Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597248

Nonparametric testing of closeness between two unknown distribution functions

Qi Li [a]
[a] Department of Economics, University of Guelph, Guelph, Ontario, Canada

Online Publication Date: 01 January 1996

To cite this Article Li, Qi(1996)'Nonparametric testing of closeness between two unknown distribution functions',Econometric Reviews,15:3,261 — 274

To link to this Article: DOI: 10.1080/07474939608800355

URL: http://dx.doi.org/10.1080/07474939608800355

ECONOMETRIC REVIEWS, 15(3), 261-274 (1996)

NONPARAMETRIC TESTING OF CLOSENESS BETWEEN TWO UNKNOWN DISTRIBUTION FUNCTIONS

Qi Li
Department of Economics
University of Guelph
Guelph, Ontario, N1G 2W1
Canada

Key words and phases: Nonparametric test; unknown densities; kernel estimation; asymptotic normality, smoothing parameter.

JEL classification number: C12; C14.

ABSTRACT

Based on the kernel integrated square difference and applying a central limit theorem for degenerate U-statistic proposed by Hall (1984), this paper proposes a consistent nonparametric test of closeness between two unknown density functions under quite mild conditions. We only require the unknown density functions to be bounded and continuous. Monte Carlo simulations show that the proposed tests perform well for moderate sample sizes.

1 Introduction

Let $f(x)$ and $g(x)$ be two probability density functions (p.d.f.) with distribution functions F and G that are absolutely continuous with respect to the Lebesgue measure in $\mathcal{R}^p$. We are interested in testing the null hypothesis H_0: $P[f(x) = g(x)] = 1$ ($f(x) = g(x)$

almost everywhere (a.e.)) against the alternative H_1: $f(x) \neq g(x)$ on a set of positive measure. One widely accepted measure of global closeness between two functions $f(x)$ and $g(x)$ is the integrated square difference

$$\begin{aligned} I &= \int [f(x) - g(x)]^2 dx = \int [f^2(x) + g^2(x) - 2f(x)g(x)]dx \\ &= \int [f(x)dF(x) + g(x)dG(x) - 2g(x)dF(x)]. \end{aligned} \tag{1}$$

The measure I has the following properties that make it a proper candidate for testing our null hypothesis H_0:

$I \geq 0$ and the equality holds if and only if $f(x) = g(x)$ a.e.

Let $X_1, ..., X_{n_1}$ be independent observations with probability density function f; and $Y_1, ..., Y_{n_2}$ be independent observations with probability density function g; also X_i and Y_j are independent for $i \neq j$. We allow the possibility that X_i and Y_i are correlated. This may be the case if we have a panel of n ($n_1 = n_2 = n$) individuals over two periods. Of course X_i and Y_i can also be independent to each other, this can be the case if we have cross-sectional data from two different regions. We will use $\bar{f}(x, y)$ to denote the joint p.d.f. of (X_i, Y_i). As will be shown in the next section, our test statistic has the same asymptotic distribution whether X_i and Y_i are independent or not. This is because the terms that involve $\bar{f}(x, y)$ have orders smaller than the leading term of an U-statistic in the test statistic. There are many examples that economists are interested in testing whether two density functions are the same, for example, comparison of income distributions across two regions. Also if f indeed equals g, one can pool the two data sets (using the $n_1 + n_2$ observations) to obtain a more efficient density estimate.

With observations $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$, we can consistently estimate the unknown functions f and g by kernel estimators:

$$f_{n_1}(x) = \frac{1}{n_1 h^p} \sum_{i=1}^{n_1} K(\frac{X_i - x}{h}) \tag{2}$$

$$g_{n_2}(x) = \frac{1}{n_2 h^p} \sum_{i=1}^{n_2} K(\frac{Y_i - x}{h}) \tag{3}$$

where $K(\cdot)$ is the kernel function and $h = h_n$ is the smoothing parameter. Our proposed test will be based on (1) with $f(x)$ and $g(x)$ replaced by $f_{n_1}(x)$ and $g_{n_2}(x)$ respectively.

Recently Fan and Gencay (1993) derived a consistent nonparametric test for testing $f(x) = g(x)$ (a.e.) for the case $n_1 = n_2 = n$.[1]. Their test statistic is based on "an affinity measure" introduced by Ahmad and Van Belle (1974), viz.,

$$\sqrt{n}(\lambda_n - 1) = \sqrt{n}[\frac{2\int f_n(x)g_n(x)dx}{\int f_n^2(x)dx + \int g_n^2(x)dx} - 1].$$

However this test statistic is degenerate under the null hypothesis, i.e., $\sqrt{n}(\lambda_n - 1) = o_p(1)$ under H_0. To avoid the degeneracy, Fan and Gencay (1993) introduced a weighting scheme which assigns different weight to odd and even observations. In this paper we will use a normalization factor $nh^{p/2}$, under the conditions given in the paper, we have $nh^{p/2}/\sqrt{n} \to \infty$ as $n \to \infty$. Hence our test statistic is expected to be asymptotically more powerful than the test proposed by Fan and Gencay (1993). The fact that our test statistic has a convergence rate faster than the $\sqrt{n}$-rate (under H_0) should not be mis-interpreted as that the test has a convergent rate faster than a parametric test. This is because our test statistic is based on the integrated *square* difference of two nonparametric density estimates. If one were to construct a parametric test based on the integrated *square* difference of two parametric density estimates, the convergent rate (under H_0) would be n rather than $\sqrt{n}$. Other works related to ours are Mammen (1992), and Anderson, Hall and Titterington (1994). Both the test statistics proposed by Mammen (1992), and Anderson, Hall and Titterington (1994) involve some center terms. We will propose a test that does not have a center term and our Monte Carlo simulations (see section 3) show that our test compare favorably with the test statistics of Mammen (1992) and Anderson, Hall and Titterington (1994). Also our test does not involve any kernel convolution and hence it is computational simpler than the test proposed by Mammen (1992).

