PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 275–289

COMPONENT TYPE TESTS WITH ESTIMATED PARAMETERS

BY

R. L. EUBANK (COLLEGE STATION, TEXAS),
V. N. LARICCIA AND J. H. SCHUENEMEYER (NEWARK, DELAWARE)

Abstract. Goodness-of-fit tests based on sums of squared components of the Cramér–von Mises statistic with a growing number of summands are studied in the case of a composite null hypothesis. The tests are seen to be related to nonparametric function estimation procedures and Neyman smooth tests. The large sample properties of the tests are examined under sequences of local alternatives and the proposed methodology is illustrated on real data sets.

1. Introduction. Neyman [12] proposed what he termed *smooth* tests for the classical goodness-of-fit hypothesis. These tests were found to provide useful diagnostic and inferential tools but, over the years, had seemingly fallen out of favor relative to omnibus tests of the Cramér–von Mises and Kolmogorov–Smirinov variety. Several recent investigations (e.g., [13] and [14]) have now brought these important statistics back into the mainstream of statistical literature. In this note we examine Neyman smooth tests for composite hypotheses from the perspective of nonparametric density estimation.

To introduce the basic ideas, we begin by discussing the problem of testing a simple goodness-of-fit hypothesis. More specifically, assume that $X_1, \ldots, X_n$ are i.i.d. random variables with common distribution function (d.f.) F. Given some specified d.f. F_0, the classical goodness-of-fit problem is concerned with testing

$$H_g\colon F(\cdot) = F_0(\cdot). \tag{1.1}$$

There are a number of alternate ways to state (1.1). Assume that F and F_0 are absolutely continuous with densities f and f_0 and set

$$F_0^{-1}(u) = \inf\{x\colon F_0(x) \geqslant u\}.$$

If we then define the *comparison density function* as

$$d(u) = f(F_0^{-1}(u))/f_0(F_0^{-1}(u)), \quad 0 \leqslant u \leqslant 1, \tag{1.2}$$

testing (1.1) becomes equivalent to

$$H_g^*\colon d(\cdot) = 1. \tag{1.3}$$

In view of (1.3) we see that the problem of assessing the goodness-of-fit of F_0 can be formulated as a problem of comparing the comparison density (1.2) to the unit function. Now

$$Y_i = F_0(X_i), \quad i = 1, \ldots, n, \tag{1.4}$$

are i.i.d. with density (1.2), and hence d can be estimated from the Y_i by using a variety of nonparametric density estimators. If $\hat{d}$ is such an estimator, then H_g^* can be tested by using some measure of the distance between $\hat{d}$ and the uniform density. In particular, one might base a test on a statistic such as

$$T = \int_0^1 (\hat{d}(u) - 1)^2\, du. \tag{1.5}$$

Statistics of this form have been studied, for example, by Bickel and Rosenblatt [4], Rosenblatt [15], and Ghorai [9].

For the purpose of this note we will be primarily interested in the case where $\hat{d}$ is a cosine series estimator. Define the sample (cosine) Fourier coefficients

$$a_j = \sqrt{\frac{2}{n}} \sum_{i=1}^{n} \cos(j\pi Y_i), \quad j = 1, 2, \ldots \tag{1.6}$$

Then a cosine series estimator of d is provided by

$$d_m(u) = 1 + n^{-1/2} \sum_{j=1}^{m} a_j \sqrt{2} \cos(j\pi u), \tag{1.7}$$

where m is some nonnegative integer that governs the smoothness or, equivalently, the bias to variance tradeoff for the estimator. If we choose $\hat{d}$ in (1.5) to be d_m in (1.7), we obtain

$$T_m = \sum_{j=1}^{m} a_j^2/n \tag{1.8}$$

as our test statistic.

It is well known that m must be allowed to grow with n if d_m is to provide a consistent estimator of d. (A typical rate of growth for m would be $m \propto n^{1/4}$ in this case if d is assumed to have a square integrable second derivative.) Thus, it is no surprise that m must also grow with n for tests based on T_m to be consistent against all alternatives. In [8] it was shown that if $n, m \to \infty$ in such a way that $m^5/n^2 \to 0$, then

$$Z_m = (nT_m - m)/\sqrt{2m} \tag{1.9}$$

has a limiting standard normal distribution under H_g^* and that tests obtained from Z_m are consistent against any fixed alternative.

The statistic T_m is related in a fundamental way to Neyman smooth tests. Neyman [12] developed a test for (1.3) versus the alternative

$$d(u) = c\exp\{\sum_{j=1}^{m} \beta_j p_j(u)\}, \quad u \in [0, 1],$$

where c is a constant that ensures $\int_0^1 d(u)\,du = 1$, the β_j are unknown constants and the p_j are Legendre polynomials. His statistic is $\sum_{j=1}^{m} b_j^2$ with

$$b_j = n^{-1/2} \sum_{i=1}^{n} p_j(Y_i).$$

Thus, T_m in (1.8) is essentially Neyman's statistic except we have used the cosine functions rather than the Legendre polynomials for our orthonormal basis. As such, this statistic is not new. Bases other than the Legendre functions have been used by a number of authors and, in particular, Rayner and Best [13], [14] discuss the use of cosine functions.

