Nonparametric Statistics, Vol. 2, pp. 389–403
Reprints available directly from the publisher
Photocopying permitted by license only

HYPOTHESES TESTING BASED ON MODIFIED NONPARAMETRIC ESTIMATION OF AN AFFINITY MEASURE BETWEEN TWO DISTRIBUTIONS

YANQIN FAN and RAMAZAN GENCAY*

Department of Economics, University of Windsor, Windsor, Ontario, Canada, N9B 3P4

(Received March 2, 1992, Revised: February 2, 1993; Accepted February 17, 1993)

Let F and G denote two distribution functions defined on the same probability space which are absolutely continuous with respect to the Lebesgue measure with probability density functions f and g, respectively. Ahmad and Van Belle (1974) proposed a measure of the closeness between F and G as follows: $\lambda = \lambda(F, G) = 2\int f(x)g(x)\,dx/[\int f^2(x)\,dx + \int g^2(x)\,dx]$. Ahmad (1980) proposed to estimate λ by $\hat{\lambda} = [\int \hat{f}(x)\,dG_n(x) + \int \hat{g}(x)\,dF_n(x)]/[\int \hat{f}^2(x)\,dx + \int \hat{g}^2(x)\,dx]$, where $F_n(x)$ and $G_n(x)$ are the empirical distribution functions of $F(x)$ and $G(x)$ respectively and $\hat{f}(x)$ and $\hat{g}(x)$ are the well-known kernel estimates of $f(x)$ and $g(x)$ respectively. This paper generalizes the estimator $\hat{\lambda}$ to a family of modified estimators of λ indexed by a constant γ, $\hat{\lambda}_\gamma$, say, where $0 \le \gamma \le 1$, which includes $\hat{\lambda}$ as a special case (when $\gamma = 0$). We derive the limiting distribution of normalized $\hat{\lambda}_\gamma$ for $0 < \gamma \le 1$ by using the theory of U-statistics and show that the limiting distribution of $\hat{\lambda}_\gamma$ for $\gamma = 0$, i.e., of $\hat{\lambda}$, when normalized, is degenerate. Consequently, $\hat{\lambda}$ cannot be used to construct an asymptotically valid goodness-of-fit test. The normalized estimator $\hat{\lambda}_\gamma$ for any $0 < \gamma \le 1$, however, does have a limiting normal distribution and therefore can be used to construct an asymptotically valid two sample goodness-of-fit test. The modifications of λ proposed by Ahmad (1980) for one sample case suffer from the same problem. So, in this paper, we also generalize Ahmad's estimators of λ for one sample case and apply the resulting estimators in hypotheses testing. All the tests proposed in this paper are shown to be consistent.

1. INTRODUCTION

Let F and G be two independent distributions defined on the same probability space which are absolutely continuous with respect to the Lebesgue measure with probability density functions f and g respectively. Ahmad and Van Belle (1974) proposed a measure of the affinity between f and g, viz,

$$\lambda = \lambda(F, G) = \frac{2\delta}{\Delta(f) + \Delta(g)} \tag{1}$$

where $\delta = \int f(x)g(x)\,dx$, $\Delta(f) = \int f^2(x)\,dx$, and $\Delta(g) = \int g^2(x)\,dx$ with $f^2(x) = [f(x)]^2$. The measure λ satisfies the property that it does not exceed 1 and it

* The first author would like to thank the Natural Sciences and Engineering Research Council of Canada for its support and the second author gratefully acknowledges financial support under a grant from University of Windsor Research Board and the Social Science Research Unit. We are very grateful to I. A. Ahmad for his many helpful comments and suggestions which greatly improved an earlier version of this paper.

equals 1 if and only if $f(x) = g(x)$ for almost all x. Therefore, it can be used as a basis for testing $H_0: f(x) = g(x)$. This idea was explored by Ahmad (1980).

Ahmad (1980) proposed a nonparametric estimator of λ based on two independent random samples from F and G. In particular, given two independent random samples, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, say, from F and G respectively, he proposed to estimate λ by

$$\hat{\lambda} = \frac{2\hat{\delta}}{\hat{\Delta}(f) + \hat{\Delta}(g)} \tag{2}$$

where

$$\hat{\delta} = \int \hat{f}(x)\, \mathrm{d}G_n(x) = \int \hat{g}(x)\, \mathrm{d}F_n(x) = \frac{1}{n^2 a_n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{X_i - Y_j}{a_n}\right) \tag{3}$$

and

$$\hat{\Delta}(f) = \int \hat{f}^2(x)\, \mathrm{d}x \quad \text{and} \quad \hat{\Delta}(g) = \int \hat{g}^2(x)\, \mathrm{d}x \tag{4}$$

In equations (3) and (4), $F_n(x)$ and $G_n(x)$ denote the empirical distribution functions of $F(x)$ and $G(x)$, respectively, and $\hat{f}(x)$ and $\hat{g}(x)$ are kernel estimators of $f(x)$ and $g(x)$ respectively, viz.,

$$\hat{f}(x) = \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{a_n}\right) \quad \text{and} \quad \hat{g}(x) = \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{x - Y_i}{a_n}\right) \tag{5}$$

where $K(.)$ is a symmetric probability density function (p.d.f.) satisfying certain assumptions to be stated later and $a = a_n$ is a bandwidth approaching zero as n approaches infinity.

Ahmad (1980) discussed some asymptotic properties of $\hat{\lambda}$, including the weak and strong consistency of $\hat{\lambda}$ under certain assumptions. In addition, he presented the limiting distribution of $\sqrt{n}\,(\hat{\lambda} - \lambda)$, on the basis of which he constructed an asymptotic test of the null hypothesis that $f(x)$ is equal to $g(x)$ almost everywhere. In the same paper, Ahmad also discussed estimation of two special forms of λ. The first is estimating λ when $g = f_0$ is known. Let $\lambda = \lambda_0$ in this case, where $\lambda_0 = 2\delta_0/[\Delta(f) + \Delta(f_0)]$ and $\delta_0 = \int f(x) f_0(x)\, \mathrm{d}x$. He proposed the following estimator for λ_0, viz.,

$$\hat{\lambda}_0 = \frac{2n^{-1} \sum_{i=1}^{n} f_0(X_i)}{\hat{\Delta}(f) + C_0}, \tag{6}$$

