

# Optimal inverse Beta(3,3) transformation in kernel density estimation

Catalina Bolancé\*

*University of Barcelona*

---

## Abstract

---

A double transformation kernel density estimator that is suitable for heavy-tailed distributions is presented. Using a double transformation, an asymptotically optimal bandwidth parameter can be calculated when minimizing the expression of the asymptotic mean integrated squared error of the transformed variable. Simulation results are presented showing that this approach performs better than existing alternatives. An application to insurance claim cost data is included.

---

MSC: 62G07, 62G32

Keywords: Kernel density estimation, transformations, Beta density, right skewness.

## 1. Introduction

Kernel density estimation is nowadays a classical approach to study the form of a density with no assumption on its global functional form.

Let  $X_1, \dots, X_n$  a random sample of *iid* observations of a random variable with density function  $f$ , then the kernel density estimator at point  $x$  is:

$$\hat{f}_c(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - X_i), \quad (1)$$

where  $b$  is the bandwidth or smoothing parameter,  $K_b(t) = \frac{1}{b} K\left(\frac{t}{b}\right)$  and  $K$  is the kernel function, usually it is a symmetric density function bounded or asymptotically bounded

---

\*Dept. of Econometrics, University of Barcelona, Diagonal, 690 08034 Barcelona, Spain.

Email: bolance@eco.ub.es. Support from grant SEJ2007-63298 is acknowledged.

Received: June 2010

Accepted: October 2010

and centred at zero. In this work I use the Epanechnikov kernel, Silverman (1986) proves that this kernel is optimal for kernel density estimator. The Epanechnikov kernel is:

$$k(t) = \begin{cases} 0.75(1-t^2) & \text{si } |t| \leq 1 \\ 0 & \text{si } |t| > 1 \end{cases}$$

Silverman (1986) or Wand and Jones (1995) provide an extensive review of classical kernel estimation. In order to implement kernel density estimation both  $K$  and  $b$  need to be chosen. The optimal choice for the value of  $b$  depends inversely on the sample size, so the larger the sample size, the smaller the smoothing parameter and conversely.

When the shape of the density to be estimated is symmetric and has a kurtosis that is similar to the kurtosis of the normal distribution, then it is possible to calculate a smoothing parameter  $b$  that provides optimal smoothness or is close to optimal smoothness over the whole domain of the distribution. However, when the density is asymmetric, it is not possible to calculate a value for the smoothing parameter which captures both the mode of the density shape and the tail behaviour. In fact, optimal smoothness in the tail is much larger than in the main mode and this is due to the fact that available sampling information in the mode is much more abundant than in the tail of the density, where there are not many observations.

The majority of economic variables that measure expenditures or costs have a strong asymmetric behaviour to the right, so that classical kernel density estimation is not efficient in order to estimate the values of the density in the right tail part of the density domain. This is due to the fact that the smoothing parameter which has been calculated for the whole domain function is too small for the density in the tail. Using a variable bandwidth can be a convenient solution, but this approach has many difficulties as discussed by Jones (1990). Our aim is to propose a double transformation kernel density estimator, where the bandwidth is optimal and can be chosen automatically. The optimal bandwidth has a straightforward expression and it is obtained by minimizing the asymptotic mean integrated squared error.

An alternative to kernel estimation defined in (1) is transformation kernel estimation that is based on transforming the data so that the density of the transformed variable has a symmetric shape, so that it can easily be estimated using a classical kernel estimation approach. We say it can be easily estimated in the sense that using a Gaussian kernel or an Epanechnikov kernel, an optimal estimate of the smoothing parameter can be obtained by minimizing an error measure over the whole density domain. In the specialized literature several transformation kernel estimators have been proposed, and their main difference is the type of transformation family that they use. For instance, Wand *et al.* (1991), Bolancé *et al.* (2003), Clements *et al.* (2003) and Buch-Larsen *et al.* (2005) propose different parametric transformation families that they all make the transformed distribution more symmetric than the original one, which in many applications has usually a strong right-hand asymmetry. Also Bolancé *et al.* (2008) used

the transformation kernel estimation to approximate the conditional tail expectation risk measure.

Given a density estimator  $\hat{f}$  of a density  $f$ , the Mean Integrated Squared Error (MISE) is defined as:

$$MISE(\hat{f}) = E \left( \int_{-\infty}^{+\infty} (\hat{f}(t) - f(t))^2 dt \right).$$

Let  $T(\cdot)$  a concave transformation, the transformed sample is  $Y_1 = T(X_1), \dots, Y_n = T(X_n)$ , the classical kernel estimator of the transformed variable is:

$$\hat{f}_c(y) = \frac{1}{n} \sum_{i=1}^n K_b(y - Y_i) = \frac{1}{n} \sum_{i=1}^n K_b(T(x) - T(X_i)) \quad (2)$$

and the transformation kernel estimator of the original variable is:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_b(T(x) - T(X_i)) T'(x). \quad (3)$$

Wand *et al.* (1991) show that there exists a relationship between the value of *MISE* obtained for the classical kernel estimator of the transformed variable and the *MISE* obtained with the transformation kernel estimator of the original variable. They also show that there exists an optimal transformation that minimizes both expressions.

Based on the work by Buch-Larsen *et al.* (2005), Bolancé *et al.* (2008) proposed a double transformation with the purpose of obtaining a transformed variable whose density is as close as possible to a density that maximizes smoothness  $\int \{f''(x)\}^2 dx$  and at the same time that minimizes the asymptotic Mean Integrated Squared Error ( $A - MISE$ ) of the kernel estimator defined in (1) and obtained with the transformed observations. Terrell and Scott (1985) showed that among the vast family of densities with domain  $D$  that have a Beta distribution, one of them has the largest possible smoothness.

