T E C H N I C A L

R E P O R T

0429

EMPIRICAL LIKELIHOOD TESTS FOR TWO-SAMPLE PROBLEMS VIA NONPARAMETRIC DENSITY ESTIMATION

CAO, R. and I. VAN KEILEGOM

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Empirical likelihood tests for two-sample problems via nonparametric density estimation

Ricardo Cao
Universidade da Coruña

Ingrid Van Keilegom
Université catholique de Louvain

October 22, 2004

Abstract

In this paper we study the problem of testing whether two populations have the same law, by comparing kernel estimators of the two density functions. The proposed test statistic is based on a local empirical likelihood approach. We obtain the asymptotic distribution of the test statistic and propose a bootstrap approximation to calibrate the test. A simulation study is carried out, in which the proposed method is compared with two competitors, and a procedure to select the bandwidth parameter is studied. The proposed test can be extended to more than two samples and to multivariate distributions.

Key Words: Bandwidth selection; Bootstrap; Comparison of two populations; Kernel method; Local empirical likelihood.

1 Motivation and background

A very common problem in statistics is testing whether two populations have the same distribution. It is typically assumed that we are given a random sample from each of these populations. This testing problem, often known as the two-sample problem, has been dealt with using many different approaches. Let us briefly mention the classical Mann-Whitney test, Smirnov's edf test, and Cramér-von Mises type tests. There exists a vast literature dealing with these type of tests, including multivariate settings, during the past decades. The papers by Kim and Foutz (1987), Einmahl and Khmaladze (2001) and Henze and Nikitin (2003) are a small sample of it.

If the two distributions are assumed to be absolutely continuous it is also natural to base a test on some functional distance between nonparametric estimates of the two densities. Assume that we are given two random samples: $X_1, \ldots, X_n$, with common distribution F_1 and density f_1, and $Y_1, \ldots, Y_m$, with distribution F_2 and density f_2. We are interested in testing the hypothesis H_0: $f_1 = f_2$, versus the alternative that the two densities are different. Anderson, Hall and Titterington (1994) have studied the behaviour of an L_2-type test based on kernel density estimates that addresses the two sample problem, while Louani (2000) uses an L_1 and L_∞ approach.

More recently, a number of papers have dealt with the two-sample problem by using an empirical likelihood approach. The method of empirical likelihood can be viewed as a nonparametric counterpart of the classical parametric likelihood theory, and as such, it enjoyes the same advantages as the parametric method. It creates e.g. confidence regions that respect the range of the parameter space, that are invariant under transformations and whose shape is determinded by the data, it avoids in many situations the estimation of the variance by studentizing internally, and is often Bartlett correctable. For a comprehensive study of this method, see Owen (2001).

Some recent papers that deal with the two-sample problem via an empirical likelihood approach are Qin (1994), Jing (1995) and Zhang (2000) (comparison of the mean of two populations), McKeague and Zhao (1996) (confidence band for the ratio of two survival functions), Einmahl and McKeague (1999) (confidence tubes for QQ-plots), Einmahl and

McKeague (2003) (comparison of the distribution of two populations) and Claeskens, Jing, Peng and Zhou (2003) (point and interval estimation for ROC curves and PP-plots).

None of these is concerned with the density view of the two-sample problem. However, for the one-sample problem, there are some papers related to empirical likelihood for density estimation, like, for instance, Hall and Owen (1993) and Chen (1996). In this paper we propose a local empirical likelihood test for the two-sample problem that is based on the pointwise comparison of kernel estimators of the density of the two populations. An omnibus test statistic is then obtained by integrating the local log-likelihood process over an appropriate interval. We propose a bootstrap procedure to approximate the distribution of the test statistic. The testing procedure can be extended in a straightforward way to the case of more than two populations and to the comparison of multivariate distributions (but the asymptotic properties are more cumbersome to obtain and are not considered in this paper).

The rest of the paper is organized as follows. The empirical likelihood density based test is constructed in Section 2, whereas the main result is stated in Section 3. Some simulations, to illustrate the theory, are included in Section 4. Finally, Section 5 contains the proofs.

2 Test statistic

The null hypothesis $H_0 : f_1 = f_2$, considered above, is equivalent to $f_1(x) = \frac{nf_1(x)+mf_2(x)}{n+m} = f_2(x)$ for all x, and, when h tends to zero, also to the condition

$$\begin{aligned} & \int K_h(x-y)\, dF_1(y) \\ = & \frac{n}{n+m}\int K_h(x-y)\, dF_1(y) + \frac{m}{n+m}\int K_h(x-y)\, dF_2(y) \\ = & \int K_h(x-y)\, dF_2(y), \end{aligned} \tag{1}$$

for all x, where K is a kernel function (typically a density function), h is the bandwidth, or smoothing parameter, and $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$. Instead of considering condition (1) for all x, we replace the middle expression in (1) by an unbiased estimator:

$$\int K_h(x-y)\,dF_1(y) = \hat{f}_{12}(x) = \int K_h(x-y)\,dF_2(y), \tag{2}$$

where

$$\hat{f}_{12}(x) = \frac{1}{n+m}\left[\sum_{i=1}^{n} K_h(x-X_i) + \sum_{j=1}^{m} K_h(x-Y_i)\right] = \frac{n\hat{f}_1(x) + m\hat{f}_2(x)}{n+m},$$

with

$$\begin{aligned} \hat{f}_1(x) &= \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i), \\ \hat{f}_2(x) &= \frac{1}{m}\sum_{j=1}^{m} K_h(x-Y_j). \end{aligned}$$

