Annals of the Institute of Statistical Mathematics (2021) 73:953–977
https://doi.org/10.1007/s10463-020-00772-1

Mellin–Meijer kernel density estimation on $\mathbb{R}^+$

Gery Geenens[1]

Received: 7 February 2018 / Revised: 6 October 2020 / Accepted: 9 October 2020 /
Published online: 24 November 2020

Abstract

Kernel density estimation is a nonparametric procedure making use of the smoothing power of the convolution operation. Yet, it performs poorly when the density of a positive variable is estimated, due to boundary issues. So, various extensions of the kernel estimator allegedly suitable for $\mathbb{R}^+$-supported densities, such as those using asymmetric kernels, abound in the literature. Those, however, are not based on any valid smoothing operation. By contrast, in this paper a kernel density estimator is defined through the Mellin convolution, the natural analogue on $\mathbb{R}^+$ of the usual convolution. From there, a class of asymmetric kernels related to Meijer G-functions is suggested, and asymptotic properties of this 'Mellin–Meijer kernel density estimator' are presented. In particular, its pointwise- and L_2-consistency (with optimal rate of convergence) are established for a large class of densities, including densities unbounded at 0 and showing power-law decay in their right tail.

Keywords Kernel density estimator · Boundary issues · Asymmetric kernels · Mellin transform · Meijer G-functions

1 Introduction

Kernel density estimation is a very popular nonparametric method which simply makes use of the smoothing power of the convolution operation for estimating an unknown probability density function. Let $\{X_k, k = 1, \dots, n\}$ be a sample drawn from a distribution F admitting a density f with respect to the Lebesgue measure, and $\mathbb{P}_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$ be its empirical measure. The conventional kernel density estimator of f is just

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-020-00772-1) contains supplementary material, which is available to authorised users.

✉ Gery Geenens
ggeenens@unsw.edu.au

[1] School of Mathematics and Statistics, UNSW Sydney, Kensington, NSW 2032, Australia

$$\hat{f}(x) = \left(K_h * \mathbb{P}_n\right)(x) = \frac{1}{n}\sum_{k=1}^{n} K_h(x - X_k), \tag{1}$$

where K is a unit-variance probability density symmetric around 0 and $K_h(\cdot) = K(\cdot/h)/h$ is the rescaled version of K with standard deviation $h > 0$. The properties of $\hat{f}$ are well understood (Wand and Jones 1995). A major downside of (1) is that it suffers from *boundary bias* when the support of f is not the whole real line: the terms $K_h(x - X_k)$'s corresponding to X_k's close to a boundary overflow beyond it in the forbidden area, preventing consistency of the estimator (Wand and Jones 1995, Section 2.11). Hence boundary corrections for (1) abound in the literature, such as the 'cut-and-normalised' method and its variants based on 'boundary kernels' (Cheng et al. 1997; Dai and Sperlich 2010; Jones 1993; Jones and Foster 1996; Müller 1991; Zhang et al. 1999) or the reflection method (Karunamuni and Alberts 2005; Schuster 1985). Those are essentially *ad hoc* manual surgeries on (1) at the boundary, though.

In the important particular case where the density f is supported on $\mathbb{R}^+ = [0, +\infty)$,[1] a more global approach has been to construct an estimator in the form

$$\hat{f}(x) = \frac{1}{n}\sum_{k=1}^{n} \mathcal{K}_h(X_k; x), \tag{2}$$

where $\mathcal{K}_h(\cdot; x)$ is an asymmetric $\mathbb{R}^+$-supported density whose parameters are functions of $x \geq 0$ and a smoothing parameter $h > 0$. Using asymmetric kernels supposedly enables the estimator to take the constrained nature of the support of f into account. In his pioneering work, Chen (2000) studied the 'first' gamma kernel density estimator[2]

$$\hat{f}(x) = \frac{1}{n}\sum_{k=1}^{n} \frac{X_k^{x/h^2} e^{-X_k/h^2}}{h^{2x/h^2+2}\Gamma\left(x/h^2 + 1\right)}$$

corresponding to (2) with $\mathcal{K}_h(\cdot; x)$ the gamma density with shape parameter $\alpha = 1 + x/h^2$ and rate $\beta = 1/h^2$. Other asymmetric kernels were investigated, e.g., log-normal (Igarashi 2016; Jin and Kawczak 2003), Birnbaum-Saunders (Igarashi and Kakizawa 2014; Jin and Kawczak 2003; Marchant et al. 2013), or inverse Gaussian (Igarashi and Kakizawa 2014; Scaillet 2004), while Hirukawa and Sakudo (2015), Igarashi and Kakizawa (2018) and Kakizawa (2018) described families of 'generalised gamma' and 'mixture inverse Gaussian' kernels in an attempt to standardise those results for a variety of asymmetric kernels.

Yet, such asymmetric kernel estimators do not entirely fix boundary issues, and they need another manual correction around 0 for performing satisfyingly ('second' or 'modified' gamma kernel estimator in Chen (2000); see also Hirukawa and Sakudo

[1] Estimation of a density supported on $[a, \infty)$ or $(-\infty, b]$, $a, b \in \mathbb{R}$, is achieved in the exact same way through a straightforward change of origin and/or reflection.

[2] We use h^2 instead of Chen's original b for the smoothing parameter.

(2015, Conditions 1 and 2) or Kakizawa (2018, Section 3))—note that even these corrected versions may show disappointing boundary behaviour (Malec and Schienle 2014; Zhang 2010). Those problems originate in that estimators like (2) are not induced by any valid smoothing operation on $\mathbb{R}^+$. Among unpleasant consequences, this implies that (2) does not automatically integrate to one, hence is not a *bona fide* density. Note that Jones and Henderson (2007) and Jeon and Kim (2013) obtained *bona fide* asymmetric kernel density estimators by swapping around the roles of x and X_k in (2), that is, estimators of the form

$$\hat{f}_0(x) = \frac{1}{n}\sum_{k=1}^{n} \mathcal{K}_h(x;X_k), \tag{3}$$

where $\mathcal{K}_h$ is this time a proper density in x whose parameters depend on X_k.

In this paper, we revisit the idea of 'asymmetric kernel density estimation' on $\mathbb{R}^+$ from a novel perspective, by defining an estimator based on the natural smoothing operation on $\mathbb{R}^+$. The convolution of two probability densities $g_1 * g_2$ is known to be the density of the sum of two independent random variables having respective densities g_1 and g_2. Hence, smoothing is achieved in (1) through 'diluting' each observation X_k by *adding* to it some continuous random noise ε with density K_h. Indeed it is incoherent if f is $\mathbb{R}^+$-supported: as $X_k + \varepsilon$ may be negative, it produces estimates $\hat{f}$ which 'spill over', which implies the boundary issues. In algebraic terms, the conventional estimator (1) is justified on $\mathbb{R}$ but not on $\mathbb{R}^+$ because $(\mathbb{R}, +)$ is a group but not $(\mathbb{R}^+, +)$. By contrast, $(\mathbb{R}^+, \times)$ is a group, which motivates an estimator realising smoothing by *multiplying* each X_k by a positive random disturbance ε.

If X_1 and X_2 are two independent positive random variables with respective densities g_1 and g_2, then the density of their product is

$$g(x) = \int_0^\infty g_1\left(\frac{x}{v}\right) g_2(v)\,\frac{\mathrm{d}v}{v}. \tag{4}$$

This operation, hereafter denoted by $g_1 *_{\mathcal{M}} g_2$, is called *Mellin convolution*, as it is strongly associated with the *Mellin transform*, the natural analytical tool for studying products of independent random variables (Epstein 1948). Consequently, for estimating a density f supported on $\mathbb{R}^+$, this paper proposes and fully investigates a 'Mellin version' of the kernel estimator, whose basic definition is

$$\hat{f}_0(x) = (L_\eta *_{\mathcal{M}} \mathbb{P}_n)(x) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_k} L_\eta\left(\frac{x}{X_k}\right) \tag{5}$$

for an $\mathbb{R}^+$-supported density L_η whose 'spread' (to make precise later) is driven by a smoothing parameter $\eta > 0$. In particular, it will be seen that kernel functions L_η that fit naturally in this framework belong to a family of distributions strongly related to Meijer G-functions (Meijer 1936)—the duality between the Mellin transform and Meijer G-functions in statistics was already elucidated in Kabe (1958). Hence we call the whole methodology *Mellin–Meijer kernel density estimation*. We will mainly study the case where L_η is a scaled-powered F-density, as well as a refined 'sample-smoothing' version of (5) allowing the parameters of that kernel to vary with X_k—see details in Sect. 4.

Clearly (5) is just the multiplicative analogue of (1), and assesses which observations X_k are local to x through the ratios x/X_k. Note that (5) shares similarities with expression (2) in Comte and Genon-Catalot (2012) and expression (2.4) in Mnatsakanov and Sarkisian (2012),[3] although they are different. In particular, neither of those two estimators integrates to one, whereas $\hat{f}_0$ always defines a *bona fide* density. Indeed (5) is an asymmetric kernel density estimator of type (3).

After reviewing the main properties of the Mellin transform useful here (Sect. 2), this paper lays the foundations of *Mellin–Meijer kernel density estimation* by defining the class of Meijer distributions, natural kernels in this framework (Sect. 3), and the estimator (Sect. 4). Asymptotic results are obtained in Sect. 5, while Sect. 6 explores a novel way of selecting the smoothing parameter η in practice, again taking advantage of the 'Mellin' perspective. Sections 7 and 8 investigate the performance of the estimator in practice through simulations and real data examples. Section 9 concludes.

