INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data

T.BOUEZMARNI and O.SCAILLET

http://www.stat.ucl.ac.be

Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data

T. Bouezmarni * and O. Scaillet †

February 11, 2003

Abstract

We consider asymmetric kernel density estimators and smoothed histograms when the unknown probability density function f is defined on $[0, +\infty)$. Uniform weak consistency on each compact set in $[0, +\infty)$ is proved for these estimators when f is continuous on its support. Weak convergence in L_1 is also established. Finally we prove that the asymmetric kernel density estimator and the smoothed histogram converge in probability to infinity at $x = 0$ when the density is unbounded at $x = 0$. Monte Carlo results and an empirical study of the shape of a highly skewed income distribution based on a large micro-data set are finally provided.

Key words and phrases : Asymmetric kernel, smoothed histogram, density estimation, weak convergence, L_1 consistency, unbounded density, boundary bias, income distribution, inequality measurement. *JEL Classification* : C13,C14. *MSC 2000* : 62G07, 62G08

We would like to thank M. P. Feser and J. Litchfield for kindly providing the Brazilian data. We are grateful to I. Gijbels, J. M. Rolin and I. Van Keilegom for their stimulating remarks, as well as to participants at the workshop on statistical modelling (UCL 2002), LAMES (Sao Paulo 2002), L1 Norm conference (Neuchatel 2002), Geneva econometrics seminar and KUL econometrics seminar. Part of this research was done when the second author was visiting THEMA and IRES.

The research of T. Bouezmarni was supported by the contract "Projet d'Actions de Recherche Concertées" n^0 98/03 − 217, and by the IAP research network n^0 P5/24 of the Belgian state.

*Institute of Statistics, Université Catholique de Louvain.

†HEC Genève and FAME, Université de Genève.

1 Introduction

The most popular nonparametric estimator of an unknown probability density function f is the standard kernel estimator. Its consistency is well documented when the support of the underlying density is unbounded. In the case of a bounded support we know that there exists a boundary bias (see for example the estimation of Figure 3 in Section 5). This problem is due to the use of a fixed kernel which assigns weight outside the support when smoothing is carried out near the boundary. It is further known that the expected value of the standard kernel density estimator at $x = 0$ converges to the half value of the underlying density when f is twice continuously differentiable on its support $[0, +\infty)$. To solve this problem many remedies have already been suggested (see e.g. Rice (1984), Schuster (1985), M*ü*ller (1991), Marron and Ruppert (1994), Jones (1993), Jones and Foster (1996)). They include use of particular kernels or bandwidths.

Recently, Chen (2000) has proposed a gamma kernel estimator while Scaillet (2001) has introduced inverse Gaussian (IG) and reciprocal inverse Gaussian (RIG) estimators for densities defined on $[0, +\infty)$. These estimators are based on asymmetric kernels which have flexible form and location on the non-negative real line. The kernel shapes are allowed to vary according to the position of the data points, thus changing the degree of smoothing in a natural way, and their support matches the support of the probability density function to be estimated. The gamma, IG and RIG kernel density estimators are simple to implement, free of boundary bias, always non-negative, and achieve the optimal rate of convergence for the mean integrated squared error (MISE) within the class of non-negative kernel density estimators. Furthermore their variance reduces as the position where the smoothing is made moves away from the boundary. This is an advantage in estimating densities that have sparse areas because more data points can be pooled to smooth in areas with fewer observations. As pointed out by Cowell (1999) "Empirical income distributions typically have long tails with sparse data". Hence it is expected that such estimators should perform well in practice on income data (this will be confirmed by our empirical results in Section 7).

In this paper we first analyse convergence of these asymmetric kernel density estimators for the class of density functions with support $[0, +\infty)$. Then we examine convergence of the smoothed histograms proposed by Gawronski and Stadtm*ü*ller (1980, 1981)) which are also free of boundary bias and achieve the same rate of convergence.

The paper is organised as follows. In Section 2, we outline the framework and present both estimators, namely the asymmetric kernel density estimator and the smoothed histogram. Particular examples are developped. Uniform weak consistency and uniform strong consistency on each compact set in $[0, +\infty)$ are proved for both estimators in Section 3. The L_1 convergence of the two estimators are established in Section 4. In Section 5, the density function f is assumed to be *unbounded* at $x = 0$, and we analyse the weak convergence of the two estimators to infinity at $x = 0$. To our best knowledge it is the first attempt of providing a consistent estimator for such a density (see however Marron and Ruppert (1994) and Bouezmarni and Rolin (2002) but for densities defined on $[0, 1]$). Section 6 provides Monte Carlo results concerning the finite sample properties of the estimators for various distributions and parameter values. An empirical illustration on a large micro-data set is provided in Section 7. We examine the shape of the Brazilian income distribution which is notoriously known to be highly skewed with an accumulation of observed points near the zero boundary. Besides a data-driven procedure based on the L_1 distance (Hall and Wand (1988)) is discussed to select the bandwidth in practical situations. Section 8 contains some concluding remarks. An appendix gathers the proofs.

2 Asymmetric kernel density estimators and smoothed histograms

Let $X_1, \cdots, X_n$ be a random sample from a probability distribution F with an unknown density function f. The most popular nonparametric estimator for the unknown probability density function f is the standard kernel estimator:

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

where the kernel $\mathcal{K}$ is a symmetric density function and h is a smoothing parameter, called the bandwidth. When the density is defined on $[0, +\infty)$ the boundary bias of the standard kernel estimator is due to weight allocation by the fixed symmetric kernel outside the support when the estimation of density is made near the boundary. To overcome the problem a simple idea is to use a flexible kernel, which never assigns weight outside the support of the density function. This is the idea behind the first estimator considered in this paper, namely the asymmetric kernel density estimator:

$$\hat{f}_b(x) = \frac{1}{n} \sum_{i=1}^{n} K(x,b)(X_i), \tag{1}$$

where b is the bandwidth and the asymmetric kernel K is either a gamma density K_G with parameters $(x/b+1,b)$, an inverse Gaussian density K_{IG} with parameters $(x,1/b)$ or a reciprocal inverse Gaussian density K_{RIG} with parameters $(1/(x-b),1/b)$. These densities correspond to:

$$K_G\left(\frac{x}{b}+1,b\right)(t) = \frac{t^{x/b}e^{-t/b}}{b^{x/b+1}\Gamma(x/b+1)}, \tag{2}$$

$$K_{IG}\left(x,\frac{1}{b}\right)(t) = \frac{1}{\sqrt{2\pi b t^3}} \exp\left(-\frac{1}{2bx}\left(\frac{t}{x}-2+\frac{x}{t}\right)\right), \tag{3}$$

$$K_{RIG}\left(\frac{1}{x-b},\frac{1}{b}\right)(t) = \frac{1}{\sqrt{2\pi b t}} \exp\left(-\frac{x-b}{2b}\left(\frac{t}{x-b}-2+\frac{x-b}{t}\right)\right). \tag{4}$$

Note that these asymmetric kernels do not take the form $\kappa(x-t,b)$ where κ is an asymmetric function (instead of a symmetric one), and thus do not belong to the class studied by Abadir and Lawford (2002).

The estimator $\hat{f}_b$ based on the gamma kernel was proposed by Chen (2000) while the IG and RIG kernel density estimators were proposed by Scaillet (2001). Figure 1 plots the shapes of the Gamma, IG and RIG kernels for some selected values of x and $b = 0.2$. It can be noticed that $K_G(x/b+1,b)(t)$ for $x = 0$ is decreasing for $t > 0$, and becomes unbounded at $t = 0$ when b shrinks to zero. This feature of the gamma kernel will be instrumental for convergence when the density is unbounded at $x = 0$ (see Section 5). Let us also remark that the asymmetric kernel density estimator is a particular case of the generalized kernel density estimator (Foldes and Revesz (1974), Walter and Blum (1979)).

The second estimator considered in this paper is another particular case of the generalized kernel density estimator, and is inspired by a well known approximation theorem for continuous distribution (Feller (1971) p. 219). It was developped by Gawronski and Stadtm$\ddot{u}$ller (1980, 1981), and is called a smoothed histogram. It is defined by:

$$\hat{f}_k(x) = k \sum_{i=0}^{+\infty} \omega_{i,k}\, p_{ki}(x), \tag{5}$$

where the weights $\omega_{i,k}$ are random. These weights are given by:

$$\omega_{i,k} = F_n\left(\frac{i+1}{k}\right) - F_n\left(\frac{i}{k}\right), \tag{6}$$

where F_n is the empirical distribution function, the integer k is the smoothing parameter and $p_{ki}(.)$ is based on use of either a family of lattice distribution or integrals of continuous distributions and satisfies the Gawronski and Stadtm*ü*ller's conditions. (see Gawronski and Stadtm*ü*ller (1980, 1981) for further details).

Example 1. $p_{ki}(x)$ corresponds to a Poisson distribution function with parameter kx, namely:

$$p_{ki}(x) = e^{-kx}\frac{(kx)^i}{i!} \qquad i = 0, 1, \ldots.$$

For this choice the smoothed histogram $\hat{f}_k$ is in fact a mixture of gamma densities:

$$\hat{f}_k(x) = \sum_{i=0}^{+\infty} \omega_{i,k}\, \Gamma(x, i+1, k),$$

where Γ is the gamma density function:

$$\Gamma(x, i+1, k) = \frac{x^i k^{i+1}}{i!} e^{-kx}.$$

Figure 2 shows the differents shapes of the gamma densities $\Gamma(x, i+1, k) = kp_{ki}(x)$ in the mixture, for $k = 3$, and $i = 0, 1, 2, 3$. Let us point out the decreasing shape of the gamma density for $i = 0$ and its unboundedness at $x = 0$ when k goes to infinity. We will come back to this characteristic in Section 5.

Example 2. $p_{ki}(x) = \int_{i/k}^{(i+1)/k} K(x, 1/k)(t)dt$ where $K(x, 1/k)(t)$ is either the IG kernel or the RIG kernel defined in (3) and (4) with a bandwidth equal to $1/k$.

Gawronski and Stadtm*ü*ller (1980, 1981) found that smoothed histograms are free of boundary bias and that their rate of convergence for the mean integrated squared error is $O(n^{-4/5})$ for f in $C^2([0, +\infty))$. Using the Hadamard product technique, Stadtm*ü*ller (1983) proved the uniform consistency in probability for the estimators when the density function f is continuous and bounded on $[0, +\infty)$. We provide a simpler proof of this type of convergence in this paper.

Remark 1 *For a large value of the parameter k in the smoothed histogram $\hat{f}_k$, we get:*

$$\begin{aligned} \hat{f}_k(x) &= k \sum_{j=0}^{+\infty} \omega_{j,k} p_{kj}(x) \\ &= \frac{k}{n} \sum_{i=1}^{n} \left(\sum_{j \geq 0} \mathbb{1}_{\left\{X_i \in \left(\frac{j}{k}, \frac{j+1}{k}\right]\right\}} p_{kj}(x) \right) \\ &\simeq \frac{k}{n} \sum_{i=1}^{n} p_{k,kX_i}(x). \end{aligned}$$

In Example 1, under $p_{ki}(x) = e^{-kx} \frac{(kx)^i}{i!}$, we get

$$\hat{f}_k(x) \simeq \frac{1}{n} \sum_{i=1}^{n} K(X_i, 1/k)(x).$$

The last formula corresponds to the expression of the asymmetric kernel density estimator based on the gamma kernel K_G with bandwidth parameter $1/k$ and permutation of the role of x and X_i.

