INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0716

NONPARAMETRIC DENSITY ESTIMATION FOR MULTIVARIATE BOUNDED DATA

BOUEZMARNI, T. and J. ROMBOUTS

NONPARAMETRIC DENSITY ESTIMATION FOR MULTIVARIATE BOUNDED DATA

Taoufik Bouezmarni* Jeroen V.K. Rombouts†

August 27, 2007

Abstract

We propose a new nonparametric estimator for the density function of multivariate bounded data. As frequently observed in practice, the variables may be partially bounded (e.g., nonnegative) or completely bounded (e.g., in the unit interval). In addition, the variables may have a point mass. We reduce the conditions on the underlying density to a minimum by proposing a nonparametric approach. By using a gamma, a beta, or a local linear kernel (also called boundary kernels), in a product kernel, the suggested estimator becomes simple in implementation and robust to the well known boundary bias problem. We investigate the mean integrated squared error properties, including the rate of convergence, uniform strong consistency and asymptotic normality. We establish consistency of the least squares cross-validation method to select optimal bandwidth parameters. A detailed simulation study investigates the performance of the estimators. Applications using lottery and corporate finance data are provided.

Key words and phrases. Asymmetric kernels, multivariate boundary bias, nonparametric multivariate density estimation, asymptotic properties, bandwidth selection, least squares cross-validation.

We would like to thank Luc Bauwens, Aurore Delaigle, Hariharan Dharmarajan, Jeffry Racine and Vanessa Sumo for useful comments.

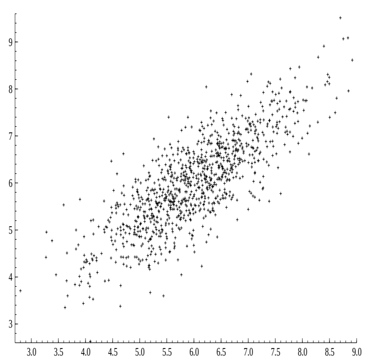
*HEC Montréal, CREF, Institute of Statistics (Université catholique de Louvain).

†Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, CORE (Université catholique de Louvain), CREF, 3000 Cote Sainte Catherine, H3T 2A7 Montréal (QC), Canada.

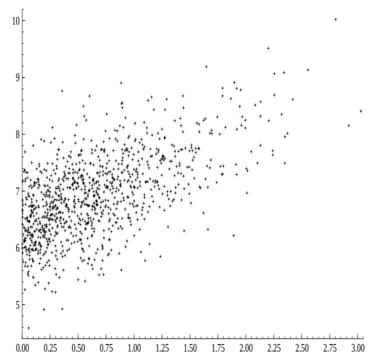
1 Introduction

Among multivariate nonparametric density estimators, the standard Gaussian kernel is the most popular. The estimator has excellent asymptotic properties; see Silverman (1986), Scott (1992), and Wand and Jones (1995) for more details. However, the estimator does not take into account the potential finite support of the variables. When the support of some variables is bounded, for example, in the case of nonnegative data, the standard kernel estimator continues to give weight outside the supports. This causes a bias in the boundary region. The boundary bias problem of the standard kernel is well documented in the univariate case. An initial solution to the boundary problem is given by Schuster (1985), who proposes the reflection method. Müller (1991), Lejeune and Sarda (1992), Jones (1993), Jones and Foster (1996), and Cheng, Fan, and Marron (1997) suggest the use of adaptive and boundary kernels at the edges and a fixed standard kernel in the interior region. Marron and Ruppert (1994) investigate some transformations before using the standard kernels, and Cowling and Hall (1996) propose a pseudodata method. Recently, Chen (2000), Bouezmarni and Scaillet (2003), and Bouezmarni and Rombouts (2006) study the gamma kernels for univariate nonnegative data. For data defined on the unit interval, Chen (1999) proposes to use a beta kernel.

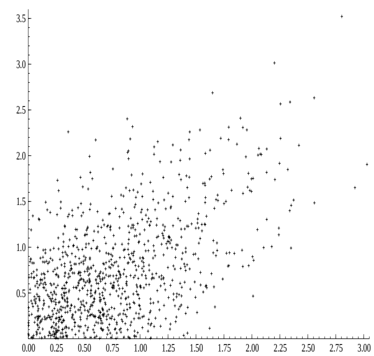
The boundary bias problem becomes more severe in the multivariate case because the boundary region increases with the dimension of the support. Boundary regions are illustrated in Figure 1 for bivariate nonnegative data. In panel (a) there is no boundary problem because the data are



(a) No boundary problem



(b) One boundary problem

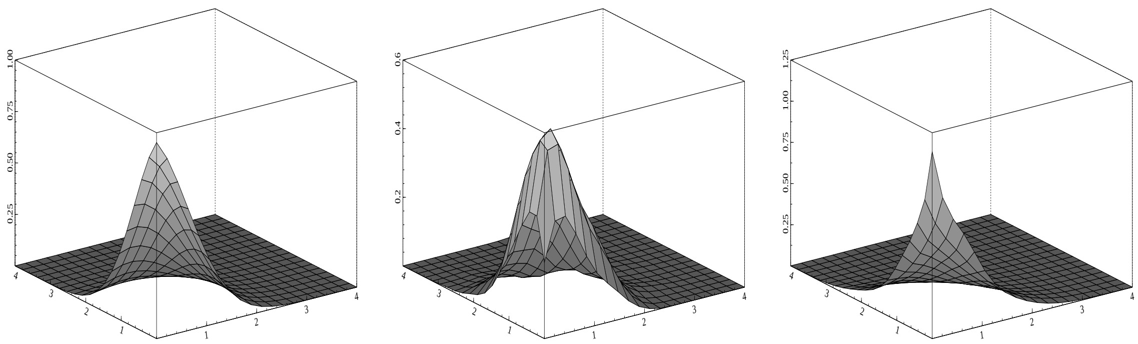


(c) Two boundary problems

Figure 1: Illustration of boundary regions for bivariate nonnegative data.

far away from zero. In this case, the standard kernel has the best performance. However, we will see in the simulations of this paper that the estimator we propose is very close to this optimal performance. In panel (b) and in particular panel (c) of Figure 1, the standard kernel has poor performance because it underestimates the density in a large area of the support. This severe under performance in the case of two boundary problems is further illustrated in Figure 2.

Although the consequences of the boundary problem in multivariate dimensions are much more



(a) True density (b) Gaussian kernel (c) Gamma kernel

Figure 2: Density estimation with Gaussian kernel, gamma kernel and local linear kernel.

severe, solutions to the problem are not well investigated. Müller and Stadtmüller (1999) propose boundary kernels for multivariate data defined on arbitrary support by selecting the kernels that minimize a variational problem. In fact, they extend the minimum variance selection principle kernel used to select the optimal kernel in the interior region, as in Epanechnikov (1969) and Granovsky and Müller (1991). In the nonparametric regression context, the problem of boundary bias is developed by Gasser, Müller, and Mammitsch (1985), and Zhang, Karunamuni, and Jones (1999) for the univariate case, and Fan and Gijbels (1992), Ruppert (1994), Staniswalis, Messer, and Finston (1993), and Staniswalis and Messer (1997) for multivariate data.

This paper proposes a nonparametric product kernel estimator for density functions of multivariate bounded data. Estimation is based on a gamma kernel or a local linear kernel when the support of the variable is nonnegative and a beta kernel when the support is a compact set. By doing so, no weight is assigned outside the support of the underlying density so that the estimators are robust to the boundary problem. The method is easy in conception and implementation. We provide the asymptotic properties of these estimators and show that the optimal rate of convergence of the mean integrated squared error is obtained. For the multivariate uniform density, we show that the estimator we propose using beta kernels is unbiased. We examine the finite sample performance in several simulations. As for any nonparametric kernel estimator, the performance is sensitive to the choice of the bandwidth parameters. We suggest the application of the least squares cross-validation method to select these parameters. We prove the consistency of this method for the proposed estimators and investigate its performance in the simulations.

The rest of the paper is organized as follows. We introduce the multivariate nonparametric estimator for multivariate bounded data in Section 2. Section 3 provides convergence properties. The consistency of the least squares cross-validation bandwidth selection method is established in Section 4. In Section 5 we investigate the finite sample properties of several kernel estimators for nonnegative bivariate data. Section 6 contains two applications, one with lottery data and another with corporate finance data. Section 7 concludes. The proofs of the theorems are presented in the

Appendix.

2 Nonparametric estimator

Let $\{(X_i^1, ..., X_i^d), i = 1, .., n\}$ be a sample of independent and identically distributed random variables with an unknown density function f. The general multivariate nonparametric density estimator is given by

$$\hat{f}(x_1, ..., x_d) = \frac{1}{nh_1...h_d} \sum_{i=1}^{n} \mathbf{K}\left(\frac{x_1 - X_{i1}}{h_1}, ..., \frac{x_d - X_{id}}{h_d}\right),$$

where $\mathbf{K}$ denotes a multivariate kernel function and $(h_1, ..., h_d)$ the vector of bandwidth parameters.

In practice the choice of $\mathbf{K}$ is especially difficult when the supports of the random variables are potentially unequal. Therefore, we propose to use the product kernel estimator with adapted and flexible kernels in order to solve the boundary bias problem. The estimator is defined as

$$\hat{f}(x_1, ..., x_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{s=1}^{d} K^s(b_s, X_{is})(x_s), \tag{1}$$

where $b_1, ..., b_d$ are the bandwidth parameters and the kernel K^s is a kernel for variable s. Throughout the paper, this superscript s will be omitted for notational convenience. As described in the introduction, the kernel for each variable is chosen to be the standard kernel, which is indeed optimal when the total support is R^d. We consider two cases for random variables with bounded support.

