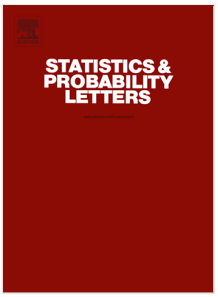
Inverse beta transformation in kernel density estimation

Catalina Bolancé, Montserrat Guillén, Jens Perch Nielsen

Accepted Manuscript



Inverse beta transformation in kernel density estimation

Catalina Bolancé, Montserrat Guillén, Jens Perch Nielsen

PII: S0167-7152(08)00037-0
DOI: 10.1016/j.spl.2008.01.028
Reference: STAPRO 4901

To appear in: *Statistics and Probability Letters*

Received date: 28 December 2007
Accepted date: 10 January 2008

Please cite this article as: Bolancé, C., Guillén, M., Nielsen, J.P., Inverse beta transformation in kernel density estimation. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.01.028

Inverse Beta Transformation in Kernel Density Estimation

Catalina Bolancé [a], Montserrat Guillén [a,*], Jens Perch Nielsen [b]

[a] *Dept. of Econometrics, RFA-IREA. University of Barcelona. Diagonal 690 08034 Barcelona, Spain.*

[b] *Cass Business School, City University London, 106 Bunhill Row London EC1Y 8TZ, UK.*

Abstract

A transformation kernel density estimator that is suitable for heavy-tailed distributions is presented. Using a double transformation, the choice of the bandwidth parameter becomes straightforward. An illustration and simulation results are presented.

Key words: Heavy-tailed distribution, bandwith choice

1 Introduction

Kernel density estimation of heavy-tailed distributions has been studied by several authors (Buch-Larsen et al., 2005, Clements et al., 2003 and Bolancé et al., 2003). They have all proposed estimators based on a transformation of the original variable. The transformation method proposed initially by Wand et al. (1991) is very suitable for asymmetrical variables, it was based on the shifted power transformation family. Some alternative transformations such as the one based on a generalization of the Champernowne distribution have been analyzed and simulation studies have shown that it is preferable to other transformation density estimation approaches for distributions that are Pareto-like in the tail. In the existing contributions, the choice of the bandwidth parameter in transformation kernel density estimation is still a problem, so we now aim

* Corresponding author. We thank the Spanish Ministry of Education and Science SEJ2007-63298 / FEDER grant.

Email addresses: bolance@ub.edu (Catalina Bolancé), mguillen@ub.edu (Montserrat Guillén), festinalente@nielsen.mail.dk (Jens Perch Nielsen).

at finding a new estimator for which the bandwidth choice is straightforward. One way of undergoing bandwidth choice is to implement the transformation approach so that transformation leads to a beta distribution, then use existing theory to optimize bandwidth parameter choice on beta distributed data and backtransform to the original scale. The main drawback is that the beta distribution may be very steep in the domain boundary, which causes numerical instability when the derivative of the inverse distribution function is needed for the backward transformation. In this work we propose to truncate the beta distribution and use the truncated version at transformation kernel density estimation. Thus, the derivative is smaller and can be inverted without causing numerical instability. The results on the optimal choice of the bandwidth for kernel density estimation of beta density are used in the truncated version directly. In the simulation study we see that our approach produces very good results for heavy-tailed data. Our results are particularly relevant for applications in insurance, where the claims amounts are analyzed and usually small claims (low cost) coexist with only a few large claims (high cost).

An alternative way to obtain a kernel density estimator for heavy-tailed distributions is using the variable kernel estimator, which consists in selecting a different bandwidth parameter depending on the point where the density is being estimated. Several authors have studied the variable kernel estimator. For instance, Jones (1990) made an extensive revision of this method and more recently Wu et al. (2007) proposed a multivariate version. In our work we study transformation kernel estimation, which is equivalent to variable kernel estimation where we guarantee that the resulting density integrates to one.

Let f be a density function, Terrell (1990) and previously Terrell and Scott (1985) analyzed several density families that minimize functionals such as $\int \left\{ f^{(p)}(x) \right\}^2 dx$, where superscript (p) refers to the p-th derivative of the density function. We will use these families in the context of transformed kernel density estimation. The results on those density families are very useful to improve the properties of the transformation kernel density estimator.