In Example 2 under $p_{ki}(x) = \int_{i/k}^{(i+1)/k} K(x, 1/k)(t)dt$ with $K = K_{IG}$ or $K = K_{RIG}$, we get

$$\hat{f}_k(x) \simeq \frac{1}{n} \sum_{i=1}^{n} K(x, 1/k)(X_i).$$

This is the expression of the asymmetric kernel density estimator based on the IG kernel or RIG kernel with bandwidth parameter $1/k$.

3 Uniform weak convergence of asymmetric kernel density estimators and smoothed histograms

In this section, we show that both estimators have the same asymptotic behaviour. In fact, we prove the uniform weak convergence on each compact set I in $[0,+\infty)$ of the asymmetric kernel estimator $\hat{f}_b$ and the smoothed histogram $\hat{f}_k$ under some conditions on the smoothing parameter. These conditions are the following:

Condition 1 (Asymmetric kernel density estimator)

$$\lim_{n\longrightarrow\infty} b = 0 \quad \textit{and} \quad \lim_{n\longrightarrow\infty} n\, b^{2a} = +\infty \ (a>0).$$

Condition 2 (Smoothed histogram)

$$\lim_{n\longrightarrow\infty} k = +\infty \quad \textit{and} \quad \lim_{n\longrightarrow\infty} n\, k^{-2} = +\infty.$$

The next proposition states the uniform convergence of the bias of the asymmetric kernel density estimator on each compact set in $[0,+\infty)$ as the bandwith converges to zero.

Proposition 3.1 *Let f be a continuous and bounded probability density function on $[0,+\infty)$, $\hat{f}_b$ the asymmetric kernel density estimator and I a compact set in $[0,+\infty)$.*
Then

$$\sup_{x\in I} \mid I\!\!E\left(\hat{f}_b(x)\right) - f(x) \mid \longrightarrow 0 \quad as \quad b \longrightarrow 0.$$

The main result for the asymmetric kernel density estimator in this section is its uniform weak consistency under Condition 1.

Theorem 3.1 (Uniform Weak Consistency of $\hat{f}_b$)
Let f be a continuous and bounded probability density function on $[0,+\infty)$, $\hat{f}_b$ the asymmetric kernel density estimator and I a compact set in $[0,+\infty)$. Then

$$\sup_{x\in I} \mid \hat{f}_b(x) - f(x) \mid \stackrel{P}{\longrightarrow} 0 \quad as \quad n \longrightarrow \infty$$

under Condition 1 with $a=1$ for the gamma kernel and $a=5/2$ for the inverse Gaussian and reciprocal inverse Gaussian kernels.

Remark 2 *We have the same result as the one of Theorem 3.1 for a density estimator based on a general kernel $k(x,b)(t)$ under the following conditions:*

1. $I\!\!E(\xi_x) = x + O(b)$ *and* $\sup_{x\in I} Var(\xi_x) = O(b)$, *for any compact set I, where ξ_x is a random variable with density $k(x,b)$.*

2. $\displaystyle\int_0^{+\infty} |dk(x,b)(t)| = O(b^{-a})$ *for a strictly positive number a.*

3. $\displaystyle\lim_{n\longrightarrow\infty} b = 0$ *and* $\displaystyle\lim_{n\longrightarrow\infty} n\, b^{2a} = +\infty.$

Under a stronger set of conditions on the bandwidth b, the next corollary states the uniform strong consistency on each compact set of the asymmetric kernel density estimator.

Corollary 3.1 (Uniform Strong Consistency of $\hat{f}_b$)

Let f be a continuous and bounded probability density function on $[0,+\infty)$, $\hat{f}_b$ the asymmetric kernel density estimator, and I a compact set in $[0,+\infty)$. Under the two conditions:

$$b \longrightarrow 0 \qquad \textit{and} \qquad \frac{nb^{2a}}{\ln n} \longrightarrow +\infty \quad \textit{as} \quad n \longrightarrow +\infty.$$

we have

$$\sup_{x\in I} \mid \hat{f}_b(x) - f(x) \mid \xrightarrow{a.s} 0 \quad \textit{as} \quad n \longrightarrow \infty$$

where $a = 1$ for the gamma kernel and $a = 5/2$ for the inverse Gaussian and reciprocal inverse Gaussian kernels.

Similar results hold for smoothed histograms. We begin with the uniform convergence of the bias of the smoothed histogram on each compact set in $[0,+\infty)$ as the smoothing parameter k goes to infinity.

Proposition 3.2 *Let f be a continuous and bounded probability density function on $[0,+\infty)$, $\hat{f}_k$ the smoothed histogram and I a compact set in $[0,+\infty)$. Then*

$$\sup_{x\in I} \mid \mathbb{E}(\hat{f}_k(x)) - f(x) \mid \longrightarrow 0 \quad \textit{as} \quad k \longrightarrow +\infty. \tag{7}$$

Now we state the main result for the smoothed histogram.

Theorem 3.2 (Uniform Weak Consistency of $\hat{f}_k$)

Let f be a continuous and bounded probability density function on $[0,+\infty)$, $\hat{f}_k$ the smoothed histogram and I a compact set in $[0,+\infty)$.
Then

$$\sup_{t\in I} \left| \hat{f}_k(x) - f(x) \right| \xrightarrow{P} 0 \quad \textit{as} \quad n \longrightarrow +\infty,$$

under Condition 2.

Uniform strong consistency on each compact set of the smoothed histogram is given in the next corollary under a stronger set of conditions on the smoothing parameter k.

Corollary 3.2 (Uniform Strong Consistency of $\hat{f}_k$)
Let f be a continuous and bounded probability density function on $[0, +\infty)$, $\hat{f}_k$ the smoothed histogram and I a compact set in $[0, +\infty)$.
Under the two conditions:

$$k \longrightarrow +\infty \qquad \textit{and} \qquad \frac{nk^{-2}}{\ln n} \longrightarrow +\infty \quad \textit{as} \quad n \longrightarrow +\infty.$$

we have

$$\sup_{x \in I} \mid \hat{f}_k(x) - f(x) \mid \xrightarrow{a.s} 0 \quad \textit{as} \quad n \longrightarrow \infty.$$

4 Weak Convergence in L_1 of asymmetric kernel density estimators and smoothed histograms

The excellent monograph of Devroye and Gyorfi (1985) contains lots of results for the standard kernel density estimator in the L_1 case (see Fenton and Gallant (1996) for SNP density estimators). In particular many equivalences (types of convergence, conditions on bandwidth, ...) are shown to hold. They advocate the L_1 approach for three main reasons. First it is a natural metric on the space of density functions. Second it is proportional to the total variation metric. Eventually it is invariant under monotone transformations. Note that Hall and Wand (1988) (see also the proposal of Devroye and Lugosi (1996) and the survey in Devroye (1997)) have proposed an algorithm wich permits minimization of the L_1 distance for different estimators: standard kernel density estimator, histogram, ... We investigate application of this type of algorithm later in the paper.

Hereafter we prove the consistency in L_1 of the asymmetric kernel density estimator and the smoothed histogram. In a first step, we establish the convergence in L_1 of the bias of both estimators in Propositions 4.1 and 4.2. In a second step, we state the weak consistency in L_1 in the two theorems of this section.

Proposition 4.1 *Let f be a probability density function on $[0,+\infty)$ and $\hat{f}_b$ the asymmetric kernel density estimator.*
Then
$$\int_0^{\infty} \mid \mathbb{E}(\hat{f}_b(x)) - f(x) \mid dx \longrightarrow 0 \quad as \quad b \longrightarrow 0.$$

Proposition 4.2 *Let f be a probability density function on $[0,+\infty)$ and $\hat{f}_k$ the smoothed histogram.*
Then
$$\int_0^{\infty} \mid \mathbb{E}(\hat{f}_k(x)) - f(x) \mid dx \longrightarrow 0 \quad as \quad k \longrightarrow \infty.$$

Theorem 4.1 (Weak Consistency in L_1 of $\hat{f}_b$)
Let f be a probability density function on $[0,+\infty)$ and $\hat{f}_b$ the asymmetric kernel density estimator.
Then
$$\int_0^{\infty} \mid \hat{f}_b(x) - f(x) \mid dx \xrightarrow{P} 0 \quad as \quad n \longrightarrow \infty.$$
under Condition 1 with $a = 1$ for the gamma kernel and $a = 5/2$ for the inverse Gaussian and reciprocal inverse Gaussian kernels.

Theorem 4.2 (Weak Consistency in L_1 of $\hat{f}_k$)
Let f be a probability density function on $[0,+\infty)$ and $\hat{f}_k$ the smoothed histogram.
Then
$$\int_0^{\infty} \mid \hat{f}_k(x) - f(x) \mid dx \xrightarrow{P} 0 \quad as \quad n \longrightarrow \infty,$$
under Condition 2.

5 Estimation of unbounded densities at $x = 0$

As already mentioned the standard kernel density estimator suffers from a boundary bias for the class of density functions defined on $[0,+\infty)$. Until now all previous methods aimed to remove this boundary bias have been developped under the assumption of a *bounded* density at $x = 0$. In this section, we consider a density function f defined on $[0,+\infty)$ and *unbounded* at $x = 0$. This obviously should induce a clustering of observations near the boundary. As shown in Figure 3 behaviors of the standard kernel density estimator and the true density can be dramatically different. This illustrative estimation has been performed on $n = 200$ data drawn from a gamma density $\Gamma(\lambda, \alpha)$ with $\alpha = 0.7$ and $\lambda = 2$. We have used here a Gaussian kernel with a bandwidth minimizing the MISE. As far as we know only two methods have been shown to accomodate the case of an unbounded density: the complicated P and PD algorithms developed by Marron and Ruppert (1994), and the Bernstein polynomial and Beta kernel estimators (Bouezmarni and Rolin (2002)). However the last estimators do not apply here since they are designed for density functions defined on $[0, 1]$ (the P, resp. PD, algorithm further requires the presence of poles at both boundaries, resp. one boundary).

Coming back to Figure 3 we may observe that the gamma kernel density estimator and the smoothed histogram based on the Poisson distribution exhibit the same behavior at the boundary point and interior points as the true gamma density function $\Gamma(0.7, 2)$. In fact these two estimators satisfy the additional sufficient conditions needed to get weak convergence of the asymmetric kernel density estimator and smoothed histogram to infinity at $x = 0$.

