The Annals of Statistics
1977, Vol. 5, No. 5, 1017–1026

ASYMPTOTIC PROPERTIES OF TESTS BASED ON LINEAR COMBINATIONS OF THE ORTHOGONAL COMPONENTS OF THE CRAMÉR-VON MISES STATISTIC[1]

BY DAVID A. SCHOENFELD

Harvard University

Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables defined on the unit interval. The generalized jth orthogonal component is defined as $V_{nj} = n^{-\frac{1}{2}} \sum_{i=1}^{n} d_j(X_i)$, where $\{1, d_1, d_2, \cdots\}$ is an orthonormal basis for $\mathscr{L}_2([0, 1])$. These statistics are a generalization of the orthogonal components of the Cramér-von Mises statistic [2]. Linear combinations of the V_{nj} are applied to the problem of testing the null hypothesis of a uniform distribution against the alternative density $p_n(x) = 1 + h(x)/n^{\frac{1}{2}} + k_n(x)/n$ where $h(x)$ is square integrable and $k_n(x)$ is dominated by a square integrable function. When $a_j = \int h(x)\, d_j(x)\, dx$, tests based on $\sum_{j=1}^{m} a_j V_{nj}$ are shown to be asymptotically most powerful as $\min(m, n) \to \infty$. The asymptotic power and efficiency of these tests are computed. A procedure is developed for choosing among possible density functions when a goodness-of-fit test rejects the null hypothesis.

1. Introduction. Durbin and Knott [2] showed that the Cramér-von Mises statistic can be partitioned into orthogonal components in a manner analogous to principal components analysis. These components can be used to test a simple null hypothesis. The jth orthogonal component of the Cramér-von Mises statistic, Z_{nj}, can be expressed as

$$Z_{nj} = n^{-\frac{1}{2}} \sum_{i=1}^{n} 2^{\frac{1}{2}} \cos(j\pi X_{ni}),$$

where $X_{n1}, X_{n2}, \cdots, X_{nn}$ are continuous i.i.d. random variables defined on the unit interval. Durbin and Knott proposed tests that reject the null hypothesis of a uniform distribution when Z_{nj} is too large.

This paper shows that a sequence of tests can be found, based on linear combinations of the orthogonal components, which is asymptotically most powerful for a given member of a large class of sequences of alternatives. The asymptotic power and efficiency of these tests are derived and a method of finding $1 - \alpha$ confidence intervals for a density function is described.

Since results about statistics based on cosine functions can easily be extended to other orthonormal bases, this paper will define components more generally. For this purpose, let $\{d_0(x), d_1(x), d_2(x), \cdots\}$ be an orthonormal basis for the space

Received May 1975; revised January 1977.

[1] This work grew out of the author's doctoral dissertation at the University of Oregon supported by NSF Grant GP 32052X2. Partial support was provided at Stanford University PHS Grant 5-TIGM25-17 and State University of New York at Buffalo NCI CA-10810-09.

AMS 1970 *subject classifications.* Primary 62G10; Secondary 62G20.

Key words and phrases. Orthogonal components, Cramér-von Mises statistic, asymptotic efficiency and power.

of square integrable functions on the unit interval. We require that $d_0(x) = 1$ so that $d_1(x), d_2(x), \cdots$ have zero integrals. Let $(b_1, b_2, \cdots, b_m)$ be real coefficients. We shall examine the asymptotic properties of tests which reject the null hypothesis when $\sum_{j=1}^m b_j[n^{-\frac{1}{2}} \sum_{i=1}^n d_j(X_{ni})]$ is too large.

The results provide a method for choosing that linear combination of the orthogonal components which is appropriate for a given testing problem as well as a method for finding a $1 - \alpha$ confidence region for a density function when one must decide which of several possible density functions fits the sample.

Section 2 contains definitions of the tests and distributions under consideration. Section 3 contains the major results and Section 4 contains the proofs of these results.

2. Definitions and assumptions. When a null hypòthesis completely specifies a continuous distribution, the probability integral transformation reduces the goodness-of-fit problem to one of testing whether or not a random variable has the uniform distribution on the unit interval. Therefore, without loss of generality we take the unit interval to be the range of random variables, the domain of functions and the region of integration.

Let $\{X_{n1}, X_{n2}, \cdots, X_{nn}\}$ be a double sequence of random variables i.i.d. in each row. Asymptotic results will be proved for a sequence of testing problems indexed by n. The null hypothesis, H_0, is that X_{ni} has a uniform distribution and the alternative, H_n, is that X_{ni} has probability density $p_n(x)$, specified in Definition 3.

