Mohamed Khider University – Biskra

Mathematics Department

A Thesis Presented for the Degree of

DOCTOR OF SCIENCES

In the Field of **Statistics**

By

Abdallah Sayah

Title

Kernel quantile estimtion for heavy-tailed distributions

Examination Committee Members

Brahim Mezerdi	**Professor**	**Biskra Univ.**	**Chairman**
Abdelhakim Necir	**Professor**	**Biskra Univ.**	**Director of thesis**
Zaher Mohdeb	**Professor**	**Constantine Univ.**	**Examinator**
Hocine Fellag	**Professor**	**Tizi Ouzou Univ.**	**Examinator**
Abderrahmane Yousfate	**Professor**	**Sidi Belabbas Univ.**	**Examinator**
Djamel Meraghni	**MC (A)**	**Biskra Univ.**	**Examinator**

Dédicace

Je dédie cette thèse à ma femme et mes enfants Anas, Selsebil, Rihab, Roufia et Taha que j'éspère qu'il aura son baccalauréat cette année.

Remerciements

Mes sincères remerciements à mon directeur de thèse le Professeur Abdelhakim Necir qui m'a accompagné avec détermination et amabilité. Il m'a constamment soutenu et encouragé à mener à bien ce travail, je le remercie chaleureusement pour son contact constructif si généreux sur le plan humain comme scientifique.

Je suis reconnaissant envers le Professeur Brahim Mezerdi qui m'a fait l'honneur de présider le Jury de ma thése.

Je remercie chaleureusement aussi les membres du Jury d'avoir accepter de juger ma thése : Mr le Professeur Zaher Mohdeb de l'université de Constantine, Mr le Professeur Hocine Fellag de l'université de Tizi-ouzou, Mr le Professeur Abderrahmane Yousfate de l'université de Sidi Belabbas et Mr. Djamel Meraghni Maitre de Conférences (A) de l'université de Biskra.

Mes remerciements à tous les membres du Laboratoire de Mathématiques Appliquées et en premier lieu à Yahia Djabrane de son aide et sa contribution à la réalisation de cette thèse, ainsi à Djamel Meraghni, Fateh Benatia et Brahim Brahimi.

Je ne peux pas passer outre dans ces remerciements tous les membres de département de Mathématiques.

Merci du fond du coeur à toute ma famille, et en particulier ma femme pour son soutien.

Contents

List of Figures

List of Tables

Notations and abbreviations

$\mathcal{S}$	Class of subexponential distribution
$\mathcal{R}_\alpha$	Class of regularly varying at infinity distributions with index α
$\mathcal{D}$	Class of dominated varying distribution functions
$\mathcal{L}$	Class of long tailed distributions
Φ	Normal distribution
$*$	Convolution product
$\underrightarrow{P}$	Convergence in probability
$\underrightarrow{\mathrm{D}}$	Convergence in distribution
i.i.d	Independent and identically distributed
EVT	Extreme value theory
GPD	Generalized Pareto Distributions
D	Domain of attraction
$\bar{X}$	Empirical mean
e.d.f	Empirical distribution function
MSE	Mean squared error
$AMSE$	Asymptotic mean squared error
$MISE$	Mean integrated squared error
$AMISE$	Asymptotic mean integrated squared error
B	Standard Brownian bridge
F_n	Empirical distribution function
Q	Quantile function
F^{-1}	Generalized inverse of F
$[x]$	Integral part of x
a.s	Almost sure
Γ	Gamma function
$o_p(1)$	Convergence in probability to zero
$k.d.f.e$	Kernel distribution function estimator
$TKQE$	Transformed kernel quantile estimator
$CKQE$	Classical kernel quantile estimator

Introduction

The estimation of quantiles of a distribution is of great interest in many applications when the parametric form of the underlying distribution is not available. In addition, extreme quantiles often seem to be the natural thing to estimate in many applications when the underlying distribution is skewed and heavy-tailed, and in particular the extreme quantiles that play an important role in applications to both statistics and probability, namely the benefits of adjustment, and the Value-at-Risk Insurance and financial risk management. In addition, a large class of actuarial measures of risk can be defined as functional quantiles.

Furthermore, estimates of extreme quantiles of the loss distribution in actuarial and financial risk management are fundamental elements of business. From the actuarial point of view, quantile extreme called extreme Value-At-Risk (VaR) which is generally defined as the maximum potential loss that should be attained with a given probability over a given time horizon.

The Value at Risk is the worst expected loss over a horizon given time for a given confidence level. In the majority of situations, the losses are small, and extreme losses occur rarely, so they are rare events. But the number and size of extreme losses can have an important influence on the benefit of the company. The most popular specifications are the lognormal, Weibull and Pareto distributions or a mixture of lognormal and Pareto distributions. The parametric and nonparametric methods work well in traditional areas of the empirical distribution where there are many observations, but they provide a poor adjustment at the extreme tail of the distribution. This is evidently a disadvantage because the extreme risk management calls for the estimation of quantiles and tails of distribution that are generally not directly observable from the data.

Most of existing quantile estimators have problems of bias or inefficiency levels of high probability. To solve this problem, we suggest using the estimation that is called transformed kernel quantile estimation, which is based on the estimation of quantiles of the transformed variable so it can easily to be estimated using a classical approach of the kernel estimation and then taking the inverse transform, this idea was first used in the context of density estimation by Devroye and Gyorfi (1985) for heavy-tailed observations. The idea is to transform the original observations $\{X_1, ..., X_n\}$ in a sample

$$\{Z_1, ..., Z_n\} := \{T(X_1), ..., T(X_n)\}$$

where T is a given function with values in $]0, 1[$.

The subject of the thesis is not about solving the problem of bias for the classical kernel estimator of extreme quantiles for heavy-tailed distributions, but we focus on the reduction of the mean squared error especially when dealing with probabilities close to one, for that we propose a new estimator of the quantile function based on the modified Champernowne transformation. we will concentrate not to estimate the quantiles of X based on the observations $\{X_1, ..., X_n\}$ but to estimate the quantiles of $Z = T(X)$ based on the sample $\{Z_1, ..., Z_n\}$ where $Z_i = T(X_i)$. Then the quantile will be estimated: $Q_{n,X}(p) = T^{-1}(Q_{n,Z}(p))$. This new estimator improves the existing results.

Buch-Larsen et al. (2005) 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 and 1952), and applied to transformed data. For the nonparametric estimation of the quantile function, the smoothing parameter controls the balance between two considerations: bias and variance. Moreover, the mean square error (MSE), which is the sum of squared bias and variance, provides a composite measure of performance of the estimator. Therefore, the optimality in the sense of MSE is not seriously affected by the choice of the kernel but is affected by that of the smoothing parameter (for details, see Wand and Jones, 1995).

The kernel estimator for heavy-tailed distributions has been studied by several authors Bolancé et al. (2003), Clements et al. (2003) and Buch-Larsen et al. (2005) 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.

Buch-Larsen et al. (2005) propose an alternative transformation such as that based on the distribution of Champernowne, where they have shown in simulation studies that this transformation is preferable to the method of transformation in the case of heavy-tailed distributions.

The thesis is organized into four chapters:

The first part of the first chapter is devoted to the presentation of the concept of heavy-tailed distributions and different classes of this type of distributions. 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. A distribution has a heavy tail 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, the supexponential distributions, the regularity varying distribution with index $\alpha > 0$ and the α stable distributions with $0 < \alpha < 2$.

The second part provides an introduction to extreme value theory. Many statistical tools are available in order to draw information concerning specific measures in a statistical distribution. We focus on the behavior of the extreme values of a data set. Assume that the data are realizations of a sample X_1,...,X_n of n independent and identically distributed random variables. The ordered data will then be denoted by $X_{(1)}$,...,$X_{(n)}$. Sample data are generally used to study the properties about the distribution function

$$F(x) := P(X \leq x),$$

or about its inverse function, the quantile function defined as

$$Q(p) := \inf\{x : F(x) \geq p\}.$$

In the classical theory, one is often interested in the behavior of the mean or average. This average will then be described through the expected value $E(X)$ of the distribution. On the basis of the law of large numbers, the sample mean $\bar{X}$ is used as a consistent estimator of $E(X)$. Furthermore, the central limit theorem yields the asymptotic behavior of the sample mean. This result can be used to provide a confidence interval for $E(X)$ in case the sample size is sufficiently large, a condition necessary when invoking the central limit theorem. What if the second moment $E(X^2)$ or even the mean $E(X)$ is not finite ? Then the central limit theorem does not apply and the classical theory, dominated by the normal distribution, is no longer relevant. Or, what if one wants to estimate $\bar{F}(x) = P(X > x)$, where $x > x_{(n)}$, and the estimate $1 - F_n(x)$, where $F_n(x)$ is the empirical distribution function. Evidently we can not simply assume that these values of x are impossible. However, the traditional technique based on the empirical distribution function, does not give useful information concerning this type of question. In terms of the empirical quantile function $Q_n(p) := \inf\{x : F_n(x) \geq p\}$, problems arise when considering the extreme quantiles $Q(1-p)$ with $p < 1/n$. These observations show that it is necessary to develop special techniques that focus on the extreme values of a sample on the extreme quantiles. In practice, these extreme values are often of crucial importance. Logically, the most pertinent information for these extreme values unobserved is contained in the most extreme values observed. When using classical statistical methods, the information (the largest) contained in the rest of the sample masks the essential information concerning the rare events. Focus on the extreme values of the data allows to select only the relevant information and therefore to better extrapolate distribution tail. It is in a first step to select (and model) the extreme values of the data, ie to determine what values the most extreme of the sample will contain appropriate information on extreme events. There exist two equivalent methods for selecting: the method of maxima and the method of excess (above a threshold), see (Coles, 2001, Embrechts et al. 1997, and Reiss et al. 1997).

This theory is based on the fundamental theorem of Fisher-Tippett (1928), and Gnedenko (1943) which describes the possible limits of the law of the maximum of n random variables independent and identically distributed (i.i.d), suitably normalized. We assume always that the law which regulates the phenomenon which we are interested is in the domain of attraction of a law of extremes (GEV) G_τ, where τ is a real parameter. If we consider a sample $X_1, ..., X_n$ with the same distribution F, this means that there exist two normalizing sequences (b_n) (in $\mathbb{R}$) and (a_n) (in $\mathbb{R}_+$) such that

$$\frac{X_{(n)} - b_n}{a_n} \xrightarrow{\mathcal{D}} G_\tau.$$

This result evidently implies that the behavior of the tail depends on a single parameter, denoted τ and called extreme value index. The sign of this parameter is a key indicator of the behavior of the tail. Indeed, three behaviors are possible. When $\tau < 0$, the distribution of X is bounded and we say that we are in the field of Weibull, when $\tau = 0$, the distribution of X present an exponential type decay in the tail of distribution, we say that we are in the field of Gumbel, and finally, the field of Fréchet, corresponding to $\tau > 0$ and an unbounded distribution of X and has a decreasing of polynomial type.

There is a strong relationship between the extreme value distributions and generalized Pareto distribution (GPD), which describes the limit distribution of exceedances of a high threshold. GPD estimate is the classical way of estimating the losses and the extreme value theory is used extensively in insurance.

The second chapter is divided into two parts. The first is devoted 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

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

This estimator is an unbiased estimator and consist of $F(x)$. In addition, among the unbiased estimators of $F(x)$, $F_n(x)$ is the unique minimum variance estimator that is $F(x)(1-F(x))/n$ (Yamato, 1973). Another estimator of F is the kernel estimator $\tilde{F}_n$ which is defined by

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

where $K\left(x\right) = \int\limits_{-\infty}^{x} k\left(t\right) dt$ and k is a kernel function and h is the smoothing parameter verifying $\lim\limits_{n\to\infty} h = 0.$

The search for the asymptotic properties of $\tilde{F}_n$ was initiated by Nadaraya (1964) and continued in a series of papers among which we mention Winter (1973, 1979), Yamato (1973), Reiss (1981) and Falk (1983). The second part provides a full introduction to the non-parametric estimation of the quantile function and the density and their asymptotic properties.

Let $X_1, ..., X_n$ be independent and identically distributed with absolutely continuous distribution function F. Let $X_{(1)} \leq ... \leq X_{(n)}$ be the corresponding order statistics. Define the quantile function

Q to be the left continuous inverse of F given by

$$Q(p) = \inf(x : F(x) \geq p\}, 0 < p < 1.$$

A basic estimator of $Q\,(p)$ is the pth sample quantile which is given by

$$Q_n\,(p) = \inf\,(x : F_n\,(x) \geq p) = X_{([np]+1)},$$

where $[np]$ denotes the integer part of np, and $F_n\,(x)$ is the empirical distribution function.

A popular class of L-estimator are kernel quantile estimator given by

$$\hat{Q}_n\,(p) = \sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x-p}{h}\right) dx,$$

where k is a density function symmetric about zero, while $h := h_n \to 0$ as n tends to infinity.

Estimating the quantile function, has been treated extensively by several authors mention among them Parzen (1979), Azzalini (1981), Falk (1983-1984), Nadaraya (1964), Yamato (1973), Ralescu and Sun (1993), Yang (1985), Padgett (1986), Harrell and Davis (1982), and Sheater et Marron (1990). But most of these estimators have a problem with bias in the case of extreme quantiles. For example, Parzen (1979), Padgett (1986), Sheather and Marron (1990), and Ralescu and Sun (1993) use kernels like Gaussian kernel. But all these estimators have a large bias when p is close to 1. To correct this bias, Harrell and Davis (1982) or Park (2006) suggest using asymmetric kernel, namely the beta kernel i.e the kernel k is the density of a beta distribution.

The third chapter focuses on the study of the transformation kernel density estimation. Kernel density estimation is nowadays a classical approach to study the form of a density with no assumption on its global functional form.

Let $X_1, ..., X_n$ a random sample of i.i.d observations of a random variable with density function f, then the kernel density estimator at point x is

$$\tilde{f}_n\,(x) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{x-X_i}{h}\right), \tag{1}$$

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

When the shape of the density to be estimated is symmetric and has a kurtosis that is similar to the kurtosis of the normal distribution, then it is possible to calculate a smoothing parameter h that provides optimal smoothness or is close to optimal smoothness over the whole domain of the distribution. However, when the density is asymmetric, it is not possible to calculate a value for the smoothing parameter which captures both the mode of the density shape and the tail behavior.

The majority of economic variables that measure expenditures or costs have a strong asymmetric behavior to the right, so that classical kernel density estimation is not efficient in order to estimate the values of the density in the right tail part of the density domain. This is due to the fact that the smoothing parameter which has been calculated for the whole domain function is too small for the density in the tail.

An alternative to kernel density estimation defined in (1) is transformation kernel estimation that is based on transforming the data so that the density of the transformed variable has a symmetric shape, so that it can easily be estimated using a classical kernel estimation approach. We say it can be easily estimated in the sense that using a Gaussian kernel or an Epanechnikov kernel, an optimal estimate of the smoothing parameter can be obtained by minimizing an error measure over the whole density domain.

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

Bolancé et al. (2008) presents a comparison of the inverse beta transformation method with the results presented by Buch–Larsen et al. (2005) based only on the modified Champernowne distribution. He show that the second transformation, that is based on the inverse of a Beta distribution, improves density estimation.

The fourth chapter focuses on the estimation of extreme quantiles using the Champernowne transformation (1936-1952) which is introduced in the work of Buch–Larcen et al. (2005) in the case of density estimation for heavy tails distributions to compare the performance of the transformed estimator of extreme quantiles from the traditional kernel estimator in the sense of mean square error, which we found an improvement in this direction.

Chapter 1

Heavy-tailed distribution and extreme value theory

1.1 Heavy-tailed distribution

Many distributions that are found in practice are thin-tailed distributions. The first example of heavy tailed distributions was found in Mandelfort (1963) 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), returns on financial markets Rachev (2003), and Embrechts et al. (1997).

Heavy tailed distribution are typical in complex multi systems: Finance and business, internet traffic, hydrology, economics and have been accepted as realistic models for various phenomena, flood levels of rivers, major insurance claims, low and high temperatures. 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 variance. (Occasionally, heavy-tailed is used for any distribution that has heavier tails than the normal 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. Concerning about large losses leads us to consider $P(X > x)$ for x large. If F is the distribution function of X, we define the tail function $\bar{F}$ by

$$\bar{F}(x) = 1 - F(x).$$

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.

The usage of the term *"heavy-tailed distribution"* varies according to the area of interest, but is frequently taken to correspond to an absence of (positive) exponential moments. There are a few different definitions of heavy tailedness of a distribution. These definitions all relate to the decay of the survivor function $\bar{F}$ of a random variable. Two widely used classes of heavy tailed distributions are the regularly varying and subexponantial distributions.
Characterizing the simplest is that based on comparison with the normal law.

Definition 1.1.1 *It is said that the distribution has heavy tail if:*

$$\Delta = \frac{\mu_4}{\mu_2^2} > 3. \tag{1.1}$$

Which is equivalent to saying that a distribution to a heavy tail if and only if its coefficient of applatissement is higher than the normal with $\Delta = 3$. The characterization given by equation (1.1) 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.

