REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 169–190

BOUNDARY KERNELS FOR DISTRIBUTION FUNCTION ESTIMATION

Author: Carlos Tenreiro
– CMUC, Department of Mathematics, University of Coimbra,
Apartado 3008, 3001–501 Coimbra, Portugal
tenreiro@mat.uc.pt

Received: January 2012 Accepted: August 2012

Abstract:

- Boundary effects for kernel estimators of curves with compact supports are well known in regression and density estimation frameworks. In this paper we address the use of boundary kernels for distribution function estimation. We establish the Chung-Smirnov law of iterated logarithm and an asymptotic expansion for the mean integrated squared error of the proposed estimator. These results show the superior theoretical performance of the boundary modified kernel estimator over the classical kernel estimator for distribution functions that are not smooth at the extreme points of the distribution support. The automatic selection of the bandwidth is also briefly discussed in this paper. Beta reference distribution and cross-validation bandwidth selectors are considered. Simulations suggest that the cross-validation bandwidth performs well, although the simpler reference distribution bandwidth is quite effective for the generality of test distributions.

Key-Words:

- *kernel distribution function estimation; boundary kernels; Chung-Smirnov property; MISE expansion; bandwidth selection.*

AMS Subject Classification:

- 62G05, 62G20.

1. INTRODUCTION

Given $X_1, ..., X_n$ independent copies of an absolutely continuous real random variable with unknown density and distribution functions f and F, respectively, a kernel estimator of F is introduced by authors such as Tiago de Oliveira [33], Nadaraya [20] or Watson and Leadbetter [35]. Such an estimator arises as an integral of the Parzen-Rosenblatt kernel density estimator (see Rosenblatt [25] and Parzen [21]) and is defined, for $x \in \mathbb{R}$, by

$$\bar{F}_{nh}(x) = \frac{1}{n} \sum_{i=1}^{n} \bar{K}\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where, for $u \in \mathbb{R}$,

$$\bar{K}(u) = \int_{]-\infty,u]} K(v)dv,$$

with K a kernel on $\mathbb{R}$, that is, a bounded and symmetric probability density function with support $[-1, 1]$ and $h = h_n$ a sequence of strictly positive real numbers converging to zero when n goes to infinity. Theoretical properties of this estimator, including bandwidth selection, have been investigated by several authors. Classical and more recent references, showing a continued interest in the subject, are, among others, Winter [36, 37], Yamato [38], Falk [7], Singh, Gasser and Prasad [28], Swanepoel [30], Jones [13], Shirahata and Chu [27], Sarda [26], Altman and Léger [1], Bowman, Hall and Prvan [2], Tenreiro [31, 32], Liu and Yang [16], Giné and Nickl [11], Mason and Swanepoel [18] and Chacón and Rodrígues-Casal [3].

If the support of f is known to be the finite interval $[a, b]$, from the continuity of F it is well known that the kernel estimator (1.1) is an asymptotically unbiased estimator of F if and only if $h \to 0$ as n goes to infinity (see Yamato [38], Lemma 1). However, if F is not smooth enough at the extreme points of the distribution support, the bias of $\bar{F}_{nh}$ does not achieve the standard h^2 order of convergence on the left and right boundary regions. In fact, assuming that the restriction of F to the interval $[a, b]$ is twice continuously differentiable, for $x = a + \alpha h$ with, $\alpha \in [0, 1]$, we have

$$\mathrm{E}\bar{F}_{nh}(x) - F(x) = hF'_+(a)\varphi_1(\alpha) + \frac{h^2}{2}F''_+(a)\varphi_2(\alpha) + o(h^2),$$

uniformly in α, with

$$\varphi_1(\alpha) = \alpha(\mu_{0,\alpha}(K) - 1) - \mu_{1,\alpha}(K), \tag{1.2}$$

$\varphi_2(\alpha) = \alpha^2(\mu_{0,\alpha}(K) - 1) - 2\alpha\mu_{1,\alpha}(K) + \mu_{2,\alpha}(K)$ and $\mu_{\ell,\alpha}(K) = \int_{-1}^{\alpha} z^\ell K(z)dz$. A similar expansion is valid for x in the right boundary region. As noticed by Gasser and Müller [9] in a regression context, this local behaviour dominates the global behaviour of the estimator which implies an inferior global order of convergence for the kernel estimator (1.1) which can be confirmed by examining the asymptotic behaviour of widely used measures of the quality of kernel estimators such as the maximum absolute deviation or the mean integrated squared error.

This type of boundary effect for kernel estimators of curves with compact supports is well-known in regression and density function estimation frameworks and several modified estimators have been proposed in the literature (see Müller [19], Karunamuni and Alberts [14], and Karunamuni and Zhang [15], and references therein). In order to improve the theoretical performance of the standard kernel distribution function estimator when the underlying distribution function F is not smooth enough at the extreme points of the distribution support, the use of the so-called boundary kernels, suggested for regression and density kernel estimators by Gasser and Müller [9], Rice [24], Gasser, Müller and Mammitzsch [10] and Müller [19], is addressed in this paper, which is organised as follows.

In Section 2, we introduce the boundary modified kernel distribution function estimator and some families of boundary kernels are presented, one of them leading to proper distribution function estimators. Contrary to the boundary modified kernel density estimators which possibly assume negative values, in a distribution function estimation framework the theoretical advantage of using boundary kernels is compatible with the natural property of obtaining a proper distribution function estimate. In Section 3 we show that the Chung-Smirnov theorem, that gives the supremum norm convergence rate of the empirical distribution function estimator, is also valid for the boundary kernel distribution function estimator. In Section 4 we present an asymptotic expansion for the mean integrated squared error of the estimator. This result illustrates the superior theoretical performance of the boundary kernel distribution function estimator over the classical kernel estimator whenever the underlying distribution function is not smooth enough at the extreme points of the distribution support. The automatic selection of the bandwidth is addressed in Section 5 where beta reference distribution and cross-validation bandwidth selectors are considered. Simulations suggest that the cross-validation bandwidth performs well, although the simpler reference distribution bandwidth is quite effective for the generality of test distributions. All the proofs can be found in Section 6. The simulations and plots in this paper were carried out using the R software [23].

2. KERNEL ESTIMATOR WITH BOUNDARY KERNELS

In order to deal with the boundary effects that occur in nonparametric regression and density function estimation, the use of boundary kernels is proposed and studied by authors such as Gasser and Müller [9], Rice [24], Gasser, Müller and Mammitzsch [10] and Müller [19]. Next we extend this approach to a distribution function estimation framework, where we assume that the support of the underlying distribution is known to be the finite interval $[a, b]$.