Theorem 5.1 *Let f be a probability density function on $[0, +\infty)$, unbounded at $x = 0$ and $\hat{f}_b$ the asymmetric kernel density estimator.*
Under Condition 1 we have

$$\hat{f}_b(0) \xrightarrow{P} +\infty \quad \textit{as} \qquad n \longrightarrow \infty,$$

if

$$\textit{for all } \delta > 0, \quad \int_0^\delta K(0, b)(t)\, dt \longrightarrow 1 \quad \textit{as} \qquad b \longrightarrow 0,$$

The additional condition in the above theorem can be checked for the asymmetric kernel density estimator based on the gamma kernel. Indeed we have for $\delta > 0$:

$$\int_0^\delta K_G(0, b)(t)\, dt = 1 - \exp(-\delta/b) \longrightarrow 1 \quad \text{as } b \longrightarrow 0.$$

Hence the gamma kernel density estimator gives almost all weight to the boundary point when the bandwidth converges to zero. This is due to the particular behaviour of the gamma kernel at $x = 0$. Note that the two other asymmetric kernels do not share this behaviour and will not be suitable for estimation of unbounded densities.

Let us now examine the case of smoothed histograms.

Theorem 5.2 *Let f be a probability density function on $[0, +\infty)$, unbounded at $x = 0$ and $\hat{f}_k$ the smoothed histogram.*
Under Condition 2, we have

$$\hat{f}_k(0) \xrightarrow{P} +\infty \quad as \qquad n \longrightarrow \infty,$$

if

$$p_{k0}(0) \longrightarrow 1 \quad as \;\; k \longrightarrow +\infty,$$

When $p_{ki}(x)$ corresponds to a Poisson distribution, we have

$$p_{k0}(0) = 1 \quad \text{for all} \quad k,$$

and the additional condition of the last theorem is fulfilled.

This means that the smoothed histogram based on the Poisson distribution gives a large weight to the boundary point. The convergence result should not come as a surprise in view of the particular behaviour of $kp_{ki}(x)$ at $i = 0$ (cf. Section 2).

6 Monte Carlo results

This section gathers some simulation results about the finite sample properties of the gamma kernel estimator and the smoothed histogram based on the Poisson distribution. We compare their properties with those of the Gaussian kernel estimator and a log-transformed gaussian kernel estimator (transformation kernel density estimator based on the Gaussian kernel and the logarithmic mapping). We consider ten test densities with a left end boundary. The group of densities contains bounded and unbounded densities with either a single mode or two modes.

(a) the standard lognormal density: $f(x) = (1/x\sqrt{2\pi}) \exp(-(\ln x)^2/2)$.

(b) the chisquare density with one degree of freedom: $f(x) = (1/\sqrt{2\pi x})e^{-x/2}$.

(c) the Maxwell's density: $f(x) = x \exp(-x^2/2)$.

(d) the gamma density with scale $\alpha = 2$ and shape $\lambda = 2$.

(e) the gamma density with scale $\alpha = 2$ and shape $\lambda = 0.7$.

(f) the standard exponential density: $f(x) = \exp -x$.

(g) the asymmetric Pareto density with parameter 3/2: $f(x) = 1/(2x^{3/2})$ on $[1, +\infty)$.

(h) the inverse exponential density: $f(x) = 1/(2x^{3/2})\exp(-1/\sqrt{x})$.

(i) a gamma mixture: 2/3 weight is put on a $\Gamma(0.7, 2)$ and 1/3 weight on a $\Gamma(20, 0.2)$.

(j) a lognormal mixture: 2/3 weight is put on a $LN(0, 1)$ and 1/3 weight on a $LN(1.5, 0.1)$.

The study is based on 100 simulations for each density. For each simulation the bandwidth minimizing the L_1 norm among a grid of values is chosen. Throughout we have a sample size of $n = 200$. Figures 4-8 plot the different densities together with the average of $\hat{f}$ on the different simulations. This gives an idea of the finite sample bias of the different estimators. Bias seems to be always smaller for the smoothed histogram and gamma kernel estimates (left column) in comparison with the standard kernel and log-transformed kernel estimates (right column).

Global performance is assessed in terms of mean, median and variance of $\int |\hat{f} - f|$ on the 100 simulations. Tables 1-3 list results. The conclusions concerning the global behaviour may be summarised as follows. Tables 1 and 2 show that the gamma kernel estimator is always dominated by the smoothed histogram in terms of mean and median of $\int |\hat{f} - f|$. The variance for the smoothed histogram is smaller for the five first densities and larger for the last five. When the shape is lognormal (density (a)) or close to (density (h)) the log-transformed kernel estimator performs better than the smoothed histogram and the gamma kernel estimator in terms of mean and median. The variance in the log-transformed case is smaller for distributions (f)-(g) and (i) when compared to the smoothed histogram and gamma kernel estimates. It is also smaller for distributions (b)-(c) with regard to gamma kernel estimates.

7 An application to income data

The empirical illustration concerns the analysis of the income distribution for Brazil in 1981. The estimation is performed on a comprehensive micro-data set ($n = 101,864$) already used in a study of the dynamics of income inequality by Cowell, Ferreira and Litchfield (1998). These authors were interested in these data because of the importance of Brazil as a major world economy (ninth largest GDP) and the presence of a strong inequality in terms of percentage shares of income accruing to the richest and to the poorest of its population. This strong inequality is in fact revealed by the abnormal skewness of the income distribution (see Table 4 for the descriptive statistics). Income should be understood as gross monthly household income per capita denominated in 1981 cruzeiros, where the income receiver is the individual. Since lots of data are located near the boundary it would not be surprising that the true density is unbounded at the boundary.

Figure 9 compares the results of alternative estimation approaches. The upper part plots the gamma kernel estimator and the smoothed histogram based on the Poisson distribution together with a pseudo maximum likelihood estimate under a parametric assumption of a gamma distribution. The smoothing parameters b and k have been choosen according to a bandwidth selection method inspired by the proposal of Hall and Wand (1988), which leads to an asymptotically optimal window in the sense of minimizing L_1 distance. For the gamma kernel density estimator it consists in setting $b^* = n^{-2/5}\,(u^*)^4$, where u^* is that value of u which minimises:

$$\lambda(u) = 2\int_{\epsilon}^{\infty}\left(u^4 B_0(x)\Phi\left(\frac{u^3 B_0(x)}{\sigma_0(x)}\right) + u^{-1}\sigma_0(x)\phi\left(\frac{u^3 B_0(x)}{\sigma_0(x)}\right)\right)dx, \tag{8}$$

where ϵ is a small strictly positive number (chosen to avoid any problems due to potential undefined derivatives at zero when performing numerical integration ; we have taken $\epsilon = 10^{-15}$ in the simulation results presented below), $B_0(x) = f'(x) + (1/2)xf''(x)$ and $\sigma_0(x) = 1/(2\sqrt{\pi})^{1/2}x^{-1/4}f(x)^{1/2}$ while ϕ and Φ denote the Normal density and distribution functions, respectively. The same procedure applies to the smoothed histogram based on the Poisson distribution by taking $k^* = (b^*)^{-1}$ with $B_0(x) = (1/2)(f'(x) + xf''(x))$. Unknown quantities in criterion (8) have been computed from the fitted gamma distribution.

Estimated values for the gamma distribution are 0.9233, resp. 13156, for the shape, resp. scale, parameter with standard deviation 1.59E-007, resp. 44.64. It is worth being emphasised that the estimate of the shape parameter yields an unbounded density at $x = 0$. The smoothing parameter values b^* and k^* based on this L_1 reference density method with a parametric assumption of a gamma distribution (see Devroye (1997) for further discussion in the context of standard kernel density estimators) are found to be $b^* = 0.02915$ and $k^* = 33$.

To check whether the chosen L_1 reference density method is a satisfactory bandwidth selection procedure in practice, we have applied it on 10 simulated samples from the distribution (e) (unbounded gamma distribution) of Section 6. Table 5 shows that the values of the data-driven bandwidth are akin to the values of the optimal bandwidth, which entails similar performance in terms of $\int |\hat{f} - f|$.

Finally, the lower part of Figure 9 plots standard nonparametric estimates performed with a Gaussian kernel on the raw data and log-transformed data (transformation kernel density estimator based on the logarithmic mapping). Bandwidth values are selected here by an L_1 reference density method with a parametric assumption of a normal distribution for the raw data or the log-transformed data. This corresponds to take $h = 2.3019\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ denotes the empirical standard deviation of the raw data or the log-transformed data. The difference between the two parts of the figure is striking and illustrates the practical relevance of the asymmetric kernel density estimator and of the smoothed histogram.

8 Concluding remarks

We have studied consistency of two types of density estimators when the density function is defined on $[0, +\infty)$. These are the asymmetric kernel density estimator and the smoothed histogram. Simulation results show that they both have good finite sample properties and are able to avoid boundary bias existing in standard kernel density estimation. We think that they should be of some help in monitoring the evolution of the shape of density functions, and should therefore be useful in applied work involving such nonparametric techniques (see Härdle and Linton (1994), Pagan and Ullah (1999) for examples). This point has already been illustrated through a nonparametric estimation of the income distribution from a Brazilian micro-data set. Nonparametric hazard rate estimation should be another important area of application (see Fernandes and Grammig (2000) for a convincing use in goodness-of-fit testing procedures for duration models).

A Appendix

A.1 Proofs of Propositions and Theorems of Section 3

Without loss of generality, we suppose that $I = [\eta_1, \eta_2]$, $0 < \eta_1 < \eta_2$ and $x \in I$.

Proof of Proposition 3.1

We note by μ_x and σ_x^2 the mean and the variance of the random variable ξ_x where ξ_x has a $K(x, b)$ density function defined in (2), (3) and (4). Note that

$$\mu_x = x \qquad \text{or} \qquad \mu_x = x + b \qquad \text{and} \qquad \sigma_x^2 = bx^3 \quad \text{or} \quad \sigma_x^2 = bx + b^2.$$

Since f is uniformly continuous on $J = [\eta_1/2, 2\eta_2]$, for $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ such that

$$| f(t) - f(x) | < \epsilon, \qquad \text{for} \qquad | x - t | < \delta \quad \text{for all } \ x \text{ and } \ t \ \text{ in } J. \tag{9}$$

For all x in I we have

$$\begin{aligned}
| I\!E\left(\hat{f}_b(x)\right) - f(x) | &= \left| \int_0^{+\infty} (f(t) - f(x))\, K(x, b)(t) dt \right| \\
&\leq \int_{|t-\mu_x|<\delta} |f(t) - f(x)|\, K(x, b)(t) dt \\
&+ \int_{|t-\mu_x|>\delta} |f(t) - f(x)|\, K(x, b)(t) dt \\
&\leq \int_{|t-\mu_x|<\delta} |f(t) - f(x)|\, K(x, b)(t) dt \\
&+ \ 2 \sup_{t\in(0,+\infty)} | f(t) | \int_{|t-\mu_x|>\delta} K(x, b)(t) dt.
\end{aligned}$$

From Equation (10) and using Tchebyshev inequality we deduce

$$\begin{aligned}
| I\!E\left(\hat{f}_b(x)\right) - f(x) | &\leq \epsilon + \left\{ 2 \sup_{t\in(0,+\infty)} | f(t) | \right\} \frac{\sigma_x^2}{\delta^2} \\
&\leq 2\epsilon, \qquad \text{for} \qquad b < \delta_1.
\end{aligned} \tag{10}$$

■

Proof of Theorem 3.1

We begin with the usual inequality

$$\begin{aligned}\sup_{x\in I} \mid \hat{f}_b(x) - f(x) \mid &\leq \sup_{x\in I} \mid \hat{f}_b(x) - I\!\!E(\hat{f}_b(x)) \mid \\ &+ \sup_{x\in I} \mid I\!\!E(\hat{f}_b(x)) - f(x) \mid .\end{aligned}$$

Since the second term is non-stochastic and converges to zero by Proposition 3.1, we only need to prove that

$$\sup_{x\in I} \mid \hat{f}_b(x) - I\!\!E(\hat{f}_b(x)) \mid \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty. \tag{11}$$

Note that

$$\begin{aligned}\hat{f}_b(x) &= \frac{1}{n}\sum_{i=1}^{n} K(x,b)(X_i) \\ &= \int_0^{+\infty} K(x,b)(t)dF_n(t),\end{aligned}$$

and

$$\begin{aligned}I\!\!E(\hat{f}_b(x)) &= \frac{1}{n}\sum_{i=1}^{n} I\!\!E(K(x,b)(X_i)) \\ &= I\!\!E(K(x,b)(Y)) \\ &= \int_0^{+\infty} K(x,b)(t)dF(t),\end{aligned}$$

where Y is a random variable with distribution F with an unknown probability density function f.