2 The Test Statistics and Their Asymptotic Distributions

2.1 The $n_1 = n_2$ case

In this subsection we will consider the case of equal sample sizes: $n_1 = n_2 = n$. The following assumptions are needed on the kernel function $K(\cdot)$ and the density functions

[1]See the references in Fan and Gencay (1993) for other related tests in the literature.

$f(\cdot)$ and $g(\cdot)$.

(A1) K is a bounded, nonnegative function on $\mathcal{R}^p$ satisfying $\int K(u)du = 1$, $\int u_i K(u)du = 0$ *and* $\int u_i u_j K(u)du = 2k\delta_{ij} < \infty$ *for each* i, *where* $k(>0)$ *does not depend on* i *(*u_i *is the ith component of* u*).*

(A2) f *and* g *are continuous and bounded in* $\mathcal{R}^p$.

Throughout the remaining part of the paper, the above assumptions **(A1)** and **(A2)** will be assumed to be true. Assumption **(A1)** requires a nonnegative second order kernel, this is purely due to the fact that second order kernel is the most popular and convenience in practice. For our result in this paper to hold, the kernel function can be of any order. Assumption **(A2)** seems minimum, we do not require f and g to be differentiable.

Our result relies on Hall's (1984) Theorem 1. We present it below as a lemma for ease of reference.

Lemma 2.1 *Assume* H_n *is symmetric,* $E[H_n(Z_1, Z_2)|z_1] = 0$ *a. e. and* $E[H_n^2(Z_1, Z_2)] < \infty$ *for each* n. *Define* $G_n(Z_1, Z_2) = E[H_n(Z_1, Z)H_n(z_2, z)]$. *If*

$\{E[G_n^2(Z_1, Z_2)] + n^{-1}E[H_n^4(Z_1, Z_2)]\}/\{E[H_n^2(Z_1, Z_2)]\}^2 \to 0$ *as* $n \to 0$, *then*

$U_n = \sum_{1\le i<j\le n} H_n(Z_i, Z_j)$ *is asymptotically normally distributed with zero mean and variance given by* $\frac{1}{2}n^2 E[H_n^2(Z_1, Z_2)]$.

Replacing $f(x)$, $F(x)$, $g(x)$ and $G(x)$ by $f_n(x)$, $F_n(x)$, $g_n(x)$ and $G_n(x)$ in (1), we obtain a feasible estimator of I.

$$\begin{aligned} I_n &= \int [f_n(x)dF_n(x) + g_n(x)dG_n(x) - 2g_n(x)dG_n(x)] \\ &= \frac{1}{n^2h^p}\sum_{i=1}^n\sum_{j=1}^n \{K(\frac{X_i - X_j}{h}) + K(\frac{Y_i - Y_j}{h}) - 2K(\frac{X_i - Y_j}{h})\} \\ &= \frac{1}{n^2h^p}\sum_{i=1}^n \{2K(0) - 2K(\frac{X_i - Y_i}{h})\} \\ &\quad + \frac{1}{n^2h^p}\sum_{i=1}^n \sum_{j\ne i, j=1}^n \{K_{ij}^x + K_{ij}^y - K_{ij}^{x,y} - K_{ij}^{y,x}\} \\ &\equiv I_{1n} + I_{2n}, \end{aligned} \tag{4}$$

where we used short hand notation $K_{ij}^x = K(\frac{X_i - X_j}{h})$, $K_{ij}^y = K(\frac{Y_i - Y_j}{h})$, $K_{ij}^{x,y} = K(\frac{X_i - Y_j}{h})$ and $K_{ij}^{y,x} = K(\frac{Y_i - X_j}{h})$. Also we have used $\int M(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^n M(X_i)$ and $\int M(x)dG_n(x) = \frac{1}{n}\sum_{i=1}^n M(Y_i)$, where $F_n(\cdot)$ and $G_n(\cdot)$ are the empirical distribution functions based on the sample data $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ respectively.

The asymptotic distribution of I_{1n} and I_{2n} are given in the next two lemmas.

Lemma 2.2 *Assuming that $h \to 0$ and $nh^p \to \infty$, then*

$I_{1n} = \frac{2K(0)}{nh^p} + O_p(n^{-1})$.

Proof: $E|\frac{1}{n^2h^p}\sum_{i=1}^n K(\frac{X_i-Y_i}{h})| = (nh^p)^{-1}E[K(\frac{X_1-Y_1}{h})] = n^{-1}\int\int \bar{f}(X_1, X_1+hu)K(u)dudX_1$ $= O(n^{-1})$. Hence $I_{1n} - \frac{2K(0)}{nh^p} = 2(n^2h^p)^{-1}\sum_{i=1}^n K(\frac{X_i-Y_i}{h}) = O_p(n^{-1})$.

Lemma 2.3 *Under H_0, also assuming that $h \to 0$ and $nh^p \to \infty$, then*

$nh^{p/2}I_{2n} \xrightarrow{d} N(0, \sigma_0^2)$, *where* $\sigma_0^2 = 2\{\int[f(x)+g(x)]^2dx\}[\int K^2(u)du]$.