2 Mellin transform

The *Mellin transform* of any locally integrable $\mathbb{R}^+$-supported function f is the function defined on the complex plane $\mathbb{C} = \{z : z = c + i\omega; c, \omega \in \mathbb{R}\}$ as

$$\mathcal{M}(f;z) = \int_0^\infty x^{z-1} f(x)\,\mathrm{d}x, \tag{6}$$

when the integral exists. If, for some $\delta > 0$ and $a < b$, $f(x) = O(x^{-(a-\delta)})$ as $x \to 0^+$ and $f(x) = O(x^{-(b+\delta)})$ as $x \to +\infty$, then (6) converges absolutely on the vertical strip of the complex plane $\mathcal{S}_f = \{z \in \mathbb{C} : a < \Re(z) < b\}$. It can be shown that $\mathcal{M}(f;\cdot)$ is holomorphic on $\mathcal{S}_f$—therefore known as the *strip of holomorphy* of $\mathcal{M}(f;\cdot)$—and uniformly bounded on any closed vertical strip contained in $\mathcal{S}_f$. There is a one-to-one correspondence between f and *the couple* $(\mathcal{M}(f;\cdot), \mathcal{S}_f)$, in the sense that two different functions may have the same Mellin transform, but defined on two non-overlapping vertical strips of $\mathbb{C}$. It is thus equivalent to know f or $\mathcal{M}(f;\cdot)$ in a given vertical strip of $\mathbb{C}$. In particular, f can be recovered from $\mathcal{M}(f;\cdot)$ by the *inverse Mellin transform*:

$$f(x) = \frac{1}{2\pi i} \int_{\Re(z)=c} x^{-z} \mathcal{M}(f;z)\,\mathrm{d}z \tag{7}$$

for any real $c \in \mathcal{S}_f$. Cauchy's residue theorem allows the integration path (a vertical line in $\mathbb{C}$) to be displaced sideways inside $\mathcal{S}_f$ without affecting the value of integral, which is independent of $c \in \mathcal{S}_f$. In addition, for any $c \in \mathcal{S}_f$,

[3] Mnatsakanov and Ruymgaart (2012) and Mnatsakanov and Sarkisian (2012) briefly mentioned estimator (5) as such, but gave up the idea and focused on a modified version.

$$\int_0^\infty x^{2c-1} f^2(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{\Re(z)=c} |\mathcal{M}(f;z)|^2\,\mathrm{d}z, \tag{8}$$

which is the Mellin version of Parseval's identity. See Paris and Kaminski (2001, Chapter 3) for a comprehensive treatment of the Mellin transform.

Now, if f is the probability density of a positive random variable X, then $\mathcal{M}(f;1) = \int_0^\infty f(x)\,\mathrm{d}x = 1$. Hence, the line $\{z \in \mathbb{C} : \Re(z) = 1\}$ is always part of $\mathcal{S}_f$, which allows f to be unequivocally represented by its Mellin transform $\mathcal{M}(f;\cdot)$. From (6), one has $\mathcal{M}(f;z) = \mathbb{E}(X^{z-1})$. Thus $\mathcal{M}(f;\cdot)$ actually defines all real, complex, integral and fractional moments of X, and $\mathcal{S}_f$ is determined by the existence (finiteness) of the real moments of f: $z \in \mathcal{S}_f$ if and only if $\mathbb{E}(X^{\Re(z)-1}) < \infty$. So, $b = \infty$ for light-tailed densities whose all positive moments exist, while $\mathcal{S}_f$ is bounded from the right ($1 < b < \infty$) for fat-tailed densities with only a certain number of finite positive moments.[4] Similarly, $a = -\infty$ for densities whose all negative moments exist—let us call such densities '*light-headed*', while $\mathcal{S}_f$ is bounded from the left ($-\infty < a < 1$) for '*fat-headed*' densities, for which some negative moments are infinite.

Let X_1, X_2 be two independent positive random variables with respective densities g_1 and g_2, and let g be the density of their product (4). Then

$$\mathcal{M}(g;z) = \mathbb{E}((X_1X_2)^{z-1}) = \mathbb{E}(X_1^{z-1})\mathbb{E}(X_2^{z-1}) = \mathcal{M}(g_1;z)\mathcal{M}(g_2;z) \tag{9}$$

for $z \in \mathcal{S}_{g_1} \cap \mathcal{S}_{g_2}$. Note that $g = (g_1 *_{\mathcal{M}} g_2) \iff \mathcal{M}(g;\cdot) = \mathcal{M}(g_1;\cdot)\mathcal{M}(g_2;\cdot)$, so Mellin transform/Mellin convolution play the same role for products of independent variables as Fourier transform/convolution for sums. From the operational properties of the Mellin transform (Geenens 2020, Section A), one can show:

Lemma 1 *Let X be a positive random variable whose density f_X has Mellin transform $\mathcal{M}(f_X;z)$ on $\mathcal{S}_{f_X} = \{z \in \mathbb{C} : a < \Re(z) < b\}$ for some $a < 1 < b$. Then, for $\nu > 0$ and $\xi \in \mathbb{R}$, the random variable $Y = \nu X^\xi$ has density f_Y whose Mellin transform is $\mathcal{M}(f_Y;z) = \nu^{z-1}\mathcal{M}\big(f_X;1+\xi(z-1)\big)$ on $\mathcal{S}_{f_Y} = \{z \in \mathbb{C} : 1 - \frac{1-a}{\xi} < \Re(z) < 1 + \frac{b-1}{\xi}\}$ ($\xi > 0$) or $\mathcal{S}_{f_Y} = \{z \in \mathbb{C} : 1 - \frac{b-1}{|\xi|} < \Re(z) < 1 + \frac{1-a}{|\xi|}\}$ ($\xi < 0$).*

For illustration, let $f_{\text{Gam}}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ be the Gamma(α,β)-density $(\alpha,\beta > 0)$. By definition, $\mathcal{M}(e^{-x};z) = \Gamma(z)$ ($\Re(z) > 0$), from which it follows

$$\mathcal{M}(f_{\text{Gam}};z) = \frac{1}{\beta^{z-1}}\frac{\Gamma(\alpha+z-1)}{\Gamma(\alpha)}, \qquad \Re(z) > 1-\alpha, \tag{10}$$

with (A.2)-(A.4) in Geenens (2020). Now, consider the inverse gamma (α,β)-density f_{IGam}, i.e., the density of $1/X$ when $X \sim \text{Gamma}(\alpha,\beta)$. By Lemma 1 with $\xi = -1$, $\mathcal{M}(f_{\text{IGam}};z) = \mathcal{M}(f_{\text{Gam}};2-z)$, that is,

[4] The qualifiers 'fat', 'heavy' or 'long' sometimes find different meanings in the literature when describing the tails of a distribution. In this paper, by 'fat-tailed' distribution we mean explicitly a distribution whose not all positive power moments are finite. Hence here the log-normal is 'light-tailed', although it is regarded as 'heavy-tailed' in many references.

$$\mathcal{M}(f_{\text{IGam}};z) = \frac{1}{\beta^{1-z}}\frac{\Gamma(\alpha+1-z)}{\Gamma(\alpha)}, \qquad \Re(z) < 1+\alpha. \tag{11}$$

3 Meijer densities

For all $\nu, \gamma, \xi > 0$ and $\theta \in (0, \pi/2)$, consider the $\mathbb{R}^+$-supported function $L_{\nu,\gamma,\xi,\theta}$ whose Mellin transform is

$$\begin{aligned}\mathcal{M}(L_{\nu,\gamma,\xi,\theta};z) &= \nu^{z-1}\left(\frac{1}{\tan^2\theta}\right)^{\xi(z-1)} \\ &\times \frac{\Gamma\left(\frac{\xi^2}{\gamma^2\cos^2\theta}+\xi(z-1)\right)\Gamma\left(\frac{\xi^2}{\gamma^2\sin^2\theta}+\xi(1-z)\right)}{\Gamma\left(\frac{\xi^2}{\gamma^2\cos^2\theta}\right)\Gamma\left(\frac{\xi^2}{\gamma^2\sin^2\theta}\right)}\end{aligned} \tag{12}$$

on

$$\mathcal{S}_{L_{\nu,\gamma,\xi,\theta}} = \left\{ z \in \mathbb{C} : 1 - \frac{\xi}{\gamma^2\cos^2\theta} < \Re(z) < 1 + \frac{\xi}{\gamma^2\sin^2\theta} \right\}. \tag{13}$$

The following result establishes that $L_{\nu,\gamma,\xi,\theta}$ is a scaled-powered-F-density:

Proposition 1 *For all $\nu, \gamma, \xi > 0$ and $\theta \in (0, \pi/2)$, the $\mathbb{R}^+$-supported function $L_{\nu,\gamma,\xi,\theta}$, whose Mellin transform is (12) on $\mathcal{S}_{L_{\nu,\gamma,\xi,\theta}}$ (13), is the density of the random variable $Y = \nu X^\xi$, where X follows the Fisher–Snedecor F-distribution with $\frac{2\xi^2}{\gamma^2\cos^2\theta}$ and $\frac{2\xi^2}{\gamma^2\sin^2\theta}$ degrees of freedom.*

Proof This follows from the characterisation of the F-distribution as a (scaled) ratio of two independent gamma random variables, (9), (10), (11) and Lemma 1; see Geenens (2020, Section C) for details. □

By taking the limit $\theta \to 0$ or $\theta \to \pi/2$, we can define

$$\mathcal{M}(L_{\nu,\gamma,\xi,0};z) = \frac{\nu^{z-1}}{\Gamma\left(\frac{\xi^2}{\gamma^2}\right)}\left(\frac{\gamma^2}{\xi^2}\right)^{\xi(z-1)}\Gamma\left(\frac{\xi^2}{\gamma^2}+\xi(z-1)\right), \quad \Re(z) > 1 - \frac{\xi}{\gamma^2}, \tag{14}$$

$$\mathcal{M}(L_{\nu,\gamma,\xi,\pi/2};z) = \frac{\nu^{z-1}}{\Gamma\left(\frac{\xi^2}{\gamma^2}\right)}\left(\frac{\gamma^2}{\xi^2}\right)^{\xi(1-z)}\Gamma\left(\frac{\xi^2}{\gamma^2}+\xi(1-z)\right), \quad \Re(z) < 1 + \frac{\xi}{\gamma^2}. \tag{15}$$

Then $L_{\nu,\gamma,\xi,0}$ is the density of νX^ξ for $X \sim \text{Gamma}\left(\frac{\xi^2}{\gamma^2}, \frac{\xi^2}{\gamma^2}\right)$ and $L_{\nu,\gamma,\xi,\pi/2}$ that of νX^ξ for $X \sim \text{InvGamma}\left(\frac{\xi^2}{\gamma^2}, \frac{\xi^2}{\gamma^2}\right)$, in agreement with the usual interpretation of the F-distribution with infinite degrees of freedom.

The strip of holomorphy (13) clarifies how the parameters γ, ξ and θ act on the lightness/fatness of the head and the tail of the density $L_{\nu,\gamma,\xi,\theta}$. Playing on γ, ξ and θ, one can produce a wide variety of different head and tail behaviours for $L_{\nu,\gamma,\xi,\theta}$. Those include exponential behaviours ($\theta = 0$ or $\theta = \pi/2$, or $\xi \to \infty$), and positiveness/unboundedness at $x = 0$, for $\xi < \gamma^2 \cos^2 \theta$.