where $C_0 = \int f_0^2(x)\, \mathrm{d}x$ is a known constant. A one sample test statistic for $H_0: f = f_0$ was constructed from $\hat{\lambda}_0$. The second is estimating

$$\lambda^* = \frac{2 \int f(x) f(-x)\, \mathrm{d}x}{\int f^2(x)\, \mathrm{d}x + \int f^2(-x)\, \mathrm{d}x} = \frac{\delta^*}{\Delta(f)} \tag{7}$$

by

$$\hat{\lambda}^* = \frac{\hat{\delta}^*}{\hat{\Delta}(f)} \tag{8}$$

where $\delta^* = \int f(x)f(-x)\,dx$, $\hat{\delta}^* = (n^2a)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K((X_i+X_j)/a)$, and the last equality of (7) follows from the fact that $\int f^2(x)\,dx = \int f^2(-x)\,dx$. Based on this estimate, a test statistic was proposed for testing H_0:f is symmetric about zero, i.e., $f(x)=f(-x)$ for all real x.

The estimator $\hat{\Delta}(f)$ of $\Delta(f)$ was first proposed by Bhattacharayya and Roussas (1969). The showed that it is a consistent estimator in mean square error sense. Schuster (1974) derived the strong consistency and the rate of strong convergence of $\hat{\Delta}(f)$. He also proposed an alternative estimator for $\Delta(f)$, i.e.,

$$\tilde{\Delta}(f) = \int \hat{f}(x)\,dF_n(x) = \frac{1}{n^2a}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_i - X_j}{a}\right) \tag{9}$$

and showed that $\tilde{\Delta}(f)$ is also a strongly consistent estimator of $\Delta(f)$ under the same conditions as those of the estimator $\hat{\Delta}(f)$. Ahmad (1976) derived the limiting normal distribution of $\sqrt{n}\,(\tilde{\Delta}(f) - \Delta(f))$. The limiting distribution of $\sqrt{n}\,(\hat{\Delta}(f) - \Delta(f))$ follows from that of $\sqrt{n}\,(\tilde{\Delta}(f) - \Delta(f))$ as seen in Lemma 2.1 to follow.

This paper generalizes the estimator $\hat{\lambda}$ proposed by Ahmad (1980) to a family of modified estimators of λ indexed by a constant γ, $\hat{\lambda}_\gamma$, say, where $0 \le \gamma \le 1$. If $\gamma = 0$, then the modified estimator $\hat{\lambda}_\gamma$ reduces to $\hat{\lambda}$. In the next section of the paper, we define the modified estimator $\hat{\lambda}_\gamma$ for $0 \le \gamma \le 1$ and investigate the limiting distribution of $\sqrt{n}\,(\hat{\lambda}_\gamma - \lambda)$ by invoking the theory of U-statistics (or V-statistics). Our result when specialized to $\gamma = 0$ shows that under the null hypothesis of $f(x) = g(x)$ the limiting distribution of $\sqrt{n}\,(\hat{\lambda} - 1)$ is degenerate. There is no scaling of $(\hat{\lambda} - 1)$ that will have a limiting distribution of known or consistently estimable form. Hence, we will not be able to construct an asymptotically valid test based on $\hat{\lambda}$. One approach to overcome the degeneracy problem is to modify $\hat{\lambda}$. The estimator $\hat{\lambda}_\gamma$ for any $0 < \gamma \le 1$ serves this purpose, since as will be shown in the next section, $\sqrt{n}\,(\hat{\lambda}_\gamma - \lambda)$ has a limiting normal distribution whether the null hypothesis is correct or not. An alternative method of dealing with the degeneracy problem is to modify the measure of affinity being used, i.e. λ, in our paper. See Ahmad and Cerrito (1989) for more details and references of work along this direction.

The same problem exists in one sample case. Thus, we also provide modified estimators of λ_0 and λ^* and derive their respective limiting distributions in the next section. The third section presents asymptotic tests based on our modified estimators. It is also shown in this section that all three tests are consistent. The size and power properties of these three tests are discussed in section four. It is demonstrated through some simulation studies that all three tests have good power against a variety of alternatives for moderate sample sizes. The last section includes a discussion of possible extensions of the modified test statistics to the testing of the residuals of a regression model and the concluding remarks.

2. MODIFIED ESTIMATORS AND THEIR LIMITING DISTRIBUTIONS

To construct an asymptotically valid test for H_0:$f(x) = g(x)$ (or H_0:$f(x) = f_0(x)$, or, H_0:$f(x) = f(-x)$), we need an estimator of λ (or λ_0, or, λ^*) that when appropriately scaled has a known or consistently estimable limiting distribution under H_0. As indicated in the introduction, the estimator $\hat{\lambda}$ (or $\hat{\lambda}_0$, or $\hat{\lambda}^*$) is not an appropriate candidate for this purpose. The intuitive reason is that both the numerator and the denominator of $\hat{\lambda}$ (or $\hat{\lambda}_0$, or $\hat{\lambda}^*$) are asymptotically equivalent to the same sample average under H_0. This suggests that we must estimate either the numerator or the denominator differently. In this paper, we choose the former and provide the following modified estimator of λ, viz.,

$$\hat{\lambda}_\gamma = \frac{2\hat{\delta}_\gamma}{\hat{\Delta}(f) + \hat{\Delta}(g)} \tag{10}$$

where $\hat{\Delta}(f)$ and $\hat{\Delta}(g)$ are given in (4) and $\hat{\delta}_\gamma$ is a generalized estimator of δ indexed by a constant $0 < \gamma \leq 1$. It is defined as[1]

$$\hat{\delta}_\gamma = \frac{1}{2n_\gamma} \sum_{i=1}^{n} C_i(\gamma)[\hat{f}(Y_i) + \hat{g}(X_i)]$$

where

$$C_i(\gamma) = \begin{cases} 1+\gamma & \text{for } i \text{ odd} \\ 1-\gamma & \text{for } i \text{ even} \end{cases} \quad \text{and} \quad n_\gamma = \begin{cases} n & \text{for } n \text{ even} \\ n+\gamma & \text{for } n \text{ odd} \end{cases}.$$

Similarly, we propose to estimate λ_0 and λ^* respectively by

$$\hat{\lambda}_{0,\gamma} = \frac{2\hat{\delta}_{0,\gamma}}{\hat{\Delta}(f) + C_0} \tag{11}$$

where $\hat{\delta}_{0,\gamma} = n_\gamma^{-1} \sum_{i=1}^{n} C_i(\gamma) f_0(X_i)$, a weighted estimator of δ_0, and

$$\hat{\lambda}_\gamma^* = \frac{\hat{\delta}_\gamma^*}{\hat{\Delta}(f)} \tag{12}$$

where $\hat{\delta}_\gamma^* = n_\gamma^{-1} \sum_{i=1}^{n} C_i(\gamma) \hat{f}(-X_i)$, and $\hat{\Delta}(f)$ and C_0 are defined in the introduction.

