AISM (2008) 60:273–300
DOI 10.1007/s10463-006-0108-y

# Plug-in bandwidth selector for the kernel relative density estimator

**Elisa María Molanes-López · Ricardo Cao**

Received: 31 March 2005 / Revised: 31 August 2006 / Published online: 31 January 2007


**Abstract** This paper is focused on two kernel relative density estimators in a two-sample problem. An asymptotic expression for the mean integrated squared error of these estimators is found and, based on it, two solve-the-equation plug-in bandwidth selectors are proposed. In order to examine their practical performance a simulation study and a practical application to a medical dataset are carried out.

**Keywords** Kernel-type estimates · Smoothing parameter · Solve-the-equation rules · Survival analysis · Two-sample problem

## 1 Introduction

The study of differences among groups or changes over time is a goal in fields such as medical research and social science research. The traditional method for this purpose is the usual parametric location and scale analysis. However, this is a very restrictive tool, since a lot of the information available in the data is unaccessible. In order to make a better use of this information it is convenient to focus on distributional analysis, i.e., on the general two-sample problem of comparing the cumulative distribution functions (cdf), $F_0$ and $F$, of two random variables, $X_0$ and $X$. Useful tools for this purpose are the relative distribution function, $R(t)$, and the relative density function, $r(t)$, of $X$ with respect to (w.r.t.)

E. M. Molanes-López (✉) · R. Cao
Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña,
Campus de Elviña s/n, 15071 A Coruña, Spain
e-mail: emolanes@udc.es

R. Cao
e-mail: rcao@udc.es

$X_0$:

$$R(t) = P(F_0(X) \leq t) = F\left(F_0^{-1}(t)\right), \quad 0 < t < 1,$$

where $F_0^{-1}(t) = \inf\{x \,|F_0(x) \geq t\}$ denotes the quantile function of $F_0$ and

$$r(t) = R^{(1)}(t) = \frac{f\left(F_0^{-1}(t)\right)}{f_0\left(F_0^{-1}(t)\right)}, \quad 0 < t < 1,$$

where $f$ and $f_0$ are the densities pertaining to $F$ and $F_0$, respectively. These two curves, as well as estimators for them, have been studied by Gastwirth (1968), Ćwik and Mielniczuk (1993), Hsieh (1995), Hsieh and Turnbull (1996), Cao et al. (2000, 2001) and Handcock and Janssen (2002).

These functions, $R$ and $r$, are closely related to other statistical methods. The ROC curve, used in the evaluation of the performance of medical tests for separating two groups, is related to $R$ through the relationship $ROC(t) = 1-R(1-t)$ [see, for instance, Holmgren (1996) and Li et al. (1996) for details] and the density ratio $h(x) = \frac{f(x)}{f_0(x)}, x \in \mathbb{R}$, used by Silverman (1978) is linked to $r$ through $h(x) = r(F_0(x))$.

Throughout the paper, we will focus on two kernel-type estimators of $r$, similar to the one already proposed by Ćwik and Mielniczuk (1993). In the following section we will give some notation and obtain an asymptotic representation for the MISE of the relative density estimators. This is a difference with respect to Ćwik and Mielniczuk (1993), where an asymptotic expression for the MISE of only the dominant part of the estimator was found. Section 3 is concerned with automatic global bandwidth selection. Two solve-the-equation plug-in bandwidth selectors based on the ideas by Sheather and Jones (1991) are proposed. A simulation study is shown in Sect. 4 where the performance of the data-driven selectors proposed in this paper is compared with the selector proposed in Ćwik and Mielniczuk (1993). A medical application is presented in Sect. 5. Finally, the proofs of the results presented in Sects. 2 and 3 are included in Sect. 6.

## 2 Kernel relative density estimators

Consider the two-sample problem with completely observed data:

$$\{X_{01}, \ldots, X_{0n}\}, \quad \{X_1, \ldots, X_m\},$$

where the $X_{0i}$'s are independent and identically distributed as $X_0$; and the $X_j$'s are independent and identically distributed as $X$. These two sequences are independent each other.

Throughout this paper all the asymptotic results are obtained if both sample sizes $m$ and $n$ tend to infinity in such a way that, for some constant $0 < \lambda < \infty$,

$$\lim_{m \longrightarrow \infty} \frac{m}{n} = \lambda.$$

We assume the following conditions on the underlying distributions, the kernels $K$ and $M$ and the bandwidths $h$ and $h_0$ to be used in the estimators (see (1), (2) and (3) below):

(F1) $F_0$ and $F$ have continuous density functions, $f_0$ and $f$, respectively.
(F2) $f_0$ is a three times differentiable density function with $f_0^{(3)}$ bounded.
(R1) $r$ is a twice continuously differentiable density with compact support contained in $[0, 1]$.
(K1) $K$ is a symmetric four times differentiable density function with compact support $[-1, 1]$ and $K^{(4)}$ bounded.
(K2) $M$ is a symmetric density and continuous function except at a finite set of points.
(B1) $h \longrightarrow 0$ and $mh^3 \longrightarrow \infty$.
(B2) $h_0 \longrightarrow 0$ and $nh_0^4 \longrightarrow 0$.

Since $\frac{1}{h} \int_0^1 K\left(\frac{t-z}{h}\right) \mathrm{d}R(z)$ is close to $r(t)$ and for smooth distributions it is satisfied that:

$$\frac{1}{h} \int_0^1 K\left(\frac{t-z}{h}\right) \mathrm{d}R(z) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{t - F_0(z)}{h}\right) \mathrm{d}F(z),$$

a natural way to define a kernel-type estimator of $r(t)$ is to replace the unknown functions $F_0$ and $F$ by some appropriate estimators. We consider two proposals:

$$\hat{r}_h(t) = \int_{-\infty}^{\infty} K_h(t - F_{0n}(z)) \, \mathrm{d}F_m(z) = \frac{1}{m} \sum_{j=1}^{m} K_h\left(t - F_{0n}(X_j)\right) \tag{1}$$

and

$$\hat{r}_{h,h_0}(t) = \int_{-\infty}^{\infty} K_h\left(t - \tilde{F}_{0n}(z)\right) \mathrm{d}F_m(z) = \frac{1}{m} \sum_{j=1}^{m} K_h\left(t - \tilde{F}_{0n}(X_j)\right), \tag{2}$$

where $K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$, $K$ is a kernel function, $h$ is the bandwidth used to estimate $r$, $F_{0n}$ and $F_m$ are the empirical distribution functions based on $X_{0i}$'s and $X_j$'s, respectively, and $\tilde{F}_{0n}$ is a kernel-type estimate of $F_0$ given by:

$$\tilde{F}_{0n} = n^{-1} \sum_{i=1}^{n} \mathbb{M}\left(\frac{x - X_{0i}}{h_0}\right) \tag{3}$$

where $\mathbb{M}$ denotes the cdf of the kernel $M$ and $h_0$ is the bandwidth used to estimate $F_0$.

Using a Taylor expansion, $\hat{r}_h(t)$ can be written as follows:

$$\begin{aligned}\hat{r}_h(t) &= \int_{-\infty}^{\infty} K_h\left(t - F_0(z)\right) \mathrm{d}F_m(z) \\ &\quad + \int_{-\infty}^{\infty} K_h^{(1)}\left(t - F_0(z)\right)\left(F_0(z) - F_{0n}(z)\right) \mathrm{d}F_m(z) \\ &\quad + \int_{-\infty}^{\infty} \left(F_0(z) - F_{0n}(z)\right)^2 \\ &\quad \times \int_0^1 (1-s) K_h^{(2)}\left(t - F_0(z) - s\left(F_{0n}(z) - F_0(z)\right)\right) \mathrm{d}s\, \mathrm{d}F_m(z).\end{aligned}$$

Let us define $\tilde{U}_n = F_{0n} \circ F_0^{-1}$ and $\tilde{R}_m = F_m \circ F_0^{-1}$. Then, $\hat{r}_h(t)$ can be rewritten in a useful way for the study of its mean integrated squared error (MISE):

$$\hat{r}_h(t) = \tilde{r}_h(t) + A_1 + A_2 + B, \tag{4}$$

where

$$\begin{aligned}\tilde{r}_h(t) &= \int_{-\infty}^{\infty} K_h\left(t - F_0(z)\right) \mathrm{d}F_m(z) = \frac{1}{m}\sum_{j=1}^{m} K_h\left(t - F_0(X_j)\right) \\ A_1 &= \int_0^1 \left(v - \tilde{U}_n(v)\right) K_h^{(1)}(t - v)\, \mathrm{d}\left(\tilde{R}_m - R\right)(v) \\ A_2 &= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left(F_0(w) - 1\left\{X_{0i} \le w\right\}\right) K_h^{(1)}\left(t - F_0(w)\right) \mathrm{d}F(w) \\ B &= \int_{-\infty}^{\infty} \left(F_0(z) - F_{0n}(z)\right)^2 \times \int_0^1 (1-s) K_h^{(2)}\left(t - F_0(z) - s\left(F_{0n}(z) - F_0(z)\right)\right) \\ &\quad \mathrm{d}s\mathrm{d}F_m(z).\end{aligned}$$

Proceeding in a similar way, we can rewrite $\hat{r}_{h,h_0}$ as follows:

$$\hat{r}_{h,h_0}(t) = \tilde{r}_h(t) + A_1 + A_2 + \hat{A} + \hat{B}, \tag{5}$$

where

$$\begin{aligned}\hat{A} &= \int \left(F_{0n}(w) - \tilde{F}_{0n}(w)\right) K_h^{(1)}\left(t - F_0(w)\right) \mathrm{d}F_m(w) \\ \hat{B} &= \int_{-\infty}^{\infty} \left(F_0(z) - \tilde{F}_{0n}(z)\right)^2 \int_0^1 (1-s) K_h^{(2)} \times \left(t - F_0(z) - s\left(\tilde{F}_{0n}(z) - F_0(z)\right)\right) \\ &\quad \times \mathrm{d}s\, \mathrm{d}F_m(z).\end{aligned}$$

Our main result is an asymptotic representation for the MISE of $\hat{r}_h(t)$ and $\hat{r}_{h,h_0}(t)$. With the purpose of simplifying the exposition of the results obtained, from here on we will denote $C(g) = \int_{-\infty}^{\infty} g^2(x)\,\mathrm{d}x$, for any square integrable function $g$.