We call a probability density whose Mellin transform is (12) (or (14)/(15)) a *Meijer density*, as it is strongly related to Meijer G-functions. These are very general functions, conveniently defined by their Mellin transforms (Prudnikov et al. 1990, Section 8.2), whose particular cases cover most of the common, useful or special functions on $\mathbb{R}^+$ (Beals and Szmigielski 2013). In particular, for $a, b \in \mathbb{R}$ such that $a - b < 1$, the G-function $G^{1,1}_{1,1}$ has Mellin transform

$$\mathcal{M}\left(G^{1,1}_{1,1}\left(\cdot \middle| \begin{matrix} a \\ b \end{matrix}\right);z\right) = \Gamma(b+z)\Gamma(1-a-z)$$

on $\{z \in \mathbb{C} : -b < \Re(z) < 1 - a\}$. All probability densities $L_{\nu,\gamma,\xi,\theta}$, whose Mellin transforms (12) are rescaled products of two gamma functions, are thus rescaled versions of $G^{1,1}_{1,1}$.

Most of the $\mathbb{R}^+$-supported probability distributions of practical interest are actually Meijer distributions; see Geenens (2020, Section B). These include gamma and inverse gamma,[5] as well as the generalised F-distribution (Cox 2008). They also admit the log-normal as limiting case as $\xi \to \infty$. All these distributions have a tractable Mellin transform (12), whereas most of them do not admit an explicit characteristic function (Fourier transform).

The following results establish that the parameter γ is asymptotically equivalent to the coefficient of variation of $L_{\nu,\gamma,\xi,\theta}$, say χ, and provide asymptotic expansions of $\mathcal{M}(L_{\nu,\gamma,\xi,\theta};z)$ and $\mathcal{M}(L^2_{\nu,\gamma,\xi,\theta};z)$ as $\gamma \to 0$.

Proposition 2 *Let $\nu, \xi > 0$ and $\theta \in [0, \pi/2]$ be fixed. The coefficient of variation χ of $L_{\nu,\gamma,\xi,\theta}$ is such that $\chi \sim \gamma$, as $\gamma \to 0$.*

Proposition 3 *Let $\xi > 0$ and $\theta \in [0, \pi/2]$* be fixed. Let $\nu = 1 + \frac{1}{2}\gamma^2\left(1 + \frac{\cos 2\theta}{\xi}\right)$. *Then, as $\gamma \to 0$,*

(i) *the Mellin transform (12) of $L_{\nu,\gamma,\xi,\theta}$ on $\mathcal{S}_{L_{\nu,\gamma,\xi,\theta}}$ (13) is such that*

[5] It appears from (12) and Lemma 1 that the class of Meijer distributions is closed under the 'inverse' operation.

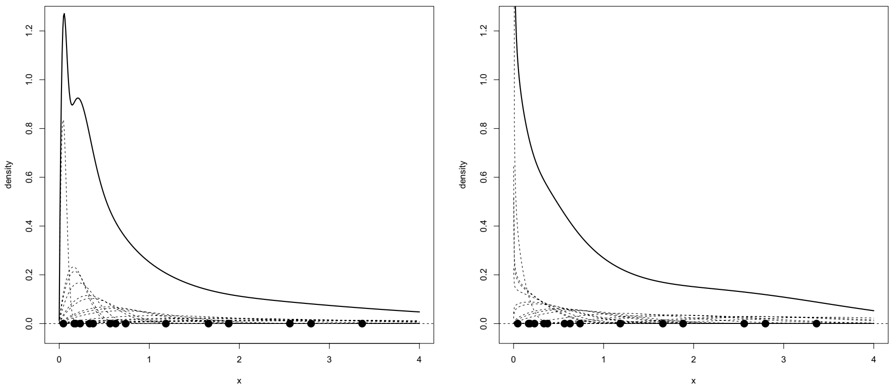

Fig. 1 Construction of the basic estimator (5) (left panel) and its refined version (19) with (18) (right panel) for an artificial sample of size $n = 15$. The observations X_k (big dots) and the associated 'bumps' $\Lambda_\eta^{(k)}$ (dashed lines) are shown. The final estimator (thick line) is the sum of those bumps. In both cases, the smoothing parameter is $\eta = 0.5$ and the kernel is a Meijer density with $\xi = 1/2$ and $\theta = 0$: in (5), $L_\eta = L_{1+\eta^2,\eta,1/2,0}$; and in (19), $L_\eta^{(k)}$ is the Meijer kernel $L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},1/2,0}$, as described below (21) and (22)

$$\mathcal{M}(L_{\nu,\gamma,\xi,\theta};z) = 1 + \frac{\gamma^2}{2} z(z-1) + \rho(\gamma, z), \tag{16}$$

where $|\rho(\gamma,z)| = O(\gamma^4(1+|z-1|)^2)$, *provided* $|z-1| = o(\gamma^{-2})$.

(ii) *the Mellin transform of* $L^2_{\nu,\gamma,\xi,\theta}$ *on* $\mathcal{S}_{L^2_{\nu,\gamma,\xi,\theta}} = \left\{ z \in \mathbb{C} : z/2 \in \mathcal{S}_{L_{\nu,\gamma,\xi,\theta}} \right\}$ *is such that*

$$\mathcal{M}(L^2_{\nu,\gamma,\xi,\theta};z) = \frac{1}{2\sqrt{\pi}\gamma}(1+\omega(\gamma,z)), \tag{17}$$

where $|\omega(\gamma,z)| = O(\gamma^2(1+|z-2|))$, *provided* $|z-2| = o(\gamma^{-2})$.

Proof See Geenens (2020, Section C). □

A notable observation is that, with the suggested value of $\nu = 1 + \frac{1}{2}\gamma^2\left(1 + \frac{\cos 2\theta}{\xi}\right)$, expansions (16) and (17) do not depend on θ or ξ, but only on γ.

4 Mellin–Meijer kernel density estimation

Like the conventional estimator (Härdle et al. 2004, Section 3.1.5), estimator (5)

$$\hat{f}_0(x) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_k} L_\eta\left(\frac{x}{X_k}\right)$$

is constructed as a sum of 'bumps' $\Lambda_\eta^{(k)}(x) \doteq \frac{1}{X_k} L_\eta\left(\frac{x}{X_k}\right)$, as illustrated by Fig. 1 (left panel) for an artificial sample of size $n = 15$. Unlike in the conventional case, though, here the 'bumps' do not have the same width: if σ_η is the standard deviation of L_η, then $\Lambda_\eta^{(k)}$ has standard deviation $\sigma_\eta^{(k)} = X_k\, \sigma_\eta$, obviously different for each k. The bumps $\Lambda_\eta^{(k)}$'s corresponding to the X_k's close to the boundary 0 are high and narrow, while those in the right tail are wide and flat. More smoothing is thus automatically applied in the tail than close to the boundary, and this is essentially how the boundary issue is addressed by (5). What is common to all the $\Lambda_\eta^{(k)}$'s is actually their *coefficient of variation* $\chi_\eta^{(k)} \equiv \chi_\eta$, that of the 'canonical bump' L_η. This points out the natural role of the coefficient of variation of the kernel L_η In this framework, suggesting to defining it as the global smoothing parameter η.

Unfortunately, estimator (5) shows very disappointing performance, though, which can be related to what Geenens and Wang (2018, Sections 2.2 and 3.1, Figure 1) described about the 'log-transformed' kernel density estimator. Specifically, that estimator first sends the observations onto the whole $\mathbb{R}$ through the log-transform, performs smoothing by adding to them some random disturbance, and takes everything back to $\mathbb{R}^+$ by exponentiation. So, on $\mathbb{R}^+$, smoothing is realised by multiplying each X_k by a positive random disturbance ε, exactly as (5), and is thus a particular case thereof. Geenens and Wang (2018) explained how and why this leads to severe undersmoothing at the boundary and oversmoothing in the tail. Here, given that $\sigma_\eta^{(k)} = X_k\, \sigma_\eta \simeq 0$ for $X_k \simeq 0$, the effective amount of smoothing applied in the boundary area is virtually nil, while it is very high in the tail area, as $\sigma_\eta^{(k)}$ gets huge. A natural fix is to operate some *smoothing transfer*: make the estimator use some of the amount of smoothing in excess in the tail for filling the shortage of smoothing at the boundary. This transfer is achieved by making the coefficient of variation $\chi_\eta^{(k)}$ of $\Lambda_\eta^{(k)}$ a decreasing function of X_k, instead of keeping it constant. Set

$$\chi_\eta^{(k)} = \frac{\eta}{\sqrt{X_k + \eta^2}}, \tag{18}$$

a choice driven by theoretical considerations (Sect. 5). Figure 1 (right panel) shows, for the same sample and with the same smoothing parameter η as in the left panel, how the bumps at the boundary are no more as narrow, and the bumps in the tail no more as flat, as in the initial case. On this empirical illustration, the final estimate of f seems rightly smooth all over $\mathbb{R}^+$. It highlights another major benefit of allowing $\chi_\eta^{(k)}$ to depend on X_k. In the basic case (5), all bumps $\Lambda_\eta^{(k)}(x)$'s are rescaled versions of L_η and have the same shape. In particular, if $L_\eta(0) = 0$, then $\hat{f}_0(0) \equiv 0$ automatically (Fig. 1, left).[6] This is no more the case when $\Lambda_\eta^{(k)}$ may have different coefficients of variation: in Fig. 1 (right), the bumps associated to the data close to 0 are no more tied down to 0. As their coefficient of variation increases, they are forced to climb along the y-axis, allowing $\hat{f}(0) \neq 0$. This justifies to define a 'refined' version of (5) as

[6] This problem of '$\hat{f}(0) \equiv 0$' is shared by many other kernel estimators, e.g., Jin and Kawczak (2003); Marchant et al. (2013); Mnatsakanov and Sarkisian (2012); Scaillet (2004).

