Chapter 1
Heavy-Tailed Distributions and Their Properties

Abstract We define the heavy-tailed distribution as distribution with infinite mathematical expectation. For such distributions the standard statistical tools—sample mean and sample standard deviation—exhibit a high instability. Some examples illustrating this conclusion are presented. We discuss relations between the maximum event $M_{max}^{(n)} = max(x_1, \ldots, x_n)$ and the total sum $S_n = x_1 + \cdots + x_n$ for heavy-tailed distributions. The asymptotical proportionality of $M_{\max}^{(n)}$ and S_n is derived for the heavy-tailed Pareto distributions (Eq. (1.18)).

The main focus of our study is made on damages from disasters whose distribution can be described by power-like laws with a heavy tail. The disasters of this type typically occur in a very broad range of scales, the rare greatest events being capable of causing losses comparable with the total loss due to all the other (smaller) events. The disasters of such a type are the most unexpected ones and they frequently entail huge losses. One of the most important results of our study is the conclusion about the instability of the "maximum possible earthquake" parameter $M_{\max}$ which is frequently used in the seismic risk assessment. Instead of focusing on the unstable parameter $M_{\max}$ we suggest a stable and convenient characteristic $M_{\max}(\tau)$ defined as the maximum size of a given phenomenon that can be recorded over a future time interval τ. The random value $M_{\max}(\tau)$ can be described by its distribution function, or equivalently by its quantiles $Q_q(\tau)$, which are stable, robust characteristics in contrast to $M_{\max}$. Besides, if $\tau \to \infty$, then $M_{\max}(\tau) \to M_{\max}$ with probability one. The method of calculation of $Q_q(\tau)$ is exposed below. In particular, we can estimate $Q_q(\tau)$ for, say, $q = 90, 95$ and 99 %, as well as for the median ($q = 50$ %) for any desirable time interval τ. Our method provides an alternative and robust way to parameterize the rightmost tail of the frequency-size relation. The final goal of our data processing consists in obtaining a family of quantiles $Q_q(\tau)$ that corresponds to damage (fatalities in a disaster, economic damage, insurance losses etc.) that will not be exceeded in future τ years

with a prescribed probability q. This characteristic seems to be particularly useful, for instance in the insurance business.

It is well known that distributions of some parameters of natural processes (earthquake energy, economic damage and casualties from natural disasters and others) are often modeled by power-like laws, such as the Pareto distribution. The Pareto distribution function and probability density are given as follows:

$$F(x) \, = \, 1 - (h/x)^{\beta}; \, f(x) \, = \frac{\beta \cdot h^{\beta}}{x^{1+\beta}}; \quad x \geq h. \tag{1.1}$$

If the exponent β of such a distribution is less than unity, $\beta \leq 1$, then the mathematical expectation of the corresponding random variable is infinite. In this case, the standard statistical tools, such as sample mean and sample standard deviation, exhibit a high instability. Such distributions are often called distributions with *heavy tails*. It should be noted here that there is no commonly accepted definition of a heavy tailed distribution. Here we apply this term to distributions $F(x)$ whose mathematical expectation is infinite:

$$Ex \, = \, \int x dF(x) = \infty. \tag{1.2}$$

Accordingly, the opposite term *light tail* will designate distributions with finite expectation.

However there is a wide class of distributions where mathematical expectation is finite but higher statistical moments are infinite. Such distributions might be named "heavy-tailed" as well.

One may note that there are alternative definitions of the terms "heavy tail" and "light tail", see Reiss and Thomas (1997), Embrechts et al. (1997).

The Pareto law is a classical example of a heavy tailed distribution provided that $\beta <= 1$. In that case its expectation is infinite, so that the Law of Large Numbers is inapplicable, and the sample mean and the sample standard deviation are unstable. In contrast with a more common case of distributions with finite expectation, the increase of sample size does not improve the accuracy of the sample mean. The large statistical scatter of values of the sample mean (see e.g. Osipov 2002) makes this widely used statistical characteristic inadequate for applications to data sets with heavy tailed distributions. Moreover, its use in such situations may lead to essential errors and incorrect conclusions. Hereafter we are giving an illustration to this affirmation.

The total number of fatalities from typhoons, hurricanes and floods in the world during the period from 1947 till 1960 is about 900,000 (the flood causing the highest fatality of 1,300,000 occurred in China in 1931). The mean annual fatality (arithmetic mean) for the period from 1947 till 1960 was $\hat{X}_{ann} = 64,300$. The mean annual fatality for the period from 1962 till 1992 is $\hat{X}_{ann} = 36,000$. These sample means differ significantly and provide no indication about the probability of such a super-catastrophic event as the 1931 flood in China.