DEFINITION 1. Let $\{1, d_1(x), d_2(x), \cdots\}$ be an orthonormal basis for $\mathscr{L}_2([0, 1])$. The jth generalized orthogonal component based on $\{1, d_1(x), d_2(x), \cdots\}$ is defined by

$$V_{nj} = V_{nj}(X_{n1}, X_{n2}, \cdots, X_{nn}) = n^{-\frac{1}{2}} \sum_{i=1}^n d_j(X_{ni}) .$$

For the special case when $d_j(x) = 2^{\frac{1}{2}} \cos j\pi x$, $V_{nj} = Z_{nj}$, the orthogonal components of the Cramér-von Mises statistic. The "smooth" test due to Neyman [6] is based on the V_{nj} when $d_j(x)$ are the generalized Legendre polynomials. We use the symbol $\mathbf{V}_{n,m}$ to denote the vector $(V_{n1}, V_{n2}, \cdots, V_{nm})'$. The tests under consideration have the following form:

DEFINITION 2. Let $\mathbf{b} = (b_1, b_2, \cdots, b_m)'$ be an m-vector. The level-α test based on linear combination $\mathbf{b}'\mathbf{V}_{n,m}$ has the test function

$$\begin{aligned} \psi_{n,m,\alpha}(\mathbf{b}, \mathbf{V}_{n,m}) &= 1 \qquad \text{when} \quad \mathbf{b}'\mathbf{V}_{n,m} > K_{n,\alpha} \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

where $K_{n,\alpha}$ is chosen so that $P(\psi_{n,m,\alpha} = 1 \mid H_0) = \alpha$.

The asymptotic properties of the these tests are determined for sequences of alternative hypotheses, $p_n(x)$ defined as follows.

DEFINITION 3. Let $p_n(x)$ be a strictly positive probability density function defined on the unit interval such that

$$p_n(x) = 1 + n^{-\frac{1}{2}}h(x) + n^{-1}k_n(x) ,$$

where $h(x)$ is a function satisfying the condition,

$$\int (h(x))^2\, dx < \infty \,, \tag{2.1}$$

and

$$|k_n(x)| < m(x) \,, \tag{2.2}$$

for some square integrable function $m(x)$.

Definition 3 is not the usual sequence used in asymptotic studies. One often assumes that $\{Y_{n1}, Y_{n2}, \cdots, Y_{nn}\}$ is a double sequence of random variables with distribution function $F(Y, \boldsymbol{\theta}_n)$, where $\boldsymbol{\theta}_n$ is a sequence of r-dimensional vector parameters with $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\gamma}$. The hypothesis usually considered is that $\boldsymbol{\gamma} = 0$. However, a probability integral transformation and Taylor's formula can often be used to transform this sequence of alternatives into the form of Definition 3.

When testing for normal location or scale shifts, one takes $r = 1$ and $F(Y, \boldsymbol{\theta}_n)$ to be $\Phi(Y - \theta_n)$ or $\Phi(Y \exp(-\theta_n))$ respectively, where $\theta_n = n^{-\frac{1}{2}}\gamma$ and Φ is the standard normal distribution function. In the first case $h(x) = \gamma\Phi^{-1}(x)$ and in the second $h(x) = \gamma((\Phi^{-1}(x))^2 - 1)$. The remainder term, $k_n(x)$, is included in Definition 3 to allow more general alternatives and does not need to be explicitly calculated. For location shifts one can show that $k_n(x)$ is dominated by a square integrable function without explicitly finding $k_n(x)$. For scale shifts it is sufficient that $n^{-\frac{1}{2}}\gamma < (\log 2)/2$ for (2.2) to hold.

3. Results. When a double sequence of random variables has the defined sequence of densities, a special form of the central limit theorem can be proved.

THEOREM 1. *Let $\{X_{n1}, X_{n2}, \cdots, X_{nn}\}$ be a double sequence of random variables* i.i.d. *in each row, and let X_{ni} have the* pdf $p_n(x)$ *in Definition* 3. *If $g(x)$ is a square integrable function whose integral over the unit interval is zero, then*

$$n^{-\frac{1}{2}} \textstyle\sum_{i=1}^{n} g(X_{ni})$$

has a limiting normal distribution with mean $\mu = \int g(x)h(x)\, dx$ and variance $\sigma^2 = \int (g(x))^2\, dx$.

Under H_0, $h(x) = 0$ so $\mu = 0$. Notice also that $g(X_{ni})$ need not have second moments when $h(x) \neq 0$ since $\int (g(x))^2 h(x)\, dx$ may not exist.

Two sequences of probability measures Q_n, Q_n' defined on the same sequence of measurable spaces are said to be contiguous if for all sequences of statistics $\{T_n\}$, $T_n \to_p 0$ under $\{Q_n\}$ if and only if $T_n \to_p 0$ under $\{Q_n'\}$. Definition 3 describes a family of sequences of probability measures. For each function $h(x)$ and sequence of functions $k_n(x)$ satisfying (2.1) and (2.2) there is a corresponding sequence of probability measures on the sequence of n-dimensional cubes. Theorem 2 states that each of these sequences of measures is contiguous with the uniform probability measure and therefore these measures are contiguous with one another.

