TESTING GOODNESS-OF-FIT BASED ON
A ROUGHNESS MEASURE

BY

LI-SHAN HUANG

FSU STATISTICS REPORT M904
Testing Goodness-of-Fit Based on a Roughness Measure

Li-Shan Huang

Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

November 20, 1995

FSU Technical Report No. M-904
Testing Goodness-of-fit Based on a Roughness Measure

Li-Shan Huang
Department of Statistics
Florida State University
Tallahassee, F.L. 32306-3033

November 20, 1995

Abstract

One popular approach for testing one sample goodness-of-fit is based on a discrepancy measure between the empirical distribution function and the hypothesized distribution function. Examples include the Kolmogorov-Smirnov test and the Cramér-von Mises test. A new test statistic for goodness-of-fit is proposed, which measures the difference, in terms of roughness, between the underlying density function and the hypothesized density function. One advantage of using a roughness measure is high power in detecting contiguous alternative densities with local features. The asymptotic distribution of the test statistic is obtained under both null and alternative hypotheses. A testing procedure is proposed and later modified for practical implementation. Comparison of the proposed test with the Kolmogorov-Smirnov and the Cramér-von Mises tests includes empirical results from simulation studies and different rates against contaminated alternatives.

Key words and phrases: Goodness-of-fit; kernel density estimation; contaminated distributions.

This paper was part of the author's doctoral dissertation under the supervision of Dr. Jianqing Fan at University of North Carolina. The author is very grateful to his generous guidance and encouragements. Li-Shan Huang also thanks Dr. E. Carlstein, Dr. P.K. Sen, and Dr. J.S. Simonoff for useful comments on an earlier version of this paper.
1 Introduction

Suppose that $X_1, \ldots, X_n$ are a random sample from an absolutely continuous distribution function (d.f.) $F(\cdot)$. Consider the problem of testing the hypothesis $F = F_0$, for some specified absolutely continuous d.f. $F_0$. After a probability integral transformation $z \rightarrow F_0(z)$, the problem is to test whether $F_0(X_1), \ldots, F_0(X_n)$ are uniformly distributed on $(0,1)$. Thus, without loss of generality, the problem considered is testing for a uniform d.f. on $(0,1)$, i.e. $H_0 : F = \text{uniform}(0,1)$.

Conventional tools for testing the one sample hypothesis $H_0$ are based on a discrepancy measure between the empirical distribution function (EDF), $\hat{F}_n$, and the uniform distribution, $F_0$. For example, the Kolmogorov-Smirnov (K-S) test is based on the sup-norm $||\hat{F}_n - F_0||_\infty$. Some other EDF-based tests include the Cramér-von Mises (CVM) test and the Anderson-Darling test, which are based on weighted $L_2$ norms. Hájek and Šidák (1967), Durbin (1973), Kendall and Stuart (1979), and D'Agostino and Stephens (1986) give useful surveys of the test statistics based on empirical distribution function. Despite the popularity of the EDF-based tests, it is known that they have poor power against non-location/scale alternatives. Bickel and Rosenblatt (1973) propose a test statistic that measures the $L_2$-distance between a kernel density estimate for the underlying density and its expected value under the null hypothesis. Recently, using nonparametric function estimation techniques for goodness-of-fit has been the subject of several papers. For example, the test statistics proposed by Eubank and LaRicca (1992) and Fan (1995) involve Fourier series estimators. Bowman (1992) considers two test statistics based on kernel density estimators.

We consider an approach to discriminate between two smooth densities by the distance

$$d_1(f, f_0) = \int (f'(x) - f'_0(x))^2 \, dx,$$

where $f(\cdot)$ is the underlying density function and $f_0(\cdot)$ is the hypothesized density function. An alternative interpretation of $d_1$ is that it measures the difference, in terms of roughness, between $f(\cdot)$ and $f_0(\cdot)$. One advantage of test statistics based on (1.1) is high power in detecting contiguous alternatives with local features, such as sharp peaks and high frequency components. Indeed, if an alternative density $f$ has a local deviation at $x_0$, as in the following expression,

$$f(x) = f_0(x) + g((x - x_0)/\sigma),$$

where $g$ is continuously differentiable up to second order and $\int g = 0$, then

$$d_1(f, f_0) = \sigma^{-1} \int [g'(x)]^2 dx.$$
It is clear that $d_1(f, f_0)$ will be large when $\sigma$ is small, i.e. when the local feature is strong. In contrast, the Bickel-Rosenblatt test is based on the distance

$$d_0(f, f_0) = \int (f(z) - f_0(z))^2 dz = \sigma \int [g(z)]^2 dx,$$

which tends to 0 as $\sigma \to 0$. Thus the Bickel-Rosenblatt test and the EDF-based tests are not sensitive to alternatives with subtle local features. Certainly, there is penalty for using the distance $d_1$, in terms of the accuracy of assessing this distance based on a random sample. For this reason we do not consider higher derivatives, such as the second or the third derivatives.

By Parseval's identity, (1.1) can be expressed as

$$d_1(f, f_0) = \frac{1}{2\pi} \int \omega^2 [\tilde{f}(\omega) - \tilde{f}_0(\omega)]^2 d\omega$$

where $\tilde{f}$ and $\tilde{f}_0$ are the Fourier transforms (characteristic functions) of $f$ and $f_0$ respectively. It is clear that more weight is given to high frequency components than to low frequency components. Thus, test statistics based on (1.1) may be better able to detect high frequency alternatives. In contrast, the Bickel-Rosenblatt test uses a uniform weighting scheme $\omega^0 = 1$, and the Cramér von Mises test weigh the high frequency components by $\omega^{-2}$.