Definition 1.1.2 *Let F be a distribution function (d.f) with a support on $[0, \infty)$, we say that the distribution F, is heavy tailed if it has no exponential moment, i.e.,*

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

Definition 1.1.3 *Let X a random variable with a distribution function F and the density f, this distribution is said to have a heavy tail if*

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

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

Remark 1.1.1 *If a distribution is heavy-tailed then its tail function is heavy-tailed.*

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 1.1.4 *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 1.1.5 *The distribution F is said to have a heavy tail if $\bar{F}(x) = S(x)x^{-\alpha}$ for some $\alpha > 0$ (called the tail index), and $S(.)$ is a slowly varying function at infinity.*

1.1.1 Examples of heavy-tailed distributions

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

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

for some scale parameter $c > 0$ and shape parameter $\alpha > 0$. Clearly we have $\bar{F}(x) \sim (x/c)^{-\alpha}$ 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.

ii) **Burr distribution (a model for losses in insurance):** This has tail function $\bar{F}$ given by

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

for parameters $\alpha, c, \tau > 0$. We have $F(x) \sim c^{\alpha} x^{-\tau\alpha}$ 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.

iii) **The Cauchy distribution on $\mathbb{R}$:** Recall that the density of the standard Cauchy distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

and its distribution function is

$$F(x) = \frac{1}{2} + \frac{\arctan x}{\pi},$$

and hence

$$\bar{F}(x) = \frac{1}{2} - \frac{\arctan x}{\pi},$$

we see that $\bar{F}(x) \approx (\pi x)^{-1}$, as $x \to \infty$, its tail goes to zero like the power function x^{-1}. All moments are infinite.

iv) **The lognormal distribution on $\mathbb{R}_+^*$:** The density of the lognormal distribution on $\mathbb{R}_+^*$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right),$$

for parameters μ and $\sigma > 0$. The tail of the distribution F is then

$$\bar{F}(x) = \bar{\Phi}\left(\frac{\log x - \mu}{\sigma}\right) \text{ for } x > 0,$$

where $\bar{\Phi}$ is the tail of the standard normal random variable. All moments of the lognormal distribution are finite.

v) **The Weibull distribution on $\mathbb{R}_+$:** This has tail function $\bar{F}$ given by

$$\bar{F}(x) = e^{-(x/c)^{\alpha}},$$

for some scale parameter $c > 0$ and shape parameter $\alpha > 0$. This is a heavy-tailed distribution if and only if $\alpha < 1$.

Another useful classes of heavy-tailed distributions are that regularity varying distribution and Subexponential distribution.

1.1.2 Regularity varying distribution functions

An important class of heavy tailed distributions is the class of regularly varying distribution functions. A more detail is found in Bingham et al. (1987).

Definition 1.1.6 *A distribution function is called regular varying at infinity with index $-\alpha$ 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 1.1.7 *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 $\alpha=0$ 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).

Theorem 1.1.1 *If X and Y are independent real-valued random variables with $\bar{F}_X\in\mathcal{R}_{-\alpha}$ and $\bar{F}_Y\in\mathcal{R}_{-\beta}$, with $\alpha,\beta>0$, then $\bar{F}_{X+Y}\in\mathcal{R}_\gamma$, where $\gamma=\min\{\alpha,\beta\}$.*

The same theorem, but with the assumption that $\alpha=\beta$ can be found in Feller (1971).

Proposition 1.1.1 *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 $G=F_1*F_2$ has a regularly varying tail such that :

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

From Proposition 1.1.1 we obtain the following result using induction on n.

Corollairy 1.1.1 *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 F^{n*} denotes the convolution of F n-times with itself. (See Embrechts et al. (1997).

Now consider an i.i.d sample $X_1, ..., X_n$ with common distribution F, and denote the partial sum by $S_n = X_1 + ... + X_n$ and the maximum by $M_n = \max\{X_1, ..., X_n\}$. Then for $n \geq 2$ we find that

$$P(S_n > x) = \overline{F^{n*}}(x)$$

$$\begin{aligned} P(M_n > x) &= \overline{F^n}(x) = 1 - F^n(x) \\ &= \overline{F}(x) \sum_{j=0}^{n-1} F^j(x) \sim n\overline{F}(x), \; x \to \infty. \end{aligned}$$

From this we find that we can rewrite the next corollary in the following way. If $\overline{F} \in \mathcal{R}_{-\alpha}$ with $\alpha \geq 0$ then we have

$$P(S_n > x) \sim P(M_n > x) \text{ as } x \to \infty.$$

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).

Table 1.1: Regularly varying distribution functions

Distribution	$\bar{F}(x)$ or $f(x)$	Index of regular variation
Pareto	$\bar{F}(x) = \left(\frac{c}{x+c}\right)^{\alpha}$	$-\alpha$
Burr	$\bar{F}(x) = \left(\frac{c}{x^{\tau}+c}\right)^{\alpha}$	$-\tau\alpha$
Log-Gamma	$f(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)} (\ln(x))^{\beta-1} x^{-\alpha-1}$	$-\alpha.$

1.1.3 Subexponential distribution functions

In the class of heavy-tailed distribution functions, subexponential distribution functions are a special class which have just the right level of generality for risk measurement in insurance and finance models. The name arises from one of their properties, that their right tail decreases more slowly than any exponential tail. This implies that large values can occur in a sample with non-negligible probability, which proposes the subexponential distribution functions as natural candidates for situations, where extremely large values occur in a sample compared to the mean size of the data.

Let (X_n) be i.i.d positive random variables with distribution function F with support $]0,\infty[$, the distribution function F is a subexponential distribution, written $F \in \mathcal{S}$, if for $n \geq 2$

$$\lim_{x\to\infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n, \tag{1.2}$$

where $\overline{F^{n*}}(x) = 1 - F^{n*}(x) = P(X_1 + ... + X_n > x)$, the tail of the n-fold convolution of F. Note that, by definition, $F \in \mathcal{S}$ entails that the support of F is $]0,\infty[$. Whereas regular varying that the sum of independent copies is asymptotically distributed as the maximum, from equation (1.2) we see that this fact characterizes the subexponential distributions

$$P(S_n > x) \quad \sim \quad P(M_n > x) \text{ as } x \to \infty \ \Rightarrow F \in \mathcal{S}.$$

Consider two independent, identically random variables X_1, X_2 with common distribution F, then F^{2*} is defined, using Lebesgue-Stieltjes integration by:

$$F^{2*}(x) \quad = \quad P(X_1 + X_2 \leq x) \quad = \quad \int F(x-y)\, dF(y).$$

Lemma 1.1.1 *If the following equation holds*

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

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

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

Lemma 1.1.2 *1°) If $F \in \mathcal{S}$, then uniformly y-sets of* $]0,\infty[$, *we have*

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1. \tag{1.3}$$

2°) If (1.3) holds then, for all $\varepsilon > 0$,

$$e^{\varepsilon x}\overline{F}(x) \to \infty \quad x \to \infty.$$

3°) If $F \in \mathcal{S}$ then, given $\varepsilon > 0$, there exists a finite constant c such that for all $n \geq 2$, we have

$$\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq c(1+\varepsilon)^n, \quad x \geq 0.$$

Proof. See Embrechts et al. (1997). ■

The following table gives a number of subexponential distribution:

Table 1.2: Subexponential distribution

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

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

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

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

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

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

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

The two class of distributions functions are the regularly varying distribution functions $\mathcal{R}$ and the subexponential $\mathcal{S}$.

1.2 Statistical extreme value theory

The last years have been characterized by significant instabilities in financial markets worldwide. This has led to numerous criticisms about the existing risk management systems and motivated the search for more appropriate methodologies able to cope with rare events that have heavy consequences. In such a situation it seems essential to rely on a well founded methodology. Extreme value theory (EVT) provides a firm theoretical foundation on which we can build statistical models describing extreme events.

Extreme Value theory has emerged as one of the most important statistical disciplines for the applied sciences. and their techniques are also becoming widely used in many other disciplines. The distinguishing feature of an extreme value analysis is the objective to quantify the stochastic behavior of a process at unusually large or small levels. In particular, extreme value analyses usually require estimation of the probability of events that are more extreme than any that have already been observed.

In many fields of modern science, engineering and insurance, extreme value theory is well established, see e.g. Embrechts et al. (1999), and Reiss and Thomas (1997). An alternative approach can be found in the extreme value theory, which comes from the statistics field. EVT has been applied to financial issues only in the past years, although it has been broadly utilized in other fields, such as insurance claims, telecommunications and engineering.

1.2.1 Fundamental results of extreme value theory

Let $X_1, .., X_n$ be identically distributed and independent random variables representing risks or losses with unknown cumulative distribution function (c.d.f), $F(x) = P(X_i \leq x)$. Examples of random risks are negative returns on financial assets or portfolios, operational losses, catastrophic insurance claims, credit losses, natural disasters, service life of items exposed to corrosion, traffic prediction in telecommunications, etc. See Coles (1999), and McNeil and Frey (2000).

A traditional statistical discussion on the mean is based on the central limit theorem and hence often returns to the normal distribution as a basis for statistical inference. The classical central limit theorem states that the distribution of

$$\sqrt{n}\frac{\bar{X} - E(X)}{\sqrt{Var(X)}} = \frac{X_1 + ... + X_n - nE(X)}{\sqrt{nVar(X)}},$$

converges for $n \to \infty$ to a standard normal distribution. In general, the central limit problem deals with the sum $S_n := X_1 + \cdots + X_n$ and tries to find constants $a_n > 0$ and b_n such that $Y_n = \dfrac{S_n - b_n}{a_n}$ tends in distribution to a non-degenerate distribution.

A first question is to determine what distributions can appear in the limit. The answer reveals that typically the normal distribution is attained as a limit for this sum (or average) S_n of independent and identically distributed random variables, except when the underlying distribution F possesses a heavy tail, specifically, Pareto-type distributions F with infinite variance will yield non-normal limits for the average. the extremes produced by such a sample will corrupt the average so that an asymptotic behavior different from the normal behavior is obtained.

In what follows, we will replace the sum S_n by the maximum that is the cornerstone of the extreme value theory. The model focuses on the statistical property of :

$$X_{(n)} = \max(X_1, ..., X_n).$$

Of course, we could just as well study the minimum rather than the maximum. Clearly, results for one of the two can be immediately transferred to the other through the relation

$$X_{(1)} = \min(X_1, ..., X_n) = -\max(-X_1, ..., -X_n).$$

In theory the distribution of $X_{(n)}$ can be derived exactly for all values of n :

$$P\left(X_{(n)} \leq x\right) = P(X_1 \leq x, ..., X_n \leq x) = F^n(x). \tag{1.4}$$

However, this is not immediately helpful in practice, since the distribution function F is unknown. One possibility is to use standard statistical techniques to estimate F from observed data, and then to substitute this estimate into (1.4). Unfortunately, very small discrepancies in the estimate of F can lead to substantial discrepancies for the probability distribution defined in (1.4).

An alternative approach is to accept that F is unknown, and to look for approximate families of models defined in (1.4), which can be estimated on the basis of the extreme data only. This is similar to the usual practice of approximating the distribution of sample means by the normal distribution, as justified by the central limit theorem.

It is natural to consider the probabilistic problem of finding the possible limit distributions of the maximum $X_{(n)}$. Hence, the main mathematical problem posed in extreme value theory concerns the search for distributions of X for which there exist a sequence of numbers $\{b_n; n \geq 1\}$ and a sequence of positive numbers $\{a_n; n \geq 1\}$ such that for all real values x (at which the limit is continuous)

$$P\left(\frac{X_{(n)} - a_n}{b_n} \leq x\right) \rightarrow G(x) \text{ as } n \rightarrow \infty. \tag{1.5}$$

This problem has been solved in Fisher and Tippett (1928), and Gnedenko (1943) by the following theorem that is an extreme value analog of the central limit theory, and was later revived and streamlined by de Haan (1970).

Theorem 1.2.1 (Fischer-Tippett, 1928 and Gnedenko, 1943) *Let (X_i) be independent identically distributed random variables with distribution function F. If there exist two real valued sequences $a_n > 0$ and $b_n \in \mathbb{R}$ and a distribution function G such that:*

$$\frac{X_{(n)} - b_n}{a_n} \xrightarrow{\mathcal{D}} G_\tau,$$

then, if $\tau > 0$

$$G_\tau(x) = \begin{cases} 0, & x \leq 0 \\ e^{(-x)^{-\tau}}, & x > 0, \end{cases}$$

if $\tau < 0$

$$G_\tau(x) = \begin{cases} e^{-(-x)^{\tau}}, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

and if $\tau = 0$

$$G_0(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$$

Remark 1.2.1 *1) The previous theorem is true for most of the usual laws.*

2)The distribution function G_τ is called Generalized Extreme Value Distribution. The parameter τ is called the extreme value index. If F verifies the precedent Theorem , we say that belongs to the domain of attraction of G_τ.
3) Within the sign of τ there are three areas of attraction.
- *If $\tau > 0$ we say that F belongs to the domain of attraction of Frechet. This domain of attraction contains the heavy tailed distribution functions (polynomial decay) such as the Cauchy distribution, the Pareto, the Burr, the inverse gamma, the log gamma distributions etc, (see Gnedenko 1943).*
- *If $\tau < 0$ we say that F belongs to the domain of attraction of Weibull. This domain of attraction contains the majority of distribution functions whose end point is finite (uniform law, Beta(p,q), Extreme value Weibull distributions etc.)*
- *If $\tau = 0$ we say that F belongs to the domain of attraction of Gumbel. This domain of attraction contains the functions of exponential decay distribution (Gaussian, exponential, gamma, lognormal, Logistic, etc.)*
- *The sequences of normalization a_n and b_n are not unique.*

The Fischer-Tippett Theorem is stating that the distribution function describing the dynamic of extreme events belongs to Maximum Domain of Attraction of a Generalized Extreme Value Distribution, that is

Definition 1.2.1 *The Generalized Extreme Value Distribution $G_{\tau,\mu,\sigma}(z)$, is defined by*

$$G_{\tau,\mu,\sigma}(z) = \begin{cases} \exp\left(-\left(1+\tau\frac{z-\mu}{\sigma}\right)^{-1/\tau}\right) & \tau \neq 0 \\ \exp\left(-\exp\left(-\frac{z-\mu}{\sigma}\right)\right) & \tau = 0, \end{cases}$$

$G_{\tau,\mu,\sigma}(z)$ is defined on $\{z : 1+\tau(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$, and the real parameter τ is a shape parameter that determines the tail behavior of $G_\tau(z)$.

Gnedenko (1943) accomplished an important excursion related to this result in 1943. He proved that The Fischer-Tippett theorem is applicable for heavy tailed distributions functions. More precisely, he shown that heavy tailed distribution functions belong to the Maximum Domain of Attraction of the Frechet Distribution.

It is here a brief introduction to the study of the asymptotic behavior of a sample of the maximum (extreme value theory). This study using the notion of a regular variation functions. It then gives a result describing the possible limits of the law of the maximum of a sample. For more details on the extreme value theory, one can refer to the works of Castillo (1988), Gumbel (1958), Resnick (1987) and Galambos (1987).

1.2.2 Characterization of domains of attraction

We will give conditions on the distribution function F for it belongs to one of three domains of attraction. In the following, we denote $x_F = \sup\{x/F(x) < 1\}$ the end point of F and
$F^{-1}(y) = \inf\{x \in \mathbb{R}/F(x) \geq y\}$.

Domain of attraction of Frechet :

Theorem 1.2.2 *The function distribution F belongs to the domain of attraction of Frechet D(Frechet) with extreme value index $\tau > 0$ if and only if $x_F = +\infty$ and $1 - F$ is a regular varying function with index $-1/\tau$. In this case the choice of the sequences a_n and b_n is*

$$a_n = F^{-1}\left(1 - \frac{1}{n}\right), \quad b_n = 0.$$

From this Theorem, we deduce that $F \in D$(Frechet) if and only if the end point x_F is infinite $\bar{F}(x) = x^{-1/\tau} S(x)$ where S is a slowly varying function at infinity and τ a positive real.
A well-known sufficient condition can be given in terms of the hazard function

$$r(x) = \frac{f(x)}{1 - F(x)},$$

where it is assumed that F has a derivative f.

Proposition 1.2.1 *Von Mises' theorem. If $x_F = \infty$ and $\lim\limits_{x \to \infty} xr(x) = \alpha > 0$, then $F \in D$(Frechet) of parameter α*

Domain of attraction of Weibull :

Theorem 1.2.3 *The function distribution F belongs to the domain of attraction of Weibull D(Weibull) with extreme value index $\tau < 0$ if and only if $x_F < +\infty$ and $1 - F^*$ is a regular varying function with index $1/\tau$, with*

$$F^*(x) = \begin{cases} F\left(x_F - x^{-1}\right) & \text{if } \ x > 0 \\ 0 & \text{if } \ x \leq 0 \end{cases}$$

and we note that $F \in D$(Weibull). In this case the choice of the sequences a_n and b_n is

$$a_n = x_F - F^{-1}\left(1 - \frac{1}{n}\right), \quad b_n = x_F.$$

From Theorem1.2.3, we deduce that $F \in D$(Weibull) if and only if the end point x_F is finite and $\bar{F}(x) = (x_F - x) S\left[(x_F - x)^{-1}\right]$ with S is a slowly varying function at infinity and τ a strictly negative real.

Proposition 1.2.2 *Von Mises' theorem. If $x_F < \infty$ and $\lim\limits_{x \to x_F} (x_F - x) r(x) = \alpha > 0$, then $F \in D$(Weibull) of parameter α*

Domain of attraction of Gumbel :

Theorem 1.2.4 *The function distribution F belongs to the domain of attraction of Gumbel D(Gumbel) if and only if for $z < x < x_F$ we have*

$$1 - F(x) = d(x) \exp\left(-\int_z^x \frac{1}{c(t)} dt\right)$$

where $d(x) \underset{x \to x_F}{\rightarrow} d > 0$ and c is positive continuous absolutely function verifying that:

$$\lim_{x \to x_F} c'(x) = 0.$$

In this case the choice of the sequences a_n and b_n is

$$a_n = F^{-1}\left(1 - \frac{1}{n}\right), \; b_n = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(y)\, dy$$

The von Mises sufficiency condition is a bit more elaborate than before.

Proposition 1.2.3 *Von Mises' theorem. If $r(x)$ is ultimately positive in the neighborhood of x_F, is differentiable there and satisfies $\lim\limits_{x \to x_F} \frac{dr(x)}{dx} = 0$, then $F \in D$(Gumbel).*

1.2.3 Extremes quantile estimation

The quantile estimation procedure is making use of EVT and is relying essentially on the papers of Smith (1987) and the one of Mc-Neil (1999) dealing with the approximation of the tail of probability distributions .The initial ideas of this estimation procedure can also be found in Hosking and Wallis (1987), where the author is presenting some results concerning the estimation of the parameters and quantile for the Generalized Pareto Distributions(GPD). This approach leads to an invertible form of the distribution function of the innovations which help to get easily the estimator of the required quantile with appealing asymptotic properties. The use of EVT and GPD as a tool in financial risk management is also developed in Mc-Neil (1999) or Embrechts (1997). This approach consists of an appropriate choice of a threshold level u and estimating the distribution function F, by its sample version below the threshold and some GPD over the chosen threshold. For that, the concept of Excess Distribution will be defined and some fundamental results of the theory of extreme value will be recalled. Such results, due to Pickand (1975) and Fischer enable to approximate accurately the Excess Distribution over the threshold level.

