these terms.

LEMMA 4.1. *Suppose* A1, A2, *and* A3 *hold. Then under* (θ_0, η_n)

$$\Delta_\sigma(\varphi_n)W_n - \Delta_\sigma(\varphi_0)W_n = o_p(1) .$$

PROOF. For each $n = 0, 1, 2, \cdots$ define

$$M_n(t_1, \cdots, t_k) = (F_1(t_1 \mid \eta_n), \cdots, F_k(t_k \mid \eta_n))$$

where F_i are the univariate marginal df's of F. For $Y_1, \cdots, Y_n$ having df $F(t \mid \eta_n)$, let $H_1, \cdots, H_n$ be the random variables constructed in the proof of Lemma 3.2 having df $G(x \mid \eta_n)$. If G_n is the empiric df of the H_i, then

$$W_n(t) = y_n(M_n(t))$$

where y_n is weakly convergent by Lemma 3.1. It is enough to show that

$$y_n(M_n(x(\varphi_n))) - y_n(M_n(x(\varphi_0))) = o_p(1)$$

for vertices $x(\varphi)$ of $I_\sigma(\varphi)$. But $x(\varphi_n) - x(\varphi_0) = o_p(1)$ by A1, and thus $M_n(x(\varphi_n)) - M_0(x(\varphi_0)) = o_p(1)$ by A3.

Since y_n converges in $(D_k \mathscr{D}_k)$ to a continuous process and convergence in $(D_k, \mathscr{D}_k)$ to a continuous limit is uniform, the usual "random change of time" argument (see Billingsley (1968), page 145) establishes the desired result.

We can now describe the limiting behavior of the vector of standardized cell frequencies under quite general conditions.

THEOREM 4.1. *If* A1, A2 *and* A3 *hold, then under* (θ_0, η_n)

$$V_n(\theta_n, \varphi_n) = V_n(\eta_n) - Bn^{\frac{1}{2}}(\theta_n - \theta_0) + B_{12}\gamma + o_p(1) .$$

PROOF. Let us use the notation

$$u_{n\sigma}(\theta, \eta, \varphi) = n^{-\frac{1}{2}}\{N_{n\sigma}(\varphi) - np_\sigma(\theta, \eta, \varphi)\} .$$

Then

$$u_{n\sigma}(\theta_n, \varphi_n) - u_{n\sigma}(\eta_n) = \Delta_\sigma(\varphi_n)W_n - \Delta_\sigma(\varphi_0)W_n + n^{\frac{1}{2}}\{p_\sigma(\eta_n, \varphi_n) - p_\sigma(\theta_n, \varphi_n)\} .$$

Taylor's Theorem and Lemma 4.1 reduce the right-hand side to

$$\left(\frac{\partial p_\sigma}{\partial \eta}\right)' n^{\frac{1}{2}}(\eta_n - \eta_0) - \left(\frac{\partial p_\sigma}{\partial \theta}\right)' n^{\frac{1}{2}}(\theta_n - \theta_0) + o_p(1) .$$

The result of the theorem follows using continuity of p_σ and the usual Mann–Wald techniques.

Theorem 4.1 permits some immediate conclusions. First, φ_n affects the large sample theory only through its p-limit φ_0. Random-cell versions of all statistics of form T_n therefore differ by $o_p(1)$ (under both null and alternative hypotheses) from the corresponding statistics with fixed cells $I_\sigma(\varphi_0)$. (This fails when the number of cells increases with n faster than $O(n^{\frac{1}{2}})$.) Second, if θ_n is superefficient, so that $\theta_n - \theta_0 = o_p(n^{-\frac{1}{2}})$, the limiting behavior of $V_n(\theta_n, \varphi_n)$ (and hence that of any of our tests) is that of

$$V_n(\eta_n) + B_{12}\gamma .$$

This is the behavior we would obtain if the true θ_0 were known and no estimation were required.

The generality of Theorem 4.1 is best appreciated by example. In the example below we consider only the large-sample behavior under the null distribution, and therefore take $\gamma \equiv 0$ and $F(x \mid \theta, \eta) \equiv F(x \mid \theta)$.

EXAMPLE 4.1. Consider the family of translated geometric distributions with mass function

$$f(x \mid p, c) = p^{x-c}(1-p) \qquad x = c, c+1, c+2, \cdots .$$

Here $\theta = (p, c)$ and

$$\Omega_1 = \{(p, c) : 0 < p < 1, -\infty < c < \infty\} .$$

The MLE is $\hat{\theta}_n = (\hat{p}_n, \hat{c}_n)$ where

$$\hat{c}_n = \min_{1 \le i \le n} Y_i \qquad \hat{p}_n = \frac{\bar{Y}\text{-min } Y_i}{\bar{Y}\text{-min } Y_i + 1} .$$

We will use random cells with boundaries $x(c_n)$ where

$$x_\sigma(\varphi) = \varphi - \tfrac{1}{2} + j_\sigma$$

and $0 < j_1 < j_2 < \cdots < j_{M-1} < \infty$ are integers. This example has several pathological properties: $\hat{c}_n$ is superefficient, while $\hat{p}_n$ is not, and the support of $f(x \mid p, c)$ changes with c. Yet when random cells as above are used, Theorem 4.1 applies.

