Statistics and Probability Letters 82 (2012) 1337–1345

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

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

Varying kernel density estimation on $\mathbb{R}_+$

Robert Mnatsakanov *, Khachatur Sarkisian

West Virginia University, USA
National Institute for Occupational Safety and Health, USA

ARTICLE INFO

Article history:
Received 7 September 2011
Received in revised form 24 March 2012
Accepted 25 March 2012
Available online 3 April 2012

AMS 2000 subject classification:
primary 62G05
secondary 62G20

Keywords:
Varying kernel density estimator
Mean Squared Error
Mean Integrated Squared Error
δ-sequence
L_1-consistency

ABSTRACT

In this article a new nonparametric density estimator based on the sequence of asymmetric kernels is proposed. This method is natural when estimating an unknown density function of a positive random variable. The rates of Mean Squared Error, Mean Integrated Squared Error, and the L_1-consistency are investigated. Simulation studies are conducted to compare a new estimator and its modified version with traditional kernel density construction.

1. Introduction

Let us assume that the support of unknown cumulative distribution function (cdf) F is the positive half-line $\mathbb{R}_+ = (0, \infty)$. To avoid an edge effect when estimating the density function of F it is common to use kernels with the same support as that of the target distribution. Recently, the constructions with asymmetric kernels have been studied for estimating a probability density function (pdf) defined on $\mathbb{R}_+$. Namely, in Chen (2000) and Scaillet (2004) the sequences of gamma kernels, and inverse and reciprocal inverse gaussian kernels have been used, respectively. See also Mnatsakanov and Ruymgaart (2012), where another varying kernel approach is suggested. Their method is based on the sequence of gamma pdfs with varying shapes.

We propose to use a sequence of inverse gamma kernels that represent the δ-sequences in L_2- and L_1-norms, see Lemmas 4.1 and 4.2, respectively. The constructions f_α^* and $\widehat{f_\alpha}$ considered in (2.3) and (2.4) (called the varying kernel density estimators (vKDEs)) are different from the traditional kernel density estimator (KDE) (see, for example, Parzen (1962), Silverman (1986), and Scott (1992)). They are also different from the ones proposed in Chen (2000) and Scaillet (2004). In the kernel density estimation the convolution is considered with respect to addition as the group operation on the entire real line $\mathbb{R}$ and with a fixed kernel. Our constructions in (2.3) and (2.4) turns out to be of kernel type provided that convolution is considered on the space of a positive half-line $(\mathbb{R}_+, dH)$ equipped with multiplication as a group operation, and with the Haar measure $dH(t) = dt/t$ (see, for example, (2.7) below). It is worth mentioning that the estimators proposed by Chen (2000) and by Scaillet (2004) cannot be viewed as convolutions as well as the densities on $\mathbb{R}_+$.

In this paper we investigated the Mean Squared Error (**MSE**) and Mean Integrated Squared Error (**MISE**) rates of convergence for proposed estimators f_α^* and $\widehat{f_\alpha}$. Note that the shape of an inverse gamma density varies according to the

* Correspondence to: Department of Statistics, P.O. Box 6330, West Virginia University, Morgantown, WV 26506, USA. Fax: +1 304 293 2272.
E-mail address: rmnatsak@stat.wvu.edu (R. Mnatsakanov).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2012.03.033

position of a point x at which the pdf $f(x)$ is estimated. This allows automatic changing the "smoothing" degree around the point x. Another feature of the constructions (2.3) and (2.4) are that they have no boundary effects (see Figs. 1 and 2) and they achieve the optimal rate of convergence for **MSE** and for **MISE** within the class of non-negative kernel density estimators. Similar results have been derived in papers: Chen (2000) and Scaillet (2004). There are differences regarding the constants appearing in the first order terms only. It is worth mentioning that in contrast with KDE, the asymptotic variances of $f_\alpha^*(x)$ and $\widehat{f}_\alpha(x)$ have the same form $n^{-4/5}f(x)/(2x\sqrt{\pi})$, as $\alpha = n^{2/5}$ (see (3.6) in Section 3), that becomes smaller as x increases. Finally, note that in the case of asymmetric gamma kernels (see Chen (2000)), the corresponding variance has the form $n^{-4/5}f(x)/(2\sqrt{x\pi})$. In Mnatsakanov and Ruymgaart (2012), the construction similar to (2.1) has been used, and, as a result, another, the so-called moment-density estimate has been proposed, and its asymptotic properties were studied as well.

The paper is organized as follows. In Section 2 the assumptions and the construction of the vKDE are introduced. In Section 3 the **MSE** of f_α^* and $\widehat{f}_\alpha$ are derived, while in Section 4 the **MISE** and L_1-consistency of $\widehat{f}_\alpha$ are investigated. In Section 5 we conducted the simulation study and compared the performances of the estimators $\widehat{f}_\alpha, f_\alpha^*$ and the traditional KDE $\widehat{f}_h$.

2. Preliminaries and assumptions

In this section we outline the main idea that yields vKDEs f_α^* and $\widehat{f}_\alpha$ in (2.3) and (2.4), respectively. Assume we would like to recover (approximate) the moment-identifiable distribution F given only the sequence of its moments. About the conditions necessary and sufficient for F to be the moment-identifiable distribution, see, for example, Stoyanov (2000) and references therein. Suppose that all negative order moments of F are finite. Define the operator $\mathcal{M}$ by

$$(\mathcal{M}F)(j) = \int_0^\infty t^{-j} dF(t) = \mu_j, \quad j = 0, 1, \ldots$$

and introduce the sequence of operators $\mathcal{M}_\alpha^{-1}$:

$$(\mathcal{M}_\alpha^{-1}\mu)(x) = 1 - \sum_{k=0}^{\alpha} \frac{(\alpha x)^k}{k!} \sum_{j=k}^{\infty} \frac{(-\alpha x)^{j-k}}{(j-k)!}\mu_j, \quad x \in \mathbb{R}_+. \tag{2.1}$$

Here $\mu = \{\mu_j, j = 0, 1, \ldots\}$ and $\alpha \to \infty$ at a rate to be specified later.