For all x,

$$\begin{aligned}\mid \hat{f}_b(x) - I\!\!E(\hat{f}_b(x)) \mid &= \left| \int_0^{+\infty} K(x,b)(t)\, d[F_n(t) - F(t)] \right| \\ &= \left| \int_0^{+\infty} (F_n(t) - F(t))\, dK(x,b)(t) \right| \\ &\leq \int_0^{\infty} |F_n(t) - F(t)|\, |dK(x,b)(t)| \\ &\leq \sup_{t\in[0,+\infty)} |F_n(t) - F(t)| \int_0^{+\infty} |dK(x,b)(t)| \\ &\leq Cb^{-a} \sup_{t\in[0,+\infty)} |F_n(t) - F(t)|\,,\end{aligned}$$

where C is a constant, $a = 1$ for the gamma kernel and $a = 5/2$ for the IG and RIG kernels.
In fact, we have for the gamma kernel:

$$\begin{aligned}\int_0^{+\infty} |dK_G(x,b)(t)| &= b^{-1}\int_0^{+\infty} |K_G(x,b)(t) - K_G(x-b,b)(t)|\, dt \\ &\leq 2b^{-1},\end{aligned}$$

while it can be found that

$$\int_0^{+\infty} |dK_{IG}(x,b)(t)| = O(b^{-5/2}),$$

and

$$\int_0^{+\infty} |dK_{RIG}(x,b)(t)| = O(b^{-5/2}).$$

Now, applying the result in Massart (1990) on the Dvoretzky-Kiefer-Wolfowitz (1956) inequality, we obtain

$$\begin{aligned}\mathbb{P}\left(\sup_{x\in I}\left|\hat{f}_b(x) - \mathbb{E}(\hat{f}_b(x))\right| \geq \epsilon\right) &\leq \mathbb{P}\left(\sup_{t\in[0,+\infty)} |F_n(t) - F(t)| \geq \frac{\epsilon b^a}{C}\right) \\ &\leq 2\exp\left(-\frac{2\epsilon^2}{C^2}\, nb^{2a}\right). \end{aligned} \tag{12}$$

■

Proof of Corollary 3.1

From the proof of Theorem 3.1, it suffices to show:

$$\sup_{x\in I} \mid \hat{f}_b(x) - \mathbb{E}(\hat{f}_b(x)) \mid \xrightarrow{a.s} 0 \quad \text{as} \quad n \longrightarrow \infty.$$

For $\epsilon_n > 0$ we have

$$\mathbb{P}\left(\sup_{x\in I}\left|\hat{f}_b(x) - \mathbb{E}(\hat{f}_b(x))\right| \geq \epsilon_n\right) \leq 2\exp\left(-\frac{2\epsilon_n^2}{C^2}\, nb^{2a}\right). \tag{13}$$

Let us take $\epsilon_n = \dfrac{\alpha}{b^a C}\sqrt{\dfrac{\ln n}{n}}$ with $\alpha > \dfrac{C}{\sqrt{2}}$. This yields

$$\sum_{n\geq 1} \exp\left(-\frac{2\epsilon_n^2}{C^2}\, nb^{2a}\right) < \infty.$$

Therefore, we conclude by application of the Borel-Cantelli Theorem the uniform strong consistency on each compact set of the asymmetric kernel density estimator.

∎

Proof of Proposition 3.2

Let δ be a small positive number and x in the compact set I

$$\begin{aligned}
\mid \mathbb{E}(\hat{f}_k(x)) - f(x) \mid &= \left| k \sum_{i=0}^{+\infty} p_{ki}(x) \int_{\frac{i}{k}}^{\frac{i}{k}+\frac{1}{k}} \{f(t) - f(x)\}\, dt \right| \\
&= \left| k\mathbb{E}_x \left(\int_{\frac{S_k}{k}}^{\frac{S_k}{k}+\frac{1}{k}} \{f(t) - f(x)\}\, dt \right) \right| \\
&\leq k\mathbb{E}_x \left[\int_{\frac{S_k}{k}}^{\frac{S_k}{k}+\frac{1}{k}} |f(t) - f(x)|\, dt \ ; \ \left| \frac{S_k}{k} - x \right| > \delta \right] \\
&+ k\mathbb{E}_x \left[\int_{\frac{S_k}{k}}^{\frac{S_k}{k}+\frac{1}{k}} |f(t) - f(x)|\, dt \ ; \ \left| \frac{S_k}{k} - x \right| < \delta \right] \\
&= B_1 + B_2. \qquad (14)
\end{aligned}$$

Then we get using Chebyshev's inequality:

$$\begin{aligned}
B_1 &= k\mathbb{E}_x \left[\int_{\frac{S_k}{k}}^{\frac{S_k}{k}+\frac{1}{k}} |f(t) - f(x)|\, dt \ ; \ \left| \frac{S_k}{k} - x \right| > \delta \right] \\
&\leq 2 \sup_t \mid f(t) \mid \mathbb{P}_x \left(\left| \frac{S_k}{k} - x \right| > \delta \right) \\
&\leq \epsilon \qquad (15)
\end{aligned}$$

For the second term B_2, we have:

$$\begin{aligned}
B_2 &= k\mathbb{E}_x \left[\int_{\frac{S_k}{k}}^{\frac{S_k}{k}+\frac{1}{k}} \mid f(t) - f(x) \mid\, dt; \mid \frac{S_k}{k} - x \mid \leq \delta \right] \\
&\leq \mathbb{E}_x \left[\int_0^1 \mid f(\frac{S_k}{k} + \frac{s}{k}) - f(x) \mid\, ds; \mid \frac{S_k}{k} - x \mid \leq \delta \right] \\
&\leq \epsilon.
\end{aligned}$$

In fact, f is uniformly continuous on $[\frac{\eta_1}{2}, 2\eta_2]$, which implies that for $\epsilon > 0$, there exists $\delta(\epsilon) > 0$, such that

$$| f(t') - f(t) | < \epsilon, \quad \text{for} \quad | t' - t | < \delta(\epsilon) \quad \text{and} \quad t, t' \in [\frac{\eta_1}{2}, 2\eta_2].$$

If we choose δ and k such that:

$$\delta \ < \ \inf\left(\frac{\delta(\epsilon)}{2}, \frac{\eta_1}{2}\right) \quad \text{and} \quad k \ > \ \sup\left(\frac{2}{\delta(\epsilon)}, \frac{2}{\eta_2}\right). \tag{16}$$

Then

$$\left|\frac{S_k}{k} + \frac{s}{k} - x\right| < \delta(\epsilon), \qquad \text{for} \qquad | \frac{S_k}{k} - x | \leq \delta \quad \text{and} \quad 0 < s < 1,$$

and

$$\frac{S_k}{k} + \frac{s}{k} \in [\frac{\eta_1}{2}, 2\eta_2].$$

$\blacksquare$

Proof of Theorem 3.2

It is clear that

$$\begin{aligned} \sup_{x\in I} | \hat{f}_k(x) - f(x) | \ &\leq \ \sup_{x\in I} | \hat{f}_k(x) - \mathbb{E}(\hat{f}_k(x)) \ | \\ &+ \ \sup_{x\in I} | \ \mathbb{E}(\hat{f}_k(x)) - f(x) \ | \,. \end{aligned} \tag{17}$$

For the nonstochastic term, by Proposition 3.2, there exists an integer $n_0(\epsilon)$ such that

$$\sup_{x\in I} | \ \mathbb{E}(\hat{f}_k(x)) - f(x) \ | < \epsilon/2 \quad \text{for all} \quad n > n_0(\epsilon).$$

Then, for all $n > n_0(\epsilon)$

$$\begin{aligned} \mathbb{P}\left(\sup_{x\in I} | \hat{f}_k(x) - f(x) | > \epsilon\right) \ &\leq \ \mathbb{P}\left(\sup_{x\in I} | \hat{f}_k(x) - \mathbb{E}(\hat{f}_k(x)) \ | \right. \\ &\qquad \left. + \sup_{x\in I} | \ \mathbb{E}(\hat{f}_k(x)) - f(x) \ | > \epsilon\right) \\ &\leq \ \mathbb{P}\left(\sup_{x\in I} | \hat{f}_k(x) - \mathbb{E}(\hat{f}_k(x)) \ | > \epsilon/2\right). \end{aligned}$$

But we have for $x \in I$:

$$
\begin{aligned}
\mid \ \hat{f}_k(x) - \mathbb{E}(\hat{f}_k(x)) \ \mid \ &= \ \left| k \sum_{j=0}^{+\infty} \left(F_n(A_{kj}) - F(A_{kj}) \right) p_{kj}(x) \right| \\
&\leq \ 2k \sup_{t \in [0,+\infty)} \left| F_n(t) - F(t) \right| . \qquad (18)
\end{aligned}
$$

where

$$
A_{kj} = \left(\frac{j}{k}, \frac{j+1}{k} \right] \qquad j = 0, \ldots .
$$

Hence the version of the inequality of Dvoretzky-Kiefer-Wolfowitz (1956) given in Massart (1990) yields:

$$
\begin{aligned}
\mathbb{P}\left(\sup_{x \in I} \mid \ \hat{f}_k(x) - \mathbb{E}(\hat{f}_k(x)) \ \mid > \varepsilon/2 \right) \ &\leq \ \mathbb{P}\left(\sup_{x \in [0,+\infty)} \mid \ F_n(x) - F(x) \ \mid > \frac{\varepsilon}{4k} \right) \\
&\leq \ 2 \exp\left(-\frac{1}{8} \epsilon^2 \, n \, k^{-2} \right) .
\end{aligned}
$$

This leads to the stated result.