First, $\hat{c}_n$ eventually equals the true c_0 with probability 1 and from this it follows that $\hat{p}_n - p_0 = O_p(n^{-\frac{1}{2}})$ as required. The cell boundaries are always distance $\frac{1}{2}$ from mass points of the distribution and so are continuity points of the df. The cell probabilities $p_\sigma(\theta, \varphi)$ are continuous in (θ, φ) in a neighborhood of (θ_0, φ_0) and are all positive at (θ_0, φ_0). (They are not continuous for all (θ, φ), but we did not require that.) Thus A1, A2 and A3 are satisfied and Theorem 4.1 holds. Applied to this example that theorem states that under (p_0, c_0)

$$v_{n\sigma}(\hat{\theta}_n, \varphi_n) = (np_\sigma)^{-\frac{1}{2}}\{N_{n\sigma} - np_\sigma\} - p_\sigma^{-\frac{1}{2}} \frac{\partial p_\sigma}{\partial p} n^{\frac{1}{2}}(\hat{p}_n - p_0) + o_p(1) ,$$

where by convention $p_\sigma = p_\sigma(p_0, c_0)$. Thus the Chernoff–Lehmann statistic $||V_n(\hat{\theta}_n, \varphi_n)||^2$ with random cells as above has the same limiting null distribution as the statistic with c_0 known, fixed cells with boundaries $c_0 - \frac{1}{2} + j_\sigma$, and p estimated by

$$\hat{p}_n = \frac{\bar{Y} - c_0}{\bar{Y} - c_0 + 1} .$$

That this should be so is "obvious", but inclusion of such examples is a test of purportedly general theorems.

To describe the limiting distribution of general χ^2 statistics of the form T_n given in (2.1) requires additional assumptions and more notation. In what follows, A, h and L are as in A5, S is as in A4 and B and B_{12} are as defined in Section 2. Define also

$$\mu = [B_{12} - BA]\gamma \qquad (M\text{-vector})$$

$$\mu_0 = S'\mu$$

$$q' = (p_1^{\frac{1}{2}}, \cdots, p_M^{\frac{1}{2}})$$

$\chi_\sigma(y)$ indicator function of $I_\sigma(\varphi_0)$ and $W(y)$ the M-vector with σth component $[\chi_\sigma(y) - p_\sigma]/p_\sigma^{\frac{1}{2}}$

$$\Sigma = I_M - qq' + BLB' - BE[h(Y)W(Y)'] - E[W(Y)h(Y)']B' \qquad (M \times M \text{ matrix})$$

$$\Sigma_0 = S'\Sigma S .$$

THEOREM 4.2. *If* A1 *through* A5 *hold with* $\eta \equiv \eta_0$ *and* $\gamma = 0$, *then under* (θ_0, η_0) *the statistic* T_n *has as its limiting distribution the distribution of*

$$\sum_{j=1}^{M} \lambda_j \chi_{1j}^2$$

where λ_j *are the characteristic roots of* Σ_0 *and the* χ_{1j}^2 *are independent* χ^2 *variables with* 1 *degree of freedom.*

If A1 *through* A6 *hold,* T_n *has as its limiting distribution under* (θ_0, η_n) *the distribution of*

$$\sum_{\lambda_j \neq 0} \lambda_j \chi_{1j}^2(\nu_j^2/\lambda_j) + \sum_{\lambda_j = 0} \nu_j^2$$

where $\chi_{1j}^2(\nu_j^2/\lambda_j)$ *are independent noncentral* χ^2 *variables with* 1 *degree of freedom and noncentrality parameter* ν_j^2/λ_j, *and* ν_j *are the components of the* M-*vector* $\nu = P'\mu_0$ *where* P *is an orthogonal matrix such that* $P'\Sigma_0 P$ *is diagonal.*

PROOF. We prove only the second part. By Theorem 4.1 and A5, under (θ_0, η_n)

$$V_n(\theta_n, \varphi_n) = V_n(\eta_n) - Bn^{-\frac{1}{2}} \sum_{i=1}^{n} h(Y_i, \eta_n) + \mu + o_p(1) .$$

