THEORY OF FINANCIAL RISKS

FROM STATISTICAL PHYSICS TO RISK MANAGEMENT

JEAN-PHILIPPE BOUCHAUD and MARC POTTERS

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/14pt. *System* LaTeX 2_ε [DBD]

A catalogue record of this book is available from the British Library

ISBN 0 521 78232 5 hardback

Contents

1

Probability theory: basic notions

All epistemologic value of the theory of probability is based on this: that large scale random phenomena in their collective action create strict, non random regularity.

(Gnedenko and Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*.)

1.1 Introduction

Randomness stems from our incomplete knowledge of reality, from the lack of information which forbids a perfect prediction of the future. Randomness arises from complexity, from the fact that causes are diverse, that tiny perturbations may result in large effects. For over a century now, Science has abandoned Laplace's deterministic vision, and has fully accepted the task of deciphering randomness and inventing adequate tools for its description. The surprise is that, after all, randomness has many facets and that there are many levels to uncertainty, but, above all, that a new form of predictability appears, which is no longer deterministic but *statistical*.

Financial markets offer an ideal testing ground for these statistical ideas. The fact that a large number of participants, with divergent anticipations and conflicting interests, are simultaneously present in these markets, leads to an unpredictable behaviour. Moreover, financial markets are (sometimes strongly) affected by external news – which are, both in date and in nature, to a large degree unexpected. The statistical approach consists in drawing from past observations some information on the frequency of possible price changes. If one then assumes that these frequencies reflect some intimate mechanism of the markets themselves, then one may hope that these frequencies will remain stable in the course of time. For example, the mechanism underlying the roulette or the game of dice is obviously always the same, and one expects that the frequency of all possible outcomes will be invariant in time – although of course each individual outcome is random.

This ‘bet’ that probabilities are stable (or better, stationary) is very reasonable in the case of roulette or dice;[1] it is nevertheless much less justified in the case of financial markets – despite the large number of participants which confer to the system a certain regularity, at least in the sense of Gnedenko and Kolmogorov. It is clear, for example, that financial markets do not behave now as they did 30 years ago: many factors contribute to the evolution of the way markets behave (development of derivative markets, world-wide and computer-aided trading, etc.). As will be mentioned in the following, ‘young’ markets (such as emergent countries markets) and more mature markets (exchange rate markets, interest rate markets, etc.) behave quite differently. The statistical approach to financial markets is based on the idea that whatever evolution takes place, this happens sufficiently *slowly* (on the scale of several years) so that the observation of the recent past is useful to describe a not too distant future. However, even this ‘weak stability’ hypothesis is sometimes badly in error, in particular in the case of a crisis, which marks a sudden change of market behaviour. The recent example of some Asian currencies indexed to the dollar (such as the Korean won or the Thai baht) is interesting, since the observation of past fluctuations is clearly of no help to predict the amplitude of the sudden turmoil of 1997, see Figure 1.1.

Hence, the statistical description of financial fluctuations is certainly imperfect. It is nevertheless extremely helpful: in practice, the ‘weak stability’ hypothesis is in most cases reasonable, at least to describe *risks*.[2]

In other words, the amplitude of the possible price changes (but not their sign!) is, to a certain extent, predictable. It is thus rather important to devise adequate tools, in order to *control* (if at all possible) financial risks. The goal of this first chapter is to present a certain number of basic notions in probability theory, which we shall find useful in the following. Our presentation does not aim at mathematical rigour, but rather tries to present the key concepts in an intuitive way, in order to ease their empirical use in practical applications.

1.2 Probabilities

1.2.1 Probability distributions

Contrarily to the throw of a dice, which can only return an integer between 1 and 6, the variation of price of a financial asset[3] can be arbitrary (we disregard

[1] The idea that science ultimately amounts to making the best possible guess of reality is due to R. P. Feynman (Seeking New Laws, in *The Character of Physical Laws*, MIT Press, Cambridge, MA, 1965).

[2] The prediction of *future returns* on the basis of past returns is however much less justified.

[3] *Asset* is the generic name for a financial instrument which can be bought or sold, like stocks, currencies, gold, bonds, etc.