THEOREM 2. *Let $\{P_n\}$ be the sequence of probability measures defined on the Borel sets of the n-dimensional unit cube by the expression $P_n(A) = \int_A \prod_{i=1}^n p_n(x_i)\, dx_i$. Let $\{Q_n\}$ be the sequence of uniform probability measures, $Q_n(A) = \int_A dx_1\, dx_2, \cdots, dx_n$. Then $\{P_n\}$ and $\{Q_n\}$ are contiguous.*

The asymptotic distribution of $\mathbf{b}'\mathbf{V}_{n,m}$ can be found when $\mathbf{b}'$ are the first m-terms of a square summable sequence and $m \to \infty$ with n.

THEOREM 3. *Let V_{nj} be as in Definition 1 and let X_{ni} have* pdf $p_n(x)$. *Suppose $m(n)$ be an integer valued function of n which approaches infinity with n. Then if*

$$a_j = \int h(x) d_j(x)\, dx\,, \tag{3.1}$$

and $\sum_{j=1}^{\infty} b_j^2 < \infty$, $\sum_{j=1}^{m(n)} b_j V_{nj}$ has an asymptotic normal distribution with mean

$$\sum_{j=1}^{\infty} a_j b_j \tag{3.2}$$

and variance

$$\sum_{j=1}^{\infty} b_j^2\,.$$

Note that (2.1) and (3.1) imply that $\{a_j\}$ is a square summable sequence so (3.2) is finite.

Theorem 3 has two immediate corollaries.

COROLLARY 1. *The vector $\mathbf{V}_{n,m}$ has an asymptotic multivariate normal distribution with mean vector $\mathbf{a} = (a_1, a_2, \cdots, a_m)'$ and variance-covariance matrix $\mathbf{I}$, the $m \times m$ identity matrix.*

COROLLARY 2. *Under the hypothesis that X_{ni} has the uniform density, the sum $\sum_{j=1}^{m(n)} b_j V_{nj}$ is asymptotically normal with zero mean and variance equal to $\sum_{j=1}^{\infty} b_j^2$.*

We shall assume that the sequence $\{a_j\}$ is defined by (3.1). If we use the first m-terms of this sequence as coefficients for the orthogonal components, we can approximate the most powerful test of H_0 versus H_n.

THEOREM 4. *Let $\psi_{n,m,\alpha}(\mathbf{a}, \mathbf{V}_{n,m})$ be the test function based on $\mathbf{a}'\mathbf{V}_{n,m}$ as in Definition 2, where a_j is given by (3.1). Assume $m \to \infty$ as $n \to \infty$. Then if $\lambda_n(X_{n1}, X_{n2}, \cdots, X_{nn})$ is any sequence of level-α test functions of H_0 versus H_n,*

$$\liminf_{n\to\infty} [P_n(\psi_{n,m,\alpha}(\mathbf{a}, \mathbf{V}_{n,m}) = 1) - P_n(\lambda_n(X_{n1}, X_{n2}, \cdots, X_{nn}) = 1)] \geqq 0\,.$$

The expression for the asymptotic power and efficiency of tests based on a linear combination of the generalized components, $\sum_{j=1}^{m} b_j V_{nj}$, have a particularly simple form.

THEOREM 5. *Let m be a fixed integer, a_j be defined as in (3.1), $\mathbf{b}$ be an arbitrary m-vector, and $\psi_{n,m,\alpha}(\mathbf{b}, \mathbf{V}_{n,m})$ be the test function defined by Definition 2. Then the following two expressions hold:*

1. $\lim_{n\to\infty} K_{n,\alpha} = z_{1-\alpha}(\sum_{j=1}^{m} b_j^2)^{\frac{1}{2}}$, *where $z_{1-\alpha}$ is the $(1-\alpha)$ 100 percentile of the normal distribution.*
2. $\lim_{n\to\infty} P_n(\psi_{n,m,\alpha}(\mathbf{b}, \mathbf{V}_{n,m}) = 1) = \Phi[(\sum_{j=1}^{m} b_j a_j)(\sum_{j=1}^{m} b_j^2)^{-\frac{1}{2}} - z_{1-\alpha}]$.

THEOREM 6. *The asymptotic efficiency* [4, page 267] *of* $\psi_{n,m,\alpha}(\mathbf{b}, \mathbf{V}_{n,m})$, *is given by*

$$e = (\sum_{j=1}^{m} a_j b_j)^2 / [(\sum_{j=1}^{\infty} a_j^2)(\sum_{j=1}^{m} b_j^2)] , \tag{3.3}$$

when $\sum_{j=1}^{m} a_j b_j > 0$. *The test would have power less than* α *if* $\sum_{j=1}^{m} a_j b_j < 0$.