We wish to estimate small probabilities or quantities whose probability observation is very low, that is to say close to zero. These quantities are called quantiles, and we talk about extreme quantile when the order of the quantile (probability of observation) converges to zero as the sample size goes to infinity.

Specifically, we consider n real random variables $\{X_i,\ i = 1, ..., n\}$ independent and identically distributed with distribution function F not necessarily continuous. From the observations of these random variables, we wish to estimate the extreme quantile of order $1-\alpha_n \to 1$ as $\alpha_n \underset{n\to\infty}{\to} 0$ defined by

$$Q(1-\alpha_n) = \inf\{x : F(x) \geq 1-\alpha_n\}.$$

In particular for n tending to infinity, we have

$$\begin{aligned} P\left(X_{(n)} < Q(1-\alpha_n)\right) &= (1-\alpha_n)^n \\ &= \exp(n \ln(1-\alpha_n)) \\ &= \exp(-n\alpha_n(1+o(1))) \text{ as } n \to \infty. \end{aligned}$$

Therefore, as $\alpha_n \to 0$, assuming that $n\alpha_n \to 0$ as $n \to \infty$ implies that

$$P\left(X_{(n)} < Q(1-\alpha_n)\right) \to 1.$$

It can not therefore be estimated $Q(1-\alpha_n)$ by reversing simply the empirical distribution function.

Several methods of estimating the extreme quantile $Q(1-\alpha_n)$ have been proposed in the literature. But before exposing they should be exposing the underlying theory to the study of a maximum of a sample.

We assume that F belongs to one of the areas attractions defined above. To summarize the estimation problem, we introduce the following result known as Poisson approximation.

Lemma 1.2.1 (Embrechts et al. 1997) *If $\alpha_n \to 0$ and $n\alpha_n \to c$ (not necessarily finite) as $n \to \infty$, then*

$$P\left(X_{(n)} < Q(1-\alpha_n)\right) \to \exp(-c).$$

Thus, by the precedent Lemma, two situations can be distinguished as a function of c when you want to estimate the quantiles of order $1-\alpha_n$.

1. If $c = \infty$, then $P\left(X_{(n)} < Q(1-\alpha_n)\right) = 0$. In this case a natural estimator of $Q(1-\alpha_n)$ is is the empirical quantile $X_{(n-[n\alpha_n]+1)}$.

2. If $c = 0$, then $P\left(X_{(n)} < Q(1-\alpha_n)\right) = 1$. Therefore we can not estimate the quantile empirically. To resolve this behavior, we have identified two main categories of methods:

- Using the relation $P\left(X_{(n)} < Q(1-\alpha_n)\right) = F^n(Q(1-\alpha_n))$ we can then estimate the extreme quantile $Q(1-\alpha_n)$ by the law for extreme values. We then have an extreme quantile estimator

$$\begin{aligned} Q(1-\alpha_n) &= \hat{a}_n x_{\alpha_n} + \hat{b}_n \\ &= \begin{cases} \hat{a}_n (n\alpha_n)^{-\hat{\tau}} + \hat{b}_n & \text{if } F \in \mathcal{D}(\text{Frechet}) \\ -\hat{a}_n (n\alpha_n)^{-\hat{\tau}} + \hat{b}_n & \text{if } F \in \mathcal{D}(\text{Weibull}) \\ -\hat{a}_n \log(n\alpha_n) + \hat{b}_n & \text{if } F \in \mathcal{D}(\text{Gumbell}), \end{cases} \end{aligned}$$

  where x_{α_n} verifying $-\log G_\tau(x_{\alpha_n}) = n\alpha_n$ and $\hat{a}_n,\ \hat{b}_n,\ \hat{\tau}$ are respectively the estimators of $a_n,\ b_n,\ \tau$.

- The method of excess is initially presented by Pickands (1975). It advocates retain only the observations above a threshold u. The law of m observations thus retained denoted by $\{X_i, i = 1, ..., m\}$ can be approached, if u is large by a generalized Pareto distribution (GPD). To estimate the quantile extreme $Q(1-\alpha_n)$, it is sufficient to use the result of Balkema and de Haan (1974), and Pickands (1975) which establishes the equivalence between the convergence in law to a law of the maximum extreme value and convergence law of excess to a GPD.

Before treating this method we start with a definition

Definition 1.2.2 *The Generalized Pareto Distribution $G_{\tau,\beta}$ is given as*

$$G_{\tau,\beta}(z) = \begin{cases} 1 - \left(1 + \dfrac{\tau z}{\beta}\right)^{-1/\tau} & \tau \neq 0 \\ 1 - \exp(-z/\beta) & \tau = 0. \end{cases}$$

where β is the scale parameter and τ is the shape parameter. The Generalized Pareto Distribution is defined under the following conditions

$$\begin{cases} 1^{\circ}) & \beta > 0 & \\ 2^{\circ}) & z \in \left[0, \dfrac{-\beta}{\tau}\right] & \textit{if } \tau < 0 \\ 3^{\circ}) & z \geq 0 & \textit{if } \tau \geq 0. \end{cases}$$

For that, the concept of Excess Distribution will be defined. Such results, due to Pickand (1975) and Fischer enable to approximate accurately the Excess Distribution over the threshold level.

1.2.4 Excess distribution function estimation

Our problem is that we consider an unknown distribution F of a random X. We are interested in estimating the distribution function F_u of values of x above a certain threshold u. The distribution function F_u is called the excess distribution function. In this approach for estimating extreme quantiles, it retains only the observations exceeding a threshold $u < x_F$. We define the excess Y of the variable X above the threshold u by $X - u$ given $X > u$. If we denote by the distribution function F_u an excess above the threshold u, we have for all $y > 0$:

$$\begin{aligned} F_u(y) &= P(Y \leq y) = P(X - u \leq y/X > u) \\ &= P(X \leq u + y/X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}. \end{aligned}$$

Definition 1.2.3 *We consider an unknown distribution function F of a random variable X supposed to be heavy tailed, the Excess Distribution Function above an appropriately high threshold u, is defined by:*

$$F_u(y) = P(X - u \leq y / X > u) = \frac{F(y+u) - F(u)}{1 - F(u)},\ 0 \leq y \leq x_F - u,$$

where $x_F = \sup\{x / F(x) < 1\} \leq \infty$ is the right end point.

When the threshold u is large, we can approximate this quantity by the function of survival of a GPD. To approximate the quantile, it is sufficient to use the result of Balkema and de Haan (1974) and Pickands (1975) which establishes the equivalence between the convergence in law to a law of the maximum extreme value and the convergence in law of excess to a GPD. This result is stated as follows.

Theorem 1.2.5 (Pickands-Balkema-de Hann, 1974–1975) *F belongs to the domain of attraction of $G_{\tau,\beta}$ if and only if*

$$\lim_{u \to x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - G_{\tau,\beta}(x)| = 0.$$

This theorem is very useful when working with observations that exceed a fixed threshold because it assures that the excess distribution function can be approximated by a generalized Pareto distribution.

Since $1 - F(x) = (1 - F(u))(1 - F_u(x - u))$. If for all $y \geq 0$ we set $Q(1 - \alpha_n) = x = u + y$, then

$$\begin{aligned} \alpha_n = 1 - F(Q(1-\alpha_n)) &= (1 - F(u))(1 - F_u(Q(1-\alpha_n) - u)) \\ &= (1 - F(u))(1 - G_{\tau,\beta}((Q(1-\alpha_n) - u)). \end{aligned}$$

For m excess above the threshold u, the approximation $1 - F(u) \simeq m/n$ leads to

$$\alpha_n = \frac{m}{n}(1 - G_{\tau,\beta}((Q(1-\alpha_n) - u)),$$

and if $\tau \neq 0$, then we approach the quantile by

$$Q(1-\alpha_n) \simeq u + \frac{\beta}{\tau}\left(\left(\frac{m}{n\alpha_n}\right)^{\tau} - 1\right).$$

We then have an estimator of type

$$\hat{Q}\left(1-\alpha_n\right)=\frac{\left(\frac{m}{n\alpha_n}\right)^{\hat{\tau}}-1}{\hat{\tau}}\hat{\beta}+u,$$

where $\hat{\beta}$ and $\hat{\tau}$ are respectively the estimators of the shape and scale parameters.

Another classical estimator is the so-called Hill estimator, based on regular variation properties of the survival distribution $\bar{F}$ of X given $X>u$, i.e.

$$\hat{Q}\left(1-\alpha_n\right)=u\left(\frac{m}{n\alpha_n^{\hat{\tau}}}\right),$$

where $\hat{\tau}=\frac{1}{k}\sum_{i=1}^{m}\log\left(X_{(n-i+1)}-\log u\right)$.

Generalized Extreme and Pareto Distribution functions play a crucial role in the study of financial market extreme events more specialty in financial market-crashes or extreme loss quantification in insurance mainly during earthquake or hurricane.

Beyond the important fact that Generalized Distributions help to estimate tails of distributions, they also provide accurate estimation tools that can be used to construct quantile estimation of heavy tailed distributions. In order to estimate the tails of the loss distribution, we resort to a theorem of Pickands-Balkema-de Hann (1974–1975) which establishes that, for a sufficiently high threshold u, $F_u\left(x\right)\approx G_{\tau,\beta}\left(x\right)$ (see Embrechts, Klüpperberg and Mikosch, 1997). By setting $x=u+y$, an approximation of $F(x)$, for $x>u$, can be obtained that

$$F\left(x\right)=\left(1-F\left(u\right)\right)G_{\tau,\beta}\left(x-u\right)+F\left(u\right).$$

The function $F(u)$ can be estimated non-parametrically using the empirical c.d.f

$$\hat{F}\left(u\right)=\frac{n-m}{n},$$

where m represents the number of exceedences over the threshold u, then we get the following estimate for $F(x)$

$$\hat{F}\left(x\right)=1-\frac{m}{n}\left(1+\frac{\hat{\tau}\left(x-u\right)}{\hat{\beta}}\right)^{-1/\tau}, \tag{1.6}$$

where $\hat{\tau}$ and $\hat{\beta}$ are estimates of τ and β, respectively, which can be obtained by the method of maximum likelihood.

For $\tau \neq 0$ the log-likelihood is given as

$$l(\tau, \beta) = -m \log \beta - \left(1 + \frac{1}{\tau}\right) \sum_{i=1}^{m} \log \left(1 + \frac{\tau Y_i}{\beta}\right).$$

In the case $\tau = 0$ the log-likelihood is given as

$$l(\tau, \beta) = -m \log \beta - \frac{1}{\beta} \sum_{i=1}^{m} Y_i.$$

Analytical maximization of the log-likelihood is not possible, so numerical techniques are again required.

Chapter 2

Nonparametric distribution and quantiles estimation

The estimate of the distribution function of a random variable is an important part of the non-parametric estimation. A common problem in statistics is that of estimating a density f or a distribution function F from a sample of variables random real $X_1, ..., X_n$ independent and identically distributed. The functions f and F, completely describe-the probability law of the observations and to know an appropriate estimation can solve many statistical problems to know an appropriate estimation that can solve many statistical problems.

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.

2.1 The empirical distribution function

Let $X_1, ..., X_n$ be independent random variables identically distributed as a random variable X whose distribution function $F(x) = P(X \leq x)$ is absolutely continuous

$$F(x) = \int_{-\infty}^{x} f(t)\, dt,$$

with probability density function $f(x)$. As an estimate of the value of the value $F(x)$ of the distribution function at a given point x. Traditionally, the estimator of F, from $X_1, ..., X_n$, is the so called empirical distribution function (e.d.f) F_n defined at some point x as

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

where

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

The e.d.f 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 e.d.f. $F_n(x)$ is

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

This estimator is widely in practice despite the known fact that smoothing can produce. Let $T_n\left(x\right) = nF_n\left(x\right)$, so that $T_n\left(x\right)$ represents the total number of sample values that are less than or equal to the specified value x. We see that $T_n\left(x\right)$ is essentially a binomially distributed random variable of parameters $\left(n, F\left(x\right)\right)$.

2.1.1 Statistical properties

Using properties of the binomial distribution, we get the following results.

Corollairy 2.1.1 *The mean and the variance of $F_n(x)$ are*

$$E\left(F_n(x)\right) = F\left(x\right) \quad \text{and} \quad V\left(F_n\left(x\right)\right) = \frac{F\left(x\right)\left(1 - F\left(x\right)\right)}{n}.$$

The corollary shows that $F_n(x)$, the proportion of sample values less than or equal to the specified value x, is an unbiased estimator of $F\left(x\right)$ and shows that the variance of $F_n(x)$ tends to zero as n tends to infinity. Thus, using Chebyshev's inequality, we can show that $F_n(x)$ is a consistent estimator of $F\left(x\right)$.

Corollairy 2.1.2 *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. A probability statement can be made simultaneously for all x, as a result of the following important theorems.

Theorem 2.1.1 (Glivenko-Cantelli Theorem) *$F_n(x)$ converge uniformly to $F(x)$ with probability* 1, *that is*

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

Theorem 2.1.2 (Dvoretsky-Kiefer-Wolfowitz) *For any $\varepsilon>0$,*

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

Another useful property of the e.d.f is its asymptotic normality, given in the following theorem.

Theorem 2.1.3 *As $n\to\infty$, the limiting probability distribution of the standardized $F_n(x)$ is standard normal, or*

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

Despite the good statistical of F_n, one could prefer in many applications a rather smooth estimate see Azzalini (1981).

2.2 Kernel distribution function estimator

Let $X_1,...,X_n$ be independent random variables identically distributed which are drawn from a continuous distribution $F(x)$ with density function $f(x)$. The kernel density estimate with appropriate kernel function $k(t)$ and $h=h_n$ is the bandwidth or the smoothing parameter

$$\tilde{f}_n(x)=\frac{1}{nh}\sum_{i=i}^{n}k\left(\frac{x-X_i}{h}\right).$$

This estimator is a popular nonparametric estimate of $f(x)$ which is introduced by Rosenblatt (1956) and Parzen (1962). The density estimator can be integrated to obtain a nonparametric alternative to $\tilde{F}_n(x)$ for smooth distribution function that said the kernel distribution function estimator k.d.f.e $\tilde{F}_n(x)$ that was proposed by Nadaraya (1964) and is defined by

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

where the function K is defined from a kernel k as

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

where k is a kernel function, and $h = h_n$ is the smoothing parameter or the bandwidth since it controls the amount of smoothness in the estimator for a given sample of size n.

We assume that the kernel function k is a continuous density such that is bounded, symmetric about zero $(k(t) = k(-t))$. Thus $k(t)$ satisfies

$$\int_{-\infty}^{\infty} k(t)\,dt = 1, \quad \int_{-\infty}^{\infty} tk(t)\,dt = 0 \quad \text{and} \int_{-\infty}^{\infty} t^2 k(t)\,dt < \infty.$$

The smoothing parameter h which tends to 0 as $n \to \infty$.

The estimate $\tilde{F}_n(x)$ has been investigated by several authors, Nadaraya (1964) has proved under mild conditions that $\tilde{F}_n(x)$ has asymptotically unbiased and has the same variance as F_n, with f is continuous Nadaraya (1964), Winter (1973), and Yamato (1973) are obtains its uniform convergence to F with probability one, and without conditions on f, Singh, et al. (1983). Winter (1979) also shows that checks the Chung-Smirnov property, that

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

with probability 1. Watson and Leadbetter (1964) proved the asymptotic normality of $\tilde{F}_n(x)$. Reiss (1981) proves that the asymptotic relative inefficiency of F_n compared to $\tilde{F}_n(x)$ tends rapidly to infinity as the sample size increases with an appropriate choice of kernel, e.g.

$$k(x) = \frac{9}{8}\left(1 - \frac{5}{3}x^2\right) I_{|x|\leq 1}.$$

Falk (1983), who has shown that the asymptotic performance of $\tilde{F}_n$ is better than that of F_n in the sense of relative deficiency for appropriately chosen kernels and sufficiently smooth c.d.f's F.

Azzalini (1981) derived also an asymptotic expression for the mean squared error MSE of $\tilde{F}_n(x)$ and determined the asymptotically optimal smoothing parameter, to have an MSE lower for F_n, for details see (Mack, 1984, and Hill, 1985), and he obtained the asymptotic expressions for the mean integrated squared error $MISE$ of $\tilde{F}_n(x)$. And some conditions verified in particular when the support of k is bounded and

$$\varphi(k) = 2\int_{-\infty}^{\infty} xk(x)\,K(x)\,dx > 0,$$

where $K(x) = \int_{-\infty}^{x} k(y)\,dy$.

Falk (1983) provides a complete solution to this problem by establishing on the representation of relative inefficiency of F_n versus $\tilde{F}_n$ under the above conditions especially when the support of k is bounded. The number $\varphi(k)$ is introduced by Falk (1984) as a measure of asymptotic performance of the kernel k. But he shows that any square integrable kernel does minimizes φ. Then he uses the number $\rho(k) = \int k^2(y)\,dy$ defined by Epanechnikov (1969) as a measure of the performance of the kernel in density estimation. In the sense of ρ, the kernel of Epanechnikov following

$$k(x) = \frac{3}{4}\left(1 - x^2\right) I_{(|x|\leq 1)},$$

is the best but the Gaussian or uniform kernels have very similar performance. Using the criterion φ the Epanechnikov kernel is then by far the best of the three.

In the sense of mean integrated squared error $MISE$, the best kernel is the uniform kernel although the performance of other kernels (Epanechnikov, normal, triangular) are, in practice, only slightly less good (Jones, 1990). It is interesting to note that this is not the best kernel in the estimation of density.

