Afrika Statistika

Afrika Statistika
Vol. 12 (3), 2017, pages 1417–1446.
DOI: http://dx.doi.org/10.16929/as/2017.1417.111

ISSN 2316-090X

The mode-dispersion approach for constructing continuous associated kernels

Francial Giscard B. LIBENGUE DOLELE-KPOKA[1,*] and Célestin C. KOKONENDJI[2]

[1]Université de Bangui, Faculté des Sciences, Département de Mathématiques et Informatique. BP 908 Bangui, République Centrafricaine
[2]Laboratoire de Mathématiques de Besançon, Université Bourgogne Franche-Comté, Besançon, France

Received June 27, 2017; Accepted October 6, 2017

Abstract. We introduce the mode-dispersion approach for constructing the (asymmetric) continuous associated kernels from suitable parametric probability density functions (p.d.f.) that we shall call the type of kernel. This leads us to value the choice of the associated kernel, since it takes into account the support of the unknown density f, to be estimated. All associated kernel density estimators must be without edge effect. For illustrating this, we introduce the extended beta kernel, which is a typical model of kernels with bounded supports. However, in the presence of a large bias of the density estimator, we propose a general but light modification in the same type of the first associated kernel; it leads to improve the mean integrated square error of the new estimator. Some properties of two estimators are investigated and compared, in particular pointwise and global (asymptotical) properties. Several forms of types of kernels and their associated kernel estimators are subsequently examined in detail. Simulation studies are made on three lognormal kernel density estimators for pointing out some behaviors at the boundaries.

Key words: Cross-validation, dispersion parameter; free of boundary effect; unimodal kernel.
AMS 2010 Mathematics Subject Classification : Primary 62G07 ; Secondary 62G99.

*Corresponding author Francial Giscard B. LIBENGUE DOLELE-KPOKA : libengue@gmail.com
Célestin C. KOKONENDJI : celestin.kokonendji@univ-fcomte.fr

Résumé. Nous introduisons l'approche mode-dispersion pour la construction des noyaux associés continus (asymétrique) à partir des fonctions de densité de probabilités (f.d.p.) paramétrées que nous appelons type de noyau. Ceci nous conduit à valoriser le choix du noyau associé puisque celui-ci tient compte du support de la densité inconnue f, à estimer. Tous les estimateurs à noyaux associés de densité sont sans effet de bord. Pour illustrer ceci, nous introduisons le noyau associé bêta étendu, qui est un modèle type des noyaux à supports bornés. Cependant, en présence d'un grand biais de l'estimateur à noyaux associés de densité, nous proposons une technique de modification générale mais légère dans le même type du premier noyau associé ; Cela conduit à réduire l'erreur quadratique moyenne intégrée du nouvel estimateur. Certaines propriétés de deux estimateurs sont étudiées et comparées, en particulier des propriétés asymptotiques ponctuelles et globales. Plusieurs formes de types de noyaux et leurs estimateurs à noyaux associés sont ensuite examinés en détail. Des études de simulation sont faites sur trois estimateurs de densité à noyau log-normal pour souligner les comportements aux bords.

1. Introduction

Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) random variables with an unknown probability density function (p.d.f.) f on $\mathbb{T}$, a subset of the real line $\mathbb{R}^d$. The kernel estimator $\widehat{f_n}$ of the p.d.f. f is classically defined by

$$\widehat{f_n}(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{T} = \mathbb{R}. \tag{1}$$

The continuous kernel function $\mathcal{K}(\cdot)$ in (1) is in general a symmetric p.d.f., independent of the target x and the bandwidth h with zero mean and finite variance; see Rosenblatt (1956), Parzen (1962), Sylverman (1986), Devroye (1987), Scott (1992), Tsybakov (2004) for a review. This popular kernel $\mathcal{K}(\cdot)$ has been imagined for estimating f with unbounded support $\mathbb{T} = \mathbb{R}$. From the works of Chen (1999) and Chen (2000) on beta and gamma kernels, respectively for densities on $\mathbb{T} = [0, 1]$ and $\mathbb{T} = [0, \infty)$, Scaillet (2004) on inverse Gaussian kernel and its reciprocal for densities on $\mathbb{T} = [0, \infty)$ and as well as Kokonendji and Senga Kiessé (2011) for the discrete case (i.e. $\mathbb{T} \subseteq \mathbb{Z}$), is borned a kind of kernel estimators (that we call *associated kernel estimators*) where the kernel function is parameterized by the estimated point x and the smoothing parameter h. In order to harmonize writing as in Kokonendji and Senga Kiessé (2011), one can define a continuous associated kernel estimator $\widehat{f_n}$ of the p.d.f. f by

$$\widehat{f_n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T} \subseteq \mathbb{R}. \tag{2}$$

It is well known that $\mathcal{K}(\cdot)$ of (1) becomes a particular case of the continuous associated kernel $K_{x,h}$ which is intrinsically dependent on x and h, through the following relation:

$$K_{x,h}(\cdot) = \frac{1}{h}\mathcal{K}\left(\frac{x - \cdot}{h}\right). \tag{3}$$

This opened the way for several authors to refine the properties of these associated kernel estimators in many particular cases. See Bouezmarni and Rolin (2003), Bouezmarni *et al.* (2005), Kokonendji *et al.* (2009), Kokonendji and Zocchi (2010), Senga Kiessé and Rivoire (2010), Bertin and Klutchnikoff (2011), Zougab *et al.* (2012), Zougab *et al.* (2013) and Libengué Dobélé-Kpoka (2013) for details. These special cases dealt show that the associated kernel estimators are without edge effects but they do not cover all types of $\mathbb{T}$. In addition, it is worth noting the absence of a general definition for classical continuous kernel and a method for constructing them. Actually, for $0 < t_1 < t_2$ with $t_2 \neq 1$, the support $\mathbb{T} = [t_1, t_2]$ corresponds, for example, to the real human size distribution between $[0.5, 2.5]$ (unit by meter) or many other observed natural distributions for which there does not exist an appropriate kernel function. This leads us to introduce, for $\mathbb{T} = [t_1, t_2]$ $(t_1 < t_2)$, the "extended beta kernel", which belongs to the family of continuous associated kernels, asymmetric, without edge effect, and very useful for smoothing some observed distributions on known compact support $\mathbb{T} \subseteq \mathbb{R}$.

In this paper, the main goal is to homogenize the theory concerning the associated kernels estimators by pointing out a systematical technique for constructing them without boundary effects for any type of support $\mathbb{T}$ of f, to be estimated. The rest of the paper is organized as follows. Section 2 gives a general definition of the continuous associated kernel which includes the classical one. A1so, we provide a principle for constructing any continuous associated kernel from a parametric p.d.f. and we illustrate this with some examples from the literature, as well as new ones. In Section 3, we apply the continuous associated kernel for smoothing an unknown given p.d.f. Some pointwise properties will be investigated, in particular the convergence in sense of asymptotic mean integrated squared error (AMISE) and the bias reduction algorithm. In Section 4, some illustrations with the types of kernels such as Pareto, lognormal, beta and its extended version, gamma and its inverse and also, inverse Gaussian and its reciprocal are derived.Section 5 provides a simulation study in which the finite-sample properties of three lognormal estimators are investigated. In particular, we explore in more detail the role of the smoothing parameter or bandwidth. The three estimators are critically compared in the sense of the MISE. Main conclusions and a final discussion are given in Section 6.

To finish, we let know the reader that all the tables are postponed after the bibliography at the end of the document.

2. Continuous associated kernels

We provide here some definitions and construction of the continuous associated kernel. We investigate their basic properties and finally give some illustrations.

2.1. Definitions and construction

Let us start by a general definition of continuous associated kernels as in Libengué Dobélé-Kpoka (2013).

Definition 1. Consider $x \in \mathbb{T} \subseteq \mathbb{R}$ and $h > 0$ with $\mathbb{T}$ the support of the p.d.f. f, to be estimated. A parametrized p.d.f. $K_{x,h}$ of support $\mathbb{S}_{x,h} \subseteq \mathbb{R}$ is called "associated kernel" if the following conditions are satisfied:

$$x \in \mathbb{S}_{x,h}, \tag{4}$$
$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + A(x,h), \tag{5}$$
$$Var(\mathcal{Z}_{x,h}) = B(x,h), \tag{6}$$

with $\mathcal{Z}_{x,h}$ a random variable with p.d.f. $K_{x,h}$, and both $A(x,h)$ and $B(x,h)$ go to 0 as h goes to 0.

Let us remark that:

(i) The support $\mathbb{S}_{x,h}$ is not necessary symmetric with respect to 0 or to x as in the classical case. It can depend or not on x and h.
(ii) The condition (4) can be replaced by $\cup_{x\in\mathbb{T}}\mathbb{S}_{x,h} \supseteq \mathbb{T}$ and implies that the associated kernel takes first into account the support $\mathbb{T}$ of the p.d.f. f, to be estimated.
(iii) If $\cup_{x\in\mathbb{T}}\mathbb{S}_{x,h} \supsetneq \mathbb{T}$, then this is the classical problem of boundary bias.
(iv) The conditions (5) and (6) indicate that the associated kernel is more and more concentrated around x as h goes to 0. This highlights the peculiarity of associated kernel which can change its shape according the target position.

The following proposition transforms all classical kernels to associated kernels, and points out the shape of their support.

Proposition 1. *Let $\mathcal{K}$ be a classical (symmetric) kernel with support $\mathbb{S}$, mean $\mu_K = 0$ and variance $\sigma_K^2 < \infty$. For a given $x \in \mathbb{T} = \mathbb{R}$ and $h > 0$, then the classical associated kernel is defined by (3) and $\mathbb{S}_{x,h} = x - h\mathbb{S}$ with*

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x \text{ and } Var(\mathcal{Z}_{x,h}) = h^2\sigma_K^2. \tag{7}$$

In other words (7) corresponds to the following characteristics A and B of classical associated kernel $K_{x,h}$:

$$A(x,h) = 0 \;\; and \;\; B(x,h) = O(h^2). \tag{8}$$

Proof. From (3), for a fixed x in $\mathbb{T}$ and for all t in $\mathbb{T}$, there exists u in $\mathbb{S}$ such that $u = (x-t)/h$. This implies that $t = x - uh$. Since $t \in \mathbb{T}$, it comes from (4) that $\mathbb{S}_{x,h} = x - h\mathbb{S}$. The last two results are derived from calculating the variance and the mean of $K_{x,h}$ by making the change of variables $u = (x-t)/h$.■

The following definition introduces the notion of *type of kernel* that we need for constructing any associated kernel.

Definition 2. A type of a (continuous) kernel K is a squared integrable p.d.f. $K = K_\theta$, depending on (parameters) $\theta \in \Theta \subseteq \mathbb{R}^2$ and with support $\mathbb{S}_K = \mathbb{S}_\theta$.

