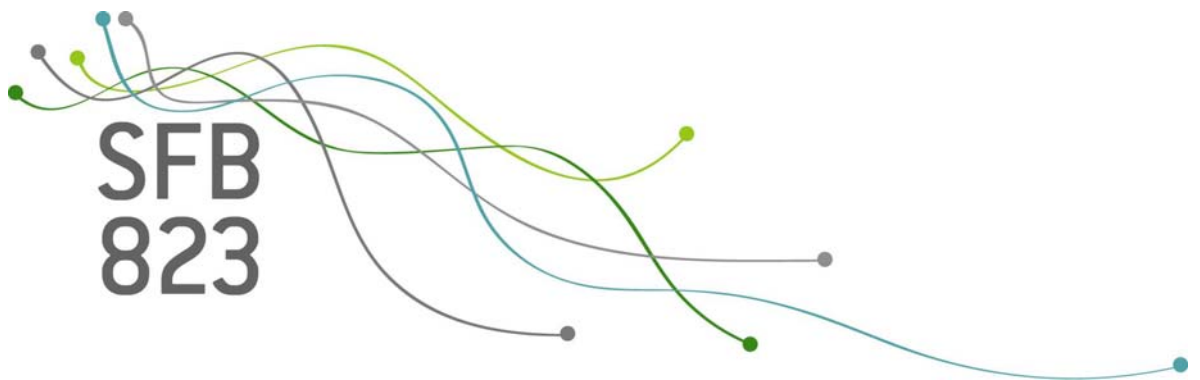


SFB
823

Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods

Benedikt Funke, Rafael Kawka

Nr. 39/2014



Discussion Paper

Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods

Benedikt Funke^{1,3,*}

Technical University of Dortmund, Department of Mathematics, Vogelpothsweg 87, 44227 Dortmund, Germany

Rafael Kawka^{2,3}

Technical University of Dortmund, Department of Statistics, Vogelpothsweg 87, 44227 Dortmund, Germany

Abstract

In this article we propose two new Multiplicative Bias Correction (MBC) techniques for nonparametric multivariate density estimation. We deal with positively supported data but our results can easily be extended to the case of mixtures of bounded and unbounded supports. Both methods improve the optimal rate of convergence of the mean squared error up to $O(n^{-8/(8+d)})$, where d is the dimension of the underlying data set. Moreover, they overcome the boundary effect near the origin and their values are always non-negative. We investigate asymptotic properties like bias and variance as well as the performance of our estimators in Monte Carlo Simulations and in a real data example.

Keywords: Asymmetric kernels, Bias correction, Multivariate density estimation

1. Introduction

In recent literature two multiplicative bias correction (“MBC”) techniques were proposed for nonparametric estimation of univariate densities with compact (Hirukawa 2010) and with bounded support (Hirukawa and Sakudo, 2014a). The applied methods were originally investigated in Jones et al. (1995, “JLN”) and in Terrell and Scott (1980, “TS”) for symmetric kernel functions. Our purpose is now the extension of these results to multivariate density estimation where the marginals possess bounded supports. As mentioned, our approach can easily be extended to the case where one has more involved supports like mixtures of bounded, compact and unbounded supports. As in Hirukawa and Sakudo (2014a) we will focus on the use of asymmetric kernels which were originally proposed in Chen (1999, 2000) for nonparametric estimation of compact as well as positively supported univariate probability densities. As kernel functions Chen suggested the use of Beta (“B”), Gamma (“G”) and the so called Modified Gamma (“MG”) densities. During the last years many other choices of kernels were proposed like Log Normal (“LN”) and Birnbaum Saunders (“BS”) (Jin and Kawczak, 2003), Inverse Gaussian (“IG”) and Reciprocal Inverse

*Corresponding author

¹benedikt.funke@math.tu-dortmund.de, Tel.: +49 231 755 5917

²kawka@statistik.tu-dortmund.de, Tel.: +49 231 755 4356

³This work has been supported by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG) which is gratefully acknowledged.

Table 1
Univariate Asymmetric Kernels

Kernel	Explicit Form
G	$K_{G,h,x}(z) = \frac{z^{x/h} \exp(-z/h)}{h^{x/h+1} \Gamma(x/h+1)}$
MG	$K_{MG,h,x}(z) = \frac{z^{\rho_h(x)-1} \exp(-z/h)}{h^{\rho_h(x)} \Gamma(\rho_h(x))}$ where $\rho_h(x) = \frac{x}{h} \mathbf{1}(x \geq 2h) + (\frac{x^2}{4h^2} + 1) \mathbf{1}(x < 2h)$
IG	$K_{IG,h,x}(z) = \frac{1}{\sqrt{2\pi h z^3}} \exp\left[-\frac{1}{2hx} (z/x - 2 + x/z)\right]$
RIG	$K_{RIG,h,x}(z) = \frac{1}{\sqrt{2\pi h z}} \exp\left[-\frac{(x-h)}{2h} (z/(x-h) - 2 + (x-h)/z)\right]$
LN	$K_{LN,h,x}(z) = \frac{1}{z\sqrt{2\pi h z}} \exp\left(-\frac{(\log(z)-\log(x))^2}{2h}\right)$
BS	$K_{BS,h,x}(z) = \frac{1}{2x\sqrt{2\pi h}} \left((x/z)^{1/2} + (x/z)^{3/2}\right) \exp\left(-\frac{1}{2h}(z/x - 2 + x/z)\right)$
NM	$K_{NM,h,x}(z) = \frac{2z^{\alpha-1} \exp(-(z/(\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2)))^2)}{(\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2))^{\alpha} \Gamma(\alpha/2)}$, where $(\alpha, \beta) = \begin{cases} (\frac{x}{h}, x), & \text{for } x \geq 2h \\ (\frac{x^2}{4h^2} + 1, \frac{x^2}{4h} + h), & \text{for } x < 2h \end{cases}$
B	$K_{B,h,x}(z) = \frac{z^{x/h} (1-z)^{(1-x)/h}}{B(x/h+1, (1-x)/h+1)}$

Gaussian (“RIG”) (Scaillet, 2004) as well as Generalized Birnbaum Saunders (“GBS”) (Marchant et. al, 2013). For the details of the derivation of Generalized Birnbaum Saunders distributions, in particular the so called generator g , see Leiva (2008). In a very recent paper Hirukawa and Sakudo (2014b) proposed the use of Generalized Gamma (“GG”) densities which are acting as a generator of asymmetric kernel functions. Due to the variety of their introduced parameters, we will only focus on a special case, namely the so called Nakagami m -kernel (“NM”).

Given a random sample of observations $\{X_i\}_{i=1}^n$ with univariate density f which is supported on $[0, \infty)$, all estimators possessing the form

$$\hat{f}_{m,h}(x) := \frac{1}{n} \sum_{i=1}^n K_{m,h,x}(X_i)$$

where $m = G, MG, IG, RIG, LN, BS, NM$ or B . K denotes the corresponding density with a certain choice of parameters.

The use of asymmetric kernel functions has many advantages like adaptive smoothing, optimal rate of convergence in the mean integrated squared error sense for kernels of order two and due to their construction they obviously avoid a boundary effect without producing negative values of the estimate. Although there are many other boundary correction methods available, this approach seems to be a good and easy implementable method for circumventing the boundary problem of nonparametric density estimation near the origin.

To our knowledge there is only one paper concerned with multivariate nonparametric density estimation where asymmetric kernels are employed, namely Bouezmarni and Rombouts (2007). Given a random sample $\{(X_{i1}, \dots, X_{id})\}_{i=1}^n$ of independent and identically distributed random vectors with positively supported marginals, they estimate the unknown joint density function f using a product kernel estimator

$$\hat{f}_{m,h}(x) = \hat{f}_{m,h}(x_1, \dots, x_d) := \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{m,h_j,x_j}(X_{ij})$$

where $h := (h_1, \dots, h_d)$ denotes the vector of bandwidths and $m = \text{G}$ or MG . They showed that the optimal rate of convergence of the mean integrated squared error for this estimator is of order $O(n^{-4/(4+d)})$. We will improve this rate with both of our proposed estimators.

The remainder of the article is organized as follows. Section 2 introduces the new estimators and determines asymptotic properties like bias and variance for several choices of the kernel function K . Section 3 includes a short discussion of the selection of the required bandwidths as well as a Monte Carlo study. Section 4 concludes and gives an outlook of what could be done in further research investigations. All proofs are postponed to the Appendix.