The first two terms on the right are $n^{-\frac{1}{2}}$ times the sum of the n M-vectors $W_n(Y_i) - Bh_n(Y_i)$ where $W_n(Y)$ has σth component $(\chi_\sigma(Y) - p_\sigma(\eta_n))/p_\sigma(\eta_n)^{\frac{1}{2}}$ and $h_n(Y) = h(Y, \eta_n)$. Each such vector has mean 0 and covariance matrix

$$\begin{aligned}\Sigma_n &= E[(W_n(Y) - Bh_n(Y))(W_n(Y) - Bh_n(Y))' \mid (\theta_0, \eta_n)] \\ &= I_M - q(\eta_n)q(\eta_n)' + BL(\eta_n)B' - BE[h_n(Y)W_n(Y)' \mid (\theta_0, \eta_n)] \\ &\quad - E[W_n(Y)h_n(Y)' \mid (\theta_0, \eta_n)]B' .\end{aligned}$$

Now as $n \to \infty$, Σ_n converges to Σ. The first two terms in Σ_n clearly approach the corresponding terms in Σ by A2 and A5. To establish convergence of the last two terms it is sufficient to consider terms of the form

$$\int g_n(y)\chi_\sigma(y)f(y \mid \eta_n)\, d\nu(y) \tag{4.1}$$

where $g_n(y)$ is one of the m components of $h_n(y)$. (We denote by ν the common dominating measure. This integral exists because A5 states that $\int g_n(y)f(y \mid \eta_n)\, d\nu(y) = 0$.) The absolute value of the integrand in (4.1) is dominated by $f_n(1 + g_n^2)$, where $f_n(y) = f(y \mid \eta_n)$. By A6, this converges a.e. (ν) and

$$\int f_n(1 + g_n^2)\, d\nu(y) \to \int f(1 + g^2)\, d\nu(y) < \infty$$

where f and g correspond to η_0. This is sufficient for convergence of (4.1) to $\int g(y)\chi_\sigma(y)f(y)\, dy$, which completes the proof that Σ_n converges to Σ.

The usual characteristic function proof of the multivariate central limit theorem (see e.g. Breiman (1968), page 238) is not affected by the presence of covariances varying with n but approaching a limiting covariance matrix. So under (θ_0, η_n), $V_n(\theta_n, \varphi_n)$ converges in law to $N(\mu, \Sigma)$, and hence $S'(\theta_n, \varphi_n)V_n(\theta_n, \varphi_n)$ converges in law to $N(\mu_0, \Sigma_0)$. Now

$$T_n = ||S'(\theta_n, \varphi_n)V_n(\theta_n, \varphi_n)||^2 .$$

If P is orthogonal and $P'\Sigma_0 P = \Lambda$ is diagonal with diagonal entries $\lambda_1, \cdots, \lambda_M$, the sum of squares of the components of $N(\mu_0, \Sigma_0)$ is well known to have the distribution stated in the theorem. This is therefore the limiting distribution of T_n.

Theorem 4.2 appears unwieldy, but we should recall that for the usual χ^2 statistics S is the identity matrix so that $\Sigma_0 = \Sigma$ and $\mu_0 = \mu$. We will see in the next section that when θ_n is $\bar{\theta}_n$ or $\hat{\theta}_n$ the general form given for Σ simplifies immediately. Finally, the characteristic roots λ_i corresponding to nonzero ν_i are strictly positive in most problems, so that the constant term in the limiting alternative distribution is usually not present.

A primary motivation for using random-cell statistics is to obtain statistics whose null distribution does not depend on the unknown parameter θ in location-scale cases. We will give a general result of this type by noting that under assumptions to be stated the statistic T_n is unchanged by linear transformations of the observations Y_i. This method is due to Dahiya and Gurland (1972), who use it in a special case. Here are our assumptions. (For the remainder of this section we will denote the jth component of a vector x by x^j.)

B1. $$F(y^1, \cdots, y^k \mid \theta) = F\left(\frac{y^1 - \theta^1}{\theta^2}, \cdots, \frac{y^k - \theta^{2k-1}}{\theta^{2k}}\right)$$

for $-\infty < \theta^{2j-1} < \infty$ and $\theta^{2j} > 0$, $j = 1, \cdots, k$.

B2. If $Z = (Z^1, \cdots, Z^k)'$ where $Z^j = \alpha_j Y^j + \beta_j$ for any $-\infty < \alpha_j < \infty$ and $\beta_j > 0$, $j = 1, \cdots, k$ then θ_n satisfies

$$\theta_n^{2j-1}(Z_1, \cdots, Z_n) = \alpha_j \theta_n^{2j-1}(Y_1, \cdots, Y_n) + \beta_j$$
$$\theta_n^{2j}(Z_1, \cdots, Z_n) = \alpha_j \theta_n^{2j}(Y_1, \cdots, Y_n) \qquad j = 1, \cdots, k .$$

B3. $r = m = 2k$ and each vertex $x(\varphi)$ and φ_n satisfy

$$x^j(\varphi_n(Z_1, \cdots, Z_n)) = \alpha_j x^j(\varphi_n(Y_1, \cdots, Y_n)) + \beta_j$$

for $j = 1, \cdots, k$.