In analysis, the transform $(\mathcal{M}F)(1-z)$, where z is a complex variable, is known as the Mellin transform. There is extensive literature investigating the problem of recovering a function from its Mellin transform. See, for instance, Tagliani (2001), Klauder et al. (2001) and Sneddon (1974), among others. In Gzyl and Tagliani (2010), and Mnatsakanov (2008a,b) the problem of recovering the cdf and corresponding density function given the moment sequence of positive orders of underlying distribution has been studied. The investigation of the properties of approximation $\mathcal{M}_\alpha^{-1}$ in (2.1) is beyond the scope of this article and will be conducted in a separate investigation.

To construct the density estimate, at first, let us approximate F by means of $\mathcal{M}_\alpha^{-1}$. A minor modification of an argument in Mnatsakanov and Ruymgaart (2003) yields

$$F_\alpha = \mathcal{M}_\alpha^{-1}\mathcal{M}F \to_w F, \quad \text{as } \alpha \to \infty. \tag{2.2}$$

Here by $\to_w$ we denote the weak convergence of corresponding cdfs.

Now, suppose we are given a sequence $X_1, \ldots, X_n$ of independent and identically distributed positive random variables from the absolutely continuous distribution function F (with pdf $f = F'$). To estimate F, let us first estimate its negative j-th order moments $\mu_j, j \geq 1$. Namely, based on (2.2), let us construct the estimate F_α^* of F by replacing the moment μ_j in (2.1) by its empirical counterpart

$$\hat{\mu}_j = \int_0^\infty t^{-j} d\hat{F}_n(t), \quad j = 0, 1, \ldots, \text{ with } \hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leq t\}.$$

Here $\hat{F}_n$ is the empirical cdf of the sample $X_1, \ldots, X_n$. After a simple algebra, we derive

$$F_\alpha^*(x) = 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{\alpha}\frac{1}{k!}\left(\frac{\alpha}{X_i}x\right)^k \exp\left(-\frac{\alpha}{X_i}x\right), \quad x \in \mathbb{R}_+.$$

To compare F_α^* with the empirical cdf $\hat{F}_n$, note that $F_\alpha^*(x) \sim \hat{F}_n(x)$ as long as α is large. This follows from the fact that for a given X_i and large α:

$$\sum_{k=0}^{\alpha}\frac{1}{k!}\left(\frac{\alpha}{X_i}x\right)^k \exp\left(-\frac{\alpha}{X_i}x\right) \sim I\{X_i > x\}.$$

Note also that $F_\alpha^*(x)$ is a continuous function of x, hence, to estimate the density $f(x)$ one can take the derivative of $F_\alpha^*(x)$:

$$f_\alpha^*(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}\cdot\frac{1}{\Gamma(\alpha)}\left(\frac{\alpha}{X_i}x\right)^\alpha \exp\left(-\frac{\alpha}{X_i}x\right), \tag{2.3}$$

and choose $\alpha = \alpha(n) \to \infty$ as $n \to \infty$. The problem of optimal choice of parameter α will be specified later. Of course $f_\alpha^*(x) \geq 0$ for each $x > 0$, and since it is easily seen that $\int_0^\infty f_\alpha^*(x)dx = 1$, the estimator is itself a probability density. The statements similar to the ones obtained in Sections 3 and 4 are valid for f_α^* as well (see for example, Theorem 3.2). To simplify the calculations below and to reduce the bias of f_α^*, let us use the modified version of f_α^*. Namely, let us increase the shape parameter of the inverse gamma kernel presented in the right hand side of (2.3) by one. Denote $S_{i,x} := \frac{1}{X_i} L_\alpha(\frac{x}{X_i})$, where $L_\alpha(u) = (\alpha u)^{\alpha+1}/\Gamma(\alpha+1)\,\exp(-\alpha u)$, $u \in \mathbb{R}_+$, and consider

$$\widehat{f}_\alpha(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{X_i} L_\alpha\left(\frac{x}{X_i}\right) = \frac{1}{n}\sum_{i=1}^n S_{i,x}. \tag{2.4}$$

Throughout the proposed estimator will be considered at a fixed point $x > 0$, where $f(x) > 0$. Also, we will assume that $F(0) = 0$, and the underlying density satisfies

$$f \in C^{(2)}(\mathbb{R}_+), \quad \text{with } \sup_{t>0} |f''(t)| = M < \infty. \tag{2.5}$$

Besides, let us denote by $g(\cdot, a_k, b_k)$ the inverse gamma density with the shape $a_k = k(\alpha+2) - 1$ and the rate $b_k = k\,\alpha x$ parameters, respectively. Namely

$$g(t; a_k, b_k) = \frac{1}{t^2} \cdot \frac{b_k^{a_k} \left(\frac{1}{t}\right)^{a_k - 1} e^{-\frac{b_k}{t}}}{\Gamma(a_k)}, \quad t > 0. \tag{2.6}$$

The mean ξ_k and variance σ_k^2 of $g(\cdot; a_k, b_k)$ have the following expressions, respectively:

$$\xi_k = \frac{b_k}{a_k - 1} = \frac{k\alpha x}{k(\alpha+2) - 2},$$
$$\sigma_k^2 = \frac{b_k^2}{(a_k-1)^2 \cdot (a_k - 2)} = \frac{k^2\alpha^2 x^2}{\{k(\alpha+2)-2\}^2 \cdot \{k(\alpha+2)-3\}}.$$

Note also that the mean of $\widehat{f}_\alpha(x)$ can be written as the convolution operator on $(\mathbb{R}_+, dH)$:

$$f_\alpha(x) = \mathbf{E}\widehat{f}_\alpha(x) = \int_0^\infty L_\alpha(x/t) f(t)\,dH(t), \quad x \in \mathbb{R}_+, \tag{2.7}$$

where $dH(t) = dt/t$. In Lemmas 4.1 and 4.2, see Section 4, it is proved that the sequence of functions $\{(1/t)\,L_\alpha(\cdot/t),\ t \in \mathbb{R}_+, \alpha \in \mathbb{N}\}$ with $L_\alpha(\cdot)$ defined in (2.4) forms the δ-sequences in L_1- and L_2-norms, as $\alpha \to \infty$.

