DISCUSSION PAPER

March 2011; revised September 2011 ■ RFF DP 11-19-REV

Fat-Tailed Distributions: Data, Diagnostics, and Dependence

Roger M. Cooke and Daan Nieboer, with
Jolanta Misiewicz

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Resources for the Future
Department of Mathematics, Delft University of Technology

Fat-Tailed Distributions: Data, Diagnostics and Dependence

Roger M. Cooke and Daan Nieboer with a chapter by Jolanta Misiewicz

September 2011

Support from US National Science Foundation grant no. 0960865 and comments from Prof. J. Misiewicz are gratefully acknowledged

Abstract
This monograph is written for the numerate nonspecialist, and hopes to serve three purposes. First it gathers mathematical material from diverse but related fields of order statistics, records, extreme value theory, majorization, regular variation and subexponentiality. All of these are relevant for understanding fat tails, but they are not, to our knowledge, brought together in a single source for the target readership. Proofs that give insight are included, but for most fussy calculations the reader is referred to the excellent sources referenced in the text. Multivariate extremes are not treated. This allows us to present material spread over hundreds of pages in specialist texts in twenty pages. Chapter 5 develops new material on heavy tail diagnostics and gives more mathematical detail. Since variances and covariances may not exist for heavy tailed joint distributions, Chapter 6 reviews dependence concepts for certain classes of heavy tailed joint distributions, with a view to regressing heavy tailed variables.

Second, it presents a new measure of obesity. The most popular definitions in terms of regular variation and subexponentiality invoke putative properties that hold at infinity, and this complicates any empirical estimate. Each definition captures some but not all of the intuitions associated with tail heaviness. Chapter 5 studies two candidate indices of tail heaviness based on the tendency of the mean excess plot to collapse as data are aggregated. The probability that the largest value is more than twice the second largest has intuitive appeal but its estimator has very poor accuracy. The *Obesity index* is defined for a positive random variable X as:

$$\mathrm{Ob}(X) = P\left(X_1 + X_4 > X_2 + X_3 | X_1 \leq X_2 \leq X_3 \leq X_4\right),\ X_i \text{ independent copies of } X.$$

For empirical distributions, obesity is defined by bootstrapping. This index reasonably captures intuitions of tail heaviness. Among its properties, if $\alpha > 1$ then $\mathrm{Ob}(X) < \mathrm{Ob}(X^\alpha)$. However, it does not completely mimic the tail index of regularly varying distributions, or the extreme value index. A Weibull distribution with shape 1/4 is more obese than a Pareto distribution with tail index 1, even though this Pareto has infinite mean and the Weibull's moments are all finite. Chapter 5 explores properties of the Obesity index.

Third and most important, we hope to convince the reader that fat tail phenomena pose real problems; they are really out there and they seriously challenge our usual ways of thinking about historical averages, outliers, trends, regression coefficients and confidence bounds among many other things. Data on flood insurance claims, crop loss claims, hospital discharge bills, precipitation and damages and fatalities from natural catastrophes drive this point home. While most fat tailed distributions are "bad", research in fat tails is one distribution whose tail will hopefully get fatter.
AMS classification 60-02, 62-02, 60-07.

Contents

Chapter 1

Fatness of Tail

1.1 Fat tail heuristics

Suppose the tallest person you have ever seen was 2 meters (6 feet 8 inches); someday you may meet a taller person, how tall do you think that person will be, 2.1 meters (7 feet)? What is the probability that the first person you meet taller than 2 meters will be more than twice as tall, 13 feet 4 inches? Surely that probability is infinitesimal. The tallest person in the world, Bao Xishun of Inner Mongolia,China is 2.36 m or 7 ft 9 in. Prior to 2005 the most costly Hurricane in the US was Hurricane Andrew (1992) at $41.5 billion USD(2011). Hurricane Katrina was the next record hurricane, weighing in at $91 billion USD(2011)[1]. People's height is a "thin tailed" distribution, whereas hurricane damage is "fat tailed" or "heavy tailed". The ways in which we reason from historical data and the ways we think about the future are - or should be - very different depending on whether we are dealing with thin or fat tailed phenomena. This monograph gives an intuitive introduction to fat tailed phenomena, followed by a rigorous mathematical treatment of many of these intuitive features. A major goal is to provide a definition of *Obesity* that applies equally to finite data sets and to parametric distribution functions.

Fat tails have entered popular discourse largely thanks to Nassim Taleb's book *The Black Swan, the impact of the highly improbable* (Taleb [2007]). The black swan is the paradigm shattering, game changing incursion from "extremistan" which confounds the unsuspecting public, the experts, and especially the professional statisticians, all of whom inhabit "mediocristan".

Mathematicians have used at least three main definitions of tail obesity. Older texts sometime speak of "leptokurtic distributions", that is, distributions whose extreme values are "more probable than normal". These are distributions with kurtosis greater than zero[2], and whose tails go to zero slower than the normal distribution.

Another definition is based on the theory of *regularly varying functions* and characterizes the rate at which the probability of values greater than x goes to zero as $x \to \infty$. For a large class of distributions this rate is polynomial. Unless otherwise indicated, we will always consider non-negative random variables. Letting F denote the distribution function of random variable X, such that $\overline{F}(x) = 1 - F(x) = Prob\{X > x\}$, we write $\overline{F}(x) \sim x^{-\alpha}, x \to \infty$ to mean $\frac{\overline{F}(x)}{x^{-\alpha}} \to 1, x \to \infty$. $\overline{F}(x)$ is called the *survivor function* of X. A survivor function with polynomial decay rate $-\alpha$, or as we shall say *tail index* α, has infinite κ^{th} moments for all $\kappa > \alpha$. The *Pareto* distribution is a special case of a regularly varying distribution where $\overline{F}(x) = x^{-\alpha}$, $x > 1$. In many cases, like the Pareto distribution, the κ^{th} moments are infinte for

[1] http://en.wikipedia.org/wiki/Hurricane_Katrina, accessed January 28, 2011

[2] Kurtosis is defined as the $(\mu_4/\sigma^4) - 3$ where μ_4 is the fourth central moment, and σ is the standard deviation. Subtracting 3 arranges that the kurtosis of the normal distribution is zero

all $\kappa \geq \alpha$. Chapter 4 unravels these issues, and shows distributions for which *all* moments are infinite. If we are "sufficiently close" to infinity to estimate the tail indices of two distributions, then we can meaningfully compare their tail heaviness by comparing their tail indices, and many intuitive features of fat tailed phenomena fall neatly into place.

A third definition is based on the idea that the sum of independent copies $X_1 + X_2+, \ldots X_n$ behaves like the maximum of $X_1, X_2, \ldots X_n$. Distributions satisfying

$$Prob\{X_1 + X_2+, \cdots + X_n > x\} \sim Prob\{Max\{X_1, X_2, \ldots X_n\} > x\}, \, x \rightarrow \infty$$

are called *subexponential*. Like regular variation, subexponality is a phenomenon that is defined in terms of limiting behavior as the underlying variable goes to infinity. Unlike regular variation, there is no such thing as an "index of subexponality" that would tell us whether one distribution is "more subexponential" than another. The set of regularly varying distributions is a strict subclass of the set of subexponential distributions. Other more exotic definitions are given in chapter 4.

There is a swarm of intuitive notions regarding heavy tailed phenomena that are captured to varying degree in the different formal definitions. The main intuitions are:

- The historical averages are unreliable for prediction
- Differences between successively larger observations increases
- The ratio of successive record values does not decrease;
- The expected excess above a threshold, given that the threshold is exceeded, increases as the threshold increases
- The uncertainty in the average of n independent variables does not converge to a normal with vanishing spread as $n \rightarrow \infty$; rather, the average is similar to the original variables.
- Regression coefficients which putatively explain heavy tailed variables in terms of covarites may behave erratically.

1.2 History and Data

A colorful history of fat tailed distributions is found in (Mandelbrot and Hudson [2008]). Mandelbrot himself introduced fat tails into finance by showing that the change in cotton prices was heavy-tailed (Mandelbrot [1963]). Since then many other examples of heavy-tailed distributions are found, among these are data file traffic on the internet (Crovella and Bestavros [1997]), returns on financial markets (Rachev [2003], Embrechts et al. [1997]) and magnitudes of earthquakes and floods (Latchman et al. [2008], Malamud and Turcotte [2006]). The website http://www.er.ethz.ch/presentations/Powerlaw_mechanisms_13July07.pdf gives examples of earthquake number per $5x5$ km grid,wildfires, solar flares, rain events, financial returns, movie sales, health car costs, size of wars, etc etc.

Data for this monograph were developed in the NSF project 0960865, and are available from $http://www.rff.org/Events/Pages/Introduction-Climate-Change-Extreme-Events.aspx$, or at public cites indicated below.

1.2.1 US Flood Insurance Claims

US flood insurance claims data from the National Flood Insurance Program (NFIP) are aggregated by county and year for the years 1980 to 2008. The data are in 2000 US dollars. Over this time period there has been substantial growth in exosure to flood risk, particularly in coastal counties. To remove the effect of growing exposure, the claims are divided by personal income estimates per county per year from the Bureau of Economic Accounts (BEA). Thus, we study flood claims per dollar income, by county and year[3].

1.2.2 US Crop Loss

US crop insurance indemnities paid from the US Department of Agriculture's Risk Management Agency are aggregated by county and year for the years 1980 to 2008. The data are in 2000 US dollars. The crop loss claims are not exposure adjusted, as a proxy for exposure is not obvious, and exposure growth is less of a concern.[4]

1.2.3 US Damages and Fatalities from Natural Disasters

The SHELDUS database, maintained by the Hazards and Vulnerability Research Group at the University of South Carolina, has county-level damages and fatalities from weather events. Information on SHELDUS is available online: http://webra.cas.sc.edu/hvri/products/SHELDUS.aspx. The damage and fatality estimates in SHELDUS are minimum estimates as the approach to compiling the data always takes the most conservative estimates. Moreover, when a disaster affected many counties, the total damages and fatalities were apportioned equally over the affected counties, regardless of population or infrastructure. These data should therefore be seen as indicative rather than as precise.

1.2.4 US Hospital Discharge Bills

Billing data for hospital discharges for a northeastern US state were collected over the years 2000 - 2008. The data is in 2000 USD.

1.2.5 G-Econ data

This uses the G-Econ database (Nordhaus et al. [2006]) showing the dependence of Gross Cell Product (GCP) on geographic variables measured on a spatial scale of one degree. At 45 latitude, a one by one degree grid cell is $[45mi]^2$ or $[68km]^2$. The size varies substantially from equator to pole. The population per grid cell varies from 0.31411 to 26,443,000. The Gross Cell Product is for 1990, non-mineral, 1995 USD, converted at market exchange rates. It varies from 0.000103 to 1,155,800 USD(1995), the units are $\$10^6$. The GCP per person varies from 0.00000354 to 0.905, which scales from \$3.54 to \$905,000. There are 27,445 grid cells. Throwing out zero and empty cells for population and GCP leaves 17,722; excluding cells with empty temperature data leaves 17,015 cells.

The data are publicly available at *http://gecon.yale.edu/world_big.html*.

1.3 Diagnostics for Heavy Tailed Phenomena

Once we start looking, we can find heavy tailed phenomena all around us. Loss distributions are a very good place to look for tail obesity, but something as mundane as hospital discharge billing data can also produce surprising evidence. Many of the features of heavy tailed phenomena would render our traditional statistical tools useless at best, dangerous at worst. Prognosticators base predictions on historical averages. Of course, on a finite sample the average and standard

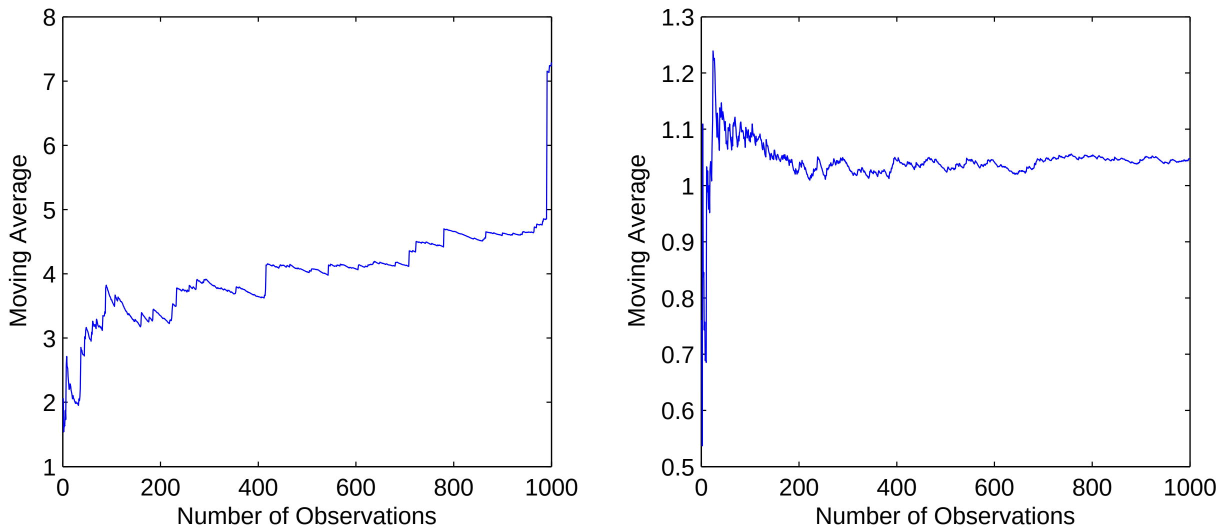
[3]Help from Ed Pasterick and Tim Scoville in securing and analysing this data is gratefully acknowledged.
[4]Help from Barbara Carter in securing and analysing this data is gratefully acknowledged.

deviation are always finite; but these may not be converging to anything and their value for prediction might be nihil. Or again, if we feed a data set into a statistical regression package, the regression coefficients will be estimated as "covariance over the variance". The sample versions of these quantities always exist, but if they aren't converging, their ratio could whiplash wildly, taking our predictions with them. In this section, simple diagnostic tools for detecting tail obesity are illustrated on mathematical distributions and on real data.

1.3.1 Historical Averages

Consider independent and identically distributed random variables with tail index $1 < \alpha < 2$. The variance of these random variables is infinite, as is the variance of any finite sum of these variables. In consequence, the variance of the average of n variables is also infinite, for any n. The mean value is finite and is equal to the expected value of the historical average, but regardless how many samples we take, the average does not converge to the variable's mean, and we cannot use the sample average to estimate the mean reliably. If $\alpha < 1$ the variables have infinite mean. Of course the average of any finite sample is finite, but as we draw more samples, this sample average tends to increase. One might mistakenly conclude that there is a time trend in such data. The universe is finite and an empirical sample would exhaust all data before it reached infinity. However, such re-assurance is quite illusory; the question is, "where is the sample average going?". A simple computer experiment suffices to convince the sceptic: sample a set of random numbers on your computer, these are approximately independent realizations of a uniform variable on the interval [0,1]. Now invert these numbers. If U is such a uniform variable, 1/U is a Pareto variable with tail index 1. Compute the moving averages and see how well you can predict the next value.

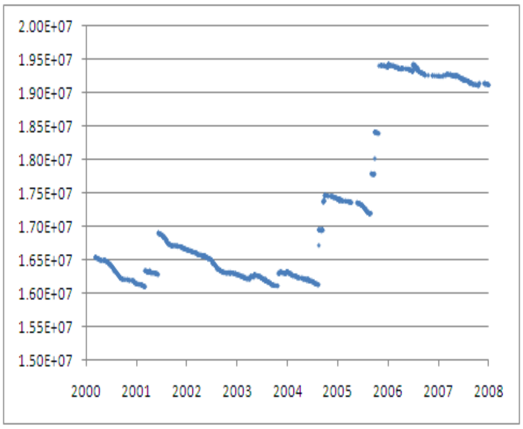
Figure 1.1 (a)–(b) shows the moving average of respectively a Pareto(1) distribution and a standard exponential distribution. The mean of the Pareto(1) distribution is infinite whereas the mean of the standard exponential distributions is equal to one.



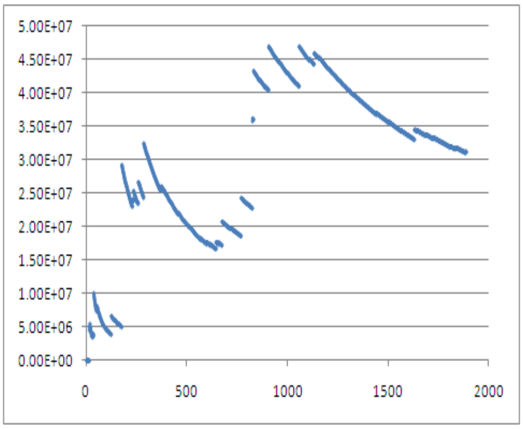
(a) Pareto(1) (b) Standard exponential

Figure 1.1: Moving average of Pareto(1) and standard exponential data

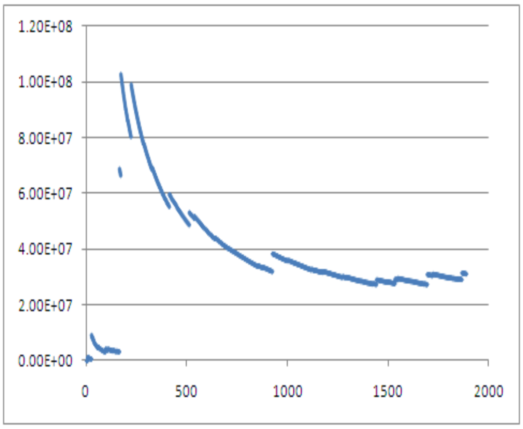
As we can see, the moving average of the Pareto(1) distribution shows an upward trend, whereas the moving average of the Standard Exponential distribution converges to the real



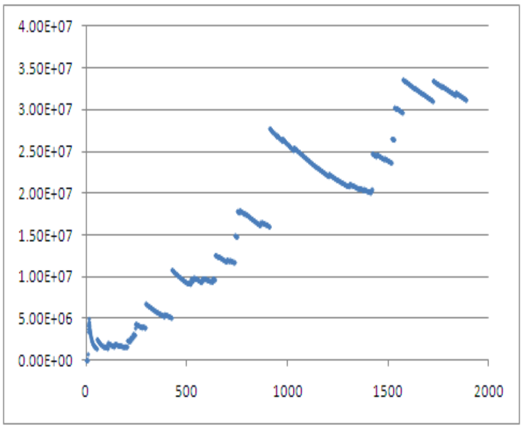
(a) Natural disaster property damage, temporal order



(b) Natural disaster property damage, random order 1



(c) Natural disaster property damage, random order 2



(d) Natural disaster property damage, random order 3

Figure 1.2: Moving average US natural disaster property damage

mean of the Standard Exponential distribution. Figure 1.2 (a) shows the moving average of US property damage from natural disasters from 2000 to 2008. We observe an increasing pattern; this might be caused by attempting to estimate an infinite mean, or it might actually reflect a temporal trend. One way to approach this question is to present the moving average in random order, as in (b),(c), (d). It is important to realize that these are simply different orderings of the same data set. Note the differences on the vertical axes. Firm conclusions are difficult to draw from single moving average plots for this reason.

1.3.2 Records

One characteristic of heavy-tailed distributions is that there are usually a few very large values compared to the other values of the data set. In the insurance business this is called the Pareto law or the 20-80 rule-of-thumb: 20% of the claims account for 80% of the total claim amount in an insurance portfolio. This suggests that the largest values in a heavy tailed data set tend to be further apart than smaller values. For regularly varying distributions the ratio between the two largest values in a data set has a non-degenerate limiting distribution, whereas for distributions like the normal and exponential distribution this ratio tends to zero as we increase the number of observations. If we order a data set from a Pareto distribution, then the ratio between two consecutive observations also has a Pareto distribution. In Table 1.1 we see the

Number of observations	standard normal distribution	Pareto(1) distribution
10	0.2343	$\frac{1}{2}$
50	0.0102	$\frac{1}{2}$
100	0.0020	$\frac{1}{2}$

Table 1.1: Probability that the next record value is at least twice as large as the previous record value for different size data sets

probability that the largest value in the data set is twice as large as the second largest value for the standard normal distribution and the Pareto(1) distribution. The probability stays constant for the Pareto distribution, but it tends to zero for the standard normal distribution as the number of observations increases.

Seeing that one or two very large data points confound their models, unwary actuaries may declare these "outliers" and discard them, re-assured that the remaining data look "normal". Figure 1.3 shows the yearly difference between insurance premiums and claims of the U.S. National Flood Insurance Program (NFIP) (Cooke and Kousky [2009]).

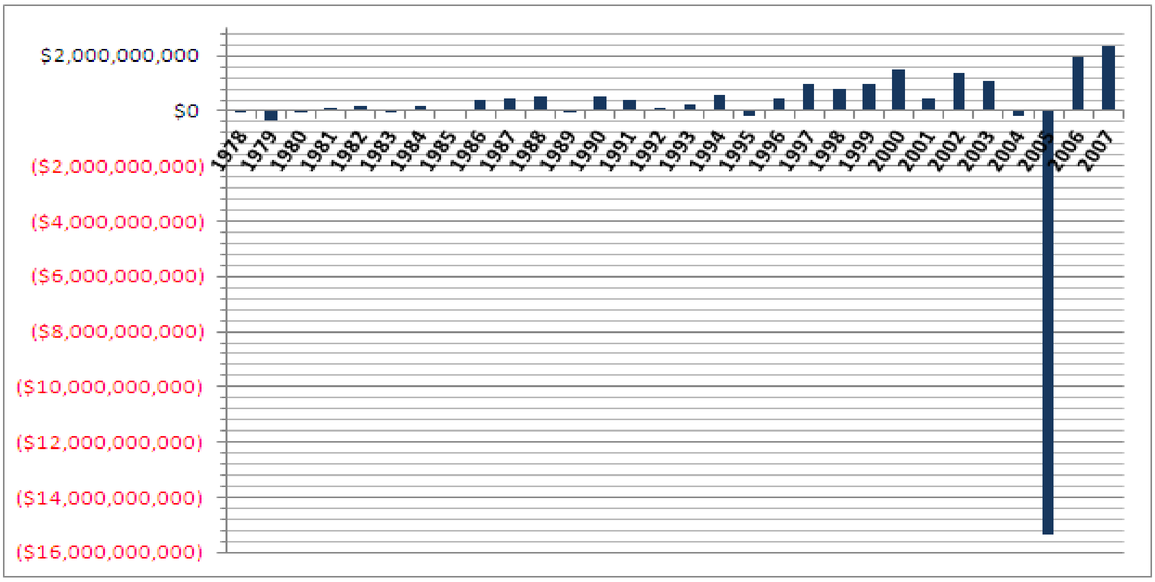


Figure 1.3: US National Flood Insurance Program, premiums minus claims

The actuaries who set NFIP insurance rates explain that their "historical average" gives 1% weight to the 2005 results including losses from hurricanes Katrina, Rita, and Wilma: "This is an attempt to reflect the events of 2005 without allowing them to overwhelm the pre-Katrina experience of the Program" (Hayes and Neal [2011] p.6)

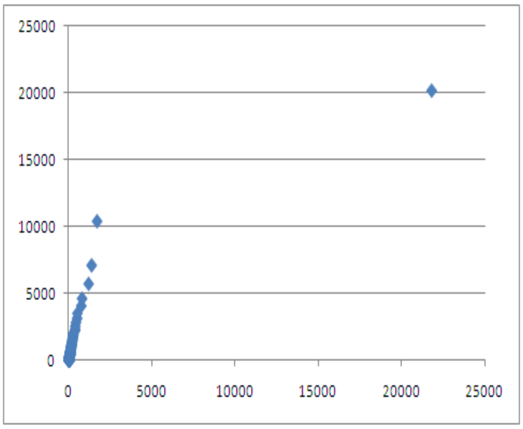
1.3.3 Mean Excess

The *mean excess function* of a random variable X is defined as:

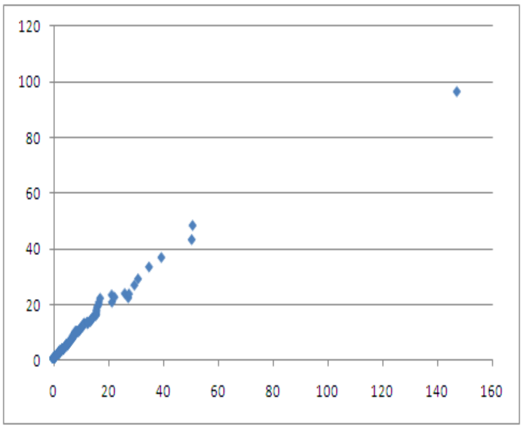
$$e(u) = E\left[X - u | X > u\right] \tag{1.1}$$

The mean excess function gives the expected excess of a random variable over a certain threshold given that this random variable is larger than the threshold. It is shown in chapter 4 that subexponential distributions' mean excess function tends to infinity as u tends to infinity. If we know that an observation from a subexponential distribution is above a very high threshold then we expect that this observation is much larger than the threshold. More intuitively, we should expect the next worst case to be much worse than the current worst case. It is also shown that regularly varying distributions with tail index $\alpha > 1$, have a mean excess function which is ultimately linear with slope $\frac{1}{\alpha-1}$. If $\alpha < 1$, then the slope is infinite and (1.1) is not useful. If we order a sample of n independent realizations of X, we can construct a *mean excess plot* as in (1.2). Such a plot will not show an infinite slope, rendering the interpretation of such plots problematic for very heavy tailed phenomena.

$$e(x_i) = \frac{\sum_{j>i} x_j - x_i}{n-i}; \; i < n, \; e(x_n) = 0; \;\; x_1 < x_2 < \ldots x_n. \tag{1.2}$$



(a) Mean excess Pareto 1



(b) Mean excess Pareto 2

Figure 1.4: Pareto mean excess plots, 5000 samples

Figure 1.4 shows mean excess plots of 5000 samples from a Pareto(1) (a) and a Pareto(2) (b). Clearly, eyeballing the slope in these plots gives a better diagnostic for (b) than for (a).