Now, the empirical likelihood pertaining to condition (2) may be written as

$$EL(x) = \frac{\sup\left\{L\left(\tilde{F}_1, \tilde{F}_2\right) : \tilde{F}_1 \text{ and } \tilde{F}_2 \text{ are cdf's and (2) holds}\right\}}{\sup\left\{L\left(\tilde{F}_1, \tilde{F}_2\right) : \tilde{F}_1 \text{ and } \tilde{F}_2 \text{ are cdf's}\right\}}, \tag{3}$$

where

$$\begin{aligned} L\left(\tilde{F}_1, \tilde{F}_2\right) &= L_1\left(\tilde{F}_1\right) \cdot L_2\left(\tilde{F}_2\right), \\ L_1\left(\tilde{F}_1\right) &= \prod_{i=1}^{n} p_i, \\ L_2\left(\tilde{F}_2\right) &= \prod_{j=1}^{m} q_j, \end{aligned}$$

with

$$\begin{aligned} p_i &= \tilde{F}_1(X_i) - \tilde{F}_1\left(X_i^-\right), \text{ for } i = 1, \ldots, n, \\ q_j &= \tilde{F}_2(Y_j) - \tilde{F}_2\left(Y_j^-\right), \text{ for } j = 1, \ldots, m. \end{aligned}$$

Standard Lagrange multiplier methods can be used to obtain the suprema in the numerator and denominator of (3), which leads to

$$\begin{aligned} -2\log EL(x) &= 2\sum_{i=1}^{n} \log\left[1 + \eta_1\left(K_{1,i} - \hat{f}_{12}(x)\right)\right] \\ &\quad +2\sum_{j=1}^{m} \log\left[1 + \eta_2\left(K_{2,j} - \hat{f}_{12}(x)\right)\right], \end{aligned} \tag{4}$$

where

$$\sum_{i=1}^{n} \frac{1}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)} = n, \tag{5}$$

$$\sum_{j=1}^{m} \frac{1}{1+\eta_2\left(K_{2,j}-\hat{f}_{12}(x)\right)} = m, \tag{6}$$

with

$$K_{1,i} = K_{1,i}(x) = K_h(x-X_i), \text{ for } i=1,2,\ldots,n, \tag{7}$$

$$K_{2,j} = K_{2,j}(x) = K_h(x-Y_j), \text{ for } j=1,2,\ldots,m. \tag{8}$$

For fixed kernel K and bandwidth h, equations (5) and (6) can be solved numerically in η_1 and η_2 to find the value of $-2\log EL(x)$ in (4). Now, a global empirical likelihood statistic can be defined as

$$\begin{aligned} T_{n,m} &= -2\int_C \log EL(x)\, d\hat{F}_{12}(x) \\ &= -\frac{2}{n+m}\left[\sum_{i=1}^{n}\log EL(X_i)\,\mathbf{1}_C(X_i)+\sum_{j=1}^{m}\log EL(Y_j)\,\mathbf{1}_C(Y_j)\right], \end{aligned}$$

for some $C \subset \mathbb{R}$, where

$$\hat{F}_{12}(x) = \frac{1}{n+m}\big[\sum_{i=1}^{n} I(X_i \le x) + \sum_{j=1}^{m} I(Y_j \le x)\big],$$

and $\mathbf{1}_C(x)$ equals 1 when x belongs to C and 0 otherwise. The integration is restricted to the set C (which satisfies condition A2 below) for technical reasons. Note that in Hall and Owen (1993) and Chen et al. (1996) a similar setup is used, whereas Einmahl and McKeague (2003) allow for integration over the full real line.

The hypothesis H_0 will be rejected for large values of $T_{n,m}$. In order to determine how large a value should be to reject the null hypothesis, the limit distribution of the test statistic, under H_0, will be found in the next section.

3 Main result

Let us consider the following assumptions :

A1. As the sample sizes tend to infinity, $\frac{n}{m} \to \kappa^2 \in (0,\infty)$, $h \to 0$, $nh^4(\log n)^{-2}(\log\log n)^{-1} \to \infty$, $n^{1-\alpha}h^{5/2} \to \infty$ and $n^{2\alpha}h \to 0$ for some $\alpha > 0$.

A2. The common density (under H_0), f, is bounded in $\mathbb{R}$ and bounded away from zero in C, which is a compact interval.

A3. The kernel, K, is a Lipschitz continuous symmetric density function of bounded variation with compact support.

A4. The common density f is twice differentiable, with f'' bounded.

We introduce the notations $R(g) = \int g^2$ and $R_C(g) = \int_C g^2$ for any square integrable function g, and

$$\mu_{n,m}(C) = \int_C d\hat{F}_{12}(x) = \frac{1}{n+m}\left(\sum_{i=1}^{n} \mathbf{1}_C(X_i) + \sum_{j=1}^{m} \mathbf{1}_C(Y_j)\right).$$

Let us start with a preliminary lemma.

Lemma 1 *Assume conditions A1-A3 and hypothesis H_0. Then, for any x in C,*

$$\begin{aligned} -2\log EL(x) &= \frac{nmh}{(n+m)R(K)f(x)}\left(\hat{f}_1(x) - \hat{f}_2(x)\right)^2 \\ &\quad + O_P\left(h + n^{-1/2}h^{-1/2} + n^{-1}h^{-2}\right). \end{aligned}$$

and,

$$\begin{aligned} -2\int_C \log EL(x)\, dF(x) &= \frac{nmh}{(n+m)R(K)}\int_C \left(\hat{f}_1(x) - \hat{f}_2(x)\right)^2 dx \\ &\quad + O_P\left(n^{\alpha}\left(h + n^{-1/2}h^{-1/2} + n^{-1}h^{-2}\right)\right) \end{aligned}$$

(with $\alpha > 0$ defined in condition A1).

We now state our main result.