In what follows, we consider only the type of uni-modal kernels $K = K_\theta$ with mode M_θ and having a dispersion parameter D_θ. Since Θ is of two dimensions, we shall denote $\theta = \theta(a,b)$, $M_\theta = M(a,b)$ and $D_\theta = D(a,b)$ where a and b are two positive reals.

Remark 1. It is important to have:

(i) The mode $M(a,b)$ always belongs in $\mathbb{S}_{\theta(a,b)}$.

(ii) Obviously, the probability of the mode is equal or always greater than the probability of the mean.
(iii) When the dispersion parameter around the mode tends to zero, this implies that the dispersion parameter around the mean goes also to zero.

There are several definitions about the concept of dispersion parameters; see Jørgensen (1997), Jørgensen *et al.* (2010) and Jørgensen and Kokonendji (2011) for further details. Now, we are able to construct some associated kernels using any p.d.f. which satisfies Definition 1. Let us give a general method of construction.

Principle (mode-dispersion). Let $K_{\theta(a,b)}$ be a type of unimodal kernel on $\mathbb{S}_{\theta(a,b)}$, with mode $M(a,b)$ and dispersion parameter $D(a,b)$. The mode-dispersion method allows the construction of the function $K_{\theta(x,h)}$ by solving in term a and b the equations

$$\begin{cases} M(a,b) = x \\ D(a,b) = h. \end{cases}$$

Then $\theta(x,h) = \theta\left(a(x,h), b(x,h)\right)$ where $a(x,h)$ and $b(x,h)$ are solutions of the previous equations, for $h > 0$ and x in $\mathbb{T}$.

In the following proposition we show that $K_{\theta(x,h)}$ satisfies Definition 1.

Proposition 2. *Let $\mathbb{T}$ be the support of the density f to be estimated. For all $x \in \mathbb{T}$ and $h > 0$, the kernel function constructed by the mode-dispersion method $K_{\theta(x,h)}$ with support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$, is such that*

$$x \in \mathbb{S}_{\theta(x,h)}, \tag{9}$$
$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = A_\theta(x,h), \tag{10}$$
$$Var(\mathcal{Z}_{\theta(x,h)}) = B_\theta(x,h), \tag{11}$$

where $\mathcal{Z}_{\theta(x,h)}$ is a random variable with p.d.f. $K_{\theta(x,h)}$ and $A_\theta(x,h) \to 0$ and $B_\theta(x,h) \to 0$ as $h \to 0$.

Proof. Firstly, from the mode-dispersion method one has $\theta(x,h) = \theta\left(a(x,h), b(x,h)\right)$ which leads to $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$. Since $K_{\theta(a,b)}$ is unimodal of mode $M(a,b) \in \mathbb{S}_{\theta(a,b)}$ (from Part (i) of Remark 1), we obtain from the mode-dispersion method the first result (9) as follows:

$$M(a,b) = x \in \mathbb{S}_{\theta(a,b)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}.$$

In addition, for a given random variable $\mathcal{Z}_{\theta(a,b)}$ associated to the type of the unimodal kernel $K_{\theta(a,b)}$, we can write $\mathbb{E}(\mathcal{Z}_{\theta(a,b)}) = M(a,b) + \varepsilon(a,b)$, where $\varepsilon(a,b)$ is the difference between the mode and mean of $K_{\theta(a,b)}$. From the mode-dispersion method, we have $M(a,b) = x$ and $\varepsilon(a,b) = \varepsilon(a(x,h), b(x,h))$. Next, we can write

$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = \varepsilon(a(x,h), b(x,h)).$$

Taking $A_\theta(x,h) = \varepsilon(a(x,h), b(x,h))$ and using the definition of the dispersion parameter around the mode, this leads to the second result (10). Finally, since $K_{\theta(a,b)}$ admits a moment of second order, this implies that the variance of $K_{\theta(x,h)}$ exists in terms of x and h. We can write it as

$$Var(\mathcal{Z}_{\theta(x,h)}) = B_{\theta(a(x,h),b(x,h))},$$

with $B_{\theta(a(x,h),b(x,h))}$ which tends to zero as $h \to 0$ from Part (iii) of Remark 1. We obtain the last result (11) by taking $B_\theta(x,h) = B_{\theta(a(x,h),b(x,h))}$.■

Note that in practice, both characteristics $A_\theta(x,h)$ and $B_\theta(x,h)$ are derived from the calculation of mean and variance of $K_{\theta(x,h)}$ in term of $a(x,h)$ and $b(x,h)$. We show how to calculate them in Section 2.2.

However, it is worth noting that some types of kernels do not satisfy the mode-dispersion method. For instance, we have the Weibul type of kernel, with shape and scale parameters $a > 1$ and $b > 0$. It has as mean and variance, respectively $b\Gamma(1 + 1/a)$ and $b^2 \{\Gamma(1 + 2/a) - (2/a)\log(2)\}$, where $\Gamma(\cdot)$ is the gamma function. This type of kernel is defined by

$$W_{\theta(a,b)}(u) = \frac{a}{b^a} u^{a-1} \exp\left\{-\left(\frac{x}{b}\right)^a\right\} \mathbb{I}_{[0,\infty)}(u).$$

Its mode and dispersion parameter are $b(1 - 1/a)^{1/a}$ and b respectively. Using the mode-dispersion method we have

$$\begin{cases} (1 - 1/a)^{1/a} = x/h \\ b = h. \end{cases}$$

Here $b(x,h) = h$ is not depending on x, but $a(x,h)$ is an implicit expression depending on x and h.

Another example is the Birnbaum-Saunders type of kernel with shape and scale parameters $a > 0$ and $b > 0$ respectively, and defined by

$$BS_{\theta(a,b)}(u) = \frac{1}{2ab\sqrt{2\pi}} \left\{\left(\frac{b}{u}\right)^{1/2} + \left(\frac{b}{u}\right)^{3/2}\right\} \exp\left(\frac{-1}{2a^2}\left[\frac{u}{b} + \frac{b}{u} - 2\right]\right) \mathbb{I}_{[0,\infty)}(u).$$

Its mean and variance are $b(1 + a^2/2)$ and $(ab)^2(1 + a^2/2)$ respectively. From the mode-dispersion method, we here have

$$\begin{cases} \arg\max_{u>0} BS_{\theta(a,h)}(u) = x \\ b = h. \end{cases}$$

However, the mode $M(a,h) = \arg\max_{u>0} BS_{\theta(a,h)}(u)$ cannot be obtained in explicit form. It has to be obtained by solving a non-linear equation in terms of the shape parameter a. In Balakrishnan *et al.* (2011), some modal values of this kernel type are calculated by varying a from 0.5 to 5.0. Let us precise here that there exists other methods to construct associated kernels; see, e.g., Jin and Kawczak (2003) for Birnbaum-Sanders and lognormal type of kernels.

2.2. *Examples of non classical kernels*

We now show in detail the case of the extended beta kernel because it has a general support from which many other cases can be deduced. The remaining examples are summarized in Table 3 which will be commented below.

2.2.1. Extended beta kernel

This kernel belongs to the family of beta densities on the support $\mathbb{S}_{EB} = [t_1, t_2]$ with $0 \leq t_1 < t_2$. It is unknown in nonparametric statistics but it is often used in Operation Research; see Grubbs (1962) for details. Introduced in Libengué Dobélé-Kpoka (2013), the type of unimodal extended beta kernel, with shape parameters $a > 1$ and $b > 1$, is defined by

$$BE_{\theta(a,b;t_1,t_2)}(u) = \frac{1}{\mathscr{B}(a,b)(t_2-t_1)^{a+b-1}}(u-t_1)^{a-1}(t_2-u)^{b-1}\mathbb{1}_{[t_1,t_2]}(u).$$

Its mode and dispersion parameter are $\{(a-1)t_2 + (b-1)t_1\}/(a+b-2)$ and $1/(a+b-2)$ respectively. From the mode-dispersion method, we obtain $a(x,h) = 1 + (x-t_1)/\{(t_2-t_1)h\}$, $b(x,h) = 1 + (t_2-x)/\{(t_2-t_1)h\}$ and

$$\theta(x,h;t_1,t_2) = \left(\frac{x-t_1}{(t_2-t_1)h} + 1, \frac{t_2-x}{(t_2-t_1)h} + 1\right), \forall x \in [t_1,t_2],\ h > 0. \tag{12}$$

This leads to the associated extended beta kernel defined on $\mathbb{S}_{EB_{\theta(x,h,t_1,t_2)}} = [t_1, t_2]$ by

$$EB_{\theta(x,h,t_1,t_2)}(u) = \frac{(u-t_1)^{(x-t_1)/\{(t_2-t_1)h\}}(t_2-u)^{(t_2-x)/\{(t_2-t_1)h\}}}{(t_2-t_1)^{1+h^{-1}}\mathscr{B}\left(1+(x-t_1)/\{(t_2-t_1)h\}, 1+(t_2-x)/\{(t_2-t_1)h\}\right)}.$$

From the mean and variance of $EB_{\theta(a,b;t_1,t_2)}$ defined respectively by $t_1 + a(t_2-t_1)/(a+b)$ and $ab(t_2-t_1)^2/\{(a+b)^2(a+b+1)\}$, it follows that

$$A_\theta(x,h;t_1,t_2) = \frac{h\{(t_1+t_2)-2x\}}{1+2h} \tag{13}$$

$$B_\theta(x,h;t_1,t_2) = \frac{h\{x-t_1+h(t_2-t_1)\}\{(t_2-x)+h(t_2-t_1)\}}{(1+2h)^2(1+3h)}. \tag{14}$$

2.2.2. Other kernels

Table 3 summarizes the results of calculation obtained from the construction of some associated kernels using mode-dispersion method. Table 1 and Table 2 provide different ingredients from which one constructs the associated kernels with the mode-dispersion method. We specify here that the quantity $\mathcal{Z}_{\theta(a,b)}$ in the tables aforementioned denotes a random variable with density $K_{\theta(a,b)}$ on $\mathbb{S}_{\theta(a,b)}$. Regarding beta and gamma kernels, our results match with those of Chen (1999) and Chen (2000). We specify in the case of inverse Gaussian kernels and its reciprocal that the version of Scaillet (2004) is in fact our modified version that we present in Table 7. Also, we introduced inverse gamma and lognormal kernels for densities with support $(0, \infty)$ as well as the Pareto kernel for extreme cases (e.g. Markovich (2007)). From Part (b) of Figure 1 one can observe the inside disfunctioning of inverse Gaussian and inverse gamma kernels.