Since $f_0$ is the uniform density on $(0, 1)$, $d_1(f, f_0)$ is reduced to $\theta_1$, where

$$\theta_\nu = \int [f^{(\nu)}(z)]^2 dz, \quad \nu = 0, 1, 2, \ldots.$$  \hspace{1cm} (1.2)

The null hypothesis to be tested is

$$H_0 : \theta_1 = 0.$$

Note that for those densities satisfying Condition (D1) (see Conditions (D) in Section 2), $\theta_1 = 0$ if and only if the underlying density is the uniform density on $(0, 1)$. Thus our test is theoretically designed for the specified class of density functions, while the EDF-based tests are consistent against all alternatives. Hall and Marron (1987) give an estimator of $\theta_1$ (see (2.2)), which is based on kernel density estimates. However, this estimator requires that the domain of the underlying density is the whole real line. It is possible, as described in Section 2, to modify the estimator of Hall and Marron (1987) so that the resulting estimator $\hat{\theta}_1$ (see (2.6)) has a support on $(0, 1)$ and is unbiased under the null hypothesis.

Note that there is cost to be paid for using a nonparametric test statistics $\hat{\theta}_1$. First, it requires a stronger assumption (Condition (D1)) of the underlying density than the EDF-based tests. Second our test can only detect smooth alternative distributions at a distance of $O(n^{-(2-5\delta)/4})$.
if the bandwidth is chosen of order $n^{-\delta}$ (see Corollary 3.1). Comparatively, the standard EDF-based tests can detect alternative distributions differing from $F_0$ by $O(n^{-1/2})$. There are other types of alternatives, e.g. contaminated distributions, for which our test is more powerful than the EDF-based tests (see Theorem 3.1). An interesting problem is how to combine the proposed testing procedure with the conventional tests, such as the Cramér-von Mises test or the Anderson-Darling test, to yield a more omnibus test.

The idea of the proposed test may be extended for testing a composite null hypothesis such as $H_0 : f = f_0 (\cdot, \sigma^2)$, where $\sigma^2$ is a nuisance parameter. An estimate $\hat{\sigma}^2$, e.g. the maximum likelihood estimator, may be used so that the null hypothesis is completely specified, as the case discussed in this paper. The extension to the case of testing a composite null hypothesis may be a problem for further research.

The motivation of an estimate $\hat{\theta}_1$ of $\theta_1$ is given in Section 2. The distribution of $\hat{\theta}_1$ is derived under both null and alternative hypotheses and a testing procedure is proposed. Asymptotic power of the proposed test against contaminated distributions are discussed in Section 3. The problem of selecting an appropriate bandwidth is also considered. In Section 4 we discuss some problems that were encountered in simulation studies. Some modifications are suggested and empirical powers are obtained using simulated data. The proofs are collected in Appendix.

2 The test statistic

The kernel density estimate $\hat{f}(x)$ for estimating $f(x)$ using data $X_1, \ldots, X_n$ is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{X_i - x}{h} \right), \quad (2.1)$$

where $K(\cdot)$ is a kernel function such that $\int K(u) du = 1$, and $h$ is the smoothing parameter. See Silverman (1986) for the literature on kernel density estimation. Hall and Marron (1987, 1991) give a kernel-based estimate of $\int_{-\infty}^{\infty} f^2(x) dx$:

$$n^{-1}(n - 1)^{-1} h^{-4} \sum_{i \neq j} \int_{-\infty}^{\infty} K' \left( \frac{X_i - x}{h} \right) K' \left( \frac{X_j - x}{h} \right) dx, \quad (2.2)$$

for densities defined on $(-\infty, \infty)$ with some conditions to eliminate the boundary bias. This estimator is motivated by differentiating $\hat{f}(x)$ in (2.1) and substituting $\hat{f}'(x)$ for $f'(x)$:

$$\int_{-\infty}^{\infty} \hat{f}'(x) dx = n^{-2} h^{-4} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} K' \left( \frac{X_i - x}{h} \right) K' \left( \frac{X_j - x}{h} \right) dx. \quad (2.3)$$
Hall and Marron (1987) argue that the diagonal terms \((i = j)\) in (2.3) do not make use of the data and hence are removed to achieve unbiasedness (see (2.2)).

However, the underlying density \(f(\cdot)\) considered here is defined on \((0,1)\). Hence, it may be useful to consider an estimator

\[
\tilde{\theta}_1 = n^{-1}(n-1)^{-1}h^{-2} \sum_{i \neq j} \int_0^1 K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_j - x}{h}\right) dx.
\]  

The following conditions on \(K\) and \(f\) are assumed throughout.

**Conditions (D):**

(D1) The underlying density \(f(\cdot)\) is defined on \((0,1)\) and is continuously differentiable to the second order with bounded derivatives.

(D2) The kernel function \(K(\cdot)\) is a symmetric probability density function. Further, \(K(\cdot)\) and its first derivative \(K'(\cdot)\) are bounded, having a support \([-1,1]\) with \(K(-1) = 0\) and \(K(1) = 0\).

One difficulty about \(\tilde{\theta}_1\) which arises from the boundaries \((0,h)\) and \((1-h,1)\) is that \(\tilde{\theta}_1\) is actually a biased estimator of \(\theta_1\) under the null hypothesis:

\[
E\left\{ \int_h^{1-h} K\left(\frac{X_1 - x}{h}\right) K\left(\frac{X_2 - x}{h}\right) dx | f = f_0 \right\} = 0,
\]

\[
E\left\{ \int_0^h K\left(\frac{X_1 - x}{h}\right) K\left(\frac{X_2 - x}{h}\right) dx | f = f_0 \right\} = h^3 \int_{-1}^0 K^2(x) dx,
\]

\[
E\left\{ \int_h^{1-h} K\left(\frac{X_1 - x}{h}\right) K\left(\frac{X_2 - x}{h}\right) dx | f = f_0 \right\} = h^3 \int_0^1 K^2(x) dx.
\]

In other words, estimation at the boundary regions creates an unacceptable amount of bias:

\[
E\{\tilde{\theta}_1 | f = f_0\} = h^{-1} \int_{-1}^1 K^2(x) dx.
\]

An attempt to correct this bias is to use higher order kernels for smoothing on the boundaries. In particular, we employ the boundary kernel functions induced by a local quadratic fit:

\[
K_c(t) = (0,1,0)S_c^{-1}(1,t,t^2)^T K(t), \quad 0 < c < 1,
\]

where \(S_c = (\mu_{i+j-2,c})_{3 \times 3}\) with

\[
\mu_{k,c} = \begin{cases} \int_{-c}^0 u^k K(u) du & \text{if } x = ch, \\ \int_c^1 u^k K(u) du & \text{if } x = 1 - ch. \end{cases}
\]

Note that \(K_c(\cdot), 0 < c < 1\) are kernel functions with order \((1,3)\), i.e. it satisfies

\[
\mu_{0,c} = \mu_{2,c} = 0, \quad \mu_{1,c} = 1, \quad \text{and } \mu_{3,c} \neq 0.
\]
See Ruppert and Wand (1994) for more details on construction of higher order kernels through local polynomial fitting.

