People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

MOHAMED KHIDER UNIVERSITY, BISKRA

FACULTY of EXACT SCIENCES, SIENCE of NATURE and LIFE

DEPARTMENT of MATHEMATICS

A thesis presented for the degree of :

Doctor in Mathematics

In the Filled of : **Statistics**

By

Madiha TOUR

Title :

On improve boundary effect in kernel distribution estimation

Members of the jury :

BRAHIM BRAHIMI	Prof.	University of Biskra	Chairman
Abdallah SAYAH	M.C.A.	University of Biskra	Supervisor
Halim ZEGHDOUDI	M.C.A	University of Annaba	Examinator
Djabrane YAHIA	M.C.A	University of Biskra	Examinator
Fatah BENATIA	M.C.A	University of Biskra	Examinator

2018

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Thèse présentée en vue de l'obtention du Diplôme :

Doctorat en Mathématiques

Option: **Statistique**

Par

Madiha TOUR

Title :

Sur l'Amélioration d'effet du bord dans estimation à noyau de la distribution

Membres du Comité d'Examen :

Brahim BRAHIMI	Prof.	Université de Biskra	Président
Abdallah SAYAH	M.C.A.	Université de Biskra	Encadreur
Halim ZEGHDOUDI	M.C.A	Université d' Annaba	Examinateur
Djabrane YAHIA	M.C.A	Université de Biskra	Examinateur
Fatah BENATIA	M.C.A	Université de Biskra	Examinateur

2018

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Dédicace

Je dédie ce travail

Mes chers père et mère, Pour tous les sacrifices qu'ils ont consentis pour me permettre de pour suivre mes études dans les meilleures conditions possibles et pour n'avoir jamais cessé de m'encourager tout au long de mes années d'étude.

A mes très chers frères et sœurs.

A toute ma famille.

REMERCIEMENTS

Mes premiers remerciements vont tout naturellement à Dieu le tout puissant pour la volonté, la santé et la patience qu'il m'a donnée pour réaliser mon travail de recherche.

Je tiens à exprimer ma profonde gratitude, mes sincères et chaleureux remerciements à mon directeur de thèse, Monsieur le Dr. Sayah Abdallah, pour la confiance qu'il m'a accordée en acceptant d'encadrer ce travail doctoral, par ses multiples conseils. J'ai été extrêmement sensible à ses qualités humaines d'écoute et de compréhension durant toutes les étapes de cette recherche.

Je remercie vivement Monsieur le Professeur Brahimi Brahim qui m'a fait le grand honneur de présider le jury. Mes remerciements vont aussi aux membres du jury, le Dr. Zeghdoudi Halim, le Dr. Yahia Djabrane et le Dr. Benatia Fatah pour avoir accepté d'examiner et bien évaluer ma thèse. Merci pour vos corrections et remarques constructives.

Je remercie chaleureusement tous les membres du Laboratoire de Mathématiques Appliquées et surtout le Professeur Meraghni Djamel de son aide.

Je tiens à remercier ma famille pour leurs encouragements, soutien et sacrifice apportés tout au long de mon cheminement scolaire et universitaire. Merci à toute personne ayant contribué directement ou indirectement à ce travail.

Abstract

In this thesis, we propose a new estimator for improve boundary effects in kernel estimator of the heavy-tailed distribution function specially the Pareto-type distributions and its bias, variance and mean squared error are determined. Kernel methods are widely used in many research areas in statistics. However, kernel estimators suffer from boundary effects when the support of the function to be estimated has finite end points. Boundary effects seriously affect the overall performance of the estimator. To remove the boundary effects, a variety of methods have been developed in the literature, the most widely used is the reflection, the transformation ... In this thesis, we introduce a new method of boundary correction when estimating the heavy-tailed distribution function. Our technique is kind of a generalized reflection method involving reflecting a transformation of the observed data by modified Champernowne distribution function.

Résumé

Dans cette thèse, nous proposons un nouveau estimateur pour améliorer les effets de bord dans l'estimateur à noyau de la fonction de distribution à queue lourde spécialement les distributions de type Pareto, son biais, variance et l'erreur quadratique moyenne de cette estimateur sont déterminées. Les méthodes du noyau sont largement utilisées dans de nombreux domaines de recherche en statistiques. Cependant, les estimateurs à noyau souffrent des problèmes d'effets aux bords de leur support. Les effets de bord affectent sérieusement la performance globale de l'estimateur. Pour corrigé ces effets de bord, une variété de méthodes ont été développées dans la littérature, la plus utilisée est la réflexion, la transformation ... Dans cette thèse, nous introduisons une nouvelle méthode de correction de l'effet de bord lors de l'estimation de la fonction de distribution à queue lourde. Notre technique est en quelque sorte une méthode de réflexion généralisée impliquant une transformation des données observées par une fonction de distribution de Champernowne modifiée.

Contents

List of Figures

List of Tables

Introduction

In the area of statistic, estimation of the unknown distribution function $F(x)$, of a population is important. Other statistical methods that are dependent on knowledge of the distribution function include hypothesis testing and confidence interval estimation. Existing methods to estimate on unknown distribution function from data can be classified into two groups, namely : parametric and nonparametric methods. Parametric methods are dependent on assumption that the functional form of the distribution function is specified. If it is known that the data are normally distributed with unknown mean and variance, the unknown parameters can be estimated from the data and the distribution function is then completely determined. If the assumption of normality cannot be made, then the parametric method of estimation cannot be used and a nonparametric method of estimation must be implemented. Here we consider only nonparametric estimators.

Nonparametric kernel smoothing belongs to a general category of techniques for nonparametric estimations including : density, distribution, regression, quantiles, ... These estimators are now popular and in wide use with great success in statistical applications. Early results on kernel density estimation are due to Rosenblatt (1956) [51] and Parzen (1962) [47]. Good references in this area are Silverman (1986) [55], and Wand and Jones (1995) [61], and the form kernel regression estimator has been proposed by Nadaraya (1964) [46] and Watson (1964) [63]. While results in a kernel distribution estimator is introduced by authors such as Nadaraya (1964) [45] or Watson and Leadbetter (1964) [62]. Such an estimator arises as an integral of the Parzen-Rosenblatt kernel density estimator.

Kernel estimates may suffer from boundary effects. This type of boundary effect for kernel estimators of curves with compact supports is well-known in regression and density function estimation frameworks. In the density estimation context, a various boundary bias correction methods have been proposed. Schuster (1999) [54] and Cline and Hart (1991) [14] considered the reflection method, which is most suitable for densities with zero derivatives near the boundaries. Boundary kernel method and local polynomial method are more general without restrictions on the shape of densities. Local polynomial method can be seen as a special case of boundary kernel method and draws much attention due to its good theoretical properties. Though early versions of these methods might produce negative estimates or inflate variance near the boundaries, remedies and refinements have been proposed, see Müller (1991) [43], Jones (1993) [31], Jones and Foster (1996) [32], Cheng (1997) [11], Zhang and Karunamuni (1998, 2000) [67], [69] and Karunamuni and Alberts (2005) [34]. Cowling and Hall (1996) [15] proposed a pseudo-data method that estimates density functions based on the original data plus pseudo-data generated by linear interpolation of order statistics. Zhang et al. (1999) [68] combined the pseudo-data, transformation and reflection methods. In the regression function estimation context, Gasser and Müller (1979) [25] identified the unsatisfactory behavior of the Nadaraya Watson regression estimator for points in the boundary region. They proposed optimal boundary kernels but did not give any formulas. However, Gasser and Müller (1979) [25] and Müller (1988) [44] suggested multiplying the truncated kernel at the boundary zone or region by a linear function. Rice (1984) [50] proposed another approach using a generalized jackknife. Schuster (1985) [53] introduced a reflection technique for density estimation. Eubank and Speckman (1991) [20] presented a method for removing boundary effects using a bias reduction theorem. Kheireddine et al. (2015) [36] produce a General method of boundary correction in kernel regression estimation, we combine the transformation and reflection boundary correction methods.

Kernel distribution estimators are not consistent near the boundary of its support. In other words, these effects seriously affect the performance of these estimators and these require good precision. A similar correction used in density estimation would be made for improve the theoretical performance of the usual kernel distribution function estimator at the boundary points. In this thesis we develop a new kernel estimator of the distribution function for heavy-tailed distributions based on the modified Champernowne transformation. We will concentrate not to estimate the distribution of X based on the samples $X_1, ..., X_n$ but to estimate the distribution of Y based on the samples $Y_1, ..., Y_n$ where $Y_i = T(X_i)$.

Buch-Larsen et al. (2005) [7] suggested to choose T so that $T(X)$ is close to the uniform distribution. They proposed a kernel estimator of the density of heavy-tailed distributions based on a transformation of set of the original data with a modified Champernowne distribution that is a heavy-tailed Pareto-type, see Champernowne (1936, 1952) [8], [10], and applied to transformed data. The kernel estimator for heavy-tailed distributions has been studied by several authors Bolancé et al. (2003) [6], Clements et al. (2003) [13] and Buch-Larsen et al. (2005) [7] propose different families of parametric transformation that they all make the transformed distribution more symmetric than the original, which in many applications are generally highly asymmetric right. Sayah et al. (2010) [52] produce a kernel quantile estimator for heavy-tailed distributions using a modification of the Champernowne distribution.

Our thesis is organized in 4 **chapters**.

• **Chapter** 1, is an introduction to the nonparametric estimation of the distribution function, a common problem in statistics is that of estimating a density f or a distribution function F from a sample of real random variables $X_1, ..., X_n$ independent and with the same unknown distribution. The functions f and F, as the characteristic function, completely describe the probability distribution of the observations and to know a convenient estimation can solve many statistical problems. The traditional estimator of the distribution function F is the empirical distribution function which is defined by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x).$$

This estimator is an unbiased estimator and consist of $F(x)$. Another estimator of F is the kernel estimator $\widehat{F}_n$ which is defined by

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right),$$

where $K(x) = \int_{-\infty}^{x} k(t)dt$ and k is a kernel function and b is the smoothing parameter. The asymptotic properties of $\widehat{F}_n$ was initiated by Nadaraya (1964) [45] and continued in a series of papers among which we mention Winter (1973, 1979) [64], [65], Yamato (1973) [66], Reiss (1981) [49].

• **Chapter** 2, we focused on the presentation of the concept of heavy-tailed distributions and different classes of this type of distributions, an important classes of heavy-tailed distributions are that subexponential distribution and the regularity varying distribution with index $\alpha > 0$. A distribution has a heavy tailed if and only if its kurtosis is higher than the normal distribution that is equal to 3. There are others definitions so that a distribution is heavy-tailed that is the distributions which the exponential moment is infinite.

• **Chapter** 3, describes the transformation in kernel density estimation. Let $X_1, ..., X_n$ a random sample of independent and identically distributed observations of a random variable with density function f, then the kernel density estimator at point x is

$$\hat{f}_n(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right),$$

where b is the bandwidth or smoothing parameter, and k is the kernel function, usually it is a symmetric density function bounded and centred at zero. Silverman (1986) [55] or Wand and Jones (1995) [61] provide an extensive review of classical kernel estimation. For heavy-tailed distributions, the kernel density estimation has been studied by several authors : Buch-Larsen et al. (2005) [7], Clements et al. (2003) [13] and Bolancé et al. (2003) [6]. 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 it is preferable to other transformation density estimation approaches for distributions that are Pareto-like in the tail.

• **Chapter** 4, in this chapter we present our result which is the estimation of heavy-tailed distributions based on a reflection method involving reflecting a transformation and using the modified Champernowne transformation which is introduced in the work of Buch-Larcen et al. (2005) [7] in the case of density estimation for heavy-tails distributions, the new approach based on the modified Champernowne distribution is the preferable method, because it has a good performance in most of the investigated situations.

Chapter 1

Nonparametric distribution estimation

Nonparametric methods are becoming increasingly popular in statistical analysis of economic problems. In most cases, this is caused by the lack of information, especially historical data, about the economic variable being analysed. Smoothing methods concerning functions, such as density or distribution function, play a special role in a nonparametric analysis of economic phenomena. Knowledge of density function or distribution function, or their estimates, allows one to characterize the random variable more completely. It is true that one can often switch from an estimator of f to an estimator of F by integration and an estimator of F to an estimator of f by derivation. However one feature is noteworthy : it is the existence the empirical distribution function F_n. Estimation of functional characteristics of random variables can be carried out using kernel methods. Nonparametric kernel distribution estimation is now popular and in wide use with great success in statistical applications.

1.1 Empirical distribution function

The best known and simplest nonparametric estimator of distribution function is the empirical distribution function (EDF). Let $X_1, ..., X_n$ be independent and identically distributed (iid) copies of the random variable (rv) X with unknown continuous distribution function (df) $F(x) = P(X \leq x)$, then the estimator of F, from $X_1, ..., X_n$, is the EDF F_n defined at some points x by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \tag{1.1}$$

where

$$I(X_i \leq x) = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{if } X_i > x. \end{cases}$$

The EDF is most conveniently defined in terms of the order statistics of a sample. Suppose that the n sample observations are distinct and arranged in increasing order so that $X_{(1)}$ is the smallest and the $X_{(n)}$ is the largest. A formal definition of the EDF $F_n(x)$ is

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < X_{(1)} \\ \dfrac{i}{n} & \text{if} \quad X_{(i)} \leq x < X_{(i+1)} \\ 1 & \text{if} \quad x \geq X_{(n)}. \end{cases}$$

1.1.1 Properties of EDF

Using properties of the binomial distribution. Note that $I(X_i \leq x)$ are independent Bernoulli random variables such that

$$I(X_i \leq x) = \begin{cases} 1, & \text{with probability } F(x) \\ 0, & \text{with probability } 1 - F(x). \end{cases}$$

Thus $nF_n(x)$, is a binomial random variable (n trials, probability $F(x)$ of success) and so

- **Bias**

$$E\left(F_n(x)\right) = \frac{1}{n}\sum_{i=1}^{n} P\left(X_i \leq x\right) = F(x).$$

- **Variance**

$$\begin{aligned} Var\left(F_n(x)\right) &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(I(X_i \leq x)\right) \\ &= \frac{1}{n} Var\left(I(X_i \leq x)\right) \\ &= \frac{F(x)\left(1 - F(x)\right)}{n} \underset{n\to\infty}{\longrightarrow} 0 \ . \end{aligned}$$

- **Mean Square Error** (MSE)

$$\begin{aligned} MSE\left(F_n(x)\right) = E\left[\left(F_n(x) - F(x)\right)^2\right] &= Bias^2 + Variance \\ = Var\left(F_n(x)\right) \underset{n\to\infty}{\longrightarrow} 0 \ . \end{aligned}$$

Thus as an estimator of $F(x)$, $F_n(x)$ is unbiased and its variance tends to 0 as $n \to \infty$.

- **Convergence in probability**

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F(x).$$

For any fixed real value x, $F_n(x)$ is a consistent estimator of $F(x)$, or, in other words, $F_n(x)$ converges to $F(x)$ in probability. The **convergence in probability** is for each value of x individually, whereas sometimes we are interested in all values of x, collectively.

- **Glivenko-Cantelli Theorem**

An even stronger proof of convergence is given by the **Glivenko-Cantelli Theorem**, the states that F can be approximatet by F_n in an uniform manner for large sample sizes such that

$$P\left(\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|=0\right)=1.$$

- **Inequality of Dvoretsky-Kiefer-Wolfowitz**

  For any $\varepsilon>0$, and $n\in\mathbb{N}$,

$$P\left(\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|>\varepsilon\right)\leq 2e^{-2n\varepsilon^2}.$$

Another useful property of the EDF is its **asymptotic normality**, given in the following theorem.

Theorem 1.1.1 *As $n\to\infty$, the limiting probability distribution of the standardized $F_n(x)$ is standard normal, or*

$$\frac{\sqrt{n}\,(F_n(x)-F(x))}{\sqrt{F(x)(1-F(x))}}\xrightarrow{L}N(0,1).$$

Despite the good statistical of F_n, the empirical distribution function is a step function, one could prefer in many applications a rather smooth estimate see Azzalini (1981) [3].