B4. $K(\theta_n(Z_1, \cdots, Z_n), \varphi_n(Z_1, \cdots Z_n)) = K(\theta_n(Y_1, \cdots, Y_n), \varphi_n(Y_1, \cdots, Y_n))$.

THEOREM 4.3. *If* B1 *through* B4 *are satisfied, the statistic* T_n *has a distribution which does not depend on the true* θ_0.

PROOF. It is easy to see that under B1 through B3 $p_\sigma(\theta_n, \varphi_n)$ is unchanged by transforming Y to Z. Moreover, the number of $Z_1, \cdots, Z_n$ in $I_\sigma(\varphi_n(Z_1, \cdots, Z_n))$ is the same as the number of $Y_1, \cdots, Y_n$ in the cell $I_\sigma(\varphi_n(Y_1, \cdots, Y_n))$. So

$V_n(\theta_n, \varphi_n)$ is invariant under all linear transformation of the Y_i's. If B4 holds, this is also true of T_n, which must therefore have a distribution not depending on the location-scale parameter θ.

Theorem 4.3 is quite general. Note that B3 is satisfied when the vertices have components $x^j(\varphi) = \varphi^{2j-1} + a_j\varphi^{2j}$ and φ_n satisfies B2, whether or not $\varphi_n = \theta_n$. Thus Theorem 4.3 covers the Dahiya–Gurland result, as well as (for example) the use of sample quantiles as cell boundaries in testing fit to a univariate location-scale family as long as θ_n satisfies B2. Since B2 holds for the MLE's $\hat{\theta}_n$ and $\bar{\theta}_n$, and it can be checked that Kambhampati's choice of K satisfies B4 (as the identity matrix obviously does), all the one-sample statistics discussed in Section 2 are covered. Notice, however, that B1 demands that F be continuous in x in order to be continuous in θ, so that Theorem 4.1 on limiting behavior and Theorem 4.3 can be applied together only when F is continuous. The usual χ^2 tests for discrete families are not parameter-free.

5. One-sample χ^2 tests. Theorem 4.2 facilitates a unified derivation of the limiting distributions of the (multivariate, random-cell) one-sample statistics

$$T_{1n} = \|V_n(\bar{\theta}_n, \varphi_n)\|^2 \qquad \text{(Pearson–Fisher)}$$
$$T_{2n} = \|V_n(\hat{\theta}_n, \varphi_n)\|^2 \qquad \text{(Chernoff–Lehmann)}$$
$$T_{3n} = V_n(\hat{\theta}_n, \varphi_n)'Q(\hat{\theta}_n, \varphi_n)V_n(\hat{\theta}_n, \varphi_n)$$

where

$$Q(\hat{\theta}_n, \varphi_n) = (I_M - B_nJ_n^{-1}B_n')^{-1}, \qquad B_n = B(\hat{\theta}_n, \varphi_n) \quad J_n = J(\hat{\theta}_n).$$

T_{3n} is Kambhampati's statistic. Kambhampati (1971) used

$$Q(\hat{\theta}_n, \varphi_n) = I_M + B_n[J_n - B_n'B_n]^{-1}B_n'$$

which may be verified to be equal to our Q. Our expression for the normalizing matrix is more convenient for the theoretical work below.

As $n \to \infty$, $Q(\hat{\theta}_n, \varphi_n)$ approaches

$$Q = (I_M - BJ^{-1}B')^{-1}$$

in probability. Clearly T_{3n} is defined (for sufficiently large n) and the remarks above are valid only if $J - B'B$, which is always nnd, is pd. Here are the regularity assumptions we now make.

C1. A1, A2, A3 and A6 hold.

C2. $m \leqq M$ and the matrix with entries $\partial p_i/\partial\theta_j$ has rank m.

C3. (2.4) holds, so that $\bar{\theta}_n$ satisfies A5 with $A = (B'B)^{-1}B'B_{12}$ and $h(y) = (B'B)^{-1}B'W(y)$.

C4. $\log f(x \mid \theta, \eta)$ is differentiable with respect to (θ, η) at (θ_0, η_0). The matrix J is pd and J_{12} is finite. $(\partial/\partial\theta)F(x \mid \theta)$ may be evaluated by differentiating $f(x \mid \theta)$ under the integral sign for all x and $\theta = \theta_0$.