Now an estimator for \( f'(\cdot) \) near the boundary point 0 is
\[
\hat{f}'(ch) = -n^{-1}h^{-2} \sum_{i=1}^{n} K_c \left( \frac{X_i - ch}{h} \right), \quad 0 < c < 1.
\]

Similarly, \( f'(\cdot) \) at the right boundary can be estimated by
\[
\hat{f}'(1-ch) = -n^{-1}h^{-2} \sum_{i=1}^{n} K_c \left( \frac{X_i - (1-ch)}{h} \right), \quad 0 < c < 1.
\]

As in (2.2), after deleting diagonal terms and adjusting the normalizing constant, the resulting estimator \( \hat{\theta}_1 \) is defined as follows:
\[
\hat{\theta}_1 = \frac{1}{n(n-1)h^4} \sum_{i \neq j} \int_{-h}^{1-h} K' \left( \frac{X_i - x}{h} \right) K' \left( \frac{X_j - x}{h} \right) dx
+ \frac{1}{n(n-1)h^3} \sum_{i \neq j} \int_{0}^{1} K_c \left( \frac{X_i - ch}{h} \right) K_c \left( \frac{X_j - ch}{h} \right) dc
+ \frac{1}{n(n-1)h^3} \sum_{i \neq j} \int_{0}^{1} K_c \left( \frac{X_i - (1-ch)}{h} \right) K_c \left( \frac{X_j - (1-ch)}{h} \right) dc.
\tag{2.6}
\]

Hall and Marron (1987) give the asymptotic bias and variance expressions of the estimator in (2.2), when \( f \) has a support \((-\infty, \infty)\). Their results are not directly applicable to our case since \( \hat{\theta}_1 \) is different from (2.2). For the purpose of hypothesis testing, the asymptotic distribution of \( \hat{\theta}_1 \) is derived in the following theorems, under both null and alternative hypotheses.

**Theorem 2.1** Assume that Conditions (D) hold and \( h = h(n) \to 0, \ nh \to \infty \) as \( n \to \infty \). If \( f(\cdot) \) is the uniform(0,1) null density, then \( \hat{\theta}_1 \) is asymptotically normally distributed with mean 0 and variance \( n^{-2}h^{-5}2\kappa_1 + o(n^{-2}h^{-5}) \) with \( \kappa_1 = \int [(K' * K')(u)]^2 du \), where \( * \) denotes convolution.

**Theorem 2.2** Assume that Conditions (D) hold and \( h = h(n) \to 0, \ nh \to \infty \) as \( n \to \infty \). If \( f(\cdot) \) is an alternative density \( f_1(\cdot) \), then \( \hat{\theta}_1 \) is asymptotically normally distributed with mean \( (\theta_1 - h^2\mu_2\theta_2 + o(h^2)) \) and variance \( (n^{-2}h^{-5}D_1 + n^{-1}D_2 + o(n^{-2}h^{-5} + n^{-1})) \), where
\[
D_1 = 2\kappa_1 \int_{0}^{1} f_1^2(z) dz,
\]
\[
D_2 = 4 \left( \int_{0}^{1} [f_1''(z)]^2 f_1(z) dz - \theta_1^2 \right), \tag{2.7}
\]
and \( \mu_2 = \int u^2 K(u) du \).
One implication of Theorem 2.1 is that an asymptotic \( \alpha \)-level test for \( H_0 \) can be obtained by rejecting the null hypothesis whenever

\[
\hat{\theta}_1 > \frac{z_{1-\alpha}}{\sqrt{n^2 h^5/2\kappa_1}},
\]

(2.8)

where \( z_{1-\alpha} \) is the \( 100(1 - \alpha) \)-th percentile of the standard normal distribution. The results from Theorem 2.2 can be used to quantify the performance of the proposed test. The asymptotic power function against an alternative \( f_1(\cdot) \) is approximately

\[
\Pi(f_1) = P\left\{ \hat{\theta}_1 > \frac{z_{1-\alpha}}{\sqrt{n^2 h^5/2\kappa_1}} | f = f_1 \right\}
\approx P\left\{ Z > \frac{z_{1-\alpha}/\sqrt{n^2 h^5/2\kappa_1} - \theta_1 + h^2 \mu_2 \theta_2}{(n^{-2}h^{-5}D_1 + n^{-1}D_2)^{1/2}} \right\},
\]

(2.9)

where \( Z \) has the standard normal distribution.

3 Asymptotic Power

To make power calculations on the test described above, we first consider a fixed alternative distribution of the form

\[
G(x) = \begin{cases} 
0, & \text{if } x \leq 0, \\
(1 - s)x + sH\left( \frac{x - x_0}{\sigma} \right), & \text{if } 0 < x < 1, \\
1, & \text{if } x \geq 1,
\end{cases}
\]

(3.1)

where \( H(\cdot) \) is an absolutely continuous distribution function defined on an interval \([a_0, b_0] \subseteq [-1, 1]\), \( x_0 \) is an interior point \((h < x_0 < 1 - h)\), and \( s \) and \( \sigma \) are fixed constants with \( 0 < s < 1 \). From (3.1) the alternative distribution function \( G(\cdot) \) is the null distribution contaminated by \( H(\cdot) \) with contamination \( s \) at scale rate \( \sigma \). Let

\[
g(x) = \left( 1 - s + s\sigma^{-1}\eta\left( \frac{x - x_0}{\sigma} \right) \right) I_{(0,1)}(x)
\]

be the density function corresponding to (3.1) with \( \eta(\cdot) \) a density function and \( H(y) = \int_{-\infty}^{y} \eta(t)dt \).