2. Nonparametric MBC estimation

2.1. Definition of multivariate JLN and TS estimator

In this section we will define our newly proposed estimators: the multivariate Jones Linton Nielsen estimator (MV-JLN) and the multivariate Terrell Scott estimator (MV-TS). Consider therefore a random sample $\{X_{i1}, \dots, X_{id}\}_{i=1}^n$ of independent and identically distributed random vectors with joint density f . The MBC technique proposed by Jones et al. (1995) is based on the identity $f(x) = \hat{f}_m(x) \left(\frac{f(x)}{\hat{f}_m(x)} \right)$. In this sense the unknown density f at a vector $x = (x_1, \dots, x_d) \in (\mathbb{R}^+)^d$ is estimated by

$$\hat{f}_{JLN,m,h}(x) := \hat{f}_{m,h}(x) \frac{1}{n} \sum_{j=1}^n \frac{\prod_{j=1}^d K_{m,h_j,x_j}(X_{ij})}{\hat{f}_{m,h}(X_i)},$$

where $X_i := (X_{i1}, \dots, X_{id})$. The second MBC technique published by Terrell and Scott is based on a geometric extrapolation which originates from numerical mathematics. Consider therefore a constant $c \in (0, 1)$ which is independent of the dimension d . The unknown density function f is now estimated by

$$\hat{f}_{TS,m,h}(x) := \left(\hat{f}_{m,h}(x) \right)^{\frac{1}{1-c}} \left(\hat{f}_{m,h/c}(x) \right)^{-\frac{c}{1-c}}.$$

In contrast to other boundary correction methods both estimators only produce positive values for f .

2.2. Bias and Variance

In this section we will develop the convergence properties of our estimators. Due to notational limitations and to the sake of brevity we will restrict ourselves to the case where $h_1 = \dots = h_d \equiv h$ and also where all components of the vector x lie in the interior or at the boundary of the support. Results for more involved locations as well as unequal bandwidths can be examined in a straight forward manner. To explore the asymptotic properties we have to impose two assumptions:

Assumption 1. f has four continuous and bounded partial derivatives and $f(x) > 0$.

Assumption 2. The bandwidth h fulfills $h \rightarrow 0$ as well as $nh^{d(r_j+1/2)+2} \rightarrow 0$ as $n \rightarrow \infty$, where

$$r_j = \begin{cases} 1/2, & \text{for } j = \text{G, MG, RIG, NM, B} \\ 1, & \text{for } j = \text{LN, BS,} \\ 3/2, & \text{for } j = \text{IG.} \end{cases}$$

Assumption 1 is usually needed for estimation via 4-th order product kernels. Assumption 2 is used to control the convergence order of remainder terms appearing in certain Taylor expansions during the

proofs.

We are now ready to formulate the main results of our paper. We will distinguish between two cases namely the case where $x_j/h_j \rightarrow \infty$ respectively $x_j/h_j \rightarrow \kappa_j$ for all $j = 1, \dots, d$. The results for other locations of x can be easily obtained and are here omitted.

Theorem 2.1. *Under Assumptions 1 and 2 the bias of the MV-TS MBC estimator is given by*

$$\text{Bias}\left(\hat{f}_{TS,m,h}(x)\right) = h^2 \frac{1}{c(1-c)} \left[\frac{a_{1,m}^2(x; f)}{f(x)} - a_{2,m}(x; f) \right] + o(h^2) := h^2 \frac{p_m(x; f)}{c(1-c)} + o(h^2),$$

where $a_{j,m}(x; f) := a_{j,m}(x_1, \dots, x_d; f)$, $j = 1, 2$ are functions depending on the choice of the kernel and their explicit forms are given in Table 2 and 3. The variance can be approximated by

$$\text{Var}\left(\hat{f}_{TS,m,h}(x)\right) = \begin{cases} n^{-1}h^{-d/2}f(x)\nu_m(x)\lambda_d(c) + o(n^{-1}h^{-d/2}), & \text{for an interior vector } x, \\ O((nh^{d(r_j+1/2)})^{-1}), & \text{for a boundary vector } x, \end{cases}$$

where $\lambda_d(c) := \frac{(1+c)^{d/2}(1+c^{(d+4)/2}) - (2c)^{(d+2)/2}}{(1-c)^2(1+c)^{d/2}}$ and $\nu_m(x) := \nu_m(x_1, \dots, x_d) = (2\sqrt{\pi})^{-d} \prod_{j=1}^d x_j^{-r_j}$.

Theorem 2.2. *Under Assumptions 1 and 2 the Bias of the MV-JLN MBC estimator is given by*

$$\text{Bias}\left(\hat{f}_{JLN,m,h}(x)\right) = -h^2 f(x) a_{1,m}(x; g) + o(h^2) := q_m(x; f) h^2 + o(h^2),$$

where $a_{1,m}(x; g)$ is defined by replacing f in the definition of $a_{1,m}(x; f)$ by $g = g(x; f) = \frac{a_{1,m}(x; f)}{f(x)}$. Furthermore the Variance can be approximated by

$$\text{Var}\left(\hat{f}_{JLN,m,h}(x)\right) = \begin{cases} n^{-1}h^{-d/2}f(x)\nu_m(x) + o(n^{-1}h^{-d/2}), & \text{for an interior vector } x, \\ O((nh^{d(r_j+1/2)})^{-1}), & \text{for a boundary vector } x, \end{cases}$$

Remark 2.3. The rate of the leading bias term for both estimators is equal to $O(h^2)$ instead of $O(h)$ as in usual asymmetric kernel density estimation. The rate of the variance remains unchanged through the imposed bias correction. Furthermore, the appealing properties of asymmetric kernel density estimation like shrinking variance in terms of the location of the components x_j and freedom of boundary bias are still persisting.

2.3. MSE and MISE convergence rates

The rate of the mean squared error (MSE) can now easily be derived using Theorems 1 and 2. Therefore let $x = (x_1, \dots, x_d)$ be a vector of interior points, then:

$$\text{MSE}\left(\hat{f}_{TS,m,h}(x)\right) = \frac{h^4 p_m^2(x; f)}{c^2(1-c)^2} + n^{-1}h^{-d/2}f(x)\nu_m(x)\lambda_d(c) + o(h^4 + n^{-1}h^{-d/2})$$

as well as

$$\text{MSE}\left(\hat{f}_{JLN,m,h}(x)\right) = h^4 q_m^2(x; f) + n^{-1}h^{-d/2}f(x)\nu_m(x) + o(h^4 + n^{-1}h^{-d/2}).$$

Which leads to the optimal smoothing parameters

$$h_{opt,TS,m} = n^{-2/(8+d)} \left(\frac{f(x)\nu_m(x)\lambda_d(c)(1-c)^2 c^2 d}{8p_m^2(x; f)} \right)^{2/(8+d)}$$

Table 2
Explicit formulas for $a_{1,m}(x; f)$

Kernel	$a_{1,m}(x; f)$
G	$\sum_{j=1}^d \left(\frac{\partial f(x)}{\partial x_j} + \frac{x_j}{2} \frac{\partial^2 f(x)}{\partial^2 x_j} \right)$
MG	$\sum_{j=1}^d \left(\frac{x_j}{2} \frac{\partial^2 f(x)}{\partial^2 x_j} 1(x_j > 2h) + \xi_h(x_j) \frac{\partial f(x)}{\partial x_j} 1(x_j \leq 2h) \right),$ where $\xi_h(x_j) := \rho_h(x_j) - \frac{x_j}{h} = O(1)$
IG	$\frac{1}{2} \sum_{j=1}^d x_j^3 \frac{\partial^2 f(x)}{\partial^2 x_j}$
RIG	$\frac{1}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x)}{\partial^2 x_j}$
LN	$\frac{1}{2} \sum_{j=1}^d x_j \left(\frac{\partial f(x)}{\partial x_j} + x_j \frac{\partial^2 f(x)}{\partial^2 x_j} \right)$
BS	$\frac{1}{2} \sum_{j=1}^d x_j \left(\frac{\partial f(x)}{\partial x_j} + \frac{\partial^2 f(x)}{\partial^2 x_j} \right)$
NM	$\sum_{j=1}^d \left(\frac{x_j}{4} \frac{\partial^2 f(x)}{\partial^2 x_j} 1(x_j > 2h) + \xi_h(x_j) \frac{\partial f(x)}{\partial x_j} 1(x_j \leq 2h) \right),$
B	$\sum_{j=1}^d \left((1 - 2x_j) \frac{\partial f(x)}{\partial x_j} + \frac{1}{2} x_j (1 - x_j) \frac{\partial^2 f(x)}{\partial^2 x_j} \right)$

and

$$h_{opt, JLN, m} = n^{-2/(8+d)} \left(\frac{f(x) \nu_m(x) d}{8q_m^2(x; f)} \right)^{2/(8+d)}.$$