■

Proof of Corollary 3.2

Along the same lines as the proof of Corollary 3.1. ■

A.2 Proofs of Propositions and Theorems of Section 4

Proof of Proposition 4.1

First, we suppose that the density f is a continuous density function vanishing outside a compact, the result is deduced from Theorem(3.1) and Scheffé's theorem (1947). It can be shown that, the Asymmetric kernel density estimator with gamma kernel,

$$
\int_0^{\infty} \left\{ \mathbb{E}(\hat{f}_b)(x) \right\} dx = \int_0^{\infty} f(t) \left\{ \int_0^{\infty} k(x,b)(t) dt \right\} dx \longrightarrow 1
$$

because

$$
\int_0^{\infty} k(x,b)(t) dx \longrightarrow 1
$$

and

$$\int_0^\infty k(x,b)(t)dx \leq c$$

where c is a universal constant.
Now let f be any probability density function. We know that, for $\epsilon > 0$ there exists a continuous density function g vanishing outside a compact set such that:

$$\| f - g \|_1 = \int | f(x) - g(x) | \, dx \; < \; \epsilon.$$

Then

$$\begin{aligned} \| I\!\!E(\hat{f}_b) - f \|_1 &\leq \| I\!\!E(\hat{f}_b) - I\!\!E(\hat{g}_b) \|_1 + \| I\!\!E(\hat{g}_b) - g \|_1 + \| f - g \|_1 \\ &\leq \int \int | f(t) - g(t) | \, K(x,b) dx dt + 2\epsilon \\ &\leq (C+2)\epsilon. \end{aligned}$$

where $C = \int K(x,b)dx = 1$ for the IG and RIG kernels and $C = \int K(x,b)dx \simeq 1$ for the gamma kernel. This completes the proof.

■

Proof of Theorem 4.1

From Proposition 4.1, it is sufficient to prove that

$$\int_0^\infty | I\!\!E(\hat{f}_b(x)) - \hat{f}_b(x) | \, dx \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty.$$

We have

$$\begin{aligned} \int_0^\infty | I\!\!E(\hat{f}_b(x)) - \hat{f}_b(x) | \, dx &= \int_0^{+\infty} \left| \int_0^{+\infty} d(F_n(t) - F(t)) \, K(x,b)(t) \right| dx \\ &= \int_0^{+\infty} \left| \int_0^{+\infty} (F_n(t) - F(t)) \, dK(x,b)(t) \right| dx \\ &\leq \int_0^{+\infty} \int_0^{+\infty} | F_n(t) - F(t) | \; dK(x,b)(t) dx, \end{aligned}$$

but

$$\int \int |dK(x,b)(t)| \, dx = O(b^{-a}),$$

where $a = 1$ for the gamma kernel and $a = 5/2$ for the IG and RIG kernels.
Then

$$\int_0^\infty | I\!\!E(\hat{f}_b(x)) - \hat{f}_b(x) | \, dx \;\leq\; b^{-a} C_1 \sup_t | F_n(t) - F(t) | \, .$$

We finally get:

$$
\begin{aligned}
\mathbb{P}\left(\int_0^{\infty} \mid \mathbb{E}(\hat{f}_b(x)) - \hat{f}_b(x) \mid dx > \epsilon\right) &\leq \mathbb{P}\left(\sup_t \mid F_n(t) - F(t) \mid > \epsilon b^a / C_1\right) \\
&\leq 2\exp\left(-2\frac{\epsilon^2}{C_1^2} n b^{2a}\right),
\end{aligned}
$$

thanks to Massart (1990).

■

Proof of Proposition 4.2

Assume that the density f is a continuous density function vanishing outside a compact, the result is deduced from Theorem(3.2) and Scheffé's theorem (1947). Now let f be any probability density function. Then for $\epsilon > 0$ there exists a continuous density function g vanishing outside a compact set such that:

$$
\| f - g \|_1 = \int \mid f(x) - g(x) \mid dx \; < \; \epsilon,
$$

and

$$
\begin{aligned}
\| \mathbb{E}(\hat{f}_k) - f \|_1 &\leq \| \mathbb{E}(\hat{f}_k) - \mathbb{E}(\hat{g}_k) \|_1 + \| \mathbb{E}(\hat{g}_k) - g \|_1 + \| f - g \|_1 \\
&\leq \sum_{i=0}^{+\infty} \int_{\frac{i}{k}}^{\frac{i+1}{k}} \mid f(t) - g(t) \mid dt \int k p_{ki}(x) dx + 2\epsilon \\
&\leq 3\epsilon.
\end{aligned}
$$

This concludes the proof.

■

Proof of Theorem 4.2

From Proposition 4.2, it is sufficient to prove that

$$
\int_0^{\infty} \mid \mathbb{E}(\hat{f}_k(x)) - \hat{f}_k(x) \mid dx \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty.
$$

First, from the proof of Theorem 3.2 we know that:

$$
\begin{aligned}
\int_0^{+\infty} \mid \; \mathbb{E}(\hat{f}_k(x)) - \hat{f}_k(x) \; \mid dx &\leq \int_0^{+\infty} k \sum_{j=0}^{+\infty} |F_n(A_{kj}) - F(A_{kj})| \, p_{kj}(x) dx \\
&\leq \sum_{j=0}^{+\infty} |F_n(A_{kj}) - F(A_{kj})|
\end{aligned}
$$

Second, from Devroye and Gyorfi (1985), we deduce that the last term converges in probability under Condition 2.

■

Proof of Theorem 5.1

From the proof of Theorem 3.1, we have

$$| \hat{f}_b(0) - \mathbb{E}(\hat{f}_b(0)) | \xrightarrow{P} 0 \quad \text{as} \quad nb^{2a} \longrightarrow \infty, \tag{19}$$

i.e $\hat{f}_b(0)$ and $\mathbb{E}(\hat{f}_b(0))$ have the same asymptotic behavior as $nb^{2a} \longrightarrow \infty$.

Now we prove that:

$$\mathbb{E}\left(\hat{f}_b(0)\right) \longrightarrow +\infty \quad \text{as} \qquad b \longrightarrow 0, \tag{20}$$

i.e for $A > 0$ there exists $\eta > 0$ such that $\mathbb{E}\left(\hat{f}_b(0)\right) > A$ for all $b < \eta$.
In fact,

$$f(t) \longrightarrow +\infty \quad \text{as} \qquad t \longrightarrow 0,$$

and thus for $A > 0$ there exists $\delta(A) > 0$ such that $f(t) > 2A$ for all $0 < t < \delta$. Now

$$\begin{aligned}
\mathbb{E}\left(\hat{f}_b(0)\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} K(0,b)(X_i)\right) \\
&= \mathbb{E}\left(K(0,b)(X_i)\right) \\
&= \int f(t)K(0,b)(t)dt \\
&\geq 2A\int_0^\delta K(0,b)(t)\,dt
\end{aligned}$$

If we suppose that

$$\text{for all } \delta > 0, \quad \int_0^\delta k(0,b)(t) \longrightarrow 1 \quad \text{as} \qquad b \longrightarrow 0,$$

we get

$$\int_0^\delta K(0,b)(t)\,dt > 1/2, \qquad \text{for all} \quad b < \eta.$$

Then, for $A > 0$ there exists η such that

$$\mathbb{E}\left(\hat{f}_b(0)\right) > A, \qquad \text{for all} \qquad b < \eta,$$

wich leads to the stated result.

■

Proof of Theorem 5.2

First, we show that:

$$\mathbb{E}(\hat{f}_k(0)) \longrightarrow +\infty \quad \text{as} \quad k \longrightarrow \infty, \tag{21}$$

i.e. for $A > 0$, there exists k_0 such that

$$\mathbb{E}(\hat{f}_k(0)) > A, \qquad \text{for all} \qquad k \geq k_0.$$

In fact, since $f(t) \longrightarrow +\infty$ as $t \longrightarrow 0$ we have

$$f(t) > 2A, \qquad \text{for all} \qquad k \geq k_1.$$

Now if

$$p_{k0}(0) \longrightarrow 1, \quad \text{as } k \longrightarrow +\infty,$$

we have

$$p_{k0}(0) > \frac{1}{2}, \qquad \text{for all} \qquad k \geq k_2.$$

Therefore,

$$\begin{aligned} \mid \mathbb{E}(\hat{f}_k(0)) \mid &= k \sum_{i=0}^{+\infty} p_{ki}(0) \int_{i/k}^{(i+1)/k} f(t)\, dt \\ &\geq k p_{k0}(0) \int_0^{1/k} f(t)\, dt \\ &> A, \qquad \text{for} \qquad k \geq k_0 = \max(k_1, k_2). \end{aligned} \tag{22}$$

From the proof of Theorem 3.2, we know that

$$\mid \hat{f}_k(0) - \mathbb{E}(\hat{f}_k(0)) \mid \xrightarrow{P} 0 \quad \text{as } nk^{-2} \longrightarrow +\infty,$$

which completes the proof.

■

References

Abadir, K. and Lawford, S. (2002). *Optimal Asymmetric Kernels*. University of York, mimeo.

Bouezmarni, T. and Rolin, J. M. (2001). *Consistency of Beta kernel Density Function Estimator*. forthcoming in Can. J. of Statist.

Bouezmarni, T. and Rolin, J. M. (2002). *Bernstein Estimator For Unbounded Density Function.* Institut de Statistique, Université Catholique de Louvain, mimeo#. <http://www.stat.ucl.ac.be>.

Chen, S. X. (1999). *A Beta Kernel Estimator for Density Functions.* Comp. Stat. and Data anal., **31**, 131-145.

Chen, S. X. (2000). *Probability Density Functions Estimation Using Gamma kernels.* Ann. Inst. Statis. Math., **52**, 471-480.

Cowell, F. (1999). *Measurement of Inequality* in Atkinson, A. and F. Bourguignon (eds), Handbook of Income Distribution, North Holland, Amsterdam.

Cowell, F., Ferreira, F. and Litchfield, J. (1998). *Income Distribution in Brazil 1981-1990: Parametric and Non-Parametric Approaches.* J. of Inc. Dist., **8**, 63-76.

Devroye, L. (1997). *Universal Smoothing Factor Selection in Density Estimation: Theory and Practice.* Test, **6**, 223-320.

Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation: The $L_1 View$.* Wiley, New York.

Devroye, L. and Lugosi, G. (1996). *A Universally Acceptable Smoothing Factor for Kernel Density Estimates.* Ann. of Statist., **24**, 2499-2512.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). *Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator.* Ann. of Math. Statist., **1**, 444-453.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, Wiley, New York.

Fenton, V. and Gallant, R. (1996). *Convergence Rates of SNP Density Estimators.* Econometrica, **64**, 719-727.

Fernandes, M. and Grammig, J. (2000). *Nonparametric Specification Tests for Conditional Duration Models.* Fondation Getulio Vargas, mimeo.