Since the density of a Beta distribution in the bounds of its domain is zero, the bias of kernel estimation near the boundaries of the domain is strictly positive, and therefore this implies a larger bias in the transformation kernel estimation in the extremes of the density of the original variable (in the right tail and in the values near the minimum). In order to correct for this positive bias, Bolancé *et al.* (2008) proposed to transform their data into a new set of data so that they have a density that is similar to the Beta density in a domain in the interior of  $D$ . Then they correct the resulting density estimate so that it integrates to one, but in their contribution they do not indicate how to optimize this second transformation. In the next section, a method based on minimizing  $A - MISE$  is proposed. One of its main features is that it can become fully automated, which is very suitable for practical applications.

Let  $g(\cdot)$  and  $G(\cdot)$  be the density and distribution functions of a Beta random variable, which we denote by  $B(\beta, \beta)$  with domain in  $[-\alpha, \alpha]$ , if  $Z$  is a random variable with a uniform distribution, then  $Y = G^{-1}(Z)$  is a random variable with distribution  $B(\beta, \beta)$ . The method proposed by Bolancé *et al.* (2008) suggests to do a first transformation on the original sample of observations  $X_1, \dots, X_n$  so that  $Z_i = T(X_i)$ ,  $i = 1, \dots, n$ . If  $T(\cdot)$  is a cumulative distribution function then  $Z_i$ ,  $i = 1, \dots, n$  can be a sample of independent observations that are close to have been generated by a uniform distribution. Then they define  $l$  as a probability close to 1, namely 0.98 or 0.99, so that  $\tilde{T}(X_i) = \tilde{Z}_i = (2l - 1)Z_i + (1 - l)$  and, therefore, the density that is associated with the data generating process  $Y_i = G^{-1}(\tilde{Z}_i)$  coincides with the density function of a Beta density,  $B(\beta, \beta)$  in a domain  $[-a, a]$ , where  $\alpha > a = G^{-1}(l)$ . Then, the resulting transformation kernel estimator, where  $'$  denotes the first derivative, is:

$$\begin{aligned} \hat{f}(x) &= \frac{1}{(2l-1)n} \sum_{i=1}^n K_b(G^{-1}(\tilde{T}(x)) - G^{-1}(\tilde{T}(X_i))) (G^{-1})'(\tilde{T}(x)) \tilde{T}'(x) \\ &= \frac{1}{n} \sum_{i=1}^n K_b(G^{-1}(\tilde{T}(x)) - G^{-1}(\tilde{T}(X_i))) (G^{-1})'(\tilde{T}(x)) T'(x). \end{aligned} \quad (4)$$

We note that the optimality of (4) depends on whether the first transformation  $T(\cdot)$  is successfully transforming the data into a sample that is likely to have been generated by a Uniform(0, 1). It is obvious that the transformation  $T(\cdot)$  must be a distribution function. Bolancé *et al.* (2008) propose to use the generalized Champernowne cdf:

$$T_{\alpha, M, c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \quad x \geq 0, \quad (5)$$

with parameters  $\alpha > 0$ ,  $M > 0$  and  $c \geq 0$ , that can be estimated by maximum likelihood. This is certainly a flexible distribution, because it can have many shapes near zero and also different behaviours in the tail. Degen and Embrechts (2008) analyzed the tail modified Champernowne distribution convergence to the tail behaviour supposed by extreme value theory, and they concluded that convergence is stronger if we compare it to the tail distribution for the Loggamma, the g-and-h and the Burr and lighter if we compare it to the Generalized Beta distribution (GB2).

In this work we propose a method to find an asymptotically optimal value for  $l$ , that is obtained when one finds the Beta truncated distribution with density  $\frac{g(\cdot)}{(2l-1)}$ , defined on  $[-a, a]$ , with  $a = G^{-1}(l)$ , whose kernel estimation minimizes *MISE* asymptotically. This result is developed in Section 2. Section 3 presents the results of a simulation study that uses the same samples as in Buch-Larsen *et al.* (2005) and in Bolancé *et al.* (2008). By means of the results of the simulation we analyze the behaviour of the estimation method that is being proposed and we see that the value of the optimal choice for  $l$  considerably reduces the distance between the true theoretical density and the density

estimate for all the asymmetric shapes that have been analyzed and, in many cases, also if the sample size is small. In Section 4 we show an application to data on costs arising from automobile insurance claims. These data were also used by Bolancé *et al.* (2009). Finally, in Section 5 we conclude.