3. Bias and MSE

Without explicit reference it will be assumed that all the conditions in Section 2 are satisfied. Let us study the bias and the second moment of the estimator $\widehat{f}_\alpha$. We have

$$\begin{aligned}
\mathbf{E}S_{i,x}^k &= \int_0^\infty \frac{1}{\{\Gamma(\alpha+1)\}^k} \left(\frac{1}{t}\right)^k \left(\frac{\alpha x}{t}\right)^{k(\alpha+1)} \exp\left(-\frac{k\alpha x}{t}\right) f(t)\,dt \\
&= \int_0^\infty \frac{\{k(\alpha+2)-2\}!(\alpha x)^{k(\alpha+1)}}{\{\Gamma(\alpha+1)\}^k\,(k\alpha x)^{k(\alpha+2)-1}} g(t; a_k, b_k) f(t)\,dt \\
&= \left(\frac{1}{\alpha x}\right)^{k-1} \frac{\{k(\alpha+2)-2\}!}{\{\Gamma(\alpha+1)\}^k} \frac{1}{k^{k(\alpha+2)-1}} \int_0^\infty g(t; a_k, b_k) f(t)\,dt.
\end{aligned} \tag{3.1}$$

In particular, for $k = 1$:

$$\mathbf{E}\widehat{f}_\alpha(x) = \mathbf{E}S_{i,x} = \int_0^\infty g(t; a_1, b_1) f(t)\,dt = f_\alpha(x). \tag{3.2}$$

This yields for the bias of $\widehat{f}_\alpha(x)$:

$$\begin{aligned}
f_\alpha(x) - f(x) &= \mathbf{Bias}\{\widehat{f}_\alpha(x)\} = \int_0^\infty g(t; a_1, b_1)\{f(t) - f(x)\}\,dt \\
&= \int_0^\infty g(t; a_1, b_1)\{f(x) + (t-x)f'(x) + \frac{1}{2}(t-x)^2 f''(\tilde{t}) - f(x)\}\,dt \\
&= \frac{1}{2}\int_0^\infty (t-x)^2 g(t; a_1, b_1) f''(x)\,dt + \frac{1}{2}\int_0^\infty (t-x)^2 g(t; a_1, b_1)\{f''(\tilde{t}) - f''(x)\}\,dt \\
&= \frac{1}{2} \cdot \frac{x^2}{\alpha - 1} \cdot f''(x) + o\left(\frac{1}{\alpha}\right), \quad \text{as } n \to \infty.
\end{aligned} \tag{3.3}$$

For the variance we have

$$\mathbf{Var}\{\widehat{f}_\alpha(x)\} = \frac{1}{n}\mathbf{Var}\, S_{i,x} = \frac{1}{n}\{\mathbf{E}\, S_{i,x}^2 - f_\alpha^2(x)\}. \tag{3.4}$$

Applying (3.1) for $k = 2$ and $B_\alpha = \alpha^{-1}\, 2^{-(2\alpha+3)}\Gamma(2\alpha+3)[\Gamma(\alpha+1)]^{-2} \sim \alpha^{1/2}/(2\sqrt{\pi})$, as $\alpha \to \infty$, yields

$$\begin{aligned}
\mathbf{E}\, S_{i,x}^2 &= \frac{1}{\alpha x}\cdot\frac{\Gamma(2\alpha+3)}{\{\Gamma(\alpha+1)\}^2}\cdot\frac{1}{2^{2(\alpha+2)-1}}\cdot\int_0^\infty g(t;a_2,b_2)\,f(t)dt \\
&= \frac{B_\alpha}{x}\int_0^\infty g(t;a_2,b_2)f(t)dt \sim \frac{1}{\alpha x\sqrt{2\pi}}\cdot\frac{e^{-2(\alpha+1)}\{2(\alpha+1)\}^{2(\alpha+1)+1/2}}{e^{-2\alpha}\cdot\alpha^{2\alpha+1}}\cdot\frac{1}{2^{2\alpha+3}} \\
&\quad\times\int_0^\infty g(t;a_2,b_2)f(t)dt \sim \frac{1}{\alpha x\sqrt{2\pi}}\frac{\alpha^{3/2}}{\sqrt{2}}\int_0^\infty g(t;a_2,b_2)f(t)dt \\
&= \frac{\sqrt{\alpha}}{2x\sqrt{\pi}}\{f(x)+o(1)\} = \frac{\sqrt{\alpha}}{2\sqrt{\pi}}\frac{f(x)}{x} + o(\sqrt{\alpha}).
\end{aligned} \tag{3.5}$$

Inserting (3.3) and (3.5) in (3.4) we obtain

$$\begin{aligned}
\mathbf{Var}\{\widehat{f}_\alpha(x)\} &= \frac{1}{n}\left[\frac{1}{2\sqrt{\pi}}\frac{\sqrt{\alpha}}{x}f(x) + o(\sqrt{\alpha}) - \left\{f(x) + O\left(\frac{1}{\alpha}\right)\right\}^2\right] \\
&= \frac{\sqrt{\alpha}}{2n\sqrt{\pi}}\frac{f(x)}{x} + o\left(\frac{\sqrt{\alpha}}{n}\right).
\end{aligned} \tag{3.6}$$

Finally, combining (3.3) and (3.6) leads to the **MSE** of $\widehat{f}_\alpha(x)$:

$$\mathbf{MSE}\{\widehat{f}_\alpha(x)\} = \mathbf{Var}\{\widehat{f}_\alpha(x)\} + \mathbf{Bias}^2\{\widehat{f}_\alpha(x)\} = \frac{\sqrt{\alpha}}{2n\sqrt{\pi}}\frac{f(x)}{x} + \frac{1}{4}\frac{x^4}{(\alpha-1)^2}\{f''(x)\}^2 + o\left(\frac{\sqrt{\alpha}}{n}\right) + o\left(\frac{1}{\alpha^2}\right). \tag{3.7}$$

For optimal rates we may take

$$\alpha = \alpha(n) = n^{2/5}. \tag{3.8}$$

By substitution (3.8) into (3.7) we find

$$\mathbf{MSE}\{\widehat{f}_\alpha(x)\} = n^{-4/5}\left[\frac{f(x)}{2x\sqrt{\pi}} + \frac{x^4\{f''(x)\}^2}{4}\right] + o(n^{-4/5}). \tag{3.9}$$

Here we have assumed that the pdf f has a continuous and bounded second derivative f'' (condition (2.5)). The following statement is valid.

