1

Fatness of Tail

1.1. Fat tail heuristics

Suppose the tallest person you have ever seen was 2 m (6 ft 8 in). Someday you may meet a taller person; how tall do you think that person will be, 2.1 m (7 ft)? What is the probability that the first person you meet taller than 2 m will be more than twice as tall, 13 ft 4 in? 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). Before 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 based on 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 book provides an intuitive introduction to fat-tailed phenomena, followed by a rigorous mathematical overview of many of these intuitive features. A major goal is

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

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 due to Nassim Taleb's book *The Black Swan: The Impact of the Highly Improbable* ([TAL 07]). 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 central definitions for tail obesity. Older texts sometime speak of "leptokurtic distributions": 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 it characterizes the rate at which the probability of values greater than x go to zero as $x \to \infty$. For a large class of distributions, this rate is polynomial. Unless indicated otherwise, 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 will say, *tail index* α, has infinite *k*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 *k*th moments are infinite for all $\kappa \geq \alpha$. Chapter 4 unravels these issues, and shows

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.

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, such that 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 $X_1 + X_2 + \cdots + X_n$ behaves like the maximum of $X_1, X_2, \ldots X_n$. Distributions satisfying

$$Prob\{X_1 + X_2 + \ldots + X_n > x\} \sim Prob\{\max\{X_1,\ X_2, \ldots X_n\} > x\}, \quad x \to \infty$$

are called *subexponential*. Like regular variation, subexponentiality 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 subexponentiality" 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 novel definitions are given in Chapter 4.

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

– 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 $n \to \infty$; rather, the average is similar to the original variables;

– regression coefficients which putatively explain heavy-tailed variables in terms of covariates may behave erratically.

1.2. History and data

A detailed history of fat-tailed distributions is found in [MAN 08]. Mandelbrot himself introduced fat tails into finance by showing that the change in cotton prices was heavy-tailed [MAN 63]. Since then many other examples of heavy-tailed distributions have been found, among these we find data file traffic on the Internet [CRO 97], financial market returns [RAC 03, EMB 97] and magnitudes of earthquakes and floods [LAT 08, MAL 06], to name a few.[3]

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

1.2.1. *US flood insurance claims*

US flood insurance claims data from the National Flood Insurance Program (NFIP) are compiled by state and year for the years 1980–2008; the data are in US dollars. Over this time period, there has been substantial growth in exposure to flood risk, particularly in coastal states. To remove the effect of growing exposure, the claims are divided by personal

3 The website http://www.er.ethz.ch/presentations/Powerlaw_mechanisms_13July07.pdf provides examples of earthquake numbers per 5×5 km grid, wildfires, solar flares, rain events, financial returns, movie sales, health care costs, size of wars, etc.

income estimates per state per year from the Bureau of Economic Accounts (BEA). Thus, we study flood claims per dollar income by state and year[4].

1.2.2. *US crop loss*

US crop insurance indemnities paid from the US Department of Agriculture's Risk Management Agency are compiled by state and year for the years 1980–2008; the data are in US dollars. The crop loss claims are not exposure adjusted, since a proxy for exposure is not easy to establish, and exposure growth is less of a concern[5].

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, registers states-level damages and fatalities from weather events[6]. The basal estimates in SHELDUS are indications as the approach to compiling the data always employs the most conservative estimates. Moreover, when a disaster affects many states, the total damages and fatalities are apportioned equally over the affected states regardless of population or infrastructure. These data should therefore be seen as indicative rather than precise.

---

4 Help from Ed Pasterick and Tim Scoville in securing and analyzing this data is gratefully acknowledged.

5 Help from Barbara Carter in securing and analyzing this data is gratefully acknowledged.

6 Information on SHELDUS is available at http://webra.cas.sc.edu/hvri/products/SHELDUS.aspx.

1.2.4. *US hospital discharge bills*

Billing data for hospital discharges in a northeast US states were collected over the years 2000–2008; the data are in US dollars.

