BOOTSTRAP BANDWIDTH SELECTORS FOR TWO KERNEL-TYPE RELATIVE DENSITY ESTIMATORS

ELISA MARÍA MOLANES-LÓPEZ[1] AND RICARDO CAO[1]

BOOTSTRAP BANDWIDTHS FOR THE RELATIVE DENSITY

Abstract. One of the main issues when estimating nonparametrically a density function is how to select the smoothing parameter based on the data. In the context of a two-sample problem we propose several data-driven selectors for choosing the bandwidth of two kernel-type relative density estimators. These selectors are based on the bootstrap technique and try to estimate the value minimizing their corresponding mean integrated squared error (MISE). The differences are in the way that the bootstrap MISE function is approximated at a grid of values, either by Monte Carlo, using a resampling scheme, or straightforwardly, based on a closed expression of the bootstrap MISE. The performance of these bootstrap selectors is checked through a simulation study. Unlike what happens in the one-sample case, this simulation study revealed the intensive computing time required to numerically approximate the closed expression of the bootstrap MISE.

Key words and phrases: grade density, mean integrated squared error, Monte Carlo approximation, optimal bandwidth, resampling, two-sample problem.

1. Introduction

It is well known that one of the main concerns in nonparametric curve estimation is the selection of the smoothing parameter. In the literature, different data-driven selectors have been proposed. Some of these proposals deal with the problem trying to estimate the value that minimizes a global error criterion such as the mean integrated squared error

(MISE). In this article we propose several data-driven selectors to help in the choice of the bandwidth for two kernel-type estimators of the relative density function. The proposed selectors are all based on the bootstrap technique. For each proposal we start designing a resampling scheme that imitates the random procedure from which the original samples were drawn. The resampling plan is used to compute a bootstrap estimate of the MISE function (either by Monte Carlo or via a closed formula) and to find its minimizer, that will be the corresponding bootstrap bandwidth selector.

There exist several articles in the literature that deal with the problem of nonparametric estimation of relative characteristics such as the relative distribution, the relative density and the relative hazard rate functions. For example, in the setting of a two-sample problem with completely observed data, Handcock and Janssen (1996) and Handcock and Morris (1999) face the problem of estimating the relative distribution function while Ćwik and Mielniczuk (1993), Handcock and Janssen (2002) and Molanes and Cao (2007) deal with the problem of estimating the relative density function. On the other hand, other settings that are typical in survival analysis were also considered in the literature. For example, Cao et al (2000, 2001) deal with the problem of estimating the relative density function when the data are right censored while Cao et al. (2005) consider the problem of estimating the relative hazard rate using left truncated and right censored data. In the same setup of left-truncated and right-censored data, Molanes and Cao (2006) proposed and studied kernel relative density estimators. It is worth mentioning here that, in all these papers, an important question is how to select the smoothing parameter in practical applications through a data-driven mechanism.

In the following section two kernel-type relative density estimators are introduced and a closed expression for the MISE function of the first one is obtained. In Section 3 the bootstrap resampling schemes are presented in detail for each selector. Section 4 includes a simulation study, where the performance of these and other selectors are compared. The proof of the theorem given in Section 2 is included in Section 5.

2. Closed expressions for the MISE

Consider two independent random samples $\{X_{01},\ldots,X_{0n_0}\}$ and $\{X_{11},\ldots,X_{1n_1}\}$ from two distributions with density functions f_0 and f_1 and distribution functions F_0 and F_1. We denote by R the relative distribution of the X_{1j}'s with respect to (wrt) the X_{0i}'s: $R(t) = P(F_0(X_{1j}) \leq t) = F_1(F_0^{-1}(t))$ and r its corresponding relative density $r(t) = R^{(1)}(t)$.

Throughout the paper we will consider two kernel-type estimators of r: $\hat{r}_h$ (see Ćwik and Mielniczuk, 1993) and a slight version of it, $\hat{r}_{h,h_0}$, in which the empirical cdf of F_0 is replaced by a smoothed estimate (see Molanes and Cao, 2007):

$$\hat{r}_h(t) = \frac{1}{h}\int K\left(\frac{t-F_{0n_0}(z)}{h}\right)\mathrm{d}F_{1n_1}(z) = \frac{1}{n_1 h}\sum_{j=1}^{n_1} K\left(\frac{t-F_{0n_0}(X_{1j})}{h}\right) \tag{2.1}$$

and

$$\hat{r}_{h,h_0}(t) = \frac{1}{h}\int K\left(\frac{t-\tilde{F}_{0h_0}(z)}{h}\right)\mathrm{d}F_{1n_1}(z) = \frac{1}{n_1 h}\sum_{j=1}^{n_1} K\left(\frac{t-\tilde{F}_{0h_0}(X_{1j})}{h}\right), \tag{2.2}$$

where K is a kernel function, h is the bandwidth, F_{0n_0} and F_{1n_1} are the empirical distribution functions of F_0 and F_1 based on, respectively, X_{0i}'s and X_{1j}'s, and $\tilde{F}_{0h_0}$ is a kernel-type estimate of F_0 based on X_{0i}'s with bandwidth h_0.

Let us consider the following assumptions:

(R) $F_0(X_1)$ is absolutely continuous.

(K) K is a bounded density function in $(-\infty,\infty)$.

The following result presents a closed expression for the MISE of the estimator $\hat{r}_h$.

