The Annals of Statistics
1992, Vol. 20, No. 4, 2071–2086

ASYMPTOTIC COMPARISON OF CRAMÉR–VON MISES AND NONPARAMETRIC FUNCTION ESTIMATION TECHNIQUES FOR TESTING GOODNESS-OF-FIT

BY R. L. EUBANK[1] AND V. N. LARICCIA

Texas A & M University and University of Delaware

Two new statistics for testing goodness-of-fit are derived from the viewpoint of nonparametric density estimation. These statistics are closely related to the Neyman smooth and Cramér–von Mises statistics but are shown to have superior properties both through asymptotic and small sample analyses. Comparison of the proposed tests with the Cramér–von Mises statistic requires the development of a novel technique for comparing tests that are capable of detecting local alternatives converging to the null at different rates.

1. Introduction. The Cramér–von Mises (CVM) statistic is one of the most popular tools for testing the one sample goodness-of-fit (GOF) hypothesis. It is even commonly presented in elementary statistics courses. Despite its popularity, it is known from empirical studies that the CVM test has poor power against essentially all but location–scale alternatives to the null hypothesis. In this paper we show that there are some simple, even naive, statistics which are both more informative and better capable of detecting nonlocation–scale alternatives than the CVM statistic.

Let $X_1, \dots, X_n$ be a random sample from an absolutely continuous *d.f.* F, and consider testing the hypothesis H_0: $F = H$ for H some specified absolutely continuous *d.f.* Set $V_i = H(X_i)$, $i = 1, \dots, n$, and let $\tilde{D}_n$ be the empirical *d.f.* of the V_i. Then the CVM statistic for H_0 is

$$C_n^2 = n\int_0^1 \left(\tilde{D}_n(u) - u\right)^2 du = n \sum_{j=1}^{\infty} \frac{\tilde{a}_{jn}^2}{(j\pi)^2} \tag{1}$$

with

$$\tilde{a}_{jn} = n^{-1} \sum_{i=1}^{n} p_j(V_i) \tag{2}$$

for

$$p_j(u) = \sqrt{2}\cos j\pi u. \tag{3}$$

The $\tilde{a}_{jn}$ in (2) are called the *components* of the CVM statistic. Note that they also provide a test for H_0, since the null hypothesis is equivalent to

---

Received July 1990; revised September 1991.

[1]Research supported by NSF Grant DMS-90-24879.

AMS 1980 *subject classifications*. Primary 62G10; secondary 62E20.

Key words and phrases. Asymptotic efficiency, density estimation, Fourier series, high frequency alternatives.

$E\tilde{a}_{jn} = 0$, $j = 1, 2, \ldots$. In particular, $\tilde{a}_{1n}$ and $\tilde{a}_{2n}$ tend to be most sensitive to location and scale departures from the null, respectively [Durbin and Knott (1972) and Eubank, LaRiccia and Rosenstein (1987)]. The $\tilde{a}_{jn}$ for $j > 2$ are useful in detecting higher frequency alternatives to the null than those of the location–scale variety. As noted by Durbin and Knott (1972), the severe down weighting of the $\tilde{a}_{jn}$ in (1) for $j > 2$ would appear to explain the poor power properties for C_n^2 against nonlocation–scale alternatives that is generally observed in simulation experiments.

The CVM statistic is an example of an omnibus test statistic. Such statistics are designed for situations, where the alternative hypotheses are vague and therefore need to be consistent against all alternatives. If, instead, specific alternatives are of special interest, directional tests can be developed which focus their power in the direction of these alternatives. This strategy has the drawback that the resulting tests will generally not be consistent against all alternatives and will have poor power against alternatives other than those they were designed to detect.

Neyman smooth tests [Neyman (1937)] represent a compromise between directional and omnibus tests. An example of a smooth type test is

$$T_{nm} = n \sum_{j=1}^{m} \tilde{a}_{jn}^2, \tag{4}$$

for m some fixed positive integer. While this test will obviously be inconsistent for alternatives having $E\tilde{a}_{jn} = 0$, $j = 1, \ldots, m$, one might hope that a judicious choice of m would lead to gains in the power of T_{nm} relative to that of C_n^2 for many alternatives due to the uniform weighting of the components in (4). Similar statements can be made concerning other smooth and omnibus tests. In support of this we note that there are numerous empirical studies [e.g., Kopecky and Pierce (1979), Miller and Quesenberry (1979) and Rayner and Best (1986)], where smooth tests have been shown to be more powerful than common omnibus test statistics over a wide range of realistic alternatives.

One goal of the present article is to shed some light on why smooth tests can have superior power to omnibus tests in many cases of interest. Although our analysis focuses on C_n^2 and statistics such as T_{nm}, we believe the basic principles carry over to much more general settings and therefore have more wide reaching implications.

The viewpoint of GOF taken in this article parallels that of Parzen (1979): Namely, testing H_0 is equivalent to testing $d(u) \equiv 1$, where d is the comparison density function

$$d(u) = f(H^{-1}(u))/h(H^{-1}(u)), \qquad 0 < u < 1, \tag{5}$$

with h the H density and $H^{-1}(u) = \inf\{x\colon H(x) \geq u\}$. Thus, one method for testing H_0 is to estimate d using a consistent estimator $\hat{d}$ and then compare $\hat{d}$ to 1 using some suitable metric. This has the advantage that when H_0 is rejected $\hat{d}$ furnishes useful information as to why rejection occurred. In fact, $\hat{d}(H(x))h(x)$ provides an estimate of the true density for the data.

One natural estimator of d in (5) is the truncated Fourier (cosine) series estimator $\hat{d}_m = 1 + \sum_{j=1}^{m} \tilde{a}_{jn} p_j$. Using the squared $L_2[0,1]$ norm as a measure of distance then gives the test statistic $n\int_0^1 (\hat{d}_m(u) - 1)^2\,du$ for H_0. This coincides with T_{nm} in (4) except now m has the interpretation of a smoothing parameter for the density estimator. Thus, we can use data-driven methods such as cross validation for selecting m and would anticipate that m needs to grow with n, at some suitable rate, for $\hat{d}_m$ to be consistent.

In Section 2 we show that if $n, m \to \infty$ at appropriate rates and T_{nm} is recentered and rescaled appropriately, it will have a normal limiting distribution under both H_0 and Pitman type alternatives approaching the null at rate $m^{1/4}/\sqrt{n}$. In contrast, C_n^2 has nontrivial power against alternatives approaching the null as fast a $1/\sqrt{n}$. Despite the disparity in rates for local alternatives, we demonstrate that, in a certain asymptotic sense, T_{nm} can be expected to have superior power to C_n^2 for higher frequency Pitman type alternatives, thereby giving an analytic explanation for the empirical results previously cited.