1.2.5. *G-Econ data*

This uses the G-Econ database [NOR 06] showing the dependence of gross cell product (GCP) on geographic variables measured on a spatial scale of 1°. At 45° latitude, a 1° by 1° grid cell is $[45\ mi]^2$ or $[68\ km]^2$. The size varies substantially from equator to pole. The population per grid cell varies from 0.31411 to 26,443,000. The GCP 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[7].

1.3. Diagnostics for heavy-tailed phenomena

If we look closely, we can find heavy-tailed phenomena all around us. Loss distributions are a very good place to look for tail obesity, but even something as mundane as hospital discharge billing data can produce surprising evidence. Many of the features of heavy-tailed phenomena would render our traditional statistical tools useless at best, and dangerous at worst. Prognosticators base their predictions on historical averages. Of course, on a finite sample the average and standard deviations are always finite; but these may not be

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

converging to anything and their value for prediction might be null. 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 do not converge, their ratio could whiplash wildly, taking our predictions with them. In this section, simple diagnostic tools for detecting tail obesity are illustrated using mathematical distributions and 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. Thereby, 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 of 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, the sample average tends to increase. One might mistakenly conclude that there is a time trend in the data; the universe is finite and an empirical sample would exhaust all data before it reached infinity. However, such reassurance is quite illusory; the question is, “where is the sample average going?”. A simple computer experiment is enough to convince sceptics: 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 shows the moving average of 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 distribution is equal to 1.

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 its real mean. Figure 1.2(a) shows the moving average of US property damage by 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) and (d). It is important to note that these are simply different orderings of the same data set (note the differences on the y-axes). Conclusive results 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 relative 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 borrow a data set from a Pareto distribution, then the ratio between two consecutive observations will also have a Pareto distribution.

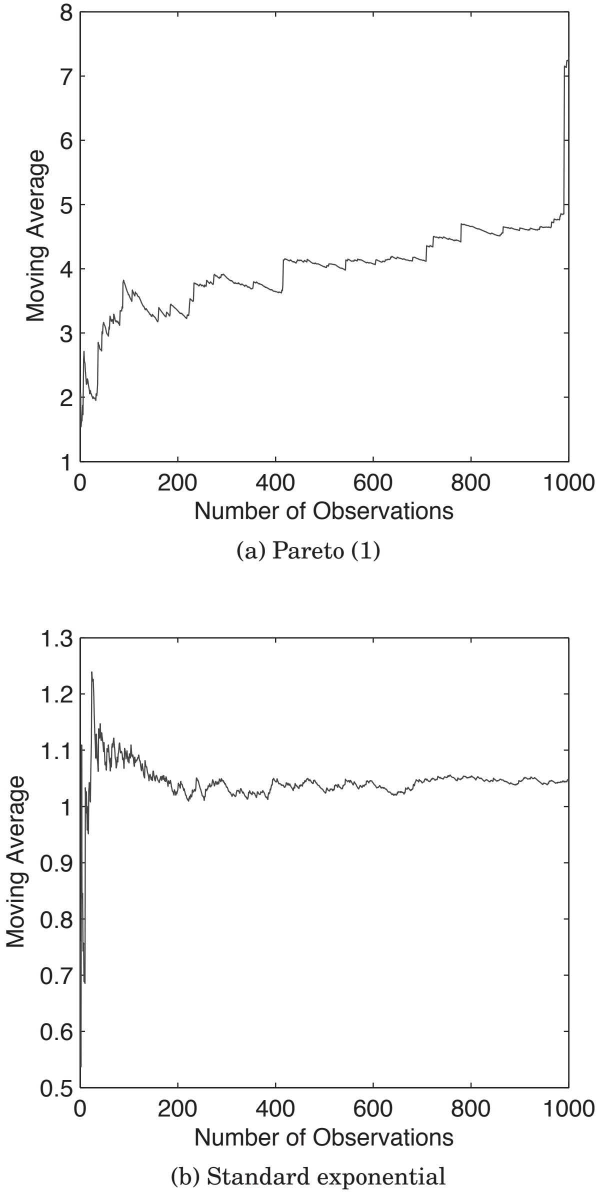
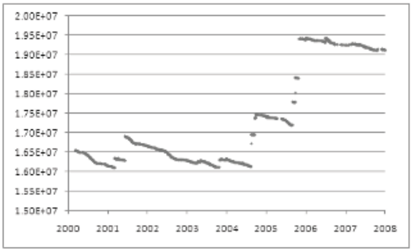
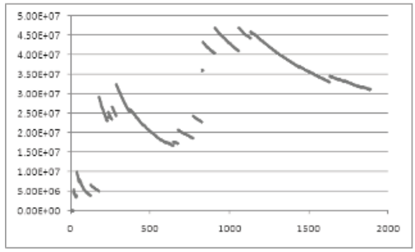


