

Density derivative estimation using asymmetric kernels

Benedikt Funke & Masayuki Hirukawa

To cite this article: Benedikt Funke & Masayuki Hirukawa (2024) Density derivative estimation using asymmetric kernels, Journal of Nonparametric Statistics, 36:4, 994-1017, DOI: [10.1080/10485252.2023.2291430](https://doi.org/10.1080/10485252.2023.2291430)

To link to this article: <https://doi.org/10.1080/10485252.2023.2291430>



Published online: 11 Dec 2023.



Submit your article to this journal [↗](#)



Article views: 217



View related articles [↗](#)



View Crossmark data [↗](#)



Density derivative estimation using asymmetric kernels

Benedikt Funke^a and Masayuki Hirukawa^b

^aInstitute for Insurance Studies, TH Köln – University of Applied Sciences, Köln, Germany; ^bFaculty of Economics, Ryukoku University, Kyoto, Japan

ABSTRACT

This paper studies the problem of estimating the first-order derivative of an unknown density with support on \mathbb{R}_+ or $[0, 1]$. Nonparametric density derivative estimators smoothed by the asymmetric, gamma and beta kernels are defined, and their convergence properties are explored. It is demonstrated that these estimators can attain the optimal convergence rate of the mean integrated squared error $n^{-4/7}$ when the underlying density has third-order smoothness. Superior finite-sample properties of the proposed estimators are confirmed in Monte Carlo simulations, and usefulness of the estimators is illustrated in two real data examples.

ARTICLE HISTORY

Received 5 October 2022
Accepted 30 November 2023

KEYWORDS

Beta kernel; density derivative estimation; digamma function; gamma kernel; trigamma function

JEL CLASSIFICATION CODES

C13; C14


MSCS 2010

62G08; 62G20

1. Introduction

The aim of this paper is to explore statistical properties of density derivative estimators using asymmetric kernels. Throughout we focus exclusively on univariate random variables supports of which have at least one boundary (e.g. \mathbb{R}_+ or $[0, 1]$). Indeed, there are many economic and financial variables values of which are either nonnegative or limited within a certain interval by construction. For example, wages, incomes, consumption expenditures, and insurance claims (or financial losses) are nonnegative. Variables expressed in the forms of shares or proportions, including expenditure and budget shares, unemployment rates, and default and recovery rates, are bounded from above and below. In addition, distributions of these nonnegative variables tend to be right-skewed and unimodal.

Asymmetric kernels are a viable device that can capture shapes of density curves with aforementioned features effectively. Because an asymmetric kernel is based on a probability density function ('pdf') having the same support as that of the curves, it is free of the boundary bias by construction. Furthermore, shapes of an asymmetric kernel vary across design points at which smoothing is made, signifying an adaptive adjustment in the degree of smoothing. In this context, asymmetric kernel smoothing exhibits similarities to, and can be considered a variation of, variable kernel (or bandwidth) methods. This adaptive

CONTACT Benedikt Funke  benedikt.funke@th-koeln.de  Institute for Insurance Studies, TH Köln – University of Applied Sciences, Gustav-Heinemann-Ufer 54, 50968 Köln, Germany

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

smoothing attribute proves particularly advantageous when the underlying distribution has both dense and sparse regions.

To make our discussion more concrete, let the random variable X be drawn from a univariate distribution having the pdf $f(\cdot)$. Because the support of X has a boundary, we employ an asymmetric kernel to estimate the pdf. Let $K_{\mathcal{J}(x,b)}(\cdot)$ be an asymmetric kernel indexed by \mathcal{J} that depends on the design point x and the smoothing parameter b . Given n i.i.d. observations $\{X_i\}_{i=1}^n$, the kernel density estimator ('KDE') of $f(x)$ smoothed by an asymmetric kernel $K_{\mathcal{J}}$ is defined as

$$\hat{f}_{\mathcal{J}}(x) = \frac{1}{n} \sum_{i=1}^n K_{\mathcal{J}(x,b)}(X_i). \quad (1)$$

This class of density estimators, introduced in the pioneered work by Chen (1999, 2000), is now popular in various practical applications.

Often practitioners' focus is not only on the density itself but also on its derivatives. From theoretical and practical standpoints, it is of particular importance and interest to estimate the first-order density derivative $f^{(1)}(x) = \partial f(x)/\partial x$ among density derivatives of all orders. A natural first-order derivative estimator implied by the KDE (1) is

$$\hat{f}_{\mathcal{J}}^{(1)}(x) := \frac{\partial}{\partial x} \hat{f}_{\mathcal{J}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x} K_{\mathcal{J}(x,b)}(X_i). \quad (2)$$

The kernel density derivative estimator of this class may be abbreviated as 'KDDE' hereinafter, whenever no confusions occur. While we deal exclusively with the first-order derivative estimation, extending the analysis in this paper to estimation of higher-order density derivatives is straightforward; actually, this task is more complicated, and these derivatives appear to be less relevant for practical issues.

Among all asymmetric kernels proposed so far, our main focuses are on the beta and gamma kernels proposed by Chen (1999, 2000), respectively. Our preference is based on (i) empirical relevance of the kernels in economics and finance and (ii) analytical tractability. For (i), Hirukawa (2018, Table 1.1) presents a list of applications of these kernels to estimation and testing problems. For (ii), unlike standard symmetric kernels, exploring statistical properties of nonparametric estimators smoothed by asymmetric kernels relies on kernel-specific and thus diversified approaches. The density derivative estimators are not an exception. The gamma function is a building block for the beta and gamma kernels. There is rich literature on approximation techniques to the gamma and related functions as they are actively studied.

Contributions of this paper can be summarised in two respects. First, this paper studies convergence properties of asymmetric KDDEs and presents conclusive bias and variance expansions for the first time in the literature, to the best of our knowledge. It is demonstrated that each of the gamma and beta KDDEs has a usual $O(b)$ bias, whereas the variance convergence of each estimator is $O(n^{-1}b^{-3/2})$ for the interior region and $O(n^{-1}b^{-3})$ for the boundary region. As in the case of derivative estimators smoothed by symmetric kernels and thus as expected, the rates of the variance are slower than those of the corresponding KDE (1); invoke that the variance of a KDE is $O(n^{-1}b^{-1/2})$ for the interior region and $O(n^{-1}b^{-1})$ for the boundary region. Second, a few new techniques are employed for

the bias and variance expansions of the gamma and beta KDDEs. These expansions involve evaluations of moments of log-transformed gamma and beta random variables. We address this issue by exploring a moment-generating function-based approach inspired by Ye and Chen (2017). Finally, combining this approach with double-inequalities on digamma and trigamma functions by Gordon (1994) helps us determine expressions of the leading bias and variance terms of the estimators.

There are potentially many applications of the KDDEs. First, these can be used to evaluate peaks, valleys and inflection points of the underlying density. A direct application arises from the field of actuarial science. Understanding the mode of an actuarial loss distribution is of particular significance for insurance companies, as the mode, representing the most probable payout amount of an insurance obligation, plays an essential role in the calculation of an appropriate premium. Second, the KDDEs can be employed to estimate density scores, which are a key ingredient for (indirect) average derivative estimation of nonparametric regression curves. Third, there are several important economic applications. The density derivative is a key component of the Engel curve, which is essential for empirically verifying the law of demand. It also plays a significant role in the estimation of elasticities; see Hildenbrand and Hildenbrand (1986) for a comprehensive understanding of these concepts, their motivations, and the economic conclusions drawn from them. Interested readers may also consult Härdle et al. (1990) for further applications.

The remainder of this paper is organised as follows. In Section 2, the gamma and beta KDDEs are defined and their convergence properties are explored under some regularity conditions. Section 3 conducts Monte Carlo simulations to investigate finite-sample properties of the KDDEs in comparison with several competing density derivative estimators. In Section 4, we apply the gamma and beta KDDEs to two real datasets to illustrate their usefulness. Section 5 concludes. All proofs are provided in the Appendix.

This paper adopts the following notational conventions: for $a > 0$, $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$ is the gamma function; for $p, q > 0$, $B(p, q) = \int_0^1 y^{p-1} (1-y)^{q-1} dy$ denotes the beta function; $\Psi(x) = \partial \log \Gamma(x) / \partial x = \Gamma^{(1)}(x) / \Gamma(x)$ and $\Psi_1(x) = \Psi^{(1)}(x) = \partial \Psi(x) / \partial x$ are the digamma and trigamma functions, respectively; $\mathbf{1}\{\cdot\}$ signifies an indicator function; and $(r)_m = \Gamma(r+m) / \Gamma(r) = r(r+1) \cdots (r+m-1)$ represents the Pochhammer symbol. The expression ' $X \stackrel{d}{=} Y$ ' reads 'A random variable X obeys the distribution Y '. The expression ' $X_n \sim Y_n$ ' is used whenever $X_n / Y_n \rightarrow 1$ as $n \rightarrow \infty$. Lastly, in order to describe different asymptotic properties of an asymmetric kernel estimator across positions of the design point $x \in \mathbb{R}_+$ ($x \in [0, 1]$), we denote by 'interior x ' and 'boundary x ' a design point x that satisfies $x/b \rightarrow \infty$ ($x/b \rightarrow \infty$ and $(1-x)/b \rightarrow \infty$) and $x/b \rightarrow \kappa$ (x/b or $(1-x)/b \rightarrow \kappa$) for some $\kappa \in (0, \infty)$ as $n \rightarrow \infty$, respectively.

2. Estimators and their large-Sample properties

2.1. The estimators

Due to their popularity in a wide range of empirical studies and analytical tractability, hereinafter we confine our discussion on kernels chosen for the KDDE (2) to the gamma and beta ones. Extending the scope of asymmetric kernels to other distribution-specific ones is, in principle, feasible but falls outside the primary focus of this paper.

