Computational Statistics and Data Analysis 54 (2010) 473–495

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval[☆]

Masayuki Hirukawa [*]

Department of Economics, Northern Illinois University, Zulauf Hall 515, DeKalb, IL 60115, USA

ARTICLE INFO

Article history:
Received 31 March 2009
Received in revised form 9 September 2009
Accepted 10 September 2009
Available online 23 September 2009

ABSTRACT

This paper demonstrates that two classes of multiplicative bias correction (MBC) techniques, originally proposed for density estimation using symmetric second-order kernels by Terrell and Scott (1980) and Jones et al. (1995), can be applied to density estimation using the beta and modified beta kernels. It is shown that, under sufficient smoothness of the true density, both MBC techniques reduce the order of magnitude in bias, whereas the order of magnitude in variance remains unchanged. Accordingly, mean squared errors of these MBC estimators achieve a faster convergence rate of $O\left(n^{-8/9}\right)$ for the interior part, when best implemented. Furthermore, the estimators always generate nonnegative density estimates by construction. To implement the MBC estimators, a plug-in smoothing parameter choice method is proposed. Monte Carlo simulations indicate good finite sample performance of the estimators.

1. Introduction

Let $X_1, \ldots, X_n$ be a random sample drawn from a univariate distribution with density f having support $[0, 1]$. In order to estimate the density at a design point $x \in [0, 1]$, we may initially consider a usual kernel density estimator

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where $K(\cdot)$ is a symmetric probability density function having support $\mathbb{R}$ or some finite interval such as $[-1, 1]$, and $h > 0$ is a smoothing parameter called the bandwidth. However, this estimator does not work well for density estimation on the unit interval without boundary correction. Indeed, there is huge literature on correction methods for boundary effects; examples include Müller (1991), Jones (1993), Jones and Foster (1996), Zhang et al. (1999), Hall and Park (2002), and Karunamuni and Alberts (2005), to name a few.

As an alternative device to these boundary correction methods, Chen (1999) proposes a class of so-called asymmetric kernel functions that have the same support as the density to be estimated has. Specifically, he considers the beta kernel density estimator

$$\hat{f}_B(x) = \frac{1}{n}\sum_{i=1}^{n} K_{B(x/b+1,(1-x)/b+1)}(X_i),$$

☆ I would like to thank an Associate Editor and two anonymous referees for helpful comments and suggestions.
* Tel.: +1 815 753 6434; fax: +1 815 752 1019.
E-mail address: mhirukawa@niu.edu.
URL: http://www.math.niu.edu/~hirukawa/.

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.09.017

and the modified beta kernel density estimator

$$\hat{f}_{MB}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{B(\rho_{b,0}(x),\rho_{b,1}(x))}(X_i),$$

where $K_{B(\alpha,\beta)}(u) = u^{\alpha-1}(1-u)^{\beta-1}\mathbf{1}\{0 \le u \le 1\}/B(\alpha,\beta)$ is the density function for the beta distribution with parameters $\alpha, \beta > 0$,

$$\rho_{b,0}(x) = \begin{cases}\rho_b(x) & \text{for } x \in [0, 2b)\\ x/b & \text{for } x \in [2b, 1],\end{cases}$$

$$\rho_{b,1}(x) = \begin{cases}(1-x)/b & \text{for } x \in [0, 1-2b]\\ \rho_b(1-x) & \text{for } x \in (1-2b, 1],\end{cases}$$

$$\rho_b(x) = 2b^2 + \frac{5}{2} - \sqrt{4b^4 + 6b^2 + \frac{9}{4} - x^2 - \frac{x}{b}},$$

and $b > 0$ is a smoothing parameter satisfying $b + nb \to 0$ as $n \to \infty$. Because the beta and modified beta kernels are defined on the unit interval, both $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ are free of boundary bias. Moreover, the shapes of these kernels vary according to the design point, and thus the kernels change the amount of smoothing in a natural way.

Strictly speaking, the beta and modified beta kernels should be referred to as kernel-type weighting functions. As discussed in Gouriéroux and Monfort (2006) and Jones and Henderson (2007a), unlike the cases of symmetric kernels, roles of the data point X and the design point x in these functions are not exchangeable, which causes $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ not to integrate to unity. The phrase 'kernel-type' signifies the fact that they lack a basic property every symmetric kernel possesses. Nevertheless, following the convention in the literature, these kernel-type functions are simply called the beta and modified beta kernels throughout.

Smoothing by beta density weights is originally proposed as an extension to Bernstein polynomial smoothing (Brown and Chen, 1999). After Chen (2000a) apply two beta kernels to nonparametric regression estimation, Bouezmarni and Rolin (2003) demonstrate that $\hat{f}_B(x)$ is consistent even if the true density becomes unbounded at a boundary. As an example of empirical work, Renault and Scaillet (2004) and Hagmann et al. (2005) apply $\hat{f}_B(x)$ for estimating recovery rate densities in credit risk modelling. Gouriéroux and Monfort (2006) point out potential issues of applying $\hat{f}_B(x)$ in recovery rate density estimation, and propose their possible solutions.

Provided that the true density f is twice continuously differentiable, the biases of $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ are $O(b)$ as $b \to 0$. This paper considers the improvements in beta kernel density estimation that reduce the order of magnitude in bias to $O(b^2)$ under sufficient differentiability of f, while the order of magnitude in variance is maintained. In the cases of symmetric kernels, this kind of rate improvements can be typically achieved by employing higher-order kernels; see Jones and Foster (1993) for methods of generating higher-order kernels from a given second-order kernel. To our best knowledge, equivalent techniques are yet to be proposed for asymmetric kernels (including two beta kernels). Instead, this paper applies two classes of multiplicative bias correction (MBC) techniques to attain the rate improvements. The first class of MBC concerned is to construct a multiplicative combination of two density estimators using different smoothing parameters. This idea is originally proposed by Terrell and Scott (1980) as an additive bias correction to the logarithms of densities, and later it is generalized and reinterpreted as an MBC technique by Koshkin (1988) and Jones and Foster (1993), respectively. The second class of MBC in the spirit of Jones et al. (1995) is based on the idea of expressing $f(x) = \hat{f}(x)\left\{f(x)/\hat{f}(x)\right\}$ and estimating the bias-correction term $f(x)/\hat{f}(x)$ nonparametrically. This MBC technique is also applied in nonparametric regression estimation (Linton and Nielsen, 1994), hazard estimation (Nielsen, 1998; Nielsen and Tanggaard, 2001), and spectral matrix estimation (Xiao and Linton, 2002; Hirukawa, 2006). When applied to beta kernel density estimation, both MBC techniques still yield the estimators that are free of boundary bias. In addition, these estimators have a practically appealing property. They always generate nonnegative density estimates everywhere by construction, as $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ do. To implement the two classes of MBC estimators, this paper also proposes a plug-in method of choosing the smoothing parameter b with a beta density used as a reference.

In the related literature, Hagmann and Scaillet (2007) apply a semi-parametric MBC technique in the spirit of Hjort and Jones (1996) (called local multiplicative bias correction, or LMBC) to the beta density estimator. Gustafsson et al. (2009) propose another semi-parametric MBC technique called local transformation bias correction (LTBC), which basically follows the idea of Rudemo (1991). In both LMBC and LTBC, the beta kernel is used in nonparametric bias correction for initial parametric density estimation. A key difference is that the bias correction is made for the original data in LMBC and for the data transformed on the unit interval in LTBC. However, unlike the MBC estimators proposed in this paper, neither LMBC nor LTBC improves the bias in order of magnitude. Furthermore, another class of asymmetric kernels, including the gamma (Chen, 2000b), inverse Gaussian and reciprocal inverse Gaussian (Scaillet, 2004) kernels, has been proposed for density estimation with support $[0, \infty)$, yielding density estimators that are free of boundary bias near the origin. Two MBC techniques considered in this paper can be readily applied for these kernels; research extensions in this direction will be addressed in a separate paper.

The remainder of the paper is organized as follows. Section 2 develops asymptotic properties of two classes of MBC density estimators. Section 3 proposes a plug-in smoothing parameter choice method, and conducts small Monte Carlo simulations to check finite sample properties of MBC estimators. Section 4 applies the MBC estimators for estimating the density of eruption durations of the Old Faithful Geyser in Yellowstone National Park, Wyoming, USA. Section 5 summarizes the main results of the paper. All proofs are given in the Appendix.

This paper adopts the following notational conventions: $B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ signifies the beta function; $\Gamma(a)=\int_0^\infty y^{a-1}\exp(-y)\,dy,\ a>0$ is the gamma function; and $\mathbf{1}\{\cdot\}$ denotes an indicator function. The expression '$X \stackrel{d}{=} Y$' reads "A random variable X obeys the distribution Y". Lastly, the expression '$X_n \sim Y_n$' is used whenever $X_n/Y_n \to 1$ as $n\to\infty$.

2. Beta MBC density estimators

2.1. Definitions

The idea of the geometric estimator by Terrell and Scott (1980, abbreviated as "TS" hereafter) can be readily extended to beta kernel density estimation. For a given kernel $j \in \{B, MB\}$, let $\hat{f}_{j,b}(x)$ and $\hat{f}_{j,b/c}(x)$ denote density estimators using smoothing parameters b and b/c, respectively, where $c\in(0,1)$ is some predetermined constant that does not depend on the design point x. The TS-MBC beta kernel density estimators are defined as

$$\tilde{f}_{\mathrm{TS},j}(x)=\left\{\hat{f}_{j,b}(x)\right\}^{\frac{1}{1-c}}\left\{\hat{f}_{j,b/c}(x)\right\}^{-\frac{c}{1-c}}.$$

The other MBC technique in the spirit of Jones et al. (1995, abbreviated as "JLN" hereafter) utilizes a single smoothing parameter b. In light of the identity $f(x)=\hat{f}_{j,b}(x)\left\{f(x)/\hat{f}_{j,b}(x)\right\}$, the JLN-MBC beta kernel density estimators are defined as

$$\tilde{f}_{\mathrm{JLN},j}(x)=\hat{f}_{j,b}(x)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{K_{(j;x,b)}(X_i)}{\hat{f}_{j,b}(X_i)}\right\},$$

where $K_{(j;x,b)}(\cdot)$ denotes the kernel j. Recognize that the term inside the brackets is a natural estimator of the bias-correction term $f(x)/\hat{f}_{j,b}(x)$. Also, by construction, both $\tilde{f}_{\mathrm{TS},j}(x)$ and $\tilde{f}_{\mathrm{JLN},j}(x)$ always generate nonnegative density estimates everywhere.

2.2. Asymptotic properties

To approximate the bias and variance of each MBC estimator, the following regularity conditions are assumed:

(A1) For a given $x\in[0,1]$, $f(x)>0$ and $f''''(x)$ is continuous and bounded in the neighborhood of x.
(A2) The smoothing parameter $b=b_n$ satisfies $b\to 0$ and $nb^3\to\infty$ as $n\to\infty$.

The smoothness condition for the true density f in (A1) is standard for consistency of density estimators using fourth-order kernels, whereas the condition that f is bounded away from zero is required for MBC. The condition (A2) implies that the convergence rate of the smoothing parameter b is slower than $O\left(n^{-1/3}\right)$. This condition is required to control the order of magnitude of remainder terms in the approximation to the bias of each MBC estimator. It will be shown shortly that the mean squared error (MSE)-optimal smoothing parameter is $b^*=O\left(n^{-2/9}\right)$ if the design point x satisfies $x/b\to\infty$ and $(1-x)/b\to\infty$, and $b^\star=O\left(n^{-1/5}\right)$ if $x/b\to\kappa$ or $(1-x)/b\to\kappa$ for some $\kappa>0$; these convergence rates are indeed within the required range.

We also call the position of x as "interior x" if $x/b\to\infty$ and $(1-x)/b\to\infty$, and "boundary x" if $x/b\to\kappa$ or $(1-x)/b\to\kappa$. The following two theorems present the approximations to bias and variance terms of two MBC estimators, whose proofs are given in the Appendix.

Theorem 1. *Suppose that the conditions* (A1) *and* (A2) *hold. Then, the biases of $\tilde{f}_{\mathrm{TS},B}(x)$ and $\tilde{f}_{\mathrm{TS},MB}(x)$ can be approximated by*

$$\begin{aligned}
bias\left\{\tilde{f}_{\mathrm{TS},B}(x)\right\} &\sim \frac{1}{c(1-c)}p_B(x)\,b^2\\
&=\frac{1}{c(1-c)}\left[\frac{1}{2}\frac{\left\{(1-2x)f'(x)+\frac{1}{2}x(1-x)f''(x)\right\}^2}{f(x)}-\left\{-2(1-2x)f'(x)+\frac{1}{2}\left(11x^2-11x+2\right)f''(x)\right.\right.\\
&\quad\left.\left.+\frac{5}{6}x(1-x)(1-2x)f'''(x)+\frac{1}{8}x^2(1-x)^2f''''(x)\right\}\right]b^2,
\end{aligned}$$

$$bias\left\{\tilde{f}_{\mathrm{TS},MB}(x)\right\} \sim \frac{1}{c(1-c)} \times \begin{cases} p_{MB}(x)\,b^2 & \text{for } x \in [2b, 1-2b] \\ p_{MB,0}(x)\,b^2 & \text{for } x \in [0, 2b) \\ p_{MB,1}(x)\,b^2 & \text{for } x \in (1-2b, 1] \end{cases}$$

$$= \frac{1}{c(1-c)} \times \begin{cases} \left[\frac{1}{8}\frac{x^2(1-x)^2\{f''(x)\}^2}{f(x)} - \left\{-\frac{1}{2}x(1-x)f''(x) \right.\right. \\ \left.\left. \quad + \frac{1}{3}x(1-x)(1-2x)f'''(x) + \frac{1}{8}x^2(1-x)^2 f''''(x)\right\}\right] b^2 & \text{for } x \in [2b, 1-2b] \\ \frac{1}{2}\left[\frac{\xi_b^2(x)\{f'(x)\}^2}{f(x)} - \left\{\xi_b^2(x) + \xi_b(x) + \frac{x}{b}\right\} f''(x)\right] b^2 & \text{for } x \in [0, 2b) \\ \frac{1}{2}\left[\frac{\xi_b^2(1-x)\{f'(x)\}^2}{f(x)} - \left\{\xi_b^2(1-x) + \xi_b(1-x) + \frac{1-x}{b}\right\} f''(x)\right] b^2 & \text{for } x \in (1-2b, 1], \end{cases}$$

where

$$\xi_b(x) = \frac{(1-x)\{\rho_b(x) - x/b\}}{1 + b\{\rho_b(x) - x/b\}} = O(1).$$

The variance of $\tilde{f}_{\mathrm{TS},B}(x)$ *can be approximated by*

$$var\left\{\tilde{f}_{\mathrm{TS},B}(x)\right\} \sim \begin{cases} \frac{\lambda(c)}{2\sqrt{\pi}\sqrt{x(1-x)}} n^{-1} b^{-1/2} f(x) & \text{for interior } x \\ \frac{1}{(1-c)^2}\left\{\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} - \frac{2c}{(1+c)^\kappa}\left(\frac{c}{1+c}\right)^{c\kappa+1}\frac{\Gamma((1+c)\kappa+1)}{\Gamma(\kappa+1)\Gamma(c\kappa+1)} \right. \\ \left. \quad + c^3\frac{\Gamma(2c\kappa+1)}{2^{2c\kappa+1}\Gamma^2(c\kappa+1)}\right\} n^{-1} b^{-1} f(x) & \text{for boundary } x, \end{cases}$$

where

$$\lambda(c) = \frac{\left(1 + c^{5/2}\right)(1+c)^{1/2} - 2\sqrt{2}c^{3/2}}{(1+c)^{1/2}(1-c)^2}.$$

Theorem 2. *Suppose that the conditions* (A1) *and* (A2) *hold. Then, the biases of* $\tilde{f}_{\mathrm{JLN},B}(x)$ *and* $\tilde{f}_{\mathrm{JLN},MB}(x)$ *can be approximated by*

$$\begin{aligned} bias\left\{\tilde{f}_{\mathrm{JLN},B}(x)\right\} &\sim q_B(x)\,b^2 \\ &= -f(x)\left[(1-2x)\left\{\frac{(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x)}{f(x)}\right\}' \right. \\ &\quad \left. + \frac{1}{2}x(1-2x)\left\{\frac{(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x)}{f(x)}\right\}''\right] b^2, \end{aligned}$$

$$bias\left\{\tilde{f}_{\mathrm{JLN},MB}(x)\right\} \sim \begin{cases} q_{MB}(x)\,b^2 & \text{for } x \in [2b, 1-2b] \\ q_{MB,0}(x)\,b^2 & \text{for } x \in [0, 2b) \\ q_{MB,1}(x)\,b^2 & \text{for } x \in (1-2b, 1] \end{cases}$$

$$= \begin{cases} -\frac{1}{4}x(1-x)f(x)\left\{\frac{x(1-x)f''(x)}{f(x)}\right\}'' b^2 & \text{for } x \in [2b, 1-2b] \\ -\xi_b(x)f(x)\left\{\frac{\xi_b(x)f'(x)}{f(x)}\right\}' b^2 & \text{for } x \in [0, 2b) \\ -\xi_b(1-x)f(x)\left\{\frac{\xi_b(1-x)f'(x)}{f(x)}\right\}' b^2 & \text{for } x \in (1-2b, 1], \end{cases}$$

where $\xi_b(x)$ *is defined in Theorem 1. The variance of* $\tilde{f}_{\mathrm{JLN},B}(x)$ *can be approximated by*

$$var\left\{\tilde{f}_{\mathrm{JLN},B}(x)\right\} \sim \begin{cases} \frac{1}{2\sqrt{\pi}\sqrt{x(1-x)}} n^{-1} b^{-1/2} f(x) & \text{for interior } x \\ \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} n^{-1} b^{-1} f(x) & \text{for boundary } x. \end{cases}$$

Asymptotic variances of $\tilde{f}_{TS,MB}(x)$ and $\tilde{f}_{JLN,MB}(x)$ are similar to those of $\tilde{f}_{TS,B}(x)$ and $\tilde{f}_{JLN,B}(x)$, except that the multipliers in front of $n^{-1}b^{-1}$ for boundary x have slightly different forms. Two theorems demonstrate that both TS- and JLN-MBC estimators are free of boundary bias. More importantly, these two MBC estimators reduce the order of magnitude in bias from $O(b)$ to $O(b^2)$, while their variances are still $O(n^{-1}b^{-1/2})$ for interior x and $O(n^{-1}b^{-1})$ for boundary x. We can also see that, for each of two MBC techniques, the modified beta kernel estimator eliminates the term involving $f'(x)$ from the bias over the interior region. Although $f'(x)$ appears in the bias only in two small regions near the boundaries, it is compensated by disappearance of higher-order derivatives.