## 2. Asymptotically optimal truncated inverse Beta transformation

Terrell and Scott (1985, Lemma 1) showed that  $B(3,3)$  defined on the domain  $(-1/2, 1/2)$  has  $\int \{g''(t)\}^2 dt$  minimal within the set of Beta densities with same support, where  $g(\cdot)$  is the pdf and is given by:

$$g(t) = \frac{15}{8} (1 - 4t^2)^2, \quad -\frac{1}{2} \leq t \leq \frac{1}{2} \quad (6)$$

and  $G(\cdot)$  is the cdf and is given by:

$$G(t) = \frac{1}{8} (4 - 9t + 6t^2) (1 + 2t)^3. \quad (7)$$

Using the Epanechnikov kernel for the upper bound (or the lower bound since the domain of the distribution  $B(3,3)$  is symmetric) the expectation of the classical kernel estimation is (see, Wand and Jones 1995, p. 47):

$$\begin{aligned} \int_{-1}^0 K(t) g\left(\frac{1}{2} - bt\right) dt &= \int_{-1}^0 \frac{3}{4} (1 - (t)^2) \frac{15}{8} \left(1 - 4\left(\frac{1}{2} - bt\right)^2\right)^2 dt \\ &= 1.2857b^4 + 3.75b^3 + 3b^2 > 0 \text{ if } b > 0. \end{aligned} \quad (8)$$

The value of the density defined in (6) in the boundaries of the domain is zero, however, as we have noted in (8), the value of the classical kernel estimation of the density is positive  $\forall b > 0$ , and therefore  $\hat{f}_c(x)$  over-estimates the beta density in the tails.

Silverman (1986) shows that asymptotically the MISE for (1) is:

$$A - MISE \{ \hat{f}_c \} = \frac{1}{4} b^4 k_2^2 \int f''(x)^2 dx + \frac{1}{nb} \int K(t)^2 dt,$$

where  $k_2 = \int t^2 K(t) dt$ . The asymptotically optimal bandwidth is:

$$b^{opt} = \left( \frac{\int K(t)^2 dt}{k_2^2 \int f''(x)^2 dx} \right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

replacing  $b^{opt}$  in  $A - MISE \{ \hat{f}_c \}$  we obtain the value of  $A - MISE$  for the asymptotically optimal bandwidth:

$$A - MISE^* (\hat{f}_c) = \frac{5}{4} k_2^{\frac{2}{5}} \left( \int K(t)^2 dt \right)^{\frac{4}{5}} \left( \int f''(x)^2 dx \right)^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (9)$$

Let  $Y$  be a transformed random variable with distribution  $B(3, 3)$ . Let  $\frac{g(y)}{2l_a-1}$ , with  $l_a = G(a)$ , be the truncated Beta density in the domain  $[-a, a]$ . If one just uses the same development that is being used to obtain (9), a value for  $A - MISE^* (\hat{f}_c(x), a)$  can easily be obtained. Replacing in Silverman's  $A - MISE$  proof  $g(y)$  by  $\frac{g(y)}{2l_a-1}$  we obtain:

$$A - MISE \{ \hat{g}_c, a \} = \frac{1}{4} b^4 \frac{k_2^2}{(2l_a - 1)^2} \int_{-a}^{+a} g''(x)^2 dx + \frac{1}{nb} \int K(t)^2 dt,$$

then

$$b^{opt}(a) = \left( \frac{\int K(t)^2 dt}{\frac{k_2^2}{(2l_a-1)^2} \int_{-a}^{+a} g''(x)^2 dx} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \quad (10)$$

and replacing  $b^{opt}(a)$  in  $A - MISE \{ \hat{f}_c, a \}$  we obtain:

$$A - MISE^* \{ \hat{g}_c, a \} = \frac{5}{4} k_2^{\frac{2}{5}} \left( \int K(t)^2 dt \right)^{\frac{4}{5}} (2l_a - 1)^{-\frac{2}{5}} \left( \int_{-a}^{+a} g''(x)^2 dx \right)^{\frac{1}{5}} n^{-\frac{4}{5}}.$$

We then analyze the behaviour of  $A - MISE^* \{ \hat{g}_c, a \}$  as a function of  $a$  in order to estimate the truncated density  $\frac{g(y)}{2l_a-1}$  whenever the objective is that the distribution of the transformed variable is  $B(3, 3)$ . Using the Epanechnikov's kernel  $K(t) = \frac{3}{4} (1 - t^2)$ ,  $|t| \leq 1$  for the density of a  $B(3, 3)$  we obtain:

$$A - MISE^* \{ \hat{g}_c, a \} = \frac{5}{4} \left( \frac{9}{125} \right)^{\frac{2}{5}} \left( \frac{360a(-40a^2 + 144a^4 + 5)}{\left( \frac{1}{4}a(-40a^2 + 48a^4 + 15) \right)^2} \right)^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (11)$$

If we also analyze the shape of expression (11), we observe that there exists a value of  $a$  that minimizes the corresponding expression for  $A - MISE^*$ . In Figure 1 we show a plot of (11) as a function of  $a$ , where we have eliminated the effect of the sample size factor ( $n^{-\frac{4}{5}}$ ).

As a result, there exists a truncated density  $\frac{g(y)}{2l_{a^*}-1}$  that depends on an optimal  $a$  which is related to  $B(3, 3)$  that minimizes (11). The objective of our proposed transformation kernel estimation method is to obtain a sample of transformed observations whose