C5. (2.2) holds, so that $\hat{\theta}_n$ satisfies A5 with $A = J^{-1}J_{12}$ and $h(y) = J^{-1}(\partial \log f(y \mid \theta, \eta)/\partial\theta)|_{\theta_0,\eta_0}$.

C6. $J - B'B$ is pd.

Let us denote by $\lambda_{M-m}, \cdots, \lambda_{M-1}$ the m roots of the determinantal equation

$$|B'B - (1 - \lambda)J| = 0 ,$$

which always satisfy $0 \leqq \lambda_j < 1$ and satisfy $0 < \lambda_j < 1$ when $J - B'B$ is pd. Finally, let

$$C = B(B'B)^{-1}B' \qquad \mu_1 = [I_M - C]B_{12}\gamma .$$

The following lemma will be applied both here and in Section 6. Denote by e the M-vector with all components 1, so that if A is any $M \times M$ diagonal matrix, the jth component $[Ae]_j$ of Ae is the jth diagonal element of A.

LEMMA 5.1. *If* C2 *holds,* J *is* pd *and all components of* q *are positive, then there exists an orthogonal matrix* P *which simultaneously diagonalizes* qq', C *and* $BJ^{-1}B'$ *and which satisfies*

$$\begin{aligned}
&(1)\quad [P'qq'Pe]_j = 0 && j = 1, \cdots, M-1 \\
&\qquad\qquad\qquad\;\; = 1 && j = M , \\
&(2)\quad [P'CPe]_j = 0 && j = 1, \cdots, M-m-1, M \\
&\qquad\qquad\qquad\;\; = 1 && j = M-m, \cdots, M-1 , \\
&(3)\quad [P'BJ^{-1}B'Pe]_j = 0 && j = 1, \cdots, M-m-1, M \\
&\qquad\qquad\qquad\qquad\quad = 1 - \lambda_j && j = M-m, \cdots, M-1 .
\end{aligned}$$

PROOF. Routine matrix manipulation shows that qq', C and $BJ^{-1}B'$ are commuting symmetric matrices, so that an orthogonal P exists which simultaneously diagonalizes them. Moreover, qq' and C are orthogonal projections having ranks 1 and m respectively, so that by proper choice of basis we can take P to satisfy (1) and (2). $BJ^{-1}B'$ has rank m and range contained in the range of C. It follows that the characteristic roots of $BJ^{-1}B'$ are 0 except for those associated with characteristic vectors in the range of C. That these are $1 - \lambda_j$ follows easily from the fact that they are roots of the determinantal equation $|BJ^{-1}B' - \beta I| = 0$.

THEOREM 5.1. *When* C1, C2 *and* C3 *hold* T_{1n} *has limiting distribution*

$$\begin{aligned}
&\chi^2_{M-m-1} && \textit{under} \quad (\theta_0, \eta_0) \\
&\chi^2_{M-m-1}(||\mu_1||^2) && \textit{under} \quad (\theta_0, \eta_n) .
\end{aligned}$$

When C1, C2, C4, C5 *and* C6 *hold,* T_{2n} *has limiting distribution*

$$\begin{aligned}
&\chi^2_{M-m-1} + \textstyle\sum_{j=M-m}^{M-1} \lambda_j \chi^2_{1j} && \textit{under} \quad (\theta_0, \eta_0) \\
&\chi^2_{M-m-1}(||\mu_1||^2) + \textstyle\sum_{j=M-m}^{M-1} \lambda_j \chi^2_{1j}(\nu_j^2/\lambda_j) && \textit{under} \quad (\theta_0, \eta_n) .
\end{aligned}$$

When C1, C2, C4, C5 *and* C6 *hold,* T_{3n} *has limiting distribution*

$$\begin{aligned}
&\chi^2_{M-1} && \textit{under} \quad (\theta_0, \eta_0) \\
&\chi^2_{M-1}(||\mu_1||^2 + \textstyle\sum_{j=M-m}^{M-1} \nu_j^2/\lambda_j) && \textit{under} \quad (\theta_0, \eta_n) .
\end{aligned}$$

PROOF. Referring to the notation of Theorem 4.2, we have using $C^2 = C$ and

$q'B = 0$ that for T_{1n}

$$\Sigma_1 = I_M - qq' - C$$
$$\mu_1 = [I_M - C]B_{12}\gamma .$$

For T_{2n} we see that $L = J^{-1}$ and hence

$$\Sigma_2 = I_M - qq' - BJ^{-1}B' \qquad \mu_2 = [B_{12} - BJ^{-1}J_{12}]\gamma .$$

Lemma 5.1 now applies. Further, if $\nu_1 = P'\mu_1$ and $\nu_2 = P'\mu_2$ we see that $\nu_{1j} = \nu_{2j}$ for $j = 1, \cdots, M - m - 1, M$ and $\nu_{1j} = 0$ for $j = M - m, \cdots, M$. The theorem now follows from Theorem 4.2, taking $\nu_j = \nu_{2j}$.