**Theorem 1** (AMISE) *Assume conditions* (*F1*), (*R1*), (*K1*) *and* (*B1*). *Then*

$$\mathrm{MISE}(\hat{r}_h) = \mathrm{AMISE}(h) + o\left(\frac{1}{mh} + h^4\right) + o\left(\frac{1}{nh}\right),$$

$$\mathrm{AMISE}(h) = \frac{1}{mh}C(K) + \frac{1}{4}h^4 d_K^2 C(r^{(2)}) + \frac{1}{nh}C(r)C(K),$$

*where* $d_K = \int_{-1}^{1} x^2 K(x)\,\mathrm{d}x$.

*If the conditions* ($F2$), ($K2$), *and* ($B2$) *are assumed as well, then the same result is satisfied for the* $\mathrm{MISE}(\hat{r}_{h,h_0})$.

*Remark 1* From Theorem 1 it follows that the optimal bandwidth, minimizing the asymptotic mean integrated squared error of any of the estimators considered for $r$, is given by

$$h_{\mathrm{AMISE}} = \left(\frac{C(K)(\lambda C(r)+1)}{d_K^2 C(r^{(2)})m}\right)^{\frac{1}{5}}. \tag{6}$$

*Remark 2* Note that AMISE($h$) derived from Theorem 1 does not depend on the bandwidth $h_0$. A more higher-order analysis should be considered to address simultaneously the bandwidth selection problem of $h$ and $h_0$.

## 3 Bandwidth selectors

### 3.1 Estimation of density functionals

It is very simple to show that, under sufficiently smooth conditions on $r$ ($r \in C^{(2\ell)}(\mathbb{R})$), the functionals

$$C\left(r^{(\ell)}\right) = \int_0^1 \left(r^{(\ell)}(x)\right)^2 \mathrm{d}x \tag{7}$$

appearing in (6), are related to other general functionals of $r$, denoted by $\Psi_{2\ell}$:

$$C\left(r^{(\ell)}\right) = (-1)^{\ell} \int_0^1 r^{(2\ell)}(x)\, r(x)\,\mathrm{d}x = (-1)^{\ell}\,\Psi_{2\ell}, \tag{8}$$

where

$$\Psi_{\ell} = \int_0^1 r^{(\ell)}(x)\, r(x)\,\mathrm{d}x = E\left[r^{(\ell)}(F_0(X))\right].$$

The equation above suggests a natural kernel-type estimator for $\Psi_\ell$ as follows

$$\hat{\Psi}_\ell(g) = \frac{1}{m}\sum_{j=1}^{m}\left[\sum_{k=1}^{m}\frac{1}{m}L_g^{(\ell)}\left(F_{0n}\left(X_j\right) - F_{0n}\left(X_k\right)\right)\right], \tag{9}$$

where $L$ is a kernel function and $g$ is a smoothing parameter called pilot bandwidth. Likewise in the previous section, this is not the unique possibility and we could consider another estimator of $\Psi_\ell$,

$$\tilde{\Psi}_\ell(g) = \frac{1}{m}\sum_{j=1}^{m}\left[\sum_{k=1}^{m}\frac{1}{m}L_g^{(\ell)}\left(\tilde{F}_{0n}\left(X_j\right) - \tilde{F}_{0n}\left(X_k\right)\right)\right], \tag{10}$$

where $F_{0n}$ in (9) is replaced by $\tilde{F}_{0n}$. Since the difference between both estimators decreases as $h_0$ tends to zero, it is expected to obtain the same theoretical results for both estimators. Therefore, we will only show theoretical results for $\hat{\Psi}_\ell(g)$.

We will obtain the asymptotic mean squared error of $\hat{\Psi}_\ell(g)$ under the following assumptions.

(R2) The relative density $r \in C^{(\ell+6)}(\mathbb{R})$.

(K3) The kernel $L$ is a symmetric kernel of order 2, $L \in C^{(\ell+7)}(\mathbb{R})$ and satisfies that $(-1)^{\frac{\ell}{2}+2} L^{(\ell)}(0)\, d_L > 0$, $L^{(\ell)}(1) = L^{(\ell+1)}(1) = 0$, with $d_L = \int_{-\infty}^{\infty} x^2 L(x)\mathrm{d}x$.

(B3) $g = g_m$ is a positive-valued sequence of bandwidths satisfying

$$\lim_{m\longrightarrow\infty} g = 0 \quad \text{and} \quad \lim_{m\longrightarrow\infty} mg^{\max\{\alpha,\beta\}} = \infty,$$

where

$$\alpha = \frac{2(\ell+7)}{5}, \quad \beta = \frac{1}{2}(\ell+1) + 2.$$

Condition (R2) implies a smooth behaviour of $r$ in the boundary of its support, contained in $[0,1]$. If this smoothness fails, the quantity $C\left(r^{(\ell)}\right)$ could be still estimated through its definition, using a kernel estimation for $r^{(\ell)}$ [see Hall and Marron (1987) for the one-sample problem setting]. Condition (K3) can only hold for even $\ell$. Observe that in condition (B3) for even $\ell$, $\max\{\alpha,\beta\} = \alpha$ for $\ell = 0, 2$ and $\max\{\alpha,\beta\} = \beta$ for $\ell = 4, 6, \ldots$

**Theorem 2** *Assume conditions* (*F1*)*,* (*R2*)*,* (*K3*) *and* (*B3*)*. Then it follows that*

$$\begin{aligned}\mathrm{MSE}\left(\hat{\Psi}_\ell(g)\right) &= \left[\frac{1}{mg^{\ell+1}}L^{(\ell)}(0)(1+\lambda\Psi_0)+\frac{1}{2}d_L\Psi_{\ell+2}g^2+O\left(g^4\right)\right.\\&\left.+o\left(\left(ng^{\ell+1}\right)^{-1}\right)\right]^2+\frac{2}{m^2g^{2\ell+1}}\Psi_0C\left(L^{(\ell)}\right)\\&+o\left(\left(m^2g^{2\ell+1}\right)^{-1}\right)+O\left(n^{-1}\right).\end{aligned} \quad (11)$$

*Remark 3* The first term in the right-hand side of (11) corresponds to the squared bias term of MSE. Note that, using (K3) and (8), the main bias term can be made to vanish by choosing $g$ as $g_\ell$

$$g_\ell = \left(\frac{2L^{(\ell)}(0)(\lambda\Psi_0+1)}{-d_L\Psi_{\ell+2}m}\right)^{\frac{1}{(\ell+3)}} = \left(\frac{2L^{(\ell)}(0)d_K^2\Psi_4}{-d_L\Psi_{\ell+2}C(K)}\right)^{\frac{1}{\ell+3}} h_{\mathrm{AMISE}}^{\frac{5}{\ell+3}}.$$

### 3.2 STE rules based on Sheather and Jones ideas

As in the context of ordinary density estimation, the practical implementation of the kernel-type estimators proposed here (see (1) and (2)), requires the choice of the smoothing parameter $h$. Our two proposals, $h_{SJ_1}$ and $h_{SJ_2}$, as well as the selector $b_{3c}$ recommended by Ćwik and Mielniczuk (1993), are modifications of Sheather and Jones (1991). Since the Sheather and Jones selector is the solution of an equation in the bandwidth, it is also known as a solve-the-equation (STE) rule. Motivated by the formula (6) for the AMISE-optimal bandwidth and the relation (8), solve-the-equation rules require that $h$ is chosen to satisfy the relationship

$$h = \left(\frac{C(K)\left(\lambda\tilde{\Psi}_0(\gamma_1(h))+1\right)}{d_K^2\cdot\tilde{\Psi}_4(\gamma_2(h))\cdot m}\right)^{\frac{1}{5}},$$

where the pilot bandwidths for the estimation of $\Psi_0$ and $\Psi_4$ are functions of $h$ ($\gamma_1(h)$ and $\gamma_2(h)$, respectively).

Motivated by Remark 3, we suggest taking

$$\gamma_1(h) = \left(\frac{2\cdot L(0)\cdot d_K^2\cdot\tilde{\Psi}_4(g_4)}{-d_L\tilde{\Psi}_2(g_2)C(K)}\right)^{\frac{1}{3}} h^{\frac{5}{3}}$$

and

$$\gamma_2(h) = \left( \frac{2 \cdot L^{(4)}(0) \cdot d_K^2 \cdot \tilde{\Psi}_4(g_4)}{-d_L \tilde{\Psi}_6(g_6)\, C(K)} \right)^{\frac{1}{7}} h^{\frac{5}{7}},$$

where $\tilde{\Psi}_j(\cdot)$, $(j = 0, 2, 4, 6)$ are kernel estimates (10). Note that this way of proceeding leads us to a never ending process in which a bandwidth selection problem must be solved at every stage. To make this iterative process feasible in practice one possibility is to propose a stopping stage in which the unknown quantities are estimated using a parametric scale for $r$. This strategy is known in the literature as the stage selection problem (see Wand and Jones, 1995). While the selector $b_{3c}$ in Ćwik and Mielniczuk (1993) used a Gaussian scale, now for the implementation of $h_{SJ_2}$, we will use a mixture of betas based on the Weierstrass approximation theorem and Bernstein polynomials associated to any continuous function on $[0, 1]$ (see Kakizawa, 2004, and references therein for the motivation of this method). Later on we will show the formula for computing the reference scale above-mentioned, together with the selector $b_{3c}$ we used in Sect. 4.

In the following we denote the Epanechnikov kernel by $K$, the uniform density in $[-1, 1]$ by $M$ and we define $L$ as follows

$$L(x) = \frac{\Gamma(18)}{2\Gamma(9)\Gamma(9)} \left( \frac{x+1}{2} \right)^8 \left( 1 - \frac{x+1}{2} \right)^8 1_{\{-1 \leq x \leq 1\}}.$$

Next, we detail the steps required in the implementation of $h_{SJ_2}$.