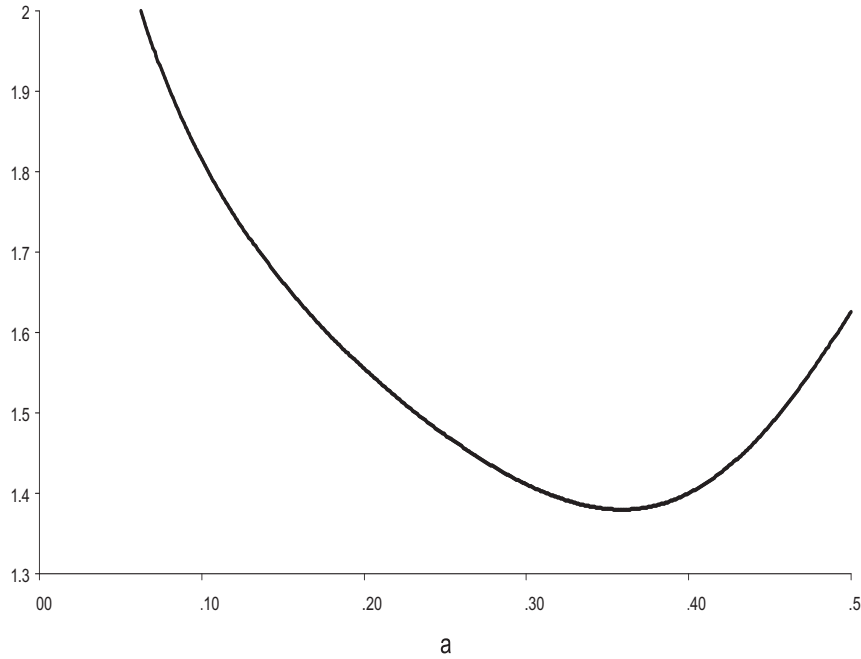


Figure 1:  $A - MISE_{B(3,3)}^{*} \{\hat{g}_c, a\} n^{\frac{4}{5}}$  vs  $a$ .

density is as close as possible to an optimally truncated Beta density, so that the optimality of the kernel estimation of the transformed variable is transferred to an optimal transformation kernel estimation of the original variable. Then we propose:

$$\begin{aligned} \hat{f}^*(x) &= \frac{1}{n} \frac{\sum_{i=1}^n K_b(G^{-1}(\tilde{T}^*(x)) - G^{-1}(\tilde{T}^*(X_i))) (G^{-1})'(\tilde{T}^*(x)) \tilde{T}^{*'}(x)}{(2l_{a^*} - 1)} \\ &= \frac{1}{n} \sum_{i=1}^n K_b(G^{-1}(\tilde{T}^*(x)) - G^{-1}(\tilde{T}^*(X_i))) (G^{-1})'(\tilde{T}^*(x)) T'(x) \end{aligned} \quad (12)$$

where  $\tilde{T}^*(X_i) = \tilde{Z}_i^* = (2l_{a^*} - 1)Z_i + (1 - l_{a^*})$ . Holding  $n$  fixed, when we minimize (11) we obtain an optimal  $a$ , which we call  $a^*$  equal to 0.389121. Therefore,  $l_{a^*} = G(0.389121) = 0.98854$ . We call the estimator defined in (12) optimal double transformation kernel density estimator or optimal Kernel Inverse Beta Modified Champernowne Estimator (KIBMCE) if we use the same name given in Bolancé *et al.* (2008).

In order to obtain the estimator in (12) the procedure is:

1. With the sample of observations  $X_1, \dots, X_n$  we estimate parameters  $\alpha$ ,  $M$  and  $c$  of the generalized Champernowne by maximum likelihood (see, for instance, Burch-Larsen *et al.*, 2005) and calculate (5)  $Z_i = T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i)$  and  $\tilde{T}^*(X_i) = \tilde{Z}_i^* = (2 \cdot 0.98854 - 1)Z_i + (1 - 0.98854)$ .

2. Calculate  $Y_i = G^{-1}(\tilde{T}^*(X_i))$  and obtain the classical kernel estimator  $\hat{f}_c(y)$  defined in (1). The smoothing parameter  $b^*$  is estimated by the value that is asymptotically optimal when estimating a  $B(3,3)$  on the domain  $(-a^*, a^*)$ , and therefore its expression is:

$$\begin{aligned} b^* &= k_2^{-\frac{2}{5}} \left( \int_{-1}^1 K(t)^2 dt \int_{-a^*}^{a^*} g(y) dy \right)^{\frac{1}{5}} \left( \int_{-a^*}^{a^*} \{g''(y)\}^2 dy \right)^{-\frac{1}{5}} n^{-\frac{1}{5}} \\ &= 0.5416079 n^{-\frac{1}{5}}. \end{aligned} \quad (13)$$

The difference between the smoothing parameter  $b^{opt}(a)$  in (10) and  $b^*$  in (13) is that first is optimal for the classical kernel estimation of truncate Beta density and second is optimal for the classical kernel estimation of Beta density in  $[-a^*, a^*]$ .

3. Obtain the optimal double transformation kernel estimator in (12) as:

$$\hat{f}^*(x) = \hat{f}_c(y) (G^{-1})'(\tilde{T}^*(x)) T'(x).$$

It is obvious that the estimator in (12) is optimal if the transformed random variable  $Z = T(X)$  is distributed as a Uniform(0, 1), and this certainly depends on the quality of the generalized Champernowne cdf defined in (5) and how well it approximates the original variable. This is going to be discussed in the next section, where simulation results are also shown.

Next we are going to present a simulation study where we show to what extend, for finite sample, and with the transformation kernel estimation expressed in (12) the results shown in Buch-Larsen *et al.* (2005) can be improved. Therefore it also improves Wand *et al.* (1991) and Clements *et al.* (2003).

### 3. Simulation study

This section presents a comparison of our inverse beta double transformation method with the results presented by Buch-Larsen *et al.* (2005) based only on the modified Champernowne distribution. Our objective is to show that the second transformation, that is based on the inverse of a Beta optimal truncated distribution, improves density estimation for a wide range of asymmetric densities that are commonly found in practice.

In this work we analyze the same simulated samples as in Buch-Larsen *et al.* (2005) and Bolancé *et al.* (2008), which were drawn from four distributions with different tails and different shapes near 0. The distributions and the chosen parameters are listed in Table 1.