THEOREM 2.1. *Assume conditions (R) and (K). Then, the MISE of the kernel relative density estimator in (2.1) can be written as follows:*

$$MISE\left(\hat{r}_h\right) = \int r^2(t)\,\mathrm{d}t - 2\sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,r(\cdot)}(i)\int K_h\left(t-\frac{i}{n_0}\right) r(t)\,\mathrm{d}t + \frac{\left(K_h * K_h\right)(0)}{n_1}$$
$$+2\frac{n_1-1}{n_1}\sum_{i=0}^{n_0}\sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,r(\cdot)}(i,j)\left(K_h * K_h\right)\left(\frac{j-i}{n_0}\right),$$

where

$$a_{n_0,r(\cdot)}(i) = \int_0^1 s^i (1-s)^{n_0-i} r(s)\,\mathrm{d}s,$$

$$b_{n_0,r(\cdot)}(i,j) = \int_0^1 (1-s_2)^{n_0-j} r(s_2) s_2^{j+1} \int_0^1 s_3^i (1-s_3)^{j-i} r(s_2 s_3)\,\mathrm{d}s_3\mathrm{d}s_2, \quad \textit{for } j \geq i,$$

$$(K_h * K_h)(t) = \int K_h(t-s) K_h(s)\,\mathrm{d}s,$$

$C_{n_0}^i = \binom{n_0}{i}$ *and* $P_{n_0}^{i,j-i,n_0-j} = \frac{n_0!}{i!(j-i)!(n_0-j)!}$ *denote, respectively, the binomial and the multinomial coefficients.*

The proof of this result can be found in Section 5.

3. Bootstrap Selectors

When using the bootstrap technique to estimate the MISE of, either $\hat{r}_h$ or $\hat{r}_{h,h_0}$, one possibility could be to approximate the distribution function of the ISE process and then compute its expectation. To this aim we first need to define a resampling scheme that imitates the procedure from which the two original samples were drawn. As pointed out in Cao (1993) for the setting of ordinary density estimation, this can be achieved replacing the role of the true target density, in this case r, by some estimator of it. Since we are in a two-sample setting we need to draw a pair of resamples of n_0 and n_1 observations respectively, the first one coming from a population, say X_0^*, and the second one coming from another one, say X_1^*. On the other hand, the relative density of X_1^* wrt X_0^* should coincide with the kernel relative density estimator of X_1 wrt X_0. Therefore, the ideas presented in Cao (1993) require some modifications to be adapted to this new setting. There exist at least two ways to proceed. Either replacing the roles of the densities, f_0 and f_1, by some appropriate estimators or considering a uniform distribution in $[0,1]$ for X_0^* and a distribution with density equal to the relative density estimator of X_1 wrt X_0 for X_1^*. The second possibility is justified by noting that the sampling distribution of $\hat{r}_h$ only depends on the two populations through their relative density, r (see also the expression for $MISE(\hat{r}_h)$ in Theorem 2.1).

We now present a bootstrap procedure to approximate $MISE(\hat{r}_h)$:

(a) Select a pilot bandwidth, g, and construct the relative density estimator $\hat{r}_g$ (see (2.1)).

(b) Draw bootstrap resamples $\left\{X_{01}^*, \ldots, X_{0n_0}^*\right\}$, from a uniform distribution in $[0,1]$, and $\left\{X_{11}^*, \ldots, X_{1n_1}^*\right\}$, with density function $\hat{r}_g$.

(c) Consider, for each $h > 0$, the bootstrap version of the kernel estimator (2.1):

$$\hat{r}_h^*(x) = n_1^{-1} \sum_{j=1}^{n_1} K_h(x - F_{0n_0}^*(X_{1j}^*)),$$

where $F_{0n_0}^*$ denotes the empirical distribution function of $\left\{X_{01}^*, \ldots, X_{0n_0}^*\right\}$.

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}_h^*) = E^*\left[\int \left(\hat{r}_h^*(x) - \hat{r}_g(x)\right)^2 \mathrm{d}x\right] \tag{3.1}$$

(e) Find the minimizer of (3.1). This value, denoted by $h_{MISE(\hat{r}_h)}^*$, is a bootstrap analogue of the MISE bandwidth for $\hat{r}_h$.

By definition, the bootstrap MISE function in (3.1) does not depend on the resamples. Therefore, in case that a closed expression could be found for it, Monte Carlo approximations could be avoided. In other words, there would be no need of drawing resamples (steps (b) and (c) in the bootstrap procedure sketched above) which always means an important computational load. In the one-sample problem this approach was plausible (see Cao et al., 1994) and yielded a considerable saving of computing time.

A bootstrap version for Theorem 2.1 can be proved using parallel arguments. For this reason, its proof is not included in the paper.

THEOREM 3.1. *Assume condition (K). Then,*

$$\begin{aligned} MISE^*\left(\hat{r}_h\right) = & \int \hat{r}_g^2\left(t\right) \mathrm{d}t - 2\sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,\hat{r}_g(\cdot)}\left(i\right) \int K_h\left(t - \frac{i}{n_0}\right) \hat{r}_g\left(t\right) \mathrm{d}t + \frac{\left(K_h * K_h\right)(0)}{n_1} \\ & + 2\frac{n_1 - 1}{n_1} \sum_{i=0}^{n_0} \sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,\hat{r}_g(\cdot)}\left(i,j\right) \left(K_h * K_h\right)\left(\frac{j-i}{n_0}\right), \end{aligned} \tag{3.2}$$

where

$$a_{n_0,\hat{r}_g(\cdot)}\left(i\right) = \int s^i \left(1-s\right)^{n_0-i} \hat{r}_g\left(s\right) \mathrm{d}s$$

and

$$b_{n_0,\hat{r}_g(\cdot)}(i,j) = \int_0^1 (1-s_2)^{n_0-j}\,\hat{r}_g(s_2)\,s_2^{j+1}\int_0^1 s_3^i\,(1-s_3)^{j-i}\,\hat{r}_g(s_2s_3)\,\mathrm{d}s_3\mathrm{d}s_2.$$

Based on the bootstrap scheme shown previously and the closed expression obtained for $MISE^*(\hat{r}_h)$, we propose two bootstrap bandwidth selectors. Both consist in approximating $MISE^*(\hat{r}_h)$ and finding its minimizer (which yields an approximation of $h^*_{MISE(\hat{r}_h)}$). While the first one, say h^*_{CE}, approximates (3.1) using the closed expression (3.2), the second proposal, say h^*_{MC}, estimates (3.1) by Monte Carlo taking a large number of resamples as described in steps (b) and (c).