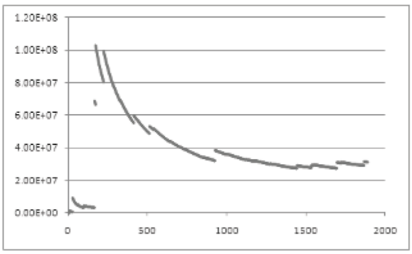
Figure 1.1. *Moving average of Pareto (1) and standard exponential data*



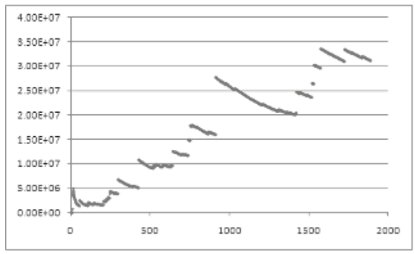
(a) Natural disaster property damage, temporal order



(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*

In Table 1.1, we see the 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 remains constant for the Pareto distribution, but it tends to zero for the standard normal distribution as the number of observations increases.

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 sized data sets*

Seeing that one or two very large data points confound their models, unwary actuaries may declare these “outliers” and discard them, reassured that the remaining data look “normal”. Figure 1.3 shows the yearly difference between insurance premiums and claims of the US NFIP [COO 09].

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” [HAY 11].

1.3.3. *Mean excess*

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

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

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:

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

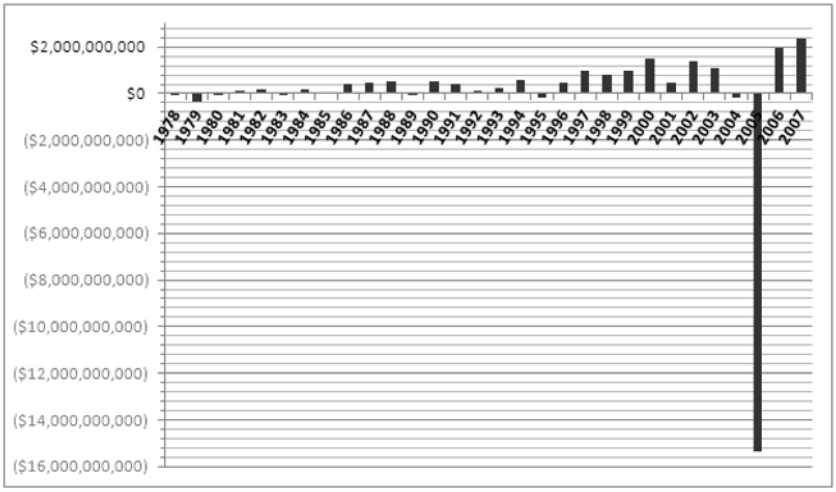


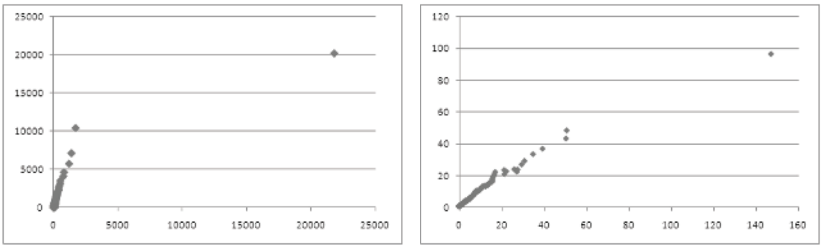
Figure 1.3. *US National Flood Insurance Program, premiums minus claims*