**Table 1:** Distributions in simulation study.

Distribution	Density	Parameters
Mixture of $p$ Lognormal $(\mu, \sigma)$ and $(1 - p)$ Pareto $(\lambda, \rho, c)$	$f(x) = p \frac{1}{\sqrt{2\pi\sigma^2}x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} + (1 - p)(x - c)^{-(\rho+1)} \rho \lambda^\rho$	$(p, \mu, \sigma, \lambda, \rho, c)$ $= (0.7, 0, 1, 1, 1, -1)$ $= (0.3, 0, 1, 1, 1, -1)$ $= (0.1, 0, 1, 1, 1, -1)$ $= (0.9, 2.5, 0.5, 1, 1, -1)$
Lognormal $(\mu, \sigma)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$	$(\mu, \sigma) = (0, 0.5)$
Weibull $(\gamma)$	$f(x) = \gamma x^{(\gamma-1)} e^{-x^\gamma}$	$\gamma = 1.5$
Truncated logistic	$f(x) = \frac{2}{s} e^{\frac{x}{s}} \left(1 + e^{\frac{x}{s}}\right)^{-2}$	$s = 1$

In Figure 2 we present the result of the ratio between the distribution function  $F(x)$  that is associated to each of the densities in Table 1 and the Champernowne distribution  $T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)$  that is estimated by means of a sample with size 1000, obtained from each of the five distribution. The right-hand plots focus on the ratio in the tail. In Table 2, we show the distance measures  $L_1$  and  $L_2$  between  $F(x)$  and  $T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)$ :

$$L_1(F, T_{\hat{\alpha}, \hat{M}, \hat{c}}) = \int_{-\infty}^{+\infty} |T_{\hat{\alpha}, \hat{M}, \hat{c}}(t) - F(t)| dt$$

and

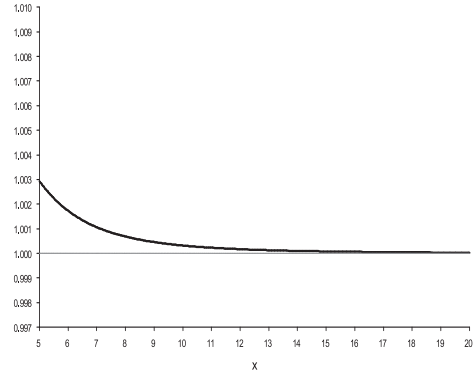
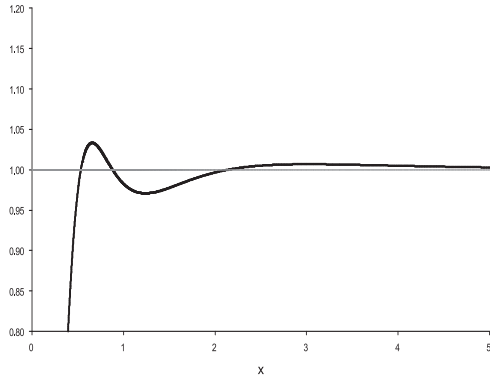
$$L_2(F, T_{\hat{\alpha}, \hat{M}, \hat{c}}) = \int_{-\infty}^{+\infty} (T_{\hat{\alpha}, \hat{M}, \hat{c}}(t) - F(t))^2 dt.$$

**Table 2:** Distance between the true distribution and the Champernowne distribution.

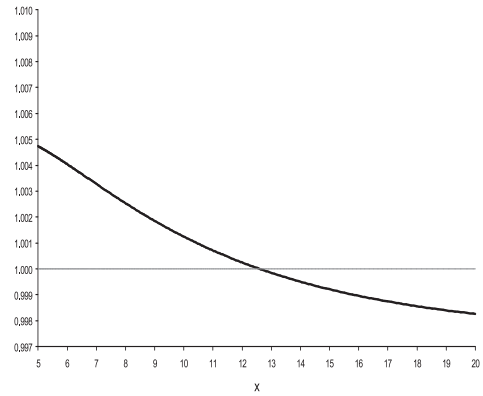
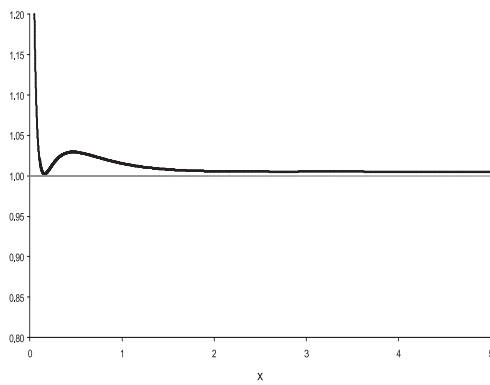
	Lognormal	Log-Pareto		Weibull	Tr. Logist.
		$p = 0.7$	$p = 0.3$		
$L_1$	0.0445	1.4423	2.1270	0.0422	0.0940
$L_2$	0.0225	0.0409	0.0544	0.0240	0.0343

It is obvious that the improvement in the KIBMCE method with respect to the Kernel Modified Champernowne Estimator (KMCE) proposed by Buch-Larsen *et al.* (2005) is larger in those cases where the shape of the true cdf is similar to the Champernowne. In