When dealing with the estimator $\hat{r}_{h,h_0}$ in (2.2), there is no hope to find a closed expression for its MISE, similar to that one in (3.2). Below we present two bootstrap procedures to approximate $MISE(\hat{r}_{h,h_0})$. The first proposal is as follows:

Smooth Uniform Monte Carlo Bootstrap resampling plan (SUMC)

(a) Select two pilot bandwidths, g and g_0, and construct the estimator $\hat{r}_{g,g_0}$ (see (2.2)) of the relative density r. Let H be the cdf of a uniform random variable in [0,1] and consider

$$\tilde{H}_b(x) = n_0^{-1}\sum_{i=1}^{n_0}\mathbb{M}\left(\frac{x-U_i}{b}\right), \tag{3.3}$$

a kernel-type estimate of H based on the uniform kernel M in $[-1,1]$ (with distribution function $\mathbb{M}$), the bandwidth parameter b and the sample $\{U_1,\ldots,U_{n_0}\}$ coming from H. Approximate the MISE function of $\tilde{H}_b$ by Monte Carlo and find its minimizer b_0.

(b) Draw bootstrap samples $\left\{X^*_{01},\ldots,X^*_{0n_0}\right\}$ and $\left\{X^*_{11},\ldots,X^*_{1n_1}\right\}$ from, respectively, a uniform distribution in $[0,1]$ and the density function $\hat{r}_{g,g_0}$.

(c) Consider, for each $h>0$, the bootstrap version of the kernel estimator (2.2):

$$\hat{r}^*_{h,b_0}(x) = n_1^{-1}\sum_{j=1}^{n_1}K_h(x-\tilde{F}^*_{0b_0}(X^*_{1j})),$$

where $\tilde{F}^*_{0b_0}$ denotes a kernel-type cdf estimate based on the bootstrap resample $\left\{X^*_{01},\ldots,X^*_{0n_0}\right\}$, the uniform kernel in $[-1,1]$ and the bandwidth parameter b_0 computed previously in (a).

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}^*_{h,b_0}) = E_* \left[\int \left(\hat{r}^*_{h,b_0}(x) - \hat{r}_{g,g_0}(x) \right)^2 \mathrm{d}x \right] \tag{3.4}$$

(e) Find a numerical approximation of the minimizer of (3.4). This value, denoted by h^*_{SUMC}, is a bootstrap version of the MISE bandwidth for $\hat{r}_{h,h_0}$.

Since we do not have a closed expression for $MISE^*(\hat{r}^*_{h,b_0})$, this function is approximated by Monte Carlo.

The second proposal is sketched below.

Smooth Monte Carlo Bootstrap resampling plan (SMC)

(a) Select three pilot bandwidths, g, g_0 and g_1, and construct the estimators $\hat{r}_{g,g_0}$ and $\tilde{f}_{0,g_1}$ of, respectively, the relative density r and the density f_0. Here, $\tilde{f}_{0,g_1}$ denotes the Parzen-Rosenblatt estimator of the ordinary density f_0 with bandwidth g_1 (see Rosenblatt, 1956, and Parzen, 1962, for more details).

(b) Draw bootstrap resamples $\left\{X^*_{01}, \ldots, X^*_{0n_0}\right\}$ from $\tilde{f}_{0,g_1}$ and $\left\{Z^*_1, \ldots, Z^*_{n_1}\right\}$ from $\hat{r}_{g,g_0}$. Define $X^*_{1j} = \tilde{F}^{-1}_{0g_1}(Z^*_j)$, $j = 1, \ldots, n_1$.

(c) Consider, for each $h > 0$, the bootstrap version of the kernel estimator (2.2):

$$\hat{r}^*_{h,h_0}(x) = n_1^{-1} \sum_{j=1}^{n_1} K_h(x - \tilde{F}^*_{0h_0}(X^*_{1j})),$$

with $\tilde{F}^*_{0h_0}$ a smooth estimate of F_0 based on the bootstrap resample $\left\{X^*_{01}, \ldots, X^*_{0n_0}\right\}$.

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}^*_{h,h_0}) = E^* \left[\int \left(\hat{r}^*_{h,h_0}(x) - \hat{r}_{g,g_0}(x) \right)^2 \mathrm{d}x \right] \tag{3.5}$$

(e) Find the minimizer of (3.5), h^*_{SMC}, which is a bootstrap analogue of the MISE bandwidth for $\hat{r}_{h,h_0}$.

Once more, a Monte Carlo approach has to be used to approximate the function in (3.5).

4. Simulations

4.1 *Practical implementation*

Although in the previous section several bootstrap selectors have been proposed, some aspects of them remained unspecified such as for example how the required pilot bandwidths, g, g_0 or g_1, are chosen. In the following we will use for K the standard Gaussian kernel.

Let us start with the proposals h^*_{CE} and h^*_{MC}. The pilot bandwidth g is selected based on the AMSE-optimal bandwidth, $g_{AMSE,4}(r)$, to estimate the value of $\Psi_4(r)$ where $\Psi_\ell(r) = \int_0^1 r^{(\ell)}(x)r(x)\mathrm{d}x$ (see Molanes and Cao, 2007, for details). Note that, under regularity assumptions on r, $\Psi_4(r)$ is equal to $C\left(r^{(2)}\right)$, the curvature of r. Here we denote $C(g) = \int g(x)^2\mathrm{d}x$. Based on the rule of thumb (Silverman, 1986), the unknown quantities depending on r that appear in $g_{AMSE,4}(r)$, are replaced by parametric estimates based on an appropriate fit for r. This procedure leads us to define g as follows

$$g = \left(\frac{-2K^{(4)}(0)(1+\kappa^2\hat{\Psi}_0^P(r))}{d_K\hat{\Psi}_6^P(r)}\right)^{\frac{1}{7}} n_1^{-\frac{1}{7}},$$

where $\hat{\Psi}_0^P(r)$ and $\hat{\Psi}_6^P(r)$ are parametric estimates of, respectively, $\Psi_0(r)$ and $\Psi_6(r)$, based on the parametric fit, $\hat{b}(x;N,R)$, considered for $r(x)$, that is introduced below.