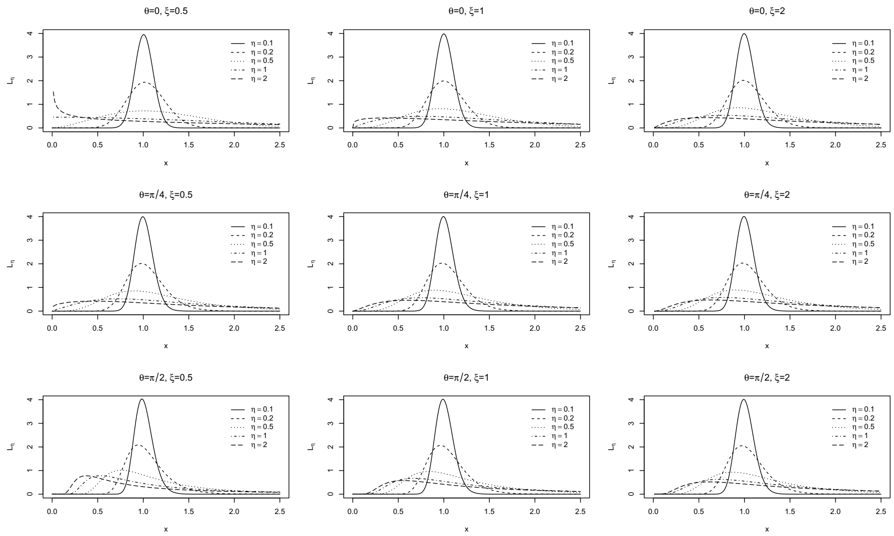

Fig. 2 Meijer kernels $L_\eta^{(k)} = L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},\xi,\theta}$ for $\theta = 0, \pi/4, \pi/2$ and $\xi = 1/2, 1, 2$ for $X_k = 1$

$$\hat{f}(x) = \frac{1}{n}\sum_{k=1}^{n}(L_\eta^{(k)} *_{\mathcal{M}} \delta_{X_k})(x) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{X_k}L_\eta^{(k)}\left(\frac{x}{X_k}\right), \tag{19}$$

where each $L_\eta^{(k)}$ has a coefficient of variation $\chi_\eta^{(k)}$ given by (18). From (4) and (9), $\hat{f}$ admits the Mellin transform

$$\mathcal{M}(\hat{f};z) = \frac{1}{n}\sum_{k=1}^{n}\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1} \tag{20}$$

on $\mathcal{S}_{\hat{f}} = \bigcap_{k=1}^{n}\mathcal{S}_{L_\eta^{(k)}}$.

Remark 1 In some sense, (19) is a 'sample-smoothing' kernel estimator (Terrell and Scott 1992), as the smoothing parameter $\chi_\eta^{(k)}$ associated with the particular bump $\Lambda_\eta^{(k)}$ varies with X_k. However, conventional 'sample-smoothing' typically requires pilot estimation of f, which is not without causing further issues (Hall et al. 1995). Here, it is deterministically articulated around (18).

The connection between the Mellin transform and Meijer densities (Sect. 3) suggests to take for $L_\eta^{(k)}$ in (19) a Meijer density $L_{\nu,\gamma,\xi,\theta}$. Fix the parameters $\xi > 0$ and $\theta \in [0, \pi/2]$. Those determine more specifically the type of kernels that will be used. For instance, with $\theta = 0$ and $\xi = 1$, then the $L_\eta^{(k)}$'s are gamma densities; with $\theta = \pi/2$ and $\xi = 1$, then the $L_\eta^{(k)}$'s are inverse gamma densities; refer to Geenens (2020, Section B) for other choices of kernels. Motivated by (18) and Propositions 2 and 3, set

$$\gamma \doteq \gamma_\eta^{(k)} = \frac{\eta}{\sqrt{X_k + \eta^2}} \tag{21}$$

for some smoothing parameter $\eta > 0$, and

$$\nu \doteq \nu_\eta^{(k)} = 1 + \frac{1}{2}\,\gamma_\eta^{(k)\,2}\left(1 + \frac{\cos 2\theta}{\xi}\right). \tag{22}$$

This ensures through (16) that, asymptotically, the behaviour of $L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},\xi,\theta}$ is independent of θ and ξ, which allows an integrated theory, not specific to a particular choice of kernel, to be developed. We call kernels $L_\eta^{(k)} = L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},\xi,\theta}$ with this parameterisation, *Meijer kernels*. Figure 2 shows examples of Meijer kernels for $\xi \in \{1/2, 1, 2\}$ and $\theta \in \{0, \pi/4, \pi/2\}$, for $X_k = 1$. As η approaches 0, the kernels concentrate around 1 with a fading effect of the values of θ and ξ on their shape, as suggested by expansion (16).

We call the estimator (19) with Meijer kernel $L_\eta^{(k)} = L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},\xi,\theta}$, the Mellin–Meijer kernel density estimator. Explicitly, following Proposition 1, it is

$$\begin{aligned}\hat{f}(x) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\xi \nu_\eta^{(k)} X_k}\left(\frac{x}{\nu_\eta^{(k)} X_k}\right)^{1/\xi - 1} \\ \times f_{\mathrm{F}}\left(\left(\frac{x}{\nu_\eta^{(k)} X_k}\right)^{1/\xi}; \frac{2\xi^2(\eta^2 + X_k)}{\eta^2\cos^2\theta}, \frac{2\xi^2(\eta^2 + X_k)}{\eta^2\sin^2\theta}\right),\end{aligned} \tag{23}$$

where $\nu_\eta^{(k)} = 1 + \frac{\eta^2 + X_k}{2\eta^2}(1 + \frac{\cos 2\theta}{\xi})$ and $f_{\mathrm{F}}(\cdot;\mathrm{df}_1, \mathrm{df}_2)$ is the density of the F-distribution with df_1 and df_2 degrees of freedom.

5 Asymptotic properties

Let us assume the following.

Assumption 1 The sample $\{X_k, k = 1, \dots, n\}$ consists of i.i.d. replications of a positive random variable X whose distribution F admits a density f twice continuously differentiable on $(0, \infty)$;

Assumption 2 There exist $\alpha, \beta \in (0, +\infty]$, with $2\alpha + \beta > 5/2$, such that $\mathbb{E}(X^{-\alpha}) < \infty$ and $\mathbb{E}(X^{\beta}) < \infty$;

Assumption 3 For all $\eta > 0$ and $k \in \{1, \dots, n\}$, $L_\eta^{(k)}$ is a Meijer kernel $L_{\nu_\eta^{(k)},\gamma_\eta^{(k)},\xi,\theta}$ with $\xi > 0$ and $\theta \in [0, \pi/2]$ such that $\frac{\xi}{\cos^2\theta} > \frac{1}{4} - \frac{\beta}{2}$ and $\frac{\xi}{\sin^2\theta} > 1 - \alpha$;

Assumption 4 The smoothing parameter $\eta \doteq \eta_n$ is such that $\eta \to 0$ and $n\eta \to \infty$ as $n \to \infty$.

Assumption 1 fixes the set-up. The requirement that f has two continuous derivatives is classical in kernel density estimation. Assumption 2 is a condition on the existence of some negative and positive moments of f. It excludes densities with *both* very fat head and tail, but it is very mild. In particular, f is allowed to be positive and even unbounded at the boundary $x = 0$ ($\alpha < 1$), and/or to have power law decay in its tail ($\beta < \infty$), provided that it does not show extreme versions of those behaviours simultaneously. Assumption 3 requires that the parameters ξ and θ of the Meijer kernels enable the estimator to properly reconstruct the head and tail behaviour of f. For instance, for $\alpha \simeq 0$ (f has a very fat head), it would not work to take ξ 'small' and $\theta \simeq \pi/2$ (lightest head for the kernel, see Fig. 2). The imposed conditions leave much freedom about the choice of ξ and θ, though, and are restrictive only in extreme cases. For instance, only for $\beta < 1/2$ (*extremely* fat tail for f) would the condition $\frac{\xi}{\cos^2\theta} > \frac{1}{4} - \frac{\beta}{2}$ not be trivially satisfied. Assumption 4 is standard.

First we study the pointwise bias and variance of estimator (23).

Theorem 1 *Under Assumptions 1–4, the Mellin–Meijer kernel density estimator (23) at any fixed $x \in (0,\infty)$ is such that, as $n \to \infty$,*

$$\mathbb{E}\big(\hat{f}(x)\big) = f(x) + \frac{1}{2}\eta^2 x f''(x) + o(\eta^2), \tag{24}$$

$$\mathbb{V}\mathrm{ar}\,\big(\hat{f}(x)\big) = \frac{f(x)}{2\sqrt{\pi}n\eta\sqrt{x}} + o((n\eta)^{-1}). \tag{25}$$

Proof From (19), we have directly

$$\begin{aligned} B(x) &= \mathbb{E}\left(\frac{1}{X_k}L_\eta^{(k)}\left(\frac{x}{X_k}\right)\right) - f(x), \\ V(x) &= \frac{1}{n}\left\{\mathbb{E}\left(\frac{1}{X_k^2}L_\eta^{(k)2}\left(\frac{x}{X_k}\right)\right) - \left(\mathbb{E}\left(\frac{1}{X_k}L_\eta^{(k)}\left(\frac{x}{X_k}\right)\right)\right)^2\right\}, \end{aligned} \tag{26}$$

where $B(x)$ and $V(x)$ denote the pointwise bias and variance of the estimator, respectively. It also holds

$$\mathcal{M}\left(\mathbb{E}\left(\frac{1}{X_k}L_\eta^{(k)}\left(\frac{\cdot}{X_k}\right)\right);z\right) = \mathbb{E}\left(\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1}\right), \text{ and}$$

$$\mathcal{M}\left(\mathbb{E}\left(\frac{1}{X_k^2}L_\eta^{(k)2}\left(\frac{\cdot}{X_k}\right)\right);z\right) = \mathbb{E}\left(\mathcal{M}(L_\eta^{(k)2};2\frac{z+1}{2})X_k^{2\frac{z+1}{2}-2}\right).$$

Hence, by the inverse Mellin transform (7), we can write

$$B(x) = \frac{1}{2\pi i}\int_{\Re(z)=c} x^{-z}\Big(\mathbb{E}\Big(\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1}\Big) - \mathcal{M}(f;z)\Big)\,\mathrm{d}z,$$
$$V(x) = \frac{1}{n}\Bigg\{\frac{1}{2\pi i}\int_{\Re(z)=2c-1} x^{-(z+1)}\mathbb{E}\Big(X_k^{z-1}\mathcal{M}(L_\eta^{(k)2};z+1)\Big)\,\mathrm{d}z - \left(\frac{1}{2\pi i}\int_{\Re(z)=c} x^{-z}\mathbb{E}\Big(\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1}\Big)\,\mathrm{d}z\right)^2\Bigg\} \tag{27}$$

for any $c \in [1-\min(\alpha, \xi/\cos^2\theta), 1+\min(\beta, \xi/\sin^2\theta)] \subseteq \mathcal{S}_{\hat{f}-f} = \mathcal{S}_{\hat{f}} \cap \mathcal{S}_f$. Indeed $\{z \in \mathbb{C} : 1-\alpha \le \Re(z) \le 1+\beta\} \subseteq S_f$ by Assumption 2, $\mathcal{S}_{\hat{f}} = \bigcap_{k=1}^n \mathcal{S}_{L_\eta^{(k)}}$ by (20), where (13) and (21) imply

$$\mathcal{S}_{L_\eta^{(k)}} = \left\{z \in \mathbb{C} : 1 - \frac{\xi(\eta^2+X_k)}{\eta^2\cos^2\theta} < \Re(z) < 1 + \frac{\xi(\eta^2+X_k)}{\eta^2\sin^2\theta}\right\} \supseteq \left\{z \in \mathbb{C} : 1 - \frac{\xi}{\cos^2\theta} \le \Re(z) \le 1 + \frac{\xi}{\sin^2\theta}\right\} \quad \text{for all } k.$$

Making use of the expansion (16) with (21) and (22), the dominant term in $B(x)$ as $n \to \infty$ is

$$B(x) \sim \frac{1}{2}\eta^2\frac{1}{2\pi i}\int_{\Re(z)=c} x^{-z}z(z-1)\mathcal{M}(f;z-1)\,\mathrm{d}z = \frac{1}{2}\eta^2\frac{1}{2\pi i}\int_{\Re(z)=c} x^{-z}\mathcal{M}(xf''(x);z)\,\mathrm{d}z = \frac{1}{2}\eta^2 xf''(x), \tag{28}$$

from (A.4) and (A.7) in Geenens (2020), provided $z-1 \in \mathcal{S}_f$, that is, $c-1 \in \mathcal{S}_f$.