In Section 3 we propose another omnibus test for H_0 and derive its asymptotic distribution theory. The test is based on the estimate $\hat{d}_\lambda(u) = 1 + \sum_{j=1}^{n} \tilde{a}_{jn} p_j/(1 + \lambda j^2)$ of $d(u)$. Here $\lambda > 0$ is a smoothing parameter to be determined by the user or, possibly, by some suitable bandwidth selector. The specific statistic that is considered is $S_{n\lambda} = n\int_0^1 (\hat{d}_\lambda(u) - 1)^2\,du$. If $\lambda \to 0$ at an appropriate rate, $S_{n\lambda}$ can detect alternatives converging to the null as fast as $1/\sqrt{n}\,\lambda^{1/8}$. For suitably chosen m and λ the Pitman asymptotic relative efficiency of $S_{n\lambda}$ to T_{nm} is found to exceed 1.46.

Finally, in Section 4, results from a small scale simulation are presented which demonstrate that the asymptotics of Sections 2 and 3 extend in principle to finite samples. The proofs of all results are collected in Section 5.

To conclude we note that tests for goodness-of-fit based on quadratic functionals of nonparametric density estimators have also been considered by Bickel and Rosenblatt (1973), Holst and Rao (1980) and Ghorai (1980). We will discuss these further in the sequel.

2. Cramér–von Mises versus sums of components. We begin by considering the performance of C_n^2 and T_{nm} under local alternatives to the null. Thus, it will be assumed that for each n a random sample $V_{1n}, \dots, V_{nn}$ is obtained having (comparison) density

$$d_n(u) = 1 + b(n)\delta(u) \tag{6}$$

with $\delta \in L_2[0,1]$ and $b(n) \to 0$ as $n \to \infty$. Classically, interest has focused on the case where $b(n) = 1/\sqrt{n}$ in (6). This derives from parametric statistics, where parameters can be estimated with $\sqrt{n}$ consistency. However, our tests are based on nonparametric estimators for which $\sqrt{n}$ consistency is no longer a possibility. Thus, we allow for more general alternatives with an initial goal being the characterization of $b(n)$ for T_{nm}.

For C_n^2 it is known that if $b(n) = n^{-1/2}$ in (6), then [Shorack and Wellner (1986)]

$$C_n^2 \to_d \sum_{j=1}^{\infty} \frac{(Z_j + \delta_j)^2}{(j\pi)^2}, \tag{7}$$

where "$\to_d$" signifies convergence in distribution, the Z_j are iid $N(0,1)$ random variables and

$$\delta_j = \int_0^1 \delta(u) p_j(u)\, du \tag{8}$$

with p_j defined in (3).

Let $\chi_a^2(b)$ denote a chi-squared random variable with a degrees of freedom and noncentrality parameter b. Now for fixed m, and $b(n) = n^{-1/2}$,

$$T_{nm} \to_d \sum_{j=1}^{m} (Z_j + \delta_j)^2 \sim \chi_m^2\left(\sum_{j=1}^{m} \delta_j^2 \right).$$

Thus, a test based on this statistic will not be consistent against any alternative for which $\delta_j = 0$ for $j = 1, \dots, m$. Consequently, we must let m grow with n.

Intuitively if m grows with n, one then has a sequence of approximate chi-squared random variables with increasing degrees of freedom. Hence, it is not entirely surprising that T_{nm} is asymptotically normal when recentered and rescaled correctly. The precise large sample properties of T_{nm} are summarized in the following theorem.

THEOREM 1. *Assume that $\delta \in L_2[0,1]$ and that $m = m(n) \to \infty$ in such a way that $m^5/n^2 \to 0$. Then, if $b(n) = m^{1/4}/\sqrt{n}$ in* (3),

$$Z_{nm} = (T_{nm} - m)/\sqrt{2m} \to_d Y_1,$$

where Y_1 is a $N(\|\delta\|^2/\sqrt{2}, 1)$ random variable.

One implication of Theorem 1 is that an asymptotic α-level test for H_0 can be obtained by rejecting the null hypothesis whenever Z_{nm} exceeds Z_α, the $100(1-\alpha)$ percentage point of the standard normal distribution. This test will have nontrivial power against alternatives converging to the null at rate $m^{1/4}/\sqrt{n}$. More specifically, its asymptotic power is a monotone increasing function of $\|\delta\|^2$, with

$$\inf_{\|\delta\|=1} \lim_{n\to\infty} P\left(Z_{nm} \geq Z_{n\alpha} | m^{1/4}\delta/\sqrt{n}\right) = 1 - \Phi\left(Z_\alpha - 1/\sqrt{2}\right), \tag{9}$$

where $Z_{n\alpha}$ is the $100(1-\alpha)$ percentile of Z_{nm} and $P(A|b(n)\delta)$ denotes the probability of the event A under the alternatives (6).

In practice a data driven procedure for selecting m might prove beneficial. For example, one could follow Hart (1985) and use the minimizer $\hat{m}$ of

$$R(m) = -(n+1)\sum_{j=1}^{m}\frac{\tilde{a}_{jn}^2}{n} + \frac{m}{n-1} + \sum_{j=1}^{m}\frac{\tilde{a}_{2j,n}}{n-1}. \tag{10}$$

Using arguments similar to those in Eubank and Hart (1992), it can be shown that Theorem 1 is no longer valid for such stochastic choices of m and that $T_{n\hat{m}}$ has a nonnormal limiting null distribution. Despite this fact $T_{n\hat{m}}$ was found to work well in the simulation study of Section 4.

Ghorai (1980) has studied the limiting distribution of statistics such as T_{nm} under the null hypothesis. Our Theorem 1 represents an extension of his Example 1 to local alternatives. Bickel and Rosenblatt (1973) give a parallel of Theorem 1 for kernel density estimators. They show their test can detect alternatives converging to the null at rate $n^{(1-\gamma/2)/2}$ when the bandwidth of the density estimator is chosen to decay like $n^{-\gamma}$. If we make the analogy that a bandwidth is like $1/m$ for a series estimator, then their local alternatives are the same as ours when the number of terms in the series is allowed to grow like n^{γ}.

Another example of nonparametric rates for $b(n)$ can be found in Holst and Rao (1980) who show that the Dixon (1940) test detects alternatives converging to the null at rate $n^{-1/4}$. The Dixon test can be viewed as a quadratic functional of a histogram type density estimator with the number of bins having the same order as the sample size. This corresponds, roughly, to taking $m = n$ in the density estimator $\hat{d}_m = 1 + \sum_{j=1}^{m}\tilde{a}_{jn}p_j$ and would result in $b(n) = n^{-1/4}$ in Theorem 1 if the condition $m^5/n^2 \to 0$ were not violated. Conventional smoothing practices would suggest choosing $m \ll n$. For example, optimal rates of growth for m in $\hat{d}_m$ are of the order $n^{1/3}$ and $n^{1/5}$ corresponding to comparison densities with square integrable derivatives and square integrable second derivatives with periodic first derivatives, respectively. Both rates satisfy the conditions of Theorem 1 and give local alternatives of the order $n^{-5/12}$ and $n^{-9/20}$ that are much closer to the $n^{-1/2}$ parametric rate.