1.2 Kernel method

The kernel method originated from the idea of Rosenblatt (1956) [51] and Parzen (1962) [47] dedicated to density estimation. The distribution function $F(x)$ is naturally estimated by the EDF (1.1). It might seem natural to estimate the density $f(x)$ as the derivative of $F_n(x)$, $\frac{d}{dx}F_n(x)$. But this estimator would be a set of mass point, not a density, and as such is not a useful estimate of $f(x)$. Instead, consider a discrete derivative. For some small $b>0$, let

$$\hat{f}_n(x)=\frac{F_n(x+b)-F_n(x-b)}{2b}.$$

We can write this as

$$\begin{aligned}\hat{f}_n(x) &= \frac{1}{2nb}\sum_{i=1}^{n} I\left(x-b \le X_i \le x+b\right)\\ &= \frac{1}{2nb}\sum_{i=1}^{n} I\left(\frac{|X_i - x|}{b} \le 1\right),\end{aligned}$$

$\hat{f}_n(x)$ is a special case of what is called a Rosenblatt-Parzen kernel density estimator is as follows (see Wand and Jones (1995) [61]; Silverman (1986) [55]):

$$\hat{f}_n(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right),$$

where $X_1, ..., X_n$ be independent random variables identically distributed which are drawn from a continuous distribution $F(x)$ with density function $f(x)$, n is the sample size and $b := b_n$ ($b \to 0$ and $nb \to \infty$) is the smoothing parameter, called the bandwidth, which controls the smoothness of the estimator, $k(.)$ is the weighting function called the kernel function. When $k(.)$ is symmetric and unimodal function and the following conditions are fulfilled:

1. $k(t) \ge 0$, $\forall t \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} k(t)dt = 1$, hence k is a density function.
3. $\int_{-\infty}^{\infty} tk(t)dt = 0$.
4. $0 < \int_{-\infty}^{\infty} t^2 k(t)dt < \infty$.

The order of a kernel, ν, is defined as the order of the first non-zero moment

$$\mu_j(k) = \int_{-\infty}^{\infty} t^j k(t)dt.$$

The order of a symmetric kernel is always even. Symmetric non-negative kernels are second-order kernels.

A kernel is higher-order kernel if $\nu > 2$. These kernels will have negative parts and are not probability densities. We refer to Hansen (2009) [29] for more details.

Some popular kernels functions used in the literature are the following (see Silverman (1986) [55])

Kernel	$k(t)$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$, for $t \in \mathbb{R}$
Epanechnikov	$\frac{3}{4}\left(1-t^2\right) I\left(\lvert t\rvert \leq 1\right)$
Quartic or Biweight	$\frac{15}{16}\left(1-t^2\right)^2 I(\lvert t\rvert \leq 1)$
Triangular or Triweight	$\frac{35}{32}\left(1-t^2\right)^3 I\left(\lvert t\rvert \leq 1\right)$

Table 1.1: Some kernel functions. I(.) denotes the indicator function.

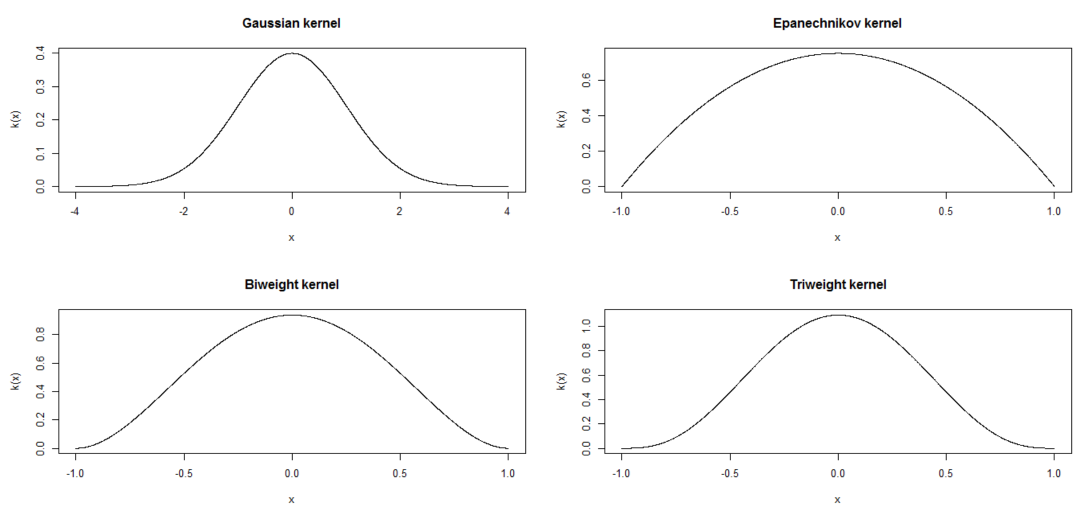

Figure 1.1: Rate of kernels : Gaussian, Epanechnikov, Biweight and Triweight .

1.2.1 Kernel distribution function estimator

Let $X_1, ..., X_n$ denote independent identically distributed random variables with an unknown density $f(.)$ function and distribution function $F(.)$, which we wish to estimate.

The density estimator can be integrated to obtain a nonparametric alternative to $\widehat{F}_n(x)$ for smooth distribution function that said the kernel distribution function estimator $\widehat{F}_n(x)$ that was proposed by Nadaraya (1964) [45] and is defined by

$$\begin{aligned} \widehat{F}_n(x) &= \int_{-\infty}^{x} \hat{f}_n(t)dt \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right), \end{aligned} \tag{1.2}$$

where the function K is defined from a kernel k as

$$K(x) = \int_{-\infty}^{x} k(t)dt.$$

Assume that k is symmetric, and has a compact support $[-1, 1]$. Let

$$\mu_i = \int_{-1}^{1} t^i k(t)dt, \ \ i = 1, 2, 3, 4.$$

In fact, $\mu_1 = \mu_3 = 0$ since k is symmetric, then the properties of function $K(x)$ are the following (see Baszczyńska,. (2016) [4]):

$$\begin{cases} \displaystyle\int_{-1}^{1} K^2(x)dx \leq \int_{-1}^{1} K(x)dx = 1, \\ \displaystyle\int_{-1}^{1} K(x)k(x)dx = \frac{1}{2}, \\ \displaystyle\int_{-1}^{1} x^i K(x)dx = \frac{1}{i+1}\left(1-\mu_{i+1}\right), \ \ i = 0, 1, 2, 3, 4. \\ \displaystyle\int_{-1}^{1} xK(x)k(x)dx = \frac{1}{2}\left[1 - \int_{-1}^{1} K^2(x)dx\right]. \end{cases}$$

Function $K(x)$ is a cumulative distribution function because $k(x)$ is a probability density function. For example, when the kernel function is Epanechnikov kernel, the function

$K(x)$ has the form:

$$K(x) = \begin{cases} 0 & \text{for} \quad x \leq -1, \\ -\frac{1}{4}x^3 + \frac{3}{4}x + \frac{1}{2} & \text{for} \quad |x| \leq 1, \\ 1 & \text{for} \quad x \geq 1. \end{cases}$$

In order to compare the kernel distribution function estimator (1.2) to the EDF (1.1), expression for the aforementioned estimator will now be derived, see Van Graan (1983) [59]. To obtain $Var\left(\widehat{F}_n(x)\right)$ note that (under certain conditions on F and K)

$$\begin{aligned} Var\left(\widehat{F}_n(x)\right) &= \frac{1}{n} Var\left[K\left(\frac{x-X_i}{b}\right)\right] \\ &= \frac{1}{n}\left[E\left\{\left[K\left(\frac{x-X_i}{b}\right)\right]^2\right\} - \left[E\left\{K\left(\frac{x-X_i}{b}\right)\right\}\right]^2\right]. \end{aligned}$$

Now

$$\begin{aligned} E\left[K\left(\frac{x-X_i}{b}\right)\right] &= \int_{-\infty}^{\infty} K\left(\frac{x-y}{b}\right) dF(y). \\ &= \int_{-\infty}^{x-b} 1.f(y)dy + \int_{x-b}^{x+b} K\left(\frac{x-y}{b}\right) f(y)dy + \int_{x+b}^{\infty} 0.f(y)dy, \end{aligned}$$

using the substitution $\frac{x-y}{b} = t$ and a Taylor series expansion it follows that

$$\begin{aligned} E\left[K\left(\frac{x-X_i}{b}\right)\right] &= F(x-b) + \int_{-1}^{1} bK(t)f(x-bt)dt \\ &= F(x) - bf(x) + \frac{1}{2}b^2 f'(x) + o(b^2) \\ &+ \int_{-1}^{1} bK(t)\left\{f(x) - btf'(x) + \frac{1}{2}b^2t^2 f''(x) + o(b^2)\right\} dt \\ &= F(x) + \frac{1}{2}b^2 f'(x)\mu_2(k) + o(b^2), \end{aligned}$$

where

$$\mu_2(k) = \int_{-1}^{1} t^2 k(t) dt.$$

Using a similar approach as above an expression for $E\left\{\left[K\left(\frac{x-X_i}{b}\right)\right]^2\right\}$ can be obtained

$$\begin{aligned}
E\left\{\left[K\left(\frac{x-X_i}{b}\right)\right]^2\right\} &= \int_{-\infty}^{\infty}\left[K\left(\frac{x-y}{b}\right)\right]^2 f(y)dy \\
&= \int_{-\infty}^{x-b} 1.f(y)dy + \int_{x-b}^{x+b} K^2\left(\frac{x-y}{b}\right) f(y)dy. \\
&= F(x-b) + \int_{-1}^{1} bK^2(t) f(x-bt)dt
\end{aligned}$$

Using the property $K(t) = 1 - K(-t)$, and Taylor series expansion we obtain

$$\begin{aligned}
E\left\{\left[K\left(\frac{x-X_i}{b}\right)\right]^2\right\} &= F(x-b) + \int_{-1}^{1} b\left(1-K(-t)\right)^2 f(x-bt)dt \\
&= F(x-b) + \int_{-1}^{1} bf(x-bt)dt + \int_{-1}^{1} bK^2(-t)f(x-bt)dt \\
&- 2\int_{-1}^{1} bK(-t)f(x-bt)dt \\
&= F(x-b) - F(x-b) + F(x+b) + \int_{-1}^{1} bK^2(t)\left\{f(x)+o\left(1\right)\right\}dt \\
&- 2\int_{-1}^{1} bK(t)\left\{f(x)+o\left(1\right)\right\}dt \\
&= F(x) + bf(x) + bf(x)\int_{-1}^{1} K^2(t)dt - 2bf(x)\int_{-1}^{1} K(t)dt + o\left(b\right) \\
&= F(x) + bf(x) - 2bf(x) + bf(x)\int_{-1}^{1} K^2(t)dt + o\left(b\right) \\
&= F(x) - bf(x) + bf(x)\int_{-1}^{1} K^2(t)dt + o\left(b\right).
\end{aligned}$$

Expression for $Var\left(\widehat{F}_n(x)\right)$ can be computed as

$$\begin{aligned}
Var\left(\widehat{F}_n(x)\right) &= \frac{1}{n}\Bigg[F(x) - bf(x) + bf(x)\int_{-1}^{1}K^2(t)dt + o\left(b\right) \\
&- \left\{F(x) + \frac{1}{2}b^2 f'(x)\mu_2(k) + o(b^2)\right\}^2\Bigg] \\
&= \frac{1}{n}F(x)\left(1 - F(x)\right) + \frac{b}{n}f(x)\left(\int_{-1}^{1}K^2(t)dt - 1\right) + o\left(\frac{b}{n}\right) \\
&= \frac{1}{n}F(x)\left(1 - F(x)\right) - \frac{b}{n}f(x)\left(2\int_{-1}^{1}tK\left(t\right)k(t)dt\right) + o\left(\frac{b}{n}\right) \\
&= \frac{1}{n}F(x)\left(1 - F(x)\right) - \frac{b}{n}f(x)\varphi\left(k\right) + o\left(\frac{b}{n}\right),
\end{aligned}$$

where

$$\varphi\left(k\right) = 2\int_{-1}^{1}tK\left(t\right)k(t)dt.$$

The previous result shows that the asymptotic variance of $\widehat{F}_n$ is smaller than the variance of the EDF. It is evident that for larger values of b, the quantity $bf(x)\varphi\left(k\right)$ increases, resulting in a smaller variance expression but larger bias. This observation has important implication for choosing the bandwith.

Several other properties of the estimator $\widehat{F}_n$ have been investigated. Nadaraya (1964)[45], Winter (1973) [64] and Yamato (1973) [66] proved almost uniform convergence of $\widehat{F}_n$ to F; Watson and Leadbetter (1964) [62] established asymptotic normality for $\widehat{F}_n$, and Winter (1979) [65] showed that $\widehat{F}_n$ has the Chung-Smirnov property, that

$$\limsup_{n\to\infty}\left\{\left(\frac{2n}{\log\log n}\right)^{1/2}\sup_{x\in\mathbb{R}}\left|\widehat{F}_n(x) - F(x)\right|\right\} \leq 1,$$

with probability 1. Reiss (1981) [49] pointed out that the loss in bias with respect to F_n is compensanted by a gain in variance. This result is referred to as the deficiency of F_n with respect to $\widehat{F}_n$ Falk (1983) [21] provided a comlete solution to the question as to which of F_n or $\widehat{F}_n$ is the better estimator of F. Using the concept of relative deficiency, conditions (as $n \to \infty$) on K and $b = b_n$ are derived, which enables the user to decide exactly whether a given kernel distribution function estimator should be preferred to the EDF.

Azzalini (1981) [3] derived also an asymptotic expression for the mean squared error MSE of $\widehat{F}_n(x)$ and determined the asymptotically optimal smoothing parameter, to have an MSE lower for F_n, and he obtained the asymptotic expressions for the mean integrated squared error $MISE$ of $\widehat{F}_n(x)$, for more details see (Mack, 1984 [39], and Hill, 1985 [30]).

In order to propose methods for estimating the bandwidth, discrepancy measures that quantify the quality of $\widehat{F}_n$ as an estimator for F must be introduced. One such measure is the mean squared error, which in the case of the kernel distribution function estimator is defined as

$$\begin{aligned} MSE\left(\widehat{F}_n(x)\right) &= E\left\{\left[\widehat{F}_n(x) - F(x)\right]^2\right\} \\ &= Bias^2\left(\widehat{F}_n(x)\right) + Var\left(\widehat{F}_n(x)\right) \\ &= \frac{1}{4} f'^2(x) h^4 \mu_2^2(k) \\ &+ \frac{1}{n} F(x)\left(1 - F(x)\right) - \frac{b}{n} f(x)\varphi\left(k\right) + o\left(b^4 + \frac{b}{n}\right), \end{aligned}$$

and the asymptotic expression of the $MSE\left(\widehat{F}_n(x)\right)$ is

$$\begin{aligned} AMSE\left(\widehat{F}_n(x)\right) &= \frac{1}{4} f'^2(x) h^4 \mu_2^2(k) \\ &\quad + \frac{1}{n} F(x)\left(1 - F(x)\right) - \frac{b}{n} f(x)\varphi\left(k\right). \end{aligned}$$

The value of b that minimizes the $AMSE\left(\widehat{F}_n(x)\right)$ is

$$\widehat{b} = \left(\frac{f(x)\varphi(k)}{n f'^2(x)\mu_2^2(k)}\right)^{1/3}.$$

The asymptotic mean integrated square error ($AMISE$) is found by integrating the

$AMSE\left(\widehat{F}_n(x)\right)$ which is

$$AMISE\left(\widehat{F}_n(x)\right) = \int_{-\infty}^{\infty}\left(\frac{1}{4}f'^2(x)h^4\mu_2^2(k) + \frac{1}{n}F(x)\left(1 - F(x)\right) - \frac{b}{n}f(x)\varphi\left(k\right)\right)dx.$$

The bandwidth which minimizes the $AMISE$ can be calculated by differentiating expression of the $AMISE\left(\widehat{F}_n(x)\right)$, setting the equation to 0 and solving it for b. The result is referred to as

$$\bar{b}=\left(\frac{\varphi(k)}{n\mu_2^2(k)\int f'^2(x)dx}\right)^{1/3}.$$

Remark 1.2.1

1. The choice of kernel k only affects the $AMISE$ through $\varphi\left(k\right)$ (larger values reduce the $AMISE$).

2. The estimator $\widehat{F}_n(x)$ is asymptotically more efficient than the $F_n(x)$ see (Swanapoel, 1988 [56]).

Chapter 2

Heavy-tailed distribution

Many distributions that are found in practice are heavy-tailed distributions. The first example of heavy-tailed distributions was found in Mandelbort (1963) [41] where it was shown that the change in cotton prices was heavy-tailed. Since then many other examples of heavy-tailed distributions are found, among these are data file in traffic on the internet Crovella and Bestavros (1997) [16], returns on financial markets Rachev (2003) [48], and Embrechts et al. (1997) [17].