Theorem 3.1. *Under the assumption (2.5) the bias of $\widehat{f}_\alpha(x)$ satisfies*

$$\mathbf{Bias}\{\widehat{f}_\alpha(x)\} = \frac{x^2 f''(x)}{2\cdot(\alpha-1)} + o\left(\frac{1}{\alpha}\right), \quad \text{as } \alpha \text{ and } n \to \infty.$$

*For the **MSE** of $\widehat{f}_\alpha(x)$ we have the expression in (3.9), provided that we choose $\alpha = \alpha(n) \sim n^{2/5}$.*

One can check very easily that the variance of vKDE f_α^* defined in (2.3) has the same form we have in the right-hand side of (3.6), while the bias of f_α^* has additional term containing f'. Applying the similar argument used in derivations of (3.3), (3.5) and (3.6), we obtain the following statement.

Theorem 3.2. *Under the assumption (2.5), the bias and **MSE** of $f_\alpha^*(x)$ have the following expressions*

$$\mathbf{Bias}\{f_\alpha^*(x)\} = \frac{xf'(x)}{\alpha-1} + \frac{x^2 f''(x)}{2}\times\frac{\alpha^2}{(\alpha-1)^2(\alpha-2)} + o\left(\frac{1}{\alpha}\right),$$

$$\mathbf{MSE}\{f_\alpha^*(x)\} = \frac{\sqrt{\alpha}}{2n\sqrt{\pi}}\frac{f(x)}{x} + \frac{x^2\{f'(x)\}^2}{(\alpha-1)^2} + \frac{x^4\{f''(x)\}^2\alpha^4}{4(\alpha-1)^4(\alpha-2)^2} + o\left(\frac{\sqrt{\alpha}}{n}\right) + o\left(\frac{1}{\alpha^2}\right),$$

*as α and $n \to \infty$. For the optimal **MSE** of $f_\alpha^*(x)$ we have*

$$\mathbf{MSE}\{f_\alpha^*(x)\} = n^{-4/5}\left[\frac{f(x)}{2x\sqrt{\pi}} + x^2\{f'(x)\}^2 + \frac{x^4\{f''(x)\}^2}{4}\right] + o(n^{-4/5}),$$

provided that we choose $\alpha = \alpha(n) \sim n^{2/5}$.

4. MISE and L_1-consistency of $\widehat{f}_\alpha$

4.1. MISE rate of convergence

Throughout this section again F concentrates mass 1 on $(0, \infty)$ but it is also supposed to have a sufficiently smooth density. Let us consider the following conditions:

$$\int_0^\infty \frac{f(x)}{x}dx = C_0 < \infty \quad \text{and,} \tag{4.1}$$

$$\int_0^\infty \{x^2 f''(x)\}^2 dx = C_1 < \infty. \tag{4.2}$$

One can very easily obtain the optimal rate $n^{-4/5}$ for $\mathbf{MISE}\{\widehat{f}_\alpha\}$ as $\alpha,\ n \to \infty$ by integrating the terms on the right-hand side of (3.7). Namely, the following statement is true.

Theorem 4.1. *Under the assumptions (2.5), (4.1) and (4.2) we have*

$$\mathbf{MISE}\{\widehat{f}_\alpha\} = \int_0^\infty \mathbf{Var}\{\widehat{f}_\alpha(x)\}dx + \int_0^\infty \mathbf{Bias}^2\{\widehat{f}_\alpha(x)\}dx \sim \frac{C_0\sqrt{\alpha}}{2n\sqrt{\pi}} + \frac{C_1}{4\alpha^2},$$

*as $\alpha, n \to \infty$. While for optimal **MISE** we have*

$$\mathbf{MISE}\{\widehat{f}_\alpha\} \sim n^{-4/5}\left(\frac{5}{4}\right)\cdot\left(\frac{C_0}{2\sqrt{\pi}}\right)^{4/5} C_1^{1/5}, \quad \text{as } \alpha,\ n \to \infty,$$

provided that we choose $\alpha = \alpha(n) = n^{2/5}(2\,C_1\sqrt{\pi}/C_0)^{2/5}$.

One can weaken the conditions on f and show that the corresponding rate is $n^{-2/3}$ under the requirement of integrability of $\{xf'(x)\}^2$. Indeed, let us denote again by $B_\alpha = \alpha^{-1}\,2^{-(2\alpha+3)}\,\Gamma(2\alpha+3)\,[\Gamma(\alpha+1)]^{-2}$ and consider the following condition (instead of (4.2)):

$$\int_0^\infty \{xf'(x)\}^2 dx = C_2 < \infty. \tag{4.3}$$

Consider the L_1- and L_2-norms of a function $\phi : \mathbb{R}_+ \to \mathbb{R}$ by

$$\|\phi\|_{L_1} = \int_0^\infty |\phi(x)|dx, \qquad \|\phi\|_{L_2} = \left\{\int_0^\infty |\phi(x)|^2 dx\right\}^{1/2},$$

respectively.