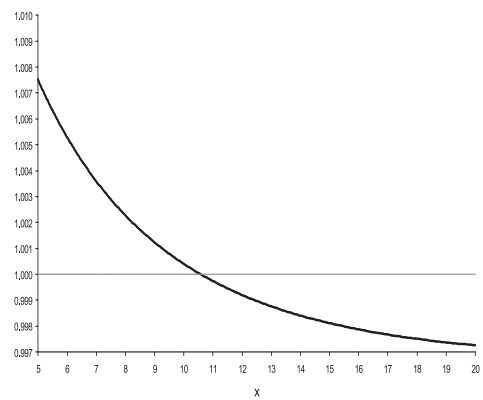
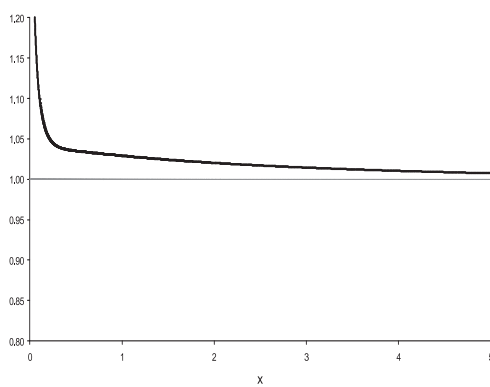
## a) Lognormal



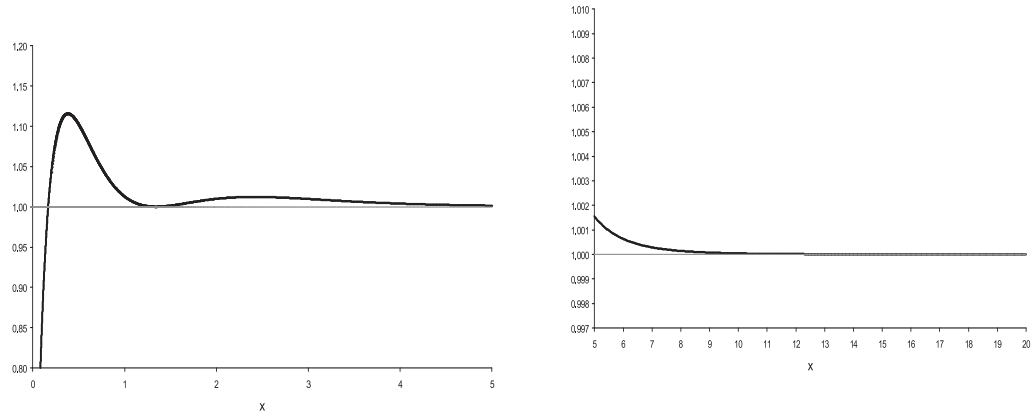
## b) 70% Lognormal-30% Pareto



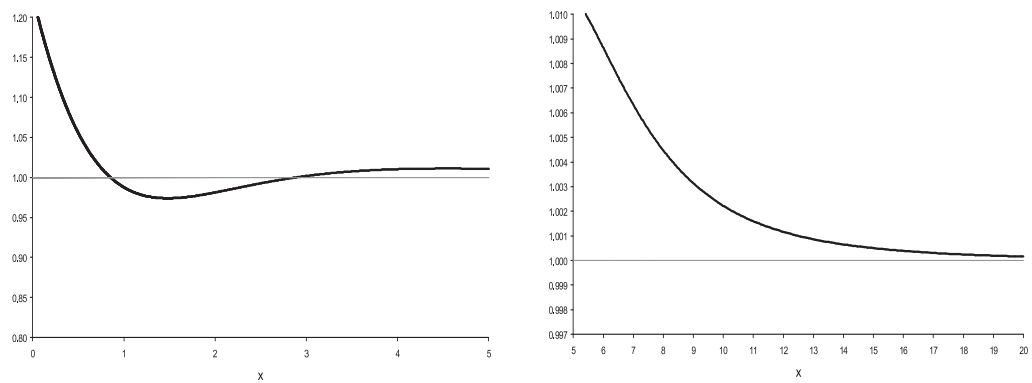
## c) 30% Lognormal-70% Pareto



## d) Weibull



## e) Truncated Logistic



**Figure 2:** Ratio of  $F(x)$  and  $T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)$  in the  $(0, 5)$  domain interval on the left and in the  $(5, 20)$  domain interval on the right, for five distributions given in Table 1.

the case of a mixture between a lognormal and a Pareto, Figures 2b and 2c show that the Champernowne distribution tends more rapidly to one than the true cdf. and this can also be seen when looking at the values of the  $L_1$  distance between the two functions. The results of Figure 1 and Table 2 show us that the improvement in KIBMCE is larger in the estimation of a density that has a Lognormal, a Weibull and a Truncated Logistic shape.

Buch-Larsen *et al.* (2005) evaluate the performance of the KMCE estimators compared to the estimator described by Clements *et al.* (2003) the estimator described by Wand *et al.* (1991) and the estimator described by Bolancé *et al.* (2003). The Champernowne transformation substantially improve the results from previous authors. Bolancé *et al.* (2008, 2009) compare his truncated inverse beta second transformation with