Likewise, making use of expansion (17) with (21) and (22), one finds that, asymptotically,

$$\mathbb{E}\Big(X_k^{z-1}\mathcal{M}(L_\eta^{(k)2};z+1)\Big) \sim \frac{1}{2\sqrt{\pi}}\frac{1}{\eta}\mathcal{M}(f;z+1/2),$$

which, plugged in $V(x)$, yields the following dominant term:

$$V(x) \sim \frac{1}{2\sqrt{\pi}}\frac{1}{n\eta}\frac{1}{2\pi i}\int_{\Re(z)=2c-1} x^{-(z+1)}\mathcal{M}(f;z+1/2)\,\mathrm{d}z = \frac{1}{2\sqrt{\pi}}\frac{1}{n\eta}\frac{1}{\sqrt{x}}\frac{1}{2\pi i}\int_{\Re(z)=2c-1/2} x^{-z}\mathcal{M}(f;z)\,\mathrm{d}z = \frac{1}{2\sqrt{\pi}}\frac{f(x)}{n\eta\sqrt{x}}, \tag{29}$$

provided $2c - 1/2 \in \mathcal{S}_f$. So, (28) and (29) are the asymptotic bias and variance of $\hat{f}(x)$, provided that there exists $c \in \big[1-\min(\alpha, \xi/\cos^2\theta), 1+\min(\beta, \xi/\sin^2\theta)\big]$ such that $1-\alpha < c-1 < 1+\beta$ and $1-\alpha < 2c-1/2 < 1+\beta$. Assumptions 2 and 3 ensure there is such a c. □

Expressions (24) and (25) are identical to the 'away-from-boundary' bias and variance of Chen (2000)'s modified gamma kernel estimator. Now, define a 'boundary point' $x_0 \doteq \kappa\eta$ for some constant $\kappa > 0$, and assume that $f(x_0)$ and $f''(x_0)$ are bounded. Then (24) and (25) show that the bias of the estimator at x_0 is of order $O(\eta^3)$, while the variance at x_0 is of order $O((n\eta^{3/2})^{-1})$ as $n \to \infty$. Balancing squared bias and variance yields the rate of convergence $O(n^{-4/5})$, as for interior points.

Next the weighted mean integrated squared error (WMISE) of the estimator (23) is investigated.

Theorem 2 *Under Assumptions 1–4, the Mellin–Meijer kernel density estimator (23) is such that*

$$\mathbb{E}\left(\int_0^\infty x^{2c-1}\left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x\right) = O(\eta^4) + O((n\eta)^{-1}) \qquad \text{as } n \to \infty, \tag{30}$$

$$\text{if } c \in \left(\max\left(2-\alpha, \frac{3-2\alpha}{4}, 1-\frac{\xi}{\cos^2\theta}\right), \min\left(\frac{3+2\beta}{4}, 1+\frac{\xi}{\sin^2\theta}\right)\right). \tag{31}$$

Proof *(outline)* (A detailed proof is given in Geenens (2020, Section C)). Write the usual bias-variance decomposition of the MSE of $\hat{f}(x)$ to obtain

$$\mathbb{E}\left(\int_0^\infty x^{2c-1}\left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x\right) \doteq \mathrm{IB}_c^2 + \mathrm{IV}_c,$$

where

$$\mathrm{IB}_c^2 = \int_0^\infty x^{2c-1} B^2(x)\,\mathrm{d}x \quad \text{and } \mathrm{IV}_c = \int_0^\infty x^{2c-1} V(x)\,\mathrm{d}x$$

are the (weighted) integrated squared-bias and integrated variance, respectively. From (27), we see that $\mathcal{M}(B;z) = \mathbb{E}\left(\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1}\right) - \mathcal{M}(f;z)$. By (8) and the expansion (16) with (21) and (22), we have

$$\begin{aligned}
\mathrm{IB}_c^2 &= \frac{1}{2\pi}\int_{\Re(z)=c}\left|\mathbb{E}\left(\mathcal{M}(L_\eta^{(k)};z)X_k^{z-1}\right) - \mathcal{M}(f;z)\right|^2 \mathrm{d}z \\
&\sim \frac{1}{2\pi}\int_{\Re(z)=c}\frac{\eta^4}{4}\left|z(z-1)\mathcal{M}(f;z-1)\right|^2 \mathrm{d}z \\
&= \frac{\eta^4}{4}\frac{1}{2\pi}\int_{\Re(z)=c}\left|\mathcal{M}(xf''(x);z)\right|^2 \mathrm{d}z = \frac{\eta^4}{4}\int_0^\infty x^{2c+1} f''^2(x)\,\mathrm{d}x
\end{aligned}$$

if $z-1 \in \mathcal{S}_f$, which is the case here as $\Re(z) = c > 2-\alpha$ by (31). Also, by (26) and the expansion (17) with (21), we can directly write

$$\begin{aligned}
\mathrm{IV}_c &\leq \frac{1}{n}\int_0^\infty x^{2c-1}\mathbb{E}\left[\left(\frac{1}{X_k}L_\eta^{(k)}\left(\frac{x}{X_k}\right)\right)^2\right]\mathrm{d}x \\
&= \frac{1}{n}\mathbb{E}\left[\int_0^\infty (yX_k)^{2c-1}\frac{1}{X_k}L_\eta^{(k)2}(y)\,\mathrm{d}y\right] = \frac{1}{n}\mathbb{E}\left[\mathcal{M}(L_\eta^{(k)2};2c)X_k^{2c-2}\right] \\
&\sim \frac{1}{2\sqrt{\pi n\eta}}\mathbb{E}\left(X^{2c-3/2}\right).
\end{aligned}$$

This is $O((n\eta)^{-1})$ if $2c - 1/2 \in \mathcal{S}_f$, which is the case under condition (31). □

Assumptions 2 and 3 ensure that (31) is a non-empty interval. This establishes the convergence to 0 of the WMISE of the estimator (23), where the range of weighting x^{2c-1} assuring convergence essentially depends on the assumed negative and positive moments of f. The L_2-consistency of the estimator follows.

Corollary 1 *Suppose $\alpha > 3/2$ in Assumption 2 and $\xi/\cos^2\theta > 1/2$ in Assumption 3. Then, $c = 1/2$ belongs to (31), and estimator (23) is such that*

$$\mathbb{E}\left(\int_0^\infty \left(\hat{f}(x) - f(x)\right)^2\mathrm{d}x\right) = O(\eta^4) + O((n\eta)^{-1}) \to 0 \qquad \text{as } n \to \infty. \tag{32}$$

The fastest rate of convergence in (32) is achieved for $\eta \sim n^{-1/5}$, which is

$$\mathbb{E}\left(\int_0^\infty \left(\hat{f}(x) - f(x)\right)^2\mathrm{d}x\right) = O(n^{-4/5}),$$

the optimal rate of convergence for nonparametric density estimation under Assumption 1. Corollary 1 establishes this when $\alpha > 3/2$, that is, $\mathbb{E}(X^{-3/2}) < \infty$. This requires $f(x) = o(\sqrt{x})$ as $x \to 0$, and in particular, $f(0) = 0$. By contrast, Chen (2000) showed the MISE-consistency of the 'modified' gamma kernel estimator under the weaker condition $\int (xf''(x))^2\,\mathrm{d}x < \infty$. Indeed there exist distributions with $\int (xf''(x))^2\,\mathrm{d}x < \infty$ but with $\mathbb{E}(X^{-3/2}) = \infty$, like the exponential distribution to cite only one.

It is worth noting that the proof of Theorem 2 is exclusively based on the properties of $\mathcal{M}(f;z)$ *inside its strip of holomorphy*. As such, (30) has been proved under conditions on the existence of moments of X only, as those define $\mathcal{S}_f$. In particular, the proof uses the general identity

$$\mathcal{M}(xf''(x);z) = z(z-1)\mathcal{M}(f;z-1) \qquad \text{for } z-1 \in \mathcal{S}_f, \tag{33}$$

and the condition '$z - 1 \in \mathcal{S}_f$', i.e., $\Re(z) = c > 2 - \alpha$, requires $\alpha > 3/2$ if one wants $c = 1/2$. Now, this condition may be relaxed if we look at the *analytic continuation* of $\mathcal{M}(f;z)$ *outside* $\mathcal{S}_f$ (Paris and Kaminski 2001, p. 86). If $f(x) \sim \sum_{k=0}^\infty a_k x^{\alpha_k - 1}$ as $x \to 0$ with $\alpha < \alpha_0 < \alpha_1 < \alpha_2 < \ldots$, then for $1 - \alpha_{k^*+1} < \Re(z) < 1 - \alpha_{k^*}$, $\mathcal{M}(f;z)$ is actually the Mellin transform of $f_{k^*} = f(x) - \sum_{k=0}^{k^*} a_k x^{\alpha_k - 1} = o(f(x))$ as $x \to 0$.