First, when the support of the variable is nonnegative, we propose the use of either the local linear kernel denoted by K_L or one of the two gamma kernels K_G, K_{NG} as shown below. Thus,

$$K_L(h, t)(x) = K_l\left(x, h, \frac{x - t}{h}\right),$$

where

$$K_l(x, h, t) = \frac{a_2(x, h) - a_1(x, h)t}{a_0(x, h)a_2(x, h) - a_1^2(x, h)} K(t),$$

K is any symmetric kernel with a compact support $[-1, 1]$ and

$$a_s(x, h) = \int_{-1}^{x/h} t^s K(t) dt.$$

The kernels K_G and K_{NG} are respectively defined as

$$K_G(b, t)(x) = \frac{t^{x/b} \exp(-t/b)}{b^{x/b+1}\Gamma(x/b + 1)},$$

and

$$K_{NG}(b, t)(x) = \frac{t^{\rho(x)-1} \exp(-t/b)}{b^{\rho(x)}\Gamma(\rho(x))},$$

where,

$$\rho(x) = \begin{cases} x/b & \text{if} \quad x \geq 2b \\ \\ \frac{1}{4}(x/b)^2 + 1 & \text{if} \quad x \in [0, 2b). \end{cases}$$

Second, when the variable has a compact support (for simplicity we take here the unit interval) we suggest the use of the beta kernel

$$K_B(b,t)(x) = B\left(x/b + 1, (1-x)/b + 1\right),$$

or a modified beta kernel

$$K_{NB}(b,t)(x) = \begin{cases} B(\rho(x), (1-x)/b) & \text{if} \quad x \in [0, 2b) \\ B(x/b, (1-x)/b) & \text{if} \quad x \in [2b, 1-2b] \\ B(x/b, \rho(1-x)) & \text{if} \quad x \in (1-2b, 1], \end{cases}$$

where $B(\alpha, \beta)$ is the beta density function with parameters α and β, b is the smoothing parameter, and $\rho(x) = 2b^2 + 2.5 - \sqrt{4b^4 + 6b^2 + 2.25 - x^2 - x/b}$.

Note that finally it may also happen that there is a point mass on the boundary with probability p. For example, Grullon and Michaely (2002) study dividends and share repurchases, which both have multiple zero observations. In this situation we suggest the estimation of the probability p by the observed proportion and the density function in $(0, +\infty)$, by using the normalized kernel

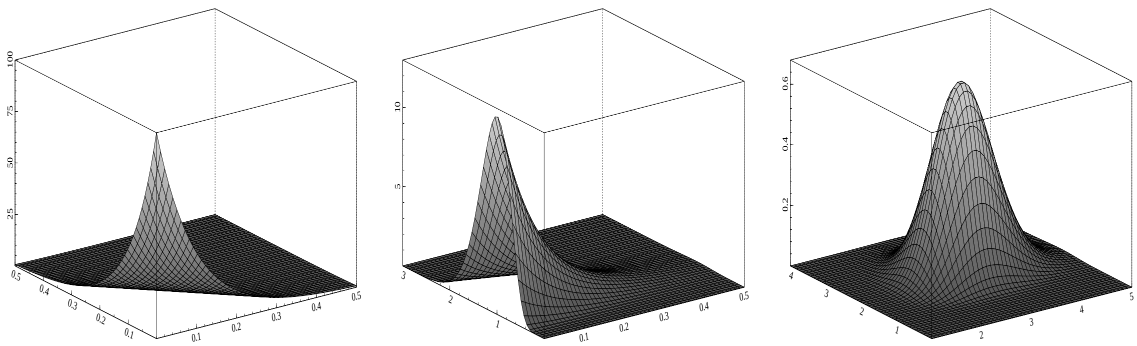
$$K^*(b,t)(x) = \frac{K(b,t)(x)}{\int_0^\infty K(b,t)(s)ds},$$

with the kernel K as defined above. This case is studied, for example, by Gourieroux and Monfort (2006) for the beta kernel estimator with applications to credit risk data.

For the bivariate nonnegative data case and for the gamma kernel, Figure 3 illustrates the flexibility of those kernels. They are asymmetric at the boundary points and symmetric away from the boundaries. The kernel never assigns weight outside the support and is therefore free of boundary bias. To conclude, the nonparametric estimator we propose allows the kernels in the product kernel to be based on the support of the underlying variables. We can, for example, combine beta with gamma kernels if the supports of the variables are the unit interval and the positive real line, respectively. Furthermore, the nonparametric estimator with a gamma or beta kernel is always nonnegative, while the Müller and Stadtmüller (1999) estimator can be negative. The latter estimator also requires an additional bandwidth parameter and a weighting function.

3 Convergence properties

In this section we establish the main asymptotic properties of the nonparametric estimator described in the previous section. We consider the case where all the variables have nonnegative



(a) (x,y)=(0,0) (b) (x,y)=(0,1) (c) (x,y)=(3,2)

Figure 3: Shape of gamma kernels at (x, y). The bandwidth parameters $b_1 = b_2 = 0.2$.

supports though the results can easily obtained for any other combination. Some assumptions on the bandwidth parameters are given first.

Assumptions on the bandwidth parameters

B1. $a_j \to 0,\ j = 1, ..., d$ and $n^{-1} \prod_{j=1}^{d} a_j^{-\frac{1}{2}} \to 0, \quad \text{as } n \to \infty.$

B2. $a_j \to 0,\ j = 1, ..., d$ and $\log(n) n^{-1} \prod_{j=1}^{d} a_j^{-2} \to 0, \quad \text{as } n \to \infty.$

The following result states the mean integrated squared error (MISE) of the nonparametric estimator.

Theorem 1. mean integrated squared error of $\hat{f}$

Suppose that f is twice differentiable. Let $\hat{f}$ be the nonparametric estimator with the gamma kernel. Under assumption B1

$$MISE = \int (\sum_{j=1}^{d} a_j B_j(x))^2 dx + \frac{1}{n} \left(\prod_{i=1}^{d} a_j^{-1/2} \right) \int V(x) dx + o\left(\sum_{j=1}^{d} a_j \right)^2 + o\left(n^{-1} \prod_{j=1}^{d} a_j^{-1/2} \right),$$

where $a_i = b_i$ and

$$B_j(x) = \left(f^j(x) + \frac{x_j f^{jj}(x)}{2} \right) \quad \text{and} \quad V(x) = (2\sqrt{\pi})^{-d} f(x) \prod_{j=1}^{d} x_j^{-1/2},$$

with $f^j = \frac{\partial f}{\partial x_j}$ and $f^{jj} = \frac{\partial^2 f}{\partial x_j^2}$.

The optimal bandwidths that minimize the asymptotic mean integrated squared error are

$$a_j^* = c_j n^{-\frac{2}{d+4}}, \quad \text{for some positive constants } c_1, ..., c_d. \tag{2}$$

Therefore, the optimal asymptotic mean integrated squared error is

$$AMISE^* = \left\{ \int (\sum_{j=1}^{d} c_j B_j(x))^2 dx + \left(\prod_{j=1}^{d} c_j^{-1/2} \right) \int V(x) dx \right\} n^{-\frac{4}{d+4}}.$$

■

Theorem 1 proves that the rate of convergence of the bias of the nonparametric estimator is uniform, hence it is free of boundary bias. The rate of convergence of the mean integrated squared error becomes slower when the dimension of the random variable increases. This is known as the curse of dimensionality. We see that in the boundary region, the variance of the product kernel estimator is larger in comparison with the variance in the interior region. However, the increase of the variance is compensated by a smaller bias in this region. Away from the boundaries, we have the opposite effect, that is, a lower variance and a slightly higher bias. Fortunately, the second derivative of the density function is negligible away from zero.

Remark 1. *If we suppose that $a = a_1 = ... = a_d$, the optimal bandwidth is*

$$a_i^* = \left(\frac{d \int V(x) dx}{4 \int B(x) dx} \right)^{\frac{2}{d+4}} n^{-\frac{2}{d+4}}$$

and the optimal asymptotic mean integrated squared error is

$$AMISE^* = (d/4+1)(d/4)^{-\frac{d}{d+4}} \left(\int V(x) dx \right)^{\frac{4}{d+4}} \left(\int B(x) dx \right)^{\frac{d}{d+4}} n^{-\frac{4}{d+4}}.$$

■

The following remark states the MISE of nonparametric estimator with local linear and the new gamma kernel.

Remark 2.

- *For the local linear estimator, the results of Theorem 1 remain valid with $a_j = h_j^2$ and*

$$B_j(x) = \frac{\kappa_2}{2} f^{jj}(x) \quad \text{and} \quad V(x) = \kappa^d,$$

  where $\kappa_2 = \int x^2 K(x) dx$ and $\kappa_2 = \int K^2(x) dx$.

- *For the new gamma estimator, the results of Theorem 1 remain valid with $a_j = b_j$ and*

$$B_j(x) = \frac{x_j f^{jj}(x)}{2} \quad \text{and} \quad V(x) = (2\sqrt{\pi})^{-d} f(x) \prod_{j=1}^{d} x_j^{-1/2}.$$

For the uniform density on $[0,1]^d$, the rate of convergence of the mean integrated square error of the standard kernel becomes $O(n^{-2/(d+2)})$. Devroye and Györfi (1985) established a corresponding result for the mean integrated absolute error in the univariate case. This lower rate is due to the decrease in the rate of convergence of the bias. For data on $[0,1]^d$, our estimator uses beta kernels. The next proposition states that our estimator is unbiased and that a large bandwidth is needed. In fact, when the bandwidth parameter tends to infinity the beta kernel becomes the uniform density.

Proposition 1. MISE when f is a uniform density

Suppose that f is the uniform density on $[0,1]^d$. Then, the nonparametric estimator $\hat{f}$ with beta kernels is an unbiased estimator for f and its integrated variance is given by:

$$IV(\hat{f}) = \left(\frac{\sqrt{\pi}}{2}\right)^d n^{-1} \prod_{i=1}^{n} b_i^{-1/2}$$

■

The following theorem establishes the uniform strong consistency of the nonparametric density estimator using gamma kernels.

Theorem 2. Uniform strong consistency of $\hat{f}$

Let f be a continuous and bounded probability density function. Under assumption B2, for any compact set I in $[0,+\infty)$, we have

$$\sup_{t\in I} \left|\hat{f}(x) - f(x)\right| \xrightarrow{a.s.} 0 \quad as \quad n \longrightarrow +\infty.$$

■

The following theorem deals with the asymptotic normality of the nonparametric density estimator of the gamma kernel estimator.