What distinguishes our proposal from the standard Neyman smooth testing paradigm is that we do not treat m as fixed but instead let m grow with n. This has an important advantage since, if m is fixed, it is easy to see that a smooth test will have only trivial power against any alternative that is orthogonal to the first m elements of the orthonormal basis used in constructing the test. If, instead, m grows with n and the basis is complete, this problem is no longer present.

Another way to view (1.8) is from the perspective of the Cramér–von Mises (CVM) test for H_g. Let

$$F_n(u) = \frac{\text{no } Y_i \leqslant u}{n}.$$

Then, the CVM statistic is

$$W^2 = n\int_0^1 (F_n(u)-u)^2\,du = \sum_{j=1}^{\infty} a_j^2/(j\pi)^2 \tag{1.10}$$

(cf. [6] or [19]). Thus, W^2 uses all the sample Fourier coefficients but downweights them according to increasing frequency. In contrast, the statistic T_m uses an increasing (with n) number of uniformly weighted Fourier coefficients in its construction. This has been shown (both analytically and empirically) to provide considerable gains in power for T_m over W^2 for many types of alternatives. See, e.g., [8] and [11].

The goal of the remainder of this paper is to extend the results in [8] to the case of a composite goodness-of-fit hypothesis. Thus, assume that the null model now involves a distribution function $F(\cdot;\theta)$ depending on a parameter θ belonging to some open subset Θ of the line and we want to test

$$H_0\colon F(\cdot) = F(\cdot;\,\theta), \quad \theta \in \Theta. \tag{1.11}$$

For this purpose define sample Fourier coefficients

$$a_j(\theta) = \sqrt{\frac{2}{n}} \sum_{i=1}^{n} \cos\big(j\pi F(X_i;\,\theta)\big), \quad j = 1, \ldots \tag{1.12}$$

Then, given an estimator $\hat{\theta}$ for θ, we propose to test H_0 in (1.11) using

$$\hat{T}_m = \sum_{j=1}^{m} a_j^2(\hat{\theta})/n. \tag{1.13}$$

In the next section we derive the asymptotic distribution theory for a standardized version of $\hat{T}_m$. This is followed by examples with real data in Section 3. Proofs of all results are collected in Section 4.

2. Limiting distribution. In this section we investigate the large sample properties of the standardized statistic

$$\hat{Z}_m = \frac{n\hat{T}_m - m}{\sqrt{2m}}. \tag{2.1}$$

We wish to accomplish this using essentially generic estimators of θ and local alternatives to the null which allow for asymptotic power calculations. Thus, we begin with a precise statement of the probability model to be employed and the specification of some requisite regularity conditions.

We now assume that for each n we have i.i.d. random variables $X_{1n}, \ldots, X_{nn}$ having common density

$$f_n(x;\,\theta_0) = f(x;\,\theta_0)\big(1 + b(n)\,\delta(x)\big) \tag{2.2a}$$

with

$$b(n) = m^{1/4}/\sqrt{n}. \tag{2.2b}$$

Here $f(\cdot;\,\theta_0)$ corresponds to the null model and θ_0 is the true value of θ when H_0 in (1.11) holds. The function $b(n)\,\delta(\cdot) f(\cdot;\,\theta_0)$ represents the departure from H_0. If $\delta = 0$, then we are dealing with the null model while, for $\delta \neq 0$ with $b(n) \to 0$, (2.2) converges to the null at rate $m^{1/4}/\sqrt{n}$. Concerning the function δ in (2.2a) we require that

(A) δ is bounded with

(i) $\lim_{u \to 0^+} \delta\big(F^{-1}(u;\,\theta_0)\big) = \lim_{u \to 1^-} \delta\big(F^{-1}(u;\,\theta_0)\big) = 0$,

(ii) $\big(\delta'(F^{-1}(\cdot;\,\theta_0))/f(F^{-1}(\cdot;\,\theta_0))\big) \in L_2[0,\,1]$, and

(iii) $\int_{-\infty}^{\infty} \delta(x) f(x;\,\theta_0)\,dx = \int_0^1 \delta\big(F^{-1}(u;\,\theta_0)\big)\,du = 0$.

Strictly speaking, we need $f_n \geqslant 0$ for all n in order for (2.2) to always be a valid probability density. However, the boundedness of δ and $b(n) \to 0$ is enough to ensure this for large n, which suffices for our analysis.

For the estimator $\hat{\theta}$ of θ we assume that

(B) $\sqrt{n}(\hat{\theta}-\theta_0) = O_p(1)$ under model (2.2).

This condition is satisfied in a variety of settings. For example, if $f(\cdot;\theta)$ is a normal distribution with mean θ, condition (B) holds under local alternatives of the form (2.2).