We consider the boundary modified kernel distribution function estimator given by

$$\tilde{F}_{nh}(x) = \frac{1}{n}\sum_{i=1}^{n} \bar{K}_{x,h}\left(\frac{x - X_i}{h}\right), \tag{2.1}$$

for $x \in]a, b[$ and $0 < h \leq (b - a)/2$, where

$$\bar{K}_{x,h}(u) = \int_{]-\infty,u]} K_{x,h}(v)dv,$$

and $K_{x,h}$ takes the form

$$K_{x,h}(u) = \begin{cases} K^L(u; (x-a)/h), & a < x < a + h \\ K(u), & a + h \leq x \leq b - h \\ K^R(u; (b-x)/h), & b - h < x < b, \end{cases}$$

where K is a bounded and symmetric probability density function with support $[-1, 1]$, and $K^L(\cdot; \alpha)$ and $K^R(\cdot; \alpha)$ are second order (left and right) boundary kernels for $\alpha \in]0, 1[$. Therefore, $K^L(\cdot; \alpha)$ and $K^R(\cdot; \alpha)$ are such that theirs supports are contained in the intervals $[-1, \alpha]$ and $[-\alpha, 1]$, respectively, and

$$\int K^\ell(u; \alpha)du = 1, \int uK^\ell(u; \alpha)du = 0 \text{ and } \int u^2 K^\ell(u; \alpha)du \neq 0,$$

for all $\alpha \in]0, 1[$, with $\ell = R, L$. Additionally we define $\tilde{F}_{nh}(x) = 0$ for $x \leq a$ and $\tilde{F}_{nh}(x) = 1$ for $x \geq b$.

If we write

$$\bar{K}^\ell(u; \alpha) = \int_{]-\infty,u]} K^\ell(v; \alpha)dv,$$

for $\ell = L, R$, the kernel $\bar{K}_{x,h}$ can be written as

$$\bar{K}_{x,h}(u) = \begin{cases} \bar{K}^L(u; (x-a)/h), & a < x < a + h \\ \bar{K}(u), & a + h \leq x \leq b - h \\ \bar{K}^R(u; (b-x)/h), & b - h < x < b. \end{cases}$$

In the following examples we present three families of boundary kernels. We will assume that $K^R(u; \alpha) = K^L(-u; \alpha)$. In this case, we have $\bar{K}^R(u; \alpha) = 1 - \bar{K}^L(-u; \alpha)$.

Example 2.1. In a density estimation setting the standard choice for K^L is

$$K^L(u;\alpha) = (A_\alpha(K) + B_\alpha(K)u)K(u)I(-1 \le u \le \alpha),$$

where $A_\alpha(K) = \mu_{2,\alpha}(K)/D_\alpha(K)$, $B_\alpha(K) = -\mu_{1,\alpha}(K)/D_\alpha(K)$ and $D_\alpha(K) = \mu_{0,\alpha}(K)\mu_{2,\alpha}(K) - \mu_{1,\alpha}(K)^2$. Despite being negative for small values of α, this type of boundary kernels is suitable for density estimation. Contrary to nonnegative boundary kernels, they allow the control of the variability of the estimator near the support distribution boundary (see Gasser and Müller [9]). In this case, we get

$$\bar{K}^L(u;\alpha) = (A_\alpha(K)\bar{K}(u) + B_\alpha(K)\mu_{1,u}(K))I(-1 \le u \le \alpha) + I(u > \alpha).$$

A local behaviour analysis of the modified kernel distribution function estimator near the end points of the distribution support reveals that this class of boundary kernels may not be especially appropriate for the estimation of a distribution function. Restricting our analysis to the left-sided boundary region, and assuming the continuity of the second derivative of F in $]a, a+h[$, for $x = a + \alpha h$, with $\alpha \in]0,1[$, we have

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = \frac{h^2}{2}\, F''(x)\mu(\alpha) + o(h^2) \tag{2.2}$$

and

$$\mathrm{Var}\,\tilde{F}_{nh}(x) = \frac{F(x)(1-F(x))}{n} - \frac{h}{n}\, F'(x)\nu(\alpha) + O(n^{-1}h^2), \tag{2.3}$$

where

$$\mu(\alpha) = \int_{-1}^{\alpha} z^2 K^L(z;\alpha)\, dz$$

and

$$\nu(\alpha) = \int_{-1}^{\alpha} z B^L(z;\alpha)\, dz,$$

with $B^L(u;\alpha) = 2\bar{K}^L(u;\alpha)\; K^L(u;\alpha)$ (see expansions (6.4) and (6.5) in Section 6).

For the previous class of kernels the quantity $\nu(\alpha)$ can be negative for small values of α, which leads to an estimator whose local variability is larger than the empirical distribution function one. Additionally, as $\mu(\alpha)$ converges to a strictly negative value, when α tends to zero, a local bias can occur for small values of α (at the order of convergence h^2). In the next examples we take for $K^L(\cdot;\alpha)$ a symmetric probability density function with support $[-\alpha,\alpha]$. In this case, F_{nh} is nonnegative and $\nu(\alpha) > 0$, for $\alpha \in]0,1[$. Therefore, the boundary kernel estimator has a local variability inferior to the empirical distribution function one. Additionally, $\mu(\alpha)$ converges to zero, as α approaches zero (for the boundary kernels of Example 2.2, this is true whenever K is continuous on a neighbourhood of the origin with $K(0) > 0$).

Example 2.2. If K is such that $\int_0^\alpha K(u)du > 0$ for all $\alpha > 0$, for

$$K^L(u;\alpha) = (2\bar{K}(\alpha) - 1)^{-1} K(u) I(-\alpha \le u \le \alpha),$$

we have

$$\bar{K}^L(u;\alpha) = (2\bar{K}(\alpha) - 1)^{-1} (\bar{K}(u) - \bar{K}(-\alpha)) I(-\alpha \le u \le \alpha) + I(u > \alpha).$$

Example 2.3. If we take

$$K^L(u;\alpha) = K(u/\alpha)/\alpha$$

we get

$$\bar{K}^L(u;\alpha) = \bar{K}(u/\alpha).$$

Finally, note that, for these two last classes of boundary kernels, $\tilde{F}_{nh}$ is, with probability one, a continuous probability distribution function. Therefore, in a distribution function estimation framework, the theoretical advantage of using boundary kernels, which we establish in the following sections, is compatible with the natural property of obtaining proper distribution function estimates.

3. UNIFORM CONVERGENCE RESULTS

The almost sure (or complete) uniform convergence of the classical kernel distribution function estimator $\bar{F}_{nh}$ to F was established by Nadaraya [20], Winter [36] and Yamato [38], whereas Winter [37] proved that, under certain regularity conditions, $\bar{F}_{nh}$ has the Chung-Smirnov law of iterated logarithm property (see also Degenhardt [5] and Chacón and Rodrígues-Casal [3]). In this section we show that these results are also valid for the boundary kernel distribution function estimator (2.1). For that, we will need the following lemma that gives upper bounds for $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$ and $||\mathrm{E}\tilde{F}_{nh} - F||$, where $||\cdot||$ denotes the supremum norm.