From (7) and Theorem 1 it may appear that the CVM test will be more effective than T_{nm} in detecting alternatives that are close to the null. However, this is not entirely true. For example, it is easily shown that $\lim_{j\to\infty}\lim_{n\to\infty} P(C_n^2 \geq c_{n\alpha}|p_j/\sqrt{n}) = \alpha$, where $c_{n\alpha}$ is the $100(1-\alpha)$ percentile of C_n^2. Hence

$$\inf_{\|\delta\|=1}\lim_{n\to\infty} P\left(C_n^2 \geq c_{n\alpha}|\delta/\sqrt{n}\right) = \alpha \tag{11}$$

and unlike Z_{nm}, the asymptotic power of C_n^2 is not monotone increasing in $\|\delta\|^2$. In fact, letting

$$C^2 = \sum_{j=1}^{\infty}\frac{(Z_j+\delta_j)^2}{(j\pi)^2},$$

we have the following theorem.

THEOREM 2. *For each* $j = 1, 2, \ldots$ *let* δ_j *be a constant, and set* $\Delta_j(\cdot) = \delta_j p_j(\cdot)$. *Further, let* β *be an arbitrary element of* $(\alpha, 1)$ *and let* c_α *be the* $100(1-\alpha)$ *percentile of* C^2. *Then,* $\lim_{j\to\infty} P(C^2 > c_\alpha | \Delta_j) \geq \beta$, *if and only if* $\delta_j \sim Aj$ *for some constant* A.

This theorem and (11) are indicative of the inability of C_n^2 to detect higher frequency alternatives to H_0. Further, along with (9) they give an indication that the power of T_{nm} might be competitive with that of C_n^2 at higher frequencies. We now describe a stronger result along these lines.

Our interest will be focused on alternatives corresponding to certain subsets of the collection of square summable sequences l_2. For each integer m and constants $0 < \gamma_1 \leq \gamma_2 < \infty$ define

$$(12)\quad \mathscr{C}_m(\gamma_1, \gamma_2) = \left\{ \mathbf{c} = (c_1, c_2, \ldots) \in l_2 \colon \sum_{j=1}^{\infty} c_j^2 \leq \gamma_2,\ \sum_{j=1}^{m} c_j^2/j^2 \geq \gamma_1 \right\}.$$

Then we are able to establish the following result.

THEOREM 3. *Given* $\beta \in (\alpha, 1)$, *there exist values of* γ_1 *and* γ_2 *such that if* $m^2/n \to 0$

$$(13)\qquad \lim_{n\to\infty} \inf_{\mathbf{c}\in\mathscr{C}_m(\gamma_1,\gamma_2)} P\left(C_n^2 \geq c_{n\alpha} \,\middle|\, \sum_{j=1}^{m} c_j j p_j/\sqrt{n} \right) \geq \beta$$

and

$$(14)\qquad \lim_{n\to\infty} \inf_{\mathbf{c}\in\mathscr{C}_m(\gamma_1,\gamma_2)} P\left(Z_{nm} \geq Z_{n\alpha} \,\middle|\, m^{1/4} \sum_{j=1}^{m} c_j p_j/\sqrt{n} \right) \geq \beta.$$

One can show that (13) and (14) hold with equality if we allow different choices of γ_1 and γ_2 for C_n^2 and T_{nm}. It is unclear whether this remains true for the situation of the theorem, where both tests are studied under the same alternatives. Nonetheless, we can still regard Theorem 2 as indicating that C_n^2 and Z_{nm} will have comparable power for alternatives of the form $d_n(u) = 1 + n^{-1/2}\sum_{j=1}^{m} j c_j p_j(u)$ and $d_n(u) = 1 + m^{1/4} n^{-1/2} \sum_{j=1}^{m} c_j p_j(u)$, at least for β near 1. However, the higher frequency portions of these two alternatives (corresponding to p_j with $j > m^{1/4}$) will actually be closer to the null for Z_{nm} than for C_n^2. This result gives some theoretical justification for the experimental conclusion that smooth type tests have better power than CVM type statistics for many nonlocation–scale alternatives.

Rosenblatt (1975) and Ghosh and Huang (1991) compared the Bickel and Rosenblatt (1973) test to others that are capable of detecting alternatives converging at parametric rates. They show that the Bickel and Rosenblatt statistic can be more effective in detecting alternatives to uniformity that have sharp peaks at some finite number of points. Such densities can also be viewed as high frequency alternatives. Thus, one might consider their results as further evidence that the Cramér–von Mises statistic is not effective against higher frequency departures from H_0.

To conclude this section we note that our basic results can be extended, with some additional labor, to include other CVM type statistics and location–scale composite hypotheses. One can show, for example, that a parallel of Theorem 1 holds for the components of the Anderson–Darling statistic if $m^9/n^2 \to 0$.

3. Another omnibus test. Motivated by the connection between estimation of $d(\cdot)$ and the test statistic T_{nm} discussed in Section 1, we now propose another possible test for H_0. This procedure is shown to be preferable to T_{nm} on the basis of Pitman asymptotic relative efficiency.

The statistic to be considered is

$$S_{n\lambda} = n \sum_{j=1}^{n} \frac{\tilde{a}_{jn}^2}{\left(1 + \lambda j^2\right)^2}, \qquad \lambda > 0.$$

It is easy to see that $S_{n\lambda} = n\int_0^1 (\hat{d}_\lambda(u) - 1)^2\, du$, where

$$\hat{d}_\lambda(u) = 1 + \sum_{j=1}^{n} \frac{\tilde{a}_{jn} p_j(u)}{1 + \lambda j^2}$$

is an estimator for the comparison density $d(u)$. This estimator is similar to those considered in Wahba (1977) and stems from work on spline smoothing. Methods discussed in Wahba (1981) can be modified to obtain adaptive procedures for choosing λ from the data.

$S_{n\lambda}$ can also be viewed as a type of compromise between T_{nm} and C_n^2. While T_{nm} gives equal weight to the first m components and ignores the rest, roughly speaking, $S_{n\lambda}$ uniformly weights the first $\lambda^{-1/2}$ components and down weights the remaining ones.

The asymptotic distribution theory for $S_{n\lambda}$ is provided in our next theorem.