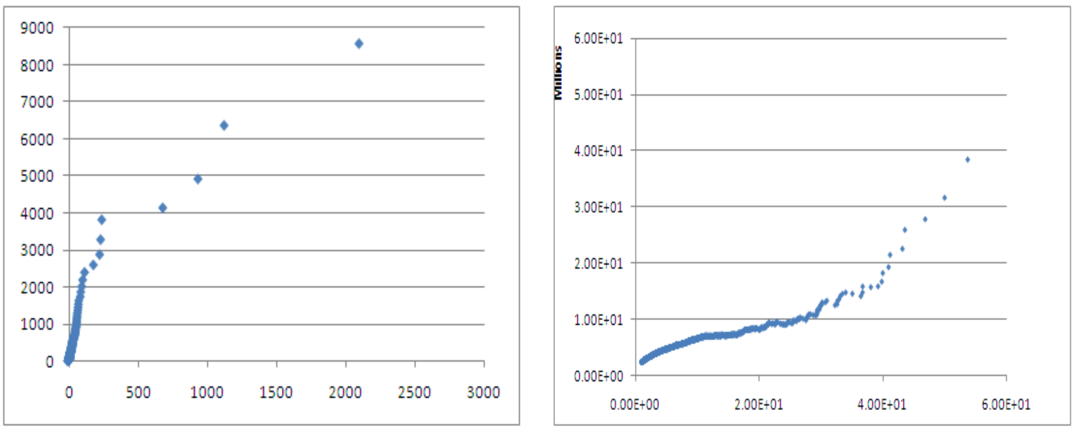
Figure 1.5 shows mean excess plots for flood claims per county per year per dollar income (a), and insurance claims for crop loss per year per county (b). Both plots are based on roughly the top 5000 entries.

1.3.4 Sum convergence: Self-similar or Normal

For regularly varying random variables with tail index $\alpha < 2$ the standard central limit theorem does not hold: The standardized sum does not converge to a normal distribution. Instead the generalized central limit theorem (Uchaikin and Zolotarev [1999]) applies: The sum of these random variables, appropriately scaled, converges to a stable distribution having the same tail index as the original random variable.

This can be observed in the mean excess plot of data sets of 5000 samples from a regularly varying distribution. In the mean excess plot the empirical mean excess function of a data set is plotted. Define the operation *aggregating by k* as dividing a data set randomly into groups of size k and summing each of these k values. If we consider a data set of size n and compare the mean excess plot of this data set with the mean excess plot of a data set we obtained through aggregating the original data set by k, then we find that both mean excess plots are very similar. For data sets from thin-tailed distributions both mean excess plots look very different.

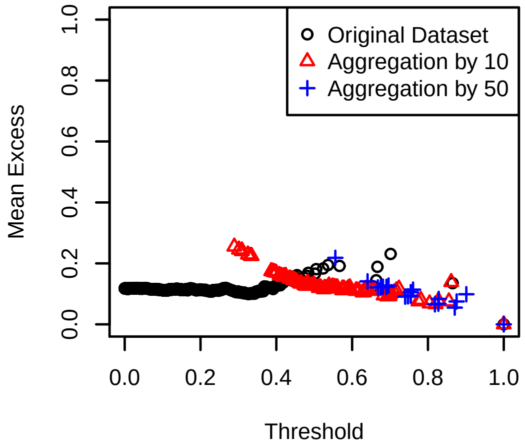
In order to compare the shapes of the mean excess plots we have standardized the data such that the largest value in the data set is scaled to one. This does not change the shape of the mean excess plot, since we can easily see that $e(cu) = ce(u)$. Figure 1.6 (a)–(d) shows the standardized mean excess plot of a sample from an exponential distribution, a Pareto(1) distribution, a Pareto(2) distribution and a Weibull distribution with shape parameter 0.5. Also shown in each plot are the standardized mean excess plots of a data set obtained through aggregating by 10 and 50. The Weibull distribution is a subexponential distribution whenever the shape parameter $\tau < 1$. Aggregating by k for the exponential distribution causes the slope of the standardized mean excess plot to collapse. For the Pareto(1) distribution, aggregating the sample does not have much effect on the mean excess plot. The Pareto(2) is the "thinnest" distribution with infinite variance, but taking large groups to sum causes the mean excess slope to collapse. Its behavior is comparable to that of the data set from a Weibull distribution with shape 0.5. This underscores an important point: Although a Pareto(2) is a very fat tailed distribution and a Weibull with shape 0.5 has all its moments and has tail index ∞, the behavior of data sets of 5000 samples is comparable. In this sense, the tail index does not tell the whole story. Figures 1.7 (a)–(b) show the standardized mean excess plot for two data sets. The standardized mean excess plot in Figure 1.7a is based upon the income- and exposure-adjusted flood claims from the National Flood Insurance program in the United States from the years 1980 to 2006. US crop loss is the second data set. This data set contains all pooled values per county with claim sizes larger than \$ 1.000.000. The standardized mean excess plot of the flood data in Figure 1.7a seems to stay the same as we aggregate the data set. This is indicative for data drawn from a distribution with infinite variance. The standardized mean excess plot of the national crop insurance data in Figure 1.7b changes when taking random aggregations, indicative of finite variance.



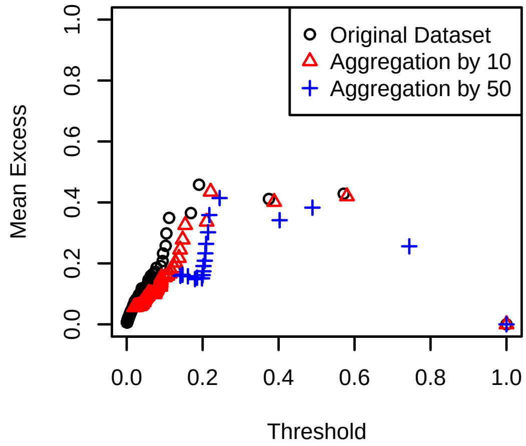
(a) Mean excess flood claims/income

(b) Mean excess crop loss

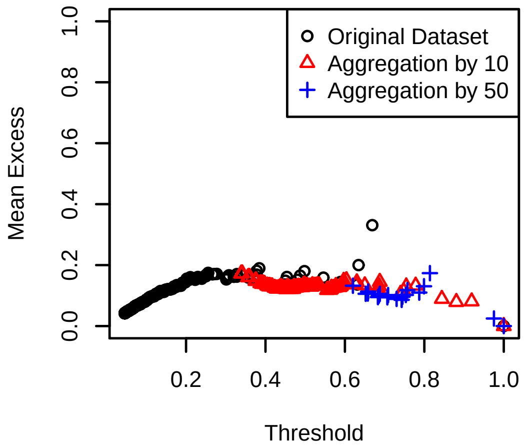
Figure 1.5: Mean excess plots, flood and crop loss



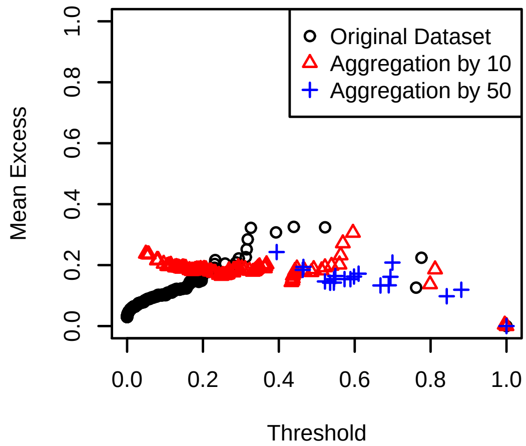
(a) Exponential distribution



(b) Pareto $\alpha = 1$



(c) Pareto $\alpha = 2$

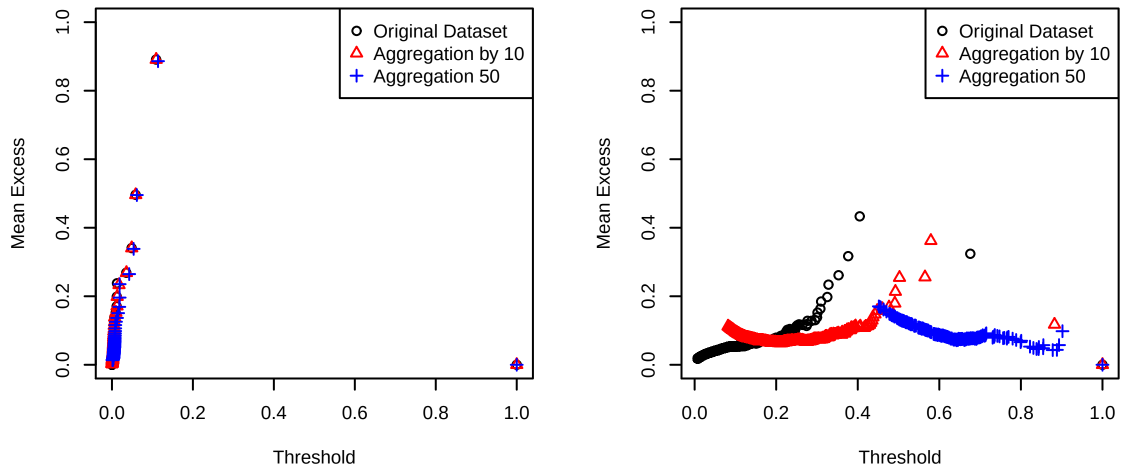


(d) Weibull distribution $\tau = 0.5$

Figure 1.6: Standardized mean excess plots

1.3.5 Estimating the Tail Index

"Ordinary" statistical parameters characterize the entire sample and can be estimated from the entire sample. Estimating a tail index is complicated by the fact that it is a parameter of a limit distribution. If independent samples are drawn from a regularly varying distribution, then the survivor function tends to a polynomial as the samples get large. We cannot estimate the degree of this polynomial from the whole sample. Instead we must focus on a small set of large values and hope that these are drawn from a distribution which approximates the limit distribution.



(a) Flood claims per income (b) National crop insurance

Figure 1.7: Standardized mean excess plots of two data sets

In this section we briefly review methods that have been proposed to estimate the tail index.

One of the simplest methods is to plot the empirical survivor function on log-log axes and fit a straight line above a certain threshold. The slope of this line is then used to estimate the tail index. Alternatively, we could estimate the slope of the mean excess plot. As noted above, this latter method will not work for tail indices less than or equal to one. The self-similarity of heavy-tailed distributions was used in Crovella and Taqqu [1999] to construct an estimator for the tail index. The ratio

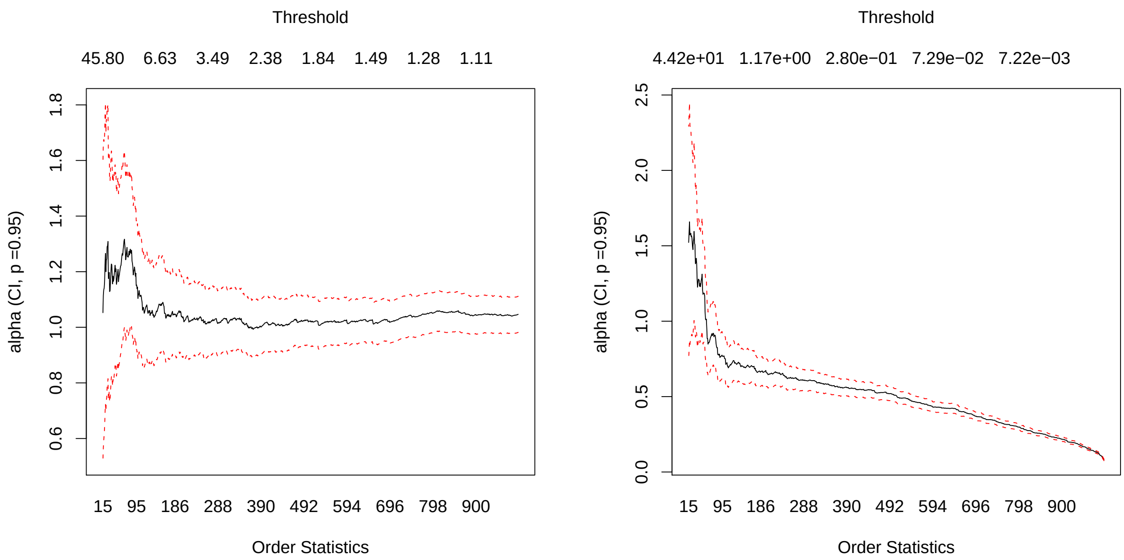
$$R(p,n) = \frac{max\{X_1^p, \dots X_n^p\}}{\sum_{i=1}^{n} X_i^p};\ X_i > 0, i = 1 \dots n$$

is sometimes used to detect infinite moments. If the $p-th$ moment is finite then $lim_{n\to\infty} R(p,n) = 0$ (Embrechts et al. [1997]). Thus if for some p, $R(p,n) \gg 0$ for large n, then this suggests an infinite $p-th$ moment. Regularly varying distributions are in the "max domain of attraction" of the Fréchet class. That is, under appropriate scaling the maximum converges to a *Fréchet* distribution: $F(x) = exp(-x^{-\alpha}), x > 0, \alpha > 0$. Note that for large x, $x^{-\alpha}$ is small and $F(X) \sim 1 - x^{-\alpha}$ The parameter $\xi = 1/\alpha$ is called the *extreme value index* for this class. There is a rich literature in estimating the extreme value index, for which we refer the reader to (Embrechts et al. [1997])

Perhaps the most popular estimator of the tail index is the Hill estimator proposed in Hill [1975] and given by

$$\mathcal{H}_{k,n} = \frac{1}{k}\sum_{i=0}^{k-1} \left(\log(X_{n-i,n}) - \log(X_{n-k,n})\right),$$

where $X_{i,n}$ are such that $X_{1,n} \le ... \le X_{n,n}$. The tail index is estimated by $\frac{1}{\mathcal{H}_{k,n}}$. The idea behind this method is that if a random variable has a Pareto distribution then the log of this random variable has an exponential distribution $S(x) = e^{-\lambda x}$ with parameter λ equal to the tail index. $\frac{1}{\mathcal{H}_{k,n}}$ estimates the parameter of this exponential distribution. Like all tail index estimators, the Hill estimator depends on the threshold, and it is not clear how it should be chosen. A useful heuristic here is that k is usually less than $0.1n$. Methods exist that choose k by minimizing the asymptotic mean squared error of the Hill estimator. Although it works very well for Pareto distributed data, for other regularly varying distribution functions the Hill estimator becomes less effective. To illustrate this we have drawn two different samples, one from the Pareto(1)



(a) Pareto distributed data (b) Burr distributed data

Figure 1.8: Hill estimator for samples of a Pareto and Burr distribution with tail index 1.

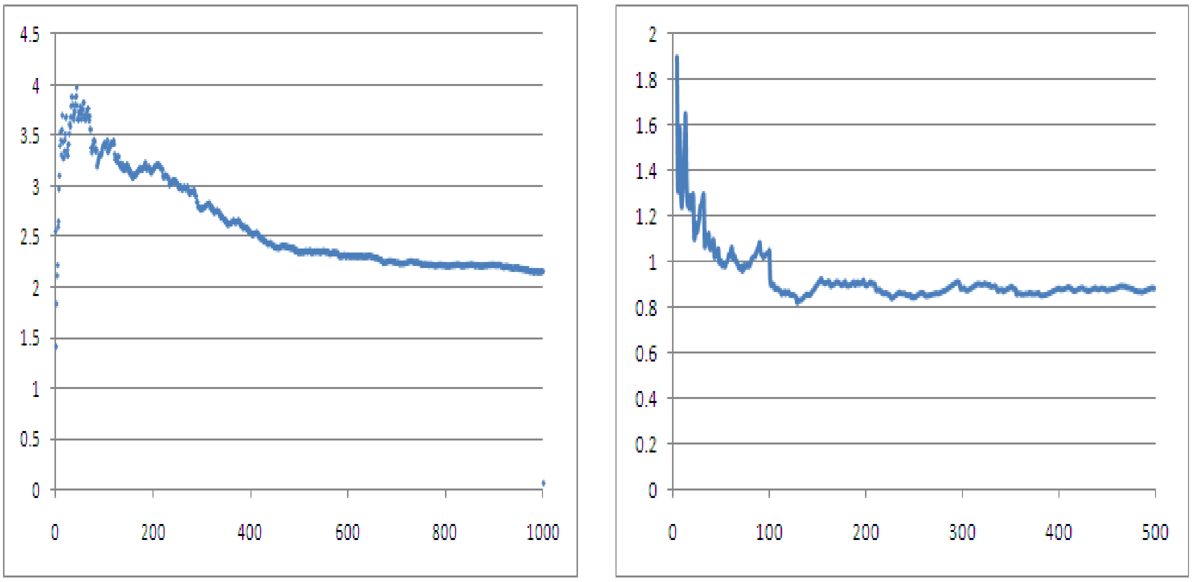
distribution and one from a Burr distribution (see Table 4.1) with parameters such that the tail index of this Burr distribution is equal to one. Figure 1.8 (a), (b) shows the Hill estimator for the two data sets together with the 95%-confidence bounds of the estimate. Note that the Hill estimate is plotted against the different values in the data set running from largest to smallest. and the largest value of the data set is plotted on the left of the x-axis. As we can see from Figure 1.8a, the Hill estimator gives a good estimate of the tail index, but from Figure 1.8b it is not clear that the tail index is equal to one. Beirlant et al. [2005] explores various improvements of the Hill estimator, but these improvements require extra assumptions on the distribution of the data set.

Figure 1.9, shows a Hill plot for crop losses (a) and natural disaster property damages (b). Figure 1.10 compares Hill plots for flood damages (a) and flood damages per income (b). The difference between these last two plots underscores the importance of properly accounting for exposure. Figure 1.9 (a) is more difficult to interpret than the mean excess plot in Figure (1.7)(b).

Hospital discharge billing data are shown in Figure 1.11; a mean excess plot (a), a mean excess plot after aggregation by 10 (b), and a Hill plot (c). The hospital billing data are a good example of a modestly heavy tailed data set. The mean excess plot and Hill plots point to a tail index in the neighborhood of 3. Although definitely heavy tailed according to all the operative definitions, it behaves like a distribution with finite variance, as we see the mean excess collapse under aggregation by 10.

1.3.6 The Obesity Index

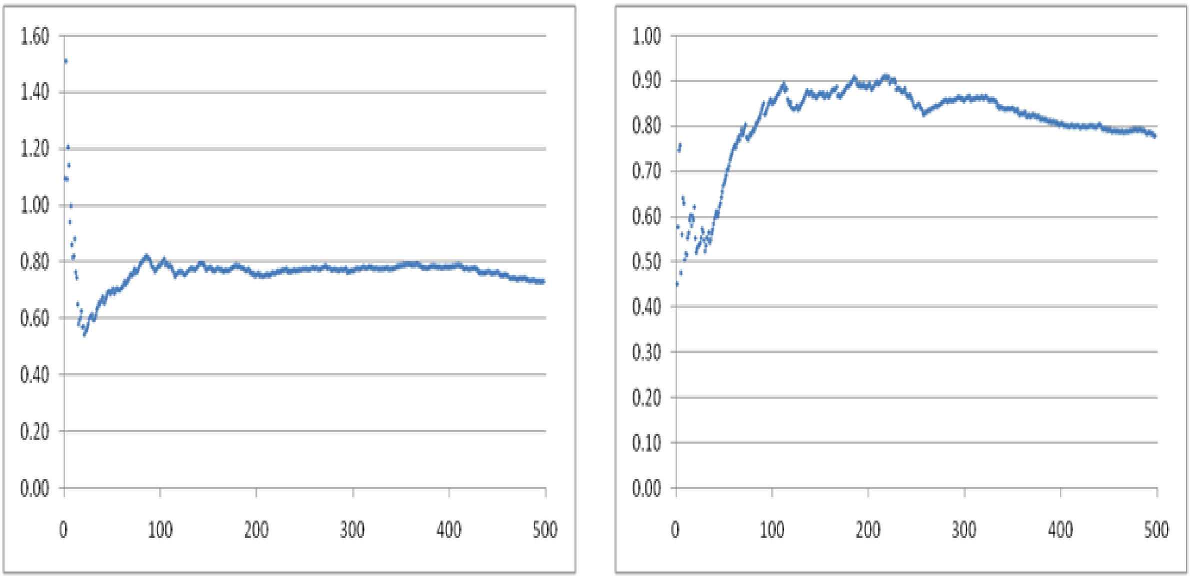
We have discussed two definitions of heavy-tailed distributions, the regularly varying distributions with tail index $0 < \alpha < \infty$ and subexponential distributions. Regularly varying distribu-



(a) Hill plot for crop loss

(b) Hill plot for property damages from natural disasters

Figure 1.9: Hill estimator for crop loss and property damages from natural disasters



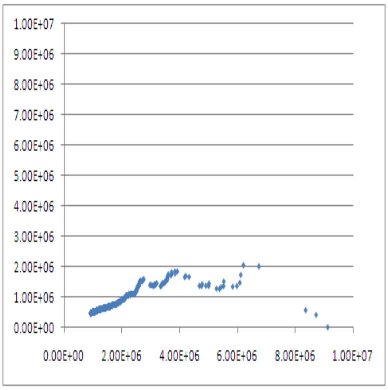
(a) Hill plot for flood claims

(b) Hill plot for flood claims per income

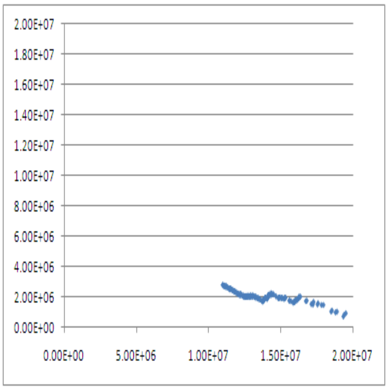
Figure 1.10: Hill estimator for flood claims

tions are a subset of subexponential distributions which have infinite moments beyond a certain point, but subexponentials include distributions all of whose moments are finite (tail index $= \infty$). Both definitions refer to limiting distributions as the value of the underlying variable goes to infinity. Their mean excess plots can be quite similar. There is nothing like a "degree of subexponentiality" allowing us to compare subexponential distributions with infinite tail index, and there is currently no characterization of obesity in finite data sets.

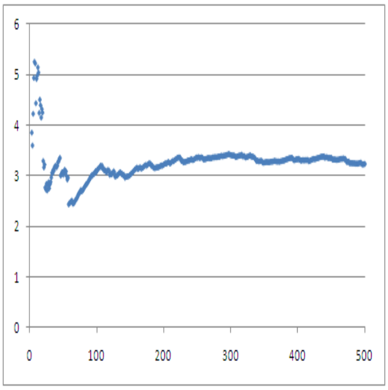
We therefore propose the following *obesity index* that is applicable to finite samples, and which can be computed for distribution functions. Restricting the samples to the higher values



(a) Mean excess plot for hospital discharge bills

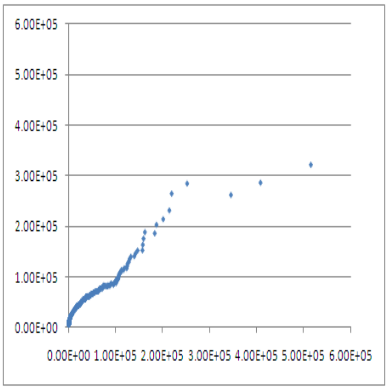


(b) Mean excess plot for hospital discharge bills, aggregation by 10

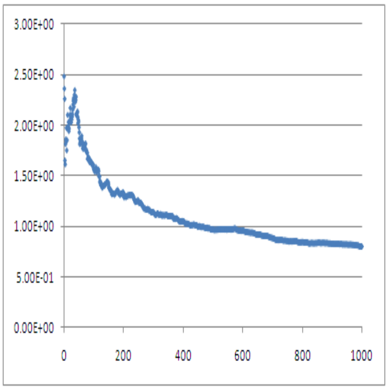


(c) Hill plot for hospital discharge bills

Figure 1.11: Hospital discharge bills, obx = 0.79



(a) Mean excess plot for Gross Cell Product (non mineral)



(b) Hill plot for Gross Cell Product (non mineral)

Figure 1.12: Gross Cell Product (non mineral) obx=0.77

then gives a tail obesity index.

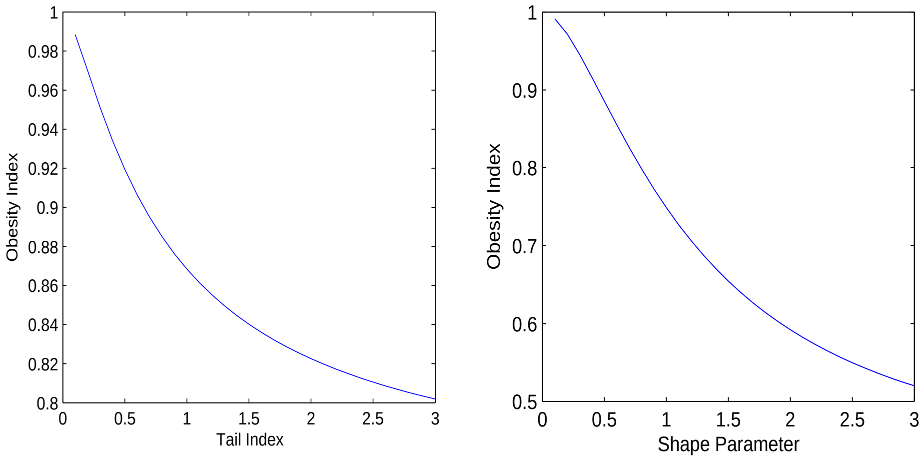
$$\begin{aligned}\mathrm{Ob}(X) &= P\left(X_1 + X_4 > X_2 + X_3 | X_1 \leq X_2 \leq X_3 \leq X_4\right); \\ &\{X_1, ... X_4\}\text{independent and identically disributed.}\end{aligned}$$

In Table 1.2 the value of the Obesity index is given for a number of different distributions. In

Distribution	Obesity index
Uniform	0.5
Exponential	0.75
Pareto(1)	$\pi^2 - 9$

Table 1.2: Obesity index for three distributions

Figure 1.13 we see the Obesity index for the Pareto distribution, with tail index α, and for the Weibull distribution with shape parameter τ.



(a) Pareto distribution (b) Weibull distribution

Figure 1.13: Obesity index for different distributions.

In chapter 5 we show that for the Pareto distribution, the obesity index is decreasing in the tail index. Figures 1.13a and 1.13b illustrate this fact. The same holds for the Weibull distribution, if $\tau < 1$; then the Weibull is a subexponential distribution and is considered heavy-tailed. The Obesity index increases as τ decreases. Note that he Weibull with shape 0.25 is much more obese than the Pareto(1).