Given a sample $X_1, ..., X_n$ of independent and identically distributed (iid) observations with density function f, the classical kernel density estimator is:

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

where b is the bandwidth or smoothing parameter and $K_b(t) = K(t/b)/b$ is the kernel. In Silverman (1986) or Wand and Jones (1995) one can find an extensive revision of classical kernel density estimation.

An error distance between the estimated density $\hat{f}$ and the theoretical density f that has widely been used in the analysis of the optimal bandwidth b is the

mean integrated squared error ($MISE$):

$$E\left\{\int\left(f\left(x\right)-\hat{f}\left(x\right)\right)^{2}dx\right\}.$$

It has been shown (see, for example, Silverman, 1986, chapter 3) that the $MISE$ is asymptotically equivalent to $A-MISE$:

$$\frac{1}{4}b^{4}\left(k_{2}\right)^{2}\int\left\{f''\left(x\right)\right\}^{2}dx+\frac{1}{nb}\int K\left(t\right)^{2}dt,$$

where $k_2 = \int t^2 K(t)\,dt$. If the second derivative of f exists (and we denote it by f''), then $\int\{f''(x)\}^2\,dx$ is a measure of the degree of smoothness because the smoother the density, the smaller this integral is. From the expression for $A-MISE$ it follows that the smoother f, the smaller the value of $A-MISE$.

We organize our presentation as follows. In section 2 we propose the Champernowne-inverse beta transformation in kernel density estimation. In section 3 we describe the estimation procedure. Finally, simulation results and a data study are shown.

2 Champernowne-inverse beta transformation in kernel density estimator

Let us consider the beta density function $Beta(\alpha,\beta)$ defined on $[\theta,\theta+D]$, parametrized as follows (Johnson et al., 1995):

$$\frac{\left(x-\theta\right)^{\alpha-1}\left(\theta+D-x\right)^{\beta-1}}{B\left(\alpha,\beta\right)D^{\alpha+\beta-1}},\theta\leq x\leq\theta+D,$$

where θ is a location parameter, D is a scale parameter, α and β are shape parameters, $B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ and $\Gamma(\cdot)$ is the Euler Gamma function.

Terrell (1990) showed that the beta distribution with parameters $\theta=-1$, $D=2$, $\alpha=\beta=4$ minimizes $\int\{f''(x)\}^2\,dx$ within the set of all densities with a given known variance. The $Beta(4,4)$ distribution will be used throughout our work. Its pdf and cdf are:

$$\begin{aligned}g\left(x\right)&=\frac{35}{32}\left(1-x^{2}\right)^{3},-1\leq x\leq1,\\G\left(x\right)&=\frac{1}{32}\left(16-29x+20x^{2}-5x^{3}\right)\left(x+1\right)^{4}.\end{aligned}\tag{2}$$

Terrell and Scott (1985, Lemma 1) showed that $Beta\,(3,3)$ defined on the domain $(-1/2, 1/2)$ minimizes the functional $\int \{f''(x)\}^2 dx$ within the set of beta densities with same support.

We assume that a transformation exists so that $T(X_i) = Z_i\ i = 1, ..., n$ can be assumed to have been produced by a random variable $\mathbf{z}$ with a $Uniform(0,1)$ distribution. We can again transform the data so that $G^{-1}(Z_i) = Y_i\ i = 1, ..., n$ is a random sample from a random variable $\mathbf{y}$ with a $Beta(4,4)$ distribution, whose pdf and cdf are defined in (2).

In this work, we use a parametric transformation $T(\cdot)$, namely the modified Champernowne cdf as proposed by Buch-Larsen et al. (2005):

$$T_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \qquad x \geq 0 \tag{3}$$

with parameters $\alpha > 0$, $M > 0$ and $c \geq 0$.