Foldes, A. and Revesz, P. (1974). *A General Method for Density Estimation.* Sc. Math. Hung., **7**, 90-94.

Gawronski, N. and Stadtmüller, U. (1980). *On Density Estimation by Means of Poisson's Distribution.* Scand. J. Statist., **7**, 90-94.

Gawronski, N. and Stadtmüller, U. (1981). *Smoothing histograms by means of lattice and continuous distributions.* Metrika, **28**, 155-164.

Härdle, W. and Linton, O. (1994). *Applied Nonparametric Methods*, in Hanbook of Econometrics, **IV**, Eds Engle R. and McFadden D., North-Holland, Amsterdam.

Hall, P. and Wand, M. (1988). *Minimizing L1 Distance in Nonparametric Density Estimation.* J. of Mult. Anal., **26**, 59-88.

Jones, M. (1993). *Simple Boundary Correction For Kernel Density Estimation.* Statist. Comp., **3**, 135-146.

Jones, M. and Foster, P. (1996). *A Simple Nonnegative Boundary Correction Method For Kernel Density Estimation.* Statist. Sin., **6**, 1005-1013.

Marron, J. S. and Ruppert, D. (1994). *Transformations to reduce boundary bias in kernel density estimation.* J. Roy. Statist. Soc. Ser. B **56** , 653-671.

Massart, P. (1990). *The Tight Constant in Dvoretzky-Kiefer-Woolfowitz Inequality.* Ann. of Prob., **18**, 1269-1283.

Müller, H. (1991). *Smooth Optimum Kernel Estimators Near Endpoints.* Biometrika, **78**, 521-530.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Rice, J. (1984). *Boundary Modification for Kernel Regression.* Comm. Statist. Theor. Meth., **13**, 893-900.

Scaillet, O. (2001). *Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels*, IRES DP 2001-17.

Scheffé, H. (1947). *A useful convergence theorem for probability distributions.* Ann. of Math. Statist., **18**, 434-458.

Schuster, E. (1985). *Incorporating Support Constraints into Nonparametric Estimators of Densities*. Comm. Statist. Ser. A, **14**, 1123-1136.

Stadtmüller, U. (1983). *Asymptotic Distribution of Smoothed Histograms*. Metrika, **30**, 145-158.

Walter, G. and Blum, J. (1979). *Probability Density Estimation Using Delta Sequences*. Ann. of Statist., **7**, 328-340.

Densities	Smoothed histogram	Gamma kernel	Log-transformed kernel
(a)	0.09525606	0.1104269	0.08356177
(b)	0.0869127	0.1048551	0.1117045
(c)	0.06243878	0.09577176	0.1969462
(d)	0.0885472	0.1011263	0.1101729
(e)	0.1171365	0.1422371	0.16949476
(f)	0.07683467	0.0960356	0.1188438
(g)	0.03779601	0.0386833	0.04050847
(h)	0.0844546	0.08636857	0.0520408
(i)	0.08000198	0.09000876	0.142668
(j)	0.06897213	0.07567178	0.08634281

Table 1: Mean of $\int |\hat{f} - f|$ on 100 simulations.

Densities	Smoothed histogram	Gamma kernel	Log-transformed kernel
(a)	0.08909575	0.106688	0.07859005
(b)	0.09128226	0.107521	0.11443
(c)	0.06018188	0.09308272	0.1943706
(d)	0.07885886	0.09322441	0.1058726
(e)	0.107452	0.130574	0.144586
(f)	0.07 245554	0.08861409	0.1200464
(g)	0.0376547	0.0375723	0.0384865
(h)	0.08151919	0.08346343	0.04933211
(i)	0.06694741	0.07701519	0.1327077
(j)	0.07049426	0.0762415	0.08064314

Table 2: Median of $\int |\hat{f} - f|$ on 100 simulations.

Densities	Smoothed histogram	Gamma kernel	Log-transformed kernel
(a)	0.000106	0.00097485	0.001203842
(b)	0.00055574	0.00067451	0.00065417
(c)	0.0002052549	0.00118107	0.000638013
(d)	0.001108625	0.00118664	0.001402027
(e)	0.000854753	0.000947512	0.002473
(f)	0.001962692	0.00140385	0.001184696
(g)	0.00064721	0.000457128	0.000187542
(h)	0.001064848	0.0009745832	0.001203842
(i)	0.001246965	0.001171683	0.000972848
(j)	0.000487383	0.0004329083	0.0005053212

Table 3: Variance of $\int |\hat{f} - f|$ on 100 simulations.

Mean income	12,147
Standard deviation	20,551
Skewness	7.737
Kurtosis	124.5
1st percentile	600
1st decile	1,733
1st quartile	3,116
Median income	6,000
3rd quartile	12,754
9th decile	26,954
99th percentile	95,097

Table 4: Descriptives statistics of Brazilian income distribution in 1981

Sample	b_{opt}	b^*	k_{opt}	k^*
Sample 1	0.035261	0.04969	28	31
Sample 2	0.021161	0.030547	30	35
Sample 3	0.031567	0.045217	45	43
Sample 4	0.016124	0.0296718	35	39
Sample 5	0.041323	0.0396519	38	39
Sample 6	0.022657	0.044917	47	50
Sample 7	0.039185	0.050694	42	44
Sample 8	0.036283	0.0382074	25	28
Sample 9	0.030289	0.0372003	29	31
Sample 10	0.021792	0.034291	38	40

Sample	$\int \lvert\hat{f}_k - f\rvert$	$\int \lvert\hat{f}_{k^*} - f\rvert$	$\int \lvert\hat{f}_b - f\rvert$	$\int \lvert\hat{f}_{b^*} - f\rvert$
Sample 1	0.1241102	0.137814	0.147992	0.172371
Sample 2	0.0893857	0.09611	0.1531942	0.208124
Sample 3	0.120691	0.13585	0.1311539	0.176137
Sample 4	0.1357063	0.15386	0.1146338	0.19931
Sample 5	0.11202135	0.131211	0.11300944	0.1766751
Sample 6	0.1381742	0.154867	0.1776474	0.207495
Sample 7	0.08927371	0.10473	0.1384143	0.17839
Sample 8	0.1147298	0.1238744	0.1257563	0.165952
Sample 9	0.11633368	0.123427	0.154371	0.184354
Sample 10	0.09401943	0.11728	0.122073	0.15041

Table 5: L_1 errors for gamma kernel estimator and smoothed histogram under optimal and data-driven bandwiths.

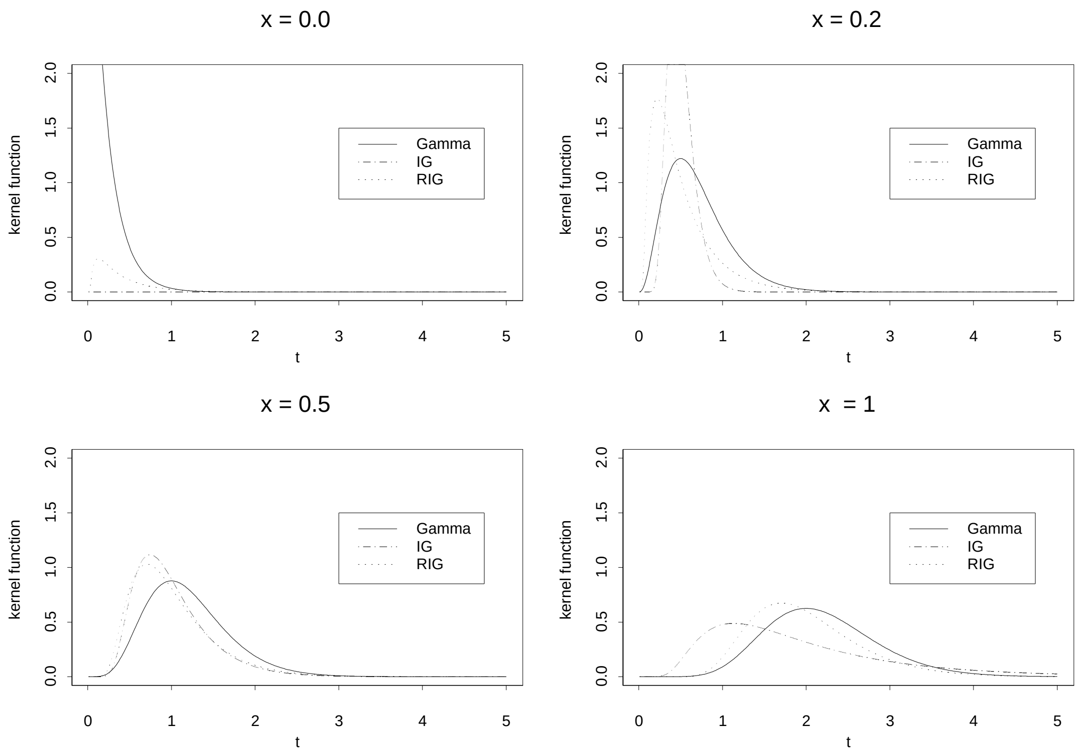


Figure 1: Gamma , IG and RIG kernels $K(x, b)$ for $b = 0.2$.

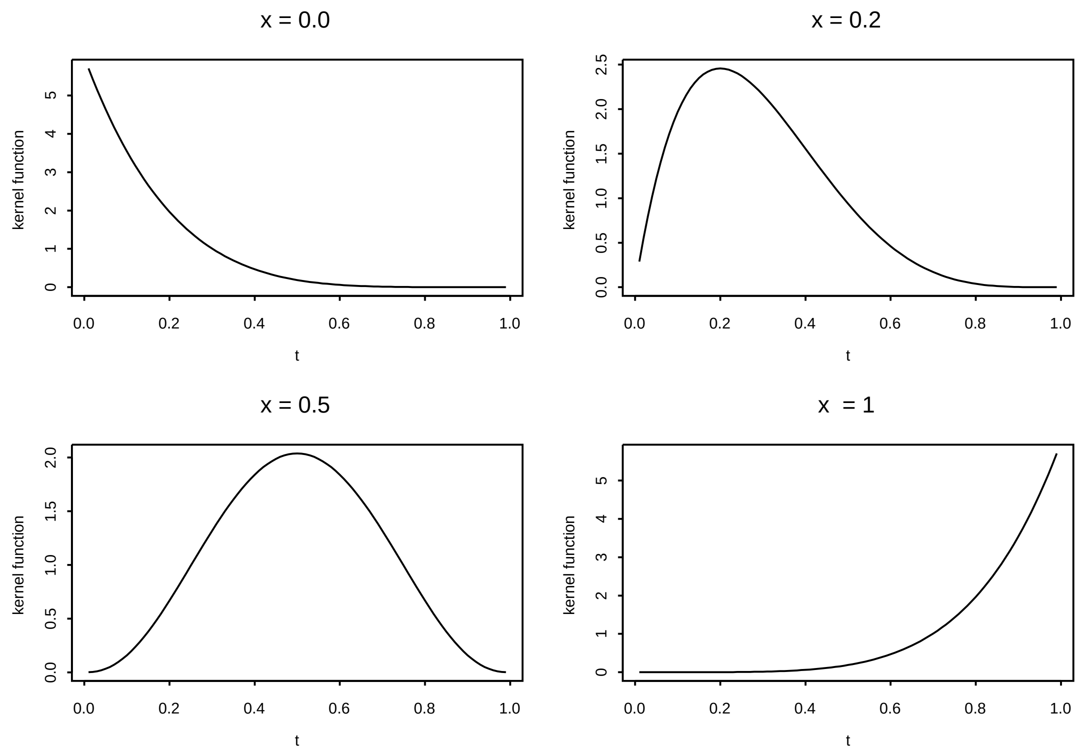


Figure 2: Gamma densities $\Gamma(x, i+1, k)$ with $k = 3$ and $i = 0, 1, 2, 3$.