Proof: $I_{2n} = \frac{2}{n^2h^p}U_n$, where $U_n = \sum_{1\le i<j\le n} H_n(Z_i, Z_j)$ $(Z_i \equiv (X_i, Y_i))$ with $H_n(Z_i, Z_j) = [K(\frac{X_i-X_j}{h}) + K(\frac{Y_i-Y_j}{h}) - K(\frac{X_i-Y_j}{h}) - K(\frac{X_j-Y_i}{h})]$. $H_n(Z_i, Z_j)$ is symmetric in Z_i and Z_j. Using $f(x) = g(x)$ a.e. under H_0, it is easy to see that $E[K(\frac{X_i-X_j}{h}) - K(\frac{X_i-Y_j}{h})|X_i] = 0$ and $E[K(\frac{Y_i-Y_j}{h}) - K(\frac{Y_i-X_j}{h})|Y_i] = 0$. Hence $E[H_n(Z_i, Z_j)|Z_i] = 0$. Define $G_n(Z_i, Z_j) = E[H_n(Z_i, Z)H_n(Z_j, Z)]$. It is easy to check that $E[H_n^2(Z_1, Z_2)] = O(h^p)$, $E[H_n^4(Z_1, Z_2)] = O(h^p)$ and $E[G_n^2(Z_1, Z_2)] = O(h^{3p})$ as $h \to 0$. Hence

$$\{E[G_n^2(Z_1, Z_2)] + n^{-1}E[H_n^4(Z_1, Z_2)]\}/\{E[H_n^2(Z_1, Z_2)]\}^2 = O(h^p + (nh^p)^{-1}) \to 0$$

as $n \to \infty$. It follows from lemma 2.1 that U_n is asymptotically normally distributed with zero mean and variance equal to $\frac{1}{2}n^2E[H_n^2(Z_1, Z_2)]$. In the appendix we show that $E[H_n^2(Z_1, Z_2)] = \frac{h^p}{2}\{\sigma_0^2 + o(1)\}$. Hence $nh^{p/2}I_{2n} \xrightarrow{d} N(0, \sigma_0^2)$.

Summarizing lemmas 2.2 and 2.3, we have

Theorem 2.4 *Under H_0, and assuming that $h \to 0$ and $nh^p \to \infty$, then we have*

(i) $J_{nc} = nh^{p/2}(I_n - c(n))/\hat{\sigma}_0 \xrightarrow{d} N(0,1)$,

(ii) $J_n = nh^{p/2}I_{2n}/\hat{\sigma}_0 \xrightarrow{d} N(0,1)$,

where $\hat{\sigma}_0^2 = \frac{2}{n^2h^p}\sum_{i=1}^n\sum_{j=1}^n[K_{ij}^x + K_{ij}^x + 2K_{ij}^{x,y}][\int K^2(u)du]$.

Proof: Given the results of lemmas 2.2 and 2.3, we only need to show that $\hat{\sigma}_0^2 = \sigma_0^2 + o_p(1)$. This follows from the fact that $\int(f_n(x) + g_n(x))^2dx = \int(f(x) + g(x))^2dx + o_p(1)$.

Remark (i) The J_{nc} test is basically the same as the test statistic as proposed by Mammen (1992) except that here we take a simpler approach and therefore we do not need to use convolution kernels to avoid the numeral integrations, while the test statistic proposed by Mammen (1992) requires the computation of four-fold convolution of the kernel function. The fact that J_{nc} has a center term is an undesirable property, because in finite sample applications, if I_{1n} is not sufficiently close to $c(n)$, the difference between the two will bring a finite sample (sometimes substantial) bias term to the J_{nc} test. In contrast, J_n does not have this problem. In section 3 we will compare the finite sample performance of J_{nc} and J_n using Monte Carlo simulations.

Remark (ii) The J_{nc} test is also related to the U_n test statistic proposed by Anderson, Hall and Titterington (1994). However a major difference is that in the U_n test of Anderson, Hall and Titterington (1994), the smoothing parameter h is fixed (h does not converge to zero as $n \to \infty$). Hence their test statistic does not have an asymptotic normal distribution.

Remark (iii) Our test statistic has the same asymptotic distribution whether X_i and Y_i are independent or not. Because the terms associated with the joint p.d.f $\bar{f}(X_i, Y_i)$ have orders smaller than $O_p((nh^{p/2})^{-1})$.

Remark (iv) Theorem 2.4 holds for a wide range of smoothing parameter choices. It allows the data to be under, optimally, or over smoothed as long as $h \to 0$, $nh^p \to \infty$. While in Fan and Gencay (1993), the data has to be under smoothed.

Remark (v) Our test is obviously a consistent test because under the alternative hypothesis, and using similar proof as in the lemma 2.3, one can easily show that , $I_{2n} \xrightarrow{p} I$ ($I > 0$ under H_1). Also note that $\hat{\sigma}_0^2 = \sigma_0^2 + o_p(1)$ under either H_0 or H_1. Hence $J_n = nh^{p/2} I_{2n}/\hat{\sigma}_0 = nh^{p/2} I/\sigma_0 + o_p(nh^{p/2})$ under H_1. So $Prob[J_n > B_n] \to 1$ as $n \to \infty$ for any nonstochastic sequence $B_n = o(nh^{p/2})$.