The first estimator employs the gamma kernel ($\mathcal{J} = G$)

$$K_{G(x,b)}(u) = \frac{u^{x/b} \exp(-u/b)}{b^{x/b+1} \Gamma(x/b+1)} \mathbf{1}\{u \geq 0\}$$

for $x \in \mathbb{R}_+$ by Chen (2000). The gamma KDDE is defined as

$$\hat{f}_G^{(1)}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x} K_{G(x,b)}(X_i) := \frac{1}{n} \sum_{i=1}^n L_{G(x,b)}(X_i) K_{G(x,b)}(X_i),$$

where

$$L_{G(x,b)}(u) = \frac{1}{b} \left\{ \log u - \log b - \Psi\left(\frac{x}{b} + 1\right) \right\}.$$

The second estimator is built on the beta kernel ($\mathcal{J} = B$)

$$K_{B(x,b)}(u) = \frac{u^{x/b} (1-u)^{(1-x)/b}}{B\{x/b+1, (1-x)/b+1\}} \mathbf{1}\{u \in [0, 1]\}$$

for $x \in [0, 1]$ by Chen (1999). The beta KDDE is given by

$$\hat{f}_B^{(1)}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x} K_{B(x,b)}(X_i) := \frac{1}{n} \sum_{i=1}^n L_{B(x,b)}(X_i) K_{B(x,b)}(X_i),$$

where

$$L_{B(x,b)}(u) = \frac{1}{b} \left\{ \log\left(\frac{u}{1-u}\right) - \Psi\left(\frac{x}{b} + 1\right) + \Psi\left(\frac{1-x}{b} + 1\right) \right\}.$$

Before proceeding, we inspect shapes of two weight functions $\partial K_{G(x,b)}(u)/\partial x$ and $\partial K_{B(x,b)}(u)/\partial x$ used for density derivative estimation. Figure 1 presents plots of the functions evaluated at five different design points; more specifically, $\partial K_{G(x,b)}(u)/\partial x$ for $x =$

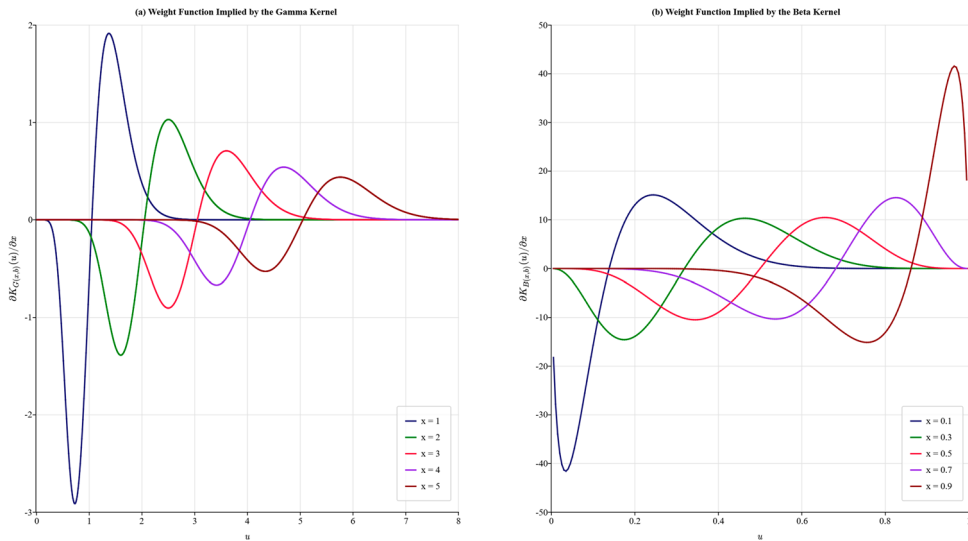


Figure 1. Shapes of Weight Functions with a Fixed Smoothing Parameter ($b = 0.1$) at Various Design Points.

1, 2, 3, 4, 5 and $\partial K_{B(x,b)}(u)/\partial x$ for $x = 0.1, 0.3, 0.5, 0.7, 0.9$ are presented. It is worth emphasising that for all plotted functions, the smoothing parameter value is fixed at $b = 0.1$.

A few remarks are in order. First, each plot produces an asymmetric S-shape, and the sign changes from negative to positive around (but not exactly at) the design point, in general. An exception is $\partial K_{B(x,b)}(u)/\partial x$ for $x = 1/2$; in this case the sign change occurs exactly at $1/2$. To make this argument more precise, let $u_{\mathcal{J}}^*$ be the interior solution to $\partial K_{\mathcal{J}(x,b)}(u)/\partial x = 0$ for a given interior design point x . Then,

$$u_{\mathcal{J}}^* = \begin{cases} x + b/2 + O(b^2) & \text{for } \mathcal{J} = G \text{ and } x \in (0, \infty) \\ x + (x - 1/2)b + O(b^2) & \text{for } \mathcal{J} = B \text{ and } x \in (0, 1/2) \cup (1/2, 1) \\ x & \text{for } \mathcal{J} = B \text{ and } x = 1/2 \end{cases} \quad (3)$$

The derivation of (3) is documented in the Appendix. Moreover, the weight functions implied by the gamma and beta kernels have a sharp contrast to the one implied by the normal kernel $K_N(u) = \exp(-u^2/2)/\sqrt{2\pi}$. The weight function $\partial K_N(u)/\partial u = -u \exp(-u^2/2)/\sqrt{2\pi}$ generates a reverse S-shape and is symmetric with respect to the origin.

Second, shapes of the weight functions vary across the design points at which smoothing is made, and they become flatter as the design point moves away from the boundary for the gamma case and from both boundaries for the beta case. Hence, it can be observed that when smoothing is made at a boundary, the weighting functions put their maximum weights (in magnitude) on the boundary. An eminent feature of an asymmetric kernel is that the kernel shape changes according to the location of the design point. The weight functions do inherit the property of altering the amount of smoothing in a locally adaptive manner by using a single smoothing parameter. This property makes them much more appealing in empirical work, especially when data is unequally spaced.

2.2. Regularity conditions

To deliver convergence properties of each KDDE, we impose the following common regularity conditions.

Assumption 2.1: The *i.i.d.* random sample $\{X_i\}_{i=1}^n$ is drawn from a univariate distribution having support either on \mathbb{R}_+ (for the gamma estimator) or on $[0, 1]$ (for the beta estimator).

Assumption 2.2: The third-order derivative of $f(\cdot)$ is uniformly bounded and Hölder-continuous of order $\varsigma \in (0, 1]$ over the entire support.

Assumption 2.3: The smoothing parameter $b(= b_n > 0)$ satisfies $b + (nb^3)^{-1} \rightarrow 0$ as $n \rightarrow \infty$.

Random sampling in Assumption 2.1 makes it easier to derive the dominant terms in bias and variance of the estimators. Observe that supports of the kernel and underlying distribution match. Therefore, no boundary bias occurs.

Boundedness and Hölder-continuity of the third-order density derivative $f^{(3)}(\cdot)$ in Assumption 2.2 are key requirements for the bias expansion of the estimators. The former

helps determine the leading bias term, whereas the latter controls the order of magnitude in the remainder term in the bias expansion. Boundedness and (Hölder or some other class of) continuity of the second-order derivative are typically assumed for density estimation. Now that our target quantity is the first-order density derivative, these regularity conditions are consistently transferred to the third-order derivative. Notice that Hölder-continuity of $f^{(3)}(\cdot)$ implies that there is a constant $C \in (0, \infty)$ such that

$$\left| f^{(3)}(u) - f^{(3)}(v) \right| \leq C |u - v|^\varsigma, \quad \forall u, v. \quad (4)$$

The condition $(nb^3)^{-1} \rightarrow 0$ in Assumption 2.3 is stronger than $(nb)^{-1} \rightarrow 0$, which is typically imposed for the variance expansion of KDEs smoothed by asymmetric kernels; see Chen (1999, 2000) for more details. It will be shown shortly that the MSE-optimal smoothing parameter for each KDDE becomes $b^* = O(n^{-2/7})$ for interior x and $b^\dagger = O(n^{-1/5})$ for boundary x ; these convergence rates are indeed within the required range.

2.3. Convergence properties of the estimators

2.3.1. Local property

Bias-Variance Tradeoff. Asymptotic properties of the gamma and beta KDDEs are explored below. We start from delivering their bias-variance tradeoffs in the following theorem.

Theorem 2.1: *If Assumptions 2.1–2.3 hold, then $\hat{f}_{\mathcal{J}}^{(1)}(x)$ for $\mathcal{J} \in \{G, B\}$ admits the following expansions as $n \rightarrow \infty$.*

(i) *Bias:*

$$E \left\{ \hat{f}_{\mathcal{J}}^{(1)}(x) \right\} := f^{(1)}(x) + \mathcal{B}_{\mathcal{J}}(x, f) b + o(b),$$

where

$$\mathcal{B}_{\mathcal{J}}(x, f) = \begin{cases} \frac{1}{2} \{ 3f^{(2)}(x) + xf^{(3)}(x) \} & \text{for } \mathcal{J} = G \\ -2f^{(1)}(x) + \frac{3}{2}(1-2x)f^{(2)}(x) + \frac{1}{2}x(1-x)f^{(3)}(x) & \text{for } \mathcal{J} = B \end{cases}.$$

(ii) *Variance:*

$$\text{Var} \left\{ \hat{f}_{\mathcal{J}}^{(1)}(x) \right\} := \begin{cases} \frac{1}{nb^{3/2}} \mathcal{V}_{\mathcal{J}}(x) f(x) + o(n^{-1}b^{-3/2}) & \text{for interior } x \\ \frac{1}{nb^3} \frac{\Gamma(2\kappa+1)f(x)}{\kappa 2^{2\kappa+2} \Gamma^2(\kappa+1)} + o(n^{-1}b^{-3}) & \text{for boundary } x \end{cases},$$

where

$$\mathcal{V}_{\mathcal{J}}(x) = \begin{cases} (4\sqrt{\pi}x^{3/2})^{-1} & \text{for } \mathcal{J} = G \\ [4\sqrt{\pi}\{x(1-x)\}^{3/2}]^{-1} & \text{for } \mathcal{J} = B \end{cases}.$$

Before discussing the above theorem in detail, we describe a few new strategies taken for its proof in the Appendix. The bias and variance expansions in the theorem involve

evaluations of moments of log-transformed gamma and beta random variables. We tackle this issue by developing a moment-generating function-based approach inspired by Ye and Chen (2017). Finally, combining this approach with Gordon's (1994) double-inequalities on digamma and trigamma functions yields the leading bias and variance terms.