**Table 3:** The estimated error measures for KMCE and KIBMCE.

			Lognormal	Log-Pareto		Weibull	Tr. Logist.	
			$l_{a^*}$	$p = 0.7$	$p = 0.3$			
N = 100	$L_1$	KIBMCE	0.9885	0.1348	0.1240	0.1202	0.1391	0.1246
		KMCE		0.1363	0.1287	0.1236	0.1393	0.1294
	$L_2$	KIBMCE	0.9885	0.1001	0.0851	0.0853	0.1095	0.0739
		KMCE		0.1047	0.0837	0.0837	0.1084	0.0786
	WISE	KIBMCE	0.9885	0.0992	0.0819	0.0896	0.0871	0.0969
		KMCE		0.1047	0.0859	0.0958	0.0886	0.0977
N = 1000	$L_1$	KIBMCE	0.9885	0.0561	0.0480	0.0471	0.0589	0.0510
		KMCE		0.0659	0.0530	0.0507	0.0700	0.0598
	$L_2$	KIBMCE	0.9885	0.0405	0.0362	0.0381	0.0482	0.0302
		KMCE		0.0481	0.0389	0.0393	0.0582	0.0339
	WISE	KIBMCE	0.9885	0.0404	0.0359	0.0402	0.0378	0.0408
		KMCE		0.0481	0.0384	0.0417	0.0450	0.0501

**Table 4:** Ratio between the error measures of KIBMCE and KMCE.

		Lognormal	Log-Pareto		Weibull	Tr. Logist.
			$p = 0.7$	$p = 0.3$		
N = 100	$L_1$	0.9888	0.9637	0.9725	0.9982	0.9629
	$L_2$	0.9563	1.0162	1.0190	1.0100	0.9398
	WISE	0.9470	0.9538	0.9350	0.9835	0.9916
N = 1000	$L_1$	0.8517	0.9059	0.9295	0.8413	0.8522
	$L_2$	0.8410	0.9295	0.9695	0.8284	0.8911
	WISE	0.8390	0.9340	0.9637	0.8392	0.8150

$l = 0.99$  and  $l = 0.98$  with Buch-Larsen method and shows that double-transformation method improves the results presented in Buch-Larsen *et al.* (2005). In this work, we compare the results obtained when using an optimal trimming parameter  $l_{a^*}$  for  $B(3, 3)$ .

We measure the performance of the estimators by the error measures based in  $L_1$  norm,  $L_2$  norm and WISE. This last weighs the distance between the estimated and the true distribution with the squared value of  $x$ . This results in an error measure that emphasizes the tail of the distribution:

$$\left( \int_0^{\infty} (\hat{f}(x) - f(x))^2 x^2 dx \right)^{1/2}.$$

The simulation results can be found in Table 3. For every simulated density and for sample sizes  $N = 100$  and  $N = 1000$ , the results presented here correspond to the following error measures  $L_1$ ,  $L_2$  and  $WISE$ . The benchmark results are labeled KMCE and they correspond to those presented in Buch-Larsen *et al.* (2005). In Table 4 we show ratios between the error measures of KIBMCE and KMCE, if this ratio is smaller than 1 then KIBMCE improves on the results of KMCE.

In Table 4 we show that for  $N = 100$  the ratios associated to  $L_1$  and  $WISE$  are always below one, and this indicates the KIBMCE method improves the fit of the density in the tail values of the density even for a small sample size. When  $N = 1000$  then KIBMCE has always smaller values than KMCE, both for  $L_1$ ,  $L_2$  and for  $WISE$ . The best results can be obtained for the Lognormal, the Weibull and the Truncated Logistic, where in all cases the errors of KMCE are reduced by more than a 10% when using the new method.

For the mixtures of a Lognormal and a Pareto the results show that when  $p = .7$  (70% Lognormal) the improvement is almost 10% for  $L_1$  and is around 7% for  $L_2$  and  $WISE$ . When  $p = .3$  (30% Lognormal)  $L_1$  is reduced by 7%, and both  $L_2$  and  $WISE$  are reduced in slightly more than 3%.

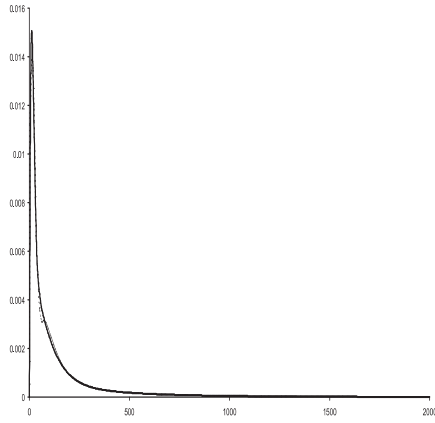
#### 4. Data analysis

In this section, we apply our estimation method to a data set that contains automobile claim costs from a Spanish insurance company for accidents occurred in 1997. It is a typical insurance cost of individual claims data set, i.e. a large sample that looks heavy-tailed. The data are divided into two age groups: claims from policyholders who are less than 30 years old, and claims from policyholders who are 30 years old or older. The first group consists of 1,061 observations in the interval  $[1;126,000]$  with mean value 402.70. The second group contains 4,061 observations in the interval  $[1;17,000]$  with mean value 243.09. Estimation of the parameters in the modified Champernowne distribution function for the two samples of is, for young drivers  $\hat{\alpha}_1 = 1.116$ ,  $\hat{M}_1 = 66$ ,  $\hat{c}_1 = 0.000$  and for older drivers  $\hat{\alpha}_2 = 1.145$ ,  $\hat{M}_2 = 68$ ,  $\hat{c}_2 = 0.000$ , respectively. We notice that  $\alpha_1 < \alpha_2$ , which indicates that the data set for young drivers has a heavier tail than the data set for older drivers.