Theorem 2 *Assume conditions A1-A4 and hypothesis H_0. Then,*

$$\frac{R(K)}{\sqrt{2hR_C(f)\,R(K*K)}}\left[T_{n,m} - \mu_{n,m}(C)\right] \xrightarrow{d} N(0,1),$$

where $K*K$ denotes the convolution of K with itself.

As an alternative to the normal approximation, we now explain how the distribution of the test statistic $T_{n,m}$ can be approximated using a smoothed bootstrap procedure. The bootstrap approximation will be used to determine the critical value of the test, whereas the shape of the rejection region is completely determined by the empirical likelihood method.

Let g be a second bandwidth, possibly different from h, and let $\hat{f}_{12,g}$ be the kernel density estimator of the pooled data :

$$\hat{f}_{12,g}(x) = \frac{1}{n+m}\left[\sum_{i=1}^{n} K_g\left(x - X_i\right) + \sum_{j=1}^{m} K_g\left(x - Y_i\right)\right].$$

Define resamples $\{X_1^*, \ldots, X_n^*\}$ and $\{Y_1^*, \ldots, Y_m^*\}$, drawn independently from $\hat{f}_{12,g}$. Using these bootstrap data, calculate the bootstrap statistic $T_{n,m}^*$ (repeat the same procedure as for the construction of $T_{n,m}$, but use the resamples instead of the original samples), and let c^* be determined by

$$P(T_{n,m}^* \leq c^* | X_1, \ldots, X_n, Y_1, \ldots, Y_m) = 1 - \alpha,$$

where α represents here the level of the test. The null hypothesis H_0 is now rejected if $T_{n,m} \geq c^*$.

4 Simulations

In this section we study the practical behavior of the proposed testing procedure by means of some simulations. We also compare our method with the test of Anderson, Hall and Titterington (1994) (based on the L_2-distance between $\hat{f}_1$ and $\hat{f}_2$) and the test of Einmahl and McKeague (2003), which is an empirical likelihood test based on the comparison of two distribution functions (whereas the proposed test is based on the comparison of two density functions). Throughout the simulation study the density f_1 is the standard normal density. For the density f_2 we consider the following models :

Model 0 : $Y \sim N(0, 1)$ (null model)

Model 1 : $Y \sim N(a, 1)$, for $a = 0.3$ and 0.6

Model 2 : $Y \sim N(0, (1 + a)^2)$, for $a = 0.5$ and 0.75

Model 3 : $Y \sim (\chi^2_{1/a} - 1/a)/\sqrt{2/a}$, for $a = 0.3$ and 0.5

Model 4 : $Y \sim t_{2/a}\sqrt{1 - a}$, for $a = 0.6$ and 0.8

Model 5 : $Y \sim (1 - a)N(0, 1) + aN(2, 1)$, for $a = 0.2$ and 0.4

Model 6 : $f_2(y) = (1 - \sin(\pi y/\sqrt{2})a)f_1(y)$, for $a = 0.5$ and 1

Graphs of these densities are shown in Figure 1. Note that for each of the models 1 to 6, the parameter a represents the departure from the null distribution, in the sense that the larger the value of a, the larger the deviation from the null model (also note that when a tends to zero, the null model is found back). In models 1 to 4 the parameter a controls the j-th moment of the density f_2 ($j = 1, 2, 3, 4$). The densities in these models are normalized in such a way that all moments up to order $j-1$ coincide with the corresponding moments of f_1.

The results are obtained for several values of n, m and h. The interval C is chosen as the union of C_1 and C_2, where $C_j = [F_j^{-1}(0.005), F_j^{-1}(0.995)]$ ($j = 1, 2$). Note that this choice of C satisfies the conditions of Section 3. The results, shown in Table 1 below, are obtained from 500 simulation runs, and for each sample 500 bootstrap samples are generated. The bandwidth g used in the bootstrap procedure equals $(n + m)^{-0.15}$, and the level of significance is 0.05.

Table 1 shows that the level is well respected by all three methods and for all bandwidths. For the different alternative models and for most sample sizes and bandwidths chosen, the results seem to suggest that the proposed test outperforms the two other tests for models 3, 4 and 6, that for models 1 and 5, the test of Einmahl and McKeague is the best, and for model 2 there is no clear winner. In summary the test by Einmahl and McKeague is very well suited for situations where one of the distributions stochastically dominates the other, while the new test is the best when one of the densities presents some local deviation with respect to the other.

Additional simulations (not shown here) indicate that the choice of the interval C does not have much impact on the level of the test, but it does influence the power. In practice we recommend to choose C as large as possible. The pilot bandwidth g used to generate resamples, seems to have only a minor effect on the rejection probabilities. The choice of

h is however important for the power, as is explained in the next remark.