Heavy-tailed distributions are probability distributions whose tails are not exponentially bounded : that is, they have heavier tails than the exponential distribution. In many applications it is the right tail of the distribution that is of interest, but a distribution may have a heavy left tail, or both tails may be heavy.

There is still some discrepancy over the use of the term heavy-tailed. There are two other definitions in use. Some authors use the term to refer to those distributions which do not have all their power moments finite, and some others to those distributions that do not have a finite variance. (Occasionally, heavy-tailed is used for any distribution that has heavier tails than the normal distribution)

2.1 Heavy-tailed distribution

We consider nonnegative random variables X, such as losses in investments or claims in insurance. For arbitrary random variables, we should consider both right and left tails.

The heavy-tailed distribution are related to extreme value theory and allow to model several phenomena encountered in different disciplines: finance, hydrology, telecommunications, geology... etc. Several definitions were associated with these distributions as a function of classification criteria. The characterization the most simple and one based on comparison with the normal distribution.

Definition 2.1.1 *It is said that the distribution has heavy tail if:*

$$\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] > 3. \tag{2.1}$$

where μ is the arithmetical mean, σ the standard deviation of rv X.

Which is equivalent to saying that a distribution to a heavy-tail if and only if its coefficient of applatissement, γ_2, is higher than normal distribution that is equal $\gamma_2 = 3$. The characterization given by equation (2.1) is very general and can be applied only if the moment of order 4 exists, therefore no discrimination, for distributions with a moment of order 4 is infinite can be made if considers that this criterion, unfortunately there is no test for all distributions under the right tail.

There are others definitions of heavy-tailed distribution. These definitions all relate to the decay of the survivor function $\overline{F}$ of a rv X.

Definition 2.1.2 (Tail function) *If F is the distribution function of X, we define the **tail function** or **survivor function** $\overline{F}$ on $\mathbb{R}_+$ by*

$$\overline{F}(x) = 1 - F(x) = P(X > x).$$

The tail of a distribution represents probability values for large values of the variable. When large values of the variable appear in a data set, their probabilities of occurrence are not zero.

Definition 2.1.3 *Let F be a df with support on $[0, \infty)$, we say that the distribution F, its corresponding nonnegative rv X, is heavy-tailed if it has no exponential moment*

$$\int_0^\infty e^{\lambda x} dF(x) = \infty, \quad \textit{for all } \lambda > 0.$$

Definition 2.1.4 *Let X a random variable with a distribution function F and the density f, this distribution is said to have a heavy tail if*

$$\overline{F}(x) = P(X > x) \sim x^{-\alpha}, \ \textit{as } x \to \infty,$$

where the parameter $\alpha > 0$ is called the tail index.

The distribution F is heavy-tailed if its **tail function** goes slowly to zero at infinity. For the next we need the following definition.

Definition 2.1.5 (Slowly varying function) *A positive measurable function S on $]0, \infty[$ is slowly varying at infinity if*

$$\lim_{x \to \infty} \frac{S(tx)}{S(x)} = 1, \ t > 0.$$

Thus, finally, here is the formal definition of heavy-tailed distributions:

Definition 2.1.6 *The distribution F is said to have a heavy tail if*

$$\overline{F}(x) = S(x) x^{-\alpha},$$

for some $\alpha > 0$ (called the tail index), and $S(.)$ is a slowly varying function at infinity.

2.1.1 Examples of heavy-tailed distributions

- **The Pareto distribution on** $\mathbb{R}_+$ **:** This has tail function $\overline{F}$ given by

$$\overline{F}(x) = \left(\frac{\lambda}{x+\lambda}\right)^{\alpha},$$

for some scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$. Clearly we have

$$\overline{F}(x) \sim (x/\lambda)^{-\alpha} \text{ as } x \to \infty,$$

and for this reason the Pareto distributions are sometimes referred to as the power law distributions. The Pareto distribution has all moments of order $\gamma < \alpha$ finite, while all moments of order $\gamma \geq \alpha$ are infinite.

- **The Burr distribution on** $\mathbb{R}_+$ **:** This has tail function $\overline{F}$ given by

$$\overline{F}(x) = \left(\frac{\lambda}{x^{\tau}+\lambda}\right)^{\alpha},$$

for parameters $\alpha, \lambda, \tau > 0$. We have

$$\overline{F}(x) \sim \lambda^{\alpha} x^{-\tau\alpha} \text{ as } x \to \infty,$$

thus the Burr distribution is similar in its tail to the Pareto distribution, of which it is otherwise a generalization. All moments of order $\gamma < \alpha\tau$ are finite, while those of order $\gamma \geq \alpha\tau$ are infinite.

- **The Cauchy distribution on** $\mathbb{R}$ **:** This is most easily given by its density function f where

$$f(x) = \frac{\lambda}{\pi\left((x-a)^2+\lambda^2\right)},$$

for some scale parameter $\lambda > 0$ and position parameter $a \in \mathbb{R}$. All moments are infinite.

- **The lognormal distribution on** $\mathbb{R}_+^*$ **:** This is again most easily given by its density function f, where
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right),$$
for parameters μ and $\sigma > 0$. The tail of the distribution F is then
$$\overline{F}(x) = \overline{\Phi}\left(\frac{\log x - \mu}{\sigma}\right) \text{ for } x > 0,$$
where $\overline{\Phi}$ is the tail of the standard normal random variable. All moments of the lognormal distribution are finite. Note that a (positive) random variable Y has a lognormal distribution with parameters μ and σ if and only if $\log Y$ has a normal distribution with mean μ and variance σ^2. For this reason the distribution is natural in many applications.

- **The Weibull distribution on** $\mathbb{R}_+$ **:** This has tail function $\overline{F}$ given by
$$\overline{F}(x) = e^{-(x/\lambda)^{\alpha}},$$
for some scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$. This is a heavy-tailed distribution if and only if $\alpha < 1$. Note that in the case $\alpha = 1$ we have the exponential distribution. All moments of the Weibull distribution are finite.

2.2 Classes of heavy-tailed distributions

An important classes of heavy-tailed distributions are that **regularity varying distribution** and **subexponential distribution.**

2.2.1 Regularity varying distribution functions

We introduce here the well-known class of heavy-tailed distributions is the class of regularly varying distribution functions.

Definition 2.2.1 (Regularity varying distribution) *A distribution function F on $\mathbb{R}$ is called regular varying at infinity with index $-\alpha < 0$ if the following limit holds*

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha},\ t > 0,$$

where $\overline{F}(x) = 1 - F(x)$ and the parameter α is called the tail index.

Definition 2.2.2 *A positive measurable function g on $]0, \infty[$ is regularly varying at infinity with index $\alpha \in \mathbb{R}$ if*

$$\lim_{x\to\infty} \frac{g(tx)}{g(x)} = t^{\alpha},\ t > 0,$$

we write $g(x) \in \mathcal{R}_\alpha$.

If $g(x) \in \mathcal{R}_\alpha$ and $\alpha = 0$ we call the function **slowly varying** at infinity. If $g(x) \in \mathcal{R}_\alpha$ we simply call the function $g(x)$ regularly varying and we can rewrite

$$g(x) = x^{\alpha} S(x),$$

where $S(x)$ is a slowly varying function.

The class of regularly varying distribution is closed under convolutions as can be found in Applebaum (2005) [1].

Proposition 2.2.1 (Regularly varying of convolution) *If F_1, F_2 are two distribution functions such that as $x \to \infty$:*

$$1 - F_i\,(x) = x^{-\alpha} S_i(x),\ \forall i = 1, 2,$$

with S_i is slowly varying, then the convolution $H = F_1 * F_2$ has a regularly varying tail such that :

$$1 - H(x) \sim x^{-\alpha} \left(S_1(x) + S_2(x) \right).$$

Remark 2.2.1 *If $\overline{F}(x) = x^{-\alpha} S(x)$ for $\alpha \geq 0$ and $S \in \mathcal{R}_0$, then for all $n \geq 1$,*

$$\overline{F^{n*}}(x) \sim n\overline{F}(x), \; x \to \infty,$$

where $\overline{F^{n*}}$ denotes the convolution of F $n-$times with itself. (See Embrechts et al. (1997) *[17]).*

An property of regularly varying distribution functions is that the $k-$th moment does not exist whenever $k \geq \alpha$, the mean and the variance can be infinite. This has a few important implications. When we consider a random variable that has a regularly varying distributions with a tail index less than one, then the mean of this random variable is infinite, and if we consider the sum of independent and identically distributed random variables that have a tail index $\alpha < 2$, the means that the variance of these random variables is infinite, and hence the central limit theorem does not hold for these random variables see Uchaikin and Zolotarev (1999) [58].

A more detail on regularly varying distribution functions is found in Bingham et al. (1987) [5].

The following table gives a particular examples of regularly varying distributions.

Distribution	$\overline{F}(x)$ **or** $f(x)$	**Index of regular variation**
Pareto	$\overline{F}(x) = \left(\dfrac{\lambda}{x+\lambda}\right)^{\alpha}$	$-\alpha$
Burr	$\overline{F}(x) = \left(\dfrac{\lambda}{x^{\tau}+\lambda}\right)^{\alpha}$	$-\tau\alpha$
Log-Gamma	$f(x) = \dfrac{\alpha^{\beta}}{\Gamma(\beta)} (\ln(x))^{\beta-1} x^{-\alpha-1}$	$-\alpha$

Table 2.1: Regularly varying distribution functions

2.2.2 Subexponential distribution functions

Subexponential distributions are a special class of heavy-tailed distributions. The name arises from one of their properties, that their tails decrease more slowly than any exponential tail, see Goldie (1978) [27]. This implies that large values can occur in a sample with non-negligible probability, and makes the subexponential distributions candidates for modelling situations where some extremely large values occur in a sample compared to the mean size of the data. Such a pattern is often seen in insurance data, for instance in fire, wind-storm or flood insurance (collectively known as catastrophe insurance). Subexponential claims can account for large fluctuations in the surplus process of a company, increasing the risk involved in such portfolios.

Definition 2.2.3 (Subexponential distribution function) *Let $X_1, ..., X_n$ be iid positive random variables with df F such that $0 < F(x) < 1$ for all $x > 0$.*
Denote

$$P(\max(X_1 + ... + X_n) > x) = \overline{F^n}(x) \sim n\overline{F}(x) \text{ as } x \to \infty,$$

and

$$P(X_1 + ... + X_n > x) = \overline{F^{n*}}(x) = 1 - F^{n*}(x),\ x \geq 0,$$

the tail of the $n-$fold convolution of F. F is a subexponential df ($F \in \mathcal{S}$) if one of the following equivalent conditions holds:

(1) $\lim\limits_{x\to\infty} \dfrac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n$ *for some (all)* $n \geq 2$,

(2) $\lim\limits_{x\to\infty} \dfrac{P(X_1 + ... + X_n > x)}{P(\max(X_1 + ... + X_n) > x)} = 1$ *for some (all)* $n \geq 2$.

Lemma 2.2.1 *If the following equation holds*

$$\limsup_{x\to\infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} = 2,$$

then $F \in \mathcal{S}$.

Proof. See Foss et al. (2013) [24]. ■

The following lemma give a few important properties of subexponential distributions:

Lemma 2.2.2 *If* F *is subexponential then for all* $t \geq 0$

$$\lim_{x\to\infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} = 1.$$

Proof. See Chistyakov (1964) [12]. ■

Lemma 2.2.3 *Let* F *be subexponential and* $r > 0$. *Then*

$$\lim_{x\to\infty} e^{rx}(\overline{F}(x)) = \infty,$$

in particular

$$\int_0^\infty e^{rx} dF(x) = \infty.$$

Proof. See Embrechts et al. (1997) [17]. ■

Next we give an upper bound for the tails of the convolutions.

Lemma 2.2.4 *Let* F *be subexponential. Then for any* $\epsilon > 0$ *there exist a* $D \in \mathbb{R}$ *such that*

$$\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq D(1+\epsilon)^n$$

for all $x > 0$ *and* $n \geq 2$.

Proof. See Embrechts et al. (1997) [17]. ∎

Remark 2.2.2

1. *Definition* (1) *goes back to Chistyakov* (1964) [*12*]. *He proved that the limit* (1) *holds for all* $n \geq 2$ *if and only if it holds for* $n = 2$. *It was shown in Embrechts and Goldie* (1982) [*19*] *that* (1) *holds for* $n = 2$ *if it holds for some* $n \geq 2$.

2. *The equivalence of* (1) *and* (2) *was shown in Embrechts and Goldie* (1980) [*18*].

3. *Definition* (2) *provides a physical in terpretation of subexponentiality : the sum of* n *iid subexponential rv is likely to be large if and only if their maximum is likely to be large. This accounts for extremely large values in a subexponential sample.*

4. *From Definition* (1) *and the fact that* $\mathcal{S}$ *is closed with respect to tail equivalence we conclude that*

$$F \in \mathcal{S} \quad \Longrightarrow \quad F^{n*} \in \mathcal{S} \quad , \ n \in \mathbb{N},$$

Furthermore, from Definition (2) *and the fact that* F^n *is the df of the maximum of* n *iid rv with df* F, *we conclude that*

$$F \in \mathcal{S} \quad \Longrightarrow \quad F^{n} \in \mathcal{S} \quad , \ n \in \mathbb{N}.$$

Hence $\mathcal{S}$ *is closed with respect to taking sums and maxima of iid random variables.*

For an more explication of subexponential distribution, one refers to, for instance, Foss et al. (2013) [24]. and Embrechts and Goldie (1980) [18]

The following table gives a number of subexponential distribution:

Distribution	$\overline{F}\ (x)$ or $f(x)$	Parameters
Weibull	$\overline{F}(x) = e^{-\lambda x^{\tau}}$	$\lambda > 0,\ 0 < \tau < 1$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\dfrac{(\log x - \mu)^2}{2\sigma^2}\right)$	$\mu \in \mathbb{R},\ \sigma > 0$
Benktender-type I	$\overline{F}\ (x) = \left(1 + 2\dfrac{\beta}{\alpha} \ln x\right) e^{-\beta(\ln x)^2 - (\alpha+1)\ln x}$	$\alpha, \beta > 0$
Benktender-type II	$\overline{F}\ (x) = e^{\dfrac{\alpha}{\beta}} x^{-(1-\beta)} e^{-\alpha \frac{x^{\beta}}{\beta}}$	$\alpha > 0,\ 0 < \beta < 1$

Table 2.2: Subexponential distribution

We give now two more classes of heavy-tailed distributions. We begin by the class of dominated varying distribution functions denoted by $\mathcal{D}$:

Definition 2.2.4 *We say that F is a dominated-varying distribution if there exists $c > 0$ such that*

$$\overline{F}(2x) \geq c\overline{F}(x) \text{ for all } x.$$

The class of dominated varying distribution functions denoted by $\mathcal{D}$

$$\mathcal{D} = \left\{ F,\ df \text{ on }]0, \infty[:\ \limsup_{x \to \infty} \frac{\overline{F}(x/2)}{\overline{F}(x)} < \infty \right\}.$$

The final class of distribution functions is the class of long tailed distributions, denoted by $\mathcal{L}$

$$\mathcal{L} = \left\{ F,\ df \text{ on }]0, \infty[:\ \lim_{x \to \infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} = 1 \text{ for all } t > 0 \right\}.$$

Chapter 3

Transformation in kernel density estimation for heavy-tailed distributions

It is well known now that kernel density estimators are not consistent when estimating a density near the finite end points of the support of the density to be estimated. This is due to boundary effects that occur in nonparametric curve estimation problems. A number of proposals have been made in the kernel density estimation context with some success. As of yet there appears to be no single dominating solution that corrects the boundary problem for all shapes of densities.

Consequently, an idea on how to include boundary corrections in these estimators is presented. The first statement implies that the density has a support which is bounded on the left hand side. Without loss of generality the support is set to be $[0, \infty)$. Concerned the kernel estimation for heavy-tailed distributions has been studied by several authors Bolancé et al. (2003) [6], Clements et al. (2003) [13] and Buch-Larsen et al. (2005) [7] propose different parametric transformation families that they all make the transformed distribution more symmetric that the original one, which in many applications has usually a strong right-hand asymmetry. Buch-Larsen et al. (2005) [7] propose an alternative transformation such as one based on the Champernowne distribution, who they have shown in studies that this transformation is preferable to other transformation in density estimation approach for heavy-tailed distribution.