2.2 The Case of $n_1 \neq n_2$

The Monte Carlo simulation results of section 3.1 suggest that, for $n_1 = n_2 = n$, J_n dominates J_{nc}. Hence in this subsection we only consider a test that does not have a center term. When $n_1 \neq n_2$, similar to I_{2n}, we define I_{2,n_1,n_2} as:

$$\begin{aligned} I_{2,n_1,n_2} &= \frac{1}{h^p}\{\sum_{i=1}^{n_1}\sum_{j\neq i, j=1}^{n_1} \frac{1}{n_1(n_1-1)} K_{ij}^{x} + \sum_{i=1}^{n_2}\sum_{j\neq i, j=1}^{n_2} \frac{1}{n_2(n_2-1)} K_{ij}^{y} \\ &\quad - \sum_{i=1}^{n_1}\sum_{j\neq i, j=1}^{n_2} \frac{1}{n_1(n_2-1)} K_{ij}^{x,y} - \sum_{i=1}^{n_2}\sum_{j\neq i, j=1}^{n_1} \frac{1}{n_1(n_2-1)} K_{ij}^{y,x}\} \\ &\equiv \frac{1}{h^p}\sum_{i}\sum_{j\neq i}\{\frac{1}{n_1(n_1-1)} K_{ij}^{x} + \frac{1}{n_2(n_2-1)} K_{ij}^{y} - \frac{1}{n_1(n_2-1)} K_{ij}^{x,y} - \frac{1}{n_1(n_2-1)} K_{ij}^{y,x}\} \end{aligned}$$

where in the second equality $\sum_i = \sum_{i=1}^{n_1}$ if the summand has X_i, and $\sum_i = \sum_{i=1}^{n_2}$ if the summand has Y_i. Similarly $\sum_{j\neq i} = \sum_{j\neq i,j=1}^{n_1}$ or $\sum_{j\neq i} = \sum_{j\neq i,j=1}^{n_2}$ depending on whether the summand has X_j or Y_j. Then similar to Theorem 2.4 (ii), we have

Theorem 2.5 *Under the same conditions as in lemma 2.3, let $\lambda_n = n_1/n_2$, assume $\lambda_n \to \lambda$ as $n_1 \to \infty$, where $0 < \lambda < \infty$ is a constant. Then as $n_1 \to \infty$, we have*

$$J_{n_1,n_2} = n_1 h^{p/2} I_{2,n_1,n_2}/\sqrt{\hat{\sigma}_\lambda^2} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}_\lambda^2 = 2\sum_i\sum_j\{\frac{1}{n_1^2}K_{ij}^x + \frac{\lambda_n^2}{n_2^2}K_{ij}^y + \frac{\lambda_n}{n_1n_2}K_{ij}^{x,y} + \frac{\lambda_n}{n_1n_2}K_{ij}^{y,x}\}[\int K^2(u)du]$ is a consistent estimator of $\sigma_\lambda^2 = 2[\int(f(x)+\lambda g(x))^2dx][\int K^2(u)du]$. Note that when $n_1 = n_2 = n$, $\lambda_n = 1$ and the result of Theorem 2.5 reduces back to that of Theorem 2.4 (ii).

Proof: The proof is similar to that of lemma 2.3, hence here we only sketch the proof of $E[I_{2,n_1,n_2}] = 0$ and $var(n_1h^{p/2}I_{2,n_1,n_2}) = \sigma_\lambda^2 + o(1)$.

Writing $I_{2,n_1,n_2} = \frac{1}{h^p}\sum_i\sum_{j\neq i}P_{n_1,n_2}(Z_i,Z_j)$, where $P_{n_1,n_2}(Z_i,Z_j) = \frac{1}{n_1(n_1-1)}K_{ij}^x + \frac{1}{n_2(n_2-1)}K_{ij}^y - \frac{1}{n_1(n_2-1)}K_{ij}^{x,y} - \frac{1}{n_1(n_2-1)}K_{ij}^{y,x}$. Then it is easy to show that $E[P_{n_1,n_2}(Z_i,Z_j)|Z_i] = 0$ by the same argument as $E[H_n(Z_i,Z_j)|Z_i] = 0$ in the proof of lemma 2.3.

Next $E[(I_{2,n_1,n_2})^2] = \frac{2}{h^{2p}}\sum_i\sum_{j\neq i}E[P_{n_1,n_2}^2(Z_i,Z_j)]$. Following the same derivation as we did for computing $E[H_n^2(Z_1,Z_2)]$ (see the appendix), one can show that $E[P_{n_1,n_2}^2(Z_1,Z_2)] = E\{[\frac{1}{n_1(n_1-1)}K_{12}^x + \frac{1}{n_2(n_2-1)}K_{12}^y - \frac{1}{n_1(n_2-1)}K_{12}^{x,y} - \frac{1}{n_2(n_1-1)}K_{12}^{y,x}]^2\} + O(n_1^{-4}h^{2p}) = h^p[\int K^2(u)du][\frac{1}{n_1^2(n_1-1)^2}\int f^2(x)dx + \frac{1}{n_2^2(n_2-1)^2}\int g^2(x)dx + \frac{1}{n_1^2(n_2-1)^2}\int f(x)g(x)dx + \frac{1}{n_2^2(n_1-1)^2}\int f(x)g(x)dx] + O(n_1^{-4}h^{2p})$. Hence $E[(n_1h^{p/2}I_{2,n_1,n_2})^2] = n_1^2h^p\{2h^{-2p}h^p[\int K^2(u)du][\frac{1}{n_1(n_1-1)}\int f^2(x)dx + \frac{1}{n_2(n_2-1)}\int g^2(x)dx + \frac{1}{n_1(n_2-1)}\int f(x)g(x)dx + \frac{1}{n_2(n_1-1)}\int f(x)g(x)dx]\} + O(h^p) = 2[\int(f(x)+\lambda_n g(x))^2dx][\int K^2(u)du] + o(1) \to 2[\int(f(x)+\lambda g(x))^2dx][\int K^2(u)du]$.