3. Associated kernel density estimators

We now investigate the main properties of the associated kernel estimators for density functions. Let $X_1, X_2, \cdots, X_n$ be a sequence of a random variable X of an unknown p.d.f. f in $\mathbb{T} \subseteq \mathbb{R}$. The associated kernel estimator $\widehat{f_n}$ of f is defined by

$$\widehat{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\theta(x,h)}(X_i), \quad \forall x \in \mathbb{T}. \tag{15}$$

This estimator (15) is similar to that of (2), except that (15) denotes the kernel estimator derived from the mode-dispersion method in the remainder of the paper.

3.1. Properties

The non-standardization of the associated kernel estimators comes, for instance, from the fact that its total mass is not always equal to 1 and it should be normalized before using for density estimation but not for regression estimation. Which does not the case with regard to the classical estimators (18).

Proposition 3. *Let $\widehat{f_n}$ be an associated kernel estimator (15) of f. For all $x \in \mathbb{T}$, $h > 0$, one has:*

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\}, \tag{16}$$

$$\int_{\mathbb{T}} \widehat{f_n}(x)dx =: \Lambda(n,h,K), \tag{17}$$

where the total mass $\Lambda(n,h,K)$ depends on the sample, type of kernel and smoothing parameter such that it is not always equal to 1.

Proof. The first result (16) is obtained in a straightforwardly way as follows:

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}} K_{\theta(x,h)}(t)f(t)dt = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}} f(t)K_{\theta(x,h)}(t)dt = \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\}.$$

The second result (**??**) stems from the fact that $K_{\theta(x,h)}$ is a p.d.f. Finally we have (17) as

$$\int_{\mathbb{T}} \widehat{f_n}(x)dx = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{T}} K_{\theta(x,h)}(X_i)dx,$$

since with respect to x, $K_{\theta(\cdot,h)}(t)$ is not necessarily a probability density for a given $t = X_i$.■

From this proposition, one can note that for non-classical kernels, the total mass $\Lambda(n,h,K) = \Lambda_n$ of $\widehat{f_n}$ is positive and fails to be equal to 1 in general see, e.g., Cherfaoui *et al.* (2015). Table 5 permit to observe that $\Lambda_n \in [0.9, 1.1]$. Since Λ_n is around 1, we study $x \mapsto \widehat{f_n}(x)$ without normalizing constant which can be used at the end of the estimation process of density. In the classical case, it is easy to check that $\Lambda_n = 1$ by making the change of variables $u = (x - X_i)/h$ as follows:

$$\int_{\mathbb{T}=\mathbb{R}} \widehat{f_n}(x)dx = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{T}=\mathbb{R}} K\left(\frac{x-X_i}{h}\right)dx = \frac{1}{h}\int_{\mathbb{T}=\mathbb{R}} K(u)\,hdu = 1. \tag{18}$$

Proposition 4. *Let $\widehat{f_n}$ be an associated kernel estimator (15) of f in the class $\mathscr{C}^2(\mathbb{T})$. For all x in $\mathbb{T}$ and $h = h_n > 0$, then*

$$Bias\left\{\widehat{f_n}(x)\right\} = A_\theta(x,h)f'(x) + \frac{1}{2}\left\{A_\theta^2(x,h) + B_\theta(x,h)\right\}f''(x) + o(h^2). \tag{19}$$

Furthermore, if f is bounded on $\mathbb{T}$ then

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)\left\|K_{\theta(x,h)}\right\|_2^2 + o\left(\frac{1}{nh^{r_2}}\right), \tag{20}$$

where $r_2 = r_2\left(K_{\theta(x,h)}\right) > 0$ is the largest real number such that $\left\|K_{\theta(x,h)}\right\|_2^2 = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)du \leq c_2(x)h_n^{-r_2}$ and $0 < c_2(x) < \infty$.

Proof. From (16) and by using Taylor's formula successively around $\mathbb{E}\left(\mathcal{Z}_{\theta(x,h)}\right)$ and x, the result (19) can be shown by

$$\begin{aligned}
Bias\left\{\widehat{f_n}(x)\right\} &= \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\} - f(x)\\
&= f\left\{\mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\} + \frac{1}{2}Var(\mathcal{Z}_{\theta(x,h)})f''\left\{\mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\} - f(x)\\
&\quad + o\left(\mathbb{E}\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right)\\
&= f\left\{x + A_\theta(x,h)\right\} + \frac{1}{2}B_\theta(x,h)f''\left\{x + A_\theta(x,h)\right\} - f(x)\\
&\quad + o\left\{B_\theta(x,h)\right\}\\
&= A_\theta(x,h)f'(x) + \frac{1}{2}\left\{A_\theta^2(x,h) + B_\theta(x,h)\right\}f''(x) + o(h^2).
\end{aligned}$$

In fact, the rest $o(h^2)$ comes from (8) and $o\left(\mathbb{E}\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right) = \mathbb{E}\left(o_P\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right)$ where $o_P(\cdot)$ is the probability rate of convergence. Concerning the variance we have

$$\begin{aligned}
Var\left\{\widehat{f_n}(x)\right\} &= \frac{1}{n}\mathbb{E}\left\{K^2_{\theta(x,h)}(X_1)\right\} - \frac{1}{n}\left[\mathbb{E}\left\{K_{\theta(x,h)}(X_1)\right\}\right]^2\\
&= \frac{1}{n}\int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)f(u)du - \frac{1}{n}\left[\mathbb{E}\left\{K_{\theta(x,h)}(X_1)\right\}\right]^2\\
&= I_1 - I_2.
\end{aligned}$$

From (16) and (19) one has the following behavior of the second term $I_2 := (1/n)\left[\mathbb{E}\left\{K_{\theta(x,h)}(X_1)\right\}\right]^2 \simeq (1/n)f^2(x) \simeq O(1/n)$ since f is bounded for all $x \in \mathbb{T}$. By using Taylor's expansion around x, the first term I_1 gives

$$I_1 := \frac{1}{n}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)f(u)du = \frac{1}{n}f(x)\int_{\mathbb{S}_{x,h}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)du + R(x,h),$$

with

$$R(x,h) = \frac{1}{n}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)\left[(u-x)f'(x) + \frac{(u-x)^2}{2}f''(x) + o\{(u-x)^2\}\right]du.$$

Under the assumption of $\left\|K_{\theta(x,h)}\right\|_2^2 \le c_2(x)h_n^{-r_2}$ we deduce successively

$$0 \le R(x,h) \le \frac{1}{nh^{r_2}}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}} c_2(x)\left\{(u-x)f'(x) + \frac{(u-x)^2}{2}f''(x)\right\}du \simeq o(n^{-1}h^{-r_2}).\blacksquare$$

An illustration of the calculation of r_2 will be given on the particular case of the modified lognormal kernel for edge points in Section 5.

Here, we define the appropriate measure for assessing the similarity of the associated kernel estimator $\widehat{f_n}$ with respect to the true density f, to be estimated. We remind the reader that the most natural measure is the mean integrated square error (MISE). Thus, we first define the mean squared error (MSE) by

$$MSE(x) = Var\left\{\widehat{f_n}(x)\right\} + Bias^2\left\{\widehat{f_n}(x)\right\}. \tag{21}$$

The integrated form of MSE on $\mathbb{T}$ and its approximate expressions are respectively given by:

$$MISE(\widehat{f}_{n,h,K,f}) = \int_{\mathbb{T}}\left[Var\left\{\widehat{f_n}(x)\right\} + Bias^2\left\{\widehat{f_n}(x)\right\}\right]dx$$

and

$$\begin{aligned} AMISE(\widehat{f}_{n,h,K,f}) = &\int_{\mathbb{T}}\left(\left[A_\theta(x,h)f'(x) + \frac{1}{2}\{A_\theta^2(x,h) + B_\theta(x,h)\}f''(x)\right]^2\right)dx \\ &+ \frac{1}{n}\left\|K_{\theta(x,h)}\right\|_2^2 f(x)dx. \end{aligned} \tag{22}$$

The next proposition gives the rate of convergence in the sense of $AMISE$.

Proposition 5. *Suppose that $f \in \mathscr{C}^2$, with first and second derivative being bounded. Then the optimal bandwidth that minimizes the AMISE is*

$$h = C(x)n^{-1/(r_2+2)}$$

with $r_2 = r_2\left(K_{\theta(x,h)}\right)$ used in (20).

Proof. Definition 1 and Proposition 1 allow to say that there exists two positive and finite constants $c_1^*(x)$ and $c_1^{**}(x)$ such that $A_\theta(x,h) \le c_1^*(x)h$ and $B_\theta(x,h) \le c_1^{**}(x)h^2$. Using Proposition 4, one has:

$$Bias\left\{\widehat{f_n}(x)\right\} \le hc_1(x) \text{ and } Var\left\{\widehat{f_n}(x)\right\} \le n^{-1}h^{-r_2}c_2(x),$$

with $c_1(x) = \sup_{x\in\mathbb{T}}|c_1^*(x)f'(x) + c_1^{**}(x)f''(x)|$. From (21), it follows that

$$MSE(x) \le h^2c_1^2(x) + n^{-1}h^{-r_2}c_2(x).$$

By integrating (22), one obtains

$$AMISE(\widehat{f}_{n,h,K,f}) \leq h^2 C_1(x) + n^{-1}h^{-r_2}C_2(x),$$

with $C_1(x)$ and $C_2(x)$ the anti-derivatives of respectively, $c_1^2(x)$ and $c_2(x)$ on $\mathbb{T}$. Taking the second member equal to 0 leads to the result.■

3.2. *Reduction of bias*

Naturally, the presence of the non-null term $A_\theta(x,h)$ in (19) increases the bias of $\widehat{f_n}(x)$. Thus, we propose in the following section an algorithm, inspired to the Chen (1999) and Chen (2000), but especially developed in Libengué Dobélé-Kpoka (2013), for eliminating the term $A_\theta(x,h)f(x)$ in the largest region of $\mathbb{T}$. For reducing the bias of $\widehat{f_n}(x)$ defined in (19), one proceeds in two steps. The first step consists to define both inside and boundary regions and the second step deals to the modification of the associated kernel which leads to the inside bias reduction.

First step : One divides this support $\mathbb{T} = [t_1, t_2]$ in two regions of order $\alpha(h) > 0$ with $\alpha(h) \to 0$ as $h \to 0$;

(i) *interior region* (the largest one in order to contain at least 95% of observations) denoted by $\mathbb{T}_{\alpha(h),0}$ and defined as interval

$$\mathbb{T}_{\alpha(h),0} = (t_1 + \alpha(h), t_2 - \alpha(h)),$$

(ii) *boundary regions* (can be empty) denoted by two intervals $\mathbb{T}_{\alpha(h),-1}$ and $\mathbb{T}_{\alpha(h),+1}$ respectively defined by

$$\mathbb{T}_{\alpha(h),-1} = [t_1, t_1 + \alpha(h)] \text{ (left boundary region)},$$

and

$$\mathbb{T}_{\alpha(h),+1} = [t_2 - \alpha(h), t_2] \text{ (right boundary region)}.$$

We can denote them as the complementary sets of the interior region: $\mathbb{T}^c_{\alpha(h),0} = \mathbb{T}_{\alpha(h),-1} \cup \mathbb{T}_{\alpha(h),+1}$.