Figure 1.4(a) shows the mean excess plots of 5,000 samples from a Pareto (1) and Figure 1.4(b) shows a Pareto (2). At a glance, the slope in these plots clearly gives a better diagnostic for (b) than for (a).

Figure 1.5 shows the mean excess plots for flood claims per state per year per dollar income (Figure 1.5(a)) and insurance claims for crop loss per year per state (Figure 1.5(b)). Both plots are based roughly on the top 5,000 entries.

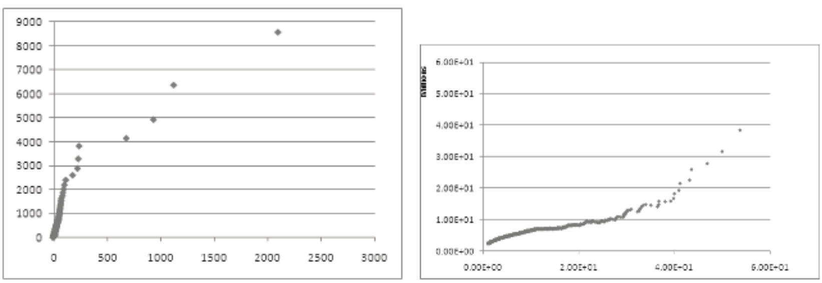
1.3.4. *Sum convergence: self-similar or normal*

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 [UCH 99] applies: the sum of these random variables, appropriately scaled, converges to a stable distribution having the same tail index as the original random variable.



(a) Mean excess Pareto 1 (b) Mean excess Pareto 2

Figure 1.4. *Pareto mean excess plots, 5,000 samples*



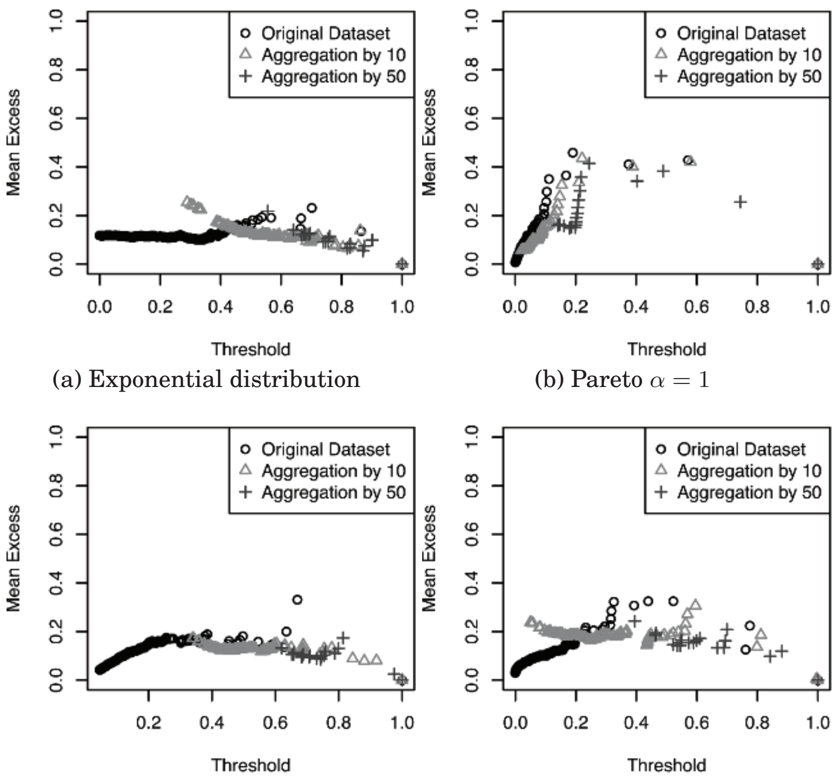
(a) Mean excess flood claims/income (b) Mean excess crop loss