3 Monte Carlo Results

This section reports some Monte Carlo simulation results for the proposed tests J_n and J_{nc}. In section 3.1 we concentrate on the case of equal sample sizes ($n_1 = n_2$) and compare the estimated size for the test statistics J_n and J_{nc} (for $p = 1$). For small sample applications, we expect J_n perform better than J_{nc} because of the possible small sample bias due to the difference between I_{1n} and $c(n)$ in the J_{nc} test. We will also report simulation results for the test statistic $T_n \equiv nh^{1/2}I_{2n}/\tilde{\sigma}_0$, where $\tilde{\sigma}_0^2 = \hat{\sigma}_0^2 -$

$4h\{2E_n[f^2(X)] - [E_n(f(X))]^2\}$, where $E_n[f(X)] = n^{-1}\sum_{i=1}^n f_n(X_i)$ and $E_n[f^2(X)] = n^{-1}\sum_{i=1}^n f_n^2(X_i)$. In the appendix, we show that $\tilde{\sigma}_0^2 = var(nh^{p/2}I_{2n}) + O_p(h^2)$, while $\hat{\sigma}_0^2 = var(nh^{p/2}I_{2n}) + O_p(h)$. Hence for small samples, we expect $\tilde{\sigma}_0^2$ to estimate $var(nh^{p/2}I_{2n})$ more accurate than $\hat{\sigma}_0^2$. In section 3.2, we will compare the finite sample performance of our test J_n, the test statistic U_n proposed by Anderson, Hall and Titterington (1994) and the test statistic $T_n^{(1)}$, where $T_n^{(1)}$ is the same as J_n except that h is a fixed value ($h = 1$ in the simulations).

3.1 Comparison of J_{nc}, J_n and T_n tests

In this subsection, we used equal sample size $n_1 = n_2 = n$ and chose $n = 50, 100, 200, 400$ and 800. We compare the estimated sizes for J_{nc}, J_n and T_n. The data generating process (DGP) is that both $f(x)$ and $g(x)$ are $\phi(x;0,1)$, where $\phi(x;\mu,\sigma^2)$ denotes the univariate standard normal probability density function corresponding to $X \sim N(\mu,\sigma^2)$. Also X_i and Y_i are independent. The number of replications are 5,000 for all cases considered. The smoothing parameter is chosen by $h = cn^{-1/5}$, $c = 0.8, 1, 1.2$ for all these three tests, the results are quite similar and hence we only report the case of $c = 1$ here to save space. Table 1 gives the estimated mean value, the estimated standard deviation and the estimated sizes for these three tests. Clearly we see that J_{nc} has a finite sample bias due to the difference between $c(n)$ and I_{1n}. Although the bias show some decrease as n gets large, J_{nc} still has a significant negative bias even for $n = 800$. As a consequence (also due to its over estimate the standard deviation of $(nh^{p/2}I_n)$), J_{nc} test is quite under sized. In contrast, both J_n and T_n have mean quite close to zero and both give much better estimated sizes than that of J_{nc}. For small samples, J_n slightly under estimate the size and T_n slightly over estimate the size. As n increases, the performances of both the J_n and the T_n tests improve. Note that as expected, $\tilde{\sigma}_0^2$ gives a more accurate estimate of $var(nh^{p/2}I_{2n})$ and hence the standard deviations of T_n are quite close to 1 for all cases considered.

3.2 Comparison of $T_n^{(1)}$, J_n and U_n tests

Anderson, Hall and Titterington (1994) proposed a test statistic U_n for testing H_0. Their test statistic U_n is similar to the test statistic J_{nc} but with smoothing parameter h being

Table 1: Empirical size for J_{nc}, J_n and T_n tests

	J_{nc}				J_n				T_n			
	mean	std.	5%	10%	mean	std.	5%	10%	mean	std.	5%	10%
$n = 50$	-.445	.682	.014	.025	.016	.712	.033	.061	.034	1.09	.085	.120
$n = 100$	-.414	.725	.018	.029	.010	.731	.037	.066	.017	1.03	.076	.109
$n = 200$	-.412	.742	.019	.032	.021	.782	.043	.069	.029	1.04	.071	.112
$n = 400$	-.398	.768	.022	.035	-.001	.810	.043	.071	-.002	1.03	.068	.104
$n = 800$	-.401	.804	.021	.036	-.002	.824	.044	.078	-.003	.998	.066	.105

fixed. As argued by Anderson, Hall and Titterington (1994), the advantage of using a fixed value of h is that the resulting test can detect local alternatives that approach to the null at the rate of $n^{-1/2}$, while the test statistics like J_n ($h \to 0$ as $n \to \infty$) can only detect local alternatives that approach to the null at the rate of $n^{-1/2}h^{-p/2}$. However disadvantages of using a fix value of h are (i) the asymptotic distribution of the test statistic is quite complicated, and the critical values of the test statistic has to be approximated using bootstrap method. Hence it is computationally more costly than the J_n test especially when the sample size is large; (ii) when the DGP is generated by a fixed alternative, due to its over smoothing procedure, the finite sample power of the U_n test is often less than that of J_n. Below we compare the small sample power performances of J_n, $T_n^{(1)}$ and U_n using both local and fixed alternatives. For the local alternative, we choose the same DGP as used by Anderson, Hall and Titterington (1994).