Second step : We modify the associated kernel $K_\theta(x,h)$ corresponding to $A(x,h)$ and $B(x,h)$ with a new kernel function noted $K_{\widetilde{\theta}(x,h)}$ corresponding to $\widetilde{A}(x,h) = \left(\widetilde{A}_{-1}(x,h), \widetilde{A}_0(x,h), \widetilde{A}_{+1}(x,h)\right)$ and $\widetilde{B}(x,h) = \left(\widetilde{B}_{-1}(x,h), \widetilde{B}_0(x,h), \widetilde{B}_{+1}(x,h)\right)$ such that, for any fixed h,

$$\widetilde{\theta}(x,h) = \left(\widetilde{\theta}_{-1}(x,h), \widetilde{\theta}_0(x,h), \widetilde{\theta}_{+1}(x,h)\right) = \begin{cases} \widetilde{\theta}_{-1}(x,h) & \text{if } x \in \mathbb{T}_{\alpha(h),-1} \\ \widetilde{\theta}_0(x,h): \quad \widetilde{A}_0(x,h) = 0 & \text{if } x \in \mathbb{T}_{\alpha(h),0} \\ \widetilde{\theta}_{+1}(x,h) & \text{if } x \in \mathbb{T}_{\alpha(h),+1} \end{cases} \tag{23}$$

must be continuous on $\mathbb{T}$ and constant on $\mathbb{T}^c_{\alpha(h),0}$ (form personal communication with Chen). Some illustrations are given in Section 4. The following proposition shows that the modified kernel function $K_{\widetilde{\theta}(x,h)}$ with support $\mathbb{S}_{\widetilde{\theta}(x,h)} = \mathbb{S}_{\theta(x,h)}$ is an associated kernel.

Proposition 6. *The function $K_{\widetilde{\theta}(x,h)}$ obtained from (23) is an associated kernel.*

Proof. We are going to show that $K_{\widetilde{\theta}(x,h)}$ satisfies all conditions of Definition 1. Since $K_{\theta(x,h)}$ is an associated kernel and from the first step of the bias reduction we have for $j \in J = \{-1, 0, +1\}$,

$$x \in \mathbb{T}_{\alpha(h),j} \Rightarrow x \in \mathbb{T} \Rightarrow x \in \mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\widetilde{\theta}(x,h)}.$$

Using Proposition 2 it follows that for a given random variable $\mathcal{Z}_{\widetilde{\theta}(x,h)}$ with pdf $K_{\widetilde{\theta}(x,h)}$ we obtain the last two condition as,

$$\mathbb{E}(\mathcal{Z}_{\widetilde{\theta}(x,h)}) = x + A_{\widetilde{\theta}(x,h)} \text{ and } Var(\mathcal{Z}_{\widetilde{\theta}(x,h)}) = B_{\widetilde{\theta}(x,h)}$$

with both $A_{\widetilde{\theta}(x,h)}$ and $B_{\widetilde{\theta}(x,h)}$ go to 0 as h goes to 0. Taking $\widetilde{A}(x,h) = A_{\widetilde{\theta}(x,h)}$ and $\widetilde{B}(x,h) = B_{\widetilde{\theta}(x,h)}$ it follows that for a fixed $j \in J$, $\widetilde{A}_j(x,h) = A_{\widetilde{\theta}_j(x,h)}$ and $\widetilde{B}_j(x,h) = B_{\widetilde{\theta}_j(x,h)}$. In particular, from (23) we have $A_{\widetilde{\theta}_0(x,h)} = 0$.■

So, we define the *modified associated kernel density estimator* $\widetilde{f_n}$ of f using $K_{\widetilde{\theta}(x,h)}$ by

$$\widetilde{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\widetilde{\theta}(x,h)}(X_i)\,. \tag{24}$$

The following proposition gives the new expressions of bias and variance of $\widetilde{f_n}$.

Proposition 7. *Let $\mathbb{T} = \cup_{j\in J}\mathbb{T}_{\alpha(h),j}$ ($J = \{-1, 0, +1\}$) be the support of the p.d.f. f to be estimated, $\widehat{f_n}$ and $\widetilde{f_n}$ the associated kernel estimators of f defined in (15) and (24) respectively. For all $x \in \mathbb{T}_{\alpha(h),0}$ and $h > 0$ then,*

$$Bias\left\{\widetilde{f_n}(x)\right\} = \frac{1}{2}\widetilde{B}_0(x,h)f''(x) + o(h^2),$$

and

$$Var\left\{\widetilde{f_n}(x)\right\} \simeq Var\left\{\widehat{f_n}(x)\right\} \text{ as } h \to 0.$$

Proof. We obtain the first result by replacing in (19) respectively, $\widehat{f_n}$, A_θ and B_θ by $\widetilde{f_n}$, $\widetilde{A}_0$ and $\widetilde{B}_0$. For the last result, considering (20) it suffices to show that

$$\left\|K_{\theta(x,h)}\right\|_2^2 \simeq \left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \text{ as } h \to 0.$$

Indeed, since $K_{\theta(x,h)}$ and $K_{\widetilde{\theta}(x,h)}$ are associated kernels of the same type K, $r_2 = r_2(K_{\theta(x,h)})$ and $\widetilde{r}_2 = \widetilde{r}_2(K_{\widetilde{\theta}(x,h)})$, there exists a common largest real number $r = r(K)$ such that

$$\left\|K_{\theta(x,h)}\right\|_2^2 \leq c_2(x)h^{-r} \text{ and } \left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \leq \widetilde{c}_2(x)h^{-r},$$

with $0 < c_2(x), \widetilde{c}_2(x) < \infty$. Taking $c(x) = \sup\{c_2(x), \widetilde{c}_2(x)\}$ we have $\left\|K_{\theta(x,h)}\right\|_2^2 \leq c(x)h^{-2r}$ and $\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \leq c(x)h^{-2r}$. Since $c(x)/nh^{2r} = o(n^{-1}h^{-2r})$ therfore $\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \simeq \left\|K_{\theta(x,h)}\right\|_2^2$.■

Thus, we define the asymptotic expression of the MISE of $\widetilde{f_n}$ on $\mathbb{T}_{\alpha(h),0}$ as follows:

$$AMISE_0(\widetilde{f}_{n,h,K,f}) = \int_{\mathbb{T}_{\alpha(h),0}} \left[\frac{1}{4}\widetilde{B}_0^2(x,h)\{f''(x)\}^2 + \frac{1}{n}\left\|K_{\theta(x,h)}\right\|_2^2 f(x)\right] dx.$$

3.3. *Choices of associated kernel and bandwidth*

In opposite to the classical associated kernel, the choice of the non-classical associated kernel is very important. Firstly, it depends on a prior knowledge of the support $\mathbb{T}$ of the density f, to be estimated, which has to coincide with the support $\mathbb{S}_\theta$ of the (non-classical) associated kernel. In addition, this choice must take into account the behavior of the associated kernel according different positions of the target x in $\mathbb{T}$. For several associated kernels with the same support $\mathbb{S}_\theta = \mathbb{T}$, the most appropriate kernel $K_{\theta\,opt}$ is one that checks for a fixed $h > 0$ if:

$$K_{\theta\,opt} = \arg\min_{K_\theta \text{ on } \mathbb{T}_{\alpha(h),0}, \mathbb{T}^c_{\alpha(h),0}} AMISE_0(\widetilde{f}_{n,h,K}).$$

For $\mathbb{T} = \mathbb{R}$, then $K_{\theta\,opt}$ is the Epanechnikov kernel. The optimal bandwidth parameter obtained in Proposition 5 cannot be used in practice because it depends on the unknown density function. Several methods exist for selecting bandwidth parameters for associated kernel density estimators. In our case, we propose to use for instance the least squares cross-validation (LSCV) method to select the bandwidth. This technique has been developed by several authors; see Rudemo (1982), Stone (1984),Bowman (1984), and Marron (1987) for details. The LSCV method is based on the minimization of the integrated squared error (ISE) which is defined as

$$ISE(h) = \int_{\mathbb{T}} \widehat{f_n}^2(x)dx - 2\int_{\mathbb{T}} \widehat{f_n}(x)f(x)dx + \int_{\mathbb{T}} f^2(x)dx.$$

Because the last term does not depend on the bandwidth parameters, minimizing the ISE boils down to minimizing the two first terms. However, we need to estimate the second term since it depends on the unknown density function f. The LSCV estimator of $ISE(h) - \int_{\mathbb{T}} f^2(x)dx$ is

$$\begin{aligned} LSCV(h) &= \int_{x\in\mathbb{T}} \left\{\widehat{f_n}(x)\right\}^2 dx - \frac{2}{n}\sum_{i=1}^n \widehat{f}_{n,-i}(X_i) \\ &= \int_{x\in\mathbb{T}} \left\{\frac{1}{n}\sum_{i=1}^n K_{\theta(x,h)}(X_i)\right\}^2 dx - \frac{2}{n}\sum_{i=1}^n \widehat{f}_{n,-i}(X_i), \end{aligned}$$

where $\widehat{f}_{n,-i}(X_i) = (n-1)^{-1}\sum_{j\neq i} K_{\theta(X_i,h)}(X_j)$ is being computed as $\widehat{f_n}(X_i)$ excluding the observation X_i. The bandwidth LSCV rule selection is defined as follows:

$$h_{cv} = \arg\min_{h>0} LSCV(h).$$

4. Illustration of modified associated kernel density estimators

Now, we illustrate the bias reduction for particular continuous associated kernels already used as examples in Section 2.2. Only the case extended beta is given in detail. Other results are provided in Table **??** and Table 7.