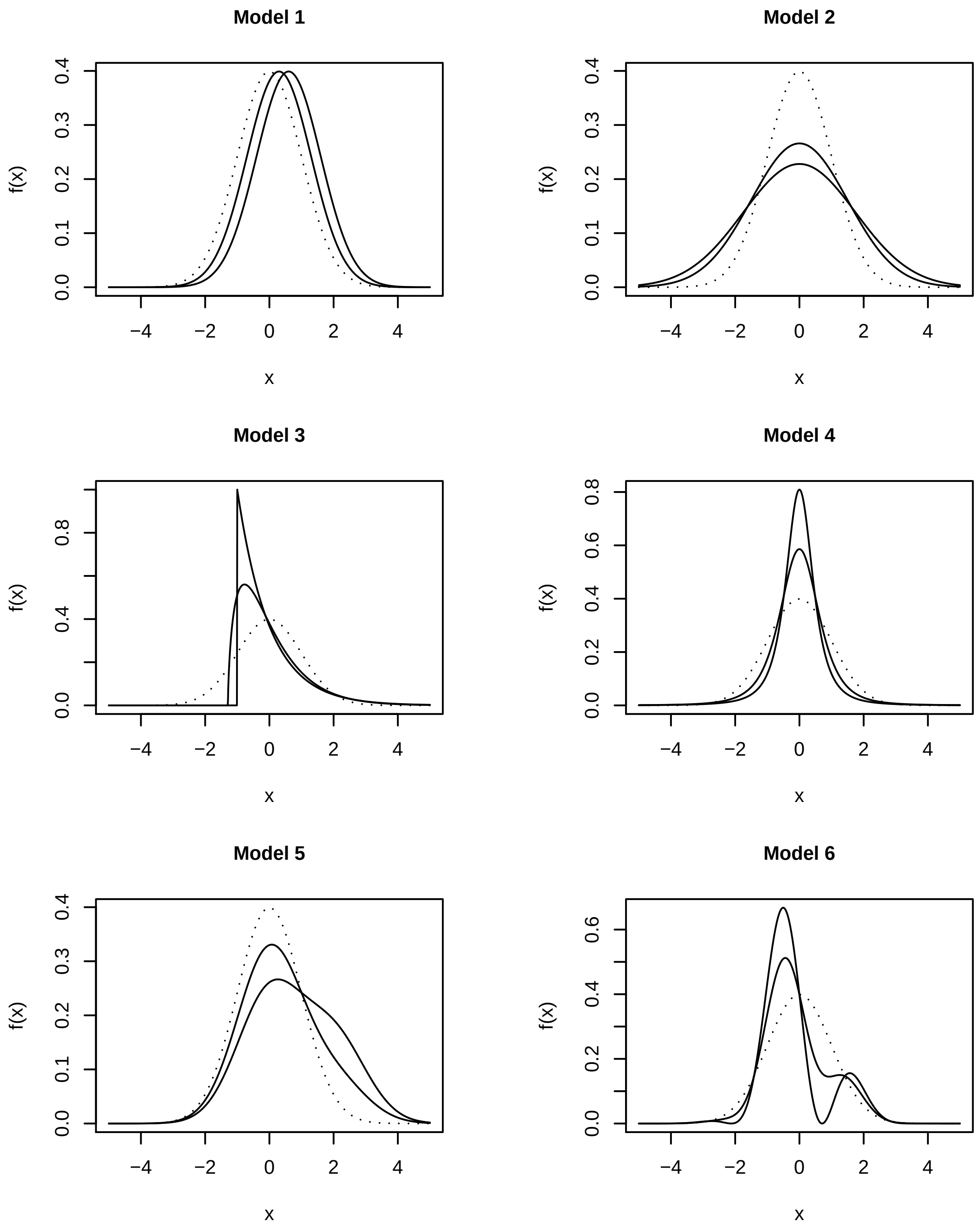

Figure 1: *Graph of the densities under models 1 to 6. The dotted curve represents the density under the null model, the full curves represent the densities under the alternative models.*

Mod.	a	h	$n = 50, m = 50$			$n = 50, m = 100$			$n = 100, m = 100$		
			New	AHT	EM	New	AHT	EM	New	AHT	EM
0	0	1	.056	.054	.050	.054	.038	.050	.044	.040	.050
		1.5	.056	.056		.042	.040		.046	.052	
1	0.3	1	.188	.186	.314	.250	.256	.402	.294	.328	.568
		1.5	.234	.252		.298	.318		.390	.426	
	0.6	1	.594	.640	.806	.710	.770	.918	.880	.928	.990
		1.5	.680	.726		.810	.842		.942	.958	
2	0.5	1	.366	.358	.242	.496	.464	.416	.702	.678	.650
		1.5	.450	.428		.606	.532		.762	.768	
	0.75	1	.598	.664	.630	.700	.788	.790	.880	.940	.962
		1.5	.708	.744		.794	.872		.922	.982	
3	0.3	1	.424	.356	.174	.572	.560	.322	.780	.732	.534
		1.5	.344	.288		.472	.430		.670	.610	
	0.5	1	.658	.632	.448	.812	.792	.622	.924	.918	.884
		1.5	.542	.464		.684	.666		.874	.826	
4	0.6	1	.290	.278	.100	.326	.332	.154	.472	.462	.230
		1.5	.246	.202		.308	.288		.432	.358	
	0.8	1	.738	.732	.496	.868	.870	.574	.960	.950	.806
		1.5	.702	.624		.846	.828		.946	.924	
5	0.2	1	.216	.198	.318	.284	.248	.526	.392	.392	.700
		1.5	.290	.258		.324	.352		.490	.518	
	0.4	1	.636	.762	.918	.726	.886	.986	.878	.984	.998
		1.5	.758	.864		.790	.940		.912	.992	
6	0.5	1	.298	.264	.128	.390	.364	.196	.546	.516	.282
		1.5	.226	.200		.294	.272		.416	.388	
	1	1	.856	.846	.464	.936	.944	.712	.992	.990	.908
		1.5	.750	.706		.860	.840		.978	.960	

Table 1: *Rejection probabilities for the new method, the method of Anderson, Hall, and Titterington (1994)(AHT) and the one of Einmahl and McKeague (2003) (EM).*

Remark (choice of h). The above simulations suggest that the bandwidth h has only a minor effect on the theoretical level of the test, but it does influence the power. We therefore propose to select h by means of the following procedure, which estimates the bandwidth that maximizes the power by means of a double-bootstrap procedure :