Let us consider another example: the total number of fatalities from earthquakes in the world during the period from 1947 till 1970 was about 151,000 (the earthquake causing the highest death toll of 240,000 occurred in China in 1976). The mean annual fatality rate (arithmetic mean) for the period from 1947 till 1970 is $\hat{X}_{ann} = 6,300$. The mean annual fatality rate for the period from 1962 till 1992 is $\hat{X}_{ann} = 18,600$. Here again, both sample means give quite an uncertain characteristic of annual fatality, which is a typical indication of a data set with the heavy-tailed distribution.

Figure 1.1 shows the non-normalized complementary sample distribution function for: *a*—annual casualties; *b*—number of victims in a single event; *c*—annual economic losses. These data are taken from the Web site of the US Geological Survey (http://www.neic.cr.usgs.gov/neis/eqlists). The Pareto law fits to the middle range and to the extreme events range are shown by dotted lines, with $\beta = 0.77 \pm 0.11$ (a); $\beta = 0.73 \pm 0.11$ (b); $\beta = 0.65 \pm 0.16$ (c). We can see that all the three distributions on Fig. 1.1 can be qualified as heavy-tailed.

Below we present some new approaches to the reliable statistical estimation in the situations where the standard statistical tools fail.

In the statistical analysis of data with heavy-tailed distributions an important role belongs to the maximum observed event $M_{max}^{(n)} = max(x_1, \ldots, x_n)$. It can be shown that for a data set with a heavy tailed distribution (according to definition (1.2)) the maximum event $M_{max}^{(n)}$ is commensurable with the total sum $S_n = x_1 + \cdots + x_n$, i.e. $M_{max}^{(n)}$ and S_n are of the same order. For example, in the case of the Pareto distribution with exponent $\beta < 1$ one has:

Fig. 1.1 Non-normalized complementary sample of world-wide losses from earthquakes distribution function for: **a** annual casualties; **b** casualties for individual event; **c** annual economic losses. The power-law functions fitting tails are shown by dotted lines, with $\beta = 0.77 \pm 0.11$ (**a**); $\beta = 0.73 \pm 0.11$ (**b**); $\beta = 0.65 \pm 0.16$ (**c**)
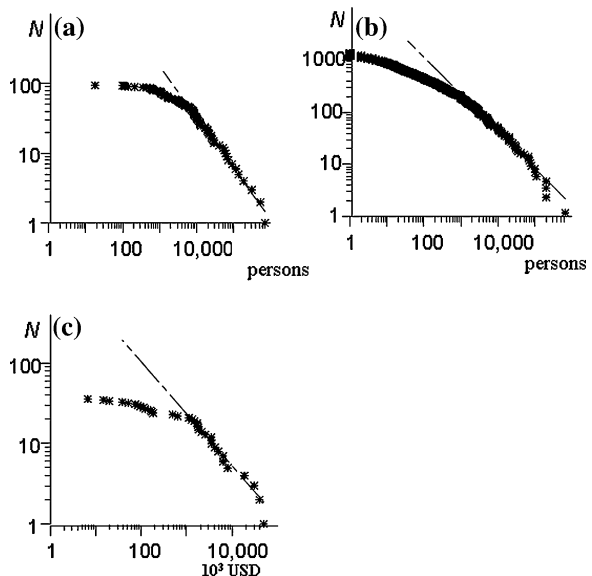
$$E\,\frac{S_n}{M_{\max}^{(n)}} \to \frac{1}{1-\beta}, \quad n \to \infty \tag{1.3}$$

Conversely, for a data set with a light tailed distribution (in which case the expectation is finite $E|x| < \infty$), the contribution of $M_{\max}^{(n)}$ in the total sum S_n decreases with the sample size n:

$$\frac{M_{\max}^{(n)}}{S_n} \to 0, \quad n \to \infty \tag{1.4}$$

If a sample is characterized by the Pareto distribution with index β, then $M_{\max}^{(n)}$ grows with the size of the sample as $n^{1/\beta}$. (We would like to emphasize that this tendency of a non-linear growth of $M_{\max}^{(n)}$ with the sample size n or, equivalently, with the observation time span τ can be incorrectly interpreted as an evidence of a non-stationarity in time (Pisarenko and Rodkin 2010). In many cases, the widespread belief that the rate of losses from natural disasters has a clear tendency of increasing with time is precisely connected with the misinterpretation of this apparent non-stationarity effect).