Remark 2.1. The modified estimators $\hat{\lambda}_{0,\gamma}$, $\hat{\lambda}_\gamma^*$, and $\hat{\lambda}_\gamma$ are indexed by a parameter γ. If γ is allowed to take the value zero, then these estimators reduce to those of Ahmad (1980). However, for testing purposes, γ must not be zero.

Common to all three estimators $\hat{\lambda}_{0,\gamma}$, $\hat{\lambda}_\gamma^*$, and $\hat{\lambda}_\gamma$ is the estimator $\hat{\Delta}(f)$ of $\Delta(f)$. We will first study the limiting distribution of $\sqrt{n}(\hat{\Delta}(f) - \Delta(f))$ and then use the result to examine the respective limiting distributions of the appropriately normalized estimators.

For ease of reference, we collect the assumptions on $K(.)$ and $a = a_n$ in the assumption (K) and the assumption (A) below. Throughout the rest of this paper, these two assumptions are assumed to be satisfied and all the limits are taken as $n \to \infty$, unless otherwise stated.

Assumptions. (K) The Kernel function $K(.)$ is a symmetric p.d.f satisfying

(K1) $\sup_u K(u) < \infty$ and $|u|\, K(u) \to 0$ as $|u| \to \infty$,
(K2) $\int uK(u)\, du = 0$ and $\int u^2 K(u)\, du < \infty$.

(A) The smoothing parameter $a = a_n$ satisfies $a \to 0$, $na^2 \to \infty$, and $na^4 \to 0$.

[1] The idea of using weighted estimators is borrowed from Robinson (1991).

The following lemma is obtained from Theorems 2.2 and 2.3[2] in Ahmad (1976) by relating $\hat{\Delta}(f)$ to $\tilde{\Delta}(f)$ given in (9).

LEMMA 2.1 If $f(x)$ is twice differentiable with bounded second derivative and $\int f^3(x)\, dx < \infty$, then

(i) $\hat{\Delta}(f) - \Delta(f) = 2n^{-1} \sum_{i=1}^{n} [f(X_i) - Ef(X_i)] + o_p(n^{-1/2})$
(ii) $\sqrt{n}\,[\hat{\Delta}(f) - \Delta(f)] \to N(0, 4\,\mathrm{Var}[f(X)])$ in distribution.

Proof.[3] The results (i) and (ii) follow from Theorems 2.2 and 2.3 in Ahmad (1976), since

$$\begin{aligned}\hat{\Delta}(f) &= \frac{1}{a^2}\int\int\int K\left(\frac{x-u}{a}\right)K\left(\frac{x-v}{a}\right) dF_n(u)\, dF_n(v)\, dx\\ &= \frac{1}{a}\int\int\left[\int K(w)K\left(w+\frac{v-u}{a}\right) dw\right] dF_n(u)\, dF_n(v)\\ &= \frac{1}{n^2 a}\sum_{i=1}^{n}\sum_{j=1}^{n} K^{(2)}\left(\frac{X_i - X_j}{a}\right)\\ &= \tilde{\Delta}(f)\end{aligned}$$

where $K^{(2)}(.)$ in the last equation is the convolution of $K(.)$ with itself and $K^{(2)}(.)$ satisfies the assumption (K). □

The limiting distribution of $\sqrt{n}\,(\hat{\lambda}_{0,\gamma} - \lambda_0)$ is the simplest to derive and stated in the following theorem.

THEOREM 2.2. If $f(x)$ has a bounded second derivative, $\int f_0^3(x) f(x)\, dx < \infty$, $\int f^4(x)\, dx < \infty$, $\int f^3(x) f_0(x)\, dx < \infty$, and $\int f^2(x) f_0^2(x)\, dx < \infty$, then $\sqrt{n}\,(\hat{\lambda}_{0,\gamma} - \lambda_0) \to N(0, \sigma_{0,\gamma}^2)$, where

$$\sigma_{0,\gamma}^2 = \frac{4\{(1+\gamma^2)\,\mathrm{Var}[f_0(X)] + \lambda_0^2\,\mathrm{Var}[f(X)] - 2\lambda_0\,\mathrm{Cov}[f_0(X), f(X)]\}}{[\Delta(f) + C_0]^2}.$$

Proof. By using Lemma 2.1, we have

$$\begin{aligned}\hat{\lambda}_{0,\gamma} - \lambda_0 &= \frac{2n_\gamma^{-1}\sum_i C_i(\gamma)[f_0(X_i) - Ef_0(X_i)] - 2\lambda_0 n^{-1}\sum_i [f(X_i) - Ef(X_i)] + o_p(n^{-1/2})}{\hat{\Delta}(f) + C_0}\\ &= \frac{2n^{-1}\sum_i \{[nn_\gamma^{-1}C_i(\gamma)f_0(X_i) - \lambda_0 f(X_i)] - E[nn_\gamma^{-1}C_i(\gamma)f_0(X_i) - \lambda_0 f(X_i)]\}}{\hat{\Delta}(f) + C_0}\\ &\quad + o_p(n^{-1/2}) \qquad (13)\end{aligned}$$

Let $Y_{ni} = nn_\gamma^{-1}C_i(\gamma)f_0(X_i) - \lambda_0 f(X_i)$ and $Z_{ni} = Y_{ni} - EY_{ni}$. Then, $EZ_{ni} = 0$ and $Z_{n1}, \ldots, Z_{nn}$ are independent and satisfy the condition $E|Z_{ni}|^3 < \infty$ by the

[2] The proof of Theorem 2.2 is also used.
[3] We are indebted to I. A. Ahmad for providing this simple proof.