Let $p_n^*(x) = 1 + n^{-\frac{1}{2}} h^*(x) + n^{-1} k_n^*(x)$ be another sequence of alternatives satisfying Definition 3, with $a_j^* = \int h^*(x) d_j(x)\, dx$. The sequence of tests $\psi_{m,n}(\mathbf{a}^*, \mathbf{V}_{n,m})$ is asymptotically most powerful for testing H_0 versus the alternative that X_{ni} has the density $p_n^*(x)$. Letting $m \to \infty$ and substituting a_j^* for b_j in (3.3), we have an expression for the asymptotic efficiency of $\psi_{m,n}(\mathbf{a}^*, \mathbf{V}_{n,m})$ when the alternative hypothesis is H_n. This can be rewritten as

$$e = [1 - \tfrac{1}{2} \sum_{j=1}^{\infty} (a_j (\sum_{j=1}^{\infty} a_j^2)^{-\frac{1}{2}} - a_j^* (\sum_{j=1}^{\infty} a_j^{*2})^{-\frac{1}{2}})^2]^2$$

or

$$e = (1 - \tfrac{1}{2} \int (h(x)/c - h^*(x)/c^*)^2\, dx)^2 , \tag{3.4}$$

where c and c^* are the $\mathscr{L}_2$ norm of $h(x)$ and $h^*(x)$ respectively.

The class of alternative sequences of densities that satisfy Definition 3 can be associated with the space $\mathscr{M}$ of square integrable functions having zero integrals. The sequence $\{p_n(x)\}$ is associated with $h(x)$. This association is "many to one" because of the second order term $k_n(x)$. Equation (3.4) shows that if two alternative sequences are close in the $\mathscr{L}_2([0, 1])$ norm on the unit sphere in $\mathscr{M}$ then the most powerful test for one of the alternatives will be fairly efficient under the other alternative. In essence the efficiency measures the "angle" between two alternatives; if this "angle" is small the same test could be used for both alternatives.

The orthogonal components can also be used to find a $1 - \alpha$ confidence set for an unknown density. When there are a finite set of possible densities it is easy to check which of these densities belong to the set.

THEOREM 7. *Assume that* $\{d_1(x), d_2(x), \cdots,\}$ *are bounded functions. Let* $p(x)$ *be a density function on* $[0, 1]$, *let* $u_i = \int d_i(x) p(x)\, dx$, *and let* $\Gamma_{ij} = \int (d_j(x) - u_j)(d_i(x) - u_i) p(x)\, dx$. *Then if* $\{X_{n1}, X_{n2}, \cdots, X_{nn}\}$ *are* i.i.d. *with density* $p(x)$, $\mathbf{V}_{n,m}$ *is asymptotically m-dimensional multivariate normal with mean vector* $\mathbf{u} = n^{\frac{1}{2}}(u_1, u_2, \cdots, u_m)'$ *and covariance matrix* $\{\Gamma_{ij}\}$ *as* $n \to \infty$ *(fixed m)*.

COROLLARY 3. *If* $\beta_{m,1-\alpha}$ *is the* $(1 - \alpha)$ 100 *percentile of a* $\chi_m^2(0)$ *distribution, the set* $\{p(x) : (\mathbf{V}_{n,m} - n^{\frac{1}{2}}\mathbf{u})\Gamma^{-1}(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\mathbf{u}) > \beta_{m,1-\alpha}\}$, *in which* $\mathbf{u}$ *and* Γ *are defined as in Theorem* 7, *is an asymptotic* $1 - \alpha$ *confidence set for the density function oj* X_{ni}.

When the densities of interest are of the form $p_\theta(x) = 1 + \theta h(x)$ there is an asymptotic confidence set for $h(x)$ which does not depend on θ. This set would be useful if the distribution was a mixture of a known and an unknown distribution and one wanted a confidence set for the unknown distribution which did not depend on the mixing parameter.