The variance of JLN-MBC estimators is first-order asymptotically equivalent to that of bias-uncorrected estimators $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ for interior x. While the beta LMBC density estimator in Hagmann and Scaillet (2007, Proposition 2) also yields the same leading variance term, it does not reduce the bias in order of magnitude. In contrast, when JLN-MBC is applied for density estimation using a symmetric second-order kernel, the leading variance term tends to be larger (although not inflated in order of magnitude) because the multiplier in the variance term involves the roughness (or squared integral) of the 'twiced' kernel (Stuetzle and Mittal, 1979) that is generated by the original kernel. Moreover, due to the multiplier $\lambda(c)$, which is increasing in $c \in (0, 1)$ and takes the value from 1 to 27/16, the asymptotic variance of TS-MBC estimators tends to be larger than those of $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ for interior x.

For interior x, MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ can be approximated by

$$MSE\left\{\tilde{f}_{TS,j}(x)\right\} = \frac{p_j^2(x)}{c^2(1-c)^2}b^4 + \frac{n^{-1}b^{-1/2}\lambda(c)f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}} + o\left(b^4 + n^{-1}b^{-1/2}\right),$$

$$MSE\left\{\tilde{f}_{JLN,j}(x)\right\} = q_j^2(x)\,b^4 + \frac{n^{-1}b^{-1/2}f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}} + o\left(b^4 + n^{-1}b^{-1/2}\right).$$

MSE-optimal smoothing parameters are

$$b^*_{TS,j} = \left\{c^2(1-c)^2\lambda(c)\right\}^{2/9}\left\{\frac{f(x)}{16\sqrt{\pi}\sqrt{x(1-x)}p_j^2(x)}\right\}^{2/9} n^{-2/9},$$

$$b^*_{JLN,j} = \left\{\frac{f(x)}{16\sqrt{\pi}\sqrt{x(1-x)}q_j^2(x)}\right\}^{2/9} n^{-2/9},$$

which yield the optimal MSEs

$$MSE^*\left\{\tilde{f}_{TS,j}(x)\right\} \sim \frac{9}{8^{8/9}}\gamma(c)\,p_j^{2/9}(x)\left\{\frac{f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}}\right\}^{8/9} n^{-8/9},$$

$$MSE^*\left\{\tilde{f}_{JLN,j}(x)\right\} \sim \frac{9}{8^{8/9}}q_j^{2/9}(x)\left\{\frac{f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}}\right\}^{8/9} n^{-8/9},$$

where

$$\gamma(c) = \left\{\frac{\left(1+c^{5/2}\right)(1+c)^{1/2} - 2\sqrt{2}c^{3/2}}{c^{1/4}(1+c)^{1/2}(1-c)^{9/4}}\right\}^{8/9}.$$

Observe that the MSE-optimal smoothing parameters are $O\left(n^{-2/9}\right) = O\left(h^{*2}\right)$, where h^* is the MSE-optimal bandwidth for fourth-order kernel estimators, or TS- or JLN-MBC estimators using symmetric second-order kernels. As a result, the optimal MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ become $O\left(n^{-8/9}\right)$, as with MBC estimation using the second-order kernels. This convergence rate is faster than $O\left(n^{-4/5}\right)$, the optimal convergence rate in the MSEs of $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ for interior x. On the other hand, for boundary x, the MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ are $O\left(b^4 + n^{-1}b^{-1}\right)$. The MSE-optimal smoothing parameter $b^\star = O\left(n^{-1/5}\right)$ achieves the optimal MSE of $O\left(n^{-4/5}\right)$, which is also faster than $O\left(n^{-2/3}\right)$, the optimal convergence rate in the MSEs of $\hat{f}_B(x)$ and $\hat{f}_{MB}(x)$ for boundary x.

The undesirable convergence rate $O\left(n^{-1}b^{-1}\right)$ over boundary regions does not affect the global properties of these estimators. By the trimming argument in Chen (1999), mean integrated squared errors (MISEs) of the MBC estimators are

$$MISE\left\{\tilde{f}_{TS,j}(x)\right\} = \frac{b^4}{c^2(1-c)^2}\int_0^1 p_j^2(x)\,dx + \frac{n^{-1}b^{-1/2}\lambda(c)}{2\sqrt{\pi}}\int_0^1 \frac{f(x)}{\sqrt{x(1-x)}}dx + o\left(b^4 + n^{-1}b^{-1/2}\right),$$

$$MISE\left\{\tilde{f}_{JLN,j}(x)\right\} = b^4\int_0^1 q_j^2(x)\,dx + \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi}}\int_0^1 \frac{f(x)}{\sqrt{x(1-x)}}dx + o\left(b^4 + n^{-1}b^{-1/2}\right),$$

provided that all integrals are finite. MISE-optimal smoothing parameters are given by

$$b_{\mathrm{TS},j}^{**} = \left\{c^2 (1-c)^2 \lambda (c)\right\}^{2/9} \left\{ \frac{\int_0^1 f(x)/\sqrt{x(1-x)}\mathrm{d}x}{16\sqrt{\pi}\int_0^1 p_j^2(x)\,\mathrm{d}x} \right\}^{2/9} n^{-2/9},$$

$$b_{\mathrm{JLN},j}^{**} = \left\{ \frac{\int_0^1 f(x)/\sqrt{x(1-x)}\mathrm{d}x}{16\sqrt{\pi}\int_0^1 q_j^2(x)\,\mathrm{d}x} \right\}^{2/9} n^{-2/9}.$$

Therefore, the optimal MISEs become

$$MISE^{**}\left\{\tilde{f}_{\mathrm{TS},j}(x)\right\} \sim \frac{9}{8^{8/9}}\gamma(c)\left\{\int_0^1 p_j^2(x)\,\mathrm{d}x\right\}^{1/9}\left\{\frac{1}{2\sqrt{\pi}}\int_0^1 \frac{f(x)}{\sqrt{x(1-x)}}\mathrm{d}x\right\}^{8/9} n^{-8/9},$$

$$MISE^{**}\left\{\tilde{f}_{\mathrm{JLN},j}(x)\right\} \sim \frac{9}{8^{8/9}}\left\{\int_0^1 q_j^2(x)\,\mathrm{d}x\right\}^{1/9}\left\{\frac{1}{2\sqrt{\pi}}\int_0^1 \frac{f(x)}{\sqrt{x(1-x)}}\mathrm{d}x\right\}^{8/9} n^{-8/9}.$$

Furthermore, the multiplier $\gamma(c)$ in the optimal MISE for the TS-MBC estimator is minimized at $c^* \approx 0.2636$; this optimal value is considered throughout.

2.3. Normalization

Neither $\tilde{f}_{\mathrm{TS},j}(x)$ nor $\tilde{f}_{\mathrm{JLN},j}(x)$ integrates to one. In general, MBC leads to lack of normalization, even if symmetric second-order kernels are employed; see Section 2.2 in JLN, for example. While lack of normalization in $\tilde{f}_{\mathrm{TS},j}(x)$ and $\tilde{f}_{\mathrm{JLN},j}(x)$ may be ignored in some applications, Gouriéroux and Monfort (2006), for example, argue that this issue should be resolved for the estimation of recovery rate densities in credit risk modelling, and propose two renormalized beta kernel density estimators. Taking the structures of $\tilde{f}_{\mathrm{TS},j}(x)$ and $\tilde{f}_{\mathrm{JLN},j}(x)$ into account, we adopt their "macro" approach to obtain the renormalized MBC estimators

$$\tilde{f}_{\mathrm{TS},j}^{R}(x) = \frac{\tilde{f}_{\mathrm{TS},j}(x)}{\int_0^1 \tilde{f}_{\mathrm{TS},j}(x)\,\mathrm{d}x},$$

$$\tilde{f}_{\mathrm{JLN},j}^{R}(x) = \frac{\tilde{f}_{\mathrm{JLN},j}(x)}{\int_0^1 \tilde{f}_{\mathrm{JLN},j}(x)\,\mathrm{d}x}.$$

Since

$$E\left\{\int_0^1 \tilde{f}_{\mathrm{TS},j}(x)\,\mathrm{d}x\right\} = 1 + \frac{b^2}{c(1-c)}\int_0^1 p_j(x)\,\mathrm{d}x + o\left(b^2\right),$$

$$E\left\{\int_0^1 \tilde{f}_{\mathrm{JLN},j}(x)\,\mathrm{d}x\right\} = 1 + b^2\int_0^1 q_j(x)\,\mathrm{d}x + o\left(b^2\right),$$

provided that $p_j(x)$ and $q_j(x)$ are integrable, biases of $\tilde{f}_{\mathrm{TS},j}^{R}(x)$ and $\tilde{f}_{\mathrm{JLN},j}^{R}(x)$ can be approximated by

$$bias\left\{\tilde{f}_{\mathrm{TS},j}^{R}(x)\right\} \sim \frac{1}{c(1-c)}\left\{p_j(x) - \int_0^1 p_j(x)\,\mathrm{d}x\right\} b^2,$$

$$bias\left\{\tilde{f}_{\mathrm{JLN},j}^{R}(x)\right\} \sim \left\{q_j(x) - \int_0^1 q_j(x)\,\mathrm{d}x\right\} b^2.$$

Their asymptotic variances are unaffected.

2.4. Further bias reduction

In principle, further bias reduction is possible after the regularity conditions are properly strengthened. Constructing a multiplicative combination of $(s+1)$ different density estimators, we can generalize the TS-MBC estimator $\tilde{f}_{\mathrm{TS},j}(x)$ as

$$\tilde{f}_{\mathrm{TS},j}^{(s)}(x) = \prod_{r=0}^{s}\left\{\hat{f}_{j,b/c_r}(x)\right\}^{\alpha_{s,r}},$$

where $c_0 = 1, c_1, \ldots, c_s \in (0,1)$ are mutually different constants, and the exponent is

$$\alpha_{s,r} = \frac{(-1)^s c_r^s}{\prod_{p=0,p\neq r}^{s}\left(c_p - c_r\right)}.$$

Similarly, the sth iterated JLN-MBC estimator can be defined as

$$\tilde{f}^{(s)}_{\text{JLN},j}(x) = \tilde{f}^{(s-1)}_{\text{JLN},j}(x)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{K_{(j;x,b)}(X_i)}{\tilde{f}^{(s-1)}_{\text{JLN},j}(X_i)}\right\},$$

where $\tilde{f}^{(0)}_{\text{JLN},j}(x) = \hat{f}_{j,b}(x)$. Provided that the true density f is $2(s+1)$ times continuously differentiable, for each of these estimators, it can be shown that the order of magnitude in bias is $O\left(b^{s+1}\right)$ while the variance remains $O\left(n^{-1}b^{-1/2}\right)$ and $O\left(n^{-1}b^{-1}\right)$ for interior and boundary x, respectively. In particular, $var\left\{\tilde{f}^{(s)}_{\text{JLN},j}(x)\right\}$ is first-order asymptotically equivalent to $var\left\{\tilde{f}^{(0)}_{\text{JLN},j}(x)\right\} = var\left\{\hat{f}_{j,b}(x)\right\}$ for interior x. Their optimal MSEs are $O\left(n^{-(4s+4)/(4s+5)}\right)$ and $O\left(n^{-(2s+2)/(2s+3)}\right)$ for interior and boundary x, respectively. Accordingly, as the number of iterations increases, global convergence rates of the iterated MBC estimators when best implemented can be arbitrarily close to the parametric one. However, it is doubtful whether there is much gain in practice from these estimators, and thus we do not pursue this issue any further.

3. Finite sample performance

3.1. Smoothing parameter selection

Choosing the smoothing parameter b is an important practical issue. In the spirit of Jones and Henderson (2007b), we propose a plug-in method with a beta density used as a reference. The plug-in smoothing parameters for $\tilde{f}_{\text{TS},MB}(x)$ and $\tilde{f}_{\text{JLN},MB}(x)$ (called "beta-referenced smoothing parameters" hereafter) are defined as the minimizers of asymptotic weighted mean integrated squared errors (AWMISEs)

$$\begin{aligned}
\hat{b}_{\text{TS}.MB} &= \arg\min_b \text{AWMISE}\left\{\tilde{f}_{\text{TS},MB}(x)\right\}\\
&= \arg\min_b \frac{b^4}{c^2(1-c)^2}\int_0^1 \tilde{p}^2_{MB}(x)\,w(x)\,\mathrm{d}x + \frac{n^{-1}b^{-1/2}\lambda(c)}{2\sqrt{\pi}}\int_0^1\frac{g(x)}{\sqrt{x(1-x)}}w(x)\,\mathrm{d}x,\\
\hat{b}_{\text{JLN},MB} &= \arg\min_b \text{AWMISE}\left\{\tilde{f}_{\text{JLN},MB}(x)\right\}\\
&= \arg\min_b b^4\int_0^1 \tilde{q}^2_{MB}(x)\,w(x)\,\mathrm{d}x + \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi}}\int_0^1\frac{g(x)}{\sqrt{x(1-x)}}w(x)\,\mathrm{d}x,
\end{aligned}$$

where $g(x) = x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}\{0\le x\le 1\}/B(\alpha,\beta)$ is the density function for the beta distribution with parameters (α,β), $\tilde{p}_{MB}(x)$ and $\tilde{q}_{MB}(x)$ can be obtained by replacing $f(x)$ with $g(x)$ in $p_{MB}(x)$ and $q_{MB}(x)$, respectively, and $w(x)$ is a weighting function that ensures finiteness of integrals. Because of the expressions of $\tilde{p}_{MB}(x)$ and $\tilde{q}_{MB}(x)$, we choose $w(x) = x^5(1-x)^5$. In practice, the parameters (α,β) are replaced by their method of moments estimates $\left(\hat{\alpha},\hat{\beta}\right)$. Analytical expressions of $\hat{b}_{\text{TS},MB}$ and $\hat{b}_{\text{JLN},MB}$, as well as the beta-referenced smoothing parameter for $\hat{f}_{MB}(x)$ in Jones and Henderson (2007b), denoted as $\hat{b}_{MB}$, are given in the Appendix.