Theorem 2.1 implies that the pointwise bias convergence of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ is $O(b)$, just like that of the corresponding KDE. Observe that while $\mathcal{B}_G(x, f)$ is free of $f^{(1)}(x)$, $\mathcal{B}_B(x, f)$ involves this quantity. This is because the moments $E\{L_{B(x, b)}(\theta_x)(\theta_x - x)^j\}$ for $\theta_x \stackrel{d}{=} \text{Beta}\{x/b + 1, (1 - x)/b + 1\}$ are rational functions of (x, b) , whereas the moments $E\{L_{G(x, b)}(\xi_x)(\xi_x - x)^j\}$ for $\xi_x \stackrel{d}{=} G(x/b + 1, b)$ are polynomials of (x, b) ; see Appendix for more details. In particular, $E\{L_{B(x, b)}(\theta_x)(\theta_x - x)\}$ is not exactly 1 unlike $E\{L_{G(x, b)}(\xi_x)(\xi_x - x)\}$, and as a consequence, the term including $f^{(1)}(x)$ is left over in $\mathcal{B}_B(x, f)$.

Moreover, the pointwise variance convergence of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ slows down from that of the corresponding KDE; it is $O(n^{-1}b^{-3/2})$ and $O(n^{-1}b^{-3})$ for interior and boundary x , respectively, to be more precise. It follows from the bias and variance expansions that $\hat{f}_{\mathcal{J}}^{(1)}(x)$ is consistent for $f^{(1)}(x)$ under the set of assumptions.

Furthermore, Chen (2000) calls the quantity $\mathcal{V}_{\mathcal{J}}(x)$ the variance coefficient. Variance coefficients for the gamma and beta KDEs are known to be $(2\sqrt{\pi}\sqrt{x})^{-1}$ and $\{2\sqrt{\pi}\sqrt{x(1-x)}\}^{-1}$, respectively. Observe that each variance coefficient decreases as the design point x moves away from the boundaries. For the gamma and beta KDDEs, this feature is preserved in that their variance coefficients are $(4\sqrt{\pi}x^{3/2})^{-1}$ and $[4\sqrt{\pi}\{x(1-x)\}^{3/2}]^{-1}$, respectively. It can be inferred from this property that the gamma KDDE works well for estimating the density derivative over sparse areas with fewer observations (e.g. the right-tail part), as in the gamma KDE. The sparse areas are typically observed in the applications of asymmetric kernel functions mentioned in Section 1, which emphasises the practical relevance of our approach once again.

Mean Squared Error ('MSE'). It follows from Theorem 2.1 that the MSE for interior x of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ is given by

$$\text{MSE}\left\{\hat{f}_{\mathcal{J}}^{(1)}(x)\right\} = b^2 \left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2 + \frac{1}{nb^{3/2}} \mathcal{V}_{\mathcal{J}}(x) f(x) + o\left(b^2 + \frac{1}{nb^{3/2}}\right). \quad (5)$$

The smoothing parameter value that minimises the two leading terms on the right-hand side of (5) is

$$b_{\mathcal{J}}^* = \left[\frac{3\mathcal{V}_{\mathcal{J}}(x)f(x)}{4\left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2} \right]^{2/7} n^{-2/7}.$$

Observe that the MSE-optimal smoothing parameter $b_{\mathcal{J}}^* = O(n^{-2/7}) = O(h^{*2})$, where h^* is the MSE-optimal bandwidth for KDDEs using nonnegative symmetric kernels (see, e.g. Wand and Jones 1995). Therefore, when best implemented, the approximation to the MSE becomes

$$\text{MSE}^*\left\{\hat{f}_{\mathcal{J}}^{(1)}(x)\right\} \sim 7 \left[\frac{1}{3} \left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2 \right]^{3/7} \left\{ \frac{1}{4} \mathcal{V}_{\mathcal{J}}(x) f(x) \right\}^{4/7} n^{-4/7}.$$

Observe that the optimal MSE of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ for interior x becomes $O(n^{-4/7})$, which is also the optimal convergence rate in the MSE of the nonnegative symmetric KDDE. On the other

hand, for boundary x , $MSE\{\hat{f}_{\mathcal{J}}^{(1)}(x)\} = O(b^2 + n^{-1}b^{-3})$, which yields the MSE-optimal smoothing parameter $b_{\mathcal{J}}^{\dagger} = O(n^{-1/5})$ and the optimal MSE of $O(n^{-2/5})$.

2.3.2. Global property

The inferior rate in the optimal MSE of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ for boundary x does not affect its global property. If $\int \{\mathcal{B}_{\mathcal{J}}(x, f)\}^2 dx$ and $\int \mathcal{V}_{\mathcal{J}}(x) f(x) dx$ are both finite, then applying the trimming argument in Chen (1999, 2000) yields the mean integrated squared error ('MISE') of $\hat{f}_{\mathcal{J}}^{(1)}(x)$ as

$$\begin{aligned} MISE\left\{\hat{f}_{\mathcal{J}}^{(1)}(x)\right\} \\ = b^2 \int \left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2 dx + \frac{1}{nb^{3/2}} \int \mathcal{V}_{\mathcal{J}}(x) f(x) dx + o\left(b^2 + \frac{1}{nb^{3/2}}\right). \end{aligned} \quad (6)$$

The smoothing parameter value that minimises the two leading terms on the right-hand side of (6) is

$$b_{\mathcal{J}}^{**} = \left[\frac{3 \int \mathcal{V}_{\mathcal{J}}(x) f(x) dx}{4 \int \left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2 dx} \right]^{2/7} n^{-2/7}.$$

Therefore, when best implemented, the approximation to the MISE becomes

$$MISE^{**}\left\{\hat{f}_{\mathcal{J}}^{(1)}(x)\right\} \sim 7 \left[\frac{1}{3} \int \left\{\mathcal{B}_{\mathcal{J}}(x, f)\right\}^2 dx \right]^{3/7} \left\{ \frac{1}{4} \int \mathcal{V}_{\mathcal{J}}(x) f(x) dx \right\}^{4/7} n^{-4/7}.$$

It is worth remarking that the rate $n^{-4/7}$ coincides with the optimal convergence rate of the MISE for a KDDE under third-order smoothness in the underlying density f that is implied by Theorem 1 of Müller and Gasser (1979) and Theorem of Stone (1980).

3. Finite-sample performance

In the Monte Carlo study below, finite-sample performances of the asymmetric KDDEs are compared with those of other competing density derivative estimators. For each estimator, how to select the tuning parameter is an important practical issue. We start our analysis from the 'oracle' method, where a tuning parameter value that minimises a performance measure is taken as the optimal one. While assessing an estimator with the oracle tuning parameter value plugged in can be interpreted as examining its performance in the best-case scenario, this method is infeasible in reality. To make the asymmetric KDDEs fully operational, we also examine a cross-validation ('CV') method.

3.1. Oracle method

3.1.1. Case #1: gamma KDDE

Our simulation study starts from appraising the gamma KDDE. Three density functions with support on \mathbb{R}_+ are examined. The first case is the *Weibull distribution* with density

and derivative

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, \quad (k, \lambda) = (2, 1) \Rightarrow f^{(1)}(x) = (-4x^2 + 2) \exp(-x^2).$$

The second case is the *beta prime distribution* (also known as the *beta distribution of the second kind*) with density and derivative

$$f(x) = \frac{1}{B(\alpha, \beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}, \quad (\alpha, \beta) = (3, 3) \Rightarrow f^{(1)}(x) = \frac{-120x^2 + 60x}{(1+x)^7}.$$

The third case is the *Lomax distribution* (also known as the *Pareto distribution of the second kind*) by Lomax (1954) with density and derivative

$$f(x) = \frac{\alpha}{\lambda} \left(1 + \frac{x}{\lambda}\right)^{-(\alpha+1)}, \quad (\alpha, \lambda) = (1, 1) \Rightarrow f^{(1)}(x) = -\frac{2}{(1+x)^3}.$$

Shapes of these densities and derivatives can be found in Figure 2. Observe that the Weibull distribution has a thin tail with an exponential decay, whereas two others have thick tails with polynomial decays. For each distribution, 1000 data sets of sample size $n \in \{250, 500\}$ are simulated.

The density derivative estimators compared are: (i) the KDDE using the normal kernel [KDDE-N]; (ii) the KDDEs obtained via the second- and third-order local polynomial density estimators using the triangular kernel $K_T(u) = (1 - |u|)\mathbf{1}\{|u| \leq 1\}$ by Cattaneo et al. (2020) [LPDE2-T, LPDE3-T]; and (iii) the KDDE using the gamma kernel [KDDE-G]. The choice of the normal kernel is based on the recommendation by Härdle et al. (1990), who study the CV bandwidth choice for density derivative estimation and report superior performance of this kernel.