1. Let $\{h_1, \ldots, h_k\}$ be a grid of h-values among which to select the optimal one.
2. Generate B resamples $X_{i1}^*, \ldots, X_{in}^*$ ($i = 1, \ldots, B$) from $\hat{f}_{1g}(\cdot) = n^{-1} \sum_{j=1}^n K_g(\cdot - X_j)$, and similarly, generate B resamples $Y_{i1}^*, \ldots, Y_{im}^*$ ($i = 1, \ldots, B$) from $\hat{f}_{2g}(\cdot)$ (note that since we aim at maximizing the power (and not the level), we generate resamples from the two samples separately).
3. For each $i = 1, \ldots, B$ and each $j = 1, \ldots, k$:
   (a) Calculate $T_{n,m}^*(i, j)$ (i.e. test statistic based on $X_{i1}^*, \ldots, X_{in}^*, Y_{i1}^*, \ldots, Y_{im}^*$ and on bandwidth h_j)
   (b) Determine critical value $c^{**}(i, j)$ by generating (second level) bootstrap samples under H_0 (see Section 3 for a detailed description of the approximation of this critical value).
4. For each $j = 1, \ldots, k$, calculate

$$\widehat{\text{power}}(h_j) = B^{-1} \sum_{i=1}^B I\{T_{n,m}^*(i, j) > c^{**}(i, j)\}$$

5. Estimate the optimal bandwidth by

$$\widehat{h}_{opt} = \text{argmax}_{1 \leq j \leq k} \widehat{\text{power}}(h_j).$$

Table 2 summarizes the results of a small simulation study, which illustrates the above procedure. Given the computational complexity of the double bootstrap procedure, we limit here to 50 simulation runs, 50 bootstrap samples and also 50 second level bootstrap samples. The sample sizes are $n = m = 50$ and the level equals $\alpha = 0.10$. The table indicates that the power for the estimated bandwidth is for all models close to the maximal power, and that the level is well respected. These results also suggest that the power of the new method shown in Table 1, is possibly underestimated, since the table only gives the power for two fixed values of h ($h = 1$ and 1.5). The power at $\widehat{h}_{opt}$ might be substantially larger.

Mod.	a	Rej. Prob(h)						$\widehat{h}_{opt}$	
		$\widehat{h}_{opt}$	1	1.25	1.5	1.75	2	Mean	Var
0	0	0.12	0.08	0.08	0.08	0.12	0.12	1.65	0.15
1	0.6	0.74	0.64	0.66	0.68	0.74	0.74	1.93	0.04
2	0.75	0.86	0.70	0.82	0.86	0.86	0.88	1.89	0.04
3	0.5	0.72	0.68	0.76	0.66	0.60	0.46	1.35	0.11
4	0.8	0.58	0.68	0.68	0.58	0.54	0.50	1.62	0.12
5	0.4	0.92	0.84	0.78	0.88	0.88	0.92	1.93	0.04
6	1	0.92	0.94	0.92	0.90	0.86	0.80	1.38	0.12

Table 2: *Rejection probabilities for the new method for five fixed bandwidths, as well as for the estimator* $\widehat{h}_{opt}$ *of the optimal bandwidth. The last two columns give the mean and variance of* $\widehat{h}_{opt}$ *over the 50 simulation runs.*

5 Proofs

Proof of Lemma 1. Equation (5) may be rewritten as

$$n = \sum_{i=1}^{n} \frac{1 + \eta_1 \left(K_{1,i} - \hat{f}_{12}(x) \right)}{1 + \eta_1 \left(K_{1,i} - \hat{f}_{12}(x) \right)} - \eta_1 \sum_{i=1}^{n} \frac{K_{1,i} - \hat{f}_{12}(x)}{1 + \eta_1 \left(K_{1,i} - \hat{f}_{12}(x) \right)},$$

which implies

$$\sum_{i=1}^{n} \frac{K_{1,i} - \hat{f}_{12}(x)}{1 + \eta_1 \left(K_{1,i} - \hat{f}_{12}(x) \right)} = 0. \tag{9}$$

Now, direct manipulations on equation (9) gives

$$\begin{aligned}0 &= \frac{\text{sign}(\eta_1)}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)-\frac{\text{sign}(\eta_1)}{n}\sum_{i=1}^{n}\frac{\left(K_{1,i}-\hat{f}_{12}(x)\right)^2\eta_1}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)}\\ &\leq \text{sign}(\eta_1)\left(\hat{f}_1(x)-\hat{f}_{12}(x)\right)\\ &\quad -\frac{|\eta_1|}{1+|\eta_1|\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|}\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2,\end{aligned}$$

which leads to

$$\begin{aligned}|\eta_1|\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2 &\leq \left(1+|\eta_1|\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right)\\ &\quad\times\text{sign}(\eta_1)\left(\hat{f}_1(x)-\hat{f}_{12}(x)\right),\end{aligned}$$

or, equivalently,

$$\begin{aligned}&|\eta_1|\left[\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2-\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\text{sign}(\eta_1)\left(\hat{f}_1(x)-\hat{f}_{12}(x)\right)\right]\\ &\leq \text{sign}(\eta_1)\left(\hat{f}_1(x)-\hat{f}_{12}(x)\right).\end{aligned} \tag{10}$$

Using H_0 we have $E\left(\hat{f}_1(x)\right)=E\left(\hat{f}_{12}(x)\right)$ and, consequently,

$$\hat{f}_1(x)-\hat{f}_{12}(x)=O_P\left(\frac{1}{\sqrt{nh}}\right). \tag{11}$$

On the other hand

$$0\leq\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\leq h^{-1}\sup_{u\in\mathbb{R}}|K(u)|+\hat{f}_{12}(x)=O_P\left(h^{-1}\right),$$

which, together with (11) gives

$$\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\text{sign}(\eta_1)\left(\hat{f}_1(x)-\hat{f}_{12}(x)\right)=O_P\left(n^{-1/2}h^{-3/2}\right). \tag{12}$$

Similar arguments lead to

$$\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2=R(K)\,h^{-1}f(x)+O_P(1)+O_P\left(n^{-1/2}h^{-3/2}\right). \tag{13}$$