We do not pursue the beta-referenced smoothing parameter for $\tilde{f}_{\text{TS},B}(x)$ or $\tilde{f}_{\text{JLN},B}(x)$; since extra terms are involved in $p_B(x)$ and $q_B(x)$, the minimizers of their AWMISEs take much more complicated forms than $\hat{b}_{\text{TS},MB}$ and $\hat{b}_{\text{JLN},MB}$. From the viewpoint of practical relevance, these smoothing parameters are simply employed for $\tilde{f}_{\text{TS},B}(x)$ and $\tilde{f}_{JLN,B}(x)$ in the Monte Carlo simulations below. Similarly, $\hat{b}_{MB}$ is chosen as the smoothing parameter for $\hat{f}_B(x)$. Strictly speaking, $\hat{b}_{\text{TS},MB}$ and $\hat{b}_{\text{JLN},MB}$ are not AWMISE-optimal choices for $\tilde{f}_{\text{TS},B}(x)$ and $\tilde{f}_{\text{JLN},B}(x)$. However, when $\hat{b}_{\text{TS},MB}$ and $\hat{b}_{\text{JLN},MB}$ are chosen, optimal convergence rates in MISEs for these estimators can be still achieved, and the Monte Carlo simulations indicate that they often outperform $\tilde{f}_{\text{TS},MB}(x)$ and $\tilde{f}_{\text{JLN},MB}(x)$.

3.2. Monte Carlo setup

Monte Carlo simulations compare the finite sample performance of the following 7 classes of estimators, or 14 estimators in total: (i) bias-uncorrected estimator $\hat{f}_j(x)$; (ii) "micro" estimator $\hat{f}^r_j(x)$ in Gouriéroux and Monfort (2006); (iii) "macro" (or renormalized) estimator $\hat{f}^R_j(x)$ in Gouriéroux and Monfort (2006); (iv) TS-MBC estimator $\tilde{f}_{\text{TS},j}(x)$; (v) renormalized TS-MBC estimator $\tilde{f}^R_{\text{TS},j}(x)$; (vi) JLN-MBC estimator $\tilde{f}_{\text{JLN},j}(x)$; and (vii) renormalized JLN-MBC estimator $\tilde{f}^R_{\text{JLN},j}(x)$, where "micro" and "macro" estimators are defined as

$$\hat{f}^r_j(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_{(j;x,b)}(X_i)}{\int_0^1 K_{(j;x,b)}(X_i)\,\mathrm{d}x},$$

Table 1
True distributions considered in Monte Carlo simulations.

	Distribution	Density function $f(x)$, $x \in [0, 1]$	Shape
1.	$U[0, 1]$	1	Horizontal
2.	$B(3, 1)$	$3x^2$	Increasing
3.	Truncated Exp (1/2)	$2\exp(-2x) / \{1 - \exp(-2)\}$	Decreasing
4.	Truncated $N(0, 1/4)$	$2\phi(2x) / \{\Phi(2) - 1/2\}$	Decreasing
5.	$B(3, 3)$	$30x^2(1-x)^2$	Bell-shaped
6.	$B(1/2, 1/2)$	$1/\left\{\pi\sqrt{x(1-x)}\right\}$	U-shaped
7.	$(1/2)B(3, 1) + (1/2)B(1, 3)$	$(3/2)\left\{x^2 + (1-x)^2\right\}$	U-shaped
8.	$B(2, 4)$	$20x(1-x)^3$	Unimodal, right-skewed
9.	$(1/2)B(5, 2) + (1/2)B(2, 5)$	$15\left\{x^4(1-x) + x(1-x)^4\right\}$	Bimodal, symmetric
10.	$(1/4)B(8, 2) + (3/4)B(2, 8)$	$18\left\{x^7(1-x) + 3x(1-x)^7\right\}$	Bimodal, right-skewed

$$\hat{f}_j^R(x) = \frac{\hat{f}_j(x)}{\int_0^1 \hat{f}_j(x)\,dx},$$

respectively. The value of the constant c in each TS-MBC estimator is set equal to the MISE-optimal $c^* = 0.2636$. Ten true distributions are considered, as listed in Table 1, and for each distribution, 1000 data sets of sample size $n = 100$ or $n = 200$ are simulated. For each simulated data set, the smoothing parameter is chosen via the following two methods: (i) a simple "rule-of-thumb" method (see Renault and Scaillet, 2004, and Gouriéroux and Monfort, 2006) such that $\hat{b}_{plain} = \hat{\sigma}n^{-2/5}$ and $\hat{b}_{mbc} = \hat{\sigma}n^{-2/9}$ for all bias-uncorrected and MBC estimators, respectively, where $\hat{\sigma}$ is the sample standard deviation; and (ii) the beta-referenced method described in the previous section, where $\hat{b}_{MB}$, $\hat{b}_{\mathrm{TS},MB}$ and $\hat{b}_{\mathrm{JLN},MB}$ are applied for all bias-uncorrected, TS- and JLN-MBC estimators, respectively. The purposes of using two plug-in smoothing parameters for a given density estimator are to speed up computation and to see how the beta-referenced method can improve the performance of each estimator in comparison with the rule-of-thumb method. Since the beta-referenced method often picks up a large value, estimated smoothing parameter values are trimmed at 1 so that their lengths may not exceed that of the support. All density estimates are evaluated on an equally spaced grid of 999 points $\{0.001, 0.002, \ldots, 0.999\}$. For each estimator $\bar{f}$,

$$ISE(\bar{f}) = \int_0^1 \{\bar{f}(x) - f(x)\}^2\,dx \approx \frac{1}{1000}\sum_{j=1}^{999}\left\{\bar{f}\left(\frac{j}{1000}\right) - f\left(\frac{j}{1000}\right)\right\}^2$$

is computed over the grids, and then it is averaged over 1000 replications. Averages of integrated squared biases and standard errors (defined as square roots of the estimates of asymptotic integrated variances) are also computed for reference.

3.3. Simulation results

Monte Carlo results, including descriptive statistics of smoothing parameters, are given in Tables 2–5. Observe that the results are qualitatively similar across two sample sizes. Influences of trimming for smoothing parameters are limited in the sense that there are very few numbers of trimmed smoothing parameter estimates. Frequencies of trimming appear to be large for Distribution 1 ($=U[0, 1]$). However, the bias-uncorrected and its renormalized estimators have no asymptotic biases (and thus bias correction is unnecessary). For this distribution, simply setting $b = O(1)$ (i.e. oversmoothing) is optimal, and numbers of trimmed smoothing parameters are not an issue.

Comparing TS- and JLN-MBC estimators with their corresponding bias-uncorrected estimators, we find that JLN-MBC estimators often improve the average ISE, whereas results are mixed for TS-MBC estimators. Even in the absence of substantial improvement in the average ISE, $\tilde{f}_{\mathrm{JLN},B}(x)$ tends to reduce the average integrated squared bias from $\hat{f}_B(x)$. In addition, as suggested by the asymptotic results, TS-MBC estimators are less efficient than JLN-MBC estimators in that standard errors of the former are larger due to the extra constant $\lambda(c) \geq 1$. Superior performance of JLN-MBC over TS-MBC is consistent with the simulation results reported in Jones and Signorini (1997).

However, the results do not necessarily support superiority of the modified beta kernel over the beta kernel. In addition, the effect of normalization appears to be marginal in many cases. This is due to the fact that renormalization induces an extra bias term. Since the sign of this term depends on the true density, it may or may not lead to a less biased estimate.

Comparing two smoothing parameter choice methods for a given estimator, we also find that for JLN-MBC estimators, the beta-referenced method works better for Distributions 1, 2, 3, 4, 6, and 7. For other distributions, the rule-of-thumb method tends to perform better, but there is not much gain in the average ISE unless the true density is bimodal (e.g. Distributions 9 and 10). In short, as far as the shape of the true density does not deviate considerably from that of a beta distribution, the beta-referenced method works well for JLN-MBC estimators. In contrast, the rule-of-thumb method tends to perform better for TS-MBC estimators. This finding is revisited shortly, in relation to poor performance of $\tilde{f}_{\mathrm{TS},MB}(x)$. The results are mixed for bias-uncorrected estimators.

Fig. 1 is prepared to visualize how two classes of MBC techniques work. Each panel displays average plots of density estimates by bias-uncorrected beta ($\hat{f}_B(x)$, denoted "Plain" in the legend), beta TS-MBC ($\hat{f}_{\mathrm{TS},B}(x)$), and beta JLN-MBC

Table 2
Average ISE computed on 1000 replications ($n = 100$; Distributions 1–5).

	True distribution									
	1		2		3		4		5	
	ROT	BR	ROT	BR	ROT	BR	ROT	BR	ROT	BR
Density estimator:										
	0.0283	0.0072	0.0382	0.0408	0.0321	0.0416	0.0298	0.0382	0.0319	0.0338
$\hat{f}_B$	(0.0000)	(0.0000)	(0.0030)	(0.0202)	(0.0027)	(0.0256)	(0.0009)	(0.0281)	(0.0073)	(0.0112)
	[0.2006]	[0.1217]	[0.2316]	[0.1766]	[0.2103]	[0.1456]	[0.2073]	[0.1299]	[0.1961]	[0.1865]
	0.0308	0.0057	0.0410	0.0324	0.0335	0.0451	0.0310	0.0494	0.0255	0.0249
$\hat{f}_{MB}$	(0.0000)	(0.0000)	(0.0009)	(0.0093)	(0.0010)	(0.0207)	(0.0001)	(0.0337)	(0.0033)	(0.0054)
	[0.2004]	[0.1217]	[0.2342]	[0.1796]	[0.2122]	[0.1472]	[0.2069]	[0.1297]	[0.1892]	[0.1787]
	0.0442	0.0187	0.0659	0.0484	0.0516	0.0369	0.0472	0.0396	0.0328	0.0362
$\hat{f}_B^r$	(0.0085)	(0.0072)	(0.0195)	(0.0224)	(0.0135)	(0.0129)	(0.0118)	(0.0260)	(0.0094)	(0.0144)
	[0.2073]	[0.1248]	[0.2417]	[0.1871]	[0.2195]	[0.1533]	[0.2137]	[0.1333]	[0.1935]	[0.1834]
	0.0322	0.0068	0.0469	0.0393	0.0384	0.0368	0.0332	0.0436	0.0239	0.0213
$\hat{f}_{MB}^r$	(0.0031)	(0.0002)	(0.0095)	(0.0197)	(0.0077)	(0.0191)	(0.0038)	(0.0289)	(0.0009)	(0.0015)
	[0.1986]	[0.1222]	[0.2305]	[0.1776]	[0.2093]	[0.1473]	[0.2051]	[0.1302]	[0.1915]	[0.1813]
	0.0282	0.0068	0.0384	0.0359	0.0323	0.0394	0.0297	0.0383	0.0324	0.0352
$\hat{f}_B^R$	(0.0000)	(0.0000)	(0.0017)	(0.0137)	(0.0021)	(0.0219)	(0.0009)	(0.0284)	(0.0091)	(0.0137)
	[0.2007]	[0.1217]	[0.2333]	[0.1792]	[0.2116]	[0.1475]	[0.2073]	[0.1297]	[0.1933]	[0.1832]
	0.0308	0.0056	0.0395	0.0317	0.0326	0.0446	0.0309	0.0494	0.0246	0.0229
$\hat{f}_{MB}^R$	(0.0000)	(0.0000)	(0.0010)	(0.0105)	(0.0011)	(0.0212)	(0.0001)	(0.0337)	(0.0012)	(0.0022)
	[0.2006]	[0.1217]	[0.2331]	[0.1786]	[0.2115]	[0.1470]	[0.2071]	[0.1297]	[0.1918]	[0.1816]
	0.0218	0.0087	0.0287	0.0325	0.0250	0.0283	0.0230	0.0306	0.0315	0.0428
$\tilde{f}_{\mathrm{TS},B}$	(0.0000)	(0.0000)	(0.0016)	(0.0060)	(0.0023)	(0.0097)	(0.0007)	(0.0184)	(0.0126)	(0.0064)
	[0.1852]	[0.1363]	[0.2153]	[0.1994]	[0.1943]	[0.1654]	[0.1922]	[0.1414]	[0.1850]	[0.2050]
	0.0250	0.0068	0.0428	0.0420	0.0335	0.0369	0.0266	0.0437	0.0206	0.0332
$\tilde{f}_{\mathrm{TS},MB}$	(0.0000)	(0.0001)	(0.0097)	(0.0112)	(0.0069)	(0.0057)	(0.0017)	(0.0210)	(0.0042)	(0.0022)
	[0.1851]	[0.1366]	[0.2209]	[0.2054]	[0.1981]	[0.1689]	[0.1925]	[0.1423]	[0.1749]	[0.1981]
	0.0216	0.0083	0.0288	0.0306	0.0249	0.0278	0.0226	0.0314	0.0326	0.0439
$\tilde{f}_{\mathrm{TS},B}^R$	(0.0000)	(0.0000)	(0.0014)	(0.0048)	(0.0018)	(0.0082)	(0.0008)	(0.0192)	(0.0157)	(0.0079)
	[0.1852]	[0.1363]	[0.2156]	[0.2002]	[0.1953]	[0.1666]	[0.1915]	[0.1406]	[0.1800]	[0.2011]
	0.0248	0.0066	0.0304	0.0305	0.0274	0.0323	0.0244	0.0436	0.0192	0.0310
$\tilde{f}_{\mathrm{TS},MB}^R$	(0.0001)	(0.0000)	(0.0031)	(0.0061)	(0.0043)	(0.0057)	(0.0012)	(0.0226)	(0.0021)	(0.0014)
	[0.1851]	[0.1363]	[0.2155]	[0.1998]	[0.1951]	[0.1663]	[0.1915]	[0.1409]	[0.1779]	[0.1997]
	0.0269	0.0144	0.0330	0.0234	0.0284	0.0215	0.0270	0.0167	0.0233	0.0270
$\tilde{f}_{\mathrm{JLN},B}$	(0.0001)	(0.0000)	(0.0005)	(0.0043)	(0.0007)	(0.0036)	(0.0001)	(0.0041)	(0.0028)	(0.0050)
	[0.1630]	[0.1269]	[0.1897]	[0.1560]	[0.1718]	[0.1432]	[0.1686]	[0.1296]	[0.1561]	[0.1502]
	0.0257	0.0116	0.0320	0.0238	0.0274	0.0227	0.0258	0.0168	0.0228	0.0268
$\tilde{f}_{\mathrm{JLN},MB}$	(0.0001)	(0.0000)	(0.0017)	(0.0084)	(0.0018)	(0.0059)	(0.0002)	(0.0057)	(0.0037)	(0.0066)
	[0.1630]	[0.1270]	[0.1892]	[0.1552]	[0.1717]	[0.1429]	[0.1682]	[0.1291]	[0.1527]	[0.1465]
	0.0272	0.0144	0.0341	0.0215	0.0289	0.0212	0.0274	0.0167	0.0235	0.0276
$\tilde{f}_{\mathrm{JLN},B}^R$	(0.0000)	(0.0000)	(0.0001)	(0.0004)	(0.0002)	(0.0018)	(0.0001)	(0.0035)	(0.0023)	(0.0045)
	[0.1635]	[0.1271]	[0.1915]	[0.1592]	[0.1731]	[0.1450]	[0.1692]	[0.1303]	[0.1567]	[0.1505]
	0.0260	0.0117	0.0326	0.0223	0.0276	0.0220	0.0262	0.0159	0.0235	0.0258
$\tilde{f}_{\mathrm{JLN},MB}^R$	(0.0001)	(0.0000)	(0.0017)	(0.0063)	(0.0017)	(0.0048)	(0.0002)	(0.0043)	(0.0021)	(0.0036)
	[0.1634]	[0.1272]	[0.1905]	[0.1576]	[0.1723]	[0.1441]	[0.1689]	[0.1304]	[0.1559]	[0.1500]
Smoothing parameter:										
Plain:										
Mean	0.0456	0.4434	0.0304	0.0891	0.0414	0.2049	0.0395	0.3090	0.0299	0.0383
Std. Dev.	0.0021	0.4960	0.0022	0.0282	0.0026	0.1758	0.0024	0.2020	0.0017	0.0064
#{Trimmed}	0	60	0	0	0	3	0	12	0	0
TS-MBC:										
Mean	0.1034	0.3880	0.0690	0.1034	0.0939	0.2007	0.0896	0.3549	0.0677	0.0537
Std. Dev.	0.0047	0.1661	0.0050	0.0562	0.0059	0.1169	0.0055	0.1619	0.0039	0.0337
#{Trimmed}	0	5	0	0	0	2	0	3	0	0
JLN-MBC:										
Mean	0.1034	0.3030	0.0690	0.1461	0.0939	0.2019	0.0896	0.2724	0.0677	0.0848
Std. Dev.	0.0047	0.1154	0.0050	0.0274	0.0059	0.0723	0.0055	0.0847	0.0039	0.0256
#{Trimmed}	0	3	0	0	0	1	0	1	0	0

Note: "ROT" and "BR" in column headings denote "rule-of-thumb" and "beta-referenced" smoothing parameter choice methods. Numbers in parentheses and brackets for density estimators are averages of integrated squared biases and standard errors (defined as square roots of the estimates of asymptotic integrated variances). "Mean", "Std. Dev.", and "#{Trimmed}" for smoothing parameters are averages, standard deviations, and numbers of smoothing parameters trimmed at one.