The idea of local polynomial density estimation by Cattaneo et al. (2020) comes from local polynomial regression estimation. Local polynomial regression-based techniques

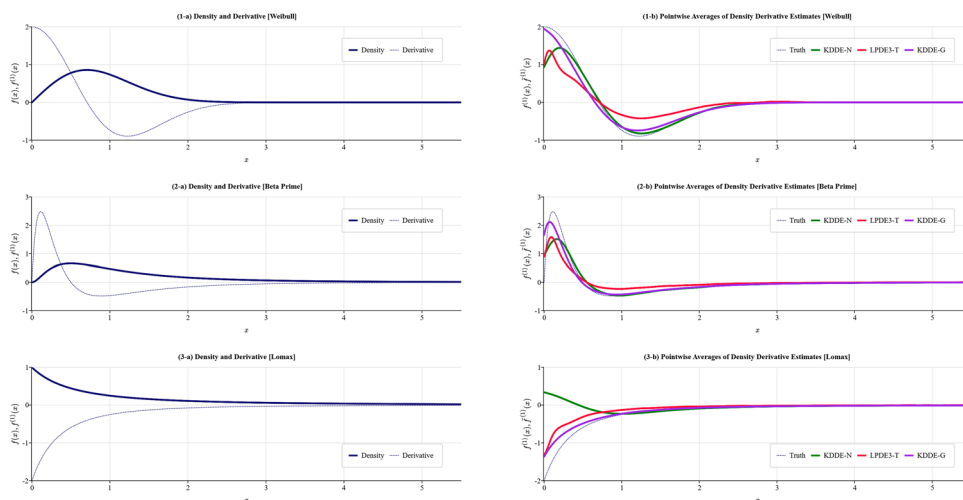


Figure 2. True Densities, True Density Derivatives and Pointwise Averages of Density Derivative Estimates (Case #1, $n = 500$).

typically require prebinning of the data, which results in choosing an additional tuning parameter (i.e. the bin width) and decreasing the sample size at the stage of density estimation. By smoothing out the empirical cumulative distribution function ('cdf') directly, Cattaneo et al. (2020) preserve the original sample size and avoid choosing an extra tuning parameter. The KDDE can be obtained from their estimation procedure in the following manner. Let $\hat{F}(x) := (1/n) \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$ be the empirical cdf, which is a consistent estimator of the unknown cdf $F(x) = \int_0^x f(u) du$. Then, using n data points $\{\hat{F}(X_i)\}_{i=1}^n$ as regressands, we run a local polynomial regression of order p (≥ 1). Also let $\mathbf{r}_p(u) := (1, u, u^2, \dots, u^p)^\top$ and $\mathbf{e}_2 := (0, 0, 1, 0, \dots, 0)^\top$ be a p th-order polynomial expansion and the third $(p+1)$ -dimensional unit vector, respectively. Finally, $f^{(1)}(x)$ can be estimated by $\hat{f}_{LP}^{(1)}(x) := \mathbf{e}_2^\top \hat{\beta}(x)$, where

$$\hat{\beta}(x) = \arg \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left\{ \hat{F}(X_i) - [\mathbf{r}_p(X_i - x)]^\top \mathbf{b} \right\}^2 K_T \left(\frac{X_i - x}{h} \right),$$

and h is the bandwidth. We consider two orders of polynomials, namely, $p = 2, 3$, because of the difference in bias convergences of $\hat{f}_{LP}^{(1)}(x)$ between the polynomial orders. To be more precise, $\hat{f}_{LP}^{(1)}(x)$ for $p = 2$ has $O(h^2)$ and $O(h)$ biases for interior and boundary x , respectively, whereas the bias convergence of the one for $p = 3$ is $O(h^2)$ everywhere. The variance convergence of the estimator for $p = 2, 3$ is $O(n^{-1}h^{-3})$; see the supplemental appendix of Cattaneo et al. (2020) for more details.

As performance measures of an estimator $\bar{f}^{(1)}(\cdot)$, we adopt the root integrated squared error ('RISE')

$$RISE(\bar{f}^{(1)}) = \sqrt{\int_0^\infty \{\bar{f}^{(1)}(x) - f^{(1)}(x)\}^2 dx}$$

and the integrated absolute deviation ('IAD')

$$IAD(\bar{f}^{(1)}) = \int_0^\infty |\bar{f}^{(1)}(x) - f^{(1)}(x)| dx,$$

where each integral is approximated by the trapezoidal rule on an equally-spaced grid of 551 points over the interval $[0.0, 5.5]$.

The value of a tuning parameter $a \in \{h, b\}$ is chosen via the oracle method. For each simulated sample the RISE is computed over a grid of a , and then a minimiser of the RISE is obtained. The minimiser is found on an equally-spaced grid of 100 points over the interval $[0.005, 0.500]$.

Table 1 reports averages and standard deviations of two performance measures and tuning parameter values over 1000 Monte Carlo samples. It can be immediately found that for each distribution, KDDE-G outperforms all other estimators. Moreover, its tuning parameter values are by far the smallest among all (and this tendency continues to hold for the beta KDDE in Case #2 below). This is not surprising, because the MISE-optimal smoothing parameter of two asymmetric KDDEs $b_{\mathcal{J}}^{**} = O(n^{-2/7})$ tends to be smaller than the MISE-optimal bandwidth for a symmetric KDDE $h^{**} = O(n^{-1/7})$. While KDDE-N performs much better than two LPDEs for Weibull and beta prime distributions, it performs poorly for the Lomax distribution, failing to capture concavity of the curve. There is slight

Table 1. Monte Carlo Results (Oracle: Case #1).

Estimator	$n = 250$			$n = 500$		
	RISE	IAD	h or b	RISE	IAD	h or b
Weibull						
KDDE-N	0.4504 (0.0756)	0.4866 (0.0999)	0.1858 (0.0264)	0.4008 (0.0591)	0.4206 (0.0776)	0.1628 (0.0220)
LPDE2-T	0.7956 (0.0956)	0.9786 (0.0986)	0.3866 (0.0781)	0.7618 (0.0689)	0.9419 (0.0718)	0.3344 (0.0758)
LPDE3-T	0.7782 (0.1443)	0.9569 (0.1268)	0.4314 (0.0815)	0.7065 (0.1015)	0.8986 (0.0881)	0.4125 (0.0846)
KDDE-G	0.3867 (0.1061)	0.4296 (0.1206)	0.0656 (0.0133)	0.3255 (0.0804)	0.3612 (0.0894)	0.0532 (0.0105)
Beta Prime						
KDDE-N	0.5519 (0.0874)	0.7395 (0.1032)	0.1647 (0.0232)	0.4881 (0.0709)	0.6546 (0.0873)	0.1401 (0.0161)
LPDE2-T	0.7843 (0.1261)	0.9856 (0.1440)	0.3430 (0.0870)	0.7365 (0.0945)	0.9091 (0.1223)	0.2939 (0.0744)
LPDE3-T	0.6782 (0.1987)	0.8424 (0.1726)	0.4299 (0.0892)	0.5907 (0.1396)	0.7303 (0.1213)	0.4030 (0.0851)
KDDE-G	0.4497 (0.1327)	0.4442 (0.1057)	0.0560 (0.0162)	0.3979 (0.1088)	0.3834 (0.0854)	0.0441 (0.0109)
Lomax						
KDDE-N	0.9875 (0.0184)	0.8223 (0.0383)	0.5000 (0.0000)	0.9869 (0.0126)	0.8054 (0.0277)	0.5000 (0.0000)
LPDE2-T	0.5343 (0.1323)	0.6678 (0.1264)	0.4210 (0.0846)	0.5043 (0.1023)	0.6069 (0.0973)	0.3829 (0.0919)
LPDE3-T	0.5302 (0.1862)	0.6507 (0.1659)	0.4199 (0.1325)	0.4727 (0.1560)	0.5601 (0.1388)	0.4230 (0.1009)
KDDE-G	0.3263 (0.1459)	0.2762 (0.0969)	0.1602 (0.0612)	0.2891 (0.1258)	0.2354 (0.0763)	0.1291 (0.0441)

Note: Simulation averages and standard deviations (in parentheses) of two performance measures ('RISE' and 'IAD') and tuning parameter values (' h ' or ' b ') are reported.

advantage of LPDE3-T over LPDE2-T, which appears to reflect better bias properties of the former. For this reason, we do not include LPDE2-T in Figure 2.

Figure 2 exhibits plots of pointwise averages of KDDE-N, LPDE3-T and KDDE-G for $n = 500$ over 1000 Monte Carlo samples. Although no significant differences can be found when $x > 3$ for all distributions, there are substantial differences over $x \in [0, 3]$. The density derivative for the Weibull distribution is initially decreasing, and KDDE-G alone successfully traces the shape. For the beta prime case, the density derivative starts from the origin and has a sharp mode in the vicinity of the origin. While all three estimators fail to start from the origin, KDDE-G captures the curvature better than two others.

KDDE-N exhibits erratic behaviour for the Lomax distribution; more specifically, the plot of KDDE-N is initially decreasing and then increasing, whereas the truth is monotone increasing everywhere. As discussed in Section 2.1, the weight function for KDDE-N $\partial K_N(u)/\partial u = -u \exp(-u^2/2)/\sqrt{2\pi}$ exhibits a reverse S-shape, unlike $\partial K_{G(x,b)}(u)/\partial x$ (or $\partial K_{B(x,b)}(u)/\partial x$). When a density $f(x)$ has support on \mathbb{R} , it dies out as $x \rightarrow \pm\infty$. KDDE-N works well in this scenario, because the corresponding density derivative takes a reverse S-shape. However, a density with support on \mathbb{R}_+ or $[0, 1]$ is allowed to be monotone decreasing everywhere, and the corresponding density derivative becomes monotone increasing in this case. Moreover, the amount of smoothing by symmetric kernels with a single bandwidth parameter is fixed everywhere. Under these conditions KDDE-N tends to mistakenly assume a reverse S-shape of the density derivative and then produces a reverse

S-curve with a fictitious decreasing part. Also observe that the oracle bandwidth for this estimator is persistently chosen at the upper bound 0.5. Such an oversmoothing bandwidth reflects that KDDE-N focuses more on smoothing out the flat part on the right tail. Notice that although KDDE-G is built on an S-shape weight function, it can successfully handle such an irregular shape by varying shapes of the function across design points.

3.1.2. Case #2: beta KDDE

Now we assess the beta KDDE. Again we investigate three density functions with support on $[0, 1]$. The first case is the *beta distribution* with density and derivative

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad (\alpha, \beta) = (2.5, 2) \Rightarrow f^{(1)}(x) = -21.875\sqrt{x}(x-0.6).$$

The second case is the *truncated normal distribution* with density and derivative

$$f(x) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}{\Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)}, \quad (\mu, \sigma) = (0.2, 0.4)$$

$$\Rightarrow f^{(1)}(x) = \frac{(-125x + 25) \exp\left\{-\frac{(5x-1)^2}{8}\right\}}{8\sqrt{2\pi} \{\Phi(2) - \Phi(-0.5)\}},$$

where $\Phi(\cdot)$ is the standard normal cdf. The third case is the *Kumaraswamy distribution* by Kumaraswamy (1980) and Jones (2009) with density and derivative

$$f(x) = \alpha\beta x^{\alpha-1} (1-x^\alpha)^{\beta-1}, \quad (\alpha, \beta) = (1, 4) \Rightarrow f^{(1)}(x) = -12(1-x)^2.$$

Shapes of these densities and derivatives can be found in Figure 3. For each distribution, 1000 data sets of sample size $n \in \{250, 500\}$ are simulated, as before.