It follows from (2.9) that

\[
\Pi(g) \approx P\left\{ Z > \frac{z_{1-\alpha}2\kappa_1 - nh^{2.5}s^2\sigma^{-3} \int \eta'(y)dy}{(2\kappa_1(1-s)^2 + 4nh^5(1-s)s^2\sigma^{-5} \int \eta''(y)dy)^{1/2}} \right\}.
\]

(3.2)

Clearly, the RHS of (3.2) tends to 1 as \( n \to \infty \). Thus our test is consistent against fixed alternatives of the form (3.1).
Suppose that, we allow the alternatives to approach $H_0$ as $n$ increases, i.e.

$$G_n(x) = (1 - s_n)x + s_nH\left(\frac{x - x_0}{\sigma_n}\right), \quad \text{for } 0 < x < 1,$$

(3.3)

where the sequences $s_n$, $h$, and $\sigma_n$ tend to 0 as $n \to \infty$ at appropriate rates to be specified. Thus the alternative density functions are of the form

$$g_n(x) = \left(1 - s_n + s_n\sigma_n^{-1}\eta\left(\frac{x - x_0}{\sigma_n}\right)\right)I_{(0,1)}(x).$$

(3.4)

Assume that $h$, $s_n$, and $\sigma_n$ are of orders $n^{-\delta}$, $n^{-\epsilon}$, and $n^{-\gamma}$ respectively with $\delta > 0$, $\epsilon > 0$, and $\gamma \geq 0$. The asymptotic power of the proposed test against contaminated alternatives (3.3) is given in the following theorem.

**Theorem 3.1** For $\delta \in (0, 2/5)$ and $0 \leq \gamma < \delta$,

$$\lim_{n \to \infty} \text{II}(g_n) = \begin{cases}
\alpha & \text{if } \epsilon > (1 + 3\gamma - 2.5\delta)/2, \\
\Phi(l) & \text{if } \epsilon = (1 + 3\gamma - 2.5\delta)/2, \\
1 & \text{if } \epsilon < (1 + 3\gamma - 2.5\delta)/2,
\end{cases}$$

(3.5)

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable and

$$l = \int \frac{[\eta'(u)]^2}{\sqrt{2\kappa_1}} du - z_{1-\alpha}.$$  

(3.6)

Furthermore, for $\delta \in (0, 2/5)$ and $0 < \delta \leq \gamma$,

$$\lim_{n \to \infty} \text{II}(g_n) = \begin{cases}
\alpha & \text{if } \epsilon > \max((1 + 3\gamma - 2.5\delta)/2, (1 + 5\gamma - 5\delta)/2), \\
\alpha & \text{if } \epsilon = (1 + 3\gamma - 2.5\delta)/2, \text{ and } \epsilon \geq (1 + 5\gamma - 5\delta)/2, \\
1 & \text{if } \epsilon < \min((1 + 3\gamma - 2.5\delta)/2, (1 + \gamma)/2).
\end{cases}$$

(3.7)

The results of Theorem 3.1 can be derived from (2.9) and (3.4). The details are available from the author.

A special case of Theorem 3.1, when against smooth alternatives ($\gamma = 0$), is stated in the following Corollary.

**Corollary 3.1** For $\delta \in (0, 2/5)$, $0 < \epsilon < 1/2$, and $\gamma = 0$,

$$\lim_{n \to \infty} \text{II}(g_n) = \begin{cases}
\alpha & \text{if } \delta > (2 - 4\epsilon)/5, \\
\Phi(l) & \text{if } \delta = (2 - 4\epsilon)/5, \\
1 & \text{if } \delta < (2 - 4\epsilon)/5.
\end{cases}$$

(3.8)
Some observations can be made from Theorem 3.1 and Corollary 3.1. First, as $h$ becomes larger ($\delta$ becomes smaller), the power against smooth alternatives improves. This feature conflicts with the intuitive reasoning that when $h$ becomes smaller at an appropriate rate, the test statistic may be more sensitive to densities with local features. Second, our test is less powerful than the EDF-based tests such as the Kolmogorov-Smirnov test when testing smooth alternatives. The Bickel-Rosenblatt test also shares the same features, as indicated by Bickel and Rosenblatt (1973) and Ghosh and Huang (1991). But, when against alternative distributions with contamination, the proposed test can have greater power with an appropriate choice of smoothing parameter. For example, when $h$ is chosen such that $\delta > \gamma > 5/6\delta_0$, our test achieves power 1 for contamination rate $\varepsilon < (1 + 3\gamma - 2.5\delta_0)/2$. An example of $\delta = 1/7$, $\gamma = .14$, and $\varepsilon < .531$ belongs to this category. It is well known that the standard tests will have power $\alpha$ for $\varepsilon > 1/2$, even when the contaminating distribution $H(\cdot)$ is discrete such as a point mass at $x_0$. On the other hand, the proposed test requires that the underlying density satisfies Condition (D1), and hence the contaminating density $\eta(\cdot)$ must be continuous and smooth.

To implement the proposed testing procedure, an immediate problem is how to choose an appropriate smoothing parameter $h$. The bandwidth which maximizes the power function (see (2.9)) may be effective, i.e. choose the bandwidth such that it minimizes

$$B(h) = (z_{1-\alpha}/\sqrt{n^2h^5/2\kappa_1 - \theta_1 + h^2\mu_2\theta_2})/(n^{-2}h^{-5}D_1 + n^{-1}D_2)^{1/2}.$$  \hspace{1cm} (3.9)

It is not difficult to observe that:

- if $nh^5 \to \infty$, then $B(h) \to -\sqrt{n}\theta_1/\sqrt{D_2}$.

- if $nh^5 \to c$, then $B(h) \to -\sqrt{nc}\theta_1/(D_1 + cD_2)^{1/2}$, which increases as $c$ increases.

Hence, the first-order asymptotically optimal smoothing parameter must satisfy $nh^5 \to \infty$. Note that the optimal bandwidth for estimating $f^\ast f(x)dx$ in Theorem 3.3 of Hall and Marron (1987) is of order $n^{-2/9}$, and therefore their optimal smoothing parameter is not appropriate in the current testing problem.