Figure 1.5. *Mean excess plots, flood and crop loss*

This can be observed in the mean excess plot of data sets of 5,000 samples from a regularly varying distribution with tail index $\alpha < 2$. Let's define the operation of *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 will find that both mean excess plots are very similar. For data sets from thin-tailed distributions, both mean excess plots look very different.

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 $e(cu) = ce(u)$. Figures 1.6(a)–(d) show 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. The standardized mean excess plots of a data set obtained through aggregating by 10 and 50 are also shown in each plot. 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 5,000 samples is comparable. In this sense, the tail index does not tell the whole story.

Figures 1.7(a) and 1.7(b) show the standardized mean excess plot for two data sets. The standardized mean excess plot in Figure 1.7(a) is based on the income- and exposure-adjusted flood claims from the US NFIP from 1980 to 2006; this is the first data set. The second data set is the US crop loss. 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 1.7(a) seems to remain 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 1.7(b) changes when taking random aggregations, indicative of finite variance.



(a) Exponential distribution

(b) Pareto $\alpha = 1$

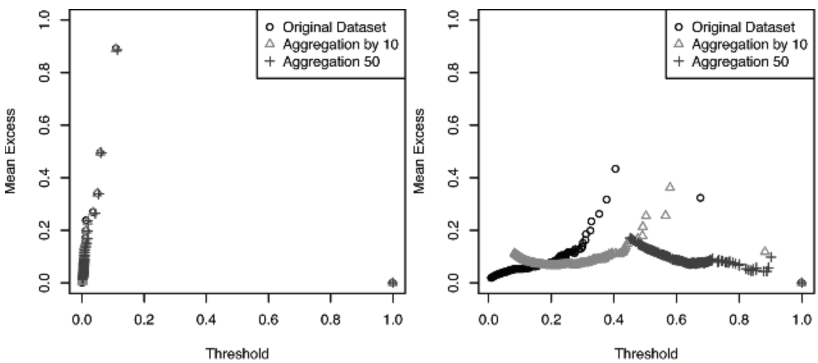
(c) Pareto $\alpha = 2$

(d) Weibull distribution $\tau = 0.5$

Figure 1.6. *Standardized mean excess plots. For a color version of this figure, see www.iste.co.uk/cooke/distributions.zip*

1.3.5. *Estimating the tail index*

Ordinary statistical parameters characterize the entire sample and can be estimated as such. Estimating a tail index is complicated by the fact that it is a parameter of a limit of a distribution function. 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. In this section, we briefly review the methods that have been proposed to estimate the tail index.



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

Figure 1.7. *Standardized mean excess plots of two data sets. For color version of this figure, see www.iste.co.uk/cooke/distributions.zip*

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 mentioned above, the latter method will not work for tail indices less than or equal to 1. The self-similarity of heavy-tailed distributions was used in [CRO 99] 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 pth moment is finite, then $lim_{n\to\infty}R(p,n) = 0$ [EMB 97]. Thus, if for some $p, R(p,n) \gg 0$ for large n, then this suggests an infinite pth 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 on the estimation of the extreme value index, for which we refer the readers to [EMB 97].