THEOREM 8. *Let $p_\theta(x) = 1 + \theta h(x)$ be a strictly positive density for $0 \leqq \theta \leqq c$ and let $\{d_1(x), d_2(x), \cdots\}$ be uniformly bounded. Define*

$$\begin{aligned} A_{ij} &= \textstyle\int d_i(x)d_j(x)h(x)\,dx\,; \\ a_i &= \textstyle\int d_i(x)h(x)\,dx\,; \\ \mathbf{a} &= (a_1, a_2, \cdots, a_m)\,; \\ \theta_n{}^* &= n^{-\frac{1}{2}}(\mathbf{a}'\mathbf{V}_{n,m}) \cdot (\mathbf{a}'\mathbf{a})^{-1}\,; \\ \tilde{\theta}_n &= \theta^* \qquad 0 < \theta^* < c \\ &= 0 \qquad \theta^* \leqq 0 \\ &= c \qquad \theta^* \geqq c\,; \\ [\Gamma_n]_{ij} &= \tilde{\theta}_n A_{ij} - (\tilde{\theta}_n)^2 a_i a_j\,, \qquad i \neq j \\ &= 1 + \tilde{\theta}_n A_{ii} - \tilde{\theta}_n{}^2 a_i^2\,, \qquad i = j\,; \\ \hat{\theta}_n &= (\mathbf{V}'_{n,m}\Gamma_n{}^{-1}\mathbf{a})(n^{\frac{1}{2}}\mathbf{a}'\Gamma_n{}^{-1}\mathbf{a})^{-1}\,. \end{aligned}$$

Then as $n \to \infty$

$$(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\hat{\theta}_n\mathbf{a})'\Gamma_n{}^{-1}(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\hat{\theta}_n\mathbf{a}) \tag{3.5}$$

has the $\chi^2_{m-1}(0)$ distribution and

$$\{h(x)\colon (\mathbf{V}_{n,m} - n^{\frac{1}{2}}\hat{\theta}_n\mathbf{a})'\Gamma_n{}^{-1}(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\theta_n\mathbf{a}) > \beta_{m-1,1-\alpha}\}$$

is a $1 - \alpha$ confidence set for $h(x)$.

Since cosine functions are standard computer software the orthogonal components of the Cramér-von Mises statistic are easy to compute. For certain special applications other sets of orthogonal components might be better. If the null hypothesis is a standard normal distribution and the components are based on $\{(j!)^{-\frac{1}{2}}H_j(\Phi^{-1}(x))\}$ where $H_j(x)$ is the jth Hermite polynomial, then the first component measures location shift, the second measures scale and each succeeding component measures the corresponding coefficient in the Type A Gram–Charlier series [5, page 156].

Monte Carlo studies were conducted using the components based on $\{\cos j\pi x\}$. A total of 10,000 samples of 50 uniform random numbers were drawn. The first five components were computed and used in 10 different linear combinations to perform tests with an asymptotic 5% significance level based on Theorem 5. From 4.4% to 5.5% of the 10,000 samples were rejected, depending on the linear combination used on the test. Monte Carlo power studies were also performed and the number of rejections was usually quite close to the number predicted by Theorem 5.2.

Monte Carlo studies of the proposed $1 - \alpha$ confidence limits were also conducted, using 100 samples of size 500 and 100 with mixing parameters .25, .5 and .75. In each study a mixture of a uniform and one of 10 different alternatives were generated and the procedure described in Theorem 8 was performed to see if each of the 10 alternatives was in the $1 - \alpha$ confidence region. The 90% confidence interval for samples of size 500 with mixing parameter .5 contained the correct alternative from 86% to 93% of the time. When the mixing parameter was changed to .25, the 90% confidence set contained the correct alternative from 61% to 91% of the time (mean 78, s.d. 11). The 90% confidence set contained the correct alternative from 72% to 92% of the time (mean 87, s.d. 6) when a .75 mixing parameter was used with samples of size 100. The procedure did not perform satisfactorily for the other mixing parameters with samples of size 100. One would expect that a procedure that discriminates between contaminating densities would require a large sample since only part of the sample comes from the density of interest.

The particular densities that were excluded in each trial depended on how similar their shape was to the density used to generate the sample. For instance, if a sample of size 500 came from a .5 mixture of the uniform density and the triangular density $f(x) = 2x$, a quadratic density, $f(x) = 3x^2$, was excluded from a 90% confidence interval 30% of the time. On the other hand, the density $f(x) = 4x$, $x \leqq \frac{1}{2}$, $f(x) = 4 - 4x$, $x \geqq \frac{1}{2}$ was always excluded and the density $f(x) = 4x^3$ was excluded 68% of the time.

4. Proofs.

PROOF OF THEOREM 1. One must check that the double sequence $\{n^{-\frac{1}{2}}g(X_{ni})\}$ obeys the three conditions of the general form of the central limit theorem [3, page 493].