4.1. Extended beta density estimators

This kernel estimator is appropriate for densities having support $\mathbb{T} = [t_1, t_2] = \mathbb{S}_{EB}$. From (13), (14) and (19) we have

$$Bias\left\{\widehat{f_{n;t_1,t_2}}(x)\right\} = \frac{(t_1+t_2-2x)h}{1+2h} f'(x) + \frac{1}{2}\left\{\frac{(t_1+t_2-2x)h}{1+2h}\right\}^2 f''(x) + \frac{\{x-t_1+(t_2-t_1)h\}\{t_1-x+(t_2-t_1)h\}}{2(1+2h)^2(1+3h)} f''(x) + o(h^2).$$

Being large, this bias must be reduced. Using the first step of bias reduction one has the three intervals

$$\mathbb{T}_{\alpha(h),-1} = [t_1, t_1+\alpha(h)], \quad \mathbb{T}_{\alpha(h),0} = (t_1+\alpha(h), t_2-\alpha(h)) \quad \text{and} \quad \mathbb{T}_{\alpha(h),+1} = [t_2-\alpha(h), t_2].$$

For the second step, we first consider the function $\psi : \mathbb{T} \to \mathbb{T}$ defined by

$$\psi(z) = \{z - \alpha(h) + 1\}\alpha(h) \ \text{ for all } \ z \geq 0. \tag{25}$$

We construct $\widetilde{\theta}(x,h) = \widetilde{\theta}(x,h;t_1,t_2)$ in the mind of (23) as follows:

$$\widetilde{\theta}(x,h;t_1,t_2) = \begin{cases} \left(\dfrac{\psi(x-t_1)}{(t_2-t_1)h}, \dfrac{x-t_1}{(t_2-t_1)h}\right) & \text{if } x \in [t_1, t_1+\alpha(h)] \\ \left(\dfrac{x-t_1}{(t_2-t_1)h}, \dfrac{t_2-x}{(t_2-t_1)h}\right) & \text{if } x \in (t_1+\alpha(h), t_2-\alpha(h)) \\ \left(\dfrac{t_2-x}{(t_2-t_1)h}, \dfrac{\psi(t_2-x)}{(t_2-t_1)h}\right) & \text{if } x \in [t_2-\alpha(h), t_2]. \end{cases}$$

By calculating the mean and variance of $BE_{\widetilde{\theta}(x,h;t_1,t_2)}$ it follows that

$$A_{\widetilde{\theta}}(x,h,t_1;t_2) = \begin{cases} \dfrac{x(x-t_1)+(1-x)\psi(x-t_1)}{x-t_1+\psi(x-t_1)} & \text{if } x \in [t_1, t_1+\alpha(h)] \\ 0 & \text{if } x \in (t_1+\alpha(h), t_2-\alpha(h)) \\ \dfrac{x(t_2-x)+(1-x)\psi(t_2-x)}{t_2-x+\psi(t_2-x)} & \text{if } x \in [t_2-\alpha(h), t_2], \end{cases}$$

and

$$B_{\widetilde{\theta}}(x,h,t_1;t_2) = \begin{cases} \dfrac{(x-t_1)\psi(x-t_1)h\{x-t_1+\psi(x-t_1)\}^{-2}}{\{x-t_1+\psi(x-t_1)+h\}} & \text{if } x \in [t_1, t_1+\alpha(h)] \\ \dfrac{(x-t_1)(t_2-x)h}{1+h} & \text{if } x \in (t_1+\alpha(h), t_2-\alpha(h)) \\ \dfrac{(t_2-x)\psi(t_2-x)\}h\{(t_2-x)+\psi(t_2-x)\}^{-2}}{\{(t_2-x)+\psi(t_2-x)+h\}} & \text{if } x \in [t_2-\alpha(h), t_2]. \end{cases}$$

This allows us to obtain the reduced expression of the bias in interval $(t_1+\alpha(h), t_2-\alpha(h))$ as

$$Bias\left\{\widetilde{f}_{n;t_1,t_2}(x)\right\} = \frac{1}{2}(x-t_1)(t_2-x)hf''(x) + o(h^2).$$

Moreover, the calculation of $\left\|BE_{\widetilde{\theta}(x,h;t_1,t_2)}\right\|_2^2$ yields:

$$\left\|BE_{\theta(x,h;t_1,t_2)}\right\|_2^2 = \frac{\mathscr{B}\left(1+2(x-t_1)/(t_2-t_1)h, 1+2(t_2-x)/(t_2-t_1)h\right)}{\left\{\mathscr{B}\left(1+(x-t_1)/(t_2-t_1)h, 1+(t_2-x)/(t_2-t_1)h\right)\right\}^2}.$$

Therefore, from (20) the variance of $\widehat{f}_{n;t_1,t_2}(x)$ is:

$$\begin{aligned} Var\left\{\widehat{f}_{n;t_1,t_2}(x)\right\} &= \frac{\mathscr{B}\left(1+2(x-t_1)/(t_2-t_1)h, 1+2(t_2-x)/(t_2-t_1)h\right)}{\left\{\mathscr{B}\left(1+(x-t_1)/(t_2-t_1)h, 1+(t_2-x)/(t_2-t_1)h\right)\right\}^2}\frac{1}{n}f(x) \\ &+ o\left(\frac{1}{nh^{r_2}}\right) \end{aligned}$$

In order to determine r_2 and an explicit expression of the variance of $\widehat{f}_{n;t_1,t_2}$, let us introduce the R function of Brown and Chen (1998) defined as follows, for $z \geq 0$:

$$R(z) = \frac{1}{\Gamma(z+1)}\left(\frac{z}{e}\right)^z \sqrt{2\pi z}. \tag{26}$$

It is shown in Brown and Chen (1998) that R is an increasing function such that $R(z) < 1$ and $R(z) \to 1$ as $z \to \infty$. Using (26) and similarly as in Chen (1999), one obtains

$$\left\|BE_{\widehat{\theta}(x,h;t_1,t_2)}\right\|_2^2 = \begin{cases} \dfrac{\Gamma(2c+1)}{2^{2c+1}\Gamma^2(c+1)} & \text{if } \dfrac{x-t_1}{(t_2-t_1)h} \text{ or } \dfrac{t_2-x}{(t_2-t_1)h} \to c \\ \dfrac{(x-t_1)^{-1/2}(t_2-x)^{-1/2}}{2\sqrt{\pi}(t_2-t_1)^{-1}h^{1/2}} & \text{if } \dfrac{x-t_1}{(t_2-t_1)h}, \dfrac{t_2-x}{(t_2-t_1)h} \to \infty, \end{cases}$$

Hence

$$Var\left\{\widehat{f}_{n;t_1,t_2}(x)\right\} = \begin{cases} \dfrac{\Gamma(2c+1)f(x)}{2^{2c+1}\Gamma^2(c+1)n} + O(n^{-1}) & \text{if } \dfrac{x-t_1}{(t_2-t_1)h} \text{ or } \dfrac{t_2-x}{(t_2-t_1)h} \to c \\ \dfrac{(x-t_1)^{-1/2}(t_2-x)^{-1/2}f(x)}{2\sqrt{\pi}(t_2-t_1)^{-1}nh^{1/2}} + o\left(\dfrac{1}{nh^{1/2}}\right) & \text{if } \dfrac{x-t_1}{(t_2-t_1)h}, \dfrac{t_2-x}{(t_2-t_1)h} \to \infty. \end{cases}$$

It follows that $r_2 = 1/2$. From Proposition 7, one obtains the variance of $\widetilde{f}_{n;t_1,t_2}$ in the interior region $(t_1 + \alpha(h), t_2 - \alpha(h))$ as

$$Var\left\{\widetilde{f}_{n;t_1,t_2}(x)\right\} = \frac{(x-t_1)^{-1/2}(t_2-x)^{-1/2}}{2\sqrt{\pi}(t_2-t_1)^{-1}nh^{1/2}} f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

Taking $t_1 = 0$ and $t_2 = 1$, one deduces similar results of the (modified) beta density estimators of Chen (1999), see Libengué Dobélé-Kpoka (2013) for more details.

4.2. Other associated kernel density estimators

Table **??** and Table 7 summarize the results of different calculation for bias reduction as in the previous case. We indicate that our parametrizations at the edges are depending on $\alpha(h)$ which is more general. The user can set the value of $\alpha(h)$ according to his objective. For example, we have $\alpha(h) = 2h$ in Chen (1999), Chen (2000) and $\alpha(h) = h$ in Zhang and Karunamuni (2009) and Zhang (2010). We clarify also that Scaillet did not use this concept in his work on inverse Gaussian kernel and its reciprocal. He directly used the reduced versions of these kernels that we give in Table 7. Also, the support of Pareto kernel should be divided into $[x, x + \alpha(h)) \cup [x + \alpha(h), \infty)$ but since the target x is always the left edge of the support, then we directly study its modified version on $[x, \infty)$.

5. Simulation studies for lognormal kernel density estimators

This section presents the results of simulation studies for three lognormal kernel density estimators $\widehat{f}_{n,LN}$, $\widetilde{f}_{n,LN}$ and $f^*_{n,LN}$ of f on $\mathbb{T} = (0, \infty)$ corresponding to the three following lognormal kernels LN_θ, $LN_{\widetilde{\theta}}$ and LN_{θ^*} respectively. Before proceeding to these simulation studies, we first describe the different associated kernels from which the estimators mentioned above are defined. In particular, we prove r_2 of Proposition 4 for the second lognormal estimator at the boundary region.

From the parametrized lognormal density function in Table 2, the lognormal kernel LN_θ is constructed by the mode-dispersion method, $LN_{\widetilde{\theta}}$ is its modified version and LN_{θ^*} is the extracted version in the lognormal density estimator of Jin and Kawczak (2003). The mode-dispersion lognormal kernel $LN_{\theta(x,h)}$ is defined by

$$LN_{\theta(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(\frac{-1}{2h^2}\left[\log u - \log\left\{x \exp\left(h^2\right)\right\}\right]\right)\mathbb{I}_{u>0}, \quad x > 0, h > 0. \tag{27}$$

For $\alpha(h) > 0$ bounding both left edge and inside regions of $\mathbb{T} = (0, \infty) = (0, \alpha(h)] \cup (\alpha(h), \infty)$, the modified version $LN_{\widetilde{\theta}(x,h)} := \widetilde{LN}_{\theta(x,h)}$ of $LN_{\theta(x,h)}$ is given by

$$\begin{aligned} LN_{\widetilde{\theta}(x,h)}(u) &= LN_{\widetilde{\theta}_{-1}(x,h)}(u) + LN_{\widetilde{\theta}_0(x,h)}(u) \\ &= \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{\frac{\alpha^3(h)\exp\left(-h^2/2\right)}{x^2}\right\}\right]^2\right)\mathbb{I}_{(0,\alpha(h)]}(u) \\ &\quad + \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{x\exp\left(-h^2/2\right)\right\}\right]^2\right)\mathbb{I}_{(\alpha(h),\infty)}(u), \end{aligned} \tag{28}$$

where the left boundary part $\widetilde{\theta}_{-1}(x,h)$ and the inside part $\widetilde{\theta}_0(x,h)$ are given in Table **??**. The third version of lognormal kernel which has been considered by Jin and Kawczak (2003) without showing the method of its construction is defined as follows:

$$LN_{\theta^*(x,h)}(u) = \frac{1}{2u\sqrt{2\pi\log(1+h)}}\exp\left\{-\frac{\log^2(u/x)}{8\log(1+h)}\right\}\mathbb{I}_{(0,\infty)}(u). \tag{29}$$