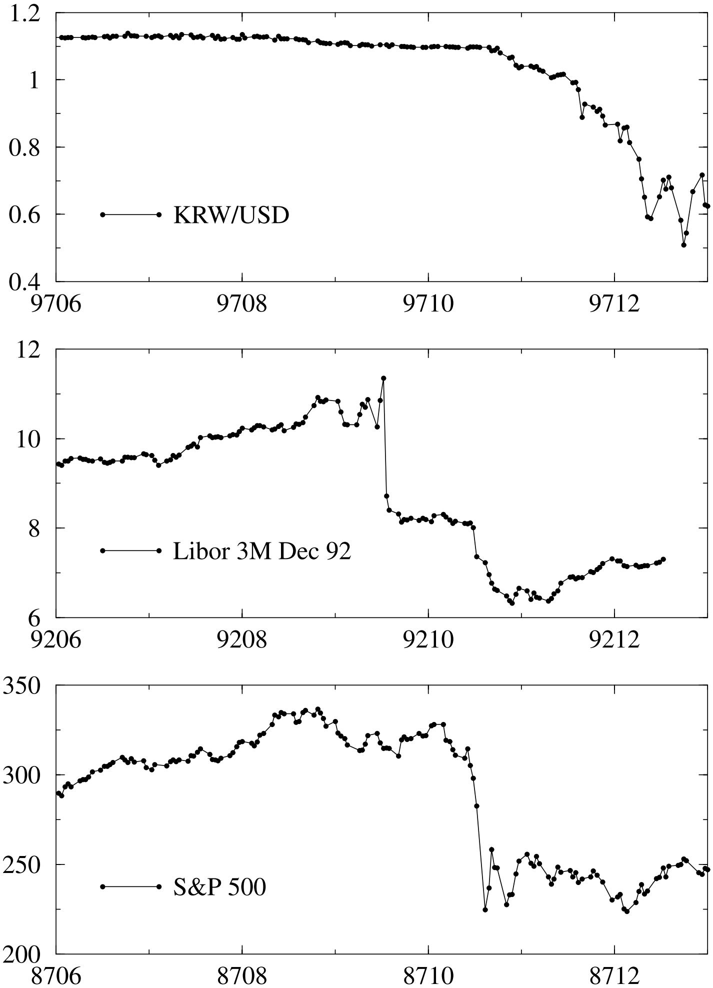

Fig. 1.1. Three examples of statistically unforeseen crashes: the Korean won against the dollar in 1997 (top), the British 3-month short-term interest rates futures in 1992 (middle), and the S&P 500 in 1987 (bottom). In the example of the Korean won, it is particularly clear that the distribution of price changes before the crisis was extremely narrow, and could not be extrapolated to anticipate what happened in the crisis period.

the fact that price changes cannot actually be smaller than a certain quantity – a 'tick'). In order to describe a random process X for which the result is a real number, one uses a probability density $P(x)$, such that the probability that X is within a small interval of width $\mathrm{d}x$ around $X = x$ is equal to $P(x)\,\mathrm{d}x$. In the following, we shall denote as $P(\cdot)$ the probability density for the variable appearing as the argument of the function. This is a potentially ambiguous, but very useful notation.

The probability that X is between a and b is given by the integral of $P(x)$ between a and b,

$$\mathcal{P}(a < X < b) = \int_a^b P(x)\,\mathrm{d}x. \tag{1.1}$$

In the following, the notation $\mathcal{P}(\cdot)$ means the probability of a given event, defined by the content of the parentheses $(\cdot)$.

The function $P(x)$ is a density; in this sense it depends on the units used to measure X. For example, if X is a length measured in centimetres, $P(x)$ is a probability density per unit length, i.e. per centimetre. The numerical value of $P(x)$ changes if X is measured in inches, but the probability that X lies between two specific values l_1 and l_2 is of course independent of the chosen unit. $P(x)\,\mathrm{d}x$ is thus invariant upon a change of unit, i.e. under the change of variable $x \to \gamma x$. More generally, $P(x)\,\mathrm{d}x$ is invariant upon any (monotonic) change of variable $x \to y(x)$: in this case, one has $P(x)\,\mathrm{d}x = P(y)\,\mathrm{d}y$.

In order to be a probability density in the usual sense, $P(x)$ must be non-negative ($P(x) \geq 0$ for all x) and must be normalized, that is that the integral of $P(x)$ over the whole range of possible values for X must be equal to one:

$$\int_{x_m}^{x_M} P(x)\,\mathrm{d}x = 1, \tag{1.2}$$

where x_m (resp. x_M) is the smallest value (resp. largest) which X can take. In the case where the possible values of X are not bounded from below, one takes $x_m = -\infty$, and similarly for x_M. One can actually always assume the bounds to be $\pm\infty$ by setting to zero $P(x)$ in the intervals $]-\infty, x_m]$ and $[x_M, \infty[$. Later in the text, we shall often use the symbol $\int$ as a shorthand for $\int_{-\infty}^{+\infty}$.

An equivalent way of describing the distribution of X is to consider its cumulative distribution $\mathcal{P}_<(x)$, defined as:

$$\mathcal{P}_<(x) \equiv \mathcal{P}(X < x) = \int_{-\infty}^{x} P(x')\,\mathrm{d}x'. \tag{1.3}$$

$\mathcal{P}_<(x)$ takes values between zero and one, and is monotonically increasing with x. Obviously, $\mathcal{P}_<(-\infty) = 0$ and $\mathcal{P}_<(+\infty) = 1$. Similarly, one defines $\mathcal{P}_>(x) = 1 - \mathcal{P}_<(x)$.

1.2.2 *Typical values and deviations*

It is quite natural to speak about ‘typical’ values of X. There are at least three mathematical definitions of this intuitive notion: the *most probable* value, the *median* and the *mean*. The most probable value x^* corresponds to the maximum of the function $P(x)$; x^* needs not be unique if $P(x)$ has several equivalent maxima.

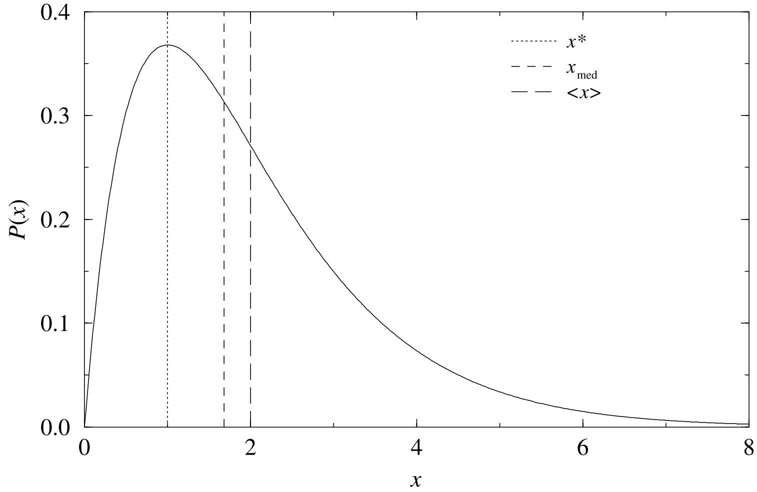

Fig. 1.2. The 'typical value' of a random variable X drawn according to a distribution density $P(x)$ can be defined in at least three different ways: through its mean value $\langle x \rangle$, its most probable value x^* or its median x_{med}. In the general case these three values are distinct.