($\hat{f}_{\mathrm{JLN},B}(x)$) estimators for $n = 100$, where average is taken over 1000 replications for each grid. For comparison, results from two smoothing parameters are provided for a given distribution. We can see that if the bias-uncorrected estimator

Table 3
Average ISE computed on 1000 replications ($n = 100$; Distributions 6–10).

	True distribution									
	6		7		8		9		10	
	ROT	BR	ROT	BR	ROT	BR	ROT	BR	ROT	BR
Density estimator:										
$\hat{f}_B$	0.2801	0.3090	0.0310	0.0255	0.0384	0.0449	0.0387	0.0542	0.1004	0.1336
	(0.2524)	(0.2837)	(0.0023)	(0.0104)	(0.0102)	(0.0210)	(0.0122)	(0.0446)	(0.0728)	(0.1147)
	[0.2015]	[0.1843]	[0.2029]	[0.1502]	[0.2106]	[0.1906]	[0.1977]	[0.1398]	[0.2089]	[0.1717]
$\hat{f}_{MB}$	0.2850	0.3184	0.0337	0.0271	0.0331	0.0379	0.0356	0.0604	0.1149	0.1657
	(0.2496)	(0.2837)	(0.0014)	(0.0100)	(0.0063)	(0.0155)	(0.0101)	(0.0510)	(0.0861)	(0.1448)
	[0.2137]	[0.1948]	[0.2060]	[0.1528]	[0.2044]	[0.1844]	[0.1920]	[0.1379]	[0.2068]	[0.1738]
$\hat{f}_B^r$	0.1382	0.1533	0.0506	0.0291	0.0430	0.0568	0.0453	0.0686	0.1140	0.1545
	(0.0700)	(0.0893)	(0.0112)	(0.0051)	(0.0148)	(0.0315)	(0.0183)	(0.0586)	(0.0837)	(0.1341)
	[0.2370]	[0.2174]	[0.2136]	[0.1599]	[0.2092]	[0.1891]	[0.1968]	[0.1391]	[0.2110]	[0.1758]
$\hat{f}_{MB}^r$	0.2422	0.2708	0.0377	0.0236	0.0315	0.0356	0.0360	0.0637	0.1222	0.1637
	(0.2009)	(0.2325)	(0.0071)	(0.0097)	(0.0049)	(0.0138)	(0.0114)	(0.0522)	(0.0957)	(0.1467)
	[0.2177]	[0.1996]	[0.2028]	[0.1527]	[0.2051]	[0.1848]	[0.1915]	[0.1369]	[0.2029]	[0.1691]
$\hat{f}_B^R$	0.2550	0.2815	0.0316	0.0243	0.0385	0.0461	0.0375	0.0527	0.0969	0.1358
	(0.2194)	(0.2482)	(0.0015)	(0.0079)	(0.0115)	(0.0234)	(0.0124)	(0.0442)	(0.0699)	(0.1165)
	[0.2139]	[0.1966]	[0.2051]	[0.1533]	[0.2083]	[0.1877]	[0.1951]	[0.1371]	[0.2080]	[0.1723]
$\hat{f}_{MB}^R$	0.2842	0.3169	0.0326	0.0267	0.0327	0.0366	0.0359	0.0602	0.1086	0.1505
	(0.2483)	(0.2817)	(0.0014)	(0.0102)	(0.0048)	(0.0133)	(0.0097)	(0.0510)	(0.0809)	(0.1315)
	[0.2149]	[0.1966]	[0.2047]	[0.1524]	[0.2062]	[0.1861]	[0.1931]	[0.1372]	[0.2051]	[0.1702]
$\tilde{f}_{\mathrm{TS},B}$	0.3222	0.2971	0.0257	0.0255	0.0411	0.0496	0.0448	0.0622	0.1342	0.1402
	(0.3015)	(0.2687)	(0.0035)	(0.0069)	(0.0189)	(0.0070)	(0.0239)	(0.0557)	(0.1125)	(0.1175)
	[0.1816]	[0.1947]	[0.1863]	[0.1658]	[0.1989]	[0.2284]	[0.1859]	[0.1367]	[0.1970]	[0.1931]
$\tilde{f}_{\mathrm{TS},MB}$	0.3412	0.3252	0.0304	0.0301	0.0363	0.0456	0.0491	0.0663	0.1969	0.2060
	(0.3135)	(0.2815)	(0.0038)	(0.0067)	(0.0147)	(0.0046)	(0.0276)	(0.0622)	(0.1722)	(0.1778)
	[0.1929]	[0.2083]	[0.1909]	[0.1693]	[0.1918]	[0.2228]	[0.1812]	[0.1363]	[0.2013]	[0.1979]
$\tilde{f}_{\mathrm{TS},B}^R$	0.2919	0.2722	0.0257	0.0248	0.0406	0.0493	0.0431	0.0599	0.1247	0.1318
	(0.2646)	(0.2354)	(0.0022)	(0.0047)	(0.0206)	(0.0076)	(0.0239)	(0.0544)	(0.1044)	(0.1104)
	[0.1944]	[0.2072]	[0.1890]	[0.1690]	[0.1942]	[0.2252]	[0.1825]	[0.1336]	[0.1949]	[0.1913]
$\tilde{f}_{\mathrm{TS},MB}^R$	0.3399	0.3220	0.0276	0.0284	0.0361	0.0456	0.0492	0.0651	0.1431	0.1505
	(0.3129)	(0.2822)	(0.0032)	(0.0069)	(0.0143)	(0.0045)	(0.0276)	(0.0614)	(0.1240)	(0.1288)
	[0.1933]	[0.2070]	[0.1881]	[0.1679]	[0.1923]	[0.2232]	[0.1814]	[0.1342]	[0.1923]	[0.1888]
$\tilde{f}_{\mathrm{JLN},B}$	0.2787	0.2617	0.0291	0.0262	0.0316	0.0373	0.0383	0.0512	0.1011	0.1163
	(0.2519)	(0.2280)	(0.0014)	(0.0029)	(0.0070)	(0.0136)	(0.0137)	(0.0383)	(0.0761)	(0.0951)
	[0.1648]	[0.1739]	[0.1659]	[0.1495]	[0.1693]	[0.1588]	[0.1603]	[0.1312]	[0.1706]	[0.1601]
$\tilde{f}_{\mathrm{JLN},MB}$	0.3182	0.3007	0.0270	0.0255	0.0346	0.0416	0.0433	0.0615	0.1123	0.1324
	(0.2948)	(0.2693)	(0.0022)	(0.0043)	(0.0105)	(0.0185)	(0.0202)	(0.0511)	(0.0902)	(0.1121)
	[0.1663]	[0.1758]	[0.1661]	[0.1496]	[0.1664]	[0.1558]	[0.1585]	[0.1308]	[0.1692]	[0.1593]
$\tilde{f}_{\mathrm{JLN},B}^R$	0.2555	0.2423	0.0298	0.0266	0.0319	0.0376	0.0380	0.0505	0.1089	0.1265
	(0.2210)	(0.1996)	(0.0006)	(0.0017)	(0.0067)	(0.0133)	(0.0139)	(0.0383)	(0.0825)	(0.1035)
	[0.1748]	[0.1838]	[0.1678]	[0.1517]	[0.1696]	[0.1590]	[0.1593]	[0.1296]	[0.1722]	[0.1622]
$\tilde{f}_{\mathrm{JLN},MB}^R$	0.3103	0.2952	0.0272	0.0256	0.0351	0.0409	0.0436	0.0613	0.1170	0.1357
	(0.2830)	(0.2590)	(0.0021)	(0.0041)	(0.0091)	(0.0158)	(0.0200)	(0.0511)	(0.0941)	(0.1147)
	[0.1724]	[0.1817]	[0.1667]	[0.1504]	[0.1689]	[0.1586]	[0.1589]	[0.1301]	[0.1703]	[0.1600]
Smoothing parameter:										
Plain:										
Mean	0.0559	0.0771	0.0502	0.1802	0.0282	0.0459	0.0424	0.2372	0.0453	0.1035
Std. Dev.	0.0021	0.0159	0.0020	0.1420	0.0019	0.0124	0.0019	0.1892	0.0027	0.0256
#{Trimmed}	0	0	0	7	0	0	0	10	0	0
TS-MBC:										
Mean	0.1268	0.1048	0.1137	0.1892	0.0639	0.0433	0.0961	0.3794	0.1026	0.1133
Std. Dev.	0.0048	0.0199	0.0045	0.0852	0.0042	0.0323	0.0043	0.1169	0.0061	0.0208
#{Trimmed}	0	0	0	1	0	0	0	3	0	0
JLN-MBC:										
Mean	0.1268	0.1084	0.1137	0.1774	0.0639	0.0879	0.0961	0.2440	0.1026	0.1355
Std. Dev.	0.0048	0.0133	0.0045	0.0606	0.0042	0.0200	0.0043	0.0623	0.0061	0.0199
#{Trimmed}	0	0	0	0	0	0	0	0	0	0

Note: "ROT" and "BR" in column headings denote "rule-of-thumb" and "beta-referenced" smoothing parameter choice methods. Numbers in parentheses and brackets for density estimators are averages of integrated squared biases and standard errors (defined as square roots of the estimates of asymptotic integrated variances). "Mean", "Std. Dev.", and "#{Trimmed}" for smoothing parameters are averages, standard deviations, and numbers of smoothing parameters trimmed at one.

underestimates (overestimates) the density, JLN-MBC corrects the estimate in an upward (downward) direction, as reported in Hirukawa (2006). Such a bias correction mechanism is not obvious for TS-MBC.

Table 4
Average ISE computed on 1000 replications ($n = 200$; Distributions 1–5).

	True distribution									
	1		2		3		4		5	
	ROT	BR	ROT	BR	ROT	BR	ROT	BR	ROT	BR
Density estimator:										
	0.0172	0.0036	0.0231	0.0257	0.0198	0.0270	0.0182	0.0296	0.0192	0.0202
$\hat{f}_B$	(0.0000)	(0.0000)	(0.0020)	(0.0146)	(0.0019)	(0.0176)	(0.0006)	(0.0245)	(0.0044)	(0.0069)
	[0.1520]	[0.0860]	[0.1757]	[0.1323]	[0.1596]	[0.1109]	[0.1569]	[0.0942]	[0.1474]	[0.1397]
	0.0188	0.0029	0.0241	0.0173	0.0206	0.0246	0.0191	0.0347	0.0156	0.0150
$\hat{f}_{MB}$	(0.0000)	(0.0000)	(0.0003)	(0.0053)	(0.0006)	(0.0113)	(0.0001)	(0.0260)	(0.0020)	(0.0032)
	[0.1519]	[0.0860]	[0.1775]	[0.1345]	[0.1608]	[0.1122]	[0.1566]	[0.0941]	[0.1433]	[0.1349]
	0.0283	0.0135	0.0461	0.0389	0.0349	0.0261	0.0305	0.0323	0.0199	0.0217
$\hat{f}_B^r$	(0.0070)	(0.0072)	(0.0191)	(0.0252)	(0.0119)	(0.0132)	(0.0092)	(0.0254)	(0.0056)	(0.0089)
	[0.1567]	[0.0882]	[0.1828]	[0.1399]	[0.1660]	[0.1168]	[0.1613]	[0.0969]	[0.1459]	[0.1378]
	0.0204	0.0038	0.0293	0.0276	0.0255	0.0227	0.0214	0.0299	0.0147	0.0129
$\hat{f}_{MB}^r$	(0.0026)	(0.0002)	(0.0070)	(0.0168)	(0.0066)	(0.0137)	(0.0033)	(0.0222)	(0.0005)	(0.0008)
	[0.1504]	[0.0864]	[0.1748]	[0.1327]	[0.1586]	[0.1118]	[0.1552]	[0.0945]	[0.1447]	[0.1365]
	0.0171	0.0034	0.0228	0.0215	0.0198	0.0249	0.0181	0.0299	0.0197	0.0214
$\hat{f}_B^R$	(0.0000)	(0.0000)	(0.0011)	(0.0097)	(0.0015)	(0.0148)	(0.0006)	(0.0248)	(0.0055)	(0.0086)
	[0.1521]	[0.0860]	[0.1768]	[0.1340]	[0.1604]	[0.1122]	[0.1569]	[0.0940]	[0.1458]	[0.1378]
	0.0188	0.0029	0.0234	0.0173	0.0202	0.0244	0.0190	0.0347	0.0149	0.0135
$\hat{f}_{MB}^R$	(0.0000)	(0.0000)	(0.0004)	(0.0061)	(0.0007)	(0.0117)	(0.0001)	(0.0260)	(0.0007)	(0.0011)
	[0.1520]	[0.0860]	[0.1768]	[0.1337]	[0.1604]	[0.1119]	[0.1568]	[0.0941]	[0.1448]	[0.1367]
	0.0123	0.0045	0.0161	0.0174	0.0145	0.0168	0.0130	0.0242	0.0194	0.0249
$\tilde{f}_{TS,B}$	(0.0000)	(0.0000)	(0.0012)	(0.0040)	(0.0017)	(0.0067)	(0.0005)	(0.0172)	(0.0088)	(0.0051)
	[0.1361]	[0.0968]	[0.1580]	[0.1458]	[0.1429]	[0.1229]	[0.1410]	[0.1010]	[0.1347]	[0.1462]
	0.0142	0.0035	0.0245	0.0242	0.0200	0.0217	0.0150	0.0344	0.0117	0.0178
$\tilde{f}_{TS,MB}$	(0.0000)	(0.0000)	(0.0065)	(0.0102)	(0.0051)	(0.0063)	(0.0010)	(0.0183)	(0.0024)	(0.0013)
	[0.1360]	[0.0969]	[0.1617]	[0.1502]	[0.1453]	[0.1256]	[0.1410]	[0.1015]	[0.1283]	[0.1415]
	0.0122	0.0042	0.0160	0.0162	0.0143	0.0161	0.0128	0.0250	0.0208	0.0263
$\tilde{f}_{TS,B}^R$	(0.0000)	(0.0000)	(0.0010)	(0.0032)	(0.0014)	(0.0055)	(0.0006)	(0.0180)	(0.0111)	(0.0063)
	[0.1361]	[0.0968]	[0.1583]	[0.1464]	[0.1435]	[0.1238]	[0.1406]	[0.1004]	[0.1316]	[0.1437]
	0.0141	0.0034	0.0174	0.0153	0.0165	0.0181	0.0141	0.0348	0.0107	0.0162
$\tilde{f}_{TS,MB}^R$	(0.0000)	(0.0000)	(0.0020)	(0.0040)	(0.0032)	(0.0049)	(0.0008)	(0.0197)	(0.0012)	(0.0007)
	[0.1360]	[0.0968]	[0.1584]	[0.1462]	[0.1435]	[0.1236]	[0.1406]	[0.1006]	[0.1301]	[0.1426]
	0.0151	0.0073	0.0182	0.0131	0.0161	0.0125	0.0152	0.0096	0.0133	0.0148
$\tilde{f}_{JLN,B}$	(0.0000)	(0.0000)	(0.0003)	(0.0032)	(0.0005)	(0.0024)	(0.0001)	(0.0034)	(0.0017)	(0.0031)
	[0.1199]	[0.0899]	[0.1395]	[0.1137]	[0.1264]	[0.1059]	[0.1239]	[0.0933]	[0.1144]	[0.1097]
	0.0147	0.0060	0.0178	0.0139	0.0160	0.0141	0.0148	0.0094	0.0130	0.0140
$\tilde{f}_{JLN,MB}$	(0.0001)	(0.0000)	(0.0008)	(0.0060)	(0.0013)	(0.0045)	(0.0001)	(0.0040)	(0.0020)	(0.0038)
	[0.1199]	[0.0899]	[0.1392]	[0.1132]	[0.1264]	[0.1058]	[0.1236]	[0.0930]	[0.1124]	[0.1074]
	0.0152	0.0072	0.0185	0.0110	0.0162	0.0120	0.0153	0.0094	0.0132	0.0149
$\tilde{f}_{JLN,B}^R$	(0.0000)	(0.0000)	(0.0000)	(0.0003)	(0.0002)	(0.0013)	(0.0000)	(0.0029)	(0.0013)	(0.0027)
	[0.1202]	[0.0900]	[0.1405]	[0.1156]	[0.1271]	[0.1070]	[0.1241]	[0.0937]	[0.1148]	[0.1100]
	0.0148	0.0060	0.0180	0.0128	0.0160	0.0138	0.0149	0.0085	0.0132	0.0130
$\tilde{f}_{JLN,MB}^R$	(0.0000)	(0.0000)	(0.0009)	(0.0048)	(0.0012)	(0.0041)	(0.0001)	(0.0030)	(0.0012)	(0.0022)
	[0.1201]	[0.0900]	[0.1399]	[0.1146]	[0.1267]	[0.1064]	[0.1240]	[0.0938]	[0.1142]	[0.1095]
Smoothing parameter:										
Plain:										
Mean	0.0346	0.4448	0.0232	0.0699	0.0315	0.1439	0.0300	0.2637	0.0227	0.0291
Std. Dev.	0.0011	0.3946	0.0012	0.0146	0.0014	0.0950	0.0013	0.1340	0.0009	0.0034
#{Trimmed}	0	62	0	0	0	2	0	7	0	0
TS-MBC:										
Mean	0.0887	0.3805	0.0595	0.0859	0.0807	0.1540	0.0770	0.3326	0.0582	0.0481
Std. Dev.	0.0029	0.1624	0.0031	0.0324	0.0036	0.0625	0.0034	0.1334	0.0024	0.0285
#{Trimmed}	0	10	0	0	0	0	0	0	0	0
JLN-MBC:										
Mean	0.0887	0.3023	0.0595	0.1303	0.0807	0.1657	0.0770	0.2503	0.0582	0.0716
Std. Dev.	0.0029	0.1137	0.0031	0.0179	0.0036	0.0467	0.0034	0.0635	0.0024	0.0145
#{Trimmed}	0	2	0	0	0	0	0	0	0	0

Note: "ROT" and "BR" in column headings denote "rule-of-thumb" and "beta-referenced" smoothing parameter choice methods. Numbers in parentheses and brackets for density estimators are averages of integrated squared biases and standard errors (defined as square roots of the estimates of asymptotic integrated variances). "Mean", "Std. Dev.", and "#{Trimmed}" for smoothing parameters are averages, standard deviations, and numbers of smoothing parameters trimmed at one.