Given two random variables X_1 and X_2 with tail indexes, α_1 and α_2, $\alpha_1 < \alpha_2$, the question arises whether the Obesity index of X_1 larger than the Obesity index of X_2. Numerical approximation of two Burr distributed random variables indicate that this is not the case. Consider X_1, a Burr distributed random variable with parameters $c = 1$ and $k = 2$, and a Burr distributed random variable with parameters $c = 3.9$ and $k = 0.5$. The tail index of X_1 is equal to 2 and the tail index of X_2 is equal to 1.95. But numerical approximation indicate that the Obesity index of X_1 is approximately equal to 0.8237 and the Obesity index of X_2 is approximately equal to 0.7463. Of course this should not come as a surprise; the obesity index in this case is applied to the whole distribution, whereas the tail index applies only to the tail.

A similar qualification applies for any distributions taking positive and negative values. For a symmetrical such as the normal or the Cauchy the Obesity index is always $\frac{1}{2}$. The Cauchy distribution is a regularly varying distribution with tail index 1 and the normal distribution is considered a thin-tailed distribution. In such cases it is more useful to apply the Obesity index separately to positive or negative values.

1.4 Conclusion and Overview of the Technical Chapters

Fat tailed phenomena are not rare or exotic, they occur rather frequently in loss data. As attested in hospital billing data and Gross Cell Product data, they are encountered in mundane economic data as well. Customary definitions in terms of limiting distributions, such as regular variation or subexponentiality, may have contributed to the belief that fat tails are mathematical freaks of no real importance to practitioners concerned with finite data sets. Good diagnostics help dispel this incautious belief, and sensitize us to the dangers of uncritically applying thin tailed statistical tools to fat tailed data: Historical averages, even in the absence of time trends may may be poor predictors, regardless of sample size. Aggregation may not reduce variation relative to the aggregate mean, and regression coefficients are based on ratios of quantities that fluctuate wildly.

The various diagnostics discussed here and illustrated with data each have their strengths and weaknesses. Running historical averages have strong intuitive appeal but may easily be confounded by real or imagined time trends in the data. For heavy tailed data, the overall impression may be strongly affected by the ordering. Plotting different moving averages for different random orderings can be helpful. Mean excess plots provide a very useful diagnostic. Since these are based on ordered data, the problems of ordering do not arise. On the downside, they can be misleading for regular varying distributions with tail indices less than or equal to one, as the theoretical slope is infinite. Hill plots, though very popular, are often difficult to interpret. The Hill estimator is designed for regularly varying distributions, not for the wider class of subexponential distributions; but even for regularly varying distributions, it may be impossible to infer the tail index from the Hill plot.

In view of the jumble of diagnostics, each with their own strengths and weaknesses, it is useful to have an intuitive scalar measure of obesity, and the obesity index is proposed here for this purpose. The obesity index captures the idea that larger values are further apart, or that the sum of two samples is driven by the larger of the two, or again that the sum tends to behave like the max. This index does not require estimating a parameter of a hypothetical distribution; in can be computed for data sets and computed, in most cases numerically, for distribution functions.

In Chapter 2 and 3 we discuss different properties of order statistics and present some results from the theory of records. These results are used in Chapter 5 to derive different properties of the index we propose. Chapter 4 discusses and compares regularly varying and subexponential distributions, and develops properties of the mean excess function. Chapter 6 opens a salient toward fat tail regression by surveying dependence concepts.

Chapter 2

Order Statistics

This chapter discusses some properties of order statistic that are used later to derive properties of the Obesity index. Most of these properties can be found in David [1981] or Nezvorov [2001]. Another useful source is Balakrishnan and Stepanov [2007] We consider only order statistics from an i.i.d. sequence of continuous random variables. Suppose we have a sequence of n independent and identically distributed continuous random variables $X_1, ..., X_n$; if we order this sequence in ascending order we obtain the order statistics

$$X_{1,n} \leq ... \leq X_{n,n}.$$

Order statistics are used in Chapter 5 to prove properties of the Obesity index; we therefore step through some fussy proofs. Readers less interested in the proofs of new results may surf this and the following chapters.

2.1 Distribution of order statistics

The distribution function of the r-th order statistic $X_{r,n}$, from a sample of a random variable X with distribution function F, is given by

$$\begin{aligned} F_{r,n}(x) &= P\left(X_{r,n} \leq x\right) \\ &= P\left(\text{at least } r \text{ of the } X_i \text{ are less than or equal to } x\right) \\ &= \sum_{m=r}^{n} P\left(\text{ exactly } m \text{ variables among } X_1, ..., X_n \ \leq \ x\right) \\ &= \sum_{m=r}^{n} \binom{n}{m} F(x)^m \left(1 - F(x)\right)^{n-m} \end{aligned}$$

Using the following relationship for the regularized incomplete Beta function[1]

$$\sum_{m=k}^{n} \binom{n}{m} y^m (1-y)^{n-m} = \int_0^y \frac{n!}{(k-1)!(n-k)!} t^{k-1} (1-t)^{n-k} dt, \qquad 0 \leq y \leq 1,$$

we get the following result

$$F_{r,n}(x) = I_{F(x)}\left(r, n-r+1\right), \tag{2.1}$$

where $I_x(p,q)$ is the regularized incomplete beta function

$$I_x\left(p,q\right) = \frac{1}{B\left(p,q\right)} \int_0^x t^{p-1} (1-t)^{q-1} dt,$$

[1] http://en.wikipedia.org/wiki/Beta_function, accessed Feb.7 2011

and $B(p,q)$ is the beta function

$$B(p,q)=\int_0^1 t^{p-1}(1-t)^{q-1}dt.$$

Now assume that the random variable X_i has a probability density function $f(x)=\frac{d}{dx}F(x)$. Denote the density function of $X_{r,n}$ as $f_{r,n}$. Using (2.1) we get the following result.

$$\begin{aligned} f_{r,n}(x) &= \frac{1}{B(r,n-r+1)}\frac{d}{dx}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}dt, \\ &= \frac{1}{B(r,n-r+1)}F(x)^{r-1}\left(1-F(x)\right)^{n-r}f(x) \end{aligned} \tag{2.2}$$

If $\{k(1),...,k(r)\}$ is a subset of the numbers $1,2,3,...,n$, $k(0)=0$, $k(r+1)=n+1$ and $1\le r\le n$, the joint density of $X_{k(1),n},...,X_{k(r),n}$ is given by

$$\begin{aligned} f_{k(1),...,k(n);n}(x_1,...,x_r) &= \frac{n!}{\prod_{s=1}^{r+1}(k(s)-k(s-1)-1)!} \\ & \prod_{s=1}^{r+1}\left(F(x_s)-F(x_{s-1})\right)^{k(s)-k(s-1)-1}\prod_{s=1}^{r}f(x_s), \end{aligned} \tag{2.3}$$

where $-\infty=x_0<x_1<...<x_r<x_{r+1}=\infty$. We prove this for $r=2$ and assume for simplicity that f is continuous at the points x_1 and x_2 under consideration. Consider the following probability

$$P(\delta,\Delta)=P\left(x_1\le X_{k(1),n}<x_1+\delta<x_2\le X_{k(2),n}<x_2+\Delta\right).$$

We show that as $\delta\to 0$ and $\Delta\to 0$ the following limit holds.

$$f(x_1,x_2)=\lim\frac{P(\delta,\Delta)}{\delta\Delta}$$

Now define the following events

$$\begin{aligned} A &= \{x_1\le X_{k(1),n}<x_1+\delta<x_2\le X_{k(2),n}<x_2+\Delta \text{ and the intervals} \\ & \quad [x_1,x_1+\delta) \text{ and } [x_2,x_2+\Delta) \text{ each contain exactly one order statistic}\}, \\ B &= \{x_1\le X_{k(1),n}<x_1+\delta<x_2\le X_{k(2),n}<x_2+\Delta \text{ and} \\ & \quad [x_1,x_1+\delta)\cup[x_2,x_2+\Delta) \text{ contains at least three order statistics}\}. \end{aligned}$$

We have that $P(\delta,\Delta)=P(A)+P(B)$. Also define the following events

$$\begin{aligned} C &= \{\text{at least two out of } n \text{ variables } X_1,...,X_n \text{ fall into } [x_1,x_1+\delta)\} \\ D &= \{\text{at least two out of } n \text{ variables } X_1,...,X_n \text{ fall into } [x_2,x_2+\Delta)\}. \end{aligned}$$

Now we have that $P(B)\le P(C)+P(D)$. We find that

$$\begin{aligned} P(C) &= \sum_{k=2}^{n}\binom{n}{k}\left(F(x_1+\delta)-F(x_1)\right)^k\left(1-F(x_1+\delta)+F(x_1)\right)^{n-k} \\ &\le \left(F(x_1+\delta)-F(x_1)\right)^2\sum_{k=2}^{n}\binom{n}{k} \\ &\le 2^n\left(F(x_1+\delta)-F(x_1)\right)^2 \\ &= O(\delta^2), \qquad \delta\to 0, \end{aligned}$$

and similarly we obtain that

$$\begin{aligned} P(D) &= \sum_{k=2}^{n} \binom{n}{k} \left(F(x_2+\Delta) - F(x_2)\right)^k \left(1 - F(x_2+\Delta) + F(x_2)\right)^{n-k} \\ &\leq \left(F(x_2+\Delta)\right) \sum_{k=2}^{n} \binom{n}{k} \\ &\leq 2^n \left(F(x_2+\Delta) - F(x_2)\right)^2 \\ &= O\left(\Delta^2\right), \qquad \Delta \to 0. \end{aligned}$$

This yields

$$\lim \frac{P(\delta,\Delta) - P(A)}{\delta\Delta} = 0 \qquad \text{as } \delta \to 0,\ \Delta \to 0.$$

It remains to note that

$$\begin{aligned} P(A) &= \frac{n!}{(k(1)-1)!(k(2)-k(1)-1)!(n-k(2))!} F(x_1)^{k(1)-1} \left(F(x_1+\delta) - F(x_1)\right) \\ &\quad \left(F(x_2) - F(x_1+\delta)\right)^{k(2)-k(1)-1} \left(F(x_2+\Delta) - F(x_2)\right) \left(1 - F(x_2)\right)^{n-k(2)}. \end{aligned}$$

From this equality we see that the limit exists and that

$$\begin{aligned} f(x_1, x_2) &= \frac{n!}{(k(1)-1)!(k(2)-k(1)-1)!(n-k(2))!} F(x_1)^{k(1)-1} \left(F(x_2) - F(x_1)\right)^{k(2)-k(1)-1} \\ &\quad \left(1 - F(x_2)\right)^{n-k(2)} f(x_1) f(x_2), \end{aligned}$$

which is the same as equation (2.3). Note that we have only found the right limit of $f(x_1+0, x_2+0)$, but since f is continuous we can obtain the other limits $f(x_1+0, x_2-0)$, $f(x_1-0, x_2+0)$ and $f(x_1-0, x_2-0)$ in a similar way.

Also note that when $r = n$ in (2.3) we get the joint density of all order statistics and that this joint density is given by

$$f_{1,\dots,n;n}(x_1, \dots, x_n) = \begin{cases} n! \prod_{s=1}^{n} f(x_s) & \text{if, } -\infty < x_1 < \dots < x_n < \infty \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

2.2 Conditional distribution

When we pass from the original random variables $X_1, \dots, X_n$ to the order statistics, we lose independence among these variables. Now suppose we have a sequence of n order statistics $X_{1,n}, \dots, X_{n,n}$, and let $1 < k < n$. In this section we derive the distribution of an order statistic $X_{k+1,n}$ given the previous order statistic $X_k = x_k, \dots, X_1 = x_1$. Let the density of this conditional random variable be denoted by $f(u|x_1, .., x_k)$. We show that this density coincides with the distribution of $X_{k+1,n}$ given that $X_{k,n} = x_k$, denoted by $f(u|x_k)$

$$\begin{aligned} f(u|x_1,...,x_k) &= \frac{f_{1,...,k+1;n}(x_1,...,x_k,u)}{f_{1,...,k;n}(x_1,...,x_k)} \\ &= \frac{\frac{n!}{(n-k-1)!}\left[1-F(u)\right]^{n-k-1}\prod_{s=1}^{k} f(x_s)f(u)}{\frac{n!}{(n-k)!}\left[1-F(x_k)\right]^{n-k}\prod_{s=1}^{k} f(x_s)} \\ &= \frac{\frac{n!}{(k-1)!(n-k-1)!}\left[1-F(u)\right]^{n-k-1}F(x_k)^{k-1}f(x_k)f(u)}{\frac{n!}{(k-1)!(n-k)!}\left[1-F(x_k)\right]^{n-k}F(x_k)^{k-1}f(x_k)} \\ &= \frac{f_{k,k+1;n}(x_k,u)}{f_{k,n}(x_k)} = f(u|x_k). \end{aligned}$$

From this we see that the order statistics form a Markov chain. The following theorem is useful for finding the distribution of functions of order statistics.

Theorem 2.2.1. *Let $X_{1,n} \leq ... \leq X_{n,n}$ be order statistics corresponding to a continuous distribution function F. Then for any $1 < k < n$ the random vectors*

$$X^{(1)} = (X_{1,n},...,X_{k-1,n}) \text{ and } X^{(2)} = (X_{k+1,n},...,X_{n,n})$$

are conditionally independent given any fixed value of the order statistic $X_{k,n}$. Furthermore, the conditional distribution of the vector $X^{(1)}$ given that $X_{k,n} = u$ coincides with the unconditional distribution of order statistics $Y_{1,k-1},...,Y_{k-1,k-1}$ corresponding to i.i.d. random variables $Y_1,...,Y_{k-1}$ with distribution function

$$F^{(u)}(x) = \frac{F(x)}{F(u)} \qquad x < u.$$

Similarly, the conditional distribution of the vector $X^{(2)}$ given $X_{k,n} = u$ coincides with the unconditional distribution of order statistics $W_{1,n-k},...,W_{n-k;n-k}$ related to the distribution function

$$F_{(u)}(x) = \frac{F(x)-F(u)}{1-F(u)} \qquad x > u.$$

Proof. To simplify the proof we assume that the underlying random variables $X_1,...,X_n$ have density f. The conditional density is given by

$$\begin{aligned} f(x_1,...,x_{k-1},x_{k+1},...,x_n|X_{k,n}=u) &= \frac{f_{1,...,n;n}(x_1,...,x_{k-1},x_{k+1,n},...,x_n)}{f_{k;n}(u)} \\ &= \left[(k-1)!\prod_{s=1}^{k-1}\frac{f(x_s)}{F(u)}\right]\left[(n-k)!\prod_{r=k+1}^{n}\frac{f(x_r)}{1-F(u)}\right]. \end{aligned}$$

As we can see the first part of the conditional density is the joint density of the order statistics from a sample size $k-1$ where the random variables have a density $\frac{f(x)}{F(u)}$ for $x < u$. The second part in the density is the joint density of the order statistics from a sample of size $n-k$ where the random variables have a distribution $\frac{F(x)-F(u)}{1-F(u)}$ for $x > u$. $\square$

2.3 Representations for order statistics

We noted that one of the drawbacks of using the order statistics is losing the independence property among the random variables. If we consider order statistics from the exponential distribution or the uniform distribution the following properties can be used to study linear combinations of the order statistics. The first is obvious.

Lemma 2.3.1. *Let $X_{1,n} \leq ... \leq X_{n,n}$, $n = 1, 2, ...$, be order statistics related to independent and identically distributed random variables with distribution function F, and let*

$$U_{1,n} \leq ... \leq U_{n,n},$$

be order statistics related to a sample from the uniform distribution on $[0, 1]$. Then for any $n = 1, 2, ...$ the vectors $(F(X_{1,n}), ..., F(X_{n,n}))$ and $(U_{1,n}, ..., U_{n,n})$ are equally distributed.

Theorem 2.3.2. *Consider exponential order statistics*

$$Z_{1,n} \leq ... \leq Z_{n,n},$$

related to a sequence of independent and identically distributed random variables $Z_1, Z_2, ...$ with distribution function

$$H(x) = \max\left(0, 1 - e^{-x}\right).$$

Then for any $n = 1, 2, ...$ we have

$$(Z_{1,n}, ..., Z_{n,n}) \stackrel{d}{=} \left(\frac{v_1}{n}, \frac{v_1}{n} + \frac{v_2}{n-1}, ..., \frac{v_1}{n} + ... + v_n\right), \tag{2.5}$$

where $v_1, v_2, ...$ is a sequence of independent and identically distributed random variables with distribution function $H(x)$.

Proof. In order to prove Theorem 2.3.2 it suffices to show that the densities of both vectors in (2.5) are equal. Putting

$$f(x) = \begin{cases} e^{-x}, & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

and substituting equation (2.6) into the joint density of the n order statistics given by

$$f_{1,2,...,n;n}(x_1, ..., x_n) = \begin{cases} n! \prod_{i=1}^{n} f(x_i), & x_1 < ... < x_n, \\ 0, & \text{otherwise,} \end{cases}$$

we find that the joint density of the vector on the LHS of equation (2.5) is given by

$$f_{1,2,...,n;n}(x_1, ..., x_n) = \begin{cases} n! \exp\{-\sum_{s=1}^{n} x_s\}, & \text{if } 0 < x_1 < ... < x_n < \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

The joint density of n i.i.d. standard exponential random variables $v_1, ..., v_n$ is given by

$$g(y_1, ..., y_n) = \begin{cases} \exp\{-\sum_{s=1}^{n} y_s\}, & \text{if } y_1 > 0, ..., y_n > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

The linear change of variables

$$(v_1, ..., v_n) = (\frac{y_1}{n}, \frac{y_1}{n} + \frac{y_2}{n-1}, \frac{y_1}{n} + \frac{y_2}{n-1} + \frac{y_3}{n-2}, ..., \frac{y_1}{n} + ... + y_n)$$

with Jacobian $\frac{1}{n!}$ which corresponds to the passage to random variables

$$V_1 = \frac{v_1}{n}, V_2 = \frac{v_1}{n} + \frac{v_2}{n-1}, ..., V_n = \frac{v_1}{n} + ... + v_n,$$

has the property that

$$v_1 + v_2 + ... + v_n = y_1 + ... + y_n$$

and maps the domain $\{y_s > 0 s = 1, ..., n\}$ into the domain $\{0 < v_1 < v_2 < ... < v_n < \infty\}$. Equation (2.8) implies that $V_1, ..., V_n$ have the joint density

$$f(v_1, ..., v_n) = \begin{cases} n! \exp\left\{-\sum_{s=1}^n v_s\right\}, & \text{if } 0 < v_1 < ... < v_n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Comparing equation (2.7) with equation (2.9) we find that both vectors in (2.5) have the same density and this proves the theorem. □

Using Theorem 2.3.2 it is possible to find the distribution of any linear combination of order statistics from an exponential distribution, since we can express this linear combination as a sum of independent exponential distributed random variables.

Theorem 2.3.3. *Let $U_{1,n} \leq ... \leq U_{n,n}$, $n = 1, 2, ...$ be order statistics from an uniform sample. Then for any $n = 1, 2, ...$*

$$(U_{1,n}, ..., U_{n,n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, ..., \frac{S_n}{S_{n+1}}\right),$$

where

$$S_m = v_1 + ... + v_m, \qquad m = 1, 2, ...,$$

and where $v_1, ..., v_m$ are independent standard exponential random variables.

2.4 Functions of order statistics

In this section we discuss different techniques that can be used to obtain the distribution of different functions of order statistics.

2.4.1 Partial sums

Using Theorem 2.2.1 we can obtain the distribution of sums of consecutive order statistics, $\sum_{i=r+1}^{s-1} X_{i,n}$. The distribution of the order statistics $X_{r+1,n}, ..., X_{s-1,n}$ given that $X_{r,n} = y$ and $X_{s,n} = z$ coincides with the unconditional distribution of order statistics $V_{1,n}, ..., V_{s-r-1}$ corresponding to an i.i.d. sequence $V_1, ..., V_{s-r-1}$ where the distribution function of V_i is given by

$$V_{y,z}(x) = \frac{F(x) - F(y)}{F(z) - F(y)}, \qquad y < x < z. \tag{2.10}$$

From Theorem 2.2.1 we can write the distribution function of the partial sum in the following way

$$\begin{aligned} P\left(X_{r+1} + ... + X_{s-1} < x\right) &= \int_{-\infty<y<z<\infty} P\left(X_{r+1} + ... + X_{s-1} < x | X_{r,n} = y, X_{s,n} = z\right) f_{r,s;n}(y,z) dy dz \\ &= \int_{-\infty<y<z<\infty} V_{y,z}^{(s-r-1)*}(x) f_{r,s;n}(y,z) dy dz, \end{aligned}$$

where $V_{y,z}^{(s-r-1)*}(x)$ denotes the $s - r - 1$-th convolution of the distribution given by (2.10).

2.4.2 Ratio between order statistics

Now we look at the distribution of the ratio between two order statistics.

Theorem 2.4.1. *For $r < s$ and $0 \leq x \leq 1$*

$$P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) = \frac{1}{B(s,n-s+1)} \int_0^1 I_{Q_x(t)}(r, s-r) t^{s-1}(1-t)^{n-s} dt, \tag{2.11}$$

where

$$Q_x(t) = \frac{F\left(xF^{-1}(t)\right)}{t}.$$

Proof.

$$\begin{aligned}
P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) &= \int_{-\infty}^{\infty} P\left(\frac{y}{X_{s,n}} \leq x | X_{r,n} = y\right) f_{X_{r,n}}(y) dy, \\
&= \int_{-\infty}^{\infty} P\left(X_{s,n} > \frac{y}{x} | X_{r,n} = y\right) f_{X_{r,n}}(y) dy, \\
&= \int_{-\infty}^{\infty} \int_{\frac{y}{x}}^{\infty} f_{X_{s,n}|X_{r,n}=y}(z) dz f_{X_{r,n}}(y) dy, \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{zx} f_{X_{r,n}}(y) f_{X_{s,n}|X_{r,n}=y}(z) dy dz, \\
&= C \int_{-\infty}^{\infty} \int_{-\infty}^{zx} F(y)^{r-1} \left[1 - F(y)\right]^{n-r} f(y) \\
&\quad \frac{\left[F(z) - F(y)\right]^{s-r-1} \left[1 - F(z)\right]^{n-s} f(z)}{\left[1 - F(y)\right]^{n-r}} dy dz,
\end{aligned}$$

where $C = \frac{1}{B(r,n-r+1)B(s-r,n-s+1)}$. We apply the transformation $t = F(z)$ from which we get the following

$$P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) = C \int_0^1 \int_{-\infty}^{xF^{-1}(t)} F(y)^{r-1} f(y) \left[t - F(y)\right]^{s-r-1} dy \, (1-t)^{n-s} \, dt.$$

Next we use the transformation $\frac{F(y)}{t} = u$.

$$\begin{aligned}
P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) &= C \int_0^1 \int_0^{\frac{F(xF^{-1}(t))}{t}} t^{r-1} u^{r-1} \left(t - tu\right)^{s-r-1} t du \, (1-t)^{n-s} \, dt, \\
&= C \int_0^1 \int_0^{\frac{F(xF^{-1}(t))}{t}} u^{r-1} \left(1 - u\right)^{s-r-1} du t^{s-1} \left(1-t\right)^{n-s} dt
\end{aligned}$$

We can rewrite the constant C in the following way

$$\begin{aligned}
C &= \frac{1}{B(r,n-r+1)B(s-r,n-s+1)} \\
&= \frac{n!}{(r-1)!(n-r)!} \frac{(n-r)!}{(s-r-1)!(n-s)!} \\
&= \frac{1}{(s-r-1)!(r-1)!} \frac{n!}{(n-s)!} \\
&= \frac{(s-1)!}{(s-r-1)!(r-1)!} \frac{n!}{(n-s)!(s-1)!} \\
&= \frac{1}{B(r,s-r)B(s,n-s+1)}.
\end{aligned}$$

If we substitute this in our integral, and define $Q_x(t) = \frac{F(xF^{-1}(t))}{t}$, we get the following

$$\begin{aligned} P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) &= \frac{1}{B(s, n-s+1)} \int_0^1 \frac{\int_0^{Q_x(t)} u^{r-1} (1-u)^{s-r-1} \, du}{B(s, s-r)} t^{s-1} (1-t)^{n-s} \, dt \\ &= \frac{1}{B(s, n-s+1)} \int_0^1 I_{Q_x(t)}(r, s-r) t^{s-1} (1-t)^{n-s} \, dt. \end{aligned}$$

□

This chapter derived properties of order statistics which are used in chapter 5 to derive properties of the obesity index and the distribution of the ratio between order statistics. Because the ratio of order statistics has intuitive appeal in characterizing fat tails, the next chapter reviews the theory of records.

Chapter 3

Records

Records are used in Chapter 5 to explore possible measures of tail obesity. Records are closely related to order statistics. This brief chapter discusses the theory of records and summarizes the main results. For a more detailed discussion see Arnold [1983],Arnold et al. [1998] or Nezvorov [2001], where most of the results we present here can be found. Records are closely related to extreme values and related material can be found in A.J. McNeil and Embrechts [2005], Coles [2001]and Beirlant et al. [2005].

3.1 Standard record value processes

Let $X_1, X_2, ...$ be an infinite sequence of independent and identically distributed random variables. Denote the cumulative distribution function of these random variables by F and assume it is continuous. An observation is called an upper record value if its value exceeds all previous observations. So X_j is an upper record if $X_j > X_i$ for all $i < j$. We are also interested in the times at which the record values occur. For convenience assume that we observe X_j at time j. The record time sequence $\{T_n, n \geq 0\}$ is defined as

$$T_0 = 1 \text{ with probability } 1$$

and for $n \geq 1$,

$$T_n = \min\left\{j : X_j > X_{T_{n-1}}\right\}.$$

The record value sequence $\{R_n\}$ is then defined by

$$R_n = X_{T_n}, \qquad n = 0, 1, 2, ...$$

The number of records observed at time n is called the record counting process $\{N_n, n \geq 1\}$ where

$$N_n = \{\text{number of records among } X_1, ..., X_n\}.$$

We have that $N_1 = 1$ since X_1 is always a record.