THEOREM 4. *Assume that $\delta' \in L_2[0,1]$ and that $n \to \infty$, $\lambda \to 0$ in such a way that $n\lambda^{5/4} \to \infty$. Then, if $b(n) = 1/n^{1/2}\lambda^{1/8}$ in* (3),

$$Z_{n\lambda} = \frac{S_{n\lambda} - \sum_{j=1}^{n}\left(1 + \lambda j^2\right)^{-2}}{\left\{2\sum_{j=1}^{n}\left(1 + \lambda j^2\right)^{-4}\right\}^{1/2}} \to_d Y_2,$$

where Y_2 is a $N(\|\delta\|^2/\sqrt{2l_2}, 1)$ random variable and $l_2 = \int_0^\infty (1 + y^2)^{-4}\, dy$.

The theorem states that for suitable sequences of λ's one can test H_0 by rejecting if $Z_{n\lambda} \geq Z_\alpha$. The resulting test will have nontrivial asymptotic power against alternatives converging to the null at the rate $1/\sqrt{n}\,\lambda^{1/8}$.

The square root of λ is essentially a bandwidth for the estimator $\hat{d}_\lambda$ and plays the same role as $1/m$ for Z_{nm} in Section 2. Thus, we may align the local alternatives in such a way that Pitman asymptotic relative efficiencies can be computed.

COROLLARY . *Assume the conditions of Theorems* 1 *and* 4 *hold and let* $m \sim C_T n^{\gamma}$ *and* $\lambda^{1/2} \sim C_S n^{-\gamma}$ *for some* $0 < \gamma < 1$. *Then, the asymptotic relative efficiency of* $Z_{n\lambda}$ *to* Z_{nm} *is* $(C_T C_S / l_2)^{1/(2-\gamma)}$.

Assume that d has a square integrable derivative. Then "optimal" choices for m and λ are provided by taking $\gamma = 1/3$, $C_T = [\|\delta'\|^2/\pi^2]^{1/3}$ and $C_S = [\pi^2 l_1/\|\delta'\|^2]^{1/3}$ with $l_1 = \int_0^{\infty}(1+y^2)^{-2}\,dy$. The asymptotic relative efficiency then becomes $(l_1^{1/3}/l_2)^{3/5} \doteq 1.46$. So, $S_{n\lambda}$ leads to a test that can be asymptotically over 40% more efficient than T_{nm}.

It is also possible to establish a parallel of Theorem 3 for comparing $S_{n\lambda}$ and C_n^2. The alternatives to be considered will now be of the form $d_n(u) = 1 + n^{-1/2}\lambda^{-1/8}\sum_{j=1}^{n} c_j(1+\lambda j^2)p_j$ and $d_n(u) = 1 + n^{-1/2}\sum_{j=1}^{n} c_j j p_j$ for $Z_{n\lambda}$ and C_n^2, respectively. The same essential conclusions hold for this case.

4. Finite sample comparisons. To determine if the asymptotic results of the previous sections reflect the finite sample properties of our tests for fixed alternatives, a simulation experiment was conducted. A variety of alternatives were examined along with different choices for m and λ.

All our experiments were based on samples of size 50. To detect the sensitivity of $S_{n\lambda}$ and T_{nm} to the selection of λ and m, several values were examined for these parameters. Specifically, $\lambda = 10^{-1}$ and 10^{-2} and $m = 3$, 6 and 9 were used. Since $n = 50$, $\lambda = 10^{-1}$ and 10^{-2} correspond roughly to uniform weightings of the first three and five Fourier coefficients, respectively. We also considered the performance of $T_{n\hat{m}}$ with $\hat{m}$ the minimizer of (10) to see the effect of a stochastic choice for m.

Critical values for C_n^2 were taken from Shorack and Wellner (1986). Appropriate 5% and 10% level critical values for T_{nm}, $S_{n\lambda}$ and $T_{n\hat{m}}$ were then found by simulation from the null distribution of these statistics. In doing this 5000 replicate samples of size 50 were used.

Once appropriate critical values had been determined the basic experiment was replicated 1000 times. A different random seed was used for each case. The results are shown in Tables 1–3 as the proportion of rejections in 1000 samples of size 50 for each test.

TABLE 1

Proportion of rejections in 1000 *samples with* $n = 50$ *and* $\alpha = 0.10$ *for cosine alternatives*

	T_{nm}			$T_{n\hat{m}}$	$S_{n\lambda}$		C_n^2
(ρ, j)	$m = 3$	$m = 6$	$m = 9$		$\lambda = 10^{-1}$	$\lambda = 10^{-2}$	
(0.5, 1)	0.664	0.530	0.444	0.609	0.757	0.569	0.795
(0.5, 2)	0.658	0.538	0.445	0.563	0.626	0.559	0.346
(1, 3)	1.000	0.998	0.995	0.999	0.992	0.997	0.614
(1, 4)	0.141	1.00	0.998	0.997	0.867	0.999	0.311
(1, 8)	0.094	0.130	0.998	0.983	0.145	0.935	0.118
(1, 9)	0.117	0.113	0.995	0.983	0.139	0.829	0.119

TABLE 2
Values of ρ for which T_{nm} and C_n^2 have equivalent empirical power

	$m = 3$		$m = 6$		$m = 9$		
j	ρ	$m^{1/4}/j$	ρ	$m^{1/4}/j$	ρ	$m^{1/4}/j$	$P(C_n^2 > c_{n\alpha} \mid \rho = 1)$
2	0.800	0.860	0.900	0.783	0.900	0.658	0.985
3	0.480	0.577	0.520	0.522	0.580	0.439	0.614
4	—	—	0.330	0.391	0.380	0.329	0.311
5	—	—	0.250	0.313	0.270	0.263	0.195
6	—	—	0.190	0.260	0.220	0.219	0.153
7	—	—	—	—	0.140	0.188	0.125
8	—	—	—	—	0.110	0.165	0.118
9	—	—	—	—	0.110	0.146	0.119

Our first type of alternative was included to observe the behavior of T_{nm}, $S_{n\lambda}$, C_n^2 and $T_{n\hat{m}}$ against high frequency alternatives to H_0. The null density for this case was the uniform with cosine alternatives,

$$d_j(x;\rho) = 1 + \rho\cos(\pi jx).$$

The choice of ρ determines the distance of the alternative from the null density, while j can be manipulated to obtain higher or lower frequency departures from uniformity. For Table 1 we chose $j = 1$, 2, 3, 4, 8 and 9, and for each j, ρ was selected so that reasonable comparisons between the tests could be made.