Tables 2–5 also indicate that either the modified beta TS-MBC estimator $\tilde{f}_{TS,MB}(x)$ or the "micro" beta estimator $\hat{f}_B^r(x)$ consistently performs inferiorly. The reasons for their poor performance can be explained as follows. First, when it comes to a large bias in $\tilde{f}_{TS,MB}(x)$, a TS-MBC estimator depends on two smoothing parameters b and b/c. Controlling both b and b/c

Table 5
Average ISE computed on 1000 replications ($n = 200$; Distributions 6–10).

	True distribution									
	6		7		8		9		10	
	ROT	BR	ROT	BR	ROT	BR	ROT	BR	ROT	BR
Density estimator:										
	0.2356	0.2662	0.0184	0.0168	0.0237	0.0280	0.0247	0.0457	0.0794	0.1077
$\hat{f}_B$	(0.2186)	(0.2507)	(0.0015)	(0.0080)	(0.0064)	(0.0141)	(0.0084)	(0.0393)	(0.0629)	(0.0969)
	[0.1552]	[0.1421]	[0.1544]	[0.1148]	[0.1585]	[0.1431]	[0.1486]	[0.1064]	[0.1574]	[0.1307]
	0.2355	0.2691	0.0198	0.0172	0.0198	0.0219	0.0219	0.0514	0.0884	0.1329
$\hat{f}_{MB}$	(0.2133)	(0.2476)	(0.0007)	(0.0068)	(0.0033)	(0.0090)	(0.0061)	(0.0446)	(0.0714)	(0.1214)
	[0.1650]	[0.1508]	[0.1565]	[0.1169]	[0.1544]	[0.1386]	[0.1444]	[0.1047]	[0.1548]	[0.1313]
	0.0946	0.1068	0.0337	0.0206	0.0262	0.0353	0.0287	0.0619	0.0862	0.1263
$\hat{f}_B^r$	(0.0539)	(0.0694)	(0.0112)	(0.0079)	(0.0091)	(0.0208)	(0.0123)	(0.0548)	(0.0686)	(0.1152)
	[0.1821]	[0.1674]	[0.1619]	[0.1222]	[0.1576]	[0.1420]	[0.1480]	[0.1060]	[0.1584]	[0.1331]
	0.1918	0.2219	0.0237	0.0178	0.0187	0.0196	0.0221	0.0551	0.0968	0.1381
$\hat{f}_{MB}^r$	(0.1668)	(0.1986)	(0.0057)	(0.0097)	(0.0023)	(0.0073)	(0.0067)	(0.0465)	(0.0810)	(0.1278)
	[0.1675]	[0.1537]	[0.1540]	[0.1162]	[0.1550]	[0.1390]	[0.1443]	[0.1039]	[0.1526]	[0.1278]
	0.2106	0.2383	0.0185	0.0155	0.0239	0.0291	0.0242	0.0448	0.0752	0.1073
$\hat{f}_B^R$	(0.1891)	(0.2183)	(0.0010)	(0.0060)	(0.0073)	(0.0157)	(0.0086)	(0.0390)	(0.0591)	(0.0966)
	[0.1640]	[0.1509]	[0.1558]	[0.1169]	[0.1571]	[0.1413]	[0.1470]	[0.1044]	[0.1566]	[0.1306]
	0.2350	0.2681	0.0192	0.0169	0.0194	0.0207	0.0219	0.0514	0.0857	0.1209
$\hat{f}_{MB}^R$	(0.2127)	(0.2466)	(0.0007)	(0.0070)	(0.0024)	(0.0073)	(0.0058)	(0.0447)	(0.0690)	(0.1101)
	[0.1655]	[0.1516]	[0.1557]	[0.1164]	[0.1556]	[0.1399]	[0.1453]	[0.1043]	[0.1543]	[0.1290]
	0.2916	0.2639	0.0146	0.0158	0.0269	0.0276	0.0320	0.0563	0.1154	0.1195
$\tilde{f}_{TS,B}$	(0.2799)	(0.2478)	(0.0026)	(0.0057)	(0.0144)	(0.0057)	(0.0199)	(0.0523)	(0.1033)	(0.1071)
	[0.1348]	[0.1451]	[0.1374]	[0.1228]	[0.1451]	[0.1635]	[0.1359]	[0.1017]	[0.1447]	[0.1424]
	0.3087	0.2867	0.0174	0.0185	0.0231	0.0250	0.0349	0.0622	0.1649	0.1730
$\tilde{f}_{TS,MB}$	(0.2927)	(0.2617)	(0.0029)	(0.0052)	(0.0110)	(0.0036)	(0.0226)	(0.0592)	(0.1512)	(0.1583)
	[0.1440]	[0.1558]	[0.1409]	[0.1255]	[0.1400]	[0.1593]	[0.1321]	[0.1012]	[0.1469]	[0.1450]
	0.2607	0.2376	0.0142	0.0147	0.0270	0.0277	0.0311	0.0545	0.1041	0.1089
$\tilde{f}_{TS,B}^R$	(0.2454)	(0.2169)	(0.0017)	(0.0039)	(0.0156)	(0.0061)	(0.0199)	(0.0510)	(0.0929)	(0.0975)
	[0.1438]	[0.1540]	[0.1391]	[0.1249]	[0.1421]	[0.1614]	[0.1336]	[0.0995]	[0.1428]	[0.1407]
	0.3081	0.2847	0.0152	0.0173	0.0228	0.0249	0.0349	0.0614	0.1206	0.1263
$\tilde{f}_{TS,MB}^R$	(0.2929)	(0.2621)	(0.0020)	(0.0052)	(0.0105)	(0.0034)	(0.0224)	(0.0586)	(0.1098)	(0.1144)
	[0.1435]	[0.1543]	[0.1387]	[0.1242]	[0.1406]	[0.1598]	[0.1326]	[0.0998]	[0.1410]	[0.1389]
	0.2455	0.2268	0.0159	0.0147	0.0186	0.0227	0.0246	0.0413	0.0811	0.0944
$\tilde{f}_{JLN,B}$	(0.2305)	(0.2079)	(0.0009)	(0.0022)	(0.0048)	(0.0099)	(0.0104)	(0.0334)	(0.0671)	(0.0828)
	[0.1225]	[0.1294]	[0.1224]	[0.1106]	[0.1241]	[0.1160]	[0.1173]	[0.0969]	[0.1251]	[0.1179]
	0.2857	0.2658	0.0151	0.0148	0.0208	0.0260	0.0290	0.0532	0.0927	0.1089
$\tilde{f}_{JLN,MB}$	(0.2724)	(0.2481)	(0.0014)	(0.0032)	(0.0072)	(0.0136)	(0.0155)	(0.0465)	(0.0800)	(0.0978)
	[0.1237]	[0.1309]	[0.1225]	[0.1107]	[0.1222]	[0.1140]	[0.1158]	[0.0965]	[0.1238]	[0.1171]
	0.2221	0.2064	0.0160	0.0145	0.0187	0.0228	0.0245	0.0410	0.0855	0.1009
$\tilde{f}_{JLN,B}^R$	(0.2031)	(0.1829)	(0.0005)	(0.0013)	(0.0047)	(0.0098)	(0.0107)	(0.0336)	(0.0710)	(0.0886)
	[0.1293]	[0.1362]	[0.1235]	[0.1119]	[0.1242]	[0.1161]	[0.1165]	[0.0957]	[0.1258]	[0.1190]
	0.2779	0.2597	0.0151	0.0148	0.0209	0.0251	0.0290	0.0531	0.0971	0.1117
$\tilde{f}_{JLN,MB}^R$	(0.2626)	(0.2396)	(0.0013)	(0.0031)	(0.0064)	(0.0119)	(0.0153)	(0.0465)	(0.0839)	(0.1002)
	[0.1279]	[0.1349]	[0.1228]	[0.1110]	[0.1237]	[0.1157]	[0.1162]	[0.0960]	[0.1246]	[0.1176]
Smoothing parameter:										
Plain:										
Mean	0.0424	0.0572	0.0380	0.1223	0.0213	0.0343	0.0321	0.1591	0.0344	0.0761
Std. Dev.	0.0011	0.0072	0.0010	0.0499	0.0010	0.0058	0.0010	0.0983	0.0014	0.0116
#{Trimmed}	0	0	0	0	0	0	0	3	0	0
TS-MBC:										
Mean	0.1088	0.0875	0.0975	0.1524	0.0547	0.0365	0.0823	0.3015	0.0882	0.0948
Std. Dev.	0.0028	0.0101	0.0027	0.0400	0.0025	0.0200	0.0026	0.0789	0.0037	0.0109
#{Trimmed}	0	0	0	0	0	0	0	1	0	0
JLN-MBC:										
Mean	0.1088	0.0919	0.0975	0.1461	0.0547	0.0749	0.0823	0.2003	0.0882	0.1142
Std. Dev.	0.0028	0.0071	0.0027	0.0303	0.0025	0.0115	0.0026	0.0395	0.0037	0.0107
#{Trimmed}	0	0	0	0	0	0	0	0	0	0

Note: "ROT" and "BR" in column headings denote "rule-of-thumb" and "beta-referenced" smoothing parameter choice methods. Numbers in parentheses and brackets for density estimators are averages of integrated squared biases and standard errors (defined as square roots of the estimates of asymptotic integrated variances). "Mean", "Std. Dev.", and "#{Trimmed}" for smoothing parameters are averages, standard deviations, and numbers of smoothing parameters trimmed at one.

is a cumbersome task. Because $0 < c < 1$, the density estimator using b/c tends to be oversmoothed, which is potentially a source of a large bias in every TS-MBC estimator. On the other hand, if we make b too short in order to have a reasonable length of b/c, additional variability is introduced to the other estimator using b due to undersmoothing. In addition, when the modified beta kernel is employed, these two smoothing parameters also play a role of determining the boundary region explicitly (e.g. $[0, 2b) \cup (1-2b, 1]$ for the density estimator using b). Unless b is short enough, there is a very narrow 'interior' region for the density estimator using b/c. This aspect is also thought to amplify the bias in $\tilde{f}_{TS,MB}(x)$. In other words, when TS-MBC estimation is applied, it is desirable to use a smoothing parameter choice method that tends to pick up a small value consistently. This explains why the rule-of-thumb method in general performs better in TS-MBC estimation than the beta-referenced method.

Second, Fig. 2 exhibits average plots of selected density estimates based on $\hat{f}_B^r(x)$. For brevity, only plots based on the rule-of-thumb method are provided. We can see from Panels (a), (b) and (d) that the density estimates tend to be upward biased substantially near the boundary, wherever the tail of the true density is thick; on the other hand, this issue is not serious if the tail is thin (e.g. left tail of Panel (a)). In fact, $\hat{f}_B^r(x)$ possesses two kinds of bias terms. Other than the same bias term as $\hat{f}_B(x)$ has $\left(=\left\{(1-2x)f'(x)+(1/2)x(1-x)f''(x)\right\}b\right)$, the estimator generates an extra $O(b)$ bias that is induced by renormalization and thus peculiar to this estimator. The extra bias comes from the fact that the normalization constant does not become unity near the boundary. More precisely, if the data point X_i lies in an $O(b)$ boundary region, $\int_0^1 K_{B(x/b+1,(1-x)/b+1)}(X_i)\,dx \sim \Phi\left(\sqrt{\kappa}\right) \in (1/2, 1)$, where $\Phi(\cdot)$ is the standard normal distribution function; the derivation is similar to the proof for the variance part of Theorem 2, and thus it is omitted. In finite samples, the $O(b)$ boundary region is not negligible, and a number of observations are likely to be concentrated near the boundary when the true density has a thick tail. Then, the 'renormalized' contributions $K_{B(x/b+1,(1-x)/b+1)}(X_i) / \int_0^1 K_{B(x/b+1,(1-x)/b+1)}(X_i)\,dx$ from these observations tend to be inflated. As a result, at a design point x toward the boundary, the inflated contributions dominate due to proximity of x to the observations in the boundary region, leading to an upward biased estimate. A similar argument can also explain superior performance of $\hat{f}_B^r(x)$ for Distribution 6 in Panel (c). Moreover, as mentioned above, the bias-uncorrected and its renormalized estimators have no asymptotic biases for Distribution 1. Nonetheless, in this case, $\hat{f}_B^r(x)$ still generates a substantially large bias; indeed, the extra bias induced by renormalization does not disappear even in this case.

4. An application: Estimating the density of eruption durations of the Old Faithful Geyser

We apply two MBC techniques to a data set of eruption durations of the Old Faithful Geyser in Yellow Stone National Park, Wyoming, USA. The data set is famous and applied frequently in the literature. Many textbooks on nonparametric statistics (e.g. Silverman, 1986; Loader, 1999) adopt it as an example of density estimation. Azzalini and Bowman (1990) also analyze the data set extensively. The data set can be found in many sources. The one used in this section has been downloaded from the web site for Wasserman (2006).

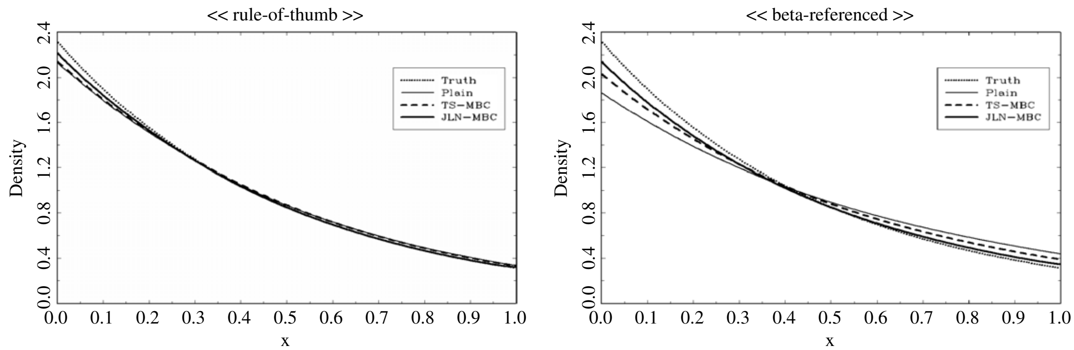
The data set contains durations of 272 consecutive eruptions, ranging from 1.6 to 5.1 minutes. As in Gu (2002, p. 191), the density of eruption durations is assumed to have a compact support [1.5, 5.25]. An extension of beta density estimation to any compact support is easy. In this example, an original observation $X_i \in [1.5, 5.25]$ can be mapped onto the unit interval by the transformation $Z_i = (X_i - 1.5)/3.75$. Then, beta density estimation can be implemented using the transformed data Z_i.