The median x_{med} is such that the probabilities that X be greater or less than this particular value are equal. In other words, $\mathcal{P}_<(x_{\text{med}}) = \mathcal{P}_>(x_{\text{med}}) = \frac{1}{2}$. The mean, or *expected value* of X, which we shall note as m or $\langle x \rangle$ in the following, is the average of all possible values of X, weighted by their corresponding probability:

$$m \equiv \langle x \rangle = \int x\, P(x)\, \mathrm{d}x. \tag{1.4}$$

For a unimodal distribution (unique maximum), symmetrical around this maximum, these three definitions coincide. However, they are in general different, although often rather close to one another. Figure 1.2 shows an example of a non-symmetric distribution, and the relative position of the most probable value, the median and the mean.

One can then describe the fluctuations of the random variable X: if the random process is repeated several times, one expects the results to be scattered in a cloud of a certain 'width' in the region of typical values of X. This width can be described by the *mean absolute deviation* (MAD) E_{abs}, by the *root mean square* (RMS) σ (or, in financial terms, the *volatility*), or by the 'full width at half maximum' $w_{1/2}$.

The mean absolute deviation from a given reference value is the average of the distance between the possible values of X and this reference value,[4]

$$E_{\mathrm{abs}} \equiv \int |x - x_{\mathrm{med}}| P(x)\,\mathrm{d}x. \tag{1.5}$$

Similarly, the *variance* (σ^2) is the mean distance squared to the reference value m,

$$\sigma^2 \equiv \langle (x-m)^2 \rangle = \int (x-m)^2 P(x)\,\mathrm{d}x. \tag{1.6}$$

Since the variance has the dimension of x squared, its square root (the RMS, σ) gives the order of magnitude of the fluctuations around m.

Finally, the full width at half maximum $w_{1/2}$ is defined (for a distribution which is symmetrical around its unique maximum x^*) such that $P(x^* \pm (w_{1/2})/2) = P(x^*)/2$, which corresponds to the points where the probability density has dropped by a factor of two compared to its maximum value. One could actually define this width slightly differently, for example such that the total probability to find an event outside the interval $[(x^* - w/2), (x^* + w/2)]$ is equal to, say, 0.1.

The pair mean–variance is actually much more popular than the pair median–MAD. This comes from the fact that the absolute value is not an analytic function of its argument, and thus does not possess the nice properties of the variance, such as additivity under convolution, which we shall discuss below. However, for the empirical study of fluctuations, it is sometimes preferable to use the MAD; it is more *robust* than the variance, that is, less sensitive to rare extreme events, which may be the source of large statistical errors.

1.2.3 Moments and characteristic function

More generally, one can define higher-order *moments* of the distribution $P(x)$ as the average of powers of X:

$$m_n \equiv \langle x^n \rangle = \int x^n P(x)\,\mathrm{d}x. \tag{1.7}$$

Accordingly, the mean m is the first moment ($n = 1$), and the variance is related to the second moment ($\sigma^2 = m_2 - m^2$). The above definition, Eq. (1.7), is only meaningful if the integral converges, which requires that $P(x)$ decreases sufficiently rapidly for large $|x|$ (see below).

From a theoretical point of view, the moments are interesting: if they exist, their knowledge is often equivalent to the knowledge of the distribution $P(x)$ itself.[5] In

[4] One chooses as a reference value the median for the MAD and the mean for the RMS, because for a fixed distribution $P(x)$, these two quantities minimize, respectively, the MAD and the RMS.

[5] This is not rigorously correct, since one can exhibit examples of different distribution densities which possess exactly the same moments, see Section 1.3.2 below.

practice however, the high order moments are very hard to determine satisfactorily: as n grows, longer and longer time series are needed to keep a certain level of precision on m_n; these high moments are thus in general not adapted to describe empirical data.

For many computational purposes, it is convenient to introduce the *characteristic function* of $P(x)$, defined as its Fourier transform:

$$\hat{P}(z) \equiv \int e^{izx} P(x)\, dx. \tag{1.8}$$

The function $P(x)$ is itself related to its characteristic function through an inverse Fourier transform:

$$P(x) = \frac{1}{2\pi} \int e^{-izx} \hat{P}(z)\, dz. \tag{1.9}$$

Since $P(x)$ is normalized, one always has $\hat{P}(0) = 1$. The moments of $P(x)$ can be obtained through successive derivatives of the characteristic function at $z = 0$,

$$m_n = (-i)^n \left. \frac{d^n}{dz^n} \hat{P}(z) \right|_{z=0}. \tag{1.10}$$

One finally defines the *cumulants* c_n of a distribution as the successive derivatives of the logarithm of its characteristic function:

$$c_n = (-i)^n \left. \frac{d^n}{dz^n} \log \hat{P}(z) \right|_{z=0}. \tag{1.11}$$

The cumulant c_n is a polynomial combination of the moments m_p with $p \leq n$. For example $c_2 = m_2 - m^2 = \sigma^2$. It is often useful to normalize the cumulants by an appropriate power of the variance, such that the resulting quantities are dimensionless. One thus defines the *normalized cumulants* λ_n,

$$\lambda_n \equiv c_n/\sigma^n. \tag{1.12}$$

One often uses the third and fourth normalized cumulants, called the *skewness* and *kurtosis* (κ),[6]

$$\lambda_3 = \frac{\langle (x-m)^3 \rangle}{\sigma^3} \qquad \kappa \equiv \lambda_4 = \frac{\langle (x-m)^4 \rangle}{\sigma^4} - 3. \tag{1.13}$$

The above definition of cumulants may look arbitrary, but these quantities have remarkable properties. For example, as we shall show in Section 1.5, the cumulants simply add when one sums independent random variables. Moreover a Gaussian distribution (or the normal law of Laplace and Gauss) is characterized by the fact that all cumulants of order larger than two are identically zero. Hence the

[6] Note that it is sometimes $\kappa + 3$, rather than κ itself, which is called the kurtosis.

cumulants, in particular κ, can be interpreted as a measure of the distance between a given distribution $P(x)$ and a Gaussian.