The second-order optimal smoothing parameter may be obtained by differentiating $B(h)$ with respect to $h$ and setting it to 0,

$$\frac{dB(h)}{dh} = \left(n^{-2}h^{-5}D_1 + n^{-1}D_2\right)^{-3/2} \left(2n^{-1}h\mu_2\theta_2 D_2 - 2.5n^{-2}h^{-6}\theta_1 D_1 + o(n^{-1}h + n^{-2}h^{-6})\right)$$

$$= 0.$$  \hspace{1cm} (3.10)
One solution to (3.10) is
\[ h = \left( \frac{5\theta_1 D_1}{4\mu_2\theta_2 D_2 n} \right)^{1/7} + o_P(n^{-1/7}). \] (3.11)

4 Testing goodness-of-fit in action

We carry out some simulations to investigate the empirical power and the level of the proposed testing procedure. The Biweight kernel \( K(u) = \frac{15}{16}(1 - u^2)^2 I_{[-1,1]}(u) \) is used throughout the simulation study. The choice of the smoothing parameter is based on the first-order optimal bandwidth for estimating \( f' \):
\[ h = \left( \frac{3 \int K^2(u) du}{\mu^2 \int f^{(3)}(x) dx} \right)^{-1/7} n^{-1/7}. \]

We then obtain a bandwidth selector with reference to the normal densities,
\[ h = 1.307 sn^{-1/7}, \]
where \( s \) is the sample standard deviation of the data.

It is the author's experience that the boundary modifications in (2.6) introduce large variation into the null hypothesis distribution of the test statistic \( \hat{\theta}_1 \). The reason is that the asymptotic theory does not take into effect for moderate sample sizes, as to be shown. Although the power against some alternatives were excellent, the significance level under the null hypothesis tends to be larger than the selected level. To account for this, we compute the asymptotic variance of \( S_{n,1}, S_{n,2} \) and \( S_{n,3} \), the three terms in (2.6) respectively. It was found that under the null hypothesis,
\[
\begin{align*}
\text{Var}\{S_{n,1}|f = f_0\} & = 7.52n^{-2}h^{-5} + o(n^{-2}h^{-5}), \\
\text{Var}\{S_{n,2}|f = f_0\} & = 1375.95n^{-2}h^{-4} + o(n^{-2}h^{-4}), \\
\text{Var}\{S_{n,3}|f = f_0\} & = 1375.95n^{-2}h^{-4} + o(n^{-2}h^{-4}).
\end{align*}
\]

Clearly the variance of \( S_{n,1} \) is not the dominant term as stated in Theorem 2.1, unless the data size is huge. We therefore consider the following two test statistics for practical implementation:
\[
\begin{align*}
T_1 & = S_{n,1} = \frac{1}{n(n-1)h^4} \sum_{i \neq j} \int_h^{1-h} K' \left( \frac{X_i - x}{h} \right) K' \left( \frac{X_j - x}{h} \right) dx, \\
T_2 & = S_{n,1} + h \hat{f}^2(h) + h \hat{f}^2(1-h),
\end{align*}
\]
where
\[
\hat{f}^2(h) = \max \left( n^{-1}(n-1)^{-1}h^{-4} \sum_{i \neq j} K' \left( \frac{X_i - h}{h} \right) K' \left( \frac{X_j - h}{h} \right), 0 \right),
\]

\[ 9 \]
and

\[
\tilde{f}^2(1-h) = \max \left( n^{-1}(n-1)^{-1}h^{-4} \sum_{i \neq j} K' \left( \frac{X_i - (1-h)}{h} \right) K' \left( \frac{X_j - (1-h)}{h} \right), 0 \right).
\]

In other words, \( \tilde{f}^2(h) \) and \( \tilde{f}^2(1-h) \) are bias-corrected estimates of \( f^2(h) \) and \( f^2(1-h) \) respectively. \( T_1 \) is basically \( \hat{\theta}_1 \) without boundary modifications. One can criticize that \( T_1 \) may fail to detect the local features present in the boundary regions. Even so, \( T_1 \) performs reasonably well in our power study. The motivation of \( T_2 \) is that instead of employing boundary kernels, one can use the simple estimates \( h f^2(h) \) and \( h f^2(1-h) \) for \( \int_0^h f^2(x)dx \) and \( \int_{1-h}^1 f^2(x)dx \) respectively.

The empirical power results for four test statistics, \( T_1, T_2, \) Kolmogorov-Smirnov (K-S) and Cramér-von Mises (CVM), were obtained by simulating from a variety of different distributions. For each example, data of sample sizes of \( n = 50 \) and \( n = 200 \) are generated. We took \( \alpha = .05 \) and hence the critical value is \( 1.645n^{-1}h^{-2.5}/\sqrt{2\kappa_1} \), with \( \kappa_1 = 3.76 \) for the Biweight kernel. The proposed test is to reject \( H_0 \) if

\[
\sqrt{n^2h^5/2\kappa_1} T_1 > 1.645, \quad \text{and} \quad \sqrt{n^2h^5/2\kappa_1} T_2 > 1.645, \quad (4.1)
\]

for \( T_1 \) and \( T_2 \) respectively. One fact of the uniform density on \((0, 1)\) is that it has the largest variance \( 1/12 \) among all nondegenerate and continuous densities on \((0, 1)\). Taking advantage of this property, the power of the proposed test can be enhanced by computing the ratio of \( 1/12 \) to the data variance, \( r = 1/(12 \times \text{Var(data)}), \) and then rejecting \( H_0 \) if

\[
\sqrt{n^2(hr)^5/2\kappa_1} T_1 > 1.645,
\]

\[
\sqrt{n^2(hr)^5/2\kappa_1} T_2 > 1.645. \quad (4.2)
\]

Proportions of rejection were calculated based on 400 simulations.