Finally, we need several smoothness conditions on $F(\cdot;\theta)$ as a function of θ. Specifically, we assume that

(C) $\sup_x |F(x;\hat{\theta})-F(x;\theta_0)| = o_p(1)$,

(D) $g(u) = \left.\dfrac{\partial F(x;\theta)}{\partial\theta}\right|_{x=F^{-1}(u;\theta_0),\,\theta=\theta_0}$ satisfies

$$\lim_{u\to 0^+} g(u) = \lim_{u\to 1^-} g(u) = 0 \quad \text{and} \quad \int_0^1 g^4(u)\,du < \infty$$

and

(E) there exists an open neighborhood $\mathscr{B}$ containing θ_0 and a function H independent of θ (but possibly depending on $\mathscr{B}$ and θ_0) such that

$$\left| \frac{\partial F(x;\theta)}{\partial\theta} - \left.\frac{\partial F(x;\theta)}{\partial\theta}\right|_{\theta=\theta_0} \right| \leqslant |\theta-\theta_0|\, H(x)$$

for all x and all $\theta\in\mathscr{B}$. The function H must satisfy

$$\int_{-\infty}^{\infty} H^4(x) f(x;\theta_0)\,dx < \infty.$$

We now state our main result.

THEOREM 1. *Assume conditions* (A)–(E) *hold true. Then, under the local alternatives* (2.2), *if* $m, n\to\infty$ *in such a way that* $\sqrt{m}(m^3/n+m/\sqrt{n})\to 0$,

$$\hat{Z}_m \xrightarrow{d} Y \sim N\left(\frac{\|\delta(F^{-1}(\cdot;\theta_0))\|^2}{\sqrt{2}}, 1\right),$$

where

$$\|\delta(F^{-1}(\cdot;\theta_0))\|^2 = \int_0^1 \delta^2(F^{-1}(u;\theta_0))\,du.$$

This theorem has a number of implications. First by taking $\delta = 0$ in (2.2) we see that $\hat{Z}_m$ has a limiting standard normal distribution under the null hypothesis. A large sample test for H_0 with asymptotic level α can therefore be obtained by comparing $\hat{Z}_m$ to Z_α, the $100(1-\alpha)$th percentage point of the standard normal distribution. In practice, we have found that $\hat{T}_m$ behaves very much like a chi-square random variable with m degrees of freedom for small to moderate sample sizes, and the normal approximation in Theorem 1 may not be satisfactory for such cases. A better approximation for the $100(1-\alpha)$th percentile of $\hat{Z}_m$ is often provided by $(\chi^2_{m;\alpha}-m)/\sqrt{2m}$, where $\chi^2_{m;\alpha}$ is the $100(1-\alpha)$th percentile for a chi-square distribution with m degrees of freedom (cf. [5]). Another approach is to use Monte Carlo methods to obtain approximate percentiles which is what we have done for the examples in Section 3.

More generally, when $\delta \neq 0$, Theorem 1 states that a test for H_0 based on $\hat{Z}_m$ can detect alternatives converging to the null at rate $m^{1/4}/\sqrt{n}$. The asymptotic power against such alternatives is

$$\lim_{n\to\infty} P(\hat{Z}_m > Z_\alpha) = 1 - \Phi\left(Z_\alpha - \frac{\|\delta(F^{-1}(\cdot;\theta_0))\|^2}{\sqrt{2}}\right),$$

where Φ is the standard normal d.f. Note that the power is uniform over departures from the null of any given size or norm. One can use this property as in [8] to argue that $\hat{Z}_m$ will be more effective against higher frequency departures from H_0 than Cramér–von Mises type tests even though it cannot detect local alternatives converging at the parametric $n^{-1/2}$ rate.

Under the assumption that the comparison density for the data has a square integrable second derivative, a mean squared error "optimal" choice for m is

$$m(n) = n^{1/4}\,[|d'(0)| + |d'(1)|]^{1/2}/\pi.$$

This is of no great help for the present situation since the true comparison density is unknown and is uniform under the null model. The formula could be used in practice by selecting a target choice for d corresponding to an alternative model that one is particularly interested in detecting. Notice that this choice for m satisfies the conditions of Theorem 1.

An alternative strategy for choosing m has been suggested by Azzalini et al. [2]. In our setting, their idea entails computing the P-values for $\hat{Z}_m$ for consecutive values of m until they begin to stabilize or become effectively constant. One then chooses a value of m to be somewhere in the region of stability. We do not know the operating characteristics of this approach but have found it to be useful from a diagnostic standpoint.

Yet another method for selecting m is to use a data-driven order selector for the estimator

$$\hat{d}_m(u) = 1 + \sqrt{2/n}\sum_{j=1}^{m} a_j(\hat{\theta})\cos(j\pi u)$$

of $d(u)$. This approach has been found to be quite effective in some empirical studies (e.g., [8] and [11]). However, work by Kim [11] for the simple goodness-of-fit hypothesis makes it highly unlikely that Theorem 1 will apply when m is chosen in this fashion. Further discussion of data driven methods for choosing m can be found in [10].

An interesting feature of $\hat{Z}_m$ is that its limiting distribution is the same as for Z_m in (1.9) under the null hypothesis. Thus, in this sense it is unaffected by the estimation of θ_0. This is not true for either the Cramér–von Mises or standard Neyman smooth statistics in general (see [7], [16], [3], [20], [13], [14]). The fact that m grows with n and the division by $\sqrt{m}$ in (2.1) is what removes (asymptotically) the dependence of our statistic on estimation of θ_0.