The KDDE using the beta kernel, labelled as KDDE-B, is compared with KDDE-N, LPDE2-T and LPDE3-T. To implement all these estimators, we choose their tuning parameters via the oracle method. The RISE and IAD are again considered as performance measures, where each integral is approximated by the trapezoidal rule on an equally-spaced grid of 101 points over the interval $[0, 1]$.

Table 2 reports averages and standard deviations of two performance measures and tuning parameter values over 1000 Monte Carlo samples. We can see the dominance of KDDE-B over the other three estimators. As before, pointwise averages of KDDE-N, LPDE3-T and KDDE-B for $n = 500$ over 1000 Monte Carlo samples are plotted in Figure 3. It can be immediately found that KDDE-B captures the entire shape of each density derivative well. KDDE-N continues to produce a reverse S-curve for each of three distributions, and plots of LPDE3-T are often off target.

3.2. CV method

Before proceeding to real data examples in the next section, we should investigate a CV method. Our construction of a CV criterion is inspired by Härdle et al. (1990). The integrated squared error ('ISE') of KDDE- \mathcal{J} using the smoothing parameter b for $\mathcal{J} \in \{G, B\}$

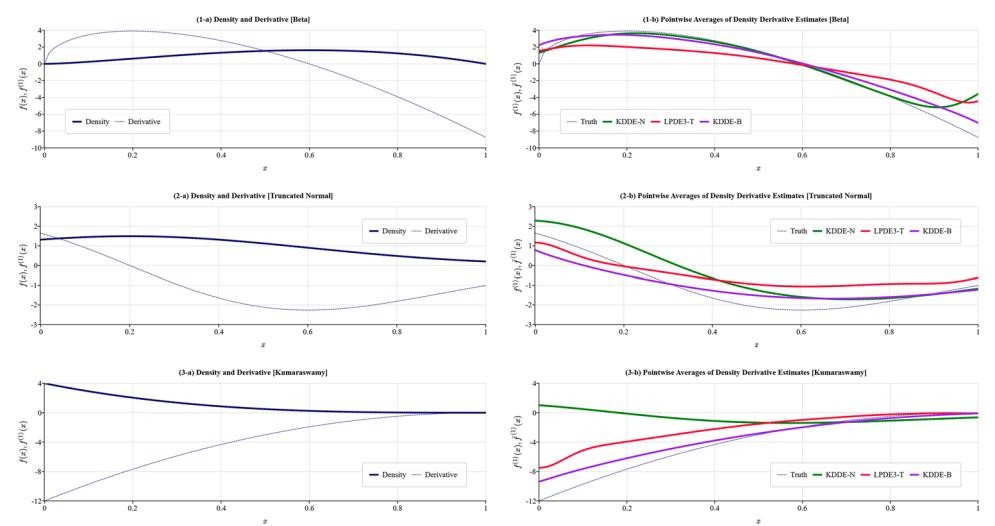


Figure 3. True Densities, True Density Derivatives and Pointwise Averages of Density Derivative Estimates (Case #2, $n = 500$).

Table 2. Monte Carlo Results (Oracle: Case #2).

Estimator	$n = 250$			$n = 500$		
	RISE	IAD	h or b	RISE	IAD	h or b
Beta						
KDDE-N	1.3051 (0.1652)	0.9098 (0.1431)	0.0969 (0.0154)	1.1891 (0.1433)	0.8086 (0.1280)	0.0831 (0.0128)
LPDE2-T	2.2846 (0.2247)	1.8768 (0.1823)	0.2402 (0.0589)	2.1695 (0.1656)	1.7873 (0.1279)	0.1931 (0.0489)
LPDE3-T	2.0021 (0.3645)	1.6521 (0.2705)	0.4164 (0.1015)	1.8529 (0.2560)	1.5679 (0.1844)	0.3887 (0.1044)
KDDE-B	1.1704 (0.2668)	0.9069 (0.2205)	0.0481 (0.0103)	1.0036 (0.2013)	0.7716 (0.1676)	0.0383 (0.0074)
Truncated Normal						
KDDE-N	0.7895 (0.0905)	0.6871 (0.0759)	0.2397 (0.0171)	0.7896 (0.0644)	0.6857 (0.0543)	0.2377 (0.0125)
LPDE2-T	0.9275 (0.1519)	0.8087 (0.1398)	0.3667 (0.0921)	0.8988 (0.1108)	0.7892 (0.1092)	0.3226 (0.0901)
LPDE3-T	0.9925 (0.1818)	0.8720 (0.1639)	0.3728 (0.1096)	0.9213 (0.1168)	0.8143 (0.1160)	0.3853 (0.1101)
KDDE-B	0.7031 (0.2409)	0.5787 (0.2047)	0.1420 (0.0606)	0.6243 (0.2020)	0.5052 (0.1610)	0.1067 (0.0326)
Kumaraswamy						
KDDE-N	5.3098 (0.0366)	3.6365 (0.1751)	0.3857 (0.0780)	5.3133 (0.0238)	3.6553 (0.1206)	0.3846 (0.0621)
LPDE2-T	2.7114 (0.4150)	1.9286 (0.2333)	0.2483 (0.0971)	2.6707 (0.3271)	1.9093 (0.1707)	0.2076 (0.0873)
LPDE3-T	2.6403 (0.5454)	1.9679 (0.3184)	0.3727 (0.1164)	2.5344 (0.4104)	1.9236 (0.2533)	0.3334 (0.1210)
KDDE-B	1.5498 (0.5570)	0.9562 (0.2980)	0.0735 (0.0215)	1.3378 (0.4523)	0.8121 (0.2314)	0.0601 (0.0167)

Note: Simulation averages and standard deviations (in parentheses) of two performance measures ('RISE' and 'IAD') and tuning parameter values (' h ' or ' b ') are reported.

is defined as

$$\begin{aligned} ISE_{\mathcal{J}}(b) &= \int \left\{ \hat{f}_{\mathcal{J},b}^{(1)}(x) - f^{(1)}(x) \right\}^2 dx \\ &= \int \left\{ \hat{f}_{\mathcal{J},b}^{(1)}(x) \right\}^2 dx - 2 \int \hat{f}_{\mathcal{J},b}^{(1)}(x) f^{(1)}(x) dx + \int \left\{ f^{(1)}(x) \right\}^2 dx, \end{aligned}$$

where $\hat{f}_{\mathcal{J},b}^{(1)}(x)$ signifies the dependence of KDDE- \mathcal{J} on b , and the third term of the right-hand side does not depend on b and thus may be safely omitted from the CV criterion. We also assume that $f(0) = 0$ for $\mathcal{J} = G$ and that $f(0) = f(1) = 0$ for $\mathcal{J} = B$, where these assumptions are called the ‘zero-boundary condition’ hereinafter. Then, by integration by parts, it holds that

$$\int \hat{f}_{\mathcal{J},b}^{(1)}(x) f^{(1)}(x) dx = - \int \hat{f}_{\mathcal{J},b}^{(2)}(x) dx.$$

Therefore, the CV criterion is given by

$$CV_{\mathcal{J}}^1(b) = \int \left\{ \hat{f}_{\mathcal{J},b}^{(1)}(x) \right\}^2 dx + \frac{2}{n} \sum_{i=1}^n \hat{f}_{\mathcal{J},b,-i}^{(2)}(X_i), \quad (7)$$

where the superscript ‘1’ in $CV_{\mathcal{J}}^1(b)$ signifies the CV criterion for the first-order density derivative estimator, and

$$\hat{f}_{\mathcal{J},b,-i}^{(2)}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n \frac{\partial^2}{\partial x^2} K_{\mathcal{J}(x,b)}(X_j)$$

is the second-order density derivative estimate using the sample with the i th observation eliminated. Specifically,

$$\begin{aligned} & \frac{\partial^2}{\partial x^2} K_{\mathcal{J}(x,b)}(u) \\ &= \begin{cases} \left\{ L_{G(x,b)}^2(u) - \frac{1}{b^2} \Psi_1\left(\frac{x}{b} + 1\right) \right\} K_{G(x,b)}(u) & \text{for } \mathcal{J} = G \\ \left[L_{B(x,b)}^2(u) - \frac{1}{b^2} \left\{ \Psi_1\left(\frac{x}{b} + 1\right) + \Psi_1\left(\frac{1-x}{b} + 1\right) \right\} \right] K_{B(x,b)}(u) & \text{for } \mathcal{J} = B \end{cases}. \end{aligned}$$

Finally, the CV smoothing parameter \hat{b} is defined as the minimiser of $CV_{\mathcal{J}}^1(b)$.

The same Weibull and beta distributions as in Section 3.1 are chosen to check finite-sample performances of KDDE-G and KDDE-B, respectively, because each distribution satisfies the zero-boundary condition. For each distribution, 1000 data sets of sample size $n \in \{250, 500\}$ are simulated. Judging from performance measures in the corresponding oracle cases, we focus exclusively on KDDE-N as a competitor, where the CV bandwidth \hat{h} is chosen via the method by Härdle et al. (1990). As before, \hat{b} and \hat{h} are found on an equally-spaced grid of 100 points over the interval $[0.005, 0.500]$.

Table 3 reports averages, standard deviations and medians of two performance measures and CV tuning parameter values over 1000 Monte Carlo samples. The reason for

Table 3. Monte Carlo Results (CV).