assumptions on $f(x)$ and $f_0(x)$. So by the Liapounov CLT we will obtain

$$\frac{n^{-1/2}\sum_i Z_{ni}}{\sqrt{n^{-1}\operatorname{Var}[\sum_i Z_{ni}]}} \to N(0,1)$$

if we can show that $\sum_i E|Z_{ni}|^3/(\sum_i \operatorname{Var} Z_{ni})^{3/2} \to 0$, or equivalently that $n^{-3/2}\sum_i E|Z_{ni}|^3 \to 0$, since $n^{-1}\operatorname{Var}[\sum_i Z_{ni}] \to [\Delta(f)+C_0]^2\sigma_{0,\gamma}^2/4$. But, it is obvious that $\sum_i E|Z_{ni}|^3 = O(n)$ which completes the proof. □

It is worth noting that Theorem 2.2 holds whether the null hypothesis that $f(x)=f_0(x)$ is true or not. For testing purposes, we only need

COROLLARY 2.3. Given the assumptions in Theorem 2.2, if $f(x)=f_0(x)$, then $\sqrt{n}(\hat{\lambda}_{0,\gamma}-1) \to N(0, \sigma_{0,\gamma,0}^2)$, where $\sigma_{0,\gamma,0}^2 = \gamma^2 \int f_0(x)[f_0(x)-C_0]^2\,dx/C_0^2$.

Proof. It follows from Theorem 2.2 and the fact that under $H_0: f(x)=f_0(x)$, $\lambda_0 = 1$ and $\operatorname{Cov}[f_0(x), f(X)] = \operatorname{Var}[f_0(X)]$. □

Remark 2.2.

(i) If $\gamma = 0$, Theorem 2.2 provides the limiting distribution of $\sqrt{n}(\hat{\lambda}_0 - \lambda_0)$ when $f(x) \neq f_0(x)$, where $\hat{\lambda}_0$ is the estimator of λ_0 proposed by Ahmad (1980).

(ii) As long as $0<\gamma\le 1$, Corollary 2.3 shows that $\sqrt{n}(\hat{\lambda}_{0,\gamma}-\lambda_0)$ will have a well defined limiting distribution under $H_0: f(x)=f_0(x)$ which enables us to construct a valid test for H_0 based on $\hat{\lambda}_{0,\gamma}$.

We now analyze the estimator $\hat{\lambda}_\gamma^*$. The first result stated in Lemma 2.4 is concerned with the asymptotic behavior of $\hat{\delta}_\gamma^*$.

LEMMA 2.4. If $f(x)$ is twice differentiable with bounded second derivative and $\int f(x)f^3(-x)\,dx < \infty$, then

(i) $\hat{\delta}_\gamma^* - \delta^* = n^{-1}\sum_i C_i^*(\gamma)[f(-X_i) - Ef(-X_i)] + o_p(n^{-1/2})$, where $C_i^*(\gamma) = [nn_\gamma^{-1}C_i(\gamma)+1]$.

(ii) $\sqrt{n}(\hat{\delta}_\gamma^* - \delta^*) \to N(0, (4+\gamma^2)\operatorname{Var}[f(-X)])$

Proof. Note that $\hat{\delta}_\gamma^*$ can be regarded as a weighted V-statistic of second order, viz.,

$$\hat{\delta}_\gamma^* = \frac{1}{n^2}\sum_i\sum_j \frac{nC_i(\gamma)}{n_\gamma} P_n(X_i, X_j)$$

where $P_n(X_i, X_j) = a^{-1}K((X_i + X_j)/a)$.

Let $P_n(X_i) = E[P_n(X_i, X_j) \mid X_i]$. Then, it is easy to show that the following result holds for all i, viz.,

$$P_n(X_i) = f(-X_i) + o_p(n^{-1/2}). \tag{14}$$

Let $\theta_n = EP_n(X_i, X_j) = EP_n(X_i)$. Then

$$\hat{\delta}_\gamma^* - \delta^* = \frac{1}{n^2}\sum_i\sum_j Q_n(X_i, X_j) + \frac{1}{n_\gamma}\sum_i C_i(\gamma)[P_n(X_i) - \theta_n] + \frac{1}{n}\sum_j [P_n(X_j) - \theta_n] + [\theta_n - \delta^*] \tag{15}$$

where

$$Q_n(X_i, X_j) = \frac{nC_i(\gamma)}{n_\gamma}[P_n(X_i, X_j) - \theta_n] - \frac{nC_i(\gamma)}{n_\gamma}[P_n(X_i) - \theta_n] - [P_n(X_j) - \theta_n].$$

Given equation (14), it is sufficient to show $\Sigma_i \Sigma_j Q_n(X_i, X_j) = o_p(n^{3/2})$ and $\theta_n - \delta^* = o(n^{-1/2})$. The latter is clearly true given (14). The former is equivalent to the following two equations, viz.,

$$\sum_i Q_n(X_i, X_i) = o_p(n^{3/2}) \tag{16}$$

$$\sum_i \sum_{j,i\neq j} Q_n(X_i, X_j) = o_p(n^{3/2}) \tag{17}$$

It is trivial to see that the first of the above two equations holds. For the second, it is sufficient to show that $E[\Sigma_i \Sigma_{j,i\neq j} Q_n(X_i, X_j)]^2 = o(n^3)$. Since the observations are independent, the left hand side equals $\Sigma_i \Sigma_{j,i\neq j} E[Q_n(X_i, X_j)]^2$ which does not exceed

$$3\sum_i \sum_{j\neq i}\left\{\frac{n^2C_i(\gamma)^2}{n_\gamma^2}E[P_n(X_i, X_j) - \theta_n]^2 + \frac{n^2C_i(\gamma)^2}{n_\gamma^2}E[P_n(X_i) - \theta_n]^2 + E[P_n(X_j) - \theta_n]^2\right\}$$

$$\leq 3\sum_i \sum_{j\neq i}\left[\frac{2n^2C_i(\gamma)^2}{n_\gamma^2} + 1\right]E[P_n(X_1, X_2)]^2$$

$$\leq 3n\left[\frac{2n^3}{n_\gamma^2}(1+\gamma^2) + n\right]E_{X_1}\int \frac{1}{a}K^2(t)f(-X_1 + at)\,dt$$

$$= O(n^2a^{-1})$$

The result in (i) then follows from the assumption that $na \to \infty$ as $n \to \infty$.