Lemma 3.1. *For all* $0 < h \le (b-a)/2$, *we have*

$$||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}|| \le C_K ||\, F_n - F|| \tag{3.1}$$

and

$$||\mathrm{E}\tilde{F}_{nh} - F|| \le C_K \sup_{x,y\in[a,b]\colon |x-y|\le h} |F(x) - F(y)|, \tag{3.2}$$

where F_n *is the empirical distribution function and*

$$C_K = \max\left(1, \max_{\ell=L,R} \sup_{\alpha\in\,]0,1[} \int |K^\ell(u;\alpha)|\, du\right).$$

Moreover, if the derivative F' is continuous on $[a,b]$, then

$$||\mathrm{E}\tilde{F}_{nh} - F|| \le h\, C_K \sup_{x,y\in[a,b]:\,|x-y|\le h} |F'(x) - F'(y)|. \tag{3.3}$$

The next results follow straightforwardly from Lemma 3.1 after separating the difference $\tilde{F}_{nh} - F$ into a stochastic component $\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}$ and a non-stochastic bias component $\mathrm{E}\tilde{F}_{nh} - F$. The first one is a consequence of a well-known exponential inequality due to Dvoretzky, Kiefer and Wolfowitz [6], which gives a bound on the tail probabilities of $||F_n - F||$, and the second one follows from the law of iterated logarithm for the empirical distribution function estimator due to Smirnov [29] and Chung [4] (see also van der Vaart [34], p. 268, and references therein). Also note that the condition imposed on the boundary kernels is trivially satisfied by nonnegative boundary kernels such as those of the Examples 2.2 and 2.3. It is also fulfilled by the boundary kernels of Example 2.1.

Theorem 3.1. *For $\ell = L, R$, let K^ℓ be such that*

$$\sup_{\alpha\in\,]0,1[} \int |K^\ell(u;\alpha)|\, du < \infty.$$

If $h \to 0$, then

$$||\tilde{F}_{nh} - F|| \to 0 \text{ almost completely.}$$

Theorem 3.2. *Under the conditions of Theorem 3.1, if F is Lipschitz and $(n/\log\log n)^{1/2} h \to 0$, then $\tilde{F}_{nh}$ has the Chung-Smirnov property, i.e.,*

$$\limsup_{n\to+\infty} (2n/\log\log n)^{1/2} ||\tilde{F}_{nh} - F|| \le 1 \text{ almost surely.}$$

Moreover, the same is true whenever F' is Lipschitz on $[a,b]$ and h satisfies the less restrictive condition $(n/\log\log n)^{1/2} h^2 \to 0$.

Remark 3.1. If F is Lipschitz and the bandwidth fulfills the more restrictive condition $n^{1/2} h \to 0$, the Chung-Smirnov property can be deduced from the strong approximation property $\sqrt{n}\, ||\tilde{F}_{nh} - F_n|| = o(1)$ *almost surely*, that can be derived by adapting the approach by Fernholz [8]. In this case, $\sqrt{n}\, ||\tilde{F}_{nh} - F||$ and the Kolmogorov statistic $\sqrt{n}\, ||F_n - F||$ have the same asymptotic distribution.

Remark 3.2. When F' is Lipschitz on $[a,b]$ and $(n/\log\log n)^{1/2} h^2 \to 0$, F_{nh} has the Chung-Smirnov property without assuming the continuity of F' at $x = a$ or $x = b$. This shows that F_{nh} improves on $\bar{F}_{nh}$ for distribution functions which are not smooth enough at the extreme points of the distribution support (cf. Winter [37], Theorem 3.2).

Remark 3.3. If F is the uniform distribution on $[a, b]$, from inequality (3.3) we deduce that $||\mathrm{E}\tilde{F}_{nh} - F|| = 0$, for all $0 < h \leq (b-a)/2$. Therefore,

$$||\tilde{F}_{nh} - F|| = ||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}|| \leq C_K ||F_n - F||,$$

and $\tilde{F}_{nh}$ has the Chung-Smirnov property even when h does not converge to zero as n goes to infinity.

Remark 3.4. In practice the bandwidth h is usually chosen on the basis of the data, that is, $h = \hat{h}(X_1, ..., X_n)$. From the proof of Lemma 3.1 we easily conclude that the so-called automatic boundary kernel estimator defined by (2.1) with $h = \hat{h}$ satisfies the inequalities

$$||\tilde{F}_{n\hat{h}} - F|| \leq C_K \Big\{ ||F_n - F|| + \sup_{x,y\in[a,b] \colon |x-y|\leq\hat{h}} |F(x) - F(y)| \Big\},$$

for any F, and

$$||\tilde{F}_{n\hat{h}} - F|| \leq C_K \Big\{ ||F_n - F|| + \hat{h} \sup_{x,y\in[a,b] \colon |x-y|\leq\hat{h}} |F'(x) - F'(y)| \Big\},$$

whenever F' is continuous on $[a, b]$. Therefore, under the conditions of Theorems 3.1 and 3.2, if the assumptions on h are replaced by their almost sure counterparts, we conclude that the automatic boundary kernel estimator, $\tilde{F}_{n\hat{h}}$, is an almost sure uniform convergent estimator of F that enjoys the Chung-Smirnov property.

4. MISE ASYMPTOTIC EXPANSION

A widely used measure of the quality of the kernel estimator is the mean integrated squared error given by

$$\begin{aligned}
\mathrm{MISE}(F; h) &= \mathrm{E} \int \{\tilde{F}_{nh}(x) - F(x)\}^2 dx \\
&= \int \mathrm{Var}\, \tilde{F}_{nh}(x) dx + \int \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2 dx \\
&=: \tilde{\mathbf{V}}(F; h) + \tilde{\mathbf{B}}(F; h),
\end{aligned}$$

where the integrals are over $\mathbb{R}$. Denoting by $\bar{\mathbf{V}}(F; h)$ and $\bar{\mathbf{B}}(F; h)$ the corresponding variance and bias terms for the classical kernel distribution function estimator (1.1), the approach followed by Swanepoel [30] leads to the following expansions whenever the restriction of F to the interval $[a, b]$ is twice continuously differentiable:

$$\bar{\mathbf{V}}(F; h) = \frac{1}{n} \int F(x)(1 - F(x)) dx - \frac{h}{n} \int u B(u) du + O\left(n^{-1}h^2\right) \tag{4.1}$$

where

$$B(u) = 2\bar{K}(u)K(u),$$

for $u \in \mathbb{R}$, and

$$\begin{aligned}(4.2)\qquad \bar{\mathbf{B}}(F;h) &= h^3\left(F'_+(a)^2 + F'_-(b)^2\right)\int_0^1 \varphi_1(\alpha)^2 d\alpha \\ &\quad + h^4\big(F'_+(a)F''_+(a) - F'_-(b)F''_-(b)\big)\int u^2K(u)\,du\int_0^1 \varphi_1(\alpha)d\alpha \\ &\quad + \frac{h^4}{4}\left(\int u^2K(u)\,du\right)^2 ||F''||_2^2 + o(h^4),\end{aligned}$$

where φ_1 is given by (1.2) and $||\cdot||_2$ is the L_2 distance in $[a,b]$.