Lemma 4.1. *If f' is bounded and condition (4.3) is satisfied, then*

$$\|f_\alpha - f\|_{L_2} \le \frac{1}{\alpha}\sqrt{C_2(\alpha+1)}.$$

Proof of the Lemma 4.1. Let us denote by η_α the r.v. with pdf $L_\alpha(t)/t$, $t \in \mathbb{R}_+$. Note also that the r.v. x/η_α has pdf $L_\alpha(x/t)/t$ and

$$\int_0^\infty L_\alpha(x/s)\frac{1}{s}ds = 1, \qquad \mathbf{E}[(1/\eta_\alpha)] = 1, \qquad \mathbf{Var}[(1/\eta_\alpha)] = \frac{1}{\alpha-1}, \tag{4.4}$$

$$f(x/\eta_\alpha) - f(x) = \int_x^{x/\eta_\alpha} f'(y)dy.$$

Then after simple algebra combined with application of the Cauchy–Schwarz's inequality we obtain

$$\begin{aligned}
\|f_\alpha - f\|_{L_2}^2 &= \int_0^\infty \mathbf{Bias}^2\{\widehat{f}_\alpha(x)\}dx = \int_0^\infty \left[\int_0^\infty \{f(s) - f(x)\}L_\alpha(x/s)\frac{1}{s}ds\right]^2 dx \\
&= \int_0^\infty [\mathbf{E}(f(x/\eta_\alpha) - f(x))]^2 dx = \int_0^\infty \left[\mathbf{E}\int_x^{x/\eta_\alpha} f'(s)ds\right]^2 dx \\
&\le \mathbf{E}\int_0^\infty \left[\int_x^{x/\eta_\alpha} (f'(s))^2 ds\, x(\eta_\alpha^{-1} - 1)\right] dx
\end{aligned}$$

$$= \mathbf{E}\left\{I_{[\eta_\alpha\le 1]}\int_0^\infty \left[(f'(s))^2\int_{s\eta_\alpha}^{s} x(\eta_\alpha^{-1}-1)dx\right]ds + I_{[\eta_\alpha\ge 1]}\int_0^\infty \left[(f'(s))^2\int_{s}^{s\eta_\alpha} x(1-\eta_\alpha^{-1})dx\right]ds\right\}$$

$$= \mathbf{E}\left\{I_{[\eta_\alpha\le 1]}\int_0^\infty (f'(s))^2\frac{1}{2}s^2(1-\eta_\alpha^2)\,(\eta_\alpha^{-1}-1)ds + I_{[\eta_\alpha\ge 1]}\int_0^\infty (f'(s))^2\frac{1}{2}\,s^2(\eta_\alpha^2-1),\,(1-\eta_\alpha^{-1})ds\right\}$$

$$= \frac{1}{2}\mathbf{E}\left(\frac{(\eta_\alpha-1)^2(\eta_\alpha+1)}{\eta_\alpha}\right)\int_0^\infty \{sf'(s)\}^2 ds. \tag{4.5}$$

But

$$\mathbf{E}\left(\frac{(\eta_\alpha-1)^2(\eta_\alpha+1)}{\eta_\alpha}\right) = \int_0^\infty \frac{(u-1)^2(u+1)}{u}L_\alpha(u)\frac{1}{u}du$$

$$= \int_0^\infty (u-1)^2\,(u+1)\cdot\frac{\alpha^{\alpha+1}u^{\alpha-1}}{\Gamma(\alpha+1)}e^{-\alpha u}du = \frac{2\,(\alpha+1)}{\alpha^2}. \tag{4.6}$$

Combination of (4.5) and (4.6) gives

$$\|f_\alpha - f\|_{L_2}^2 = \int_0^\infty \mathbf{Bias}^2\{\widehat{f}_\alpha(x)\}dx \le \frac{\alpha+1}{\alpha^2}\int_0^\infty \{sf'(s)\}^2 ds. \tag{4.7}$$

Lemma 4.1 is proved. □

Theorem 4.2. *If f' is bounded and the conditions (4.1) and (4.3) are satisfied, then*

$$\mathbf{MISE}\{\widehat{f}_\alpha\} \le \frac{B_\alpha C_0}{n} + \frac{C_2}{\alpha} + \frac{C_2}{\alpha^2},\quad \alpha > 1,$$

$$\mathbf{MISE}\{\widehat{f}_\alpha\} \le \frac{C_0\sqrt{\alpha}}{2n\sqrt{\pi}} + \frac{C_2}{\alpha} + o\left(\frac{1}{\alpha}\right),$$

*as $\alpha, n \to \infty$. While for optimal **MISE** we have*

$$\mathbf{MISE}\{\widehat{f}_\alpha\} \le n^{-2/3}\frac{3}{2^{2/3}}\cdot\left(\frac{C_0}{2\sqrt{\pi}}\right)^{2/3} C_2^{1/3} + o(n^{-2/3}),\quad \text{as } n\to\infty,$$

provided that we choose $\alpha = \alpha(n) = n^{2/3}(4\,C_2\sqrt{\pi}/C_0)^{2/3}$.

Proof. Let us study the variance term. According to the definitions of the inverse gamma $g(\cdot, a_k, b_k)$ in (2.6) and gamma $h(\cdot, shape, rate)$ densities, we have

$$\int_0^\infty \frac{1}{x}g(t; a_2, b_2)dx = \frac{1}{t}\int_0^\infty h(x, 2\alpha+3, 2\alpha/t)dx = \frac{1}{t}.$$

So that integration of the both sides of the first equation in (3.5) combined with $B_\alpha \sim \alpha^{1/2}/(2\sqrt{\pi})$, as $\alpha \to \infty$, yields

$$\frac{1}{n}\int_0^\infty \mathbf{E}S_{i,x}^2 dx = \frac{B_\alpha}{n}\int_0^\infty \frac{1}{x}\left[\int_0^\infty g(t; a_2, b_2)f(t)dt\right]dx$$

$$= \frac{B_\alpha}{n}\int_0^\infty f(t)\left[\int_0^\infty \frac{1}{x}g(t; a_2, b_2)dx\right]dt$$

$$= \frac{B_\alpha}{n}\int_0^\infty \frac{f(t)}{t}dt \sim \frac{\sqrt{\alpha}}{2n\sqrt{\pi}}\int_0^\infty \frac{f(t)}{t}dt,\quad \text{as } \alpha\to\infty. \tag{4.8}$$