(ii) follows from (i), the Liapounov CLT, and the fact that $n^{-1}\Sigma_i C_i^{*2}(\gamma) \to (4+\gamma^2)$ as $n \to \infty$. □

Lemma 2.4 along with Lemma 2.1 yields

THEOREM 2.5. Under the assumptions stated in Lemma 2.4, if $\int f^4(x)\,dx < \infty$, $\int f^2(x)f^2(-x)\,dx < \infty$, and $\int f^2(x)f(-x)\,dx < \infty$, then $\sqrt{n}(\hat{\lambda}_\gamma^* - \lambda^*) \to N(0, \sigma_\gamma^{*2})$, where

$$\sigma_\gamma^{*2} = \frac{4+\gamma^2}{\Delta^2(f)}\mathrm{Var}[f(-X)] + \frac{4\lambda^{*2}}{\Delta^2(f)}\mathrm{Var}[f(X)] - \frac{8\lambda^*}{\Delta^2(f)}\mathrm{Cov}[f(-X), f(X)].$$

Proof. From Lemma 2.1 and Lemma 2.4, we obtain

$$\hat{\lambda}_\gamma^* - \lambda^* = \frac{[\hat{\delta}_\gamma^* - \delta^*] - \lambda^*[\hat{\Delta}(f) - \Delta(f)]}{\hat{\Delta}(f)}$$

$$= \frac{n^{-1}\Sigma_i C_i^*(\gamma)[f(-X_i) - Ef(-X_i)] - 2\lambda^* n^{-1}\Sigma_i [f(X_i) - Ef(X_i)]}{\hat{\Delta}(f)} + o_p(n^{-1/2})$$

$$= \frac{n^{-1}\Sigma_i \{[C_i^*(\gamma)f(-X_i) - 2\lambda^* f(X_i)] - E[C_i^*(\gamma)f(-X_i) - 2\lambda^* f(X_i)]\}}{\hat{\Delta}(f)} + o_p(n^{-1/2})$$

Therefore, $\sqrt{n}\,(\hat{\lambda}_\gamma^* - \lambda^*) \rightarrow N(0, \sigma_\gamma^{*2})$ by the Liapounov CLT, where

$$\sigma_\gamma^{*2} = \lim_{n\to\infty}\left\{\frac{1}{\Delta^2(f)n}\sum_i \operatorname{Var}[C_i^*(\gamma)f(-X_i) - 2\lambda^* f(X_i)]\right\}$$

$$= \frac{4+\gamma^2}{\Delta^2(f)}\operatorname{Var}[f(-X)] + \frac{4\lambda^{*2}}{\Delta^2(f)}\operatorname{Var}[f(X)] - \frac{8\lambda^*}{\Delta^2(f)}\operatorname{Cov}[f(-X), f(X)] \quad \square$$

To construct a test of the null hypothesis that $f(x)$ is symmetric around zero, we need the following special case of Theorem 2.5.

COROLLARY 2.6. Under the same assumptions as in Theorem 2.5, $\sqrt{n}\,(\hat{\lambda}_\gamma^* - 1) \rightarrow N(0, \sigma_{\gamma,0}^{*2})$, if $f(x) = f(-x)$, where

$$\sigma_{\gamma,0}^{*2} = \frac{\gamma^2 \int f(x)[f(x) - \delta^*]^2\,dx}{\Delta^2(f)} = \frac{\gamma^2 \int f(x)[f(x) - \Delta(f)]^2\,dx}{\Delta^2(f)}$$

Finally, we consider the general estimator $\hat{\lambda}_\gamma$.

LEMMA 2.7. If $f(x)$ and $g(x)$ have bounded second derivatives, and if $\int f(x)g^3(x)\,dx < \infty$ and $\int f^3(x)g(x)\,dx < \infty$, then

(i) $\hat{\delta}_\gamma - \delta = (2n)^{-1}\sum_i C_i^*(\gamma)[g(X_i) - Eg(X_i) + f(Y_i) - Ef(Y_i)] + o_p(n^{-1/2})$, where $C_i^*(\gamma) = [nn_\gamma^{-1}C_i(\gamma) + 1]$.

(ii) $\sqrt{n}\,(\hat{\delta}_\gamma - \delta) \rightarrow N(0, (4+\gamma^2)\operatorname{Var}[g(X) + f(Y)]/4)$.

Proof. It is important to note that $\hat{\delta}_\gamma$ can be regarded as a weighted V-statistic of the following form,

$$\hat{\delta}_\gamma = \frac{1}{n^2}\sum_i\sum_j \frac{nC_i(\gamma)}{2n_\gamma} P_n(Z_i, Z_j),$$

where $P_n(Z_i, Z_j) = a^{-1}[K((X_i - Y_j)/a) + K((Y_i - X_j)/a)]$, and $Z_i = (X_i, Y_i)'$. Let $P_n(Z_i) = E[P_n(Z_i, Z_j) \mid Z_i]$. Then, it can be shown that for all i

$$P_n(Z_i) = f(Y_i) + g(X_i) + o_p(n^{-1/2}). \quad (18)$$

Let $\theta_n = EP_n(Z_i) = EP_n(Z_i, Z_j)$. Then

$$\hat{\delta}_\gamma - \delta = \frac{1}{n^2}\sum_i\sum_j Q_n(Z_i, Z_j) + \frac{1}{2n_\gamma}\sum_i C_i(\gamma)[P_n(Z_i) - \theta_n] + \frac{1}{2n}\sum_j [P_n(Z_j) - \theta_n] + \left[\frac{\theta_n}{2} - \delta\right] \quad (19)$$

where

$$Q_n(Z_i, Z_j) = \frac{nC_i(\gamma)}{2n_\gamma}[P_n(Z_i, Z_j) - \theta_n] - \frac{nC_i(\gamma)}{2n_\gamma}[P_n(Z_i) - \theta_n] - \tfrac{1}{2}[P_n(Z_j) - \theta_n]$$

Following the same argument as in the proof of Lemma 2.4, we can show that the first term in equation (19) is $o_p(n^{-1/2})$. The result in (i) then follows from (18) and (19).