DGP1: $g(x) = \phi(x; 0, 1)$ and $f(x) = (1-p)\phi(x; 0, 1) + \bar{p}\phi(x; 0, \sigma^2)$ with $p = cn_2^{-1/2}$, $c = 2, 4, 6$ and $\sigma^2 = 2, 4$.

Table 2 reports the empirical powers of U_n ($h = 1$), J_n ($h = n^{-1/5}$) and $T_n^{(1)}$ ($h = 1$). Note that for a fixed value of h, $T_n^{(1)}$ does not have an asymptotically normal distribution. In fact $T_n^{(1)}$ has the same asymptotic distribution as that of U_n (see Anderson, Hall and Titterington (1994)). For all these three tests, the critical values are obtained using bootstrap procedure (resampling from the pooled sample). The number of replications is 1000, and in each replication, 200 bootstrap resamples were generated to give the 1%, 5% and 10% critical values for the corresponding tests. The estimated power reported in Table 2 are the percentage of rejecting the null hypothesis (in 1000 replications). We observe that $T_n^{(1)}$ is the most powerful test among the three. The fact that $T_n^{(1)}$

Table 2: **Empirical Power of J_n, $T_n^{(1)}$ and U_n**

	Size n		1%			5%			10%		
Test (σ^2)	n_1	n_2	$c=2$	$c=4$	$c=6$	$c=2$	$c=4$	$c=6$	$c=2$	$c=4$	$c=6$
J_n ($\sigma^2=2$)	20	60	0.007	0.013	0.035	0.055	0.084	0.135	0.105	0.157	0.240
$T_n^{(1)}$ ($\sigma^2=2$)	20	60	0.007	0.015	0.032	0.054	0.093	0.145	0.111	0.156	0.238
U_n ($\sigma^2=2$)	20	60	0.007	0.016	0.022	0.046	0.083	0.126	0.104	0.149	0.219
J_n ($\sigma^2=4$)	20	60	0.018	0.080	0.235	0.079	0.227	0.484	0.159	0.322	0.602
$T_n^{(1)}$ ($\sigma^2=4$)	20	60	0.016	0.084	0.270	0.097	0.262	0.552	0.168	0.397	0.686
U_n ($\sigma^2=4$)	20	60	0.012	0.059	0.228	0.078	0.233	0.501	0.151	0.367	0.647

is more powerful than J_n supports the use of a fixed value of h when the DGP of a local alternative is very close to the DGP of the null model (at the rate $n^{-1/2}$). The result that $T_n^{(1)}$ is more powerful than U_n is similar to the finding of Anderson, Hall and Titterington (1994), who found U_n (which subtract a center term from the original test) is more powerful than a test that does not subtract a center term. Here we observe that the $T_n^{(1)}$ test which does not include a center term, is more powerful than the U_n test that includes a center term.

Table 3 gives the empirical of the J_n ($h = n^{-1/5}$) and the $T_n^{(1)}$ ($h = 1$) tests under the following fixed alternatives: $g(x) = \phi(x;0,1)$ and

$DGP2:$ $f_2(x) = t(5)$,

$DGP3:$ $f_1(x) = \chi^2(4)$,

$DGP4:$ $f_3(x) = U(x;0,1)$,

$DGP5:$ $f_4(x) = \mu^{1/2}\phi(x;0,1) + (1-\mu)^{1/2}\chi^2(2)$ ($\mu = 0.2$, mixture distribution)

where $U(x;0,1)$ is the p.d.f. of a uniform distribution with mean zero and variance one. All the random variables are standardized to have zero mean and unit variance. We observe from Table 3 that while $T_n^{(1)}$ is more powerful than J_n for $DGP2$, it is less powerful than J_n for $DGP3,4,5$. Note that $t(5)$ is the most close to $\phi(x;0,1)$ among the four fixed alternatives. This result show that while the test with a fixed h ($h = 1$) is indeed more powerful than the test with $h \to 0$ (as $n \to \infty$) for alternatives that are close to the null, it is often less powerful when the alternatives are not close to the null. Thus the optimal choice of h seems to choose a fixed value of h when the alternative is sufficiently close to the null, and to choose a variable h ($h \to 0$ as $n \to \infty$) if the

Table 3: **Empirical Power of J_n and $T_n^{(1)}$ with fixed alternatives**

	$n_1 = n_2 = 50$						$n_1 = n_2 = 100$					
	J_n $(h = n^{-1/5})$			$T_n^{(1)}$ $(h = 1)$			J_n $(h = n^{-1/5})$			$T_n^{(1)}$ $(h = 1)$		
α-level	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$t(5)$	.013	.078	.137	.019	.081	.163	.017	.085	.153	.022	.113	.199
$\chi^2(4)$	.128	.310	.438	.040	.150	.246	.368	.622	.745	.107	.294	.407
Uniform	.058	.180	.290	.014	.072	.133	.133	.373	.524	.013	.055	.141
Mixture	.301	.582	.712	.080	.242	.386	.745	.930	.970	.213	.476	.628

Table 4: **Empirical Power of J_n and $T_n^{(1)}$ for DGP6, 7**

			$h = \sigma n_2^{-1/5}$			$h = \sigma$			$h = 1$		
DGP	n_1	n_2	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP6 $(\sigma^2 = 0.1)$	20	60	.235	.484	.602	.270	.552	.686	.065	.224	.376
DGP7 $(\sigma^2 = 100)$	20	60	.235	.484	.602	.270	.552	.686	.063	.181	.294

alternative is not close to the null. This also suggest an important research topic: how to design some automatic data driven procedure such that h can be chosen optimally in the sense that the power is maximized under both the local and the fixed alternatives. The answer for this question is beyond the scope of this paper.