Hence, it is proved

$$\int_0^\infty \mathbf{Var}\{\widehat{f}_\alpha(x)\}dx \le \int_0^\infty \frac{1}{n}\{\mathbf{E}\,S_{i,x}^2\}dx \sim \frac{\sqrt{\alpha}}{2n\sqrt{\pi}}\int_0^\infty \frac{f(t)}{t}dt, \tag{4.9}$$

as $n, \alpha \to \infty$. Finally, from (4.7)–(4.9) we obtain the statements of Theorem 4.2. □

4.2. L_1-consistency

In this subsection let us consider the condition

$$\int_0^\infty x^2|f''(x)|dx = C_3 < \infty. \tag{4.10}$$

Consider the L_1-distance $\|f_\alpha - f\|_{L_1}$ between f_α and f (with respect to the Lebesgue measure λ on $\mathbb{R}_+$). Here $f_\alpha(x) = \mathbf{E}\widehat{f_\alpha}(x) = \mathbf{E}f(x/\eta_\alpha)$ with η_α defined in the proof of Lemma 4.1. One can show that the functions $\{(1/t)\,L_\alpha(\cdot/t),\ t > 0\}$ form a δ-sequence in L_1-norm as well, as $\alpha \to \infty$. Namely, the following statement is true.

Lemma 4.2. *If f'' is bounded and the condition (4.10) is satisfied, then*

$$\|f_\alpha - f\|_{L_1} \le C_3\left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\right).$$

Proof. Combination of (4.4), (4.10) and the following equations

$$\int_0^\infty L_\alpha(x/s)\frac{1}{s}ds = 1, \qquad \mathbf{E}[x(\eta_\alpha^{-1} - 1)] = 0,$$

$$f(x/\eta_\alpha) - f(x) = f'(x)(x/\eta_\alpha - x) + \int_x^{x/\eta_\alpha} ds \int_x^s f''(y)dy,$$

gives

$$\begin{aligned}\|f_\alpha - f\|_{L_1} &= \int_0^\infty \left|\int_0^\infty \{f(s) - f(x)\}L_\alpha(x/s)\frac{1}{s}ds\right| dx \\ &= \int_0^\infty |\mathbf{E}(f(x/\eta_\alpha) - f(x))|dx = \int_0^\infty \left|\mathbf{E}\int_x^{x/\eta_\alpha} ds \int_x^s f''(y)dy\right| dx \\ &\le \int_0^\infty \left[\mathbf{E}I_{[\eta_\alpha<1]}(x/\eta_\alpha - x)\int_x^{x/\eta_\alpha} |f''(y)|dy + \mathbf{E}I_{[\eta_\alpha>1]}(x - x/\eta_\alpha)\int_{x/\eta_\alpha}^x |f''(y)|dy\right] dx. \end{aligned} \tag{4.11}$$

Now in a similar way as we did in (4.5) and (4.6), changing the integrations in (4.11) yields

$$\|f_\alpha - f\|_{L_1} \le \frac{1}{2}\int_0^\infty y^2|f''(y)|dy\mathbf{E}\left(\frac{(\eta_\alpha - 1)^2(\eta_\alpha + 1)}{\eta_\alpha}\right) = \left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\right)\int_0^\infty y^2|f''(y)|dy.$$

Lemma 4.2 is proved. □

Theorem 4.3. *If f'' is bounded and the conditions (4.1) and (4.10) are satisfied, then*

$$\mathbf{E}\|\widehat{f_\alpha} - f\|_{L_1} = \mathbf{E}\int_0^\infty |\widehat{f_\alpha}(x) - f(x)|dx \to 0, \quad \text{as } \sqrt{\alpha}/n \to 0, \alpha,\ n \to \infty. \tag{4.12}$$

Proof. Under the assumptions (4.10) we have from Lemma 4.2 that $\|f_\alpha - f\|_{L_1} \to 0$, as $\alpha \to \infty$. Hence, to prove (4.12) it is sufficient to show

$$F\{A_n(\delta)\} = F\left\{x : \int_0^\infty \frac{1}{t^2}L_\alpha^2(x/t)f(t)\,dt \ge n\delta\right\} \to 0,$$

for any $\delta > 0$ and $\alpha,\ n \to \infty$ (see, Theorem 1 in Mnatsakanov and Khmaladze (1981)). But F is an absolutely continuous distribution with respect to Lebesgue measure λ, so, let us establish $\lambda\{A_n(\delta)\} \to 0$, for any $\delta > 0$ and $\alpha,\ n \to \infty$. Indeed, application of (4.8) yields

$$\begin{aligned}\lambda\{A_n(\delta)\} &\le \frac{1}{n\delta}\int_{A_n(\delta)} dx \int_0^\infty \frac{1}{t^2}L_\alpha^2(x/t)f(t)dt \le \frac{1}{n\delta}\int_0^\infty \mathbf{E}S_{i,x}^2 dx = \frac{B_\alpha}{n\delta}\int_0^\infty \frac{f(t)}{t}dt \\ &\sim \frac{\sqrt{\alpha}}{2n\delta\sqrt{\pi}}\int_0^\infty \frac{f(t)}{t}dt, \quad \text{as } \alpha \to \infty. \end{aligned} \tag{4.13}$$

The proof of Theorem 4.3 follows from (4.1), (4.13), and $\sqrt{\alpha}/n \to 0$. □

Remark 4.1. Taking $\alpha = h^{-2}$ one can see that the condition $\sqrt{\alpha}/n \to 0$ from Theorem 4.3 corresponds to the condition $nh \to \infty$ in traditional kernel density estimation.

Table 1
The values of $\alpha_{\rm cv}$, $\alpha^*_{\rm cv}$, and $h_{\rm cv}$ and corresponding $\widehat{\mathbf{MISE}}$s of vKDEs and KDE.