**Example 1:**

\[
H_0 : F = N(0, 1) \quad \text{versus} \quad F = 0.3N(0, \frac{\sigma^2}{0.3\sigma^2+0.7}) + 0.7N(0, \frac{1}{0.3\sigma^2+0.7}), \quad \sigma = 4 \times 1.4^{-i}, \ i = 0, \ldots, 10.
\]

Note that the mean and standard deviation of this family are the same as the standard normal distribution. This example is designed to examine the power against a kurtosis departure; that is, a departure from normality that preserves symmetry. The densities and the transformed densities
are plotted in Figures 1(a) and 1(b) respectively. It is clear that the underlying densities become sharper as $\sigma$ becomes smaller. The proportions of rejecting the null hypothesis in 400 simulations are shown in Figures 1(c) and 1(d) for $n = 50$ and $n = 200$ respectively. Both $T_1$ and $T_2$ statistics have higher power, in comparison with the Kolmogorov-Smirnov and Cramér-von Mises tests. This example illustrates the capability of $T_1$ and $T_2$ in detecting sharp peaks.

**Example 2:**

\[
H_0 : F = N(0,1) \quad \text{versus} \quad 0.3N\left(\frac{-\mu}{0.3\sigma}, 1/\sigma\right) + 0.7N\left(\frac{\mu}{0.7\sigma}, 1/\sigma\right),
\]

with $\sigma = \sqrt{1 + \mu^2(1/0.3 + 1/0.7)}$ for $\mu = 0, 0.1, \ldots, 1$.

**Put Figure 2 about here**

We examine a mean-mixture family (Figure 2(a)) in this example. Again, the mean and standard deviation of this family are the same as the standard normal density. From Figures 2(c) and 2(d), the $T_2$ statistic shows better power than $T_1$. Figure 2(b) shows that there is sharpness in the left boundary region. This indicates that for some densities, boundary regions $(0, h)$ and $(1-h, 1)$ may contain information that is necessary to discriminate from the null density.

**Example 3:**

\[
H_0 : \text{Uniform}(0,1) \quad \text{versus} \quad f(x) = aI_{[0<x<a]} + (1+a)I_{[a<x<1]}, \quad a = 0.1, 0.2, \ldots, 1.
\]

**Put Figure 3 about here**

As shown in Example 1, the $T_1$ and $T_2$ test statistics have high power in detecting densities with sharp peaks. It will be interesting to see the performance of $T_1$ and $T_2$ against a family with more uniform features, as plotted in Figure 3(a), although this family does not satisfy Condition (D1). It can be observed from Figures 3(b) and 3(c) that the EDF-based tests outperform $T_1$ and $T_2$, particularly for the sample size $n = 50$.

Since the critical values of the Kolmogorov-Smirnov and the Cramér von Mises tests are different for sample sizes 50 and 200, we may use the empirical critical values of $T_1$ and $T_2$ for a fair comparison. The 95% percentiles of normalized $T_1$ and $T_2$ (see (4.1)) are given in the following table. The results are that the proportions of rejection for $T_1$ and $T_2$ drop a bit, but a similar phenomenon is observed.
<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>1.8804</td>
<td>2.1756</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.8684</td>
<td>1.9530</td>
</tr>
</tbody>
</table>

In summary, the simulation results support the asymptotic analysis. They indicate that the proposed test is to be preferred for discriminating densities with sharp features. It may not be consistent against all alternatives and may be less powerful against some alternatives. We feel that the testing procedures $T_1$ and $T_2$ have promise as competitors at least for moderate sample sizes.

5 Appendix—Proofs

Proof of Theorem 2.1: Recall

$$
\hat{\theta}_1 = S_{n,1} + S_{n,2} + S_{n,3},
$$

where $S_{n,1}$, $S_{n,2}$ and $S_{n,3}$ are the three terms in (2.6) respectively. Denote

$$
S_{n,1} = n^{-1}(n-1)^{-1}h^{-4}\sum_{i\neq j}H_n(X_i, X_j)
$$

with

$$
H_n(X_i, X_j) = \int_h^{1-h} K'\left(\frac{X_i - x}{h}\right) K'\left(\frac{X_j - x}{h}\right) dx.
$$

It is clear that $S_{n,1}$ is a one-sample U-statistic and $H_n$ is symmetric with $E\{H_n(X_1, X_2)|X_1\} = 0$ under the null hypothesis. Some calculations show that

$$
E\{H_n^2(X_1, X_2)|f = f_0\} = h^3 \int [(K' * K')(u)]^2 du + o(h^3)
$$

and

$$
E\{H_n^4(X_1, X_2)|f = f_0\} = O(h^5),
$$

$$
E\{G_n^2(X_1, X_2)|f = f_0\} = O(h^7),
$$

where $G_n(x, y) = E\{H_n(X_1, X)H_n(X_1, y)\}$. Checking the condition in Theorem 1 of Hall (1984),

$$
\frac{E\{G_n^2(X_1, X_2)|f = f_0\} + n^{-1}E\{H_n^4(X_1, X_2)|f = f_0\}}{(E\{H_n^2(X_1, X_2)|f = f_0\})^2} \to 0,
$$

12
as \( n \to \infty \). It follows that \( \sum_{1 \leq i < j \leq n} H_n(X_i, X_j) \) is asymptotically normally distributed with mean 0 and variance \( n^2 E\{H_n^2(X_1, X_2)|f = f_0\}/2 \). Therefore \( S_{n,1} \) is asymptotically normally distributed with mean 0 and variance \( n^{-2}h^{-5}2\kappa_1 + o(n^{-2}h^{-5}) \). \( S_{n,2} \) and \( S_{n,3} \) are asymptotically negligible in comparison with \( S_{n,1} \). Thus the asymptotic normality of \( \hat{\theta}_1 \) under the null hypothesis is proved. \( \square \)