These parameters lead us now to the optimal rate of convergence of the MSE for an interior vector x :

$$MSE_{opt} \left(\hat{f}_{TS, m, h}(x) \right) \sim n^{-8/(8+d)} \frac{(d+8)}{(88/d_d)^{d/(8+d)}} \gamma_d(c) p_m^{2d/(8+d)}(x; f) (\nu_m(x) f(x))^{8/(8+d)}$$

and

$$MSE_{opt} \left(\hat{f}_{JLN, m, h}(x) \right) \sim n^{-8/(8+d)} \frac{(d+8)}{(88/d_d)^{d/(8+d)}} q_m^{2d/(8+d)}(x; f) (\nu_m(x) f(x))^{8/(8+d)}$$

where

$$\gamma_d(c) := \frac{\lambda_d(c)^{8/(8+d)}}{(c(1-c))^{2d/(8+d)}} = \left(\frac{(1+c)^{d/2} (1+c^{(4+d)/2}) - (2c)^{(2+d)/2}}{c^{d/4} (1+c)^{d/2} (1-c)^{(8+d)/4}} \right)^{8/(8+d)}.$$

Remark 2.4. Observe that Assumption 2 guarantees that the smoothing parameter h converges slower than $O(n^{-1/((r_j+1/2)d+2)})$. As we can see, the optimal bandwidth parameter h_{opt} is of order $O(n^{-8/(8+d)})$ for both estimators. This means that our requirement is fulfilled for the G, MG, RIG, NM and B kernel only in the case $d < 4$. For other kernels, h_{opt} does not fulfill this assumption. From a practical point of view this should not be a big problem due to the fact that in finite sample examples the bandwidth is often chosen by a data driven method and also in dimensions higher than 3, nonparametric density estimation suffers extremely from the curse of dimensionality.

Similarly, we can derive the rates for boundary vectors. The MSE of both estimators is in this case of order $O(h^4 + (nh^{d(r_j+1/2)})^{-1})$ and the optimal bandwidth parameter h_{opt}^* fulfills consequently $h_{opt}^* = O(n^{-1/(4+d(r_j+1/2))})$ which leads to an optimal mean squared error for boundary vectors of order

$$MSE^* \left(\hat{f}_{JLN, m, h}(x) \right) = MSE^* \left(\hat{f}_{TS, m, h}(x) \right) = O(n^{-4/((4+d(r_j+1/2)))}).$$

It is natural to include also a global performance criterion like the mean integrated squared error (MISE). As in Chen (2000) suggested, a trimming argument yields that the unwanted slower rates near the origin

Table 3Explicit formulas for $a_{2,m}(x; f)$

Kernel	$a_{2,m}(x; f)$
G	$\sum_{j=1}^d \left(\frac{\partial^2 f(x)}{\partial^2 x_j} + \frac{5}{6} x_j \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{1}{8} x_j \frac{\partial^4 f(x)}{\partial^4 x_j} \right) + \sum_{i \neq j} \left(\frac{\partial^2 f(x)}{\partial x_j \partial x_i} + \frac{x_i x_j}{4} \frac{\partial^4 f(x)}{\partial^2 x_i \partial^2 x_j} \right)$
MG	$\sum_{j=1}^d \left(\frac{x_j}{3} \frac{\partial^2 f(x)}{\partial^2 x_j} + \frac{5}{6} x_j \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^2}{8} \frac{\partial^4 f(x)}{\partial^4 x_j} \right) 1(x_j > 2h)$ $+ \frac{1}{2} \sum_{j=1}^d \left(\xi_h^2(x_j) + \xi_h(x_j) + \frac{x_j}{h} \right) \frac{\partial^2 f(x)}{\partial^2 x_j} 1(x_j \leq 2h)$ $+ \sum_{i \neq j} \frac{x_i x_j}{4} \frac{\partial^4 f(x)}{\partial^4 x_j} 1(x_i > 2h; x_j > 2h)$ $+ \sum_{i \neq j} \xi_h(x_j) \xi_h(x_i) \frac{\partial^2 f(x)}{\partial x_j \partial x_i} 1(x_j \leq 2h; x_i \leq 2h)$
IG	$\frac{1}{2} \sum_{j=1}^d \left(x_j^5 \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^2}{4} \frac{\partial^4 f(x)}{\partial^4 x_j} \right) + \sum_{i \neq j} \frac{x_i^3 x_j^3}{4} \frac{\partial^4 f(x)}{\partial^2 x_j \partial^2 x_i}$
RIG	$\frac{1}{2} \sum_{j=1}^d \left(\frac{\partial^2 f(x)}{\partial^2 x_j} + x_j \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^2}{4} \frac{\partial^4 f(x)}{\partial^4 x_j} \right) + \sum_{i \neq j} \frac{x_i x_j}{4} \frac{\partial^4 f(x)}{\partial^2 x_j \partial^2 x_i}$
LN	$\sum_{j=1}^d \left(\frac{x_j}{8} \frac{\partial f(x)}{\partial x_j} + \frac{7x_j}{8} \frac{\partial^2 f(x)}{\partial^2 x_j} + \frac{3x_j^3}{4} \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^4}{8} \frac{\partial^4 f(x)}{\partial^4 x_j} \right)$ $+ \sum_{i \neq j} x_i x_j \left(\frac{\partial^2 f(x)}{\partial x_j \partial x_i} + \frac{x_i x_j}{4} \frac{\partial^4 f(x)}{\partial^2 x_i \partial^2 x_j} \right)$
BS	$\sum_{j=1}^d \left(\frac{3x_j^2}{4} \frac{\partial^2 f(x)}{\partial^2 x_j} + \frac{3x_j^3}{4} \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^4}{8} \frac{\partial^4 f(x)}{\partial^4 x_j} \right) + \sum_{i \neq j} x_i x_j \left(\frac{\partial^2 f(x)}{\partial x_j \partial x_i} + \frac{x_i x_j}{4} \frac{\partial^4 f(x)}{\partial^2 x_i \partial^2 x_j} \right)$
NM	$\frac{1}{8} \sum_{j=1}^d \left(\frac{\partial^2 f(x)}{\partial^2 x_j} + \frac{x_j}{3} \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{x_j^2}{4} \frac{\partial^4 f(x)}{\partial^4 x_j} \right) 1(x_j > 2h) + \sum_{i \neq j} \frac{x_i x_j}{16} \frac{\partial^4 f(x)}{\partial^2 x_j \partial^2 x_i} 1(x_i > 2h; x_j > 2h)$ $+ \sum_{j=1}^d \left(\left(\xi_h(x_j) + \frac{x_j}{h} \right)^2 \frac{\Gamma\left(\frac{\xi_h(x_j) + \frac{x_j}{h}}{2}\right) \Gamma\left(\frac{\xi_h(x_j) + \frac{x_j}{h}}{2} + 1\right)}{\Gamma^2\left(\frac{\xi_h(x_j) + \frac{x_j}{h}}{2} + 1\right)} - \frac{2x_j \xi_h(x_j)}{h} + \left(\frac{x_j}{h}\right)^2 \right) 1(x_j \leq 2h)$ $+ \sum_{i \neq j} \xi_h(x_j) \xi_h(x_i) \frac{\partial^2 f(x)}{\partial x_j \partial x_i} 1(x_j \leq 2h; x_i \leq 2h)$
B	$\sum_{j=1}^d \left(-2(1 - 2x_j) \frac{\partial f(x)}{\partial x_j} + \frac{1}{2}(11x_j^2 - 11x_j + 2) \frac{\partial^2 f(x)}{\partial^2 x_j} \right)$ $+ \sum_{j=1}^d \left(\frac{5}{6} x_j (1 - x_j) (1 - 2x_j) \frac{\partial^3 f(x)}{\partial^3 x_j} + \frac{1}{8} x_j^2 (1 - x_j)^2 \frac{\partial^4 f(x)}{\partial^4 x_j} \right)$ $+ \sum_{i \neq j} \left((1 - 2x_i) (1 - 2x_j) \frac{\partial^2 f(x)}{\partial x_j \partial x_i} + \frac{1}{2} (1 - 2x_i) x_j (1 - x_j) \frac{\partial^3 f(x)}{\partial x_i \partial^2 x_j} \right)$ $+ \sum_{i \neq j} \frac{1}{4} x_i (1 - x_i) x_j (1 - x_j) \frac{\partial^4 f(x)}{\partial^2 x_i \partial^2 x_j}$

do not affect the global performance. One can easily use this argument in each direction due to the product form of the chosen kernel (cf. Bouezmarni and Rombouts (2010)). Therefore we have the following rates for the MISEs of the proposed estimators:

$$MISE\left(\hat{f}_{TS,m,h}(x)\right) = \frac{h^4}{c^2(1-c)^2} \int_0^\infty p_m^2(x) dx + \frac{\lambda_d(c)}{nh^{d/2}} \int_0^\infty f(x) \nu_m(x) dx + o(h^4 + n^{-1}h^{-d/2})$$

as well as

$$MISE\left(\hat{f}_{JLN,m,h}(x)\right) = h^4 \int_0^\infty q_m^2(x) dx + \frac{1}{nh^{d/2}} \int_0^\infty f(x) \nu_m(x) dx + o(h^4 + n^{-1}h^{-d/2}),$$

provided that all appearing integrals exist.