**Step 1.** Obtain $\hat{\Psi}_j^{PR}$ $(j = 0, 4, 6, 8)$, parametric estimates for $\Psi_j$ $(j = 0, 4, 6, 8)$, with the replacement of $r$ in $C\left(r^{(j/2)}\right)$ (see (7)), by a mixture of betas, $b(x)$, as it will be explained later on (see (12)).

**Step 2.** Compute kernel estimates for $\Psi_j$ $(j = 2, 4, 6)$, by using $\tilde{\Psi}_j(g_j^{PR})$ $(j = 2, 4, 6)$, with

$$g_j^{PR} = \left( \frac{2 \cdot L^{(j)}(0) \left( \lambda \hat{\Psi}_0^{PR} + 1 \right)}{-d_L \cdot \hat{\Psi}_{j+2}^{PR} \cdot m} \right)^{\frac{1}{j+3}}, \quad (j = 2, 4, 6).$$

**Step 3.** Estimate $\Psi_0$ and $\Psi_4$, using (10), by means of $\tilde{\Psi}_0(\hat{\gamma}_1(h))$ and $\tilde{\Psi}_4(\hat{\gamma}_2(h))$, where

$$\hat{\gamma}_1(h) = \left( \frac{2 \cdot L(0) \cdot d_K^2 \cdot \tilde{\Psi}_4\left(g_4^{PR}\right)}{-d_L \tilde{\Psi}_2\left(g_2^{PR}\right) C(K)} \right)^{\frac{1}{3}} h^{\frac{5}{3}}$$

and

$$\hat{\gamma}_2(h) = \left( \frac{2 \cdot L^{(4)}(0) \cdot d_K^2 \cdot \tilde{\Psi}_4\left(g_4^{PR}\right)}{-d_L \tilde{\Psi}_6\left(g_6^{PR}\right) C(K)} \right)^{\frac{1}{7}} h^{\frac{5}{7}}.$$

**Step 4.** Select the bandwidth $h_{SJ_2}$ as the one that solves the following equation in $h$:

$$h = \left( \frac{C(K)\left(\lambda \tilde{\Psi}_0\left(\hat{\gamma}_1(h)\right) + 1\right)}{d_K^2 \cdot \tilde{\Psi}_4\left(\hat{\gamma}_2(h)\right) \cdot m} \right)^{\frac{1}{5}}.$$

In order to solve the equation above, it will be necessary to use a numerical algorithm. In the simulation study we will use the false-position method. The main reason is that the false-position algorithm does not require the computation of the derivatives, what simplifies considerably the implementation of the proposed bandwidth selectors. At the same time, this algorithm presents some advantages over others because it tries to combine the speed of methods such as the secant method with the security afforded by the bisection method.

Unlike the Gaussian parametric reference, used to obtain $b_{3c}$, the selector $h_{SJ_2}$ uses in Step 1 a mixture of betas as follows:

$$b(x) = \sum_{j=1}^{N} \left( \tilde{R}_{n,m}\left(\frac{j}{N}\right) - \tilde{R}_{n,m}\left(\frac{j-1}{N}\right) \right) \beta(x, j, N-j+1), \tag{12}$$

where

$$\tilde{R}_{n,m}(x) = m^{-1} \textstyle\sum_{j=1}^{m} \mathbb{M}\left(\frac{x - \tilde{F}_{0n}(X_j)}{g}\right), \tag{13}$$

$$g = \left( \frac{2\int_{-\infty}^{\infty} x M(x) \mathbb{M}(x) \mathrm{d}x}{m d_M^2 C(r^{(1)})} \right)^{\frac{1}{3}}, \tag{14}$$

$\beta(x, a, b)$ stands for 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 $N$ is the number of betas in the mixture.

Since we are trying to estimate a density with support in $[0, 1]$ it seems more suitable to consider a parametric reference with this support. A mixture of betas is an appropriate option because it is flexible enough to model a large variety of relative densities, when derivatives of order 1, 3 and 4 are also required.

Note that, for the sake of simplicity, we are using above the AMISE-optimal bandwidth ($g$) for estimating a distribution function in the setting of a one-sample problem [see Polansky and Baker (2000) for more details in the kernel-type estimate of a distribution function]. The use of this bandwidth requires the previous estimation of the unknown functional, $C\left(r^{(1)}\right)$. We will consider a quick and dirty method, the rule of thumb, that uses a parametric reference for $r$ to estimate the above-mentioned unknown quantity. More specifically, our reference scale will be a beta with parameters $(p, q)$ estimated from the smoothed relative sample $\left\{\tilde{F}_{0n}(X_j)\right\}_{j=1}^{m}$, using the method of moments.

Following the same ideas as for (13) and (14), the bandwidth selector used for the kernel-type estimator $\tilde{F}_{0n}$ introduced in (3) is based on the AMISE-optimal bandwidth in the one-sample problem:

$$h_0 = \left( \frac{2\int_{-\infty}^{\infty} xM(x)\,\mathbb{M}(x)\,\mathrm{d}x}{nd_M^2 C\left(f_0^{(1)}\right)} \right)^{\frac{1}{3}}.$$

As it was already mentioned above, in most of the cases this methodology will be applied to survival analysis, so it is natural to assume that our samples come from distributions with support on the positive real line. Therefore, a gamma reference distribution, Gamma$(\alpha, \beta)$, has been considered, where the parameters $(\alpha, \beta)$ are estimated from the smoothed relative sample $\left\{\tilde{F}_{0n}(X_j)\right\}_{j=1}^{m}$, using the method of moments.

For the implementation of $h_{SJ_1}$, we proceed analogously to that of $h_{SJ_2}$ above. The only difference now is that throughout the previous discussion, $\tilde{\Psi}_j(\cdot)$ and $\tilde{F}_{0n}(\cdot)$ are replaced by, respectively, $\hat{\Psi}_j(\cdot)$ and $F_{0n}(\cdot)$.

As a variant of the selector that Ćwik and Mielniczuk (1993) proposed, $b_{3c}$ is obtained as the soluction to the following equation:

$$b_{3c} = \left( \frac{C(K)\left(1 + \lambda\hat{\Psi}_0(a)\right)}{d_K^2 \hat{\Psi}_4\left(\alpha_2\left(b_{3c}\right)\right) m} \right)^{\frac{1}{5}},$$

where $a = 1.781\hat{\sigma}m^{-\frac{1}{3}}$, $\hat{\sigma} = \min\left\{s_m, \widehat{IQR}/1.349\right\}$, $s_m$ and $\widehat{IQR}$ denote, respectively, the empirical standard deviation and the sample interquartile range of the relative data, $\left\{F_{0n}(X_j)\right\}_{j=1}^{m}$,

$$\alpha_2\left(b_{3c}\right) = 0.7694\left( \frac{\hat{\Psi}_4\left(g_4^{GS}\right)}{-\hat{\Psi}_6\left(g_6^{GS}\right)} \right)^{\frac{1}{7}} b_{3c}^{\frac{5}{7}},$$

where $GS$ stands for standard Gaussian scale,

$$g_4^{GS} = 1.2407\hat{\sigma}m^{-\frac{1}{7}}, \quad g_6^{GS} = 1.2304\hat{\sigma}m^{-\frac{1}{9}}$$

and the estimates $\hat{\Psi}_j$ (with $j = 0, 4, 6$) were obtained using (9), with $L$ replaced by the standard Gaussian kernel and with data driven bandwidth selectors derived from reducing the two-sample problem to a one-sample problem.

It is interesting to note that all the kernel-type estimators presented previously ($\hat{r}_h(t), \hat{r}_{h,h_0}(t), \tilde{R}_{n,m}(x)$ and $\tilde{F}_{0n}(x)$) were not corrected to take into account, respectively, the fact that $r$ and $R$ have support on $[0, 1]$ instead of on the whole real line, and the fact that $f_0$ is supported only on the positive real line. Therefore, in order to correct the boundary effect in practical applications we will use the well known reflecting method to modify $\hat{r}_h(t)$, $\hat{r}_{h,h_0}(t)$, $\tilde{R}_{n,m}(x)$ and $\tilde{F}_{0n}(x)$, where needed.

## 4 Simulations

We compare, through a simulation study, the performance of the bandwidth selectors $h_{SJ_1}$ and $h_{SJ_2}$, proposed in Sect. 3, with the standard competitor $b_{3c}$ recommended by Ćwik and Mielniczuk (1993). Although we are aware that the smoothing parameter $N$ introduced in (12) should be selected by some optimal way based on the data, this issue goes beyond the scope of this article. Consequently, from here on, we will consider $N = 14$ components in the beta mixture reference scale model given by (12).

We will consider the first sample coming from the random variate $X_0 = W^{-1}(U)$ and the second sample coming from the random variate $X = W^{-1}(S)$, where $U$ denotes a uniform distribution in the compact interval $[0, 1]$, $W$ is the distribution function of a Weibull distribution with parameters $(2, 3)$ and $S$ is a random variate from one of the following five different populations (see Fig. 1):

(a) A beta distribution with parameters 14 and 17 ($\beta(14, 17)$).
(b) A mixture consisting of $V_1$ with probability $\frac{4}{5}$ and $V_2$ with probability $\frac{1}{5}$, where $V_1 = \beta(14, 37)$ and $V_2 = \beta(14, 20)$.
(c) A mixture consisting of $V_1$ with probability $\frac{1}{3}$ and $V_2$ with probability $\frac{2}{3}$, where $V_1 = \beta(34, 15)$ and $V_2 = \beta(15, 30)$.

Choosing different values for the pair of sample sizes $m$ and $n$ and under each of the models presented above, we start drawing 500 pair of random samples and, according to every method, we select the bandwidths $\hat{h}$. Then, in order to check their performance we approximate by Monte Carlo the mean integrated squared error, $EM$, between the true relative density and the kernel-type estimate for $r$, given by (1) when $\hat{h} = b_{3c}, h_{SJ_1}$ or by (2) when $\hat{h} = h_{SJ_2}$.