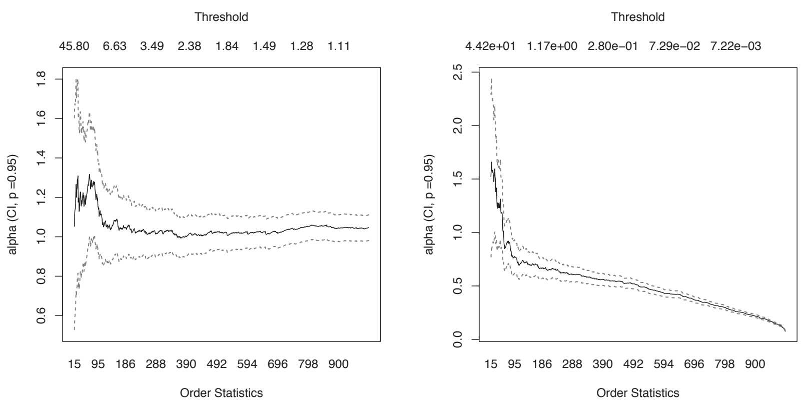
Perhaps the most popular estimator of the tail index is the Hill estimator, which is proposed in [HIL 75] and given by

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

where $X_{i,n}$ are such that $X_{1,n} \leq ... \leq X_{n,n}$. 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 $0.1n$. There are methods that choose k by minimizing the asymptotic mean squared error of the Hill estimator. Although the Hill estimator works very well for Pareto distributed data, for other regularly varying distribution functions it is less effective.

To illustrate this, we have drawn two different samples, one from the Pareto (1) distribution and the other from a Burr distribution (see Table 4.1) with parameters such that the tail index of the Burr distribution is equal to 1.

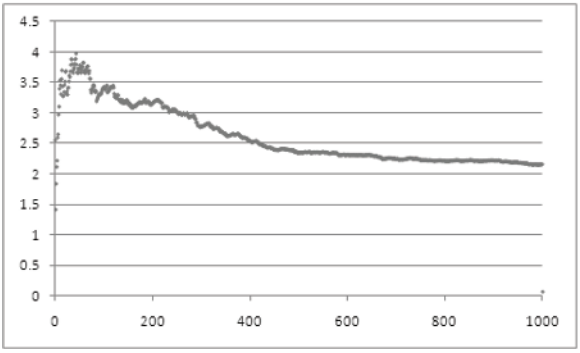
Figures 1.8(a) and 1.8(b) show the Hill estimator for the two data sets together with the 95%-confidence bounds of the estimate. Note that the Hill estimate is plotted against the different values in the data set running from the largest to the smallest, and the largest value of the data set is plotted on the left of the x-axis. As can be seen in Figure 1.8(a), the Hill estimator gives a good estimate of the tail index, but in Figure 1.8(b) it is not clear that the tail index is equal to 1. Beirlant *et al.* [BEI 05] explored various improvements of the Hill estimator, but these improvements require extra assumptions on the distribution of the data set.



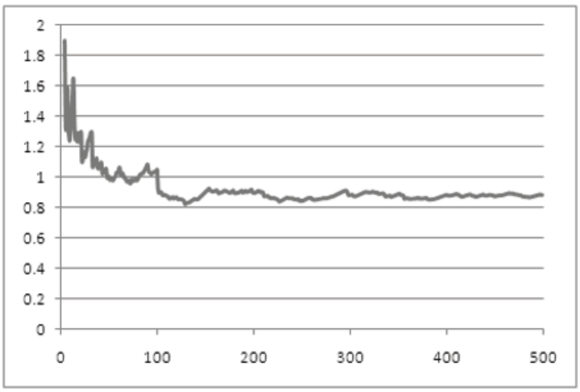
(a) Pareto distributed data (b) Burr distributed data

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

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).