The optimal smoothing parameters of the rate for the MISE are therefore given by

$$\bar{h}_{opt,TS,m} = n^{-2/(8+d)} \left(\lambda_d(c) (1-c)^2 c^2 d \frac{\int_0^\infty f(x) \nu_m(x) dx}{8 \int_0^\infty p_m^2(x; f)} \right)^{2/(8+d)}$$

and

$$\bar{h}_{opt,JLN,m} = n^{-2/(8+d)} \left(\frac{d \int_0^\infty f(x) \nu_m(x) dx}{8 \int_0^\infty q_m^2(x; f) dx} \right)^{2/(8+d)}.$$

Consequently, the optimal MISEs are of order

$$MISE_{opt}(\hat{f}_{TS,m,h}(x)) \sim n^{-8/(8+d)} \frac{(d+8)\gamma_d(c)}{(8^{8/d}d)^{d/(8+d)}} \left(\int_0^\infty p_m^2(x; f) dx \right)^{2/(8+d)} \left(\int_0^\infty \nu_m(x) f(x) dx \right)^{8/(8+d)}$$

and

$$MISE_{opt}(\hat{f}_{JLN,m,h}(x)) \sim n^{-8/(8+d)} \frac{(d+8)}{(8^{8/d}d)^{d/(8+d)}} \left(\int_0^\infty q_m^2(x; f) dx \right)^{2/(8+d)} \left(\int_0^\infty \nu_m(x) f(x) dx \right)^{8/(8+d)}.$$

We remark that the function $\gamma_d(c)$ for our relevant cases is minimized at $c \approx 0.283$ for $d = 2$. We will use this constant exclusively in our subsequent analysis.

The optimal MISE rate for both estimators is therefore $O(n^{-8/(8+d)})$ which is faster than the rate $O(n^{-4/(4+d)})$ obtained in Bouezmarni and Rombouts (2010).

Remark 2.5. We can easily transmit the results when bias and variance at a generic vector $x \in (\mathbb{R}^d)^+$ have to be established. As we have seen, the bias remains unchanged and is uniform over the whole support of order $O(h^2)$. The variance has now the following form:

$$Var(\hat{f}_{JLN,m,h}(x)) = Var(\hat{f}_{TS,m,h}(x)) = O\left(n^{-1} \prod_{l=1}^d h^{-(1/2+r_j 1_l)}\right),$$

which depends on the location of the respective components and furthermore where $1_l := 1(x_l/h \rightarrow \kappa_l > 0)$ is a function which indicates if a component lies in the boundary region.

3. Finite sample performance

3.1. Monte Carlo Simulation Setup

In this section we compare the performance of the bivariate JLN and TS- estimator with the classical (“C”) bivariate nonparametric Gamma kernel density estimator

$$\hat{f}_{C,m,h}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^n K_{G,h_1,x_1}(X_{i1}) K_{G,h_2,x_2}(X_{i2})$$

in a Monte Carlo simulation study. The smoothing parameters are chosen by the simple “rule of thumb” method $h_i = \sigma(X_i) n^{-1/5}$, $i = 1, 2$ with $\sigma(\cdot)$ denoting the sample standard deviation. We consider data samples drawn from the following distributions:

- Bivariate independent Gamma distribution with shape parameter $\alpha = 1.5$ and scale parameter 1.
- Bivariate independent Weibull distribution with equal shape and scale parameters $\alpha = \beta = 1.5$.
- Bivariate independent Log-normal distribution with log-scale parameter $\mu = 0$ and shape parameter $\sigma = 0.75$.

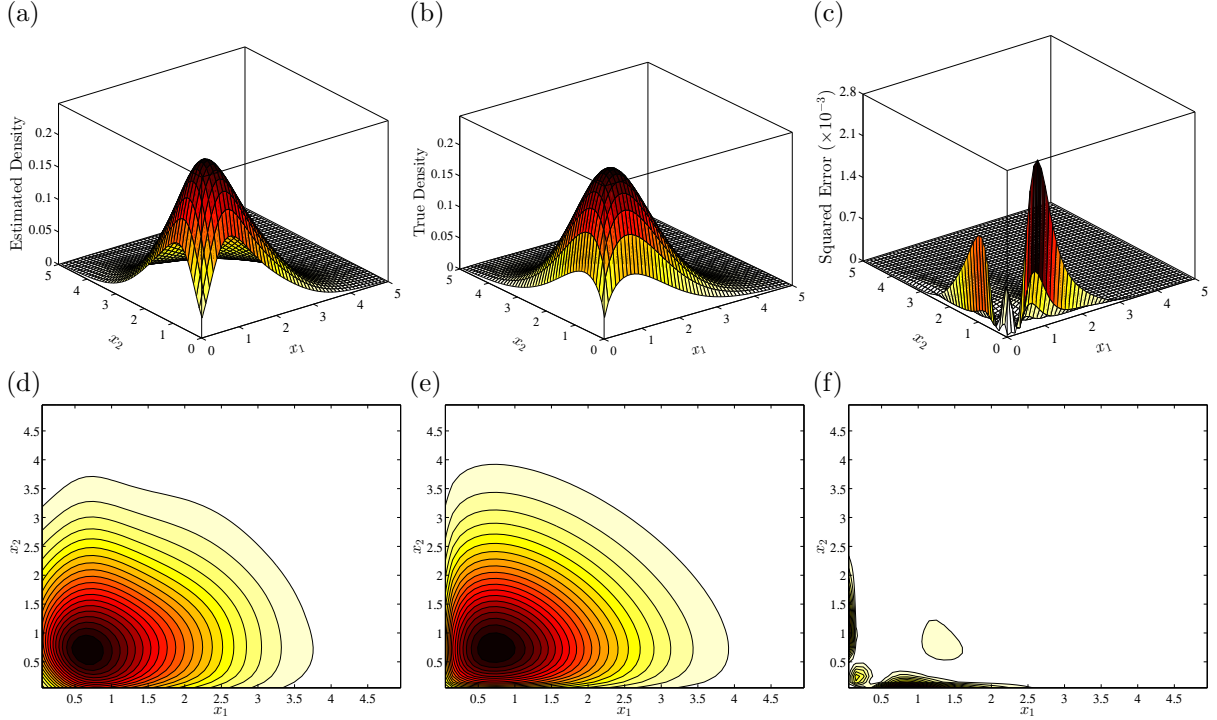


Fig. 1. Surface and contour plots of estimated Weibull(1.5,1.5) density. (a),(d): JLN-estimated density with G kernel. (b),(e): True density. (c),(f): Squared error.

- Bivariate independent Inverse Gaussian distribution with mean parameter $\mu = 0.8$ and shape parameter $\lambda = 1$.

For each distribution 1000 data sets of sample sizes $n = 250$ and $n = 500$ were simulated. We performed the kernel estimations with the G and MG kernel, since both became quite popular in recent literature. The constant c in the TS- estimator is set to 0.283, since this value minimizes the function $\gamma_2(c)$ over the interval $(0, 1)$ as we have already mentioned above. The performance is measured in terms of the mean root integrated squared error (“MRISE”), which is defined by

$$\text{MRISE}(\hat{f}_{g,m,h}) = E \left[\left(\int (\hat{f}_{g,m,h}(x) - f(x))^2 dx \right)^{1/2} \right]$$

as well as the integrated absolute bias (“IAB”)

$$\text{IAB}(\hat{f}_{g,m,h}) = \int |E[\hat{f}_{g,m,h}(x)] - f(x)| dx$$

with $g = \text{TS}$, JLN or C . Thereby the expected values are approximated with the sample mean. Additionally we provide the standard deviation of the root integrated squared error. The integrals are approximated on an equidistant grid of 50^2 , 100^2 and 200^2 points over the square $(0, 5) \times (0, 5)$, since the probability mass of the given distributions is negligible outside this rectangle.

As an example we plotted the true density of a Weibull(1.5,1.5) distribution together with the estimated density and the corresponding squared error in Figure 1.