Let $\beta(x,a,b)$ be the beta density

$$\beta(x,a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \quad x\in[0,1],$$

and let $B(x;N,G)$ be the Bernstein polynomial associated to any continuous function G on the closed interval $[0,1]$

$$B(x;N,G) = (N+1)^{-1}\sum_{j=1}^{N+1} G\left(\frac{j-1}{N}\right)\beta(x,j,N-j+2).$$

Applying Weierstrass's theorem it is known that $B(x;N,G)$ converges to $G(x)$ uniformly in $x\in[0,1]$ as $N\to\infty$. For a distribution function, G, on $[0,1]$, it follows that $B(x;N,G)$ is a proper distribution function with density function $b(x;N,G) = B^{(1)}(x;N,G)$, i.e.

$$b(x;N,G) = \sum_{j=1}^{N}\left(G\left(\frac{j}{N}\right) - G\left(\frac{j-1}{N}\right)\right)\beta(x,j,N-j+1). \tag{4.1}$$

Based on this idea, we propose for $r(x)$ the following parametric fit, $\hat{b}(x;N,R)$, where the unknown relative distribution function R in (4.1) (when $G=R$) is replaced by a smooth estimate, $\tilde{R}_{g_{\tilde{R}}}$, as follows:

$$\hat{b}(x;N,R)=\sum_{j=1}^{N}\left(\tilde{R}_{g_{\tilde{R}}}\left(\frac{j}{N}\right)-\tilde{R}_{g_{\tilde{R}}}\left(\frac{j-1}{N}\right)\right)\beta(x,j,N-j+1) \tag{4.2}$$

where

$$\tilde{R}_{g_{\tilde{R}}}(x)=n_1^{-1}\sum_{j=1}^{n_1}\mathbb{M}\left(\frac{x-F_{0n_0}(X_{1j})}{g_{\tilde{R}}}\right), \tag{4.3}$$

$$g_{\tilde{R}}=\left(\frac{2\int_{-\infty}^{\infty}xM\left(x\right)\mathbb{M}\left(x\right)\mathrm{d}x}{n_1 d_M^2\hat{C}^P\left(r^{(1)}\right)}\right)^{\frac{1}{3}}.$$

In practice we have used a number of $N=2n_1$ beta distributions in the mixture (4.2). We refer to the interested reader to Kakizawa (2004) for more details with respect to this choice. Note that bandwidth $g_{\tilde{R}}$ is based on the AMISE-optimal bandwidth, $g_{AMISE}(R)$, for a kernel type estimator of the distribution function in a one-sample problem (see Polansky and Baker, 2000, for more details). As before, by means of the rule of thumb, a Gaussian fit using the method of moments is considered for r based on the relative sample $\{F_{0n_0}(X_{1j})\}_{j=1}^{n_1}$, and then the unknown quantity, $C(r^{(1)})$, appearing in $g_{AMISE}(R)$ is approximated parametrically by $\hat{C}^P\left(r^{(1)}\right)$.

For implementing h^*_{SUMC} and h^*_{SMC}, g is obtained using a similar procedure as explained above for selectors h^*_{CE} and h^*_{MC}. The differences now are that the Gaussian fit considered for r to estimate the unknown quantity, $C(r^{(1)})$, appearing in $g_{AMISE}(R)$, is based on the smoothed relative sample $\left\{\tilde{F}_{0g_0}(X_{1j})\right\}_{j=1}^{n_1}$, and F_{0n_0} in (4.3), is replaced by the smooth estimate

$$\tilde{F}_{0g_0}\left(x\right)=n_0^{-1}\sum_{i=1}^{n_0}\mathbb{M}\left(\frac{x-X_{0i}}{g_0}\right)$$

$$g_0=\left(\frac{2\int_{-\infty}^{\infty}xM\left(x\right)\mathbb{M}\left(x\right)\mathrm{d}x}{n_0 d_M^2\tilde{C}^P\left(f_0^{(1)}\right)}\right)^{\frac{1}{3}},$$

where $\tilde{C}^P\left(f_0^{(1)}\right)$ denotes a parametric estimate of $C\left(f_0^{(1)}\right)$ based on a gamma fit for f_0, using the method of moments and the original sample $\{X_{01},\ldots,X_{0n_0}\}$. Note that

the definiton of g_0 follows the same strategy as the definition of $g_{\tilde{R}}$ given above. The only difference now is that the target distribution is F_0 rather than R and therefore a parametric scale needs to be assumed for f_0 (not for r).

The selector h^*_{SMC} entails a third pilot bandwidth, g_1. In this case, we consider the AMSE-optimal bandwidth, $g_{AMSE,4}(f_0)$, to estimate $\Psi_4(f_0)$ in a one-sample problem (see Wand and Jones, 1995). Note that, under regularity conditions on f_0, $\Psi_4(f_0)$ is equal to $C\left(f_0^{(2)}\right)$, the curvature of f_0. Using again the rule of thumb, the bandwidth g_1 is defined as follows

$$g_1 = \left(\frac{-2K^{(4)}(0)}{d_K \hat{\Psi}_6^P(f_0)}\right)^{\frac{1}{7}} n_1^{-\frac{1}{7}},$$

where $\hat{\Psi}_6^P(f_0)$ is a parametric estimate of $\Psi_6(f_0)$, based on a gamma fit for f_0. As before, the parameters of this gamma distribution are estimated by the method of moments, using the original sample $\{X_{01}, \ldots, X_{0n_0}\}$.

Note that the computation of the selector h^*_{SMC} requires the selection of another bandwidth, denoted by h_0 in (3.5), for every bootstrap resample of size n_0 drawn as in step (b) of the SMC resampling plan. The expression of h_0 is the same as the one given previously for g_0. The only difference now is that the gamma fit used to approximate the unknown functional $C\left(f_0^{(1)}\right)$ is obtained using the bootstrap resample rather than the original sample.

It is worth mentioning here that all the kernel type estimates required in the computation of the pilot bandwidths are corrected by the boundary effect using the well-known reflection method (see Schuster, 1985). Likewise, all the kernel estimates required in steps (a) and (c) of the resampling plans SUMC and SMC, are boundary corrected. However, for selectors h^*_{CE} and h^*_{MC} only $\hat{r}_g$ (in step (a)) is boundary corrected.