The results in Table 1 for the cosine alternatives are as predicted by the asymptotic analysis. That is, for each value of m the power of T_{nm} is both

TABLE 3
Proportion of rejections in 1000 *samples with* $n = 50$ *and* $\alpha = 0.05$ *for beta and sine alternatives*

	T_{nm}			$T_{n\hat{m}}$	$S_{n\lambda}$		C_n^2
(a,b)	$m = 3$	$m = 6$	$m = 9$		$\lambda = 10^{-1}$	$\lambda = 10^{-2}$	
			Beta Alternatives				
(3, 3)	0.997	0.993	0.968	0.982	0.995	0.993	0.809
(2, 2)	0.735	0.644	0.537	0.608	0.666	0.664	0.206
(1.5, 1.5)	0.267	0.203	0.157	0.209	0.208	0.207	0.060
(0.5, 0.5)	0.735	0.834	0.813	0.836	0.734	0.873	0.410
(2, 3)	0.985	0.972	0.917	0.951	0.992	0.974	0.967
(3, 2)	0.990	0.966	0.915	0.945	0.993	0.972	0.958
(1.5, 2)	0.689	0.570	0.466	0.561	0.718	0.609	0.555
(2, 1.5)	0.665	0.546	0.447	0.548	0.717	0.598	0.556
(0.8, 1.5)	0.916	0.827	0.751	0.857	0.955	0.868	0.964
(1, 0.5)	0.979	0.979	0.977	0.977	0.991	0.986	0.988
			Sine Alternatives				
$j = 2$	1.00	1.00	1.00	1.00	1.0	1.00	0.872
$j = 4$	0.397	0.986	0.998	0.999	0.416	0.998	0.115

stable and high over the first m frequencies. It then drops off dramatically. The same remarks hold for the power of $S_{n\lambda}$, although the drop in power is more gradual. It therefore appears that both T_{nm} and $S_{n\lambda}$ can have excellent power against trigonometric type alternatives if m and λ are chosen correctly. On the other hand, the power of C_n^2 decays drastically for $j > 1$. The adaptive test $T_{n\hat{m}}$ appears to have power near that of the best of the three tests based on deterministic selection of m without their associated drops in power.

The second part of the simulation was aimed at detecting the degree to which the implications of Theorem 3 would be realized in finite samples. According to the theorem, T_{nm} should have about the same power against $d_j(\cdot, m^{1/4}/j)$ as C_n^2 has against $d_j(\cdot, 1)$ for $m^{1/4} < j < m$. Table 2 records the values of ρ for which T_{nm} has the same (or greater) empirical power under alternative $d_j(\cdot, m^{1/4}/j)$ as C_n^2 obtained against $d_j(\cdot, 1)$. For example, when $m = 4$, we see that C_n^2 has power 0.311 against $d_4(\cdot, 1)$ while T_{n6} had power at least that large against the alternatives $d_4(\cdot, 0.33)$. This is actually better than predicted by Theorem 3 since $0.33 < m^{1/4}/j = 0.391$, in this case. The values of ρ in the table are frequently less than $m^{1/4}/j$ and in all cases less than 1 so that T_{nm} is in fact detecting alternatives that are closer to the null than those for C_n^2.

Finally, cases in which the alternative did not lie in the direction of any specific cosine function were considered. Since $d(\cdot)$ is a density on $[0, 1]$ a wide class of alternatives is provided by choosing $d(\cdot)$ to be a beta density $d(u) = (\Gamma(a + b)/\Gamma(a)\Gamma(b))u^{a-1}(1 - u)^{b-1}$. To obtain multimodal alternatives we also considered comparison densities of the form $d(u) = 1 + \sin(\pi j u)$ with $j = 2$ and 4. The results of these experiments are reported in Table 3. For some cases where $d(\cdot)$ was J shaped or skewed unimodal (i.e., $a \neq b$) C_n^2 had slightly higher power than the nonparametric function estimation type tests. However, when $d(\cdot)$ was either symmetric, U shaped or multimodal ($a = b$ for the beta and all the sine functions), T_{nm}, $S_{n\lambda}$ and $T_{n\hat{m}}$ all significantly outperformed C_n^2.

In summary the experimental results support the asymptotic analysis. They indicate that, for anything but lower frequency alternatives to H_0, either $S_{n\lambda}$ or T_{nm} are to be preferred to C_n^2. However, the power of both $S_{n\lambda}$ and T_{nm} can be quite dependent on the choice of λ and m. The simulation suggests that it may be feasible to use data driven methods for choosing the smoothing parameters to overcome this problem.

5. Proofs. In this section we establish Theorems 1–4. We begin with some notational preliminaries.

For any density of the form (6) we denote its *d.f.* by $D_n(u) = u + b(n)\int_0^u \delta(t)\,dt$. The local alternatives being considered are then sequences of random samples $V_{1n}, \dots, V_{nn}$ with *d.f.*'s D_n. We also use the notation $c = \sum_{j=1}^{\infty}(j\pi)^{-2}$ and

$$a_{jn} = E\tilde{a}_{jn} = b(n)\delta_j \tag{15}$$

in all that follows.

PROOFS OF THEOREMS 1 AND 4. The proofs of Theorems 1 and 4 rely on two lemmas. The first lemma can be verified by direct calculation while the second is a consequence of Proposition 3.2 in de Jong (1987).

LEMMA 1. *For local alternatives of the form* (6):

(i) $a_{jn} = O(b(n))$, *uniformly in* j,

(ii) $$\operatorname{Cov}(\tilde{a}_{jn}, \tilde{a}_{kn}) = \begin{cases} (1 + O(b(n)))/n, & j = k, \\ O(b(n)/n), & j \neq k, \end{cases}$$

uniformly in j *and* k *and*

(iii) $\int_0^1 (p_j(u) - a_{jn})^4 d_n(u)\,du = O(1)$ *uniformly in* j.

LEMMA 2. *Let* $\{Y_{in}\}_{i=1}^n$, $n = 1, 2, \dots$ *be a triangular array of random variables that are* iid *within rows. Set* $w_{ijn} = w_{ijn}(Y_{in}, Y_{jn}) + w_{jin}(Y_{jn}, Y_{in})$ *for some function* $w_{ijn}(\cdot,\cdot)$ *and assume that* $Ew_{ijn}|Y_{in} = 0$ *for all* $i, j \le n$. *Define*

$$w(n) = \sum_{1 \le i < j \le n} w_{ijn},$$

$$\sigma(n)^2 = \operatorname{Var} w(n) = \sum_{1 \le i < j \le n} Ew_{ijn}^2,$$

$$G_{\mathrm{I}} = \sum_{1 \le i < j \le n} Ew_{ijn}^4,$$

$$G_{\mathrm{II}} = \sum_{1 \le i < j < k \le n} \left[Ew_{ijn}^2 w_{ikn}^2 + Ew_{jin}^2 w_{jkn}^2 + Ew_{kin}^2 w_{kjn}^2 \right],$$

and

$$G_{\mathrm{IV}} = \sum_{1 \le i < j < k < l \le n} \left[Ew_{ijn} w_{ikn} w_{ljn} w_{lkn} + Ew_{ijn} w_{iln} w_{kjn} w_{kln} + Ew_{iln} w_{ikn} w_{jkn} w_{jln} \right].$$