Finally, (11), (12) and (13) can be applied to (10) to derive the rate

$$|\eta_1|=O_P\left(n^{-1/2}h^{1/2}\right). \tag{14}$$

We now start from (9) to obtain

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)\left[1-\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)+\frac{\eta_1^2\left(K_{1,i}-\hat{f}_{12}(x)\right)^2}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)}\right] \\
&= \hat{f}_1(x)-\hat{f}_{12}(x)-\eta_1\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2 \\
&\quad +\eta_1^2\frac{1}{n}\sum_{i=1}^{n}\frac{\left(K_{1,i}-\hat{f}_{12}(x)\right)^3}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)}. \qquad (15)
\end{aligned}
$$

Its last term can be easily bounded:

$$
\begin{aligned}
&\left|\eta_1^2\frac{1}{n}\sum_{i=1}^{n}\frac{\left(K_{1,i}-\hat{f}_{12}(x)\right)^3}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)}\right| \\
&\le \eta_1^2\left(\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\left(K_{1,i}-\hat{f}_{12}(x)\right)^2}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)} \\
&\le \eta_1^2\left(\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right) \\
&\quad\times\left[\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2+\frac{1}{n}|\eta_1|\sum_{i=1}^{n}\frac{\left|K_{1,i}-\hat{f}_{12}(x)\right|^3}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\le \eta_1^2\left(\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right)\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2 \\
&\quad+|\eta_1|^3\left(\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right)^4\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\eta_1\left(K_{1,i}-\hat{f}_{12}(x)\right)} \\
&\le \eta_1^2\left(\max_i\left|K_{1,i}-\hat{f}_{12}(x)\right|\right)\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2+O_P\left(n^{-3/2}h^{-5/2}\right) \\
&= O_P\left(n^{-1}h^{-1}\right)+O_P\left(n^{-3/2}h^{-5/2}\right). \qquad (16)
\end{aligned}
$$

Now (13), (15) and (16), lead to

$$
\eta_1=\frac{\hat{f}_1(x)-\hat{f}_{12}(x)}{\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i}-\hat{f}_{12}(x)\right)^2}+O_P\left(n^{-1}\right)+O_P\left(n^{-3/2}h^{-3/2}\right). \qquad (17)
$$

Similar calculations for the second sample give

$$\eta_2 = \frac{\hat{f}_2(x) - \hat{f}_{12}(x)}{\frac{1}{m}\sum_{j=1}^{m}\left(K_{2,j} - \hat{f}_{12}(x)\right)^2} + O_P\left(m^{-1}\right) + O_P\left(m^{-3/2}h^{-3/2}\right). \tag{18}$$

Starting from expression (4), using Taylor expansions, equations (17) and (18) and direct derivations it is not difficult to conclude

$$\begin{aligned} -2\log EL(x) &= \frac{n\left(\hat{f}_1(x) - \hat{f}_{12}(x)\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(K_{1,i} - \hat{f}_{12}(x)\right)^2} + \frac{m\left(\hat{f}_2(x) - \hat{f}_{12}(x)\right)^2}{\frac{1}{m}\sum_{j=1}^{m}\left(K_{2,j} - \hat{f}_{12}(x)\right)^2} \\ &\quad + A_1 + A_2 + O_P\left(n^{-1/2}h^{-1/2}\right) + O_P\left(n^{-1}h^{-2}\right), \end{aligned} \tag{19}$$

where

$$\begin{aligned} A_1 &= \frac{2}{3}\eta_1^3 \sum_{i=1}^{n} \frac{\left(K_{1,i} - \hat{f}_{12}(x)\right)^3}{\left(1+\xi_{1,i}\right)^3}, \\ A_2 &= \frac{2}{3}\eta_2^3 \sum_{j=1}^{m} \frac{\left(K_{2,j} - \hat{f}_{12}(x)\right)^3}{\left(1+\xi_{2,i}\right)^3}, \end{aligned}$$

with $\xi_{1,i}$ some intermediate point between 0 and $\eta_1\left(K_{1,i} - \hat{f}_{12}(x)\right)$ and $\xi_{2,j}$ some intermediate point between 0 and $\eta_2\left(K_{2,j} - \hat{f}_{12}(x)\right)$. Now equations (13) and (14) and standard manipulations can be used to obtain

$$A_1 = O_P\left(n^{-1/2}h^{-1/2}\right) \tag{20}$$

and similarly

$$A_2 = O_P\left(m^{-1/2}h^{-1/2}\right). \tag{21}$$

Now, using (13), (20) and (21) in (19) gives, after straightforward calculations,

$$-2\log EL(x) = \frac{\frac{nmh}{n+m}\left(\hat{f}_1(x) - \hat{f}_2(x)\right)^2}{R(K)\,f(x)} + O_P\left(h + n^{-1/2}h^{-1/2} + n^{-1}h^{-2}\right),$$

which proves the first statement in the lemma.

The proof of the second part relies on showing that

$$-2\log EL(x) = \frac{\frac{nmh}{n+m}\left(\hat{f}_1(x) - \hat{f}_2(x)\right)^2}{R(K)\,f(x)} + O_P\left(n^{\alpha}\left(h + n^{-1/2}h^{-1/2} + n^{-1}h^{-2}\right)\right), \tag{22}$$

uniformly in $x \in C$. A careful inspection of the rates obtained in the first part of the proof leads to parallel uniform orders but with the extra factor n^α. To do this, condition A3 and Bernstein inequality can be used as in Stone (1984). This is illustrated by stating and proving next, only one of the uniform results needed. ■

Lemma 3 *Assume the conditions in Lemma 1. Then,*

$$\sup_{x\in C}\left|\hat{f}_1(x)-\hat{f}_{12}(x)\right| = O_P\left(\frac{n^\alpha}{\sqrt{nh}}\right).$$