In order to compare with other data driven selectors proposed in the literature, we have considered one of the proposals given by Molanes and Cao (2007), say h_{SJ_2}, and the slightly modified version of the selector introduced by Ćwik and Mielniczuk (1993), b_{3c} (see Molanes and Cao, 2007, for more details).

4.2 *Models and results*

The simulations were carried out for different sample sizes and the performance of the different data-driven selectors was examined for seven populational models for r (see (a)-(g) below).

Put Figure 1 about here.

Let U_0 denote a uniform distribution in the interval $[0,1]$ and let W be the Weibull cumulative distribution function with parameters $(2,3)$. The first sample was drawn from the random variate $X_0 = W^{-1}(U_0)$ and the second sample from the random variate $X_1 = W^{-1}(S)$, where S is a random variate from one of the following populations (see Figures 1 and 2):

(a) $V = \frac{1}{4}(U_1 + U_2 + U_3 + U_4)$, where U_1, U_2, U_3, U_4 are iid $U[0,1]$.

(b) A mixture consisting of V_1 with probability $\frac{1}{2}$ and V_2 with probability $\frac{1}{2}$, where $V_1 = \frac{V}{2}$, $V_2 = \frac{V+1}{2}$ and V as for model (a).

(c) A beta distribution with parameters 4 and 5 ($\beta(4,5)$).

(d) A mixture consisting of V_1 with probability $\frac{1}{2}$ and V_2 with probability $\frac{1}{2}$, where $V_1 \overset{d}{=} \beta(15,4)$ and $V_2 \overset{d}{=} \beta(5,11)$.

(e) A beta distribution with parameters 14 and 17 ($\beta(14,17)$).

(f) A mixture consisting of V_1 with probability $\frac{4}{5}$ and V_2 with probability $\frac{1}{5}$, where $V_1 \overset{d}{=} \beta(14,37)$ and $V_2 \overset{d}{=} \beta(14,20)$.

(g) A mixture consisting of V_1 with probability $\frac{1}{3}$ and V_2 with probability $\frac{2}{3}$, where $V_1 \overset{d}{=} \beta(34,15)$ and $V_2 \overset{d}{=} \beta(15,30)$.

Put Figure 2 about here.

For each one of the relative populations listed above, a large number (250 or 500) of pairs of samples were taken. For each pair of samples, the six bandwidth selectors (h^*_{CE}, h^*_{MC}, h^*_{SUMC}, h^*_{SMC}, h_{SJ_2} and b_{3c}), let say $\hat{h}$, were computed and, based on each

one, the kernel-type relative density estimate, (2.1) or (2.2), was computed. While for b_{3c} and h_{SJ_2} 500 estimations of r were computed, for each bootstrap bandwidth selector only 250 estimations of r were obtained, due to the important computational load that their implemetation requires. Based on them, the following global error measure was approximated by Monte Carlo:

$$EM = E\left[\int \left(\hat{r}(t) - r(t)\right)^2 \mathrm{d}t\right],$$

where $\hat{r}$ denotes $\hat{r}_{\hat{h}}$, for selectors h^*_{CE}, h^*_{MC} and b_{3c}, and the boundary corrected version of $\hat{r}_{\hat{h},h_0}$ for selectors h^*_{SUMC}, h^*_{SMC} and h_{SJ_2}.

Put Table 1 about here.

When implementing h^*_{CE}, numerical estimates of $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$, are required. Using the binomial formula, these integrals can be rewritten as follows:

$$a_{n_0,\hat{r}_g(\cdot)}(i) = \sum_{k=0}^{n_0-i}(-1)^k \binom{n_0-i}{k} \int s^{i+k}\hat{r}_g(s)\mathrm{d}s \tag{4.4}$$

and

$$b_{n_0,\hat{r}_g(\cdot)}(i,j) = \sum_{q=0}^{j-i}\sum_{p=0}^{n_0-j}(-1)^q\binom{j-i}{q}(-1)^p\binom{n_0-j}{p}\int\int_{s_1} s_1^{i+q}s_2^{j-i-q+p}\hat{r}_g(s_2)\hat{r}_g(s_1)\mathrm{d}s_2\mathrm{d}s_1. \tag{4.5}$$

Therefore a possible strategy to estimate $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$, could be based on numerical estimates of the terms in the right hand side of (4.4) and (4.5). However, from a practical point of view, this procedure presents the disadvantage of having to take into account approximation errors for plenty of terms, which finally leads to worse approximations of $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$. Therefore, in the simulation study, we estimate these quantities in a direct way.

Put Table 2 about here.

While the practical implementation of h^*_{CE} is very time consuming, there exists a way to make the simulation study faster for a pair of fixed sample sizes (n_0, n_1). The fact is that there are some computations that do not need to be carried out every time that a pair of samples are drawn because they only depend on the sample sizes but not

on the observations themselves. Therefore, carrying out these computations previously and using them with any pair of samples of the same sample sizes, (n_0, n_1), can lead to a considerable decrease in computing time. In the simulation study carried out here, we take advantage of this fact and proceed in a direct way. Table 3 collects the median CPU time in seconds required per trial for implementing the proposed bootstrap bandwidth selectors, h^*_{CE}, h^*_{MC}, h^*_{SUMC} and h^*_{SMC}, and the plug-in selectors, b_{3c} and h_{SJ_2}. These values reveal that the bootstrap selector h^*_{CE} requires at about one hour and 20 or 25 minutes per trial when the sample sizes are $n_0 = n_1 = 50$. The fastest implementation is achieved by the plug-in bandwidth selector proposed by Ćwik and Mielniczuk (1993), b_{3c}, which is followed by the plug-in selector, h_{SJ_2}, and the bootstrap selectors, h^*_{MC}, h^*_{SUMC} and h^*_{SMC}.

Put Table 3 about here.