Finally one caution to the applied researchers is that in practice the choice of h should depend on x_{sd}, the standard deviation of the data $\{X_i\}_{i=1}^n$. All the DGP's we used above have standard deviation 1. In practice the value of x_{sd} should take into account so that the testing result should be invariant to the scale of measuring x. For example for the univariate case, one can use $h = x_{sd}$ or $h = x_{sd}n^{-1/5}$ (rather than $h = 1$ and $h = n^{-1/5}$) for a 'fixed' value of h or a variable h. This is because when x_{sd} is much larger (smaller) than 1, $h = 1$ will under (over) smooth the data too much, and this will lead to poor power performance of the test. As an illustration, Table 4 gives the estimated power for two such cases:

*DPG*6: $g(x) = \phi(x; 0, \sigma^2)$ and $f(x) = (1-p)\phi(x; 0, \sigma^2) + p\phi(x; 0, 4\sigma^2)$ with $p = 6n_2^{-1/2}$ and $\sigma^2 = 0.1$.

*DGP*7: same as the DGP6 but with $\sigma^2 = 100$.

We see that while the results for $h = \sigma$ and $h = \sigma n_2^{-1/5}$ are invariant to σ (they are the same to the corresponding results in Table 2). The choice of $h = 1$ gives poor estimated power due to its over (under) smooth for DGP6 (DGP7).

4 Conclusion

Under mild conditions that f and g are bounded and continuous in $\mathcal{R}^p$, we proposed a simple test statistic for testing the closeness between two unknown density functions. Our test statistics have a simple asymptotic distribution for a wide range of smoothing parameter choices and it is interesting to observe that the assumptions on the unknown density functions f and g considered in this paper are weaker than the corresponding assumption on f when testing whether f belongs to a known parametric family (Fan (1994)). Our Monte Carlo results show that the proposed tests performs well for sample size $n \geq 50$ (when $n_1 = n_2 = n$). An important question that has not been answered by this paper is: how to design some automatic data driven procedure such that h can be chosen optimally in the sense that the power is maximized under both the local and the fixed alternatives. The answer for this question is left for future research.

ACKNOWLEDGMENTS

I would like to thank Yanqin Fan, Aman Ullah and a referee for very helpful comments that greatly improve the paper, and Boping Yan for his excellent research assistance. This research is supported by the Social Sciences and Humanities Research Council of Canada.

References

[1] Ahmad, I.A. and Van Belle, G. (1974) "Measuring affinity of distributions," *Reliability and Biometry, Statistical Analysis of Life Testing* (eds. Proschan and R.J. Serfling). SIAM, Philadelphia, 651-668.

[2] Anderson, N.H., Hall, P. and Titterington, D.M. (1994) "Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates," *Journal of Multivariate Analysis* 50, 41-54.

[3] Fan, Y. (1994) "Testing goodness-of-fit of a parametric density function by kernel method" *Econometric Theory* 10, 316-356.

[4] Fan, Y. and Gencay, R. (1993) "Hypotheses testing based on modified nonparametric estimation of an affinity measure between two distributions," *Journal of Nonparametric Statistics* 4, 389-403.

[5] Hall, P. (1984) "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Annals of Statistics* 14, 1-16.

[6] Mammen, E. (1992) When does bootstrap work: Asymptotic results and Simulations. Lecture notes in Statistics. 77. Springer, Berlin.

Appendix

(i) Proof of $E[H_n^2(Z_1, Z_2)] = \frac{h^p}{2}\{\sigma_0^2 + o_p(1)\}$.

Use the short hand notation $K_{12}^x = K(\frac{X_i - X_j}{h})$, $K_{12}^y = K(\frac{Y_i - Y_j}{h})$, $K_{12}^{x,y} = K(\frac{X_i - Y_j}{h})$ and $K_{12}^{y,x} = K(\frac{Y_i - X_j}{h})$, it is easy to see that

$$\begin{aligned} E(H_n^2(Z_1, Z_2)) &= E\{\,[K_{12}^x + K_{12}^y - K_{12}^{x,y} - K_{12}^{y,x}]^2\} \\ &= E[(K_{12}^x)^2 + (K_{12}^y)^2 + (K_{12}^{x,y})^2 + (K_{12}^{y,x})^2] + O(h^{2p}). \end{aligned}$$

$E[(K_{12}^x)^2] = h^p \int\int f(x_1) f(x_1 + hu) K^2(u) du dx_1 = h^p\{[\int f^2(x)dx][\int K^2(u)du] + o(1)\}$.
Similar one can easily show that $E[(K_{12}^y)^2] = h^p\{[\int g^2(x)dx][\int K^2(u)du] + o(1)\}$, $E[(K_{12}^{x,y})^2]$ $= E[(K_{12}^{y,x})^2] = h^p\{[\int f(x)g(x)dx][\int K^2(u)du] + o(1)\}$.

Hence $E[H_n^2(Z_1, Z_2)] = h^p\{\int[f(x) + g(x)]^2 dx\}[\int K^2(u)du] + o(1)\} = \frac{h^p}{2}\{\sigma_0^2 + o(1)\}$.