It is widely believed that the density is bimodal, which would be the closest to Distribution 10 in our Monte Carlo simulations. Tables 3 and 5 indicate that for this distribution, the beta kernel and the rule-of-thumb method perform better than the modified beta kernel and the beta-referenced method, respectively. Renormalization does not necessarily decrease the average ISE. Therefore, the following three density estimators are considered: (i) bias-uncorrected beta estimator $\hat{f}_B(x)$; (ii) beta TS-MBC estimator $\tilde{f}_{TS,B}(x)$ with the MISE-optimal $c^* = 0.2636$; and (iii) beta JLN-MBC estimator $\tilde{f}_{JLN,B}(x)$. Smoothing parameters are chosen via the rule-of-thumb method for the transformed data, which picks $\hat{b}_{plain} = 0.0323$ and $\hat{b}_{mbc} = 0.0876$.

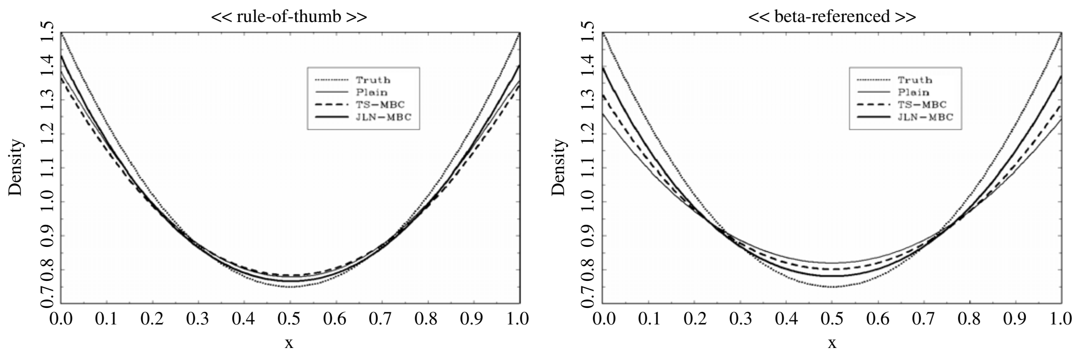
The three density estimates are plotted in Fig. 3. Note that "**×**" symbols are data points. Each of three density estimates successfully captures the bimodal shape. As the true density is unknown, it is hard to judge superiority among three estimators. However, we can see that compared with the bias-uncorrected estimate, two MBC estimates tend to make the heights of two peaks lower without changing their locations much.

5. Conclusion

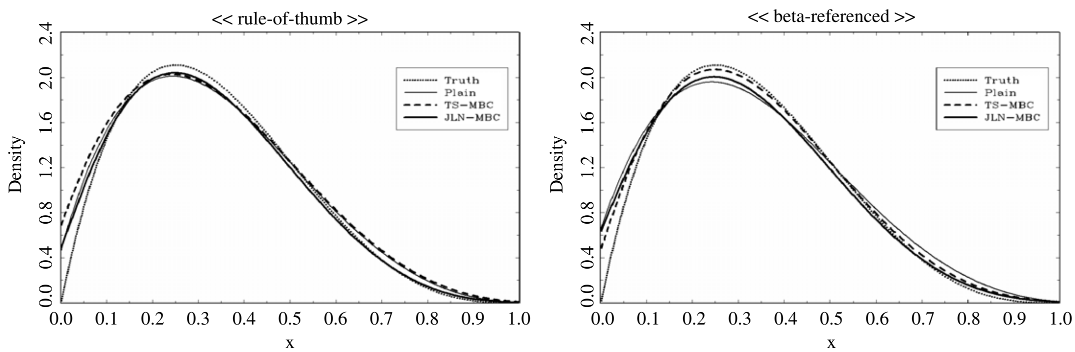
This paper has demonstrated that two well-known bias correction techniques can be applied to density estimation using the beta and modified beta kernels. Under sufficient smoothness of the true density, both bias reduction methods are shown to improve the order of magnitude in bias from $O(b)$ to $O\left(b^2\right)$, while the order of magnitude in variance remains unchanged. Two classes of bias-corrected density estimators are by construction nonnegative, and establish a faster convergence rate of $O\left(n^{-8/9}\right)$ in MSE for the interior part when best implemented, as with symmetric second-order kernels. Monte Carlo simulations indicate superior performance of JLN-type estimators in particular, compared to corresponding bias-uncorrected estimators.



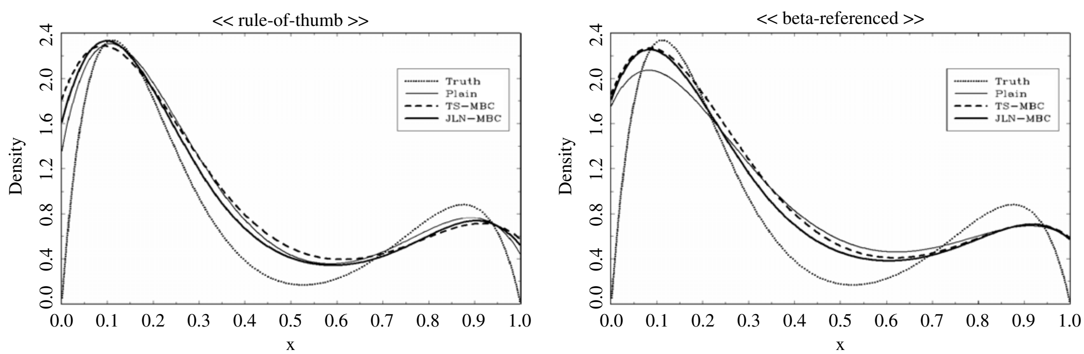
(a) Truncated Exp(1/2) [Distribution 3].



(b) $(1/2)B(3, 1) + (1/2)B(1, 3)$ [Distribution 7].

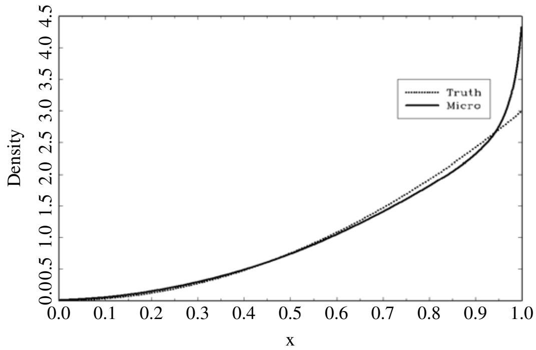


(c) $B(2, 4)$ [Distribution 8].

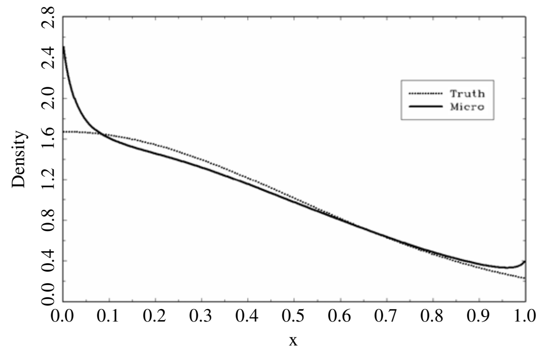


(d) $(1/4)B(8, 2) + (3/4)B(2, 8)$ [Distribution 10].

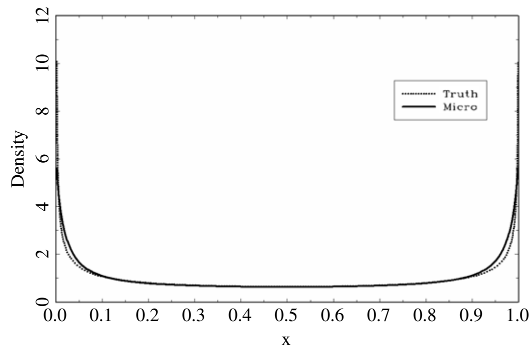
Fig. 1. Average plots of MBC beta density estimates ($n = 100$).



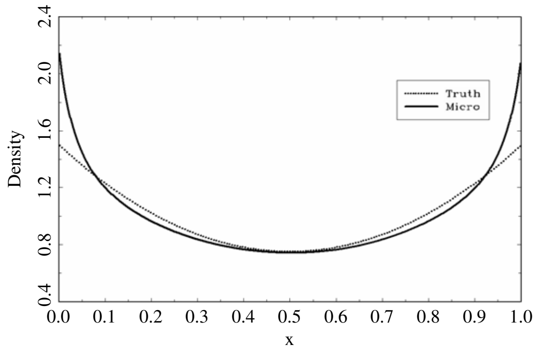
(a) $B(3, 1)$ [Distribution 2].



(b) Truncated $N(0, 1/4)$ [Distribution 4].



(c) $B(1/2, 1/2)$ [Distribution 6].



(d) $(1/2)B(3, 1) + (1/2)B(1, 3)$ [Distribution 7].

Fig. 2. Average plots of "micro" beta density estimates ($n = 100$; rule-of-thumb method).

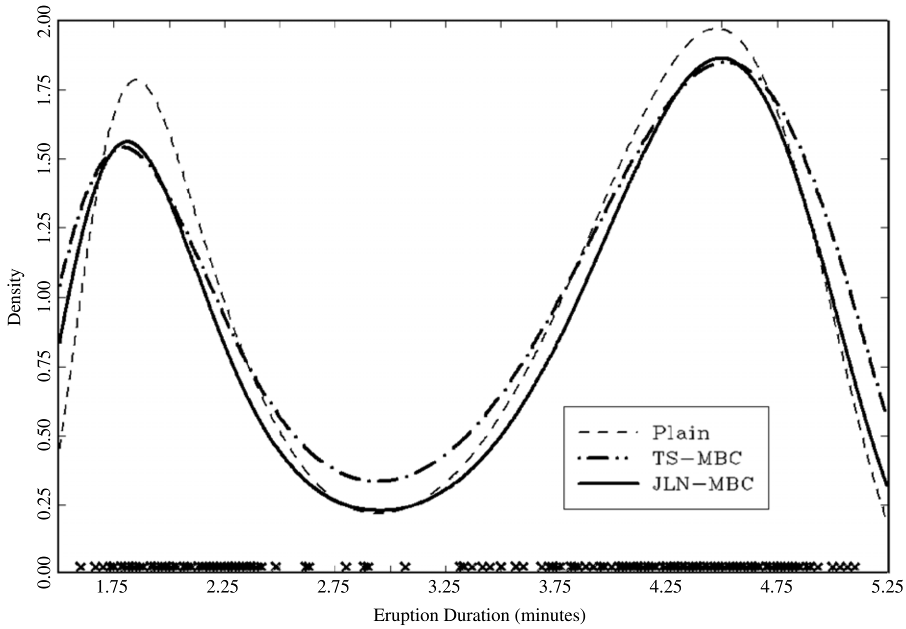


Fig. 3. Density estimates of eruption durations of the Old Faithful Geyser.

Appendix

A.1. Proof of Theorem 1

We present only the proof when the beta kernel is employed. The proof when the modified beta kernel is used is similar, and thus it is omitted. Accordingly, we can safely suppress the subscript *j*.

A.1.1. Bias

The proof largely follows TS. Let $\theta_x \overset{d}{=} B\left(x/b+1, (1-x)/b+1\right)$. Taking a fourth-order Taylor expansion around $\theta_x = x$ for

$$I_b(x) = E\left\{\hat{f}_b(x)\right\} = \int_0^1 K_{B(x/b+1,(1-x)/b+1)}(u) f(u)\,\mathrm{d}u = E\left\{f(\theta_x)\right\},$$

we have

$$I_b(x) = f(x) + \sum_{j=1}^{4} \frac{f^{(j)}(x)}{j!} E(\theta_x - x)^j + o\left\{E(\theta_x - x)^4\right\}.$$

By the property of the beta random variables and a Taylor expansion around $b = 0$,

$$\begin{aligned}
E(\theta_x - x) &= (1-2x)\,b - 2(1-2x)b^2 + O\left(b^3\right),\\
E(\theta_x - x)^2 &= x(1-x)\,b + (11x^2 - 11x + 2)b^2 + O\left(b^3\right),\\
E(\theta_x - x)^3 &= 5x(1-x)(1-2x)\,b^2 + O\left(b^3\right),\\
E(\theta_x - x)^4 &= 3x^2(1-x)^2\,b^2 + O\left(b^3\right).
\end{aligned}$$

An argument similar to the proof of LemmaB2(a)-(ii) in Gospodinov and Hirukawa (2007) can also demonstrate that $E(\theta_x - x)^r = O\left(b^3\right)$ for $r \geq 5$. Hence,

$$I_b(x) = f(x)\left\{1 + \frac{a_1(x)}{f(x)}b + \frac{a_2(x)}{f(x)}b^2 + o\left(b^2\right)\right\}, \tag{1}$$

where

$$\begin{aligned}
a_1(x) &= (1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x),\\
a_2(x) &= -2(1-2x)f'(x) + \frac{1}{2}\left(11x^2 - 11x + 2\right)f''(x) + \frac{5}{6}x(1-x)(1-2x)f'''(x) + \frac{1}{8}x^2(1-x)^2 f''''(x).
\end{aligned}$$

Taking the logarithm on both sides of (1) for an arbitrarily small $b > 0$ and then using a Taylor expansion, we have

$$\log I_b(x) = \log f(x) + \frac{a_1(x)}{f(x)}b + \frac{1}{2}\left\{\frac{2a_2(x)f(x) - a_1^2(x)}{f^2(x)}\right\}b^2 + o\left(b^2\right). \tag{2}$$

Similarly, $\log I_{b/c}(x) = E\left\{\hat{f}_{b/c}(x)\right\}$ can be approximated by

$$\log I_{b/c}(x) = \log f(x) + \frac{1}{c}\left\{\frac{a_1(x)}{f(x)}\right\}b + \frac{1}{2c^2}\left\{\frac{2a_2(x)f(x) - a_1^2(x)}{f^2(x)}\right\}b^2 + o\left(b^2\right). \tag{3}$$

Then,

$$\frac{1}{1-c}\log I_b(x) - \frac{c}{1-c}\log I_{b/c}(x) = \log f(x) - \frac{1}{c(1-c)}\left\{\frac{a_2(x)f(x) - \frac{1}{2}a_1^2(x)}{f^2(x)}\right\}b^2 + o\left(b^2\right),$$

which yields

$$\left\{I_b(x)\right\}^{\frac{1}{1-c}}\left\{I_{b/c}(x)\right\}^{-\frac{c}{1-c}} = f(x) + \frac{1}{c(1-c)}\left[\frac{1}{2}\left\{\frac{a_1^2(x)}{f(x)}\right\} - a_2(x)\right]b^2 + o\left(b^2\right) \tag{4}$$

after taking the exponential on both sides and using a Taylor expansion.

Define $Z = \hat{f}_b(x) - I_b(x)$ and $W = \hat{f}_{b/c}(x) - I_{b/c}(x)$. Then, $E(Z) = E(W) = 0$, and each of $E\left(Z^2\right) = var\left\{\hat{f}_b(x)\right\}$, $E\left(W^2\right) = var\left\{\hat{f}_{b/c}(x)\right\}$ and $E(ZW) = cov\left\{\hat{f}_b(x), \hat{f}_{b/c}(x)\right\}$ is at most $O\left(n^{-1}b^{-1}\right)$, as shown in the variance part below. Again, by a Taylor expansion,

$$\begin{aligned}
\tilde{f}_{TS}(x) &= \left\{I_b(x)\right\}^{\frac{1}{1-c}}\left\{1 + \frac{Z}{I_b(x)}\right\}^{\frac{1}{1-c}}\left\{I_{b/c}(x)\right\}^{-\frac{c}{1-c}}\left\{1 + \frac{W}{I_{b/c}(x)}\right\}^{-\frac{c}{1-c}}\\
&= \left\{I_b(x)\right\}^{\frac{1}{1-c}}\left\{I_{b/c}(x)\right\}^{-\frac{c}{1-c}} + \frac{1}{1-c}Z\left\{\frac{I_b(x)}{I_{b/c}(x)}\right\}^{\frac{c}{1-c}} - \frac{c}{1-c}W\left\{\frac{I_b(x)}{I_{b/c}(x)}\right\}^{\frac{1}{1-c}} + O\left\{(Z+W)^2\right\}.
\end{aligned} \tag{5}$$

Since (2) and (3) imply that $I_b(x)/I_{b/c}(x) = 1 + O(b)$, substituting (4) finally yields

$$E\left\{\tilde{f}_{TS}(x)\right\} = \{I_b(x)\}^{\frac{1}{1-c}}\left\{I_{b/c}(x)\right\}^{-\frac{c}{1-c}} + O\left(n^{-1}b^{-1}\right)$$
$$= f(x) + \frac{1}{c(1-c)}\left[\frac{1}{2}\left\{\frac{a_1^2(x)}{f(x)}\right\} - a_2(x)\right]b^2 + o\left(b^2\right) + O\left(n^{-1}b^{-1}\right),$$

where the remainder term $O\left(n^{-1}b^{-1}\right) = o\left(b^2\right)$ by (A2).