To conclude this section we note that Theorem 1 can be extended in several directions. Perhaps of most interest is the case of more than one estimated parameter. The basic proof in Section 4 extends in a straightforward manner to include this situation under the natural extensions of conditions (B)–(E) to the multiparameter setting.

3. Examples. Two data sets are now analyzed by using the methodology proposed in the previous section. One goal of these analyses is to illustrate the utility of the comparison density estimator,

$$d_m(u) = 1+n^{-1/2}\sum_{j=1}^{m} a_j(\hat{\theta})\sqrt{2}\cos(j\pi u), \tag{3.1}$$

as a diagnostic tool for assessing the nature of departures from the null model when H_g^* is rejected.

For both examples θ was a scale parameter that was estimated using maximum likelihood (or asymptotically equivalent) estimators. The scale equivariance of the estimators has the consequence that for these particular cases the distributions of the $a_j(\hat{\theta})$'s and the $\hat{Z}_m$'s depend only on F and not on θ for each value of n. This made it quite simple to determine very precise critical values for the tests using Monte Carlo methods and, accordingly, that was the approach we employed to compute the P-values mentioned in the discussions below.

EXAMPLE 1. We first consider the Angus [1] data set of $n = 20$ operational lifetimes. The null hypothesis of exponentiality for this data (i.e., $F(x;\theta) = 1-e^{-x/\theta}$ for $x > 0$ and $\theta > 0$) was tested by Rayner and Best [13] by using a smooth type test. After some experimentation we chose $m = 4$ for our analysis and obtained $\hat{Z}_4 = 2.858$ for which the approximate P-value is 0.01.

Fig. 1 gives a plot of $d_4(u)-1$. (Note that this function should be identically zero under the null hypothesis.) The graph reveals that, when compared to an exponential, the true distribution for the data may have lower probability in both tails and higher probability in the interior region right of the median. This suggests that the data would be better fit by either a Weibull distribution with shape parameter $c > 1$ and very large scale parameter or a two-parameter exponential distribution. The comparison density estimators that are obtained from these two choices for F with $m = 4$ are also shown in Fig. 1. In both cases the fit appears to be much better with corresponding test statistic values of -0.171 and 0.385 for the Weibull and two-parameter exponential, respectively.

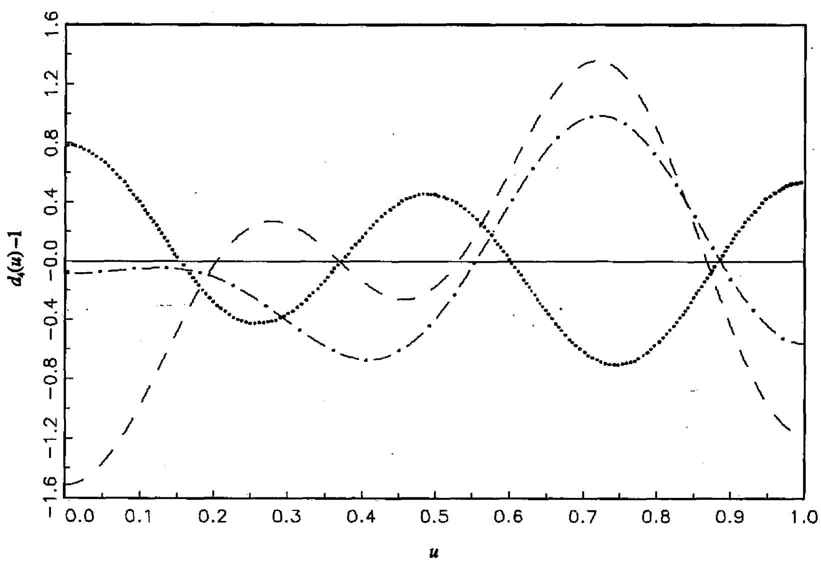

Fig. 1. Comparison density estimators for the Angus data

The dashed line is for the exponential model, the dotted line is for the Weibull model, and the dotted and dashed line is for the two-parameter exponential model

EXAMPLE 2. This example illustrates the use of our procedure on a large data set. The data are the sizes of oil and gas fields (in log base 2 units) in the Frio strandplain exploration play located in the central coastal plain of Texas (see [18]).

It is generally believed that, at least in the early stages of exploration of a play, the lognormal distribution provides a reasonable model to fit the observed size distribution of oil and gas fields. Thus, we begin by analyzing the 318 discoveries made through 1959. The P-values corresponding to the $\hat{Z}_m$ for this data set are fairly erratic for small values of m, oscillating from 0.03 to 0.35. However, for $m \geqslant 54$ they stabilize at around 0.26 suggesting, as expected, that the lognormal distribution provides an adequate model for the data.