Estimator	$n = 250$			$n = 500$		
	RISE	IAD	\hat{h} or \hat{b}	RISE	IAD	\hat{h} or \hat{b}
Weibull (Case #1)						
KDDE-N	4.7562 (10.9551) [0.7318]	5.1578 (11.4842) [0.8514]	0.1277 (0.0905) [0.1200]	3.4975 (7.4160) [0.6075]	3.8651 (8.0525) [0.6965]	0.1141 (0.0775) [0.1150]
KDDE-G	0.4767 (0.1397) [0.4550]	0.5202 (0.1285) [0.5167]	0.0685 (0.0245) [0.0750]	0.3815 (0.0903) [0.3712]	0.4213 (0.0893) [0.4181]	0.0570 (0.0174) [0.0600]
Beta (Case #2)						
KDDE-N	9.9746 (17.7338) [1.9212]	7.4058 (13.1133) [1.5355]	0.0680 (0.0469) [0.0650]	6.6297 (11.3958) [1.6902]	4.9283 (8.5476) [1.3296]	0.0606 (0.0398) [0.0550]
KDDE-B	1.8429 (0.9501) [1.4146]	1.4384 (0.7288) [1.1383]	0.0443 (0.0243) [0.0500]	1.4427 (0.6113) [1.1894]	1.1237 (0.4790) [0.9381]	0.0361 (0.0181) [0.0400]

Note: Simulation averages, standard deviations (in parentheses) and medians (in brackets) of two performance measures ('RISE' and 'IAD') and CV tuning parameter values (\hat{h} or \hat{b}) are reported.

presenting medians is that KDDE-N generates extremely large averages and standard deviations of the performance measures. This phenomenon can be attributed to the fact that its CV bandwidths are systematically shorter than the corresponding oracle ones. Such short bandwidths introduce additional variability to the resulting derivative estimates and makes them highly volatile and unstable. In contrast, CV smoothing parameters for KDDE-G and KDDE-B are very close to the corresponding oracle ones. As a consequence, these estimators do not induce extra bias or variability even after they are implemented by the CV smoothing parameters.

4. Real data examples

The first-order density derivative serves as a tool for finding locations of peaks and valleys in the underlying pdf. Consequently, we are motivated to see how precisely two asymmetric KDDEs can detect such locations. For this exercise, these estimators are applied to two datasets. Strictly speaking, features of curve estimates including peaks, valleys and their locations vary across resolutions (i.e. smoothing parameter values). Accordingly, it is desirable to conduct a scale-space analysis like the SiZer map by Chaudhuri and Marron (1999). However, for simplicity, we exclusively examine KDEs and KDDEs implemented by CV smoothing parameters.

4.1. Data descriptions

The first dataset, used in Vella and Verbeek (1998), comprises a sample of 545 full-time working males who completed their schooling by 1980 over the period from 1980 to 1987, and it is originally taken from the National Longitudinal Survey (Youth Sample). We utilise hourly wages, labelled as 'Wage', in year 1980. Since all observations are recorded in the log-transformed scale, they are exponentiated to restore the original scale on \mathbb{R}_+ before the analysis. The second one is the dataset on the Old Faithful Geyser in Yellow Stone National Park, Wyoming, USA. While the literature suggests that several different versions of the

Table 4. Descriptive Statistics.

n	Mean	SD	SK	Min.	Q1	Q2	Q3	Max.
(a) US Hourly Wages (in US dollars)								
545	4.59	2.20	1.06	0.33	3.21	4.25	5.71	16.81
(b) Eruption Durations (in minutes)								
272	3.49	1.14	-0.42	1.60	2.16	4.00	4.46	5.10

Note: n = sample size; Mean = average; SD = standard deviation; SK = skewness; Min. = minimum value; Q1 = first quartile; Q2 = median (i.e. second quartile); Q3 = third quartile; and Max. = maximum value.

data exist, we employ what is used in Wasserman (2006). Our focus is on durations of 272 consecutive eruptions. Both datasets analysed in this section are openly available; see the data availability statement for the URLs at which these can be downloaded.

4.2. Estimation methods

We estimate densities and density derivatives of wages and eruption durations nonparametrically using the gamma and beta kernels, respectively. Table 4 presents summary statistics of two variables. While observations of wages may be directly smoothed by the gamma kernel, those of eruption durations are not confined within $[0, 1]$. Considering that the longest eruption duration is less than six minutes, we map observations in the original scale, denoted as X , onto the unit interval by dividing them by six, before employing the beta kernel. Then, the density and density derivative estimates evaluated at a design point x in the original scale become $\hat{f}_B(x/6)/6$ and $\hat{f}_B^{(1)}(x/6)/36$, respectively, where $\hat{f}_B(\cdot)$ and $\hat{f}_B^{(1)}(\cdot)$ are the beta KDE and KDDE using the transformed observations $U = X/6 \in [0, 1]$.

Density and density derivative estimations are implemented by CV. We may safely apply the CV criterion for density derivative estimation (7), as Table 4 indicates that the zero boundary condition is fulfilled for densities of both wages (in the original scale) and eruption durations (in the transformed scale). Furthermore, the CV criterion for KDE- \mathcal{J} is

$$CV_{\mathcal{J}}^0(b) = \int \left\{ \hat{f}_{\mathcal{J},b}(x) \right\}^2 dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{\mathcal{J},b,-i}(X_i),$$

where $\hat{f}_{\mathcal{J},b,-i}(x)$ is the density estimate using the sample with the i th observation eliminated. CV smoothing parameters are found on an equally-spaced grid of 200 points over the interval $[0.0025, 0.5000]$.

4.3. Results

Figure 4 displays plots of density and density derivative estimates of two variables. First, CV smoothing parameter values for the KDE-G and KDDE-G using the wage data are (coincidentally) 0.1375 and 0.1375, respectively. The density estimate looks single-peaked, having the peak located around 3.57. The density derivative estimate supports the location of the peak in that its sign changes from positive to negative between 3.56 and 3.57.

Second, CV smoothing parameter values for the KDE-B and KDDE-B using the eruption duration data (in the transformed scale on $[0, 1]$) are 0.0025 and 0.0050, respectively.

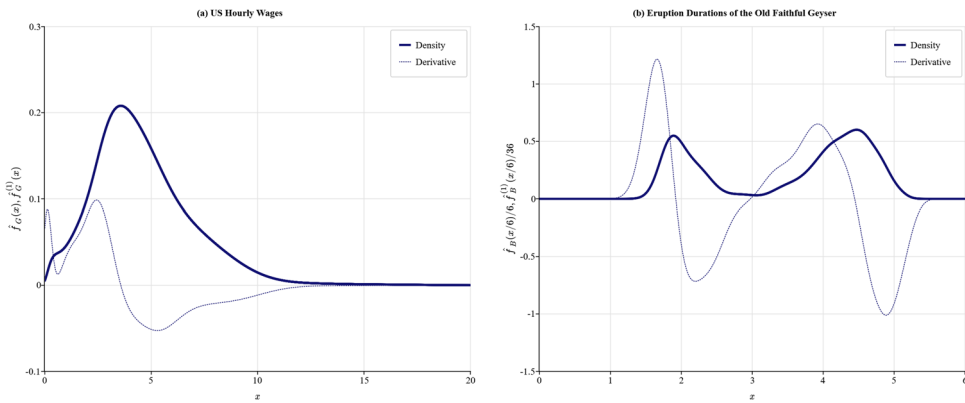


Figure 4. Estimates of Densities and Density Derivatives from Real Data.

The distribution of eruption durations is empirically known to be bimodal. The density estimate captures this shape successfully and indicates that two peaks are located around 1.89 and 4.47, whereas a valley exists around 3.06. The density derivative estimate changes its sign from positive to negative between 1.92 and 1.93 and between 4.44 and 4.45, and from negative to positive between 2.97 and 2.98. Again, locations of two peaks and a valley of the density estimate roughly coincide with the information provided by the density derivative estimate. All in all, both KDDEs work fine in practice in the sense that each derivative estimate can well detect locations of local maximums and minimums of the corresponding density estimate.

5. Conclusion

In this paper, we have proposed to employ asymmetric kernels to estimate first-order density derivatives. The gamma and beta KDDEs are defined, and their large-sample properties including the novel bias and variance expansions are explored. It is demonstrated that these estimators can attain the optimal convergence rate of the MISE $n^{-4/7}$ if the underlying density has third-order smoothness. Superior finite-sample properties of these estimators have been confirmed in Monte Carlo simulations, and two real data examples have illustrated their usefulness.

It is of importance and interest to see whether these estimators can be applicable to other problems than detection of locations of peaks and valleys in a pdf. In particular, incorporating them into (indirect) average derivative estimation of nonparametric regression curves appears to be promising. We leave this task for our future research.

Acknowledgments

The authors would like to thank the editor and two anonymous referees for their constructive comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Japan Society for the Promotion of Science under grant numbers 19K01595 and 23K01340 (M. Hirukawa).

Data availability

The datasets used in Section 4 are openly available at the following URLs:

- <http://qed.econ.queensu.ca/jae/1998-v13.2/vella-verbeek/> (US hourly wages)
- <https://www.stat.cmu.edu/~larry/all-of-nonpar/data.html> (eruption durations of the Old Faithful Geyser).