A.1.2. Variance

It follows from (5) that

$$var\left\{\tilde{f}_{TS}(x)\right\} = E\left(\frac{1}{1-c}Z - \frac{c}{1-c}W\right)^2 + O\left(n^{-1}\right)$$
$$= \frac{1}{(1-c)^2}\left[var\left\{\hat{f}_b(x)\right\} - 2c\cdot cov\left\{\hat{f}_b(x), \hat{f}_{b/c}(x)\right\} + c^2\cdot var\left\{\hat{f}_{b/c}(x)\right\}\right] + O\left(n^{-1}\right).$$

For interior x,

$$var\left\{\hat{f}_b(x)\right\} = n^{-1}b^{-1/2}\frac{f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}} + o\left(n^{-1}b^{-1/2}\right),$$
$$var\left\{\hat{f}_{b/c}(x)\right\} = n^{-1}b^{-1/2}c^{1/2}\frac{f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}} + o\left(n^{-1}b^{-1/2}\right).$$

Now,

$$cov\left\{\hat{f}_b(x), \hat{f}_{b/c}(x)\right\} = n^{-1}E\left\{K_{B(x/b+1,(1-x)/b+1)}(X_1)\,K_{B(cx/b+1,c(1-x)/b+1)}(X_1)\right\} + O\left(n^{-1}\right).$$

In particular,

$$E\left\{K_{B(x/b+1,(1-x)/b+1)}(X_1)\,K_{B(cx/b+1,c(1-x)/b+1)}(X_1)\right\} = G_b(x)\{f(x) + o(1)\},$$

where

$$G_b(x) = \frac{B\left(\frac{(1+c)x}{b}+1, \frac{(1+c)(1-x)}{b}+1\right)}{B\left(\frac{x}{b}+1, \frac{1-x}{b}+1\right)B\left(\frac{cx}{b}+1, \frac{c(1-x)}{b}+1\right)}$$
$$= \frac{\Gamma\left(\frac{(1+c)x}{b}+1\right)\Gamma\left(\frac{(1+c)(1-x)}{b}+1\right)\Gamma\left(\frac{1}{b}+2\right)\Gamma\left(\frac{c}{b}+2\right)}{\Gamma\left(\frac{1+c}{b}+2\right)\Gamma\left(\frac{x}{b}+1\right)\Gamma\left(\frac{1-x}{b}+1\right)\Gamma\left(\frac{cx}{b}+1\right)\Gamma\left(\frac{c(1-x)}{b}+1\right)}. \quad (6)$$

Define $R(z) = \sqrt{2\pi}z^{z+1/2}\exp(-z)/\Gamma(z+1)$ for $z > 0$, as in Lemma 3 of Brown and Chen (1999). Then, following the argument in Chen (2000a, p. 87),

$$G_b(x) = \left(\frac{1+c}{c}\right)^{1/2}\frac{b^{1/2}}{\sqrt{2\pi}\sqrt{x(1-x)}}\frac{\left(1+\frac{1}{b}\right)\left(1+\frac{c}{b}\right)}{1+\frac{1+c}{b}}S(b,x),$$

where

$$S(b,x) = \frac{R\left(\frac{1+c}{b}\right)R\left(\frac{x}{b}\right)R\left(\frac{1-x}{b}\right)R\left(\frac{cx}{b}\right)R\left(\frac{c(1-x)}{b}\right)}{R\left(\frac{(1+c)x}{b}\right)R\left(\frac{(1+c)(1-x)}{b}\right)R\left(\frac{1}{b}\right)R\left(\frac{c}{b}\right)} \to 1,$$
$$\frac{\left(1+\frac{1}{b}\right)\left(1+\frac{c}{b}\right)}{1+\frac{1+c}{b}} = \left(\frac{c}{1+c}\right)b^{-1}\{1+O(b)\}.$$

Therefore,

$$cov\left\{\hat{f}_b(x), \hat{f}_{b/c}(x)\right\} = n^{-1}b^{-1/2}\frac{\sqrt{2}c^{1/2}}{(1+c)^{1/2}}\frac{f(x)}{2\sqrt{\pi}\sqrt{x(1-x)}} + o\left(n^{-1}b^{-1/2}\right),$$

which establishes the result for interior x.

On the other hand, as $x/b \to \kappa$ when $x \to 0$,

$$var\left\{\hat{f}_b(x)\right\} = n^{-1}b^{-1}\frac{\Gamma(2\kappa+1)f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o\left(n^{-1}b^{-1}\right),$$

$$var\left\{\hat{f}_{b/c}(x)\right\} = n^{-1}b^{-1}c\frac{\Gamma(2c\kappa+1)f(x)}{2^{2c\kappa+1}\Gamma^2(c\kappa+1)} + o\left(n^{-1}b^{-1}\right).$$

It follows from (6) that

$$G_b(x) \sim \frac{1}{(1+c)^{\kappa}}\left(\frac{c}{1+c}\right)^{c\kappa}\frac{\Gamma((1+c)\kappa+1)}{\Gamma(\kappa+1)\Gamma(c\kappa+1)}\frac{\left(1+\frac{1}{b}\right)\left(1+\frac{c}{b}\right)}{1+\frac{1+c}{b}}\frac{R\left(\frac{1+c}{b}\right)R\left(\frac{1-x}{b}\right)R\left(\frac{c(1-x)}{b}\right)}{R\left(\frac{(1+c)(1-x)}{b}\right)R\left(\frac{1}{b}\right)\Gamma\left(\frac{c}{b}\right)}$$

$$\sim \frac{1}{(1+c)^{\kappa}}\left(\frac{c}{1+c}\right)^{c\kappa+1}\frac{\Gamma((1+c)\kappa+1)b^{-1}}{\Gamma(\kappa+1)\Gamma(c\kappa+1)}.$$

Therefore,

$$cov\left\{\hat{f}_b(x), \hat{f}_{b/c}(x)\right\} = n^{-1}b^{-1}\frac{1}{(1+c)^{\kappa}}\left(\frac{c}{1+c}\right)^{c\kappa+1}\frac{\Gamma((1+c)\kappa+1)f(x)}{\Gamma(\kappa+1)\Gamma(c\kappa+1)} + o\left(n^{-1}b^{-1}\right),$$

which yields the stated result. The proof for $(1-x)/b \to \kappa$ is similar. ■

A.2. Proof of Theorem 2

Again, only the proof when the beta kernel is employed is given. Because JLN-MBC uses a single smoothing parameter, the subscript b is also suppressed throughout.

A.2.1. Bias

The proof largely follows JLN. Let $\hat{\alpha}(x) = n^{-1}\sum_{i=1}^{n} K_{B(x/b+1,(1-x)/b+1)}(X_i)/\hat{f}(X_i)$ and rewrite $\tilde{f}_{JLN}(x)$ as

$$\tilde{f}_{JLN}(x) = \hat{f}(x)\hat{\alpha}(x) = f(x)\left\{1+\frac{\hat{f}(x)-f(x)}{f(x)}\right\}\left[1+\left\{\hat{\alpha}(x)-1\right\}\right].$$

Then,

$$E\left\{\tilde{f}_{JLN}(x)\right\} = f(x) + f(x)\left[E\left\{\frac{\hat{f}(x)-f(x)}{f(x)}\right\} + E\left\{\hat{\alpha}(x)-1\right\} + E\left\{\left(\frac{\hat{f}(x)-f(x)}{f(x)}\right)\left(\hat{\alpha}(x)-1\right)\right\}\right]. \quad (7)$$

We provide approximations to three expectations inside the brackets separately. First, we approximate $E\left\{\hat{\alpha}(x)-1\right\}$. By a geometric expansion,

$$\hat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)}\left[1-\left\{\frac{\hat{f}(X_i)-f(X_i)}{f(X_i)}\right\}+\left\{\frac{\hat{f}(X_i)-f(X_i)}{f(X_i)}\right\}^2\right] + o_p\left(b^2+n^{-1}b^{-1}\right),$$

where the remainder term is $o_p\left(b^2+n^{-1}b^{-1}\right) = o_p\left(b^2\right)$ by (A2). We now evaluate the expectation of the ith summand conditional on X_i. Using (1),

$$E\left\{\frac{\hat{f}(X_i)-f(X_i)}{f(X_i)}\middle| X_i\right\} = \frac{a_1(X_i)}{f(X_i)}b + \frac{a_2(X_i)}{f(X_i)}b^2 + o\left(b^2\right) \equiv h_1(X_i)b + h_2(X_i)b^2 + o\left(b^2\right), \quad (8)$$

where $a_1(x)$ and $a_2(x)$ are defined in the proof of Theorem 1. In addition, by (A2),

$$E\left[\left\{\frac{\hat{f}(X_i)-f(X_i)}{f(X_i)}\right\}^2\middle| X_i\right] = h_1^2(X_i)b^2 + o\left(b^2\right) + O\left(n^{-1}b^{-1}\right) = h_1^2(X_i)b^2 + o\left(b^2\right).$$

Hence, the conditional expectation of the ith summand can be approximated by

$$\frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)}\left[1-h_1(X_i)b+\left\{h_1^2(X_i)-h_2(X_i)\right\}b^2\right] + o\left(b^2\right).$$

It follows that

$$\begin{aligned} E\left\{\hat{\alpha}(x)\right\} &= E\left\{\frac{K_{B(x/b+1,(1-x)/b+1)}(X_1)}{f(X_1)}\right\} - E\left\{\frac{K_{B(x/b+1,(1-x)/b+1)}(X_1)}{f(X_1)} h_1(X_1)\right\} b \\ &\quad + E\left[\frac{K_{B(x/b+1,(1-x)/b+1)}(X_1)}{f(X_1)}\left\{h_1^2(X_1) - h_2(X_1)\right\}\right] b^2 + o\left(b^2\right) \\ &= 1 - E\left\{h_1(\theta_x)\right\} b + E\left\{h_1^2(\theta_x) - h_2(\theta_x)\right\} b^2 + o\left(b^2\right), \end{aligned}$$

where $\theta_x \overset{d}{=} B(x/b+1, (1-x)/b+1)$. By the property of the beta random variables and a Taylor expansion around $\theta_x = x$, we have

$$\begin{aligned} &E\left\{h_1(\theta_x)\right\} = h_1(x) + \left\{(1-2x) h_1'(x) + \frac{1}{2} x(1-x) h_1''(x)\right\} b + o(b), \\ &E\left\{h_1^2(\theta_x) - h_2(\theta_x)\right\} = h_1^2(x) - h_2(x) + O(b), \end{aligned}$$

so that

$$E\left\{\hat{\alpha}(x)\right\} = 1 - h_1(x) b + \left[h_1^2(x) - h_2(x) - \left\{(1-2x) h_1'(x) + \frac{1}{2} x(1-x) h_1''(x)\right\}\right] b^2 + o\left(b^2\right). \tag{9}$$

Second, it immediately follows from (8) that

$$E\left\{\frac{\hat{f}(x) - f(x)}{f(x)}\right\} = h_1(x) b + h_2(x) b^2 + o\left(b^2\right). \tag{10}$$

Third, by the Cauchy–Schwarz inequality and (A2),

$$E\left[\left\{\frac{\hat{f}(x) - f(x)}{f(x)}\right\}\left\{\hat{\alpha}(x) - 1\right\}\right] = -h_1^2(x) b^2 + o\left(b^2\right). \tag{11}$$

Finally, substituting (9)–(11) into (7) yields

$$\begin{aligned} E\left\{\tilde{f}_{\mathrm{JLN}}(x)\right\} &= f(x) - f(x)\left\{(1-2x) h_1'(x) + \frac{1}{2} x(1-x) h_1''(x)\right\} b^2 + o\left(b^2\right) \\ &= f(x) - f(x)\left[(1-2x)\left\{\frac{a_1(x)}{f(x)}\right\}' + \frac{1}{2} x(1-x)\left\{\frac{a_1(x)}{f(x)}\right\}''\right] b^2 + o\left(b^2\right). \end{aligned}$$

A.2.2. Variance

Observe that

$$\begin{aligned} \tilde{f}_{\mathrm{JLN}}(x) &= f(x)\left\{1 + \frac{\hat{f}(x) - f(x)}{f(x)}\right\} \frac{1}{n} \sum_{i=1}^{n} \frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)}\left[1 - \frac{\hat{f}(X_i) - f(X_i)}{f(X_i)} + o\left\{\frac{\hat{f}(X_i) - f(X_i)}{f(X_i)}\right\}\right] \\ &\sim f(x) \frac{1}{n} \sum_{i=1}^{n} \frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)}\left\{2 - \frac{\hat{f}(X_i)}{f(X_i)}\right\}, \end{aligned}$$

where

$$\begin{aligned} &\sum_{i=1}^{n} \frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)}\left\{2 - \frac{\hat{f}(X_i)}{f(X_i)}\right\} \\ &\quad = \sum_{i=1}^{n}\left\{\frac{2K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)} - \frac{1}{n} \sum_{j=1}^{n} \frac{K_{B(x/b+1,(1-x)/b+1)}\left(X_j\right) K_{B\left(X_j/b+1,\left(1-X_j\right)/b+1\right)}(X_i)}{f^2\left(X_j\right)}\right\} \\ &\quad \equiv \sum_{i=1}^{n}\left\{\varsigma_1(X_i) - \varsigma_2(X_i)\right\} \\ &\quad \equiv \sum_{i=1}^{n} \varsigma(X_i). \end{aligned}$$

We approximate $\varsigma_2(X_i)$ for a given data point X_i. The idea closely follows Section 3.2 of Jones and Henderson (2007b). For $X_i \neq 0, 1$,

$$\begin{aligned}
\varsigma_2(X_i) &\sim \varphi(X_i) \\
&\equiv E\left\{\left.\frac{K_{B(x/b+1,(1-x)/b+1)}\left(X_j\right)K_{B\left(X_j/b+1,\left(1-X_j\right)/b+1\right)}(X_i)}{f^2\left(X_j\right)}\right| X_i\right\} \\
&= \int_0^1 \frac{K_{B(x/b+1,(1-x)/b+1)}(u)\,K_{B(u/b+1,(1-u)/b+1)}(X_i)}{f(u)}\,\mathrm{d}u \\
&= \int_0^1 \frac{u^{\frac{x}{b}}(1-u)^{\frac{1-x}{b}}\exp\left\{\left(\frac{u}{b}\right)\log X_i + \left(\frac{1-u}{b}\right)\log(1-X_i) - \log B\left(\frac{u}{b}+1, \frac{1-u}{b}+1\right)\right\}}{B\left(\frac{x}{b}+1, \frac{1-x}{b}+1\right)f(u)}\,\mathrm{d}u.
\end{aligned}$$

By Stirling's approximation,

$$\begin{aligned}
\log B\left(\frac{u}{b}+1, \frac{1-u}{b}+1\right) &= \frac{1}{b}\left\{u\log u + (1-u)\log(1-u)\right\} + \frac{1}{2}\log b \\
&\quad + \frac{1}{2}\log 2\pi + \frac{1}{2}\log u(1-u) - \frac{b\left(u^2-u+1\right)}{12u(1-u)} + O\left(b^{3/2}\right),
\end{aligned}$$

and thus

$$\varphi(X_i) = \frac{b^{-1/2}}{\sqrt{2\pi}}\int_0^1 \psi(u)\exp\left\{\left(\frac{u}{b}\right)\log\frac{X_i}{u} + \left(\frac{1-u}{b}\right)\log\left(\frac{1-X_i}{1-u}\right) - \frac{b\left(u^2-u+1\right)}{12u(1-u)} + O\left(b^{3/2}\right)\right\}\mathrm{d}u,$$

where

$$\psi(u) = \frac{u^{\frac{x}{b}-\frac{1}{2}}(1-u)^{\frac{1-x}{b}-\frac{1}{2}}}{f(u)\,B\left(\frac{x}{b}+1, \frac{1-x}{b}+1\right)} = \frac{K_{B(x/b+1,(1-x)/b+1)}(u)}{\sqrt{u(1-u)}f(u)}.$$