3.1 Kernel density estimator and boundary effects

Nonparametric kernel density estimation is now popular and in wide use with great success in statistical applications. Kernel density estimates are commonly used to display the shape of a data set without relying on a parametric model, not to mention the exposition of skewness, multimodality, dispersion, and more. Early results on kernel density estimation are due to Rosenblatt (1956) [51] and Parzen (1962) [47]. Since then, much research has been done in the area; see the monographs of Silverman (1986) [55], and Wand and Jones (1995) [61].

Consider a density function f which is continuous on $[0,\infty)$ and is 0 for $x<0$. Given a bandwidth b, the interval $[0,b]$ is defined to be the boundary interval and $]b,a-b]$, $0<a\leq\infty$, the interior interval, and consider nonparametric estimation of the unknown density function f based on a random sample $X_1,...,X_n$. Suppose that f' and f'' are the first and second derivatives of f, exists and is continuous on $[0,b]$. Then the standard kernel estimator of f is given by

$$\hat{f}_n(x)=\frac{1}{nb}\sum_{i=1}^{n}k\left(\frac{x-X_i}{b}\right), \tag{3.1}$$

where k is a symmetric density function with support $[-1,1]$ and b is the bandwidth. The basic properties of $\hat{f}_n(x)$ at interior points are well-known see Silverman (1986) [55], and under some smoothen assumptions these include, for $b<x\leq a-b$, $0<a\leq\infty$,

$$E\left(\hat{f}_n(x)\right)-f(x)=\frac{1}{2}\mu_2(k)f''(x)\,b^2+o\left(b^2\right),$$

and

$$Var\left(\hat{f}_n(x)\right) = \frac{1}{nb} f(x) \int_0^\infty k^2(x)dx + o\left(\frac{1}{nb}\right).$$

The bias of $\hat{f}_n(x)$ is of order $o\left(b^2\right)$, whereas at boundary points, for $x \in [0, b] \cup (a-b, a]$, $\hat{f}_n$ is not even consistent. In nonparametric curve estimation problems this phenomenon is referred to as the "boundary effects". Problems will arise if x is smaller than the chosen bandwidth b. This fact can be clearly seen by examining the behavior of $\hat{f}_n(x)$ inside the left boundary region $[0, b]$. Let x be a point in the left boundary, $x \in [0, b]$. Then we can write for $x = sb$, $0 \leq s \leq 1$:

$$\begin{aligned} E\left(\hat{f}_n(x)\right) &= E\left(\frac{1}{b} k\left(\frac{x - X_i}{b}\right)\right) \\ &= \frac{1}{b} \int_0^\infty k\left(\frac{x-z}{b}\right) f(z)dz. \end{aligned}$$

We used the change of variable $t = (x - z)/b$, we have

$$E\left(\hat{f}_n(x)\right) = \int_{-1}^{s} k\left(t\right) f\left(x - bt\right) dt.$$

Assuming that f'' exists and is continuous in a neighborhood of x, the density in the integral can be approximated by its second order Taylor expansion evaluated at x:

$$\begin{aligned} f(x - bt) = {} & f(x) + (x - bt - x) f'\left(x\right) \\ & + \frac{1}{2}\left(x - bt - x\right)^2 f''\left(x\right) + o\left(b^2\right), \end{aligned}$$

given for $b \to 0$ and $t \in [-1, 1]$,

$$E\left(\hat{f}_n(x)\right) = f(x)\int_{-1}^{s} k(t)\,dt - bf'(x)\int_{-1}^{s} tk(t)\,dt + \frac{b^2}{2}f''(x)\int_{-1}^{s} t^2 k(t)\,dt + o\left(b^2\right),$$

and

$$Var\left(\hat{f}_n(x)\right) = \frac{1}{nb}f(x)\int_{-1}^{s} k^2(t)\,dt + o\left(\frac{1}{nb}\right).$$

It is now clear that the bias of $\hat{f}_n(x)$ is of order $o(b)$ instead of $o(b^2)$, the variance isn't much changed.

Example 3.1.1 *The boundary problem can be detected in figure* (3.1)*. The theoretical curve is that of the pareto density.*

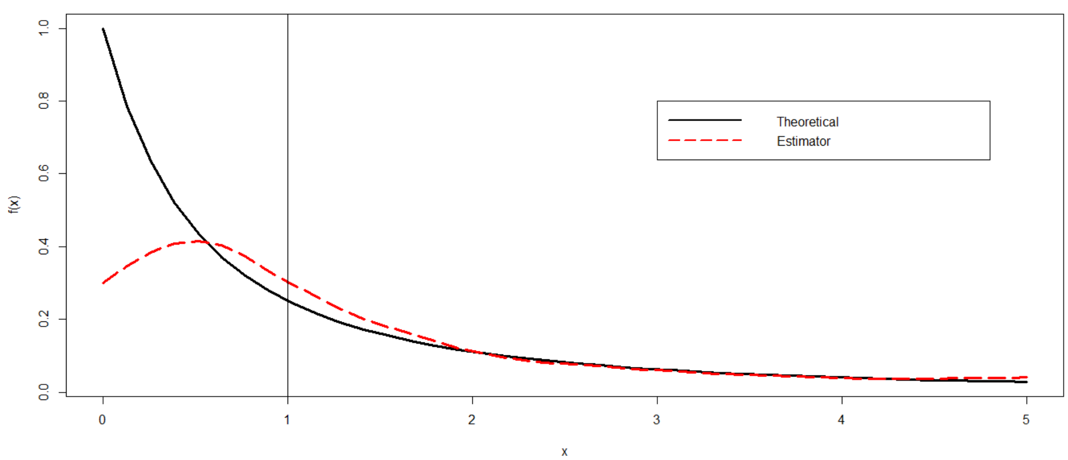

Figure 3.1: Boundary effect in kernel density estimation

3.2 Methods for removing boundary effects

The properties of the classical kernel methods are satisfactory, but when the support of the variable is bounded, kernel estimates may suffer from boundary effects. Therefore, the so-called boundary correction is needed in kernel estimation. Removing boundary effects in kernel density estimation can be done in various methods. Some methods were selected which seemed to be reasonable. There were methods which were rather complicated and others which on the other hand felt quite natural.

- **The reflection method**

  The reflection method is introduced by Schuster (1985) [53], then study by Cline and Hart (1991) [14]. this method specifically designed for the case $f'(0) = 0$, where f' denotes the first derivative of f. Simplest way is to reflect the data points $X_1, ..., X_n$ at the origin, just add $-X_1, ..., -X_n$ to the data set. This is usually referred to as the reflection estimator and it can also be formulated as

$$\widehat{f}_R(x) = \frac{1}{nb}\sum_{i=1}^{n}\left\{k\left(\frac{x - X_i}{b}\right) + k\left(\frac{x + X_i}{b}\right)\right\}, \text{ for } x \geq 0,$$

  for $x < 0$, $\widehat{f}_n(x) = 0$.

- **Transformation of data method**

  The transformation idea is based on transforming the original data $X_1, ..., X_n$ to $g(X_1), ..., g(X_n)$, where g is a non-negative, continuous and monotonically increasing function from $[0, \infty)$ to $[0, \infty)$. Based on the transformed data, the estimator (3.1) becomes:

$$\hat{f}_T(x) = \frac{1}{nb}\sum_{i=1}^{n}k\left(\frac{x - g(X_i)}{b}\right).$$

  Note this isn't really estimating the density function of X, but instead of $g(X)$

- **Pseudo-Data Methods**

  The pseudo-data method estimator is defined (see Cowling and Hall (1996) [15]), this generates data beyond the left endpoint of the support of the density.

$$\widehat{f}_{CH}(x) = \frac{1}{nb}\left\{\sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right) + \sum_{i=1}^{m} k\left(\frac{x + X_{(-i)}}{b}\right)\right\}.$$

where

$$X_{(-i)} = -5X_{(i/3)} - 4X_{(2i/3)} + \frac{10}{3}X_{(i)},\ \ i = 1, 2, ..., n,$$

and $X_{(i)}$ is the ith-order statistic of sample $X_1, ..., X_n$, and m is an integer such that $nb < m < n$.

- **Boundary kernel method**

The boundary kernel method is more general than the reflection method in the sense that it can adapt to any shape of density. However, a drawback of this method is that the estimates might be negative near the endpoints; especially when $f(0) \approx 0$.The boundary kernel and related methods usually have low bias but the price for that is an increase in variance. The boundary kernel estimator with bandwidth variation is defined (see Zhang and Karunamuni (1998) [67]) as

$$\hat{f}_B(x) = \frac{1}{nb_s}\sum_{i=1}^{n} k_{(s/\varphi(s))}\left(\frac{x - X_i}{b_s}\right),$$

where $s = \min\{x/b, 1\}$, $k_{(s/\varphi(s))}$ is a boundary kernel satisfying $k_{(1)}(t) = k(t)$, and $b_s = \varphi(s)b$ with $\varphi(s) = 2 - s$. Also

$$k_{(s/\varphi(s))}(t) = \frac{12}{(1+s)^4}(1+t)\left\{(1-2s)\,t + \frac{3s^2 - 2s + 1}{2}\right\} I_{\{-1 \le t \le s\}}.$$

- **Reflection and transformation methods**

The reflection estimator computes the estimate density based on the original and the reflected data points. Unfortunately, this does not always yield a satisfying result since this estimator enforces the shoulder condition and still contains a bias of order b if the density does not fulfill this condition. The generalized reflection and transformation density estimators introduce by Karunamuni and Alberts (2005) [34] and is given by

$$\widehat{f}_{RT}(x) = \frac{1}{nb}\sum_{i=1}^{n}\left\{k\left(\frac{x+g\left(X_i\right)}{b}\right)+k\left(\frac{x-g\left(X_i\right)}{b}\right)\right\}.$$

where g is a transformation that need to be determined.

We refer to Baszczyńska (2016) [4], Karunamuni and Alberts (2005) [34] and Koláček and Karunamuni (2009) [38] for more details about this methods and for other methods see Zhang et al. (1999) [68].

Now for remove the boundary effect in density estimation of heavy-tail distributions, we investigate a new class of estimators based on a transformation of set of the original data by the Champernowne distribution function.

3.3 Champernowne distribution

Buch-Larsen et.al. (2005) [7] used modified Champernowne distribution to estimate loss distributions in insurance which is categorically heavy-tailed distributions. Some time it is difficult to find a parametric model which is simple and fit for all values of claim in the insurance industry. Gustafsson et.al. (2007) [28] used asymmetric kernel density estimation to estimate actuarial loss distributions. The new estimator of density function is obtained by transforming the data using generalized Champernowne distribution function, because it produces good results in all the studied situations and it is straightforward to apply.

The original Champernowne distribution has density

$$f(x) = \frac{C}{x\left((1/2)\,(x/M)^{-\alpha} + \lambda + (1/2)\,(x/M)^{\alpha}\right)},\ \ x \geq 0,$$

where C is a normalizing constant and α, λ and M are parameters. The distribution was mentioned for the first time in 1936 by D.G. Champernowne when he spoke on The Theory of Income Distribution at the Oxford Meeting of the Econometric Society see, Champernowne (1936) [8], Champernowne (1937) [9]. Later, he gave more details about the distribution in Champernowne (1952) [10], and its application to economics. When λ equals to one and the normalizing constant c equals $(1/2)\alpha$, the density of the original distribution is simply called the Champernowne. Champernowne cumulative distribution function is defined on $x \geq 0$ and has the form

$$F(x) = \frac{x^{\alpha}}{x^{\alpha} + M^{\alpha}},$$

with parameter $\alpha > 0$, $M > 0$, and density function is of the form

$$f(x) = \frac{\alpha M^{\alpha} x^{\alpha-1}}{\left(x^{\alpha} + M^{\alpha}\right)^{2}}.$$

The Champernowne distribution converges to a Pareto distribution in the tail, while looking more like a lognormal distribution near 0 when $\alpha > 1$. Its density is either 0 or infinity at 0 (unless $\alpha = 1$).

3.3.1 Modified Champernowne distribution

We generalize the Champernowne distribution with a new parameter c. This parameter ensures the possibility of a positive finite value of the density at 0 for all α.

Definition 3.3.1 *The modified Champernowne cumulative df is defined for $x \geq 0$ and* *has the form*

$$T(x) = \frac{(x+c)^{\alpha} - c^{\alpha}}{(x+c)^{\alpha} + (M+c) - 2c^{\alpha}}, \ \forall x \in \mathbb{R}_+,$$

with parameter $\alpha > 0$, $M > 0$ *and* $c \geq 0$, *and its density is*

$$t(x) = \frac{\alpha\,(x+c)^{\alpha-1}\left((M+c)^{\alpha} - c^{\alpha}\right)}{\left((x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}\right)^2}, \ \forall x \in \mathbb{R}_+.$$

Corresponding to the Champernowne distribution, the modified Champernowne distribution converges to a Pareto distribution in the tail, for the large values of x:

$$t(x) \to \frac{\alpha\left(\left((M+c)^{\alpha} - c^{\alpha}\right)^{1/\alpha}\right)^{\alpha}}{x^{\alpha+1}}.$$

A crucial step when using the Champernowne distribution, is the choice of parameter estimators. As described in Buch-Larsen et al. (2005) [7], a natural way is to recognize that $T(M) = 1/2$ and therefore estimate the parameter M as the empirical median, and then estimate (α, c) by maximizing the log-likelihood function

$$\begin{aligned} l(\alpha, c) = n\log(\alpha) + n\log((M+c)^{\alpha} - c^{\alpha}) + (\alpha - 1)\sum_{i=1}^{n}\log(X_i + c) \\ - 2\sum_{i=1}^{n}\log\left((X_i + c) + (M+c)^{\alpha} - 2c^{\alpha}\right). \end{aligned}$$

The choice of M as the empirical median, especially for heavy-tailed distributions, and the maximum likelihood estimates of $(\alpha,\, c)$ ensures the best over-all fit of the distribution.

Remark 3.3.1 *The effect of the additional parameter* c *is different for* $\alpha > 1$ *and for* $\alpha < 1$. *The parameter* c *has some scale parameter properties: when* $\alpha < 1$, *the derivative of the cumulative df becomes larger for increasing* c, *and conversely, when* $\alpha > 1$, *the derivative of the df becomes smaller for increasing* c. *When* $\alpha \neq 1$, *the choice of* c *affects the density in three ways.*

First, c *changes the density in the tail. When* $\alpha < 1$, *positive* c *result in lighter tails, and* *the opposite when $\alpha > 1$.*

Secondly, *c changes the density in 0. A positive c provides a positive finite density in 0 :*

$$0 < t(0) = \frac{\alpha c^{\alpha-1}}{(M+c)^{\alpha} - c^{\alpha}} < \infty, \ \textit{when } c > 0.$$

Thirdly, *c moves the mode. When $\alpha > 1$, the density has a mode, and positive c shift the mode to the left. We therefore see that the parameter c also has a shift parameter effect. When $\alpha = 1$, the choice of c has no effect.*

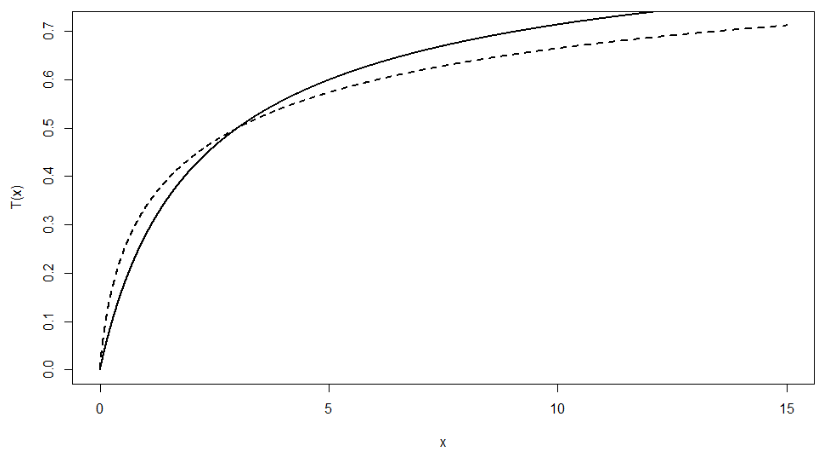

Figure 3.2: Modified Champernowne distribution function, $(M = 3; \alpha = 0.5)$. $c = 0$ dashed line and $c = 2$ solid line.