The computation of the kernel-type estimations can be very time consuming by using a direct algorithm. Therefore, we will use linear binned approximations that, thanks to their discrete convolution structures, can be fast computed by using the fast Fourier transform (FFT) (see Wand and Jones 1995 for more details).

For all the models, the values of this criterion for the three bandwidth selectors, $h_{SJ_1}$, $h_{SJ_2}$ and $b_{3c}$, can be found in Table 1.

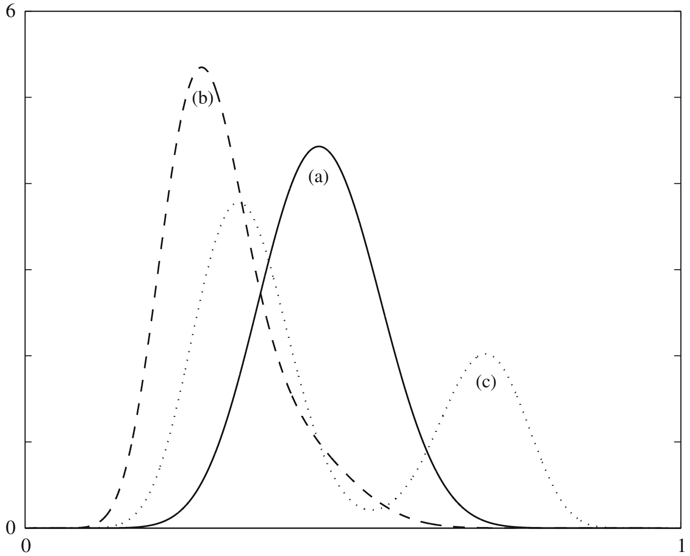


**Fig. 1** Plots of the relative densities (a)–(c)

**Table 1** Values of *EM* for $h_{SJ_1}$, $h_{SJ_2}$ and $b_{3c}$ for models (a)–(c)

| EM | Model (a) | | | Model (b) | | | Model (c) | | |
|---|---|---|---|---|---|---|---|---|---|
| $(n,m)$ | $h_{SJ_1}$ | $h_{SJ_2}$ | $b_{3c}$ | $h_{SJ_1}$ | $h_{SJ_2}$ | $b_{3c}$ | $h_{SJ_1}$ | $h_{SJ_2}$ | $b_{3c}$ |
| $(50,50)$ | 0.8437 | 0.5523 | 1.2082 | 1.1278 | 0.7702 | 1.5144 | 0.7663 | 0.5742 | 0.7718 |
| $(100,100)$ | 0.5321 | 0.3717 | 0.6654 | 0.6636 | 0.4542 | 0.7862 | 0.4849 | 0.3509 | 0.4771 |
| $(200,200)$ | 0.2789 | 0.2000 | 0.3311 | 0.4086 | 0.2977 | 0.4534 | 0.2877 | 0.2246 | 0.2830 |
| $(100,50)$ | 0.5487 | 0.3804 | 0.7162 | 0.6917 | 0.4833 | 0.8796 | 0.4981 | 0.3864 | 0.4982 |
| $(200,100)$ | 0.3260 | 0.2443 | 0.3949 | 0.4227 | 0.3275 | 0.4808 | 0.3298 | 0.2601 | 0.3252 |
| $(400,200)$ | 0.1739 | 0.1346 | 0.1958 | 0.2530 | 0.1924 | 0.2731 | 0.1830 | 0.1490 | 0.1811 |
| $(50,100)$ | 0.8237 | 0.5329 | 1.1189 | 1.1126 | 0.7356 | 1.4112 | 0.7360 | 0.5288 | 0.7135 |
| $(100,200)$ | 0.5280 | 0.3627 | 0.6340 | 0.6462 | 0.4288 | 0.7459 | 0.4568 | 0.3241 | 0.4449 |
| $(200,400)$ | 0.2738 | 0.1923 | 0.3192 | 0.3926 | 0.2810 | 0.4299 | 0.2782 | 0.2099 | 0.2710 |

A careful look at the table points out that the new selector $h_{SJ_2}$ presents a much better behaviour than the selector $b_{3c}$, especially when the sample sizes are equal or when $m$ is larger than $n$. The improvement is even larger for unimodal relative densities (model (a) and (b)). On the other hand, it is observed that the other proposal, $h_{SJ_1}$, presents only a moderate improvement over $b_{3c}$ for unimodal relative densities (model (a) and (b)) and performs only slightly better or even worse than $b_{3c}$ for bimodal relative densities (model (c)). The ratio $\frac{m}{n}$ produces an important effect on the behaviour of any of the three selectors considered. For instance, it is clearly seen an asymmetric behaviour of the selectors in terms of the sample sizes.

Other proposals for selecting $h$ have been investigated. For instance, versions of $h_{SJ_2}$ were considered in which either the unknown functionals $\Psi$ are estimated from the viewpoint of a one-sample problem or the STE rule is modified in such a way that only the function $\hat{\gamma}_2$ is considered in the equation to be solved (see Step 4 in Sect. 3). After a simulation study similar to the one detailed here, but now carried out for these versions of $h_{SJ_2}$, it was observed a similar practical performance to that observed for $h_{SJ_2}$. However, a worse behaviour was observed when, in the implementation of these versions of $h_{SJ_2}$, the smooth estimate of $F_0$ is replaced by the empirical distribution function $F_{0n}$. Therefore, although $h_{SJ_2}$ requires the selection of two bandwidth parameters, a clear better practical behaviour is observed when considering the smoothed relative data instead of the non-smoothed ones.

## 5 A medical application

In this section we apply the plug-in STE selector $h_{SJ_2}$ detailed above, to estimate the relative density for a real data set concerned with prostate cancer (PC).

The data consist of 599 patients suffering from PC (+) and 835 patients PC-free (−). For each patient the illness status has been determined through a prostate biopsy carried out for first time in Hospital Juan Canalejo (Galicia, Spain) between January of 2002 and September of 2005.

In the literature, there exists an increasingly interest in finding a good diagnostic test that helps in the early detection of PC and avoids the need of undergoing a prostate biopsy. There are several studies in which, through ROC curves, it was investigated the performance of different diagnostic tests based on some analytic measurements such as the total prostate specific antigen (tPSA), the free PSA (fPSA) or the complexed PSA (cPSA).

As it was mentioned in Sect. 1, there exists a close relation between the concepts of ROC curve and relative density. Relative density estimates can provide more detailed information about the performance of a diagnostic test which can be useful not only in comparing different tests but also in designing an improved one. This issue goes beyond the scope of this article and therefore it will not be investigated here.

In this section we compare from a distributional point of view the above mentioned measurements (tPSA, fPSA and cPSA) among the two groups in the data set (PC+ and PC−). To this end we start computing the appropriate bandwidths using the data-driven bandwidth selector $h_{SJ_2}$ and then the corresponding relative density estimates are computed using (2). These estimates are shown in Fig. 2.

It is clear from Fig. 2 that the relative density estimate is above one in the upper interval accounting for a probability of about 30% of the PC– distribution for the variables tPSA and cPSA. In the case of fPSA the 25% left tail of the PC– group and an interval in the upper tail that starts approximately at the quantile 0.7 of the PC– group, show as well that the relative density estimate is slightly above one. However, this effect is less remarkable that in the case of the variables tPSA and cPSA.

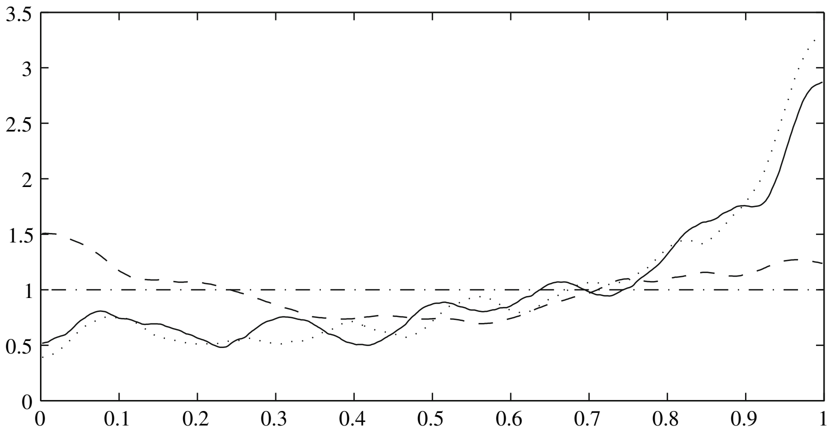


**Fig. 2** Relative density estimate of the PC+ group w.r.t. the PC− group for the variables tPSA (*solid line*, $h_{SJ_2} = 0.0645$), cPSA (*dotted line*, $h_{SJ_2} = 0.0625$) and fPSA (*dashed line*, $h_{SJ_2} = 0.1067$)

## 6 Proofs

The proof of Theorem 1 will be a direct consequence of some previous lemmas where each one of the terms that result from expanding the expression for the MISE are studied. Some of them will produce dominant parts in the final expression for the MISE while others will yield negligible terms.

**Lemma 1** *Assume the hypothesis above. Then*

(i) $\int_0^1 E[(\tilde{r}_h(t)-r(t))^2]\mathrm{d}t = \frac{1}{mh}C(K)\int_0^1 r(t)\,\mathrm{d}t + \frac{1}{4}h^4 d_K^2 C(r^{(2)}) + o\left(\frac{1}{mh}+h^4\right)$.

(ii) $\int_0^1 E[A_2^2]\mathrm{d}t = \frac{1}{nh}C(r)C(K) + o\left(\frac{1}{nh}\right) = O\left(\frac{1}{nh}\right)$.

(iii) $\int_0^1 E[A_1^2]\mathrm{d}t = o\left(\frac{1}{nh}\right)$.

(iv) $\int_0^1 E[B^2]\mathrm{d}t = o\left(\frac{1}{nh}\right)$.

(v) $\int_0^1 E[2A_1(\tilde{r}_h - r)(t)]\mathrm{d}t = 0$.

(vi) $\int_0^1 E[2A_2(\tilde{r}_h - r)(t)]\mathrm{d}t = 0$.