Table 4
Simulation results (50×50 grid points)

Distribution:		Gamma		Weibull		Log-normal		Inv-Gaussian	
Kernel:		G	MG	G	MG	G	MG	G	MG
Panel 1: $n = 250$									
TS	MRISE	0.0669 (0.0111)	0.0739 (0.0098)	0.0738 (0.0145)	0.0706 (0.0105)	0.1751 (0.0125)	0.1796 (0.0121)	0.3344 (0.0169)	0.3546 (0.0155)
	IAB	0.1172	0.1765	0.1445	0.1629	0.2937	0.3854	0.3187	0.4474
JLN	MRISE	0.0602 (0.0125)	0.0608 (0.0086)	0.0644 (0.0140)	0.0653 (0.0107)	0.1512 (0.0151)	0.1683 (0.0133)	0.2988 (0.0216)	0.3337 (0.0190)
	IAB	0.1197	0.0867	0.1336	0.0934	0.2759	0.2944	0.3045	0.3542
C	MRISE	0.0933 (0.0097)	–	0.0984 (0.0111)	–	0.2078 (0.0117)	–	0.3871 (0.0163)	–
	IAB	0.2282	–	0.2463	–	0.3923	–	0.4251	–
Panel 2: $n = 500$									
TS	MRISE	0.0589 (0.0082)	0.0615 (0.0064)	0.0645 (0.0106)	0.0587 (0.0079)	0.1625 (0.0095)	0.1633 (0.0096)	0.3158 (0.0130)	0.3279 (0.0117)
	IAB	0.1023	0.1429	0.1282	0.1395	0.2686	0.3437	0.2983	0.4003
JLN	MRISE	0.0514 (0.0091)	0.0526 (0.0064)	0.0527 (0.0109)	0.0537 (0.0073)	0.1357 (0.0116)	0.1528 (0.0106)	0.2732 (0.0157)	0.3125 (0.0151)
	IAB	0.0991	0.0768	0.1084	0.0808	0.2427	0.2672	0.2714	0.3337
C	MRISE	0.0845 (0.0071)	–	0.0890 (0.0081)	–	0.1978 (0.0088)	–	0.3684 (0.0123)	–
	IAB	0.2077	–	0.2250	–	0.3724	–	0.4002	–

Note: The C estimator only uses the G kernel.

3.2. Simulation Results

The results of the Monte Carlo simulations are provided in Tables 4, 5 and 6, separated into the different grids sizes for numerical integral evaluation. One recognizes the similarity of the results across all grid and sample sizes. Obviously the classical Gamma kernel estimator C is outperformed by the TS and the JLN- estimators by far, what is particularly observable in the IAB. For instance, the IAB of the C estimator using the Gamma kernel of the Weibull distribution using a grid size of 100×100 and a sample size of 500 is 0.2257, whereas the IAB of the TS and JLN- estimator accounts for 0.1276 and 0.1097 respectively. This huge discrepancy is in fact rather the rule than the exception. Only the C estimator exhibits a smaller IAB than the TS- estimator with MG kernel for the Inverse Gaussian distribution, what is relativized by the bad performance of the mean root integrated squared error in that case and the superior performance of the TS- estimator with Gamma kernel. In terms of MRISE and IAB, the JLN- estimator is – and that is indeed worth to be mentioned – nearly always outperforming the TS and the C estimator.

The question, which kernel behaves more beneficial, is much harder to answer. In terms of MRISE the JLN- estimator performs marginally better using the Gamma kernel. On the other hand the IAB of the

Table 5
Simulation results (100×100 grid points)

Distribution:		Gamma		Weibull		Log-normal		Inv-Gaussian	
Kernel:		G	MG	G	MG	G	MG	G	MG
Panel 1: $n = 250$									
TS	MRISE	0.0697 (0.0112)	0.0770 (0.0096)	0.0759 (0.0141)	0.0735 (0.0104)	0.1735 (0.0122)	0.1799 (0.0125)	0.3309 (0.0168)	0.3512 (0.0146)
	IAB	0.1186	0.1802	0.1460	0.1692	0.2901	0.3871	0.3258	0.4487
JLN	MRISE	0.0638 (0.0124)	0.0642 (0.0082)	0.0654 (0.0143)	0.0665 (0.0096)	0.1513 (0.0154)	0.1668 (0.0137)	0.2962 (0.0207)	0.3311 (0.0192)
	IAB	0.1241	0.0905	0.1333	0.0960	0.2746	0.2920	0.3170	0.3552
C	MRISE	0.0945 (0.0097)	–	0.0998 (0.0107)	–	0.2092 (0.0118)	–	0.3837 (0.0176)	–
	IAB	0.2294	–	0.2486	–	0.3958	–	0.4281	–
Panel 2: $n = 500$									
TS	MRISE	0.0620 (0.0083)	0.0650 (0.0062)	0.0655 (0.0106)	0.0609 (0.0074)	0.1620 (0.0094)	0.1629 (0.0093)	0.3107 (0.0130)	0.3243 (0.0120)
	IAB	0.1040	0.1457	0.1276	0.1420	0.2673	0.3410	0.3024	0.4061
JLN	MRISE	0.0540 (0.0090)	0.0559 (0.0058)	0.0536 (0.0109)	0.0560 (0.0069)	0.1352 (0.0108)	0.1538 (0.0111)	0.2701 (0.0157)	0.3090 (0.0145)
	IAB	0.1012	0.0808	0.1097	0.0828	0.2414	0.2676	0.2806	0.3270
C	MRISE	0.0869 (0.0071)	–	0.0903 (0.0081)	–	0.1964 (0.0087)	–	0.3650 (0.0124)	–
	IAB	0.2106	–	0.2257	–	0.3695	–	0.4051	–

Note: The C estimator only uses the G kernel.

JLN- estimator is smaller for the Gamma and Weibull distribution using the MG kernel, whereas the opposite is observable for the Log-normal and Inverse Gaussian distributions. The TS- estimator exhibits an essential better performance in terms of IAB when the Gamma kernel is used, while the MRISE seems to behave very similar for both kernels.

3.3. Real Data Example

For an illustration purpose we have applied our introduced methods to a bivariate data set containing the eruption waiting times (X_1) and eruption duration times (X_2) of the Old Faithful Geyser in Yellowstone National Park, Wyoming, USA¹. This data set contains 272 bivariate samples and has been investigated by Hirukawa (2010), who estimated the marginal density of the eruption times using univariate Beta kernel based JLN- and TS-estimators. We extend his results to the estimation of the bivariate joint probability density, including the waiting times data set.

¹We used the data provided from <http://www.stat.cmu.edu/~larry/all-of-statistics/=data/faithful.dat>

Table 6
Simulation results (200×200 grid points)

Distribution:		Gamma		Weibull		Log-normal		Inv-Gaussian	
Kernel:		G	MG	G	MG	G	MG	G	MG
Panel 1: $n = 250$									
TS	MRISE	0.0703 (0.0107)	0.0791 (0.0099)	0.0774 (0.0139)	0.0745 (0.0104)	0.1749 (0.0129)	0.1795 (0.0123)	0.3299 (0.0165)	0.3510 (0.0156)
	IAB	0.1190	0.1794	0.1500	0.1688	0.2938	0.3845	0.3249	0.4511
JLN	MRISE	0.0648 (0.0125)	0.0653 (0.0082)	0.0661 (0.0149)	0.0678 (0.0100)	0.1506 (0.0155)	0.1679 (0.0138)	0.2951 (0.0206)	0.3310 (0.0189)
	IAB	0.1235	0.0916	0.1331	0.0966	0.2745	0.2951	0.3141	0.3547
C	MRISE	0.0962 (0.0096)	–	0.1005 (0.0111)	–	0.2085 (0.0121)	–	0.3840 (0.0168)	–
	IAB	0.2320	–	0.2496	–	0.3945	–	0.4309	–
Panel 2: $n = 500$									
TS	MRISE	0.0634 (0.0085)	0.0664 (0.0057)	0.0665 (0.0106)	0.0618 (0.0070)	0.1621 (0.0094)	0.1639 (0.0099)	0.3104 (0.0127)	0.3243 (0.0119)
	IAB	0.1044	0.1460	0.1289	0.1439	0.2686	0.3437	0.3013	0.4053
JLN	MRISE	0.0552 (0.0092)	0.0572 (0.0054)	0.0544 (0.0103)	0.0570 (0.0070)	0.1356 (0.0118)	0.1533 (0.0108)	0.2686 (0.0160)	0.3082 (0.0153)
	IAB	0.1011	0.0826	0.1102	0.0835	0.2425	0.2681	0.2792	0.3267
C	MRISE	0.0871 (0.0067)	–	0.0911 (0.0079)	–	0.1961 (0.0091)	–	0.3643 (0.0125)	–
	IAB	0.2096	–	0.2273	–	0.3694	–	0.4047	–

Note: The C estimator only uses the G kernel.