1.2.4 Divergence of moments – asymptotic behaviour

The moments (or cumulants) of a given distribution do not always exist. A necessary condition for the nth moment (m_n) to exist is that the distribution density $P(x)$ should decay faster than $1/|x|^{n+1}$ for $|x|$ going towards infinity, or else the integral, Eq. (1.7), would diverge for $|x|$ large. If one only considers distribution densities that are behaving asymptotically as a power-law, with an exponent $1 + \mu$,

$$P(x) \sim \frac{\mu A_\pm^\mu}{|x|^{1+\mu}} \text{ for } x \to \pm\infty, \tag{1.14}$$

then all the moments such that $n \geq \mu$ are infinite. For example, such a distribution has no finite variance whenever $\mu \leq 2$. [Note that, for $P(x)$ to be a normalizable probability distribution, the integral, Eq. (1.2), must converge, which requires $\mu > 0$.]

The characteristic function of a distribution having an asymptotic power-law behaviour given by Eq. (1.14) is non-analytic around $z = 0$. The small z expansion contains regular terms of the form z^n for $n < \mu$ followed by a non-analytic term $|z|^\mu$ (possibly with logarithmic corrections such as $|z|^\mu \log z$ for integer μ). The derivatives of order larger or equal to μ of the characteristic function thus do not exist at the origin ($z = 0$).

1.3 Some useful distributions

1.3.1 Gaussian distribution

The most commonly encountered distributions are the ‘normal’ laws of Laplace and Gauss, which we shall simply call Gaussian in the following. Gaussians are ubiquitous: for example, the number of *heads* in a sequence of a thousand coin tosses, the exact number of oxygen molecules in the room, the height (in inches) of a randomly selected individual, are all approximately described by a Gaussian distribution.[7] The ubiquity of the Gaussian can be in part traced to the Central Limit Theorem (CLT) discussed at length below, which states that a phenomenon resulting from a large number of small independent causes is Gaussian. There exists however a large number of cases where the distribution describing a complex phenomenon is *not* Gaussian: for example, the amplitude of earthquakes, the velocity differences in a turbulent fluid, the stresses in granular materials, etc., and, as we shall discuss in the next chapter, the price fluctuations of most financial assets.

[7] Although, in the above three examples, the random variable cannot be negative. As we shall discuss below, the Gaussian description is generally only valid in a certain neighbourhood of the maximum of the distribution.

A Gaussian of mean m and root mean square σ is defined as:

$$P_{\mathrm{G}}(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right). \tag{1.15}$$

The median and most probable value are in this case equal to m, whereas the MAD (or any other definition of the width) is proportional to the RMS (for example, $E_{\mathrm{abs}} = \sigma\sqrt{2/\pi}$). For $m = 0$, all the odd moments are zero and the even moments are given by $m_{2n} = (2n-1)(2n-3)\ldots\sigma^{2n} = (2n-1)!!\,\sigma^{2n}$.

All the cumulants of order greater than two are zero for a Gaussian. This can be realized by examining its characteristic function:

$$\hat{P}_{\mathrm{G}}(z) = \exp\left(-\frac{\sigma^2 z^2}{2} + \mathrm{i}mz\right). \tag{1.16}$$

Its logarithm is a second-order polynomial, for which all derivatives of order larger than two are zero. In particular, the kurtosis of a Gaussian variable is zero. As mentioned above, the kurtosis is often taken as a measure of the distance from a Gaussian distribution. When $\kappa > 0$ (*leptokurtic* distributions), the corresponding distribution density has a marked peak around the mean, and rather 'thick' tails. Conversely, when $\kappa < 0$, the distribution density has a flat top and very thin tails. For example, the uniform distribution over a certain interval (for which tails are absent) has a kurtosis $\kappa = -\frac{6}{5}$.

A Gaussian variable is peculiar because 'large deviations' are extremely rare. The quantity $\exp(-x^2/2\sigma^2)$ decays so fast for large x that deviations of a few times σ are nearly impossible. For example, a Gaussian variable departs from its most probable value by more than 2σ only 5% of the times, of more than 3σ in 0.2% of the times, whereas a fluctuation of 10σ has a probability of less than 2×10^{-23}; in other words, it never happens.

1.3.2 Log-normal distribution

Another very popular distribution in mathematical finance is the so-called 'log-normal' law. That X is a log-normal random variable simply means that $\log X$ is normal, or Gaussian. Its use in finance comes from the assumption that the *rate of returns*, rather than the absolute change of prices, are independent random variables. The increments of the logarithm of the price thus asymptotically sum to a Gaussian, according to the CLT detailed below. The log-normal distribution density is thus defined as:[8]

$$P_{\mathrm{LN}}(x) \equiv \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\log^2(x/x_0)}{2\sigma^2}\right), \tag{1.17}$$

the moments of which being: $m_n = x_0^n \mathrm{e}^{n^2\sigma^2/2}$.

In the context of mathematical finance, one often prefers log-normal to Gaussian distributions for several reasons. As mentioned above, the existence of a random rate of return, or random interest rate, naturally leads to log-normal statistics. Furthermore, log-normals account for the following symmetry in the problem of exchange rates:[9] if x is the rate of currency A in terms of currency B, then obviously, $1/x$ is the rate of currency B in terms of A. Under this transformation, $\log x$ becomes $-\log x$ and the description in terms of a log-normal distribution (or in terms of any other even function of $\log x$) is independent of the reference currency. One often hears the following argument in favour of log-normals: since the price of an asset cannot be negative, its statistics cannot be Gaussian since the latter admits in principle negative values, whereas a log-normal excludes them by construction. This is however a red-herring argument, since the description of the fluctuations of the price of a financial asset in terms of Gaussian or log-normal statistics is in any case an *approximation* which is only valid in a certain range. As we shall discuss at length below, these approximations are totally unadapted to describe extreme risks. Furthermore, even if a price drop of more than 100% is in principle possible for a Gaussian process,[10] the error caused by neglecting such an event is much smaller than that induced by the use of either of these two distributions (Gaussian or log-normal). In order to illustrate this point more clearly, consider the probability of observing n times 'heads' in a series of N coin tosses, which is exactly equal to $2^{-N}C_N^n$. It is also well known that in the neighbourhood of $N/2$, $2^{-N}C_N^n$ is very accurately approximated by a Gaussian of variance $N/4$; this is however not contradictory with the fact that $n \geq 0$ by construction!

Finally, let us note that for moderate volatilities (up to say 20%), the two distributions (Gaussian and log-normal) look rather alike, especially in the 'body' of the distribution (Fig. 1.3). As for the tails, we shall see below that Gaussians substantially underestimate their weight, whereas the log-normal predicts that large

8 A log-normal distribution has the remarkable property that the knowledge of all its moments is not sufficient to characterize the corresponding distribution. It is indeed easy to show that the following distribution: $\frac{1}{\sqrt{2\pi}} x^{-1} \exp\left[-\frac{1}{2}(\log x)^2\right][1 + a\sin(2\pi \log x)]$, for $|a| \leq 1$, has moments which are independent of the value of a, and thus coincide with those of a log-normal distribution, which corresponds to $a = 0$ ([Feller] p. 227).

9 This symmetry is however not always obvious. The dollar, for example, plays a special role. This symmetry can only be expected between currencies of similar strength.

10 In the rather extreme case of a 20% annual volatility and a zero annual return, the probability for the price to become negative after a year in a Gaussian description is less than one out of 3 million.

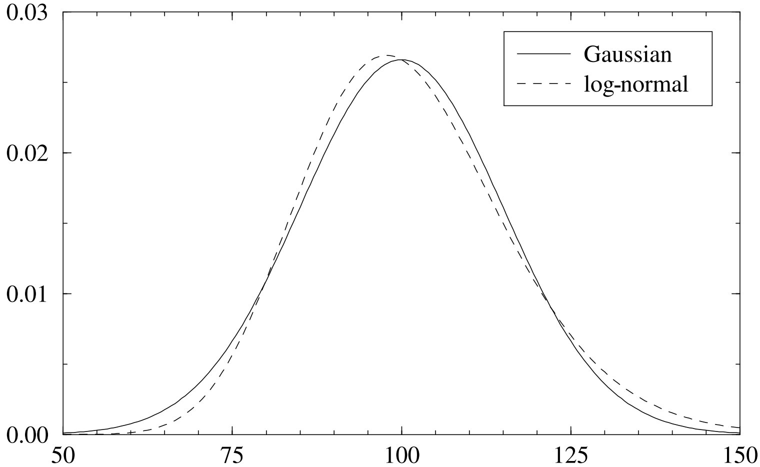

Fig. 1.3. Comparison between a Gaussian (thick line) and a log-normal (dashed line), with $m = x_0 = 100$ and σ equal to 15 and 15% respectively. The difference between the two curves shows up in the tails.

positive jumps are more frequent than large negative jumps. This is at variance with empirical observation: the distributions of absolute stock price changes are rather symmetrical; if anything, large negative draw-downs are more frequent than large positive draw-ups.

1.3.3 Lévy distributions and Paretian tails

Lévy distributions (noted $L_\mu(x)$ below) appear naturally in the context of the CLT (see below), because of their stability property under addition (a property shared by Gaussians). The tails of Lévy distributions are however much 'fatter' than those of Gaussians, and are thus useful to describe multiscale phenomena (i.e. when both very large and very small values of a quantity can commonly be observed – such as personal income, size of pension funds, amplitude of earthquakes or other natural catastrophes, etc.). These distributions were introduced in the 1950s and 1960s by Mandelbrot (following Pareto) to describe personal income and the price changes of some financial assets, in particular the price of cotton [Mandelbrot]. An important constitutive property of these Lévy distributions is their power-law behaviour for large arguments, often called 'Pareto tails':

$$L_\mu(x) \sim \frac{\mu A_\pm^\mu}{|x|^{1+\mu}} \text{ for } x \to \pm\infty, \tag{1.18}$$

where $0 < \mu < 2$ is a certain exponent (often called α), and $A_\pm^\mu$ two constants which we call *tail amplitudes*, or *scale parameters*: $A_\pm$ indeed gives the order of magnitude of the large (positive or negative) fluctuations of x. For instance, the probability to draw a number larger than x decreases as $\mathcal{P}_>(x) = (A_+/x)^\mu$ for large positive x.

One can of course in principle observe Pareto tails with $\mu \geq 2$; but, those tails do not correspond to the asymptotic behaviour of a Lévy distribution.

In full generality, Lévy distributions are characterized by an *asymmetry parameter* defined as $\beta \equiv (A_+^\mu - A_-^\mu)/(A_+^\mu + A_-^\mu)$, which measures the relative weight of the positive and negative tails. We shall mostly focus in the following on the symmetric case $\beta = 0$. The fully asymmetric case ($\beta = 1$) is also useful to describe strictly positive random variables, such as, for example, the time during which the price of an asset remains below a certain value, etc.

An important consequence of Eq. (1.14) with $\mu \leq 2$ is that the variance of a Lévy distribution is formally infinite: the probability density does not decay fast enough for the integral, Eq. (1.6), to converge. In the case $\mu \leq 1$, the distribution density decays so slowly that even the mean, or the MAD, fail to exist.[11] The scale of the fluctuations, defined by the width of the distribution, is always set by $A = A_+ = A_-$.

There is unfortunately no simple analytical expression for symmetric Lévy distributions $L_\mu(x)$, except for $\mu = 1$, which corresponds to a Cauchy distribution (or 'Lorentzian'):

$$L_1(x) = \frac{A}{x^2 + \pi^2 A^2}. \tag{1.19}$$

However, the characteristic function of a symmetric Lévy distribution is rather simple, and reads:

$$\hat{L}_\mu(z) = \exp\left(-a_\mu |z|^\mu\right), \tag{1.20}$$

where a_μ is a certain constant, proportional to the tail parameter A^μ.[12] It is thus clear that in the limit $\mu = 2$, one recovers the definition of a Gaussian. When μ decreases from 2, the distribution becomes more and more sharply peaked around the origin and fatter in its tails, while 'intermediate' events lose weight (Fig. 1.4). These distributions thus describe 'intermittent' phenomena, very often small, sometimes gigantic.

Note finally that Eq. (1.20) does not define a probability distribution when $\mu > 2$, because its inverse Fourier transform is not everywhere positive.

In the case $\beta \neq 0$, *one would have:*

$$\hat{L}_\mu^\beta(z) = \exp\left[-a_\mu |z|^\mu \left(1 + \mathrm{i}\beta \tan(\mu\pi/2)\frac{z}{|z|}\right)\right] \quad (\mu \neq 1). \tag{1.21}$$

[11] The median and the most probable value however still exist. For a symmetric Lévy distribution, the most probable value defines the so-called 'localization' parameter m.

[12] For example, when $1 < \mu < 2$, $A^\mu = \mu\Gamma(\mu - 1)\sin(\pi\mu/2)a_\mu/\pi$.

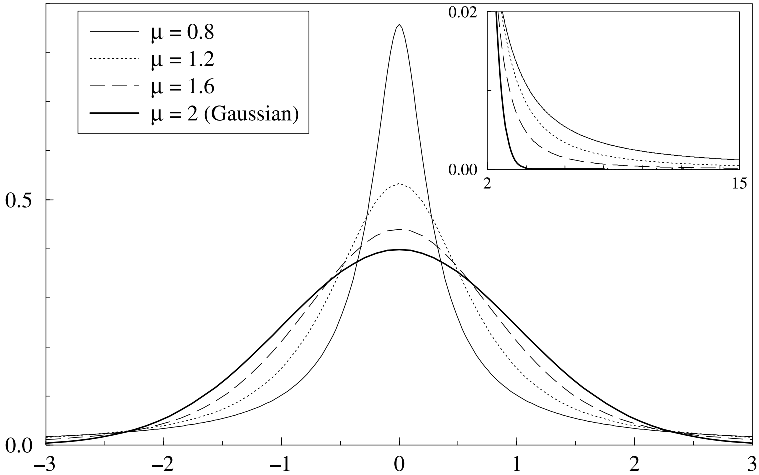

Fig. 1.4. Shape of the symmetric Lévy distributions with $\mu = 0.8,\ 1.2,\ 1.6$ and 2 (this last value actually corresponds to a Gaussian). The smaller μ, the sharper the 'body' of the distribution, and the fatter the tails, as illustrated in the inset.

It is important to notice that while the leading asymptotic term for large x is given by Eq. (1.18), there are subleading terms which can be important for finite x. The full asymptotic series actually reads:

$$L_\mu(x) = \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{\pi n!} \frac{a_\mu^n}{x^{1+n\mu}} \Gamma(1+n\mu)\sin(\pi\mu n/2). \tag{1.22}$$

The presence of the subleading terms may lead to a bad empirical estimate of the exponent μ based on a fit of the tail of the distribution. In particular, the 'apparent' exponent which describes the function L_μ for finite x is larger than μ, and decreases towards μ for $x \to \infty$, but more and more slowly as μ gets nearer to the Gaussian value $\mu = 2$, for which the power-law tails no longer exist. Note however that one also often observes empirically the opposite behaviour, i.e. an apparent Pareto exponent which *grows* with x. This arises when the Pareto distribution, Eq. (1.18), is only valid in an intermediate regime $x \ll 1/\alpha$, beyond which the distribution decays exponentially, say as $\exp(-\alpha x)$. The Pareto tail is then 'truncated' for large values of x, and this leads to an effective μ which grows with x.

An interesting generalization of the Lévy distributions which accounts for this exponential cut-off is given by the 'truncated Lévy distributions' (TLD), which will be of much use in the following. A simple way to alter the characteristic function

Eq. (1.20) to account for an exponential cut-off for large arguments is to set:[13]

$$\hat{L}_{\mu}^{(t)}(z) = \exp\left[-a_{\mu}\frac{(\alpha^2+z^2)^{\frac{\mu}{2}}\cos\left(\mu \arctan(|z|/\alpha)\right) - \alpha^{\mu}}{\cos(\pi\mu/2)}\right], \qquad (1.23)$$

for $1 \leq \mu \leq 2$. The above form reduces to Eq. (1.20) for $\alpha = 0$. Note that the argument in the exponential can also be written as:

$$\frac{a_{\mu}}{2\cos(\pi\mu/2)}\left[(\alpha+\mathrm{i}z)^{\mu} + (\alpha-\mathrm{i}z)^{\mu} - 2\alpha^{\mu}\right]. \qquad (1.24)$$

Exponential tail: a limiting case

Very often in the following, we shall notice that in the formal limit $\mu \to \infty$, the power-law tail becomes an exponential tail, if the tail parameter is simultaneously scaled as $A^{\mu} = (\mu/\alpha)^{\mu}$. Qualitatively, this can be understood as follows: consider a probability distribution restricted to positive x, which decays as a power-law for large x, defined as:

$$\mathcal{P}_{>}(x) = \frac{A^{\mu}}{(A+x)^{\mu}}. \qquad (1.25)$$

This shape is obviously compatible with Eq. (1.18), and is such that $\mathcal{P}_{>}(x=0) = 1$. If $A = (\mu/\alpha)$, one then finds:

$$\mathcal{P}_{>}(x) = \frac{1}{[1+(\alpha x/\mu)]^{\mu}} \underset{\mu\to\infty}{\longrightarrow} \exp(-\alpha x). \qquad (1.26)$$

1.3.4 Other distributions (*)

There are obviously a very large number of other statistical distributions useful to describe random phenomena. Let us cite a few, which often appear in a financial context:

- The discrete Poisson distribution: consider a set of points randomly scattered on the real axis, with a certain density ω (e.g. the times when the price of an asset changes). The number of points n in an arbitrary interval of length ℓ is distributed according to the Poisson distribution:

$$P(n) \equiv \frac{(\omega\ell)^n}{n!}\exp(-\omega\ell). \qquad (1.27)$$

- The hyperbolic distribution, which interpolates between a Gaussian 'body' and exponential tails:

$$P_{\mathrm{H}}(x) \equiv \frac{1}{2x_0 K_1(\alpha x_0)}\exp -[\alpha\sqrt{x_0^2+x^2}], \qquad (1.28)$$

  where the normalization $K_1(\alpha x_0)$ is a modified Bessel function of the second

[13] See I. Koponen, Analytic approach to the problem of convergence to truncated Lévy flights towards the Gaussian stochastic process, *Physical Review E*, **52**, 1197 (1995).

kind. For x small compared to x_0, $P_H(x)$ behaves as a Gaussian although its asymptotic behaviour for $x \gg x_0$ is fatter and reads $\exp(-\alpha|x|)$.

From the characteristic function

$$\hat{P}_H(z) = \frac{\alpha x_0 K_1(x_0\sqrt{1+\alpha z})}{K_1(\alpha x_0)\sqrt{1+\alpha z}}, \tag{1.29}$$

we can compute the variance

$$\sigma^2 = \frac{x_0 K_2(\alpha x_0)}{\alpha K_1(\alpha x_0)}, \tag{1.30}$$

and kurtosis

$$\kappa = 3\left(\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)}\right)^2 + \frac{12}{\alpha x_0}\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)} - 3. \tag{1.31}$$

Note that the kurtosis of the hyperbolic distribution is always between zero and three. In the case $x_0 = 0$, one finds the symmetric exponential distribution:

$$P_E(x) = \frac{\alpha}{2}\exp(-\alpha|x|), \tag{1.32}$$

with even moments $m_{2n} = (2n)!\,\alpha^{-2n}$, which gives $\sigma^2 = 2\alpha^{-2}$ and $\kappa = 3$. Its characteristic function reads: $\hat{P}_E(z) = \alpha^2/(\alpha^2 + z^2)$.

- The Student distribution, which also has power-law tails:

$$P_S(x) \equiv \frac{1}{\sqrt{\pi}}\frac{\Gamma((1+\mu)/2)}{\Gamma(\mu/2)}\frac{a^\mu}{(a^2+x^2)^{(1+\mu)/2}}, \tag{1.33}$$

which coincides with the Cauchy distribution for $\mu = 1$, and tends towards a Gaussian in the limit $\mu \to \infty$, provided that a^2 is scaled as μ. The even moments of the Student distribution read: $m_{2n} = (2n-1)!!\,\Gamma(\mu/2 - n)/\Gamma(\mu/2)\,(a^2/2)^n$, provided $2n < \mu$; and are infinite otherwise. One can check that in the limit $\mu \to \infty$, the above expression gives back the moments of a Gaussian: $m_{2n} = (2n-1)!!\,\sigma^{2n}$. Figure 1.5 shows a plot of the Student distribution with $\kappa = 1$, corresponding to $\mu = 10$.

1.4 Maximum of random variables – statistics of extremes

If one observes a series of N independent realizations of the same random phenomenon, a question which naturally arises, in particular when one is concerned about risk control, is to determine the order of magnitude of the *maximum* observed value of the random variable (which can be the price drop of a financial asset, or the water level of a flooding river, etc.). For example, in Chapter 3, the so-called ‘value-at-risk’ (VaR) on a typical time horizon will be defined as the possible maximum loss over that period (within a certain confidence level).