Theorem 3. Asymptotic normality of $\hat{f}$

Suppose that $f_1, ..., f_d$ are twice differentiable at x. Also suppose that the bandwidth parameters satisfy (2). Then we have

$$\sigma^{*-1} n^{\frac{1}{2}} \prod_{j=1}^{d} b_j^{\frac{1}{4}} \left(\hat{f}(x) - f(x) - \mu^*\right) \xrightarrow{\mathcal{D}} N(0,1), \tag{3}$$

where

$$\sigma^*(x) = \left(f(x)(\prod_{j\in I} (2\sqrt{\pi})^{-1} x_j^{-1/2})(\prod_{l\in I^C} \left(\frac{\Gamma(2\kappa_j+1)}{2^{2\kappa_j+1}\Gamma^2(\kappa_j+1)}\right) b_j^{-1/2}) \right)^{\frac{1}{2}}$$

with $I = \{j, x_j/b_j \to \infty\}$ *and* I^C *its complement, and*

$$\mu^* = \sum_{j=1}^{d} b_j B_j(x).$$

■

The next remark gives the asymptotic normality of the product kernel with the local linear kernel and the new gamma kernel.

Remark 3. *The asymptotic normality in (3) remains valid*

- *With the local linear kernel, with* $b_j = h_j^2$

$$B_j = \frac{s_2^2(p_j) - s_1(p_j)s_3(p_j)}{s_2(p_j)s_0(p_j) - s_1^2(p_j)} \frac{f''(x_j)}{2} \quad and \quad \sigma^*(x) = \sqrt{f(x)\frac{s_2^2(p_j) - 2s_2(p_j)s_1(p_j)e_1(p_j) + s_1^2(p_j)e_2(p_j)}{(s_2(p_j)s_0(p_j) - s_1^2(p_j))^2}},$$

  where $p_j = x_j/h_j$, $s_i(p) = \int_{-1}^{p} u^i K(u)du$ *and* $e_i = \int_{-1}^{p} u^i K^2(u)du$, *and*

- *With the new gamma kernel, with the same* σ^* *and*

$$B_j == \begin{cases} \frac{1}{2} x_j f''(x_j) & \text{if } x_j \geq 2b_j \\ \\ \xi_{b_j}(x_j) f'(x_j) & x_j < 2b_j, \end{cases}$$

  where $\xi_b(x) = (1-x)(\rho(2,x) - x/b)/(1 + b\rho(2,x) - x)$.

4 Bandwidth Selection

Theorem 1 establishes the optimal bandwidth parameter, which cannot be used in practice because it depends on the unknown density function. In this section, we propose to use the least squares cross-validation (LSCV) method to select the bandwidth. This technique has been developed by several authors. For the Gaussian kernel estimator, its consistency is investigated by Rudemo (1982), Hall (1983), Stone (1984), Bowman (1984), Härlde and Marron (1985), and Marron and Härlde (1986). In this section, we show the performance of the LSCV method for the product kernel estimator based on the gamma kernel. We first explain how the method works. The LSCV method is based on the minimization of the integrated squared error which is defined as

$$ISE_h = \int \hat{f}_h^2(x)dx - 2\int \hat{f}(x) f_h(x)dx + \int f^2(x)dx.$$

Because the last term does not depend on the bandwidth parameter, minimizing the integrated squared error boils down to minimizing the two first terms. However, we need to estimate the second term since it depends on the unknown density function f. The LSCV estimator of $ISE_h - \int f^2$ is

$$LSCV_h = \int \hat{f}_h^2(x)dx - \frac{2}{n^2} \sum_{i \neq j} K(b, X_i)(X_j)$$

where

$$K(b, X_i)(X_j) = K(b_1, X_i^1)(X_j^1)...K(b_d, X_i^d)(X_j^1).$$

The bandwidth LSCV rule selection is defined as follows

$$\hat{b} = \text{argmin}_h \; LSCV_h.$$

Our aim is to prove that this choice is asymptotically optimal in terms of the mean integrated squared error. To establish this result, we need some additional assumptions. Because the optimal bandwidth parameter is of order $O(n^{-2/(d+2)})$, we suppose that $\hat{b} \in H_n$, where

$$H_n = \left\{ b, \alpha_1 n^{-2/(d+4)} < b_i < \alpha_2 n^{-2/(d+4)}, i = 1, .., d \right\}.$$

We assume that

$$\#(H_n) \leq An^{\alpha}, \text{where } A \text{ and } \alpha \text{ are positive constants} \tag{4}$$

and

$$\int \int \frac{(\sum_{i=1}^{\delta} x_i)^{\frac{1}{b}\sum_{i=1}^{\delta} x_i + \frac{1}{2}}}{\prod_{i=1}^{\delta} x_i^{\frac{x_i}{b}+\frac{1}{2}} \delta^{\frac{1}{b}\sum_{i=1}^{\delta} x_i + 1}} f\left(\frac{\sum_{i=1}^{\delta} x_i}{\delta}\right) dx_1..dx_\delta \; < \; \text{constant for some integer } \delta. \tag{5}$$

The following theorem states that $\hat{b}$ is asymptotically optimal.

Theorem 4. *Under condition (4) and (5), we have*

$$\frac{MISE_{\hat{b}}}{MISE_{b_0}} \longrightarrow 1, \quad \textit{almost surely,}$$

where b_0 is the bandwidth that minimizes the mean integrated squared error.

■

5 Finite sample properties

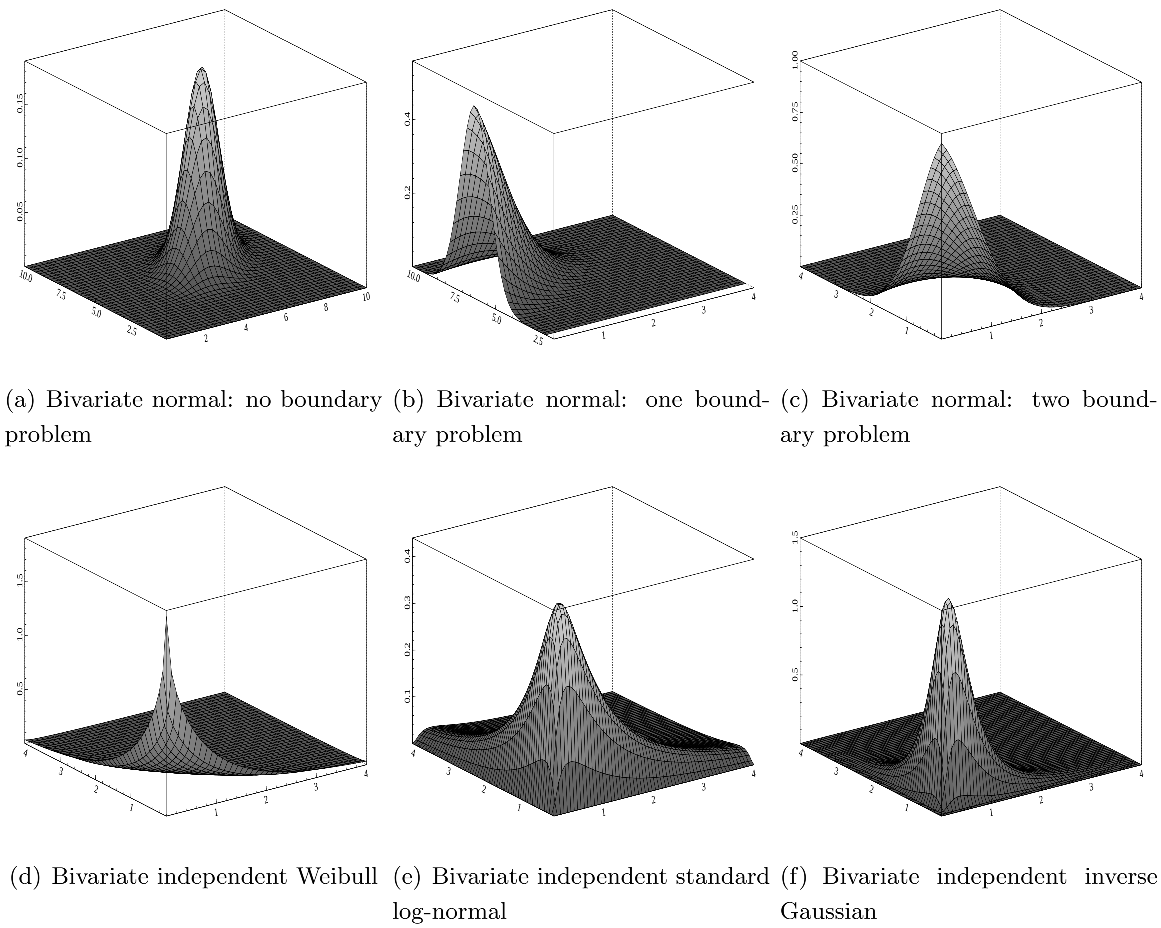
In this section we study the finite sample properties for the nonparametric density estimator for bivariate data with non-negative supports. We compare the two first moments of the mean integrated squared error distribution of the nonparametric product kernel estimator using the following kernels: Gaussian, Gaussian with log-transformation, local linear, gamma, and modified gamma. We consider the following six data generating processes:

- Model A: no boundary problem, bivariate normal density with mean $(\mu_1, \mu_2) = (6, 6)$ and variance $(\sigma_1^2, \sigma_2^2) = (1, 1)$ and correlation $r = 0.5$.

- Model B: one boundary problem, truncated bivariate normal density with mean $(\mu_1, \mu_2) = (-0.5, 6)$ and variance $(\sigma_1^2, \sigma_2^2) = (1, 1)$ and correlation $r = 0.5$.

- Model C: two boundary problems, truncated bivariate normal density with mean $(\mu_1, \mu_2) = (-0.5, -0.5)$ and variance $(\sigma_1^2, \sigma_2^2) = (1, 1)$ and correlation $r = 0.8$.
- Model D: two boundary problems, bivariate independent Weibull density with shape parameter 0.91 and scale parameter 1.
- Model E: bivariate independent standard log-normal.
- Model F: bivariate independent inverse Gaussian with mean $\mu = 0.8$ and the scaling parameter $\lambda = 1$.

From Figure 4, which displays the densities for the six models, we observe that we cover a wide range of shapes. In simulations, we consider the sample sizes 250 and 500 and perform 100



(a) Bivariate normal: no boundary problem

(b) Bivariate normal: one boundary problem

(c) Bivariate normal: two boundary problem

(d) Bivariate independent Weibull

(e) Bivariate independent standard log-normal

(f) Bivariate independent inverse Gaussian

Figure 4: Density functions considered for simulations.

replications for each model. In each replication the bandwidth is chosen such that the integrated squared error is minimized. For each model, the support of integration is specified such that the density is negligible outside this support. We report the mean and the standard deviation of the

Table 1: Mean of L_2 error for the density function estimators.