Let us define the kernel estimator of the density function for the transformed variable:

$$\hat{g}(y) = \frac{1}{n}\sum_{i=1}^{n} K_b(y - Y_i), \tag{4}$$

which should be as close as possible to a $Beta(4,4)$. We can obtain an exact value for the bandwidth parameter that minimizes $A - MISE$ of $\hat{g}$. If $K(t) = (3/4)(1-t^2)\,1(|t| \leq 1)$ is the Epanechnikov kernel, where $1(\cdot)$ equals one when the condition is true and zero otherwise, then we show that the optimal smoothing parameter for $\hat{g}$ if $\mathbf{y}$ follows a $Beta(4,4)$ is:

$$b = k_2^{-\frac{2}{5}} \left(\int_{-1}^{1} K(t)^2\, dt\right)^{\frac{1}{5}} \left(\int_{-1}^{1} \{g''(y)\}^2\, dy\right)^{-\frac{1}{5}} n^{-\frac{1}{5}} = \left(\frac{1}{5}\right)^{-\frac{2}{5}} \left(\frac{3}{5}\right)^{\frac{1}{5}} (35)^{-\frac{1}{5}}\, n^{-\frac{1}{5}},$$

Finally, in order to estimate the density function of the original variable, since $y = G^{-1}(z) = G^{-1}\{T(x)\}$, the transformation kernel density estimator is:

$$\begin{aligned} \hat{f}(x) &= \hat{g}(y)\left[G^{-1}\{T(x)\}\right]' T'(x) = \\ &= \frac{1}{n}\sum_{i=1}^{n} K_b\left(G^{-1}\{T(x)\} - G^{-1}\{T(X_i)\}\right)\left[G^{-1}\{T(x)\}\right]' T'(x). \end{aligned} \tag{5}$$

The estimator in (5) asymptotically minimizes $MISE$.

The properties of the transformation kernel density estimation (5) follow directly from the properties in finite sample of the kernel density estimation for the pdf of the $Beta(4,4)$. The expectation and variance are:

$$\mathbb{E}\left[\widehat{f}(x)\right] = \left[G^{-1}\{T(x)\}\right]' T'(x)\, \mathbb{E}\left[\hat{g}(G^{-1}\{T(x)\})\right] \tag{6}$$

$$\mathbb{V}\left[\widehat{f}(x)\right] = \left(\left[G^{-1}\left\{T\left(x\right)\right\}\right]' T'\left(x\right)\right)^2 \mathbb{V}\left[\hat{g}(G^{-1}\left\{T\left(x\right)\right\})\right]. \tag{7}$$

Besides, since g is known (we impose a beta distribution), it follows that (see, for example, Wand and Jones, 1992, p.20) $\mathbb{E}\left[\hat{g}\left(y\right)\right] = \int K\left(t\right) g\left(y - bt\right) dt$.

When the transformation in kernel estimation is based on a $Beta\left(4,4\right)$ distribution and $K\left(\cdot\right)$ is the Epanechnikov kernel the expectation is:

$$\begin{aligned}\mathbb{E}\left[\widehat{g}\left(y\right)\right] &= B\left[\widehat{g}\left(y\right)\right] + g\left(y\right) \\ &= \frac{1}{96}b^2\left(5b^4 - 27b^2 + 135b^2y^2 + 315y^4 - 378y^2 + 63\right) + \frac{35}{32}\left(1 - y^2\right)^3,\end{aligned}$$

where $B\left[\cdot\right]$ is the bias. If $n \to \infty$ and consequently $b \to 0$, we see that $B\left[\hat{g}\left(y\right)\right] \to 0$ that is of order b^2.

Terrell (1990) showed that since the $Beta(4,4)$ distribution is estimated by $\widehat{g}\left(y\right)$ and this estimator has some optimality properties, then it follows that the properties of the transformation kernel estimation defined in (5) inherit the optimality. Nevertheless, some difficulties arise when implementing the transformation kernel estimator defined in (5), because it depends on the estimator of the kernel density estimator of the transformed variable $\widehat{g}\left(y\right)$. When the density to be estimated is symmetric, defined on a bounded interval and its values decrease rapidly to zero in the boundaries, then the kernel estimation overestimates the density in the upper and lower bounds of the domain of the random variable. If we use a bounded the Epanechnikov kernel that is defined on the $[-1,1]$ interval, the expectation in the boundaries is negative and since the theoretical density in the extremes is zero, we deduce that the estimated density is above the true one:

$$\begin{aligned}\mathbb{E}\left[\widehat{g}\left(1\right)\right] = \mathbb{E}\left[\widehat{g}\left(-1\right)\right] &= \int_{-1}^{1/b} K\left(t\right) g\left(1 - bt\right) dt \\ &= -\left(\frac{5}{192}b^6 + \frac{105}{512}b^5 + \frac{9}{16}b^4 + \frac{35}{64}b^3\right)\end{aligned}$$

Besides, the derivative $\left[G^{-1}\left\{T\left(x\right)\right\}\right]'$ is very large for the extreme values of the domain of the beta distribution and therefore, the bias of the transformed kernel density estimator defined in (6) increases enormously.