(vii) $\int_0^1 E[2B(\tilde{r}_h - r)(t)]\mathrm{d}t = o\left(\frac{1}{mh}+h^4\right)$.

(viii) $\int_0^1 E[2A_1A_2]\mathrm{d}t = o\left(\frac{1}{mh}+h^4\right)$.

(ix) $\int_0^1 E[2A_1B]\mathrm{d}t = o\left(\frac{1}{mh}+h^4\right)$.

(x) $\int_0^1 E[2A_2B]\mathrm{d}t = o\left(\frac{1}{nh}+h^4\right)$.

**Lemma 2** *Assume the hypothesis above. Then*

(i) $\int_0^1 E[\hat{A}^2]\mathrm{d}t = o\left(\frac{1}{nh}\right)$.

(ii) $\int_0^1 E[\hat{B}^2]\mathrm{d}t = o\left(\frac{1}{nh}\right)$.

*Proof of Lemma 1* The proof of (i) is not included here because it is a classical result in the setting of ordinary density estimation in a one-sample problem (see Wand and Jones, 1995 for details).

We next prove (ii). Standard algebra gives

$$E[A_2^2] = \frac{1}{n^2h^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left[(F_0(w_1) - 1_{\{X_{0i}\le w_1\}})(F_0(w_2) - 1_{\{X_{0j}\le w_2\}})\right]$$
$$\times K^{(1)}\left(\frac{t - F_0(w_1)}{h}\right) K^{(1)}\left(\frac{t - F_0(w_2)}{h}\right) \mathrm{d}F(w_1)\mathrm{d}F(w_2).$$

Due to the independence between $X_{0i}$ and $X_{0j}$ for $i \neq j$, and using the fact that

$$\mathrm{Cov}(1_{\{F_0(X_{0i})\le u_1\}}, 1_{\{F_0(X_{0i})\le u_2\}}) = (1 - u_1)(1 - u_2)g_0(u_1 \wedge u_2),$$

where $g_0(t) = \frac{t}{1-t}$, the previous expression can be rewritten as follows

$$E[A_2^2] = \frac{2}{nh^4}\int_0^1\int_{u_2}^1 (1 - u_1)(1 - u_2)g_0(u_2)K^{(1)}\left(\frac{t - u_1}{h}\right) K^{(1)}$$
$$\times \left(\frac{t - u_2}{h}\right) r(u_1)r(u_2)\mathrm{d}u_1\mathrm{d}u_2$$
$$= -\frac{1}{nh^4}\int_0^1 g_0(u_2)\mathrm{d}\left[\int_{u_2}^1 (1 - u_1)K^{(1)}\left(\frac{t - u_1}{h}\right) r(u_1)\mathrm{d}u_1\right]^2.$$

Now, using integration by parts, it follows that

$$E[A_2^2] = \frac{-1}{nh^4}\lim_{u_2\to 1^-} g_0(u_2)G(u_2)^2 + \frac{1}{nh^4}\lim_{u_2\to 0^+} g_0(u_2)G(u_2)^2$$
$$+\frac{1}{nh^4}\int_0^1 G(u_2)^2 g_0^{(1)}(u_2)\mathrm{d}u_2, \tag{15}$$

where

$$G(u_2) = \int_{u_2}^1 (1 - u_1)K^{(1)}\left(\frac{t - u_1}{h}\right) r(u_1)\mathrm{d}u_1.$$

Since $G$ is a bounded function and $g_0(0) = 0$, the second term in the right hand side of (15) vanishes to zero. On the other hand, due to the boundedness of $K^{(1)}$ and $r$, it follows that $|G(u_2)| \le \left\|K^{(1)}\right\|_\infty \|r\|_\infty \frac{(1-u_2)^2}{2}$, which let us conclude that the first term in (15) is zero as well. Therefore,

$$E[A_2^2] = \frac{1}{nh^4}\int_0^1 G(u_2)^2 g_0^{(1)}(u_2)\mathrm{d}u_2. \tag{16}$$

Now, using integration by parts, it follows that

$$\begin{aligned}
G(u_2) &= \left[ -(1-u_1)r(u_1)hK\left(\frac{t-u_1}{h}\right)\right]_{u_2}^{1} \\
&\quad + \int_{u_2}^{1} hK\left(\frac{t-u_1}{h}\right)[-r(u_1)+(1-u_1)r^{(1)}(u_1)]\mathrm{d}u_1 \\
&= h(1-u_2)r(u_2)K\left(\frac{t-u_2}{h}\right) \\
&\quad + h\int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)]\mathrm{d}u_1,
\end{aligned}$$

and plugging this last expression in (16), it is concluded that

$$E[A_2^2] = \frac{1}{nh^2}(I_{21}+2I_{22}+I_{23}),$$

where

$$\begin{aligned}
I_{21} &= \int_0^1 r^2(u_2)K^2\left(\frac{t-u_2}{h}\right)\mathrm{d}u_2 \\
I_{22} &= \int_0^1 \frac{1}{(1-u_2)} r(u_2)K\left(\frac{t-u_2}{h}\right) \\
&\quad \times \int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)]\mathrm{d}u_1\mathrm{d}u_2 \\
I_{23} &= \int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^{1}\int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)] \\
&\quad \times K\left(\frac{t-u_1^*}{h}\right)[(1-u_1^*)r^{(1)}(u_1^*)-r(u_1^*)]\mathrm{d}u_1\,\mathrm{d}u_1^*\mathrm{d}u_2.
\end{aligned}$$

Therefore,

$$\int_0^1 E[A_2^2]\mathrm{d}t = \int_0^1 \frac{1}{nh^2}I_{21}\mathrm{d}t + 2\int_0^1 \frac{1}{nh^2}I_{22}\mathrm{d}t + \int_0^1 \frac{1}{nh^2}I_{23}\mathrm{d}t. \tag{17}$$

Next, we will study each summand in (17) separately. The first term can be handled by using changes of variable and a Taylor expansion:

$$\int_0^1 \left(\frac{1}{nh^2}I_{21}\right)\mathrm{d}t = \frac{1}{nh}\int_0^1 r^2(u_2)\left(\int_{-\frac{u_2}{h}}^{\frac{1-u_2}{h}} K^2(s)\mathrm{d}s\right)\mathrm{d}u_2.$$

Let us define $\mathbb{K}_2(x) = \int_{-\infty}^{x} K^2(s)\,\mathrm{d}s$ and rewrite the previous term as follows

$$\int_0^1 \left(\frac{1}{nh^2} I_{21}\right) \mathrm{d}t = \frac{1}{nh} \int_0^1 r^2(u_2) \left(\mathbb{K}_2\left(\frac{1-u_2}{h}\right) - \mathbb{K}_2\left(-\frac{u_2}{h}\right)\right) \mathrm{d}u_2.$$

Now, by splitting the integration interval into three subintervals: $[0, h]$, $[h, 1-h]$ and $[1-h, 1]$, using changes of variable and the fact that

$$\mathbb{K}_2(x) = \begin{cases} C(K) & \forall x \geq 1, \\ 0 & \forall x \leq -1, \end{cases}$$

it is easy to show that

$$\int_0^1 \left(\frac{1}{nh^2} I_{21}\right) \mathrm{d}t = \frac{1}{nh} C(K) \cdot C(r) + O\left(\frac{1}{n}\right).$$

Below, we will study the second term in the right hand side of (17). By using changes of variable, Cauchy–Schwarz inequality and conditions $\|r\|_\infty < \infty$, $\left\|r^{(1)}\right\|_\infty < \infty$ and $C(K) < \infty$, straightforward calculations lead to

$$\int_0^1 \left(\frac{1}{nh^2} I_{22}\right) \mathrm{d}t = O\left(\frac{1}{n} + \frac{1}{nh^{\frac{1}{2}}}\right) = O\left(\frac{1}{nh^{\frac{1}{2}}}\right).$$

Similar arguments give

$$\int_0^1 \left(\frac{1}{nh^2} I_{23}\right) \mathrm{d}t = O\left(\frac{1}{nh^{\frac{1}{2}}}\right).$$

Therefore, it has been shown that

$$\int_0^1 E[A_2^2]\mathrm{d}t = \frac{1}{nh} C(r) C(K) + O\left(\frac{1}{n}\right) + O\left(\frac{1}{nh^{\frac{1}{2}}}\right).$$

Finally the proof of (ii) concludes using condition (B1).

We now prove (iii). Direct calculations lead to

$$E[A_1^2] = \frac{1}{h^4} E\left[I_1\right], \tag{18}$$

where

$$I_1 = E\Bigg[\int_0^1 \int_0^1 (v_1 - \tilde{U}_n(v_1))(v_2 - \tilde{U}_n(v_2)) K^{(1)}\left(\frac{t - v_1}{h}\right) \times K^{(1)}\left(\frac{t - v_2}{h}\right) \mathrm{d}(\tilde{R}_m - R)(v_1) \mathrm{d}(\tilde{R}_m - R)(v_2) / X_{01}, \ldots, X_{0n}\Bigg]. \tag{19}$$

To tackle with (18) we first study the conditional expectation (19). It is easy to see that

$$I_1 = \mathrm{Var}[V/X_{01}, \ldots, X_{0n}],$$

where

$$V = \frac{1}{m} \sum_{j=1}^{m} \left( X_j - \tilde{U}_n(X_j) \right) K^{(1)} \left( \frac{t - X_j}{h} \right).$$

Thus

$$I_1 = \frac{1}{m} \left\{ \int_0^1 \left[ (v - \tilde{U}_n(v)) K^{(1)} \left( \frac{t-v}{h} \right) \right]^2 \mathrm{d}R(v) \right.$$
$$\left. - \left[ \int_0^1 (v - \tilde{U}_n(v)) K^{(1)} \left( \frac{t-v}{h} \right) \mathrm{d}R(v) \right]^2 \right\}$$

and

$$E[A_1^2] = \frac{1}{mh^4} \int_0^1 E\left\{ \left[ (v - \tilde{U}_n(v)) \right]^2 \right\} \left[ K^{(1)} \left( \frac{t-v}{h} \right) \right]^2 \mathrm{d}R(v) - \frac{1}{mh^4}$$
$$\int_0^1 \int_0^1 E\left[ (v_1 - \tilde{U}_n(v_1))(v_2 - \tilde{U}_n(v_2)) \right] K^{(1)} \left( \frac{t-v_1}{h} \right) K^{(1)} \left( \frac{t-v_2}{h} \right) \mathrm{d}R(v_1)\mathrm{d}R(v_2).$$

Taking into account that

$$E\left[ \sup_v |(\tilde{U}_n(v) - v)|^2 \right] = \int_0^\infty P\left( \sup_v |(\tilde{U}_n(v) - v)|^2 > c \right) \mathrm{d}c,$$

we can use the Dvoretzky–Kiefer–Wolfowitz inequality, to conclude that

$$E\left[ \sup_v |(\tilde{U}_n(v) - v)|^2 \right] \leq \int_0^\infty 2\mathrm{e}^{-(2nc)} \mathrm{d}c = \frac{2}{n} \int_0^\infty y \mathrm{e}^{-y^2} \mathrm{d}y = O\left( \frac{1}{n} \right). \quad (20)$$

Consequently, using (20) and the conditions $\|r\|_\infty < \infty$ and $\left\|K^{(1)}\right\|_\infty < \infty$ we obtain that $E[A_1^2] = O\left( \frac{1}{mnh^4} \right)$. The proof of (iii) is concluded using condition (B1).