(ii) follows from (i) and the Liapounov CLT. $\square$

Combining the results in Lemma 2.1 and Lemma 2.7, we have

THEOREM 2.8. Suppose that the assumptions of Lemma 2.7 are satisfied and $\int f(x)g^3(x)\,dx < \infty$, $\int f^3(x)g(x)\,dx < \infty$, $\int g^4(x)\,dx < \infty$, and $\int f^4(x)\,dx < \infty$, then $\sqrt{n}\,(\hat{\lambda}_\gamma - \lambda) \to N(0, \sigma_\gamma^2)$, where

$$\sigma_\gamma^2 = \frac{(4+\gamma^2)\,\mathrm{Var}[g(X)+f(Y)] + 4\lambda^2\,\mathrm{Var}[f(X)+g(Y)]}{[\Delta(f)+\Delta(g)]^2} - \frac{8\lambda\{\mathrm{Cov}[f(X), g(X)] + \mathrm{Cov}[f(Y), g(Y)]\}}{[\Delta(f)+\Delta(g)]^2}$$

Proof. Let $Z_i^* = C_i^*(\gamma)[g(X_i) + f(Y_i)] - 2\lambda[f(X_i) + g(Y_i)]$. Then the result follows from the following expression obtained from Lemma 2.1 and Lemma 2.7, and the Liapounov CLT,

$$\hat{\lambda}_\gamma - \lambda = \frac{n^{-1}\sum_i [Z_i^* - EZ_i^*] + o_p(n^{-1/2})}{\hat{\Delta}(f) + \hat{\Delta}(g)}. \qquad \square$$

From Theorem 2.8, we have

COROLLARY 2.9. Under $H_0: f(x) = g(x)$, $\sqrt{n}\,(\hat{\lambda}_\gamma - 1) \to N(0, \sigma_{\gamma,0}^2)$, where

$$\sigma_{\gamma,0}^2 = \frac{\gamma^2 \int f(x)[f(x) - \delta]^2\,dx}{\Delta^2(f)}.$$

Remark 2.3. Corollary 2.9 implies that $\hat{\lambda}_\gamma$ can be used to construct a two sample goodness-of-fit test as long as the parameter $0 < \gamma \le 1$.

3. TEST STATISTICS AND THEIR PROPERTIES

This section develops tests of $H_0: f(x) = f_0(x)$, $H_0': f(x) = f(-x)$, and $H_0'': f(x) = g(x)$, respectively, where $f_0(x)$ is assumed to be known. The alternative hypotheses for these tests are that the corresponding null hypotheses are false. The basic idea is to use affinity measures λ_0, λ^*, and λ defined in the introduction to measure the "closeness" between $f(x)$ and $f_0(x)$, $f(x)$ and $f(-x)$, and $f(x)$ and $g(x)$, respectively. Since λ_0, λ^*, and λ take the value 1 if and only if the respective null hypotheses hold, and take values less than 1 if and only if the alternative hypotheses hold, we will reject the null hypotheses if the estimates $\hat{\lambda}_{0,\gamma}$, $\hat{\lambda}_\gamma^*$, and $\hat{\lambda}_\gamma$ are smaller than 1 by an appropriate amount. To determine the rejection regions for the corresponding null hypotheses at a given significance level α, say, we will make use of the limiting distributions given in Corollaries 2.3, 2.6, and 2.9.

Recall from Corollary 2.3, under H_0, $\sqrt{n}\,(\hat{\lambda}_{0,\gamma} - 1) \to N(0, \sigma_{0,\gamma,0}^2)$, where $\sigma_{0,\gamma,0}^2 = \gamma^2 \int f_0(x)[f_0(x) - C_0]^2\,dx / C_0^2$. Let $\hat{T}_{\gamma,0} = \sqrt{n}\,(\hat{\lambda}_{0,\gamma} - 1)/\sigma_{0,\gamma,0}$ be the test statistic for H_0. We have

THEOREM 3.1. If $f(x)$ has bounded second derivative and $\int f_0^4(x)\,dx < \infty$, then under H_0, $\hat{T}_{\gamma,0} \to N(0, 1)$ in distribution.

It follows from Theorem 3.1 that H_0 will be rejected at significance level α if $\hat{T}_{\gamma,0} < -Z_\alpha$, where Z_α is such that $P(Z > Z_\alpha) = \alpha$ for $Z \sim N(0, 1)$.

The next result concerns the consistency of the test $\hat{T}_{\gamma,0}$. It ensures that the probability of rejecting the null hypothesis of $f(x)=f_0(x)$ when in fact it is not true approaches 1 as sample size approaches infinity. This result is obtained by noting that under the alternative hypothesis of $f(x)\neq f_0(x)$, $n^{-1/2}\hat{T}_{\gamma,0}\to(\lambda_0-1)\sigma_{0,\gamma,0}^{-1}<0$ in probability. So, for a given significance level α, we can always choose n sufficiently large such that $\hat{T}_{\gamma,0}<-Z_\alpha$.

THEOREM 3.2. The test for H_0 given by $\hat{T}_{\gamma,0}$ and the left tail of the standard normal distribution is consistent.

Similarly, we can construct an asymptotically valid test for $H_0':f(x)=f(-x)$ by using Corollary 2.6 which states that under H_0', $\sqrt{n}\,(\hat{\lambda}_\gamma^*-1)\to N(0,\sigma_{\gamma,0}^{*2})$, where $\sigma_{\gamma,0}^{*2}=\gamma^2\int f(x)[f(x)-\delta^*]^2\,dx/\Delta^2(f)$. The variance $\sigma_{\gamma,0}^{*2}$ can be consistently estimated by

$$\hat{\sigma}_{\gamma,0}^{*2}=\frac{\gamma^2\sum_i[\hat{f}(X_i)-\hat{\delta}_\gamma^*]^2}{n\hat{\Delta}^2(f)},$$

which suggests the following test statistic for H_0', viz., $\hat{T}_{\gamma,*}=\sqrt{n}\,(\hat{\lambda}_\gamma^*-1)/\hat{\sigma}_{\gamma,0}^*$. The consistency of $\hat{\sigma}_{\gamma,0}^*$ and Corollary 2.6 lead to

THEOREM 3.3. If $f(x)$ is twice differentiable with bounded second derivative and $\int f^4(x)\,dx<\infty$, then under H_0', $\hat{T}_{\gamma,*}\to N(0,1)$.