Depending on the smoothness of F on $\mathbb{R}$, we see that two different orders of convergence to zero for the mean integrated square error can be obtained. In the smooth case, that is, when $F'_+(a) = F'_-(b) = 0$, the previous expansions agree with the classical ones (cf. Jones [13]). However, in the non-smooth case an inferior global order of convergence occurs and a different order of convergence for the optimal bandwidth, in the sense of minimising the asymptotic MISE, takes place.

Next we show that, even when F is not smooth at the extreme points of the distribution support, the leading terms of the MISE expansion of the boundary kernel estimator agree with those given in Jones [13] for the classical kernel distribution function estimator. This shows the theoretical advantage of the boundary kernel distribution function estimator over the classical kernel estimator. Next define $B^\ell(u;\alpha) = 2\bar{K}^\ell(u;\alpha)K^\ell(u;\alpha)$, for $u \in \mathbb{R}$, $\alpha \in]0,1[$ and $\ell = L,R$.

Theorem 4.1. *For $\ell = L,R$, let K^ℓ be such that*

$$\int_0^1\left(\int |K^\ell(u;\alpha)|du\right)^2 d\alpha < \infty,$$

and assume that the restriction of F to the interval $[a,b]$ is twice continuously differentiable. We have

$$\tilde{\mathbf{V}}(F;h) = \frac{1}{n}\int F(x)(1-F(x))dx - \frac{h}{n}\int uB(u)du + O\left(n^{-1}h^2\right)$$

and

$$\tilde{\mathbf{B}}(F;h) = \frac{h^4}{4}\left(\int u^2K(u)du\right)^2 ||F''||_2^2 + o\left(h^4\right).$$

Note that the previous assumptions on the boundary kernels are trivially satisfied by nonnegative boundary kernels such as those of Examples 2.2 and 2.3, and also by the boundary kernels of Example 2.1. Next we give the asymptotically optimal choice for the bandwidth in the sense of minimising the leading terms in the expansion of the MISE.

Theorem 4.2. *Under the conditions of Theorem 4.1, let us assume that* $C_B > 0$ *where*

$$C_B = 2\int uB(u)\,du - \int_0^1\!\!\int u\left(B^L(u;\alpha) + B^R(u;\alpha)\right)\,du d\alpha.$$

Then the asymptotically optimal bandwidth is given by

$$h_0 = \min\left(\delta(K)||F''||_2^{-2/3} n^{-1/3}\,,\ \frac{b-a}{2}\min\left(1\,,\int uB(u)du/C_B\right)\right), \tag{4.3}$$

where

$$\delta(K) = \left(\int uB(u)\,du\right)^{1/3}\left(\int u^2K(u)du\right)^{-2/3}.$$

Remark 4.1. Following the approach by Marron and Jones [17], and taking into account the results of Swanepoel [30] and Jones [13], we conclude that the uniform density on $[-1,1]$ is the optimal kernel in the sense of minimising the asymptotic MISE. However, as noticed by Jones [13], other suboptimal kernels, such as the Epanechnikov kernel on $[-1,1]$, have a performance very close to the optimal one.

Remark 4.2. For the boundary kernels of Example 2.3, we have $C_B = \int uB(u)du > 0$ and the asymptotically optimal bandwidth is simply given by $h_0 = \min\left(\delta(K)||F''||_2^{-2/3}n^{-1/3}, (b-a)/2\right)$.

5. BANDWIDTH SELECTION

In a kernel estimation setting the bandwidth is usually chosen on the basis of the data. For the classical kernel distribution function estimator (1.1) and assuming that f is a smooth function over the whole real line, two main approaches for the automatic selection of h can be found in the literature. Cross-validation methods are discussed in Sarda [26], Altman and Léger [1] and Bowman, Hall and Prvan [2], and direct plug-in methods, including normal reference distribution methods, are proposed by Altman and Léger [1], Polansky and Baker [22] and Tenreiro [32]. In the following subsections we consider two fully automatic bandwidth selectors for the boundary kernel distribution function estimator. The first one is a reference distribution method based on the beta distribution family. The second one is a cross-validation bandwidth selector inspired in the approach of Bowman, Hall and Prvan [2].

5.1. A reference distribution method

A commonly used quick and simple method for choosing the bandwidth involves using the asymptotically optimal bandwidth for a fixed reference distribution having the same mean and scale as that estimated for the underlying distribution. In what follows a beta distribution over the interval $[a,b]$ with both shape parameters greater than or equal to 2 is taken as reference distribution. The restriction on the shape parameters values takes into account the assumptions on F imposed in Theorem 4.1. If X has a beta distribution over the interval $[a,b]$ with shape parameters p and q, the expected value of X is given by

$$\mathrm{E}(X) = a + (b-a)\frac{p}{p+q}$$

and the variance of X by

$$\mathrm{Var}(X) = (b-a)^2 \frac{pq}{(p+q)^2(p+q+1)}$$

(see Johnson, Kotz and Balakrishnan [12], p. 222). Taking the sample mean $\bar{X}$ and the sample variance S^2 as estimators of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$, respectively, the method of moments estimators for the parameters p and q are given by

$$\tilde{p} = \tilde{X}(\tilde{X}(1-\tilde{X})\tilde{S}^{-2} - 1) \quad \text{and} \quad \tilde{q} = (1-\tilde{X})(\tilde{X}(1-\tilde{X})\tilde{S}^{-2} - 1),$$

where $\tilde{X} = (\bar{X} - a)/(b-a)$ and $\tilde{S}^2 = S^2/(b-a)^2$. Thus, denoting by $\hat{F}$ the beta distribution over the interval $[a,b]$ with shape parameters $\hat{p} = \max(2,\tilde{p})$ and $\hat{q} = \max(2,\tilde{q})$, the considered beta optimal bandwidth, which we denote by $\hat{h}_{\mathrm{BR}}$, is defined by (4.3) with $||\hat{F}''||_2$ in place of $||F''||_2$ where

$$||\hat{F}''||_2^2 = \frac{(\hat{p}-1)(\hat{q}-1)\mathrm{B}(2\hat{p}-3, 2\hat{q}-3)}{(b-a)(2(\hat{p}+\hat{q})-5)\mathrm{B}(\hat{p},\hat{q})^2},$$

and $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the beta function.