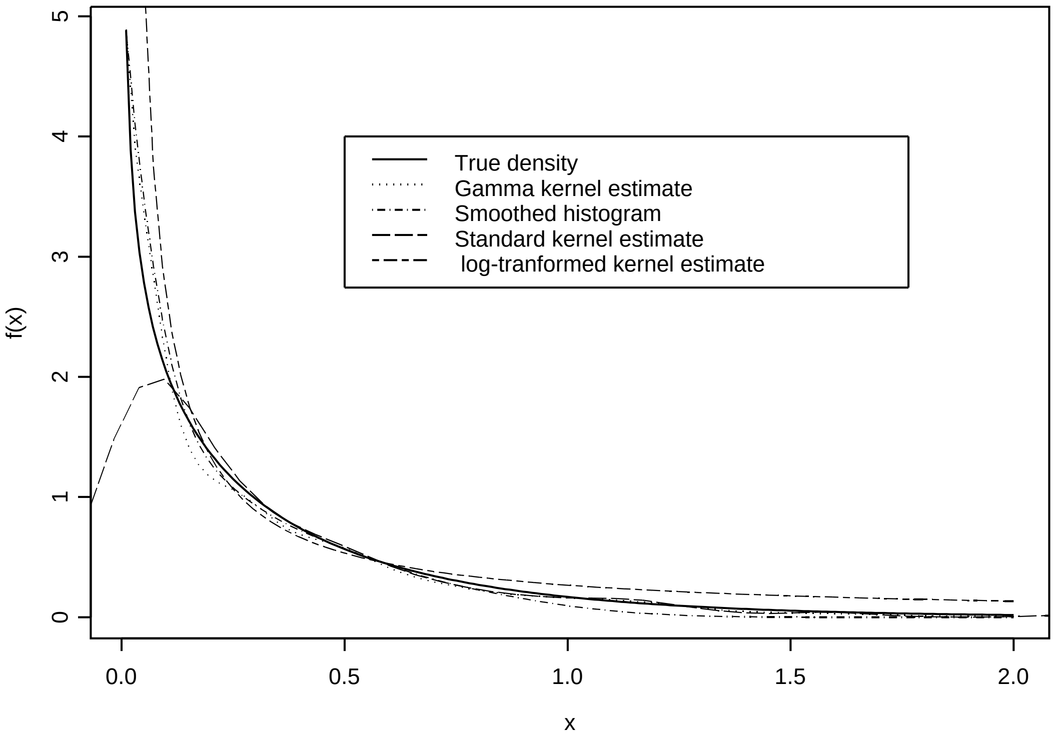


Figure 3: True density $\Gamma(0.7, 2)$ together with its gamma kernel, smoothed histogram, standard kernel and log-tranformed kernel estimates, each based on 200 observations

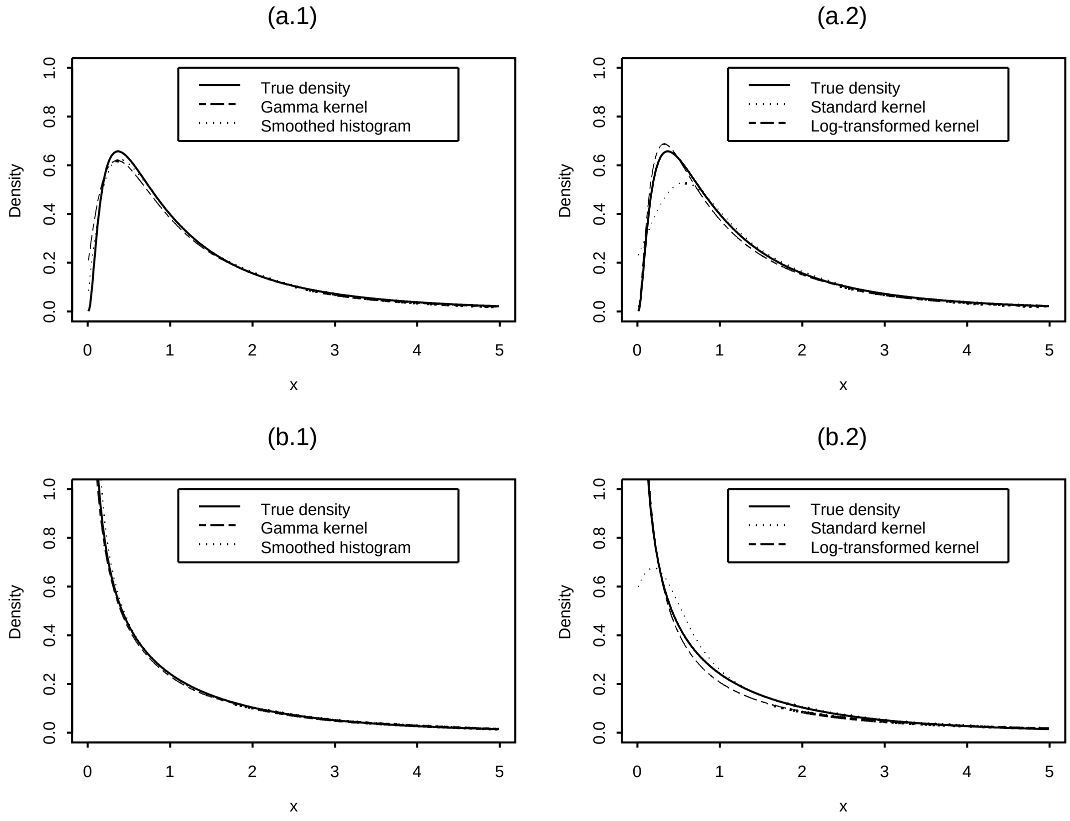


Figure 4: Gamma kernel, smoothed histogram, standard kernel and log-transformed kernel estimates for (a) lognormal, (b) chisquare.

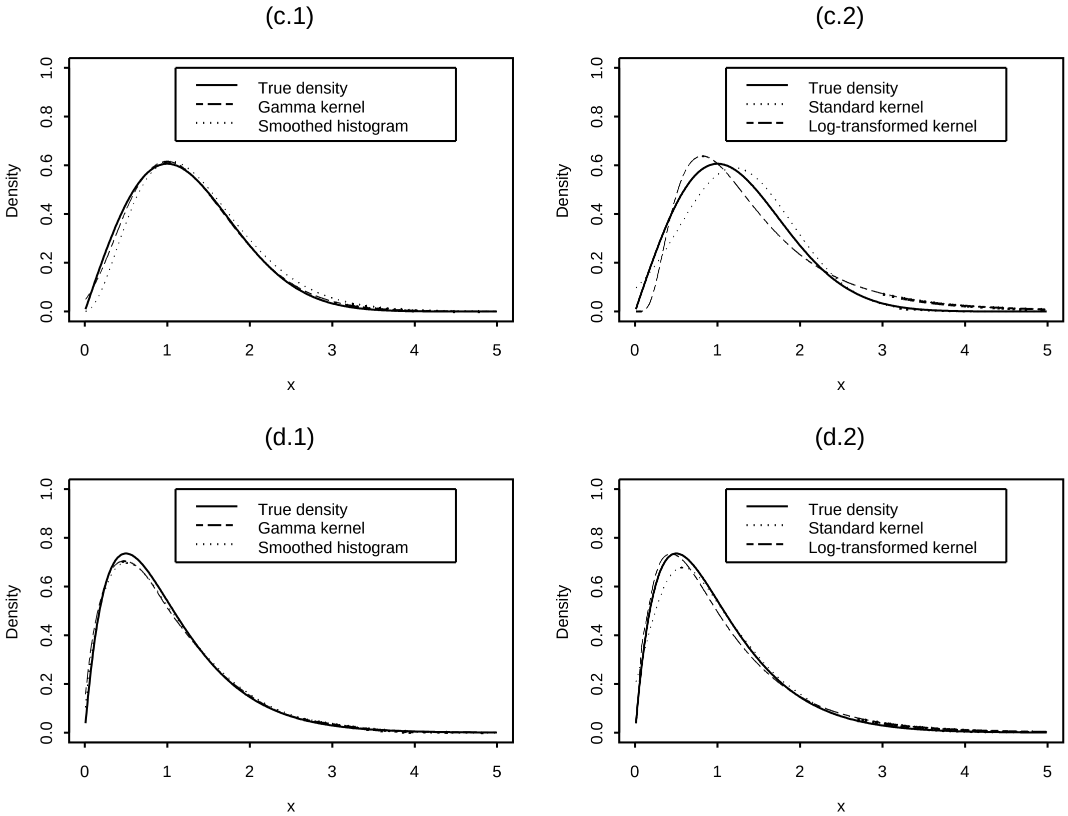


Figure 5: Gamma kernel, smoothed histogram, standard kernel and log-transformed kernel estimates for (c) Maxwell's (d) $\Gamma(2,2)$.