The asymptotic expression of $MISE$. is also studied by SwanPoel (1988). For a continuous function f, he proves that the best kernel is the uniform kernel $k(x) = (1/2\xi)\,I_{[-\xi,\xi]}(x)$ for an arbitrary constant $\xi > 0$ (indicating that the criteria for Falk to define an optimal kernel are really not adapted to the distribution function), whereas for discontinuous f in a finite number of points, the exponential kernel $k(x) = \frac{c}{2} \exp(-c|x|)$ for an arbitrary constant $c > 0$, $\tilde{F}_n(x)$ is again more efficiency than F_n for $h_n = o\left(n^{-1/2}\right)$ and $\varphi(k) > 0$. However, $\tilde{F}_n(x)$ does not always provide a better estimate than F_n. Indeed, in the case of a uniformly Lipschitz function F, Fernholz (1991) obtain that

$$\sqrt{n}\left\|\tilde{F}_n(x) - F_n(x)\right\|_\infty \to 0 \text{ a.s,}$$

and that $\sqrt{n}\left\|\tilde{F}_n(x) - F(x)\right\|_\infty$ and $\sqrt{n}\left\|F_n(x) - F(x)\right\|_\infty$ have the same asymptotic distribution. More, Shirahata and Chu (1992) show that under certain hypotheses on F, the integrated square error $ISE = \int \left(\tilde{F}_n(x) - F(x)\right) dF(x)$ for $\tilde{F}_n(x)$ is almost certainly higher than that of $F_n(x)$.

2.2.1 Mean squared error

We first obtain the MSE. The assumptions used by Azzalini (1981) which are that f is continuous and differentiable with finite mean and square integrable derivatives, $h \to 0$ and $nh \to \infty$ as $n \to \infty$, and the kernel satisfying the above assumptions. We have

$$E\left(\tilde{F}_n(x)\right) = F(x) + \frac{h^2 f'(x)\mu_2(k)}{2} + o\left(h^2\right),$$

$$Bias^2\left(\tilde{F}_n(x)\right) = \frac{h^4 f'^2(x)\mu_2^2(k)}{4} + o\left(h^4\right),$$

and

$$V\left(\tilde{F}_n(x)\right) = n^{-1}F(x)(1 - F(x)) - n^{-1}hf(x)\varphi(k) + o\left(\frac{h}{n}\right),$$

where

$$\mu_2(k) = \int_{-\infty}^{\infty} x^2 k(x)\,dx, \text{ and } \varphi(k) = 2\int_{-\infty}^{\infty} xk(x)K(x)\,dx.$$

The $MSE\left(\tilde{F}_n(x)\right)$ is given by

$$\begin{aligned} MSE\left(\tilde{F}_n(x)\right) &= Bias^2\left(\tilde{F}_n(x)\right) + V\left(\tilde{F}_n(x)\right) \\ &= \frac{h^4 f'^2(x)\mu_2^2(k)}{4} \\ &+ n^{-1}F(x)(1 - F(x)) - n^{-1}hf(x)\varphi(k) + o\left(h^4 + \frac{h}{n}\right), \end{aligned}$$

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

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

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

$$\hat{h} = \left(\frac{f(x)\varphi(k)}{nf'^2(x)\mu_2^2(k)}\right)^{1/3}$$

and the associated asymptotic mean squared error is given by :

$$n^{-1}\left[F(x)(1-F(x)) - \frac{3}{4}\left(\frac{f^4(x)\varphi^4(k)}{nf'^2(x)\mu_2^2(k)}\right)^{1/3}\right].$$

The $AMISE\left(\tilde{F}_n(x)\right)$ is found by integrating the $AMSE\left(\tilde{F}_n(x)\right)$ which is

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

The value of h that minimizes the $AMISE\left(\tilde{F}_n(x)\right)$ is

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

and the optimal $AMISE\left(\tilde{F}_n(x)\right)$ is given by

$$n^{-1}\left[\int F(x)(1-F(x)) - \frac{3}{4}\left(\frac{\varphi^4(k)}{n\mu_2^2(k)\int f'^2(x)\,dx}\right)^{1/3}\right].$$

which is lower than of the (e.d.f). From this expression we learn the following.

1. While the improvement over the $\tilde{F}_n(x)$ disappears as $n \to \infty$, it does so the slow rate of $n^{-1/3}$, which suggests that the $\tilde{F}_n(x)$ may have meaningful finite sample gain over the $F_n(x)$.

2. The improvement of $\tilde{F}_n(x)$ is inversely proportional to $\rho(f') = \int f'^2(x)\,dx$. Thus we expect the gains to minimal when the density is steep.

3. The choice of kernel k only affects the $AMISE$ through φ (larger values reduce the $AMISE$).

4. The estimator $\tilde{F}_n(x)$ is asymptotically more efficient than the $F_n(x)$ see (Swanapoel 1988).

2.3 Quantiles estimation

Quantile estimation plays an important role in a wide range of statistical application: the Q-Q plot,Value at risk, in financial risk management, etc. The estimation of population quantiles is of great interest when a parametric form for the underlying distribution is not avaible. In addition, quantiles often arise as the natural thing to estimate when the underlying distribution is skewed. The quantile function estimation can be broken down into two approaches, parametric and nonparametric.

2.3.1 Parametric estimation

Assume that the distribution F_X is continuous and belongs to some parametric distribution family $\mathcal{F} = \left\{F_\theta, \theta \in \Theta \subset \mathbb{R}^k\right\}$. The idea of parametric estimation is to assume that any statistical quantity can be seen as a function of θ. Then, the natural estimator of the quantile $Q_X(p) = F_\theta^{-1}(p)$ is obtained by substituting some parameter estimator $\hat{\theta}$ for θ, and the natural estimator would be

$$\hat{Q}_X(p) = F_{\hat{\theta}}^{-1}(p).$$

This method is convenient for practical purposes, since several techniques exist for obtaining $\hat{\theta}$ (maximum likelihood, moment method...), but the choice of $\mathcal{F}$ is crucial. A natural idea (that can be found in classical financial models) is to assume Gaussian distributions : if $X \sim N(\mu, \sigma)$, then the quantile $Q_X(p)$ is simply

$$Q_X(p) = \mu + \Phi^{-1}(p)\,\sigma,$$

where Φ^{-1} is the inverse of a normal distribution.

Definition 2.3.1 *Let $X_1, ..., X_n$ be independent and normally distributed with distribution function Φ, the (Gaussian) parametric estimation of the p-quantile $Q_X(p)$ is*

$$\hat{Q}_n(p) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(p),$$

where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$, *and* $\hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$.

A parametric models actually, is the Gaussian model does not fit very well, it is still possible to use Gaussian approximation. If the variance is finite, $(X - E(X))/\sigma$ might be closer to the Gaussian distribution, and thus, consider the so-called Cornish-Fisher approximation, i.e.

$$Q_X(p) \sim E(X) + \sigma\hat{z}_p,$$

where

$$\begin{aligned} \hat{z}_p &= \Phi^{-1}(p) + \frac{\hat{\xi}_1}{6}\left(\left(\Phi^{-1}(p)\right)^2 - 1\right) \\ &+ \frac{\hat{\xi}_2}{24}\left(\left(\Phi^{-1}(p)\right)^3 - 3\Phi^{-1}(p)\right) \\ &- \frac{\hat{\xi}_1^2}{36}\left(2\left(\Phi^{-1}(p)\right)^3 - 5\Phi^{-1}(p)\right), \end{aligned}$$

where $\hat{\xi}_1$ is the natural estimator of the skewness ξ_1 of X, and $\hat{\xi}_2$ is the natural estimator of the the excess kurtosis ξ_2 of X, i.e.

$$\hat{\xi}_1 = \frac{\sqrt{n(n-1)}}{n-2} \frac{\sqrt{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^3}{\left(\sum_{i=1}^{n}(X_i - \hat{\mu})^2\right)^{3/2}},$$

and

$$\hat{\xi}_2 = \frac{n-1}{(n-2)(n-3)}\left((n+1)\hat{\xi}_2' + 6\right),$$

where

$$\hat{\xi}_2' = \frac{n\sum_{i=1}^{n}(X_i - \hat{\mu})^4}{\left(\sum_{i=1}^{n}(X_i - \hat{\mu})^2\right)^2} - 3.$$

Definition 2.3.2 *Given a n sample* $\{X_1, ..., X_n\}$*, the Cornish-Fisher estimation of the p-quantile* $Q_X(p)$ *is*

$$\hat{Q}_n(p) = \hat{\mu} + \hat{\sigma}\hat{z}_p$$

2.3.2 Nonparametric estimation

Let $X_1, ..., X_n$ be independent and identically distributed with absolutely continuous distribution function F. Let $X_{(1)} \leq ... \leq X_{(n)}$ be the corresponding order statistics. Define the quantile function Q to be the left continuous inverse of F given by

$$Q(p) = \inf(x : F(x) \geq p\}, 0 < p < 1.$$

A basic estimator of $Q(p)$ is the pth sample quantile which is given by

$$Q_n(p) = \inf(x : F_n(x) \geq p) = X_{([np]+1)},$$

where $[np]$ denotes the integer part of np, and $F_n(x)$ is the empirical distribution function.

Making use of the fact that $Q_n(p)$ is the inverse of the empirical distribution function. However, one can improve on the estimator $Q_n(p)$ of $Q(p)$ by averaging over the order statistics, using suitable weights w_i

$$L_n = \sum_{i=1}^{n} w_i X_{(i)} \text{ where } \sum_{i=1}^{n} w_i = 1$$

These estimators are called L-estimators.

Notice that $Q_n(p)$ is an L-estimator with $w_{[np]+1} = 1$ and $w_i = 0$ for $i \neq [np] + 1$.

One has, under mild regularity conditions

$$\sqrt{n}\left(Q_n(p) - Q(p)\right) \xrightarrow{\text{D}} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(Q(p))}\right),$$

where f is the density of F. See (Serfling, 1980).

A popular kernel quantile estimator, is based on the Nadaraya (1964) type kernel distribution function estimator $\tilde{F}_n(x)$ for $F(x)$ that defined as

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

where the function K is defined from a kernel k as

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

and k is a density function, $h = h_n$ is the smoothing parameter (or the bandwidth) since it controls the amount of smoothness in the estimator, and satisfy $h := h_n \to 0$ as $n \to \infty$.

The corresponding estimator of the quantile function $Q = F^{-1}$ is then defined by

$$\tilde{Q}_n(p) = \inf\{x : \tilde{F}_n(x) \geq p\}, 0 < p < 1.$$

Nadaraya (1964) showed under some assumptions for $k,\ f$ and h, $\tilde{Q}_n(p)$ has an asymptotic standard normal distribution. The almost sure consistency, was obtained by Yamato (1973). Ralescu and Sun (1993) obtained the necessary and sufficient conditions for the asymptotic normality of $\tilde{Q}_n(p)$.

A popular class of L-estimator are kernel quantile estimator given by

$$\hat{Q}_n(p) = \sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x - p}{h}\right) dx.$$

Here k is a density function symmetric about zero, while $h := h_n \to 0$ as n tends to infinity.

This form can be traced to Parzen (1979), Falk (1984) investigated the asymptotic relative deficiency of the sample quantile with respect to $\hat{Q}_n(p)$, and showed that the asymptotic performance of $\hat{Q}_n(p)$ is better than that of the empirical sample quantile, Yang (1985) established the asymptotic normality and mean squared consistency of $\hat{Q}_n(p)$, Padgett (1986) generalized the definition of $\hat{Q}_n(p)$ to right-censored data.

In studies of $\hat{Q}_n(p)$, Yang (1985) and Padgett (1986) examined several kernel functions including the triangular functions given by

$$k_Y(u) = \begin{cases} 1-|u| & \text{if } |u \leq 1| \\ 0 & \text{otherwise.} \end{cases}$$

The kernel functions k_Y was shown to have optimal properties by Sachs and Ylvisaker (1981) in nonparametric density estimation functions. There are many similarities between the smooth kernel estimator $\hat{Q}_n(p)$ of $Q(p)$ and the kernel method used in density estimation. If the kernel function k is replaced by its derivative in the definition of $\hat{Q}_n(p)$, Yang (1985) suggested that the resulting statistic would be useful estimator of the derivative of $Q(p)$.

An analogous result for the derivative density estimation Silvermann (1978). It is useful to have a good estimator of $Q'(p) = 1/f(Q(p))$ in order to estimate the variance of $Q_n(p)$ and $\hat{Q}_n(p)$. When the sample size n is large , both $\hat{Q}_n(p)$ has the same approximate standard deviation that $Q_n(p)$ that equal to $\dfrac{p(1-p)}{f(Q(p))}$.

For instance, Parzen (1979), Padgett (1986), Sheather and Marron (1990), and Ralescu and Sun (1993) considered Gaussian kernels. But all those estimators have a large bias when p is close to 1. In order to correct this bias, Harrell and Davis (1982) or Park (2006) suggest to use asymmetric kernel, namely the Beta-type kernel that is the following

$$HD_n(p) = \frac{\Gamma(n+1)}{\Gamma((n+1)p)\,\Gamma((n+1)(1-p))}\int_0^1 F_n^{-1}(y)\, y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}\,dy,$$

where $F_n^{-1}(x)$ is the inverse of the empirical distribution function that is defined by

$$F_n^{-1}(y) = \begin{cases} X_{(i)} & \text{if } (i-1)/n < y \leq i/n \\ X_{(n)} & \text{if } 1-1/n < y < 1. \end{cases}$$

The $HD_n(p)$ estimator can be expressed as an L-estimator

$$HD_n(p) = \sum_{i=1}^{n} w_{n,i}(p)\, X_{(i)},$$

where

$$w_{n,i}(p) = \frac{\Gamma(n+1)}{\Gamma((n+1)p)\,\Gamma((n+1)(1-p))} \int_{(i-1)/n}^{i/n} y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}\,dy.$$

Notice that the expected value of the kth order statistic is given by

$$E\left(X_{(k)}\right) = \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)} \int_0^1 Q(y)\,y^{k-1}(1-y)^{n-k}\,dy,$$

where $\Gamma(.)$ is the gamma function that is defined by

$$\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx, \quad k > 0.$$

Observing that

$$E\left(X_{((n+1)p)}\right) \to Q(p) \quad \text{as } n \to \infty \text{ for } p \in \left]0,1\right[,$$

see (David, 1981).

Asymptotic behavior of HD estimator

For an absolutely continuous distribution function F, with a strictly positive density function f, we can follow van der Vaart and Weller (1996) to get

$$\sqrt{n}\left(Q_n(p) - Q(p)\right) \overset{\mathrm{D}}{\to} \frac{B(F(Q(p))}{f(Q(p))} \overset{\mathrm{D}}{=} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(Q(p))}\right), \quad \text{as } n \to \infty,$$

where B is a standard Brownian bridge.

We give now the theorem of the central limit for the HD estimator

Theorem 2.3.1 (Harrel and Davis (1982), Zelterman (1990)) *Let F be an absolutely continuous distribution function with a strictly positive continuous density function f, such that*

$$\int_{\mathbb{R}} |x^\alpha|\, f(x)\,dx < \infty \quad \textit{for some } \alpha > 0.$$

The HD estimator satisfies the same central limit theorem as does Q_n:

$$\sqrt{n}\left(HD_n\left(p\right)-Q\left(p\right)\right)\xrightarrow{\text{D}}\mathcal{N}\left(0,\frac{p\left(1-p\right)}{f^2\left(Q\left(p\right)\right)}\right),\quad \textit{as } n\to\infty \textit{ for } p\in\left]0,1\right[.$$

Kaigh and Lachenbruch (1982) proposed an L-estimator for $Q\left(p\right)$ that may be written as

$$KL\left(p\right)=\sum_{i=r}^{n+r-m}\frac{C_{r-1}^{i-1}C_{m-r}^{m-i}}{C_k^n}X_{(i)},$$

where $r=\left[p\left(m+1\right)\right]$, and m is an integer valued parameter satisfying $1\leq m\leq n$. The $KL\left(p\right)$ for general values of m are explored by Kaigh and Driscoll (1987), Kaigh and Lachenbruch (1982) remarked that the weight $\frac{C_{r-1}^{i-1}C_{m-r}^{n-i}}{C_m^n}$ correspond to the mass function of the negative hypergeometric distribution.

For more details concerning the kernel quantile estimators,see (Sheater and Marron, 1990).

2.3.3 Asymptotic properties

Theorem 2.3.2 *Suppose that $f^{'}$ is bounded and continuous in a neighborhood of $Q\left(p\right)$,with $Q\left(p\right)\neq 0$ and the kernel function k is a continuous bounded density, symmetric about zero,and satisfies*

$$\mu_2\left(k\right)=\int_{-\infty}^{\infty}t^2k\left(t\right)dt<\infty.$$

Then for all $p\in\left]0,1\right[$, the mean squared error of $\tilde{Q}_n(p)$ is

$$\begin{aligned} MSE\left(\tilde{Q}_n(p)\right) &= \frac{p\left(1-p\right)}{nf^2\left(Q\left(p\right)\right)}+\frac{h^4\left(f^{'}\left(Q\left(p\right)\right)\right)^2}{4f^2\left(Q\left(p\right)\right)} \\ &- \frac{h}{nf\left(Q\left(p\right)\right)}\varphi\left(k\right)+o(\frac{h}{n}+h^4), \end{aligned}$$

where $\varphi\left(k\right)=2\int yk\left(y\right)K\left(y\right)dy$.

Corollairy 2.3.1 *The optimal bandwidth for $AMSE\left(\tilde{Q}_n(p)\right)$ is*

$$\tilde{h}_{opt}=\left(\frac{f\left(Q\left(p\right)\right)\varphi\left(k\right)}{n\left(f^{'}\left(Q\left(p\right)\right)\right)^2\mu_2^2\left(k\right)}\right)^{1/3}$$

and the asymptotic mean squared error associated for this $\tilde{h}_{opt}$ is:

$$AMSE_{opt}\left(\tilde{Q}_n(p)\right) = n^{-1}\left[\frac{p\left(1-p\right)}{f^2\left(Q\left(p\right)\right)} - \frac{3}{4}\left(\frac{\varphi^4\left(k\right)}{nf^2\left(Q\left(p\right)\right)\left(f'\left(Q\left(p\right)\right)\right)^2\mu_2^2\left(k\right)}\right)^{1/3}\right].$$

Theorem 2.3.3 *Suppose that $Q''(p)$ is continuous in a neighborhood of p and that k is a compactly supported density, symmetric about zero. Then the mean squared error of $\hat{Q}_n(p)$ as follows if F is not symmetric or F is symmetric but $p \neq 1/2$*

$$\begin{aligned} MSE\left(\hat{Q}_n\left(p\right)\right) &= \frac{p\left(1-p\right)}{n}\left(Q'\left(p\right)\right)^2 + \frac{h^4}{4}\left(Q''\left(p\right)\right)^2\mu_2^2\left(k\right) \\ &- \frac{h}{n}\left(Q'\left(p\right)\right)^2\varphi\left(k\right) + o\left(n^{-1}h + h^4\right) \end{aligned}$$

When F is symmetric and $p = 1/2$ then

$$\begin{aligned} MSE\left(\hat{Q}_n\left(p\right)\right) &= n^{-1}\left(Q'\left(1/2\right)\right)^2\left[0.25 - 0.5h\varphi\left(k\right) + \left(nh\right)^{-1}\rho\left(k\right)\right] \\ &+ o\left(n^{-1}h + \left(nh\right)^{-2}\right) \end{aligned}$$

where $\rho(k) = \int k^2(x)\,dx$.