The results appearing in items (iv)–(x) can be proved by first conditioning to some appropriate random variables and then handling the conditional moments using standard arguments. For this reason their proofs are not included here. □

*Proof of Lemma 2* We start proving (i). Let us define $D_n(w) = \tilde{F}_{0n}(w) - F_{0n}(w)$, then

$$
\begin{aligned}
E[\hat{A}^2] &= E[E[\hat{A}^2/X_1,\dots,X_m]] \\
&= E\bigg[\iint E[D_n(w_1)D_n(w_2)] \\
&\quad \times K_h^{(1)}(t-F_0(w_1))K_h^{(1)}(t-F_0(w_2))\mathrm{d}F_m(v_1)\mathrm{d}F_m(v_2)\bigg].
\end{aligned}
$$

Based on the results set for $D_n(w)$ in Hjort and Walker (2001), the conditions (F2) and (K2) and since $E[D_n(w_1)D_n(w_2)] = \mathrm{Cov}(D_n(w_1), D_n(w_2)) + E[D_n(w_1)]E[D_n(w_2)]$, it follows that $E[D_n(w_1)D_n(w_2)] = O\left(\frac{h_0^4}{n}\right) + O(h_0^4)$.

Therefore, for any $t \in [0,1]$, we can bound $E[\hat{A}^2]$, using suitable constants $C_2$ and $C_3$ as follows

$$
\begin{aligned}
E[\hat{A}^2] = C_2\frac{h_0^4}{h^4}\frac{1}{m}&\int\left(K^{(1)}\left(\frac{t-F_0(z)}{h}\right)\right)^2 f(z)\mathrm{d}z \\
+C_3\frac{h_0^4}{h^4}\frac{(m-1)}{m}&\iint\left|K^{(1)}\left(\frac{t-F_0(z_1)}{h}\right)\right| \\
\times&\left|K^{(1)}\left(\frac{t-F_0(z_2)}{h}\right)\right| f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2.
\end{aligned}
$$

Besides, the condition (R1) allows us to conclude that $\int\left(K^{(1)}\left(\frac{t-F_0(z)}{h}\right)\right)^2 f(z)\mathrm{d}z = O(h)$ and $\iint\left|K^{(1)}\left(\frac{t-F_0(z_1)}{h}\right)\right|\left|K^{(1)}\left(\frac{t-F_0(z_2)}{h}\right)\right| f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2 = O(h^2)$ for all $t \in [0,1]$. Therefore, $\int_0^1 E\left[\hat{A}^2\right]\mathrm{d}t = O\left(\frac{h_0^4}{nh^3}\right) + O\left(\frac{h_0^4}{h^2}\right)$, which, taking into account conditions (B1) and (B2), implies (i).

We next prove (ii). The proof is parallel to that of item (iv) in Lemma 1. The only difference now is that instead of requiring $E[\sup|F_{0n}(x) - F_0(x)|^p] = O(n^{-\frac{p}{2}})$, where $p$ is an integer larger than 1, it is required that

$$
E\left[\sup\left|\tilde{F}_{0n}(x) - F_0(x)\right|^p\right] = O(n^{-\frac{p}{2}}). \tag{21}
$$

To conclude the proof, below we show that (21) is satisfied. Define $H_n = \sup|\tilde{F}_{0n}(x) - F_0(x)|$, then, as it is stated in Ahmad (2002), it follows that $H_n \le E_n + W_n$ where $E_n = \sup|F_{0n}(x) - F_0(x)|$ and $W_n = \sup|E\tilde{F}_{0n}(x) - F_0(x)| = O(h_0^2)$. Using the binomial formula it is easy to obtain that, for any integer $p \ge 1$, $H_n^p \le \sum_{j=0}^p C_j^p W_n^{p-j}E_n^j$, where the constants $C_j^p$'s (with $j \in \{0,1,\dots,p-1,p\}$) are the binomial coefficients. Therefore, since $E[E_n^j] = O(n^{-\frac{j}{2}})$ and $W_n^{p-j} = O(h_0^{2(p-j)})$, condition (B2) leads to $W_n^{p-j}E[E_n^j] = O(n^{-\frac{p}{2}})$.

As a straightforward consequence (21) holds and the proof of (ii) is concluded. □

*Proof of theorem 2* Below, we will briefly detail the steps followed to study the asymptotic behaviour of the mean squared error of $\hat{\Psi}_\ell(g)$ defined in (9). First of all, let us observe that

$$\hat{\Psi}_\ell(g) = \frac{1}{m} L_g^{(\ell)}(0) + \frac{1}{m^2} \sum_{j=1}^{m} \sum_{k=1, j\neq k}^{m} L_g^{(\ell)}\left(F_{0n}\left(X_j\right) - F_{0n}\left(X_k\right)\right),$$

which implies:

$$E\left[\hat{\Psi}_\ell(g)\right] = \frac{1}{mg^{\ell+1}} L^{(\ell)}(0) + \left(1 - \frac{1}{m}\right) E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right) - F_{0n}\left(X_2\right)\right)\right].$$

Starting from the equation

$$\begin{aligned}
&E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right) - F_{0n}\left(X_2\right)\right)\right] \\
&\quad = E\left[E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right) - F_{0n}\left(X_2\right)\right) / X_{01}, \ldots, X_{0n}\right]\right] \\
&\quad = E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L_g^{(\ell)}\left(F_{0n}\left(x_1\right) - F_{0n}\left(x_2\right)\right) f\left(x_1\right) f\left(x_2\right) \mathrm{d}x_1 \mathrm{d}x_2\right] \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left[L_g^{(\ell)}\left(F_{0n}\left(x_1\right) - F_{0n}\left(x_2\right)\right)\right] f\left(x_1\right) f\left(x_2\right) \mathrm{d}x_1 \mathrm{d}x_2
\end{aligned}$$

and using a Taylor expansion, we have

$$E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right) - F_{0n}\left(X_2\right)\right)\right] = \sum_{i=0}^{7} I_i, \tag{22}$$

where

$$\begin{aligned}
I_0 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{g^{\ell+1}} L^{(\ell)}\left(\frac{F_0\left(x_1\right) - F_0\left(x_2\right)}{g}\right) f\left(x_1\right) f\left(x_2\right) \mathrm{d}x_1\, \mathrm{d}x_2 \\
I_i &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{i! g^{\ell+i+1}} L^{(\ell+i)}\left(\frac{F_0\left(x_1\right) - F_0\left(x_2\right)}{g}\right) \\
&\quad \times E\left[\left(F_{0n}\left(x_1\right) - F_0\left(x_1\right) - F_{0n}\left(x_2\right) + F_0\left(x_2\right)\right)^i\right] f\left(x_1\right) f\left(x_2\right) \mathrm{d}x_1\, \mathrm{d}x_2
\end{aligned}$$

$i = 1, \ldots, 6$

$$\begin{aligned}
I_7 = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{7! g^{\ell+7+1}} E\left[L^{(\ell+7)}\left(\xi_n\right)\left(F_{0n}\left(x_1\right) - F_0\left(x_1\right) - F_{0n}\left(x_2\right) + F_0\left(x_2\right)\right)^7\right] \\
&\times f\left(x_1\right) f\left(x_2\right) \mathrm{d}x_1\, \mathrm{d}x_2
\end{aligned}$$

and $\xi_n$ is a value between $\frac{F_0(x_1) - F_0(x_2)}{g}$ and $\frac{F_{0n}(x_1) - F_{0n}(x_2)}{g}$.