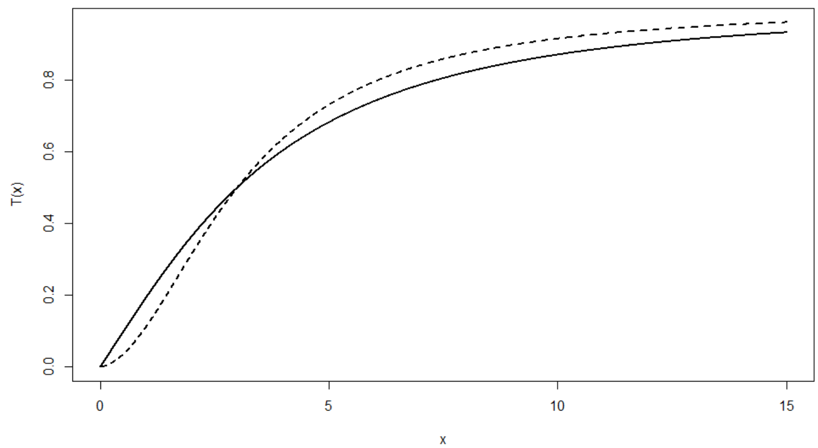

Figure 3.3: Modified Champernowne distribution function, $(M = 3; \alpha = 2)$. $c = 0$ dashed line and $c = 2$ solid line.

3.4 Density estimation using Champernowne transformation

Consider a sample random of size n, $X_1, ..., X_n$, from unknown df, F or density function f. We will make a detailed derivation of the density estimator based on the modified Champernowne distribution. This estimator is obtained by transforming the data set with a parametric estimator. The estimator of M is the empirical median and the likelihood estimator of α and c are the values which maximize likelihood function and afterwards estimating the density of the transformed data set using the classical kernel density estimator (3.1). The estimator of the original density is obtained by back-transformation.

Lemma 3.4.1 *Using transformation $y = T(x)$, then*

$$g(y) = f\left(T^{-1}(y)\right) \frac{1}{\left|t\left(T^{-1}(y)\right)\right|},$$

and

$$f(x) = g\left(T(x)\right) t\left(x\right) = g(T(x)) \frac{1}{\left|\left(T^{-1}\right)'(x)\right|},$$

where $t\left(x\right) = T'(x)$.

Proof. For $y = T(x)$, $x = T^{-1}\left(y\right)$ and $t\left(x\right) = \dfrac{dT(x)}{dx}$. The density function of variable X is $f(x)$ and $F(x)$ its cumulative df. Note that $G\left(y\right)$ cumulative df of variable Y and $g(y)$ its density function, then

$$\begin{aligned} G\left(y\right) &= P\left(Y \leq y\right) \\ &= P\left(X \leq T^{-1}\left(y\right)\right) \\ &= F\left(T^{-1}\left(y\right)\right), \end{aligned}$$

and

$$\begin{aligned} g(y) &= \frac{dG(y)}{dy} \\ &= \frac{dF\left(T^{-1}\left(y\right)\right)}{dy} \\ &= f\left(T^{-1}\left(y\right)\right) \left|\frac{dT^{-1}\left(y\right)}{dy}\right| \\ &= f\left(T^{-1}\left(y\right)\right) \frac{1}{\left|\frac{dT(x)}{dx}\right|} \\ &= f\left(T^{-1}\left(y\right)\right) \frac{1}{\left|t\left(T^{-1}\left(y\right)\right)\right|}. \end{aligned}$$

For $x = T^{-1}\left(y\right), y = T(x)$,

$$\begin{aligned} f(x) &= g(T(x)) \left|\frac{dT(x)}{dx}\right| \\ &= g(T(x)) \left|t(x)\right| \\ &= g(T(x)) \frac{1}{\left|\left(T^{-1}\right)'(x)\right|}. \end{aligned}$$

This achieves the proof of Lemma 3.4.1. ■

Theorem 3.4.1 *Given a set of data $X_1, ..., X_n$, cumulative df T, is the modified Champernowne distribution function, then*

$$Y_i = T(X_i),\ i = 1, ..., n,$$

are new variable, Y_i is in the interval $[0, 1]$ and uniform distributed, then the density function for transform data is

$$g(y) = f\left(T^{-1}(y)\right) \frac{1}{\left|t\left(T^{-1}(y)\right)\right|}.$$

and the formulation of the kernel density estimation for transform data $Y_1, ..., Y_n$ is

$$\widetilde{g}_n(y) = \frac{1}{nb} \sum_{i=1}^{n} k\left(\frac{y - Y_i}{b}\right),$$

where $k(.)$ is kernel function.

Boundary correction, is needed since y are in the interval $[0, 1]$, it is necessary to have a boundary correction to ensure that the kernel density estimator for transform data is a consistent estimator at the boundary. We use a simple renormalization method, as described in Jones (1993) [31] which ensures that each kernel function integrates to 1. The formula kernel density estimator for transform data $Y_1, ..., Y_n$ with the boundary correction is so

$$\widetilde{g}_n(y) = \frac{1}{nbk_y} \sum_{i=1}^{n} k\left(\frac{y - Y_i}{b}\right),$$

where

$$k_y = \int_{\max(-1, -y/b)}^{\max(1, (1-y/b))} k(u)du.$$

Using Theorem 3.4.1 kernel density estimation for data X_i, $i = 1, ..., n$ is;

$$\widetilde{f}_n(x) = \frac{\widetilde{g}_n(T(x))}{|(T^{-1})'(x)|}.$$

The formula of transformation kernel density estimation is

$$\widetilde{f}_n(x) = \frac{1}{nbk_{T(x)}}\sum_{i=1}^{n} k_b\left(\frac{T(x) - T(X_i)}{b}\right) T'(x)$$

3.4.1 Asymptotic theory for the transformation kernel density estimator

We investigate the asymptotic theory of the transformation kernel density estimator. Buch-Larsen et.al. (2005) [7], presented a theorem about the asymptotic theory of the transformation kernel density estimator in general (asymptotic bias and variance).

Theorem 3.4.2 *Let $X_1, ..., X_n$ be independent identically distributed variables with density f. Let $\widetilde{f}_n(x)$ be the transformation kernel density estimator of $f(x)$*

$$\widetilde{f}_n(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{T(x) - T(X_i)}{b}\right) T'(x),$$

where $T(\cdot)$ is the transformation function.
Then the bias and the variance of $\widetilde{f}_n(x)$ are given by

$$E\left(\widetilde{f}_n(x)\right) - f(x) = \frac{1}{2}\mu_2(k)\, b^2 \left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'(x)}\right)' + o\left(b^2\right),$$

and

$$Var\left(\widetilde{f}_n(x)\right) = \frac{1}{nb} R(k)\, T'(x) f(x) + o\left(\frac{1}{nb}\right).$$

as $n \to \infty$, where $\mu_2(k) = \int u^2 k(u)dx$ and $R(k) = \int k^2(u)dx$.

Proof. The variable transformation $Y_i = T(X_i)$ has the density g such as

$$g(y) = \frac{f\left(T^{-1}(y)\right)}{T'(T^{-1}(y))}.$$

Let $\widetilde{g}_n(y)$ be the classical kernel density estimator of $g(y)$

$$\widetilde{g}_n(y) = \frac{1}{nb}\sum_{i=n}^{n} k\left(\frac{y - Y_i}{b}\right).$$

The mean and variance of the classical kernel density estimator

$$E\left(\widetilde{g}_n(y)\right) = g(y) + \frac{1}{2}\mu_2(k)\, b^2 g''(y) + o\left(b^2\right),$$

and

$$Var\left(\widetilde{g}_n(y)\right) = \frac{1}{nb}R(k)g(y) + o\left(\frac{1}{nb}\right).$$

The expression of the kernel estimator of density through the transformation by the standard kernel estimator of density is:

$$\widetilde{f}_n(x) = T'(x)\,\widetilde{g}_n(T(x)).$$

Then

$$\begin{aligned} E\left(\widetilde{f}_n(x)\right) &= T'(x)\, E\left(\widetilde{g}_n(T(x))\right) \\ &= T'(x)\left[g(T(x)) + \frac{b^2}{2}g''(T(x))\,\mu_2(k) + o\left(b^2\right)\right], \end{aligned}$$

we have

$$\begin{aligned} g\left(T\left(x\right)\right) &= \frac{f(x)}{T'\left(x\right)} \\ g'\left(T\left(x\right)\right) &= \frac{dg\left(T\left(x\right)\right)}{dT\left(x\right)} \\ &= \frac{dg\left(T\left(x\right)\right)}{dx} . \frac{dx}{dT\left(x\right)} \\ &= \left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'\left(x\right)}, \end{aligned}$$

and

$$\begin{aligned} g''\left(T\left(x\right)\right) &= \frac{d}{dT\left(x\right)}\left(g'\left(T\left(x\right)\right)\right) \\ &= \frac{d}{dx}\left(g'\left(T\left(x\right)\right)\right)\frac{dx}{dT\left(x\right)} \\ &= \left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'\left(x\right)}\right)\frac{1}{T'\left(x\right)}, \end{aligned}$$

$$E\left(\widetilde{f}_n\left(x\right)\right) - f(x) = \frac{1}{2}\mu_2\left(k\right)b^2\left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'\left(x\right)}\right)' + o\left(b^2\right),$$

and

$$\begin{aligned} Var\left(\widetilde{f}_n\left(x\right)\right) &= \left(T'\left(x\right)\right)^2 Var\left(\widetilde{g}_n\left(T(x)\right)\right) \\ &= \left(T'\left(x\right)\right)^2\left[\frac{1}{nb}R(k)g(y) + o\left(\frac{1}{nb}\right)\right] \\ &= \frac{1}{nb}T'\left(x\right)R\left(k\right)f(x) + o\left(\frac{1}{nb}\right). \end{aligned}$$

This completes the proof of Theorem 3.4.2. ■

Example 3.4.1 *Taking boundary problem for rv X with pareto distribution with parameter $(\alpha, \beta) = (1, 1)$ and sample size $n = 500$. Graphical output figure (3.4) illustrates the boundary correction by the transformation method.*

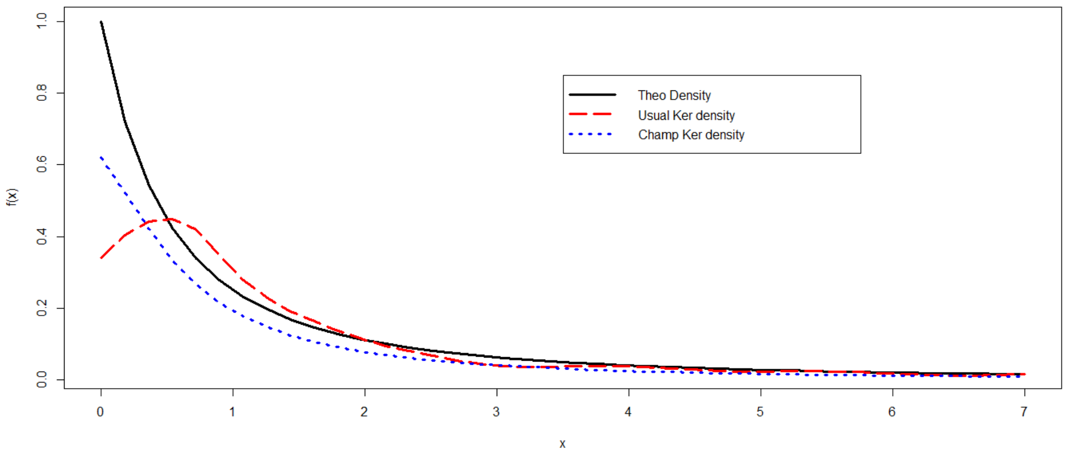

Figure 3.4: Usual kernel estimator and Champernowne transformation estimator.

Chapter 4

A modified Champernowne transformation to improve boundary effect in kernel distribution estimation

Abstract. Kernel distribution estimators are not consistent when estimating a distribution function near the boundary of its support. This problem is due to boundary effects. Several solutions to this problem have already been proposed. In this paper, we propose an estimator for heavy-tailed distributions using the boundary kernel distribution estimator by transforming the data set with a modification of the Champernowne distribution function. The asymptotic bias, variance and mean squared error of the proposed estimator are determined. In a simulation studies, we show that the proposed method performs quite well when compared with the existing methods.

Key words: Transformation; Boundary effect; Kernel distribution estimation; Mean Square Error; MeanIntegrated Equare Error.

AMS 2010 Subject Classification: 62G07; 62G20.

4.1 Introduction

Let X be a real random variable (rv) with unknown continuous distribution function (cdf) F and density function f. An important statistical problem is the estimation of a cdf F. A simple or the classic nonparametric estimator of the cdf is the empirical distribution function estimator. But, these estimators are step functions, and therefore, they have undesirable properties. To overcome these disadvantages, smoothing versions of them are often used. Among them kernel smoothing is most widely used because it is easy to derive and has good properties. Kernel smoothing has received a lot of attention in density estimation contex (see, e.g., Silverman (1986) [55], Wand and Jones (1995) [61]). Specifically, let $X_1, ..., X_n$ be a sample of size $n \geq 1$ from the rv X. The popular nonparametric kernel estimator of f which is introduced by Rosenblatt (1956) [51] and Parzen (1962) [47] and has the form

$$\widehat{f}_n(x) = \frac{1}{nb} \sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right),$$

where $b := b_n$ is the bandwidth or the smoothing parameter ($b \longrightarrow 0$, as $n \longrightarrow \infty$) and k is a nonnegative symmetric kernel function such that it is bounded and has finite support. The kernel distribution function estimator $\widehat{F}_n(x)$ was proposed by Nadaraya (1964) [45]. Such an estimator arises as an integral of the Parzen-Rosenblatt kernel density estimator (see Reiss (1981) [49] and Tenreiro (2013)) [57] and is defined for $x \in \mathbb{R}$, by

$$\widehat{F}_n(x) = \int_{-\infty}^{x} \widehat{f}_n(t)dt = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right), \tag{4.1}$$

where

$$K(x) := \int_{-\infty}^{x} k(t)dt,$$

is the integrated kernel. However, several properties of $\widehat{F}_n(x)$ have been investigated, Azzalini (1981) [3] have derived an asymptotic expression for the mean squared error of

$\widehat{F}_n(x)$, and determined also the asymptotically optimal smoothing parameter. Winter (1979) [65] and Yamato (1973) [66] proved the uniform convergence of $\widehat{F}_n(x)$ to $F(x)$ with probability one, the asymptotic normality of $\widehat{F}_n(x)$ is established by Watson and Leadbetter (1964) [62].

The problems of boundary effect for kernel estimators with compact supports is well-known in regression and density function estimation and several modified estimators have been proposed in the literature (see Gasser and Müller (1979) [25], Karunamuni and Alberts (2005) [34], Zhang and Karunamuni (1999) [68], and references therein). A similar correction would be made for improve the theoretical performance of the usual kernel distribution function estimator (4.1), at the boundary points. More specifically the performance of $\widehat{F}_n(x)$ at boundary points, for $x \in [0, b] \cup (a-b, a]$, $0 < a \leq \infty$, however differs from the interior points due to so-called "boundary effects" that occur in nonparametric curve estimation problems. The bias of $\widehat{F}_n(x)$ is of order $o(b)$ instead of $o(b^2)$ at boundary points, while the variance of $\widehat{F}_n(x)$ is of order $o\left(\frac{b}{n}\right)$. This fact can be clearly seen by examining the behavior of $\widehat{F}_n$ inside the left boundary region $[0, b]$. Let x be a point in the left boundary region, $x \in [0, b]$. The bias and variance of $\widehat{F}_n(x)$ at $x = sb$, $0 \leq s \leq 1$ are

$$\begin{aligned} Bias\left(\widehat{F}_n(x)\right) &= bf\left(0\right)\int_{-1}^{-s} K(t)dt \\ &\quad + b^2 f'\left(0\right)\left\{\frac{s^2}{2} + s\int_{-1}^{-s} K(t)dt - \int_{-1}^{s} tK(t)dt\right\} + o\left(b^2\right), \end{aligned} \tag{4.2}$$

and

$$Var\left(\widehat{F}_n(x)\right) = \frac{F\left(x\right)\left(1-F\left(x\right)\right)}{n} + \frac{b}{n}f\left(0\right)\left\{\int_{-1}^{s} K^2(t)dt - s\right\} + o\left(\frac{b}{n}\right). \tag{4.3}$$

To remove those boundary effects in kernel distribution estimator, a variety of methods have been developed in the literature. We briefly mention reflection of data (see, e.g., Silverman (1986) [55]), transform of data (see, Marron and Ruppert (1994) [42]), pseudo-data method (see Cowling and Hall (1996) [15]) and also the boundary kernel method (Gasser et al. (1985) [26], Zhang and Karunamuni (2000) [69]). For more details about this techniques one refers to Karunamuni and Alberts (2005) [34]; Karunamuni and Alberts (2004) [33].