			Gaussian	Log-Trans.	Gamma	Mod. gam.	local linear
A	n=250	Mean	0.00305	0.00337	0.00318	0.00312	0.00288
		Std dev	0.00109	0.00104	0.00104	0.00103	0.00100
	n=500	Mean	0.00190	0.00233	0.00207	0.00203	0.00186
		Std dev	0.00066	0.00067	0.00066	0.00067	0.00064
B	n=250	Mean	0.02075	0.01647	0.01241	0.00998	0.00943
		Std dev	0.00514	0.00822	0.00393	0.00413	0.00386
	n=500	Mean	0.01733	0.01102	0.00960	0.00785	0.00773
		Std dev	0.00319	0.00435	0.00265	0.00267	0.00271
C	n=250	Mean	0.04011	0.03679	0.02373	0.01471	0.01483
		Std dev	0.00659	0.00873	0.00582	0.00462	0.00411
	n=500	Mean	0.03261	0.02648	0.01742	0.01087	0.01096
		Std dev	0.00398	0.00567	0.00401	0.00328	0.00282
D	n=250	Mean	0.06883	0.03212	0.01974	0.01554	0.01075
		Std dev	0.00476	0.01931	0.00482	0.00511	0.00280
	n=500	Mean	0.06216	0.02457	0.01626	0.01118	0.00841
		Std dev	0.00498	0.01407	0.00401	0.00381	0.00308
E	n=250	Mean	0.01729	0.00592	0.01054	0.01030	0.01554
		Std dev	0.00374	0.00221	0.00268	0.00320	0.00372
	n=500	Mean	0.01216	0.00424	0.00694	0.00653	0.01072
		Std dev	0.00228	0.00149	0.00170	0.00208	0.00247
F	n=250	Mean	0.02944	0.03145	0.02597	0.02352	0.02949
		Std dev	0.00808	0.01373	0.00786	0.00898	0.00749
	n=500	Mean	0.02138	0.02672	0.02107	0.01758	0.02119
		Std dev	0.00451	0.01178	0.00606	0.00501	0.00448

A: bivariate normal, B: truncated bivariate normal with one boundary problem, C: truncated bivariate normal with two boundary problems, D: two independent Weibull with two boundary problems, E: two independent standard log-normal, F: two independent inverse Gaussian. For each replication, the bandwidth parameter is chosen such that the integrated squared error is minimized. Std dev: standard deviation, Log-Trans.: log-transform estimator and Mod. gam.: modified gamma estimator.

mean integrated squared error in Table 1. As a general remark we observe that the mean and the variance of the MISE decreases for all models, as expected. Also as expected, the mean MISE increases when the boundary region becomes larger. For example, for $n = 500$, the mean MISE of the Gaussian kernel is 0.01733 in the case of one boundary problem (model B) and it becomes 0.06216 when there are two boundary problems (model D).

Next, we summarize the main findings for each model separately. For model A, the Gaussian kernel estimator is the best since there are no observations in the boundary region. In terms of mean MISE, the gamma and the modified gamma kernel estimators perform almost the same as the Gaussian kernel estimator, and they are better than the log-transform estimator. In terms of variance, the estimators have also almost the same performance. For example, for $n = 500$, the mean MISE for the Gaussian and gamma kernel is 0.0019 and 0.00207, respectively, and the standard deviation for both is 0.0066. For models B and C, the local linear and the modified gamma kernels have the same performance and dominate the other kernels in terms of mean and variance MISE, while the gamma kernel and especially the log-transform Gaussian kernel underperform. Obviously, the Gaussian kernel underperforms since there is a boundary bias problem. For model D, where the density function is unbounded at zero, the local linear estimator dominates especially for $n = 250$, followed by the modified gamma and then the gamma kernel estimator. The log-transformed Gaussian kernel is not a good estimator for this model. For example, for $n = 500$, the mean MISE for the log-transformed and the local linear kernels are 0.02457 and 0.00841, respectively. The performance of the log-transformed Gaussian kernel improves only mildly when the sample size increases. As expected, for the standard log normal density of model E, the log-transformed Gaussian kernel performs better than the others. For this model, the two gamma estimators dominate the local linear estimator. For the inverse Gaussian density (model F), the modified gamma kernel dominates clearly for both samples sizes, followed by the gamma kernel estimator. For $n = 500$, the gamma kernel, the local linear and the Gaussian kernel estimators have perform similarly and dominate the log-transformed kernel estimator.

For each replication in the simulation we also computed the optimal bandwidth by LSCV method whose consistency was established in the previous section. Tables 2 and 3 report details for the theoretical and LSCV bandwidth parameters for models A to D (models E and F are left out for the sake of brevity). First, we can observe that the local linear estimator bandwidths are larger than the bandwidths used by the Gaussian kernel. The same is true for the bandwidths of the modified gamma estimator compared with the gamma kernel estimator. We also find as expected that the variance decreases with the sample size. Comparing the bandwidths obtained from LSCV procedure and that implied by minimization of the theoretical MISE, we can observe that the means are quite close to each other. In terms of variance, the LSCV bandwidth estimates are more variable than the theoretical ones. This is not surprising, in particular for the local linear estimator; see Hall and Marron (1987), Scott and Terrell (1987), and Chiu (1991). For example, for model D and $n = 250$, the standard deviation becomes $(3.074, 2.067)$ instead of $(5.541, 5.665)$ for the theoretical case. We also remark that the two gamma kernel estimators are more stable than the local linear for the LSCV bandwidth selection procedure. The variation coefficient of the gamma kernel estimators are smaller than that of the local linear estimator. For example, for model B and $n = 250$, the variation coefficient for both bandwidths is (0.741,0.855) and (0.447, 0.803) for the local linear estimator and the modified gamma kernel estimator, respectively. We also want to draw attention to what happens with the bandwidths when we go from model A (no boundary problem) to model B (one boundary problem). We find that with the second bandwidth (where we have the boundary problem), the Gaussian kernel estimator substantially decreases in order to correct the bias while the first remains invariant. For example, for $n = 250$, the bandwidths are $(0.121, 0.392)$ for model B, in comparison with bandwidths used in model A $(0.372, 0.369)$. For the two gamma kernel estimators, we remark the opposite effect. These use a larger bandwidth when there is a concentration of data in the boundary region. For example, for $n = 250$, the mean bandwidth of the gamma kernel estimator is $(0.240, 0.221)$ in model A and $(0.828, 0.181)$ in model B. By doing this, the two estimators reduce the variance in the boundary region, which is related to Theorem 1 of the previous section. As a final illustration of the LSCV procedure, Figure 5 shows the mean over all the replications of the theoretical MISE and its LSCV estimator for two gamma kernel estimators. We observe that in all cases there is a global minimum for the bandwidth parameters and that the theoretical MISE and LSCV MISE surfaces are quite similar.

6 Applications

We have two illustrations. In the first illustration we reproduce the density estimates of the second example in Müller and Stadtmüller (1999). The data come from the 1970 US draft lottery data and are available on the Statlib website. We have 365 pairs of observations. The first element is the day of the year $(1, 2, \ldots, 365)$ and the second element is a priority score assigned to that day. If the priority scores are randomly assigned to the days, the density should be flat over the support. Figure 6 demonstrates clearly that this is not the case. Indeed, we observe that lower scores are assigned to those born early in the year. A formal test could be conducted to check that this surface is flat.

The second example illustrates the two boundary problem for nonnegative data. We collect data for 620 companies from Compustat for the year 1986. The first variable (Compustat item 24) is the price of the stock of the company when the books are closed at the end of the accounting year. The second variable (Compustat item 25) is the number of shares that can be bought on the stock market. Figure 7 displays the scatter plot and the nonparametric density estimates. There is clearly a high concentration close to the origin, which would result in a serious boundary problem if the standard Gaussian kernel were used. We notice that the price of the stock and the number of stocks in the market seem to be positively associated.

Table 2: Mean and standard deviation ($\times 10^{-1}$) of the theoretical and LSCV bandwidth.

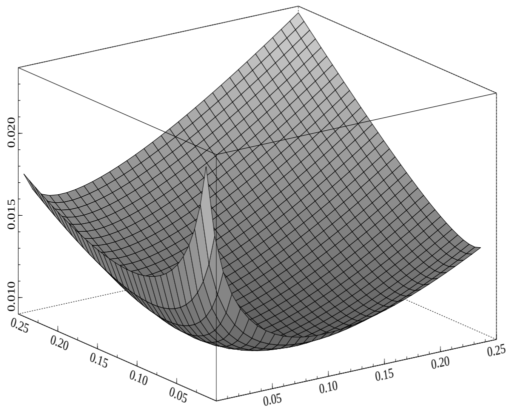
			Gaussian		local linear	
			Optimal	LSCV	Optimal	LSCV
A	n=250	Mean	(0.372,0.369)	(0.360,0.365)	(0.814,0.783)	(0.675,0.759)
		Std dev	(0.392,0.398)	(1.086,1.085)	(0.097,0.089)	(3.326,3.124)
	n=500	Mean	(0.332,0.324)	(0.301,0.328)	(0.719,0.698)	(0.585,0.694)
		Std dev	(0.337,0.366)	(1.004,0.947)	(0.765,0.780)	(2.884,2.845)
B	n=250	Mean	(0.121,0.392)	(0.065,0.417)	(0.758,0.851)	(0.781,0.671)
		Std dev	(0.447,0.822)	(0.573,1.192)	(2.407,1.718)	(6.681,4.973)
	n=500	Mean	(0.093,0.364)	(0.052,0.366)	(0.630,0.769)	(0.689,0.583)
		Std dev	(0.438,0.734)	(0.421,0.959)	(2.179,1.269)	(5.940,4.375)
C	n=250	Mean	(0.153,0.137)	(0.123,0.123)	(0.623,0.606)	(0.711,0.573)
		Std dev	(0.448,0.402)	(1.013,0.993)	(1.670,1.702)	(6.060,5.688)
	n=500	Mean	(0.129,0.122)	(0.085,0.118)	(0.525,0.499)	(0.518,0.609)
		Std dev	(0.310,0.339)	(0.718,0.825)	(1.560,1.395)	(5.456,5.697)
D	n=250	Mean	(0.146,0.139)	(0.109,0.152)	(0.756,0.750)	(0.824,0.725)
		Std dev	(0.521,0.514)	(0.956,1.091)	(3.074,2.067)	(5.541,5.665)
	n=500	Mean	(0.128,0.110)	(0.107,0.079)	(0.716,0.641)	(0.693,0.618)
		Std dev	(0.325,0.352)	(0.797,0.583)	(2.883,2.399)	(5.865,5.645)

A: bivariate normal, B: truncated bivariate normal with one boundary problem, C: truncated bivariate normal with two boundary problems, D: two independent Weibull with two boundary problems. Std dev means standard deviation.