It is linked to the expression (27) by

$$LN_{\theta^*(x,h)} = LN_{\theta(x\exp(h^2),2\sqrt{\log(1+h)})}, \quad x>0, h>0;$$

thus, the expression $2\sqrt{\log(1+h)})$ corresponds to h in the first one and both are proportional for small h. Denoting by $\widehat{r}_2 = r_2(\widehat{f}_{n,LN})$, $\widetilde{r}_2 = r_2(\widetilde{f}_{n,LN})$ and $r_2^* = r_2(f^*_{n,LN})$ the r_2 of these lognormal kernel estimators as in Proposition 4, we have the following results:

$$\widehat{r}_2 = \widetilde{r}_{2,0} = 2r_2^* = 1 \text{ and } \widetilde{r}_{2,-1} = \widetilde{r}_{2,-1}(\alpha); \tag{30}$$

in particular, if $\alpha(h) = \alpha_1 h^\beta$ then $\widetilde{r}_{2,-1}(\alpha) = (3\beta\sqrt{2}+2)/2$ for $\beta > -\sqrt{2}/3$. Indeed, we only prove the last part of (30): $\widetilde{r}_{2,-1} = \widetilde{r}_{2,-1}(\alpha)$. Since

$$\mathbb{E}(\mathcal{Y}^m) = \exp\left[m\left\{h^2+\log(x)\right\} + (mh)^2/2\right], \ \forall m \in \mathbb{R},$$

for any lognormal random variable $\mathcal{Y} \sim LN_{\theta(x,h)}$, one gets consecutively

$$\begin{aligned}
\left\|LN_{\widetilde{\theta}_{-1}(x,h_n)}\right\|_2^2 &= \int_0^{\{\alpha(h_n)\}^{\sqrt{2}}} \frac{v^{-\sqrt{2}/2}dv}{vh_n^2\pi\sqrt{8}}\exp\left(\frac{-1}{2h_n^2}\left[\log v - \log\left\{\frac{\{\alpha(h_n)\}^{3\sqrt{2}}}{x^{2\sqrt{2}}e^{\sqrt{2}h_n^2/2}}\right\}\right]^2\right)\\
&\leqslant \frac{1}{2h_n\sqrt{2\pi}}\mathbb{E}\left(\mathcal{Y}_\star^{-\sqrt{2}/2}\right) = \frac{1}{2h_n\sqrt{2\pi}}\left\{\frac{x^2}{\alpha^3(h_n)}\right\}^{\sqrt{2}/2}\exp\left(\frac{\sqrt{2}+1}{4}h_n^2\right)\\
&\leqslant \frac{x^{\sqrt{2}/2}\left[1+\left\{\left(\sqrt{2}+1\right)/4\right\}h_n^2\right]}{2\sqrt{\pi}h_n\{\alpha(h_n)\}^{3\sqrt{2}/2}},
\end{aligned}$$

with $\mathcal{Y}_\star \sim LN_{\theta(x^\star,h_n)}$ such that

$$x^\star = x^{-\sqrt{2}/2}\{\alpha(h_n)\}^{3\sqrt{2}}\exp\left(-\sqrt{2}h_n^2/2 - h_n^2\right);$$

and, therefore, $\widetilde{r}_{2,-1}(\alpha)$ is the largest power in h_n of $h_n\{\alpha(h_n)\}^{3\sqrt{2}/2}$.

These three associated kernels are positive and more flexible than the associated gamma kernel in terms of their construction. We illustrate the finite sample behavior of these three lognormal estimators, through simulation studies. We analyze the influence of the bandwidth in the estimators mean integrated squared errors (MISEs), and we measure the amount of efficiency which is gained through the bias reduction. We study both situations of boundary and interior behaviors of these estimators on two different models. The first one is essentially the truncated normal density function $\mathcal{N}_t(\mu, \sigma; a, b)$ on interval $[a, b]$

$$\text{Model 1} : X \sim \mathcal{N}_t(0.5, 0.15; a, b);$$

and the second is a mixture of truncated normal with exponential densities

$$\text{Model 2} : X \sim 0.5\mathcal{N}_t(0.25, 0.15; a, b) + 0.5\mathcal{E}xp(2).$$

For each Model, we simulate 100 replications of size $n = 50$, 100, 500 or 1000, by taking $a = 0.01$ and $b = 0.5$ for boundary region and $a = 0.5$ and $b = 8.5$ for inside region.

In Table 8 and Table 9, we report the optimal bandwidths (in the sense of the MISE) with the corresponding times. It is seen that the optimal bandwidths decrease in general (for both models) when increasing the sample size; besides, the modified lognormal estimator $\widetilde{f}_{n,LN}$ has the smallest optimal bandwidths than the two other lognormal estimators $\widehat{f}_{n,LN}$ and $f^*_{n,LN}$.
Table 10 and Table 11 provide the biases and the variances of the these estimators at some selected points, corresponding to the boundary and inside points, respectively for Model 1 and Model 2. It is seen that for the small size sample, the modified lognormal estimator $\widetilde{f}_{n,LN}$ behaves better than $\widehat{f}_{n,LN}$ and $f^*_{n,LN}$ at the boundary and it becomes worse than them when increasing the sample size. However, taking into consideration the bias-variance tradeoff, we see that the lognormal estimator $\widehat{f}_{n,LN}$ constructed by the mode-dispersion method is better at the boundary in comparison with other and this, whatever the size of the sample. For both Models 1 and 2, we find that there are no significant differences between the pointwise variances of these three estimators. This consolidates the results of Proposition 7 of Subsection 3.2, since the three lognormal kernels belong to the same family. Also, observing the pointwise bias of these estimators, it is seen that the bias of the first two estimators $\widehat{f}_{n,LN}$ and $f^*_{n,LN}$ increase more and more inside. This is due to the presence of non-zero term $A(x, h)f'(x)$ in (19). The exact difference between their bias and those of $\widetilde{f}_{n,LN}$ shows the interest of bias reduction algorithm proposed by in Subsection 3.2. So that the modified lognormal estimator $\widetilde{f}_{n,LN}$ undoubtedly remains better in the interior region.
Finally, we report in Table 12 the minimum MISEs of these estimators. We remind the reader that the theoretical MISE function is well approximated by the average of the ISEs which can be defined for $N = 100$ (number of trials) as

$$\widehat{ISE} = \frac{1}{N}\sum_{n=1}^{N}\int_{\mathbb{T}}\left\{\widehat{f}_n(x) - f(x)\right\}^2 dx.$$

Comparing these models, one can appreciate that the density corresponding to the first one is underestimated and the second is overestimated by these three estimators. This is because we have negative and positive biases in first and the second model, respectively. However, according to Model 1, the MISEs of these estimators are smallest than those according to Model 2. The best estimators remains $\widetilde{f}_{n,LN}$ inside and $\widehat{f}_{n,LN}$ at the edges.

6. Concluding remarks

This paper proposes a new family of nonparametric estimators for density functions having compact support or not. The estimator is based on associated kernels which are dependent intrinsically on the estimated point x and the smoothing parameter h. All associated kernels are free of boundary effects but they have a bias slightly different from that Rosenblatt (1956) and Parzen (1962). We have provided a general definition and a technique of construction called "mode-dispersion" for these associated kernels as well as an algorithm of the reduction of their biases. We have illustrated this construction method and the bias reduction algorithm on the extended beta kernel which is a more general case with two boundary regions. We have examined the finite sample performance in several simulations particularly on three lognormal kernel density estimators $\widehat{f}_{n,LN}$, $\widetilde{f}_{n,LN}$ and $f^{*}_{n,LN}$ where $LN_{\theta(x,h)}$ is constructed by mode-dispersion method, $LN_{\widetilde{\theta}(x,h)}$ is its modified version and $LN_{\theta^{*}(x,h)}$ is the version proposed by Jin and Kawczak (2003). In these simulations, the optimal bandwidths are obtained by the least squares cross validation method and we found that the modified lognormal kernel density estimator $\widetilde{f}_{n,LN}$ has the smallest optimal bandwidths. In fact, it has been found that the modified lognormal kernel estimator $\widetilde{f}_{n,LN}$ is undoubtedly better in the interior region and its first version $\widehat{f}_{n,LN}$ resulting from the mode-dispersion method is better at the boundary region. This leads us to strongly recommend the use of $\widehat{f}_{n,LN}$ at the edge and $\widetilde{f}_{n,LN}$ in the interior region.

It would be interesting to compare the performance of all associated kernels with support $(0, \infty)$ at the edges and inside on densities satisfying the conditions of Shoulder as done in Zhang (2010). Since the bandwidth h plays a very important role in the performance of the associated kernel density estimators, an interesting topic for future research is to investigate automatic bandwidth selectors. Bandwidth selectors can be derived from other possible criteria as Bayesian (see, e.g., Zougab *et al.* (2012)) or the adaptative modified Lepski methods. An other interesting topic is to extend this construction to the multivariate data defined on more involved support. Finally, it would be important to combine the present results with those obtained in Kokonendji and Senga Kiessé (2011) for estimating densities defined on univariate and then multivariate time-scale sets by taking into account or not the correlation for multivariate cases.

Acknowledgements. We sincerely thank the *Agence Universitaire de la Francophonie (AUF)* for funding this work.

References

Balakrishnan, N., Gupta R.C., Kundu D., Leiva V. and Sanhueza A.(2011) On some mixture models base on Birnbaum-Saunder distribution and associated inference. *Journal of Statistical Planning and Inference*, **141**, 2175–2190.

Bertin K. and Klutchnikoff N.(2011) Minimax properties of beta kernel estimators. *Journal of Statistical Planning and Inference*, **141**, 2287–2297.

Bouezmarni T. and Rolin J.M.(2003) Consistency of the beta kernel density function estimator. *The Canadian Journal of Statistics*, **31**, 89–98.

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

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimate. *Biometrika*, **71**, 352-360.

Brown, B.M. and Chen, S.X. (1998). 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 and Data Analysis*, **31**, 131-145.

Chen, S.X. (2000). Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics*, **52**, 471-480.

Cherfaoui, M. Boualem, M. Aïssani, D. and Adjabi, S. (2015). Quelques propriétés des estimateurs à noyaux gamma pour des échantillons de petites tailles. *Afrika Statistika*, 10**(1)**, 763-776.

Devroye, L. (1987). *A Course in Density Estimation*. Boston: Birkhäuser.

Epanechnikov, V. (1969). Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications*, **14**, 153-158.

Grubbs, F.E. (1962). Attempts to validate certain PERT statistics or picking on PERT. *Operations Research*, **10**, 912-915.

Jin, X. and Kawczak, J. (2003). Birnbaum-Saunder and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, **4**, 103-1024.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman and Hall.

Jørgensen, B. and Kokonendji, C.C. (2011). Dispersion models for geometric sums. *Brazilian Journal of Probability and Statistics*, **25**, 263-293.