Since we want to avoid the difficulties of the estimator defined in (5), we will construct the transformation so that we avoid the extreme values of the beta distribution domain. An easy way to accomplish this is that the random variable z, which results form the initial transformation $z = T\left(x\right)$, is corrected in such a way that its distribution is uniform on an interval that it strictly included in the $(0,1)$ interval.

3 Estimation procedure

Let $z = T(x)$ be a $Uniform(0,1)$, we define a new random variable in the interval $[1-l, l]$, where $1/2 < l < 1$. The values for l should be close to 1. The new random variable is:

$$z^* = T^*(x) = (1-l) + (2l-1)\,T(x). \tag{8}$$

We will discuss later the value of l.

The pdf of the new variable $y^* = G^{-1}(z^*)$ is proportional to the $Beta(4,4)$ pdf, but it is in $[-a, a]$ interval, where $a = G^{-1}(l)$. Finally, our proposed transformation kernel density estimation is:

$$\begin{aligned}\hat{f}(x) &= \frac{\hat{g}(y^*)\left[G^{-1}\{T^*(x)\}\right]' T^{*\prime}(x)}{(2l-1)} = \hat{g}(y^*)\left[G^{-1}\{T^*(x)\}\right]' T'(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} K_b\left(G^{-1}\{T^*(x)\} - G^{-1}\{T(X_i)\}\right)\left[G^{-1}\{T^*(x)\}\right]' T'(x)\end{aligned} \tag{9}$$

The value of $A-MISE$ associated to the kernel estimation $\hat{g}(y^*)$, where the random variable y^* is defined on an interval that is smaller than $Beta(4,4)$ domain is:

$$A-MISE_a = \frac{1}{4}b^4 (k_2)^2 \int_{-a}^{a} \{g''(y)\}^2 dy + \frac{1}{nb}\int_{-a}^{a} g(y)\,dy \int K(t)^2\,dt. \tag{10}$$

And finally, the optimal bandwidth parameter based on the asymptotic mean integrated squared error measure for $\hat{g}(y^*)$ is:

$$\begin{aligned}b_g^{opt} &= k_2^{-\frac{2}{5}}\left(\int_{-1}^{1} K(t)^2\,dt \int_{-a}^{a} g(y)^2\,dy\right)^{\frac{1}{5}}\left(\int_{-a}^{a}\{g''(y)\}^2\,dy\right)^{-\frac{1}{5}} n^{-\frac{1}{5}} \\ &= \left(\frac{1}{5}\right)^{-\frac{2}{5}}\left(\frac{3}{5}\left(-\frac{5}{16}a^7 + \frac{21}{16}a^5 - \frac{35}{16}a^3 + \frac{35}{16}a\right)\right)^{\frac{1}{5}} \\ &\quad \times\left(\frac{35}{128}a\left(-1260a^2 + 2898a^4 - 2700a^6 + 875a^8 + 315\right)\right)^{-\frac{1}{5}} n^{-\frac{1}{5}}.\end{aligned} \tag{11}$$

The difficulty that arises when implementing the transformation kernel estimation expressed in (9) is the selection of the value of l. This value can be chosen subjectively and in the simulation study below we give results for some possible values of l.

Let X_i, $i = 1, ..., n$, be iid observations from a random variable with an unknown density f. The following describes in detail the transformation kernel

density estimator of f. The resulting transformation kernel density estimator of f is called KIBMCE (kernel inverse beta modified Champernowne estimator).