References

- Cattaneo, M.D., Jansson, M., and Ma, X. (2020), 'Simple Local Polynomial Density Estimators', *Journal of the American Statistical Association*, 115, 1449–1455.
- Chaudhuri, P., and Marron, J.S. (1999), 'SiZer for Exploration of Structures in Curves', *Journal of the American Statistical Association*, 94, 807–823.
- Chen, S.X. (1999), 'Beta Kernel Estimators for Density Functions', *Computational Statistics & Data Analysis*, 31, 131–145.
- Chen, S.X. (2000), 'Probability Density Function Estimation Using Gamma Kernels', *Annals of the Institute of Statistical Mathematics*, 52, 471–480.
- Gordon, L. (1994), 'A Stochastic Approach to the Gamma Function', *American Mathematical Monthly*, 101, 858–865.
- Härdle, W., Marron, J.S., and Wand, M.P. (1990), 'Bandwidth Choice for Density Derivatives', *Journal of the Royal Statistical Society, Series B*, 52, 223–232.
- Hildenbrand, K., and Hildenbrand, W. (1986), 'On the Mean Income Effect: A Data Analysis of the U.K. Family Expenditure Survey', in *Contributions to Mathematical Economics: in Honor of Gérard Debreu*, ed. W. Hildenbrand and A. Mas-Colell, Amsterdam: North-Holland, pp. 247–268.
- Hirukawa, M. (2018), *Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance*, Singapore: Springer.
- Jones, M.C. (2009), 'Kumaraswamy's Distribution: A Beta-Type Distribution with Some Tractability Advantages', *Statistical Methodology*, 6, 70–81.
- Kumaraswamy, P. (1980), 'A Generalized Probability Density Function for Double-Bounded Random Processes', *Journal of Hydrology*, 46, 79–88.
- Lomax, K.S. (1954), 'Business Failures: Another Example of the Analysis of Failure Data', *Journal of the American Statistical Association*, 49, 847–852.
- Müller, H.-G., and Gasser, T. (1979), 'Optimal Convergence Properties of Kernel Estimates of Derivatives of a Density Function', in *Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2–4, 1979*, ed. T. Gasser and M. Rosenblatt, Berlin: Springer-Verlag, pp. 144–154.
- Stone, C.J. (1980), 'Optimal Rates of Convergence for Nonparametric Estimators', *Annals of Statistics*, 8, 1348–1360.
- Vella, F., and Verbeek, M. (1998), 'Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men', *Journal of Applied Econometrics*, 13, 163–183.
- Wand, M.P., and Jones, M.C. (1995), *Kernel Smoothing*, Boca Raton, FL: Chapman & Hall/CRC.
- Wasserman, L. (2006), *All of Nonparametric Statistics*, New York: Springer.
- Ye, Z.-S., and Chen, N. (2017), 'Closed-Form Estimators for the Gamma Distribution Derived From Likelihood Equations', *American Statistician*, 71, 177–181.

Appendix

A.1. A list of useful results

Before proceeding, we summarise useful results. All results but two recursive formulae are stated as lemmata. First two lemmata refer to the moments of log-transformed gamma and beta random variables. The third lemma provides approximations to certain types of digamma and trigamma functions. These are key ingredients for the bias and variance expansions of $\hat{f}_{\mathcal{J}}^{(1)}(x)$.

A.1.1. Recursive formulae for digamma and trigamma functions

For $z > 0$,

$$\Psi(z+1) = \Psi(z) + \frac{1}{z}, \quad \text{and} \quad (\text{A1})$$

$$\Psi_1(z+1) = \Psi_1(z) - \frac{1}{z^2}. \quad (\text{A2})$$

A.1.2. Lemma A.1

Let $X \stackrel{d}{=} G(\alpha, \beta)$. Then,

$$E(X^m \log X) = \beta^m (\alpha)_m \{\log \beta + \Psi(\alpha + m)\}, \quad m \in \{0, 1, 2, \dots\}, \quad \text{and} \\ E(\log^2 X) = \{\log \beta + \Psi(\alpha)\}^2 + \Psi_1(\alpha).$$

A.1.3. Lemma A.2

Let $X \stackrel{d}{=} \text{Beta}(\gamma, \delta)$. Then,

$$E\left\{X^m \log\left(\frac{X}{1-X}\right)\right\} = \frac{(\gamma)_m}{(\gamma + \delta)_m} \{\Psi(\gamma + m) - \Psi(\delta)\}, \quad m \in \{0, 1, 2, \dots\}, \quad \text{and} \\ E\left\{\log^2\left(\frac{X}{1-X}\right)\right\} = \{\Psi(\gamma) - \Psi(\delta)\}^2 + \Psi_1(\gamma) + \Psi_1(\delta).$$

A.1.4. Lemma A.3

For $z > 0$,

$$\Psi\left(\frac{z}{b} + 1\right) = \log\left(\frac{z}{b}\right) + \frac{b}{2z} + O(b^2), \quad \text{and} \\ \Psi_1\left(\frac{z}{b} + 1\right) = \frac{b}{z} + O(b^2),$$

as $b \rightarrow 0$.

A.1.5. Proof of Lemma A.1

$E(X \log X)$ is available in the proof of Theorem 3.1 of Ye and Chen (2017). $E(\log^2 X)$ are also provided in the proof of Theorem 3.2 of Ye and Chen (2017). Finally, $E(X^m \log X)$ for $m \geq 2$ can be derived analogously to the proof of Lemma A.2 below. ■

A.1.6. Proof of Lemma A.2

The proof strategy largely comes from Ye and Chen (2017). Observe that the moment generating function of $\log\{X/(1-X)\}$ for $X \stackrel{d}{=} \text{Beta}(\gamma, \delta)$ is given by

$$M_{\log\left(\frac{X}{1-X}\right)}(z) = E\left[\exp\left\{z \log\left(\frac{X}{1-X}\right)\right\}\right] = E\left\{\left(\frac{X}{1-X}\right)^z\right\} = \frac{\Gamma(\gamma + z) \Gamma(\delta - z)}{\Gamma(\gamma) \Gamma(\delta)}.$$

Therefore,

$$\begin{aligned} E \left\{ \log \left(\frac{X}{1-X} \right) \right\} &= \Psi(\gamma) - \Psi(\delta), \quad \text{and} \\ E \left\{ \log^2 \left(\frac{X}{1-X} \right) \right\} &= \{\Psi(\gamma) - \Psi(\delta)\}^2 + \Psi_1(\gamma) + \Psi_1(\delta). \end{aligned}$$

To obtain $E[X^m \log\{X/(1-X)\}]$ for $m \geq 1$, consider that

$$\begin{aligned} E \left\{ X^m \log \left(\frac{X}{1-X} \right) \right\} &= \int_0^1 u^m \log \left(\frac{u}{1-u} \right) \frac{u^{\gamma-1} (1-u)^{\delta-1}}{B(\gamma, \delta)} du \\ &= \frac{B(\gamma+m, \delta)}{B(\gamma, \delta)} \int_0^1 \log \left(\frac{u}{1-u} \right) \frac{u^{\gamma+m-1} (1-u)^{\delta-1}}{B(\gamma+m, \delta)} du, \end{aligned}$$

where the integral corresponds to $E[\log\{Y/(1-Y)\}]$ for $Y \stackrel{d}{=} \text{Beta}(\gamma+m, \delta)$. Then, the result immediately follows. ■

A.1.7. Proof of Lemma A.3

By Theorems 4 and 5 of Gordon (1994),

$$\begin{aligned} \log y - \frac{1}{2y} - \frac{1}{12y^2} &< \Psi(y) < \log y - \frac{1}{2y} - \frac{1}{12(y+1/14)^2}, \quad \text{and} \\ \frac{1}{y} + \frac{1}{2y^2} + \frac{1}{6(y+1/14)^3} &< \Psi_1(y) < \frac{1}{y} + \frac{1}{2y^2} + \frac{1}{6y^3} \end{aligned}$$

for all $y > 0$. The stated results can be demonstrated by putting $y = z/b + 1$ and using (A1) and (A2). ■

A.2. Derivation of (3)

(a) For $\mathcal{J} = G$:

It suffices to solve

$$L_{G(x,b)}(u) = \frac{1}{b} \left\{ \log u - \log b - \Psi \left(\frac{x}{b} + 1 \right) \right\} = 0$$

for u . It follows from Lemma A.3 that

$$\log u = \log b + \Psi \left(\frac{x}{b} + 1 \right) = \log x + \frac{b}{2x} + O(b^2).$$

Then,

$$u_G^* = x \exp \left\{ \frac{b}{2x} + O(b^2) \right\} = x \left\{ 1 + \frac{b}{2x} + O(b^2) \right\} = x + \frac{b}{2} + O(b^2).$$

(b) For $\mathcal{J} = B$:

Again, we solve

$$L_{B(x,b)}(u) = \frac{1}{b} \left\{ \log \left(\frac{u}{1-u} \right) - \Psi \left(\frac{x}{b} + 1 \right) + \Psi \left(\frac{1-x}{b} + 1 \right) \right\} = 0$$

for u . If $x = 1/2$, then this equation reduces to $\log\{u/(1-u)\} = 0$ and thus $u_B^* = 1/2 = x$ trivially holds. If $x \neq 1/2$, then, by Lemma A.3,

$$\log \left(\frac{u}{1-u} \right) = \Psi \left(\frac{x}{b} + 1 \right) - \Psi \left(\frac{1-x}{b} + 1 \right) = \log \left(\frac{x}{1-x} \right) - \left\{ \frac{1-2x}{2x(1-x)} \right\} b + O(b^2),$$

so that

$$\begin{aligned}\frac{u}{1-u} &= \left(\frac{x}{1-x}\right) \exp\left[-\left\{\frac{1-2x}{2x(1-x)}\right\}b + O(b^2)\right] \\ &= \left(\frac{x}{1-x}\right) \left[1 - \left\{\frac{1-2x}{2x(1-x)}\right\}b + O(b^2)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}u_B^* &= \frac{x - \left\{\frac{1-2x}{2(1-x)}\right\}b + O(b^2)}{1 - \left\{\frac{1-2x}{2(1-x)}\right\}b + O(b^2)} \\ &= \left[x - \left\{\frac{1-2x}{2(1-x)}\right\}b + O(b^2)\right] \left[1 + \left\{\frac{1-2x}{2(1-x)}\right\}b + O(b^2)\right] \\ &= x + \left(x - \frac{1}{2}\right)b + O(b^2).\end{aligned}$$

■

A.3. Proof of Theorem 2.1

A.3.1. Proof of (i)

(a) For $\mathcal{J} = G$:

Let $\xi_x \stackrel{d}{=} G(x/b + 1, b)$. Because the gamma kernel is the pdf of $G(x/b + 1, b)$, we may write

$$E\left\{\hat{f}_G^{(1)}(x)\right\} = \int_0^\infty L_{G(x,b)}(u) K_{G(x,b)}(u) f(u) du = E\left\{L_{G(x,b)}(\xi_x) f(\xi_x)\right\}. \quad (\text{A3})$$