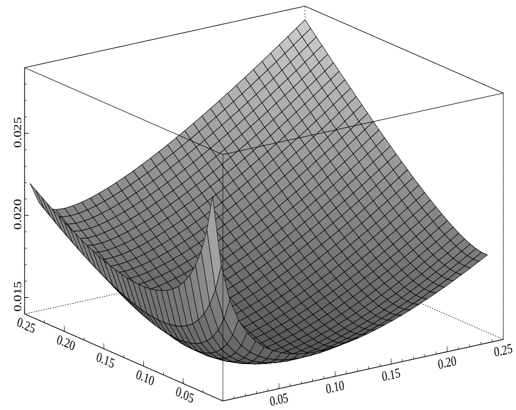
Table 3: Mean and standard deviation ($\times 10^{-2}$) of the theoretical and LSCV bandwidth.

			Gamma		Modified Gamma	
			Optimal	LSCV	Optimal	LSCV
	n=250	Mean	(0.024,0.022)	(0.022,0.025)	(0.024,0.022)	(0.023,0.025)
		Std dev	(0.060,0.059)	(0.124,0.124)	(0.061,0.049)	(0.129,0.127)
A	n=500	Mean	(0.018,0.017)	(0.016,0.019)	(0.018,0.017)	(0.016,0.019)
		Std dev	(0.040,0.037)	(0.084,0.087)	(0.043,0.039)	(0.084,0.089)
	n=250	Mean	(0.082,0.018)	(0.068,0.019)	(0.117,0.026)	(0.081,0.021)
		Std dev	(0.262,0.052)	(0.542,0.101)	(0.394,0.089)	(0.652,0.098)
B	n=500	Mean	(0.062,0.015)	(0.058,0.014)	(0.087,0.022)	(0.071,0.016)
		Std dev	(0.236,0.045)	(0.443,0.071)	(0.379,0.059)	(0.485,0.071)
	n=250	Mean	(0.064,0.061)	(0.073,0.059)	(0.112,0.109)	(0.105,0.110)
		Std dev	(0.246,0.221)	(0.446,0.489)	(0.416,0.355)	(0.507,0.486)
C	n=500	Mean	(0.050,0.045)	(0.058,0.048)	(0.089,0.088)	(0.088,0.095)
		Std dev	(0.175,0.177)	(0.356,0.392)	(0.282,0.259)	(0.387,0.383)
	n=250	Mean	(0.064,0.091)	(0.080,0.109)	(0.115,0.112)	(0.115,0.095)
		Std dev	(0.507,0.476)	(0.540,0.759)	(0.516,0.398)	(0.421,0.752)
D	n=500	Mean	(0.060,0.052)	(0.055,0.075)	(0.087,0.083)	(0.076,0.087)
		Std dev	(0.313,0.250)	(0.469,0.448)	(0.347,0.266)	(0.492,0.556)

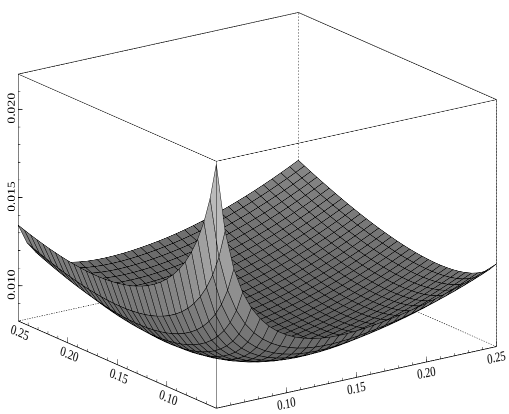
A: bivariate normal, B: truncated bivariate normal with one boundary problem, C: truncated bivariate normal with two boundary problems, D: two independent Weibull with two boundary problems. Std dev means standard deviation.



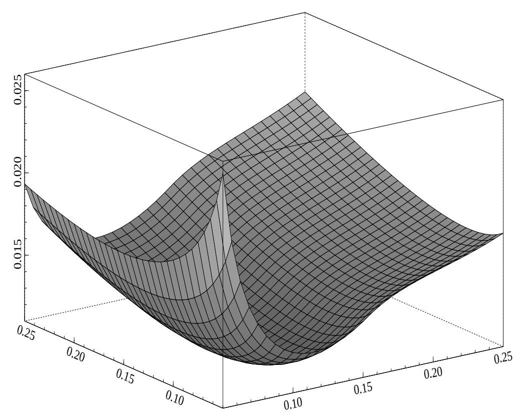
(a) Theoretical MISE for gamma estimator



(b) LSCV MISE for gamma estimator

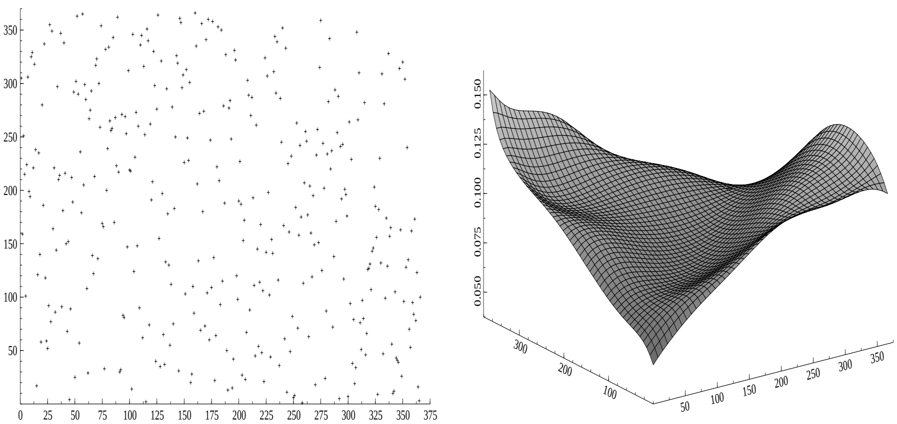


(c) Theoretical MISE for modified gamma estimator



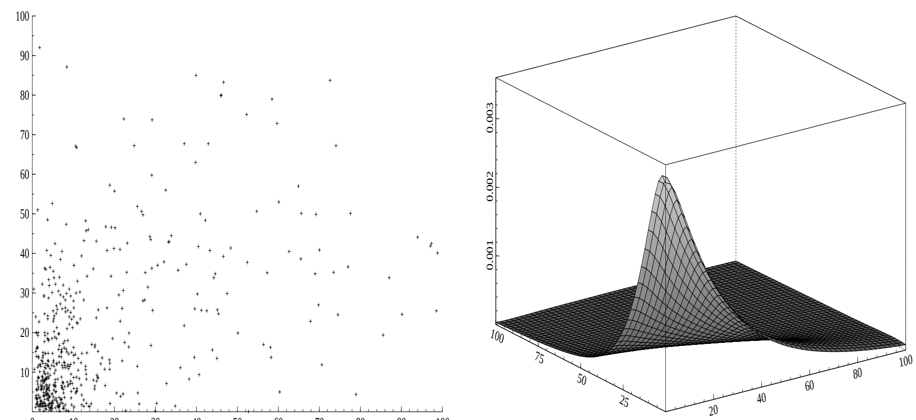
(d) LSCV MISE for modified gamma estimator

Figure 5: Theoretical mean integrated squared error and its LSCV estimator for gamma and modified gamma estimator. The data are from model B ($n = 500$), which is a truncated gamma density with one boundary bias problem.



(a) Scatter plot of the lottery data. (b) Nonparametric estimator of the lottery data.

Figure 6: Scatter plot and nonparametric estimator with gamma kernel of the lottery data.



(a) Scatter plot of the corporate finance data. (b) Nonparametric estimator of the corporate finance data.

Figure 7: Scatter plot and nonparametric estimator with gamma kernel of the corporate finance data.

7 Conclusion

This paper proposes a nonparametric estimator for density functions of multivariate bounded data. The estimator is based on a gamma kernel or a local linear kernel when the support of the variable is nonnegative, and we use the beta kernel when the support is a compact set. By using boundary kernels, no weight is assigned outside the support of the underlying density so that the estimators are robust to the boundary problem. We provide the asymptotic properties of the estimator and show that the optimal rate of convergence of the mean integrated squared error is obtained. We examine the finite sample performance in several simulations. In fact, we find that the estimators we propose perform almost as well as the standard Gaussian estimator when there are no boundary problems. With respect to the choice of the bandwidth parameters, we suggest to apply the least squares cross-validation method for which we prove consistency. In the simulations we find indeed that the distributions of the bandwidth parameters are close to the theoretical distributions. Further research on this work can be done on different angles. It would be interesting to perform another detailed simulation analysis to investigate alternative bandwidth selection methods (i.e. biased cross validation or bootstrap) and to compare our estimator with the Müller and Stadtmüller (1999) estimator. The results can also be extended to the multivariate time series case, the censored data case or further developed for multivariate nonparametric regression and for multivariate data defined on more involved supports.

Appendix

We give the proofs for the nonparametric estimator using gamma kernels.

Proof of Theorem 1

We start with the bias of the nonparametric gamma estimator

$$\begin{aligned} E_X(\hat{f}(x)) &= E\left(K(b_1, X_1)(x_1)...K(b_d, X_d)(x_d)\right) \\ &= \int K(b_1, t_1)(x_1)...K(b_d, t_d)(x_d) f(t_1, ..., t_d)\, dt_1...dt_d \\ &= E_Y(f(Y_1, ..., Y_d)) \end{aligned}$$

where the random variables Y_j are independent and gamma distributed $G(x_j/b_j + 1, b_j)$ with mean $\mu_1 = x_j + b_j$ and variance $\sigma_j^2 = x_j b_j + b_j^2$.