When the size distribution of oil and gas fields is viewed at a later stage of exploration, it is the contention of Schuenemeyer and Drew [17] that this distribution is *not* lognormal. The discovery data available through 1985 ($n = 695$) was tested against this assumption. The P-values associated with the $\hat{Z}_m$ are now extremely stable. In fact, for all $m > 3$ the P-values are less than 0.0001. Thus, we chose $m = 10$ and the resulting test statistic $\hat{Z}_{10} = 19.122$ is clearly significant. The graph of $d_{10}(u)-1$ is presented in Fig. 2. This figure shows that the number of small fields is overestimated by the lognormal model since the graph is near -1.0 for u small. We also see from this that the lognormal model slightly underestimates the probability in the right tail.

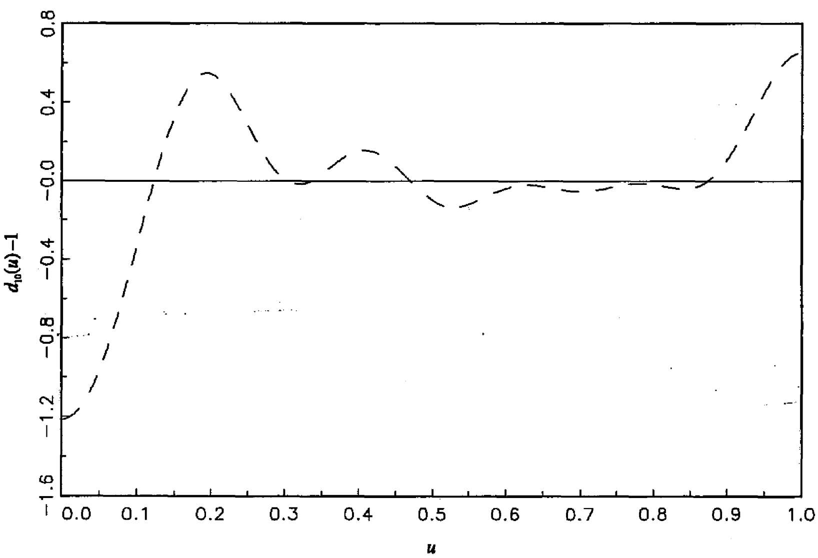

Fig. 2. Comparison density estimator for the Frio strandplain data

4. Proof. In this section we prove Theorem 1. We first give an outline of the steps that are involved and then focus on the specific details.

Since we are working with the local alternatives (2.2), this means the sample Fourier coefficients are

$$a_j(\theta) = n^{-1/2} \sum_{i=1}^{n} \sqrt{2}\cos\left(j\pi F(X_{in};\, \theta)\right).$$

Set $Z_m = \left(\sum_{j=1}^{m} a_j^2(\theta_0) - m\right)/\sqrt{2m}$. Then we can write $\hat{Z}_m$ in (2.1) as

$$\hat{Z}_m = Z_m + \frac{1}{\sqrt{2m}} \sum_{j=1}^{m} \left[a_j^2(\hat{\theta}) - a_j^2(\theta_0)\right]. \tag{4.1}$$

The limiting distribution of Z_m was shown in [8] to be normal with mean $\|\delta(F^{-1}(\cdot;\, \theta_0))\|^2/\sqrt{2}$ and variance 1 under the conditions of our theorem. Thus, Theorem 1 is proved when we show that $\hat{Z}_m - Z_m \xrightarrow{P} 0$.

Now

$$\sqrt{2}(\hat{Z}_m - Z_m) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \Delta_j^2 + \frac{2}{\sqrt{m}} \sum_{j=1}^{m} \Delta_j a_j(\theta_0) \tag{4.2}$$

with

$$\Delta_j = a_j(\hat{\theta}) - a_j(\theta_0). \tag{4.3}$$

To show that $\hat{Z}_m - Z_m$ is asymptotically negligible we therefore concentrate on the Δ_j and show, in Lemma 3, that they behave much like a sequence of square summable constants under the conditions of our theorem. This will then be seen to imply the desired result.

To simplify the expressions that follow we introduce some additional notation. Let

$$F_0(\cdot) = F(\cdot;\theta_0), \quad \hat{F}(\cdot) = F(\cdot;\hat{\theta}), \quad F_0^{-1}(\cdot) = F^{-1}(\cdot;\theta_0),$$
$$f_0(\cdot) = f(\cdot;\theta_0), \quad D_0(\cdot) = \left.\frac{\partial F(\cdot;\theta)}{\partial\theta}\right|_{\theta=\theta_0}.$$

We will also require the following two lemmas.

LEMMA 1. *Let g_1 be such that*

$$\lim_{u\to 0^+} g_1(F_0^{-1}(u)) = \lim_{u\to 1^-} g_1(F_0^{-1}(u)) = 0$$

and $(g_1'(F_0^{-1}(\cdot))/f_0(F_0^{-1}(\cdot)))\in L_2[0,1]$. Set $g_2 = g_1\cdot\delta$. Then

$$\frac{\sqrt{2}}{n}\sum_{i=1}^{n} g_1(X_{in})\sin(j\pi F_0(X_{in})) = \frac{1}{j\pi}(C_{1j}+b(n)C_{2j}) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in j, with

$$C_{ij} = \sqrt{2}\int_0^1 \frac{g_i'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}\cos(j\pi u)\,du, \qquad i=1,2,$$

and $\sum_{j=1}^{\infty} C_{ij}^2 < \infty$, $i=1,2$.