6. Two-sample χ^2 tests. The two-sample statistics of Chase (1972) and Murty–Gafarian (1970) require only minor modifications of the arguments already given, as do the k-sample analogs which we will not consider. Chase considers the case in which $\theta_n = \theta_n(Y_{n+1}, \cdots, Y_{n+m(n)})$ where $Y_{n+1}, \cdots, Y_{n+m(n)}$ is a sample from $F(x \mid \theta, \eta)$ independent of $Y_1, \cdots, Y_n$. We allow $\varphi_n = \varphi_n(Y_1, \cdots, Y_{n+m(n)})$ as well. Theorem 4.1 continues to apply to $V_n(\theta_n, \varphi_n)$ and Theorem 4.2 is modified only by the fact that in A5 n must now be replaced by $m(n)$ since θ_n is based on a sample of size $m(n)$. Combining this version of A5 with Theorem 4.1 and assuming that as $n \to \infty$, $m(n) \to \infty$ and $n/m(n) \to \tau \geqq 0$ establishes that under (θ_0, η_n)

$$V_n(\theta_n, \varphi_n) \to_{\mathscr{L}} N(\mu_\tau, \Sigma_\tau)$$

where

$$\mu_\tau = [B_{12} - \tau^{\frac{1}{2}}BA]\gamma$$
$$\Sigma_\tau = I_M - qq' + \tau BLB' .$$

Theorem 4.2 then applies with μ, Σ replaced by μ_τ, Σ_τ. For the special cases $\hat{\theta}$ and $\bar{\theta}$ we therefore have the following generalization of Chase's result, to which we have added a two-sample version of Kambhampati's statistic.

THEOREM 6.1. *Suppose that as $n \to \infty$, $n/m(n) \to \tau$. If* C1, C2 *and* C3 *with n replaced by $m(n)$ hold, then*

$$T_{4n} = ||V_n(\bar{\theta}_{m(n)}, \varphi_n)||^2$$

has limiting distribution

$$\chi^2_{M-m-1} + (1 + \tau)\chi_m^{\ 2} \qquad under \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-m-1}(||\mu_1||^2) + (1 + \tau)\chi_m^{\ 2}(||\mu_\tau - \mu_1||^2) \qquad under \quad (\theta_0, \eta_n) .$$

If C1, C2, C4, C6 *and* C5 *with n replaced by $m(n)$ hold, then*

$$T_{5n} = ||V_n(\hat{\theta}_{m(n)}, \varphi_n)||^2$$

has limiting distribution

$$\chi^2_{M-m-1} + \textstyle\sum_{j=M-m}^{M-1} \alpha_j \chi^2_{1j} \qquad under \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-m-1}(||\mu_1||^2) + \textstyle\sum_{j=M-m}^{M-1} \alpha_j \chi^2_{1j}(\beta_j^{\ 2}/\alpha_j) \qquad under \quad (\theta_0, \eta_n)$$

where

$$\alpha_j = 1 + (1 - \lambda_j)\tau$$
$$\textstyle\sum_{j=M-m}^{M-1} \beta_j^{\ 2} = ||\mu_\tau - \mu_1||^2 .$$

If C1, C2, C4, C6 *and* C5 *with* n *replaced by* $m(n)$ *hold, and* $Q_n = (I_M + (n/m(n))B_n J_n^{-1} B_n')^{-1}$ *then*

$$T_{6n} = V_n(\hat{\theta}_{m(n)}, \varphi_n)' Q_n(\hat{\theta}_{m(n)}, \varphi_n) V_n(\hat{\theta}_{m(n)}, \varphi_n)$$

has limiting distribution

$$\chi^2_{M-1} \qquad \textit{under} \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-1}(||\mu_1||^2 + \textstyle\sum_{j=M-m}^{M-1} \beta_j^2/\alpha_j) \qquad \textit{under} \quad (\theta_0, \eta_n) .$$

PROOF. When $\theta_n = \bar{\theta}_{m(n)}(Y_{n+1}, \cdots, Y_{n+m(n)})$ calculation shows that

$$\Sigma_\tau = I_M - qq' + \tau C$$
$$\mu_\tau = \mu_1 + (1 - \tau^{\frac{1}{2}}) C B_{12} \gamma .$$

When $\theta_n = \hat{\theta}_{m(n)}(Y_{n+1}, \cdots, Y_{n+m(n)})$, we see that

$$\Sigma_\tau = I_M - qq' + \tau B J^{-1} B'$$
$$\mu_\tau = \mu_2 + (1 - \tau^{\frac{1}{2}}) B J^{-1} J_{12} \gamma .$$

Applying Lemma 5.1 and Theorem 4.2 as in Theorem 5.1 completes the proof.