Jørgensen, B., Goegebeur, Y. and Martnez, J. R. (2010). Dispersion models for extremes.*Extremes*, **13**, 399-437.

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, **8**, 497-516.

Kokonendji, C.C., Senga Kiessé, T. and Balakrishnan, N. (2009). Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference*, **139**, 3625-3638.

Kokonendji, C.C. and Zocchi, S.S. (2010). Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters*, **80**, 1655-1662.

Libengué Dobélé-Kpoka, F. G. B. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*. PhD thesis manuscript (in French), Université de Franche-Comté, Besançon, France & Université de Ouagadougou, Burkina Faso, June, LMB no. 14334, Besançon, France.

Markovich, N. (2007). *Nonparametric analysis of univariate heavy-tailed data: Research and practice*. Moscow: Wiley and sons.

Marron, J.S. (1987). A comparison of cross-validation techniques in density estimation. *The Annals of Statistics*, **15**, 152–162.

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

R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. *http://www.R-project.org*.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832-837.
Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65-78.
Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics*, **16**, 217-226.
Scott, D.W. (1992). *Multivariate Density Estimation-Theory, Practice, and Visualization*. New York: Wiley.
Senga Kiessé, T. and Rivoire, M.(2010). Discrete semiparametric regression models with associated kernel and applications. *Journal of Nonparametric Statistics*, **23**, 927-941.
Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistitics*, **12**, 1285-1297.
Sylverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.
Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*. Paris: Springer.
Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary. *Statistics and Probability Letters*, **80**, 548-557.
Zhang, S. and Karunamuni, R.J. (2009). Boundary performance of the beta-kernel estimator. *Journal of Nonparametric Statistics*, **22**, 81-104.
Zougab, N., Adjabi, S. & Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics*, **24**, 783-795.
Zougab, N., Adjabi, S. & Kokonendji, C.C. (2013). A Bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, **7**, 8-23.

Now, we list all the tables and figures which are referenced in the text. In table 7, the value of the function A is given by

$$A = \frac{h(1-x)\{1-x+\psi(1-x)\}^{-2}}{\{\psi(1-x)\}^{-1}\{1-x+\psi(1-x)+h\}}.$$

Table 1. Some types of kernels (supplemented with Table 2)

	Extended beta	Beta	Gamma	Inverse gamma
$\mathbb{S}_{\theta(a,b)}$	$[t_1, t_2], \quad 0 \leq t_1 < t_2$	$[0,1]$	$[0,\infty)$	$(0,\infty)$
$K_{\theta(a,b)}$	$\dfrac{(u-t_1)^{a-1}(t_2-u)^{b-1}}{\mathscr{B}(a,b)(t_2-t_1)^{a+b-1}}$	$\dfrac{u^{a-1}(1-u)^{b-1}}{\mathscr{B}(a,b)}$	$\dfrac{b^{-a}u^{a-1}\exp\left(-\frac{u}{b}\right)}{\Gamma(a)}$	$\dfrac{b^{a}u^{-(a+1)}\exp\left(-\frac{b}{u}\right)}{\Gamma(a)}$
$M(a,b)$	$\dfrac{(a-1)t_2+(b-1)t_1}{a+b-2}$	$\dfrac{a-1}{a+b-2}$	$(a-1)b$	$\dfrac{b}{a+1}$
$D(a,b)$	$(a+b-2)^{-1}$	$(a+b-2)^{-1}$	b	b^{-1}
$\mathbb{E}\left(\mathcal{Z}_{\theta(a,b)}\right)$	$t_1+\dfrac{a(t_2-t_1)}{a+b}$	$\dfrac{a}{a+b}$	ab	$\dfrac{b}{a-1}$
$Var\left(\mathcal{Z}_{\theta(a,b)}\right)$	$\dfrac{ab(t_2-t_1)^2}{(a+b)^2(a+b+1)}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$	ab^2	$\dfrac{b^2}{(a-1)^2(a-2)}$

Table 2. Some types of kernels (topped with Table 1)

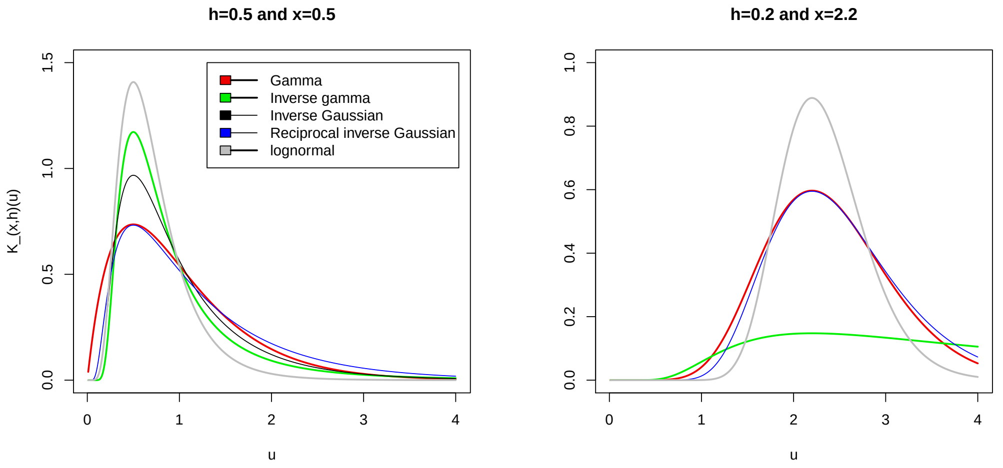
	Inverse Gaussian	Reciprocal inverse Gaussian	Pareto	Lognormal
$\mathbb{S}_{\theta(a,b)}$	$(0,\infty)$	$(0,\infty)$	$[a,\infty)$	$(0,\infty)$
$K_{\theta(a,b)}$	$\dfrac{\sqrt{b}\exp\left\{-\frac{b}{2a}\left(\frac{u}{a}-2+\frac{a}{u}\right)\right\}}{\sqrt{2\pi t^3}}$	$\dfrac{\sqrt{b}\exp\left\{-\frac{b}{2a}\left(au-2+\frac{1}{au}\right)\right\}}{\sqrt{2\pi u}}$	$\dfrac{ba^b}{u^{b+1}}$	$\dfrac{1}{ub\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\log(u)-a}{b}\right)^2\right\}$
$M(a,b)$	$a\left\{\left(1+\dfrac{9a^2}{4b^2}\right)^{1/2}-\dfrac{3a}{2b}\right\}$	$\dfrac{1}{a}\left\{\left(1+\dfrac{a^2}{4b^2}\right)^{1/2}-\dfrac{a}{2b^1}\right\}$	a	$\exp(a-b^2)$
$D(a,b)$	b^{-1}	b^{-1}	b^{-1}	b
$\mathbb{E}\left(\mathcal{Z}_{\theta(a,b)}\right)$	a	$a^{-1}+b^{-1}$	$\dfrac{ab}{b-1}$	$e^{a+b^2/2}$
$Var\left(\mathcal{Z}_{\theta(a,b)}\right)$	a^3/b	$1/(ab)+2/b^2$	$\dfrac{a^2b}{(b-1)^2(b-2)}$	$\left(e^{b^2}-1\right)e^{2a+b^2}$

Table 3. Some non-classical associated kernels for non-standard density estimators

$K_{\theta(x,h)}$	Ext. beta	Beta	Gamma	Inv. gamma	Inv. Gaussian
$\mathbb{S}_{\theta(x,h)}$	$[t_1, t_2]$	$[0,1]$	$(0,\infty)$	$(0,\infty)$	$(0,\infty)$
$\theta(x,h)$	(12)	$\left(\frac{x}{h}+1, \frac{1-x}{h}+1\right)$	$\left(\frac{x}{h}+1, h\right)$	$\left(\frac{1}{xh}-1, \frac{1}{h}\right)$	$\left(\frac{x}{\sqrt{1-3xh}}, \frac{1}{h}\right)$
$A_\theta(x,h)$	(13)	$\frac{h(1-2x)}{1+2h}$	h	$\frac{2x^2h}{1-2xh}$	$x\left\{\frac{1}{\sqrt{1-3xh}}-1\right\}$
$B_\theta(x,h)$	(14)	$\frac{h\{x(1-x)+h+h^2\}}{(1+3h)(1+2h)^2}$	$h(x+h)$	$\frac{x^3h(1-3xh)^{-1}}{(1-2xh)^2}$	$\frac{x^3h}{(1+3xh)^{3/2}}$
$r_2(x)$	1/2	1/2	1/2	1/2	1/2

Table 4. Some non-classical associated kernels for non-standard density estimators (Continuation of Table 3)

$K_{\theta(x,h)}$	Rec. inv. Gaussian	Pareto	Lognormal
$\mathbb{S}_{\theta(x,h)}$	$(0,\infty)$	$[x,\infty)$	$(0,\infty)$
$\theta(x,h)$	$\left(\frac{1}{\sqrt{x^2+xh}}, \frac{1}{h}\right)$	$\left(x, \frac{1}{h}\right)$	$(\log(x)+h^2, h)$
$A_\theta(x,h)$	$\sqrt{x^2+xh}-x+h$	$\frac{xh}{1-h}$	$x\left(e^{(3h^2)/2}-1\right)$
$B_\theta(x,h)$	$h\{\sqrt{x^2+xh}+2h\}$	$\frac{(xh)^2(1-2h)^{-1}}{(1-h)^2}$	$x^2e^{(3h^2)}\left(e^{h^2}-1\right)$
$r_2(x)$	1/2	1	1



(a) (b)

Fig. 1. Behaviors of some non-classical kernels on positive real line; in (a) the behaviors at the edges and in (b) the inside behaviors

Table 5. Some values of the total mass of $\widehat{f_n}$, $\Lambda_n \neq 1$ defined in (17) for $h = 0.05$ and $n = 1000$

Type of kernel	*sample* 1	*sample* 2	*sample* 3	*sample* 4	*sample* 5
Beta	1.028727	1.037655	1.042822	1.045671	1.046681
Gamma	0.978753	0.982778	1.009463	1.056642	0.912753
Inverse gamma	0.970302	0.922467	0.976362	0.978275	0.967052
Inverse Gaussian	1.247213	0.978136	1.034702	1.003275	1.062435
Reciprocal inverse Gaussian	0.937992	0.998741	0.993740	0.990672	0.983422
Lognormal	1.056480	0.947233	0.887284	1.139463	0.874912

Table 6. Some modified associated kernels for non-standard density estimators(supplemented with Table 7)