5.2. A cross-validation method

An alternative approach for bandwidth selection can be based on the cross-validation ideas of Bowman, Hall and Prvan [2]. The cross-validation function proposed by these authors is a mean over all the observations of the integrated squared error between the indicator function $I(X_i \le x)$ associated to the observation X_i, and the boundary kernel estimator constructed from the data with observation X_i omitted, that is,

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \int \{I(X_i \le x) - \tilde{F}_{-ih}(x)\}^2 dx,$$

where

$$\tilde{F}_{-ih}(x) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \bar{K}_{x,h}\left(\frac{x - X_j}{h}\right).$$

The cross-validation bandwidth, which we denote by $\hat{h}_{\mathrm{CV}}$, is the minimiser of $\mathrm{CV}(h)$. The main motivation for this method comes from the equality

$$\mathrm{E}\left(\mathrm{CV}(h) - \frac{1}{n}\sum_{i=1}^{n}\int\{I(X_i \leq x) - F(x)\}^2 dx\right) = \mathrm{E}\int\{\tilde{F}_{n-1,h}(x) - F(x)\}^2 dx,$$

which shows that the criterion function $\mathrm{CV}(h)$ provides an unbiased estimator of $\mathrm{MISE}(F;h)$ for a sample size $n-1$, shifted vertically by an unknown term which is independent of h. Although the asymptotic behaviour of the cross-validation bandwidth is not discussed in this paper, it will be of interest to know whether $\hat{h}_{\mathrm{CV}}$ is asymptotically equivalent to the asymptotically optimal bandwidth h_0. As shown in Bowman, Hall and Prvan [2], this property is valid for the standard kernel distribution function estimator.

5.3. A simulation study

In order to analyse the finite sample performance of the bandwidth selectors $\hat{h}_{\mathrm{BR}}$ and $\hat{h}_{\mathrm{CV}}$, a simulation study was carried out for a set of beta mixture distributions with support $[0,1]$ that represents different shapes and boundary behaviours. Their weights and shape parameters are given in Table 1 and the corresponding probability density and cumulative distribution functions are shown in Figure 1.

Table 1: Beta mixture test distributions.

	Beta mixture distribution $\sum_i w_i\, B(p_i, q_i)$		
	Weights w	1st shape parameters p	2nd shape parameters q
#1	$(1/4, 3/4)$	$(1, 6)$	$(6, 1)$
#2	$(1/10, 7/10, 2/10)$	$(1, 2, 3)$	$(2, 2, 1)$
#3	$(1/10, 7/10, 2/10)$	$(1, 2, 6)$	$(2, 6, 1)$
#4	$(5/16, 5/16, 3/16, 2/16, 1/16)$	$(1, 25, 160, 320, 800)$	$(10, 60, 100, 80, 90)$

From each distribution we generated 500 samples of sizes $n = 25, 50, 100$ and 200, and we calculated the integrated squared error $\mathrm{ISE}(F;h) = \int\{\tilde{F}_{nh}(x) - F(x)\}^2 dx$ for $h = \hat{h}_{\mathrm{BR}}$ and $h = \hat{h}_{\mathrm{CV}}$ as a measure of the performance of each bandwidth selector. The integrated squared error associated to the asymptotically optimal bandwidth h_0 was also evaluated for the sake of comparison.

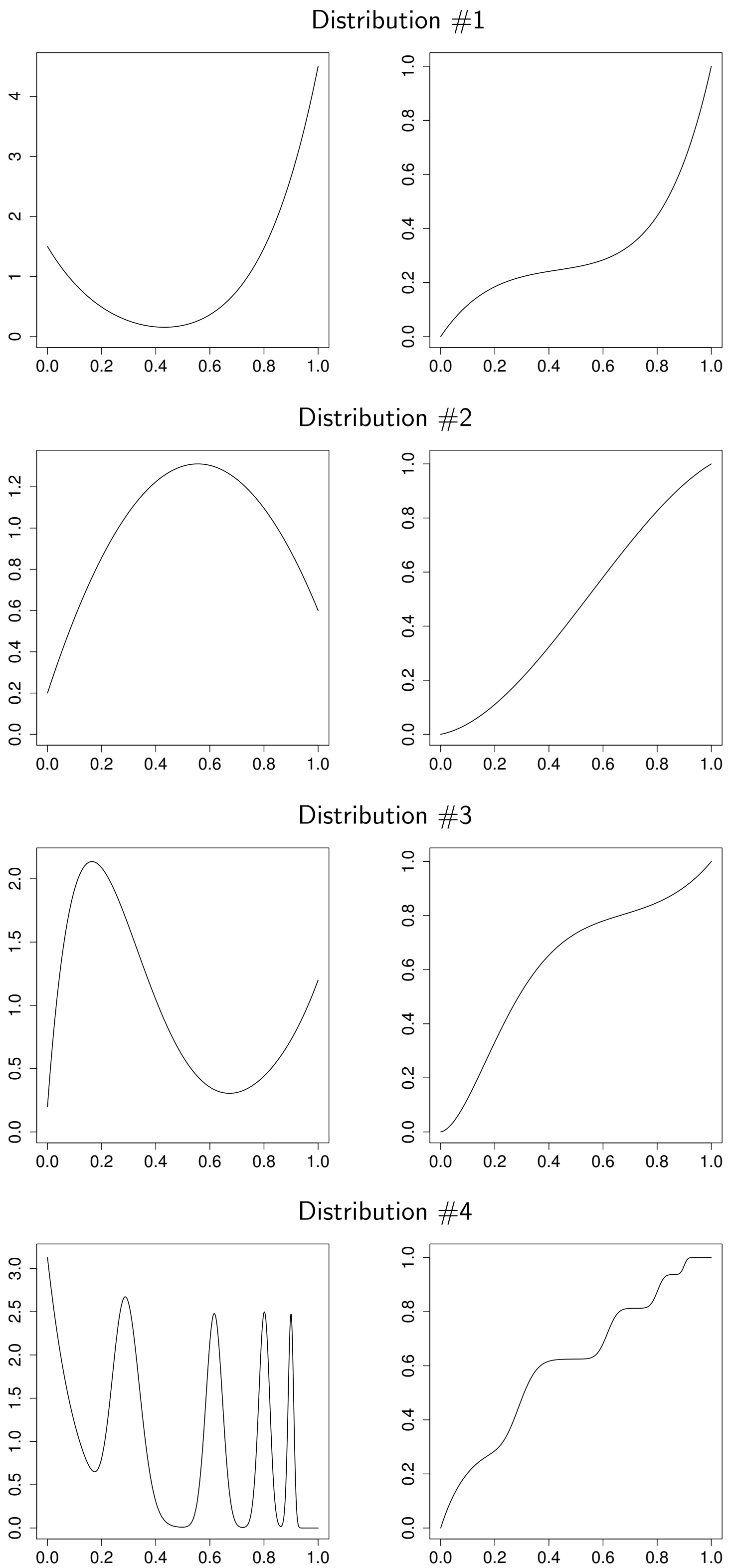

Figure 1: Beta mixture test density and cumulative distribution functions.