Thus, if $-1/2 \in (1-\alpha_{k^*+1}, 1-\alpha_{k^*})$, then $z(z-1)\mathcal{M}(f;z-1)$ is the Mellin transform of $xf''_{k^*}(x)$. If this is well-behaved, the proof of Theorem 2 would carry over without the condition $c > 2-\alpha$. For instance, the exponential density $f(x) = e^{-x}\mathbb{1}_{x\geq 0}$ has Mellin transform $\mathcal{M}(f;z) = \Gamma(z)$ for $\Re(z) > 0$. As $e^{-x} = \sum_{k=0}^{\infty} \frac{(-x)^k}{k!}$ we have that, for $\Re(z) \in (-1,0)$, $\Gamma(z)$ is the Mellin transform of $f_1(x) = e^{-x} - 1$. By inversion, $xf_1''(x) = \frac{1}{2\pi i}\int_{\Re(z)=c} x^{-z} z(z-1)\Gamma(z-1)\,\mathrm{d}z$ for $c \in (0,1)$. As $f_1'' \equiv f''$, (33) holds true for $\Re(z) = 1/2$ even with $\alpha < 3/2$. Geenens (2020, Proposition C.1) establishes (32) under the condition $\int_0^\infty (xf''(x))^2\,\mathrm{d}x < \infty$, as in Chen (2000).

6 Smoothing parameter selection

Here we propose a data-driven way of selecting η in (23). Although cross-validation could be used, Mellin transform ideas provide an easy *plug-in* selector. From (24) and (25), the asymptotically dominant terms in the WMISE (30) are $\frac{1}{4}\eta^4 \int_0^\infty x^{2c+1} f''^2(x)\,\mathrm{d}x + \frac{1}{2\sqrt{\pi}}\frac{1}{n\eta}\int_0^\infty x^{2c-3/2} f(x)\,\mathrm{d}x$, for any c in (31). Balancing the two terms, the asymptotically optimal value of η is

$$\eta_{\mathrm{opt},c} = \left(\frac{(2\sqrt{\pi})^{-1}\int_0^\infty x^{2c-3/2} f(x)\,\mathrm{d}x}{\int_0^\infty x^{2c+1} f''^2(x)\,\mathrm{d}x} \right)^{1/5} n^{-1/5}. \tag{34}$$

Plug-in methods attempt to estimate the unknown factors in (34) for approximating $\eta_{\mathrm{opt},c}$. Estimating $\int_0^\infty x^{2c-3/2} f(x)\,\mathrm{d}x = \mathbb{E}(X^{2c-3/2})$ is straightforward, but estimating the denominator involving f'' is less obvious. Usually this step requires estimating higher derivatives of f, which in turn requires selecting pilot smoothing parameters and/or resorting to a 'reference distribution' (Sheather and Jones 1991). Here, combining (8) and (33) yields, for $c-1 \in \mathcal{S}_f$,

$$\int_0^\infty x^{2c+1} f''^2(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{\Re(z)=c} |z(z-1)\mathcal{M}(f;z-1)|^2\,\mathrm{d}z. \tag{35}$$

$\mathcal{M}(f;z-1)$ can be naturally estimated by $\mathcal{M}(\mathbb{P}_n;z-1) = n^{-1}\sum_{k=1}^n X_k^{z-2}$. Now, if $z = c + i\omega$, $|z(z-1)|^2 = (c(c-1)-\omega^2)^2 + (2c-1)^2\omega^2$, and

$$\begin{aligned}\left|\sum_{k=1}^n X_k^{z-2}\right|^2 &= \sum_{k=1}^n X_k^{z-2} \times \sum_{k'=1}^n X_{k'}^{z^*-2} \qquad (\cdot^* \text{ denotes complex conjugation})\\ &= \sum_k \sum_{k'} (X_k X_{k'})^{c-2}\left(\frac{X_k}{X_{k'}}\right)^{i\omega} = \sum_k \sum_{k'} (X_k X_{k'})^{c-2}\cos\left(\omega \log \frac{X_k}{X_{k'}}\right).\end{aligned}$$

This suggests to approximate (35) by

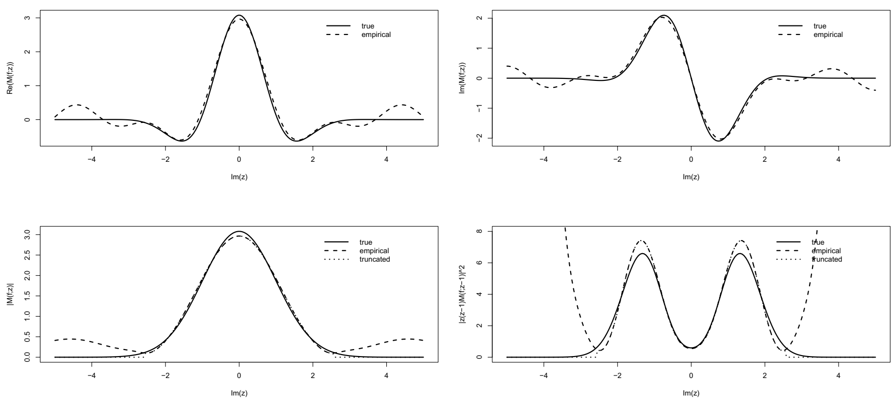

Fig. 3 $\Re(\mathcal{M}(f;z))$, $\Im(\mathcal{M}(f;z))$, $|\mathcal{M}(f;z)|$ and $|z(z-1)\mathcal{M}(f;z-1)|^2$ along the vertical line $\Re(z)=1/2$ for f the standard log-normal density (plain line). Dashed lines show the empirical approximations $\Re(\mathcal{M}(\mathbb{P}_n;z))$, $\Im(\mathcal{M}(\mathbb{P}_n;z))$, $|\mathcal{M}(\mathbb{P}_n;z)|$ and $|z(z-1)\mathcal{M}(\mathbb{P}_n;z-1)|^2$ from a typical sample of size $n=500$. Dotted lines show the truncated versions of the last 2 when $\mathcal{M}(\mathbb{P}_n;z)$ is set to 0 outside $(-T_0,T_0)$, where T_0 is the location of the first local minimum of $|\mathcal{M}(\mathbb{P}_n;z)|$ away from $\Im(z)=0$

$$\widehat{I}_c(T) \doteq \frac{1}{2\pi n^2}\sum_k\sum_{k'}(X_kX_{k'})^{c-2} \int_{-T}^{T}\left((c(c-1)-\omega^2)^2+(2c-1)^2\omega^2\right)\cos\left(\omega\log\frac{X_k}{X_{k'}}\right)d\omega \tag{36}$$

for some value T. Note that $\left((c(c-1)-\omega^2)^2+(2c-1)^2\omega^2\right)\cos\left(\omega\log\frac{X_k}{X_{k'}}\right)$ has closed-form antiderivative, which makes evaluating the integral very easy. However, (36) actually diverges for $T\to\infty$, as it would essentially reflect the integrated squared 'second derivative' of $\mathbb{P}_n=n^{-1}\sum_k\delta_{X_k}$. It is, therefore, paramount to select an appropriate value of T.

Note from (6) that, for a fixed $\Re(z)=c$, $|\mathcal{M}(f;z)|$ is symmetric around $\Im(z)=0$, i.e., the real axis, and always reaches its maximum at $\Im(z)=0$. In addition, $|\mathcal{M}(f;z)|$ typically tends to 0 quickly as one moves away from the real axis. In particular, $|\Gamma(z)|$ is known to be $O(e^{-\frac{1}{2}\pi|z|})$ as $\Im(z)\to\infty$ (Paris and Kaminski 2001, Lemma 3.2), implying a similar exponential behaviour for the modulus of (12). It turns out that $\mathcal{M}(\mathbb{P}_n;z)$ is remarkably accurate at reconstructing $\mathcal{M}(f;z)$ over a substantial set of values of $\Im(z)$ around the real axis, that is, where it matters. The approximation badly deteriorates as $\Im(z)$ grows, but there we know that $|\mathcal{M}(f;z)|\simeq 0$ anyway. This is illustrated in Fig. 3 for the case of the log-normal distribution.

Therefore, it is sensible to truncate the integral in (36) at some T for which $|\mathcal{M}(\mathbb{P}_n;z)|$ is already 'small' but before the empirical oscillations start. A reasonable choice is T_0, the location of the first local minimum in $|\mathcal{M}(\mathbb{P}_n;z)|$ away from $\Im(z)=0$; see Fig. 3. This suggests to take, finally,

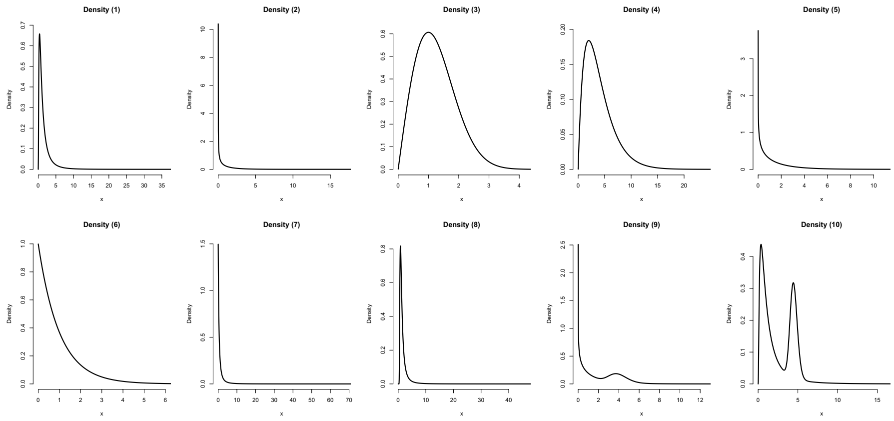

Fig. 4 Densities used in the simulation study

$$\eta = \left(\frac{(2\sqrt{\pi})^{-1} \frac{1}{n} \sum_{k=1}^{n} X_k^{2c-3/2}}{\widehat{I}_c(T_0)} \right)^{1/5} n^{-1/5} \tag{37}$$

for some c in (31) (which guarantees that all involved quantities are finite). The theory establishes that the exact value of c inside (31) has little importance. This is confirmed by simulation in the next section.

7 Simulation study

Inspired by Bouezmarni and Scaillet (2005), we consider the following 10 test densities, as shown in Fig. 4: [1] standard log-normal; [2] chi-squared with $k = 1$ degree of freedom; [3] Nakagami with $m = 1$ and $\Omega = 2$; [4] gamma with $\alpha = 2$ and $\beta = 1/2$; [5] gamma with $\alpha = 0.7$ and $\beta = 1/2$; [6] standard exponential; [7] generalised Pareto with $\sigma = 2/3$ and $\zeta = 2/3$; [8] inverse Weibull with $\mu = 1$ and $\eta = 2$; [9] mixture of gammas: $2/3 \times \Gamma(0.7, 1/2) + 1/3 \times \Gamma(20, 5)$; [10] mixture of log-normals: $2/3 \times \text{log-}\mathcal{N}(0, 1) + 1/3 \times \text{log-}\mathcal{N}(1.5, 0.1)$.