Using a second order Taylor expansion

$$\begin{aligned} f(Y_1, ..., Y_d) &= f(\mu_1, ..., \mu_d) + \sum_{j=1}^{d}(y_j - \mu_j)\frac{\partial f}{\partial x_j} + \frac{1}{2}\sum_{j=1}^{d}(y_j - \mu_j)^2\frac{\partial^2 f}{\partial x_j^2} \\ &+ \sum_{l \neq j}^{d}(y_l - \mu_l)(y_j - \mu_j)\frac{\partial^2 f}{\partial x_l \partial x_j} + O\left(\sum_{j=1}^{d} b_j^2\right). \end{aligned}$$

Then

$$
\begin{aligned}
E(\hat{f}(x)) &= f(\mu_1, ..., \mu_d) + \sum_{j=1}^{d} (x_j b_j + b_j^2) \frac{\partial^2 f}{\partial x_j^2} + O\left(\sum_{j=1}^{d} b_j^2\right) \\
&= f(x_1, ..., x_d) + \sum_{j=1}^{d} b_j \frac{\partial f}{\partial x_j} + \sum_{j=1}^{d} (x_j b_j + b_j^2) \frac{\partial^2 f}{\partial x_j^2} + O\left(\sum_{j=1}^{d} b_j^2\right) \\
&= f(x_1, ..., x_d) + \sum_{j=1}^{d} b_j \left(\frac{\partial f}{\partial x_j} + \frac{1}{2} x_j \frac{\partial^2 f}{\partial x_j^2}\right) + O\left(\sum_{j=1}^{d} b_j^2\right) \\
&= f(x) + \sum_{j=1}^{d} b_j B_j(x) + O\left(\sum_{j=1}^{d} b_j^2\right).
\end{aligned}
$$

Hence

$$
\int (E(\hat{f}(x)) - f(x))^2 dx = \int (\sum_{j=1}^{d} b_j dx B_j(x))^2 + O\left(\sum_{j=1}^{d} b_j^2\right). \tag{6}
$$

Now, the variance of the nonparametric gamma estimator is

$$
\begin{aligned}
n\, var(\hat{f}(x)) &= E\left(K(b_1, X_1)(x_1)...K(b_d, X_d)(x_d)\right)^2 + O(1) \\
&= \prod_{j=1}^{d} B_j(x_j, b_j) E(f(\varphi_x)) + o(1)
\end{aligned}
$$

where $\varphi_x = (Z_1, ..., Z_d)$ and the random variables Z_j are independent and gamma distributed $G(2x_j/b_j + 1, b_j/2)$ and

$$
B_j(x_j, b_j) = \frac{b_j^{-1}\Gamma(2x_j/b_j + 1)}{2^{2x_j/b_j+1}\Gamma^2(x_j/b_j + 1)}
$$

where Γ is the gamma function. Define

$$
R(x) = \sqrt{2\pi} x^{x+1/2} e^{-x} / \Gamma(x+1), \quad x > 0.
$$

Let us recall the following properties of R:

$$
R(x) < 1, \text{ for all } x > 0
$$

and

$$
R(x) \longrightarrow 1 \text{ as } x \longrightarrow \infty.
$$

Combining these properties with Stirling's formula we can show that

$$
B_j(x_j, b_j) = \begin{cases} \frac{1}{2\sqrt{(\pi)}} b_j^{-1/2} x_j^{-1/2} & \text{if } x_j/b_j \to \infty \\ \\ \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} b_j^{-1} & x_j/b_j \to \kappa \end{cases}
$$

and from a Taylor expansion

$$E(f(\varphi_x)) = f(x) + O(b).$$

Therefore

$$n\, var(\hat{f}(x)) = f(x_1, ..., x_d) \prod_{j\in I} (\frac{1}{2\pi^{1/2}} b_j^{-1/2} x_j^{-1/2}) \prod_{j\in I^C} \left(\frac{\Gamma(2\kappa_j + 1)}{2^{2\kappa_j+1}\Gamma^2(\kappa_j + 1)} \right) b^{-1}$$

where $I = \{j, x_j/b_j \to \infty\}$ and I^C its complement.

The second product disappears in the integrated variance. Let $\delta_j = b_j^{1-\epsilon}$, $0 < \epsilon < 1/2$ and $\delta = (\delta_1, ..., \delta_d)$, then

$$\begin{aligned}
\int var(\hat{f}(x)) &= n^{-1} \left\{ \int_0^\delta var(\hat{f}(x))dx + \int_\delta^\infty var(\hat{f}(x))dx \right\} \\
&= O\left(\prod_{j\in I} b_j^{1/2-\epsilon} \prod_{j\in I^C} b_j^{-\epsilon} \right) + \int_0^\infty \prod_{j=1}^d \left(\frac{1}{2\pi^{1/2}} b_j^{-1/2} x_j^{-1/2} \right) f(x)dx \\
&= o\left(\prod_{j=1}^d b_j^{-1/2} \right) + \frac{1}{2^d \pi^{d/2}} \int_0^\infty \prod_{j=1}^d \left(b_j^{-1/2} x_j^{-1/2} \right) f(x)dx \\
&= \prod_{j=1}^d b_j^{-1/2} \int V(x)dx + o\left(\prod_{j=1}^d b_j^{-1/2} \right). \qquad (7)
\end{aligned}$$

By combining (7) and (6), we obtain the mean integrated squared error of the nonparametric estimator with the gamma kernel.

Proof of theorem 2

We denote by $\mu_x = (\mu_{x_1}, ..., \mu_{x_d})$ where μ_{x_j} is the mean of a gamma random variable with parameter $(1/b_j, x_j/b_j + 1)$.

Since f is continuous, $\mu_x = x + b$ and $b \to 0$, with $b = (b_1, ..., b_d)$ we have, for any ϵ there exists δ such that

$$|f(t) - f(x)| < \epsilon, \quad \text{for} \quad ||t - x|| < \delta. \qquad (8)$$

We start with

$$\begin{aligned}
\left| E\left(\hat{f}(x) \right) - f(x) \right| &\leq \int_{||t-x||\leq\delta} (K_G(b_1, t_1)(x_1)...K_G(b_d, t_d)(x_d))|f(t) - f(x)|dt \\
&+ \int_{||t-x||\geq\delta} (K_G(b_1, t_1)(x_1)...K_G(b_d, t_d)(x_d))|f(t) - f(x)|dt \\
&= I + II.
\end{aligned}$$

From (8),

$$I < \epsilon \int_{||t-x|| \leq \delta} (K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))|f(t)-f(x)|dt < \epsilon. \quad (9)$$

On the other hand, using Chebyshev's inequality and that $\mathrm{Var}(K_G(b_j,t_j)(x_j)) = x_j b_j + b_j^2$, we obtain

$$\begin{aligned} II &\leq 2\sup_x |f(x)| \int_{||t-x|| \geq \delta} (K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))dt \\ &= 2\sup_x |f(x)| \prod_j (x_j b_j + b_j^2) \\ &= o(1). \end{aligned} \quad (10)$$

Hence, from (9) and (10),

$$\left| E\left(\hat{f}(x)\right) - f(x) \right| \to 0.$$

Now, it remains to prove that the variation term $\left|E\left(f_{np}(x) - \hat{f}(x)\right)\right|$ converges almost surely to zero.

Using integration by parts,

$$\begin{aligned} \left|\hat{f}(x) - E\left(\hat{f}(x)\right)\right| &= \left|\int (K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))d(F_n(t)-F(t))\right| \\ &\leq \left|\int (F_n(t)-F(t))d(K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))\right| \\ &\leq \sup_{x \in R^{+d}} |F_n(x)-F(x)| \left|\int d(K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))\right|. \end{aligned}$$

We can see that

$$\left|\int d(K_G(b_1,t_1)(x_1)...K_G(b_d,t_d)(x_d))\right| \leq 2^d (\prod_j b_j^{-1}).$$

Therefore, and from Kiefer (1961),

$$\begin{aligned} P\left(\left|\hat{f}(x) - E\left(\hat{f}(x)\right)\right| > \epsilon\right) &\leq P\left(\sup_{x \in R^{+d}} |F_n(x)-F(x)| > \epsilon 2^{-d} (\prod_j b^1)\right) \\ &\leq C(d)\exp\left(-c2^{-2d}\epsilon^2 n(\prod_j b_j^2)\right), \end{aligned}$$

with some constants $c < 2$ and $C(d)$ depending on the dimension d. Let us take $\epsilon_n = \frac{\alpha}{\sqrt{c}}\sqrt{\frac{\log(n)}{n}}\prod_j b_j^{-2}$, with $\alpha \geq 2^d$. This implies that

$$\sum_n \exp\left(-c2^{-2d} n(\prod_j b_j^2)\right) < \infty.$$

Therefore, $\left|E\left(f_{np}(x) - \hat{f}(x)\right)\right|$ converges almost surely. This concludes the proof of theorem 2. ■

Proof of Theorem 3

We start with the classical decomposition and using the expression of the asymptotic bias:

$$\begin{aligned}\left(\hat{f}_{sp}(x) - f(x)\right) &= \left(\hat{f}_{sp}(x) - E(\hat{f}_{sp}(x))\right) + \left(E(\hat{f}_{sp}(x)) - f(x)\right) \\ &= \left(\hat{f}_{sp}(x) - E(\hat{f}_{sp}(x))\right) + \sum_{j=1}^{d} b_j B_j + O(\sum_{j=1}^{d} b_j^2)\end{aligned}$$

Hence, and using that $b_j = O(n^{-\frac{2}{d+2}})$

$$\sigma^{*-1} n^{\frac{1}{2}} \prod_{j=1}^{d} b_j^{\frac{1}{4}} \left(\hat{f}_{sp}(x) - f(x) - \mu^*\right) = \sum_{i=1}^{n} Z_i + O(n^{-\frac{2}{d+2}}),$$

where

$$Z_i = \sigma^{*-1} n^{-\frac{1}{2}} \prod_{j=1}^{d} b_j^{\frac{1}{4}} \left(K(b_1, X_{i1})(x_1)...K(b_d, X_{id})(x_d) - E(K(b_1, X_{i1})(x_1)...K(b_d, X_{id})(x_d))\right).$$

Now, we apply Liapunov central limit theorem to prove the asymptotic normality of $S_n^* = \sum_{i=1}^{n} Z_i$. From (7),

$$\text{Var}(S_n^*) = 1 + o(1)$$

Using (7) and that $b_j = O(n^{-\frac{2}{d+2}})$,

$$E(|Z_i|^3) = o(n^{-1}).$$

Therefore, $S_n^* = \sum_{i=1}^{n} Z_i \xrightarrow{\mathcal{D}} N(0, 1)$. This concludes the proof of theorem 3.