Let $\varepsilon > 0$, $S_n = \{x: |g(x)| > \varepsilon n^{\frac{1}{2}}\}$, and denote the complement of S_n by $\bar{S}_n$. The condition on a double sequence that imply a limiting normal distribution reduce to the condition that as $n \to \infty$,

$$n \int_{S_n} p_n(x) \to 0 ;$$
$$\int_{\bar{S}_n} (g(x))^2 p_n(x)\, dx \to \int (g(x))^2\, dx ;$$
$$n^{\frac{1}{2}} \int_{\bar{S}_n} g(x) p_n(x)\, dx \to \int g(x)h(x)\, dx .$$

By using the representation of $p_n(x)$ given in Definition 3, these conditions hold if the following integrals approach zero as $n \to \infty$.

$$n \int_{S_n} dx ; \tag{4.1}$$
$$n^{\frac{1}{2}} \int_{S_n} h(x)\, dx ; \tag{4.2}$$
$$\int_{S_n} k_n(x)\, dx ; \tag{4.3}$$
$$n^{-\frac{1}{2}} \int_{\bar{S}_n} (g(x))^2 h(x)\, dx ; \tag{4.4}$$
$$n^{-1} \int_{\bar{S}_n} (g(x))^2 k_n(x)\, dx ; \tag{4.5}$$
$$n^{\frac{1}{2}} \int_{\bar{S}_n} g(x)\, dx ; \tag{4.6}$$
$$n^{-\frac{1}{2}} \int_{\bar{S}_n} g(x) k_n(x)\, dx . \tag{4.7}$$

Let $I_A(x)$ denote the function that is 1 on the set A and zero elsewhere. The integrands of (4.1) — (4.3) are

$$nI_{S_n}(x) , \qquad n^{\frac{1}{2}}h(x)I_{S_n}(x) , \qquad k_n(x)I_{S_n}(x)$$

respectively. By the definition of S_n, $I_{S_n}(x) \leqq |g(x)|/n^{\frac{1}{2}}\varepsilon$ and so $(g(x))^2/\varepsilon^2$, $h(x)|g(x)|/\varepsilon$ and $m(x)g(x)/\varepsilon$ dominate these integrands. Furthermore each of the integrands has a pointwise limit of zero, since x is eventually in $\bar{S}_n$, so the Lebesgue dominated convergence theorem, LDCT, can be used to show that (4.1)–(4.3) approach zero. On $\bar{S}_n$, $|g(x)|/n^{\frac{1}{2}} \leqq \varepsilon$, which can be used with the LDCT to show that (4.4), (4.5) and (4.7) have the desired limit. The function $g(x)$ has a zero integral so (4.6) equals $-n^{\frac{1}{2}} \int_{S_n} g(x)\, dx$, which can be shown to approach zero by the same arguments used for (4.2).

The proof of Theorem 2 is facilitated by two lemmas.

LEMMA 1. *Let $\{X_{n1}, X_{n2}, \cdots, X_{nn}\}$ be a double sequence of uniform random variables independent in each row and let $|U_n(x)| < f(x)$, with $f(x)$ integrable. Then*

$$\lim_{n\to\infty} E(U_n(x)) = 0 \Rightarrow \frac{1}{n} \sum_{i=1}^n U_n(X_{ni}) \to_p 0 .$$

PROOF. One needs to check the three conditions of the law of large numbers, that for any $\varepsilon > 0$ with $A_n = \{x : |U_n(x)| > \varepsilon n\}$, specifically that

$$n \int_{A_n} dx \to 0 , \tag{4.8}$$

$$n^{-1} \int_{\bar{A}_n} (U_n(x))^2\, dx \to 0 , \tag{4.9}$$

$$\int_{\bar{A}_n} U_n(x)\, dx \to 0 , \tag{4.10}$$

as $n \to \infty$. The first integrand is dominated by $f(x)/\varepsilon$ and the second by $\varepsilon f(x)$; hence, the LDCT implies that the first two conditions hold. The last integral can be rewritten as $E(U_n(x)) - \int_{A_n} U_n(x)\, dx$. The first term approaches zero as a consequence of the hypotheses, and the second approaches zero by the LDCT.

LEMMA 2. *Making the same assumptions as in Lemma 1 and letting $\gamma = \int (h(x))^2\, dx$ we have that,*

$$\sum_{i=1}^n \log (p_n(X_{ni})) + \frac{\gamma}{2} - n^{-\frac{1}{2}} \sum_{i=1}^n h(X_{ni}) \to_p 0 . \tag{4.11}$$

PROOF. Let $r_n(x) = n^{-\frac{1}{2}}h(x) + n^{-1}k_n(x)$, so that $p_n(x) = 1 + r_n(x)$. The proof is based on the fact that $\log p_n(x)$ is approximately equal to $r_n(x) - \frac{1}{2}(r_n(x))^2$. Rewrite (4.11) as

$$\begin{aligned} &n^{-1} \sum_{i=1}^n k_n(X_{ni}) - (2n)^{-1} \sum_{i=1}^n [(h(X_{ni}))^2 - \gamma] \\ &\quad - n^{-1} \sum_{i=1}^n h(X_{ni})k_n(X_{ni})n^{-\frac{1}{2}} - (2n)^{-1} \sum_{i=1}^n (k_n(X_{ni}))^2 n^{-1} \\ &\quad + \sum_{i=1}^n [\log p_n(X_{ni}) - r_n(X_{ni}) + \tfrac{1}{2}(r_n(X_{ni}))^2] . \end{aligned} \tag{4.12}$$