In this paper, we develop a new kernel type estimator of the heavy-tailed distributions functions that improved boundary effects near the points at left boundary region, for $x \in [0, b]$. This estimator is based on a new transformation on boundary corrected kernel estimator ideas of Koláček and Karunamuni (2009) [38], Buch-Larsen et al. (2005) [7], developed for boundary correction in kernel density estimation. The basic technique of construction of the proposed estimator is kind of a generalized reflection method involving reflecting a transformation of the observed data, we used two transformations. First, a transformation g is selected from a parametric family, second we propose to use a transformation T based on the little-known Champernowne distribution function, which produces good results in all situations studied and it is straightforward to apply.

Theoretical properties of boundary kernel distribution estimator are introduced in Section 4.2. In Section 4.3 the proposed estimator is given and its bias and variance are computed. In Section 4.4, simulation studies are done to see the performance of the proposed estimator, and compare it with the "usual" and "boundary" distribution function estimators. Finally, all Proofs are referred to Section 4.5.

4.2 Boundary kernel distribution estimator

In order to deal with the boundary effects that occur in nonparametric regression and density function estimation, the use of boundary kernels is proposed and studied by authors such as Gasser and Müller (1979) [25], Karunamuni and Alberts (2004) [33]. Next we extend this approach to a distribution function estimator framework. The structure of this estimator is the same type of that in density estimation case which has been discussed in Karunamuni and Alberts (2007) [35], for more details see Zhang and Karunamuni (1999) [68]. This method of estimating combines the transformation and the reflection methods, consisting of three steps:

Step 1. Transform the initial data $X_1, ..., X_n$ to $g(X_1), ..., g(X_n)$, where g is a nonnegative, continuous, and monotonically increasing function from $[0, \infty)$ to $[0, \infty)$.

Step 2. Reflect $g(X_1), ..., g(X_n)$ around the origin, so we get $-g(X_1), ..., -g(X_n)$.

Step 3. The estimator of F is based on the enlarged data sample $-g(X_1), ..., -g(X_n), g(X_1), ..., g(X_n)$. Then the boundary kernel distribution estimator of the distribution function for $x \in [0, b]$, is given by

$$\overline{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left\{K\left(\frac{x-g(X_i)}{b}\right) - K\left(-\frac{x+g(X_i)}{b}\right)\right\}, \tag{4.4}$$

where K is a distribution of the kernel function k as in (4.1).

This estimator generates a class of boundary corrected estimators. We need to obtain explicit forms of the bias, variance and asymptotic mean square error expressions of the estimator (4.4).

Lemma 4.2.1 *Assume that $f'(.)$ and $g''(.)$ exist and are continuous. Further, assume that $g^{-1}(0) = 1$ and $g'(0) = 0$, where g^{-1} the inverse function of g and f' and g'' are the first and second derivatives of f and g respectively. Then for $x = sb$, $0 \leq s \leq 1$, we have*

$$\begin{aligned} Bias\left(\overline{F}_n(x)\right) = b^2 \Bigg\{ & f'(0)\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right) \\ & - f(0)\, g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\Bigg\} + o\left(b^2\right), \end{aligned} \tag{4.5}$$

and

$$Var\left(\overline{F}_n(x)\right) = \frac{F(x)(1-F(x))}{n} + \frac{b}{n} f(0) \left\{ 2\int_{-1}^{-s} K^2(t)dt - s \right.$$
$$\left. + \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt \right\} + o\left(\frac{b}{n}\right). \tag{4.6}$$

Accordingly, the asymptotic mean squared error is

$$AMSE\left(\overline{F}_n(x)\right) = b^4 \left\{ f'(0) \left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt \right) \right.$$
$$\left. - f(0)\, g''(0) \left(\int_{-1}^{s} (s-t)K(t)dt + \int_{-1}^{-s} (s+t)K(t)dt \right) \right\}^2$$
$$+ \frac{F(x)(1-F(x))}{n} + \frac{b}{n} f(0) \left\{ 2\int_{-1}^{-s} K^2(t)dt - s \right.$$
$$\left. + \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt \right\}. \tag{4.7}$$

Remark 4.2.1 *Functions satisfying conditions* $g^{-1}(0) = 1$ *and* $g'(0) = 0$ *are easy to construct. The trivial choice is* $g(y) = y$*, which represents the "classical" reflection method estimator. The following transformation adapts well to various shapes of distributions:*

$$g(y) = y + \frac{1}{2} I_s y^2,$$

for $y \geq 0$ *and* $0 \leq s \leq 1$*, where* $I_s = \int_{-1}^{-s} K(t)\, dt$*.*

Remark 4.2.2 *Some discussion on the above choice of g and other various improvements that can be made would be appropriate here. It is possible to construct functions g that improve the bias under some additional conditions. For instance, if one examines the right hand side of bias expansion, then it is not difficult to see that the coefficient of* b^2 *can be made equal to zero if g is appropriately chosen,* (*see Koláček and Karunamuni* (2009) [*38*])*.*

Remark 4.2.3 *It is easy to see that for* $x > b$*, the estimator* (4.4) *reduces to* (4.1)*, which* *is the usual kernel distribution estimator. So* (4.4) *is a natural boundary continuation of the usual estimator.*

4.3 The proposed estimator

We now have all the necessary tools to introduce our estimator of heavy tailed cdf F, based on ideas of Koláček and Karunamuni (2009) [38], Buch Larsen et al. (2005) [7] and we insert a new transformation. We shall assume that the unknown cdf F has support $[0, \infty)$. The transformation idea is based on transforming the original data by a new parametric transformation T, chosen by modified Champernowne distribution function. The modified Champernowne distribution is defined on $x \geq 0$, and formulated as

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

with parameter $\alpha > 0$, $M > 0$ and $c \geq 0$, and its density is

$$t(x) = \frac{\alpha\,(x+c)^{\alpha-1}\left((M+c)^{\alpha} - c^{\alpha}\right)}{\left((x+c)^{\alpha} + (M+c) - 2c^{\alpha}\right)^{2}}, \ x \geq 0.$$

The modified Champernowne distribution converges to a Pareto distribution in the tail:

$$t_{\alpha,M,c}(x) \to \frac{\alpha\left((M+c)^{\alpha} - c^{\alpha}\right)}{x^{\alpha+1}} \text{ as } x \longrightarrow \infty.$$

For more details about modified Champernowne distribution see for instance Buch Larsen et al. (2005) [7], Champernowne (1952)[10].

The following steps describes the techniques using for obtain the proposed estimator of F :

Step 1. Estimate the parameters $(\widehat{\alpha}, \widehat{M}, \widehat{c})$ of the modified Champernowne distribution to obtain the transformation function. In the modified Champernowne distribution, we notice that $T(x) = 0.5$. This suggests that M can be estimated as the empirical median of the data set. Then to estimate the pair (α, c) which maximizes the log likelihood function :

$$l(\alpha, c) = n\log(\alpha) + n\log((M+c)^\alpha - c^\alpha) + (\alpha - 1)\sum_{i=1}^{n}\log(X_i + c)$$
$$-2\sum_{i=1}^{n}\log\left((X_i + c) + (M+c)^\alpha - 2c^\alpha\right).$$

Step 2. Transform the initial data $X_1, ..., X_n$, with the transformation function,

$$Y_i = T(X_i), i = 1, ..., n,$$

are new rv's, Y_i is in the interval $(0, 1)$ and uniform distributed.

Step 3. Calculate the boundary kernel distribution estimator of the transformed data, $Y_1, ..., Y_n$:

$$\widetilde{H}_n(y) = \frac{1}{n}\sum_{i=1}^{n}\left\{K\left(\frac{y - g(Y_i)}{b}\right) - K\left(-\frac{y + g(Y_i)}{b}\right)\right\}, \tag{4.8}$$

where g is the same transformation as in (4.4).

Step 4. The final form of our estimator of the original data set, $X_1, ..., X_n$ is defined as, for $x = sb$, $0 \leq s \leq 1$,

$$\widetilde{F}_n(x) = \widetilde{H}_n(T(x)). \tag{4.9}$$

Thus $\widetilde{F}_n(x)$ is a natural boundary continuation of the usual kernel distribution estimator (4.1). An important adjustment in the estimator (4.9) is that it is based on a new transformation T. Furthermore, it is important to remark here that the transform kernel distribution estimator (4.9) is nonnegative (provided K is nonnegative).

The next theorem establishes the bias and variance of the proposed estimator (4.9).

Theorem 4.3.1 *Assume that F is a heavy-tailed distribution function. Under the same conditions on the transformation function g. Then for $x = sb$, $0 \le s \le 1$ the bias and variance of $\widetilde{F}_n(x)$ are respectively*

$$\begin{aligned} Bias\left(\widetilde{F}_n(x)\right) = b^2 \Bigg\{ &\left(\frac{f}{T'}\right)'(0)\frac{1}{T'(0)}\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right) \\ &- \frac{f(0)}{T'(0)} g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\Bigg\} + o\left(b^2\right), \end{aligned} \quad (4.10)$$

and

$$\begin{aligned} Var\left(\widetilde{F}_n(x)\right) = &\frac{F(x)(1-F(x))}{n} + \frac{b}{n}\frac{f(0)}{T'(0)}\Bigg\{2\int_{-1}^{-s} K^2(t)dt - s \\ &+ \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt\Bigg\} + o\left(\frac{b}{n}\right). \end{aligned} \quad (4.11)$$

The asymptotic mean squared error is

$$\begin{aligned} AMSE\left(\widetilde{F}_n(x)\right) = b^4 \Bigg\{ &\left(\frac{f}{T'}\right)'(0)\frac{1}{T'(x)}\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right) \\ &- \frac{f(0)}{T'(0)} g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\Bigg\}^2 \\ &+ \frac{F(x)(1-F(x))}{n} + \frac{b}{n}\frac{f(0)}{T'(0)}\Bigg\{2\int_{-1}^{-s} K^2(t)dt - s \\ &+ \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt\Bigg\}. \end{aligned} \quad (4.12)$$

Remark 4.3.1 *By comparing expressions* (4.2), (4.3), (4.10), *and* (4.11) *at boundary points we can see that the bias of $\widetilde{F}_n(x)$ is of order $o\left(b^2\right)$, while the variance of $\widetilde{F}_n(x)$ is of the same order of $\widehat{F}_n(x)$. So the proposed estimator improved boundary effects in kernel distribution estimator since the bias at boundary points is of the same order as the bias at the interior points.*

4.4 Simulation Studies

To compare the performance of our proposed estimator $\widetilde{F}_n$ against the boundary kernel estimator $\overline{F}_n$ and the usual $\widehat{F}_n$ estimator described by Nadaraya (1964) [45], we made some simulation studies. We simulate data from three different heavy tailed distributions : Pareto type I, Pareto type II and Pareto type III. The distributions and the chosen parameters are listed in table 1.

Distribution	$F(x)$ for $x \geq 0$	Parameters
Pareto type I	$\dfrac{1}{1+(x/\lambda)^{-\beta}}$	$(\beta,\, \lambda) = (1, 1)$
Pareto type II	$1 - \left(1 + \dfrac{x}{\rho}\right)^{-\alpha}$	$(\rho = 1, \alpha = 2)$
Pareto type III	$1 - \left(1 + \left(\dfrac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right)^{-1}$	$(\mu, \sigma, \gamma) = (0, 0.7, 1)$

Table 4.1: Distributions used in the simulation studies.

We measure the performance of the estimators by the error measures $AMSE$ and $AMISE$. The simulation is based on 1000 replications. In each replication the sample sizes: $n = 50$, $n = 200$ and $n = 400$ was used. For the kernel, we choosing the Epanechnikov kernel $k(t) = 3/4(1-t^2)I(|t| \leq 1)$, where $I(.)$ denotes the indicator function, has been observed in Silverman (1986) [55], that this kernel possesses the maximum efficiency, in the sense that it produces the minimal $AMISE$. The choice of bandwidth is very important for the good performance of any kernel estimator. In all cases, we select the asymptotic optimal global bandwidth of the estimator $\overline{F}_n$ by minimizing the $AMISE$, because this is much more likely to be used in application and gave reliably good results. We have

$$b_{opt} = \left(\frac{2f\,(0)\,A(s)}{5\left[f'(0)B(s) - f\,(0)\,g''\,(0)\,C(s)\right]^2}\right)^{1/3} n^{-1/3},$$

where

$$A(s) := \left(2\int_{-1}^{s} K(t)\,K(t-2s)\,dt + s - 2\int_{-1}^{-s} K^2(t)\,dt - \int_{-s}^{s} K^2(t)\,dt\right),\ 0 \le s \le 1,$$

$$B(s) := \left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right),\ 0 \le s \le 1,$$

and

$$C(s) := \left(\int_{-1}^{s} (s-t)K(t)dt + \int_{-1}^{-s} (s+t)K(t)dt\right),\quad 0 \le s \le 1.$$

The comparison is based on data simulated from the four distributions described in table 4.1. Firstly, for each value of $s \in \{0.35, 0.45, 0.55\}$ we have calculated the absolute bias, variance and the $AMSE$ values of the three estimators and have displayed the results in a tables 4.2, 4.3 and 4.4. Secondly, for different values of s we calculated the $AMISE$ values for each estimator over the whole boundary region $[0, b]$. The values of $AMISE$ are tabulated in table 4.5. The comparison show that the values of the $AMSE$ and the $AMISE$ were smallest in case of the proposed estimator, this is due to the fact that the proposed estimator is locally adaptive..

Discussion : For **Pareto type I** distribution, close examination of tables of $AMSE$ clear by shows that, we have the proposed estimator $\widetilde{F}_n$ and boundary kernel distribution estimator $\overline{F}_n$ show the best performance, but the estimator $\widetilde{F}_n$ out performs the estimator $\overline{F}_n$ for all n. Also, in terms of $AMISE$ for each sample size, the $AMISE$ of the estimator $\widetilde{F}_n$ is smaller than that of $\overline{F}_n$. the performance of usual kernel distribution estimator $\widehat{F}_n$ is worse than the performance of the estimator $\widetilde{F}_n$.

For the **Pareto type II** distribution, much the best, in terms of both $AMSE$ and $AMISE$, is the proposed estimator $\widetilde{F}_n$. Next much the worst, although with performance, is the usual kernel distribution estimator $\widehat{F}_n$.

For **Pareto type III** distribution, the estimator $\widehat{F}_n$ also is overall clearly the worst. The proposed estimator and boundary kernel distribution estimator have rather different performances in this case. Clearly best in terms of $AMSE$ and $AMISE$ terms is the estimator $\widetilde{F}_n$.

In conclution: the main resultant of our simulation studies is that the proposed estimator is recommended for it improved boundary effet for heavy tailed distributions. We see that overall $\widetilde{F}_n$ is the best choice amony the three estimators considered. Indeed, the performance of boundary kernel distribution estimator $\overline{F}_n$ is very disappointing, and this estimator can not be recommended for use. The usual kernel distribution estimator $\widehat{F}_n$ is clearly the worst estimator for the three heavy tailed distribution considered. This is clearly due to the boundary effect.