The treatment of heavy-tailed data is often facilitated by using logarithms of original values. Switching to logarithms (which can be done only when the original numerical values are positive) ensures almost always that all the statistical moments exist, and hence the Law of Large Numbers and the Central Limit Theorem are applicable to the sums of logarithms. We are advocating the use of the logarithmic transform of the original data values as one of the main tools to prevent complications in the processing of heavy-tailed data sets. As we shall see later, the use of the logarithmic transform turned to be efficient in all the considered cases except one (the data set on the fatality of floods in USA from 1995 till 2011).

It should be remarked that if X has the Pareto distribution, then $\log(X)$ has exactly exponential distribution. We shall use this fact below.

Another statistical tool helping to overcome difficulties connected with processing heavy tailed data is *the order statistics*: sample median, sample quantiles, interquartile range etc. The order statistics can be used also for construction of confidence intervals. The main statistical tool suggested in this short-monograph for description of distribution tail—the family of quantiles $Q_q(\tau)$—is in fact a continuous analog of the sample ordered statistics. This explains its robustness and stability.

In our approach we model the sequence of disaster occurrences by well-known Poisson point process, see Embrechts et al. (1997). The disaster effects can be different: fatalities and economic losses due to natural catastrophes, earthquake seismic moments and ground acceleration at a fixed point etc. These effects are "marks" assigned to occurrence times of the point Poisson process. Thus, the whole construction is called the marked point process. In our applications we shall model our catalog usually by a stationary Poisson process with intensity λ events

per year (but we devote the special section *Non-stationarity of natural processes, "operational time" method* to the case of non-stationary Poisson processes). The random number of events in stationary process for T years is a random Poisson variable with the mean λT. Let us denote the maximum event for this time period as M_T and the total sum as $\sum_T$. We are now going to derive some relations between M_T and $\sum_T$. It is well known (see, e.g., Gumbel 1958) that for distributions with a light, exponential tail M_T increases as the logarithm of T:

$$M_T \cong c \cdot \log(\lambda T), \tag{1.5}$$

where c is some constant, whereas $\sum_T$ increases as T, in accordance with the CLT:

$$\sum\nolimits_T \cong \lambda T \cdot b + \zeta \cdot \sigma \cdot (\lambda T)^{1/2}, \tag{1.6}$$

where ζ is some standard normal rv, b is the expectation of a single event (here we assume $b > 0$), and σ is the standard deviation (std) of a single event. We consider the ratio $R(T)$ of the total sum to maximum event:

$$R(T) = \sum\nolimits_T / M_T.$$

It follows from (1.5), (1.6) that

$$R(T) \cong (b/c) \cdot (T/\log(\lambda T)). \tag{1.7}$$

Hence it follows that the $R(T)$ ratio increases with T linearly for the cases of light tail if we disregard the slowly varying $\log(\lambda T)$. A quite different behavior of $R(T)$ appears for heavy-tailed distributions. For such distributions $R(T)$ grows much slower and can even have a finite expectation (cf. with (1.3)). In other words, in this case the total and the maximum event become comparable, i.e., the total sum is determined in a large extent by the single maximum event.

In order to illustrate this assertion, we are going to derive lower and upper bounds for $R(T)$ corresponding to the Pareto distribution. Taking logarithms and their expectation we get

$$E \log \sum\nolimits_T = E \log R(T) + E \log M_T \tag{1.8}$$

Using Jensen's inequality for concave functions, we get:

$$E \log R(T) \leq \log E R(T). \tag{1.9}$$

As shown in Pisarenko (1998),

$$[1 - (\lambda T)^{1-1/\beta} \cdot \gamma(1/\beta; \lambda T)]/(1-\beta); \qquad \beta \neq 1; \tag{1.10}$$

$ER(T) =$

$$\log(\lambda T) - \exp(-\lambda T) \cdot (\log \lambda T - 1); \qquad \beta = 1, \tag{1.11}$$

Table 1.1 Expectation $E\,R(T)$ as function of λT and β

β	λT (or n)						
	10	50	100	300	500	1000	∞
3	5.77	17.93	28.8	60.7	85.7	136.6	∞
2	4.6	11.52	16.7	29.7	38.6	55	∞
1.3	3.48	6.55	8.26	11.6	13.5	16.42	∞
1.1	3.1	5.16	6.15	7.85	8.7	9.91	∞
1.0	2.88	4.49	5.18	6.28	6.78	7.48	∞
0.9	2.59	3.81	4.26	4.92	5.2	5.56	10
0.8	2.45	3.29	3.57	3.91	4.04	4.19	4.3
0.6	2.02	2.33	2.4	2.45	2.46	2.48	2.5

where $\gamma(\alpha; t)$ is the incomplete gamma function. The function $ER(T)$ is tabulated in Table 1.1.