Using Lemma 1, one can easily show that the first 4 terms in (4.12) approach zero in probability as $n \to \infty$. To show that the last term approaches zero in probability we must verify that conditions (4.8), (4.9) and (4.10) hold with $U_n(x)$ replaced by $n[\log p_n(x) - r_n(x) + \frac{1}{2}(r_n(x))^2]$. Let $n[\log p_n(x) - r_n(x) + \frac{1}{2}(r_n(x))^2]$ be denoted by $W_n(x)$. By Taylor's formula on $\log (1 + r_n(x))$, one has that

$$|W_n(x)| \leqq n(r_n(x))^3/3|1 + y|^3 ,$$

where y has the same sign as $r_n(x)$ and $|y| \leqq |r_n(x)|$. The set A_n, as in the proof of Lemma 1, is $\{x : |W_n(x)| > n\varepsilon\}$. Remember that $|r_n(x)| = |n^{-\frac{1}{2}}h(x) + n^{-1}k_n(x)| \leqq n^{-\frac{1}{2}}(|h(x)| + m(x))$ and let $B_n = \{x : |r_n(x)| < \frac{1}{2}\}$. On B_n, $\frac{1}{2} \leqq |1 + y| < \frac{3}{2}$, so if $|W_n(x)| > n\varepsilon$ we have that $\varepsilon < (r_n(x))^3(\frac{8}{3}) \leqq (\frac{1}{2})^2(\frac{8}{3})n^{-\frac{1}{2}}(|h(x)| + m(x))$. On $\bar{B}_n$, $\frac{1}{2} \leqq n^{-\frac{1}{2}}(|h(x)| + m(x))$. In either case $nI_{A_n}(x)$ is dominated by an integrable function so (4.8) holds. Similarly, on $\bar{A}_n \cap B_n$

$$|W_n(x)| \leqq \tfrac{4}{3}(|h(x)| + m(x))^2$$
$$n^{-1}|W_n(x)|^2 \leqq \varepsilon\tfrac{4}{3}(|h(x)| + m(x))^2, \qquad \text{and}$$
$$|W_n(x)| \leqq n^{-\frac{1}{2}}(\tfrac{8}{3})(|h(x)| + m(x))^3 .$$

On $\bar{A}_n \cap \bar{B}_n$

$$|W_n(x)| \leqq 4\varepsilon(|h(x)| + m(x))^2 ,$$
$$n^{-1}|W_n(x)|^2 \leqq 4\varepsilon^2(|h(x)| + m(x))^2 , \qquad \text{and}$$
$$I_{\bar{A}_n \cap \bar{B}_n}(x) \to 0 \quad \text{as} \quad n \to \infty .$$

Therefore $|W_n(x)|$ and $n^{-1}|W_n(x)|^2$ are dominated by integrable functions on $\bar{A}_n$ and $|W_n(x)|$ and $n^{-1}|W_n(x)|^2$ approach zero on $\bar{A}_n$. The LDCT implies (4.9) and (4.10).

Proof of Theorem 2. We use a corollary to Le Cam's first lemma [4, page 203] to show that P_n, the joint distribution of $\{X_{ni}\}_{i=1}^n$ is contiguous to Q_n, the uniform probability measure on $[0, 1]^n$. This result states that P_n and Q_n are contiguous if the likelihood ratio statistic for testing Q_n versus P_n is asymptotically log normal with mean $-\sigma^2/2$ and variance σ^2 under Q_n. By Theorem 1, $n^{-\frac{1}{2}} \sum_{i=1}^n h(X_{ni})$ has an asymptotic normal distribution with mean 0 and variance γ. Therefore, Lemma 2 implies that $\prod_{i=1}^n p_n(X_{ni})$ is asymptotically log normal with mean $-\gamma/2$ and variance γ.

Proof of Theorem 3. Let $g(x) = \sum_{j=1}^\infty b_j d_j(x)$. Then

$$E(\textstyle\sum_{j=1}^m b_j V_{nj} - n^{-\frac{1}{2}} \sum_{i=1}^n g(X_{ni}))^2 = E(n^{-\frac{1}{2}} \sum_{i=1}^n [\sum_{j=1}^m b_j d_j(X_{ni}) - g(X_{ni})])^2 . \tag{4.13}$$

Under H_0 the quantity in square brackets has zero expectation, hence expression (4.13) equals

$$\textstyle\int (\sum_{j=1}^m b_j d_j(X_{ni}) - g(X_{ni}))^2 \, dx$$

which approaches zero as $m \to \infty$. Convergence in quadratic mean implies convergence in probability. Theorem 2 shows that convergence in probability occurs under $\{H_n\}$ as well. The result follows from the asymptotic distribution of $n^{-\frac{1}{2}} \sum_{i=1}^n g(X_{ni})$, given by Theorem 1.