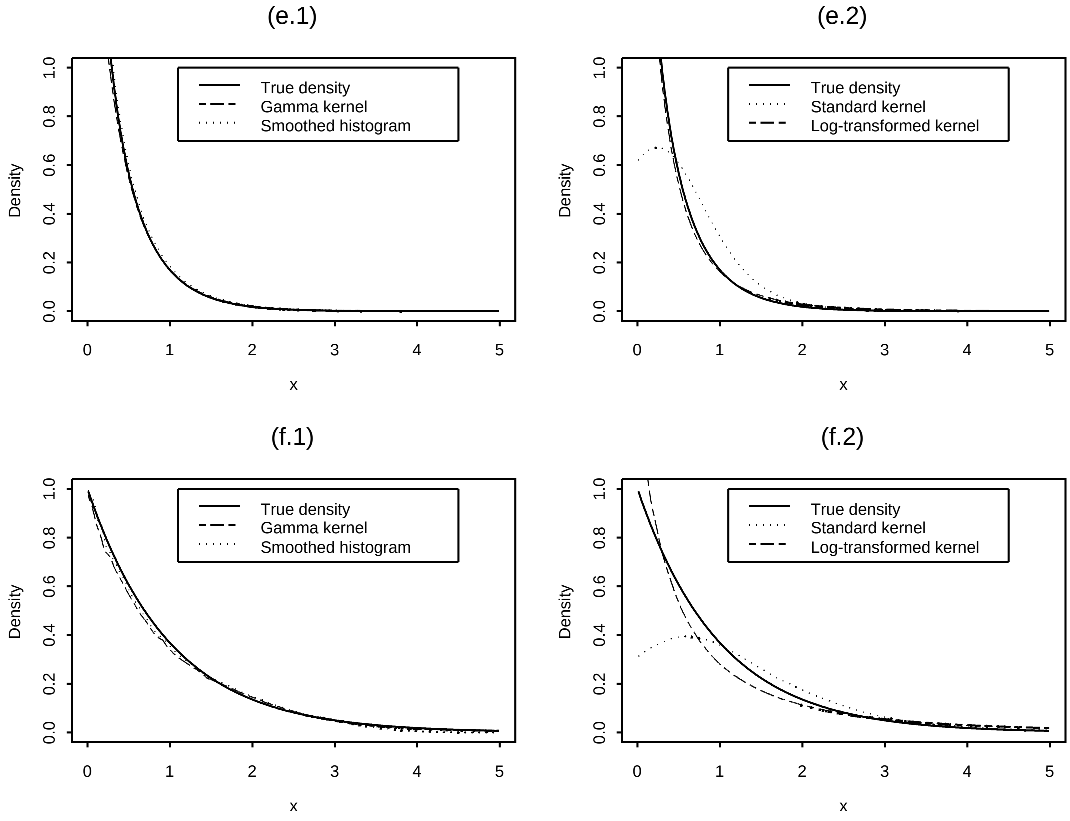


Figure 6: Gamma kernel, smoothed histogram, standard kernel and log-transformed kernel estimates for (e) $\Gamma(0.7, 2)$ (f) exponential.

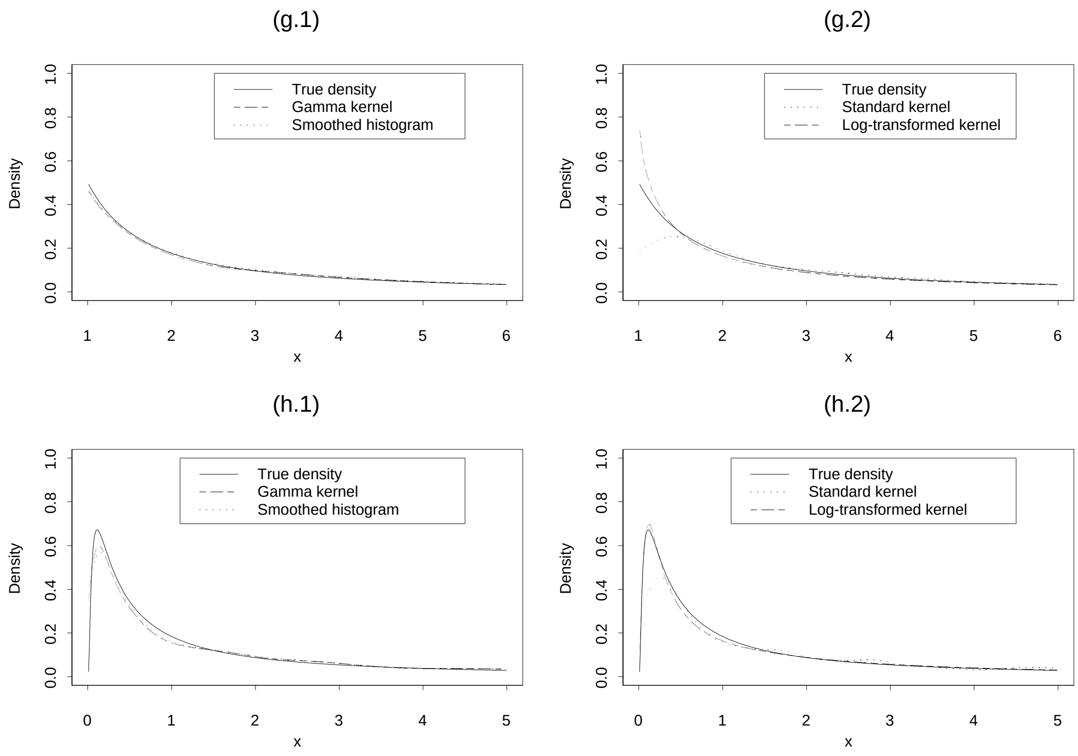


Figure 7: Gamma kernel, smoothed histogram, standard kernel and log-transformed kernel estimates for (g) asymmetric Pareto (h) inverse exponential.

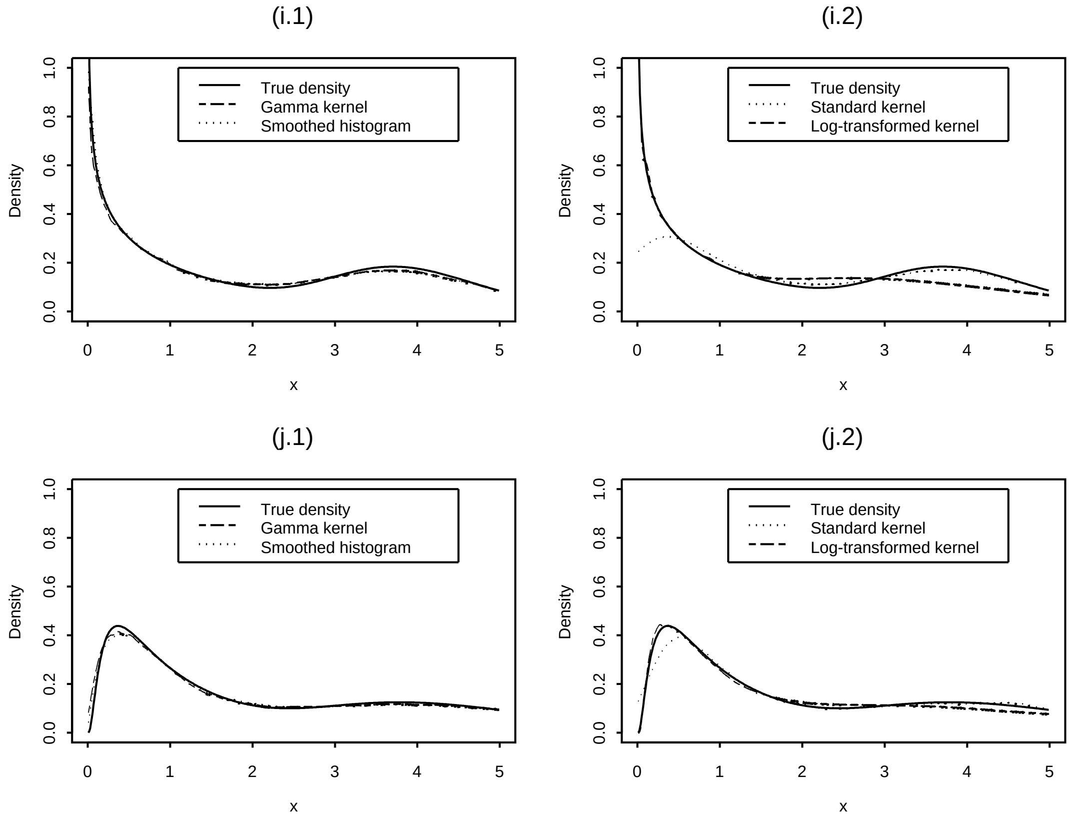
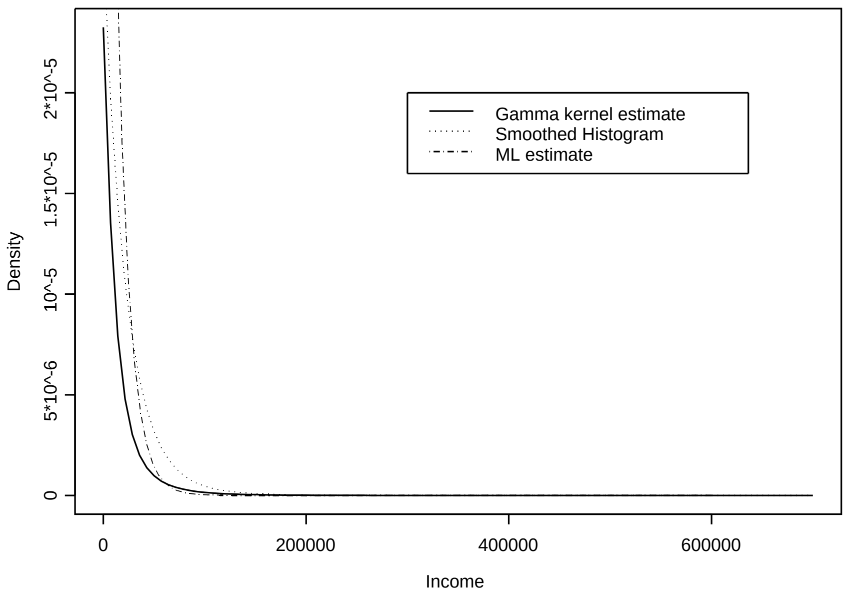


Figure 8: Gamma kernel, smoothed histogram, standard kernel and log-transformed kernel estimates for (i) gamma mixture (j) lognormal mixture.



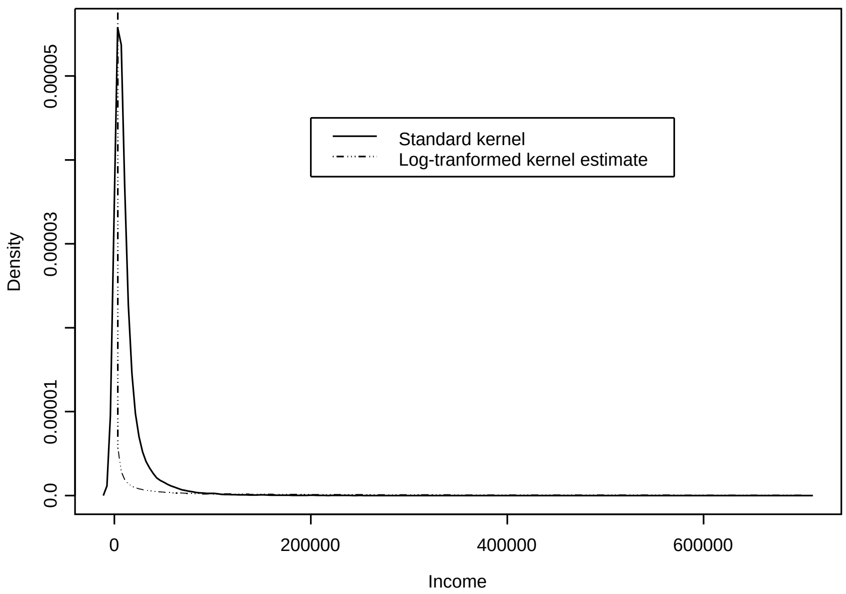


Figure 9: The upper part plots the gamma kernel, smoothed histogram and pseudo maximum likelihood estimates and the lower parts plots the standard kernel and log-transformed kernel estimates for the Brazilian income distribution in 1981