These 10 densities exhibit various behaviours at 0 (light head: [1], [8], [10]; fat head ($\mathbb{E}(X^{-3/2}) < \infty$): [3], [4]; *very* fat head ($\mathbb{E}(X^{-3/2}) = \infty$): [6], [7] (bounded), [2], [5], [9] (unbounded)) and in the tail (light tail: [1]–[6], [9], [10]; fat tail: [7], [8]). From each of these distributions, independent samples of size $n = 100$ and $n = 500$ were generated, with $M = 1,000$ Monte-Carlo replications for each sample size. On each of them, the density was estimated by the estimator (23), where the

Table 1 (approximated) MISE relative to the MISE of the modified gamma kernel estimator (second row), $n = 100$. Bold values show the minimum MISE for the corresponding density (non-significantly different values are underlined)

Density	c	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Gamma		1.0000	1.0000	1.0000	1.0000	1.0000	**1.0000**	**1.0000**	1.0000	**1.0000**	1.0000
		2.52	*15.04*	*27.89*	*1.59*	*5.37*	*9.39*	*0.02*	*3.08*	*6.70*	*8.04*
MM-1	$\frac{1}{2}$	1.1423	60.5327	0.8151	0.8668	16.3165	7.1636	7.1307	0.5167	10.5386	**0.7328**
MM-2		1.1365	60.5346	0.8712	0.8507	16.2919	7.0522	7.1313	0.5144	10.5138	0.7334
MM-3		1.1459	60.5320	0.8235	0.8812	16.3201	7.1824	7.1304	0.5178	10.5421	0.7335
MM-1	1	0.9929	6.0001	0.9142	0.8704	3.0499	1.7499	1.8415	**0.4838**	1.7679	1.0435
MM-2		1.0608	5.9855	1.0394	0.9920	2.9871	1.6653	1.8440	0.4851	1.7170	1.1043
MM-3		**0.9865**	6.0018	0.9144	0.8738	3.0635	1.7788	1.8405	0.4848	1.7790	1.0352
MM-1	$\frac{3}{2}$	1.0267	**0.4933**	0.8263	**0.8310**	0.9904	1.4479	1.1517	0.5052	1.0581	1.5039
MM-2		1.1962	0.5274	**0.8116**	0.8611	1.1069	1.5037	1.0181	0.5304	1.1575	1.6735
MM-3		1.0057	0.4975	0.8402	0.8466	**0.9898**	1.4613	1.1706	0.5034	1.0502	1.4776

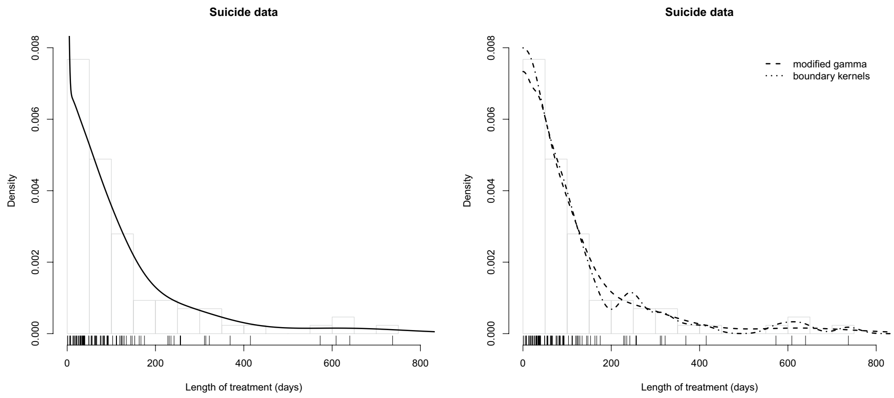

Fig. 5 ‘Suicide’ data set: Mellin–Meijer kernel density estimator with $\theta = 0$, $\xi = 1/2$ and $\eta = 4.74$ (left panel); modified gamma kernel estimator and ‘boundary-corrected’ conventional kernel estimator (right panel)

basic parameters (ξ, θ) of the Meijer kernel were set to $(1, \pi/4)$, (1/2, 0) and $(2, \pi/2)$ (‘MM-1’, ‘MM-2’ and ‘MM-3’ in Table 1). For each case, the smoothing parameter η was selected according to (37), where three values c were used: $c = 1/2$, $c = 1$ and $c = 3/2$. For comparison, the ‘modified’ gamma kernel estimator (‘Gamma’ in Table 1), with bandwidth chosen as in Hirukawa and Sakudo (2014), was included in the study as well.[7]

The densities were estimated on a fine grid of $N = 1,000$ points between $q_{0.9999}/1000$ and $q_{0.9999}$, where $q_{0.9999}$ is the quantile of level 0.9999 of the relevant density. The MISE of a given estimator $\hat{f}$ was then approximated by $\widehat{\mathrm{MISE}}(\hat{f}) = \frac{1}{M}\sum_{q=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}_{[q]}\left(\frac{i \times q_{0.9999}}{N}\right) - f\left(\frac{i \times q_{0.9999}}{N}\right)\right)^2$, where $M = 1,000$ is the number of Monte-Carlo replications and $\hat{f}_{[q]}$ is the estimate obtained from the qth replicated sample. The results are reported in Table 1 for $n = 100$. The results for $n = 500$ show a very similar pattern and are omitted. For ease of reading, all the values in Table 1 are relative to the MISE of the gamma kernel estimator, which is taken as benchmark owing to its reference role among the asymmetric kernel density estimators. Its effective MISE ($\times 10^4$) is reported in italics in the second row of the table.

Table 1 confirms the potential of Mellin–Meijer kernel estimation. There is a Mellin–Meijer kernel estimator which outperforms (Dens. [2], [3], [4], [8], [10]), sometimes by a large extent (half MISE for [2] and [8]), or is on par with (Dens. [1], [5], [7], [9]) the modified gamma kernel estimator. An exception is the exponential [6], for which the modified gamma kernel estimator does better. The modified gamma kernel estimator is actually so designed for staying bounded at $x = 0$ in finite samples (Chen 2000, p. 473).[8] So it is especially good at estimating densities f such

[7] This estimator was computed using the `dbckden` function in the R package `evmix`.

[8] That estimator remains consistent for unbounded densities, though; see Bouezmarni and Scaillet (2005).

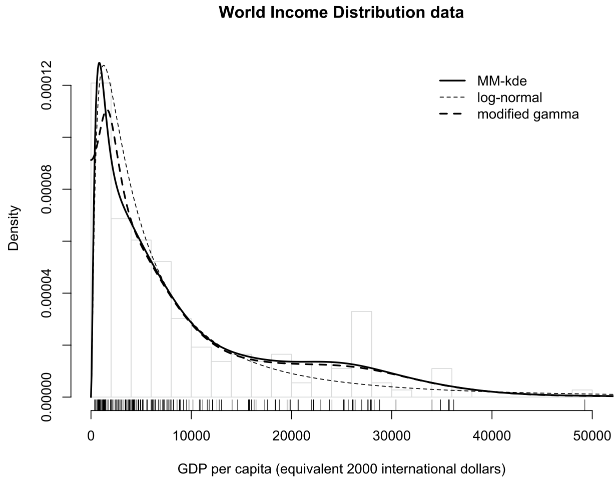

Fig. 6 'World Distribution of Income' data set: Mellin–Meijer kernel density estimator with $\theta = \pi/4$, $\xi = 1$ and $\eta = 28.54$(plain line), maximum likelihood log-normal parametric fit (dashed thin line), modified gamma kernel estimator (dashed thick line)

that $0 < f(0) < \infty$, such as the exponential. This may sometimes be counter-productive, though, see next Section. Nota that the Mellin–Meijer estimates are always *bona fide* densities, which is not the case for the modified gamma estimates.

The values of ξ and θ have little influence on the MISE of the estimator, as expected from the theory. The results evidence as well that the selector (37) is good at picking a right value of η. Of course, we get huge MISE's for Densities [2], [5], [6], [7] and [9] if (37) is computed with $c = 1/2$, in agreement with the theory: those are the densities such that $\mathbb{E}(X^{-3/2}) = \infty$, hence $c = 1/2$ does not belong to (31). For those densities, the selector is doing very good with $c = 3/2$. For the other densities, the value of c is less important. In practice, the choice of θ and ξ may be driven by a basic visual analysis of the sample, in particular a qualitative appreciation of the likely head and tail behaviour of the density f. As an example, if 'many' observations fall close to the boundary, suggesting a fat-headed density f, then it may be meaningful to take $\theta = 0$ and $\xi = 1/2$ (kernel with the fattest head, see Fig. 2). If no clear indication of that type may be drawn, then it seems reasonable to take $\theta = \pi/4$ and $\xi = 1$ ('balanced' kernel) as default choice. Likewise, for bandwidth selection, it seems wise to take $c = 3/2$, as it always belongs to (31) under the mild moment conditions $\alpha \geq 1/2$ and $\beta \geq 3/2$ (and $\xi \geq 1/2$), so we avoid the above problem and the returned bandwidth is always meaningful.

8 Real data analyses

In this section, the Mellin–Meijer kernel estimator (23) is applied on two real data sets. The first is the 'suicide' data set, which gives the lengths (in days) of $n = 86$ spells of psychiatric treatment undergone by patients in a study of suicide risks.

Among others, it was studied in Silverman (1986) and Chen (2000) in relation to boundary issues. Visual inspection (raw data at the bottom of the graph, histogram) reveals that the density should be positive, if not unbounded, at $x = 0$. Hence, we take $\theta = 0$ and $\xi = 1/2$ for the Meijer kernels. The smoothing parameter returned by (37) with $c = 3/2$ is $\eta = 4.74$. Figure 5 (left panel) shows the estimated density. The estimate shows a spike at the 0 boundary: there are 3 observations exactly equal to 1 in the data set, and at this scale, this is pretty much 'on the boundary'. Hence, the estimator attempts to put a positive probability mass atom there, producing a meaningful spike. Away from the boundary, the estimate decays readily and smoothly.

For comparison, the modified gamma kernel estimator, with bandwidth chosen by reference rule (Hirukawa and Sakudo 2014), as well as the 'boundary-corrected' conventional estimator (Jones and Foster 1996) with 'SJ-bandwidth' (Sheather and Jones 1991), are shown in the right panel. While the modified gamma kernel estimator behaves very similarly to the Mellin–Meijer kernel estimator in the tail, its behaviour at the boundary is not satisfactory as it seems to underestimate f there, compared to the other estimates and the histogram. This is typical of the modified gamma kernel estimator, as discussed in Zhang (2010) and Malec and Schienle (2014). The boundary-corrected kernel estimate may do better at $x = 0$, but exhibits numerous 'spurious bumps' in the right tail.