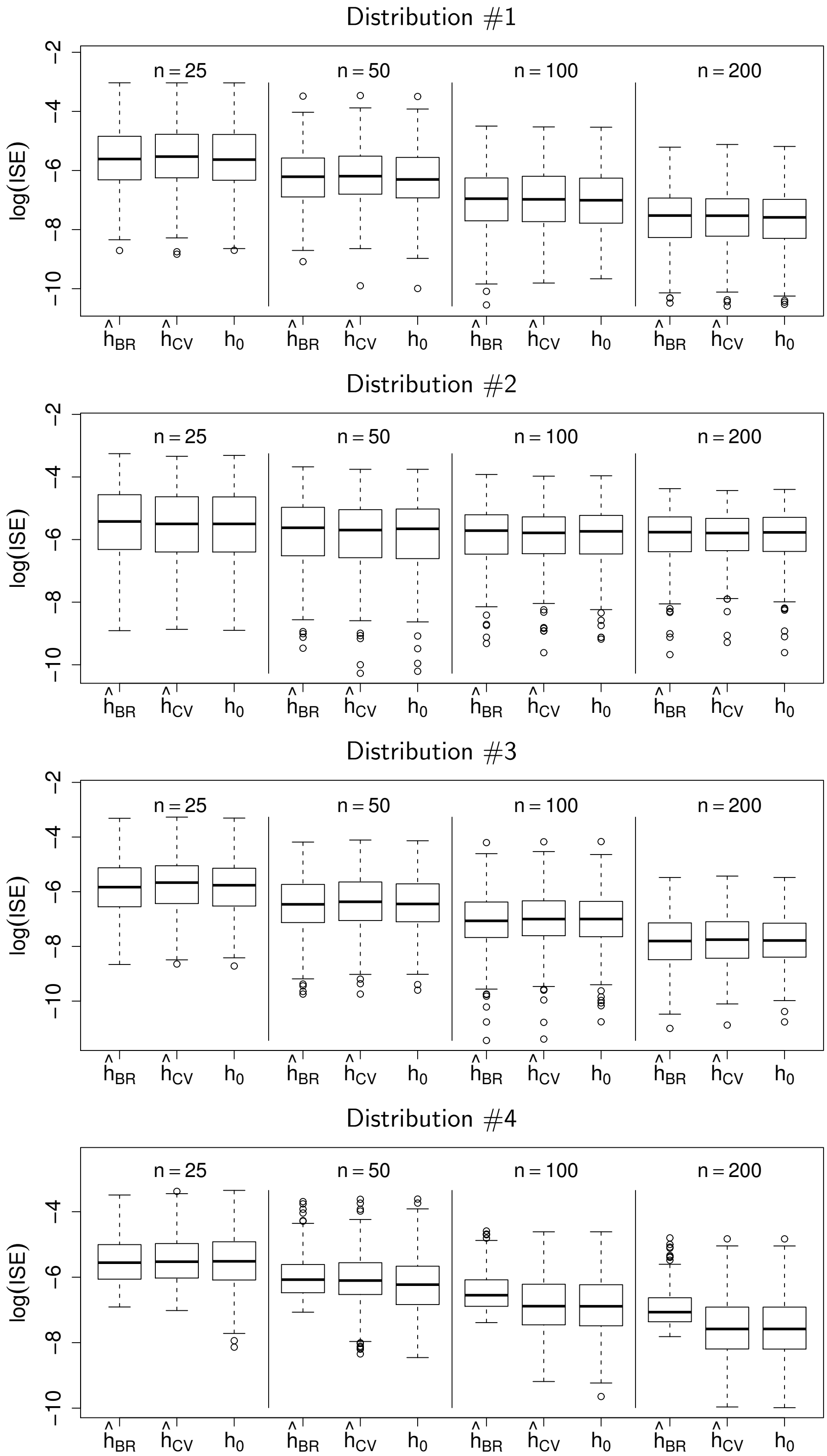

Figure 2: Integrated squared error results for the smoothing parameters $h = \hat{h}_{\mathrm{BR}}$, $h = \hat{h}_{\mathrm{CV}}$ and $h = h_0$ and sample sizes $n = 25, 50, 100$ and 200. K is the Epanechnikov density function. The number of replications for each case is 500.

In the implementation of cross-validation method the minimisation of $\mathrm{CV}(h)$ was confined to the interval $[\hat{h}_{\mathrm{BR}}/10, 1/2]$. The previous integrals have been numerically evaluated using the composite Simpson's rule. The Epanechnikov density $K(t) = \frac{3}{4}(1-t^2)I(|t| \leq 1)$ was taken as kernel function and we restrict our attention to the boundary kernels defined by $K^\ell(u,\alpha) = K(u/\alpha)/\alpha$ for $\ell = L, R$ (see Example 2.3). The integrated squared error empirical distributions (log scale) are presented in Figure 2.

For all the considered test distributions, Figure 2 suggests that the cross-validation bandwidth performs quite well showing a performance close to that one of the oracle estimator with bandwidth h_0. Additionally, for distributions #1, #2 and #3 there is no indication of significant differences between the bandwidths $\hat{h}_{\mathrm{CV}}$ and $\hat{h}_{\mathrm{BR}}$. This can be seen as an evidence of the well-known fact that smoothing has only a second order effect in kernel distribution function estimation. For the beta mixture #4 the cross-validation approach is clearly more effective than the beta optimal smoothing for large sample sizes. This distribution presents features that are not revealed until the sample size is above some threshold which explains the fact that both methods performed similarly for small sample sizes but not for large ones. In this latter case the cross-validation method is able to adapt to distributional shape while the beta distribution reference method does not reveal such a property.

In conclusion, we can say that the cross-validation bandwidth reveals a very good performance, although the simpler and less time consuming beta reference distribution bandwidth shows it self to be quite effective for the generality of test distributions.