By a change of variable $w = (X_i - u)/b^{1/2}$ and a Taylor expansion,

$$\begin{aligned}
\varphi(X_i) &= \frac{1}{\sqrt{2\pi}}\int_{-(1-X_i)/b^{1/2}}^{X_i/b^{1/2}} \psi\left(X_i - b^{1/2}w\right)\exp\left\{-\frac{w^2}{2X_i(1-X_i)}\right\} \\
&\quad\times\left\{1 - \frac{b^{1/2}(1-2X_i)w^3}{6X_i^2(1-X_i)^2} - \frac{b\left(3X_i^2-3X_i+1\right)w^4}{12X_i^3(1-X_i)^3} - \frac{b\left(X_i^2-X_i+1\right)}{12X_i(1-X_i)} + \frac{b(1-2X_i)^2w^6}{72X_i^4(1-X_i)^4} + O\left(b^{3/2}\right)\right\}\mathrm{d}w.
\end{aligned}$$

By another change of variable $v = w/\sqrt{X_i(1-X_i)}$,

$$\begin{aligned}
\varphi(X_i) &= \sqrt{X_i(1-X_i)}\int_{-\sqrt{(1-X_i)/X_i}/b^{1/2}}^{\sqrt{X_i/(1-X_i)}/b^{1/2}} \psi\left(X_i - b^{1/2}\sqrt{X_i(1-X_i)}v\right)\phi(v) \\
&\quad\times\left\{1 - \frac{b^{1/2}(1-2X_i)v^3}{6\sqrt{X_i(1-X_i)}} - \frac{b\left(3X_i^2-3X_i+1\right)v^4}{12X_i(1-X_i)} - \frac{b\left(X_i^2-X_i+1\right)}{12X_i(1-X_i)} + \frac{b(1-2X_i)^2v^6}{72X_i(1-X_i)} + O\left(b^{3/2}\right)\right\}\mathrm{d}v \\
&\sim \frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)} \times \begin{cases} 1 & \text{if } X_i/b \to \infty \text{ and } (1-X_i)/b \to \infty \\ \Phi\left(\sqrt{\kappa}\right) & \text{if } X_i/b \to \kappa \text{ or } (1-X_i)/b \to \kappa, \end{cases}
\end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution functions. Therefore,

$$\begin{aligned}
\varsigma(X_i) &\sim \varsigma_1(X_i) - \varphi(X_i) \\
&\sim \frac{K_{B(x/b+1,(1-x)/b+1)}(X_i)}{f(X_i)} \times \begin{cases} 1 & \text{if } X_i/b \to \infty \text{ and } (1-X_i)/b \to \infty \\ 2-\Phi\left(\sqrt{\kappa}\right) & \text{if } X_i/b \to \kappa \text{ or } (1-X_i)/b \to \kappa \end{cases} \\
&\equiv \zeta(X_i).
\end{aligned}$$

It follows that

$$var\left\{\tilde{f}_{\mathrm{JLN}}(x)\right\} \sim f^2(x)\frac{1}{n}var\left\{\zeta(X_1)\right\} = f^2(x)\left[\frac{1}{n}E\left\{\zeta^2(X_1)\right\} + O\left(n^{-1}\right)\right].$$

Pick $\delta = b^{1-\epsilon}$ where $\epsilon \in (0, 1/2)$. Following the trimming argument as in Chen (1999),

$$\begin{aligned} E\left\{\zeta^2(X_1)\right\} &= \int_0^1 \zeta^2(u) f(u)\,\mathrm{d}u \\ &= \int_0^\delta + \int_\delta^{1-\delta} + \int_{1-\delta}^1 \zeta^2(u) f(u)\,\mathrm{d}u \\ &= A_b(x) \int_0^1 \frac{K_{B(2x/b+1,2(1-x)/b+1)}(u)}{f(u)}\mathrm{d}u + O\left(b^{-\epsilon}\right), \end{aligned}$$

where

$$A_b(x) = \frac{B\left(\frac{2x}{b}+1, \frac{2(1-x)}{b}+1\right)}{B^2\left(\frac{x}{b}+1, \frac{1-x}{b}+1\right)} \sim \begin{cases} \dfrac{b^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} & \text{if } x/b \to \infty \text{ and } (1-x)/b \to \infty \\ \dfrac{\Gamma(2\kappa+1)\, b^{-1}}{2^{2\kappa+1}\Gamma^2(\kappa+1)} & \text{if } x/b \to \kappa \text{ or } (1-x)/b \to \kappa. \end{cases}$$

The stated result follows from recognizing that

$$\int_0^1 \frac{K_{B(2x/b+1,2(1-x)/b+1)}(u)}{f(u)}\mathrm{d}u = E\left\{f^{-1}(\vartheta_x)\right\} = f^{-1}(x) + O(b),$$

where $\vartheta_x \overset{d}{=} B(2x/b+1, 2(1-x)/b+1)$. ■

A.3. Formulae for beta-referenced smoothing parameters

The analytical expression of $\hat{b}_{\mathrm{TS},MB}$ is

$$\hat{b}_{\mathrm{TS},MB} = \left\{c^2(1-c)^2\lambda(c)\right\}^{2/9} \left\{\frac{B(\alpha,\beta)\, B(\alpha+9/2, \beta+9/2)}{16\sqrt{\pi} C_{\mathrm{TS},MB}(\alpha,\beta)}\right\}^{2/9} n^{-2/9},$$

where

$$\begin{aligned} C_{\mathrm{TS},MB}(\alpha,\beta) = \; & \frac{1}{\Gamma(2\alpha+2\beta+8)} \left\{\psi_1^2\Gamma(2\alpha)\Gamma(2\beta+8) - 2\psi_1\psi_2\Gamma(2\alpha+1)\Gamma(2\beta+7)\right. \\ & + \left(\psi_2^2 + 2\psi_1\psi_3\right)\Gamma(2\alpha+2)\Gamma(2\beta+6) - 2(\psi_1\psi_4+\psi_2\psi_3)\Gamma(2\alpha+3)\Gamma(2\beta+5) \\ & + \left(\psi_3^2 + 2\psi_1\psi_5 + 2\psi_2\psi_4\right)\Gamma(2\alpha+4)\Gamma(2\beta+4) \\ & - 2(\psi_2\psi_5+\psi_3\psi_4)\Gamma(2\alpha+5)\Gamma(2\beta+3) + \left(\psi_4^2 + 2\psi_3\psi_5\right)\Gamma(2\alpha+6)\Gamma(2\beta+2) \\ & \left. - 2\psi_4\psi_5\Gamma(2\alpha+7)\Gamma(2\beta+1) + \psi_5^2\Gamma(2\alpha+8)\Gamma(2\beta)\right\}, \\ \psi_1 = \; & \frac{1}{8}(\alpha-1)^2(\alpha-2)^2 - \frac{1}{3}(\alpha-1)(\alpha-2)(\alpha-3) - \frac{1}{8}(\alpha-1)(\alpha-2)(\alpha-3)(\alpha-4), \\ \psi_2 = \; & \frac{1}{2}(\alpha-1)^2(\alpha-2)(\beta-1) - \frac{1}{2}(\alpha-1)(\alpha-2) - \frac{1}{3}(\alpha-1)(\alpha-2)(\alpha-3) \\ & - (\alpha-1)(\alpha-2)(\beta-1) - \frac{1}{2}(\alpha-1)(\alpha-2)(\alpha-3)(\beta-1), \\ \psi_3 = \; & \frac{1}{2}(\alpha-1)^2(\beta-1)^2 - \frac{1}{2}(\alpha-1)(\alpha-2)(\beta-1)(\beta-2) - (\alpha-1)(\beta-1) \\ & - (\alpha-1)(\alpha-2)(\beta-1) - (\alpha-1)(\beta-1)(\beta-2), \\ \psi_4 = \; & \frac{1}{2}(\alpha-1)(\beta-1)^2(\beta-2) - \frac{1}{2}(\beta-1)(\beta-2) - \frac{1}{3}(\beta-1)(\beta-2)(\beta-3) \\ & - (\alpha-1)(\beta-1)(\beta-2) - \frac{1}{2}(\alpha-1)(\beta-1)(\beta-2)(\beta-3), \\ \psi_5 = \; & \frac{1}{8}(\beta-1)^2(\beta-2)^2 - \frac{1}{3}(\beta-1)(\beta-2)(\beta-3) - \frac{1}{8}(\beta-1)(\beta-2)(\beta-3)(\beta-4). \end{aligned}$$

Furthermore, $\hat{b}_{\mathrm{JLN},MB}$ takes the form of

$$\hat{b}_{\mathrm{JLN},MB} = \left\{\frac{B(\alpha,\beta)\, B(\alpha+9/2, \beta+9/2)}{4\sqrt{\pi} C_{\mathrm{JLN},MB}(\alpha,\beta)}\right\}^{2/9} n^{-2/9},$$

where

$$C_{\mathrm{JLN},MB}(\alpha,\beta) = \frac{1}{\Gamma(2\alpha+2\beta+8)}\Big[(\alpha-1)^2(\alpha-2)^2 \times \{\Gamma(2\alpha)\Gamma(2\beta+8)+2\Gamma(2\alpha+1)\Gamma(2\beta+7)+\Gamma(2\alpha+2)\Gamma(2\beta+6)\} + 2(\alpha-1)(\alpha-2)(\beta-1)(\beta-2) \times \{\Gamma(2\alpha+3)\Gamma(2\beta+5)+2\Gamma(2\alpha+4)\Gamma(2\beta+4)+\Gamma(2\alpha+5)\Gamma(2\beta+3)\} + (\beta-1)^2(\beta-2)^2\{\Gamma(2\alpha+6)\Gamma(2\beta+2)+2\Gamma(2\alpha+7)\Gamma(2\beta+1)+\Gamma(2\alpha+8)\Gamma(2\beta)\}\Big].$$

On the other hand, Jones and Henderson (2007b) define the beta-referenced smoothing parameter for $\hat{f}_{MB}(x)$ as

$$\hat{b}_{MB} = \arg\min_b \mathrm{AWMISE}\left\{\hat{f}_{MB}(x)\right\} = \arg\min_b \frac{b^2}{4}\int_0^1 \left\{x(1-x)g''(x)\right\}^2 v(x)\,\mathrm{d}x + \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi}}\int_0^1 \frac{g(x)}{\sqrt{x(1-x)}}v(x)\,\mathrm{d}x,$$

where the weighting function $v(x)$ is chosen as $v(x) = x^3(1-x)^3$ to ensure finiteness of integrals. It follows that $\hat{b}_{MB}$ can be expressed as

$$\hat{b}_{MB} = \left\{\frac{B(\alpha,\beta)B(\alpha+5/2,\beta+5/2)}{2\sqrt{\pi}C_{MB}(\alpha,\beta)}\right\}^{2/5} n^{-2/5},$$

where

$$C_{MB}(\alpha,\beta) = \frac{1}{\Gamma(2\alpha+2\beta+4)}\Big\{(\alpha-1)^2(\alpha-2)^2\Gamma(2\alpha)\Gamma(2\beta+4) - 4(\alpha-1)^2(\alpha-2)(\beta-1)\Gamma(2\alpha+1)\Gamma(2\beta+3) + 2(\alpha-1)(\beta-1)(3\alpha\beta-4\alpha-4\beta+6)\Gamma(2\alpha+2)\Gamma(2\beta+2) - 4(\alpha-1)(\beta-1)^2(\beta-2)\Gamma(2\alpha+3)\Gamma(2\beta+1) + (\beta-1)^2(\beta-2)^2\Gamma(2\alpha+4)\Gamma(2\beta)\Big\}. \quad \blacksquare$$

References

Azzalini, A., Bowman, A.W., 1990. A look at some data on the Old Faithful Geyser. Applied Statistics 39, 357–365.
Bouezmarni, T., Rolin, J.-M., 2003. Consistency of the beta kernel density function estimator. Canadian Journal of Statistics 31, 89–98.
Brown, B.M., Chen, S.X., 1999. Beta-Bernstein smoothing for regression curves with compact support. Scandinavian Journal of Statistics 26, 47–59.
Chen, S.X., 1999. Beta kernel estimators for density functions. Computational Statistics & Data Analysis 31, 131–145.
Chen, S.X., 2000a. Beta kernel smoothers for regression curves. Statistica Sinica 10, 73–91.
Chen, S.X., 2000b. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 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.
Gouriéroux, C., Monfort, A., 2006. (Non)Consistency of the beta kernel estimator for recovery rate distribution. CREST Discussion Paper No. 2006-31.
Gu, C., 2002. Smoothing Spline ANOVA Models. Springer-Verlag, New York.
Gustafsson, J., Hagmann, M., Nielsen, J.P., Scaillet, O., 2009. Local transformation kernel density estimation of loss distributions. Journal of Business & Economic Statistics 27, 161–175.
Hagmann, M., Renault, O., Scaillet, O., 2005. Estimation of recovery rate densities: Nonparametric and semi-parametric approaches versus industry practice. In: Altman, E.I., Resti, A., Sironi, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Publications, London, pp. 323–346.
Hagmann, M., Scaillet, O., 2007. Local multiplicative bias correction for asymmetric kernel density estimators. Journal of Econometrics 141, 213–249.
Hall, P., Park, B.Y., 2002. New methods for bias correction at endpoints and boundaries. Annals of Statistics 30, 1460–1479.
Hirukawa, M., 2006. A modified nonparametric prewhitened covariance estimator. Journal of Time Series Analysis 27, 441–476.
Hjort, N.L., Jones, M.C., 1996. Locally parametric nonparametric density estimation. Annals of Statistics 24, 1619–1647.
Jones, M.C., 1993. Simple boundary correction for kernel density estimation. Statistics and Computing 3, 135–146.
Jones, M.C., Foster, P.J., 1993. Generalized jackknifing and higher order kernels. Journal of Nonparametric Statistics 3, 81–94.
Jones, M.C., Foster, P.J., 1996. A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica 6, 1005–1013.
Jones, M.C., Henderson, D.A., 2007a. Kernel-type density estimation on the unit interval. Biometrika 24, 977–984.
Jones, M.C., Henderson, D.A., 2007b. Kernel-type density estimation on the unit interval. Statistics Group Technical Report 07/01, Open University.
Jones, M.C., Linton, O., Nielsen, J.P., 1995. A simple bias reduction method for density estimation. Biometrika 82, 327–338.
Jones, M.C., Signorini, D.F., 1997. A comparison of higher-order bias kernel density estimators. Journal of the American Statistical Association 92, 1063–1073.
Karunamuni, R.J., Alberts, T., 2005. A generalized reflection method of boundary correction in kernel density estimation. Canadian Journal of Statistics 33, 497–509.
Koshkin, G.M., 1988. Improved non-negative kernel estimate of a density. Theory of Probability and its Applications 33, 759–764.
Linton, O., Nielsen, J.P., 1994. A multiplicative bias reduction method for nonparametric regression. Statistics & Probability Letters 19, 181–187.
Loader, C., 1999. Local Regression and Likelihood. Springer-Verlag, New York.
Müller, H.-G., 1991. Smooth optimum kernel estimators near endpoints. Biometrika 78, 521–530.
Nielsen, J.P., 1998. Multiplicative bias correction in kernel hazard estimation. Scandinavian Journal of Statistics 25, 541–553.
Nielsen, J.P., Tanggaard, C., 2001. Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics 28, 675–698.
Renault, O., Scaillet, O., 2004. On the way to recovery: A nonparametric bias free estimation of recovery rate densities. Journal of Banking & Finance 28, 2915–2931.
Rudemo, M., 1991. Comment on "Transformations in Density Estimation" by M.P. Wand, J.S. Marron and D. Ruppert. Journal of the American Statistical Association 86, 353–354.
Scaillet, O., 2004. Density estimation using inverse and reciprocal inverse Gaussian kernels. Journal of Nonparametric Statistics 16, 217–226.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Stuetzle, W., Mittal, Y., 1979. Some comments on the asymptotic behavior of robust smoothers. In: Gasser, T., Rosenblatt, M. (Eds.), Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2–4, 1979. Springer-Verlag, Berlin, pp. 191–195.
Terrell, G.R., Scott, D.W., 1980. On improving convergence rates for nonnegative kernel density estimators. Annals of Statistics 8, 1160–1163.
Wasserman, L., 2006. All of Nonparametric Statistics. Springer-Verlag, New York.
Xiao, Z., Linton, O., 2002. A nonparametric prewhitened covariance estimator. Journal of Time Series Analysis 23, 215–250.
Zhang, S., Karunamuni, R.J., Jones, M.C., 1999. An improved estimator of the density function at the boundary. Journal of the American Statistical Association 94, 1231–1241.