Proof. *See (Sheater and Marron, 1990).*

Corollairy 2.3.2 *If F is not symmetric or F is symmetric but $p \neq 1/2$, the expression for the asymptotic mean squared error of $\hat{Q}_n(p)$ is*

$$AMSE\left(\hat{Q}_n\left(p\right)\right) = \frac{p\left(1-p\right)}{n}\left(Q'\left(p\right)\right)^2 + \frac{h^4}{4}\left(Q''\left(p\right)\right)^2\mu_2^2\left(k\right) - \frac{h}{n}\left(Q'\left(p\right)\right)^2\varphi\left(k\right).$$

The optimal bandwidth for $AMSE\left(\hat{Q}_n(p)\right)$ is

$$\hat{h}_{opt} = \left(\frac{\left(Q'\left(p\right)\right)^2\varphi\left(k\right)}{n\left(Q''\left(p\right)\right)^2\mu_2^2\left(k\right)}\right)^{1/3},$$

and the asymptotic mean squared error associated for this $\hat{h}_{opt}$ is:

$$AMSE_{opt}\left(\hat{Q}_n\left(p\right)\right) = n^{-1}\left[p\left(1-p\right)\left(Q'\left(p\right)\right)^2 - \frac{3}{4}\left(\frac{\left(Q'\left(p\right)\right)^8\varphi^4\left(k\right)}{n\left(Q''\left(p\right)\right)^2\mu_2^2\left(k\right)}\right)^{1/3}\right]$$

$$AMSE_{opt}\left(\hat{Q}_n(p)\right) = n^{-1}p(1-p)\left(Q'(p)\right)^2 + O\left(n^{-4/3}\right).$$

If F is symmetric and $p = 1/2$ then

$$AMSE\left(\hat{Q}_n(p)\right) = n^{-1}\left(Q'(1/2)\right)^2\left[0.25 - 0.5h\varphi(k) + (nh)^{-1}\rho(k)\right].$$

■

Remark 2.3.1 *In this case, there is no single optimal bandwidth minimizing the $AMSE\left(\hat{Q}_n(p)\right)$.*

Theorem 2.3.4 *Suppose that:*

1) $F(x)$ has a p.d.f $f(x)$ witch is continuous and positive in some neighborhood of $Q(p)$.

2) $f'(x)$ exists and is continuous at $Q(p)$.

3) The kernel $k(x)$ is a p.d.f symmetric about zero with finite support.

4) $\lim_{n\to\infty} n^{1/4}h \to 0$.

Then

$$n^{1/2}\left(\hat{Q}_n(p) - Q(p)\right) = -n^{-1/2}\left(F_n(Q(p)) - p\right)/f(Q(p)) + o_p(1),$$

where $o_p(1)$ converge to zero in probability as $n\to\infty$ and F_n is the empirical distribution function.

Proof. *See (Yang, 1985)* ■

Asymptotic normality of $\hat{Q}_n(p)$

Using the precedent theorem and the multivariate central limit theorem, we have the following corollary

Corollairy 2.3.3 *Let $0 < p_1 < ... < p_m < 1$. Then the asymptotic joint distribution of*

$$n^{1/2}\left(\hat{Q}_n(p_1) - Q(p_1), ..., \left(\hat{Q}_n(p_m) - Q(p_m)\right)\right),$$

is m-dimensional normal with a zero mean vector and a covariance matrix with element

$$p_i(1-p_j)/f(Q(p_i))f(Q(p_j)) \ \ (i,j = 1,...,m).$$

Choice of the bandwidth

We interest for the choice of the smoothing parameter h of $\hat{Q}_n(p)$, for all p, apart if F is symmetric and $p = 1/2$.

Definition 2.3.3 *A kernel is said to be of order m for some $m \geq 2$ if*

$$\int t^j k(t)\,dt = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j = 1, ..., m-1 \\ \mu_m & \text{if } j = m \end{cases}$$

where $\mu_m \neq 0$.

We see from Corollary 2.3.2 that for a given choice of k, the asymptotically optimal value of h depends on the first and second derivatives of the quantile function $Q(p)$. Thus estimates of $Q'(p)$ and $Q''(p)$ are necessary for the choice of h. If the first and second derivatives of k exist, then we can estimate these quantities by the first and second derivatives of $\hat{Q}_n(p)$. This results in the estimator

$$\hat{Q}'(p) = \sum_{i=1}^{n}\left[\int_{(i-1)/n}^{i/n} a^{-2}k'\left(a^{-1}(x-p)\right)dx\right] X_{(i)},$$

and

$$\hat{Q}''(p) = \sum_{i=1}^{n}\left[\int_{(i-1)/n}^{i/n} b^{-3}k''\left(b^{-1}(x-p)\right)dx\right] X_{(i)}$$

where k is a kernel of order m.

The resulting estimate of the asymptotically optimal bandwidth h is given by

$$\hat{h}_{opt} = \left(\frac{\left(\hat{Q}'(p)\right)^2 \varphi(k)}{n\left(\hat{Q}''(p)\right)^2 \mu_2^2(k)}\right)^{1/3}.$$

The problem is then to choose values for the bandwidth a and b that results in an asymptotically efficient $\left(\hat{Q}'(p)/\hat{Q}''(p)\right)^{2/3}$.

Theorem 2.3.5 *Suppose that $Q^{(m+2)}$ is continuous in a neighborhood of p and that k is a compactly supported kernel of order m, symmetric about zero. Then the asymptotically optimal band-*

width for $\hat{Q}'(p)$ is given by

$$a_{opt} = \left(\frac{Q'(p)}{Q^{(m+1)}(p)}\right)^{2/m+2} \left[\frac{(m!)^2 \int k^2(t)\,dt}{2m\left(\int t^m k(t)\,dt\right)^2}\right]^{1/2m+1} n^{-1/(2m+1)}.$$

and the asymptotically optimal bandwidth for $\hat{Q}''(p)$ is given by

$$b_{opt} = \left(\frac{Q'(p)}{Q^{(m+2)}(p)}\right)^{2/m+3} \left[\frac{3(m!)^2 \int k'^2(t)\,dt}{2m\left(\int t^m k(t)\,dt\right)^2}\right]^{1/2m+3} n^{-1/(2m+3)}.$$

2.4 Quantile density function estimation

Let P be a probability measure on the real line with distribution function F. The estimation of the p-quantile $Q(p) = F^{-1}(p)$ of P is closely related to the quantile density function $\left(F^{-1}\right)'(p)$, since the asymptotic variance of a nonparametric estimator of $Q(p)$ is usually given by

$$\sigma^2 = \frac{p(1-p)}{f^2(Q(p))} = p(1-p)\left(F^{-1}\right)'^2(p).$$

For the kernel quantile estimator

$$\hat{Q}_n(p) = \sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x-p}{h}\right) dx = \int_0^1 F_n^{-1}(x) \frac{1}{h} k\left(\frac{p-x}{h}\right) dx,$$

if we want to construct, for example, confidence intervals of asymptotic level $1-\alpha$ for the underlying p-quantile, we are usually concerned with the problem of estimating $Q'(p)$.

We define a histogram type estimator that has the form

$$H_n(p) := \frac{F_n^{-1}(p+h) - F_n^{-1}(p-h)}{2h}, \quad h > 0,$$

this histogram was suggested by Siddiki (1960) and investigated by Bloch and Gastwirth (1968), and Bofinger (1975). For a brief discussion of their results and an asymptotic expansion of the distribution of $H_n(p)$ see (Reiss, 1978).

2.4.1 Asymptotic properties

We begin by the asymptotic normality of the histogram type estimator $H_n(p)$.

Theorem 2.4.1 *Suppose that F^{-1} is twice differentiable near p with bounded second derivative. Then if $0 < h \underset{n\to\infty}{\to} 0,\ nh \underset{n\to\infty}{\to} \infty$ we have*

$$(2nh)^{1/2}\left(H_n(p) - \left[Q(p+h) - Q(p-h)\right]/2h\right) \overset{\mathrm{D}}{\to} \mathcal{N}\left(0, Q'^2(p)\right)$$

Proof. *See (Falk, 1986).* ■

Theorem 2.4.2 *we assume that F^{-1} is three times derivable near p and that the third derivative is bounded and continuous. Then if $h \underset{n\to\infty}{\to} 0,\ nh \underset{n\to\infty}{\to} \infty$, we have*

$$E\left(\left(H_n(p) - Q'(p)\right)^2\right) \underset{n\to\infty}{\to} 0.$$

Proof. *We have*

$$E(H_n(p)) = \zeta_n(p)/2h \underset{n\to\infty}{\to} Q'(p),$$

where

$$\zeta_n(p) = (Q(p+h) - Q(p-h))/2h,$$

then

$$\begin{aligned} E\left(H_n(p) - Q'(p)\right)^2 &= E\left(H_n(p) + \zeta_n(p) - \zeta_n(p) - Q'(p)\right)^2 \\ &= E(H_n(p) - \zeta_n(p))^2 + \left(\zeta_n(p) - Q'(p)\right)^2, \end{aligned}$$

from the Taylor formula and the theorem 2-4-1

$$\zeta_n(p) - Q'(p) = \frac{h^2}{6}Q^{(3)}(p) + o\left(h^2\right),$$

and

$$E\left(H_n(p) - Q'(p)\right)^2 \sim Q'^2(p)/2nh + \left(\frac{h^2}{6}Q^{(3)}(p)\right)^2.$$

■

Corollairy 2.4.1 *Under the assumptions of the theorem 2-4-2, and if $Q^{'}(p)\,Q^{(3)}(p)\neq 0$. Then the optimal bandwidth for $AMSE\left(H_{n}(p)\right)$ is*

$$h^{*}=\left(\frac{3Q^{'}(p)}{\sqrt{2}Q^{(3)}(p)}\right)^{2/5}n^{-1/5},$$

and the asymptotic mean squared error associated for this h^{*} is:

$$AMSE_{opt}\left(H_{n}(p)\right)=5\left(Q^{(3)}(p)/6\right)^{2/5}\left(Q^{'2}(p)/8\right)^{4/5}n^{-4/5}.$$

Another estimator of $Q^{'}(p)$ is suggested by means of the corresponding kernel quantile estimator $\hat{Q}_{n}(p)$ as follows.

Falk (1985) proved that under appropriate conditions on $F,\ k$ and h

$$\hat{Q}_{n}(p)\overset{P}{\rightarrow}Q(p)\quad\text{as } n\rightarrow\infty.$$

Consequently, one might expect that

$$\hat{Q}_{n}^{'}(p)=\int_{0}^{1}F_{n}^{-1}(x)\,\frac{1}{h^{2}}k^{'}\left(\frac{p-x}{h}\right)dx\overset{P}{\rightarrow}Q^{'}(p)\quad\text{as } n\rightarrow\infty.$$

From this idea we can define a kernel estimator of $Q^{'}(p)$ by

$$A_{n}(p)=\int_{0}^{1}\frac{1}{h^{2}}F_{n}^{-1}(x)\,l\left(\frac{p-x}{h}\right)dx,$$

where $l:\mathbb{R}\rightarrow\mathbb{R}$ is a kernel function has bounded support and verifying that

$$\int x^{i}l(x)=\begin{cases}0 & \text{if } i=0,2\\ -1 & \text{if } i=1.\end{cases}\tag{2.1}$$

Related kernel estimators of the quantile density were proposed by Parzen (1979) and by Csörgõ (1983). Moreover observe that $A_{n}(p)$ is a linear combination of order statistics $\sum_{i=1}^{n}a_{in}X_{(i)}$, where a_{in} are given by

$$a_{in}=\int_{\frac{i-1}{n}}^{\frac{i}{n}}\frac{1}{h^{2}}l\left(\frac{p-x}{h}\right)dx.$$

Theorem 2.4.3 *Let $0 < p < 1$ and suppose that F^{-1} is twice derivable near p with bounded second derivative. Then if l has a bounded support and verifying 2.1 and $h \underset{n\to\infty}{\longrightarrow} 0$, $nh^2 \underset{n\to\infty}{\longrightarrow} \infty$ we have*

$$(nh)^{1/2} (A_n (p) - \upsilon_n) \xrightarrow{\mathrm{D}} \mathcal{N} \left(0, \left(F^{-1}\right)^{\prime 2} (p) \int L^2 (y) \, dy\right),$$

where

$$L (y) = \int_{-\infty}^{y} l (t) \, dt,$$

and

$$\upsilon_n = \int_0^1 \frac{1}{h^2} F^{-1} (x) \, l \left(\frac{p - x}{h}\right) dx.$$

Proof. *See (Falk, 1986).* ■

If we assume that $nh^3 \underset{n\to\infty}{\longrightarrow} 0$, we can replace υ_n in precedent theorem by$\left(F^{-1}\right)' (p)$, and if we suppose that the third derivative of F^{-1} is continuous at p, then if the kernel l verifying 2.1, the mean squared error of $A_n (p)$ is given by

$$\begin{aligned} MSE (A_n (p)) &= E \left(\left[A_n (p) - \left(F^{-1}\right)' (p)\right]^2\right) \\ &= \frac{1}{nh} Q^{\prime 2} (p) \int L^2 (y) \, dy + \left(\frac{h^2}{3!} Q^{(3)} (p) \int y^3 l (y) \, dy\right)^2. \end{aligned}$$

The optimal bandwidth which minimize $E \left(\left[A_n (p) - \left(F^{-1}\right)' (p)\right]^2\right)$ is

$$h^{**} = \left[\frac{3! Q' (p) \left(\int L^2 (y) \, dy\right)^{1/2}}{2 Q^{(3)} (p) \int y^3 l (y) \, dy}\right]^{2/5} n^{-1/5},$$

and the asymptotic mean squared error associated for this h^{**} is

$$AMSE_{opt} (A_n (p)) = \frac{5}{4} \left(Q^{\prime 2} (p) \int L^2 (y) \, dy\right)^{4/5} \left(\frac{1}{3} Q^{(3)} (p) \int y^3 l (y) \, dy\right)^{2/5} n^{-4/5}.$$

Take for example $l(x) = -\xi_2^{'}(x)/2$, where ξ denotes the Legendre polynomial of degree 2 on $[-1,1]$, then

$$l(x) = \frac{-1}{16}\left(\left(x^2-1\right)^2\right)^{(3)} = -3x/2.$$

Then we have $\int l(x)\,dx = 0$, $\int xl(x)\,dx = -1$ and

$$\int L^2(y)\,dy = 3/5.$$

The $A_n(p)$ is a linear combination of order statistics $\sum_{i=1}^{n} a_{in}X_{(i)}$, where a_{in} are given by

$$a_{in} = \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h^2} l\left(\frac{p-x}{h}\right) dx.$$

We have

$$\begin{aligned} a_{in} &\sim (nh^2)^{-1} l\left(\frac{p-i/n}{h}\right) \\ &\sim \frac{-3}{2nh^2}\left(\frac{p-i/n}{h}\right). \end{aligned}$$

From $-1 \leq \frac{p-i/n}{h} \leq 1$, then $n(p-h) \leq i \leq n(p+h)$,

In addition, we have $\int_{-1}^{0} l(x)\,dx = \int_{0}^{1} l(x)\,dx$. Therefore,

$$A_n(p) \sim \frac{3}{nh^3} \sum_{i=[n(p-h)]}^{[n(p+h)]} (i/n - p)\, X_{(i)}.$$

Chapter 3

Transformation in kernel density estimation

The kernel density estimator

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

has the disadvantage that h is not locally adjusted and the performance of the kernel density estimator deteriorates when f becomes less smooth or heavy tailed.

We can alleviate the problem by estimating the density of transformed random variable, that is based on transforming the data so that the density of the transformed variable has a symmetric shape, so that it can easily be estimated using a classical kernel estimation approach, and then taking the inverse transform. The transformation method proposed initially by Wand et al. (1991), is very suitable for asymmetrical variables. In the specialized literature several transformation kernel estimators have been proposed, and their main difference is the type of transformation family that they use.

Concerned the kernel estimation for heavy tailed distributions has been studied by several authors Bolancé et al. (2003), Clements et al. (2003) and Buch-Larsen et al. (2005) propose different parametric transformation families that they all make the transformed distribution more symmetric that the original one, which in many applications has usually a strong right-hand asymmetry. Buch-Larsen et al. (2005) propose an alternative transformation such as one based on the Champernowne distribution, who they have shown in simulation studies that this transformation is preferable to other transformation density estimation approach for heavy tailed distribution.

The transform kernel density estimator (see Devroye et al. 1983) is based upon a transformation $T : \mathbb{R} \rightarrow [0, 1]$ which is strictly monotonically increasing, continuously differentiable, and which has a continuous differentiable inverse. The transformed data sequence is $Y_1, ...Y_n$ where $Y_i = T(X_i)$.