	Log-normal (0, 1)						Gamma (2, 1)					
n	$\widehat{f}_\alpha$	$[\alpha_{\rm cv}]$	f^*_α	$[\alpha^*_{\rm cv}]$	$\widehat{f}_h$	$[h_{\rm cv}]$	$\widehat{f}_\alpha$	$[\alpha_{\rm cv}]$	f^*_α	$[\alpha^*_{\rm cv}]$	$\widehat{f}_h$	$[h_{\rm cv}]$
200	0.0092	[11]	0.0066	[7]	0.0166	[0.14]	0.0060	[14]	0.0046	[10]	0.0080	[0.30]
400	0.0057	[14]	0.0043	[9]	0.0103	[0.11]	0.0033	[18]	0.0026	[14]	0.0045	[0.25]
600	0.0039	[17]	0.0030	[11]	0.0075	[0.10]	0.0020	[22]	0.0016	[16]	0.0030	[0.22]
800	0.0029	[19]	0.0022	[12]	0.0059	[0.09]	0.0018	[24]	0.0015	[18]	0.0026	[0.19]

5. Simulations

In this section we study the performances of f^*_α and $\widehat{f}_\alpha$ defined in (2.3) and (2.4), respectively. In particular, we compare them with KDE $\widehat{f}_h$ when the kernel function K is assumed to be a standard normal density function. Let us consider the case when the optimal choice of h, $h = h_{\rm cv}$, is based on the least-squares cross validation (CV) algorithm that minimizes the expression $M_1(h)$ defined by Eq. (3.39) in Silverman (1986).

In our simulation studies we plotted the curves of vKDEs $\widehat{f}_\alpha$ and f^*_α, when the optimal $\alpha = \alpha_{\rm cv}$ and, respectively, $\alpha = \alpha^*_{\rm cv}$, are chosen via the least-squares CV algorithm as well (cf. with Mnatsakanov and Ruymgaart (2012)), and compared them with corresponding curve of KDE $\widehat{f}_h$, when $h = h_{\rm cv}$ (see Figs. 1 and 2). In particular, we simulated the r.v.'s X_i, $i = 1, \ldots, n$, from two different distributions: Log-normal (0, 1) and Gamma (2, 1) with different sample sizes $n = 200k$, $1 \le k \le 4$. In addition, we repeated these simulations $N = 500$ times and studied the performances of $\widehat{f}_\alpha$, f^*_α, and $\widehat{f}_h$ using the **MISE**. Namely, we used the estimated **MISE**:

$$\widehat{\mathbf{MISE}} := \hat{E}(\mathbf{ISE})\{\widehat{f}\} = \frac{1}{N}\sum_{j=1}^{N}\int_0^\infty [\widehat{f}_{(j)}(x) - f(x)|^2 dx.$$

Here the expectation $\hat{E}$ is calculated with respect to the empirical cdf of $N = 500$ values of **ISE**s, while $\widehat{f}_{(j)}$ denotes the vKDEs or KDE used on the j-th replication. The optimal $\alpha = \alpha_{\rm cv}$ minimizes the expression $M_2(\alpha)$, i.e.

$$\alpha_{\rm cv} = \mathrm{argmin}_\alpha M_2(\alpha) = \mathrm{argmin}_\alpha \left[\int_0^\infty [\widehat{f}_\alpha(x)]^2 dx - 2\int_0^\infty \widehat{f}_\alpha(x) d\hat{F}_n(x)\right], \tag{5.1}$$

where $\alpha \in \{1, \ldots, 40\}$ for each $n = 200k$, $1 \le k \le 4$. In the second term of the right hand side of (5.1) let us apply the leave-one-out construction instead of $\widehat{f}_\alpha$. This yields the following expression of

$$M_2(\alpha) = \frac{\Gamma(2\alpha+3)}{n^2\alpha\Gamma^2(\alpha+1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(X_iX_j)^{\alpha+1}}{(X_i+X_j)^{2\alpha+3}} - \frac{2}{n(n-1)\Gamma(\alpha+1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{X_j}\left(\frac{\alpha X_i}{X_j}\right)^{\alpha+1} e^{-\frac{\alpha X_i}{X_j}}.$$

In the case of vKDE f^*_α, we choose the optimal CV parameter $\alpha = \alpha^*_{\rm cv}$ that minimizes the function

$$M_3(\alpha) = \frac{\Gamma(2\alpha+1)}{n^2\alpha\Gamma^2(\alpha)}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(X_iX_j)^{\alpha}}{(X_i+X_j)^{2\alpha+1}} - \frac{2}{n(n-1)\Gamma(\alpha)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{X_j}\left(\frac{\alpha X_i}{X_j}\right)^{\alpha} e^{-\frac{\alpha X_i}{X_j}}.$$

During the simulation study, we found out that $\widehat{\mathbf{MISE}}$s of vKDEs are decreasing functions of n when the parameters $\alpha = \alpha_{\rm cv}$, $\alpha = \alpha^*_{\rm cv}$, and $\alpha = n^{2/5}$. In Table 1, we recorded the values of $\alpha_{\rm cv}$, $\alpha^*_{\rm cv}$, and $h_{\rm cv}$ and corresponding $\widehat{\mathbf{MISE}}$ for Log-normal (0, 1) and Gamma (2, 1) distributions for four different sample sizes. We see that the values of $\widehat{\mathbf{MISE}}$ for f^*_α are smaller than corresponding values of $\widehat{\mathbf{MISE}}$ for $\widehat{f}_\alpha$ and $\widehat{f}_h$. To illustrate the performances of vKDEs graphically, we plotted the graphs of estimators $\widehat{f}_{\alpha_{\rm cv}}$ (the dashed curves) with $\alpha_{\rm cv} = 11$ and 24 and $\widehat{f}_h$ (the dotted curve) with $h = h_{\rm cv}$, in Fig. 1(a) and 2(a) when the sampled distributions are Log-normal (0, 1) (with $n = 200$) and Gamma (2, 1) (with $n = 800$), respectively. For the same samples, in Fig. 1(b) and 2(b) we plotted the graphs of estimators $f^*_{\alpha^*_{\rm cv}}$ (the dashed curve) and $\widehat{f}_h$ when $\alpha^*_{\rm cv} = 7$ and 18 and $h = h_{\rm cv}$, respectively. In each model the sampled pdf f (the solid curve) is plotted as well. Based on the records in Table 1, we conclude that the performances of vKDEs are better compared to the one based on KDE $\widehat{f}_{h_{\rm cv}}$. After conducting many simulations we can say that the asymptotic behavior of $f^*_{\alpha^*_{\rm cv}}$ and its modified version $\widehat{f}_{\alpha_{\rm cv}}$ are similar to each other, and their performances around the origin and on the right tail are much better than that of KDE $\widehat{f}_{h_{\rm cv}}$. For the small sample sizes we suggest to use $\widehat{f}_{\alpha_{\rm cv}}$ instead of $\widehat{f}_{h_{\rm cv}}$ and $f^*_{\alpha^*_{\rm cv}}$.