Proof of Theorem 4. Definition 3, of $p_n(x)$, implies that $\int h(x)\,dx = 0$, and so for $h(x) = \sum_{j=1}^\infty a_j d_j(x)$. Using the previous argument we find that

$$\textstyle\sum_{j=1}^m a_j V_{nj} \to_p n^{-\frac{1}{2}} \sum_{i=1}^n h(X_{ni}) .$$

By Lemma 2, $\sum_{j=1}^m a_j V_{nj} \to_p \sum_{i=1}^n \log (p_n(X_{ni})) + \gamma/2$. The most powerful test of H_0 versus H_n is that which rejects when $\sum_{i=1}^n \log p_n(X_{ni}) > k$.

PROOF OF THEOREM 5. Statements 1 and 2 are a consequence of Theorem 3 where one has chosen the sequence $\{b_j\}$ with $b_j = 0$ when $j > m$.

PROOF OF THEOREM 6. The definition for asymptotic efficiency used is due to Hájek and Šidák [4, page 267]. The proof is straightforward.

PROOF OF THEOREM 7. It is necessary to show that $\sum_{j=1}^{m} b_j V_{nj}$ has the asymptotic normal distribution with mean $\sum_{j=1}^{m} n^{\frac{1}{2}} b_j u_j$ and variance $\sum_{j=1}^{m} \sum_{k=1}^{m} b_j \Gamma_{ik} b_k$ for all possible sequences $b_1, b_2, \cdots, b_m$. This is a straightforward application of the Lindeberg–Feller theorem [1, page 187], noting the fact that $\sum_{j=1}^{m} b_j d_j(x)$ is bounded.

PROOF OF THEOREM 8. Since the statistic $\theta_n{}^*$ is a consistent estimate of θ, so is $\tilde{\theta}_n$, the restriction of $\theta_n{}^*$ to the parameter space of θ. We use $\tilde{\theta}_n$ rather than $\theta_n{}^*$ to insure that Γ_n remains invertible.

Expression (3.5) is a continuous function of $(V_{n1}, V_{n2}, \cdots, V_{nm}, \tilde{\theta}_n)$. Let Γ be the matrix obtained by substituting θ for $\tilde{\theta}_n$ in the definition of Γ_n, and let $\bar{\theta}_n$ be the function obtained by substituting Γ for Γ_n in the definition of $\hat{\theta}_n$. Since $\tilde{\theta} \rightarrow_p \theta$, the expression (3.5) approaches

$$(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\bar{\theta}_n \mathbf{a})'\Gamma^{-1}(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\bar{\theta}_n \mathbf{a}) .$$

This expression can be written as

$$(4.14) \qquad (\mathbf{V}_{n,m} - n^{\frac{1}{2}}\theta\mathbf{a})'(\Gamma^{-1} - \Gamma^{-1}\mathbf{a}(\mathbf{a}'\Gamma^{-1}\mathbf{a})^{-1}\mathbf{a}'\Gamma^{-1})(V_{n,m} - n^{\frac{1}{2}}\theta\mathbf{a}) .$$

The vector $(\mathbf{V}_{n,m} - n^{\frac{1}{2}}\theta\mathbf{a})$ is asymtotically multivariate normal with zero mean and covariance matrix Γ. By diagonalizing Γ and transforming by an appropriate orthogonal matrix, expression (4.14) can be written as the sum of squares of m-1 independent asymptotically normal variates with zero mean and unit variance.

Acknowledgment. I would like to thank my thesis advisor, Professor Fred Andrews, for his advice and encouragement.

REFERENCES

[1] CHUNG, K. L. (1968). *A Course in Probability Theory*. Harcourt, Brace & World, New York.
[2] DURBIN, J. and KNOTT, M. (1972). Components of Cramér-von Mises statistics I. *J. Roy. Statist. Soc. Ser. B* **34** 290–307.
[3] EISEN, M. (1969). *Introduction to Mathematical Probability Theory*. Prentice-Hall, Englewood Cliffs.
[4] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
[5] KENDALL, M. G. and STUART, A. (1963). *The Advanced Theory of Statistics I*. Hafner, New York.
[6] NEYMAN, J. (1937). Smooth test for goodness of fit. *Skand. Aktuarietidskr*. **20** 149–199.

EPI-STAT DIVISION
SIDNEY FARBER CANCER INSTITUTE
44 BINNEY STREET
BOSTON, MASSACHUSETTS 02115