(1) Calculate the parameters $\left(\widehat{\alpha}, \widehat{M}, \widehat{c}\right)$ of the modified Champernowne distribution defined in (3). We use the same method as proposed by Buch-Larsen et al. (2005) which is based on maximum likelihood estimation.
(2) Transform the data set X_i, $i = 1, ..., n$, with the transformation function, T:

$$Z_i = T_{\widehat{\alpha},\widehat{M},\widehat{c}}(X_i), \qquad i = 1, ..., n.$$

The transformation function transforms data into the interval $(0, 1)$, and the parameter estimation is designed to make the transformed data as close to a uniform distribution as possible.
(3) We fix l equal to 99% and 98%.
(4) Rescale the transformed variable in step 2, so that instead of being defined on the $(0, 1)$ interval, it is defined on the $(1 - l, l)$ interval:

$$T^*_{\widehat{\alpha},\widehat{M},\widehat{c}}(X_i) = Z^*_i = (1 - l) + (2l - 1) Z_i, \qquad i = 1, ..., n.$$

(5) Transform de data Z^*_i with transformation function, $Y_i = G^{-1}(Z^*_i)$, where G is the $Beta(4, 4)$ distribution cdf. The density of the transformed variable is not defined on the same domain as $Beta(4, 4)$ distribution, its domain depends on the value of l. In other words, the domain $Beta(4, 4)$ distribution defined in (2) is $[-1, 1]$ and the density of the transformed variable is defined on the interval $[-a, a]$, where $a = G^{-1}(l) < 1$.
(6) Calculate the classical kernel density estimator on the transformed data, $Y_i, i = 1, ..., n$:

$$\hat{g}(y) = \frac{1}{n} \sum_{i=1}^{n} K_b(y - Y_i), \tag{12}$$

where $K_b(\cdot) = \frac{1}{b} K(\cdot)$ and $K(\cdot)$ is the kernel function, we use the Epanechnikov kernel.
(7) The classical kernel density estimator of the transformed data set results in the KIBMCE estimator on the transformed scale. Therefore the estimator of the density of the original data set, X_i, $i = 1, ..., n$ is calculated by transforming the smoothed distribution of the transformed data with $G\left(T^{*-1}_{\widehat{\alpha},\widehat{M},\widehat{c}}(z)\right) = x$, the expression of the resulting KIBMCE estimator of f therefore is:

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_b(T^*_{\widehat{\alpha},\widehat{M},\widehat{c}}(x) - T^*_{\widehat{\alpha},\widehat{M},\widehat{c}}(X_i)) \left(G^{-1}\right)'(z^*) T'_{\widehat{\alpha},\widehat{M},\widehat{c}}(x). \tag{13}$$

4 Simulation study

This section presents a comparison of our inverse beta 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 distribution, improves density estimation.

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

Table 1: Distributions in simulation study.

Distribution	Density	Parameters
Mixture of pLognormal(μ,σ) and $(1-p)$Pareto(λ,ρ,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(μ,σ)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$	$(\mu,\sigma) = (0,0.5)$
Weibull(γ)	$f(x) = \gamma x^{(\gamma-1)}e^{-x^\gamma}$	$\gamma = 1.5$
Normal(μ,σ)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$(\mu,\sigma) = (5,1)$

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. Here we see that if the second transformation based on the inverse beta transformation improves the results presented in Buch-Larsen, et al. (2005), this means that the double-transformation method presented here is a substantial gain with respect to existing methods.

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, which is very

relevant in practice when dealing with income or cost data:

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

The simulation results can be found in Table 2. 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 $WISE$ for different values of the trimming parameter $l = 0.99, 0.98$. The benchmark results are labelled KMCE and they correspond to those presented in Buch-Larsen, et al. (2005).

In general, we can conclude that after a second transformation based on the inverse of a certain Beta distribution cdf the error measures diminish with respect to the KMCE method. In some situations the errors diminish quite substantially with respect to the existing approaches. Table 3 presents the ratio between the error measure of each method and the error measure of the KMCE method. When the ratio is below(above) 1, it means that the KIBMCE estimator has less(more) error than the KMCE estimator.

We can see that the error measure that shows improvements when using the KIBMCE estimator is the $WISE$, which means that this new approach is fitting the tail of positive distributions better than existing alternatives. The $WISE$ error measure is always smaller for the KIBMCE than for the KMCE, at least for one of the two possible value of l that have been used in this simulation study. This would make the KIBMCE estimator specially suitable for positive heavy-tailed distributions. When looking more closely at the results for the mixture of a lognormal distribution and a Pareto tail, we see that larger values of l are needed to improve the error measures that were encountered with the KMCE method.