Proof. Set $b_j = \sqrt{2}n^{-1}\sum_{i=1}^{n} g_1(X_{in})\sin(j\pi F_0(X_{in}))$. Then

$$\begin{aligned}
Eb_j &= \sqrt{2}\int_{-\infty}^{\infty} g_1(x)\sin(j\pi F_0(x))f_0(x)(1+b(n)\delta(x))\,dx \\
&= \sqrt{2}\int_0^1 g_1(F_0^{-1}(u))\sin(j\pi u)\,du + b(n)\int_0^1 g_2(F_0^{-1}(u))\sin(j\pi u)\,du \\
&= \frac{1}{j\pi}(C_{1j}+C_{2j}b(n)),
\end{aligned}$$

after integration by parts. The proof is completed by observing that, because of condition (A),

$$\operatorname{Var} b_j \leqslant \frac{2}{n}\int g_1^2(x)f_0(x)(1+b(n)\delta(x))\,dx \leqslant \frac{2}{n}\int g_1^2(F_0^{-1}(u))\,du\,(1+O(b(n))). \ \blacksquare$$

LEMMA 2. *Let g be a function on the line with $(g'(F_0^{-1}(\cdot))/f_0(F_0^{-1}(\cdot)))\in L_2[0,1]$. Then*

$$n^{-1}\sum_{i=1}^{n} g(X_{in})\cos(j\pi F_0(X_{in})) = \frac{1}{j\pi}(s_j+O(b(n))) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in j for

$$s_j = -\int_0^1 \frac{g'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}\sin(j\pi u)\,du \quad \text{and} \quad \sum_{j=1}^{\infty} s_j^2 < \infty.$$

Proof. The proof parallels that for Lemma 1 and is therefore omitted. ∎

The next lemma gives the essential approximation we need for the Δ_j in (4.3).

LEMMA 3. *Let*

$$C_j = \sqrt{2}\int_0^1 \frac{D_0(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}\sin(j\pi u)\,du. \tag{4.4}$$

Then, if $m^3/n \to 0$,

$$\Delta_j = \sqrt{n}(\hat{\theta}-\theta_0)C_j + O_p\left(\frac{m^3}{n} + \frac{m}{\sqrt{n}}\right)$$

uniformly in $j \leqslant m$.

Proof. Since

$$\cos x - \cos y = 2\sin\left(\frac{x-y}{2}\right)\cos\left(\frac{x+y}{2}\right),$$

we have

$$\Delta_j = \frac{2\sqrt{2}}{\sqrt{n}}\sum_{i=1}^{n}\sin\left[j\pi\frac{(\hat{F}(X_{in})-F_0(X_{in}))}{2}\right]\cos\left[j\pi\left(\frac{\hat{F}(X_{in})+F_0(X_{in})}{2}\right)\right].$$

Now, for $|x| < 1$, $|\sin x - x| \leqslant |x|^3/6$. Thus, for n sufficiently large

$$\Delta_j = \frac{2\sqrt{2}}{\sqrt{n}}\sum_{i=1}^{n} j\pi\left(\frac{\hat{F}(X_{in})-F_0(X_{in})}{2}\right)\cos\left[j\pi\frac{(\hat{F}(X_{in})+F_0(X_{in}))}{2}\right]+r_n$$

with

$$|r_n| \leqslant (m\pi)^3\frac{\sqrt{2}}{3\sqrt{n}}\sum_{i=1}^{n}|\hat{F}(X_{in})-F_0(X_{in})|^3 = O_p\left(\frac{m^3}{n}\right)$$

due to assumptions (A)–(E).

Using similar arguments, the trigonometric identity $\cos(x+y) = \cos x\sin y + \sin x\cos y$ and the bound $|\cos x - 1| \leqslant x^2/2$ for $|x| < 1$, we obtain

$$\frac{2\sqrt{2}}{\sqrt{n}}\sum_{i=1}^{n} j\pi\left(\frac{\hat{F}(X_{in})-F_0(X_{in})}{2}\right)\cos\left[j\pi\left(\frac{\hat{F}(X_{in})-F_0(X_{in})}{2}\right)+j\pi F_0(X_{in})\right]$$
$$= \Delta_{1j}+\Delta_{2j}+O_p\left(\frac{m^4}{n^{3/2}}+\frac{m^3}{n}\right)$$

with

$$\Delta_{1j} = \frac{\sqrt{2}}{\sqrt{n}} j\pi \sum_{i=1}^{n} [\hat{F}(X_{in}) - F_0(X_{in})] \sin[j\pi F_0(X_{in})]$$

and

$$\Delta_{2j} = \frac{1}{\sqrt{2n}} (j\pi)^2 \sum_{i=1}^{n} [\hat{F}(X_{in}) - F_0(X_{in})]^2 \cos[j\pi F_0(X_{in})].$$