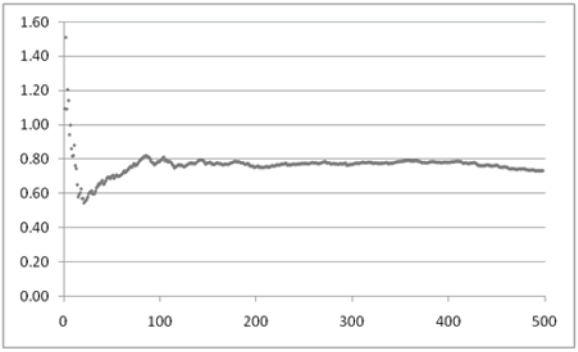
(a) Hill plot for crop loss



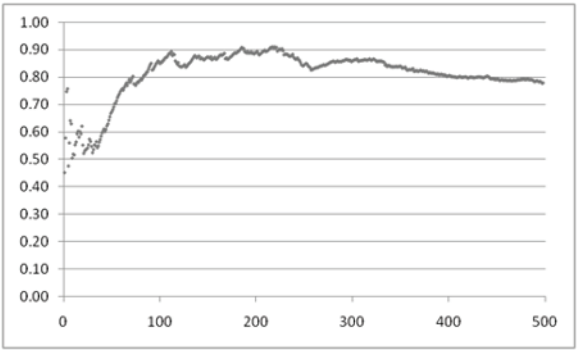
(b) Hill plot for property damages from natural disasters

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

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.



(a) Hill plot for flood claims

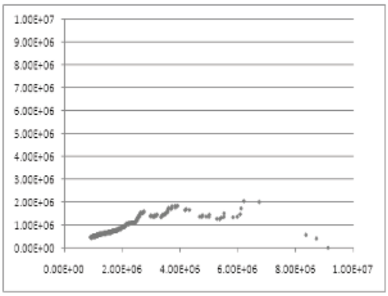


(b) Hill plot for flood claims per income

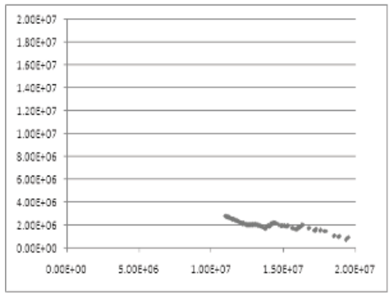
Figure 1.10. *Hill estimator for flood claims*

1.3.6. *The obesity index*

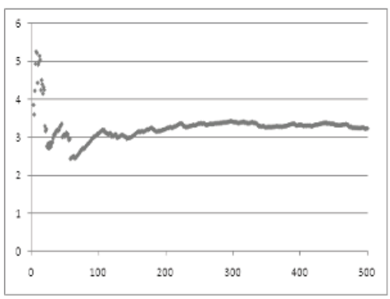
We have discussed two definitions of heavy-tailed distributions, the regularly varying distributions with tail index $0 < \alpha < \infty$ and subexponential distributions. Regularly varying distributions are a subset of subexponential distributions which have infinite moments beyond a certain point, but subexponentials include distributions all of whose moments are finite (tail index = ∞). 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.



(a) Mean excess plot for hospital discharge bills



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



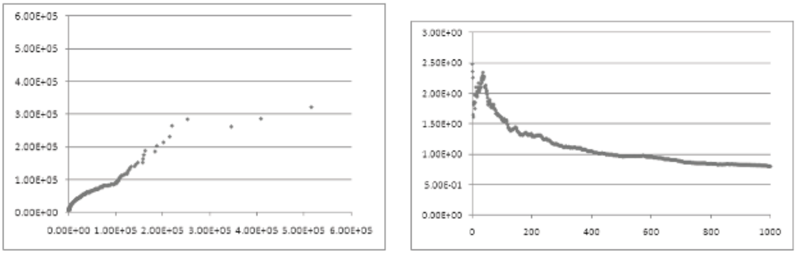
(c) Hill plot for hospital discharge bills

Figure 1.11. *Hospital discharge bills, obx = 0.79*

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

$$\text{Ob}(X) = P(X_1 + X_4 > X_2 + X_3 | X_1 \leq X_2 \leq X_3 \leq X_4);$$
$$\{X_1, ...X_4\} \text{ independent and identically distributed.}$$