Now, consider the first term, $I_0$, in (22). It is easy to see that

$$\begin{aligned} I_0 &= \int_0^1 \int_0^1 L_g^{(\ell)}(z_1 - z_2) r(z_1) r(z_2) \mathrm{d}z_1 \, \mathrm{d}z_2 \\ &= \int_0^1 \int_0^1 L_g(x) r(z_1 - z_2) r^{(\ell)}(z_2) \mathrm{d}z_1 \, \mathrm{d}z_2 \\ &= \int_0^1 \int_0^{(1-z_2)/g} L(x) r(z_2 + gx) r^{(\ell)}(z_2) \mathrm{d}x \, \mathrm{d}z_2, \end{aligned}$$

hence using a Taylor expansion, we have $I_0 = \Psi_\ell + (1/2) d_L \Psi_{\ell+2} g^2 + O\left(g^4\right)$. Assume $x_1 > x_2$ and define $Z = \sum_{i=1}^n 1_{\{x_2 < X_{0i} \leq x_1\}}$. Then, the random variable $Z$ has a $Bi(n,p)$ distribution with $p = F_0(x_1) - F_0(x_2)$ and mean $\mu = np$. It is easy to show that, for $i = 1, \ldots, 6$,

$$I_i = 2 \int_{-\infty}^{\infty} \int_{x_2}^{\infty} \frac{1}{i! g^{\ell+i+1}} L^{(\ell+i)} \left( \frac{F_0(x_1) - F_0(x_2)}{g} \right) f(x_1) f(x_2) \frac{1}{n^i} \mu_i(Z) \, \mathrm{d}x_1 \, \mathrm{d}x_2,$$

where

$$\begin{aligned} \mu_r(Z) &= E\left[(Z - E[Z])^r\right] \\ &= \sum_{j=0}^r (-1)^j \binom{r}{j} m_{r-j} \mu^j, \\ m_k &= E\left[Z^k\right] = \sum_{j=0}^k \frac{S(k,j)\, n! p^j}{(n-j)!}, \\ S(m,n) &= \frac{\sum_{j=0}^n \binom{n}{j} (-1)^j (n-j)^m}{n!}. \end{aligned}$$

Noting $\mu_1(Z) = 0$ and $\mu_2(Z) = n(F_0(x_1) - F_0(x_2))(1 - F_0(x_1) + F_0(x_2))$, we have $I_1 = 0$ and

$$\begin{aligned} I_2 &= \frac{1}{n g^{\ell+1+2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L^{(\ell+2)} \left( \frac{F_0(x_1) - F_0(x_2)}{g} \right) f(x_1) f(x_2) \\ &\quad \times (F_0(x_1) - F_0(x_2))(1 - F_0(x_1) + F_0(x_2)) \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \frac{1}{n g^{\ell+1+2}} \int_0^1 \int_v^1 L^{(\ell+2)} \left( \frac{u - v}{g} \right) (u - v)(1 - u - v) r(u) r(v) \mathrm{d}u \, \mathrm{d}v \\ &= \frac{1}{n g^{\ell+1}} \int_0^1 \int_0^{(1-v)/g} L^{(\ell+2)}(x) x (1 - gx) r(v + gx) r(v) \mathrm{d}x \, \mathrm{d}v. \qquad (23) \end{aligned}$$

Using a Taylor expansion of $(1 - gx)r(v + gx)$ and noting $\int_0^1 xL^{(\ell+2)}(x)\mathrm{d}x = L^{(\ell)}(0)$ (see condition (K3)), we have from (23)

$$I_2 = \frac{1}{ng^{\ell+1}}\Psi_0 L^{(\ell)}(0) + O\left((ng^{\ell})^{-1}\right).$$

Similar arguments can be used to handle $I_i = O((n^2g^{\ell+2})^{-1})$ for $i = 3, 4$ and $I_i = O((n^3g^{\ell+3})^{-1})$ for $i = 5, 6$. Coming back to the last term in (22) and using Dvoretzky–Kiefer–Wolfowitz inequality and condition (K3), it is easy to show that $I_7 = O((n^{\frac{7}{2}}g^{\ell+8})^{-1})$. Therefore,

$$E\left[\hat{\Psi}_\ell(g)\right] = \Psi_\ell + \frac{1}{2}d_L\Psi_{\ell+2}g^2 + \frac{1}{mg^{\ell+1}}L^{(\ell)}(0) + \frac{1}{ng^{\ell+1}}L^{(\ell)}(0)\,\Psi_0 + O\left(g^4\right) + o\left((ng^{\ell+1})^{-1}\right).$$

In order to study the variance of $\hat{\Psi}_\ell(g)$, note that

$$\mathrm{Var}\left[\hat{\Psi}_\ell(g)\right] = \sum_{i=1}^{3} c_{n,i}V_{\ell,i}, \tag{24}$$

where

$$c_{n,1} = \frac{2(m-1)}{m^3},$$
$$c_{n,2} = \frac{4(m-1)(m-2)}{m^3},$$
$$c_{n,3} = \frac{(m-1)(m-2)(m-3)}{m^3},$$

$$V_{\ell,1} = \mathrm{Var}\left[L_g^{(\ell)}\left(F_{0n}(X_1) - F_{0n}(X_2)\right)\right], \tag{25}$$
$$V_{\ell,2} = \mathrm{Cov}\left[L_g^{(\ell)}\left(F_{0n}(X_1) - F_{0n}(X_2)\right), L_g^{(\ell)}\left(F_{0n}(X_2) - F_{0n}(X_3)\right)\right], \tag{26}$$
$$V_{\ell,3} = \mathrm{Cov}\left[L_g^{(\ell)}\left(F_{0n}(X_1) - F_{0n}(X_2)\right), L_g^{(\ell)}\left(F_{0n}(X_3) - F_{0n}(X_4)\right)\right]. \tag{27}$$

Therefore, in order to get an asymptotic expression for the variance of $\hat{\Psi}_\ell(g)$, we will start getting asymptotic expressions for the terms (25), (26) and (27) in (24). To deal with the term (25), we will use

$$V_{\ell,1} = E\left[\left(L_g^{(\ell)}\left(F_{0n}(X_1) - F_{0n}(X_2)\right)\right)^2\right] - E^2\left[L_g^{(\ell)}\left(F_{0n}(X_1) - F_{0n}(X_2)\right)\right] \tag{28}$$

and study separately each term in the right-hand side of (28). Note that the expectation of $L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)$ has been already studied when dealing with the expectation of $\hat{\Psi}_\ell(g)$. Next we study the first term in the right-hand side of (28). Using a Taylor expansion, the term:

$$\begin{aligned}
&E\left[\left(L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)\right)^2\right] \\
&\quad=E\left[E\left[\left(L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)\right)^2/X_{01},\ldots,X_{0n}\right]\right] \\
&\quad=E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}L_g^{(\ell)^2}\left(F_{0n}\left(x\right)-F_{0n}\left(y\right)\right)f\left(x\right)f\left(y\right)\mathrm{d}x\,\mathrm{d}y\right] \\
&\quad=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\left[L_g^{(\ell)^2}\left(F_{0n}\left(x\right)-F_{0n}\left(y\right)\right)\right]f\left(x\right)f\left(y\right)\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

can be decomposed in a sum of six terms that can be bounded easily. The first term in that decomposition can be rewritten as $\frac{1}{g^{2\ell+1}}\Psi_0 C\left(L^{(\ell)}\right)+o\left(\frac{1}{g^{2\ell+1}}\right)$ after applying some changes of variable and a Taylor expansion. The other terms can be easily bounded using Dvoretzky–Kiefer–Wolfowitz inequality and standard changes of variable. These bounds and condition (B3) prove that the order of these terms is $o\left(\frac{1}{g^{2\ell+1}}\right)$. Consequently,

$$\begin{aligned}
V_{\ell,1}&=\frac{1}{g^{2\ell+1}}\Psi_0 C\left(L^{(\ell)}\right)+o\left(\frac{1}{g^{2\ell+1}}\right)-\left(\Psi_\ell+o\left(1\right)\right)^2 \\
&=\frac{1}{g^{2\ell+1}}\Psi_0 C\left(L^{(\ell)}\right)-\Psi_\ell^2+o\left(\frac{1}{g^{2\ell+1}}\right)+o\left(1\right).
\end{aligned}$$

The term (26) can be handled using

$$\begin{aligned}
V_{\ell,2}=&\,E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)L_g^{(\ell)}\left(F_{0n}\left(X_2\right)-F_{0n}\left(X_3\right)\right)\right] \\
&-E^2\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)\right].
\end{aligned}\tag{29}$$

As for (28), it is only needed to study the first term in the righthand side of (29).

Note that

$$E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)L_g^{(\ell)}\left(F_{0n}\left(X_2\right)-F_{0n}\left(X_3\right)\right)\right]$$
$$=E\left[E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)L_g^{(\ell)}\left(F_{0n}\left(X_2\right)-F_{0n}\left(X_3\right)\right)/X_{01,\ldots,}X_{0n}\right]\right]$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\left[L_g^{(\ell)}\left(F_{0n}\left(y\right)-F_{0n}\left(z\right)\right)L_g^{(\ell)}\left(F_{0n}\left(z\right)-F_{0n}\left(t\right)\right)\right]$$
$$\times f\left(y\right)f\left(z\right)f\left(t\right)\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}t.$$

Taylor expansions, changes of variable, Cauchy–Schwarz inequality and Dvoretzky–Kiefer–Wolfowitz inequality, give:

$$E\left[L_g^{(\ell)}\left(F_{0n}\left(X_1\right)-F_{0n}\left(X_2\right)\right)L_g^{(\ell)}\left(F_{0n}\left(X_2\right)-F_{0n}\left(X_3\right)\right)\right]$$
$$=\int_0^1 r^{(\ell)^2}\left(z\right)r\left(z\right)\mathrm{d}z+O\left(\frac{1}{n}\right)+O\left(\frac{1}{n^2}\right)+O\left(\frac{1}{n^3}\right)+O\left(\frac{1}{n^4g^{2((\ell+1)+4)}}\right).$$

Consequently, using (B3) and (29), $V_{\ell,2}=O(1)$.
To study the term $V_{\ell,3}$ in (27), let us define

$$A_\ell=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[L_g^{(\ell)}\left(F_{0n}\left(y\right)-F_{0n}\left(z\right)\right)-L_g^{(\ell)}\left(F_0\left(y\right)-F_0\left(z\right)\right)\right]f\left(y\right)f\left(z\right)\mathrm{d}y\,\mathrm{d}z.$$

It is easy to show that:

$$V_{\ell,3}=\mathrm{Var}\left(A_\ell\right).$$

Now a Taylor expansion gives

$$\mathrm{Var}\left(A_\ell\right)=\sum_{k=1}^N\mathrm{Var}\left(T_k\right)+\sum_{k=1}^N\sum_{\substack{\ell=1\\k\neq\ell}}^N\mathrm{Cov}\left(T_k,T_\ell\right), \tag{30}$$

where

$$A_{\ell} = \sum_{k=1}^{N} T_k,$$

$$T_k = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{k!g^{\ell+1}} L^{(\ell+k)}\left(\frac{F_0(y)-F_0(z)}{g}\right) f(y) f(z)$$
$$\times \left(\frac{F_{0n}(y)-F_{0n}(z)-(F_0(y)-F_0(z))}{g}\right)^k \mathrm{d}y\,\mathrm{d}z, \quad \mathit{for}\ k=1,\ldots,N-1,$$

$$T_N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{N!g^{\ell+1}} L^{(\ell+N)}(\xi_n) f(y) f(z)$$
$$\times \left(\frac{F_{0n}(y)-F_{0n}(z)-(F_0(y)-F_0(z))}{g}\right)^N \mathrm{d}y\,\mathrm{d}z,$$

for some positive integer $N$. We will only study each one of the first $N$ summands in (30). The rest of them will be easily bounded using Cauchy–Schwarz inequality and the bounds obtained for the first $N$ terms.