Then, if G_{I}, G_{II} *and* G_{IV} *are all of smaller order than* $\sigma(n)^4$,

$$w(n)/\sigma(n) \to_d N(0,1).$$

The proof of Theorem 1 now proceeds in the following fashion. Write

$$Z_{nm} = \left(n \sum_{j=1}^m (\tilde{a}_{jn} - a_{jn})^2 - m \right) \Big/ \sqrt{2m} + R_n$$

with

$$R_n = \left[2n \sum_{j=1}^m (\tilde{a}_{jn} - a_{jn}) a_{jn} + n \sum_{j=1}^m a_{jn}^2 \right] \Big/ \sqrt{2m}.$$

We handle the first term using Lemma 2 and show that $R_n \to_P \|\delta\|^2/\sqrt{2}$, where " $\to_P$ " denotes convergence in probability.

To see that $R_n \to_P \|\delta\|^2/\sqrt{2}$, note that

$$ER_n = n\sum_{j=1}^{m}\frac{a_{jn}^2}{\sqrt{2m}} = \frac{nb(n)^2}{\sqrt{2m}}\sum_{j=1}^{m}\delta_j^2 = \sum_{j=1}^{m}\frac{\delta_j^2}{\sqrt{2}} \to \frac{\|\delta\|^2}{\sqrt{2}}$$

while the variance of R_n is $O(m^{-1/2} + m^{5/4}/\sqrt{n})$. The result then follows from Chebychev's inequality.

Now write

$$\left(n\sum_{j=1}^{m}(\tilde{a}_{jn} - a_{jn})^2 - m\right)\Big/\sqrt{2m} = (w_1(n) + w(n))/\sqrt{2m}$$

for

$$w_1(n) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(p_j(V_{in}) - a_{jn})^2 - m$$

and

$$w(n) = \sum_{1\le i<j\le n} w_{ijn}$$

with

$$w_{ijn} = \frac{2}{n}\sum_{s=1}^{m}(p_s(V_{in}) - a_{sn})(p_s(V_{jn}) - a_{sn}).$$

We see that $w_1(n)/\sqrt{m} \to_P 0$ since $Ew_1(n) = O(m^{5/4}/\sqrt{n})$ and $\operatorname{Var} w_1(n) = O(m^{5/2}/n)$, by Lemma 1. The random variable $w(n)$ satisfies the conditions of Lemma 2 and tedious calculations reveal that $\sigma(n)^2 = 2m(1 + o(1))$, $G_{\mathrm{I}} = O((m/n)^4)$, $G_{\mathrm{II}} = O(m^2/n)$ and $G_{\mathrm{IV}} = O(m)$. For example, we have

$$Ew_{ijn}^2 w_{ikn}^2 = \frac{16}{n^4}\sum_r\sum_s\sum_l\sum_t EA_{ri}A_{rj}A_{si}A_{sj}A_{li}A_{lk}A_{ti}A_{tk}$$

with $A_{\alpha\nu} = p_\alpha(V_{\nu n}) - a_{\alpha n}$. Now use the independence of the V_{jn}, the Cauchy–Schwarz inequality and Lemma 1 to see that this expression is $O(m^2/n^4)$ if $m^5/n^2 \to 0$ and verify our claim for G_{II}. Thus, Theorem 1 has been proved.

The proof of Theorem 4 proceeds along similar lines. Set $e_{1n} = \sum_{j=1}^n (1 + \lambda j^2)^{-2}$, $e_{2n} = \sum_{j=1}^n(1 + \lambda j^2)^{-4}$, $l_1 = \int_0^\infty(1 + y^2)^{-2}\,dy$ and $l_2 = \int_0^\infty(1 + y^2)^{-4}\,dy$. Then write

$$Z_{n\lambda} = \left(n\sum_{j=1}^{n}\frac{(\tilde{a}_{jn} - a_{jn})^2}{(1 + \lambda j^2)^2} - e_{1n}\right)\Big/\sqrt{2e_{2n}} + R_n$$

with

$$R_n = \left(2n\sum_{j=1}^{n}\frac{(\tilde{a}_{jn} - a_{jn})a_{jn}}{(1 + \lambda j^2)^2} + n\sum_{j=1}^{n}\frac{a_{jn}^2}{(1 + \lambda j^2)^2}\right)\Big/\sqrt{2e_{2n}}.$$

We first show that $ER_n \to \|\delta\|^2/\sqrt{2l_2}$.

Arguing as in Wahba (1975) we can show that $e_{jn} = \lambda^{-1/2} l_j(1 + o(1))$ with $o(1) \to 0$ as $\lambda \to 0$. Therefore,

$$ER_n \sim \lambda^{1/4} n \sum_{j=1}^{n} a_{jn}^2/(1 + \lambda j^2)^2 \sqrt{2l_2} = \sum_{j=1}^{n} \delta_j^2/(1 + \lambda j^2)^2 \sqrt{2l_2}.$$

Let $g_n = \sum_{j=1}^n \delta_j p_j$ and observe that $\|g_n - g\|^2 + (\lambda/\pi^2)\|g'\|^2$ is minimized over all functions g with $g' \in L_2[0, 1]$ by $g_{n\lambda} = \sum_{j=1}^n \delta_j p_j/(1 + \lambda j^2)$. Thus, for any g with $g' \in L_2[0, 1]$, $\|g_n - g_{n\lambda}\|^2 \le \|g_n - g\|^2 + (\lambda/\pi^2)\|g'\|^2$. In particular, we have $\|g_n - g_{n\lambda}\|^2 \le (\lambda/\pi^2)\|g_n\|^2 \le (\lambda/\pi^2)\|\delta\|^2$. Therefore,

$$\left| \sum_{j=1}^{n} \delta_j^2 - \sum_{j=1}^{n} \delta_j^2/(1 + \lambda j^2)^2 \right| \le \|g_n - g_{n\lambda}\| + 2\|g_n - g_{n\lambda}\| \, \|\delta\| = O(\sqrt{\lambda}).$$

To finish the first part of the proof note that $n\sum_{j=1}^n (\tilde{a}_{jn} - a_{jn}) a_{jn}/\sqrt{e_{2n}}$ has zero expectation and variance of order $\lambda^{1/4}$. Consequently, $R_n \to_P \|\delta\|^2/\sqrt{2l_2}$.