**Proof of Theorem 2.2:** Again we evaluate the three terms \( S_{n,1}, S_{n,2}, \) and \( S_{n,3} \) in the sequel. Write

\[
S_{n,1} = n^{-1}(n-1)^{-1}h^{-4} \sum_{i \neq j} \int_h^{1-h} \left( K' \left( \frac{X_i - x}{h} \right) - E \left\{ K' \left( \frac{X_i - x}{h} \right) \right\} \right) dx
\]

\[
\times \left( K' \left( \frac{X_j - x}{h} \right) - E \left\{ K' \left( \frac{X_j - x}{h} \right) \right\} \right) dx
\]

\[
+ 2n^{-1}(n-1)^{-1}h^{-4} \sum_{i \neq j} \int_h^{1-h} \left( K' \left( \frac{X_i - x}{h} \right) - E \left\{ K' \left( \frac{X_i - x}{h} \right) \right\} \right) dx
\]

\[
\times \left( \frac{X_i - x}{h} \right) dx
\]

\[
+ n^{-1}(n-1)^{-1}h^{-4} \sum_{i \neq j} \int_h^{1-h} E \left\{ K' \left( \frac{X_i - x}{h} \right) \right\} E \left\{ K' \left( \frac{X_j - x}{h} \right) \right\} dx.
\]

(5.1)

The last term in (5.1) is purely deterministic and under the alternative hypothesis \( f(\cdot) = f_1(\cdot) \),

\[
\int_h^{1-h} E \left\{ K' \left( \frac{X_i - x}{h} \right) \right\} E \left\{ K' \left( \frac{X_j - x}{h} \right) \right\} dx = h^4 \left( \int_h^{1-h} f_1^2(x)dx - h^2 \mu_2 \theta_2 \right) + o(h^6).
\]

(5.2)

Observe that the second term (\( I_{n,2} \)) in the RHS of (5.1) can be written as a sum of i.i.d. random variables.

\[
I_{n,2} = 2n^{-1}h^{-4} \sum_{i=1}^{n} Z_{n,i}
\]

\[
= 2n^{-1}h^{-4} \sum_{i=1}^{n} (Y_{n,i} - E\{Y_{n,i}\}),
\]

where \( Y_{n,i} = \int_h^{1-h} K' \left( \frac{X_i - x}{h} \right) E \left\{ K' \left( \frac{X_i - x}{h} \right) \right\} dx \). One can show that

\[
E\{Y_{n,i}\} = h^4 \int_0^1 f_1^2(x)dx + o(h^4),
\]

\[
E\{Y_{n,i}^2\} = h^8 \int_0^1 f_1(x)f_1^2(x)dx + o(h^8),
\]

\[
E\{Y_{n,i}^3\} = O(h^{12}),
\]

\[
E\{Y_{n,i}^4\} = O(h^{16}),
\]

13
and so that

\[ E\{Z_{n,i}^2\} = h^8 \left( \int_0^1 f_1(x)f_1''(x)dx - \theta_1^2 \right) + o(h^8), \]

\[ E\{Z_{n,i}^4\} = O(h^{16}). \]

The first term (\(\equiv I_{n,1}\)) in the RHS of (5.1) may be expressed as

\[ I_{n,1} = 2n^{-1}(n-1)^{-1}h^{-4} \sum_{i=2}^{n} \sum_{j=1}^{n-i} \tilde{H}_n(X_i, X_j), \tag{5.3} \]

where

\[ \tilde{H}_n(X_i, X_j) = \int_0^{1-h} \left( K' \left( \frac{X_i-x}{h} \right) - E\left\{ K' \left( \frac{X_i-x}{h} \right) \right\} \right) \times \left( K' \left( \frac{X_j-x}{h} \right) - E\left\{ K' \left( \frac{X_j-x}{h} \right) \right\} \right) dx. \]

Let \(X_0 = 0, W_{n,1} = 0,\) and \(W_{n,i} = \sum_{j=1}^{i-1} \tilde{H}_n(X_i, X_j), i = 2, \ldots, n.\) Then

\[ I_{n,1} + I_{n,2} = 2n^{-1}(n-1)^{-1}h^{-4} \sum_{i=1}^{n} (W_{n,i} + (n-1)Z_{n,i}) \tag{5.4} \]

and \(Z_{n,i}\) and \(W_{n,i}\) are uncorrelated. Note that \(I_{n,1} + I_{n,2}\) is a martingale array with respect to the \(\sigma\)-fields \(\sigma(X_0, \ldots, X_i), i = 0, \ldots, n.\) To show \(I_{n,1} + I_{n,2}\) is asymptotically normally distributed, by the martingale central limit theorem (see Hall and Heyde (1980)), we check the following two conditions:

\[ \sum_{i=1}^{n} E\{(W_{n,i} + (n-1)Z_{n,i})^4\} = o(s_n^4) \tag{5.5} \]

and

\[ s_n^{-2}V_n^2 \longrightarrow 1 \quad \text{in probability,} \tag{5.6} \]

as \(n \to \infty,\) where \(V_n^2 = \sum_{i=1}^{n} E\{(W_{n,i} + (n-1)Z_{n,i})^2|X_0, \ldots, X_{i-1}\}\) and

\[ s_n^2 = \sum_{i=1}^{n} \text{Var}(W_{n,i} + (n-1)Z_{n,i}) \]

\[ = n(n-1)^2h^8 \left( \int_0^1 f_1(x)f_1''(x)dx - \theta_1^2 \right) + n(n-1)h^3\kappa_1 \int_0^1 f^2(x)dx/2 \]

\[ + o(n^3h^8 + n^2h^3). \]

Some calculations show that for \(i = 2, \ldots, n,\)

\[ E\{Z_{n,i}^2W_{n,i}\} = 0, \]
\[ E\{Z_{n,i}^2 W_{n,i}^2\} = (i - 1)E\{Z_{n,i}^2 \tilde{H}_n^2(X_1, X_2)\} = (i - 1)O(h^8), \]
\[ E\{Z_{n,i} W_{n,i}^3\} = (i - 1)E\{Z_{n,i} \tilde{H}_n^3(X_1, X_2)\} = (i - 1)O(h^7), \]
\[ E\{W_{n,i}^4\} = (i - 1)E\{\tilde{H}_n(X_1, X_2)\} + 4(i - 1)(i - 2)(E\{\tilde{H}_n^2(X_1, X_2)\})^2 \\
= (i - 1)O(h^5) + 4(i - 1)(i - 2)O(h^6). \]

And hence
\[ \sum_{i=1}^{n} E\{(W_{n,i} + (n - 1)Z_{n,i})^4\} = O(n^2 h^5 + n^3 h^6 + n^3 h^7 + n^4 h^9 + n^5 h^{16}) \]
\[ = o(s_n^4). \tag{5.7} \]