6. PROOFS

Proof of Lemma 3.1: We start by the analysis of the stochastic component $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$. For that we follow the approach by Winter [37]. In order to deal with kernels that could have negative values, we need the following version of the integration by parts result presented by Winter [37, Lemma 2.1].

Lemma 6.1. *If Φ is a probability distribution function and*

$$\Psi(u) = \int_{]-\infty,u]} \psi(v)dv,$$

where ψ is a Lebesgue integrable function with $\int \psi(v)dv = 1$, then

$$\int \Phi d\Psi + \int \Psi d\Phi = 1.$$

Proof: Denoting by μ_Φ and μ_Ψ the finite signed measures defined by $\mu_\Phi(]-\infty,x]) = \Phi(x)$ and $\mu_\Psi(]-\infty,x]) = \Psi(x)$, for all $x \in \mathbb{R}$, it is enough to apply Fubini's theorem to the indicator function $(s,t) \to I(s > t)$ which is integrable with respect to the product measure $\mu_\Phi \otimes \mu_\Psi$. $\square$

Returning to the proof of Lemma 3.1, for $x \in]a, a+h[$, we have

$$\tilde{F}_{nh}(x) = \int \bar{K}^L((x-y)/h; (x-a)/h) dF_n(y) = 1 - \int \Psi_{x,h}(y) dF_n(y),$$

$$\mathrm{E}\tilde{F}_{nh}(x) = \int \bar{K}^L((x-y)/h; (x-a)/h) dF(y) = 1 - \int \Psi_{x,h}(y) dF(y),$$

and

$$\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x) = \int \Psi_{x,h}(y)(dF(y) - dF_n(y)),$$

where $\Psi_{x,h}(u) = \int_{]-\infty,u]} \psi_{x,h}(v) dv$ with $\psi_{x,h}(v) = K^L((x-v)/h; (x-a)/h)/h$.

From Lemma 6.1 we get

$$\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x) = \int \{F_n(y) - F(y)\} d\Psi_{x,h}(y),$$

and therefore

$$\sup_{x \in]a,a+h[} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F|| \sup_{\alpha \in]0,1[} \int |K^L(u;\alpha)|\, du \tag{6.1}$$

because

$$\sup_{x \in]a,a+h[} \int d|\Psi_{x,h}|(y) = \sup_{x \in]a,a+h[} \int |\psi_{x,h}(u)|\, du \leq \sup_{\alpha \in]0,1[} \int |K^L(u;\alpha)|\, du.$$

Similarly, we get

$$\sup_{x \in]b-h,b[} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F|| \sup_{\alpha \in]0,1[} \int |K^R(u;\alpha)|\, du, \tag{6.2}$$

and the standard approach (see Winter [37]) can be used for $x \in [a+h, b-h]$, in order to obtain

$$\sup_{x \in [a+h,b-h]} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F||. \tag{6.3}$$

Finally, from (6.1), (6.2) and (6.3) we obtain the upper bound (3.1) for $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$.

In the analysis of the bias component $||\mathrm{E}\tilde{F}_{nh} - F||$, we first note that, for $x \in]a, a+h[$, the expectation of $\tilde{F}_{nh}(x)$ is given by

$$\begin{aligned}\mathrm{E}\tilde{F}_{nh}(x) &= \int \bar{K}^L((x-y)/h; (x-a)/h) f(y)\,dy \\ &= \iint K^L(u; (x-a)/h) f(y) I(y \le x - uh)\,dudy \\ &= \int F(x-uh) K^L(u; (x-a)/h)\,du.\end{aligned}$$

Therefore,

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = \int \{F(x-uh) - F(x)\} K^L(u; (x-a)/h)\,du, \tag{6.4}$$

which leads to

$$\sup_{x\in]a,a+h[} |\mathrm{E}\tilde{F}_{nh}(x) - F(x)| \le \sup_{x,y\in[a,b]:\,|x-y|\le h} |F(x) - F(y)| \sup_{\alpha\in]0,1[} \int |K^L(u;\alpha)|\,du.$$

Additionally, if F' is continuous on $[a, b]$, from the Taylor formula we have

$$F(x-uh) - F(x) = -uhF'(x) - uh \int_0^1 \{F'(x-tuh) - F'(x)\}\,dt.$$

Using the fact that $\int uK^L(u;\alpha)du = 0$, for all $\alpha \in]0,1[$, from (6.4) we get

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = -h \int \{F'(x-tuh) - F'(x)\} uK^L(u; (x-a)/h)\,du$$

which leads to

$$\sup_{x\in]a,a+h[} |\mathrm{E}\tilde{F}_{nh}(x) - F(x)| \le h \sup_{x,y\in[a,b]:\,|x-y|\le h} |F'(x) - F'(y)| \sup_{\alpha\in]0,1[} \int |K^L(u;\alpha)|\,du.$$

A similar analysis can be carried out for the cases $x \in [a+h, b-h]$ and $x \in]b-h, b[$, leading to the bounds (3.2) and (3.3) for the bias term $||\mathrm{E}\tilde{F}_{nh} - F||$. □

Proof of Theorem 4.1: We start by the analysis of the bias term $\tilde{\mathbf{B}}(F;h) = \int \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2 dx$. By using the continuity of the second derivative of F and the Taylor expansion

$$F(x-uh) - F(x) = -uhF'(x) + u^2h^2 \int_0^1 (1-t) F''(x-tuh)\,dt,$$

from (6.4) we get

$$\begin{aligned}&\int_a^{a+h} \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2\,dx \\ &= h^5 \int_0^1 \left(\iint_0^1 (1-t) F''(a + \alpha h - tuh) u^2 K^L(u;\alpha)\,dtdu \right)^2 d\alpha \\ &\le h^5\, ||F''||^2 \int_0^1 \left(\int |K^L(u;\alpha)| du \right)^2 d\alpha = O\left(h^5\right).\end{aligned}$$

A similar upper bound can be obtained for the term $\int_{b-h}^{b}\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2dx$. The stated expansion for $\tilde{\mathbf{B}}(F;h)$ follows now from the dominated convergence theorem:

$$\begin{aligned}
&\int_{a+h}^{b-h}\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2dx \\
&=\int_{a+h}^{b-h}\left(\int\{F(x-uh)-F(x)\}K(u)\,du\right)^2dx \\
&=h^4\int_{a+h}^{b-h}\left(\int\!\!\int_0^1(1-t)F''(x-tuh)u^2K(u)\,dtdu\right)^2dx \\
&=\frac{h^4}{4}\left(\int u^2K(u)\,du\right)^2||F''||_2^2+o\left(h^4\right).
\end{aligned}$$