We will focus first on approximating Δ_{2j}. By a Taylor expansion,

$$\hat{F}(X_{in}) - F_0(X_{in}) = (\hat{\theta} - \theta_0) D_0(X_{in}) + \hat{r}_i.$$

Thus,

$$\begin{aligned}\Delta_{2j} = {} & \frac{(j\pi)^2}{\sqrt{2}} \sqrt{n}(\hat{\theta} - \theta_0)^2 \frac{1}{n} \sum_{i=1}^{n} D_0(X_{in})^2 \cos[j\pi F_0(X_{in})] \\ & + \sqrt{2}(j\pi)^2 \sqrt{n}(\hat{\theta} - \theta_0) \frac{1}{n} \sum_{i=1}^{n} D_0(X_{in}) \hat{r}_i \cos[j\pi F_0(X_{in})] \\ & + \frac{(j\pi)^2}{\sqrt{2}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{r}_i^2 \cos[j\pi F_0(X_{in})].\end{aligned}$$

Assumption (E) ensures that for n sufficiently large the last term in Δ_{2j} is bounded by a constant multiple of

$$\frac{m^2}{\sqrt{n}} (\hat{\theta} - \theta_0)^4 \sum_{i=1}^{n} |H(X_{in})|^2 = O_p\left(\frac{m^2}{n^{3/2}}\right),$$

using condition (A). The Cauchy–Schwarz inequality, assumption (D) and a similar argument reveal that the cross-product term is bounded in magnitude by

$$\sqrt{2}(m\pi)^2 \sqrt{n} |\hat{\theta} - \theta_0| \left[\left(\frac{1}{n} \sum_{i=1}^{n} D_0^2(X_{in})\right)\left(\frac{1}{n} \sum_{i=1}^{n} \hat{r}_i^2\right)\right]^{1/2} = O_p\left(\frac{m^2}{n}\right).$$

Finally, apply Lemma 2 to $n^{-1} \sum_{i=1}^{n} D_0(X_{in})^2 \cos(j\pi F_0(X_{in}))$ with $g(\cdot) = D_0^2(\cdot)$ and use assumption (D) to see that the first term in Δ_{2j} is of the form

$$\frac{j\pi}{\sqrt{2}} (s_j + O(b(n))) \sqrt{n}(\hat{\theta} - \theta_0)^2 + O_p\left(\frac{m^2}{n}\right) \quad \text{with} \ \sum_{j=1}^{\infty} s_j^2 < \infty.$$

Since $|s_j| \leqslant (\sum_{k=1}^{\infty} s_k^2)^{1/2}$ for all j, we have

$$\Delta_{2j} = O_p\left(\frac{m}{\sqrt{n}} + \frac{m^2}{n}\right) + O_p\left(\frac{m^2}{n}\right) + O_p\left(\frac{m^2}{n^{3/2}}\right) = O_p\left(\frac{m}{\sqrt{n}}\right).$$

The approximation for Δ_{1j} is similar to that for Δ_{2j}. One uses Lemma 1 with $g_1(\cdot) = D_0(\cdot)$ and assumptions (D) and (E) to obtain

$$\Delta_{1j} = \sqrt{n}(\hat{\theta}-\theta_0)(C_j+O(b(n)))+O_p\left(\frac{m}{\sqrt{n}}\right) = \sqrt{n}(\hat{\theta}-\theta_0)C_j+O_p\left(\frac{m}{\sqrt{n}}\right).$$

The proof is finished by combining all our approximations. ∎

We can now complete the proof of Theorem 1. Referring again to (4.2) and the definition of C_j in (4.4), we see that

$$\text{(4.5)} \quad \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \Delta_j^2 = \frac{n(\hat{\theta}-\theta_0)^2}{\sqrt{m}}\sum_{j=1}^{m} C_j^2+O_p\left(\frac{m^3}{n}+\frac{m}{\sqrt{n}}\right)+O_p\left(\sqrt{m}\left(\frac{m^3}{n}+\frac{m}{\sqrt{n}}\right)^2\right)$$

and

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m} a_j(\theta_0)\Delta_j = \left(\frac{1}{\sqrt{m}}\sum_{j=1}^{m} a_j(\theta_0)C_j\right)\sqrt{m}(\hat{\theta}-\theta_0)+\frac{1}{\sqrt{m}}\sum_{j=1}^{m} a_j(\theta_0)r_j$$

with $r_j = \Delta_j-\sqrt{n}(\hat{\theta}-\theta_0)C_j$, $j = 1, \ldots, m$. To handle this last expression we need results from [8] that we state here in the form of a lemma.

LEMMA 4 (Eubank and LaRiccia [8]). *Under condition* (A) *and the local alternatives* (2.2) *we have*

$$\text{(4.6)} \qquad \mathrm{E}a_j(\theta_0) = \frac{m^{1/4}}{j\pi}s_j, \quad j = 1, 2, \ldots,$$

with

$$\text{(4.7)} \qquad s_j = -\sqrt{2}\int_0^1 \frac{\delta'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}\sin(j\pi u)\,du, \quad j = 1, 2, \ldots,$$

and, uniformly in $j, k = 1, \ldots,$

$$\mathrm{Cov}(a_j(\theta_0), a_k(\theta_0)) = \delta_{jk}+O(b(n)).$$

If $n, m \to \infty$ *in such a way that* $m^5/n^2 \to 0$, *then*

$$\text{(4.8)} \qquad \frac{1}{\sqrt{2m}}\left(\sum_{j=1}^{m} a_j(\theta_0)^2-m\right) \xrightarrow{d} Y,$$

where Y *is a normal random variable with mean* $\|\delta(F_0^{-1}(\cdot))\|^2/\sqrt{2}$ *and variance* 1.