4.5 Proofs

Proof of (4.2)**.** For $x = sb$, $0 \leq s \leq 1$, using the property $K(t) = 1 - K(-t)$, $-s \leq t \leq s$, and a Taylor expansion of order 1. First note that

$$Bias\left(\widehat{F}_n(x)\right) = E\widehat{F}_n(x) - F(x),$$

then,

$$\begin{aligned} E\widehat{F}_n(x) &= EK\left(\frac{x - X_i}{b}\right) \\ &= \int_0^{\infty} K\left(\frac{x-z}{b}\right) f(z)dz. \end{aligned}$$

To calculate the mean of $\widehat{F}_n$, we used the change of variable $t = (x - z)/b$, we have

$$\begin{aligned} E\widehat{F}_n(x) &= b\int_{-1}^{s} K\left(t\right) f((s-t)b)dt \\ &= b\int_{-1}^{-s} K\left(t\right) f((s-t)b)dt + b\int_{-s}^{s} \left(1 - K\left(-t\right)\right) f((s-t)b)dt \\ &= b\int_{-1}^{-s} K\left(t\right) f((s-t)b)dt + F(2sb) - b\int_{-s}^{s} K\left(t\right) f((s+t)b)dt. \end{aligned}$$

Using a Taylor expansion of order 2 on the function $F(.)$ we have

$$F(2sb) = F(0) + f(0)2sb + f'(0)2s^2b^2 + o\left(b^2\right).$$

By the existence and continuity of $f'(.)$ near 0, we obtain for $x = sb$

$$\begin{aligned} F(0) &= F(x) - f(x)sb + \frac{1}{2}f'(x)\, s^2b^2 + o(b^2) \\ f(x) &= f(0) + f'(0)\, sb + o(b) \\ f'(x) &= f'(0) + o(1). \end{aligned}$$

Therefore,

$$F(2sb) = F(x) + f(0)sb + \frac{3}{2}f'(0)\, s^2b^2 + o\left(b^2\right).$$

We obtain

$$\begin{aligned} & Bias\left(\widehat{F}_n(x)\right) \\ &= b\int_{-1}^{-s} K(t)\left\{f(0) + f'(0)(s-t)b + o(b)\right\} dt + f(0)sb + \frac{3}{2}f'(0)\, s^2b^2 + o(b^2) \\ &-b\int_{-s}^{s} K(t)\left\{f(0) + f'(0)(s+t)b + o(b)\right\} dt \\ &= b\left\{f(0)s + f(0)\int_{-1}^{-s} K(t)\, dt - f(0)\int_{-s}^{s} K(t)\, dt\right\} + b^2\left\{\frac{3}{2}f'(0)\, s^2\right. \\ &\left. + f'(0)\int_{-1}^{-s}(s-t)K(t)\, dt - f'(0)\int_{-s}^{s}(s+t)K(t)\, dt\right\} + o(b^2). \end{aligned}$$

From the symmetry of k and the definition $K(x)$, one can write $K(x) = 1/2 + r(x)$, where $r(x) = -r(-x)$ for all x such that $|x| \leq 1$. Thus $\int_{-s}^{s} K(t)dt = s$ and after some algebra we obtain the bias expression as

$$\begin{aligned} Bias\left(\widehat{F}_n(x)\right) &= bf(0)\int_{-1}^{-s} K(t)\, dt \\ &+ b^2f'(0)\left\{\frac{s^2}{2} + s\int_{-1}^{-s} K(t)dt - \int_{-1}^{s} tK(t)dt\right\} + o\left(b^2\right). \end{aligned}$$

This completes the proof of expression (4.2). ■

Proof of (4.3)**.** Observe that for $x = sb$, $0 \leq s \leq 1$, we have

$$\begin{aligned} Var\left(\widehat{F}_n(x)\right) &= \frac{1}{n^2} Var\left\{\sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right)\right\} \\ &= \frac{1}{n} E\left\{K\left(\frac{x - X_i}{b}\right)\right\}^2 - \frac{1}{n}\left\{E\left\{K\left(\frac{x - X_i}{b}\right)\right\}\right\}^2 \\ &= I_1 - I_2, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \frac{1}{n} E\left\{K\left(\frac{x - X_i}{b}\right)\right\}^2 \\ &= \frac{1}{n}\int_0^{\infty} K^2\left(\frac{x - z}{b}\right) f(z)dz \\ &= \frac{b}{n}\int_{-1}^{s} K^2\left(t\right) f((s-t)b)dt \\ &= \frac{b}{n}\int_{-1}^{-s} K^2\left(t\right) f((s-t)b)dt + \frac{b}{n}\int_{-s}^{s} K^2\left(t\right) f((s-t)b)dt. \\ &= I_{11} + I_{12}. \end{aligned}$$

It can be shown that

$$\begin{aligned} I_{11} &= \frac{b}{n}\int_{-1}^{-s} K^2\left(t\right) f((s-t)b)dt \\ &= \frac{b}{n}\int_{-1}^{-s} K^2\left(t\right)\left\{f(0) + o(1)\right\} dt. \end{aligned}$$

We use the property $K(t) = 1 - K(-t)$ and similarly as in the last proof we obtain I_{12}

$$
\begin{aligned}
I_{12} &= \frac{b}{n}\int_{-s}^{s} K^2(t)\, f((s-t)b)dt \\
&= \frac{b}{n}\int_{-s}^{s} \left(1 - 2K(-t) + K^2(-t)\right) f((s-t)b)dt \\
&= \frac{b}{n}\int_{-s}^{s} f((s-t)b)dt - 2\frac{b}{n}\int_{-s}^{s} K(t)\, f((s+t)b)dt + \frac{b}{n}\int_{-s}^{s} K^2(t)\, f((s+t)b)dt \\
&= \frac{F(2sb)}{n} - 2\frac{b}{n}\int_{-s}^{s} K(t)\{f(0)+o(1)\}\, dt + \frac{b}{n}\int_{-s}^{s} K^2(t)\{f(0)+o(1)\}\, dt \\
&= \frac{F(x)}{n} - f(0)s\frac{b}{n} + \frac{b}{n}f(0)\int_{-s}^{s} K^2(t)\, dt + o(\frac{b}{n}),
\end{aligned}
$$

and now combine I_{11} and I_{12} to obtain the express I_1 as

$$
\begin{aligned}
I_1 &= I_{11} + I_{12} \\
&= \frac{b}{n}\int_{-1}^{-s} K^2(t)\{f(0)+o(1)\}\, dt + \frac{F(x)}{n} - f(0)s\frac{b}{n} + \frac{b}{n}f(0)\int_{-s}^{s} K^2(t)\, dt + o(\frac{b}{n}) \\
&= \frac{F(x)}{n} + \frac{b}{n}f(0)\left\{\int_{-1}^{s} K^2(t)\, dt - s\right\} + o(\frac{b}{n}).
\end{aligned}
$$

With the expression obtained for the bias we get the expression for I_2 as

$$
\begin{aligned}
I_2 &= \frac{1}{n}\left\{E\left\{K\left(\frac{x - X_i}{b}\right)\right\}\right\}^2 \\
&= \frac{1}{n}\left\{E\widehat{F}_n(x)\right\}^2 \\
&= \frac{1}{n}F^2(x) + o(\frac{b}{n}).
\end{aligned}
$$

Finally, we obtain the variance of the estimator $\widehat{F}_n(x)$ as

$$
\begin{aligned}
Var\left(\overline{F}_n(x)\right) &= I_1 - I_2 \\
&= \frac{F(x)}{n} + \frac{b}{n}f(0)\left\{\int_{-1}^{s} K^2(t)\, dt - s\right\} - \frac{1}{n}F^2(x) + o(\frac{b}{n}) \\
&= \frac{F(x)\,(1 - F(x))}{n} + \frac{b}{n}f(0)\left\{\int_{-1}^{s} K^2(t)\, dt - s\right\} + o(\frac{b}{n}).
\end{aligned}
$$

This completes the proof of expression (4.3). ■

proof of Lemma 4.2.1. The proof is the same as On boundary correction in kernel estimation of ROC curves, (see Koláček and Karunamuni (2009) [38]). It suffices to replace the tansformations g_1 and g_2 by g. Then for $x = sb$, $0 \leq s \leq 1$, we have

$$\begin{aligned} Bias\left(\overline{F}_n(x)\right) = b^2 \Bigg\{ & f'(0)\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right) \\ & - f(0)\, g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\Bigg\} + o\left(b^2\right), \end{aligned}$$

and

$$\begin{aligned} Var\left(\overline{F}_n(x)\right) = & \frac{F(x)(1-F(x))}{n} + \frac{b}{n} f(0)\left\{2\int_{-1}^{-s} K^2(t)dt - s \right. \\ & \left. + \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt\right\} + o\left(\frac{b}{n}\right). \end{aligned}$$

Hence, the MSE of $\overline{F}_n(x)$ is

$$MSE\left(\overline{F}_n(x)\right) = Bias^2\left(\overline{F}_n(x)\right) + Var\left(\overline{F}_n(x)\right).$$

The asymptotic MSE of $\overline{F}_n(x)$ is

$$\begin{aligned} AMSE\left(\overline{F}_n(x)\right) = b^4 \Bigg\{ & f'(0)\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right) \\ & - f(0)\, g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\Bigg\}^2 \\ & + \frac{F(x)(1-F(x))}{n} + \frac{b}{n} f(0)\left\{2\int_{-1}^{-s} K^2(t)dt - s \right. \\ & \left. + \int_{-s}^{s} K^2(t)dt - 2\int_{-1}^{s} K(t)K(t-2s)dt\right\}. \end{aligned}$$

Here completes the proof of Lemma 4.2.1. ■

Proof of Theorem 4.3.1. We have $X_1, ..., X_n$ are independent identically distributed variables with density f and cdf F. the Transform kernel distribution estimator of $F(x)$

is

$$;\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left\{K\left(\frac{T(x)-g(T(X_i))}{b}\right) - K\left(-\frac{T(x)+g(T(X_i))}{b}\right)\right\},$$

where $T(\cdot)$ is the transformation function. Let the transformed variable $Y_i = T(X_i)$, have distribution H:

$$H(y) = F\left(T^{-1}(T(x))\right) = F(x),$$

and the density of $H(y)$ as

$$h(y) = \frac{f\left(T^{-1}(y)\right)}{T'\left(T^{-1}(y)\right)},$$

so the boundary kernel distribution estimator of $H(y)$ is

$$\widetilde{H}_n(y) = \frac{1}{n}\sum_{i=1}^{n}\left\{K\left(\frac{y-g(Y_i)}{b}\right) - K\left(-\frac{y+g(Y_i)}{b}\right)\right\}.$$

The transform kernel distribution estimator can be expressed by :

$$\widetilde{F}_n(x) = \widetilde{F}_n\left(T^{-1}(T(x))\right) = \widetilde{H}_n(y),$$

implying The Bias of the transform kernel distribution estimator is

$$\begin{aligned}
Bias\left(\widetilde{F}_n(x)\right) &= Bias\left(\widetilde{F}_n\left(T^{-1}(T(x))\right)\right) \\
&= Bias\left(\widetilde{H}_n(T(x))\right) \\
&= b^2\left\{h'(T(0))\left(\frac{s^2}{2} + 2s\int_{-1}^{-s} K(t)dt - \int_{-s}^{s} tK(t)dt\right)\right. \\
&\left. - h(T(0))g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt + \int_{-1}^{-s}(s+t)K(t)dt\right)\right\} + o\left(b^2\right),
\end{aligned}$$

note that

$$h(T(x)) = \frac{f(x)}{T'(x)},\ h'(T(x)) = \left(\frac{f(x)}{T'(x)}\right)'\frac{1}{T'(x)}.$$

then

$$h\left(T\left(0\right)\right)=\frac{f\left(0\right)}{T'\left(0\right)},\ h'(T\left(0\right))=\left(\frac{f}{T'}\right)'(0)\frac{1}{T'\left(0\right)},$$

which are used to find the mean of the transform kernel distribution estimator

$$\begin{aligned}Bias\left(\widetilde{F}_n\left(x\right)\right)&=b^2\left\{\left(\frac{f}{T'}\right)'(0)\frac{1}{T'\left(0\right)}\left(\frac{s^2}{2}+2s\int_{-1}^{-s}K(t)dt-\int_{-s}^{s}tK(t)dt\right)\right.\\&\left.-\frac{f\left(0\right)}{T'\left(0\right)}g''(0)\left(\int_{-1}^{s}(s-t)K(t)dt+\int_{-1}^{-s}(s+t)K(t)dt\right)\right\}+o\left(b^2\right).\end{aligned}$$

By the same idea we calculated the variance

$$\begin{aligned}Var\left(\widetilde{F}_n\left(x\right)\right)&=Var\left(\widetilde{F}_n\left(T^{-1}\left(T\left(x\right)\right)\right)\right)\\&=Var\left(\widetilde{H}_n\left(T(x)\right)\right)\\&=\frac{H\left(y\right)\left(1-H\left(y\right)\right)}{n}+\frac{b}{n}h(T\left(0\right))\left\{2\int_{-1}^{-s}K^2(t)dt-s\right.\\&\left.+\int_{-s}^{s}K^2(t)dt-2\int_{-1}^{s}K(t)K(t-2s)dt\right\}+o\left(\frac{b}{n}\right).\\&=\frac{F\left(x\right)\left(1-F\left(x\right)\right)}{n}+\frac{b}{n}\frac{f\left(0\right)}{T'\left(0\right)}\left\{2\int_{-1}^{-s}K^2(t)dt-s\right.\\&\left.+\int_{-s}^{s}K^2(t)dt-2\int_{-1}^{s}K(t)K(t-2s)dt\right\}+o\left(\frac{b}{n}\right).\end{aligned}$$

This completes the proof of Theorem 4.3.1. ■

			Pareto type I			**Pareto type II**			**Pareto type III**		
	Estimator	s	.35	.45	.55	.35	.45	.55	.35	.45	.55
	$\widetilde{F}_n$		6.3049	6.3733	6.6385	14.830	16.289	17.605	4.1028	4.3000	4.5112
$\lvert Bias\rvert$	$\overline{F}_n$		26.904	28.668	31.311	28.840	31.802	34.524	26.783	28.780	31.482
	$\widehat{F}_n$		38.987	48.741	54.245	38.659	45.524	53.794	39.962	48.336	48.841
	$\widetilde{F}_n$		0.1042	0.2428	0.1299	0.2403	0.2759	0.1587	0.1183	0.3229	0.2383
Var	$\overline{F}_n$		0.3714	0.5461	0.4918	0.4391	0.5249	0.4619	0.3802	0.5384	0.4985
	$\widehat{F}_n$		0.9980	1.1071	1.3112	1.0359	1.3299	1.5280	0.9695	1.1331	1.3767
	$\widetilde{F}_n$		0.1439	0.2834	0.1740	0.4602	0.5412	0.4686	0.1352	0.3414	0.2587
$AMSE$	$\overline{F}_n$		1.0952	1.3679	1.4722	1.2709	1.5363	1.6539	1.0976	1.3667	1.4896
	$\widehat{F}_n$		2.5180	3.4828	4.2537	2.5305	3.4023	4.4219	2.5665	3.4695	3.7622

Table 4.2: Bias, Var and AMSE Values Over the Boundary Region for sample size n=50. Results are re-scaled by the factor 0.001.

			Pareto type I			**Pareto type II**			**Pareto type III**		
	Estimator	s	.35	.45	.55	.35	.45	.55	.35	.45	.55
	$\widetilde{F}_n$		2.0016	2.0856	2.1298	5.0417	5.3146	5.8175	0.8145	0.8062	0.8240
$\lvert Bias\rvert$	$\overline{F}_n$		10.489	11.443	12.596	11.811	12.531	13.786	10.623	11.250	12.468
	$\widehat{F}_n$		13.168	18.879	24.637	13.276	19.884	32.944	15.899	21.445	23.283
	$\widetilde{F}_n$		0.0611	0.0929	0.1029	0.0533	0.0879	0.0935	0.0593	0.1022	0.1247
Var	$\overline{F}_n$		0.0931	0.1309	0.1490	0.1025	0.1452	0.1647	0.0961	0.1344	0.1489
	$\widehat{F}_n$		0.1719	0.2346	0.2498	0.2329	0.2664	0.2778	0.1932	0.2092	0.2640
	$\widetilde{F}_n$		0.0651	0.0972	0.1075	0.0787	0.1161	0.1273	0.0600	0.1028	0.1253
$AMSE$	$\overline{F}_n$		0.2031	0.2618	0.3077	0.2420	0.3022	0.3547	0.2089	0.2609	0.3043
	$\widehat{F}_n$		0.3453	0.5910	0.8568	0.4091	0.6617	1.3631	0.4459	0.6691	0.8061

Table 4.3: Bias, Var and AMSE Values Over the Boundary Region for sample size n=200. Results are re-scaled by the factor 0.001.