The analysis of the variance term, $\tilde{\mathbf{V}}(F;h)=\int \mathrm{Var}\,\tilde{F}_{nh}(x)dx$, can be made easy by considering the expansion

$$\begin{aligned}
n\,\mathrm{Var}\,\tilde{F}_{nh}(x)&=F(x)(1-F(x))+\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,du \\
&\quad-\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2-2\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}F(x),
\end{aligned}\tag{6.5}$$

where $B_{x,h}$ is defined as $K_{x,h}$ with K replaced by B. In fact, from the first part of the proof we conclude that the integral over $[a,b]$ of the last two terms is of order $O(h^2)$, and from standard arguments we get

$$\int_{a+h}^{b-h}\!\!\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,dudx=-h\int uB(u)\,du+O\left(h^2\right)$$

and

$$\begin{aligned}
&\left|\int_a^{a+h}\!\!\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,dudx\right| \\
&\qquad\le h^2\,||F'||\int_0^1\!\!\int|u||B^L(u;\alpha)|\,dud\alpha \\
&\qquad\le h^2||F'||\int_0^1\left(\int|K^L(u;\alpha)|du\right)^2d\alpha=O\left(h^2\right).
\end{aligned}$$

Taking into account that the same order of convergence can be obtained for the term $\int_{b-h}^{b}\int\{F(x-uh)-F(x)\}B_{x,h}(u)dudx$, we finally get the stated expansion for $\tilde{\mathbf{V}}(F;h)$. □

Proof of Theorem 4.2: We shall restrict our attention to the case where F is the uniform distribution function on the interval $[a,b]$. From Remark 3.3 and equality (6.5) we get

$$\mathrm{MISE}(F;h)=\frac{b-a}{6n}-\frac{h}{n}\left(\int uB(u)\,du-h\,\frac{C_B}{b-a}\right),$$

for $0 < h \leq (b-a)/2$. It is now easy to conclude that

$$h_0 = \frac{b-a}{2} \min\left(1\,, \int uB(u)\,du/C_B\right)$$

is the minimiser of $\mathrm{MISE}(F;h)$, for $0 < h \leq (b-a)/2$. $\square$

ACKNOWLEDGMENTS

Research partially supported by the Centre of Mathematics of the University of Coimbra funded by the European Regional Development Fund through the program COMPETE and by the Portuguese Government through the Fundação para a Ciência e Tecnologia under the project PEst-C/MAT/UI0324/2011.

REFERENCES

[1] Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation, *Journal of Statistical Planning and Inference*, **46**, 195–214.

[2] Bowman, A.; Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions, *Biometrika*, **85**, 799–808.

[3] Chacón, J.E. and Rodríguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator, *Statistics & Probability Letters*, **80**, 1414–1419.

[4] Chung, K.-L. (1949). An estimate concerning the Kolmogoroff limit distribution, *Transactions of the American Mathematical Society*, **67**, 36–50.

[5] Degenhardt, H.J.A. (1993). Chung-Smirnov property for perturbed empirical distribution functions, *Statistics & Probability Letters*, **16**, 97–101.

[6] Dvoretzky, A.; Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Annals of Mathematical Statistics*, **27**, 642–669.

[7] Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statistica Neerlandica*, **37**, 73–83.

[8] Fernholz, L.T. (1991). Almost sure convergence of smoothed empirical distribution functions, *Scandinavian Journal of Statistics*, **18**, 255–262.

[9] Gasser, T. and Müller, H.-G. (1979). *Kernel estimation of regression functions*. In “Smoothing techniques for curve estimation” (T. Gasser and M. Rosenblatt, Eds.), Lecture Notes in Mathematics 757, 23–68.

[10] Gasser, T.; Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society. Series B. Methodological*, **47**, 238–252.

[11] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation, *Probability Theory and Related Fields*, **143**, 569–596.

[12] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions 2*, Wiley, New York.

[13] Jones, M.C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statistics & Probability Letters*, **9**, 129–132.

[14] Karunamuni, R.J. and Alberts, T. (2005). On boundary correction in kernel density estimation, *Statistical Methodology*, **2**, 191–212.

[15] Karunamuni, R.J. and Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator, *Statistics & Probability Letters*, **78**, 499–507.

[16] Liu, R. and Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function, *Journal of Nonparametric Statistics*, **20**, 661–677.

[17] Marron, J.S. and Nolan, D. (1989). Canonical kernels for density estimation, *Statistics & Probability Letters*, **7**, 191–195.

[18] Mason, D.M. and Swanepoel, J.W.H. (2011). A general result on the uniform in bandwidth consistency of kernel-type function estimators, *TEST*, **20**, 72–94.

[19] Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, **78**, 521–530.

[20] Nadaraya, E.A. (1964). Some new estimates for distribution functions, *Theory of Probability and Its Applications*, **9**, 497–500.

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

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

[23] R Development Core Team (2011). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`

[24] Rice, J. (1984). Boundary modification for kernel regression, *Communications in Statistics – Theory and Methods*, **13**, 893–900.

[25] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Annals of Mathematical Statistics*, **27**, 832–837.

[26] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference*, **35**, 65–75.

[27] Shirahata, S. and Chu, I-S. (1992). Integrated squared error of kernel-type estimator of distribution function, *Annals of the Institute of Statistical Mathematics*, **44**, 579–591.

[28] Singh, R.S.; Gasser, T. and Prasad, B. (1983). Nonparametric estimates of distribution function, *Communications in Statistics – Theory and Methods*, **12**, 2095–2108.

[29] Smirnov, N.V. (1944). Approximate laws of distribution of random variables from empirical data (in Russian), *Uspehi Matematičeskih Nauk*, **10**, 179–206.

[30] Swanepoel, J.W.H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function, *Communications in Statistics – Theory and Methods*, **17**, 3785–3799.

[31] Tenreiro, C. (2003). On the asymptotic behaviour of the ISE for automatic kernel distribution estimators, *Journal of Nonparametric Statistics*, **15**, 485–504.

[32] Tenreiro, C. (2006). Asymptotic behaviour of multistage plug-in bandwidth selectors for kernel distribution function estimators, *Journal of Nonparametric Statistics*, **18**, 101–116.

[33] Tiago de Oliveira, J. (1963). Estatística de densidades: resultados assintóticos, *Revista da Faculdade de Ciências de Lisboa*, **9**, 111–206.

[34] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[35] Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis II, *Sankhyā Ser. A*, **26**, 101–116.

[36] Winter, B.B. (1973). Strong uniform consistency of integrals of density estimators, *Canadian Journal of Statistics*, **1**, 247–253.

[37] Winter, B.B. (1979). Convergence rate of perturbed empirical distribution functions, *Journal of Applied Probability*, **16**, 163–173.

[38] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, *Bulletin of Mathematical Statistics*, **15**, 69–78.