(ii) Proof of $\tilde{\sigma}^2 = var(nh^{p/2} I_{2n}) + O_p(h^2)$.

We need to assume X_i and Y_i are independent of each other, i.e., $\bar{f}(x, y) = f(x)g(y)$, and that both $f(x)$ and $g(y)$ are twice differentiable and their derivatives are bounded by integrable functions. We will only prove the case of $p = 1$. Write

$$\begin{aligned} E(H_n^2(z_1, z_2)) &= E\{\,[K_{12}^x + K_{12}^y - K_{12}^{x,y} - K_{12}^{y,x}]^2\} \\ &= E[K_{12}^x - K_{12}^{x,y}]^2 + E[K_{12}^y - K_{12}^{y,x}]^2 + 2E[K_{12}^x - K_{12}^{x,y}][K_{12}^y - K_{12}^{y,x}] \\ &\equiv S_{1n} + S_{2n} + 2S_{3n}. \end{aligned}$$

We will first compute S_{1n}. $S_{1n} = E[(K_{12}^x)^2] + E[(K_{12}^{x,y})^2] - 2E[K_{12}^x K_{12}^{x,y}] \equiv S_{1n1} + S_{1n2} - 2S_{1n3}$.

$S_{1n1} = h^p\{[\int f^2(x)dx][\int K^2(u)du] + O(h^2)\}$ by the same proof of (i). The remaining term $O(h^2)$ follows from the fact $\int K^2(u)du = 0$ and $f(x)$ is twice differentiable.

Similarly $S_{1n2} = h\{[\int f(x)g(x)dx][\int K^2(u)du] + O(h^2)\}$, and $S_{1n3} = \int\int\int f(x_1)f(x_2)g(y_2)K_{12}^{x}K_{12}^{x,y}dx_1dx_2dy_2 = h^2\int\int\int f(x_1)f(x_1+hu)g(x_1+hv)K(u)K(v)dx_1dudv = h^2\{\int f^2(x)g(x)dx + O(h)\}$. Hence $S_{1n} = h\{[\int(f^2(x) + f(x)g(x))dx][\int K^2(u)du] - 2h\int f^2(x)g(x)dx + O(h^2)\}$.

Interchange x with y in S_{1n}, we get $S_{2n} = h\{[\int(g^2(x) + f(x)g(x))dx][\int K^2(u)du] - 2h\int f(x)g^2(x)dx + O(h^2)\}$.

Finally we consider S_{3n}. $S_{3n} = E\{[K_{12}^{x} - K_{12}^{x,y}][K_{12}^{y} - K_{12}^{y,x}]\} = \{E[K_{12}^{x}K_{12}^{y}] - E[K_{12}^{x}K_{12}^{y,x}] - E[K_{12}^{x,y}K_{12}^{y}] + E[K_{12}^{x,y}K_{12}^{y,x}]\} \equiv S_{3n1} - S_{3n2} - S_{3n3} + S_{3n4}$.

$S_{3n1} = E[K_{12}^{x}K_{12}^{y}] = \int\int\int\int f(x_1)g(y_1)f(x_2)g(y_2)K_{12}^{x}K_{12}^{y}dx_1dx_2dy_1dy_2 = h^2\int\int\int\int f(x_1)\cdot g(y_1)f(x_1 + hu)g(y_1 + hv)K(u)K(v)dx_1dudy_1dv = h^2\{\int\int f^2(x)g^2(y)dxdy + O(h)\}$.

$S_{3n2} = E[K_{12}^{x}K_{12}^{y,x}] = \int\int\int f(x_1)g(y_1)f(x_2)K_{12}^{x}K_{12}^{y,x}dx_1dy_1dx_2$

$= h^2\{\int\int\int f(x_2+hu)g(x_2+hv)f(x_2)K(u)K(v)dudvdx_2 = h^2\{\int f^2(x)g(x)dx+O(h)\}$.

Interchanging x and y in S_{3n2}, we get $S_{3n3} = h^2\{\int f(x)g^2(x)dx + O(h)\}$. Also similar to the derivation of S_{3n1}, on can show that $S_{3n4} = h^2\{[\int f(x)g(x)dx]^2 + O(h)\}$.

Summarizing above, we have shown that $E[H_n^2(Z_1, Z_2)] = S_{1n} + S_{2n} + 2S_{3n} = h\{\int(f(x)+g(x))^2dx][\int K^2(u)du] - 4h[\int(f^2(x)g(x)+g(x)f^2(x))dx] + 2h\{[\int f^2(x)dx][\int g^2(x)dx] + [\int f(x)g(x)dx]^2\} + O(h^2)\} = h\{\sigma_0^2 - 4h\{2E[f^2(X)] - [E(f(X))]^2\} + O(h^2)\}$, where the last equality used the fact that $f = g$ (a.e.) under H_0. Hence $\{\sigma_0^2 - 4h(2E[f^2(X)] - [E(f(X))]^2)\} = h^{-1}E[H_n(Z_1, Z_2)] + O(h^2)$. This result together with the facts that $\hat{\sigma}_0^2 = \sigma_0^2 + O_p(h)$ and $E_n[f(X)] = E[f(X)] + O_p(h)$ and $E_n[f^2(X)] = E[f^2(X)] + O_p(h)$ implies that $\tilde{\sigma}_0^2 = var(nh^{p/2}I_{2n}) + O_p(h^2)$.