We can see that for the Truncated logistic distribution, the lognormal distribution and the Weibull distribution, the method presented here is clearly better than the existing KMCE. We can see in Table 3 that for $N = 1000$, the KIBMCE $WISE$ is about 20% lower than the KMCE $WISE$ for these distributions. A similar behavior is shown by the other error measures. L_1 and L_2,for $N = 1000$, are about 15% lower for the KIBMCE.

Note that the KMCE method was studied in Buch-Larsen, et al. (2005) and the simulation study showed that it improved on the error measures for the existing methodological approaches (Clements, et al., 2003 and Wand, et al., 1991).

Table 2: The estimated error measures for KMCE and KIBMCE
with $l = 0.99$ and $l = 0.98$ for sample size 100 and 1000 based on 2000 repetitions

			Log-Normal	Log-Pareto		Weibull	Tr. Logist.
				p =.7	p =.3		
N=100	L1	KMCE	0.1363	0.1287	0.1236	0.1393	0.1294
		l=0.99	0.1335	0.1266	0.1240	0.1374	0.1241
		l=0.98	0.1289	0.1215	0.1191	0.1326	0.1202
	L2	KMCE	0.1047	0.0837	0.0837	0.1084	0.0786
		l=0.99	0.0981	0.0875	0.0902	0.1085	0.0746
		l=0.98	0.0956	0.0828	0.0844	0.1033	0.0712
	WISE	KMCE	0.1047	0.0859	0.0958	0.0886	0.0977
		l=0.99	0.0972	0.0843	0.0929	0.0853	0.0955
		l=0.98	0.0948	0.0811	0.0909	0.0832	0.0923
N =1000	L1	KMCE	0.0659	0.0530	0.0507	0.0700	0.0598
		l=0.99	0.0544	0.0509	0.0491	0.0568	0.0497
		l=0.98	0.0550	0.0509	0.0522	0.0574	0.0524
	L2	KMCE	0.0481	0.0389	0.0393	0.0582	0.0339
		l=0.99	0.0394	0.0382	0.0393	0.0466	0.0298
		l=0.98	0.0408	0.0385	0.0432	0.0463	0.0335
	WISE	KMCE	0.0481	0.0384	0.0417	0.0450	0.0501
		l=0.99	0.0393	0.0380	0.0407	0.0358	0.0393
		l=0.98	0.0407	0.0384	0.0459	0.0369	0.0394

Table 3: Ratio between the error measures of KIBMCE and KMCE for sample size 100 and 1000 based on 2000 repetitions

			Log-Normal	Log-Pareto		Weibull	Tr. Logist.
				p =.7	p =.3		
N =100	L1	l=0.99	0.9795	0.9837	1.0032	0.9864	0.9590
		l=0.98	0.9457	0.9441	0.9636	0.9519	0.9289
	L2	l=0.99	0.9370	1.0454	1.0777	1.0009	0.9491
		l=0.98	0.9131	0.9892	1.0084	0.9530	0.9059
	WISE	l=0.99	0.9284	0.9814	0.9697	0.9628	0.9775
		l=0.98	0.9054	0.9441	0.9489	0.9391	0.9447
N =1000	L1	l=0.99	0.8255	0.9604	0.9684	0.8114	0.8311
		l=0.98	0.8346	0.9604	1.0296	0.8200	0.8763
	L2	l=0.99	0.8191	0.9820	1.0000	0.8007	0.8791
		l=0.98	0.8482	0.9897	1.0992	0.7955	0.9882
	WISE	l=0.99	0.8170	0.9896	0.9760	0.7956	0.7844
		l=0.98	0.8462	1.0000	1.1007	0.8200	0.7864

5 Data study

In this section, we apply our semiparametric estimation method to a data set that contains automobile claim costs from a Spanish insurance company for accidents occurred in 1997. This data set was analyzed in detail by Bolancé et al. (2003). It is a typical insurance claims amount 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 $\widehat{\alpha}_1 = 1.116$, $\widehat{M}_1 = 66$, $\widehat{c}_1 = 0.000$ and for older drivers $\widehat{\alpha}_2 = 1.145$, $\widehat{M}_2 = 68$, $\widehat{c}_2 = 0.000$, 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.