3.2 Distribution of record values

Let the record increment process be defined by

$$J_n = R_n - R_{n-1}, \qquad n > 1,$$

with $J_0 = R_0$. It can easily be shown that if we consider the record increment process from a sequence of i.i.d. standard exponential random variables then all the J_n are independent and

J_n has a standard exponential distribution. Using the record increment process we are able to derive the distribution of the n-th record from a sequence of i.i.d. standard exponential distributed variables.

$$\begin{aligned} P\left(R_n < x\right) &= P\left(R_n - R_{n-1} + R_{n-1} - R_{n-2} + R_{n-2} - \ldots + R_1 - R_0 + R_0 < x\right) \\ &= P\left(J_n + J_{n-1} + \ldots + J_0 < x\right) \end{aligned}$$

Since $\sum_{i=0}^{n} J + i$ is the sum of $n+1$ standard exponential distributed random variables we find that the record values from a sequence of standard exponential distributed random variables has the gamma distribution with parameters $n+1$ and 1.

$$R_n \sim \text{Gamma}(n+1, 1), \; n = 0, 1, 2, \ldots$$

A Gamma(n, λ) distributed random variable X has density function

$$f_X(x) = \begin{cases} \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{\Gamma(n)}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

We can use the result above to find the distribution of the n-th record corresponding to a sequence $\{X_i\}$ of i.i.d. random variables with continuous distribution function F. If X has distribution function F then

$$H(X) \equiv -\log(1 - F(X))$$

has a standard exponential distribution function. We also have that $X \stackrel{d}{=} F^{-1}(1 - e^{-X^*})$ where X^* is a standard exponential random variable. Since X is a monotone function of X^* we can express the n-th record of the sequence $\{X_j\}$ as a simple function of the n-th record of the sequence $\{X^*\}$. This can be done in the following way

$$R_n \stackrel{d}{=} F^{-1}(1 - e^{-R_n^*}), \qquad n = 0, 1, 2, \ldots$$

Using the following expression of the distribution of the n-th record from a standard exponential sequence

$$P\left(R_n^* > r^*\right) = e^{-r^*} \sum_{k=0}^{n} \frac{(r^*)^k}{k!}, \qquad r^* > 0,$$

the survival function of the record from an arbitrary sequence of i.i.d. random variables with distribution function F is given by

$$P\left(R_n > r\right) = \left[1 - F(r)\right] \sum_{k=0}^{n} \frac{-\left[\log(1 - F(r))\right]^k}{k!}.$$

3.3 Record times and related statistics

The definition of the record time sequence $\{T_n, n \geq 0\}$ was given by

$$T_0 = 1, \text{ with probability } 1,$$

and for $n \geq 1$

$$T_n = \min\{j : X_j > X_{T_{n-1}}\}.$$

In order to find the distribution of the first n non-trivial record times $T_1, T_2, \ldots, T_n$ we first consider the sequence of record time indicator random variables:

$$I_1 = 1 \text{ with probability } 1,$$

and for $n > 1$

$$I_n = \mathbb{1}_{\{X_n > \max\{X_1, \ldots, X_{n-1}\}\}}$$

So $I_n = 1$ if and only if X_n is a record value. We assume that the distribution function F is continuous. It is easily verified that the random variables I_n have a Bernoulli distribution with parameter $\frac{1}{n}$ and are mutually independent. The joint distribution for the first m record times can be obtained using the record indicators. For integers $1 < n_1 < ... < n_m$ we have that

$$\begin{aligned} P\left(T_1 = n_1, ..., T_m = n_m\right) &= P\left(I_2 = 0, ..., I_{n_1 - 1} = 0, I_{n_1} = 1, I_{n_1+1} = 0, ..., I_{n_m} = 0\right) \\ &= \left[(n_1 - 1)(n_2 - 1)...(n_m - 1)n_m\right]^{-1}. \end{aligned}$$

In order to find the marginal distribution of T_k we first review some properties of the record counting process $\{N_n, n \geq 1\}$ defined by

$$\begin{aligned} N_n &= \{\text{number of records among } X_1, ...X_n\} \\ &= \sum_{j=1}^{n} I_j. \end{aligned}$$

Since the record indicators are independent we can immediately write down the mean and the variance for N_n.

$$\begin{aligned} E\left[N_n\right] &= \sum_{j=1}^{n} \frac{1}{j}, \\ \mathrm{Var}(N_n) &= \sum_{j=1}^{n} \frac{1}{j}\left(1 - \frac{1}{j}\right). \end{aligned}$$

We obtain the exact distribution of N_n using the probability generating function. We have the following result.

$$\begin{aligned} E\left[s^{N_n}\right] &= \prod_{j=1}^{n} E\left[s^{I_j}\right] \\ &= \prod_{j=1}^{n}\left(1 + \frac{s-1}{j}\right) \end{aligned}$$

From this we find that

$$P(N_n = k) = \frac{S_n^k}{n!}$$

where S_n^k is a Stirling number of the first kind:

$$(x)_n = \sum_{k=0}^{n} S_n^k x^k,$$

where $(x)_n = x(x-1)(x-2)...(x-n+1)$. The record counting process N_n obeys the central limit theorem.

$$\frac{N_n - \log(n)}{\sqrt{\log(n)}} \xrightarrow{d} N(0,1)$$

We can use the information about the record counting process to obtain the distribution of T_k. Note that the events $\{T_k = n\}$ and $\{N_n = k+1, N_{n-1} = k\}$ are equivalent. From this we obtain:

$$\begin{aligned} P\left(T_k = n\right) &= P\left(N_n = k+1, N_{n-1} = k\right) \\ &= P\left(I_n = 1, N_{n-1} = k\right) \\ &= \frac{1}{n}\frac{S_{n-1}^k}{(n-1)!} \\ &= \frac{S_{n-1}^k}{n!}. \end{aligned}$$

We also have asymptotic log-normality for T_k.

$$\frac{\log(T_k) - k}{\sqrt{k}} \xrightarrow{d} N(0,1)$$

3.4 k-records

There are two different sequences that are called k-record values in the literature. We discuss both definitions here. First define the sequence of initial ranks ρ_n given by

$$\rho_n = \#\{j : j \leq n \text{ and } X_n \leq X_j\}, \qquad n \geq 1.$$

We call X_n a Type 1 k-record value if $\rho_n = k$, when $n \geq k$. Denote the sequence that is generated through this process by $\left\{R_n^{(k)}\right\}$. The Type 2 k-record sequence is defined in the following way, let $T_{0(k)} = k$, $R_{0(k)} = X_{n-k+1,k}$ and

$$T_{n(k)} = \min\left\{j : j > T_{(n-1)(k)}, X_j > X_{T_{(n-1)(k)}-k+1, T_{(n-1)(k)}}\right\},$$

and define $R_{n(k)} = X_{T_{n(k)}-k+1}$ as the n-th k-record. Here a k record is established whenever $\rho_n \geq k$. Although the corresponding X_n does not need to be a Type 2 k-record, unless $k = 1$, the observation eventually becomes a Type 2 k-record value. The sequence $\left\{R_{n(k)}, n \geq 0\right\}$ from a distribution F is identical in distribution to a record sequence $\{R_n, n \geq 0\}$ from the distribution function $F_{1,k}(x) = 1 - (1 - F(x))^k$. So all the distributional properties of the record values and record counting statistics do extend to the corresponding k-record sequences.

The difference between the Type 1 and Type 2 k-records can also be explained in the following way. We only observe a new Type 1 k-record whenever an observation is exactly the k-th largest seen yet. We observe a new Type 2 k-record whenever we observe a new value that is larger than the previous k-th largest.

Chapter 4

Regularly Varying and Subexponential Distributions

This chapter develops properties of the two most important classes of heavy tailed distributions, namely the regularly varying and the subexponential distributions, introduced informally in Chapter 1. Regularly varying distributions have tails with polynomial decay; all sufficiently high moments are infinite. Moments of subexponential distributions may be all finite, and yet they may exhibit very fat tail behavior. The goal is to derive diagnostic properties for these classes, the most important of which are related to their mean excess functions. In a nutshell, the mean excess function of a light tailed distribution is non increasing, for subexponential distributions the mean excess function goes to infinity and for regularly varying functions with finite expectation the function is asymptotically linear with slope determined by the polynomial decay rate.

The following notation is used throughout: if $X_1, ..., X_n$ is an *iid* sample, then $S_n = X_1 + ... + X_n$ is the partial sum with distribution function F^{n*}, and $M_n = \max\{X_1, ..., X_n\}$ is the maximum with distribution function F^n. This material is found in Embrechts et al. [1997], Bingham et al. [1987], Rachev [2003]and Resnick [2005].

4.0.1 Regularly varying distribution functions

An important class of heavy-tailed distributions is the class of regularly varying distribution functions.

Definition 4.0.1. *A distribution function F is called regularly varying at infinity with tail index α if:*

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha},\ \alpha \in (0,\infty) \tag{4.1}$$

where the survivor function $\overline{F}(x) = 1 - F(x)$. If equation (4.1) holds with $\alpha = \infty$, then F is rapidly varying. In either case we write $F \in \mathcal{R}_{-\alpha}$, and if F is the distribution function of X, $X \in \mathcal{R}_{-\alpha}$.

The definition of regular variation implies that if $\alpha \in (0,\infty)$ then for every $\epsilon > 1$; there is an $x_\epsilon(t)$ depending on t, such that for all $x \geq x_\epsilon$, $\epsilon t^{-\alpha}\overline{F}(x) > \overline{F}(tx) > \frac{t^{-\alpha}}{\epsilon}\overline{F}(x)$. However, for survivor functions proposition (4.0.1) gives uniform convergence in t for x_ϵ sufficiently large. We first present properties of regularly varying functions, and then develop some properties of regularly varying distributions.

Regularly varying functions

The following results concern regularly varying functions, for details see Bingham et al. [1987].

Definition 4.0.2. *A positive measurable function h on $(0,\infty)$ is regularly varying at infinity with index $\alpha \in \mathbb{R}$ if:*

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^{\alpha}, \qquad t > 1. \tag{4.2}$$

We write $h(x) \in \mathcal{R}_\alpha$. If $\alpha = 0$, $h(x)$ is slowly varying at infinity; if $\alpha = -\infty, h(x)$ is rapidly varying at infinity.

If $h(x) \in \mathcal{R}_\alpha$ then we can rewrite the function $h(x)$ as:

$$h(x) = x^{\alpha} L(x), \tag{4.3}$$

where $L(x)$ is slowly varying. In the case of polynomial decay, convergence is uniform in the following sense:

Proposition 4.0.1. *For $\alpha > 0$, and $h(x) \in \mathcal{R}_{-\alpha}, lim_{x\to\infty} \frac{h(t,x)}{h(x)} = t^{-\alpha}$, uniformly in t for $x \in [b,\infty), b > 0$, $t > 1$.*

Karamata's theorem is an important tool for studying the behavior of regularly varying functions, the 'direct half' is important for this chapter; the converse is omitted.

Theorem 4.0.2. *Suppose $h \in \mathcal{R}_\alpha$, for $\alpha \in \mathcal{R}$, and that h is locally bounded in $[x_0,\infty)$ for some $x_0 \geq 0$. Then*

1. *for $\alpha > -1$,*

$$lim_{x\to\infty} \frac{\int_{x_0}^{x} h(t)dt}{xh(x)} = \frac{1}{1+\alpha},$$

2. *for $\alpha < -1$*

$$lim_{x\to\infty} \frac{\int_{x}^{\infty} h(t)dt}{xh(x)} = -\frac{1}{1+\alpha},$$

Properties of regularly varying distribution functions

The regularly varying distributions have attractive properties. This class is closed under convolutions as can be found in Applebaum [2005], where the result was attributed to G. Samorodnitsky.

Theorem 4.0.3. *If X and Y are independent real-valued random variables with $\overline{F}_X \in \mathcal{R}_{-\alpha}$ and $\overline{F}_Y \in \mathcal{R}_{-\beta}$, where $\alpha, \beta > 0$, then $\overline{F}_{X+Y} \in \mathcal{R}_\rho$, where $\rho = \min\{\alpha, \beta\}$.*

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

Proposition 4.0.4. *If F_1 and F_2 are two distribution functions such that as $x \to \infty$*

$$1 - F_i(x) \sim x^{-\alpha} L_i(x) \tag{4.4}$$

with L_i slowly varying, then the convolution $G = F_1 * F_2$ satisfies, as $x \to \infty$:

$$1 - G(x) \sim x^{-\alpha} \left(L_1(x) + L_2(x)\right). \tag{4.5}$$

From Proposition 4.0.4 we obtain the following result using induction on n.

Corollary 4.0.1. *If $\overline{F}(x) = x^{-\alpha}L(x)$ for $\alpha \geq 0$ and $L \in \mathcal{R}_0$, then for all $n \geq 1$,*

$$\overline{F^{n*}}(x) \sim n\overline{F}(x), \qquad \text{as } x \to \infty, \tag{4.6}$$

$$P(S_n > x) \sim P(M_n > x) \qquad \text{as } x \to \infty. \tag{4.7}$$

Proof (Embrechts et al. [1997]) (4.6) follows directly from Proposition 4.0.4. To prove (4.7), for all $n \geq 2$ we have:

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

Thus, we have

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

□

To establish moment properties of regularly varying distributions, we first need:

Lemma 4.0.5. *Let X be a positive random variable, and let g be an increasing differentiable function with $g(0) = 0$ such that $\infty > E(g(X))$; then*

1. $lim_{x\to\infty} g(x)\overline{F(x)} = 0$
2. $E(g(X)) = \int_0^\infty (dg/dx)\overline{F(x)}dx$.

Proof

1. $\int_0^\infty g(x)d\overline{F(x)} - \int_0^{x_o} g(x)d\overline{F(x)} \geq g(x_o)\overline{F(x_o)} \geq 0$, since g is increasing from zero. The LHS goes to zero as $x_o \to \infty$.
2. Using the above and integrating by parts, $0 = g(x)\overline{F(x)}|_0^\infty = \int_0^\infty (dg/dx)\overline{F(x)}dx - \int_0^\infty g(x)dF(x)$. Since $E(g(X)) < \infty$, the terms on the RHS are finite and equal.

□

Proposition 4.0.6. *Let $X \in \mathcal{R}_{-1}$, and $\beta > 0$:*

1. *if $\beta < 1$ then $E(X^\beta) < \infty$*
2. *if $\beta > 1$ then $E(X^\beta) = \infty$*

Proof: It is easy to check that $Y = X^\beta \in \mathcal{R}_{-1/\beta}$.

1. For Y, we are in case (2) of Theorem (4.0.2). Using proposition (4.0.1), for any $\epsilon > 0$, $t > 1$ and y_o sufficiently large:

$$(ty_o)\overline{F}_Y(ty_o) < y_o t^{(\beta-1)/\beta}\overline{F}_Y(y_o)(1+\epsilon).$$

It follows that $lim_{y\to\infty} y\overline{F}_Y(y) = 0$, and by Theorem (4.0.2), $lim_{y\to\infty} \int_y^\infty \overline{F}_y dy \to 0$. Since $\overline{F}_Y \leq 1, E(Y) = \infty$ only if $\int_{y_o}^\infty \overline{F}_y dy = \infty$ for all y_o; therefore $E(Y) < \infty$.

2. If $\beta > 1$ then, by a similar argument

$$\infty = lim_{t\to\infty} y_o t^{(\beta-1)/\beta}\overline{F}_Y(y_o)(1-\epsilon) \leq lim_{t\to\infty}(ty_o)\overline{F}_Y(ty_o).$$

With lemma(4.0.5)(1), this contradicts $E(Y) < \infty$.

□

If $X \in R_{-1}$ then $E(X)$ can be either finite or infinite, depending on the slowly varying function L in equation (4.3).

Remark 4.0.1. *The following facts are easily checked:*

- *If $X \in \mathcal{R}_{-\alpha}$ then $E(X^\beta) < (=)\infty$ if $\beta < (>)\alpha$. If $\beta = \alpha$, then $E(X^\beta)$ can be either finite or infinite.*
- *For $\alpha = 1$ the Pareto(1) is an example with infinite mean.*
- *For an example of $X \in \mathcal{R}_{-\alpha}$ with $E(X^\alpha) < \infty$, consider the distribution function F on (e, ∞) with the density $f(t) = Ct^{-\alpha-1}(\ln(t))^{-s-1}, \alpha > 0, s > 0$, where C is the appropriate constant. By L'Hopital's rule, $\frac{\overline{F}(tx)}{\overline{F}(x)} \sim \frac{tf(xt)}{f(x)} \to t^{-\alpha}$ as $x \to \infty$. On the other hand $\int_e^\infty t^\alpha f(t)dt = \left[-\frac{1}{s}(\ln(t))^{-s}\right]_e^\infty = \frac{1}{s} < \infty$.* [1]
- *if $\alpha = 0, s > 0$ in the above example then the corresponding random variable $X \in R_0$. $\overline{F}(x) = (C/s)(ln(x))^{-s}$. For all $b > 0, E(X^b) = \infty$.*
- *if $\alpha = 0, s = 0$ then $f(x) = c(xln(x))^{-1}, x \geq e$ is not a density, as its integral $ln(ln(x))$ is infinite on this domain.*
- *$\overline{F}(x) = e(xln(x))^{-1}$, $x \geq e$, is a survivor function for $X \in R_{-1}; lim_{x\to\infty} x\overline{F}(x) = 0$, but $\int_e^\infty \overline{F}(x)dx = eln(ln(x))|_e^\infty = \infty$. This shows that the converse of lemma 4.0.5(1) is false.*
- *If X is exponential, then $1/X \in R_{-1}$.*
- *If X is normal with mean μ and variance σ^2 then $1/X$ R_{-1}; if $\mu = 0, E(1/X) = 0$.*

Other distribution functions in the class of regularly varying distributions are given in Table 4.1.

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

Table 4.1: Regularly varying distribution functions

The tail index of a regularly varying distribution gives a convenient characterization of the degree of tail heaviness; if $X \in \mathcal{R}_{-\alpha}$ and $Y \in \mathcal{R}_{-\beta}$ with $\beta < \alpha$ then Y is heavier tailed than X.

[1] We are grateful to Prof. Misiewicz for these examples.

4.0.2 Subexponential distribution functions

The class of subexponential distribution functions contains the class of regularly varying distribution functions and captures the feature that 'the sum behaves like the max'. This section develops several properties of distributions with subexponential tails.

Definition 4.0.3. *A distribution function F with support $(0, \infty)$ is a subexponential distribution, written $F \in \mathcal{S}$, if for all $n \geq 2$,*

$$\lim_{x \to \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n. \tag{4.8}$$

Note that, by definition, $F \in \mathcal{S}$ entails that F is supported on $(0, \infty)$. Whereas regular variation entails that the sum of independent copies is asymptotically disributed as the maximum, from equation (4.8) we see that this fact characterizes the subexponential distributions:

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

Noting that $t^{-\alpha} \sim \frac{\overline{F}(ln(tx))}{\overline{F}(ln(x))}$ if and only if $\frac{\overline{F}(ln(t)+ln(x))}{\overline{F}(ln(x))} \sim e^{-\alpha ln(t)}$, we find the following is a useful property:

Lemma 4.0.7. *A distribution function F with support $(0, \infty)$ satisfies*

$$\frac{\overline{F}(z + x)}{\overline{F}(z)} \sim e^{-\alpha x}, \qquad \text{as } z \to \infty, \;\; \alpha \in [0, \infty] \tag{4.9}$$

if and only if

$$\overline{F} \circ ln \in \mathcal{R}_{-\alpha} \tag{4.10}$$

.

The first equation is equivalent to $\frac{\overline{F}(x-y)}{\overline{F}(x)} \sim e^{\alpha y}, \qquad \text{as } x \to \infty.$

In order to check if a distribution function is a subexponential distribution we do not need to check equation (4.8) for all $n \geq 2$. Lemma 4.0.8 gives a sufficient condition for subexponentiality.

Lemma 4.0.8. *If*

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

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

Lemma 4.0.9 gives a few important properties of subexponential distributions, (Embrechts et al. [1997]).

Lemma 4.0.9. *1. If $F \in \mathcal{S}$, then uniformly in compact y-sets of $(0, \infty)$,*

$$\lim_{x \to \infty} \frac{\overline{F}(x - y)}{\overline{F}(x)} = 1. \tag{4.11}$$

2. If (4.11) holds then, for all $\varepsilon > 0$,

$$e^{\varepsilon x} \overline{F}(x) \to \infty, \qquad x \to \infty$$

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

$$\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq K(1 + \varepsilon)^n, \qquad x \geq 0. \tag{4.12}$$

Proof. The proof of the first statement involves an interesting technique. If $X_1, \dots X_{n+1}$ are positive *i.i.d.* variables, then

$$\begin{aligned}
F_{n+1}(x) - F^{(n+1)*}(x) &= P\{\bigcup_{t \le x}\{\omega | X_{n+1}(\omega) = t\ ;\ \sum_{i=1}^{n} X_i(\omega) > x - t\}\} \\
&= \int_0^x \overline{F^{n*}}(x-t)dF(t). \\
\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} &= \frac{1 - F^{2*}(x)}{\overline{F}(x)} \\
&= \frac{\overline{F}(x) + F(x) - F^{2*}(x)}{\overline{F}(x)} \\
&= 1 + \frac{\int_0^x \overline{F}(x-t)dF(t)}{\overline{F}(x)} \\
&= 1 + \frac{\int_0^y \overline{F}(x-t)dF(t)}{\overline{F}(x)} + \frac{\int_y^x \overline{F}(x-t)dF(t)}{\overline{F}(x)}.
\end{aligned}$$

Since $\frac{\overline{F}(x-t)}{\overline{F}(x)} > 1$ and $\frac{\overline{F}(x-t)}{\overline{F}(x)} > \frac{\overline{F}(x-y)}{\overline{F}(x)}$ we have

$$\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} \ge 1 + F(y) + \frac{\overline{F}(x-y)}{\overline{F}(x)}(F(x) - F(y)).$$

Re-arranging gives

$$\frac{\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} - 1 - F(y)}{F(x) - F(y)} \ge \frac{\overline{F}(x-y)}{\overline{F}(x)} \ge 1.$$

The proof of the first statement concludes by using (4.0.8) to show that the left hand side converges to 1 as $x \to \infty$. Uniformity follows from monotonicity in y. For the second statement, note that by lemma (4.0.7) $\overline{F} \circ ln \in \mathcal{R}_0$, which implies that $x^{\varepsilon}\overline{F}(ln(x)) \to \infty$. The third statement is a fussy calculation for which we refer the reader to (Embrechts et al. [1997] p.42). □

Note that if $F \in \mathcal{S}$, then $\alpha = 0$ in lemma (4.0.7).

Table 4.2 gives a number of subexponential distributions. Unlike the class of regularly

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

Table 4.2: Distributions with subexponential tails.

varying distributions the class of subexponential distributions is not closed under convolutions, a counterexample was provided in Leslie [1989]. Moreover, there is nothing corresponding to a "degree of subexponentiality".

4.0.3 Related classes of heavy-tailed distributions

For the sake of completeness, this section introduces two other classes of heavy tailed distributions, the dominatedly varying denoted by $\mathcal{D}$ and the long tailed denoted by $\mathcal{L}$

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

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

The relations between these, the regularly varying distribution functions ($\mathcal{R}$) and the subexponential distribution functions ($\mathcal{S}$) are as follows:

1. $\mathcal{R} \subset \mathcal{S} \subset \mathcal{L}$ and $\mathcal{R} \subset \mathcal{D}$,
2. $\mathcal{L} \cap \mathcal{D} \subset \mathcal{S}$,
3. $\mathcal{D} \not\subseteq \mathcal{S}$ and $\mathcal{S} \not\subseteq \mathcal{D}$.

4.1 Mean excess function

The mean excess function is a popular diagnostic tool. For subexponential distribution functions the mean excess function tends to infinity. For regularly varying distributions with tail index $\alpha > 1$ it tends to a straight line with slope $1/(\alpha - 1)$. For the exponential distribution the mean excess function is a constant and for the normal distribution it tends to zero. In insurance $e(u)$ is called the mean excess loss function where it is interpreted as the expected claim size over some threshold u. In reliability theory or in the medical field $e(u)$ is often called the mean residual life function. An accessible discussion is found in Beirlant and Vynckier.

The mean excess function of a random variable X with finite expectation is defined as:

Definition 4.1.1. *Let X be a random variable with right endpoint $x_F \in (0,\infty]$ and $E(X) < \infty$, then*

$$e(u) = E\left[X - u | X > u\right], \qquad 0 \leq u \leq x_F,$$

is the mean excess function of X.

In data analysis one uses the empirical counterpart of the mean excess function which is given by

$$\widehat{e}_n(u) = \frac{\sum_{i=1}^{n} X_{i,n} \mathbb{1}_{X_{i,n}>u}}{\sum_{i=1}^{n} \mathbb{1}_{X_{i,n}>u}} - u.$$

The empirical version is usually plotted against the values $u = x_{i,n}$ for $k = 1, ..., n-1$.

4.1.1 Properties of the mean excess function

For positive random variables the mean excess function can be calculated using the following proposition:

Proposition 4.1.1. *The mean excess function of a positive random variable with survivor function $\overline{F}$ can be calculated as:*

$$e(u) = \frac{\int_u^{x_F} \overline{F}(x)dx}{\overline{F}(u)}, \qquad 0 < u < x_F \leq \infty,$$

where x_F is the endpoint of the distribution function F.

Distribution	Mean excess function
Exponential	$\frac{1}{\lambda}$
Weibull	$\frac{x^{1-\tau}}{\beta\tau}$
Log-Normal	$\frac{\sigma^2 x}{\ln(x)-\mu}(1+o(1))$
Pareto	$\frac{\kappa+u}{\alpha-1}, \qquad \alpha > 1$
Burr	$\frac{u}{\alpha\tau-1}(1+o(1)), \qquad \alpha\tau > 1$
Loggamma	$\frac{u}{\alpha-1}(1+o(1))$

Table 4.3: Mean excess functions of distributions

The mean excess function uniquely determines the distribution.

Proposition 4.1.2. *For any continuous distribution function F with density f supported on $(0, \infty)$, if the mean excess function is everywhere finite, then:*

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\left\{ -\int_0^x \frac{1}{e(u)} du \right\}. \tag{4.13}$$

Proof The hazard rate $r(u) = \frac{f(u)}{\overline{F}(u)}$ determines the distribution via

$$\overline{F}(x) = e^{-\int_0^x r(u)du}.$$

Differentiate both sides of $\overline{F}(u)e(u) = \int_u^{x_F} \overline{F}(x)dx$ to obtain, for some constant A

$$r(u) = \frac{1 + d_u e(u)}{e(u)} \tag{4.14}$$

$$-\int_0^x r(u)du = -\int_0^x \frac{1}{e(u)} du - ln(e(x)) + A \tag{4.15}$$

$$\overline{F}(x) = \frac{e^A}{e(x)} e^{-\int_0^x \frac{1}{e(u)}du}. \tag{4.16}$$

Since $\overline{F}(0) = 1$, it follows that $e^A = e(0)$.□

Table 4.3 gives the first order approximations of the mean excess function for different distribution functions. The popularity of the mean excess function as a diagnostic derives from the following two propositions.