If $\lambda T \to \infty$, then in (1.10) the incomplete gamma function $\gamma(1/\beta;\ \lambda T)$ tends to the standard gamma function $\Gamma(1/\beta)$. It is possible to derive an explicit expression for $E \log(M_T)$ (Pisarenko 1998):

$$E \log M_T = [\log(\lambda T) + C - Ei(-\lambda T)/\beta] \tag{1.12}$$

where C is the Euler constant ($C = 0.577\ldots$), and $Ei(-\lambda T)$ is the integral exponential function

$$Ei(x) = \int_{-\infty}^{x} \frac{\exp(z)}{z} dz; \quad x < 0.$$

Combining Eqs. (1.8)–(1.12), we derive an upper bound on $E \log \Sigma_T$:

$$E \log \sum\nolimits_T \leq \log \frac{1 - (\lambda T)^{1-1/\beta}\gamma(1/\beta; \lambda T)}{1 - \beta} + [\log(\lambda T) + c - Ei(-\lambda T)]/\beta. \tag{1.13}$$

If $\lambda T \gg 1$, then (1.13) can be simplified. Keeping only terms growing with (λT), we get:

$$E \log \sum\nolimits_T \leq max\left(1;\ \frac{1}{\beta}\right) \cdot \log(\lambda T) \tag{1.14}$$

Now we derive a lower bound for $E \log \Sigma_T$. If $\beta > 1$ then by virtue of the Law of Large Numbers $\frac{1}{n}\Sigma_T \to 1/\beta$ and we get:

$$E \log \sum\nolimits_T = E \log\left(n \frac{1}{n}\sum\nolimits_T\right) = E \log(n) + 1/\beta \cong \log(\lambda T) + 1/\beta \tag{1.15}$$

If $\beta < 1$, then we just drop the first term in rhs of (1.8) and get:

$$E \log \sum\nolimits_T > E \log M_T = [\log(\lambda T) + C - Ei(-\lambda T)/\beta] \qquad (1.16)$$

We can put down (1.15) and (1.16) in one relation that holds true for any β and disregards terms of lower order:

$$E \log \sum\nolimits_T \geq \max\left(1; \frac{1}{\beta}\right) \cdot \log(\lambda T) \qquad (1.17)$$

We get from (1.14) and (1.17) that asymptotically (as $\lambda T \to \infty$) up to terms of lower order for any β:

$$E \log \sum\nolimits_T = \max\left(1; \frac{1}{\beta}\right) \cdot \log(\lambda T). \qquad (1.18)$$

Thus, one can say that for $\beta < 1$ random quantities $\log \Sigma_T$ and $\log M_T$ are comparable in value and both grow as $\log(\lambda T)^{1/\beta}$. This fact might be interpreted as a nonlinear growth of Σ_T and M_T at the same rate as $(\lambda T)^{1/\beta}$ does, since their logarithms are asymptotically proportional. We recall that both random quantities have infinite expectations if $\beta < 1$. If $\beta > 1$ then Σ_T increases linearly with T (cf. with (1.6)), whereas M_T increases more slowly, as $T^{1/\beta}$. The relation (1.18) holds true for any probability density that decreases asymptotically as a power.

References

Embrechts P, Klueppelberg C, Mikosch T (1997) Modelling extremal events. Springer, Berlin

Gumbel EJ (1958) Statistics of extremes. Columbia University Press, New York

Osipov VI (2002) Natural hazards control. Vestnik RAN 8:678–686 (in Russian)

Pisarenko VF (1998) Non-linear growth of cumulative flood losses with time. Hydrol Process 12:461–470

Pisarenko VF, Rodkin MV (2010) Heavy-tailed distributions in disaster analysis. Springer, Dordrecht

Reiss RD, Thomas M (1997) Statistical analysis of extreme values. Birkhauser, Basel