Note that the transformed variable has density

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

The density g is estimated by the classical kernel density estimator

$$\tilde{g}_n(y) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{y - Y_i}{h}\right),$$

and f is estimated by

$$\tilde{f}_n(x) = \tilde{g}_n(T(x))\, T'(x) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{T(x) - T(X_i)}{h}\right) T'(x).$$

Buch-Larsen et al. (2005) introduced an alternative large loss estimation approach based on nonparametric statistics. They recommended an estimator based on the classical kernel density estimator

$$\tilde{f}_n(x) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right),$$

where $X_1, ..., X_n$ is the data set, whose density we want to estimate, and k is a kernel function and h is a bandwidth. They showed that, when introducing a tail flattening transformation, inspired by the work of Wand et al. (1991), the Champernowne c.d.f with maximum likelihood estimated parameters, this estimator has promising tail performance at the same time as being an estimator on the entire axis. When the transformation function is an estimated cumulative distribution function, this estimator corresponds to a poorly parametric estimated distribution with a non-parametric correction, as described in Buch-Larsen et al. (2005). The resulting transformation kernel density estimator has the form

$$\tilde{f}_n(x) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{T(x) - T(X_i)}{h}\right) T'(x).$$

where $T(x)$ is the transformation function.

When the transformation function returns values on a compact interval, if this is a c.d.f, it is necessary to have a boundary correction to ensure that the transformation kernel density estimator is a consistent estimator at the boundary. We use a simple renormalization method, as described in Jones (1993) which ensures that each kernel function integrates to 1. With the notation from Chen (1999) the transformation kernel density estimator with the renormalizing boundary correction is

$$\tilde{f}_n(x) = \frac{1}{n a_{01}(T(x), h)} \sum_{i=1}^{n} k\left(\frac{T(x) - T(X_i)}{h}\right) T'(x),$$

where

$$a_{sm}(T(x), h) = \begin{cases} \int_{-1}^{T(x)/h} y^s k^m(y)\, dy & \text{if} \quad 0 \le T(x) \le 1 - h \\ \int_{-(1-T(x))/h}^{1} y^s k^m(y)\, dy & \text{if} \quad 1 - h \le T(x) \le 1. \end{cases}$$

When the transformation function $T(x)$ is a c.d.f of a parametric distribution estimated to the data set under investigation, then the kernel density approach can be interpreted as a nonparametric correction to this estimated parametric distribution.

3.1 Asymptotic theory for the transformation kernel density estimator

Now we investigate the asymptotic theory of the transformation kernel density estimator in general, and we derive its asymptotic bias and variance.

Theorem 3.1.1 *Let $X_1, ..., X_n$ be independent identically distributed variables with density f . Let $\tilde{f}_n(x)$ be the transformation kernel density estimator of $f(x)$*

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

where $T(\cdot)$ is the transformation function.
Then the bias and variance of $f_n(x)$ is given by

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

and

$$Var\left(\tilde{f}_n(x)\right) = \frac{1}{nh}\rho(k)\, T'(x)\, f(x) + o\left(\frac{1}{nh}\right).$$

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

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

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

$$\tilde{g}_n(y) = (nh)^{-1} \sum_{i=1}^{n} k\left(\frac{y - Y_i}{h}\right).$$

The the mean and variance of the classical kernel density estimator $\tilde{g}_n(y)$

$$bias\left(\tilde{g}_n(y)\right) = \frac{h_n^2}{2} g''(y) \int t^2 k(t)\, dt + o\left(h^2\right),$$

and

$$\begin{aligned} Var\left(\tilde{g}_n(y)\right) &= (nh_n)^{-1} \left(g(y) \int k^2(t)\, dt + o(1)\right) \\ &= (nh_n)^{-1} g(y) \int k^2(t)\, dt + o\left((nh_n)^{-1}\right). \end{aligned}$$

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

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

then

$$E\left(\widetilde{f}_n(x)\right) = T'(x)\, E(\tilde{g}_n(T(x)) = T'(x)\left[g(T(x)) + \frac{h^2}{2} g''(y) \int t^2 k(t)\, dt + o(h^2)\right],$$

we have

$$g(T(x)) = \frac{f(x)}{T'(x)}$$

$$g'(T(x)) = \frac{dg(T(x))}{dT(x)} = \frac{dg(T(x))}{dx} \cdot \frac{dx}{dT(x)} = \left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'(x)},$$

and

$$g''(T(x)) = \frac{d}{dT(x)}\left(g'(T(x))\right) = \frac{d}{dx}\left(g'(T(x))\right) \frac{dx}{dT(x)} = \left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'(x)}\right)' \frac{1}{T'(x)}$$

$$E\left(\widetilde{f}_n(x)\right) = f(x) + \frac{h^2}{2}\left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'(x)}\right)' \int t^2 k(t)\, dt + o\left(h^2\right),$$

and

$$Var\left(\widetilde{f}_n(x)\right) = (T'(x))^2\, Var(\tilde{g}_n(T(x)) = (T'(x))^2 \left[(nh)^{-1} g(T(x)) \int k^2(t)\, dt + o\left((nh)^{-1}\right)\right] = (nh)^{-1} T'(x)\, \rho(k)\, f(x) + o\left((nh)^{-1}\right).$$

■

3.1.1 Mean Squared Error

Now we derive its asymptotic mean squared error, the optimal bandwidth h^* and the asymptotic mean squared error associated for this h^*. We call the mean squared error at point x the quantity:

$$MSE\left(\widetilde{f}_n(x)\right) = \left[\frac{h^2}{2}\left(\left(\frac{f(x)}{T'(x)}\right)' \frac{1}{T'(x)}\right)' \mu_2(k) + o(h^2)\right]^2 + (nh)^{-1} T'(x)\, \rho(k)\, f(x) + o\left((nh)^{-1}\right),$$

and the asymptotic squared error in point x the quantity :

$$AMSE\left(\widetilde{f}_n(x)\right) = \left[\frac{h^2}{2}\left(\left(\frac{f(x)}{T'(x)}\right)'\frac{1}{T'(x)}\right)'\mu_2(k)\right]^2 + (nh)^{-1}T'(x)\rho(k)f(x).$$

Seek a value h^* that minimizes the asymptotic mean squared error $AMSE\left(\widetilde{f}_n(x)\right)$.

If the density $f(x)$ is twice differentiable on $\mathbb{R}$, the second derivative $f''(x)$ is absolutely continuous on $\mathbb{R}$, and k is a symmetric kernel, that as $\int t^2K(t)\,dt < \infty$, and if $f(x)f''(x) \neq 0$, the value h that minimizes $AMSE\left(\widetilde{f}_n(x)\right)$ is

$$h^* = \left(T'(x)\rho(k)f(x)\left[\left(\left(\frac{f(x)}{T'(x)}\right)'\frac{1}{T'(x)}\right)'\mu_2(k)\right]^{-2}\right)^{1/5} n^{-1/5},$$

and the associated mean squared error is given by :

$$\begin{aligned} AMSE^*\left(\widetilde{f}_n(x)\right) &= 5\left(\frac{1}{4n}T'(x)\rho(k)f(x)\right)^{4/5} \\ &\times \left(\left[\frac{1}{2}\left(\left(\frac{f(x)}{T'(x)}\right)'\frac{1}{T'(x)}\right)'\mu_2(k)\right]^2\right)^{\frac{1}{5}} \\ &= \frac{5}{4}\left(\left[\left(\left(\frac{f(x)}{T'(x)}\right)'\frac{1}{T'(x)}\right)'\mu_2(k)\right]\right)^{2/5} \\ &\times \left(T'(x)\rho(k)f(x)\right)^{4/5} n^{-4/5}. \end{aligned}$$

3.1.2 Mean Integrated Squared Error

An error distance between the estimated density $\widetilde{f}_n(x)$ and the theoretical density f that has widely been used in analysis of the optimal bandwidth h is the mean integrated squared error

$$MISE\left(\widetilde{f}_n(x)\right) = \int E\left(\tilde{f}_n(x) - f(x)\right)^2 dx.$$

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

The asymptotic mean integrated squared error is

$$AMISE\left(\tilde{f}_n(x)\right) = (nh)^{-1}\rho(k) + \frac{h^4}{4}\mu_2^2(K)\int f''^2(x)\,dx,$$

and the optimal bandwidth minimizes the asymptotic mean squared error

$$\tilde{h} = \left[\frac{\rho(k)}{\mu_2^2(k)\int (f''(x))^2\,dx.}\right]^{1/5} n^{-1/5},$$

but the bandwidth h can be chosen to make $AMISE\left(\tilde{f}_n(x)\right)$ minimal.

We remark that the bandwidth $\tilde{h}$ is inversely proportional to the roughness $\rho\left(f''\right) = \int \left(f''(x)\right)^2 dx$ of $f(x)$.

Let us consider the beta density function $B(\alpha,\beta)$ defined on the interval $[\delta, \delta + a]$ parametrized as follows Johnson et al. (1995)

$$\frac{(x-\delta)^{\alpha-1}(\delta+a-x)^{\beta-1}}{a^{\alpha+\beta-1}B(\alpha,\beta)}, \quad -1 \leq x \leq 1,$$

with $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, and $\Gamma(.)$ is the Euler Gamma function.

Terrel and Scott (1985) analyzed several density families that minimizes the measure of smoothness $\rho\left(f''\right)$. Terrel (1990) showed that the beta distribution defined on $[-1,1]$ minimizes $\rho\left(f''\right)$ within the set of all densities with a given known variance, and showed that $B(3,3)$ defined on $]-1/2, 1/2[$ minimizes $\rho\left(f''\right)$ within the set of beta density with same support.

Based on the work by Buch-Larsen et al. (2005), Bolancé et al (2008) proposed a double transformation with the purpose of obtaining a transformed variable whose density is as close as possible to a density that maximizes smoothness $\rho\left(f''\right)$ and at the same time that minimizes the asymptotic mean integrated squared error $AMISE$ of the classical kernel estimator and obtained with the transformed observations.

Bolancé (2010) used a double transformation Champernowne-inverse beta in kernel density estimator for heavy-tailed distributions, and calculated the asymptotically optimal bandwidth parameter when minimizing the expression of the asymptotic mean integrated squared error of the transformed variable.

3.1.3 Champernowne distribution

In Buch-Larsen et al. (2005) the Champernowne distribution is proposed as transformation function. The Champernowne c.d.f is a heavy tailed, quite flexible three- parameter distribution and has the form

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

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

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

The Champernowne distribution is a heavy tailed distribution converging to the Pareto distribution

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

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

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

The choice of M as the empirical median gives a stable estimator, especially for heavy-tailed distributions, and the maximum likelihood estimates of (α, c) ensures the best over-all fit of the distribution.

Remark 3.1.1 *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 cdf becomes larger for increasing c, and conversely, when $\alpha > 1$, the derivative of the c.d.f becomes smaller for increasing c. When $\alpha = 1$, 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

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

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.*

Chapter 4

Champernowne transformation in kernel quantile estimation for heavy-tailed distributions

Abstract[1]. By transforming a data set with a modification of the Champernowne distribution function, a kernel quantile estimator for heavy-tailed distributions is given. The asymptotic mean squared error (AMSE) of the proposed estimator and related asymptotically optimal bandwidth are evaluated. Some simulations are drawn to show the performance of the obtained results.

Keywords: Bandwidth; Champernowne distribution; Heavy tails; Kernel estimator; Quantile function.

4.1 Introduction

The estimation of population quantiles is of great interest when a parametric form for the underlying distribution is not available. It plays an important role in both statistical and probabilistic applications, namely: the goodness-of-fit, the computation of extreme quantiles and Value-at-Risk in insurance business and financial risk management. Also, a large class of actuarial risk measures can be defined as functional of quantiles (see, Denuit *et al.* 2005).

Quantile estimation has been intensively used in many fields, see Azzalini (1981), Harrel and

[1] This chapter is a paper appeared in Journal Afrika Statistika. Vol. 5, N° 12, 2010, page 288–296. (authers : Sayah A., Yahia D., Necir A.)

Davis (1982), Sheather and Marron (1990), Ralescu and Sun (1993), and Chen and Tang (2005). Most of the existing estimators suffer from either a bias or an inefficiency for high probability levels. To solve this inconvenience, we suggest to use the so-called transformed kernel estimate, firstly used in the density estimation context, by Devroye and Györfi (1985) for heavy-tailed observations. The idea is to transform the initial observations $\{X_1, ..., X_n\}$ into a sample $\{Z_1, ..., Z_n\} := \{T(X_1), ..., T(X_n)\}$, where T is a given function having values in $(0, 1)$. Buch-Larsen *et al.* (2005) suggested to choose T so that $T(X)$ is close to the uniform distribution. They proposed a kernel density estimation of heavy-tailed distributions based on a transformation of the original data set with a modification of the Champernowne cumulative distribution function (c.d.f) (see, Champernowne, 1936 and 1952). While Bolancé *et al.* (2008) proposed the Champernowne-inverse beta transformation in kernel density estimation to model insurance claims and showed that their method is preferable to other transformation density estimation approaches for distributions that are Pareto-like.

Recently, in order to correct the bias problems, Charpentier and Oulidi (2010) suggested several nonparametric quantile estimators based on the beta-kernel and applied them to transformed data. For nonparametric estimation, the bandwidth controls the balance between two considerations: bias and variance. Furthermore, the mean squared error (MSE) which is the sum of squared bias and variance, provides a composite measure of performance. Therefore, optimally in the sense of MSE is not seriously swayed by the choice of the kernel but is affected by that of the bandwidth (for more details, see Wand and Jones, 1995). In this paper, we propose a new estimator of the quantile function, based on the modified Champernowne transformation and we obtain an expression for the value of the smoothing parameter that minimizes the AMSE of the obtained estimator. The use of this transformation in kernel estimation of quantile functions for heavy-tailed distributions improves the already existing results.

The rest of the paper is organized as follows. In Section 2, the kernel quantile estimation is given. Section 3 is devoted to the Champernowne transformation and the estimation procedure. In Section 4, we propose an asymptotically optimal bandwidth selection. A simulation study is carried out in Section 5. Finally we outline some concluding remarks in Section 6.

4.2 Kernel quantile estimation

Let $X_1, X_2, ...,$ be independent and identically distributed (i.i.d) random variables (rv's) drawn from an absolutely continuous (c.d.f) F with probability density function (p.d.f) f. For each integer n, let $X_{1,n} \leq ... \leq X_{n,n}$ denote the order statistics pertaining to the sample $X_1, ..., X_n$. We define the pth quantile $Q_X(p)$ as the left-continuous inverse of F as

$$Q_X(p) := \inf \{x \in \mathbb{R} : F(x) \geq p\}, \quad 0 < p < 1.$$

A basic estimator of $Q_X(p)$, is the sample quantile $Q_n(p) = X_{[np]+1,n}$ where $[x]$ denotes the integer part of $x \in \mathbb{R}$. Suppose that K is a p.d.f symmetric about 0 and $h := h_n$ is a sequence of real numbers (called bandwidth) such that $h \to 0$ as $n \to \infty$. The classical kernel quantile estimator (CKQE) was introduced by Parzen (1979) in the following form:

$$\tilde{Q}_{n,X}(p) := \sum_{i=1}^{n} X_{i,n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} K_h(x-p)\,dx, \tag{4.1}$$

where $K_h(t) := K(t/h)/h$. Yang (1985) established the asymptotic normality and the mean squared consistency of $\tilde{Q}_{n,X}(p)$, while Falk (1984) showed that the asymptotic performance of $\tilde{Q}_{n,X}(p)$ is better than that of the empirical sample quantile. Sheather and Marron (1990) gave the AMSE of $\tilde{Q}_{n,X}(p)$. For further details on kernel-based estimation, see Silverman (1986) and Wand and Jones (1995).

4.3 Champernowne transformation and estimation procedure

In the context of quantile estimation, if T is strictly increasing, the p^{th} quantile of $T(X)$ is equal to $T(Q_X(p))$. Firstly, we use a parametric transformation T, namely the modified Champernowne c.d.f as proposed by Buch-Larsen *et al.* (2005) when fitting insurance claims:

$$T_{\alpha,M,c}(x) := \frac{(x+c)^{\alpha} - c^{\alpha}}{(x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}}, \quad x \geq 0, \tag{4.2}$$

with parameters $\alpha > 0$, $M > 0$ and $c \geq 0$. The associated p.d.f is

$$t_{\alpha,M,c}(x) := \frac{\alpha (x+c)^{\alpha-1} ((M+c)^{\alpha} - c^{\alpha})}{((x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha})^2}, \quad x \geq 0.$$

This distribution is of Pareto type, that is

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

The idea is to transform the initial data $\{X_1, ..., X_n\}$ into $\{Z_1, ..., Z_n\}$, where $Z_i := T(X_i)$, $i = 1, ..., n$. This can be assumed to have been produced by a $(0,1)$-uniform rv Z. Thus, (4.1) yields the transformed kernel quantile estimator (TKQE)

$$\hat{Q}_{n,X}(p) := T^{-1}\left(\hat{Q}_{n,Z}(p)\right),$$

where T^{-1} is the inverse of T and

$$\hat{Q}_{n,Z}(p) := \sum_{i=1}^{n} Z_{i,n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} K_h(z-p)\, dz. \tag{4.3}$$

The estimation procedure is described as follows:

1. Compute the estimates $\left(\hat{\alpha}, \hat{M}, \hat{c}\right)$ of the parameters of the modified Champernowne distribution (4.2). Notice that $T_{\alpha,M,0}(M) = 0.5$, this suggests that M can be estimated by the empirical median (see Lehmann, 1991). Then, estimate the pair (α, c) which maximizes the log-likelihood function (see, Buch-Larsen *et al.* 2005):

$$\begin{aligned} l(\alpha, c) = {} & n \log \alpha + n \log ((M+c)^{\alpha} - c^{\alpha}) + (\alpha - 1) \sum_{i=1}^{n} \log (X_i + c) \\ & - 2 \sum_{i=1}^{n} \log ((X_i + c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}). \end{aligned} \tag{4.4}$$

2. Transform the data $X_1, ..., X_n$ into $Z_1, ..., Z_n$ by

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

   The resulting transformed data belong to the interval $(0,1)$.

3. Using (4.3), calculate the kernel quantile estimator $\hat{Q}_{n,Z}(p)$ of the transformed data: $Z_1, ..., Z_n$.

4. The resulting TKQE of the original data $X_1, ..., X_n$ is given by

$$\hat{Q}_{n,X}(p) = T^{-1}_{\hat{\alpha},\hat{M},\hat{c}}\left(\hat{Q}_{n,Z}(p)\right). \tag{4.5}$$

4.4 Asymptotic theory and bandwidth selection

Let $X_1, ..., X_n$ be i.i.d rv's with c.d.f F and p.d.f f. For each p in $(0,1)$, let $\hat{Q}_{n,X}(p)$ be the TKQE (4.5) of $Q_X(p)$.