Now set $w_{ijn} = (2/n)\sum_{k=1}^n (p_k(V_{in}) - a_{kn})(p_k(V_{jn}) - a_{kn})/(1 + \lambda k^2)^2$, $w(n) = \sum_{1 \le i < j \le n} w_{ijn}$ and define $w_1(n) = n^{-1}\sum_{i=1}^n w_{iin} - e_{1n}$. Using Lemma 1 and the fact that $\sum_{j=1}^n (1 + \lambda j^2)^{-k} = O(\lambda^{-1/2})$ for $k = 1, 2, \dots$, we find that $Ew_1(n) = O(1/\sqrt{n}\lambda^{5/8})$ and $\operatorname{Var} w_1(n)/e_{2n} = O(1/\sqrt{n}\lambda^{5/8})$. So, $w_1(n) \to_P 0$ if $n\lambda^{5/4} \to \infty$.

To employ Lemma 2, calculations reveal that $\sigma(n)^2 \sim \lambda^{-1/2} l_2$, $G_{\mathrm{I}} = O((n\lambda)^{-2})$, $G_{\mathrm{II}} = O((n\lambda)^{-1})$ and $G_{\mathrm{IV}} = O(\lambda^{-1/2})$. This completes the proof of Theorem 4. □

Proof of Theorem 2. The sufficiency of $\delta_j \sim Aj$ follow from the inequality

$$P\left(C^2 > c_\alpha | \Delta_j\right) \ge P\left((Z_j + \delta_j)^2 > c_\alpha (j\pi)^2\right) = 1 - \Phi\left(\sqrt{c_\alpha} j\pi - \delta_j\right) + \Phi\left(-\sqrt{c_\alpha} j\pi - \delta_j\right).$$

Take $\delta_j = \sqrt{c_\alpha} j\pi - \Phi^{-1}(1 - \beta)$ to get β as a limiting lower bound.

To see the necessity of the condition set $C_0^2 = \sum_{j=1}^\infty Z_j^2/(j\pi)^2$ and note that, for any $\varepsilon > 0$,

$$P\left(C^2 > c_\alpha | \Delta_j\right) \le P\left(C_0^2 > c_\alpha - \varepsilon^2, |Z_j + \delta_j|/j\pi \le \varepsilon\right) + P\left(|Z_j + \delta_j|/j\pi > \varepsilon\right) \le P\left(C_0^2 > c_\alpha - \varepsilon^2\right) + 1 - \Phi(\varepsilon j\pi - \delta_j) + \Phi(-\varepsilon j\pi - \delta_j).$$

If $|\delta_j|$ grows slower than j this upper bound limits to $P(C_0^2 > c_\alpha - \varepsilon^2)$. Now let ε tend to zero to finish the proof. □

Proof of Theorem 3. We begin by establishing (12). For this purpose let $\delta_m(\cdot; \mathbf{c}) = \sum_{j=1}^n jc_j p_j$. Then, for $\mathbf{c} \in \mathscr{C}_m(\gamma_1, \gamma_2)$,

$$P\left(C_n^2 \ge c_{n\alpha} | n^{-1/2}\delta_m(\cdot; \mathbf{c})\right) \ge P\left(A_{nm} \ge (\pi^2 c_{n\alpha} - \gamma_1)/2 | n^{-1/2}\delta_m(\cdot; \mathbf{c})\right)$$

with $A_{nm} = \sum_{j=1}^m \sqrt{n}(\tilde{a}_{jn} - jc_j/\sqrt{n})c_j/j$.

Now $A_{nm}/\sqrt{n}$ is the average of n iid random variables $Y_{in} = \sum_{j=1}^{m} c_j[p_j(V_{in}) - jc_j/\sqrt{n}]/j$, $i = 1, \dots, n$. The Y_{in} have mean zero and variance of order $\sum_{j=1}^{m} c_j^2/j^2 + O(m^{3/2}/\sqrt{n})$, because $\sum_{j=1}^{\infty} c_j^2 \le \gamma_2$. Also,

$$\left\{E|Y_{in}|^3\right\}^{1/3} \le \sum_{j=1}^{m}\left(\frac{|c_j|}{j}\right)\left[\int_0^1 \left|p_j(u) - jc_j/\sqrt{n}\right|^3 \left(1 + n^{-1/2}\delta_m(u;\mathbf{c})\right) du\right]^{1/3}.$$

Since

$$\left(\int_0^1 \left|p_j(u) - jc_j/\sqrt{n}\right|^3 \left(1 + n^{-1/2}\delta_m(u;\mathbf{c})\right) du\right)^{1/3} = 1 + O\left(m^{3/2}/\sqrt{n}\right),$$

$$E|Y_{in}|^3 \le \left(\gamma_2 \pi^2 c\right)^{1/2}\left(1 + O\left(m^{3/2}/\sqrt{n}\right)\right),$$

uniformly over $\mathbf{c} \in \mathscr{C}_m(\gamma_1, \gamma_2)$. An application of the Berry–Esseén theorem [Serfling (1980) page 33] then shows that

$$\sup_t \left| P\left(\sqrt{n}\sum_{i=1}^{n} Y_{in}/(\operatorname{Var} Y_{1n})^{1/2} \ge t\right) - 1 + \Phi(t)\right|$$

$$\le \left(\gamma_2\pi^2 c\right)^{1/2}\left(1 + O\left(m^{3/2}/\sqrt{n}\right)\right)/\sqrt{n}\left(\gamma_1 + O\left(m^{3/2}/\sqrt{n}\right)\right)^{3/2},$$

for each value of n. Thus,

$$P\left(A_{nm} \ge (\pi^2 c_{n\alpha} - \gamma_1)/2 \,|\, \delta_m(\cdot;\mathbf{c})/\sqrt{n}\right)$$

$$\ge 1 - \Phi\left((\pi^2 c_{n\alpha} - \gamma_1)/2(\operatorname{Var} Y_{in})^{1/2}\right) + O(n^{-1/2})$$

uniformly over $\mathbf{c} \in \mathscr{C}_m(\gamma_1, \gamma_2)$. Taking limits we then find that $\lim_{n\to\infty} \inf_{\mathbf{c}\in\mathscr{C}_m(\gamma_1,\gamma_2)} P(C_n^2 \ge c_{n\alpha}|n^{-1/2}\delta_m(\cdot;\mathbf{c}))$ is bounded below by the smaller of $1 - \Phi((\pi^2 \lim c_{n\alpha} - \gamma_1)/2\sqrt{\gamma_1})$ or $1 - \Phi((\pi^2 \lim c_{n\alpha} - \gamma_1)/2\sqrt{\gamma_2})$ and (12) has been shown.