Proof. Let us define the random variables

$$Z_r(x) = \begin{cases} \frac{m}{n}\left(K_{1,r}-f(x)\right) & \text{if} \quad r\in\{1,2,\ldots,n\} \\ f(x)-K_{2,r-n} & \text{if} \quad r\in\{n+1,n+2,\ldots,n+m\} \end{cases}$$

It is not difficult to prove that $|Z_r(x)| \leq \max\left\{1,\frac{m}{n}\right\}\|K\|h^{-1}$ and $Var\left(Z_r(x)\right) \leq \left(\max\left\{1,\frac{m}{n}\right\}\right)^2 R(K)\|f\|h^{-1}$, where $\|g\| = \sup_{x\in\mathbf{R}}|g(x)|$. Now, Bernstein inequality applies to the random variables $Z_r(x)$, $r=1,2,\ldots,n+m$, to obtain

$$P\left(\left|\hat{f}_1(x)-\hat{f}_{12}(x)\right|>t\right) = P\left(\left|(n+m)^{-1}\sum_{r=1}^{n+m}Z_r(x)\right|>t\right) \tag{23}$$
$$\leq 2\exp\left(-\frac{(n+m)t^2}{2\left(\max\left\{1,\frac{m}{n}\right\}\right)^2 R(K)\|f\|h^{-1}+\frac{2}{3}\max\left\{1,\frac{m}{n}\right\}\|K\|h^{-1}t}\right)$$
$$\leq 2\exp\left(-\frac{(n+m)ht^2}{3\left(\max\left\{1,\frac{m}{n}\right\}\right)^2 R(K)\|f\|}\right),$$

whenever $t \leq \frac{3R(K)\|f\|}{2\|K\|}$. On the other hand, using condition A3, there exists some $\lambda > 0$ such that $|K(x)-K(y)| \leq \lambda|x-y|$. As a consequence,

$$\left|\hat{f}_1(x)-\hat{f}_{12}(x)-\left(\hat{f}_1(y)-\hat{f}_{12}(y)\right)\right| \leq \frac{2\lambda}{h^2}|x-y|.$$

Let us now fix some $\varepsilon > 0$, and define $\delta_n = \frac{\varepsilon n^\alpha h^2}{4\lambda\sqrt{nh}}$. Using conditions A2 and A3, we may select $x_1, x_2, \ldots, x_{k_n} \in C$, with $k_n = O\left(\frac{\sqrt{nh}}{n^\alpha h^2}\right)$, such that for all $x \in C$, there exists some $i \in \{1,2,\ldots,k_n\}$, with $|x-x_i| \leq \delta_n$, which, therefore, implies that

$$\left|\hat{f}_1(x)-\hat{f}_{12}(x)-\left(\hat{f}_1(x_i)-\hat{f}_{12}(x_i)\right)\right| \leq \frac{\varepsilon n^\alpha}{2\sqrt{nh}}.$$

As a consequence,

$$\sup_{x\in C}\left|\hat{f}_1(x)-\hat{f}_{12}(x)\right|\le\frac{\varepsilon n^{\alpha}}{2\sqrt{nh}}+\max_{1\le i\le k_n}\left|\hat{f}_1(x_i)-\hat{f}_{12}(x_i)\right|$$

which, using (23), leads to

$$\begin{aligned}&P\left(\sup_{x\in C}\left|\hat{f}_1(x)-\hat{f}_{12}(x)\right|>\frac{\varepsilon n^{\alpha}}{\sqrt{nh}}\right)\le P\left(\max_{1\le i\le k_n}\left|\hat{f}_1(x_i)-\hat{f}_{12}(x_i)\right|>\frac{\varepsilon n^{\alpha}}{2\sqrt{nh}}\right)\\&\le\sum_{i=1}^{k_n}P\left(\left|\hat{f}_1(x_i)-\hat{f}_{12}(x_i)\right|>\frac{\varepsilon n^{\alpha}}{2\sqrt{nh}}\right)\\&\le 2k_n\exp\left(-\frac{\left(1+\frac{m}{n}\right)\varepsilon^2n^{2\alpha}}{12\left(\max\left\{1,\frac{m}{n}\right\}\right)^2R(K)\|f\|}\right).\end{aligned}$$

The last term tends to zero, using the order of k_n and A1. ■

Remark 4 *The previous result is an extension of expression (11) to a uniform setting in $x\in C$. It can also be derived from one of the results in Silverman (1978).*

Proof of Theorem 2. First, we will prove that $T_{n,m}+2\int_C\log EL(x)\,dF(x)=o_P\left(h^{1/2}\right)$. To do this we start writing

$$T_{n,m}+2\int_C\log EL(x)\,dF(x)=-2\int_C\log EL(x)\,d\left(\hat{F}_{12}-F\right)(x).$$

Now bounding the absolute value of the integral in the right hand side by some integral with respect to the total variation measure, the Dvoretzky-Kiefer-Wolfowitz inequality and the uniform result (22) apply to get

$$\begin{aligned}T_{n,m}+2\int_C\log EL(x)\,dF(x)&=O_P\left(n^{1/2}h\sup_{x\in C}\frac{\left(\hat{f}_1(x)-\hat{f}_2(x)\right)^2}{f(x)}\right)\\&\quad+O_P\left(n^{\alpha-1/2}\left(h+n^{-1/2}h^{-1/2}+n^{-1}h^{-2}\right)\right).\end{aligned}\tag{24}$$

On the other hand using arguments parallel to those in the proof of Lemma 3, it is easy to prove that

$$\sup_{x\in C}\left(\hat{f}_1(x)-\hat{f}_2(x)\right)^2=O_P\left(\frac{n^{\alpha}}{nh}\right).$$