Now the variance of $T_k$ is studied. First of all, note that

$$\mathrm{Var}(T_k) \le E\left[T_k^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{k!g^{\ell+k+1}}\right)^2 f(y_1)f(z_1)f(y_2)f(z_2)$$
$$\times L^{(\ell+k)}\left(\frac{F_0(y_1)-F_0(z_1)}{g}\right) L^{(\ell+k)}\left(\frac{F_0(y_2)-F_0(z_2)}{g}\right)$$
$$\times h_k(y_1,z_1,y_2,z_2)\mathrm{d}y_1\,\mathrm{d}z_1\,\mathrm{d}y_2\,\mathrm{d}z_2,$$

where

$$h_k(y_1,z_1,y_2,z_2) = E\left\{\left[F_{0n}(y_1)-F_{0n}(z_1)-(F_0(y_1)-F_0(z_1))\right]^k \times \left[F_{0n}(y_2)-F_{0n}(z_2)-(F_0(y_2)-F_0(z_2))\right]^k\right\}.$$

Using changes of variable we can rewrite $E[T_k^2]$ as follows:

$$E\left[T_k^2\right] = \int_0^1\int_0^1\int_0^1\int_0^1 \left(\frac{1}{k!}\right)^2 r(s_1)r(t_1)r(s_2)r(t_2)L_g^{(\ell+k)}(s_1-t_1)\,L_g^{(\ell+k)}(s_2-t_2)$$
$$\times h_k(F_0^{-1}(s_1),F_0^{-1}(t_1),F_0^{-1}(s_2),F_0^{-1}(t_2))\mathrm{d}s_1\,\mathrm{d}t_1\,\mathrm{d}s_2\,\mathrm{d}t_2$$
$$= \int_0^1\int_{s_2-1}^{s_2}\int_0^1\int_{s_1-1}^{s_1}\left(\frac{1}{k!}\right)^2 r(s_1)r(s_1-u_1)r(s_2)r(s_2-u_2)L_g^{(\ell+k)}(u_1)\,L_g^{(\ell+k)}(u_2)$$
$$\times h_k(F_0^{-1}(s_1),F_0^{-1}(s_1-u_1),F_0^{-1}(s_2),F_0^{-1}(s_2-u_2))\mathrm{d}u_1\,\mathrm{d}s_1\,\mathrm{d}u_2\,\mathrm{d}s_2.$$

Note that closed expressions for $h_k$ can be obtained using the expressions for the moments of order $\mathbf{r} = (r_1,r_2,r_3,r_4,r_5)$ of $\mathbf{Z}$, a random variable with multinomial distribution with parameters $(n; p_1, p_2, p_3, p_4, p_5)$. Based on these expressions, the condition (R2) and the use of integration by parts we can rewrite $E\left[T_k^2\right]$ as follows:

$$E\left[T_k^2\right] = \int_0^1 \int_{s_2-1}^{s_2} \int_0^1 \int_{s_1-1}^{s_1} \left(\frac{1}{k!}\right)^2 L_g(u_1) L_g(u_2) \\ \times \frac{\partial^{2(\ell+k)}}{\partial u_1^{\ell+k} \partial u_2^{\ell+k}} (\tilde{h}_k(u_1, s_1, u_2, s_2)) \mathrm{d}u_1, \mathrm{d}s_1, \mathrm{d}u_2\, \mathrm{d}s_2,$$

where

$$\tilde{h}_k(u_1, s_1, u_2, s_2) = r(s_1) r(s_1 - u_1) r(s_2) r(s_2 - u_2) \\ \times h_k(F_0^{-1}(s_1), F_0^{-1}(s_1 - u_1), F_0^{-1}(s_2), F_0^{-1}(s_2 - u_2)).$$

Besides, based on the multinomial moments we can show that $\sup_{\mathbf{z} \in \Re^4} |h_k(\mathbf{z})| = O\left(\frac{1}{n^k}\right)$.This result and condition (R2) allow us to conclude that $\mathrm{Var}(T_k) \leq E\left[T_k^2\right] = O\left(\frac{1}{n^k}\right)$, for $1 \leq k < N$, which implies that $\mathrm{Var}(T_k) = o\left(\frac{1}{n}\right)$, for $2 \leq k < N$.

A Taylor expansion of order $N = 6$, gives $\mathrm{Var}(T_6) = O\left(\frac{1}{n^N g^{2(N+\ell+1)}}\right)$, which using condition (B3), proves $\mathrm{Var}(T_6) = o\left(\frac{1}{n}\right)$. Consequently,

$$\mathrm{Var}\left[\hat{\Psi}_\ell(g)\right] = \frac{2}{m^2 g^{2\ell+1}} \Psi_0 C\left(L^{(\ell)}\right) + o\left(\left(m^2 g^{2\ell+1}\right)^{-1}\right) + O\left(n^{-1}\right).$$

*Remark 4* If Eq. (22) is replaced by a three-term Taylor expansion $\sum_{i=1}^2 I_i + I_3^*$, where

$$I_3^* = \frac{1}{3! g^{\ell+4}} \iint E\left[L^{(\ell+3)}(\zeta_n)(F_{0n}(x_1) - F_0(x_1) - F_{0n}(x_2) + F_0(x_2))^3\right] \\ \times f(x_1) f(x_2) \mathrm{d}x_1\, \mathrm{d}x_2,$$

and $\zeta_n$ is a value between $\frac{F_0(x_1) - F_0(x_2)}{g}$ and $\frac{F_{0n}(x_1) - F_{0n}(x_2)}{g}$, then $I_3^* = O\left(\frac{1}{n^{\frac{3}{2}} g^{\ell+4}}\right)$ and we would have to ask for the condition $ng^6 \longrightarrow \infty$ to conclude that $I_3^* = o\left(\frac{1}{ng^{\ell+1}}\right)$. However, this condition is very restrictive because it is not satisfied by the optimal bandwidth $g_\ell$ with $\ell = 0, 2$, which is $g_\ell \sim n^{-\frac{1}{\ell+3}}$. We could consider $\sum_{i=1}^3 I_i + I_4^*$ and then we would need to ask for the condition $ng^4 \longrightarrow \infty$. However, this condition is not satisfied by $g_\ell$ with $\ell = 0$. In fact, it follows that $ng_\ell^4 \longrightarrow 0$ if $\ell = 0$ and $ng_\ell^4 \longrightarrow \infty$ if $\ell = 2, 4, \ldots$ Something similar happens when we consider $\sum_{i=1}^4 I_i + I_5^*$ or $\sum_{i=1}^5 I_i + I_6^*$, i.e., the condition required in $g$, it is not satisfied by the optimal bandwidth when $\ell = 0$. Only when we stop in $I_7^*$, the required condition, $ng^{\frac{14}{5}} \longrightarrow \infty$, it is satisfied for all even $\ell$.

If Eq. (30) is reconsidered by the mean-value theorem, and then we consider that $A_\ell = T_1^*$ with

$$T_1^* = \iint \frac{1}{g^{\ell+2}} L^{(\ell+1)}(\zeta_n)\left[F_{0n}(y) - F_{0n}(z) - (F_0(y) - F_0(z))\right] f(x_1) f(x_2)\, \mathrm{d}y\, \mathrm{d}z,$$

it follows that $\mathrm{Var}(A_\ell) = O\left(\frac{1}{ng^{2(\ell+2)}}\right)$. However, assuming that $g \longrightarrow 0$, it is impossible to conclude from here that $\mathrm{Var}(A_\ell) = o\left(\frac{1}{n}\right)$. □

**Acknowledgments** Research supported by Grants BES-2003-1170 (EU ESF support included) for the first author, XUGA PGIDIT03PXIC10505-PN for the second author and BFM2002-00265 and MTM2005-00429 (EU ERDF support included) for both authors. The authors would like to thank two anonymous referees and an associate editor whose comments have helped to improve the paper substantially. Thanks are also due to Sonia Pértega Díaz and Francisco Gómez Veiga from the "Hospital Juan Canalejo" in A Coruña for providing the prostate cancer data set.

## References

Ahmad, I.A. (2002). On moment inequalities of the supremum of empirical processes with applications to kernel estimation. *Statistics & Probability Letters*, *57*, 215–220.

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

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

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

Gastwirth, J. L. (1968). The first-median test: A two-sided version of the control median test. *Journal of the American Statistical Association*, *63*, 692–706.

Hall, P., Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters*, *6*, 109–115.

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

Hjort, N. L., Walker, SG. (2001). A note on kernel density estimators with optimal bandwidths. *Statistics & Probability Letters*, *54*, 153–159.

Holmgren, EB. (1996). The *P–P* plot as a method for comparing treatment effects. *Journal of the American Statistical Association*, *90*, 360–365.

Hsieh, F. (1995). The empirical process approach for semiparametric two-sample models with heterogenous treatment effect. *Journal of the Royal Statistical Society. Series B*, *57*, 735–748.

Hsieh, F., Turnbull, BW. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics*, *24*, 25–40.

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

Li, G., Tiwari, R. C., Wells, M. T. (1996). Quantile comparison functions in two-sample problems with applications to comparisons of diagnostic markers. *Journal of the American Statistical Association*, *91*, 689–698.

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

Sheather, S. J., Jones, MC. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B*, *53*, 683–690.
Silverman, B. W. (1978). Density ratios, empirical likelihood and cot death. *Applied Statistics*, *27*, 26–33.
Wand, M.P., Jones, M.C. (1995). *Kernel Smoothing*, London: Chapman and Hall.