From the simulation study carried out here (see results in Tables 1 and 2), we conclude that all the proposed bootstrap selectors improve the one proposed by Ćwik and Mielniczuk (1993). However, only two of them, h^*_{SUMC} and h^*_{SMC}, show a similar behaviour to the plug-in selector, h_{SJ_2}, with very good performance, studied in Molanes and Cao (2007). Sometimes, it is even observed a slight improvement over h_{SJ_2}. However, this is not always the case. These facts and the intensive computing time required for any of the bootstrap selectors compared to the time required for any of the plug-in selectors make h_{SJ_2} a good choice in this setting.

5. Proof of Theorem 2.1

Standard bias-variance decomposition of MSE gives:

$$MISE\left(\hat{r}_h\right) = \int \left[E\left[\hat{r}_h\left(t\right)\right] - r\left(t\right)\right]^2 \mathrm{d}t + \int Var\left[\hat{r}_h\left(t\right)\right] \mathrm{d}t. \tag{5.1}$$

For the first term, it is easy to check that

$$E\left[\hat{r}_h\left(t\right)\right] = E\left[K_h\left(t - F_{0n_0}\left(X_1\right)\right)\right] = E\left[E\left[K_h\left(t - F_{0n_0}\left(X_1\right)\right) / X_{01}, \ldots, X_{0n_0}\right]\right]$$

$$= E\left[\int K_h\left(t - F_{0n_0}\left(y\right)\right) f_1\left(y\right) \mathrm{d}y\right] = \int E\left[K_h\left(t - F_{0n_0}\left(y\right)\right) f_1\left(y\right)\right] \mathrm{d}y$$
$$= \sum_{i=0}^{n_0} K_h\left(t - \frac{i}{n_0}\right) \binom{n_0}{i} \int F_0\left(y\right)^i \left(1 - F_0\left(y\right)\right)^{n_0 - i} f_1\left(y\right) \mathrm{d}y$$
$$= \sum_{i=0}^{n_0} K_h\left(t - \frac{i}{n_0}\right) \binom{n_0}{i} \int s^i \left(1 - s\right)^{n_0 - i} r\left(s\right) \mathrm{d}s. \tag{5.2}$$

On the other hand, it is straightforward to prove that

$$\begin{aligned} Var\left[\hat{r}_h\left(t\right)\right] &= \frac{1}{n_1} Var\left[K_h\left(t - F_{0n_0}\left(X_1\right)\right)\right] \\ &\quad + \frac{n_1 - 1}{n_1} Cov\left[K_h\left(t - F_{0n_0}\left(X_{11}\right)\right), K_h\left(t - F_{0n_0}\left(X_{12}\right)\right)\right] \\ &= \frac{1}{n_1} E\left[K_h^2\left(t - F_{0n_0}\left(X_1\right)\right)\right] - \frac{1}{n_1} E^2\left[K_h\left(t - F_{0n_0}\left(X_1\right)\right)\right] \\ &\quad + \frac{n_1 - 1}{n_1} Var\left[\int K_h\left(t - F_{0n_0}\left(y\right)\right) f_1\left(y\right) \mathrm{d}y\right] \\ &= \frac{1}{n_1} E\left[K_h^2\left(t - F_{0n_0}\left(X_1\right)\right)\right] - E^2\left[K_h\left(t - F_{0n_0}\left(X_1\right)\right)\right] \\ &\quad + \frac{n_1 - 1}{n_1} E\left[\left(\int K_h\left(t - F_{0n_0}\left(y\right)\right) f_1\left(y\right) \mathrm{d}y\right)^2\right] \end{aligned} \tag{5.3}$$

In order to get a more explicit expression for the variance, we study the expectations in the right hand-side of the expression above.

The first expectation is

$$\begin{aligned} E\left[K_h^2\left(t - F_{0n_0}\left(X_1\right)\right)\right] &= E\left[E\left[K_h^2\left(t - F_{0n_0}\left(X_1\right)\right) / X_{01}, \ldots, X_{0n_0}\right]\right] \\ &= \int E\left[K_h^2\left(t - F_{0n_0}\left(y\right)\right)\right] f_1\left(y\right) \mathrm{d}y \\ &= \sum_{i=0}^{n_0} K_h^2\left(t - \frac{i}{n_0}\right) \binom{n_0}{i} \int F_0\left(y\right)^i \left(1 - F_0\left(y\right)\right)^{n_0 - i} f_1\left(y\right) \mathrm{d}y \\ &= \sum_{i=0}^{n_0} K_h^2\left(t - \frac{i}{n_0}\right) \binom{n_0}{i} \int s^i \left(1 - s\right)^{n_0 - i} r\left(s\right) \mathrm{d}s \end{aligned} \tag{5.4}$$

The last expectation can be written as

$$E\left[\left(\int K_h\left(t - F_{0n_0}\left(y\right)\right) f_1\left(y\right) \mathrm{d}y\right)^2\right]$$
$$= E\left[\int\int K_h\left(t - F_{0n_0}\left(y_1\right)\right) K_h\left(t - F_{0n_0}\left(y_2\right)\right) f_1\left(y_1\right) f_1\left(y_2\right) \mathrm{d}y_1 \mathrm{d}y_2\right] = 2A,$$