■

Proof of Theorem 4

To show the results of the theorem, it suffices to establish that:

$$I = \lim \max_{b,b' \in H_n} \frac{|MISE_{b'} - MISE_b - (LSCV_{b'} - LSCV_b)|}{MISE_{b'} + MISE_b} \to 0, \quad \text{a.s.}$$

Because $MISE_{\hat{b}} - MISE_{b_0} > 0$ and $LSCV_{\hat{b}} - LSCV_{b_0} < 0$, we have $\left|1 - \frac{MISE_{b_0}}{MISE_{\hat{b}}}\right| \leq 2I$.

To show that I converges almost surely, we state that:

$$\frac{ISE_b}{MISE_b} \to 1, \quad \text{a.s. for all } b \in H_n, \tag{11}$$

and that

$$I = \lim \max_{b,b' \in H_n} \frac{|ISE_{b'} - ISE_b - (LSCV_{b'} - LSCV_b)|}{ISE_{b'} + ISE_b} \to 0, \quad \text{a.s.} \tag{12}$$

For the remainder and without loss of generality we consider that $MISE_b = Cn^{-\frac{4}{d+4}}$ and that $b = C'n^{-\frac{2}{d+4}}$.

We start to prove (11),

$$\begin{aligned}
MISE_b &= E\left(\int (\hat{f}_b - f_b)^2(x)dx\right) + \int (f_b - f)^2(x)dx \\
&= n^{-1}\int k(b,t)^2(x)\, dF(t)dx - n^{-1}\int\int K_b(t,s)dF(t)dF(s) + \int B^2(x)dx,
\end{aligned}$$

where $K_b(t,s) = \int K(b,t)(x)K(b,s)\,dx$, $f_b = E(\hat{f}_b)$ and $B(x)$ the bias of $\hat{f}_b$ at x.
On the other hand,

$$\begin{aligned}
ISE_b &= \int (\hat{f}_b - f_b)^2(x)dx + 2\int (\hat{f}_b - f_b)(f_b - f)(x)dx + \int (f_b - f)^2(x)dx \\
&= \int\int K(t,s)d(F_n - F)(t)d(F_n - F)(s) + 2\int\int K(b,t)(x)B(x)d(F_n - F)(t)dx + \int B^2(x)dx \\
&= \int\int_{t\neq s} K(t,s)d(F_n - F)(t)d(F_n - F)(s) + n^{-1}\int K(b,t)^2(x)d(F_n - F)(t)dx \\
&+ n^{-1}\int K(b,t)^2(x)dF(t)dx + 2\int\int K(b,t)(x)B(x)d(F_n - F)(t)dx + \int B^2(x)dx,
\end{aligned}$$

where F_n denotes the empirical distribution function.
Hence, we have

$$\begin{aligned}
ISE - MISE &= \int\int_{t\neq s} K(t,s)d(F_n - F)(t)d(F_n - F)(s) + n^{-1}\int K(b,t)^2(x)d(F_n - F)(t)dx \\
&\quad +2\int\int K(b,t)(x)B(x)d(F_n - F)(t)dx + n^{-1}\int\int K_b(t,s)dF(t)dF(s) \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

First, for the non-random term I_4, let us show that there is a constant $\gamma > 0$, so that for all $k = 2, 3, ..$, there exist constants A_k such that

$$\left(\frac{I_4}{MISE}\right)^{2k} \leq A_k n^{-\gamma k}.$$

Using that $MISE_b = Cn^{-\frac{4}{d+4}}$ and that $\int K(b,x)(t)dF(t) = f(x) + O(b)$, we obtain

$$\begin{aligned}
\left(\frac{I_4}{MISE}\right)^{2k} &= MISE^{-2k}n^{2k}\left(\int\int K_b(t,s)dF(t)dF(s)\right)^{2k} \\
&= MISE^{-2k}n^{2k}\left(\int f^2(x) + O(b)\right)^{2k} \\
&= O(n^{-\frac{8k}{d+4}}n^{2k}) = O(n^{-\frac{2d}{d+4}k}).
\end{aligned}$$

■

Second, we show that the two terms $\frac{I_2}{MISE}$ and $\frac{I_3}{MISE_h}$ converges to zero almost surely and uniformly on H_n.

Lemma 1. *Under condition (4),*

$$\sup_{b\in H_n}\left|\frac{I_l}{MISE_b}\right| \to 0, \quad a.s. \quad for\ l=2,3.$$

■

I_2 and I_3 can re-expressed as $\frac{1}{n}\sum W_i$, where for I_2

$$W_i = \int K^2(b,x)(X_i)\,dx - \int\int K^2(b,x)(t)dF(t)\,dx$$

and for I_3

$$W_i = n^{-1}\int B(x)K(b,x)(X_i)\,dx - n^{-1}\int\int B(x)K(b,x)(t)dF(t)\,dx.$$

The mean of W_i is zero. Now, we apply Bernstein's inequality and (4), we obtain

$$\begin{aligned} P\left(\sup_{b\in H_n}\left|\frac{I_l}{MISE_b}\right| > \epsilon\right) &\leq A\,n^{\alpha}P\left(\sum_{i=1}^{n}W_i > \epsilon.n.MISE_b\right)\\ &\leq A\,n^{\alpha}\exp\left(-\frac{\epsilon^2n^2MISE_b^2}{\sum_{i=1}^{n}E(W_i^2)+M\epsilon nMISE_b/3}\right). \end{aligned}$$

where $|W_i| < M$. Then, it suffices to calculate $E(W_i^2)$.

For I_2, from Chen (2000), we can see that $\int\int K^2(b,x)(t)dF(t)\,dx = O(b^{-1})$. On the other hand,

$$\begin{aligned} E\left(\int K^2(b,x)(X_i)\,dx\right)^2 &= \int\int\int K^2(b,x)(t)K^2(b,y)(t)dF(t)\,dx\,dy\\ &= \int\int B_b(x,y)E(f(\xi_b))\,dx\,dy \end{aligned}$$

where ξ_b is a gamma random variable with parameter $2(x+y)/b+1$ and $b/4$ and

$$B_b(x,y) = b^{-3}\frac{\Gamma(2(x+y)/b+1)}{\Gamma^2(x/b+1)\Gamma^2(y/b+1)}\frac{1}{4^{2(x+y)/b+1}}.$$

For small bandwidth, $\Gamma(x+1) = \sqrt{2\pi}e^{-x}x^{x+1/2}$ and that $E(f(\xi_b)) = O(1)$, we can show that

$$E\left(\int K^2(b,x)(X_i)\,dx\right)^2 = O((b^{-3/2}),$$

which is negligible in comparison with the order of the second term ($O(n^{-2})$). Then $E(W_i)^2 = O(n^{-2}b^{-2})$. Therefore,

$$P\left(\sup_{b\in H_n}\left|\frac{I_l}{MISE_b}\right| > \epsilon\right) \leq A\,n^{\alpha}\exp(-n^{\frac{3d}{d+4}}),$$

which concludes the proof for I_2.

For I_3, we use that the bias of gamma kernel estimator is of order $O(b)$ and that $\int K(b,x)(t)dx < C$ for some constant C, so that we can show,

$$\begin{aligned} E(\int B(x)K(b,x)(X_i)\,dx)^2 &= \int\int\int K(b,x)(t)B(x)K(b,y)(t)B(y)f(t)dtdxdy \\ &= O(b^2). \end{aligned}$$

Therefore, as for I_2 term

$$P\left(\sup_{b\in H_n}\left|\frac{I_l}{MISE_b}\right| > \epsilon\right) \leq A\,n^{\alpha}\exp\left(-n^{\frac{d}{d+4}}\right),$$

which concludes the proof for I_3.

■

For the term I_1, applying Chebyshev's inequality and (4), we obtain

$$P\left(\sup_{b\in H_n}\left\|\frac{I_1}{MISE_b}\right\| > \epsilon\right) \leq A\,n^{\alpha}E\left(\frac{I_1}{\epsilon.MISE_b}\right)^{2k}.$$

Remark that I_1 can be re-expressed as follows

$$I_1 = \frac{1}{n}\sum_{i\neq j}^{n} W_{i,j},$$

where

$$W_{i,j} = K_b(X_i,X_j) - \int K_b(t,X_j)dF(t) - \int K_b(X_i,s)dF(s) + \int\int K_b(t,s)dF(t)dF(s).$$

Note that for $i \neq j$, the mean of $W_{i,j}$ is zero. Now, from the linearity of cumulants, it suffices to show that there exists a constant α, so that for $k = 2,3,\cdots$, there are a constants α_k such that

$$n^{-2k}MISE^{-k}\left|\sum_{i_1,j_1\cdots i_k,j_k} cum_k(W_{i_1,j_1},\cdots,W_{i_k,j_k})\right| \leq \alpha_k n^{-\alpha k}. \tag{13}$$

Let m denote the number of $i_1, j_1, \cdots i_{2k}, j_{2k}$ that are unique and that for $m = 2,\cdots,k$, the number of element of cumulants with m distinct elements is bounded by Cn^m

$$\sum_{i_1,j_1\cdots i_k,j_k} cum_k(W_{i_1,j_1},\cdots,W_{i_k,j_k}) \leq C\sum_{m=2}^{k} n^m cum_k(W_{i_1,j_1},\cdots,W_{i_k,j_k})\ m \text{ distinct indices.} \tag{14}$$

Now, for m distinct indices, i.e. $t_{i1}, t_{j1}\cdots t_{i_k}, t_{j_k} \in \{t_1\cdots t_m\}$, and by definition of $K_b(t_i,t_j)$,

$$
\begin{aligned}
E\left(W_{i_1,j_1}\cdots W_{i_k,j_k}\right) &= \int K_b(t_{i1},t_{j1})\cdots K_b(t_{i_k},t_{j_k})dF(t_1)\cdots dF(t_m)\\
&= \int K(b,t_i)(x_1)K(b,t_j)(x_1)\cdots K(b,t_{i_k})(x_k)K(b,t_{j_k})(x_k)dF(t_1)\\
&\quad \cdots dF(t_m)dx_1\cdots dx_k
\end{aligned}
$$