This rate and conditions A1 and A2, can be used in (24), to conclude

$$\begin{aligned}T_{n,m}+2\int_C\log EL(x)\,dF(x)&=O_P\left(n^{\alpha-1/2}+n^{\alpha-1/2}\left(h+n^{-1/2}h^{-1/2}+n^{-1}h^{-2}\right)\right)\\&=o_P\left(h^{1/2}\right).\end{aligned}$$

In view of this representation the theorem will be proved by showing

$$\frac{R(K)}{\sqrt{2hR_C(f)\,R(K*K)}}\left[-2\int_C \log EL(x)\,dF(x)-\mu_{n,m}(C)\right]\xrightarrow{\text{d}} N(0,1),$$

which, using Lemma 1, condition A1 and the Dvoretzky-Kiefer-Wolfowitz inequality, is equivalent to proving

$$\frac{\frac{nmh}{n+m}\int_C\left(\hat{f}_1(x)-\hat{f}_2(x)\right)^2dx-R(K)\,\mu(C)}{\sqrt{2hR_C(f)\,R(K*K)}}\xrightarrow{\text{d}} N(0,1), \tag{25}$$

with $\mu(C)=\int_C dF(x)$.

To prove (25), we first decompose the main term as follows

$$\int_C\left(\hat{f}_1(x)-\hat{f}_2(x)\right)^2dx=B_1+B_2-2B_3, \tag{26}$$

with

$$\begin{aligned} B_1 &= \int_C\left(\hat{f}_1(x)-E\hat{f}_1(x)\right)^2dx,\\ B_2 &= \int_C\left(\hat{f}_2(x)-E\hat{f}_2(x)\right)^2dx,\\ B_3 &= \int_C\left(\hat{f}_1(x)-E\hat{f}_1(x)\right)\left(\hat{f}_2(x)-E\hat{f}_2(x)\right)dx. \end{aligned}$$

Using Theorem 4.1 in Bickel and Rosenblatt (1973) (note however that condition (2.16) in the statement of that theorem should be (2.15)) and conditions A1-A4 it is not difficult to prove that

$$\begin{aligned} &\frac{1}{\sqrt{2hR_C(f)\,R(K*K)}}\left[nhB_1-R(K)\,\mu(C)\right]\xrightarrow{\text{d}} N(0,1),\\ &\frac{1}{\sqrt{2hR_C(f)\,R(K*K)}}\left[mhB_2-R(K)\,\mu(C)\right]\xrightarrow{\text{d}} N(0,1). \end{aligned}$$

However, a closer look at the proof of Theorem 4.1 in Bickel and Rosenblatt (1973) shows that

$$\begin{aligned} B_1 &= \int_C {}_2L_{n,1}^2(x)dx+O_P(h),\\ B_2 &= \int_C {}_2L_{m,2}^2(x)dx+O_P(h),\\ B_3 &= \int_C {}_2L_{n,1}\;{}_2L_{m,2}(x)dx+O_P(h) \end{aligned}$$

where ${}_2L_{n,1}(x)$ and ${}_2L_{m,2}(x)$ are defined as in page 1079 of Bickel and Rosenblatt (1973) in terms of two independent Wiener processes, $Z_1(\bullet)$ and $Z_2(\bullet)$, as in expression (2.4) in that paper. As a consequence, using (26) we have

$$\int_C \left(\hat{f}_1(x) - \hat{f}_2(x)\right)^2 dx = \int_C \left({}_2L_{n,1} - {}_2L_{m,2}\right)^2(x)dx + O_P(h)$$

and (25) can be proved following the same arguments used by Bickel and Rosenblatt (1973) to prove their Theorem 4.1. ■

Acknowledgments. Research of the first author was supported in part by MCyT Grant BFM2002-00265 (EU ERDF support included) and XUGA Grant PGIDIT03PXIC10505-PN. The second author acknowledges support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy).

References

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. *J. Multivariate Anal.* 50, 41–54.

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

Chen, S.X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* 83, 329–341.

Chen, S.X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. R. Statist. Soc. - Series B* 65, 663–678.

Claeskens, G., Jing, B.Y., Peng, L. and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *Canad. J. Statist.* 31, 173–190.

Einmahl, J.H.J. and Khmaladze, E.V. (2001). The two-sample problem in $\mathbb{R}^m$ and measure-valued martingales. In *State of the art in probability and statistics*, pp. 434–463, IMS Lecture Notes Monogr. Ser., 36, Inst. Math. Statist., Beachwood, Ohio.

Einmahl, J.H.J. and McKeague, I.W. (1999). Confidence tubes for multiple quantile plots via empirical likelihood. *Ann. Statist.* 27, 1348–1367.

Einmahl, J.H.J. and McKeague, I.W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli* 9, 267–290.

Hall, P. and Owen, A. B. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Statist.* 2 , 273–289.

Henze, N. and Nikitin, Y.Y. (2003). Two-sample tests based on the integrated empirical process. *Comm. Statist. Theory Methods* 32, 1767–1788.

Jing, B.Y. (1995). Two-sample empirical likelihood method. *Statist. Probab. Lett.* 24, 315–319.

Kim, K.K. and Foutz, R. V. (1987). Tests for the multivariate two-sample problem based on empirical probability measures. *Canad. J. Statist.* 15, 41–51.

Louani, D. (2000). Exact Bahadur efficiencies for two-sample statistics in functional density estimation. *Statist. Decisions* 18, 389–412.

McKeague, I.W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statist. Probab. Lett.* 60, 405–415.

Qin, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Ann. Inst. Statist. Math.* 46, 117–126.

Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* 6, 177–184.

Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* 12, 1285–1297.

Zhang, B. (2000). Estimating the treatment effect in the two-sample problem with auxiliary information. *J. Nonparam. Statist.* 12, 377–389.