Theorem 4.4.1 *Assume that $Q_Z(\cdot)$ is two-times differentiable in a neighbourhood of $p \in (0,1)$ with continuous second derivative. Assume further that the kernel K has compact support and fulfills:*

$$\int K(t)dt = 1, \quad \int tK(t)dt = 0 \quad \text{and} \quad \int t^2K(t)dt < \infty.$$

Then the bias and the variance of $\hat{Q}_{n,X}(p)$ are respectively

$$Bias\left(\hat{Q}_{n,X}(p)\right) = \frac{h^2}{2}\left[\left(T^{-1}\right)''(Q_Z(p))\,Q_Z'^2(p) + \left(T^{-1}\right)'(Q_Z(p))\,Q_Z''(p)\right]\mu_2(K) + o\left(h^2\right),$$

and

$$Var\left(\hat{Q}_{n,X}(p)\right) = \left(\left(T^{-1}\right)'(Q_Z(p))\,Q_Z'(p)\right)^2\left(\frac{p(1-p)}{n} - \frac{h}{n}\varphi(K)\right) + o\left(\frac{h}{n}\right),$$

where $\mu_2(K) := \int t^2K(t)\,dt$, $\varphi(K) := 2\int tK(t)\left(\int_{-\infty}^{t} K(s)ds\right)dt$, Q_Z' and Q_Z'' are the first and the second derivatives of Q_Z. The value of h that minimizes the $AMSE$ of $\hat{Q}_{n,X}(p)$ is

$$h_{opt,X} := \left(\frac{\left(\left(T^{-1}\right)'(Q_Z(p))\,Q_Z'(p)\right)^2\varphi(K)}{n\Psi_{T,Q}^2(p)\,\mu_2^2(K)}\right)^{1/3}, \tag{4.6}$$

where

$$\Psi_{T,Q}(p) := \left(T^{-1}\right)''(Q_Z(p))\,Q_Z'^2(p) + \left(T^{-1}\right)'(Q_Z(p))\,Q_Z''(p).$$

The associated AMSE is

$$AMSE_{h_{opt,X}} := n^{-1}\left\{p(1-p)\left(\left(T^{-1}\right)'(Q_Z(p))\,Q_Z'(p)\right)^2 - \frac{3}{4}\left(\left(\left(T^{-1}\right)'(Q_Z(p))\,Q_Z'(p)\right)^8 \varphi^4(K)\left(n\Psi^2_{T,Q}(p)\,\mu_2^2(k)\right)^{-1}\right)^{1/3}\right\}.$$

Proof. The proof is the same as for the classical kernel quantile estimator, (see Falk, 1984 and Sheater and Marron, 1990). It suffices to replace $Q_X(p)$ by $T^{-1}(Q_Z(p))$. Suppose that Z has p.d.f g and c.d.f G. In the cases where g is not symmetric or symmetric with $p \neq 0.5$, Sheater and Marron (1990) gave the $AMSE$ of $\hat{Q}_{n,Z}(p)$:

$$AMSE\left(\hat{Q}_{n,Z}(p)\right) = \frac{p(1-p)}{n}Q_Z'^2(p) + \frac{1}{4}h^4 Q_Z''^2(p)\,\mu_2^2(K) - \frac{h}{n}Q_Z'^2(p)\,\varphi(K).$$

If $Q_Z'(p) > 0$, the asymptotically optimal bandwidth for $\hat{Q}_{n,Z}(p)$ is

$$h_{opt,Z} = \left(\frac{Q_Z'^2(p)\,\varphi(K)}{nQ_Z''^2(p)^2\,\mu_2(K)^2}\right)^{1/3}. \tag{4.7}$$

When g is symmetric and $p = 0.5$, we have

$$AMSE\left(\hat{Q}_{n,Z}(0.5)\right) = \frac{1}{n}Q_Z'^2(0.5)\left\{0.25 - 0.5h\varphi(K) + \frac{1}{nh}\int K^2(t)\,dt\right\}.$$

Remark 4.4.1 *If $Q_X'(p) > 0$, the asymptotically optimal bandwidth for the CKQE $\tilde{Q}_{n,X}(p)$ is*

$$h_{opt,C} = \left(\frac{Q_X'^2(p)\,\varphi(K)}{nQ_X''^2(p)^2\,\mu_2(K)^2}\right)^{1/3}. \tag{4.8}$$

Remark 4.4.2 *The first and the second derivatives of Q_Z are*

$$Q_Z'(p) = \frac{1}{g(Q_Z(p))} = \frac{T'(Q_X(p))}{f(Q_X(p))},$$

and

$$\begin{aligned} Q''_Z(p) &= \frac{-g'(Q_Z(p))}{g^3(Q_Z(p))} \\ &= -\frac{f'(Q_X(p))T'(Q_X(p)) - f(Q_X(p))T''(Q_X(p))}{f^3(Q_X(p))}. \end{aligned}$$

4.5 Simulation study

The main purpose of this section is to compare the CKQE $\tilde{Q}_{n,X}(p)$ and the TKQE $\hat{Q}_{n,X}(p)$. The distributions used in simulation are described in Table 4.1.

Table 4.1: Distributions used in the simulation study

Distribution	Density for $x > 0$	Parameters
Burr (α, γ, θ)	$\dfrac{\alpha\gamma(x/\theta)^\gamma}{x(1+(x/\theta)^\gamma)^{\alpha+1}}$	$(\alpha, \gamma, \theta) = (2, 3, 1)$
Paralogistic (α, θ)	$\dfrac{\alpha^2(x/\theta)^\alpha}{x(1+(x/\theta)^\alpha)^{\alpha+1}}$	$(\alpha, \theta) = (3, .5)$
Mixture of ρ log-normal(μ, σ)	$\rho\dfrac{1}{\sqrt{2\pi\sigma^2}x}\exp\left\{-\dfrac{(\log x-\mu)^2}{2\sigma^2}\right\}$	$(\rho, \mu, \sigma, \alpha, \theta) = (0.7, 0, 1, 1, 1)$
and $(1-\rho)$ Pareto(α, θ)	$+(1-\rho)\dfrac{\alpha(x/\theta)}{x(1+(x/\theta))^{\alpha+1}}$	

Note that, the mixture of log-normal and Pareto distributions was previously used in Buch-Larsen *et al.* (2005) and Charpentier and Oulidi (2010). The performance of the estimators is measured by the $AMSE$ criteria:

$$AMSE := \frac{1}{N}\sum_{s=1}^{N}\left(\hat{Q}^{(s)}_{n,X}(p) - Q(p)\right)^2,$$

where $\hat{Q}^{(s)}_{n,X}(p)$ is the quantile corresponding to the s^{th} simulated sample $\left\{X_1^{(s)}, ..., X_n^{(s)}\right\}$ and N is the number of replications. The algorithm used to estimate the quantile function with level $p \in (0, 1)$ is described as follows:

1. Generate a sample $X_1, ..., X_n$ of size n.

2. Estimate M by the empirical median $\hat{M}$, solution of $T_{\alpha,M,0}(M) = 0.5$.

3. Estimate the pair (α, c) maximizing the log-likelihood function (4.4).

4. Transform $X_1, ..., X_n$ into $Z_1, ..., Z_n$:

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

5. Compute the estimate $\hat{Q}_{n,Z}(p)$ by choosing the Epanechnikov kernel: $K(t) = \frac{3}{4}(1-t^2)\mathbf{1}_{(|t|<1)}$.

6. The resulting TKQE of the original data is

$$\hat{Q}_{n,X}(p) = T^{-1}_{\hat{\alpha},\hat{M},\hat{c}}\left(\hat{Q}_{n,Z}(p)\right).$$

7. The CKQE is directly obtained from the original data, where the bandwidth $h := h_{opt,C}$ is such as in (4.8).

We draw from the four distributions samples of size 50, 100, 500 and compute the TKQE and CKQE for different values of p in $(0,1)$. In Figures 4.1–4.4 , the solid (black), dashed (red) and dotted (blue) lines, respectively, represent the true quantile $Q(p)$, the CKQE and the TKQE. On these figures, we observe that our TKQE is always better than the CKQE, especially when p is close to 1.

Secondly, we fix the sample size at 200 and compute both the TKQE and CKQE for probability levels $p \in \{.05, .10, .25, .50, .75, .90, .95\}$. We repeat the process $N = 200$ times and we take the average. The results are summarized in Tables 4.2–4.5 where we see that the TKQE is better than the CKQE for high probability levels $p \in \{.75, .90, .95\}$. Table 4.4 is based on the mixture 30% log-normal and 70% Pareto distributions. Both estimators are equal for $p \in \{.05, .10, .25, .50\}$.

Next, we sample, 200 times, from the four distributions sets of sizes 50, 100, 500 and compute the TKQE and CKQE with their $AMSE's$ for levels $p \in \{.75, .90, .95\}$. The respective results are given in Tables 4.6, 4.7 and 4.8. It is clear that, for large probability levels, the transformation-based approach gives results of higher quality with respect to the classical procedure. Note that, under the classical estimation, some $AMSE's$ are seriously bad when samples come from mixture distributions, especially when 70% of Pareto distribution is considered. The same remark is observed in Charpentier and Oulidi (2010) (see their table's 13-18 pages 52–53).

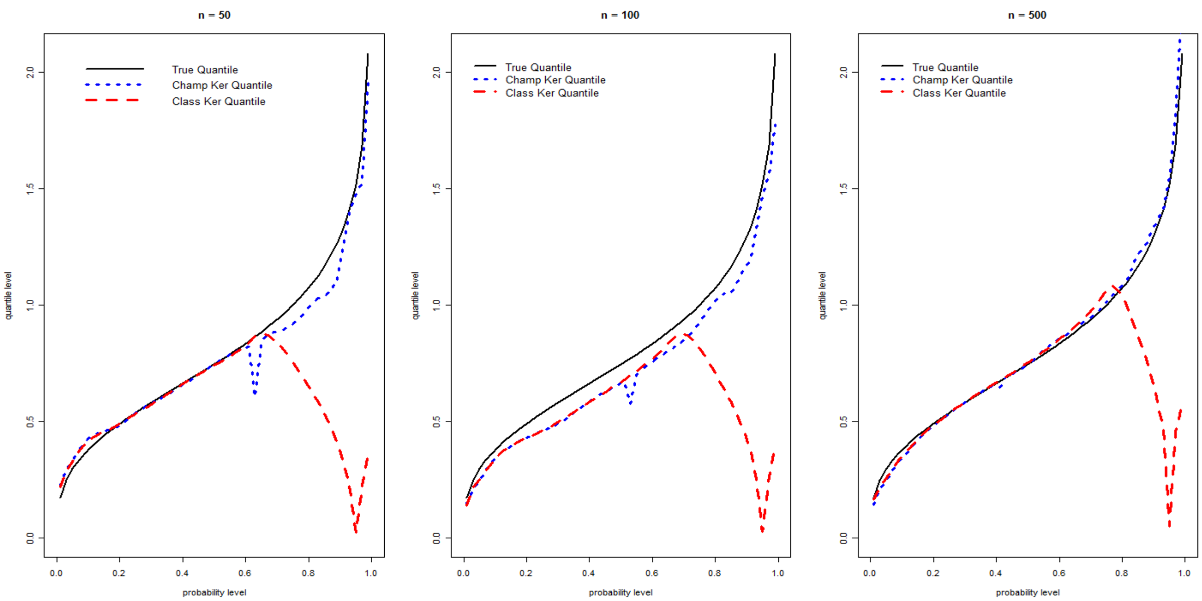

Figure 4-1: True quantile, classical and transformed p^{th} quantile estimators : Burr distribution, $n = 50$, 100 and 500, $p \in (0, 1)$.

Table 4.2: Burr distribution, 200 samples of size 200.

p	0.05	0.1	0.25	0.5	0.75	0.9	0.95
$Q(p)$	0.2962	0.3782	0.5368	0.7454	1.0000	1.2931	1.5143
$TKQE$	0.2966	0.3728	0.5345	0.7480	0.9946	1.2928	1.5150
$CKQE$	0.2988	0.3741	0.5345	0.7503	0.9852	0.5464	0.0367

Table 4.3: Paralogistic distribution, 200 samples of size 200.

p	0.05	0.1	0.25	0.5	0.75	0.9	0.95
$Q(p)$	0.1075	0.1551	0.2622	0.4291	0.6667	0.9803	1.2422
$TKQE$	0.7983	0.1278	0.2526	0.4263	0.6705	0.9676	1.1626
$CKQE$	0.1088	0.1547	0.2641	0.4330	0.7024	0.6079	0.4421

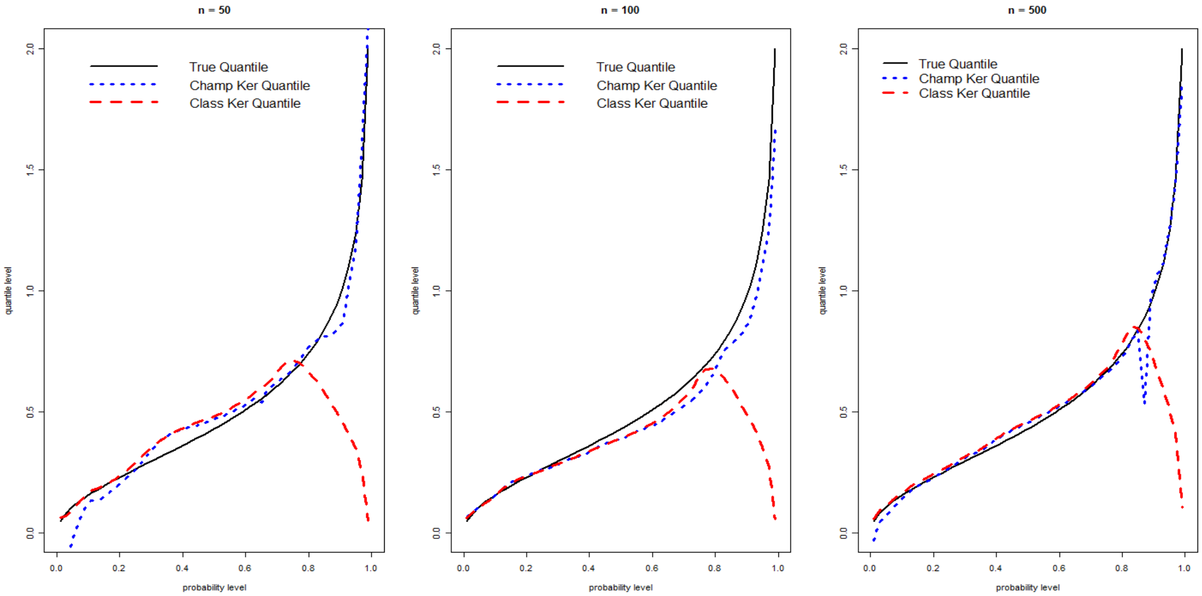

Figure 4-2: True quantile, classical and transformed p^{th} quantile estimators : Paralogistic distribution, $n = 50$, 100 and 500, $p \in (0, 1)$

Table 4.4: Mixtures (rho= 0.3) distribution, 200 samples of size 200.

p	0.05	0.1	0.25	0.5	0.75	0.9	0.95
$Q(p)$	0.0948	0.1611	0.3862	1.0000	2.6889	7.3807	14.8541
$TKQE$	0.2380	0.3391	0.6213	1.2560	2.7743	7.2812	15.2085
$CKQE$	0.2350	0.3380	0.6273	1.3246	16.4845	28.9263	21.5483

Table 4.5: Mixtures (rho= 0.7) distribution, 200 samples of size 200.

p	0.05	0.1	0.25	0.5	0.75	0.9	0.95
$Q(p)$	0.1509	0.2277	0.4566	1.0000	2.2741	5.2216	9.3262
$TKQE$	0.2987	0.4200	0.7230	1.3483	2.5389	5.1070	8.4522
$CKQE$	0.3239	0.3981	0.7293	1.3805	2.6514	6.6738	29.6183

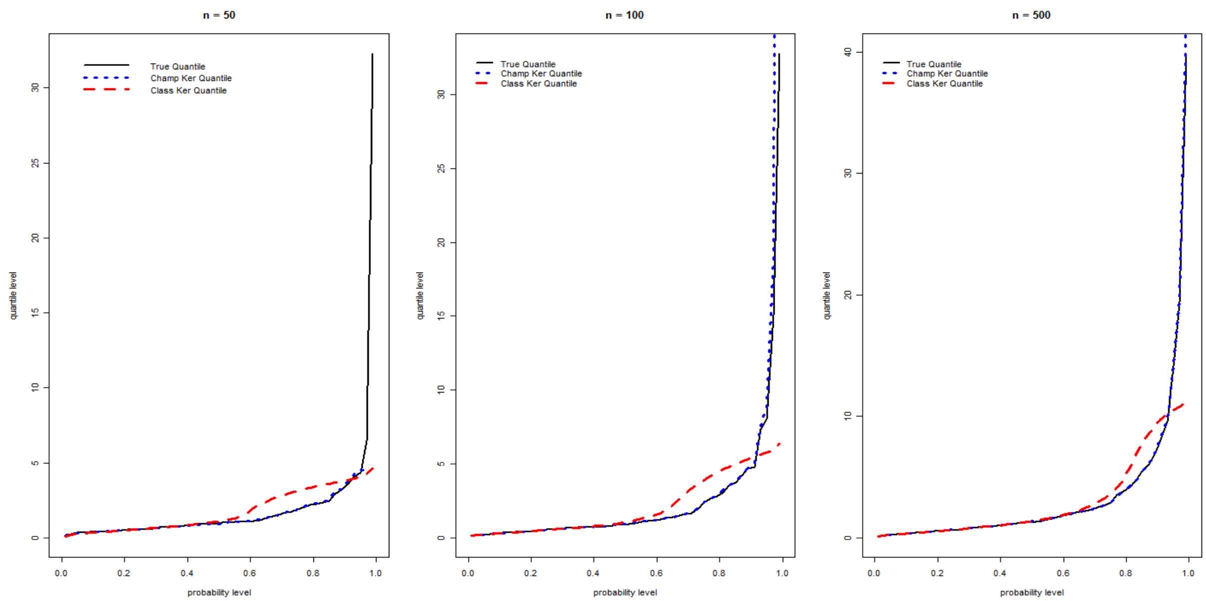

Figure 4-3: True quantile, classical and transformed p^{th} quantile estimators : Mixtures distribution ($\rho = .3$), $n = 50$, 100 and 500, $p \in (0, 1)$