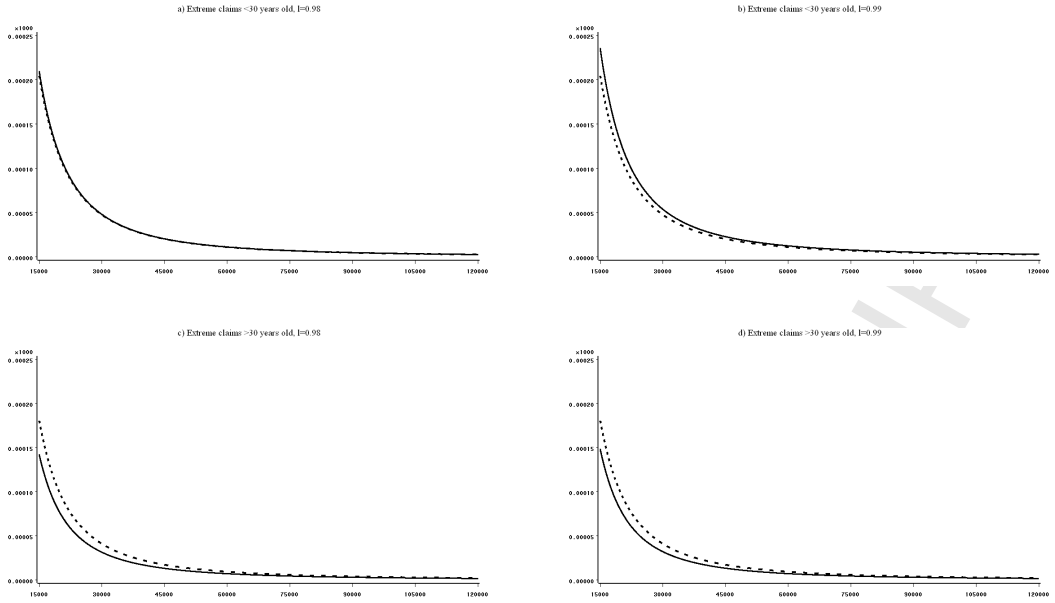


Fig. 1. KIBMCE with $l = 0.99$ and $l = 0.98$ (solid line) and KMCE (dashed line) estimators of automobile claims from an insurance company. In a) and b) younger policyholders and in c) and d) older policyholders.

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; ∞). Figure 1 only presents extreme claims. It shows the KMCE and the KIBMCE estimator for the young and older individuals. The figure illustrates that the tail in the estimated density of young policyholders is heavier than the tail of the estimated density of older policyholders. This can be taken as evidence that young drivers are more likely to claim a large amount than older drivers. The KIBMCE method produces a heavier tail for the young drivers, i.e. high risk groups, and the tail for the older drivers group is lower than the one estimated with the KMCE method. Based on the results in the simulation study presented above, we believe that the KIBMCE method improves the estimation of the density in the extreme claims class.

References

[1] Bolancé, C., Guillén, M. and Nielsen, J.P. (2003). Kernel density estimation of actuarial loss functions, *Insurance: Mathematics and Economics*, 32, 19-36.

[2] 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.

[3] 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.

[4] Jones, M.C. (1990). Variable kernel density estimation and variable kernel density estimation, *Australian Journal of Statistics, 32, 361-371.*

[5] Johnson, N.L., Jotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Vol. 1. New York: John wiley & Sons, Inc.

[6] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

[7] Terrell, G.R. (1990). The maximal smoothing principle in density estimation, *Journal of the American Statistical Association*, 85, 270-277.

[8] Terrell, G.R. and Scott, D.W. (1990). Oversmoothed nonparametric density estimates, *Journal of the American Statistical Association*, 80, 209-214.

[9] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* London: Chapman & Hall.

[10] Wand, P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the Americal Statistical Association,* 86, 414, 343-361.

[11] Wu, T.-J., Chen, C.-F. and Chen, H.-Y. (2007). A variable bandwidth selector in multivariate kernel density estimation, *Statistics & Probability Letters*, 77, 462-467.