That proves (5.5). To show (5.6), it suffices to check
\[ s_n^{-4}E\{(V_n^2 - s_n^2)^2\} \to 0, \quad \text{as } n \to \infty. \tag{5.8} \]

Let \( v_{n,i} = E\{(W_{n,i} + (n - 1)Z_{n,i})^2|X_0, \ldots, X_{i-1}\}, \) \( G_n(x, y) = E\{\tilde{H}_n(X_1, x)\tilde{H}_n(X_1, y)\} \) and \( L_n(x) = E\{\tilde{H}_n(X_1, x)Z_1\}. \) Then for \( j_1 \leq k_1, j_2 \leq k_2, \)
\[ E\{G_n(X_{j_1}, X_{k_1})G_n(X_{j_2}, X_{k_2})\} = \left\{ \begin{array}{ll} E\{G_n^2(X_{1}, X_{1})\} & \text{if } j_1 = k_1 = j_2 = k_2, \\
(E\{G_n(X_{1}, X_{1})\})^2 & \text{if } j_1 = k_1 \neq j_2 = k_2, \\
E\{G_n^2(X_{1}, X_{2})\} & \text{if } j_1 = j_2, k_1 = k_2, j_1 < k_1, \\
0 & \text{otherwise.} \end{array} \right. \]

For \( j_2 \leq k_2, \)
\[ E\{L_n(X_{j_1})G_n(X_{j_2}, X_{k_2})\} = \left\{ \begin{array}{ll} E\{L_n(X_{1})G_n(X_{1}, X_{1})\} & \text{if } j_1 = j_2 = k_2, \\
0 & \text{otherwise.} \end{array} \right. \]

And for \( j_1 \leq j_2, \)
\[ E\{L_n(X_{j_1})L_n(X_{j_2})\} = \left\{ \begin{array}{ll} E\{L_n^2(X_{1})\} & \text{if } j_1 = j_2, \\
0 & \text{otherwise.} \end{array} \right. \]

Now for \( i_1 \leq i_2, \)
\[ E\{v_{n,i_1} v_{n,i_2}\} = (n - 1)^4(E\{Z_{n,i_1}^2\})^2 + (n - 1)^2E\{Z_{n,i_1}^2\}E\{W_{n,i_2}^2\} \\
+ (n - 1)^2E\{Z_{n,i_2}^2\}E\{W_{n,i_1}^2\} + 4(n - 1)(i_1 - 1)E\{L_n^2(X_{1})\} \\
+ 4(n - 1)(i_1 - 1)E\{L_n(X_{1})G_n(X_{1}, X_{1})\} + (i_1 - 1)\text{Var}\{G_n^2(X_{1}, X_{1})\} \\
+ (i_1 - 1)(i_2 - 1)(E\{G_n(X_1, X_1)\})^2 \\
+ 2(i_1 - 1)(i_1 - 2)E\{G_n^2(X_{1}, X_{2})\}. \]
Consequently,

\[
E\{(V_n^2 - s_n^2)^2\} = \sum_{i,j=1}^{n} (E\{v_{n,i}v_{n,j}\} - E\{v_{n,i}\}E\{v_{n,j}\})
\]

\[
= 2 \sum_{2 \leq i < j \leq n} E(v_{n,i}v_{n,j}) + \sum_{i=1}^{n} E(v_{n,i}^2) - \left(\sum_{i=1}^{n} E\{v_{n,i}\}\right)^2
\]

\[
= 4(n-1)^2 \sum_{i=2}^{n} (i-1)(2n-2i+1)E\{L_n^2(X_1)\}
\]

\[
+ 4(n-1) \sum_{i=2}^{n} (i-1)(2n-2i+1)E\{L_n(X_1)G_n(X_1, X_1)\}
\]

\[
+ 2 \sum_{i=2}^{n} (i-1)(i-2)(2n-2i+1)E\{G_n^2(X_1, X_2)\}
\]

\[
+ \sum_{i=2}^{n} (i-1)(2n-2i+1)\text{Var}\{G_n(X_1, X_1)\}
\]

and

\[
s_n^{-4} E\{(V_n^2 - s_n^2)^2\} \leq \text{const} \frac{n^5h^{14} + n^4h^8 + n^4h^7 + n^3h^6}{(O(n^3h^8 + n^2h^3))^2} \rightarrow 0, \quad (5.9)
\]

as \(n \rightarrow \infty\).

From (5.1), (5.4), (5.7), and (5.9), \(S_{n,1}\) is asymptotically normal with mean \((\int_0^h f_{1-h}^1 f_1^h(x)dx - h^2 \mu_2 \theta_2 + o(h^2))\) and variance \((n^{-2}h^{-5}D_1 + n^{-1}D_2 + o(n^{-2}h^{-5} + n^{-1}))\). The other two terms in (2.6) are

\[
S_{n,2} = \int_0^h f_{1-h}^1 f_1^h(x)dx + O_P(h^3 + n^{-1}h^{-2} + n^{-1/2}h^{1/2}),
\]

\[
S_{n,3} = \int_{1-h}^1 f_1^h(x)dx + O_P(h^3 + n^{-1}h^{-2} + n^{-1/2}h^{1/2}).
\]

Therefore under the alternative hypothesis, \(\hat{\theta}_1\) has an asymptotic normal distribution with mean \((\theta_1 - h^2 \mu_2 \theta_2 + o(h^2))\) and variance \((n^{-2}h^{-5}D_1 + n^{-1}D_2 + o(n^{-2}h^{-5} + n^{-1}))\).

Reference


17

Figure 1: Plots of Example 1: (a) Densities; (b) Transformed densities; (c) Proportions of rejection for $n = 50$; (d) Proportions of rejection for $n = 200$. 
Figure 2: Plots of Example 2: (a) Densities; (b) Transformed densities; (c) Proportions of rejection for \( n = 50 \); (d) Proportions of rejection for \( n = 200 \).
Figure 3: Plots of Example 3: (a) Densities; (b) Proportions of rejection for $n = 50$; (c) Proportions of rejection for $n = 200$. 