based on the symmetry of the integrand, where

$$\begin{aligned}
A &= \int\int_{y_2>y_1} E\left[K_h\left(t-F_{0n_0}\left(y_1\right)\right)K_h\left(t-F_{0n_0}\left(y_2\right)\right)\right] f_1\left(y_2\right) f_1\left(y_1\right) \mathrm{d}y_2 \mathrm{d}y_1 \\
&= \sum_{i=0}^{n_0}\sum_{\substack{j=0 \\ i\leq\ j}}^{n_0} K_h\left(t-\frac{i}{n_0}\right) K_h\left(t-\frac{j}{n_0}\right) \frac{n_0!}{i!\left(j-i\right)!\left(n_0-j\right)!} \\
&\quad \int\int_{y_2>y_1} F_0\left(y_1\right)^i \left(F_0\left(y_2\right)-F_0\left(y_1\right)\right)^{j-i} \left(1-F_0\left(y_2\right)\right)^{n_0-j} f_1\left(y_2\right) f_1\left(y_1\right) \mathrm{d}y_2 \mathrm{d}y_1 \\
&= \sum_{i=0}^{n_0}\sum_{\substack{j=0 \\ i\leq\ j}}^{n_0} K_h\left(t-\frac{i}{n_0}\right) K_h\left(t-\frac{j}{n_0}\right) \frac{n_0!}{i!\left(j-i\right)!\left(n_0-j\right)!} \\
&\quad \int\int_{s_2>s_1} s_1^i \left(s_2-s_1\right)^{j-i} \left(1-s_2\right)^{n_0-j} r\left(s_2\right) r\left(s_1\right) \mathrm{d}s_2 \mathrm{d}s_1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&E\left[\left(\int K_h\left(t-F_{0n_0}\left(y\right)\right) f_1\left(y\right) \mathrm{d}y\right)^2\right] \\
&= 2\sum_{i=0}^{n_0}\sum_{\substack{j=0 \\ i\leq\ j}}^{n_0} K_h\left(t-\frac{i}{n_0}\right) K_h\left(t-\frac{j}{n_0}\right) \frac{n_0!}{i!\left(j-i\right)!\left(n_0-j\right)!} \\
&\int\int_{s_2>s_1} s_1^i \left(s_2-s_1\right)^{j-i} \left(1-s_2\right)^{n_0-j} r\left(s_2\right) r\left(s_1\right) \mathrm{d}s_2 \mathrm{d}s_1. \qquad (5.5)
\end{aligned}$$

Using (5.4), (5.2) and (5.5) in (5.3) and (5.2) and (5.3) in (5.1) gives

$$\begin{aligned}
MISE\left(\hat{r}_h\right) = &\int\left(\sum_{i=0}^{n_0} K_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i} a_{n_0,r(\cdot)}\left(i\right)-r\left(t\right)\right)^2 \mathrm{d}t \\
&+\frac{1}{n_1}\sum_{i=0}^{n_0}\binom{n_0}{i} a_{n_0,r(\cdot)}\left(i\right)\int K_h^2\left(t-\frac{i}{n_0}\right)\mathrm{d}t \\
&-\int\left(\sum_{i=0}^{n_0} K_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i} a_{n_0,r(\cdot)}\left(i\right)\right)^2 \mathrm{d}t \\
&+2\frac{n_1-1}{n_1}\sum_{i=0}^{n_0}\sum_{j=i}^{n_0}\frac{n_0!}{i!\left(j-i\right)!\left(n_0-j\right)!} b_{n_0,r(\cdot)}\left(i,j\right) \\
&\int K_h\left(t-\frac{i}{n_0}\right) K_h\left(t-\frac{j}{n_0}\right)\mathrm{d}t
\end{aligned}$$

and consequently we get that

$$
\begin{aligned}
MISE\left(\hat{r}_h\right) = & \int \left(\sum_{i=0}^{n_0} K_h \left(t - \frac{i}{n_0} \right) C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right) - r\left(t\right) \right)^2 \mathrm{d}t \\
& + \frac{1}{n_1} \sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right) \left(K_h * K_h\right)\left(0\right) \\
& - \int \left(\sum_{i=0}^{n_0} K_h \left(t - \frac{i}{n_0} \right) C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right) \right)^2 \mathrm{d}t \\
& + 2 \frac{n_1 - 1}{n_1} \sum_{i=0}^{n_0} \sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,r(\cdot)}\left(i,j\right) \left(K_h * K_h\right) \left(\frac{j-i}{n_0} \right).
\end{aligned}
$$

Some simple algebra concludes the proof.

Table 1. Values of EM for b_{3c}, h_{SJ_2}, h^*_{CE}, h^*_{MC}, h^*_{SUMC} and h^*_{SMC} for models (a)-(g).

EM		Model						
(n_0, n_1)	Selector	(a)	(b)	(c)	(d)	(e)	(f)	(g)
$(50, 50)$	b_{3c}	0.3493	0.5446	0.3110	0.3110	1.2082	1.5144	0.7718
$(50, 50)$	h_{SJ_2}	0.1746	0.4322	0.1471	0.2439	0.5523	0.7702	0.5742
$(50, 50)$	h^*_{CE}	0.2791	0.5141	0.2095	0.2630	0.7905	1.0784	0.7675
$(50, 50)$	h^*_{MC}	0.2404	0.4839	0.1951	0.2911	0.8103	0.9545	0.7195
$(50, 50)$	h^*_{SUMC}	0.1719	0.3990	0.1473	0.2517	0.6392	0.7246	0.5917
$(50, 50)$	h^*_{SMC}	0.1734	0.3996	0.1486	0.2372	0.5984	0.7093	0.5918

Table 2. Values of EM for h_{SJ_2}, h^*_{SUMC} and h^*_{SMC} for models (a)-(g).

EM		Model						
(n_0, n_1)	Selector	(a)	(b)	(c)	(d)	(e)	(f)	(g)
$(50, 100)$	h_{SJ_2}	0.1660	0.3959	0.1256	0.2075	0.5329	0.7356	0.5288
$(50, 100)$	h^*_{SUMC}	0.1565	0.3733	0.1276	0.2076	0.5386	0.7199	0.5576
$(50, 100)$	h^*_{SMC}	0.1580	0.3716	0.1376	0.2015	0.5655	0.7296	0.5509
$(100, 50)$	h_{SJ_2}	0.1241	0.3319	0.1139	0.1897	0.3804	0.4833	0.3864
$(100, 50)$	h^*_{SUMC}	0.1291	0.3056	0.1020	0.1914	0.4403	0.5129	0.4310
$(100, 50)$	h^*_{SMC}	0.1297	0.3060	0.1021	0.1845	0.4324	0.5048	0.4393
$(100, 100)$	h_{SJ_2}	0.1208	0.2831	0.1031	0.1474	0.3717	0.4542	0.3509
$(100, 100)$	h^*_{SUMC}	0.1095	0.2718	0.0905	0.1467	0.3789	0.4894	0.3741
$(100, 100)$	h^*_{SMC}	0.1105	0.2712	0.0871	0.1499	0.3686	0.4727	0.3639
$(100, 150)$	h_{SJ_2}	0.1045	0.2528	0.0841	0.1318	0.3305	0.4648	0.3345
$(100, 150)$	h^*_{SUMC}	0.1136	0.2552	0.0951	0.1344	0.3719	0.4887	0.3528
$(100, 150)$	h^*_{SMC}	0.1147	0.2543	0.0863	0.1439	0.3814	0.4894	0.3477

Table 3. Median time (in seconds) required to compute a realization of b_{3c}, h_{SJ_2}, h^*_{CE}, h^*_{MC}, h^*_{SUMC} and h^*_{SMC} for models (a)-(d) for sample sizes $n_0 = n_1 = 50$.