Proposition 4.1.3. *If a positive random variable X has a regularly varying distribution function with tail index $\alpha > 1$, then*

$$e(u) \sim \frac{u}{\alpha - 1}, \qquad as\ x \to \infty.$$

Proof: Since X is positive, proposition 4.1.1 yields

$$e(u) = \frac{\int_u^\infty \overline{F}(x)dx}{\overline{F}(u)} \tag{4.17}$$

Since $\overline{F} \in \mathcal{R}_{-\alpha}$ there exists a slowly varying function $l(x)$ such that

$$\overline{F}(x) = x^{-\alpha} l(x) \tag{4.18}$$

from which we find that

$$e(u) = \frac{\int_u^\infty x^{-\alpha} l(x)dx}{u^{-\alpha} l(u)}. \tag{4.19}$$

From theorem 4.0.2:
$$\frac{\int_u^\infty x^{-\alpha} l(x)dx}{u^{-\alpha} l(u)} \sim \frac{u}{\alpha - 1}, \qquad u \to \infty \square$$

Remark 4.1.1. *A direct calculation with the survivor function of a Pareto distribution with finite mean, $\overline{F}(x) = (\frac{k}{k+x})^\alpha$; $\alpha > 1$, shows that $e(u) = \frac{k+u}{\alpha-1}$. In other words, $e(u)$ is linear with intercept $\frac{k}{\alpha-1}$ and slope $\frac{1}{\alpha-1}$*

Proposition 4.1.4. *Let F be the distribution function of a positive unbounded continuous random variable X with finite mean. If for all $y \in \mathbb{R}$*

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = e^{\gamma y}, \tag{4.20}$$

for some $\gamma \in [0, \infty]$, then
$$\lim_{u\to\infty} e(u) = \frac{1}{\gamma}.$$

Proof. By lemma (4.0.7) and (**??**), $\overline{F} \circ ln \in \mathcal{R}_{-\gamma}$ which implies $\overline{F}(x) \sim x^{-\gamma} L(x)$; $\frac{L(tx)}{L(x)} \sim 1$. We have

$$\begin{aligned} e(u) &= \frac{\int_u^\infty \overline{F}(x)dx}{\overline{F}(u)} \\ &= \int_{e^u}^\infty \frac{z^{-1}\overline{F}(ln(z))dz}{\overline{F}(ln(e^u))} \\ &\sim \int_{e^u}^\infty \frac{z^{-(1+\gamma)}L(z)dz}{e^{-\gamma u}L(e^u)}. \end{aligned}$$

By (4.0.2.2) this is equal to $1/\gamma$ $\square$

Remark 4.1.2. *If $F \in \mathcal{S}$, then equation (4.20) is satisfied with $\gamma = 0$. If a distribution function is subexponential, the mean excess function $e(u)$ tends to infinity as $u \to \infty$. If a distribution function is regularly varying with a tail index $\alpha > 1$ then we know that the mean excess function of this distribution is eventually linear with slope $\frac{1}{\alpha-1}$.*

One drawback of the mean excess function as a diagnostic is that it does not exist for regularly varying distributions with tail index $\alpha < 1$. Since the empirical mean excess function always has finite slope, the plots can be misleading in such cases. Using Remark (**??**) we could search for $1 > \beta > 0$ such that $E(X^\beta) < \infty$. If the mean excess function of X^β tends to a straight line with slope $1/(\gamma - 1)$, then we could infer that $X \in \mathcal{R}_{-\gamma/\beta}$. In Chapter 5 a new diagnostic applicable to all distributions will be proposed.

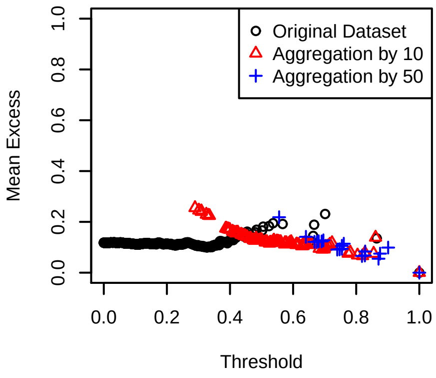
Chapter 5

Indices and Diagnostics of Tail Heaviness

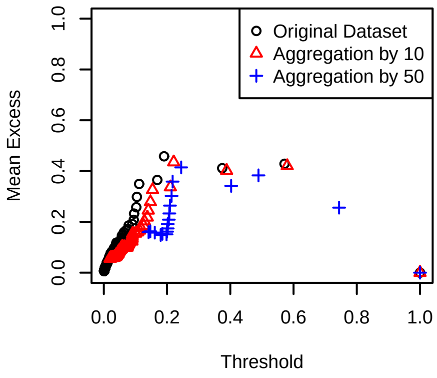
This chapter studies diagnostics for tail obesity using the self-similarity - or lack thereof - of the mean excess plot. We examine how the mean excess plot changes when we aggregate a data set by k. From this we define two new diagnostics; the first is the ratio of the largest to the second largest observation in a data set. The second, termed the Obesity index, is the probability that the sum of the largest and the smallest of four observations is larger than the sum of the other two observations. A note on terminology: *heuristic* suggests a (possibly inaccurate) shortcut, a *diagnostic* is a way of identifying something already defined, a *measure* entails a definition, but not necessarily a method of estimation. *Index* is less precise but suggests all of these. *Tail index* is already preempted, moreover we seek a characterization which applies equally to selected data, eg the largest values, as well as to entire distributions, heavy tailed or otherwise. The term *Obesity index* is enlisted for this purpose. It defines a property of empirical or theoretical distributions and in this sense includes a method of estimation. For the rest, the terms "heuristic" and "diagnostic" are used indiscriminately.

5.1 Self-similarity

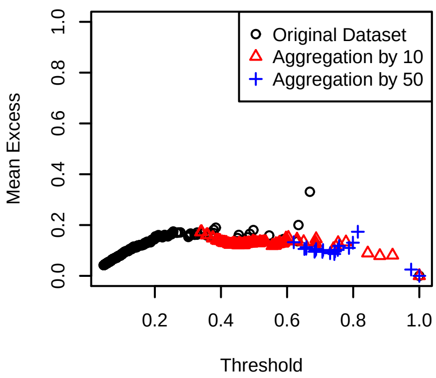
One of the heuristics discussed in Chapter 1 was the self-similarity under aggregation of heavy-tailed distributions and how this could be seen in the mean excess plot of a distribution. Consider a data set of size n and create a new data set by dividing the original data set randomly into groups of size k and summing each of the k members of each group. We call this operation "aggregation by k". If we compare the mean excess plots of regularly varying distribution function with tail index $\alpha < 2$, then the mean excess plot of the original data set and the data set obtained through aggregating by k look very similar. For distributions with a finite variance the mean excess plots of the original sample and the aggregated sample look very different.



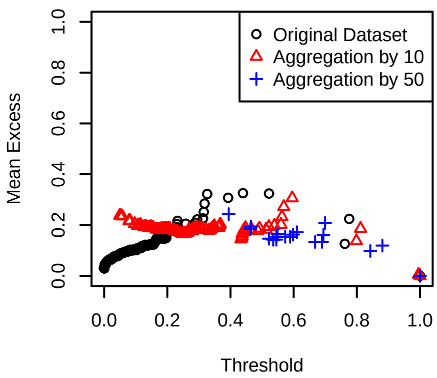
(a) Exponential distribution



(b) Pareto $\alpha = 1$



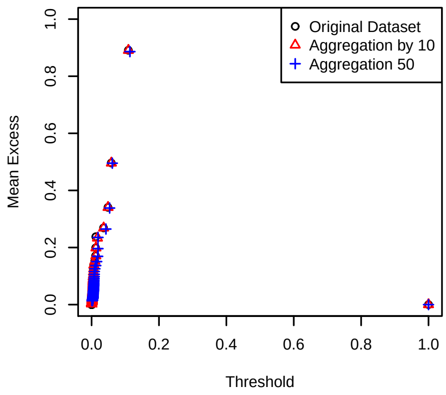
(c) Pareto $\alpha = 2$



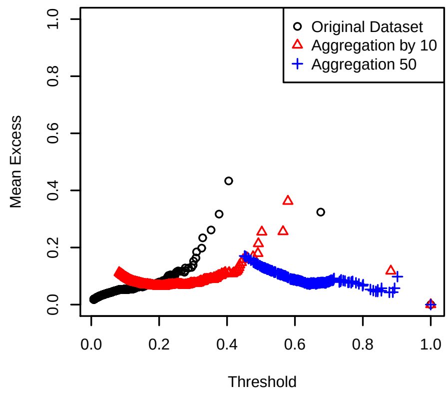
(d) Weibull distribution $\tau = 0.5$

Figure 5.1: Standardized mean excess plots

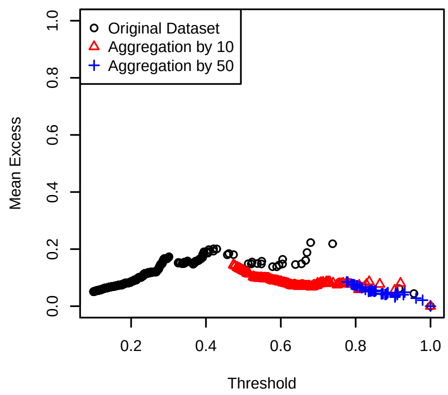
This can be explained through the generalized central limit theorem: normalized sums of regularly varying random variables with a tail index $\alpha < 2$ converge to a stable distribution with the same tail index. If $\alpha > 2$ then the normalized sums converge to a standard normal distribution whose mean excess function tends to zero. In Figures 5.1 we see the standardized mean excess plot of a number of simulated data sets of size 1000. As we can see, the mean excess plots of the exponential data set quickly collapses under random aggregations.The mean excess plot of the Pareto(2) and Weibull data sets collapse more slowly and the mean excess plot of the Pareto(1) does not change much when aggregated by 10; aggregation by 50 leads to a shift in the mean excess plot but the slope stays approximately the same. Of course, aggregation by k is a probabilistic operation, and different aggregations by k will produce somewhat different pictures. Although we might anticipate this behavior from the generalized central limit theorem, it is afterall simply a property of finite sets of numbers. Figures 5.2 (a)–(b) (also in the introduction) are the standardized mean excess plots of the NFIP database and the national crop insurance data. The standardized mean excess plot in Figure 5.2c is based upon a data set that consists of the amount billed to a patient upon discharge. Note that each of the mean excess plots in Figure 5.2 shows some evidence of tail-heaviness since each mean excess plot is



(a) NFIP



(b) National crop insurance



(c) Hospital

Figure 5.2: Standardized mean excess plots of a few data sets

increasing. The NFIP data set shows very heavy-tailed behavior, the other data sets appear less heavy, as the mean excess plot collapses under aggregation. The NFIP data behaves as if it is drawn from a distribution with infinite variance and that the two other data sets behave as if they are drawn from a finite variance distribution.

Denote the largest value in a data set of size n by M_n and the largest value in the data set obtained through aggregation by k by $M_{n(k)}$. By definition $M_n < M_{n(k)}$, but for regularly varying distributions with a small tail index the maximum of the aggregated data set does not differ much from the original maximum. This indicates that M_n is a member of the group which produced $M_{n(k)}$. In general it is quite difficult to calculate the probability that the maximum of a data set is contained in the group that produces $M_{n(k)}$. But we do know that, for positive random variables, whenever the largest observation in a data set is at least k times as large as the second largest observation, then the group that contains M_n produces $M_{n(k)}$. Let us then focus on the distribution of the ratio of the largest to the second largest value in a data set.

5.1.1 Distribution of the ratio between order statistics

In Theorem 2.4.1 we derived the distribution of the ratio between two order statistics in the general case given by

$$P\left(\frac{X_{r,n}}{X_{s,n}} \leq x\right) = \frac{1}{B(s,n-s+1)} \int_0^1 I_{Q_x(t)}(r,s-r)t^{s-1}(1-t)^{n-s}dt, \qquad (r<s) \tag{5.1}$$

where $B(x,y)$ is the beta function, $I_x(r,s)$ the incomplete beta function and $Q_x(t) = \frac{F(tF^{-1}(x)}{t}$. We are interested in the case that $r = n-1$ and $s = n$ so the distribution function in equation (5.1) simplifies to

$$P\left(\frac{X_{n-1,n}}{X_{n,n}} \leq x\right) = n(n-1)\int_0^1 I_{Q_x(t)}(n-1,1)t^{n-1}dt.$$

It turns out there is a much simpler form for the distribution function of the ratio of successive order statistics from a Pareto distribution.

Proposition 5.1.1. *When $X_{1,n},...,X_{n,n}$ are order statistics from a Pareto distribution then the ratio between two consecutive order statistics, $\frac{X_{i+1,n}}{X_{i,n}}$, also has a Pareto distribution with parameter $(n-i)\alpha$.*

Proof. The distribution function of $\frac{X_{i+1,n}}{X_{i,n}}$ can be found by conditionalizing on $X_{i,n}$ and using Theorem 2.2.1 to find the distribution of $X_{i+1,n}|X_{i,n} = x$.

$$\begin{aligned}
P\left(\frac{X_{i+1,n}}{X_{i,n}} > z\right) &= \int_1^\infty P\left(X_{i+1,n} > zx | X_{i+1,n} = x\right) f_{X_{i,n}}(x)dx \\
&= \int_1^\infty \left(\frac{1-F(zx)}{1-F(x)}\right)^{n-i} \frac{1}{B(i,n-i+1)} F(x)^{i-1}\left(1-F(x)\right)^{n-i} f(x)dx \\
&= \frac{1}{B(i,n-i+1)} \int_1^\infty (1-F(zx))^{n-i}F(x)^{i-1}f(x)dx \\
&= \frac{1}{B(i,n-i+1)} z^{-(n-i)\alpha} \int_1^\infty x^{-(n-i)\alpha}(1-x^{-\alpha})^{i-1}\alpha x^{-\alpha-1}dx \\
&= z^{-(n-i)\alpha}\frac{1}{B(i,n-i+1)} \int_0^1 u^{n-i}(1-u)^{i-1}du \qquad (u = x^{-\alpha}) \\
&= z^{-(n-i)\alpha}\frac{1}{B(i,n-i+1)} B(i,n-i+1) \\
&= z^{-(n-i)\alpha}.
\end{aligned}$$

□

One might ask whether the converse of proposition 5.1.1 also holds, i.e. if for some k and n the ratio $\frac{X_{k+1,n}}{X_{k,n}}$ has a Pareto distribution, is the parent distribution of the order statistics also a Pareto distribution? That this is not the case is shown by a counter-example of Arnold [1983]: Let Z_1 and Z_2 be two independent $\Gamma(\frac{1}{2},1)$ random variables, and let $X = e^{Z_1-Z_2}$. If one considers a sample of size 2, then we find that X_1 and X_2 are not Pareto distributed, but that the ratio $\frac{X_{2,2}}{X_{1,2}}$ does have a Pareto distribution.

One needs to make additional assumptions, for example, that the ratio of two successive order statistics have a Pareto distribution for all n, as was shown in H.J.Rossberg [1972]. Here we will give a different proof of this result. The following lemma is needed to prove the result[1].

[1]Result was found on 1 February 2010 at http://at.yorku.ca/cgi-bin/bbqa?forum=ask_an_analyst_2006;task=show_msg;msg=1091.0001

Lemma 5.1.2. *If $f(x)$ is a continuous function on $[0,1]$, and if for all $n \geq 0$*

$$\int_0^1 f(x)x^n dx = 0, \tag{5.2}$$

then $f(x)$ is equal to zero.

Proof. Since equation (5.2) holds, we know that for any polynomial $p(x)$ the following equation holds

$$\int_0^1 f(x)p(x)dx = 0.$$

From this we find that for any polynomial $p(x)$

$$\begin{aligned}\int_0^1 f(x)^2 dx &= \int_0^1 ([f(x) - p(x)]\, f(x) + f(x)p(x))dx \\ &= \int_0^1 [f(x) - p(x)]\, f(x)dx.\end{aligned}$$

Since $f(x)$ is a continuous function on $[0,1]$ we find by the Weierstrass theorem that for any $\varepsilon > 0$ there exists a polynomial $P(x)$ such that

$$\sup_{x\in[0,1]} |f(x) - P(x)| < \varepsilon.$$

By the Min-Max theorem there exists a constant M such that $|f(x)| \leq M$ for all $0 \leq x \leq 1$. From this we find that for any $\varepsilon > 0$ there exists a polynomial $P(x)$ such that

$$\begin{aligned}|\int_0^1 f(x)^2 dx| &= |\int_0^1 [f(x) - P(x)]\, f(x)dx|, \\ &\leq \int_0^1 |f(x) - P(x)||f(x)|dx, \\ &\leq \varepsilon M. \end{aligned} \tag{5.3}$$

But since equation (5.3) holds for all $\varepsilon > 0$ we find that

$$\int_0^1 f(x)^2 dx = 0. \tag{5.4}$$

Since f is continuous on $[1,0]$, it follows that $f(x) = 0, x \in [0,1]$ □

Theorem 5.1.3. *For positive continuous random variable X with invertible distribution function F, if there exists $\alpha > 0$ such that for all $n \geq 2$ and all $x > 1$:*

$$P\left(\frac{X_{n,n}}{X_{n-1,n}} > x\right) = x^{-\alpha}, \tag{5.5}$$

then for some $\kappa > 0$

$$F(x) = 1 - \left(\frac{\kappa}{x}\right)^{\alpha},$$

for $x > \kappa$.

Proof. The two largest of n values can be chosen in $n(n-1)$ ways, thus

$$P(X_{n,n} > X_{n-1,n}x) = n(n-1)\int_0^\infty (1-F(zx))\, F(z)^{n-2} f(z)dz = x^{-\alpha}. \tag{5.6}$$

Since $n(n-1)(1-F(z))F(z)^{n-2}f(z)$ is the density of the $n-1$-th order statistic from a sample of n:

$$n(n-1)\int_0^\infty (1-F(z))F(z)^{n-2}f(z)dz = 1, \tag{5.7}$$

Divide (5.6) by $x^{-\alpha}$ and subtract from (5.7) to find:

$$n(n-1)\int_0^\infty \left(x^\alpha \overline{F}(xz) - \overline{F}(z)\right) F(z)^{n-2} f(z)dz = 0, \tag{5.8}$$

Since equation (5.8) holds for all $n \geq 2$ we can apply lemma 5.1.2 and find that

$$x^\alpha \overline{F}(xz) = \overline{F}(z).$$

This is a variant of the Cauchy equation, whose solution may be written as $\overline{F}(x) = \kappa^\alpha x^{-\alpha}$, $x > \kappa$. □

As in the above proof, the distribution of the ratio between two upper order statistics can be obtained by evaluating the following integral:

$$P\left(\frac{X_{n,n}}{X_{n-1,n}} > x\right) = n(n-1)\int_{-\infty}^\infty (1-F(zx))\, F(z)^{n-2} f(z)dz \tag{5.9}$$

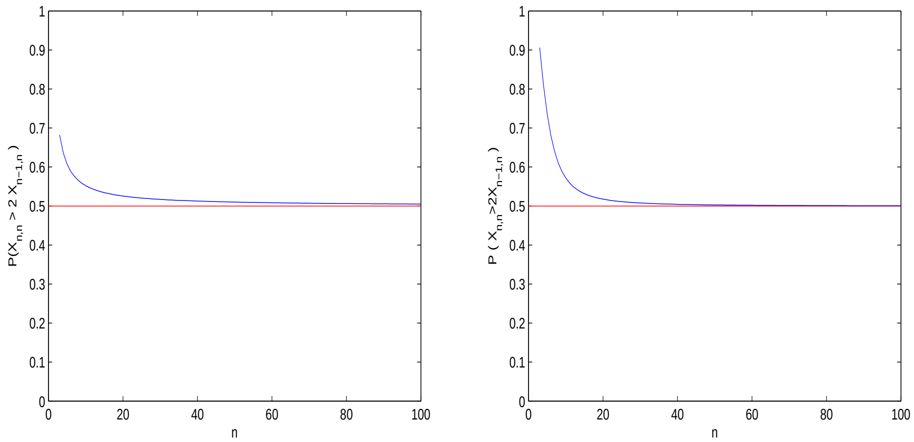
For the Weibull distribution an analytic expression for the integral in equation (5.9) is:

$$\begin{aligned}
P\left(\frac{X_{n,n}}{X_{n-1,n}} > x\right) &= n(n-1)\int_0^\infty (1-F(zx))\, F(z)^{n-2} f(z)dz \\
&= n(n-1)\int_0^\infty \left(e^{-(\lambda x z)^\tau}\right)\left(1-e^{-(\lambda z)^\tau}\right)^{n-2} \tau\lambda^\tau z^{\tau-1} e^{-(\lambda z)^\tau} dz \\
&= n(n-1)\int_0^1 u^{x^\tau}(1-u)^{n-2} du \\
&= n(n-1)B(x^\tau+1, n-1).
\end{aligned}$$

Figures 5.3 (a)–(b) show the approximation of the probability in equation (5.9), with $x = 2$, for the Burr distribution with parameters $c = 1$ and $k = 1$ and the Cauchy distribution. This Burr distribution and the Cauchy distribution both have tail index one. The probability that the largest order statistic is at least twice as large as the second largest order statistic seems to converge to a half. This is exactly the probability that the the ratio of the largest to the second largest order statistic from a Pareto(1) is at least one half, suggesting that the distribution of the ratio of the two largest order statistics of a regularly varying distribution converges to the distribution of that ratio from a Pareto distribution with the same tail index. The next theorem proves this statement. We first recall some results from the theory of regular variation. If the following limit exists for $z > 1$

$$\lim_{x\to\infty} \frac{\overline{F}(zx)}{\overline{F}(x)} = \beta(z) \in [0,1]\,, \quad z > 1$$

then the three following cases are possible:



(a) Burr distribution, $c = 1$, $k = 1$

(b) Cauchy distribution

Figure 5.3: $P(X_{n,n} > 2X_{n-1,n})$ for a few distributions

1. if $\beta(z) = 0$ for all $z > 1$, then $\overline{F}$ is called a rapidly varying function,
2. if $\beta(z) = z^{-\alpha}$ for all $z > 1$, where $\alpha > 0$, then $\overline{F}$ is called a regularly varying distribution function,
3. if $\beta(z) = 1$ for all $z > 1$, then $\overline{F}$ is called a slowly varying function.

The above suggestion is proved in Balakrishnan and Stepanov [2007].

Theorem 5.1.4. *Let F be a distribution function such that $F(x) < 1$ for all x. If $1 - F$ is rapidly varying and $0 < l \leq k$, then*

$$\frac{X_{n-k+l,n}}{X_{n-k,n}} \xrightarrow{P} 1, \qquad (n \to \infty)$$

If $1 - F$ is regularly varying with index $-\alpha$ and $0 < l \leq k$, then

$$P\left(\frac{X_{n-k+l,n}}{X_{n-k,n}} > z\right) \to \sum_{i=0}^{l-1} \binom{k}{i} (1 - z^{-\alpha})^i z^{-\alpha(k-i)}, \qquad (z > 1)$$

If $1 - F$ is a slowly varying distribution function and $0 < l \leq k$, then

$$\frac{X_{n-k+l,n}}{X_{n-k,n}} \xrightarrow{P} \infty, \qquad (n \to \infty).$$

The converse of Theorem 5.1.4 is also true, as was shown in Smid and Stam [1975].

Theorem 5.1.5. *If for some $j \geq 1$, $z \in (0, 1)$ and $\alpha \geq 0$,*

$$\lim_{n\to\infty} P\left(\frac{X_{n-j,n}}{X_{n-j+1,n}} < z\right) = z^{j\alpha} \tag{5.10}$$

then

$$\lim_{y\to\infty} \frac{1 - F\left(\frac{y}{z}\right)}{1 - F(y)} = z^{\alpha}$$

From this theorem we get the following corollary

Corollary 5.1.1. *If (5.10) holds for all $z \in (0,1)$, then $1 - F(x)$ is regularly varying of order $-\alpha$ as $x \to \infty$.*

Theorem 5.1.5 was generalized and extended in Bingham and Teugels [1979].