			Pareto type I			Pareto type II			Pareto type III		
	Estimator	s	.35	.45	.55	.35	.45	.55	.35	.45	.55
	$\widetilde{F}_n$		0.9937	1.0213	1.0449	2.5560	2.7109	2.9718	0.7311	0.7523	0.7888
$\lvert Bias \rvert$	$\overline{F}_n$		6.5212	7.2401	7.9143	7.2751	7.7788	8.5954	6.6332	7.1435	7.9339
	$\widehat{F}_n$		8.7913	10.2731	16.296	11.057	17.428	20.437	7.4034	12.070	14.831
	$\widetilde{F}_n$		0.0336	0.0450	0.0629	0.0255	0.0401	0.0407	0.0281	0.0422	0.0357
Var	$\overline{F}_n$		0.0424	0.0559	0.0759	0.0512	0.0705	0.0790	0.0448	0.0630	0.0644
	$\widehat{F}_n$		0.0719	0.1025	0.1113	0.0926	0.1066	0.1420	0.0768	0.0956	0.1193
	$\widetilde{F}_n$		0.0345	0.0461	0.0640	0.0320	0.0474	0.0496	0.0286	0.0428	0.0363
$AMSE$	$\overline{F}_n$		0.0849	0.1084	0.1386	0.1042	0.1310	0.1529	0.0888	0.1141	0.1274
	$\widehat{F}_n$		0.1492	0.2081	0.3769	0.2149	0.4103	0.5597	0.1316	0.2413	0.3393

Table 4.4: Bais, Var and AMSE Values Over the Boundary Region for sample size n=400. Results are re-scaled by the factor 0.001.

			Pareto type I			Pareto type II			Pareto type III		
	Estimator	s	.35	.45	.55	.35	.45	.55	.35	.45	.55
	$\widetilde{F}_n$		0.2015	0.1062	0.1014	0.2353	0.1148	0.0317	0.3175	0.2504	0.1805
$n = 50$	$\overline{F}_n$		0.5882	0.5437	0.4142	0.4024	0.3262	0.2826	0.3998	0.3456	0.2947
	$\widehat{F}_n$		1.3379	1.2503	1.3200	0.7429	0.7232	0.6875	0.8923	0.7979	0.8473
	$\widetilde{F}_n$		0.0405	0.0241	0.0101	0.0262	0.0178	0.0065	0.0450	0.0396	0.0337
$n = 200$	$\overline{F}_n$		0.0748	0.0657	0.0619	0.0523	0.0468	0.0432	0.0495	0.0444	0.0398
	$\widehat{F}_n$		0.2028	0.1927	0.1799	0.1281	0.1045	0.0995	0.1542	0.1249	0.1216
	$\widetilde{F}_n$		0.0160	0.0103	0.0072	0.0088	0.0036	0.0078	0.0157	0.0133	0.0108
$n = 400$	$\overline{F}_n$		0.0258	0.0225	0.0229	0.0195	0.0157	0.0135	0.0177	0.0156	0.0133
	$\widehat{F}_n$		0.0806	0.0729	0.0639	0.0512	0.0374	0.0383	0.0509	0.0469	0.0453

Table 4.5: AIMSE Values Over the Boundary Region. Results are re-scaled by the factor 0.001.

Conclusion

Estimation in the boundary points suffer a large bias, however a special treatment is needed. For heavy-tailed distributions, It is well known now that kernel distribution estimators are not consistent when estimating a distribution near the finite end points of the support. This is due to boundary effects that occur in nonparametric curve estimation problems. A number of proposals have been made in the kernel density estimation context with some success. In this thesis, we have introduced a new kernel type estimator of the heavy tailed distributions functions by using a new approach based on the modified Champernowne distribution function.

On the other hand, the present approach can be viewed as an generalized reflection method involving reflecting a transformation of the observed data with a modification of the Champernowne distribution function. The proposed estimator possesses a number of desirable properties, including the non-negativity of the estimator. Each estimator has certain advantages and works well at certain times. The proposed method seems to have inherited the best of both transformation and reflection methods and that improved boundary effects near the points at left boundary region.

Bibliography

[1] Applebaum, D., 2005. Levy-type stochastic integrals with regurlary varying tails. Stochastic Analysis and Applications.

[2] Asmussen, S., 1997. Ruin Probabilities. World Scientific. Singapore. To appear.

[3] Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika, vol. 68, pp. $326-328$.

[4] Baszczyńska, A., 2016. Kernel estimation of cumulative distribution function of a random vatriable with bounded support. Statistics in transition, vol. 17, pp.$541-556$.

[5] Bingham, N., Goldie, C. and Teugels, J., 1987. Regular Variation. Cambridge University Press.

[6] Bolancé, C., Guillén, M. and Nilsen, J. P., 2003. Kernel density estimator of actuarial loss functions. Insurance. Mathematics and Economics, vol. 32, pp. $19-36$.

[7] Buch-Larsen, T., Guillen, M., Neilson, J.P. and Bolance, C., 2005. Kernel density estimation for heavy-tailed distributions using the Chempernowne transformation, Statistics, vol. 39, pp. $503-518$.

[8] Champernowne, D.G., 1936. The Oxford Meeting. September $25-29$, Econometrica, 5, October 1937.

[9] Champernowne, D.G., 1937. The theory of income distribution. Econometrica, vol. 5, pp. $379-381$.

[10] Champernowne, D.G., 1952. The graduation of income distributions. Econometrica, vol. 20, pp. 591 – 615.

[11] Cheng, M.-Y., 1997. Boundary aware estimators of integrated density derivative products. Journal of the Royal Statistical Society. Series B (Methodological), vol. 59, pp. 191 – 203.

[12] Chistyakov, V.P., 1964. A theorem on sums of independent positive random variables and its applications to branching random processes. Theory Probab. Appl, vol. 9, pp. 640 – 648.

[13] Clemens, A. E, Hurn, A.S., and Lindsay, K.A., 2003. Mö bius-like mappings and their use in kernel density estimation using the Champernowne transformation, J. Amer. Statist. Asso c, vol. 98, pp. 993 – 1000.

[14] Cline, D.B.H and Hart, J.D., 1991. Kernel Estimation of Densities of Discontinuous-Derivatives. Statistics, vol. 22, pp. 69 – 84.

[15] Cowling, A., and Hall, P., 1996. "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation". Journal of the Royal Statistical Society, Ser. B, vol. 58, pp. 551 – 563.

[16] Crovella, M., and Bestavros, A., 1997. Self-similarity in world wide web traffic: Evidence and possible causes. IEEE/ACM Transactions on Networking.

[17] Embrechts, P., Klueppelberg ,C., and Mikosch T., 1997. Modelling Extremal Events for Insurance and Finance. Springer.

[18] Embrechts, P., and Goldie, C.M., 1980. On closure and factorization theorems for subexponential and related distributions, J. Austral. Math. Soc. Ser. A, vol. 29 , pp. 243 – 256.

[19] Embrechts, P. and Goldie, C.M., 1982. On convolution tails. Stoch. Proc. Appl, vol. 13, pp. 263 – 278.

[20] Eubank, R. L., and Speckman, P. L., 1991. A bias reduction theorem withn applications in nonparametric regression. Scandinavian Journal of Statistics, pp. 211 – 222.

[21] Falk, M., 1983. Relative efficiency and deficiency of kernels types estimators of smooth distribution functions. Statist, Neer, vol. 37, pp. 73 – 83.

[22] Fan, J., 1993. Local linear regression smoothers and their minimax efficiencies. The Annals of Statistics.

[23] Fan, J. and Gijbels, I., 1996. Local polynomial modelling and its applications: monographs on statistics andapplied probability 66 (vol. 66). CRC Press.

[24] Foss, S., Korshunov, D., Zachary, S., 2013. An Introduction toHeavy-Tailed and Subexponential Distributions. Springer Series in Operations Research and Financial Engineering.

[25] Gasser, T., Müller, H. G., 1979. Kernel estimation of regression functions. Springer Berlin Heidelberg.

[26] Gasser, T., Müller, H. G. and Mammitzsch, V., 1985. Kernels for Nonparametric Curve Estimation. Journal of the Royal Statistical Society Ser. B, vol. 47, pp. 238 – 256.

[27] Goldie, C.M., 1978. Subexponential distributions and dominated-variation tails. J. Appl. Probab, vol. 15, pp. 440 – 442.

[28] Gustafsson, J.M., Hagmann, J.P. Nielsen, Scallet, O., 2007. Local transformation kernel density estimation of loss distributions, National Centre of Competence in research Financial Valuation and Risk Management.

[29] Hansen, B.E., 2009. Lecture Notes on Nonparametrics. University of Wisconsin Spring.

[30] Hill, P. D., 1985. Kernel estimation of a distribution function. Comm. Statist. Theory Metho ds, vol. 14, pp. 605 – 620.

[31] Jones, M., 1993. Simple boundary correction for kernel density estimation. Statistics and Computing, vol. 3, pp. 135 – 146.

[32] Jones, M. and Foster, P., 1996. A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica, vol. 6, pp. 1005 – 1013.

[33] Karunamuni, R.J. and Alberts, T., 2004. On boundary Correction in Kernel Density Estimation. New York, NY, USA.

[34] Karunamuni, R.J. and Alberts, T., 2005. A Locally Adaptive Transformation Method of Boundary Correction in Kernel Density Estimation. New York, NY, USA.

[35] Karunamuni, R.J. and Alberts, T., 2007. Boundary Correction Methods in Kernel Density Estimation. University of Alberta.

[36] Kheireddine, S., Sayah, A. and Yahia, D., 2015. General metho d of b oundary correction in kernel regression estimation. Afrika Statistika, vol. 10, pp. 739 – 750.

[37] Kim, C., Kim, S. and Park, M., 2006. A bias reducing technique in kernel distribution function estimation. Computational Statistics, vol. 21, pp. 589 – 601.

[38] Koláček, J. and Karunamuni, R.J., 2009. On boundary correction in kernel estimation of ROC curves. Austrian Journal of Statistics, vol. 38, pp. 17 – 32.

[39] Mack, Y. P., 1984. Remarks on some smoothed empirical distribution functions and processes, Bull. Inform. Cybernet, vol. 21, pp. 29 – 35.

[40] Malec, P. Schienle, M., 2012. Nonparametric Kernel Density Estimation Near the Boundary, "Economic Risk", Institute for Statistics and Econometrics, Humboldt-Universitat zu Berlin.

[41] Mandelbrot, B., 1963. The variation of certain speculative prices. The Journal of Business, vol. 36, pp. 394 – 419.

[42] Marron, J.S. and Ruppert, D., 1994. Transformations to Reduce Boundary Bias in Kernel Density Estimation. Journal of the Royal Statistical Society, vol. 56, pp. 653 – 671.

[43] Müller, H. G., 1991. Smooth optimum kernel estimators near endpoints. Biometrika, vol. 78, pp. 521 – 530.

[44] Müller, H. G., 1988. Nonparametric Analysis of Longitudinal Data (Lecture Notes in Statistics 46).

[45] Nadaraya, E.A., 1964. Some new estimates for distribution functions. Theory of Probability and its Application, vol. 15, pp. 497 – 500.

[46] Nadaraya, E. A., 1964. On estimating regression. Theory of Probability and its Applications, vol. 9, pp. 141 – 142.

[47] Parzen, E., 1962. On estimation of a probability density function and mode. Annals of Mathematical Statistics, vol. 33, pp. 1065 – 1076.

[48] Rachev, S., 2003. Handbook of heavy-tailed distributions in finance. Elsevier.

[49] Reiss, R.D., 1981. Nonparametric estimation of smooth distribution functions. Scandinavian Journal of Statistics, vol. 8, pp. 116 – 119.

[50] Rice, J., 1984. Boundary modification for kernel regression. Communications in Statistics-Theory and Methods, vol. 13, pp. 893 – 900.

[51] Rosenblatt, M., 1956. Remarks on Some Nonparametric Estimates of a Density Function. Annals of Mathematical Statistics, vol. 27, pp. 832 – 837.

[52] Sayah, A., Yahia, D. and Necir, A., 2010. Champernowne transformation in kernel quantile estimation for heavy-tailed distributions. Afrika Statistika, vol. 5, pp. 288 – 296.

[53] Schuster, E. F., 1985. Incorporating support constraints into nonparametric estimators of densities. Communications in Statistics-Theory and methods, vol. 14, pp. 1123 – 1136.

[54] Schuster, E. F., 1999. Incorporating support constraints into nonparametric estimators of densities. Communications in Statistics - Theory and Methods, vol. 14, pp. 1123 – 1136.

[55] Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

[56] Swanepoel, J. W. H., 1988. Mean intgrated squared error properties and optimal kernel when estimating a distribution function. Comm. Statist. Theory Metho ds, vol. 17, pp. 3785 – 3799.

[57] Tenreiro, C., 2013. Boundary kernels for distribution function estimation. Statistical Journal, CMUC, vol. 11, pp. 169 – 190.

[58] Uchaikin, V. and V. Zolotarev., 1999. Chance and stability: stable distributions and their applications. VSP International Science Publishers.

[59] Van Grann, F.C., 1983. Nie-parametriese beraming van verdelings funksies. Master's thesis, P.U for C.H.E, Potchefstroom.

[60] Viljoen. H.J., 2007. A comparative study of non-parametric distribution function estimators. Potchefstroom.

[61] Wand, M.P. and Jones, M.C., 1995. Kernel Smoothing. London: Chapman and Hall.

[62] Watson, G.S. and Leadbetter, M.R., 1964. Hazard analysis II, Sankhyā Ser. A, vol. 26, pp. $101 - 116$.

[63] Watson, G. S., 1964. Smooth regression analysis, Sankhyā: The Indian Journal of Statistics. Series A, pp. $359 - 372$.

[64] Winter, B. B., 1973. Strong uniform consistency of integrals of density estimators, Canad. J.Statist, vol. 1, pp. $247 - 253$.

[65] Winter, B.B., 1979. Convergence rate of perturbed empirical distribution functions. Journal of Applied Probability, vol. 16, pp. $163 - 173$.

[66] Yamato, H., 1973. Uniform convergence of an estimator of a distribution function. Bulletin of Mathematical Statistics, vol. 15, pp. $69 - 78$.

[67] Zhang, S., and Karunamuni, R. J., 1998. On kernel density estimation near endpoints. Journal of Statistical Planning and Inference, vol. 70, pp. $301 - 316$.

[68] Zhang, S., Karunamuni, R.J. and Jones, M.C., 1999. An improved estimator of the density function at the boundary. Journal of the American Statistical Association, vol. 94, pp. $1231 - 1241$.

[69] Zhang, S. and Karunamuni, R.J., 2000, On Nonparametric Density Estimation at the Boundary. Nonparametric Statistics, vol. 12, pp. $197 - 221$.

Abbreviations and Notations

We list the notations that will be used in this thesis.

X	random variable
$X_1, ..., X_n$	sample of n observations of X
I_A	indicator function of set A
F	distribution function
f	density function
k	kernel function
K	Distribution of kernel function
b	bandwidth or smoothing parameter
$f',\ f''$	the first and second derivatives of f
$\widehat{f}_n$	standard kernel density estimator
$\widetilde{f}_n$	transformation density estimator
g	transformation function
g^{-1}	the inverse function of g
$o(.)$	$f(x) = o(g(x))$ as $x \to x_0 : \lim f(x)/g(x) = 0$
df	distribution function
P	law of probability

T	the modified Champernowne cumulative distribution function
t	the density of modified Champernowne distribution
iid	independent and identically distributed
$[0,\infty)$	positive interval
$E(X)$	Esperance of X
$Var(X)$	variance of X
F_n	empirical distribution function
$\widehat{F}_n$	usual kernel distribution estimator
$\overline{F}_n$	the generalized reflection and transformation distribution estimator
$\widetilde{F}_n$	transformation kernel distribution estimator
rv	random variable
Φ	standard normal distribution
$\xrightarrow{L}$	convergence in law
$\mathcal{S}$	class of subexponential distribution
$\mathcal{R}_\alpha$	class of regularly varying with index α
$\mathcal{D}$	class of dominated varying distribution functions
$\mathcal{L}$	class of long tailed distributions
EDF	Empirical distribution function
MSE	Mean Squared Error
$AMSE$	Asymptotic Mean Squared Error
$AMISE$	Asymptotic Mean Integrated Squared Error