Table 4.6: Classical and transformed pth quantile estimators, p= .75 and 200 replications

Distribution			Burr	Paralogistic	ρlog normal and $(1-\rho)$Pareto	
					$\rho = 30\%$	$\rho = 70\%$
$p = .75$		$Q(p)$	**1.0000**	**0.6667**	**2.6889**	**2.2741**
$n = 50$	*value*	$TKQE$	0.9623	0.6622	2.7235	2.6067
		$CKQE$	0.7963	0.7059	7.0750	3.0655
	$AMSE$	$TKQE$	0.0150	0.0059	1.0175	0.3999
		$CKQE$	0.0445	0.0080	106.46	1.1058
$n = 100$	*value*	$TKQE$	0.9912	0.6627	2.8756	2.5885
		$CKQE$	0.8922	0.7256	46.800	2.7845
	$AMSE$	$TKQE$	0.0048	0.0029	0.3518	0.2383
		$CKQE$	0.0135	0.0069	30501	0.4163
$n = 500$	*value*	$TKQE$	1.0027	0.6664	2.7815	2.5781
		$CKQE$	1.0479	0.6825	3.2990	2.6369
	$AMSE$	$TKQE$	0.0008	0.0006	0.0553	0.1151
		$CKQE$	0.0030	0.0008	0.4490	0.1522

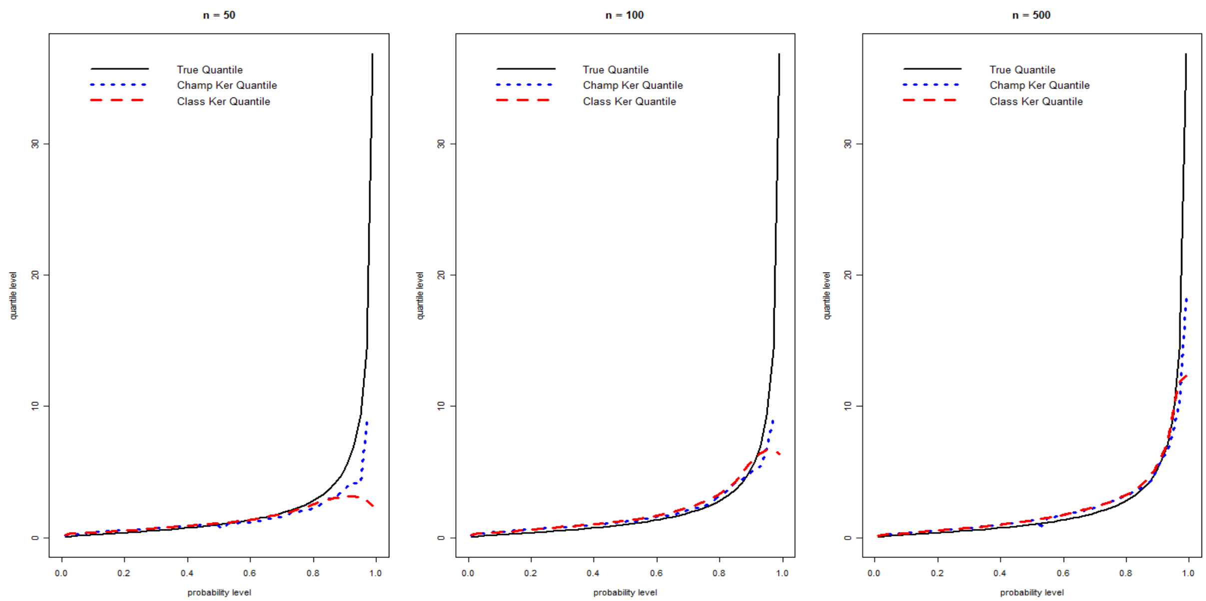

Figure 4-4: True quantile, classical and transformed p^{th} quantile estimators : Mixtures distribution ($\rho = .7$), $n = 50$, 100 and 500, $p \in (0, 1)$

Table 4.7: Classical and transformed pth quantile estimators, p= .9 and 200 replications

Distribution			Burr	Paralogistic	ρlog normal and $(1-\rho)$Pareto	
					$\rho = 30\%$	$\rho = 70\%$
$p = .90$		$Q(p)$	**1.2931**	**0.9803**	**7.3807**	**5.2216**
$n = 50$	*value*	$TKQE$	1.2941	0.9796	7.8530	5.2474
		$CKQE$	0.3864	0.4683	10.668	9.5797
	$AMSE$	$TKQE$	0.0201	0.0277	15.545	3.2335
		$CKQE$	0.8230	0.2655	298.59	179.86
$n = 100$	*value*	$TKQE$	1.2985	0.9819	7.3484	5.1982
		$CKQE$	0.4690	0.5341	12.540	11.3100
	$AMSE$	$TKQE$	0.0084	0.0113	5.3956	1.5319
		$CKQE$	0.6798	0.2012	352.99	324.23
$n = 500$	*value*	$TKQE$	1.2996	0.9773	6.9729	4.9967
		$CKQE$	0.6399	0.7219	22.028	5.3940
	$AMSE$	$TKQE$	0.0020	0.0021	1.0575	0.2473
		$CKQE$	0.4269	0.0679	698.79	0.2868

Table 4.8: Classical and transformed pth quantile estimators, p= .95 and 200 replications

Distribution			Burr	Paralogistic	ρlog normal and $(1-\rho)$Pareto	
					$\rho = 30\%$	$\rho = 70\%$
$p = .95$		$Q(p)$	**1.5143**	**1.2422**	**14.8541**	**9.3262**
$n = 50$	*value*	$TKQE$	1.5506	1.0945	16.6389	9.0187
		$CKQE$	0.0232	0.3396	12.2710	12.0748
	$AMSE$	$TKQE$	0.0443	0.0751	165.422	19.7341
		$CKQE$	2.2232	0.8165	1025.83	466.674
$n = 100$	*value*	$TKQE$	1.5332	1.1352	14.8011	8.6076
		$CKQE$	0.0291	0.3889	16.0566	17.5289
	$AMSE$	$TKQE$	0.0211	0.0702	42.2056	4.8286
		$CKQE$	2.2057	0.7294	1129.14	669.036
$n = 500$	*value*	$TKQE$	1.5181	1.1740	14.4662	8.1453
		$CKQE$	0.0498	0.5174	28.3102	27.2626
	$AMSE$	$TKQE$	0.0038	0.0468	9.4011	2.8212
		$CKQE$	2.1447	0.5259	2123.63	9055.37

Conclusion and perspectives

For heavy-tailed distributions, bias or inefficiency problems may occur in the classical kernel quantile estimation when considering high probability levels. In this paper, we have solved this incontinence by using a new approach based on the modified Champernowne distribution which behaves as the Pareto distribution. Therefore it can capture the thick-tail feature exhibited by empirical loss data. The transformation step can also be seen as a kind of variance stabilization procedure as traditionally used in statistic sampling. Our main conclusion is that the transformed kernel quantile estimator is recommended for heavy-tailed models.

By transforming a data set with a modification of the Champernowne distribution function, a kernel quantile estimator for heavy-tailed distributions gives better results in the sense of the mean square error compared with the classical estimator that defined by

$$\hat{Q}_n\left(p\right) = \sum_{i=1}^{n} X_{(i)} \int\limits_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\tfrac{x-p}{h}\right) dx.$$

Another kernel estimator of quantile is defined by

$$\tilde{Q}_n(p) = \inf\{x : \tilde{F}_n(x) \geq p\}, 0 < p < 1,$$

where $\tilde{F}_n(x)$ is the kernel distribution function estimator.

We want to compare these two Champernowne transformed estimators in the sense of the mean squared error and conclude what is the best quantiles estimators for heavy-tailed distributions, when the probability level is close to 1.

We will especially interested in the behavior of the transformed estimator $\hat{Q}_n\left(p\right)$ if we use the Beta kernel and we also compare these transformed estimators.

The law of the iterated logarithm for transformed kernel quantile function for heavy-tailed distributions is another research perspective.

Finally, the use and comparison of these kernel estimation methods for real data is an interest subject of research.

Bibliography

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

[2] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika. 68, 326-328.

[3] Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. Annals of Probability. 2: 792–804.

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

[5] Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function. Ann. Math. Statist. 39, 1083-1085.

[6] Bofinger, E. (1975). Estimation of a density function using order statistics. Austral. J. Statist. 17, 1-7.

[7] Bolancé,C., Guillén, M. and Nilsen, J. P. (2003). Kernel density estimator of actuarial loss functions. Insurance. Mathematics and Economics 32, 19-36.

[8] Bolancé, C., Guillen, M. and Nielsen, J. P. (2008). Inverse Beta transformation in kernel density estimation. Statistics & Probability Letters, 78, 1757-1764.

[9] Bolancé, C. (2010). Optimal inverse Beta(3,3) transformation in kernel density estimation. SORT 34 (2), 223-238.

[10] Buch-Larsen, T, Nielsen, J.P, Guillen, M. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. Statistics. 39 (6): 503-518.

[11] Castillo, E. (1988). Extreme value theory in engineering. Statistical Modeling and Decision Science. Academic Press, Inc.

[12] Champernowne, D.G., (1936). The Oxford Meeting, September 25–29, Econometrica, 5, October 1937.

[13] Champernowne D.G., (1952). The graduation of income distributions. Econometrica, 20, 591–615.

[14] Charpentier, A. and Oulidi, A. (2010). Beta kernel quantile estimators of heavy-tailed loss distributions. Stat. Comput. 20: 35–55.

[15] Chen. S. X. (1999). Beta kernel estimators for density functions. Comput. Statist. Data Anal. 31(2):131-145.

[16] Chen, S. X. and Tang, C.Y. (2005). Nonparametric inference of values at risk for dependent financial returns. J. Financ. Econom. 3, 227–255.

[17] 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. Assoc. 98, 993-1000.

[18] Coles S., Heffernan J. and Tawn J. (1999). Dependence Measures for Extreme Value Analyses. Extremes 2:4, 339-365.

[19] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, New York.

[20] Csörgõ, M. (1983). Quantile process with statistical applications (SIAM, Philadelphia).

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

[22] David, H.A. (1981). Order Statistics, 2nd Edition, New York: John Wiley.

[23] Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (2005). Actuarial Theory for Dependent Risk: Measures, Orders and Models. Wiley, New York.

[24] Devroye, L, Machell, F. and Penrod, C.S. (1983). The transformed kernel estimate. Technical Report, Applied Reseach Laboratories, University of Texas. Austin. Texas.

[25] Devroye, L. and Gyorfi, L. (1985). Nonparametric density estimation : The L_1 View. New York: Wiley.

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

[27] Embrechts, P. Mc Neil and Straumann. (1999). Correlation: Pitfalls and alternatives A short, non-technical article, RISK Magazine, May, 69-71.

[28] Epanechnikov, V. A. (1969). Nonparametric Estimation of a Multivariate Probability Density. Theory of Probability and its Applications. 14 (1): 153-158.

[29] Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, Statist. Neerlandica. 37, 73-83.

[30] Falk, M. (1984). Relative Deficiency of Kernel Type Estimators of Quantiles. The Annals of Statistics. 12, 261-268.

[31] Falk, M. (1985). Asymptotic normality of a kernel quantile estimator. Ann. Statist. 13, 428-433.

[32] Falk, M. (1986). On the estimation of the quantile density estimation. Statistics & Probability Letters. 4, 69-73.

[33] Feller, W. (1971). An Introduction to Probability Theory and Its Applications, volume II. John Wiley Sons, 2nd edition edition.

[34] Fernholz, L. (1991). Almost sure convergence of smoothed empirical distribution functions. Scandinavian Journal of Statistics. 18, 255–262.

[35] Fisher, R. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of largest or smallest member of a sample. Proceedings of the Cambridge philosophical society. 24:180–190.

[36] Galambos, J. (1987). The Asymptotic Theory of Extreme Order Statistics. Krieger, Malabar, Florida.

[37] Gnedenko, B. V. (1943). Sur la distribution limite du terme d'une série aléatoire. Annals of Mathematics. 44:423-453.

[38] Gumbel, E. J. (1958). Statistics of extremes. Columbia University Press.

[39] de Haan L. (1970). On Regular Variation and its Applications to the Weak Convergence of Sample Extremes. Mathematical Centre Tract, 32, Amsterdam.

[40] Harrell, F. E. and Davis, C. E. (1982). A New Distribution-Free Quantile Estimator. Biometrika. 69, 635-640.

[41] Hill, P. D. (1985). Kernel estimation of a distribution function. Comm. Statist. Theory Methods. 14, 605-620.

[42] Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalised Pareto distribution. Technometrics 29, 339-349.

[43] Johnson, N. L. Jotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distribution, vol. 1. John Wiley & Sons, Inc., New York.

[44] Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. Statistics and Probability Letters, 9 :129-132.

[45] Jones, M.C. (1993). Simple boundary correction for kernel density estimation. Statistics and Computing. 3, 135-146.

[46] Kaigh, W. D. and Lachenbruch, P. A. (1982). A Generalized Quantile Estimator,Communications in Statistics. Part A - Theory and Methods, 11 , 2217-2238.

[47] Kaigh, W. D. and Driscoll, M. F. (1987). Numerical and graphical data summary using O-Statistics. Amer, Statist. 41, 25-32.

[48] Lehmann, E. L. (1991). Theory of Point Estimation. Wadsworth and Brooks/Cole, Belmont.

[49] Mack, Y. P. (1984). Remarks on some smoothed empirical distribution functions and processes, Bull. Inform. Cybernet. 21, 29-35.

[50] Mandelbrot, B. (1963). The variation of certain speculative prices. The Journal of Business. 36: 394-419.

[51] Mc Neil, A.J. (1999). Extreme value theory for risk managers. Internal Modelling and CAD II published by RISK Books. 93-113.

[52] McNeil, A. J. and Frey, R. (2000). "Estimation of Tail-Related Risk Measures for Heterocedastic Financial Times Series: an Extreme Value Approach". Journal of Empirical Finance. 7, 271-300.

[53] Nadaraya, E. A. (1964). Some new estimates for distribution function. Theory of Probab. Appl. 9, 497-500.

[54] Padgett, W. J. (1986). A Kernel-Type Estimator of a Quantile Function from Right-Censored Data, Journal of the American Statistical Association. 81, 215-222.

[55] Park, C. (2006). Smooth nonparametric estimation of a quantile function under right censoring using beta kernels. Technical Report (TR 2006-01-CP), Department of Mathematical Sciences, Clemson University.

[56] Parzen, E. (1962). On the estimation of a probability density and mode, Ann. Math. Statist. 33, 1065-1076.

[57] Parzen, E. (1979). Nonparametric Statistical Data Modelling. Journal of the American Statistical Association. 74, 105-131.

[58] Pickands, J. (1975). Statistical inference using extreme order statistics. Annals of Statistics. 3, 119-131.[35].

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

[60] Ralescu, S. S. and Sun, S. (1993). Necessary and sufficient conditions for theasymptotic normality of perturbed sample quantiles. J. Statist. Plann. Inference. 35, 55-64.

[61] Resnick, S. I. (1987). Extreme Values, regular Variation, and Point Process. Springer-Verlag, New-York.

[62] Richard A. Tapia, James R. Thompson (1990). Nonparametric function estimation, modeling, and simulation /Society for Industrial and Applied Mathematics (SIAM).

[63] Reiss, R. D. (1978). Approximate distribution of the maximum deviation of histogram. Metrika. 25, 9-26.

[64] Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions, Scand. J. Statist. 8, 116-119.

[65] Reiss, R. D. and Thomas, M. (1997). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.

[66] Ronsenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math.Statist. 27, 832-837.

[67] Sacks, J, and Ylvisaker, D. (1981). Asymptotically optimal kernels for density estimation at a point. Ann. Statist. 9, 334-346.

[68] Sefling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley, NewYork.

[69] Sheater, S. J. and Marron, J. S. (1990). Kernel quatile estimtors. Journal of the American Statistical Association. 85, 410-416.

[70] Shirahata, S and Chu I. S. (1992). Integrated squared error of kernel-type estimator of distribution function. Annals of the Institute of Statistical Mathematics. 44: 579–591.

[71] Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. J. Res. Nat. Bur. Standards. 64 B, 145-150.

[72] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6, 177-184.

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

[74] Singh, R. S., Gasser, T. and Prasad, B. (1983). Nonparametric estimates of distributions functions. Communication in Statistics - Theory and Methods. 12 :2095-2108.

[75] Smith, R.L. (1987). Estimating tails of probability distributions. A. Statist. 15, 1174-1207.

[76] Swanepoel, J. W. H. (1988). Mean intgrated squared error properties and optimal kernel when estimating a distribution function. Comm. Statist. Theory Methods. 17, 3785-3799.

[77] Terrel, G. R. and Scott, D.W. (1985). Oversmoothed nonparametric density estimates. Journal of the American Statistical Association. 80, 209-214.

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

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

[80] van der Vaart, A. W. and Weller, J.A. (1996). Weak convergence and empirical Process. Spring Verlag, New York.

[81] Wand, M. P, Marron, J. S. and Ruppert, D. (1991). Transformations in density estimation. J. Amer. Statist. Assoc. 86 (414): 343-361.

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

[83] Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis II, Sankhyd Set. A, 26, 101-116.

[84] Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators, Canad. J.Statist. 1, 247-253.

[85] Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions, J. Appl. Probab. 16, 163-173.

[86] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, Bull. Math. Statist. 15, 69-78.

[87] Yang, S. S. (1985). A Smooth Nonparametric Estimator of a Quantile Function. Journal of the American Statistical Association. 80, 1004-1011.

[88] Zelterman, D. (1990). Nonparametric quantile estimation. J. Statist. Plann. and inference. 26, 339-352.