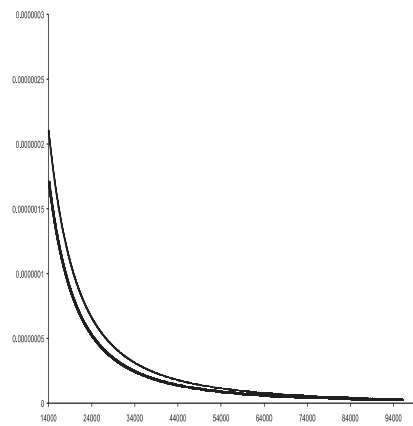
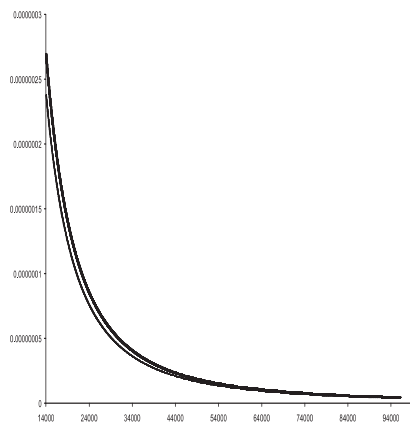
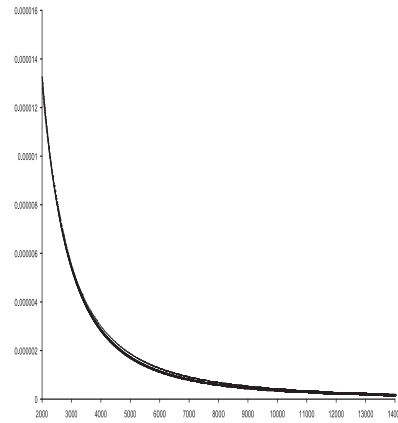
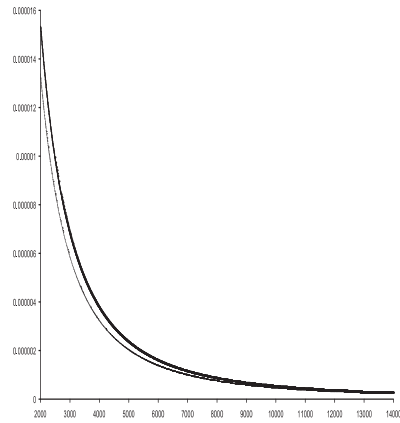
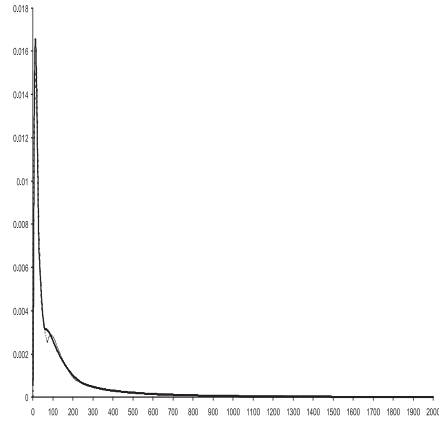
To produce the graphics, the claims have been split into three categories: *Small claims* in the interval  $(0; 2,000)$ , *moderately sized claims* in the interval  $[2,000; 14,000)$ , and *extreme claims* in the interval  $[14,000; \infty)$ . In Figure 3 we show the density function in the three categories for younger and older drivers.

Figure 3 shows that the KIBMCE method corrects the results of KMCE. In general, the density that is estimated using a KIBMCE method is smoother and larger in the mode, if compared with the KMCE estimate. In the tail and compared to the KMCE, the KIBMCE estimates a larger density in the tail, when the tails heavier, as for younger drivers, and it also estimates a smaller density when the tail is lighter, as for older drivers.

a) Younger drivers



b) Older drivers



**Figure 3:** Optimal KIBMCE estimates (thick) versus KMCE estimates (light) of insurance claims cost densities. Upper plots show small claims, middle plots show moderate claims and lower plots show large claims.

## 5. Conclusions

In this work we have proposed a transformation kernel density estimator that can provide good results when the density to be estimated is very asymmetric and has extreme values. Moreover, the method presented here has a very straightforward method to calculate the smoothing parameter. This method provides a rule of thumb method to calculate the bandwidth in the context of transformation kernel density estimation that is comparable to Silverman's rule of thumb in the context of classical kernel density estimation. For large sample sizes, like the ones shown in the application, the simulation study shows that this method outperforms existing alternatives.

## References

- Bolancé, C., Guillén, M. and Nielsen, J. P. (2009). Transformation kernel estimation of insurance claim cost distribution, in Corazza, M. and Pizzi, C. (Eds). *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, Springer, 223-231.
- Bolancé, C., Guillén, M. and Nielsen, J. P. (2008). Inverse Beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78, 1757-1764.
- Bolancé, C., Guillén, M. and Nielsen, J. P. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, 32, 19-36.
- Bolancé, C., Guillén, M., Pelican, E. and Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for CTE risk measure. *Insurance: Mathematics and Economics*, 43, 386-393.
- Buch-Larsen, T., Guillen, M., Nielsen, J. P. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39, 503-518.
- Clements, A. E., Hurn, A. S. and Lindsay, K. A. (2003). Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, 98, 993-1000.
- Degen, M. and Embrechts, P. (2008). EVT-based estimation of risk capital and convergence of high quantiles. *Advances in Applied Probability*, 40, 696-715.
- Jones, M. C. (1990). Variable kernel density estimation and variable kernel density estimation. *Australian Journal of Statistics*, 32, 361-371.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, 85, 270-277.
- Terrell, G. R. and Scott, D. W. (1985). Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, 80, 209-214.
- Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. London: Chapman & Hall.
- Wand, P., Marron, J. S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86, 414, 343-361.