So we reject H_0' at significance level α if $\hat{T}_{\gamma,*}<-Z_\alpha$. The previous argument for the consistency of $\hat{T}_{\gamma,0}$ also applies to $\hat{T}_{\gamma,*}$. Thus, we have

THEOREM 3.4. The test for H_0' given by $\hat{T}_{\gamma,*}$ is consistent.

The third test is a two sample goodness-of-fit test, i.e., $H_0'':f(x)=g(x)$. Under H_0'', Corollary 2.9 states that $\sqrt{n}\,(\hat{\lambda}_\gamma-1)\to N(0,\sigma_{\gamma,0}^2)$, where $\sigma_{\gamma,0}^2=\gamma^2\int f(x)[f(x)-\delta]^2\,dx/\Delta^2(f)$. The unknown parameter $\sigma_{\gamma,0}^2$ can be consistently estimated by

$$\hat{\sigma}_{\gamma,0}^2=\frac{\gamma^2\sum_i[\hat{f}(Y_i)-\hat{\delta}_\gamma]^2}{n\hat{\Delta}^2(f)}.$$

Let $\hat{T}_\gamma=\sqrt{n}\,(\hat{\lambda}_\gamma-1)/\hat{\sigma}_{\gamma,0}$. Then, the test: reject H_0'' at significance level α if $\hat{T}_\gamma<-Z_\alpha$ is an asymptotically valid test, since

THEOREM 3.5. If $f(x)$ and $g(x)$ are twice differentiable with bounded second derivatives and $\int f^4(x)\,dx<\infty$ and $\int g^4(x)\,dx<\infty$, then under H_0'', $\hat{T}_\gamma\to N(0,1)$.

It is clear that the following result also holds.

THEOREM 3.6. The above two sample goodness-of-fit test is consistent.

4. DESCRIPTION OF THE MONTE CARLO EXPERIMENTS

In this section, we will investigate the size and power characteristics of the tests presented in section 3 for moderate sample sizes. These tests are the goodness-of-fit test, $\hat{T}_\gamma$; the symmetry test, $\hat{T}_{\gamma,*}$; and the one sample goodness-of-fit test, $\hat{T}_{\gamma,0}$.

In the evaluation of the probability density function at a given point in its domain, the true probability density function of the corresponding random variable will be utilized. Hence, the power and the size calculations presented here are the best that can be achieved given the sample size. Therefore, these results can be considered as the benchmark performance of the test statistics.

The detailed size and power calculations with nonparametric kernel density estimation will be carried out in future research. The estimation of the density points with nonparametric kernel density estimation technique requires rather careful choice of the bandwidth parameter.

All test statistics are calculated for sample sizes of $n = 50$ and $n = 100$. To measure the power of the test statistics, four different alternative distributions are chosen. These are: t_5, Student's t with five degrees of freedom;[4] lognormal distribution; χ^2_2, chi-squared with two degrees of freedom and the standard exponential distribution. The mean and the variance of the underlying normal distribution are chosen to be one for the lognormal distribution. In each case the random variables are standardized to have expectation zero and variance one.

t_5 is chosen to capture a heavier tail structure. t_5 has a kurtosis of 9. Lognormal, χ^2_2, and standard exponential distribution are skewed and have higher kurtosis than standard normal. χ^2_2 has a kurtosis of 9 and skewness of 2; lognormal has a kurtosis of 113.94 and skewness of 6.19; and χ^2_2 and standard exponential probability density functions both have a kurtosis of 9 and skewness of 2.

For each experiment 200 replications are performed. All pseudo-random generates are obtained from the IMSL (The International Mathematical and Statistical Libraries 1987).

The size of each statistic is calculated at the five and ten percent levels by using the tabulated percentage points for the true statistics.

The calculation of the test statistic requires a choice for the γ parameter. In small samples, γ should be chosen such that the empirical size of the test is within its nominal size. We experimented various choices for γ and found that when $0.55 < \gamma < 0.70$ all three test statistics have good empirical sizes for both $n = 50$ and $n = 100$. Accordingly, we set $\gamma = 0.65$ in this study.

This section is organized in the following manner. In Section 4.1, the finite sample properties of the one sample goodness-of-fit test are analysed. Section 4.2 covers the finite sample properties of the symmetry test. The finite sample properties of the goodness-of-fit test are explained in 4.3.

4.1. The One Sample Goodness-of-Fit Test

If $g = f_0$ is chosen to be a normal distribution then $\hat{T}_{\gamma,0}$ becomes a test for normality. Here we specify f_0 to be the standard normal distribution and in this subsection, we will refer to $\hat{T}_{\gamma,0}$ as a test for normality. The size and the calculated moments of the test are given in Table 1.
It is evident from Table 1 that the test has a good empirical size both at the 5%

[4] The kurtosis of the Student's t distribution is not defined for degrees of freedom less than five which led us to choose t_5.

Table 1. The moments and the size of the normality test with 200 replications

No. of observations	$n = 50$	$n = 100$
Size at 5%	0.045	0.055
Size at 10%	0.095	0.100
Mean	0.092	0.034
Variance	1.171	1.107
Skewness	0.041	−0.064
Kurtosis	2.601	2.981
Minimum	−2.561	−3.028
Maximum	2.752	2.933

and 10% levels. The moments of the test statistic are very close to the moments of the asymptotic distribution of the test statistic.

The power of our normality test is presented in Table 2 below. Each cell in this table reports the ratio of the number of rejections to the number of replications. The critical value of the standard normal distribution at the 5% level is −1.645 and at the 10% level is −1.280.

Notice that the test statistic has very good power against all alternative distributions except t_5 for sample size as small as 50 observations. The performance of the test against t_5 improves for larger number of data points. This is summarized in Table 3 below.

From Table 3 we can conclude that the normality test reaches to a good power against t_5 distribution with 2000 observations.