Theorem 5.1.6. *Let $s \in \{0,1,2,...\}$, $r \in \{1,2,...\}$ be fixed integers. Let F be concentrated on the positive half-line. If $\frac{X_{n-r-s,n}}{X_{n-s,n}}$ converges in distribution to a non-degenerate limit, then for some $\rho > 0$, $1 - F(x)$ varies regularly of order $-\rho$ as $x \to \infty$.*

5.2 The ratio as index

In the previous section we showed that if the ratio between the two largest order statistics converges in distribution to some non-degenerate limit, then the parent distribution is regularly varying. This raises the question whether we can use this as a measure for tail-heaviness of a distribution function. This section considers estimating the following probability:

$$P\left(\frac{X_{n,n}}{X_{n-1,n}} > k\right), \tag{5.11}$$

from a data set. Consider the following estimator: Given a data set of size n, start with $ntrials = 0, nsuccess = 0$. If the third observation is larger than the previous second largest value take $ntrials = ntrials + 1$ and if the largest value is larger than k times the second largest value in the data set take $nsucces = nsucces + 1$. Repeat this until we have observed all values. The estimator of $P\left(\frac{X_{n,n}}{X_{n-1,n}} > k\right)$ is defined by $\frac{nsucces}{ntrials}$. Note that the trials are not independent. We have not proven that this is a consistent estimator, but simulations show that for the Pareto distribution the estimator behaves consistently. Note that $ntrials$ is the number of observed type 2 2-record values, the probability that a new observation is a type 2 2-record value is equal to

$$P\left(X_n > X_{n-2,n-1}\right) = \int_{-\infty}^{\infty} P\left(X > y\right) f_{n-2,n}(y)dy = \frac{2}{n}$$

This means that if we have a data set of size n then the expected number of observed Type 2 2-records equals

$$\sum_{j=3}^{n} \frac{2}{j} = 2\left(\sum_{i=1}^{n} \frac{1}{j} - 1.5\right) \approx 2(\log(n) + \gamma - 1.5).$$

where γ is the Euler-Mascheroni constant and approximately equal to 0.5772. In figure 5.4 we see the expected number of 2-records plotted against the size of the data set. In a data set of size 10000 we only expect to see 16 2-records. Since we do not observe many such records we cannot expect the estimator to be very accurate. We have used the estimate for $P\left(\frac{X_{n,n}}{X_{n-1,n}} > k\right)$ on a number of simulated data sets, all of size 1000. Figures 5.5 (a)–(d) show histograms of this ratio estimator. We calculated the estimator 1000 times by reordering the data set and calculating the estimator for the reordered data set. For the Pareto(0.5) distribution, figure 5.5a shows that on average the estimator seems to be accurate but the estimator ranges from as low as 0.4 to as high as 1. For a Weibull distribution with shape parameter $\tau = 0.5$ we see that the estimate of the probability is much larger than its limit value of zero. This is due to the slow convergence of the probability to zero and to the fact that we expect to see more Type 2 2-records early in a data set. The influence of these initial values takes a very long time to wear off. Table 5.1 summarizes the results of applying the estimators to a Pareto(0.5) distribution,

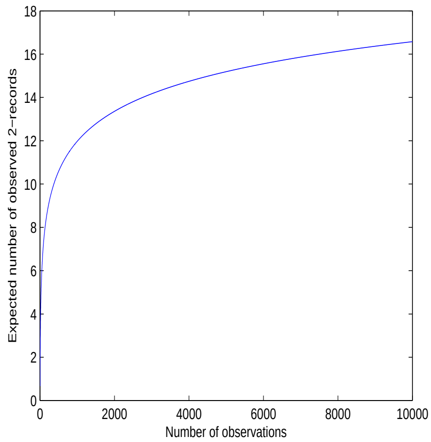


Figure 5.4: Expected number of observed 2-records

Distribution	Expected Value	Mean Estimate
Pareto $\alpha = 0.5$	0.7071068	0.7234412
Pareto $\alpha = 3$	0.125	0.09511546
Weibull $\tau = 0.5$	0	0.3865759
Exponential	0	0.1104443

Table 5.1: Mean estimate of $P\left(\frac{X_{n,n}}{X_{n-1,n}} > 2\right)$

a Pareto(3) distribution, a Weibull distribution with shape parameter $\tau = 0.5$ and a standard exponential distribution. We also applied these estimators to the NFIP data, the national crop insurance data and the hospital data; the results are shown in table 5.2. Figure 5.6a shows that the estimate of the probability $P\left(\frac{X_{n,n}}{X_{n-1,n}} > 2\right)$ suggests more heavy-tailed behavior than the estimate of the probability of the national crop insurance data and the hospital data, a conclusion supported by the mean excess plots of these data sets. We have bootstrapped the data set by reordering the data in order to calculate more than one realization of the estimator. Again, the estimator gives a nice result on average but that the individual values seem to be very spread out. To summarize, a diagnostic based on the relative size of record values is intuitively appealing, but has three serious drawbacks: the consistency of our estimator is unproven, its accuracy on large datasets is poor, and the estimator depends on the order in which the samples are drawn.

Dataset	Mean estimate
NFIP	0.5857
National crop insurance	0.2190
Hospital	0.0882

Table 5.2: Mean estimate of $P\left(\frac{X_{n,n}}{X_{n-1,n}} > 2\right)$

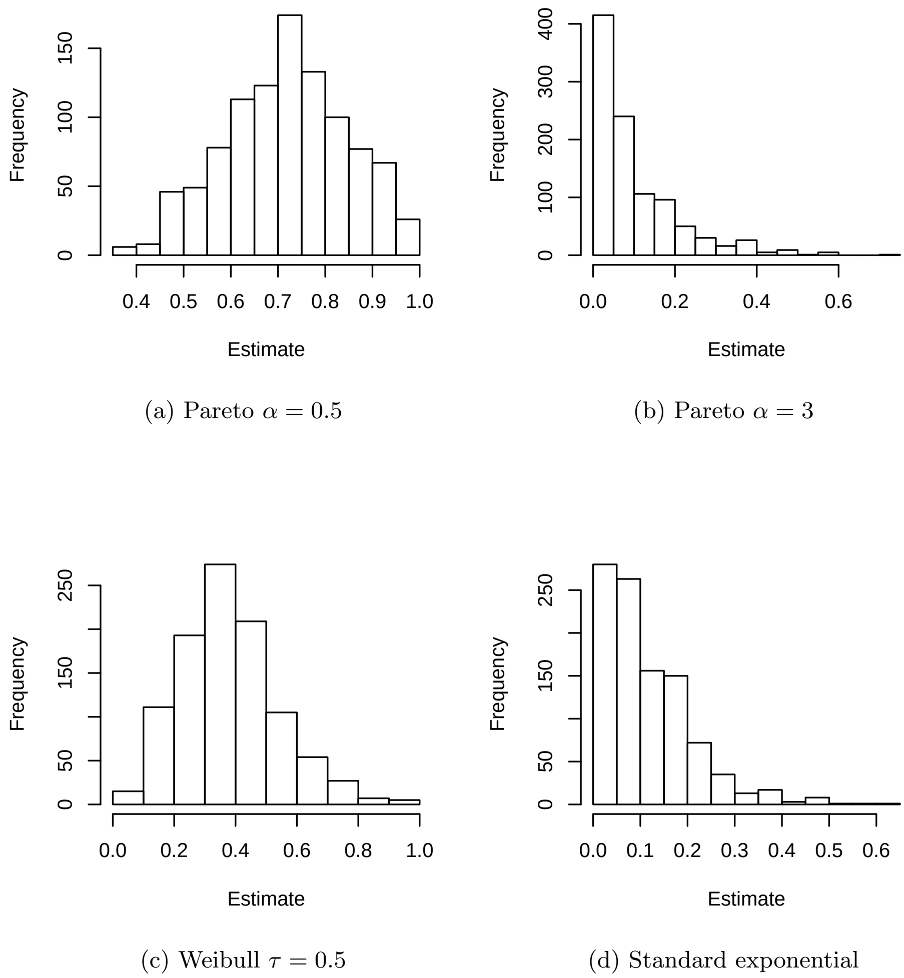


Figure 5.5: Histograms of the estimate of $P\left(\frac{X_{n,n}}{X_{n-1,n}} > x\right)$

5.3 The Obesity Index

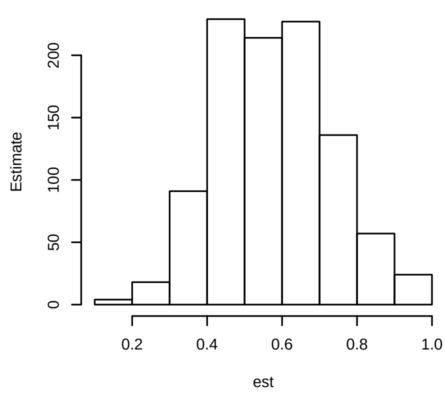
This section opens a different line of attack based on the probability that under aggregation by k, the maximum of the aggregated data set is the sum of the group containing the maximum value of the original data set. Consider aggregation by 2 in a data set of size 4 containing the observation X_1, X_2, X_3, X_4 with $X_1 < X_2 < X_3 < X_4$. By definition we have that $X_4 + X_2 > X_3 + X_1$ and $X_4 + X_3 > X_2 + X_1$, so the only interesting case arises whenever we sum X_4 with X_1. Now define the Obesity index by

$$\text{Ob}(X) = P\left(X_4 + X_1 > X_2 + X_3 \,|\, X_1 \leq X_2 \leq X_3 \leq X_4\right), \qquad X_i iid \text{ copies of } X. \tag{5.12}$$

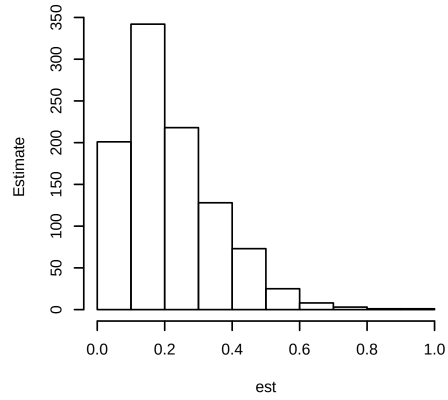
We expect that this probability is larger for heavy-tailed distribution than for thin-tailed distributions. We can rewrite the inequality in the probability in equation (5.12) as:

$$X_4 - X_3 > X_2 - X_1,$$

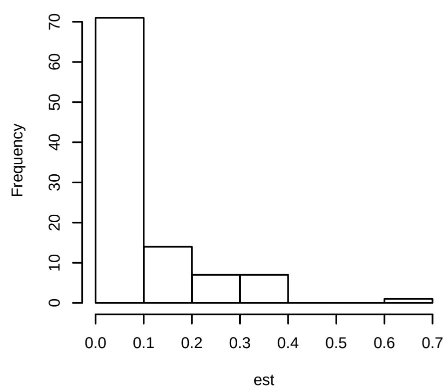
which was one of the heuristics of heavy-tailed distributions we discussed in Chapter 1, i.e. the fact that larger observations lie further apart than smaller observations. Note that the Obesity



(a) NFIP



(b) National crop insurance



(c) Hospital

Figure 5.6: Histograms of the estimate of $P\left(\frac{X_{n,n}}{X_{n-1,n}} > 2\right)$

index is positive affine invariant, that is: $\mathrm{Ob}(aX + b) = \mathrm{Ob}(X)$ for $a > 0$ and $b \in \mathbb{R}$. The Obesity index may be calculated for a finite data set, or for a random variable X by considering independent copies of X in equation (5.12). The following propositions calculate the Obesity index for a number of distributions. First note if $P(X = C) = 1$ where C is a constant, then $Ob(X) = 0$.

Proposition 5.3.1. *The obesity index of the uniform distribution is* $\frac{1}{2}$.

Proof. The obesity index can be rewritten as:

$$P\left(X_4 - X_3 > X_2 - X_1 | X_1 < X_2 < X_3 < X_4\right) = P\left(X_{4,4} - X_{3,4} > X_{2,4} - X_{1,4}\right). \tag{5.13}$$

Using theorem 2.3.3 we can calculate the probability in equation (5.13):

$$P\left(X_{4,4} - X_{3,4} > X_{2,4} - X_{1,4}\right) = P\left(X > Y\right), \tag{5.14}$$

where X and Y are standard exponential random variables. Since the random variables X and Y in equation (5.14) are independent and identically distributed random variables this probability is equal to $\frac{1}{2}$. $\square$

Proposition 5.3.2. *The obesity index of the exponential distribution is $\frac{3}{4}$.*

Proof. Again we rewrite the obesity index:

$$P\left(X_4 - X_3 > X_2 - X_1 | X_1 < X_2 < X_3 < X_4\right) = P\left(X_{4,4} - X_{3,4} > X_{2,4} - X_{1,4}\right). \tag{5.15}$$

Using theorem 2.3.2:

$$P\left(X_{4,4} - X_{3,4} > X_{2,4} - X_{1,4}\right) = P\left(X > \frac{Y}{3}\right), \tag{5.16}$$

where X and Y are independent standard exponential random variables. We can calculate the probability on the RHS in equation (5.16):

$$\begin{aligned} P\left(X > \frac{Y}{3}\right) &= \int_0^\infty P\left(X > y\right) f_{\frac{Y}{3}}(y) dy \\ &= \int_0^\infty e^{-y} 3e^{-3y} dy \\ &= \frac{3}{4}. \end{aligned}$$

□

Proposition 5.3.3. *If X is a symmetrical random variable with respect to zero, $X \stackrel{d}{=} -X$, then the obesity index is equal to $\frac{1}{2}$.*

Proof. If $X \stackrel{d}{=} -X$, then $F_X(x) = 1 - F_X(-x)$, and $f_X(x) = f_X(-x)$. The joint density of $X_{3,4}$ and $X_{4,4}$ is now given by

$$\begin{aligned} f_{3,4;4}(x_3, x_4) &= \frac{24}{2} F(x_3)^2 f(x_3) f(x_4), \quad x_3 < x_4 \\ &= \frac{24}{2} (1 - F(-x_3))^2 f(-x_3) f(-x_4), \quad -x_4 < -x_3 \\ &= f_{1,2;4}(-x_4, -x_3) \end{aligned}$$

This is equal to the joint density of $-X_{1,4}$ and $-X_{2,4}$, and from this we find that

$$X_{4,4} - X_{3,4} \stackrel{d}{=} X_{2,4} - X_{1,4}.$$

Hence the obesity index $\frac{1}{2}$. □

From proposition 5.3.3 we find that for a distribution which is symmetric with respect to some constant μ, the Obesity index is $\frac{1}{2}$. Indeed, if X is symmetric about μ, $X - \mu$ is symmetric about zero and with positive affine invariance, $\mathrm{Ob}(X) = \mathrm{Ob}(X - \mu)$. This means that the Obesity index of both the Cauchy and the Normal distribution is $\frac{1}{2}$. The Cauchy distribution has a regularly varying distribution function with tail index 1, and the Normal distribution is thin-tailed distribution on any definition. Evidently the Obesity index must be restricted to positive random variables.

Theorem 5.3.4. *The Obesity index of a random variable X with distribution function F and density f can be calculated by evaluating the following integral,*

$$24 \int_{-\infty}^{\infty} \int_{x_1}^{\infty} \int_{x_2}^{\infty} \overline{F}(x_2 + x_3 - x_1) f(x_1) f(x_2) f(x_3) dx_3 dx_2 dx_1.$$

Proof. The obesity index can be rewritten as:

$$P\left(X_1 + X_4 > X_2 + X_3 | X_1 \leq X_2 \leq X_3 \leq X_4\right).$$

Recall that the joint density of all n order statistics from a sample of n is:

$$f_{1,2,\ldots,n;n}(x_1, x_2, \ldots, x_n) = \begin{cases} n! \prod_{i=1}^{n} f(x_i), & x_1 < x_2 < \ldots < x_n, \\ 0, & \text{otherwise,} \end{cases}$$

In order to calculate the obesity index we need to integrate this joint density over all numbers such that

$$x_1 + x_4 > x_2 + x_3, \text{ and } x_1 < x_2 < x_3 < x_4.$$

We then must evaluate the following integral.

$$\mathrm{Ob}(X) = 24 \int_{-\infty}^{\infty} f(x_1) \int_{x_1}^{\infty} f(x_2) \int_{x_2}^{\infty} f(x_3) \int_{x_3+x_2-x_1}^{\infty} f(x_4) dx_4 dx_3 dx_2 dx_1$$

The innermost integral is the probability that the random variable X is larger than $x_3 + x_2 - x_1$ so this expression simplifies to

$$\mathrm{Ob}(X) = 24 \int_{-\infty}^{\infty} \int_{x_1}^{\infty} \int_{x_2}^{\infty} \overline{F}(x_2 + x_3 - x_1) f(x_1) f(x_2) f(x_3) dx_3 dx_2 dx_1.$$

□

Using Theorem 5.3.4 we calculate the Obesity index whenever the parameter α is an integer. We have done this using Maple, in Table 5.3 the exact and approximate value of the Obesity index for a number of α are given. From Table 5.3 we can observe that the Obesity index

α	Exact value	Approximate value
1	$\pi^2 - 9$	0.8696
2	$593 - 60\pi^2$	0.8237
3	$\frac{-124353}{5} + 2520\pi^2$	0.8031
4	$\frac{19150997}{21} - 92400\pi^2$	0.7912

Table 5.3: Obesity index of Pareto(α) distribution for integer α

increases as the tail index decreases, as expected. Properties for the Obesity index of a Pareto random variable are derived using the theory of majorization.

5.3.1 Theory of Majorization

The theory of majorization is used to give a mathematical meaning to the notion that the components of one vector are less spread out than the components of another vector whose components have the same mean value.

Definition 5.3.1. *A vector* $\mathbf{y} \in \mathbb{R}^n$ *majorizes a vector* $\mathbf{x} \in \mathbb{R}^n$ *if*

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i,$$

and

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \qquad k = 1, ..., n$$

where $x_{[1]} \ldots x_{[n]}$ *is the decreasing arrangement of the vector* $\mathbf{x}$ *such that*

$$x_{[1]} \geq ... \geq x_{[n]}.$$

We denote this by $\mathbf{x} \prec \mathbf{y}$.

Schur-convex functions preserve the majorization ordering.

Definition 5.3.2. *A function* $\phi : \mathcal{A} \to \mathbb{R}$, *where* $\mathcal{A} \subset \mathbb{R}^n$, *is called Schur-convex (-concave) on* $\mathcal{A}$ *if*

$$\mathbf{x} \prec \mathbf{y} \ \text{on } \mathcal{A} \Rightarrow \phi(x) \leq (\geq) \phi(y) \tag{5.17}$$

The following proposition gives sufficient conditions for a function ϕ to be Schur-convex (concave).

Proposition 5.3.5. *If* $I \subset \mathbb{R}$ *is an interval and* $g : I \to \mathbb{R}$ *is convex (concave), then*

$$\phi(x) = \sum_{i=1}^{n} g(x_i),$$

is Schur-convex (-concave)on I^n.

We prove two theorems about the inequality in the obesity index.

Theorem 5.3.6. *Let* X_α *be Pareto*(α) *distributed, for* $\alpha > 0$. *Then* $lim_{\alpha \to 0} Ob(X_\alpha) = 1$.

Proof. $X(\alpha)$ has the same distribution as $U^{-1/\alpha}$ From Hardy et al. [1934] we know that

$$\lim_{p \to \infty} \left(\sum_{i=1}^{n} u_i^p \right)^{\frac{1}{p}} = \max \{u_1, ..., u_n\}.$$

Suppose that $u_1 < u_2 < u_3 < u_4$, we have

$$\lim_{\alpha \to 0} \left(u_1^{-1/\alpha} + u_4^{-1/\alpha} \right) = \max \left\{ u_1^{-1}, u_4^{-1} \right\}^{\frac{1}{\alpha}} = \max \left\{ u_1^{-\frac{1}{\alpha}}, u_4^{-\frac{1}{\alpha}} \right\}.$$

This means that as α tends to 0, $u_1^{-1/\alpha} + u_4^{-1/\alpha}$ tends to $\max \left\{ u_1^{-1/\alpha}, u_4^{-1/\alpha} \right\}$, which is $u_1^{-1/\alpha}$. The same limit holds for $u_2^{-1/\alpha} + u_3^{-1/\alpha}$, where the maximum of these two by definition is equal to $u_2^{-1/\alpha}$. The proof concludes by substituting $X = U^{-\alpha}$. $\square$

There is no comparable result for general regularly varying distributions since obesity depends on the whole distribution, not just the tail. For example, let X be uniform on $[0, 1]$ with probability $(1/2)^{1/4}$ and be Pareto(1) with probability $1 - (1/2)^{1/4}$. Clearly $X \in \mathcal{R}_{-1}$. However, with probability 1/2 four independent copies of X are all drawn from the uniform distribution with obesity 1/2. Hence the obesity of X cannot exceed 0.75. The second result shows that obesity is increased by applying a convex increasing function.

Lemma 5.3.7. *If* $0 < y_1 < y_2 < y_3 < y_4$ *and*

$$y_4 + y_1 > y_2 + y_3$$

then for any increasing convex function $g : \mathbb{R}_+ \to \mathbb{R}$

$$g(y_4) + g(y_1) > g(y_3) + g(y_2)$$

Proof. Note that $(y_2, y_3) \prec (y_1, y_3 + y_2 - y_1)$. The function

$$\phi(x_1, x_2) = \sum_{i=1}^{2} g(x_i),$$

is Schur-convex on $\mathbb{R}^2$ by Proposition 5.3.5. Equation (5.17) and the fact that g is increasing give:

$$\begin{aligned} \phi(y_1, y_2) &= g(y_1) + g(y_2) \\ &\leq g(y_3 + y_2 - y_1) + g(y_1) \\ &\leq g(y_4) + g(1). \end{aligned}$$

□

Theorem 5.3.8. *For all positive random variables* X *and all* $g : \mathbb{R}_+ \to \mathbb{R}_+$*, convex and increasing,* $Ob(X) < Ob(g(X))$.

Proof. By lemma 5.3.7, since $0 < x_1 < x_2 < x_3 < x_4$ and $x_1 + x_4 > x_2 + x_3$ imply that $0 < g(x_1) < g(x_2) < g_(x_3) < g(x_4)$ and $g(x_1) + g(x_4) > g_(x_2) + g(x_3)$
;it follows that

$$P\{X_4 + X_1 > X_2 + X_3 \,|\, X_1 \leq X_2 \leq X_3 \leq X_4\} \leq$$
$$P\{g(X_4) + g(X_1) > g(x_2) + g(X_3) \,|\, g(X_1) \leq g(X_2) \leq g(X_3) \leq g(X_4)\}.$$

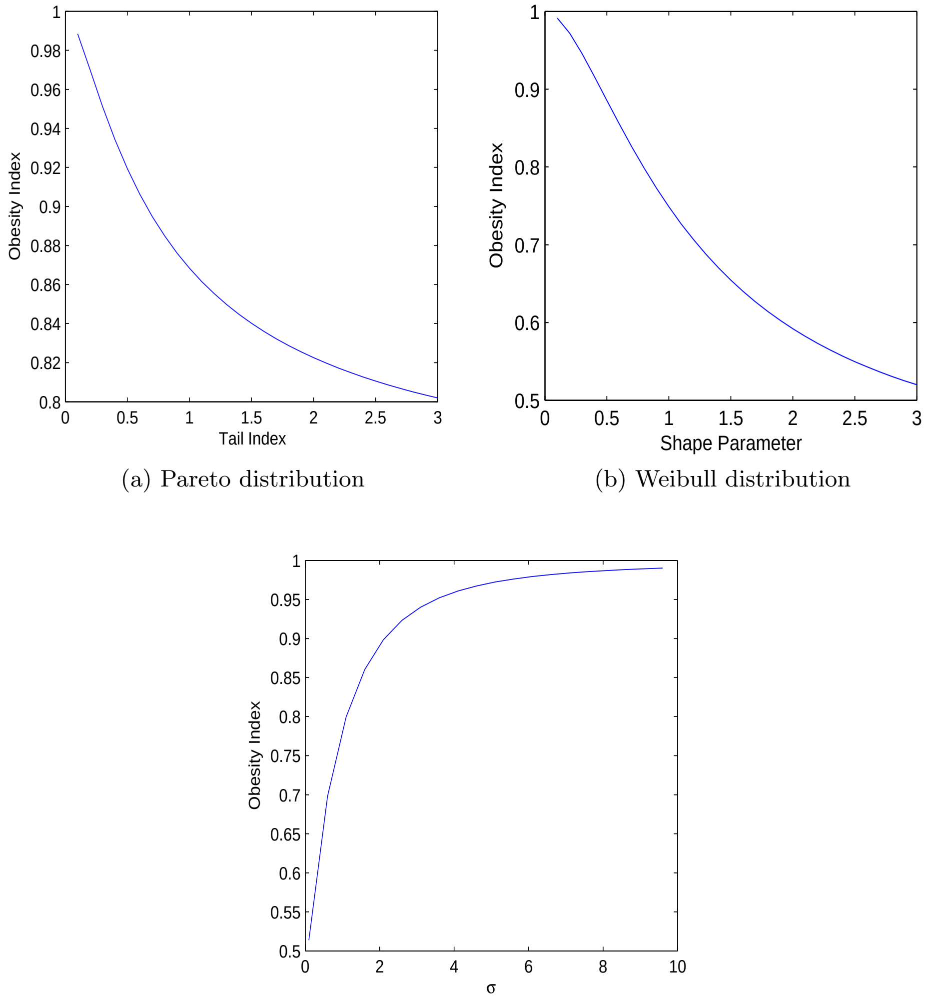
□

Remark 5.3.1. *The following facts are easily checked and show that certain transformations affect obesity and the tail index in similar ways.*

- *The functions* $g(x) = x^\alpha; \alpha > 1, g(x) = e^x, g(x) = \frac{1}{xln(x)}$ *increase obesity.*
- *If* $Ob(X) < Ob(g(X))$ *and* $Ob(X) < Ob(h(X))$*, then* $Ob(h(X)) < Ob(g(X))$ *if* $g \circ h^{-1}$ *is convex increasing.*
- *if If* $X \in \mathcal{R}_{-\alpha}$ *and* $\beta > 1$ *then* $X^\beta \in \mathcal{R}_{-\alpha/\beta}$
- *If* $X \in \mathcal{R}_{-\alpha}$ *then* $e^X \in \mathcal{R}_0$

Using Theorem 5.3.4 we have approximated the obesity index of the Pareto distribution, the Weibull distribution, the Log-normal distribution, the Generalized Pareto distribution and the Generalized Extreme Value distribution. The *Generalized Extreme Value distribution* is defined by

$$F(x; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi} \right\},$$



(a) Pareto distribution

(b) Weibull distribution

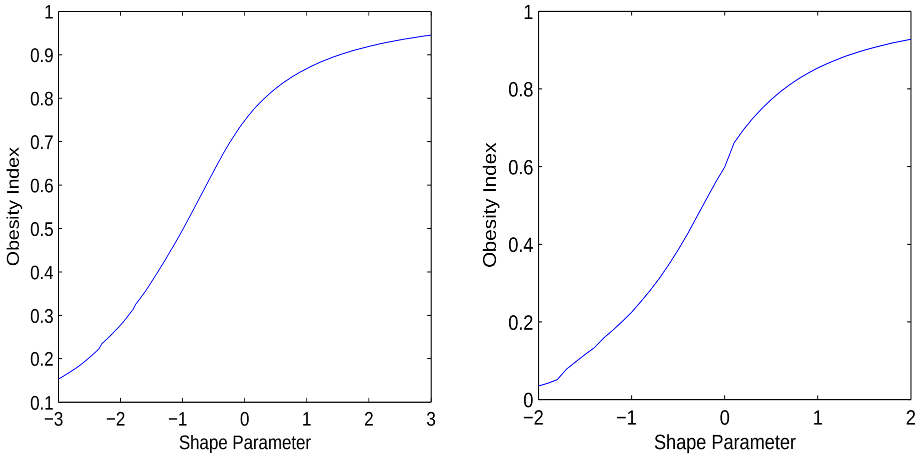
(c) Log-Normal distribution

Figure 5.7: Obesity index for different distributions.

for $1 + \xi(x - \mu)/\sigma > 0$, with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$. In the case $\xi = 0$ the generalized extreme value distribution corresponds to the Gumbel distribution. The *Generalized Pareto distribution* is defined by

$$F(x; \mu, \sigma, \xi) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left\{-\frac{x-\mu}{\sigma}\right\} & \text{for } \xi = 0, \end{cases}$$

for $x \geq \mu$ when $\xi \geq 0$, and $x \leq \mu - \frac{\sigma}{\xi}$ when $\xi < 0$, with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$. If $\xi > 0$ the generalized extreme value distribution and the Generalized Pareto distribution are regularly varying distribution functions with tail index $\frac{1}{\xi}$. As shown in Figures 5.7 and 5.8 the Obesity index of all the distributions considered here behaves nicely. This is due to the fact that if random variable X that has one of these distributions, then X^a also has the same distribution but with different parameters. In these figures we have plotted the Obesity index against the parameter that changes when considering X^a and that cannot be changed through adding a constant to X^a or by multiplying X^a with a constant. The figures demonstrate that a Weibull or lognormal distribution can be much more



(a) Generalized Pareto distribution (b) Generalized extreme value distribution

Figure 5.8: Obesity index for different distributions.

obese than a Pareto, depending on the choice of parameters.