$K_{\widetilde{\theta}(x,h)}$	Beta: $\psi(z) = \{z - \alpha(h) + 1\}\alpha(h)$ for all $z \geq 0$.			Pareto	Lognormal	
$\mathbb{T}_{\alpha(h),j}$	$[0, \alpha(h)]$	$(\alpha(h), 1 - \alpha(h))$	$[1 - \alpha(h), 1]$	$[x, \infty)$	$[0, \alpha(h)]$	$(\alpha(h), \infty)$
$\widetilde{\theta}(x,h)$	$\left(\frac{\psi(x)}{h}, \frac{x}{h}\right)$	$\left(\frac{x}{h}, \frac{1-x}{h}\right)$	$\left(\frac{1-x}{h}, \frac{\psi(1-x)}{h}\right)$	$\left(x(1-h), \frac{1}{h}\right)$	$\left(\frac{\alpha^3(h)}{x^2}\exp(-3h^2/2), h\right)$	$(x\exp(-3h^2/2), h)$
$A_{\widetilde{\theta}}(x,h)$	$\frac{(1-x)\psi(x) + x^2}{x + \psi(x)}$	0	$\frac{(1-x)\{x - \psi(1-x)\}}{1 - x + \psi(1-x)}$	0	$\frac{\alpha^3(h) - x^3}{x^2}$	0
$B_{\widetilde{\theta}}(x,h)$	$\frac{hx\psi(x)\{x+\psi(x)\}^{-2}}{\{x + \psi(x) + h\}}$	$\frac{hx(1-x)}{1+h}$	A	$\frac{(xh)^2}{(1-2h)}$	$\frac{\alpha^6(h)}{x^4}\left(e^{h^2} - 1\right)$	$x^2(e^{h^2} - 1)$
$\widetilde{r}_2(x)$	1	1/2	1	1	$\widetilde{r}_2\{\alpha(h)\}$ (30)	1

Table 7. Some modified associated kernels for non-standard density estimators(topped with Table **??**)

$K_{\widetilde{\theta}(x,h)}$	Gamma		Inverse gamma		Inverse Gaussian		Reciprocal inverse Gaussian	
$\mathbb{T}_{\alpha(h),j}$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$
$\widetilde{\theta}(x,h)$	$\left(\frac{x^2}{2h\alpha(h)}+1,h\right)$	$\left(\frac{x}{h},h\right)$	$\left(1+\frac{x}{h\alpha^2(h)},\frac{1}{h}\right)$	$\left(1+\frac{1}{xh},\frac{1}{h}\right)$	$\left(\frac{\alpha(h)}{x^2-h\alpha(h)},\frac{1}{h}\right)$	$\left(x,\frac{1}{h}\right)$	$\left(\frac{\alpha(h)}{x^2-h\alpha(h)},\frac{1}{h}\right)$	$\left(\frac{1}{x-h},\frac{1}{h}\right)$
$A_{\widetilde{\theta}}(x,h)$	$\frac{x^2+2\alpha(h)(h-x)}{2\alpha(h)}$	0	$\frac{\alpha^2(h)-x^2}{x}$	0	$\frac{2x^2-\alpha(h)\{\alpha(h)+x\}}{\alpha(h)}$	0	$\frac{x(x-h\alpha(h))}{\alpha(h)}$	0
$B_{\widetilde{\theta}}(x,h)$	$\frac{h\{x^2+2h\alpha(h)\}}{2\alpha(h)}$	xh	$\frac{h\alpha^6(h)}{x^2(x-h\alpha^2(h))}$	$\frac{x^3h}{1-xh}$	$\frac{h\{2x^2-\alpha(h)\}^3}{\alpha^3(h)}$	x^3	$\frac{h\{x^2+h\alpha(x)\}}{\alpha(h)}$	$\frac{1}{x-h}+2h$
$\widetilde{r}_2(x)$	1	1/2	1	1/2	1	1/2	1	1/2

Table 8. Boundary optimal bandwidths by LSCV for three lognormal kernel density estimators

Model	Size	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
	n	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$
1	50	0.121	0.456	0.412	0.360	0.002	2.153
	100	0.111	0.494	0.376	0.416	0.003	2.829
	500	0.110	1.409	0.367	1.090	0.004	3.922
	1000	0.104	2.893	0.351	2.458	0.003	7.543
2	50	0.374	0.554	0.089	0.579	0.088	0.572
	100	0.366	0.553	0.087	0.772	0.087	0.906
	500	0.364	1.976	0.056	2.734	0.055	2.846
	1000	0.327	4.251	0.025	5.858	0.025	5.664

Table 9. Optimal bandwidths in interior region by LSCV for three lognormal kernel density estimators

Model	Size	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
	n	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$
1	50	0.470	0.394	0.997	0.429	0.996	0.471
	100	0.429	0.410	0.996	0.524	0.996	0.573
	500	0.406	1.151	0.991	1.513	0.993	1.695
	1000	0.400	2.605	0.991	3.456	0.991	3.718
2	50	0.456	0.639	0.997	0.630	0.996	0.622
	100	0.429	0.582	0.996	0.812	996	0.849
	500	0.410	2.021	0.994	2.570	993	2.951
	1000	0.397	4.124	0.992	4.889	991	6.039

Table 10. Some pointwise biases and variances of the three lognormal kernel density estimators following Model 1

Size	Target	$f^{*}_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
n	x	*Bias*	*Variance*	*Bias*	*Variance*	*Bias*	*Variance*
100	0.25	0.189 e+1	0.450 e-2	−0.207 e+1	0.248 e-2	−0.383 e+1	0.912 e-2
	0.50	−0.111 e+1	0.683 e-2	−0.986 e+0	0.192 e-2	−0.330 e+1	0.820 e-2
	2.50	−0.331 e+1	0.215 e-2	−0.329 e+1	0.533 e-2	−0.390 e+1	0.121 e-2
	3.50	−0.381 e+1	0.265 e-2	−0.410 e+1	0.127 e-2	−0.396 e+1	0.368 e-2
500	0.25	0.126 e+1	0.781 e-2	−0.325 e+1	0.534 e-2	−0.168 e+1	0.091 e-2
	0.50	−0.138 e+1	0.752 e-2	−0.519 e+1	0.186 e-2	−0.122 e+1	0.431 e-2
	2.50	−0.374 e+1	0.504 e-2	−0.370 e+1	0.101 e-2	−0.481 e+1	0.136 e-2
	3.50	−0.347 e+1	0.674 e-2	−0.529 e+1	0.307 e-2	−0.501 e+1	0.477 e-2
1000	0.25	0.123 e+1	0.912 e-2	−0.185 e+1	0.100 e-2	−0.124 e+1	0.081 e-2
	0.50	−0.159 e+1	0.811 e-2	−0.208 e+1	0.487 e-2	−0.113 e+1	0.119 e-3
	2.50	−0.376 e+1	0.785 e-2	−0.372 e+1	0.537 e-2	−0.584 e+1	0.652 e-2
	3.50	−0.372 e+1	0.354 e-2	−0.641 e+1	0.154 e-2	−0.549 e+1	0.845 e-2

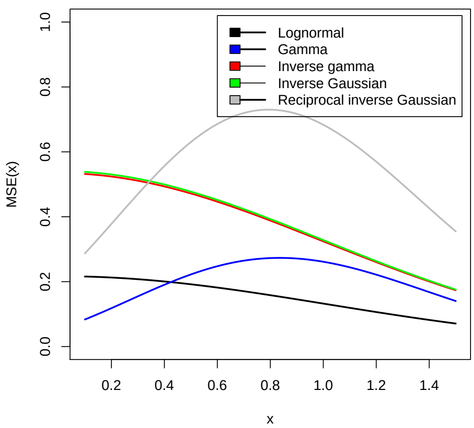


Fig. 2. MSE values for x between 0 and 1.4 for five associated kernel density estimators

Table 11. Some pointwise biases and variances of the three lognormal kernel density estimators following Model 2

Size	Target	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
n	x	*Bias*	*Variance*	*Bias*	*Variance*	*Bias*	*Variance*
100	0.25	0.265 e-1	0.137 e-3	0.207 e-1	0.248 e-3	0.191 e-1	0.721 e-3
	0.50	0.273 e-1	0.176 e-3	0.212 e-1	0.292 e-3	0.217 e-1	0.714 e-3
	2.50	0.230 e-1	0.148 e-3	0.195 e-1	0.283 e-3	0.183 e-1	0.698 e-3
	3.50	0.237 e-1	0.132 e-3	0.173 e-1	0.257 e-3	0.162 e-1	0.672 e-3
500	0.25	0.125 e-1	0.342 e-3	0.121 e-1	0.434 e-3	0.258 e-1	0.575 e-3
	0.50	0.141 e-1	0.364 e-3	0.109 e-1	0.486 e-3	0.262 e-1	0.562 e-3
	2.50	0.143 e-1	0.219 e-3	0.108 e-1	0.367 e-3	0.103 e-1	0.740 e-3
	3.50	0.113 e-1	0.205 e-3	0.095 e-1	0.326 e-3	0.091 e-1	0.725 e-3
1000	0.25	0.060 e-1	0.428 e-3	0.058 e-1	0.600 e-3	0.321 e-1	0.379 e-3
	0.50	0.054 e-1	0.454 e-3	0.046 e-1	0.687 e-3	0.347 e-1	0.360 e-3
	2.50	0.038 e-1	0.325 e-3	0.034 e-1	0.437 e-3	0.031 e-1	0.795 e-3
	3.50	0.035 e-1	0.318 e-3	0.021 e-1	0.414 e-3	0.019 e-1	0.781 e-3

Table 12. The average of the ISEs for three lognormal kernel density estimators

Model	Size	$\widehat{ISE}$ at the edges			$\widehat{ISE}$ in interior		
	n	$f^*_{n,LN}$	$\widehat{f}_{n,LN}$	$\widetilde{f}_{n,LN}$	$\widehat{f}_{n,LN}$	$f^*_{n,LN}$	$\widetilde{f}_{n,LN}$
1	50	0.146 e-1	0.141 e-2	0.359 e-2	0.184 e-2	0.152 e-2	0.481 e-3
	100	0.120 e-1	0.134 e-2	0.475 e-2	0.179 e-2	0.146 e-2	0.444 e-3
	500	0.116 e-1	0.132 e-2	0.482 e-2	0.168 e-2	0.135 e-2	0.438 e-3
	1000	0.828 e-2	0.121 e-2	0.485 e-2	0.165 e-2	0.132 e-2	0.416 e-3
2	50	0.124 e-2	0.340 e-3	0.277 e-3	0.312 e-3	0.396 e-3	0.808 e-4
	100	0.917 e-3	0.305 e-3	0.288 e-3	0.302 e-3	0.311 e-3	0.786 e-4
	500	0.891 e-3	0.287 e-3	0.308 e-3	0.298 e-3	0.287 e-3	0.751 e-4
	1000	0.837 e-3	0.267 e-3	0.310 e-3	0.272 e-3	0.252 e-3	0.322 e-4