As a result of (4.6)–(4.7) we have

$$\mathrm{E}\left(\sum_{j=1}^{m} a_j(\theta_0)C_j\right) = m^{1/4}\sum_{j=1}^{m} C_j s_j/(j\pi) = O(\sqrt{n}b(n))$$

and

$$\mathrm{Var}\left(\sum_{j=1}^{m} a_j(\theta_0)C_j\right) = \sum_{j=1}^{m} C_j^2+O(mb(n)).$$

Also from the Cauchy–Schwarz inequality we obtain

$$\Big|\sum_{j=1}^{m} a_j(\theta_0) r_j\Big| \leqslant \Big[\sum_{j=1}^{m} r_j^2 \sum_{j=1}^{m} a_j(\theta_0)^2\Big]^{1/2} = O_p\left(m\left(\frac{m^3}{n}+\frac{m}{\sqrt{n}}\right)\right)$$

because of Lemma 3 and (4.8). Thus,

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m} a_j(\theta_0)\Delta_j = O_p\big(m^{-1/4}+\sqrt{b(n)}\big)+O_p\left(\sqrt{m}\left(\frac{m^3}{n}+\frac{m}{\sqrt{n}}\right)\right). \tag{4.9}$$

Combining (4.5) and (4.9) gives

$$\hat{Z}_m - Z_m = O_p\left(\sqrt{m}\left(\frac{m^3}{n}+\frac{m}{\sqrt{n}}\right)+m^{-1/4}+\sqrt{b(n)}\right).$$

The proof is completed by using (4.8) of Lemma 4 and Slutsky's Theorem.

Acknowledgements. Eubank's research was supported in part under National Science Foundation grant DMS-93.00918.

REFERENCES

[1] J. E. Angus, *Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation*, J. Statist. Plann. Inference 6 (1982), pp. 241–251.

[2] A. Azzalini, A. W. Bowman and W. Härdle, *On the use of nonparametric regression for model checking*, Biometrika 76 (1989), pp. 1–11.

[3] D. E. Barton, *Neyman's Ψ_k^2 test of goodness of fit when the null hypothesis is composite*, Skand. Aktuarietidskr. 39 (1956), pp. 216–246.

[4] P. J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist. 1 (1973), pp. 1071–1095.

[5] M. J. Buckley and G. K. Eagleson, *An approximation to the distribution of quadratic forms in normal random variables*, Austral. J. Statist. 30A (1988), pp. 150–159.

[6] J. Durbin and M. Knott, *Components of Cramér–von Mises statistics, I*, J. Roy. Statist. Soc. Ser. B 37 (1972), pp. 216–237.

[7] – and C. C. Taylor, *Components of Cramér–von Mises statistics, II*, ibidem 34 (1975), pp. 290–307.

[8] R. Eubank and V. N. LaRiccia, *Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit*, Ann. Statist. 20 (1992), pp. 2071–2086.

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

[10] W. C. M. Kallenberg and T. Ledwina, *Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests*, Ann. Statist. (1995), to appear.

[11] J. T. Kim, *Testing Goodness-of-Fit via Order Selection Criteria*, unpublished Ph. D. Thesis, Department of Statistics, Texas A&M University, 1992.

[12] J. Neyman, *Smooth test for goodness-of-fit*, Skand. Aktuarietidskr. 20 (1937), pp. 149–199.

[13] J. Rayner and D. Best, *Neyman-type smooth tests for location-scale families*, Biometrika 73 (1986), pp. 437–446.

[14] – *Smooth Tests of Goodness of Fit*, Oxford University Press, New York 1989.

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

[16] D. Schoenfeld, *Tests based on linear combinations of the Cramér–von Mises statistic when the parameters are estimated*, ibidem 8 (1980), pp. 1017–1022.

[17] J. H. Schuenemeyer and L. J. Drew, *A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation*, Math. Geol. 15 (1983), pp. 145–161.

[18] – *A forecast of undiscovered oil and gas in the trio strandplain trend: the unfolding of a very large exploration play*, AAPG Bulletin (1991).

[19] G. Shorack and J. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, New York 1986.

[20] D. R. Thomas and D. A. Pierce, *Neyman's smooth goodness-of-fit test when the hypothesis is composite*, J. Amer. Statist. Assoc. 74 (1979), pp. 441–445.

Dept. of Statist.
Texas A&M University
College Station, Texas 77843-3143, U.S.A.

Dept. of Math. Sci.
University of Delaware
Newark, Delaware 19711, U.S.A.

Received on 11.1.1994