In the second example, we estimate the World Distribution of Income from data giving the GDP per capita (in constant 2000 international dollars) of $n = 182$ countries in 2003 obtained from the World Bank Database. This is important as various measures of poverty rates, income inequality or welfare at the scale of the world are based on this distribution (a). Raw data are shown in Fig. 6 with an histogram and the estimated density by the Mellin–Meijer kernel estimator. We set $\theta = \pi/4$ and $\xi = 1$, and the value returned by (37) with $c = 3/2$ was $\eta = 28.54$.

A log-normal parametric density, fitted by Maximum Likelihood ($\hat{\mu} = 8.58$, $\hat{\sigma} = 1.20$), is also shown in Fig. 6. (a) strongly advocated in favour of the log-normal model for these data. However, the (mostly) unconstrained Mellin–Meijer estimate reveals that the peak close to 0 is actually narrower than the 'log-normal peak', whereas there are much more countries with GDP in the range 15,000 - 40,000 than what the log-normal distribution prescribes. In other words, analysis through the log-normal model is likely to underestimate poverty and income inequality at the world level. Chen (2000)'s modified gamma kernel estimator is also shown. Here that estimator largely overestimates f at $x = 0$: it is so designed to take a finite, non-zero value at $x = 0$, whereas here it seems clear from the raw data that $f(0) = 0$. The peak at around $x = 1000$ is also largely oversmoothed.

9 Concluding remarks

Within his seminal works on compositional data (data on the simplex), Aitchison (2001, Section 4.1) noted: "*For every sample space there are basic group operations which, when recognised, dominate clear thinking about data analysis.*" He continued: "*In $\mathbb{R}^d$,the two operations, translation and scalar multiplication are so familiar that their fundamental role is often overlooked*", implying that, when not in

$\mathbb{R}^d$, there is no reason to blindly stick to those operations. The methodology developed in this paper perfectly aligns with this stance. It has apparently been largely overlooked in earlier literature that the 'boundary issues' of the conventional kernel density estimator find their very origin in that $\mathbb{R}^+$ equipped with the addition $+$ is not a group. Noting that the natural group operation on $\mathbb{R}^+$ is the multiplication $\times$, we have investigated a new kind of kernel estimation for $\mathbb{R}^+$-supported probability densities which achieves smoothing through 'multiplicative dilution'. The construction gives rise to an asymmetric kernel density estimator, although of different nature to the other estimators known under that name, such as the gamma kernel estimator (Chen 2000). Unlike those, our estimator is based on a valid smoothing operation on $\mathbb{R}^+$, the Mellin convolution, which avoids any inconsistency in the definition and the behaviour of the estimator. We have defined a huge class of distributions supported on $\mathbb{R}^+$ which, defined in terms of Meijer G-functions, perfectly fit within the 'Mellin' framework. Using those 'Meijer densities' as kernels produces an integrated theory with general features no more specific to a particular choice of kernel.

Interestingly, Aitchison (2001, Section 6.3) already introduced the Mellin transform as the suitable analytical tool for simplicial distributions. More generally, the Mellin transform of a density f ought to be a fundamental function in statistics and probability, as it explicitly returns *all* the moments (real, complex, integral and fractional) of f. It is, therefore, rather surprising that it has stayed this inconspicuous in the statistical literature so far. Historically, one can find statistical applications of the Mellin transform only intermittently over decades (Dolan 1964; Epstein 1948; Kabe 1958; Lomnicki 1967; Subrahmaniam 1970). Only recently has the Mellin transform made a (discreet) resurgence in the statistical literature, e.g., in Balakrishnan and Stepanov (2014); Belomestny and Schoenmakers (2016). Those papers testify of the appropriateness of the Mellin transform and Mellin convolution in any multiplicative framework, such as problems of multiplicative censoring.

Acknowledgements The author would like to thank the Editor, the Associate Editor and three anonymous referees for their helpful suggestions which greatly helped to improve the quality of the paper.

References

Aitchison, J. (2001). Simplicial Inference. In M. A. G. Viana & D. S. P. Richards (Eds.), *Algebraic Methods in Statistics and Probability, Contemporary Mathematics Series 287*, American Mathematical Society, Providence, Rhodes Island, pp. 1–22.

Balakrishnan, N., Stepanov, A. (2014). On the use of bivariate Mellin transform in bivariate random scaling and some applications. *Methodology and Computing in Applied Probability*, *16*, 235–244.

Beals, R., Szmigielski, J. (2013). Meijer G-functions: A gentle introduction. *Notices of the American Mathematical Society*, *60*, 866–872.

Belomestny, D., Schoenmakers, J. (2016). Statistical inference for time-changed Lévy processes via Mellin transform approach. *Stochastic Processes and their Applications*, *126*, 2092–2122.

Bouezmarni, T., Scaillet, O. (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, *21*, 390–412.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, *52*, 471–480.

Cheng, M. Y., Fan, J., Marron, J. S. (1997). On automatic boundary corrections. *Annals of Statistics*, *25*, 1691–1708.

Comte, F., Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *Journal of Statistical Planning and Inference*, *142*, 1698–1715.
Cox, C. (2008). The generalized F distribution: An umbrella for parametric survival analysis. *Statistics in Medicine*, *27*, 4301–4312.
Dai, J., Sperlich, S. (2010). Simple and effective boundary correction for kernel densities and regression with an application to the world income and Engel curve estimation. *Computational Statistics and Data Analysis*, *54*, 2487–2497.
Dolan, B. A. (1964). The Mellin transform for moment-generation and for the probability density of products and quotients of random variables. *Proceedings of the IEEE*, *52*, 1745–1746.
Epstein, B. (1948). Some applications of the Mellin transform in statistics. *Annals of Mathematical Statistics*, *19*, 370–379.
Geenens, G. (2020). Supplementary Material to 'Mellin-Meijer kernel density estimation on $\mathbb{R}^+$'. *Annals of the Institute of Statistical Mathematics*.
Geenens, G., Wang, C. (2018). Local-likelihood transformation kernel density estimation for positive random variables. *Journal of Computational and Graphical Statistics*, *27*, 822–835.
Hall, P., Hu, T., Marron, J. S. (1995). Improved variable window kernel estimates of probability densities. *Annals of Statistics*, *23*, 1–10.
Härdle, W., Müller, M., Sperlich, S., Werwatz, A. (2004). *Nonparametric and Semiparametric Models*: *An Introduction*. Berlin: Springer.
Hirukawa, M., Sakudo, M. (2014). Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics and Data Analysis*, *75*, 112–123.
Hirukawa, M., Sakudo, M. (2015). Family of generalised gamma kernels: A generator of asymmetric kernels for nonnegative data. *Journal of Nonparametric Statistics*, *27*, 41–63.
Igarashi, G. (2016). Weighted log-normal kernel density estimation. *Communications in Statistics - Theory and Methods*, *45*, 6670–6687.
Igarashi, G., Kakizawa, Y. (2014). Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum-Sauders kernel estimators. *Statistics & Probability Letters*, *84*, 235–246.
Igarashi, G., Kakizawa, Y. (2018). Generalised gamma kernel density estimation for nonnegative data and its bias reduction. *Journal of Nonparametric Statistics*, *30*, 598–639.
Jeon, Y., & Kim, J. H. T. (2013). A gamma kernel density estimation for insurance loss data. *Insurance*: *Mathematics and Economics*, *53*, 569–579.
Jin, X., Kawczak, J. (2003). Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, *4*, 103–124.
Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, *3*, 135–146.
Jones, M. C., Foster, P. J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, *6*, 1005–1013.
Jones, M. C., Henderson, D. A. (2007). Kernel-type density estimation on the unit interval. *Biometrika*, *94*, 977–984.
Kabe, D. G. (1958). Some applications of Meijer-G Functions to distribution problems in statistics. *Biometrika*, *45*, 578–580.
Kakizawa, Y. (2018). Nonparametric density estimation for nonnegative data, using symmetrical-based inverse and reciprocal inverse Gaussian kernels through dual transformation. *Journal of Statistical Planning and Inference*, *193*, 117–135.
Karunamuni, R. J., Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *Canadian Journal of Statististics*, *33*, 497–509.
Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *29*, 513–524.
Malec, P., Schienle, M. (2014). Nonparametric kernel density estimation near the boundary. *Computational Statistics and Data Analysis*, *72*, 57–76.
Marchant, C., Bertin, K., Leiva, V., Saulo, H. (2013). Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data. *Computational Statistics and Data Analysis*, *63*, 1–15.
Meijer, C. S. (1936). Uber Whittakersche bzw. *Besselsche Funktionen und deren Produkte, Nieuw Archief voor Wiskunde*, *18*, 10–39.
Mnatsakanov, R., Ruymgaart, F. H. (2012). Moment density estimation for positive random variables. *Statistics*, *46*, 215–230.
Mnatsakanov, R., Sarkisian, K. (2012). Varying kernel density estimation on $\mathbb{R}_+$. *Statistics & Probability Letters*, *82*, 1337–1345.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, *78*, 521–530.
Paris, R. B., Kaminski, D. (2001). *Asymptotics and Mellin-Barnes Integrals*. Cambridge: Cambridge University Press.
Pinkovskiy, M. and Sala-i-Martin, X. (2009), Parametric estimations of the world distribution of income, Working Paper 15433, *National Bureau of Economic Research*, Cambridge.
Prudnikov, A.P., Marichev, O.I. & Brychkov, Yu. A. (1990). *Integrals and Series, Vol. 3: More Special Functions*, Gordon and Breach, Newark.
Scaillet, O. (2004). Density estimation using inverse Gaussian and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics*, *16*, 217–226.
Schuster, E. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods*, *14*, 1123–1136.
Sheather, S. J., Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *53*, 683–690.
Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall/CRC.
Subrahmaniam, K. (1970). On some applications of Mellin transform to statistics: dependent random variables. *SIAM Journal on Applied Mathematics*, *19*, 658–662.
Terrell, G. R., Scott, D. W. (1992). Variable kernel density estimation. *Annals of Statistics*, *20*, 1236–1265.
Wand, M. P., Jones, M. C. (1995). *Kernel Smoothing*. New York: Chapman and Hall/CRC.
Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary. *Statistics & Probability Letters*, *80*, 548–557.
Zhang, S., Karunamuni, R. J., Jones, M. C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, *94*, 1231–1241.