For simplicity, let us regroup the terms concerning t_1 and suppose that there are δ_1 terms in $\{i_1,j_1,\cdots i_k,j_k\}$. We get

$$
\int \frac{t_1^{\frac{1}{b}\sum x_i}e^{-\delta_1\frac{t_1}{b}}}{\prod\Gamma(\frac{x_i}{b}+1)b^{\frac{1}{b}\sum x_i+\delta_1}}dF(t_1) \quad = \quad B_b(x)E(\eta_b),
$$

where η_b is a gamma random variable with parameter $(\frac{\sum x_i}{b}+1, b/\delta_1)$, and

$$
B_b(x) = \frac{\Gamma(\frac{\sum x_i}{b}+1)}{\prod\Gamma(\frac{1}{b}x_i+1)\delta_1^{\frac{\sum x_i}{b}+1}}b^{1-\delta_1}.
$$

For small bandwidth, using that $\Gamma(x+1)=\sqrt{2\pi}e^{-x}x^{x+1/2}$ and that the mean of η_b is $\frac{1}{b}\sum x_i+\frac{b}{\delta_1}$

$$
\int \frac{t_1^{\frac{1}{b}\sum x_i}e^{-\delta_1\frac{t_1}{b}}}{\prod\Gamma(\frac{x_i}{b}+1)b^{\frac{1}{b}\sum x_i+\delta_1}}dF(t_1) \quad \leq \quad (2\pi)^{1-\delta_1}\frac{(\sum x_i)^{\frac{1}{b}\sum x_i+\frac{1}{2}}}{\prod x_i^{\frac{x_i}{b}+\frac{1}{2}}\delta_1^{\frac{1}{b}\sum x_i+1}}f\left(\frac{\sum x_i}{\delta_1}\right)\ b^{\frac{1}{2}(1-\delta_1)}.
$$

Then, and from condition (5)

$$
\int\int \frac{t_1^{\frac{1}{b}\sum x_i}e^{-\delta_1\frac{t_1}{b}}}{\prod\Gamma(\frac{x_i}{b}+1)b^{\frac{1}{b}\sum x_i+\delta_1}}dF(t_1)dx_1\cdots dx_{\delta_1} \quad \leq \quad const\ b^{\frac{1}{2}(1-\delta_1)}
$$

Using that $\sum\delta_i = k$,

$$
E\left(W_{i_1,j_1}\cdots W_{i_k,j_k}\right) \leq const\ b^{\frac{m}{2}-\frac{k}{2}}. \tag{15}
$$

Therefore, from (13), (13), (14) and (15), we conclude the almost sure convergence of $I_1/MISE_b$.

Now we prove (12). Let us see that

$$
\begin{aligned}
ISE_b - LSCV_b - \int f^2 - 2G_n &= 2(G_{bn}-G_n) + 2\int\int_{x\neq y}K(b,x)(y)d(F_n-F)(x)d(F_n-F)(y)\\
&= I + II,
\end{aligned}
$$

where $G_{bn}=\frac{1}{n}\sum f_b(X_i)-(f_b(X))$ and $G_n=\frac{1}{n}\sum f(X_i)-(f(X))$.

The first term can be expressed as

$$
I = \frac{1}{n}\sum_{i=1}^{n}W_i,
$$

where

$$W_i = f_b(X_i) - f(X_i) - E(f_b(X_i) - f(X_i)).$$

The expression $f_b(t) - f(t)$ is the bias of the gamma kernel estimator and is of order $O(b)$, then $E(W_i^2) = Const\, b^2$. Therefore, as in Lemma 1, we can see that

$$\begin{aligned}
P\left(\sup_{b\in H_n}\left|\frac{I}{MISE_b}\right| > \epsilon\right) &\leq A\, n^{\alpha} P\left(\sum_{i=1}^{n} W_i > \epsilon.n.MISE_b\right) \\
&\leq A\, n^{\alpha}\exp\left(-\frac{\epsilon^2 n^2 MISE_b^2}{\sum_{i=1}^{n} E(W_i^2) + M\epsilon n MISE_b/3}\right) \\
&\leq A\, n^{\alpha}\exp\left(-\epsilon^2 n^{d/(d+4)}\right),
\end{aligned}$$

which states the almost sure convergence of $I/MISE_b$.

The second term can be expressed as

$$II = \frac{1}{n}\sum_{i\neq j}^{n} W_{i,j},$$

where

$$W_{i,j} = K(b, X_i)(X_j) - \int K(b, X_i)(y)dF(y) - \int K(b, x)(X_j)dF(x) + \int\int K(b, x)(y)dF(x)dF(y).$$

As for I_1 term, it suffices to calculate $E(W_{i_1,j_1}, \cdots W_{i_k,j_k})$. Using that $K(b, x)(y) \leq Cb^{-1/2}$ and that each $i_1, j_1, \cdots i_k, j_k$ may appear twice, we can see that $E(W_{i_1,j_1}, \cdots W_{i_k,j_k}) = O(b^{-\frac{\sigma k}{2}})$ for $0 < \sigma < 1$. Then

$$n^{-2k} MISE^{-k}\left|\sum_{i_1,j_1\cdots i_k,j_k} cum_k(W_{i_1,j_1}, \cdots, W_{i_k,j_k})\right| \leq \alpha_k n^{-\epsilon k},$$

which concludes the almost sure convergence of $II/MISE_b$, and therefore also concludes the proof of Theorem 4.

■

References

Bouezmarni, T., and J. Rombouts (2006): "Nonparametric Density Estimation for Positive Time Series," *CORE discusion paper 2006/85.*

Bouezmarni, T., and O. Scaillet (2003): "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data," *Econometric Theory*, 21, 390–412.

Bowman, A. (1984): "An Alternative Method of Cross-validation for the Smoothing of Density Estimates.," *Biometrika*, 71, 353–360.

Chen, S. (1999): "A Beta Kernel Estimation for Density Functions," *Computational Statistics and Data analysis*, 31, 131–145.

——— (2000): "Probability Density Functions Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

Cheng, M., J. Fan, and J. Marron (1997): "On Automatic Boundary Corrections," *Annals of Statistics*, 25, 1691–1708.

Chiu, S. (1991): "Bandwidth Selection for Kernel Density Estimation," *Annals of Statistics*, 19, 1883–1905.

Cowling, A., and P. Hall (1996): "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B*, 58, 551–563.

Devroye, L., and L. Györfi (1985): *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Epanechnikov, V. (1969): "Nonparametric Estimation of a Multivariate Probability Density," *Theory of Probability and its Applications*, 14, 153–158.

Fan, J., and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smothers," *Annals of Statistics*, 20, 2008–2036.

Gasser, T., H. Müller, and V. Mammitsch (1985): "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society, Series B*, 47, 238–252.

Gourieroux, C., and A. Monfort (2006): "(Non)Consistency of Beta Kernel Estimator of Recovery Rates," *CREST-DP*, 31.

Granovsky, B., and H. Müller (1991): "Optimizing Kernel Methods: a Unifying Variational Principle," *International statistical review*, 59, 373–388.

Grullon, G., and R. Michaely (2002): "Dividends, Share Repurchases, and the Substitution Hypothesis," *The Journal of Finance*, 57, 1649–1684.

Hall, P. (1983): "Large Sample Optimality of Least Squares Cross-Validation in Density Estimation," *Annals of Statistics*, 11, 1156–1174.

Hall, P., and J. Marron (1987): "Extent to which Least-Squares Cross-Validation Minimises Integrated Square Error in Nonparametric Density Estimation," *Probability Theory and Related Fields*, 74, 567–581.

Härlde, W., and J. Marron (1985): "Optimal Bandwidth Selection in Nonparametric Regression Function Estimation," *Journal of Multivariate Analysis*, 20, 91–113.

Jones, M. (1993): "Simple Boundary Correction for Kernel Density Estimation," *Statistical Computing*, 3, 135–146.

Jones, M., and P. Foster (1996): "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation," *Statistica Sinica*, 6, 1005–1013.

Kiefer, J. (1961): "On Large Deviations of the Empiric d.f. of Vector Chance Varaiables and a Law of the Iterated Logarithm," *Pacific Journal of Mathematics*, 11, 649–660.

Lejeune, M., and P. Sarda (1992): "Smooth Estimators of Distribution and Density Functions," *Computational Statistics and Data Analysis*, 14, 457–471.

Marron, J., and W. Härlde (1986): "Random Approximation to Some Measures of Accuracy in Nonparametric Curve Estimation," *Journal of Multivariate Analysis*, 20, 91–113.

Marron, J., and D. Ruppert (1994): "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B*, 56, 653–671.

Müller, H. (1991): "Smooth Optimum Kernel Estimators near Endpoints," *Biometrika*, 78, 521–530.

Müller, H., and U. Stadtmüller (1999): "Multivariate Boundary Kernels and a Continuous Least Squares Principle," *Journal of the Royal Statistical Society, Series B*, 61, 439–458.

Rudemo, M. (1982): "Empirical Choice of Histograms and Kernel Density Estimates," *Scandinavian Journal of Statistics*, 9, 65–78.

Ruppert, D., W. M. (1994): "Multivariate Locally Weighted Least squares Regression," *Annals of Statistics*, 22, 1346–1370.

Schuster, E. (1985): "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics - Theory and Methods*, 14, 1123–1136.

Scott, D. (1992): *Multivariate Density Estimation: Theory, Practice and Visualization.* J. Wiley and Song, New York.

Scott, D., and G. Terrell (1987): "Biased and Unbiased Cross-Validation in Density Estimation," *Journal of the American Statistical Association*, 82, 1131–1146.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Staniswalis, J., and K. Messer (1997): "Addendum to 'Kernel Estimators for Multivariate Regression'," *Journal of Nonparametric Statistics*, 7, 67–68.

Staniswalis, J., K. Messer, and D. Finston (1993): "Kernel Estimators for multivariate Regression," *Journal of Nonparametric Statistics*, 3, 103–121.

Stone, C. (1984): "An Asymtotically Optimal Window Selection Rule for Kernel Density Estimates," *Annals of Statistics*, 12, 1285–1297.

Wand, M., and M. Jones (1995): *Kernel Smoothing.* Chapman and Hall, New York.

Zhang, S., R. Karunamuni, and M. Jones (1999): "An Improved Estimator of the Density Function at the Boundary," *Journal of the American Statistical Association*, 94, 1231–1241.