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



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

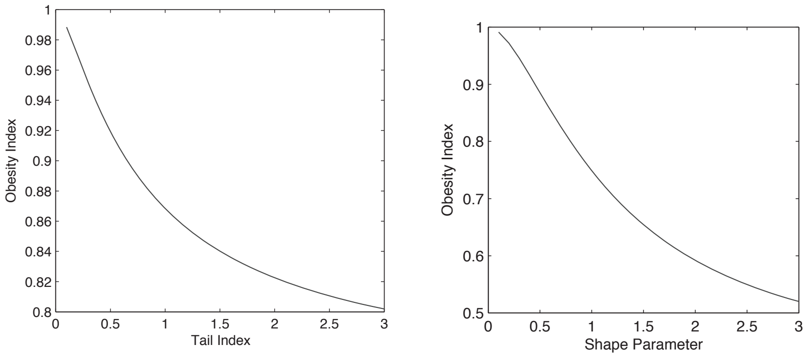
Figure 1.12. *Gross Cell Product (non-mineral) obx=0.77*

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

Table 1.2. *Obesity index for three distributions*

In 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.

In Chapter 5, we show that for the Pareto distribution, the obesity index is decreasing in the tail index. Figures 1.13(a) and 1.13(b) 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 the Weibull with shape 0.25 is much more obese than the Pareto (1).



(a) Pareto distribution (b) Weibull distribution

Figure 1.13. *Obesity index for different distributions*

Given two random variables X_1 and X_2 with tail indices, α_1 and α_2, $\alpha_1 < \alpha_2$, the question arises whether the obesity index of X_1 is larger than that of X_2. Numerical approximation of two Burr distributed random variables indicates that this is not the case. Consider X_1, a Burr distributed random variable with parameters $c = 1$ and $k = 2$, and X_2, 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 indicates 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 distribution that takes positive and negative values. For a symmetrical distribution, 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. Relation to reliability theory

Loss distributions (the subject of this book) and life distributions (the subject of reliability) are mathematical siblings, yet with very different personalities. It is therefore useful to point out some of their relationships.

Both reliability and loss analysis are concerned with non-negative random variables. In reliability, these are life variables whose realizations represent time of death or time of failure. Infinite expected lifetimes are very unusual. However, Pareto distributions do arise as mixtures of exponential distributions. The gamma density over λ with shape υ and scale α

$$f(\lambda) = \frac{\alpha^{\upsilon}\lambda^{\upsilon-1}e^{-\alpha\lambda}}{\Gamma(\upsilon)}$$

has mean υ/α and variance υ/α^2. If this is interpreted as a mixing distribution over possible failure rates λ, then the mixture density for an exponential life variable X with failure rate λ drawn from density $f(\lambda)$ is

$$p(x) = \int_{\lambda\in(0,\infty)} \frac{\lambda e^{-\lambda x}\alpha^{\upsilon}\lambda^{\upsilon-1}e^{-\alpha\lambda}}{\Gamma(\upsilon)} d\lambda.$$

Integrating over λ gives

$$p(x) = \frac{\upsilon\alpha^{\upsilon}}{(\alpha+x)^{\upsilon+1}}$$

which is a Pareto density with tail index υ. All the diagnostics for regularly varying distributions would apply in this case.

The mean excess curve is another point of contact between reliability and loss analysis. In reliability, the mean excess is called the mean residual life. Again, all the mean excess diagnostics will apply for mean residual life. The theory of records has not yet had a large footprint in reliability theory, but may someday. This may arise, for example, if we are interested in estimating the maximal service time of some component in a system of independent and identically distributed components. Finally, the obesity index may find applications in reliability as a non-parametric data-driven method for identifying sub- or superexponential behavior. These are among the research topics that remain to be explored.

1.5. 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 GCP 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 imprudent 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 be poor predictors regardless of the 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 1, as the theoretical slope is infinite. Hill plots, although 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; it can be computed for data sets and , in most cases numerically, for distribution functions.

In Chapters 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 proposed by us. 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.