Note that when $\tau = 0$ the limiting distribution of T_{4n}, T_{5n} and T_{6n} is $\chi^2_{M-1}(||B_{12}\gamma||^2)$ since $\mu_\tau = B_{12}\gamma$ then. This is the same as the limiting distribution of $||V_n(\theta_0, \varphi_0)||^2$, the statistic when θ_0 is known and no estimation is required.

Murty and Gafarian consider the case in which $\theta_n = \theta_n(Y_1, \cdots, Y_{n+m(n)})$ is based on both samples and again only $Y_1, \cdots, Y_n$ are tested for fit. Let $N(n) = n + m(n)$ and assume that $n/N(n) \to \tau$ as $n \to \infty$. Note that $0 \leqq \tau \leqq 1$ always. Arguments similar to those employed in Theorem 6.1 show that the mean μ_τ of the limiting distribution of $V_n(\theta_n, \varphi_n)$ is as above and establish the following theorem.

THEOREM 6.2. *Suppose that as* $n \to \infty$, $n/N(n) \to \tau$. *If* C1, C2 *and* C3 *with* n *replaced by* $N(n)$ *hold, then*

$$T_{7n} = ||V_n(\bar{\theta}_{N(n)}, \varphi_n)||^2$$

has limiting distribution

$$\chi^2_{M-m-1} + (1 - \tau)\chi_m^2 \qquad \textit{under} \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-m-1}(||\mu_1||^2) + (1 - \tau)\chi_m^2(||\mu_\tau - \mu_1||^2) \qquad \textit{under} \quad (\theta_0, \eta_n) .$$

If C1, C2, C4, C6 *and* C5 *with* n *replaced by* $N(n)$ *hold, then*

$$T_{8n} = ||V_n(\hat{\theta}_{N(n)}, \varphi_n)||^2$$

has limiting distribution

$$\chi^2_{M-m-1} + \textstyle\sum_{j=M-m}^{M-1} \delta_j \chi^2_{1j} \qquad \textit{under} \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-m-1}(||\mu_1||^2) + \textstyle\sum_{j=M-m}^{M-1} \delta_j \chi^2_{1j}(\beta_j^2/\delta_j) \qquad \textit{under} \quad (\theta_0, \eta_n)$$

where

$$\delta_j = 1 - (1 - \lambda_j)\tau$$

and the β_j *are as before. If* C1, C2, C4, C6 *and* C5 *with* n *replaced by* $N(n)$

hold and

$$Q_n = \left(I_M - \frac{n}{N(n)} B_n J_n^{-1} B_n'\right)^{-1}, \qquad \textit{then}$$

$$T_{9n} = V_n(\hat{\theta}_{N(n)}, \varphi_n)' Q_n(\hat{\theta}_{N(n)}, \varphi_n) V_n(\hat{\theta}_{N(n)}, \varphi_n)$$

has limiting distribution

$$\chi^2_{M-1} \quad \textit{under} \quad (\theta_0, \eta_0)$$
$$\chi^2_{M-1}(\|\mu_1\|^2 + \textstyle\sum_{j=M-m}^{M-1} \beta_j^2/\delta_j) \quad \textit{under} \quad (\theta_0, \eta_n).$$

When $\tau = 0$, these distributions are again the same as the case in which θ_0 is known. When $\tau = 1$, we have no additional information (in the limit) from the second sample and as expected T_{7n} is asymptotically equivalent to T_{1n}, T_{8n} to T_{2n} and T_{9n} to T_{3n}.

7. Comparison of χ^2 tests. The limiting null and alternative distributions which we have presented can be used computationally to obtain critical points of χ^2 tests and power against various alternatives. Methods of computation and results in this direction are given by Moore in Section 4 of [16] and especially by Dahiya and Gurland (1972, 1973). Explicit computation is unfortunately somewhat complicated.

Some general comparisons of these statistics are possible in special cases, as Chibisov (1971) noted for T_{1n} and T_{2n}. Chibisov displays examples of each of the following special cases.

Case 1. $\|\mu_1\|^2 = 0$, but some $\nu_j^2/\lambda_j > 0$.

	H_0	H_1
T_{1n}	χ^2_{M-m-1}	χ^2_{M-m-1}
T_{2n}	$\chi^2_{M-m-1} + \sum_{M-m}^{M-1} \lambda_j \chi^2_{1j}$	$\chi^2_{M-m-1} + \sum_{M-m}^{M-1} \lambda_j \chi^2_{1j}(\nu_j^2/\lambda_j)$
T_{3n}	χ^2_{M-1}	$\chi^2_{M-1}(\sum_{M-m}^{M-1} \nu_j^2/\lambda_j)$.