Since the support of the data is embedded in the set $[43, 96] \times [1.6, 5.1]$, we have used the following transformation to get a $[0, 1] \times [0, 1]$ supported data set:

$$Z_i = \frac{X_i}{1.1 \max(X_i)}, \quad i = 1, 2.$$

We performed a TS and JLN- estimation on the transformed data using Beta kernels and the least squares cross-validation method (LSCV) to select the bandwidth. The idea behind the LSCV method is minimizing the integrated squared error, which is defined as

$$\begin{aligned} \text{ISE}(h) &= \int (\hat{f}_{g,m,h}(x) - f(x))^2 dx \\ &= \int \hat{f}_{g,m,h}^2(x) dx - 2 \int \hat{f}_{g,m,h}(x) f(x) dx + \int f^2(x) dx \\ &= \int \hat{f}_{g,m,h}^2(x) dx - 2E[\hat{f}_{g,m,h}(X)] + \int f^2(x) dx \end{aligned}$$

with $h = (h_1, h_2)$ and $g = \text{TS}$ or JLN . Since the last term is independent of h_1 and h_2 , we are only interested in finding a minimum of the remaining two terms. The LSCV selected bandwidth h is therefore defined as

$$\text{LSCV}(h) = \underset{h_1, h_2}{\text{argmin}} \left(\int \hat{f}_{g,m,h}^2(x) dx - 2\hat{E}[\hat{f}_{g,m,h}(X_1, X_2)] \right),$$

where \hat{E} denotes the expectation with respect to the empirical distribution. The last expression is determined via the so called “leave one out” estimator, which we define as

$$\begin{aligned} \hat{E}[\hat{f}_{TS,m,h}(X_1, X_2)] &= \frac{1}{n} \sum_{i=1}^n \hat{f}_{TS,m,h,-i}(X_{i1}, X_{i2}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \left(\sum_{\substack{j=1 \\ j \neq i}}^n K_{m,h_1,X_{i1}}(X_{j1}) K_{m,h_2,X_{i2}}(X_{j2}) \right)^{\frac{1}{1-c}} \left(\sum_{\substack{j=1 \\ j \neq i}}^n K_{m,h_1/c,X_{i1}}(X_{j1}) K_{m,h_2/c,X_{i2}}(X_{j2}) \right)^{-\frac{c}{1-c}} \end{aligned}$$

for the TS estimator and as

$$\begin{aligned} \hat{E}[\hat{f}_{JLN,m,h}(X_1, X_2)] &= \frac{1}{n} \sum_{i=1}^n \hat{f}_{JLN,m,h,-i}(X_{i1}, X_{i2}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \left(\sum_{\substack{j=1 \\ j \neq i}}^n K_{m,h_1,X_{i1}}(X_{j1}) K_{m,h_2,X_{i2}}(X_{j2}) \sum_{\substack{j=1 \\ j \neq i}}^n \frac{K_{m,h_1,X_{i1}}(X_{j1}) K_{m,h_2,X_{i2}}(X_{j2})}{\sum_{k=1}^n K_{m,h_1,X_{j1}}(X_{k1}) K_{m,h_2,X_{j2}}(X_{k2})} \right) \end{aligned}$$

for the JLN estimator.

We calculated bandwidths using the LSCV method for the Geyser data on a grid with 50 different values for each h . The resulting shapes are displayed in Figure 2 and we obtained $\hat{h} = (0.0164, 0.0155)$ for the TS estimator and $\hat{h} = (0.0078, 0.0116)$ for the JLN estimator. The results are presented in Figure 3 and both methods provide similar results.

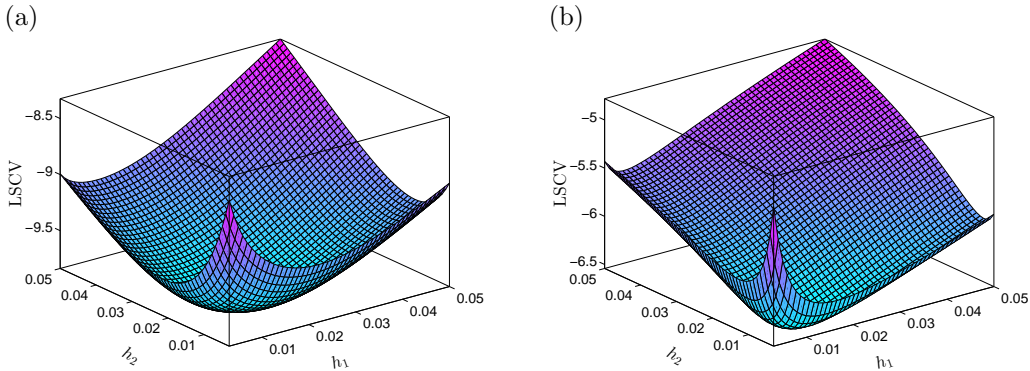


Fig. 2. LSCV shapes for $(h_1, h_2) \in (0.003, 0.05) \times (0.003, 0.05)$, calculated on an equidistant grid with 50×50 grid points. (a) TS-estimator with B kernel. (b) JLN-estimator with B kernel.

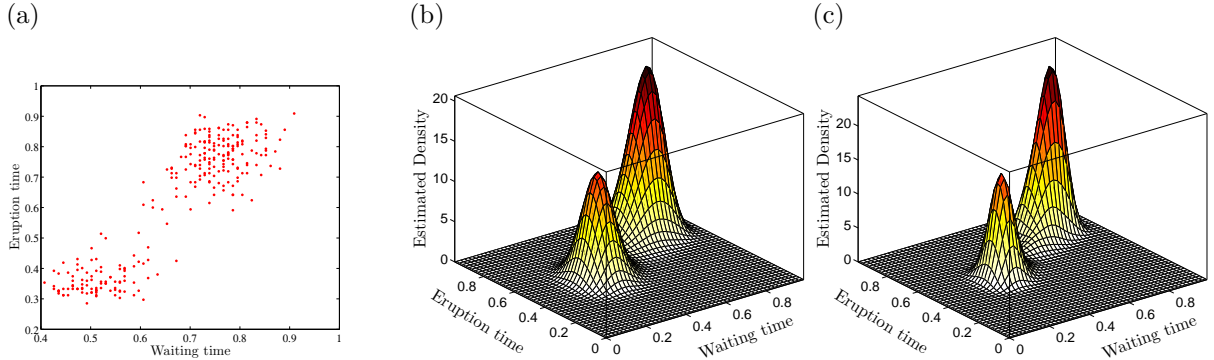


Fig. 3. Estimated density of Old Faithful Geyser data. (a) Scatter plot of transformed data. (b) TS-estimated density with B kernel. (c) JLN-estimated density with B kernel.

4. Conclusion

This paper has proposed two new MBC techniques for nonparametric density estimation of multivariate bounded data which reduce the order of magnitude of the bias from $O(h)$ to $O(h^2)$. It is demonstrated that both classes of estimators which were originally investigated in Hirukawa (2010) for compact supported and in Hirukawa and Sakudo (2014a) for positively supported data can be extended to the multivariate setting. The variance is not influenced by this procedure and both estimators possess a MSE of order $O(n^{-8/(8+d)})$.

Moreover the results of the Monte Carlo simulations reveal the superior performance of the JLN estimator especially compared to the classical one. This observation coincides with the univariate results from Hirukawa (2010) as well as Hirukawa and Sakudo (2014a).

Further research can be done in applying these methods in nonparametric multivariate regression models as it was proposed in Hjort and Glad (1995) in the univariate setting. Moreover it could be interesting to use the proposed Beta kernel estimators for nonparametric estimation of copula densities, see for example Charpentier et al. (2007). Both topics are under investigation and will be covered in separate papers.

5. Appendix

In this section we will derive the above rates for bias and variance. The arguments are mainly based on Taylor expansions and distribution specific properties. Moreover we only focus on the proof for the use of the Gamma kernel due to the sake of brevity. We start with the MV-TS estimator. The corresponding proof is based on Hirukawa (2010).