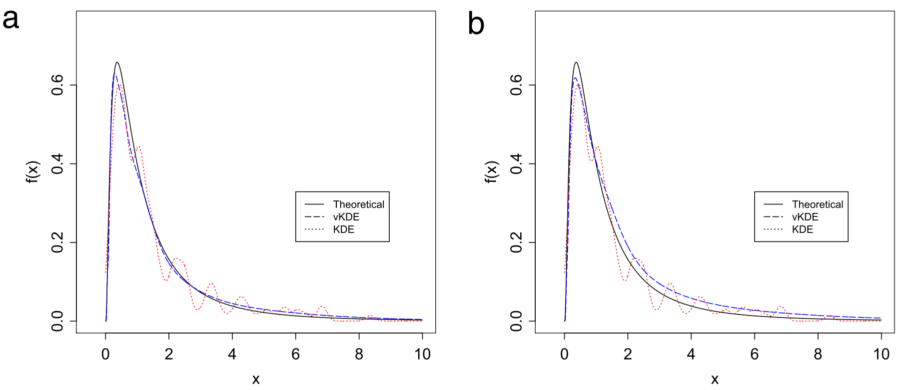

Fig. 1. Estimation of Log-normal(0, 1) density function f (solid curve) by $\widehat{f}_{h_{cv}}$ with $h_{cv} = 0.14$ and by (a) $\widehat{f}_{\alpha_{cv}}$ with $\alpha_{cv} = 11$; (b) $f^*_{\alpha^*_{cv}}$ with $\alpha^*_{cv} = 7$. In both plots $n = 200$.

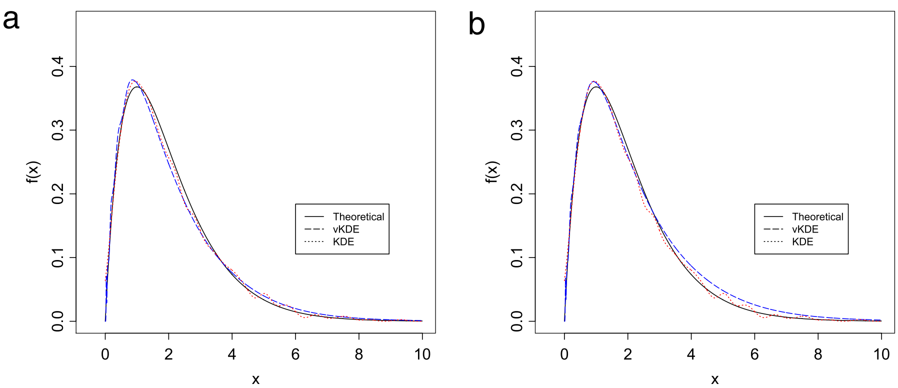

Fig. 2. Estimation of Gamma(2, 1) density function f (solid curve) by $\widehat{f}_{h_{cv}}$ with $h_{cv} = 0.19$ and by (a) $\widehat{f}_{\alpha_{cv}}$ with $\alpha_{cv} = 24$; (b) $f^*_{\alpha^*_{cv}}$ with $\alpha^*_{cv} = 18$. In both plots $n = 800$.

Acknowledgments

The authors are thankful to Estate Khmaladze and Cecil Burchfiel for helpful discussions, and the referee and the associate editor for their suggestions that led to a better presentation. The research was supported by NSF grant DMS-0906639. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

References

Chen, S.X., 2000. Probability density function estimation using gamma kernels. Ann. Inst. Statist. Math. 52, 471–480.
Gzyl, H., Tagliani, A., 2010. Hausdorff moment problem and fractional moments. Appl. Math. Comput. 216, 3319–3328.
Klauder, J.R., Penson, K.A., Sixdeniers, J.-M., 2001. Constructing coherent states through solutions of Stieltjes and Hausdorff moment problems. Phys. Rev. A 64, 013817.
Mnatsakanov, R.M., 2008a. Hausdorff moment problem: reconstruction of distributions. J. Statist. Probab. Lett. 78, 1612–1618.
Mnatsakanov, R.M., 2008b. Hausdorff moment problem: reconstruction of probability density functions. J. Statist. Probab. Lett. 78, 1869–1877.
Mnatsakanov, R.M., Khmaladze, E.V., 1981. On L_1-convergence of statistical kernel estimators of distribution densities. Soviet Math. Dokl. 23, 633–636.
Mnatsakanov, R.M., Ruymgaart, F.H., 2012. Moment-density estimation for positive random variables. Statistics 46, 215–230.
Mnatsakanov, R.M., Ruymgaart, F.H., 2003. Some properties of moment-empirical cdf's with application to some inverse estimation problems. Math. Methods Statist. 12, 478–495.
Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076.
Scaillet, O., 2004. Density estimation using inverse and reciprocal inverse gaussian kernels. Nonparametric Stat. 16, 217–226.
Scott, D.W., 1992. Multivariate Density Estimation. John Wiley and Sons, New York.
Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York.
Sneddon, I.N., 1974. The Use of Integral Transforms. McGraw-Hill, New York.
Stoyanov, J., 2000. Krein condition in probabilistic moment problems. Bernoulli 6, 939–949.
Tagliani, A., 2001. Recovering a probability density function from its Mellin transform. Appl. Math. Comput. 118, 151–159.