Substituting

$$f(\xi_x) = f(x) + f^{(1)}(x)(\xi_x - x) + \frac{f^{(2)}(x)}{2!}(\xi_x - x)^2 + \frac{f^{(3)}(\bar{x})}{3!}(\xi_x - x)^3$$

with $\bar{x} = t\xi_x + (1-t)x$ for some $t \in [0, 1]$ into the right-hand side of (A3) yields

$$\begin{aligned}E\left\{\hat{f}_G^{(1)}(x)\right\} &= \sum_{j=0}^3 \frac{f^{(j)}(x)}{j!} E\left\{L_{G(x,b)}(\xi_x) (\xi_x - x)^j\right\} \\ &\quad + E\left[\frac{\{f^{(3)}(\bar{x}) - f^{(3)}(x)\}}{3!} L_{G(x,b)}(\xi_x) (\xi_x - x)^3\right] \\ &= \sum_{j=0}^3 A_G(j) + R_G(\text{say}).\end{aligned}$$

It also follows from Lemma A.1 and a property of gamma random variables that

$$E\left\{L_{G(x,b)}(\xi_x) \xi_x^m\right\} = b^{m-1} \left(\frac{x}{b} + 1\right)_m \left\{\Psi\left(\frac{x}{b} + m + 1\right) - \Psi\left(\frac{x}{b} + 1\right)\right\}.$$

By this and (A1),

$$\begin{aligned}E\left\{L_{G(x,b)}(\xi_x)\right\} &= 0, \\ E\left\{L_{G(x,b)}(\xi_x) (\xi_x - x)\right\} &= 1, \\ E\left\{L_{G(x,b)}(\xi_x) (\xi_x - x)^2\right\} &= 3b, \quad \text{and} \\ E\left\{L_{G(x,b)}(\xi_x) (\xi_x - x)^3\right\} &= 3xb + 11b^2.\end{aligned}$$

Therefore,

$$\sum_{j=0}^3 A_G(j) = f^{(1)}(x) + \frac{1}{2} \left\{ 3f^{(2)}(x) + xf^{(3)}(x) \right\} b + O(b^2).$$

The remaining task is to demonstrate that $R_G = o(b)$. It follows from (4) and the Cauchy-Schwarz inequality that

$$|R_G| \leq \frac{Ct^\zeta}{6} E \left\{ |L_{G(x,b)}(\xi_x)| |\xi_x - x|^{3+\zeta} \right\} \leq \frac{C}{6} \left[E \left\{ L_{G(x,b)}^2(\xi_x) \right\} \right]^{1/2} \left\{ E |\xi_x - x|^{2(3+\zeta)} \right\}^{1/2}.$$

Lemmata A.1 and A.3 jointly imply that

$$E \left\{ L_{G(x,b)}^2(\xi_x) \right\} = \frac{1}{b^2} \Psi_1 \left(\frac{x}{b} + 1 \right) = \frac{1}{b^2} O(b) = O(b^{-1}).$$

Also, by Hölder's inequality and $E|\xi_x - x|^8 = O(b^4)$,

$$E |\xi_x - x|^{2(3+\zeta)} \leq \left\{ E |\xi_x - x|^8 \right\}^{(3+\zeta)/4} = O(b^{3+\zeta}).$$

Hence,

$$|R_G| \leq O(b^{-1/2}) O(b^{(3+\zeta)/2}) = O(b^{1+\zeta/2}) = o(b),$$

and thus the bias expansion is demonstrated.

(b) For $\mathcal{J} = B$:

Let $\theta_x \stackrel{d}{=} \text{Beta}\{x/b + 1, (1-x)/b + 1\}$. Since the beta kernel is the pdf of $\text{Beta}\{x/b + 1, (1-x)/b + 1\}$, it holds that $E\{\hat{f}_B^{(1)}(x)\} := \sum_{j=0}^3 A_B(j) + R_B$, where

$$A_B(j) = \frac{f^{(j)}(x)}{j!} E \left\{ L_{B(x,b)}(\theta_x) (\theta_x - x)^j \right\},$$

$$R_B = E \left[\frac{\{f^{(3)}(\bar{x}) - f^{(3)}(x)\}}{3!} L_{B(x,b)}(\theta_x) (\theta_x - x)^3 \right],$$

and $\bar{x} = t\theta_x + (1-t)x$ for some $t \in [0, 1]$ as before. It also follows from Lemma A.2 and a property of beta random variables that

$$E \left\{ L_{B(x,b)}(\theta_x) \theta_x^m \right\} = b^{-1} \frac{(x/b + 1)_m}{(1/b + 2)_m} \left\{ \Psi \left(\frac{x}{b} + m + 1 \right) - \Psi \left(\frac{x}{b} + 1 \right) \right\}.$$

Combining this with (A1) yields

$$E \left\{ L_{B(x,b)}(\theta_x) \right\} = 0,$$

$$E \left\{ L_{B(x,b)}(\theta_x) (\theta_x - x) \right\} = \frac{1}{1 + 2b} = 1 - 2b + O(b^2),$$

$$E \left\{ L_{B(x,b)}(\theta_x) (\theta_x - x)^2 \right\} = \frac{2x + 3b}{(1 + 2b)(1 + 3b)} - \frac{2x}{1 + 2b}$$

$$= 3(1 - 2x)b + O(b^2), \quad \text{and}$$

$$E \left\{ L_{B(x,b)}(\theta_x) (\theta_x - x)^3 \right\} = \frac{3x^2 + 12xb + 11b^2}{(1 + 2b)(1 + 3b)(1 + 4b)}$$

$$- \frac{3x(2x + 3b)}{(1 + 2b)(1 + 3b)} + \frac{3x^2}{1 + 2b}$$

$$= 3x(1 - x)b + O(b^2).$$

It can be also shown that $|R_B| \leq O(b^{1+\zeta/2}) = o(b)$, which establishes the stated result.

A.3.2. Proof of (ii)(a) For $\mathcal{J} = G$:

It follows from

$$\text{Var} \left\{ \hat{f}_G^{(1)}(x) \right\} = \frac{1}{n} E \left\{ L_{G(x,b)}^2(X) K_{G(x,b)}^2(X) \right\} + O(n^{-1})$$

that we focus on approximating $E\{L_{G(x,b)}^2(X)K_{G(x,b)}^2(X)\}$. Let $\eta_x \stackrel{d}{=} G(2x/b + 1, b/2)$. Then, as in the proof of (i),

$$E \left\{ L_{G(x,b)}^2(X) K_{G(x,b)}^2(X) \right\} := B_G E \left\{ L_{G(x,b)}^2(\eta_x) f(\eta_x) \right\},$$

where, by Chen (2000, p. 474),

$$B_G = \frac{b^{-1}\Gamma(2x/b + 1)}{2^{2x/b+1}\Gamma^2(x/b + 1)} \sim \begin{cases} b^{-1/2}/(2\sqrt{\pi}\sqrt{x}) & \text{if } x/b \rightarrow \infty \\ b^{-1}\Gamma(2\kappa + 1)/\{2^{2\kappa+1}\Gamma^2(\kappa + 1)\} & \text{if } x/b \rightarrow \kappa \end{cases}.$$

In addition,

$$E \left\{ L_{G(x,b)}^2(\eta_x) f(\eta_x) \right\} \sim f(x) E \left\{ L_{G(x,b)}^2(\eta_x) \right\},$$

and it follows from Lemma A.1 that the expectation can be simplified as

$$E \left\{ L_{G(x,b)}^2(\eta_x) \right\} = \frac{1}{b^2} \left[\left\{ \Psi\left(\frac{x}{b} + 1\right) - \Psi\left(\frac{2x}{b} + 1\right) + \log 2 \right\}^2 + \Psi_1\left(\frac{2x}{b} + 1\right) \right].$$

It follows from Lemma A.3 that

$$E \left\{ L_{G(x,b)}^2(\eta_x) \right\} = \frac{b^{-1}}{2x} + O(1) \sim \begin{cases} b^{-1}/(2x) & \text{if } x/b \rightarrow \infty \\ b^{-2}/(2\kappa) & \text{if } x/b \rightarrow \kappa \end{cases},$$

which leads to the variance expansion.

(b) For $\mathcal{J} = B$:

By

$$\text{Var} \left\{ \hat{f}_B^{(1)}(x) \right\} = \frac{1}{n} E \left\{ L_{B(x,b)}^2(X) K_{B(x,b)}^2(X) \right\} + O(n^{-1}),$$

we have

$$E \left\{ L_{B(x,b)}^2(X) K_{B(x,b)}^2(X) \right\} := B_B E \left\{ L_{B(x,b)}^2(\vartheta_x) f(\vartheta_x) \right\},$$

where $\vartheta_x \stackrel{d}{=} \text{Beta}\{2x/b + 1, 2(1-x)/b + 1\}$. Lemma of Chen (1999) implies that

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

In addition, $E\{L_{B(x,b)}^2(\vartheta_x)f(\vartheta_x)\} \sim f(x)E\{L_{B(x,b)}^2(\vartheta_x)\}$, where the expectation reduces to

$$\begin{aligned} & E \left\{ L_{B(x,b)}^2(\vartheta_x) \right\} \\ &= \frac{1}{b^2} \left[\left\{ \Psi\left(\frac{x}{b} + 1\right) - \Psi\left(\frac{1-x}{b} + 1\right) - \Psi\left(\frac{2x}{b} + 1\right) + \Psi\left(\frac{2(1-x)}{b} + 1\right) \right\}^2 \right. \\ & \quad \left. + \Psi_1\left(\frac{2x}{b} + 1\right) + \Psi_1\left\{\frac{2(1-x)}{b} + 1\right\} \right] \end{aligned}$$

by Lemma A.2. Then, Lemma A.3 implies that

$$E \left\{ L_{B(x,b)}^2(\vartheta_x) \right\} = \frac{b^{-1}}{2x(1-x)} + O(1) \\ \sim \begin{cases} b^{-1}/\{2x(1-x)\} & \text{if } x/b \rightarrow \infty \text{ and } (1-x)/b \rightarrow \infty \\ b^{-2}/(2\kappa) & \text{if } x/b \rightarrow \kappa \text{ or } (1-x)/b \rightarrow \kappa \end{cases},$$

which completes the proof. ■