Proof of Theorem 2.1. For the bias of the MV-TS estimator, we start with a multivariate Taylor expansion up to order 4. Therefore let Y_i , $i = 1, \dots, d$ be independent and Gamma-distributed random variables such that $Y_i \sim G(x_i/h + 1, h)$ with mean $\mu_i = x_i + h$. Using the smoothness assumption on the unknown density f as well as the fact that $E[(Y_i - \mu_i)^r] = O(h^3)$, $r \geq 5$ (cf. Gospodinov and Hirukawa 2007), we

can derive that

$$\begin{aligned}
I_h(x) &:= E[\hat{f}_{G,h}(x)] = \int_{\mathbb{R}^+} \dots \int_{\mathbb{R}^+} \prod_{j=1}^d K_{G,h,x_j}(y_j) f(y_1, \dots, y_d) dy_1 \dots dy_d = E[f(Y_1, \dots, Y_d)] \\
&= f(\mu_1, \dots, \mu_d) + \sum_{j=1}^d \frac{\text{Var}(Y_j)}{2} \frac{\partial^2 f(\mu_1, \dots, \mu_d)}{\partial^2 x_j} + \sum_{j=1}^d \frac{E[(Y_j - \mu_j)^3]}{6} \frac{\partial^3 f(\mu_1, \dots, \mu_d)}{\partial^3 x_j} \\
&= f(x_1, \dots, x_d) + h \left(\sum_{j=1}^d \left(\frac{\partial f(x_1, \dots, x_d)}{\partial x_j} + \frac{x_j}{2} \frac{\partial^2 f(x_1, \dots, x_d)}{\partial^2 x_j} \right) \right) \\
&+ h^2 \left(\sum_{j=1}^d \left(\frac{\partial^2 f(x_1, \dots, x_d)}{\partial^2 x_j} + \frac{5x_j}{6} \frac{\partial^3 f(x_1, \dots, x_d)}{\partial^3 x_j} + \frac{1}{8} \frac{\partial^4 f(x_1, \dots, x_d)}{\partial^4 x_j} \right) \right) \\
&+ \sum_{i \neq j} \left(\frac{\partial^2 f(x_1, \dots, x_d)}{\partial x_i \partial x_j} + x_i x_j \frac{\partial^4 f(x_1, \dots, x_d)}{\partial^4 x_j} \right) + o(h^2) \\
&:= f(x_1, \dots, x_d) \left(1 + h \frac{a_{1,G}(x; f)}{f(x)} + h^2 \frac{a_{2,G}(x; f)}{f(x)} + o(h^2) \right).
\end{aligned}$$

In an analogous manner we can handle the term $I_{h/c}(x)$. Using a Taylor expansion of the exponential function, we find that

$$(I_h(x))^{\frac{1}{1-c}} (I_{h/c}(x))^{\frac{-c}{1-c}} = f(x) + \frac{h^2}{c(1-c)} \left(\frac{a_{1,G}^2(x; f)}{2f(x)} - a_{2,G}(x; f) \right) + o(h^2).$$

To use this expression we will at first decompose our estimator as follows. Let $Z := \hat{f}_{G,h}(x) - I_h(x)$ and $W := \hat{f}_{G,h/c}(x) - I_{h/c}(x)$. It will be seen that $E[Z^2]$, $E[W^2]$ and $E[ZW]$ are all at most of order $O(n^{-1}h^{-d}) = o(h^2)$ due to Assumption 1. Using a Taylor expansion it holds that

$$\begin{aligned}
\hat{f}_{TS,G,h}(x) &= (I_h(x))^{\frac{1}{1-c}} \left(1 + \frac{Z}{I_h(x)} \right)^{\frac{1}{1-c}} (I_{h/c}(x))^{\frac{-c}{1-c}} \left(1 + \frac{W}{I_{h/c}(x)} \right)^{\frac{-c}{1-c}} \\
&= (I_h(x))^{\frac{1}{1-c}} (I_{h/c}(x))^{\frac{-c}{1-c}} + \frac{Z}{1-c} \left(\frac{I_h(x)}{I_{h/c}(x)} \right)^{\frac{1}{1-c}} - \frac{cW}{1-c} \left(\frac{I_h(x)}{I_{h/c}(x)} \right)^{\frac{1}{1-c}} + O(n^{-1}h^{-d}). \tag{5.1}
\end{aligned}$$

Using the asymptotic equivalence of $I_h(x)$ and $I_{h/c}(x)$ we finally derive that

$$\begin{aligned}
E[\hat{f}_{TS,G,h}(x)] &= (I_h(x))^{\frac{1}{1-c}} (I_{h/c}(x))^{\frac{-c}{1-c}} + o(h^2) \\
&= f(x) + \frac{h^2}{c(1-c)} \left(\frac{a_{1,G}^2(x; f)}{2f(x)} - a_{2,G}(x; f) \right) + o(h^2).
\end{aligned}$$

Using (4.1) we are able to decompose the variance as follows

$$\begin{aligned}
\hat{f}_{TS,G,h}(x) &= E \left[\left(\frac{Z}{1-c} - \frac{cW}{1-c} \right)^2 \right] + O(n^{-1}) \\
&= \frac{1}{(1-c)^2} \left(\text{Var}(\hat{f}_{G,h}(x)) - 2c \text{Cov}(\hat{f}_{G,h}(x), \hat{f}_{G,h/c}(x)) + c^2 \text{Var}(\hat{f}_{G,h/c}(x)) \right) + O(n^{-1}).
\end{aligned}$$

We will restrict ourselves only to the case where x is a vector including interior components x_j for all $j = 1, \dots, d$. Using the results in Bouezmarni and Rombouts (2010) we can conclude that

$$\begin{aligned} \text{Var}(\hat{f}_{G,h}(x)) &= n^{-1}h^{-d/2}(2\sqrt{\pi})^{-d}f(x_1, \dots, x_d) \prod_{j=1}^d x_j^{-1/2} + o(n^{-1}h^{-d/2}) \\ \text{Var}(\hat{f}_{G,h/c}(x)) &= n^{-1}h^{-d/2}c^{d/2}(2\sqrt{\pi})^{-d}f(x_1, \dots, x_d) \prod_{j=1}^d x_j^{-1/2} + o(n^{-1}h^{-d/2}) \\ \text{Cov}(\hat{f}_{G,h}(x), \hat{f}_{G,h/c}(x)) &= n^{-1}E \left[\prod_{j=1}^d K_{G,h,x_j}(X_{1j}) \prod_{k=1}^d K_{G,h/c,x_k}(X_{1k}) \right] + O(n^{-1}) \\ &= n^{-1} \prod_{j=1}^d C_b(x_j)(f(x_1, \dots, x_d) + o(1)) + O(n^{-1}), \end{aligned}$$

where

$$C_b(x_j) := \frac{\Gamma\left(\frac{x_j(1+c)}{h} + 1\right) \left(\frac{h}{1+c}\right)^{\frac{x_j(1+c)}{h} + 1}}{\Gamma\left(\frac{x_j}{h} + 1\right) h^{\frac{x_j}{h} + 1} \Gamma\left(\frac{cx_j}{h} + 1\right) \left(\frac{h}{c}\right)^{\frac{cx_j}{h} + 1}}.$$

Using a result in Hirukawa (2010) for the function $C_b(x_j)$, the covariance can be approximated by

$$\text{Cov}(\hat{f}_{G,h}(x), \hat{f}_{G,h/c}(x)) = n^{-1}h^{-d/2} \left(\frac{2c}{1+c}\right)^{d/2} f(x_1, \dots, x_d) \prod_{j=1}^d x_j^{-1/2} + o(n^{-1}h^{-d/2}).$$

Bringing all three parts of the variance of the MV-TS estimator together we are finally able to deduce that

$$\begin{aligned} \text{Var}(\hat{f}_{TS,G,h}(x)) &\doteq \frac{1}{(1-c)^2} \left(n^{-1}h^{-d/2} \frac{f(x_1, \dots, x_d)}{(2\sqrt{\pi})^d \prod_{j=1}^d x_j^{1/2}} \left(1 - 2c \left(\frac{2c}{1+c}\right)^{d/2} + c^{2+d/2} \right) \right) \\ &= n^{-1}h^{-d/2} \frac{f(x_1, \dots, x_d)}{(2\sqrt{\pi})^d \prod_{j=1}^d x_j^{1/2}} \left(\frac{(1+c)^{d/2}(1+c^{(d+4)/2}) - (2c)^{(d+2)/2}}{(1-c)^2(1+c)^{d/2}} \right) \\ &= n^{-1}h^{-d/2} \frac{f(x_1, \dots, x_d)\lambda_d(c)}{(2\sqrt{\pi})^d \prod_{j=1}^d x_j^{1/2}}. \end{aligned}$$

□

Now we turn to the MV-JLN- estimator. The proof is again based on the results in Hirukawa (2010).