EM	Model			
Selector	(a)	(b)	(c)	(d)
b_{3c}	0.0470	0.1560	0.0620	0.0300
h_{SJ_2}	4.7500	4.8440	4.8045	3.3795
h^*_{CE}	5063.5	5091.2	4740.0	5056.1
h^*_{MC}	180.3	177.4	177.3	181.6
h^*_{SUMC}	255.4	252.9	259.2	242.9
h^*_{SMC}	421.1	421.9	639.3	628.4

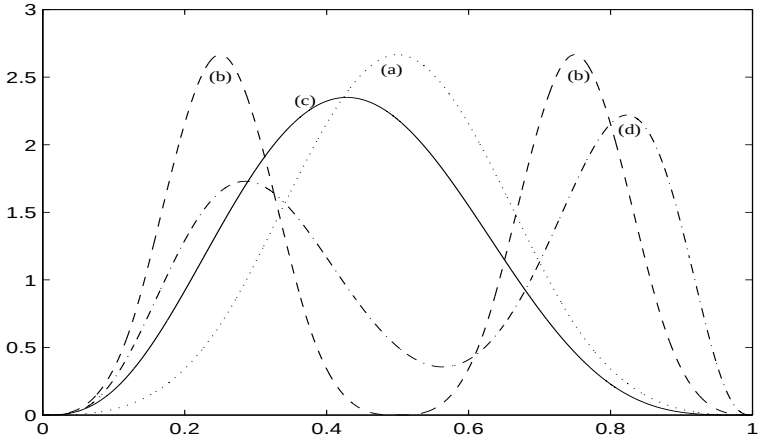

Fig. 1. Plots of the relative densities (a)-(d).

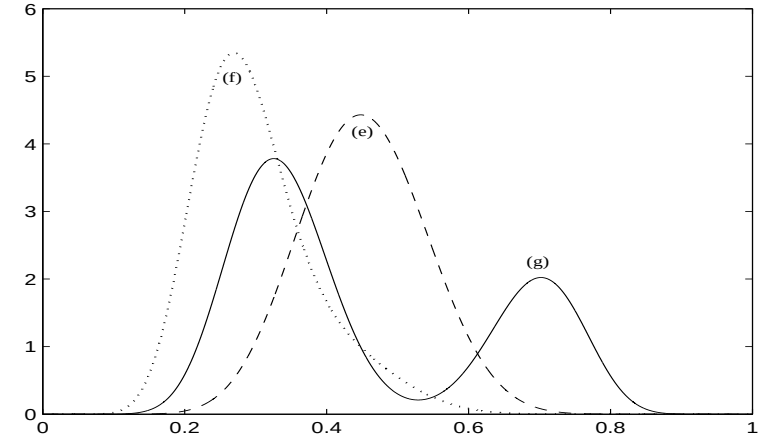

Fig. 2. Plots of the relative densities (e)-(g).

References

[1] Cao, R. (1993). Bootstrapping the mean integrated squared error, *Journal of Multivariate Analysis*, **45**, 137–160.

[2] Cao, R., Cuevas, A. and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation, *Computational Statistics & Data Analysis*, **17**, 153–176.

[3] Cao, R., Janssen, P. and Veraverbeke, N. (2000). Relative density estimation with censored data, *The Canadian Journal of Statistics*, **28**, 97–111.

[4] Cao, R., Janssen, P. and Veraverbeke, N. (2001). Relative density estimation and local bandwidth selection for censored data, *Computational Statistics & Data Analysis*, **36**, 497–510.

[5] Cao, R., Janssen, P. and Veraverbeke, N. (2005). Relative hazard rate estimation for right censored and left truncated data, *Test*, **14**, 257–280.

[6] Ćwik, J. and Mielniczuk, J. (1993). Data-dependent bandwidth choice for a grade density kernel estimate, *Statistics & Probability Letters*, **16**, 397–405.

[7] Handcock, M.S. and Janssen, P. (1996). Statistical inference for the relative distribution, *Technical Report, Department of Statistics and Operations Research, New York University, New York.*

[8] Handcock, M.S. and Janssen, P. (2002). Statistical inference for the relative density, *Sociological Methods & Research*, **30**, 394–424.

[9] Handcock, M.S. and Morris, M. (1999). Relative distribution methods in social sciences. Springer, New York.

[10] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation, *Journal of Nonparametric Statistics*, **16**, 709–729.

[11] Molanes, E.M. and Cao, R. (2006). Relative density estimation for right censored and left truncated data, *Technical Report, Departamento de Matemáticas, Universidade da Coruña, Spain*. Available at http://www.udc.es/dep/mate/Dpto_Matematicas/Investigacion/ie/ie_prepublicaciones.htm.

[12] Molanes, E.M. and Cao, R. (2007). Plug-in bandwidth selector for the kernel relative density estimator, To appear in *Annals of the Institute of Statistical Mathematics*.

[13] Parzen, E. (1962). On estimation of a probability density and mode, *Annals of Mathematical Statistics*, **33**, 1065–1076.

[14] Polansky, A.M. and Baker, E.R. (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation & Simulation*, **65**, 63–80.

[15] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Annals of Mathematical Statistics*, **27**, 642–669.

[16] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics — Theory and Methods*, **14**, 1123–1136.

[17] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Acknowledgements

Research supported by Grants BES-2003-1170 (EU ESF support included) for the first author, XUGA PGIDIT03PXIC10505-PN for the second author and MTM2005-00429 (EU ERDF support included) for both authors.

Affiliation

[1]Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain.