Chapter 4

Heavy-tailed distributions for agent-based economic modelling

4.1 Introduction

This chapter is devoted to the parametric statistical distributions of economic size phenomena of various types.

Probability distributions of size variables are usually taken as the first quantitative characterization of complex systems, allowing one to detect the possible occurrence of regularities and to identify the underlying mechanisms at their origin—and thus at the origin of the behaviour of the complex system under study.

A rapid survey covers the class of "heavy-tailed" distributions decreasing slower than exponentially at infinity. The fascination for "power laws" is then explained, starting from the statistical approaches for quantifying and testing a power-law distribution from your data, and ending with a (not exhaustive) list of mechanisms leading to power-law distributions. The description of distributions is ultimately enlarged by proposing the Laplace distribution, which has both tails—the upper and the lower—heavier than a standard Gaussian.

Exposition is conveyed by means of computer-based examples designed to assist the reader in understanding of the broad topics that are touched on in other parts of this book. Full code for examples using the `R` software environment (R Core Team, 2015) can be found on the book's website.

Here are some notations and vocabulary that will be useful in the remainder of the chapter.

- *Probability density function.* Consider a random variable X whose outcome is a real number x. The probability density function (PDF) of X, $p(x)$, is such that the probability that X is found in a small interval Δ around x is $p(x)\,\Delta$. The probability that X is between x and $x + \Delta$ is therefore given by the integral of $p(x)$ between x and $x+\Delta$, that is

$$p(x) = \lim_{\Delta \to 0^+} \frac{\Pr(x \le X < x + \Delta)}{\Delta} = \int_x^{x+\Delta} p(x)\, \mathrm{d}\, x. \tag{4.1}$$

The PDF $p(x)$ depends on the measurement unit used to quantify x and has the dimension of the inverse of x, such that $p(x)\Delta$—being a probability, i.e. a number between 0 and 1—is dimensionless. The empirical estimation of the PDF $p(x)$ is usually plotted with the horizontal axis scaled as a graded series for the variable under consideration and the vertical axis scaled for the values of $p(x)$.

- *Cumulative distribution function.* In many cases it is useful to consider the cumulative distribution function (CDF) of X, $P_{\le}(x)$, defined by

$$P_{\le}(x) = \Pr(X \le x) = \int_{-\infty}^{x} p(y)\, \mathrm{d}\, y. \tag{4.2}$$

  The CDF $P_{\le}(x)$ gives the fraction of events with values less than (or equal) to x and increases monotonically with x from 0 to 1.

- *Complementary cumulative distribution function.* This is defined by

$$P_{>}(x) = \Pr(X > x) = 1 - \Pr(X \le x) = \int_{x}^{\infty} p(y)\, \mathrm{d}\, y. \tag{4.3}$$

  The complementary cumulative distribution function (CCDF) gives the fraction of events with values greater than x and decreases monotonically with x from 1 to 0.

4.2 Heavy-tailed distributions

Heavy-tailed distributions are probability distributions whose tails decay to 0 far slower than exponentially. In many applications it is the right tail of the distribution that is of interest, but a distribution may have a heavy left tail or both of its tails might be heavy.

The exponential distribution given by the PDF

$$p(x) = \lambda e^{-\lambda x}, \quad x \ge 0, \quad \lambda > 0, \tag{4.4}$$

is often considered as the boundary between classes of heavy-tailed and light-tailed distributions—although occasionally the term "heavy-tailed" is used for any distribution that has heavier tails than the Gaussian distribution. Typical examples of heavy- and light-tailed distributions are given in Table 4.1. Their (right) tail behaviour is compared in Figure 4.1.

The class of heavy-tailed distributions comprises, among others, the subset of distributions with *regularly varying* tails (Bingham et al., 1987). Intuitively, a "regularly varying" distribution of a real variable is a distribution

Table 4.1 Examples of heavy- and light-tailed distributions

Heavy-tailed	Lognormal, Weibull with exponent less than 1, power laws (with regularly varying tails)
Light-tailed	Exponential, Gaussian, Weibull with exponent greater than 1

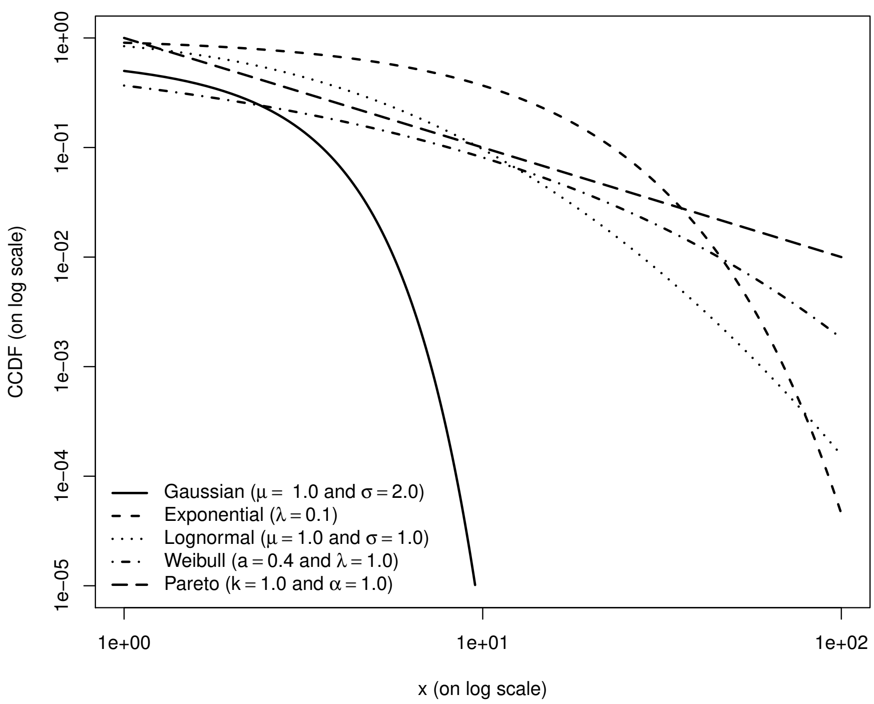


Figure 4.1 Heavy- and light-tailed distributions: comparison of right tail behaviour

whose behaviour near infinity is similar to the behaviour of a power-law function. Regularly varying distributions (also commonly referred to as "fat-tailed" distributions) are always heavy-tailed. Some distributions, however, have a tail which goes to zero slower than an exponential function (meaning they are heavy-tailed) but faster than a power (meaning they are not fat-tailed). An example is the lognormal distribution. Many other heavy-tailed distributions—*in primis* the power-law distributions—are, instead, also fat-tailed.

For light-tailed distributions and distributions with no regularly varying heavy tails, all moments exist and are finite. In contrast, for regularly varying distributions the moments exist only up to (and excluding) their tail indices.

Some more details about specific examples of heavy-tailed distributions are given in the following list.

- The *lognormal distributionLognormal distribution*. This is given by the PDF

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}, \quad x > 0, \tag{4.5}$$

for parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. All moments of the lognormal distribution are finite. The obvious advantage of the lognormal distribution is that, following a simple transformation, the enormous armature of inference for Gaussian distributions is readily available: indeed, a (positive) random variable X has a lognormal distribution with parameters μ and σ if and only if $\ln(X)$ has a Gaussian distribution with mean μ and variance σ^2. Hence, by taking the logarithm of the data points, the techniques developed for the Gaussian distribution can be used to estimate the parameters of the lognormal distribution. In particular, the location parameter μ is equal to the mean of the logarithm of the data points, and the shape parameter σ is equal to the standard deviation of the data set after transformation. The lognormal distribution has been widely used in many applications.[1]

- The *Weibull distribution*. It has the PDF

$$p(x) = a\lambda x^{a-1} e^{-\lambda x^a}, \quad x \geq 0, \tag{4.6}$$

where $a, \lambda > 0$. The exponent a controls the thin versus heavy nature of the right tail; in particular:

  - for $a = 2$, the Weibull distribution has the same asymptotic right tail as the Gaussian distribution;
  - for $a = 1$, expression (4.6) recovers the pure exponential distribution;
  - for $a < 1$, the right tail of the Weibull is fatter than an exponential.

Hence, the Weibull is a versatile distribution "interpolating" between thin-tailed distributions—such as the Gaussian or the exponential—when $a \geq 1$ and heavy-tailed distributions if (and only if) $a < 1$. Although there are many ways to estimate the parameters a and λ based on a random sample $X = (x_1, \ldots, x_n)$, maximum likelihood estimation is generally the most popular method. However, the maximum likelihood estimators of the Weibull parameters are not available in closed form, which implies that no explicit solution to the likelihood equations exists and numerical

[1] In the economic field, for example, Gibrat (1931) observed that the size distribution of French manufacturing establishments closely resembled the lognormal distribution. This led him to suggest a "law of proportionate effect" (see Section 4.3.5). Later, Aitchison and Brown (1954, 1957) argued that the lognormal hypothesis is particularly appropriate for the distribution of incomes, although much depends on the definition of income and the particular part of the distribution in which one happens to be interested—indeed, the above-cited authors consider the lognormal functional form most appropriate for modelling the distribution of earnings in fairly homogeneous sections of the workforce, but when examining the distribution of income from all sources they find in many instances that lognormality remains a reasonable assumption for the bulk of the incomes, whereas the upper tail appears to be governed by a different probabilistic law.

methods have to be used. In contrast, notwithstanding its heavy-tailed nature, the Weibull distribution has all its moments finite.[2]

- The *power-law distributions*. These are the canonical example of heavy-tailed distributions. Power-law distributions occur in many areas of scientific interest, including economic analysis, and their identification in empirical data is often interpreted as evidence for (or suggestion of) complex underlying processes. Although power law-distributions are attractive for their simplicity—they are straight lines on log-log plots—a technical difficulty is that not all moments exist for these distributions, which means, among other things, that the central limit theorem can no longer hold. Also, demonstrating that data do indeed follow a power-law distribution is not as straightforward as it might appear, as it involves more than simply fitting. Indeed, several alternative functional forms can appear to follow a power-law form over some extent, such as the above-mentioned lognormal and Weibull distributions. Thus, validating a power-law distribution as a possible adequate representation of a given data set requires more sophisticated statistics. For these reasons, in the following section we shall review the basic definitions and properties of power-law distributions as well as the commonly used statistical methods for discerning and quantifying them in empirical data. We shall also focus on some of the underlying generative models that lead to these distributions.

Table 4.2 summarizes the basic properties for the above-listed distributions and other parametric models that will be touched upon later on.

4.3 Power-law distributions

A first goal of research on economic complexity has been the determination of the ways in which complex systems represent an alternative to standard (neoclassical) economic theory. At the same time, economic complexity has from its inception been strongly motivated by the desire to explain empirical phenomena.

One of the main areas of work on the complexity/empirical interface consists of the identification of data patterns that conform with the features of complex systems (Durlauf, 2005). Specifically, a major effort of this work has been devoted to detect where *power laws*, which represent a particular class of probability distributions, occur in various economic data (Brock, 1999).

The power-law literature has identified a number of interesting statistical properties of different economic data and has made a valuable contribution in

[2] The Weibull distribution has received maximum attention in the engineering literature. In physics, it is known as the “stretched exponential distribution”. In the economic literature the Weibull is probably less prominent, but D’Addario (1974) noticed its potential for income data—although it has been used only sporadically as an income distribution (some applications can be found in Bartels and van Metelen, 1975, Bartels, 1977, Espinguet and Terraza, 1983, McDonald, 1984, Atoda et al., 1988, Bordley et al., 1996, Brachmann et al., 1996, and Tachibanaki et al., 1997). The Weibull has also been used by Di Guilmi et al. (2004, 2005) to study the empirical distribution of business cycle phases, that is of expansions and contractions. The authors find that in both cases the best fitting distribution is Weibull, although the parameters identifying the Weibull fitting models differ between upturn and downturn episodes.

Table 4.2 Statistical distributions mentioned in this chapter. For each distribution we give the basic functional form of the probability density/mass function, $p(x)$, the cumulative distribution function, $P_{\leq}(x)$, and the raw moment, $\langle X^r \rangle$

	Distribution	$p(x)$	$P_{\leq}(x)$	$\langle X^r \rangle$
Continuous	Gaussian[a]	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0$	$\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^r e^{-u^2} \, \mathrm{d}\, u$, $r \geq 0$
	Exponential[b]	$\lambda e^{-\lambda x}$, $x \geq 0, \quad \lambda > 0$	$1 - e^{-\lambda x}$	$\frac{\Gamma(r+1)}{\lambda^r}$, $r \geq 0$
	Lognormal[a]	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}$, $x > 0, \quad \mu \in \mathbb{R}, \quad \sigma > 0$	$\frac{1}{2}\left\{1 + \mathrm{erf}\left[\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right]\right\}$	$e^{r\mu + \frac{r^2\sigma^2}{2}}$, $r \geq 0$
	Weibull[b]	$a\lambda x^{a-1} e^{-\lambda x^a}$ $x \geq 0, \quad a > 0, \quad \lambda > 0$	$1 - e^{-\lambda x^a}$	$\lambda^{-\frac{r}{a}} \Gamma\left(1 + \frac{r}{a}\right)$, $r \geq 0$
	Pareto	$\frac{\alpha k^\alpha}{x^{\alpha+1}}$, $x \geq k, \quad k, \alpha > 0$	$1 - \left(\frac{x}{k}\right)^{-\alpha}$	$\frac{\alpha k^r}{\alpha - r}$, $r < \alpha$
	Laplace[c]	$\frac{1}{2\lambda} e^{-\frac{\lvert x-\theta \rvert}{\lambda}}$, $x \in \mathbb{R}, \quad \theta \in \mathbb{R}, \quad \lambda > 0$	$x \leq \theta : \frac{1}{2} e^{-\frac{\lvert x-\theta \rvert}{\lambda}}$ $x \geq \theta : 1 - \frac{1}{2} e^{-\frac{\lvert x-\theta \rvert}{\lambda}}$	$r! \sum_{j=0}^{r} \frac{1+(-1)^{j+r}}{2j!} \theta^j \lambda^{r-j}$, $r \geq 0$
Discrete	Zipf[d]	$\frac{x^{-s}}{\zeta(s)}$, $x \in \mathbb{Z}_{>0}, \quad s > 1$	$\frac{H_{x,s}}{\zeta(s)}$	$\frac{\zeta(s-r)}{\zeta(s)}$, $r < s - 1$

[a] erf $(\cdot)$ denotes the *error function* (http://mathworld.wolfram.com/Erf.html)
[b] $\Gamma(\cdot)$ denotes the *gamma function* (http://mathworld.wolfram.com/GammaFunction.html)
[c] $r!$ denotes the *factorial* of the non-negative integer r, defined as $r! \equiv r \times (r-1) \times \cdots \times 2 \times 1$ (http://mathworld.wolfram.com/Factorial.html)
[d] $H_{x,\alpha}$ is a *generalized harmonic number* (http://mathworld.wolfram.com/HarmonicNumber.html); $\zeta(\cdot)$ is the *Riemann zeta function* (http://mathworld.wolfram.com/RiemannZetaFunction.html)

identifying a range of "facts" that should help constrain theoretical modelling, such as the presence of skewed and thick tailed densities in data that points toward the existence of pervasive heterogeneity over individual agents. The attempt to identify the presence of such statistical properties in economic data has to a substantial extent led by physicists, whose attention was primarily focused on analysis of financial markets because of the large quantities of data available at high frequencies.[3] However, power laws have also fascinated economists of successive generations, due to their occurrence in varied economic contexts such as income and wealth, the size of cities and firms, stock market returns, trading volume, international trade and executive pay.[4]

Power laws have both practical importance and theoretical implications for economic research. For instance, a central argument maintained by Gabaix (2011) is that in an economy with fat-tailed distributions of agents' characteristics individual shocks do not die out in the aggregate, but create an amount of volatility that can be held responsible for a large part of macroeconomic fluctuations. This contrasts with existing research, which has focused on using

[3]Indeed, starting in the mid-1990s a new field of research has emerged within the physics community that has come to be known as "econophysics", where a major research activity is represented by efforts to find power laws in different socio-economic data sets. This literature is well surveyed by Săvoiu and Simăn (2013).

[4]See Gabaix (2009) for a survey of various power laws both within and outside economics.

economy-wide shocks to explain fluctuations of economic aggregates, arguing by the central limit theorem that shocks at the micro-level average out in the aggregate.

Roughly, the *central limit theorem* (*CLT*) states that the mean of n independent and identically distributed (i.i.d.) terms with a finite variance converges to a Gaussian distribution for an asymptotically large value of n regardless of the original distribution (e.g. Feller, 1971). To be more precise, let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. The CLT states that the distribution of the centred and normalized mean will have the standard Gaussian shape with zero mean and unit variance as n goes to infinity, that is

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1), \tag{4.7}$$

where $\bar{X}_n = (X_1 + \cdots + X_n)/n$ and $\xrightarrow{d}$ denotes convergence in distribution. Hence, in an economy where n agents have independent shocks, aggregate fluctuations should have a size proportional to $n^{-\frac{1}{2}}$. Given that modern economies can have millions of agents, this suggests that micro-level fluctuations will have a negligible aggregate effect if the fundamental hypotheses of the CLT are verified—i.e. if both independence and finite variance conditions of the n random variables X_i are satisfied—as aggregate volatility decays according to $n^{-\frac{1}{2}}$.

When agents' characteristics are power-law distributed, however, the finite variance assumption cannot always be achieved and the main conditions under which the CLT holds are not satisfied. Therefore, other existing limit theorems must be considered when fat-tailedness makes the classical CLT break down. Lévy (1954) discovered that, in addition to the Gaussian law, there is a rich class of probability distributions allowing for fat tails and sharing the convergence condition. Accordingly, a less well-known version of the CLT, the *generalized central limit theorem* (*GCLT*), shows that if the finite variance assumption is dropped, the only possible resulting limit is a "Lévy α-stable law" (Nolan, 2015). To be more rigorous, suppose $X_1, \ldots, X_n$ is a sequence of non-negative i.i.d. random variables having power-law tails with exponent $1 < \alpha \leq 2$ (implying finite mean but infinite variance). Then, as $n \to +\infty$,

$$\frac{\bar{X}_n - \mu}{a_n} \xrightarrow{d} L_\alpha(1, \beta, 0), \tag{4.8}$$

where the normalizing coefficient can be taken as $a_n \propto n^{-\left(1-\frac{1}{\alpha}\right)}$ (Uchaikin and Zolotarev, 1999, p. 62). In the equation above, $L_\alpha(1, \beta, 0)$ is the four-parameter Lévy α-stable distribution with location parameter $\delta \in \mathbb{R}$ equal to zero and scale parameter $\gamma > 0$ equal to 1. $\beta \in [-1, 1]$ is the skewness parameter quantifying the asymmetry of the distribution, which is symmetric around zero when $\beta = 0$. The exponent $\alpha \in (0, 2]$, a.k.a. the index of stability, determines the rate at which the tails of the distribution taper off. The GCLT (4.8) applies under the same restrictions (except for the finiteness of the variance when $\alpha > 2$) of independence and of large n, but an important difference is observed between the Gaussian and stable non-Gaussian attractors: as $n \to +\infty$, the Gaussian law "attracts" all the probability density functions decaying as $|x|^{-\alpha}$ at large $|x|$ for $\alpha \geq 2$; similarly, as n gets large, all probability density functions whose tails decay according to a power law with exponent

$\alpha < 2$ are attracted to the Lévy law. Therefore, there is an abrupt change in the tail behaviour of Lévy α-stable laws at the borderline case with exponent $\alpha = 2$: while for $\alpha < 2$ all the Lévy distributions are heavy-tailed, the case $\alpha = 2$ is special and represents the familiar, not heavy-tailed, Gaussian distribution—i.e. $L_2(1,0,0) = N(0,1)$, and the classical CLT (4.7) is recovered with the replacement $\alpha = 2$ in Equation (4.8).

By the above reasoning, the fat-tailed (power-law) nature of individual-level shocks is important theoretically, as it determines whether the classical CLT applies. Indeed, Equation (4.7) states that if the distribution of agents' characteristics has thin tails, then aggregate volatility decays according to $n^{-\frac{1}{2}}$. In contrast, Equation (4.8) states that if the agents' characteristics distribution has fat tails ($\alpha < 2$), then aggregate volatility decays much more slowly than $n^{-\frac{1}{2}}$: it decays as $n^{-\left(1-\frac{1}{\alpha}\right)}$. Hence, purely microeconomic shocks do not die out in the aggregate, but have a rather concrete possibility to propagate throughout the economy so as to generate non-trivial macroeconomic fluctuations.

The rest of this section is structured as follows. Section 4.3.1 provides some basis for the analysis of power-law distributed data. Section 4.3.2 illustrates how to recognize power laws in empirical data, whereas Section 4.3.3 presents the main statistical techniques for identifying the lower bound to the power-law behaviour (Section 4.3.3) and fitting the power-law form to empirical distributions (section 4.3.3). Then, Section 4.3.4 shows how to test the power-law hypothesis quantitatively and how to compare it with alternative distributions. Finally, Section 4.3.5 describes some of the mechanisms by which power-law behaviour can arise.

4.3.1 Some basic definitions and properties

A non-negative random variable X is said to obey a *power law* if its realizations are drawn from a probability distribution

$$p(x) \propto x^{-(\alpha+1)}, \tag{4.9}$$

where $\alpha > 0$ is a constant parameter of the distribution known as the *scaling exponent*.[5] At the most basic level, there are two types of power-law distributions: *continuous*, governing continuous real numbers, and *discrete*, where the quantity of interest can take only a discrete set of values—typically positive integers.

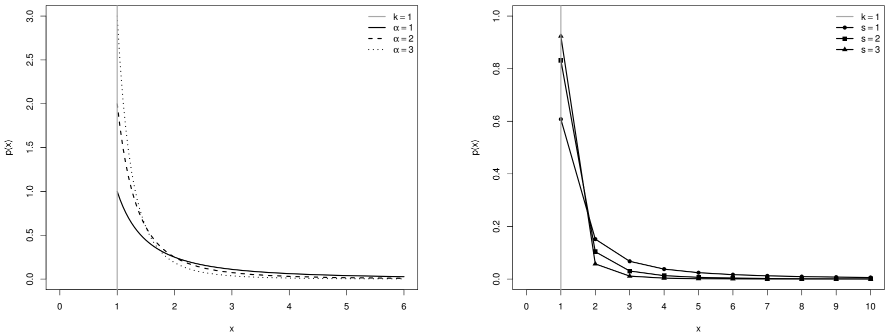
The continuous version, known as the *Pareto distribution*, is well studied in the literature and has the PDF

$$p(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \quad x \geq k, \tag{4.10}$$

[5] What is usually called a power-law distribution is simply the PDF associated with the CDF of the Pareto distribution, i.e.

$$p(x) \propto x^{-(\alpha+1)} = x^{-\zeta}.$$

Then, the exponent of the power-law distribution is $\zeta = \alpha + 1$, where α is the Pareto distribution shape parameter. To add to the confusion, some authors call $\alpha + 1$ the scaling exponent, that is the exponent of the density (4.9), while others find it easier to work with the exponent α of the CCDF $P_{>}(x) \propto x^{-\alpha}$. For the purposes of this chapter, we will not change here the nomenclature but will refer to the Pareto distribution shape parameter α as the power-law scaling exponent, with the hope that this caveat is sufficient to avoid confusion.



(a) Pareto **(b)** Zipf

Figure 4.2 Power-law probability density/mass functions with scale parameter $k = 1$ and different values of the shape parameters α and s. In panel (b), the connecting lines do not indicate continuity

where $k > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter (e.g. Kleiber and Kotz, 2003, and Arnold, 2015).[6] The discrete case has probability mass function (PMF)

$$p(x) = \Pr(X = x) = \frac{x^{-(s+1)}}{\zeta(s+1, k)}, \quad x \in \mathbb{Z}_{>0}, \quad k > 0, \quad s > 0, \tag{4.11}$$

where

$$\zeta(s+1, k) = \sum_{n=0}^{\infty} \frac{1}{(n+k)^{s+1}} \tag{4.12}$$

is the generalized zeta function.[7] For $k = 1$, $\zeta(s+1, k)$ reduces to the standard zeta function

$$\zeta(s+1) = \sum_{n=1}^{\infty} n^{-(s+1)} \tag{4.13}$$

and the distribution (4.11) becomes known as the *Zipf (or zeta) distribution*—in honour of the American linguist Zipf (1949).

Figure 4.2 visualizes the Pareto (a) and Zipf (b) probability density/mass functions with scale parameter $k = 1$ and different values of the shape parameters α and s. Clearly, both the distributions are highly skewed to the right with a heavy tail. It is therefore reasonable to assume that a random variable following the Pareto/Zipf distribution contains extreme values. The effect of changing the shape parameters α and s is visible in the plots at the scale parameter k:

[6] The statistical distribution (4.10), usually referred to as the *strong Pareto law* (Mandelbrot, 1960), has 100-year-plus history that dates back to the work of the Italian economist Pareto (1895). In his pioneering contributions at the end of the nineteenth century, Pareto (1896, 1897a,b) also suggested two variants of his distribution, occasionally called the three-parameter Pareto distributions. These further Pareto distributions, however, have not been used much in empirical economic (and other) studies.

[7] See http://mathworld.wolfram.com/HurwitzZetaFunction.html.

the higher (lower) α and s, the higher (lower) the probability density/mass at k, and thus the thinner (fatter) the right tail.[8]

In many cases it is useful to consider the cumulative distribution function (CDF) of power-law distributed variables. In the continuous (Pareto) case, the CDF has the relatively simple structure

$$P_{\leq}(x) = 1 - \left(\frac{x}{k}\right)^{-\alpha}, \tag{4.14}$$

whilst for the discrete (Zipf) version it is given by

$$P_{\leq}(x) = \frac{H_{x,s+1}}{\zeta(s+1)}, \tag{4.15}$$

where $H_{x,s+1}$ is a generalized harmonic number.[9] Figure 4.3 charts the Pareto/Zipf CDF for different values of α and s. As can be seen, the higher (lower) the value of α and s, the more (less) rapidly the CDF trends to 1.

The moments of power-law distributions are also of particular interest. For instance, in the Pareto case the r^{th} moment equals

$$\langle X^r \rangle = \frac{\alpha k^r}{\alpha - k} \tag{4.16}$$

and exists only if $r < \alpha$. Therefore, when

- $0 < \alpha \leq 1$, the mean and higher-order moments diverge, i.e. $\langle X \rangle = \infty$;
- $1 < \alpha \leq 2$, the second and higher-order moments diverge, i.e. $\left\langle X^2 \right\rangle = \infty$;
- $2 < \alpha \leq r$, the r^{th} and higher-order moments diverge, i.e. $\langle X^r \rangle = \infty$.

To understand the meaning of Paretian (power) laws with infinite population moments, one can use the *sequential moment estimation* method (Mandelbrot, 1963; Willinger et al., 2004). This method plots the *running moment estimates*, that is the value of a moment estimate of the data is plotted as a function of

[8] At times, a power-law distribution is also called a *scale-free* distribution, because it is the only distribution that is the same whatever scale one looks at it on (e.g. Sornette, 2012). To illustrate this fact, one can suppose to have some probability distribution $p(x)$ for a quantity x and to discover or somehow deduce that it satisfies the property that

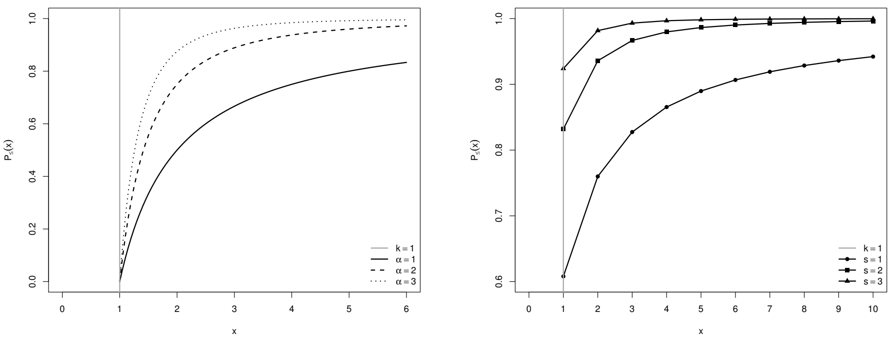
$$p(kx) = kp(x)$$

for any k. That is, if one increases the scale or unit by which x is measured by a factor of k, the shape of the distribution $p(x)$ is unchanged, except for an overall multiplicative constant. Thus, for instance, an immediate consequence of the presence of the power-law tail in the probability distribution of incomes is that the probability that a random person from the richer part of the society is k times richer than another person with income x is independent of x, i.e.

$$\frac{p(kx)}{p(x)} = k^{-(\alpha+1)}.$$

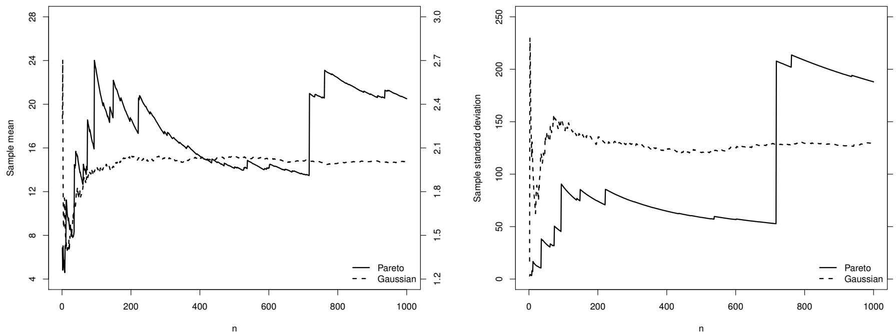
Therefore, the power-law distribution is scale-free, reflecting a certain "self-similarity" of the structure of the richest class. The scale appears in the problem through the parameter k: for example, if $k = 10$ and $\alpha = 2$, then the power law predicts that the number of people ten times richer is roughly one thousand (10^{-3}) times smaller. The suppression factor is very sensitive to α. If the value of α moves toward unity, the suppression factor decreases, and for $k = 10$ it is only 10^{-2}. In other words, in economies with a smaller value of α the tail of the distribution is fatter. This leaves more space for rich individuals.

[9] See http://mathworld.wolfram.com/HarmonicNumber.html.



(a) Pareto **(b)** Zipf

Figure 4.3 Power-law cumulative distribution functions with scale parameter $k = 1$ and different values of the shape parameters α and s. In panel (b), the connecting lines do not indicate continuity



(a) **(b)**

Figure 4.4 Running moment estimates corresponding to same-sized ($N = 1,000$) distinct samples of a Paretian random variable with $k = 2$ and $\alpha = 1$ (left y-axis scale) and a Gaussian random variable with $\mu = 2$ and $\sigma = 1$ (right y-axis scale): (a) mean; (b) standard deviation. The figure gives an idea of how erratic and sample-dependent the moments of Paretian variables can be expected to be

the number of observations used in the estimation of the moment. For example, Figure 4.4 shows the sequential mean and standard deviation plots corresponding to same-sized ($N = 1,000$) distinct samples of a Paretian random variable with $k = 2$ and $\alpha = 1$ and a Gaussian random variable with $\mu = 2$ and $\sigma = 1$, obtained by simply inverting series of random variables uniformly distributed in the interval $(0, 1)$.[10] As one can easily recognize, while in the Gaussian case the mean and standard deviation estimates exist, are finite, and converge robustly to their theoretical value as the number of observations increases, in the Pare-

[10] The "inversion method" relies on the principle that continuous CDFs range uniformly over the open interval $(0, 1)$. If u is a uniform random number on $(0, 1)$, then $x = P^{-1}(u)$ generates a random number x from any continuous distribution with the specified CDF P.

tian case the sample mean and standard deviation estimates do not converge as n increases, even if they always exist for any fixed n. This fact confirms that for a Paretian random variable with $\alpha = 1$ the first and second moments do not exist, i.e the computer-simulated data set is a sample from an underlying distribution having mean and variance infinite.

Finally, power laws are conserved with respect to addition, multiplication, power transformation, minimization and maximization (e.g. Gabaix, 2009, and Sornette, 2012). In particular, it can be proven that when two power-law distributed random variables are combined via the above algebraic transformations a new power law is generated from old ones whose exponent is preserved from the fatter-tailed power law, that is the one with the smallest exponent (Jessen and Mikosch, 2006). For instance, if X is a power-law distributed random variable with $\alpha_X < \infty$ and Y is another power-law variable with an exponent $\infty > \alpha_Y \geq \alpha_X$, then $X + Y$, $X \cdot Y$ or $\max(X, Y)$ are still power laws with the same exponent α_X. This property also holds when Y is normal, lognormal or exponential, whose right tail is thinner than any power law.

4.3.2 Recognizing power-law distributions

Given a sample of data, the most common approach followed to probe for the power-law form of the empirical distribution is from visual inspection of a plot where the logarithm of data is plotted against the logarithm of their complementary cumulative probabilities, defined as

$$\hat{P}^n_>(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{X_i > x\}, \tag{4.17}$$

where

$$\mathbf{1}\{X_i > x\} = \begin{cases} 1 & \text{if } X_i > x, \\ 0 & \text{otherwise.} \end{cases} \tag{4.18}$$

This is a natural estimator of the true CCDF (4.3) and it is essentially the CCDF of a distribution that puts mass $1/n$ on each data point.

If in doing a log-log plot of $\hat{P}^n_>(x)$ as a function of x one discovers a distribution that approximately falls on a straight line, then one can assert that the distribution follows a power law. For the specific case of a Paretian random variable, the plot will be exactly linear, as

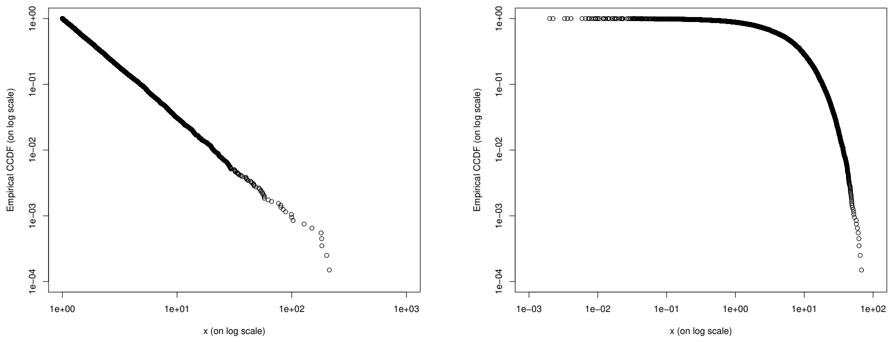
$$\ln[P_>(x)] = \alpha \ln(k) - \alpha \ln(x) = C_\alpha - \alpha \ln(x), \tag{4.19}$$

where

$$P_>(x) = \left(\frac{x}{k}\right)^{-\alpha} \tag{4.20}$$

is the CCDF for the Pareto distribution.

Figure 4.5 shows the log-log scaled empirical CCDFs of two large data sets ($N = 10,000$) drawn from a Pareto distribution with $k = 1$ and $\alpha = 1.5$ and an exponential distribution with parameter $\lambda = 0.125$. As one can easily recognize, for the Paretian random variable the log-log plot exhibits a linear decrease by several orders of magnitude, whereas for the exponential case it shows a faster-than-linear decrease of the distribution, especially at the right-end tail.



(a) **(b)**

Figure 4.5 Log-log scaled CCDFs of two large samples ($N = 10,000$) drawn from different continuous distributions: (a) Pareto distribution with $k = 1$ and $\alpha = 1.5$; (b) exponential distribution with parameter $\lambda = 0.125$

Another graphical method for the identification of power-law probability distributions using random samples is the *Pareto quantile plot* (e.g. Beirlant et al., 1996a, 2004). Since a log-transformed Pareto random variable is exponentially distributed,[11] the Pareto quantile plot compare the log-transformed data to the corresponding quantiles of a standard exponential distribution (i.e. an exponential distribution with parameter $\lambda = 1$) by plotting the former versus the latter. The quantiles of the standard exponential distribution are given by

$$-\ln\left(1 - \frac{i}{N+1}\right), \quad i = 1, \ldots, N, \tag{4.21}$$

thus yielding as coordinates for the Pareto quantile plot

$$\left\{-\ln\left(1 - \frac{i}{N+1}\right), \ln\left(x_i\right)\right\}, \quad i = 1, \ldots, N. \tag{4.22}$$

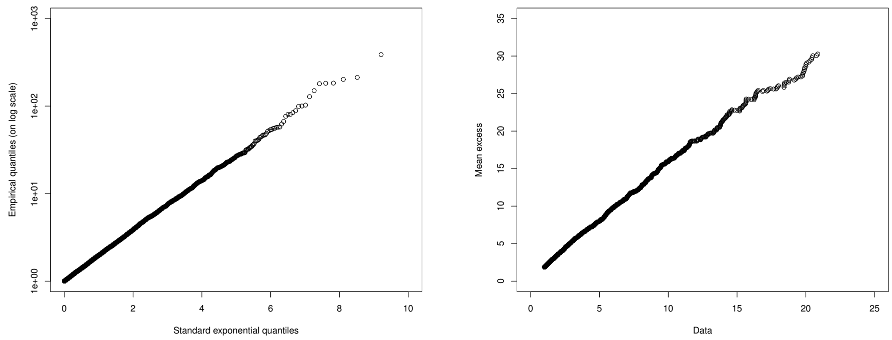
If the resulting plot suggests that the plotted points converge to a straight line, as in Figure 4.6(a), then a Pareto power-law distribution should be suspected.

The *mean excess plot* is an alternative way of graphically examining power-law probability distributions—a detailed description can be found in Beirlant et al. (1996b, 2004). This method consists of first sorting the data, such that $x_1 \leq \ldots \leq x_n$, and then computing for each observation x_i the empirical excess function

$$e_n\left(x_i\right) := \frac{1}{N-i} \sum_{j=i+1}^{N} \left(x_j - x_i\right), \quad i = 1, \ldots, N-1, \tag{4.23}$$

where N is the size of the random sample. The values of (4.23) are thus plotted against the corresponding x_i, and if the data follows a Pareto power-law distri-

[11] Formally, the Pareto distribution is related to the exponential distribution as follows: if X is Pareto-distributed with scale k and shape α, then $Y = \ln\left(\frac{X}{k}\right)$ is exponentially distributed with parameter $\lambda = 1/\alpha$.



(a) **(b)**

Figure 4.6 Pareto quantile plot (a) and mean excess plot (b) for a large sample ($N = 10,000$) drawn from a Pareto distribution with $k = 1$ and $\alpha = 1.5$

bution the plotted points will exhibit a positive linear trend, as shown in Figure 4.6(b).

4.3.3 Estimating power-law distributional parameters

Finding the threshold

When power laws are used in practice, it is usually the case that only the upper *tail* of the empirical distribution follows a power law. Therefore, before estimating the scaling parameter α, all observations below some minimum point $x_{\min}$ need to be discarded, so that one is left with only those observations for which the power-law model is a valid one. Estimating the scale parameter k of the Pareto/Zipf distribution and finding the threshold $x_{\min}$ directly corresponds with each other, and since the estimate of α will be *conditioned* on the choice of $x_{\min}$—choosing too low a value for $x_{\min}$ reduces the statistical error on the scaling parameter α because more data are used, but it increases its bias because the power law normally holds only in the tail—it becomes clear that some care must be taken when choosing this value.

The most common ways of estimating $x_{\min}$ are either to select visually the point beyond which the empirical CCDF of the distribution becomes roughly straight on a log-log plot, or using a *Hill plot* as in Figure 4.7(a), which involves estimating the α parameter for all candidate values of $x_{\min}$ and choosing the smallest value of $x_{\min}$ beyond which the estimated parameter stabilizes to a constant.[12] Clearly, these graphical methods are highly subjective and error prone, and the same can be held true when judging by eye the degree of linearity on a Pareto quantile plot or mean excess plot and using as an estimate of

[12]See e.g. Drees et al. (2000) for more details about the Hill plot. The Hill estimator and other methods for estimating the scaling parameter α once the threshold $x_{\min}$ has been fixed will be discussed in Section 4.3.3. Figure 4.7(a) was obtained using random numbers from a *composite* lognormal-Pareto distribution, which takes a lognormal density up to an unknown threshold value θ and a two-parameter Pareto density with scaling parameter α thereafter. The idea of such a composite model comes from Cooray and Ananda (2005), to whom we refer the reader for technical details.

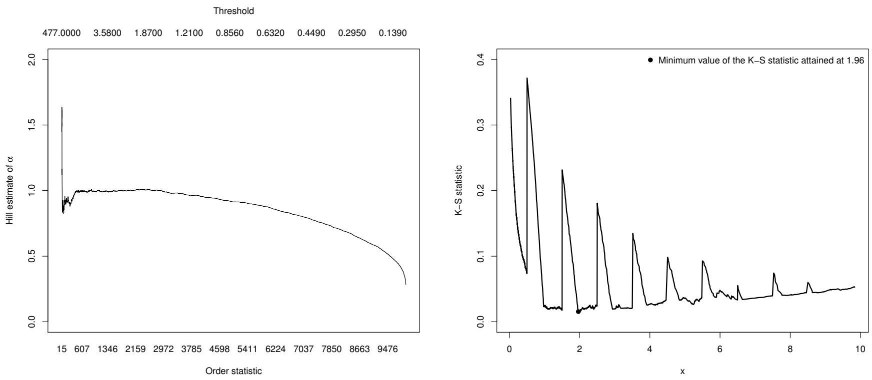


Figure 4.7 Methods for finding the threshold. (a) Hill plot for 10,000 random observations from the composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$. Notice that the left side of the graph clearly indicates the correct value of α in proximity of a threshold value equal to (or more than) θ. (b) K-S statistic versus ordered sample values for 5,000 random observations from the composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$. Notice that the K-S statistic yielding the best fit to the tail data is very close to the true value of θ

the threshold $x_{\min}$ the leftmost point beyond which these plots form almost a straight line.

In the case of a strong discontinuity at the threshold, using traditional graphical diagnostics would lead to a discernible kink so that it would be easy to choose (with little uncertainty) a lower bound $x_{\min}$ for power-law behaviour in the data. In the more realistic case of a smooth transition at the threshold, however, the traditional diagnostics would be harder to interpret, and hence their results should not be trusted. Therefore, more objective and accurate approaches have been sought. Among these, the method put forward by Clauset et al. (2007, 2009) determines the optimal choice of $x_{\min}$ by minimizing the distance between the probability distributions of the data and the best-fit power-law model. The measure used for quantifying the distance between the two probability distributions is the Kolmogorov-Smirnov (K-S) statistic

$$D = \max_{x \geq x_{\min}} \left| \hat{P}^n_{\leq}(x) - P_{\leq}(x) \right|, \tag{4.24}$$

which is simply the maximum distance between the data and fitted model CDFs (for $x \geq x_{\min}$). The optimal estimate of the lower bound is then the value of $x_{\min}$ where D attains the minimum.[13] An example of application of this technique is given in Figure 4.7(b), which shows the K-S statistic as a function of the assumed value of $x_{\min}$ for 5,000 random observations drawn from a composite

[13] As suggested by Clauset et al. (2009), the uncertainty in the estimate for $x_{\min}$ can be derived by making use of a non-parametric bootstrap method (Efron and Tibshirani, 1993). That is, given a data set with sample size N, one can generate a synthetic data set by drawing uniformly at random from the original data a new sequence of points x_i, $i = 1, \ldots, N$. Using the method described above, one then estimates $x_{\min}$ for this surrogate data set. By taking the standard deviation of all the estimates over a large number of repetitions of this process, the uncertainty in the original estimated parameter can thus be quantified.

lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$ (see footnote 12).[14] As the figure shows, the K-S statistic that yields the best fit to the tail data is very close to the true value of θ.[15]

Estimation of the scaling parameter

To estimate the scaling parameter α once the threshold $x_{\min}$ has been fixed is relatively straightforward.[16] Typically this parameter is extracted by observing that the Pareto power law implies the linear form (4.19).[17] Thus, if the doubly logarithmic plot of the empirical CCDF appear roughly straight above $x_{\min}$, then a straight line can be fitted to the data points beyond $x_{\min}$ using least squares (LS) linear regression, with the estimate of the scaling parameter α given by the absolute slope of the fitted straight line.[18] For example, the LS fit of a straight line to Figure 4.8(a) gives $\alpha = 0.985$, which is clearly compatible with the known value of $\alpha = 1$ from which the data were generated.

An alternative, simple and reliable method for extracting α is to employ the formula

$$\alpha_{\text{Hill}} = \frac{k}{\sum_{i=1}^{k} \ln\left(x_{N-k+i}\right) - k \ln\left(x_{N-k}\right)}, \quad i = 1, \ldots, N, \quad k \leq N. \tag{4.25}$$

Here the quantities x_i are the sample elements put in descending order and k is the number of observations larger than $x_{\min}$. Equation (4.25) was introduced by Hill (1975) and is referred to as the *Hill estimator*, which is well-known to be asymptotically normal (Hall, 1982) and consistent—meaning that $\alpha_{\text{Hill}} \to \alpha$ as $N \to \infty$ (Mason, 1982). It is constructed as a maximum likelihood estimator

---

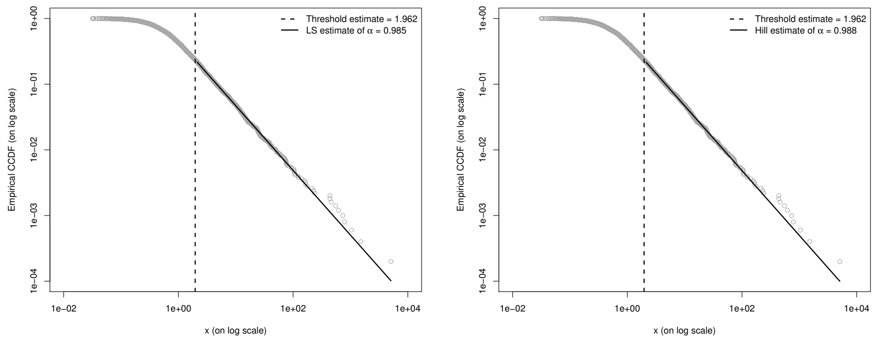
[14] For the ease of presentation, the limits of the horizontal axis in panel (b) of the figure have been adjusted so as to visually magnify the first part of the graph.

[15] Several other statistically principled methods for the estimation of the threshold $x_{\min}$ have been proposed in the literature. For instance, Beirlant et al. (1996a,b) developed a procedure that analytically determines the optimal choice of $x_{\min}$ by minimizing a non-parametric estimate of the asymptotic mean squared error (AMSE) for the maximum likelihood estimator of α proposed by Hill (1975; see also Section 4.3.3). Furthermore, Danielsson et al. (2001) proposed a bootstrap method to find the optimal $x_{\min}$ for the Hill estimator with respect to the AMSE, which has less analytical requirements than the approach proposed by Beirlant et al. (1996a,b). Finally, a robust criterion for choosing $x_{\min}$ and estimating the scaling parameter was developed by Dupuis and Victoria-Feser (2006). Nevertheless, these techniques are computationally time consuming and sometimes unstable, and are therefore not further discussed here.

[16] Parameter estimation for power-law type distributions is covered in depth in Johnson et al. (1994) and Arnold (2015). Here we shall only include the classical regression-type estimator and maximum likelihood estimation. For the method of moments, quantile and Bayes estimators, as well as methods based on order statistics, we refer the interested reader to the above-mentioned works. Some recent developments are also discussed by Alfons et al. (2013).

[17] Here we limit our attention to the methods for estimating the scaling parameter that are specific to *continuous* data. The discrete counterpart follows similar arguments, except for some additional technical conditions. A review of estimation methods that work for discrete data is contained in Clauset et al. (2009). We do not even consider the case of binned data sets —i.e. sequences of counts of observations over sets of non-overlapping ranges—which occur when direct measurements are unavailable (because impractical or impossible) and this is simply the form of the data received, or when one recovers measurements from existing histograms. Readers interested in pursuing the topic further are encouraged to consult the work by Virkar and Clauset (2014), who adapt the statistical framework of Clauset et al. (2009) to the less common but important case of binned empirical data

[18] Many software packages exist that can perform this kind of fitting, provide estimates and standard errors for the slope, as well calculate indicators for the quality of the fit.



(a) LS regression

(b) Hill estimator

Figure 4.8 Estimation of the scaling parameter. Points represent the complementary cumulative distribution for 5,000 random observations distributed according to a composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$. Solid lines represent the Pareto power-law fits to the data using the methods described in the text. Dashed lines denote the lower threshold estimated using a Kolmogorov-Smirnov approach (Clauset et al., 2007, 2009)

conditional on some known threshold level $x_{\min}$, and coincides exactly with the maximum likelihood estimator for the shape parameter of the Pareto distribution when $x_{\min}$ corresponds to the smallest value of x, i.e. when $k = N$. An error estimate for α_{Hill} can be derived by exploiting the asymptotic distribution theory of the Hill estimator as $\frac{\alpha_{\text{Hill}}}{\sqrt{k}}$ (e.g. Lux, 1996). Applying Equation (4.25) to 5,000 lognormal-Pareto distributed random numbers as in Figure 4.8(b) gives an estimate of $\alpha = 0.988$, which agrees well with the known value of 1.

4.3.4 Testing a set of data for power-law distribution

Since it is possible to fit a power-law distribution to *any* data set, it is appropriate to test whether the observed data could *plausibly* be considered to follow a power law. Many empirical studies of power-law distributed data have attempted to test the power-law hypothesis *qualitatively*, based for instance on graphical visualizations of the data such as a log-log plot of their complementary cumulative distribution. But this could lead to claims of power-law behaviour that do not hold up under closer scrutiny, because even data that are drawn from a non-power-law distribution can produce samples that resemble power-law distributions on a log-log plot. This is the case, for instance, with data drawn from a lognormal distribution, where a log-log plot may appear linear for several orders of magnitude if the variance of the corresponding normal distribution is large (Mitzenmacher, 2004). Therefore, to say with certainty whether a power law is a plausible hypothesis for the data, a more *quantitative* approach is desirable.

A standard way for testing empirical data against a hypothesized power-law distribution is to use a *goodness-of-fit* test based on the K-S statistic (4.24), which generates a p-value that quantifies the plausibility of the hypothesis.[19]

[19]Other measures for quantifying the “distance” between the distribution of the em-

Since the distribution of the K-S statistic is hard to obtain analytically, the p-value for the fit can be found by bootstrapping, i.e. by generating a large number of power-law distributed data sets with scaling parameter α and lower bound $x_{\min}$ equal to those of the distribution that best fits the observed data and then "re-inferring" the model parameters (Clauset et al., 2009).[20] When testing against the power-law distribution, the hypotheses are

$$\begin{aligned} &H_0 : \text{data are generated from a power-law distribution,} \\ &H_1 : \text{data are not generated from a power-law distribution.} \end{aligned} \tag{4.26}$$

Therefore, if the resulting p-value is larger than the chosen significance level, then the power law is a plausible hypothesis for the data; otherwise, if it is smaller, the model is rejected and another distribution may be more appropriate.

As a demonstration of this approach, we consider a data set with $N = 1,000$ observations drawn at random from a composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$. For fitting a power-law model to the distribution's upper tail, we use the distance-based method for automatically identifying the point $x_{\min}$ above which the power-law behaviour holds and the Hill estimator for α (see Section 4.3.3). We calculate that $x_{\min} = 1.01$ and $\alpha_{\text{Hill}} = 1.00$. Given this hypothesized model, we compute the distance D between the estimated model and the original data set using the K-S goodness-of-fit statistic (4.24), which yields $D = 0.02$. Then, using a semi-parametric bootstrap, we generate 100 data sets with N values that follows a Pareto power-law distribution with parameter α_{Hill} at and above the estimated threshold $x_{\min}$, but follows the original distribution below $x_{\min}$. After fitting the Pareto power-law model to each generated data set and calculated the associated K-S statistic, we are left with a p-value for the goodness-of-fit test equal to 0.78, meaning that the original data can be firmly considered to follow the Pareto power-law distribution in the upper tail at any of the usual significance levels (1%, 5%, and 10%).[21]

A large p-value for the power-law model, however, provides no information about whether some other distributions might be an equally plausible (or even a better) explanation. Hence, demonstrating that such alternatives are worse models of the data can strengthen the statistical argument in favour of the power law. As a way to accomplish this task, the test proposed by Vuong (1989) examines if the differences between a fitted model and the others are statistically significant. The approach is based on testing the null hypothesis H_0 that the competing models are equally close to the true data generating process against the alternative hypothesis H_1 that one model is closer. The test statistic is

$$\mathcal{R} = \sum_{i=1}^{N} \ln \left[\frac{p_1 (x_i; \theta_1)}{p_2 (x_i; \theta_2)} \right], \tag{4.27}$$

pirical data and the hypothesized model have been proposed and are in common use. Readers interested in pursuing the subject further may wish to consult the review by D'Agostino and Stephens (1986).

[20] For each generated data set the K-S statistic must to be computed relative to the best-fit power-law model for *that* data set, not relative to the original distribution from which the data set was drawn. In this way, it is ensured that the same calculation that has been performed for the real data set is performed also for each generated data set. This is a crucial requirement if an unbiased estimate of the p-value is sought (Capasso et al., 2009).

[21] The p-value is simply the fraction of the times the resulting K-S statistic is larger than the value for the original data.

$\theta_{j=1,2}$ being the maximum likelihood estimator of the unknown parameters for the model j, and is asymptotically distributed as a standard normal under the null hypothesis, that is

$$\text{under } H_0 : V = \frac{\mathcal{R}}{\sigma\sqrt{N}} \xrightarrow{d} N(0,1)\,, \tag{4.28}$$

where

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[\frac{p_1(x_i;\theta_1)}{p_2(x_i;\theta_2)}\right]\right\}^2 - \left\{\frac{1}{N}\sum_{i=1}^{N}\ln\left[\frac{p_1(x_i;\theta_1)}{p_2(x_i;\theta_2)}\right]\right\}^2}\,. \tag{4.29}$$

Hence, chosen a critical value z from the standard normal distribution corresponding to the desired level of significance, if $|V| \le z$ the null that models are the same cannot be rejected, whereas if $V > z$ model 1 can be considered better than model 2, and the reverse is true if $V < -z$.

An example of comparing competing distributions is shown in Figure 4.9, where the best-fit Pareto, lognormal and exponential models (all with $x_{\min} = 1.30$) have been found for the randomly generated data of the previous example.[22] As can be seen, the Pareto and lognormal distributions perform equally well, whereas the exponential model is a poor fit for the tail data. Investigating this formally by means of the method proposed by Vuong (1989) gives a p-value of 0.90 and a test statistic that is close to zero (-0.13), which means we cannot reject the null hypothesis that the Pareto and lognormal models are observationally equivalent, whereas in the case of the exponential distribution the Vuong test gives a p-value and a test statistic of, respectively, 0.00 and 4.14, meaning we can reject the null hypothesis and conclude that the Pareto model is closer to the true distribution than the exponential one.

4.3.5 Models of generation of power-law distributions

There is a large body of models capable of explaining the processes of generation of power-law distributions that span from physics (Bouchaud, 2001; Mitzenmacher, 2004; Newman, 2005; Sornette, 2006) to economics (Gabaix, 2009). Here we shall explore the role of random multiplicative processes as key mechanisms that explain distributional power laws. Readers interested in pursuing the topic further may consult the above references for a review of other mechanisms by which power-law behaviour can arise.

[22] Parameter estimation for the Pareto distribution uses the Hill estimator conditional on $x_{\min} = 1.01$ (see Section 4.3.3). For the lognormal and exponential distributions, an appropriate normalization constant C is used so that $\int_{x_{\min}}^{\infty} Cp(x)\,\mathrm{d}\,x = 1$. The normalization constant does not count as a parameter, because it is fixed once the values of the other parameters are estimated, and $x_{\min}$ does not count as a parameter because we know its value from the distance-based approach proposed by Clauset et al. (2007, 2009). Normalization constants for both the lognormal and the exponential distributions can be found in Clauset et al. (2009). Once we are given with the value of $x_{\min}$, the method of "matching moments"—which sets the distribution moments equal to the data moments and solves to obtain estimates for the distribution parameters—is used to calculate our estimates. For the lognormal, the location parameter μ is equal to the mean of the logarithm of the data points, and the shape parameter σ is equal to the standard deviation of the data set after transformation (see Section 4.2). For the exponential, the estimate of λ is just the reciprocal of the sample mean.

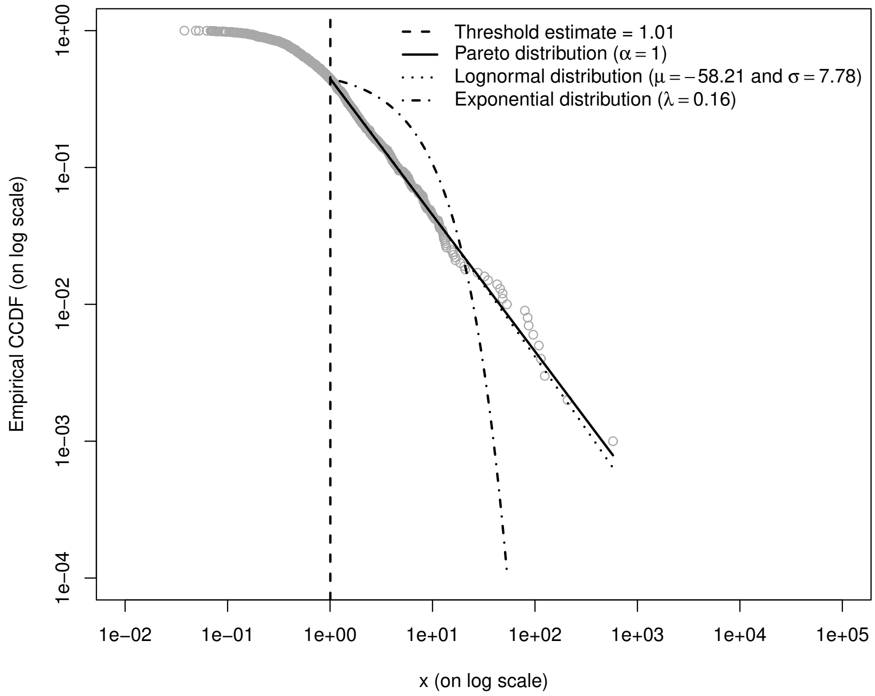


Figure 4.9 Log-log scaled CCDF for 1,000 random observations distributed according to a composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$. Red, blue and green solid lines represent, respectively, the Pareto, lognormal and exponential best fits. Dashed line denotes the lower threshold estimated using a Kolmogorov-Smirnov approach (Clauset et al., 2007, 2009)

Multiplicative processes

Power-law distributions can be obtained as steady-state solutions of stochastic processes. The stochastic theory, one of the oldest (and still popular) theories of distribution, relies for the skewed shape of economic distributions mainly or solely on chance, luck and random occurrences. The main authorship of this theory is attributed to Gibrat (1931), who viewed the evolution of firms' size (sales, employees, valued added, assets) as a multiplicative random process in which the product of a large number of individual random variables tends to the *lognormal* distribution.

The connection between multiplicative processes and the lognormal distribution can be described as follows. Suppose that

$$x_t^i = \left(1 + r_{t-1}^i\right) x_{t-1}^i, \tag{4.30}$$

where x_t^i is the size of firm i at time t and $\left\{r_t^i\right\}_{t=0,1,\ldots,T-1}$ are the per-period rates of growth in firm's size. Denoting the firm's per-period growth factors by

$\lambda_t^i = \left(1 + r_t^i\right)$, the above expression can be rewritten as

$$x_t^i = \lambda_{t-1}^i x_{t-1}^i, \tag{4.31}$$

where $\left\{\lambda_t^i\right\}_{t=0,1,\ldots,T-1}$ are independent, identically distributed, continuous and non-negative random variables. Taken literally with no other ingredient, expression (4.31) leads to an ensemble of values x_t^i over all possible realizations of the multiplicative factors $\left\{\lambda_t^i\right\}_{t=0,1,\ldots,T-1}$ which is distributed according to the lognormal distribution. Indeed, by iterating (4.31), the firm's size at time T is

$$x_T^i = x_0^i \lambda_0^i \lambda_1^i \lambda_2^i \cdots \lambda_{T-1}^i = x_0^i \prod_{j=0}^{T-1} \lambda_j^i, \tag{4.32}$$

where x_0^i represents the starting size of firm i. If the λ_t^i's are all governed by independent lognormal distributions $\Lambda_i(\lambda)$, then x_T^i is approximately lognormal as the product of lognormal distributions is again lognormal. However, lognormal distributions may be obtained even if the λ_t^i's are not themselves lognormal by the additive form of the CLT. Indeed, from the logarithm of (4.32)

$$\ln\left(x_T^i\right) = \ln\left(x_0^i\right) + \sum_{j=0}^{T-1} \ln\left(\lambda_j^i\right) \tag{4.33}$$

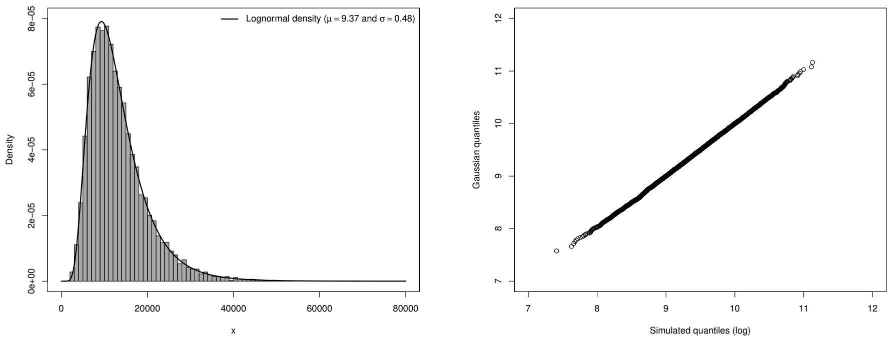
one gets via the CLT that $\sum_{j=0}^{T-1} \ln\left(\lambda_j^i\right)$ converges to a Gaussian distribution for sufficiently large t if the $\ln\left(\lambda_j^i\right)$'s are independent and identically distributed variables with finite mean and variance; hence, for large times t, x_T^i is well approximated by a lognormal distribution.

In the industrial organization and economic geography literature this result is also known as the Gibrat "law of proportionate effect", stating that if growth rates of firms' size in a fixed population (i.e. abstracting from entry and exit dynamics) are independent of size and uncorrelated, the resulting distribution is lognormal. This is clearly seen in Figure 4.10, which exhibits the simulated distribution for the multiplicative system (4.31) at time $T = 100$, where λ_t^i is a random draw from a Gaussian distribution with $\mu = 1.02$ and $\sigma = 0.05$.[23] The simulation has been run for $N = 10,000$ firms, whose size has been initialized to 100. As one can easily recognize, both the panels of Figure 4.10 show every sign of being lognormal.[24]

Lognormal and power-law distributions are intrinsically connected. Indeed, small variations in the underlying model can change the result from one to the

[23] The value $\mu = 1.02$ ensures that the experimentally observed firms' size increases in the overall system at each time step of the simulation, whereas the standard deviation $\sigma = 0.05$ is large enough to enable occasionally values which are smaller than 1.

[24] Since a random variable X has the lognormal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$ if $\ln(X)$ has the Gaussian distribution with mean μ and standard deviation σ, the lognormal density in panel (a) of Figure 4.10 uses parameter estimates obtained by first taking logarithms of the simulated data and then equating μ and σ to the mean and standard deviation of the log-transformed simulated data (see Section 4.2). The lognormal quantile plot in panel (b), instead, has been drawn by first taking logarithms of the simulated data and then using the Gaussian quantile plot with the same parameter estimates.



(a) Histogram **(b)** Lognormal quantile plot

Figure 4.10 The final time step distribution obtained from the simulation of the multiplicative stochastic process (4.31). The solid line in panel (a) denotes the lognormal density with parameter values obtained by "matching moments" (see footnote 24)

other.[25] As a consequence, many variations of the model (4.31) have been developed in the literature. The main result of this strand of research was to show that even small variations from the pure multiplicative stochastic process lead to a *power-law* distribution. As a matter of example, Champernowne (1953) offered an explanation for the Pareto power-law distribution of income similar in character to the "law of proportionate effect". His model, later on generalized and extended by Simon (1955), views income determination as a Markov process—income for the current period depends only on one's income for the last period and random influence—and relies upon the subdivision of incomes into discrete ranges as well as the specification of a constant matrix of transition probabilities—otherwise, no stationary distribution will emerge from the Markov process. Champernowne showed that if the income intervals defining each class are assumed to form a geometric progression, then the equilibrium distribution of income tends to that given by a discrete Pareto distribution.

The main difference between the multiplicative model (4.31) and the Champernowne model is that while in the former income can become arbitrarily close to zero through successive decreases, in the latter model there is a minimum income corresponding to the lowest class below which one cannot fall. Generally, for each economic system one can assume the existence of a positive cut-off $x^*_{t-1} > 0$ for the minimal income of each individual, i.e. there is some threshold income which one has to posses to fulfil minimal needs and function in the system. In welfare economies it is provided by the social security system through the economic effects of subsidies, securities, and services.

A very general extension to Champernowne model is contained in Levy and Solomon (1996a,b), who showed that a power-law distribution can be obtained by adding a reflection condition to the stochastic multiplicative model

[25] A rich and long history about generative models leading to either power-law or lognormal distributions, spanning many fields, can be found in work from decades ago. See, for instance, Aitchison and Brown (1957) and Sahota (1978); see also Kleiber and Kotz (2003), Mitzenmacher (2004) and Newman (2005) for pointers to some of the recent and historically relevant scientific literature on the subject.

(4.31), i.e. by assuming that each x^i_{t-1} is bounded from below to a threshold

$$x^*_{t-1} = \frac{c}{N}\sum_{i=1}^{N} x^i_{t-1} = c\left\langle X_{t-1}\right\rangle \tag{4.34}$$

proportional to current average income $\langle X_{t-1}\rangle$. Operationally, each time the generic multiplicative process (4.31) returns a value x^i_t smaller than x^*_{t-1}, the actual value of x^i_t is restored to

$$x^i_t = x^*_{t-1} = c\left\langle X_{t-1}\right\rangle, \tag{4.35}$$

where the fraction c is fixed in time.[26] This mechanism can be also viewed as a way of killing off an individual and introducing a new one at the same time, thus incorporating a perfectly balancing "birth and death" process. Then, the multiplicative model (4.31) formally changes to

$$x^i_t = \begin{cases} \lambda^i_{t-1}x^i_{t-1} & \text{if } x^i_t \geq c\left\langle X_{t-1}\right\rangle, \\ c\left\langle X_{t-1}\right\rangle & \text{if } x^i_t < c\left\langle X_{t-1}\right\rangle. \end{cases} \tag{4.36}$$

Numerical results probing the properties of the multiplicative stochastic model with reflecting lower bound are shown in Figure 4.11 for the final time step $T = 100$ and values $c = 0.9$, $N = 10,000$ and $x^i_0 = 100$. As in the previous simulation, the random factor λ^i_t is extracted from a Gaussian probability distribution with $\mu = 1.02$ and $\sigma = 0.05$. One can see that the lower cut-off in fact works, and the resulting distribution is no longer lognormal but can be approximated by a Pareto power-law function.

Kesten (1973) considered the following mixture of multiplicative and additive processes

$$x^i_t = \lambda^i_{t-1}x^i_{t-1} + b^i_{t-1}, \tag{4.37}$$

with $\left\{\lambda^i_t\right\}_{t=0,1,\ldots,T-1}$ and $\left\{b^i_t\right\}_{t=0,1,\ldots,T-1}$ being positive independent random variables. Clearly, for $b^i_{t-1} = 0$ the simple linear stochastic equation (4.37) recovers the model (4.31); for $b^i_{t-1} \neq 0$, it generates an ensemble of values x^i_t for which the power-law tail behaviour

$$p\left(x_t\right) \propto x_t^{-(1+\alpha)} \tag{4.38}$$

can be observed. The term b^i_{t-1} can be thought of as an effective repulsion from the origin—i.e. a re-injection of the dynamics—and thus acts similarly to the barrier x^*_{t-1} in the previous model (4.36), making the multiplicative process with the reflective barrier and the Kesten model deeply related.[27] The reconstructed final time step ($T = 100$) density of the Kesten process (4.37) for λ^i_{t-1} and b^i_{t-1} uniformly sampled in the intervals $[0.48, 1.48]$ and $[0, 1]$, respectively, and values $N = 10,000$ and $x^i_0 = 100$ is shown in Figure 4.12, where a clear power-law behaviour in the tail can be observed from the double logarithmic plot of the cumulative distribution.

[26] Clearly, if c is time independent, x^* varies in time.

[27] See Sornette (1998) on this subject; see also Sornette and Cont (1997) and Takayasu et al. (1997).

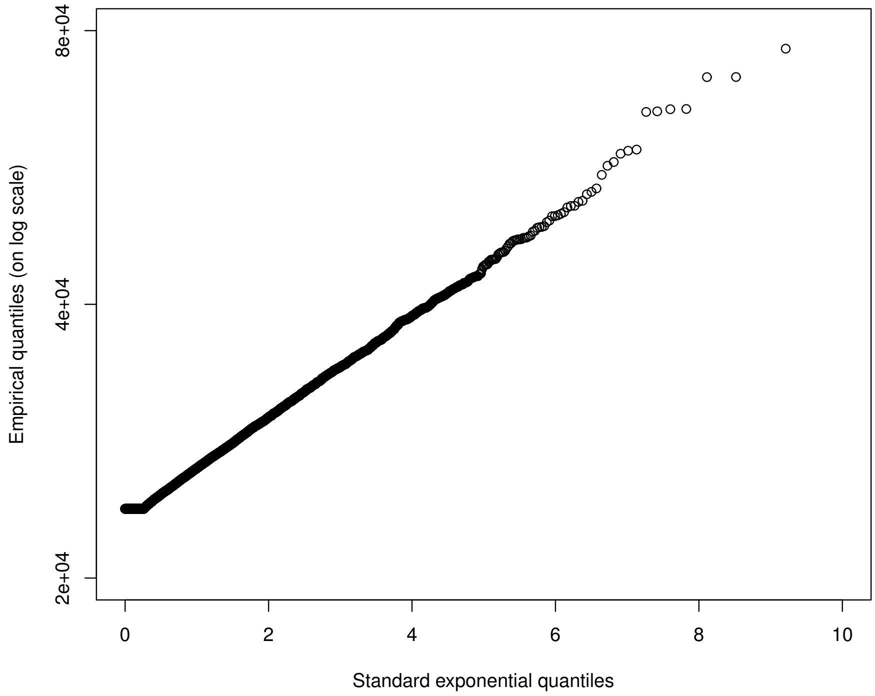


Figure 4.11 Pareto quantile plot for the final time step distribution obtained from the simulation of the multiplicative stochastic process (4.36)

Finally, Blank and Solomon (2000) incorporate both entry and exit dynamics by assuming that agents disappear from the system if they fall below the threshold (4.34) and that at each period t

$$\Delta N = N_{t+1} - N_t = K \left(\sum_{i=1}^{N_{t+1}} x_{t+1}^i - \sum_{i=1}^{N_t} x_t^i \right) = K \left(x_{t+1}^{\text{tot}} - x_t^{\text{tot}} \right) \tag{4.39}$$

new individuals enter the system with the size x_{t-1}^*. This model—with a variable number of components whose size evolves according to the multiplicative stochastic rule—also leads to a power-law distribution, as one can easily recognize from Figure 4.13, which shows the Pareto quantile plot comparing the distribution of the simulated data to the power-law distribution for $T = 100$.

4.4 The Laplace distribution

In recent years part of the research in macroeconomics has analysed the statistical properties of aggregate output growth-rate distribution in a cross section of countries, i.e. ignoring time dimension. In particular, works belonging to this line of research have shown that GDP log growth rates have a *tent-shaped* distribution, which is characterized by a high singular mode and heavy

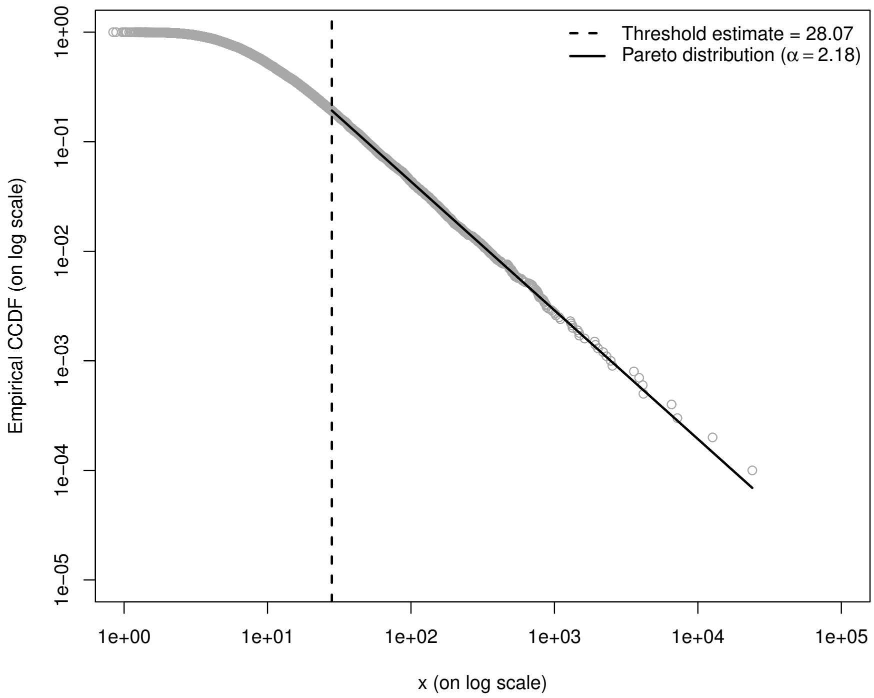


Figure 4.12 The final time step distribution obtained from the simulation of the Kesten process (4.37). The solid line represents the Pareto power-law fit to the simulated data using the Hill estimator described in Section 4.3.3. The dashed line denotes the lower threshold estimated using a Kolmogorov-Smirnov approach (Clauset et al., 2007, 2009)

tails similar to the Laplace distribution (e.g. Canning et al., 1998, Lee et al., 1998, and Castaldi and Dosi, 2004).[28] This kind of evidence was also found at lower levels of aggregation, namely in the case of the cross-sectional distribution of firms' growth rates, where the observed tent-shaped Laplace distribution appears robust to various measures of growth indicators—including value added, sales and employment—as well as over different levels of industry aggregation (see, among others, Stanley et al., 1996, Amaral et al., 1997, Lee et al., 1998, Bottazzi and Secchi, 2003a,b, Castaldi and Dosi, 2004, Fu et al., 2005, Sapio and Thoma, 2006, and Dosi and Nelson, 2010).[29] Since the Laplace has

[28]More recently, the research on the subject has expanded to investigate the distributional properties of aggregate output growth-rate *time series*, i.e. for individual countries over time (see Fagiolo et al., 2007a,b, 2008). The main finding from these studies is that in many industrialized countries the growth-rate time-series distribution can be well approximated by a Laplace density with tails much fatter than those of a Gaussian distribution. This implies that the pattern of economic growth over time tends to be quite lumpy: large growth events, either positive or negative, seem to be more frequent than what a Gaussian model would predict.

[29]There are, however, some documented variations in the observed tent-shaped distribution of firms' growth rates. In a study on Danish firms, for instance, Reichstein and Jensen (2005)

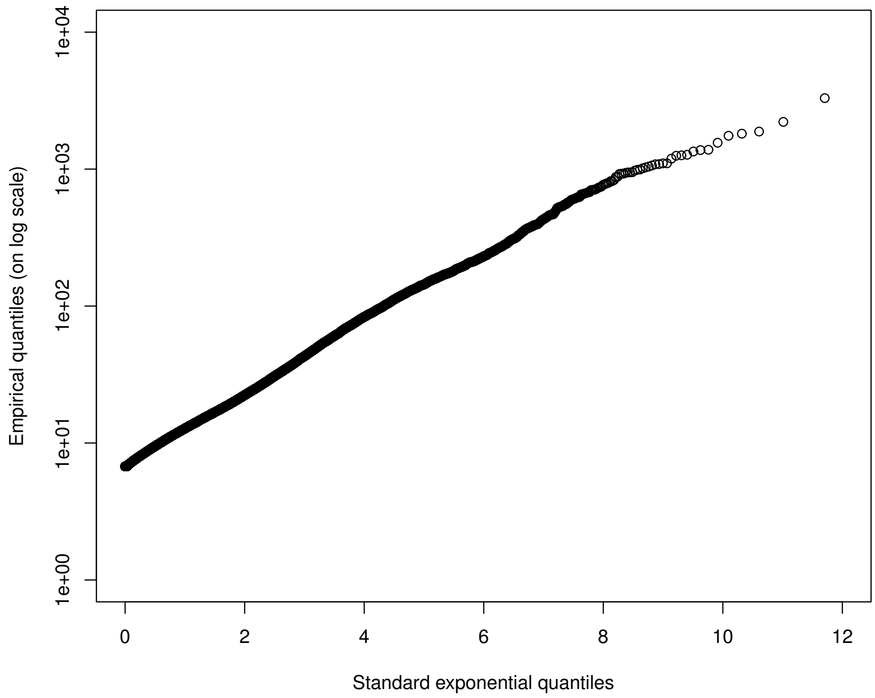


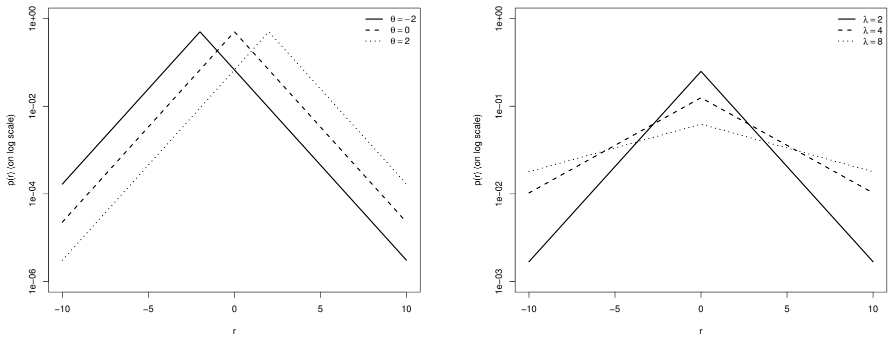
Figure 4.13 Pareto quantile plot for the final time step distribution of the simulated Blank and Solomon (2000) model with $c = 0.4$, $K = 0.1$, $x_0^i = 100$ and $N_0 = 100$. The λ_t^i's have been drawn from a Gaussian distribution with $\mu = 1.02$ and $\sigma = 0.05$. The final system size is $N = 121,185$

higher spike and thicker tails compared to the Gaussian distribution, these results indicate that—no matter the level of aggregation—the growth dynamics of complex organizations such as countries and firms is characterized by gains or losses from growth episodes that play a role statistically more significant than expected from a Gaussian distribution.[30]

More precisely, if the annual growth rate is $r_t^i = \ln\left(\frac{x_t^i}{x_{t-1}^i}\right)$, where x_{t-1}^i is

provide evidence of substantial skewness along with signs of heavier tails than are accounted for by the Laplace distribution—especially for the right tail containing the fastest growing firms. Tails fatter than those of a Laplace were also found in studies such as Bottazzi et al. (2011), who remark that «[...] the Laplace distribution of growth rates cannot be considered as a universal property valid for all sectors. Looking at French manufacturing, we observe growth rates distributions with tails that are consistently fatter than those of the Laplace» (Bottazzi et al., 2011, p. 2). In their investigation, Bottazzi et al. (2011) use a more general group of probability densities, known as the Subbotin (1923) family of distributions, which allows for both skewness and "super-Laplace" tails in growth-rate distributions and encompasses both the Laplace and the Gaussian functions as special cases.

[30] Delli Gatti et al. (2005) present an agent-based model that replicates the Laplace distribution of both the firms' and aggregate output growth rates. They show that the power-law of firms' size implies that growth is Laplace distributed and also that small micro-shocks can aggregate into macro-shocks to generate recessions.



(a) $\lambda = 1$ **(b)** $\theta = 0$

Figure 4.14 Laplace density function for different values of location (a) and scale (b). Note the appearance of a "tent shape" when the y-axis is expressed in logs

the GDP of a country or the size of a firm in year $t-1$, then for all years the probability density of r is consistent with a Laplace distribution given by the function

$$p(r) = \frac{1}{2\lambda} e^{-\frac{|r-\theta|}{\lambda}}, \quad -\infty < r < \infty, \tag{4.40}$$

where $\theta \in (-\infty, \infty)$ and $\lambda > 0$ are location and scale parameters, respectively. From Figure 4.14(a), it is apparent that changing the location simply shifts the probability density curve to the right or to the left. From Figure 4.14(b), we see that the scale parameter determines the width of the distribution.[31]

The distribution is symmetric about θ, i.e. for any real r we have

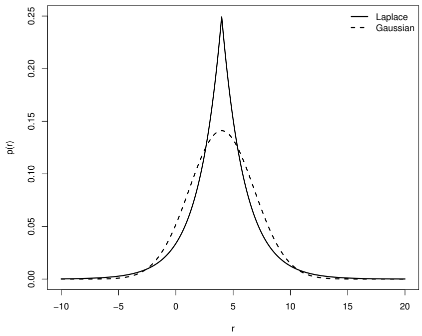
$$p(\theta - r) = p(\theta + r). \tag{4.41}$$

Consequently, as in the case of other symmetrical distributions, Laplace's location is the same as its mean, median, and mode. Figure 4.15 compares the Laplace against the Gaussian distribution. Recall that the Gaussian distribution has an expected value of μ and a variance equal to σ^2. Suppose we fix the mean of the Gaussian to equal the mean of the Laplace distribution, and then also match the variances of the two. In Figure 4.15, both the distributions have an expected value of 4 and variance equal to 8.[32] As can be seen, the Laplace has higher spike and slightly thicker tails compared to the Gaussian. The latter property is particularly visible in panel (b) of the figure, which provides a magnification of the right tail on $(8, 20)$.

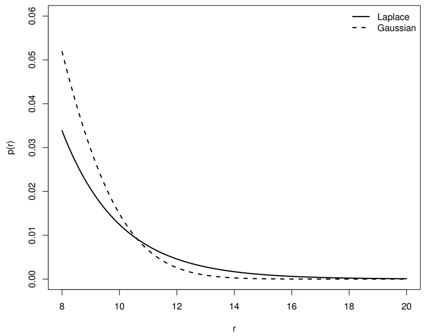
Parameter estimation for the Laplace model presents few difficulties. Given a sample $R = (r_1, \ldots, r_N)$ coming from the Laplace distribution (4.40), the maximum likelihood estimator of θ is the sample median, whereas the maximum

[31] The Laplace distribution is the distribution of differences between two independent variates with identical exponential distributions (e.g. Kotz et al., 2001). It is also sometimes called the *double exponential distribution*, because its shape reminds one of two exponential distributions (with an additional location parameter) spliced together back-to-back.

[32] The variance of the Laplace distribution is $2\lambda^2$, hence the scale parameter has been taken equal to 2 for the Laplace density of Figure 4.15.



(a)



(b)

Figure 4.15 Laplace and Gaussian densities

likelihood estimator of λ is

$$\lambda = \frac{1}{N}\sum_{i=1}^{N} |r_i - \theta| \, . \tag{4.42}$$

Figure 4.16 provides an example using 10,000 random observations generated from a Laplace-distributed population with parameters $\theta = 1$ and $\lambda = 2$. Clearly, the estimated values of the Laplace parameters are very close to the "true" values from which the data were generated. For comparison, a Gaussian distribution with the same mean and standard deviation as the simulated data is superimposed on the plot. However, while the Gaussian distribution is expressed in terms of the squared difference from the mean, the Laplace density is expressed in terms of the absolute difference from the median. Consequently, the Laplace distribution has a steeper peak and tails that asymptotically approach zero more slowly than the Gaussian.

To test whether a given data set is plausibly drawn from a Laplace distribution, one can use the K-S statistic (4.24). The hypothesis regarding the Laplace distributional form is rejected if the statistic D is greater than the critical value obtained from a table. For example, the calculated value of the K-S statistic that quantifies the distance between the estimated Laplace model and the simulated data of Figure 4.16 is 0.72, whereas the critical value at the 5% significance level is approximately 0.91.[33] Hence, as expected, the null hypothesis is not rejected for the Laplace-distributed data, while it is rejected for the Gaussian distribution—the corresponding value of D is 6.71, which is far beyond the 5% critical value of around 0.9.[34]

Finally, distribution selection criteria can be used to distinguish between the Laplace and a number of alternative models. Among these criteria, the negative log-likelihood value $l = -\ln(L)$, minimized to determine the values for the free parameters, as well as the values of Akaike (1973) and Schwarz (1978) information criteria (AIC and BIC) typically provide a way to judge whether

[33]For the tables of critical values, see Puig and Stephens (2000).

[34]For a practical guide to testing for normality by means of the K-S statistic, see Stephens (1974).

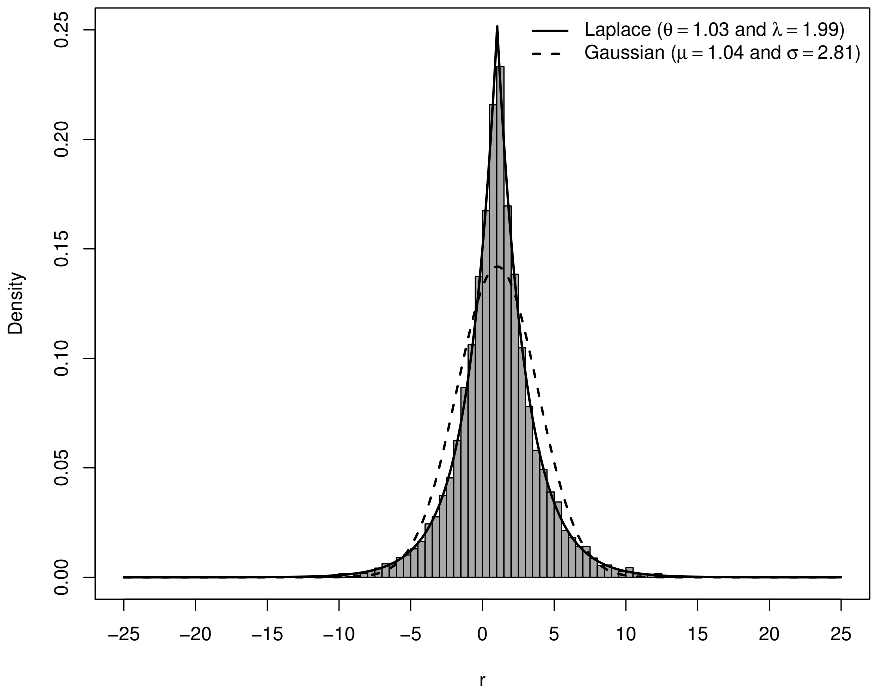


Figure 4.16 Histogram with superimposed Laplace and Gaussian densities for 10,000 random observations from the Laplace distribution with $\theta = 1$ and $\lambda = 2$

a distribution is a better explanation of the data than some other reasonable alternatives. Model selection criteria such as the AIC and BIC will select, when comparing models with the same number of parameters, the model with the largest l according to the formula

$$(2 \times l) + (d \times k)\,, \tag{4.43}$$

where k represents the number of parameters in the fitted model and $d = 2$ for the usual AIC or $d = \ln(N)$, N being the number of observations, for the so-called BIC. When comparing models fitted by maximum likelihood to the same data, the smaller the AIC or BIC the better the fit. When comparing models using the log-likelihood criterion, the larger the l the better the fit. As a matter of example, Table 4.3 reports the values of l, AIC and BIC obtained for the same simulated data as in Figure 4.16. As expected, it emerges that all the selection criteria agree on the Laplace distribution as the one to be preferred over the Gaussian.

We want to conclude this part by trying to answer a question the reader may have at this point. The literature seems often to suggest the Pareto distribution as a good approximation of firms' size and the Laplace distribution for growth rates. Is there a link between the two or are they separated phenomena? The second explanation may certainly be a possibility that we have followed in the

Table 4.3 Selection criteria for Laplace and Gaussian distributions using 10,000 random observations generated from a Laplace-distributed population with parameters $\theta = 1$ and $\lambda = 2$

	Laplace	Gaussian
l	-23,797	-24,526
AIC	47,597	49,056
BIC	47,612	49,070

presentation of this chapter, as there are separate stochastic processes generating the two distributions.

Regarding the first possibility, the existence of a link, a possible starting point is a couple of well-known statistical theorems that we have briefly mentioned before: (i) the logarithm of a Pareto random variable follows an exponential distribution (see footnote 11); (ii) the difference of two independent exponential random variables generates a Laplace distribution (see footnote 31). In other words, if we generate two independent Pareto distribution, say X_1 in period 1 and X_2 in period 2, then taking the logs $x_1 = \ln(X_1)$ and $x_2 = \ln(X_2)$ the difference

$$l = x_2 - x_1 \tag{4.44}$$

follows a Laplace (or double exponential) distribution. The problem with this theorem is the independence assumption. Can this theorem be generalized to the case involving *dependent* random variables? The analysis in Palestrini (2007) shows that the way of generating a double exponential random variable[35] by subtracting two exponential random variables can be generalized to dependent random variables having the $(t, t+1)$ joint distribution

$$\Pr(x_t > x_1, x_{t+1} > x_2) = e^{-\alpha_1 x_1 - \alpha_2 x_2 - \lambda \max(x_1, x_2)}, \tag{4.45}$$

where λ is a measure of dependence. In other terms, $\lambda = 0$ means independence, whereas with $\lambda > 0$ there is positive dependence as it is the case with firms' *log* size.

The distribution (4.45) is known as the *Marshall-Olkin bivariate exponential distribution* after the analysis in Marshall and Olkin (1967), who proved that (i) this joint distribution has exponential marginals and, more important, (ii) it is the distribution satisfying the bivariate version of the "memoryless property" of exponential random variables, i.e.

$$\begin{aligned} \Pr(x_t > \bar{x}_1 + k, x_{t+1} > \hat{x}_1 + k | x_t > k, x_{t+1} > k) = \\ = \Pr(x_t > \bar{x}_1, x_{t+1} > \hat{x}_1). \end{aligned} \tag{4.46}$$

Summarizing the theorem in Palestrini (2007), the Pareto distribution together with the memoryless property are compatible with double exponential growth rates. As noted by the author, this theorem is very sensitive to the hypotheses. The scaling free property is not always a feature of the entire support of the firms' size distribution and the memoryless property may be a bad

[35] This double exponentiality means that the tails have exponential shape but the distribution has a positive probability mass at zero because the generating joint probability distribution explained below has a singularity when $x_2 = x_1$. The probability mass at zero can be computed as in Bottazzi (2008).

approximation of actual data. For this reason, we have chosen in this chapter to explain these two important stylized facts as separated.

Exercises

Exercise 4.1 Suppose to have a set of 5,000 observations, which have been randomly generated from a composite lognormal-Pareto distribution with $\theta = 2$ and $\alpha = 1$ (see footnote 12 and various examples throughout the text; also look at the source codes of the relevant examples available on the book's website). Obtain parameters via a maximum-likelihood estimation procedure for the distributions listed in Table 4.1. Plot the complementary CDF of the data and the fitted models on log-log axes. [Hint: in `R`, functions for fitting by maximum likelihood the non-power-law distributions of Table 4.1 are implemented in the package `fitdistrplus` (Delignette-Muller and Dutang, 2015), the documentation of which the reader is referred to for more details. For the power-law model, use the methods of Section 4.3.3.]

Exercise 4.2 In `R`, load the example data set `Labour` from package `Ecdat`. The data set consists of 569 Belgian firms and includes information for 1996 on the total number of employees, their average wage, the amount of capital and a measure of output (Verbeek, 2012, ch. 4). Consider both the total number of employees and the amount of capital as proxies for size of firms and perform some exploratory graphical analysis of the data using the methods of Section 4.3.2. Does the data appear visually to follow a power law?

Exercise 4.3 Consider the same data of the previous exercise and assume the amount of capital—measured by total fixed assets (in million euros) at end of 1995—as a proxy of firm size. Estimate the lower bound $x_{\min}$ on power-law behaviour using the distance-based approach described in Section 4.3.3 and the scaling parameter α using the Hill estimator (4.25). Next, calculate the goodness of fit between the original data and the estimated power law using the method described in Section 4.3.4. Is the fit a good match to the tail data? Finally, fit the lognormal and exponential distributions to the data above the value of $x_{\min}$ for the power-law model and compare the latter with such alternative hypotheses using the method proposed by Vuong (1989). Can these alternatives be ruled out as a fit to your data or, if neither is ruled out, which one is the better fit? [Hint: see footnote 22 and look at the source code for Figure 4.9 available on the book's website.]

Exercise 4.4 Repeat the same exercise as above, but assume the total number of employees as a proxy of firm size. Summarize/discuss the results you get. [Hint: the number of workers in a factory is a *discrete* numerical variable, hence the appropriate discrete counterparts of the methods presented in the relevant parts of the chapter should be used. These are readily implemented in the `R` software package `poweRlaw` (Gillespie, 2015), the documentation of which the reader is referred to for more details. For further guidance, see also Clauset et al. (2009).]

Exercise 4.5 Using the `R` script available from the book's website, simulate the multiplicative stochastic process (4.31) for $N = 10,000$, $T = 100$, $\mu = 1.02$ and $\sigma = 0.05$. Provide a graphical comparison of the density curves estimated for time steps $t = 25$, $t = 50$, $t = 75$ and $T = 100$. Do they resemble a lognormal distribution? Which difference is more apparent when comparing the shape of the plotted densities? [Hint: find the parameter values of the lognormals by "matching moments" (see Section 4.2 and footnote 24).]

Exercise 4.6 In `R`, load the example data set `siemens` from package `evir`. These data are the daily log returns on Siemens share price from Tuesday 2$^{\text{nd}}$ January 1973 until Tuesday 23$^{\text{rd}}$ July 1996. Find which distribution the data fits more between a Laplace and a Gaussian. Provide the corresponding histogram with fitted density curves and compare the candidate distributions using selection criteria. [Hint: to improve performance of fitting functions, consider centring and scaling your data.]

References

J. Aitchison and J. A. C. Brown. On criteria for descriptions of income distribution. *Metroeconomica*, 6:88–107, 1954.

J. Aitchison and J. A. C. Brown. *The Lognormal Distribution with Special Reference to its Use in Economics*. Cambridge University Press, New York, 1957.

H. Akaike. Information theory and an extension of the likelihood ratio principle. In B. N. Petrov and F. Csaki, editors, *Proceedings of the Second International Symposium of Information Theory*, pages 257–281. Akademiai Kiado, Budapest, 1973.

A. Alfons, M. Templ, and P. Filzmoser. Robust estimation of economic indicators from survey samples based on Pareto tail modeling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62:271–286, 2013.

L. A. N. Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, H. E. Stanley, and M. H. Stanley. Scaling behavior in economics: the problem of quantifying company growth. *Physica A: Statistical and Theoretical Physics*, 244:1–24, 1997.

B. C. Arnold. *Pareto Distributions*. CRC Press, Boca Raton FL, 2$^{\text{nd}}$ edition, 2015.

N. Atoda, T. Suruga, and T. Tachibanaki. Statistical inference of functional forms for income distribution. *Economic Studies Quarterly*, 39:14–40, 1988.

C. P. A. Bartels. *Economic Aspects of Regional Welfare: Income Distribution and Unemployment*. Martinus Nijhoff, Leiden, 1977.

C.P.A. Bartels and H. van Metelen. *Alternative Probability Density Functions of Income: A Comparison of the Lognormal-, Gamma- and Weibull-distribution with Dutch Data*. Research Memorandum 29, Faculty of Economics, Vrije Universiteit, Amsterdam, 1975.

J. Beirlant, P. Vynckier, and J. L. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association*, 31:1659–1667, 1996a.

J. Beirlant, P. Vynckier, and J. L. Teugels. Excess functions and estimation of the extreme-value index. *Bernoulli*, 2:293–318, 1996b.

J. Beirlant, Y. Goegebeur, and J. L. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, West Sussex UK, 2004.

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

A. Blank and S. Solomon. Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A: Statistical Mechanics and its Applications*, 287:279–288, 2000.

R. F. Bordley, J. B. McDonald, and A. Mantrala. Something new, something old: Parametric models for the size distribution of income. *Journal of Income Distribution*, 6:91–103, 1996.

G. Bottazzi. On the relationship between firms' size and growth rate. *Economics Bulletin*, 3:1–7, 2008.

G. Bottazzi and A. Secchi. Common properties and sectoral specificities in the dynamics of U.S. manufacturing firms. *Review of Industrial Organization*, 23: 217–232, 2003a.

G. Bottazzi and A. Secchi. Why are distributions of firm growth rates tent-shaped? *Economic Letters*, 80:415–420, 2003b.

G. Bottazzi, A. Coad, N. Jacoby, and A. Secchi. Corporate growth and industrial dynamics: evidence from French manufacturing. *Applied Economics*, 43:103–116, 2011.

J.-P. Bouchaud. Power laws in economics and finance: some ideas from physics. *Quantitative Finance*, 1:105–112, 2001.

K. Brachmann, A. Stich, and M. Trede. Evaluating parametric income distribution models. *Allgemeines Statistisches Archiv*, 80:285–298, 1996.

W. A. Brock. Scaling in economics: A reader's guide. *Industrial and Corporate Change*, 8:409–446, 1999.

D. Canning, L. A. N. Amaral, Y. Lee, M. Meyer, and H. E. Stanley. Scaling the volatility of GDP growth rates. *Economic Letters*, 60:335–341, 1998.

M. Capasso, L. Alessi, M. Barigozzi, and G. Fagiolo. On approximating the distributions of goodness-of-fit test statistics based on the empirical distribution function: The case of unknown parameters. *Advances in Complex System*, 12:157–167, 2009.

C. Castaldi and G. Dosi. *Income Levels and Income Growth: Some New Cross-Country Evidence and Some Interpretative Puzzles*. Working Paper 18, Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa, 2004. Available at http://www.lem.sssup.it/WPLem/files/2004-18.pdf.

D. G. Champernowne. A model of income distribution. *The Economic Journal*, 63:318–351, 1953.

A. Clauset, M. Young, and K. S. Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51:58–88, 2007.

A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661–703, 2009.

K. Cooray and M. M. A. Ananda. Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 2005:321–334, 2005.

R. D'Addario. Intorno ad una funzione di distribuzione. *Giornale degli Economisti e Annali di Economia*, 33:205–214, 1974.

R. B. D'Agostino and M. A. Stephens. *Goodness-of-Fit Techniques*. Marcel Dekker, New York, 1986.

J. Danielsson, L. de Haan, L. Peng, and C. de Vries. Using a bootstrap method to choose sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001.

M. L. Delignette-Muller and C. Dutang. `fitdistrplus`: an `R` package for fitting distributions. *Journal of Statistical Software*, 64:1–34, 2015.

D. Delli Gatti, C. Di Guilmi, E. Gaffeo, G. Giulioni, M. Gallegati, and A. Palestrini. A new approach to business fluctuations: heterogenous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior & Organization*, 56:489–512, 2005.

C. Di Guilmi, E. Gaffeo, and M. Gallegati. Empirical results on the size distribution of business cycle phases. *Physica A: Statistical Mechanics and its Applications*, 333:325–334, 2004.

C. Di Guilmi, E. Gaffeo, M. Gallegati, and A. Palestrini. International evidence on business cycle magnitude dependence: an analyisis of 16 industrialized countries, 1881–2000. *International Journal of Applied Econometrics and Quantitative Studies*, 2:5–16, 2005.

G. Dosi and R. R. Nelson. Technical change and industrial dynamics as evolutionary processes. In B. H. Hall and N. Rosenberg, editors, *Handbook of the Economics of Innovation*, volume 1, pages 51–127. North-Holland, Amsterdam, 2010.

H. Drees, L. de Haan, and S. Resnick. How to make a Hill plot. *The Annals of Statistics*, 28:254–274, 2000.

D. J. Dupuis and M.-P. Victoria-Feser. A robust prediction error criterion for Pareto modelling of upper tails. *The Canadian Journal of Statistics*, 34:639–658, 2006.

S. N. Durlauf. Complexity and empirical economics. *The Economic Journal*, 115:225–243, 2005.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.

P. Espinguet and M. Terraza. Essai d'extrapolation des distributions de salaires français. *Economie Appliquée*, 36:535–561, 1983.

G. Fagiolo, M. Napoletano, and A. Roventini. Sulle distribuzioni dei tassi di crescita dell'output aggregato: un'analisi per Sati Uniti e Italia. *Rivista Internazionale di Scienze Sociali*, 115:215–241, 2007a.

G. Fagiolo, M. Napoletano, and A. Roventini. The european physical journal b. *How do output growth rate distributions look like? Some time-series evidence on OECD countries*, 57:205–211, 2007b.

G. Fagiolo, M. Napoletano, and A. Roventini. Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal Of Applied Econometrics*, 23:639–669, 2008.

W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley & Sons, New York, 2^{nd} edition, 1971.

D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Yamasaki, K. Matia, and H. E. Stanley. The growth of business firms: theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 102:18801–18806, 2005.

X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1:255–294, 2009.

X. Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79: 733–772, 2011.

R. Gibrat. *Les inégalités économiques. Applications: aux inégalités des richesses, à la concentration des entreprises, aux population des villes, aux statistiques des familles, etc., d'une loi nouvelle: la loi de l'effet proportionnel*. Librairie du Recueil Sirey, Paris, 1931.

C. S. Gillespie. Fitting heavy tailed distributions: the `poweRlaw` package. *Journal of Statistical Software*, 64:1–16, 2015.

P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44:37–42, 1982.

B. M. Hill. A simple approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.

A. H. Jessen and T. Mikosch. Regularly varying functions. *Publications de l'Institut Mathématique*, 80:171–192, 2006.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1. John Wiley & Sons, New York, 2^{nd} edition, 1994.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.

C. Kleiber and S. Kotz. *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, New York, 2003.

S. Kotz, T. J. Kozubowski, and K. Podgórski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Basel, 2001.

Y. Lee, L. A. N. Amaral, D. Canning, M. Meyer, and H. E. Stanley. Universal features in the growth dynamics of complex organizations. *Physical Review Letters*, 81:3275–3278, 1998.

M. Levy and S. Solomon. Power laws are logarithmic Boltzmann laws. *International Journal of Modern Physics C*, 7:595–601, 1996a.

M. Levy and S. Solomon. Spontaneous scaling emergence in generic stochastic systems. *International Journal of Modern Physics C*, 7:745–751, 1996b.

P. Lévy. *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris, 1954.

T. Lux. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Applied Financial Economics*, 6: 463–475, 1996.

B. Mandelbrot. The Pareto-Lévy law and the distribution of income. *International Economic Review*, 1:79–106, 1960.

B. Mandelbrot. New methods in statistical economics. *The Journal of Political Economy*, 71:421–440, 1963.

A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44, 1967.

D. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764, 1982.

J. B. McDonald. Some generalized functions for the size distribution of income. *Econometrica*, 52:647–665, 1984.

M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:226–251, 2004.

M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46:323–351, 2005.

J. P. Nolan. *Stable Distributions – Models for Heavy Tailed Data*. Birkhäuser, Boston MA, 2015. In progress, Chapter 1 online at http://academic2.american.edu/~jpnolan/stable/chap1.pdf.

A. Palestrini. Analysis of industrial dynamics: a note on the relationship between firms' size and growth rate. *Economics Letters*, 94:367–371, 2007.

V. Pareto. La legge della domanda. *Giornale degli Economisti*, 10:59–68, 1895.

V. Pareto. La courbe de la répartition de la richesse. 1896. Reprinted in G. Busino, editor, *Œeuvres complètes de Vilfredo Pareto, Tome 3: Écrits sur la courbe de la répartition de la richesse*, pages 1–15, Librairie Droz, Geneva, 1965.

V. Pareto. *Cours d'économie politique*. Macmillan, London, 1897a.

V. Pareto. Aggiunta allo studio della curva delle entrate. *Giornale degli Economisti*, 14:15–26, 1897b.

P. Puig and M. A. Stephens. Tests of fit for the Laplace distribution, with applications. *Technometrics*, 42:417–424, 2000.

R Core Team. *R: A Language and Environment for Statistical Computing*, 2015. https://www.R-project.org/.

T. Reichstein and M. B. Jensen. Firm size and firm growth rate distributions—The case of Denmark. *Industrial and Corporate Change*, 14:1145–1166, 2005.

G. S. Sahota. Theories of personal income distribution: A survey. *Journal of Economic Literature*, 16:1–55, 1978.

S. Sapio and G. Thoma. *The Growth of Industrial Sectors: Theoretical Insights and Empirical Evidence from U.S. Manufacturing*. Working Paper 9, Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa, 2006. Available at http://www.lem.sssup.it/WPLem/files/2006-09.pdf.

G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6: 461–464, 1978.

H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.

D. Sornette. Multiplicative processes and power laws. *Physical Review E*, 57: 4811–4813, 1998.

D. Sornette. *Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Springer, Berlin and New York, 2nd edition, 2006.

D. Sornette. Probability distributions in complex systems. In A. R. Meyers, editor, *Computational Complexity: Theory, Techniques, and Applications*, pages 2286–2300. Springer, New York, 2012.

D. Sornette and R. Cont. Convergent multiplicative processes repelled from zero: power laws and truncated power laws. *Journal de Physique I France*, 7: 431–444, 1997.

M. H. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and H. E. Stanley. Scaling behavior in the growth of companies. *Nature*, 379:804–806, 1996.

M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

M. T. Subbotin. On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301, 1923.

G. Săvoiu and I. I. Simăn. History and role of econophysics in scientific research. In G. Săvoiu, editor, *Econophysics: Background and Applications in Economics, Finance, and Sociophysics*, pages 3–16. Academic Press, Oxford UK and Waltham MA, 2013.

T. Tachibanaki, T. Suruga, and N. Atoda. Estimations of income distribution parameters for individual observations by maximum likelihood method. *Journal of the Japan Statistical Society*, 27:191–203, 1997.

H. Takayasu, A.-H. Sato, and M. Takayasu. Stable infinite variance fluctuations in randomly amplified Langevin systems. *Physical Review Letters*, 79:966–969, 1997.

V. V. Uchaikin and V. M. Zolotarev. *Chance and Stability: Stable Distributions and Their Applications*. VSP International Science Publishers, Utrecht, 1999.

M. Verbeek. *A Guide to Modern Econometrics*. John Wiley & Sons, Chichester UK, 4th edition, 2012.

Y. Virkar and A. Clauset. Power-law distributions in binned empirical data. *Annals of Applied Statistics*, 8:89–119, 2014.

Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333, 1989.

W. Willinger, D. Alderson, J. C. Doyle, and L. Li. More "normal" than normal: Scaling distributions and complex systems. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 130–141, Piscataway, 2004. IEEE Press.

G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge MA, 1949.

Subject index