Proof of Theorem 2.2. We start by decomposing the MV-JLN- estimator as follows. Denote by

$$\hat{\alpha}(x) = \hat{\alpha}(x_1, \dots, x_d) := \frac{1}{n} \sum_{i=1}^n \frac{\prod_{j=1}^d K_{G,h,x_j}(X_{ij})}{\hat{f}_{G,h}(X_i)}$$

such that

$$\begin{aligned} E[\hat{f}_{JLN}(x)] &= f(x) + f(x) \left(E \left[\frac{\hat{f}_G(x) - f(x)}{f(x)} \right] + E[(\hat{\alpha}(x) - 1)] + E \left[\frac{(\hat{f}_G(x) - f(x))(\hat{\alpha}(x) - 1)}{f(x)} \right] \right) \\ &:= f(x) + f(x)(I + II + III). \end{aligned}$$

We will handle the terms *I*, *II*, *III* separately and start with term *II*. Using a geometric expansion around $f(x)$, $\hat{\alpha}(x)$ can be approximated by

$$\hat{\alpha}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\prod_{j=1}^d K_{G,h,x_j}(X_{ij})}{f(X_i)} \left(1 - \frac{\hat{f}_{G,h}(X_i) - f(X_i)}{f(X_i)} + \left(\frac{\hat{f}_{G,h}(X_i) - f(X_i)}{f(X_i)} \right)^2 \right) + o(h^2 + (nh^{d/2})^{-1}).$$

Notice that the order of the remainder term is of order $o(h^2 + (nh^{d/2})^{-1}) = o(h^2)$ due to Assumption 2. We now make again use of a multivariate Taylor expansion to handle these terms separately. Remember that

$$g = g(x; f) = \frac{a_{1,G}(x)}{f(x)} = \frac{\sum_{j=1}^d \left(\frac{\partial f}{\partial x_j} + \frac{1}{2} x_j \frac{\partial^2 f}{\partial^2 x_j} \right)}{f(x)}$$

and additionally define

$$g_2 = g_2(x; f) := \frac{a_{2,G}(x)}{f(x)} = \frac{\sum_{j=1}^d \left(\frac{\partial^2 f}{\partial^2 x_j} + \frac{5x_j}{6} \frac{\partial^3 f}{\partial^3 x_j} + \frac{1}{8} \frac{\partial^4 f}{\partial^4 x_j} \right) + \sum_{i \neq j} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{x_i x_j}{4} \frac{\partial^4 f}{\partial^2 x_i \partial^2 x_j} \right)}{f(x)}.$$

Using these abbreviations, the term *II* can be approximated by

$$II = E[\hat{\alpha}(x) - 1] = -hg(x; f) - h^2 \left(\sum_{j=1}^d \left(\frac{\partial g(x; f)}{\partial x_j} + \frac{x_j}{2} \frac{\partial^2 g(x; f)}{\partial^2 x_j} \right) + g_2(x; f) - g^2(x; f) \right) + o(h^2).$$

Under the previous results, term *I* has the following form

$$I = E \left[\frac{(\hat{f}_{G,h}(x) - f(x))}{f(x)} \right] = hg(x; f) + h^2 g_2(x; f) + o(h^2).$$

Finally by the use of the Cauchy-Schwarz inequality and the derived results for terms *I* and *II* one has for the last term *III* that

$$III = E \left[\frac{(\hat{f}_{G,h}(x) - f(x))}{f(x)} (\hat{\alpha}(x) - 1) \right] = -h^2 g^2(x; f) + o(h^2).$$

Applying all results one has the following form for the Bias of our proposed estimator

$$E[\hat{f}_{JLN,G,h}(x)] = f(x) - h^2 f(x) \sum_{j=1}^d \left(\frac{\partial g(x; f)}{\partial x_j} + \frac{x_j}{2} \frac{\partial^2 g(x; f)}{\partial^2 x_j} \right) + o(h^2).$$

We now turn to the variance of the MV-JLN- estimator. At first approximate the estimator by making again use of a geometric expansion and focus only on the first term:

$$\begin{aligned} \hat{f}_{JLN,G,h}(x) &= \hat{f}_{G,h}(x) \left(\frac{1}{n} \sum_{i=1}^n \frac{\prod_{j=1}^d K_{G,h,x_j}(X_{ij})}{\hat{f}_{G,h}(X_i)} \right) \\ &= f(x) \left(1 + \frac{\hat{f}_{G,h}(x) - f(x)}{f(x)} \right) \left(\frac{1}{n} \sum_{i=1}^n \frac{\prod_{j=1}^d K_{G,h,x_j}(X_{ij})}{f(X_i)} \right) \left(1 - \frac{\hat{f}_{G,h}(X_i) - f(X_i)}{f(X_i)} + o(\text{s.o.}) \right) \\ &= f(x) \left(\frac{1}{n} \sum_{i=1}^n \frac{\prod_{j=1}^d K_{G,h,x_j}(X_{ij})}{f(X_i)} \right) \left(2 - \frac{\hat{f}_{G,h}(X_i)}{f(X_i)} \right) := f(x) \frac{1}{n} \sum_{l=1}^n \zeta(X_l). \end{aligned}$$

Now by the trimming argument in Chen (2000) and the use of Stirling's Formula for approximations of the Gamma function, the leading term of the variance of the right-hand side is of the following form:

$$\text{Var}(\hat{f}_{JLN,G,h}(x)) \doteq n^{-1} f(x) \prod_{j=1}^d A(x_j, h),$$

where

$$A(x_j, h) := \frac{h^{-1} \Gamma(2x_j/h + 1)}{2^{2x_j/h+1} \Gamma^2(x_j/h + 1)} = \begin{cases} \frac{h^{-1/2}}{2\sqrt{\pi x_j}} & , \text{ if } x_j/h \rightarrow \infty \\ \frac{h^{-1} \Gamma(2\kappa_j+1)}{2^{2\kappa_j+1} \Gamma^2(\kappa_j+1)} & , \text{ if } x_j/h \rightarrow \kappa_j. \end{cases}$$

Therefore one finally has for a vector $x = (x_1, \dots, x_d)$ possessing exclusively interior components that

$$\text{Var}(\hat{f}_{JLN}(x_1, \dots, x_d)) \doteq \frac{f(x)}{nh^{d/2}} (2\sqrt{\pi})^{-d} \prod_{j=1}^d x_j^{-1/2}.$$

□

References

- Bouezmarni, T., Rombouts, J.V.K. (2010): *Nonparametric density estimation for multivariate bounded data*, Journal of Statistical Planning and Inference, vol. 140, no. 1, 139-152.
- Charpentier A., Fermanian, J.D., Scaillet O. (2007): *The estimation of copulas: theory and practice*, in: J. Rank (Ed.), Copulas: From Theory to Application in Finance, Risk Books, London, 35-60.
- Chen, S.X. (1999): *Beta kernel estimators for density functions*, Computational Statistics and Data Analysis, 31, 131-145.
- Chen, S.X. (2000): *Probability density function estimation using Gamma kernels*, Annals of the Institute of Statistical Mathematics, Vol. 52, 471-480.
- Gospodinov, N., Hirukawa, M. (2007): *Time series nonparametric regression using asymmetric kernels with an application to estimation of scalar diffusion processes*, CIRJE Discussion Paper F-573, University of Tokyo.
- Hirukawa, M. (2010): *Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval*, Computational Statistics and Data Analysis, Vol. 54, Issue 2, February 2010, 473-495.
- Hirukawa, M., Sakudo, M. (2014a): *Non-negative bias reduction methods for density estimation Using Asymmetric Kernels*, Computational Statistics and Data Analysis, Vol. 75, 112-123.
- Hirukawa, M., Sakudo, M. (2014b): *Family of the generalized Gamma kernels: a generator of asymmetric kernels for non-negative data*, submitted paper.
- Hjort, N.L., Glad, I.K. (1995): *Nonparametric density estimation with a parametric start*, The Annals of Statistics, Vol. 23, No. 3, 882-904.
- Jin, X., Kawczak, J. (2003): *Birnbaum-Saunders and lognormal kernel estimators for modeling duration in high frequency financial data*, Annals of Economics and Finance, Vol. 4, 103-124.
- Jones, M.C., Linton, O., Nielsen, J.P. (1995): *A simple bias reduction method for density estimation*, Biometrika, 82, 327-338.
- Marchant, C., Bertin, K., Leiva, V., Saulo, H. (2013): *Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data*, Computational Statistics and Data Analysis, Vol. 63, 1-15.
- Sanhueza, A., Leiva, V., Balakrishnan, N., (2008): *The generalized Birnbaum-Saunders distribution and its theory, methodology, and application*, Communications in Statistics: Theory and Methods, Vol. 37, 645-670.
- Scaillet, O. (2004): *Density estimation using inverse and reciprocal inverse Gaussian kernels*, Journal of Nonparametric Statistics, Vol. 16, 217-226.
- Terrell, G.R., Scott, D.W. (1980): *On improving convergence rates for non-negative kernel density estimation*, The Annals of Statistics, Vol. 8, 1160-1163.