One can ask whether the Obesity index of a regularly varying distribution increases as the tail index of this distribution decreases. The following numerical approximation indicates that the this is not the case in general.

If X has a Pareto distribution with parameter k, then the following random variable has a Burr distribution with parameters c and k

$$Y \stackrel{d}{=} (X-1)^{\frac{1}{c}}.$$

This holds since when X has a Pareto(k) distribution then

$$P\left((X-1)^{\frac{1}{c}} > x\right) = P\left(X > x^c + 1\right) = \left(x^c+1\right)^{-k}.$$

Table 4.1 shows that the tail index of the Burr distribution is equal to ck. This means that the Obesity index of a Burr distributed random variable with parameters k and $c = 1$, equals the Obesity index of a Pareto random variable with parameter k. From this we find that the Obesity index of a Burr distributed random variable X_1 with parameters $c = 1, k = 2$ is equal to $593 - 60\pi^2 \approx 0.8237$. If we now consider a Burr distributed random variable X_2 with parameters $c = 3.9$ and $k = 0.5$ and we approximate the Obesity index numerically we find that the Obesity index of this random variable is approximately equal to 0.7463, which is confirmed by simulations. Although the tail index of X_1 is larger than the tail index of X_2, we have that $\mathrm{Ob}(X_1) > \mathrm{Ob}(X_2)$. In Figure 5.9 the Obesity index of the Burr distribution is plotted for different values of c and k.

5.3.2 The Obesity Index of selected Datasets

In this section we estimate the Obesity index of a number of data sets, and compare the Obesity index and the estimate of the tail index. Table 5.4 shows the estimate of the Obesity index based upon 250 bootstrapped values, and the 95%-confidence bounds of the estimate. From Table 5.4 we get that the NFIP data set is heavier-tailed than the National Crop Insurance data

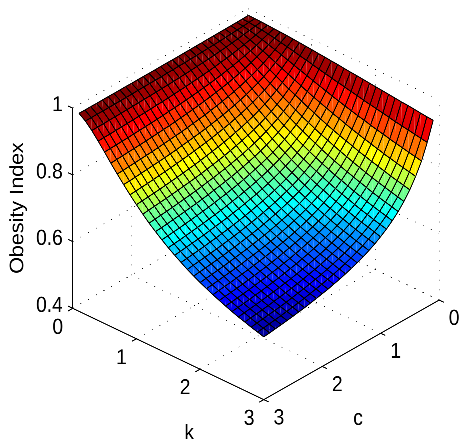
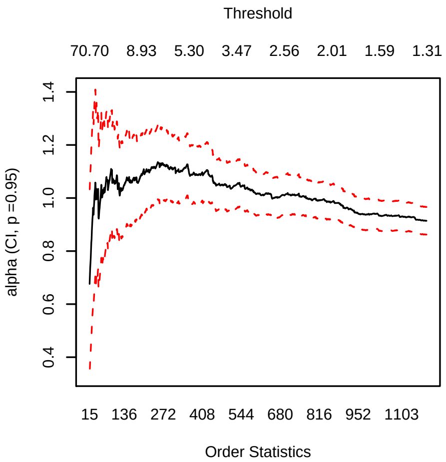


Figure 5.9: The Obesity index of the Burr distribution

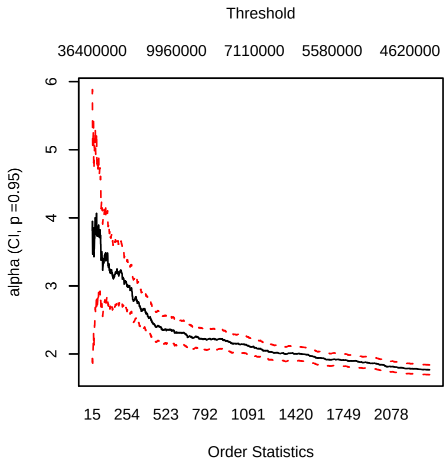
Dataset	Obesity Index	Confidence Interval
Hospital Data	0.8	(0.7928,0.8072)
NFIP	0.876	(0.8700,0.8820)
National Crop Insurance	0.808	(0.8009,0.8151)

Table 5.4: Estimate of the Obesity index

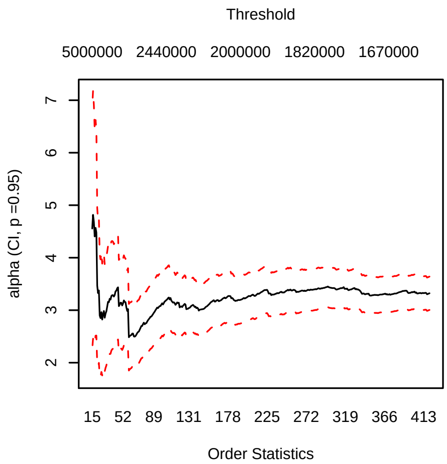
and the Hospital data. These conclusions are supported by the mean excess plots of these data sets and the Hill estimates. Figure 5.10 displays the Hill estimates based upon the top 20% of the observations of each data set. Note that the Hill plots in Figures 5.10a and 5.10c are quite stable, but that the Hill plot of the national crop insurance data in Figure 5.10b is not.



(a) NFIP



(b) National Crop Insurance



(c) Hospital

Figure 5.10: Hill estimator of a number of data sets

The final data set we consider is the G-econ database from Nordhaus et al. [2006]. This data set consists of environmental and economical characteristics of cells of 1 degree latitude and 1 degree longitude of the earth. One of the entries is the average precipitation. From the mean excess plot of this data set it is unclear whether this is a heavy-tailed distribution. In figure 5.11 the mean excess plot first decreases and then increases. The obesity index of this data set is estimated as 0.728 with 95%-confidence bounds $(0.6728, 0.7832)$. This estimate suggests a thin-tailed distribution. This conclusion is supported if we look at the exponential QQ-plot of the data set which shows that the data follows a exponential distribution almost perfectly.

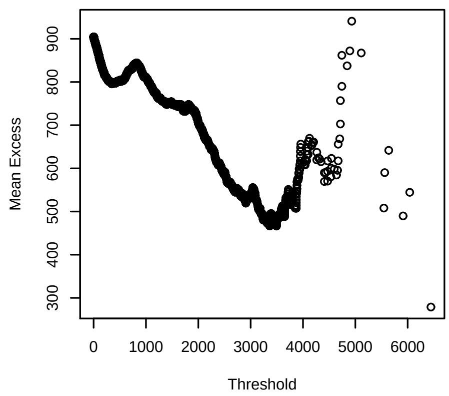


Figure 5.11: Mean excess plot average precipitation.

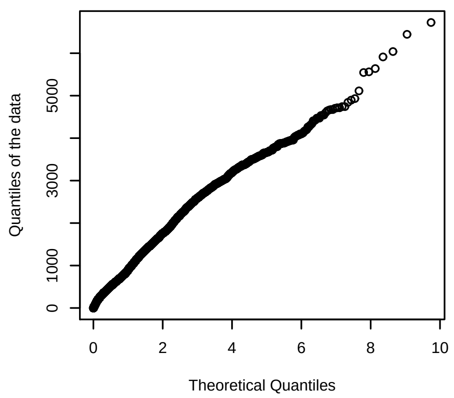


Figure 5.12: Exponential QQ-plot for the average precipitation.

Chapter 6

Dependence

[1]

6.1 Definition and main properties

We often want to explain or predict one variable based on values of other variables. Heavy tailed variables pose problems in this regard, as covariances may not exist. This chapter focuses on the class of distributions for which these problems are most acute, which we term "essentially heavy tailed". James orthogonality is proposed as a suitable replacement of orthogonality in the space of square integrable functions which underlies classical regression.

Definition 6.1.1. *A random variable X or its distribution μ has essentially heavy tail if $\varkappa(\mu) < 2$, where*

$$\varkappa(\mu) = sup\left\{p > 0 \colon \mathbf{E}|X|^p = \int |x|^p \mu(dx) < \infty\right\}.$$

In our convention the supremum over the empty set is zero. The constant $\varkappa(\mu)$ is called the essential order of the distribution μ or of the random variable X.

Note that if $\varkappa$ is the essential order of X then $\mathbf{E}|X|^p < \infty$ for each $p \in (0, \varkappa)$, but it can happen that $\mathbf{E}|X|^{\varkappa} = \infty$ which is different from the usual meaning of the order of random variable.

If $\mathbf{X}$ is a random vector in $\mathbb{R}^n$ with the distribution $\mu \in \mathcal{P}(\mathbb{R}^n)$ then the essential order of $\mathbf{X}$ (of μ) is defined by

$$\varkappa(\mu) = sup\left\{p \in (0, 2] \colon \mathbf{E}|< a, \mathbf{X} >|^p = \int |< a, x >|^p \mu(dx) < \infty \,\forall\, a \in \mathbb{R}^n\right\}.$$

In the following we briefly describe two classes of essentially heavy tail multidimensional probability distributions which seem to be natural candidates for modeling and simulating essentially heavy tail behavior of some multidimensional real processes. We propose also some measures of dependence for essentially heavy tail multidimensional random vectors.

6.2 Isotropic distributions

The construction described in this section is very natural and simple. By K we denote a compact star set in $\mathbb{R}^n$, i.e. a compact set for which the following condition holds

$$\mathbf{x} \in K \quad \Rightarrow \quad \forall\, t \in [0, 1] \quad t\mathbf{x} \in K.$$

[1] This chapter is authored by Prof. J. Misiewicz

To avoid the situation that the set K is in fact contained in some linear subspace of $\mathbb{R}^n$ we assume also that K contains an open neighborhood of zero. By ∂K we denote the boundary of the set K. Sometimes it is more convenient to define first a quasi-norm q on $\mathbb{R}^n$ and then the star set $K = \{\mathbf{x} \in \mathbb{R}^n \colon q(x) \leqslant 1\}$.

Definition 6.2.1. *A random vector* $\mathbf{X}$ *is isotropic if it can be written in the following form*

$$\mathbf{X} \stackrel{d}{=} \mathbf{Z} \cdot \theta,$$

where θ *is a non-negative random variable independent of the random vector* $\mathbf{Z}$ *with the uniform distribution on the boundary* ∂K *of a compact star subset* K *of* $\mathbb{R}^n$*.*

Notice that if $\mathbf{X}$ has a density, then this density has level curves of the form

$$c \cdot \partial K, \quad c > 0.$$

It is also easy to see that for every choice of half-lines starting at zero $\ell_1, \ell_2 \in \mathbb{R}^n$ the distributions of the following conditional random variables

$$R_1 := \big(\|\mathbf{X}\|_2 \big| \mathbf{X} \in \ell_1\big), \quad R_2 := \big(\|\mathbf{X}\|_2 \big| \mathbf{X} \in \ell_2\big)$$

are equal up to a scale parameter, i.e. there exists $c(\ell_1, \ell_2)$ such that

$$R_1 \stackrel{d}{=} c(\ell_1, \ell_2) R - 2.$$

Moreover the essential order of the random vector $\mathbf{X}$ is equal to the essential order of the variable θ since the random vector $\mathbf{Z}$ has the strong second moment finite ($\mathbf{E}\|\mathbf{Z}\|^2 < \infty$).

The construction of isotropic random vectors is maybe the simplest (except vectors with independent coordinates) construction of multidimensional distributions. Assuming for simplicity that the distribution of $\mathbf{X}$ is symmetric it is natural to define in this case the correlation coefficient

$$\widetilde{\varrho}\,(X_1, X_2) \stackrel{def}{=} \varrho(Z_1, Z_2) = \frac{\operatorname{Cov}(Z_1, Z_2)}{\sqrt{\operatorname{Var}(Z_1)\operatorname{Var}(Z_2)}}.$$

This coincides with the usual correlation coefficient if $\mathbf{E}\theta^2 < \infty$; i.e. in this case we have

$$\widetilde{\varrho}\,(X_1, X_2) \equiv \varrho\,(X_1, X_2).$$

Pareto distributions. The most convenient in precise calculations example of essentially heavy tail distributions are Pareto distributions π_α with densities given by

$$f_\alpha(x) = \frac{\alpha}{x^{\alpha+1}}\, \mathbf{1}_{[1,\infty)}(x).$$

In this case we have $\varkappa(\pi_\alpha) = \alpha$. The n-dimensional version of Pareto distribution has the density function given by

$$f_\alpha(\mathbf{x}) = \frac{C(q,\alpha)}{q(\mathbf{x})^{\alpha+n}}\, \mathbf{1}_{[1,\infty)}(q(\mathbf{x})).$$

where $q \colon \mathbb{R}^n \mapsto [0, \infty)$ is a quasi-norm on $\mathbb{R}^n$ and $C(q, \alpha)$ is the normalizing constant. Notice that the level curves of n-dimensional version of Pareto distribution are of the form

$$\{\mathbf{x} \in \mathbb{R}^n \colon q(\mathbf{x}) = c\}, \quad c > 0.$$

Here we have $K = \{\mathbf{x} \in \mathbb{R}^n : q(\mathbf{x}) \leqslant 1\}$ and the random vector $\mathbf{Z}$ is uniformly distributed on the set $\partial K = \{\mathbf{x} \in \mathbb{R}^n : q(\mathbf{x}) =\}$ and the random variable θ has Pareto distribution π_α with the density f_α. In the case

$$q(\mathbf{x})^p = \sum_{k=1}^{n} |x_k|^p$$

for some $p > 0$ we have

$$C(q, \alpha) = 2^{-n+1} p^{n-1} \Gamma\left(n/p\right).$$

Elliptically contoured distributions The best known isotropic distributions are spherically invariant (rotationally invariant) distributions with

$$q(\mathbf{x})^2 = \sum_{k=1}^{n} x_k^2 = \|\mathbf{x}\|_2, \quad \partial K = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\},$$

and elliptically contoured distributions with

$$q(\mathbf{x})^2 = \mathbf{x}\Sigma^{-1}\mathbf{x}^T, \quad \partial K = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}\Sigma^{-1}\mathbf{x}^T = 1\right\},$$

for some positive definite symmetric $n \times n$-matrix Σ. The elliptically contoured distributions are playing a very important role in multidimensional probability, stochastic processes, statistics and stochastic modeling.

Dirichlet and Liouville-type distributions. The classical Dirichlet random vector takes values in a positive rectangle of $\mathbb{R}^n$, however here we consider sign-symmetric Dirichlet distribution, where

Definition 6.2.2. *A random vector $\mathbf{U} = (U_1, \ldots, U_n)$ has a sign-symmetric Dirichlet-type distribution with parameters $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ (notation $\mathbf{U} \sim \mathcal{SD}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n)$) if the following conditions hold*

(i) U_n is a symmetric random variable;

(ii) $\sum_{i=1}^{n} |U_i|^{\alpha_i} = 1$ almost surely;

(iii) the joint density function of $(U_1, \ldots, U_{n-1})$ is

$$\frac{\Gamma(p_n)}{\Gamma(\beta_n/\alpha_n)} \prod_{i=1}^{n-1} \frac{\alpha_i}{2\Gamma(\beta_i/\alpha_i)} |u_i|^{\beta_i - 1} \left(1 - \sum_{i=1}^{n-1} |u_i|^{\alpha_i}\right)_+^{(\beta_n/\alpha_n)-1},$$

where $p_i = \sum_{j=1}^{i} \beta_j/\alpha_j$ for $i = 1, \ldots, n$.

The sign symmetric Dirichlet random vector can be obtained by the following construction: Let $\Gamma_1, \ldots, \Gamma_n$ be mutually independent real random variables where the probability density function of Γ_i is

$$\frac{\alpha_i}{2\Gamma(\beta_i/\alpha_i)} |z_i|^{\beta_i - 1} \exp\{-|z_i|^{\alpha_i}\}$$

for $z_i \in \mathbb{R}$. Then

$$(U_1, \ldots, U_n) \stackrel{d}{=} \left(\frac{\Gamma_1}{\left(\sum_{j=1}^{n} |\Gamma_j|^{\alpha_j}\right)^{1/\alpha_i}}, \ldots, \frac{\Gamma_n}{\left(\sum_{j=1}^{n} |\Gamma_j|^{\alpha_j}\right)^{1/\alpha_i}} \right).$$

Since

$$\mathrm{Var}(U_i) = \frac{\Gamma(p_n)\Gamma(\beta_i/\alpha_i + 2/\alpha_i)}{\Gamma(p_n + 2/\alpha_i)\Gamma(\beta_i/\alpha_i)}$$

we see that

$$\mathrm{Cov}\left(\sum_{i=1}^n x_i U_i, \sum_{i=1}^n y_i U_i\right) = \Gamma(p_n) \sum_{i=1}^n x_i y_i \frac{\Gamma(\beta_i/\alpha_i + 2/\alpha_i)}{\Gamma(p_n + 2/\alpha_i)\Gamma(\beta_i/\alpha_i)}.$$

Of course the Dirichlet distributions are supported on a compact subset of $\mathbb{R}^n$ and thus they do not have essentially heavy tails. In order to get essentially heavy-tail distribution based on Dirichlet random vector we shall consider the following:

Definition 6.2.3. *A random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is said to have a sign-symmetric Liouville-type distribution if there exists a sign-symmetric Dirichlet-type random vector $\mathbf{U} \sim \mathcal{SD}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n)$ and a non-negative random variable θ such that*

$$(X_1, \ldots, X_n) \stackrel{d}{=} \left(U_1 \theta^{1/\alpha_1}, \ldots, U_n \theta^{1/\alpha_n}\right).$$

Notation: $\mathbf{X} \sim \mathcal{SL}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n; \theta)$.

The random vector $\mathbf{X} \sim \mathcal{SL}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n; \theta)$ has essentially heavy tail if and only if the corresponding variable θ has essentially heavy tail, however the dependence structure of coordinates of $\mathbf{X}$ and corresponding $\mathbf{U}$ are the same up to scale mixture. More about sign-symmetric Dirichlet type and Liouville type distributions one can find e.g. in R.D. Gupta and Richards [1996].

6.3 Pseudo-isotropic distributions

Another class of handy in calculations essentially heavy tail distributions are symmetric multi-dimensional distributions having all one-dimensional projections (not only projections onto the axis) the same up to a scale parameter. A wide description of such distributions one can find in J. K. Misiewicz and Scheffer. [1990] and J. K. Misiewicz.

The main example of such distributions are symmetric stable distributions and elliptically contoured distributions.

Stable distributions. For a long time applicability of stable distributions was rather theoretical, since there are only three cases $\alpha = 2, 1, \frac{1}{2}$ in which the density of α-stable random variable is given in an explicit formula. Now we can deal with most of difficulties with the help of computer calculations and the role of stable distributions and processes in stochastic modeling is growing.

In the following we assume that the considered random vectors are living in $\mathbb{R}^n$, however in full generality we could assume that they are taking values in any separable Banach space $\mathbb{E}$.

Definition 6.3.1. *A random vector $\mathbf{Y} \in \mathbb{R}^n$ is stable if for every $a, b > 0$ there exist $c = c(a,b) > 0$ and $d = d(a,b) \in \mathbb{R}^n$ such that*

$$a\mathbf{Y} + b\mathbf{Y}' \stackrel{d}{=} c\mathbf{Y} + d,$$

where $\mathbf{Y}'$ is an independent copy of $\mathbf{Y}$ and $\stackrel{d}{=}$ denotes equality of distributions.

If $d(a,b) = 0$ for each choice of $a, b > 0$ then the random vector $\mathbf{Y}$ is called strictly stable. Every symmetric stable random vector, i.e. such vector that $\mathbf{Y} \stackrel{d}{=} -\mathbf{Y}$, is strictly stable.

It is known that if $\mathbf{Y}$ is stable then there exists a constant $\alpha \in (0,2]$ such that $c(a,b)^\alpha = a^\alpha + b^\alpha$. The constant α is called the characteristic exponent and it coincides with the essential order $\varkappa(\mu)$ for $\mu = \mathcal{L}(\mathbf{Y})$, the distribution of $\mathbf{Y}$. A stable random vector $\mathbf{Y}$ with the essential order α is called α-stable. A symmetric α-stable vector we denote by SαS.

If $\mathbf{Y} \in \mathbb{R}^n$, $n \geqslant 2$, is SαS then there exists a symmetric finite measure Γ on the unit sphere $S_{n-1} \subset \mathbb{R}^n$ such that

$$\mathbf{E}e^{i<\xi,\mathbf{Y}>} = \exp\Big\{ - \int \cdots \int S_{n-1} |<\xi, \mathbf{u}>|^\alpha \, \Gamma(du) \Big\}. \qquad (*)$$

The measure Γ is called the spectral measure for SαS random vector. This measure is uniquely determined if only $\alpha < 2$. For $\alpha = 2$ we do not have uniqueness, but this is just the Gaussian case and we have many other methods of dealing with this distribution.

By Y_α we denote the canonical SαS random variable with the characteristic function $\exp\{-|t|^\alpha\}$. By $(*)$ we see that if $\mathbf{Y}$ is SαS random vector in $\mathbb{R}^n$ then for each $\xi \in \mathbb{R}^n$ we have

$$<\xi, \mathbf{Y}> = \sum_{k=1}^{n} \xi_k Y_k \stackrel{d}{=} C(\xi) Y_\alpha,$$

where the constant $C(\xi)$ is given by

$$C(\xi)^\alpha = \int \cdots \int S_{n-1} |<\xi, \mathbf{u}>|^\alpha \, \Gamma(du).$$

Pseudo-isotropic distributions. The SαS distributions are the main examples of pseudo-isotropic distributions, where

Definition 6.3.2. *A symmetric random vector $\mathbf{X}$ taking values in $\mathbb{R}^n$ is pseudo-isotropic if all its one-dimensional projections have the same distributions up to a scale parameter, i.e.*

$$\forall\, \xi \in \mathbb{R}^n \ \exists\, C(\xi) \geqslant 0 \qquad <\xi, \mathbf{X}> = \sum_{k=1}^{n} \xi_k X_k \stackrel{d}{=} C(\xi)\, X_1.$$

The function $C\colon \mathbb{R}^n \to [0,\infty)$ defines a quasi-norm on $\mathbb{R}^n$.

It is known that if μ denotes the distribution of a pseudo-isotropic random vector $\mathbf{X}$ with the quasi-norm C and $\varkappa(\mu) > 0$ then $\varkappa(\mu) \leqslant \alpha(\mu)$, where

$$\alpha(\mu) \stackrel{def}{=} \sup\Big\{ p \in (0,2] \colon \ \exp\{-C(\xi)^p\} \ \text{ is positive definite on } \mathbb{R}^n \Big\}.$$

The constant $\alpha(\mu)$ is called the index of stability for pseudo-isotropic distribution μ. The formulation: $\exp\{-C(\xi)^p\}$ *is positive definite on* $\mathbb{R}^n$ means simply that $\exp\{-C(\xi)^p\}$ is a characteristic function of some random vector, which (it is easy to check) is SpS. By the Lévy-Cramér theorem we have that $\exp\{-C(\xi)^{\alpha(\mu)}\}$ is also a characteristic function of some S$\alpha(\mu)$S random vector $\mathbf{Y}$. Of course $\mathbf{Y}$ is also pseudo-isotropic and $<\xi, \mathbf{Y}>$ has the same distribution as $C(\xi)Y_\alpha$. Consequently for each $1 < p < \alpha(\mu)$ we have

$$\mathbf{E}\big|<\xi, \mathbf{Y}>\big|^p = C(\xi)^p \, \mathbf{E}|Y_\alpha|^p.$$

If the pseudo-isotropic random vector $\mathbf{X} = (X_1, \ldots, X_n)$ with distribution μ and the quasi-norm C is such that $1 < \varkappa(\mu) \leqslant \alpha(\mu)$ then for each $\xi \in \mathbb{R}^n$ and each $1 < p < \varkappa(\mu)$

$$\mathbf{E}\big| < \xi, \mathbf{X} > \big|^p = C(\xi)^p \, \mathbf{E}|X_1|^p.$$

Consequently for all $1 < p < \varkappa(\mu) \leqslant \alpha(\mu)$ we have

$$\frac{\mathbf{E}\big| < \xi, \mathbf{Y} > \big|^p}{\mathbf{E}|Y_\alpha|^p} = \frac{\mathbf{E}\big| < \xi, \mathbf{X} > \big|^p}{\mathbf{E}|X_1|^p}.$$

We see also that if a pseudo-isotropic random vector $\mathbf{X} \sim \mu$ is such that $\varkappa(\mu) > 0$ then there exists a symmetric finite measure Γ on the unit sphere $S_{n-1} \subset \mathbb{R}^n$ such that

$$C(\xi)^{\alpha(\mu)} = \int \cdots \int S_{n-1} | < \xi, \mathbf{u} > |^{\alpha(\mu)} \, \Gamma(du).$$

The measure Γ we shall call the spectral measure of pseudo-isotropic distribution of positive essential order.

Remak. It is known (see Misiewicz and Tabisz) that the distribution which is both isotropic and pseudo-isotropic must be elliptically contoured.

6.3.1 Covariation as a measure of dependence for essentially heavy tail jointly pseudo-isotropic variables

The covariation of random variables was originally defined for jointly stable variables. We can define it also for jointly pseudo-isotropic variables.

Definition 6.3.3. *Let (X_1, X_2) be pseudo-isotropic with parameter $\alpha > 1$ and let Γ be the spectral measure of the corresponding SαS random vector (Y_1, Y_2). The covariation of X_1 on X_2 is the real number*

$$[X_1, X_2]_\alpha \equiv [Y_1, Y_2]_\alpha \overset{def}{=} \int_{S_1} u_1 u_2^{<\alpha-1>} \Gamma(d\mathbf{u}),$$

where $a^{<p>} = |a|^p \, sign(a)$.