Table 2. The power of the normality test with 200 replications

	$n = 50$		$n = 100$	
Distribution	5%	10%	5%	10%
χ_2^2	1.00	1.00	1.00	1.00
t_5	0.060	0.115	0.100	0.180
Lognormal	1.00	1.00	1.00	1.00
Exponential	1.00	1.00	1.00	1.00

Table 3. The power of the normality test against t_5 distribution with 200 replications

n	%5	%10
1000	0.510	0.710
1500	0.660	0.825
2000	0.855	0.940
2500	0.975	1.000

Table 4. The moments and the size of the symmetry test with 200 replications

No. of observations	$n = 50$	$n = 100$
Size at 5%	0.045	0.060
Size at 10%	0.105	0.100
Mean	0.002	−0.030
Variance	1.141	1.195
Skewness	0.162	−0.009
Kurtosis	3.276	2.839
Minimum	−2.968	−2.801
Maximum	3.623	2.897

4.2. The Symmetry Test

In this section the size and the power of the symmetry test, $\hat{T}_{Y,*}$ are summarized. With the symmetry test, the null hypothesis is composite for all symmetric distributions. The size and the calculated moments of the test statistic with the standard normal random deviates are given in Table 4 above.

The test statistic has good size and the calculated moments accurately capture the moments of its asymptotic distribution. The power of the test statistic is presented in Table 5.

Since the null hypothesis is composite, we would expect that the power of the test statistic should be equal to its size for all symmetric distributions. This situation is presented with t_5 distribution in Table 5. Although the test slightly overrejects, the percentage of rejections are within their nominal sizes. For all non-symmetric distributions the test has a power of 100.0%.

4.3. The Goodness-of-Fit Test

In this last subsection, the size and power properties of the goodness-of-fit test, $\hat{T}_Y$ are investigated. As in the case of the symmetry test, the null hypothesis of the goodness-of-fit test is composite. In Table 6 below, the size of this test is calculated with random numbers generated from standard normal and t_5. In the calculation of the size of the test, two sets of independent random numbers are generated from each distribution.

The goodness-of-fit test has good empirical size both at 5% and 10% levels for $n = 50$ and $n = 100$. The power of the test statistic is presented in Table 7 below.

Table 5. The power of the symmetry test with 200 replications

	$n = 50$		$n = 100$	
Distribution	5%	10%	5%	10%
χ^2_2	1.00	1.00	1.00	1.00
t_5	0.085	0.150	0.070	0.130
Lognormal	1.00	1.00	1.00	1.00
Exponential	1.00	1.00	1.00	1.00

Table 6. The moments and the size of the goodness-of-fit test with 200 replications

No. of observations	$n = 50$		$n = 100$	
Distribution	Standard Normal	t_5	Standard Normal	t_5
Size at 5%	0.050	0.040	0.060	0.050
Size at 10%	0.080	0.090	0.110	0.095
Mean	0.082	0.023	−0.069	−0.039
Variance	1.116	0.973	0.945	0.961
Skewness	−0.084	−0.074	−0.037	−0.094
Kurtosis	3.056	3.043	2.941	3.051
Minimum	−2.967	−2.434	−2.544	−2.411
Maximum	2.770	2.900	2.160	2.951

We tried various combinations of distributions which satisfy $f(x) \neq g(x)$. As presented in Table 7, the test has excellent power for all comparisons of distributions except standard normal versus t_5 and exponential versus χ_2^2.

Table 7. The power of the goodness-of-fit test with 200 replications

	$n = 50$		$n = 100$	
Distribution	5%	10%	5%	10%
Standard Normal, t_5	0.040	0.100	0.060	0.115
Standard Normal, Exponential	1.00	1.00	1.00	1.00
Standard Normal, χ_2^2	1.00	1.00	1.00	1.00
Standard Normal, Lognormal	1.00	1.00	1.00	1.00
Exponential, χ_2^2	0.085	0.115	0.090	0.120
Exponential, Lognormal	0.865	0.895	0.965	0.965
t_5, χ_2^2	1.00	1.00	1.00	1.00

One explanation for this is that exponential and χ_2^2 distributions have the same skewness and kurtosis values and this similarity plays an important role for the power calculation. As we compared standard exponential versus χ_3^2, which in this case skewness and kurtosis values between these two distributions are different, the power of the test statistic goes up to 100% even at $n = 50$. As for the comparison of standard normal versus t_5, the power goes up as the number of observations is increased. However, the gain in the power is relatively less than the case with the normality test which is presented in Table 3.

5. CONCLUSION

In this study, we proposed a modified estimator of λ and showed that this new estimator, when appropriately normalized, has a limiting normal distribution. Therefore, it can be used to construct asymptotic tests of one sample goodness-of-fit, symmetry and goodness-of-fit.

Applied statisticians as well as econometricians often need information on the characterization of the probability density of residuals of a regression model. Particularly in a linear regression model, the normality of the errors is needed to ensure the efficiency of the least squares estimators in the Cramer-Rao lower bound sense. An extension of the normality test to the residual analysis can provide information about whether the distribution of the residuals of a linear regression model is normal or of another symmetric or an asymmetric distribution.

The goodness-of-fit test can be used in applications with cross-section data. An extension of this test with time dependent observations would make possible the comparison of time-series observations from two different populations.

References

1. I. A. Ahmad (1976). On Asymptotic Properties of an Estimate of a Functional of a Probability Density. *Scand. Actuarial J.* 176–181.
2. I. A. Ahmad (1980). Nonparametric Estimation of an Affinity Measure between Two Absolutely Continuous Distributions with Hypotheses Testing Applications. *Ann. Inst. Statist. Math.*, **32,** Part A, 223–240.
3. I. A. Ahmad and P. B. Cerrito (1989). Goodness of Fit Tests Based on the L_2-norm of Multivariate Density Functions. Submitted.
4. I. A. Ahmad and G. Van Belle (1974). Measuring Affinity of Distributions. *Reliability and Biometry, Statistical Analysis of Life Testing,* (eds. Proschan and R. J. Serfling), SIAM, Philadelphia, 651–668.
5. G. K. Bhattacharyya and G. G. Roussas (1969). Estimation of a Certain Functional of a Probability Density Function. *Skand. AktuarTidskr.*, **52,** 201–206.
6. P. M. Robinson (1991). Consistent Nonparametric Entropy-Based Testing. *The Review of Economic Studies*, **58,** 437–453.
7. E. F. Schuster (1974). On the Rate of Convergence of an Estimate of a Functional of a Probability Density. *Scand. Actuarial J.*, 103–107.