In this case, the T_{1n} test of (limiting) size α has power α so that T_{2n} and T_{2n} are both more powerful.

Case 2. $\|\mu_1\|^2 > 0$ but all $\nu_j = 0$.

	H_0	H_1
T_{1n}	χ^2_{M-m-1}	$\chi^2_{M-m-1}(\|\mu_1\|^2)$
T_{2n}	$\chi^2_{M-m-1} + \sum_{M-m}^{M-1} \lambda_j \chi^2_{1j}$	$\chi^2_{M-m-1}(\|\mu_1\|^2) + \sum_{M-m}^{M-1} \lambda_j \chi^2_{1j}$
T_{3n}	χ^2_{M-1}	$\chi^2_{M-1}(\|\mu_1\|^2)$.

Here the unique most powerful size α test of H_0 vs. H_1 based on T_{1n} is the size α upper tail test. Each of T_{2n} and T_{3n} has the distribution of $T_{1n} + V$, where V has the same distribution independent of T_{1n} under both H_0 and H_1. Thus the upper tail T_{2n} and T_{3n} tests can be reproduced as randomized tests based on T_{1n} and are less powerful than the T_{1n} test.

We see that T_{1n} can be either more or less powerful (limiting power against Pitman alternatives) than both T_{2n} and T_{3n}, so that no uniform dominance exists involving T_{1n}. When T_{2n} and T_{3n} are compared by means of approximate Bahadur slope against non-contiguous alternatives, Spruill (1973) shows that T_{3n} is uniformly at least as good as T_{2n}. The analogous result for power against Pitman alternatives is not apparent.

One further comparison is possible. When $||\mu_1||^2 > 0$ and $\tau = 1$ in T_{4n} (both samples are of equal size in the limit), then $\mu_\tau = \mu_1$ and reference to Theorem 6.1 shows that T_{1n} is more powerful than T_{4n} by the argument used is Case 2 above.

REFERENCES

[1] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Bofinger, Eve (1973). Goodness-of-fit test using sample quantiles. *J. Roy. Statist. Soc. Ser. B* **35** 277–284.

[4] Breiman, Leo (1968). *Probability.* Addison-Wesley, Reading, Mass.

[5] Chase, G. R. (1972). Chi-square test when parameters are estimated independently of the sample. *J. Amer. Statist. Assoc.* **67** 609–611.

[6] Chernoff, H. 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.

[7] Chibisov, D. M. (1971). Certain chi-square type tests for continuous distributions. *Theor. Probability Appl.* **16** 1–22.

[8] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[9] Dahiya, R. C. and Gurland, J. (1972). Pearson chi-square test of fit with random intervals. *Biometrika* **59** 147–153.

[10] Dahiya, R. C. and Gurland, J. (1973). How many classes in the Pearson chi-square test? *J. Amer. Statist. Assoc.* **68** 707–712.

[11] Davidson, R. R. and Lever, W. E. (1970). The limiting distribution of the likelihood ratio statistic under a class of local alternatives. *Sankhyā* **32** 209–224.

[12] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

[13] Kambhampati, C. (1971). A chi-square statistic for goodness-of-fit tests. Thesis, Cornell Univ.

[14] Kempthorne, O. (1968). The classical problem of inference-goodness of fit. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 235–249.

[15] Moore, D. S. (1970). On multivariate chi-square statistics with random cell boundaries. Purdue Statistics Department Mimeo Series No. 246.

[16] Moore, D. S. (1971). A chi-square statistic with random cell boundaries. *Ann. Math. Statist.* **42** 147–156.

[17] Murty, V. K. and Gafarian, A. V. (1970). Limiting distributions of some variations of the chi-square statistics. *Ann. Math. Statist.* **41** 188–194.

[18] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

[19] Neuhaus, G. (1973). Zur Verteilungskonvergenz einiger Varianten der Cramér-von Mises Statistik. *Math. Operationsforsch. u. Statist.* **4** 473–484.

[20] Roy, A. R. (1959). On χ^2 statistics with variable intervals. Technical Report No. 1, Stanford Univ. Department of Statistics.

[21] Spruill, M. C. (1973). Large-sample properties of chi-square goodness-of-fit tests. Thesis, Purdue Univ.

[22] WITTING, H. (1959). Über einen χ^2-Test, dessen Klassen durch geordnete Stichprobenfunktionen festgelegt werden. *Ark. Mat.* **10** 468–479.

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
W. LAFAYETTE, INDIANA 47907

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332