Equivalently we can say that

$$\begin{aligned} [X_1, X_2]_\alpha \equiv [Y_1, Y_2]_\alpha &= \frac{1}{\alpha} \frac{\partial}{\partial r} \int_{S_1} |ru_1 + su_2|^\alpha \, \Gamma(d\mathbf{u}) \Big|_{r=0,s=1} \\ &= \frac{1}{\alpha} \frac{\partial}{\partial r} C((r,s))^\alpha \Big|_{r=0,s=1}, \end{aligned}$$

where C is the quasi-norm C for (X_1, X_2).

Lemma 6.3.1. *If the random vector (X_1, X_2) is pseudo-isotropic and $1 < p < \varkappa \leqslant \alpha$ then*

$$\frac{[X_1, X_2]_\alpha}{C(0,1)^\alpha} = \frac{\mathbf{E} X_1 X_2^{<p-1>}}{\mathbf{E}|Y_2|^p}. \qquad (**)$$

Proof. Let (Y_1, Y_2) be the corresponding SαS random vector. Since $\mathbf{E}|rY_1 + sY_2|^p = C(r,s)^p \, \mathbf{E}|Y_\alpha|^p$ for $p < \alpha$ and $\mathbf{E}|rX_1 + sX_2|^p = C(r,s)^p \, \mathbf{E}|X_1|^p$ for $p < \varkappa$ then for each $p < \varkappa$ we have

$$\frac{\partial}{\partial r} \mathbf{E}|rX_1 + sX_2|^p = \frac{\partial}{\partial r} C(r,s)^p \, \mathbf{E}|X_1|^p = p \, \mathbf{E}|X_1|^p \, C(r,s)^{p-\alpha} \frac{1}{\alpha} \frac{\partial}{\partial r} C(r,s)^\alpha.$$

Consequently, for each $1 < p < \varkappa$ we have

$$\begin{aligned} [X_1, X_2]_\alpha &= \left[\frac{C(r,s)^{\alpha-p}}{p\,\mathbf{E}|X_1|^p} \frac{\partial}{\partial r} \mathbf{E}|rX_1 + sX_2|^p \right]_{r=0,s=1} \\ &= \frac{C(0,1)^{\alpha-p}}{\mathbf{E}|X_1|^p} \mathbf{E} X_1 X_2^{<p-1>}. \end{aligned}$$

Now it is enough to notice that $\mathbf{E}|X_2|^p = C(0,1)^p\, \mathbf{E}|X_1|^p$. □

Example 1. Consider an rotationally invariant random vector $(Y_1, Y_2) = (U_1, U_2)\theta$, where (U_1, U_2) has uniform distribution on the unit sphere $S_1 \subset \mathbb{R}^2$, θ is a random variable with the Pareto distribution with parameter $\alpha \in (1,2)$ independent of (U_1, U_2). Denote by μ the distribution of (Y_1, Y_2). It is easy to check that the quasi-norm C is given by $C(r,s)^2 = r^2 + s^2$, thus $\alpha(\mu) = 2$ and the corresponding symmetric stable vector is Gaussian with independent increments. On the other hand

$$\mathbf{E}|rY_1 + sY_2|^p = \mathbf{E}|rU_1 + sU_2|^p\, \mathbf{E}\theta^p,$$

thus $\varkappa(\mu) = \sup\{p \in (0,2] \colon \mathbf{E}\theta^p < \infty\} = \alpha$ and (Y_1, Y_2) has essentially heavy tail distribution. Evidently $[Y_1, Y_2]_2 = 0$ in spite of the fact that Y_1 and Y_2 are not independent.

Remark. Of course the covariation $[Z_1, Z_2]_p$, $p \in (1,2)$, can be defined for any symmetric random vector (Z_1, Z_2) with the property $\mathbf{E}|rZ_1 + sZ_2|^p < \infty$ for all $r, s \in \mathbb{R}$. However the chosen parameter p does not have to be the maximal with this property. Moreover if $\varkappa$ is the essential order of (Z_1, Z_2) then it may happen that $\mathbf{E}|Z_1|^\varkappa = \infty$. This makes a difference because without pseudo-isotropy we do not have equality

$$(\mathbf{E}|rZ_1 + sZ_2|^p)^{1/p} = c(p,q)\ (\mathbf{E}|rZ_1 + sZ_2|^q)^{1/q}$$

for every $p, q < \varkappa$ and a suitable constant $c(p,q)$. Consequently the substitutes of inner product $[Z_1, Z_2]_p$ and $[Z_1, Z_2]_q$ are different for the same random vector (Z_1, Z_2). In particular, between two-dimensional Pareto distributions only Bessel–Pareto (rotationally invariant Pareto) distribution is pseudo-isotropic.

In the case of non pseudo-isotropic random vector (Z_1, Z_2) of essential order $\varkappa \in (1,2)$ we could use the following construction: It is known that for each $p \in (1, \varkappa)$

$$\varphi_p(r,s) = \exp\{-\mathbf{E}|rZ_1 + sZ_2|^p\}$$

is the characteristic function of some symmetric p-stable random vector. If we assume that $\mathbf{E}|Z_1|^p = \mathbf{E}|Z_2|^p =: m_p$ (e.g. if Z_1 and Z_2 have the same distribution) then

$$\varphi_p\left(\frac{(r,s)}{m_p}\right) = \exp\left\{-\mathbf{E}\left|\frac{rZ_1 + sZ_2}{m_p}\right|^p\right\}$$

converges for $p \nearrow \varkappa$ to the characteristic function of some symmetric $\varkappa$-stable random vector (Y_1, Y_2) since

$$\left(\mathbf{E}\left|\frac{rZ_1 + sZ_2}{m_p}\right|^p\right)^{1/p} \leqslant \left(\mathbf{E}\left|\frac{rZ_1}{m_p}\right|^p\right)^{1/p} + \left(\mathbf{E}\left|\frac{sZ_2}{m_p}\right|^p\right)^{1/p} \leqslant |r| + |s|.$$

Then we can define uniquely the covariation of Z_1 on Z_2 by

$$[Z_1, Z_2]_\varkappa := [Y_1, Y_2]_\varkappa,$$

however in practice it is difficult to approximate this parameter from data.

The following lemma was proven in Samorodnitsky and Taqqu [1994] only in the case of SαS random vectors. The proof in the case of pseudo-isotropic vectors is almost the same and will be omitted.

Lemma 6.3.2. *Let $(X_1, \ldots, X_n)$ be a pseudo-isotropic random vector with the characteristic exponent $\alpha > 1$ and spectral measure Γ_n. If*

$$W = \sum_{k=1}^{n} a_k X_k \quad \textit{and} \quad Z = \sum_{k=1}^{n} b_k X_k,$$

then

$$[W, Z]_\alpha = \int_{S_{n-1}} \Big(\sum_{k=1}^{n} a_k u_k \Big) \Big(\sum_{k=1}^{n} b_k u_k \Big)^{<\alpha-1>} \Gamma_n(du).$$

Now we are able to define an analog of the covariance coefficient for essentially heavy tail pseudo-isotropic variables.

Definition 6.3.4. *Let $\mathbf{X} = (X_1, X_2)$ be a pseudo-isotropic random vector with $\alpha > 1$ and the quasi-norm C. Then the covariation norm of the random variable $X = rX_1 + sX_2$ is defined by*

$$\|X\|_\alpha^\alpha \stackrel{def}{=} [X, X]_\alpha = C(r, s)^\alpha.$$

The covariation coefficient of X_1 on X_2 is

$$\rho_\alpha(X_1, X_2) := \frac{[X_1, X_2]_\alpha}{\|X_1\|_\alpha \|X_2\|_\alpha^{\alpha-1}} = \frac{[X_1, X_2]_\alpha}{C(0,1)^{\alpha-1}}$$

Notice that by our definition $C(1, 0) = 1$.

Example 2. Consider $Y_0, Y_1, \ldots$ independent identically distributed SαS random variables with $\alpha > 1$. First we will discuss the covariation of $Z_1 = Y_1 \pm \varepsilon Y_0$ on $Z_2 = Y_2 + \varepsilon Y_0$. The random vector (Y_0, Y_1, Y_2) has the spectral measure Γ_3 with 6 atoms $(u_1, u_2, u_3) = (\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$ of the weight $1/2$ each. Thus

$$\begin{aligned} [Z_1, Z_2]_\alpha &= \int_{S_2} (u_1 \pm \varepsilon u_3)\,(u_2 + \varepsilon u_3)^{<\alpha-1>}\, \Gamma_3(du) \\ &= \pm \frac{1}{2} \Big(\varepsilon \cdot (\varepsilon)^{<\alpha-1>} - \varepsilon \cdot (-\varepsilon)^{<\alpha-1>} \Big) = \pm \varepsilon^\alpha. \end{aligned}$$

The covariation norms of these variables are as follows

$$\|Z_1\|_\alpha^\alpha = \|Z_2\|_\alpha^\alpha = 1 + \varepsilon^\alpha,$$

thus

$$\rho_\alpha(Z_1, Z_2) = \frac{\pm \varepsilon^\alpha}{1 + \varepsilon^\alpha}.$$

We see that it can be taken really small. Consider now the following:

$$Z_1^{(n)} = n^{-1/\alpha} \sum_{k=1}^{n} (Y_k \pm \varepsilon Y_0) \quad \text{and} \quad Z_2^{(n)} = n^{-1/\alpha} \sum_{k=n+1}^{2n} (Y_k + \varepsilon Y_0)\,.$$

It is well known that

$$(Z_1^{(n)}, Z_2^{(n)}) \stackrel{d}{=} \left(Y_1 \pm n^{1-1/\alpha}\varepsilon Y_0, Y_2 + n^{1-1/\alpha}\varepsilon Y_0\right),$$

thus the covariation $Z_1^{(n)}$ on $Z_2^{(n)}$ is given by

$$[Z_1^{(n)}, Z_2^{(n)}]_\alpha = \pm\, n^{\alpha-1}\varepsilon^\alpha.$$

The covariation norm

$$\|Z_1^{(n)}\|_\alpha^\alpha = 1 + n^{\alpha-1}\varepsilon^\alpha,$$

thus

$$\rho(Z_1^{(n)}, Z_2^{(n)}) = \frac{\pm\, n^{\alpha-1}\varepsilon^\alpha}{1 + n^{\alpha-1}\varepsilon^\alpha} \to 1 \quad \text{when} \quad n \to \infty.$$

However if we put $\varepsilon = \varepsilon_n = cn^{1/\alpha-1}$ then for each $n \in \mathbb{N}$

$$[Z_1^{(n)}, Z_2^{(n)}]_\alpha = \pm\, c^\alpha, \quad \|Z_1^{(n)}\|_\alpha^\alpha = 1 + c^\alpha, \quad \rho(Z_1^{(n)}, Z_2^{(n)}) = \frac{\pm\, c^\alpha}{1 + c^\alpha},$$

and it can attain every value from the interval $(-1, 1)$.

6.3.2 Codifference

Unfortunately for $\alpha \leqslant 1$ the idea of covariation and the corresponding covariation norm do not have required properties since the triangular inequality does not hold. In such case we shall use *codifference*:

$$\begin{aligned} \tau(X_1, X_2) &= \|X_1\|_\alpha^\alpha + \|X_2\|_\alpha^\alpha - \|X_1 - X_2\|_\alpha^\alpha \\ &= C(1,0)^\alpha + C(0,1)^\alpha - C(1,-1)^\alpha. \end{aligned}$$

Notice that for the variables described in the previous example (for $\alpha < 1$) we have

$$\begin{aligned} \tau(Z_1^{(n)}, Z_2^{(n)}) &= 2\left(1 + n^{\alpha-1}\varepsilon^\alpha\right) - 2 - \left(1 + (-1)^{\pm 1}\right)^\alpha n^{\alpha-1}\varepsilon^\alpha \\ &= \left(2 - \left(1 + (-1)^{\pm 1}\right)^\alpha\right) n^{\alpha-1}\varepsilon^\alpha \to 0 \quad \text{for} \quad n \to \infty, \end{aligned}$$

and for $\varepsilon_n = cn^{1/\alpha-1}$

$$\tau(Z_1^{(n)}, Z_2^{(n)}) = \left(2 - \left(1 + (-1)^{\pm 1}\right)^\alpha\right) c^\alpha.$$

6.3.3 The linear regression model for essentially heavy tail distribution

In many classical problems of statistics we want to approximate value of variable Y by the value aX with the best possible constant a in the sense of minimizing the following function

$$\mathbf{E}\,(Y - aX)^2.$$

Let $L_0^2(\Omega, \mathcal{F}, \mathbf{P})$ be the space of all random variables on $(\Omega, \mathcal{F})$ with mean zero and finite second moment. The space $L_0^2(\Omega, \mathcal{F}, \mathbf{P})$ is a Hilbert space with the scalar product

$$(X, Y) := \mathrm{Cov}(X, Y) = \mathbf{E}XY.$$

If $X, Y \in L_0^2(\Omega, \mathcal{F}, \mathbf{P})$ then minimizing the value of $\mathbf{E}(Y - aX)^2$ is the same as finding the value $a \in \mathbb{R}$ such that X is orthogonal to $(Y - aX)$, thus

$$a = \frac{(Y, X)}{\|X\|_2^2}.$$

This construction is not possible if $\mathbf{E}X^2 = \mathbf{E}Y^2 = \infty$. However in the space $L_0^p(\Omega, \mathcal{F}, \mathbf{P})$ of the variables with the finite pth moment, $p \in (1,2)$, equipped with the norm

$$\|X\|_p^p := \mathbf{E}|X|^p$$

(see James [1947]) it is possible to consider James orthogonality in the following way:
X is James orthogonal to Y (notation $X \perp_J Y$) if for every real λ

$$\|X + \lambda Y\|_p \geqslant \|X\|_p.$$

The next proposition was proven in Samorodnitsky and Taqqu [1994] for jointly $S\alpha S$ random vectors.

Proposition 6.3.3. *Let (X,Y) be pseudo-isotropic with the essential order $\varkappa \in (1,2)$ and index of stability $\alpha \in [\varkappa, 2)$. Then*

$$[X,Y]_\alpha = 0$$

if and only if Y is James orthogonal to X in $L_0^p(\Omega, \mathcal{F}, \mathbf{P})$ for each $p \in (1, \varkappa)$.

Proof. Let (Z_1, Z_2) be the corresponding $S\alpha S$ random vector with the same as (X,Y) quasi-norm c. It was shown in Samorodnitsky and Taqqu [1994] that

$$[Z_1, Z_2]_\alpha = 0 \quad \text{iff} \quad \|\lambda Z_1 + Z_2\|_\alpha \geqslant \|Z_2\|_\alpha$$

for every real λ. Now it is enough to notice that

$$\|\lambda X + Y\|_p^p = \mathbf{E}|\lambda X + Y|^p = c(\lambda,1)^p\, \mathbf{E}|X|^p = \|\lambda X + Y\|_\alpha^p\, \mathbf{E}|X|^p$$

for each real λ and that by definition

$$[X,Y]_\alpha = [Z_1, Z_2]_\alpha.$$

$\square$

Coming back to the linear regression problem for essentially heavy tail distributions assume that (X,Y) is pseudo-isotropic with pseudo-norm c, distribution μ of the essential order $\varkappa(\mu) \in (1,2)$ and the corresponding index of stability $\alpha(\mu) < 2$. We want to find $a \in \mathbb{R}$ such that

$$X \perp_\alpha (Y - aX).$$

Then the random vector $(Y - aX, X)$ is also pseudo-isotropic with the quasi-norm $c_1(r,s) = c(s - ar, s)$ since

$$r(Y - aX) + sX = (s - ar)X + rY \stackrel{d}{=} c(s - ar, s)X.$$

By Lemma 1 we have that

$$0 = [Y - aX, X]_\alpha \;=\; \frac{c_1(0,1)^{\alpha - p}}{\mathbf{E}|X|^p}\left(\mathbf{E}YX^{<p-1>} - a\,\mathbf{E}|X|^p\right),$$

for each $1 < p \leqslant \varkappa(\mu)$. Consequently

$$a = \frac{\mathbf{E}YX^{<p-1>}}{\mathbf{E}|X|^p}.$$

The next proposition describes the consistent statistics for linear regression in this case.

Proposition 6.3.4. *Let (X, Y) be pseudo-isotropic with pseudo-norm c, distribution μ of the essential order $\varkappa(\mu) \in (1, 2)$ and the corresponding index of stability $\alpha(\mu) < 2$ and let $p \in (1, \varkappa(\mu))$. If $X \perp_J (Y - aX)$, i.e.*

$$\mathbf{E}|Y - aX|^p = \inf\left\{\mathbf{E}|Y + \lambda X|^p \colon \lambda \in \mathbb{R}\right\}$$

then

$$Q_n = \frac{\sum_{k=1}^n Y_k X_k^{<p-1>}}{\sum_{k=1}^n |X_k|^p}$$

is a consistent estimator of a.

Proof. Let $m_p = \mathbf{E}|X|^p = c(1, 0)^p$ and $\mu(Y, X) = \mathbf{E} Y X^{<p-1>}$. Take $\gamma, \varepsilon, \delta > 0$, $\delta \ll m_p$ and such that

$$\frac{\delta(\mu(Y, X) + m_p)}{m_p(m_p - \delta)} < \gamma.$$

We define

$$\begin{aligned} A_n &= \left\{\omega \colon m_p - \delta < \frac{1}{n}\sum_{k=1}^n |X_k|^p < m_p + \delta\right\}, \\ B_n &= \left\{\omega \colon \mu(Y, X) - \delta < \frac{1}{n}\sum_{k=1}^n Y_k X_k^{<p-1>} < \mu(Y, X) + \delta\right\}. \end{aligned}$$

By the Weak Law of Large Numbers there exists $n_0 \in \mathbb{N}$ such that for each $n \geqslant n_0$ we have $\mathbf{P}(A_n) > 1 - \varepsilon/2$ and $\mathbf{P}(B_n) > 1 - \varepsilon/2$ and consequently $\mathbf{P}(A_n \cap B_n) > 1 - \mathbf{P}(A_n^c) - \mathbf{P}(B_n^c) > 1 - \varepsilon$. On the set $A_n \cap B_n$ we have that

$$\frac{\mu(Y, X)}{m_p} - \frac{\delta(\mu(Y, X) + m_p)}{m_p(m_p + \delta)} \leqslant Q_n \leqslant \frac{\mu(Y, X)}{m_p} + \frac{\delta(\mu(Y, X) + m_p)}{m_p(m_p - \delta)}.$$

Consequently for each $n \geqslant n_0 = n_0(\gamma, \varepsilon)$ we have

$$\mathbf{P}\left\{\omega \colon \left|Q_n - \frac{\mu(Y, X)}{m_p}\right| > \gamma\right\} \leqslant \varepsilon.$$

Now it is enough to take $\varepsilon = \varepsilon_n \to 0$ for $n \to \infty$. $\square$

Chapter 7

Conclusions and Future Research

The data sets discussed in chapter 1 hopefully persuade the reader that heavy tail phenomena are not incomprehensible bolts from "extremistan", but arise in many commonplace data sets. They are not incomprehensible, but they cannot be understood with traditional statistical tools. Using the wrong tools *is* incomprehensible. This monograph reviewed various notions, intuitions and definitions of tail heaviness. The most popular definitions in terms of regular variation and subexponentiality invoke putative properties that hold at infinity, and this complicates any empirical estimate. Each definition captures some but not all of the intuitions associated with tail heaviness.

Chapter 5 studied two different candidates to characterize the tail-heaviness of a data set based upon the behavior of the mean excess plot under aggregations by k. We first considered the ratio of the largest to the second largest observations. It turned out this ratio has a non-degenerate limit if and only if the distribution is regularly varying. An estimate of the probability that this ratio exceeds 2 based on the observed Type 2 2-records in a data set was very inaccurate: the expected number of Type 2 2-records in a data set of size 10000 is 16.58. For thin-tailed distributions the estimator is biased, since the probability in question decreases to zero but the estimator is non-negative on initial segments. This is due to the fact that most 2-records will be observed early in the data set, and then relative frequency of the largest observation exceeding twice the second largest is still quite large. This motivated the search for for another characterization of heavy-tailedness. The Obesity index of a random variable was defined as:

$$\mathrm{Ob}(X) = P\left(X_1 + X_4 > X_2 + X_3 | X_1 \leq X_2 \leq X_3 \leq X_4\right), \; X_i \text{i.i.d. copies of } X.$$

This index reasonably captures intuitions on tail heaviness and can be calculated for distributions and computed for data sets. However, it does not completely mimic the tail index. We saw that the obesity index of two Burr distributions could reverse the order of their tail indices. When applied to various data sets we saw that the Obesity index and the Hill estimator both gave roughly similar results, in those cases where the Hill estimator gave a clear signal.

As with any new notion, it is easy to think of interesting research questions. We mention two. The notion of a multivariate Obesity index immediately suggests itself by considering a joint probability

$$\begin{aligned}
&\mathrm{Ob}(X,Y)\mathrm{Ob}(X)\mathrm{Ob}(Y) = \\
&P\left(X_1 + X_4 > X_2 + X_3 \cap Y_1 + Y_4 > Y_2 + Y_3 \,|\, X_1 \leq X_2 \leq X_3 \leq X_4 \cap Y_1 \leq Y_2 \leq Y_3 \leq Y_4\right), \\
&X_i \text{ i.i.d. copies of } X, \; Y_i \text{ i.i.d. copies of } Y.
\end{aligned}$$

The second question concerns covariates. We might like to explain tail heaviness in terms of independent variables. An obvious idea is to regress the rank of the dependent variable on the independent variables; that might get at "tail" but not "tail heaviness". The problem is to define an appropriate notion of orthogonality, and James orthogonality is is proposed in chapter 6. We may safely speculate that research in fat tails is itself a fat tailed phenomenon.

Bibliography

R. Frey A.J. McNeil and P. Embrechts. *Quantitative Risk Management.* Princeton University Press, 2005.

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

B.C. Arnold. *Pareto Distributions.* Fairland International Co-operative Publishing, 1983.

B.C. Arnold, N. Balakrishnan, and H.N. Nagaraja. *Records.* John Wiley Sons, 1998.

N. Balakrishnan and A. Stepanov. Asymptotic properties of the ratio of order statistics. *Statistics and Probability letters*, 78:301–310, 2007.

J. Beirlant, G. Dierckx, and A. Guillou. Estimation of the extreme-value index and generalized quantle plots. *Bernoulli*, 11:949–970, 2005.

J.L. Beirlant, J. Teugels and title = Practical Analysis of Extreme Values publisher = Leuven University Press year = 1996 Vynckier, P.

N.H. Bingham and J.L. Teugels. Conditions implying domains of attraction. In *Proceedings of the Sixth conference on Probability Theory*, pages 23–24, 1979.

N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Cambridge University Press, 1987.

S. Coles. *An Introduction to Statistical Modelling of Extreme Values.* Springer-Verlag, 2001.

R.M. Cooke and C. Kousky. Are catgastrophes insurable? *Resources*, 172:18–23, 2009.

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

M.E. Crovella and M.S. Taqqu. Estimating the heavy tail index from scaling properties. *Methodology and Computing in Applied Probability*, 1:55–79, 1999.

Herbert A. David. *Order Statistics.* John Wiley Sons, 2nd edition edition, 1981.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremeal Events for Insurance and Finance.* Springer, 1997.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley Sons, 2nd edition edition, 1971.

G.H. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 1934.

T.L. Hayes and D.A. Neal. *Actuarial Rate Review In Support of the Recommended October 1, 2010, Rate and Rule Changes*. FEMA, 2011.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals Of Statistics*, 3, 1975.

H.J.Rossberg. Charcterization of distribution functions by the independence of certain functions of order statistics. *Saknhya Ser A*, 134, pages = 111-120, 1972.

R.C. James. Orthogonality and linear functionals in normed linear spaces. *Trans. Amer. Math. Soc.*, 61:265–292, 1947.

J.L. Latchman, F.D.O. Morgan, and W.P. Aspinall. Temporal changes in the cumulative piecewise gradient of a variant of the gutenbergrichter relationship, and the imminence of extreme events. *Earth-Science Reviews*, 87:94–112, 2008.

J.R. Leslie. On the non-closure under convolution of the subexponential family. *Journal of Applied Probability*, 1989.

B.D. Malamud and D.L. Turcotte. The applicability of power-law frequency statistics to floods. *Journal of Hydrology*, 322:168–180, 2006.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36:394–419, 1963.

B.B. Mandelbrot and R.L. Hudson. *The (Mis)behaviour of Markets*. Profile Books Ltd, 2008.

J. K. Misiewicz and K. Tabisz. When an isotropic random vector is pseudo-isotropic? In *High Dimensional Probability II (Seatle, WA.1999)*. Birkhäuser Boston, Boston, MA.

Valery B. Nezvorov. *Records: Mathematical Theory*. American Mathematical Society, 2001.

W. Nordhaus, Q. Azam, D. Corderi, K. Hood, N. Makarova Victor, Mukhtar Mohammed, Alexandra Miltner, and Jyldyz Weiss. The g-econ database on gridded output: Methods and data. *Yale university*, 2006.

S.T. Rachev. *Handbook of heavy-tailed distributions in finance*. Elsevier, 2003.

J.K. Misiewicz R.D. Gupta and D.St.P. Richards. Infinite sequences with sign-symmetric liouville distributions. *Probab. and Math. Statist.*, 16:29–44, 1996.

S.I. Resnick. *Heavy-tail phenomena : probabilistic and statistical modeling*. Springer series in operations research and financial engineering. Springer, 2005.

G. Samorodnitsky and M.S. Taqqu. *Stable non-Gaussian Random Processes; Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.

J. K. Misiewicz. Sub-stable and pseudo-isotropic processes. connections with the geometry of sub-spaces of l_α-spaces. *Dissertationes Mathematicae*, CCCLVIII.

J. K. Misiewicz and C. L. Scheffer. Pseudo-isotropic measures. Nieuw Archief voor Wiskunde, 8:111–152, 1990.

B. Smid and A.J. Stam. Convergence in distribution of quotients of order statistics. *Stochastic Processes and their Applications*, 1975.

N.N. Taleb. *The Black Swan, the impact of the highly improbable.* Random House, 2007.

V.V. Uchaikin and V.M. Zolotarev. *Chance and stability: stable distributions and their applications.* VSP International Science Publishers, 1999.