To prove (13) set $\delta_m(\cdot;\mathbf{c}) = \sum_{j=1}^{m} c_j p_j(\cdot)$ and observe that

$$P\left(Z_{nm} \ge Z_\alpha \middle| m^{1/4}\delta_m(\cdot;\mathbf{c})/\sqrt{n}\right)$$

$$\ge P\left(n\left(\sum_{j=1}^{m}\left(\tilde{a}_{jn} - a_{jn}\right)^2 - m\right)/\sqrt{2m}\right.$$

$$\left. + \sqrt{n}\,\overline{Y}_n \ge Z_\alpha - \gamma_1/\sqrt{2}\,\middle| m^{1/4}\delta_m(\cdot;\mathbf{c})/\sqrt{n}\right)$$

with

$$\sqrt{n}\,\overline{Y}_n = -2\sum_{i=1}^{n}\sum_{j=1}^{m} c_j\left(p_j(V_{in}) - c_j m^{1/4}/\sqrt{n}\right)/\sqrt{2nm^{1/2}}.$$

Now $E\overline{Y}_n = 0$ and $\operatorname{Var}\overline{Y}_n = O(1/\sqrt{m}\,n)$ uniformly in c. Consequently, it suffices to work with $A_{nm} = (\sum_{j=1}^{m}(\tilde{a}_{jn} - a_{jn})^2 - m)/\sqrt{2m}$.

Let $U_1, U_2, \dots$ be iid uniform $(0, 1)$ random variables and take G_n to be the empirical $d.f.$ of $U_1, \dots, U_n$. Defining the functional L by

$$L(G) = \left(\sum_{j=1}^{m} \left(\int_0^1 P_j(u) G(u)\, du \right)^2 \Big/ (j\pi)^2 - m \right) \Big/ \sqrt{2m}$$

for $P_j(u) = \sqrt{2} \sin \pi j u$ and $G \in L_\infty[0, 1]$, an integration by parts reveals that $A_{nm} = L(\sqrt{n}(\tilde{D}_n - D_n))$ with $\tilde{D}_n$ the empirical $d.f.$ of $V_{1n}, \dots, V_{nn}$. Since $L(\sqrt{n}(\tilde{D}_n - D_n))$ and $L(\sqrt{n}(G_n \circ D_n - D_n))$ have the same distribution, we can restrict attention to the properties of $L(\sqrt{n}(G_n \circ D_n - D_n))$ in what follows.

A direct calculation using the Gateaux derivative of L shows that

$$\begin{aligned} &|L(G_1) - L(G_2)| \\ &\quad \le (4c/\sqrt{m}) \sup_s |G_1(s) - G_2(s)| \big(\sup_s |G_1(s) - G_2(s)| + \sup_s |G_1(s)|\big). \end{aligned}$$

It is known [Csörgő and Révész (1981) Theorem 4.4.1] that there exists a sequence of Brownian bridge processes $\{B_n(\cdot)\}$ such that $\sup_s |G_n(s) - B_n(s)| = O_p(\log n / \sqrt{n})$. Thus, we obtain

$$\big| L\big(\sqrt{n}(G_n \circ D_n - D_n)\big) - L(B_n \circ D_n) \big| \le O_p\big(\log n / \sqrt{nm}\big).$$

Similarly, using the modulus of continuity of the Brownian bridge process [Csörgő and Révész (1981) Theorem 1.4.1] and the fact that $\sup|t - D_n(t)| \le m^{1/4}(\gamma_2 c)^{1/2}/\sqrt{n}$, we find that $|L(B_n \circ D_n) - L(B_n)| = O_p((\log n + \log m)/m^{1/4}\sqrt{n})$. We note in passing that these calculations can be used in place of Lemma 2 in the proof of Theorem 1.

Combining all our approximations we find that, given $\varepsilon > 0$, there exists an n_0 such that for all $n \ge n_0$,

$$P\big(Z_{nm} \ge Z_\alpha | m^{1/4} \delta_m(\cdot\,; \mathbf{c})/\sqrt{n}\big) \ge P\left(\frac{\chi_m^2 - m}{\sqrt{2m}} \ge Z_\alpha - \frac{\gamma_1}{\sqrt{2}} + \varepsilon \right) + \frac{o(1)}{\varepsilon^2}$$

with $o(1) \to 0$ as $n \to \infty$ uniformly over $\mathbf{c} \in \mathscr{C}_m(\gamma_1, \gamma_2)$ and χ_m^2 a central chi-squared random variable having m degrees of freedom. By taking the infimum, letting $n \to \infty$ and then letting $\varepsilon \to 0$, the desired result is obtained. □

REFERENCES

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

CSÖRGŐ, M. and RÉVÉSZ (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

DIXON, W. J. (1940). A criterion for testing the hypothesis that two samples are from the same population. *Ann. Math. Statist.* **11** 199–204.

DURBIN, J. and KNOTT, M. (1972). Components of Cramér–von Mises statistics, I. *J. Roy. Statist. Soc. Ser. B* **34** 290–307.

EUBANK, R., LARICCIA, V. N. and ROSENSTEIN, R. (1987). Test statistics derived as components of Pearson's phi-squared distance measure. *J. Amer. Statist. Assoc.* **82** 816–825.

EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425

GHORAI, J. (1980). Asymptotic normality of a quadratic measure of orthogonal series type density estimate. *Ann. Inst. Statist. Math.* **32** 341–350.

GHOSH, B. K. and HUANG, W. (1991). The power and optimal kernel of the Bickel–Rosenblatt test for goodness-of-fit. *Ann. Statist.* **19** 999–1009.

HART, J. (1985). On the choice of a truncation point in Fourier series density estimation. *J. Statist. Comput. Simulation* **21** 95–116.

HOLST, L. and RAO, J. S. (1980). Asymptotic theory for some families of two-sample nonparametric statistics. *Sankhyā Ser. A* **42** 19–52.

KOPECKY, K. and PIERCE, D. (1979). Efficiency of smooth goodness-of-fit tests. *J. Amer. Statist. Assoc.* **74** 393–397.

MILLER, F. and QUESENBERRY, G. (1979). Power studies of tests for uniformity II. *Comm. Statist. B—Simulation Comput.* **8** 271–290.

NEYMAN, J. (1937). Smooth test for goodness-of-fit. *Skand. Aktuarietidskr.* **20** 149–199.

PARZEN, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74** 105–131.

RAYNER, J. and BEST, D. (1986). Neyman-type smooth tests for location-scale families. *Biometrika* **73** 437–446.

ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SHORACK, G. and WELLNER, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

WAHBA, G. (1975). Smoothing noisy data with spline functions. *Numer. Math.* **24** 383–393.

WAHBA, G. (1977). Optimal smoothing of density estimates. In *Classification and Clustering* (J. Van Ryzin, ed.) 423–458. Academic, New York.

WAHBA, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.* **9** 146–156.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19711