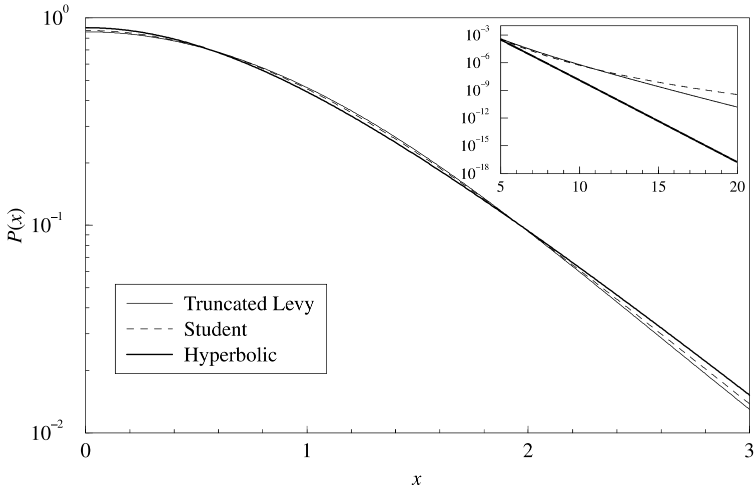

Fig. 1.5. Probability density for the truncated Lévy ($\mu = \frac{3}{2}$), Student and hyperbolic distributions. All three have two free parameters which were fixed to have unit variance and kurtosis. The inset shows a blow-up of the tails where one can see that the Student distribution has tails similar to (but slightly thicker than) those of the truncated Lévy.

The law of large numbers tells us that an event which has a probability p of occurrence appears on average Np times on a series of N observations. One thus expects to observe events which have a probability of at least $1/N$. It would be surprising to encounter an event which has a probability much smaller than $1/N$. The order of magnitude of the largest event, $\Lambda_{\max}$, observed in a series of N independent identically distributed (iid) random variables is thus given by:

$$\mathcal{P}_>(\Lambda_{\max}) = 1/N. \tag{1.34}$$

More precisely, the full probability distribution of the maximum value $x_{\max} = \max_{i=1,N}\{x_i\}$, is relatively easy to characterize; this will justify the above simple criterion Eq. (1.34). The cumulative distribution $\mathcal{P}(x_{\max} < \Lambda)$ is obtained by noticing that if the maximum of all x_i's is smaller than Λ, all of the x_i's must be smaller than Λ. If the random variables are iid, one finds:

$$\mathcal{P}(x_{\max} < \Lambda) = [\mathcal{P}_<(\Lambda)]^N. \tag{1.35}$$

Note that this result is general, and does not rely on a specific choice for $P(x)$. When Λ is large, it is useful to use the following approximation:

$$\mathcal{P}(x_{\max} < \Lambda) = [1 - \mathcal{P}_>(\Lambda)]^N \simeq e^{-N\mathcal{P}_>(\Lambda)}. \tag{1.36}$$

Since we now have a simple formula for the distribution of $x_{\max}$, one can invert it in order to obtain, for example, the median value of the maximum, noted Λ_{med}, such that $\mathcal{P}(x_{\max} < \Lambda_{\text{med}}) = \frac{1}{2}$:

$$\mathcal{P}_>(\Lambda_{\text{med}}) = 1 - \left(\tfrac{1}{2}\right)^{1/N} \simeq \frac{\log 2}{N}. \tag{1.37}$$

More generally, the value Λ_p which is greater than $x_{\max}$ with probability p is given by

$$\mathcal{P}_>(\Lambda_p) \simeq -\frac{\log p}{N}. \tag{1.38}$$

The quantity $\Lambda_{\max}$ defined by Eq. (1.34) above is thus such that $p = 1/\text{e} \simeq 0.37$. The probability that $x_{\max}$ is even *larger* than $\Lambda_{\max}$ is thus 63%. As we shall now show, $\Lambda_{\max}$ also corresponds, in many cases, to the *most probable value* of $x_{\max}$.

Equation (1.38) will be very useful in Chapter 3 to estimate a maximal potential loss within a certain confidence level. For example, the largest daily loss Λ expected next year, with 95% confidence, is defined such that $\mathcal{P}_<(-\Lambda) = -\log(0.95)/250$, where $\mathcal{P}_<$ is the cumulative distribution of daily price changes, and 250 is the number of market days per year.

Interestingly, the distribution of $x_{\max}$ only depends, when N is large, on the asymptotic behaviour of the distribution of x, $P(x)$, when $x \to \infty$. For example, if $P(x)$ behaves as an exponential when $x \to \infty$, or more precisely if $\mathcal{P}_>(x) \sim \exp(-\alpha x)$, one finds:

$$\Lambda_{\max} = \frac{\log N}{\alpha}, \tag{1.39}$$

which grows very slowly with N.[14] Setting $x_{\max} = \Lambda_{\max} + (u/\alpha)$, one finds that the deviation u around $\Lambda_{\max}$ is distributed according to the Gumbel distribution:

$$P(u) = \text{e}^{-\text{e}^{-u}}\text{e}^{-u}. \tag{1.40}$$

The most probable value of this distribution is $u = 0$.[15] This shows that $\Lambda_{\max}$ is the most probable value of $x_{\max}$. The result, Eq. (1.40), is actually much more general, and is valid as soon as $P(x)$ decreases more rapidly than any power-law for $x \to \infty$: the deviation between $\Lambda_{\max}$ (defined as Eq. (1.34)) and $x_{\max}$ is always distributed according to the Gumbel law, Eq. (1.40), up to a scaling factor in the definition of u.

The situation is radically different if $P(x)$ decreases as a power-law, cf. Eq. (1.14). In this case,

$$\mathcal{P}_>(x) \simeq \frac{A_+^\mu}{x^\mu}, \tag{1.41}$$

[14] For example, for a symmetric exponential distribution $P(x) = \exp(-|x|)/2$, the median value of the maximum of $N = 10\,000$ variables is only 6.3.

[15] This distribution is discussed further in the context of financial risk control in Section 3.1.2, and drawn in Figure 3.1.

and the typical value of the maximum is given by:

$$\Lambda_{\max} = A_+ N^{\frac{1}{\mu}}. \tag{1.42}$$

Numerically, for a distribution with $\mu = \frac{3}{2}$ and a scale factor $A_+ = 1$, the largest of $N = 10\,000$ variables is on the order of 450, whereas for $\mu = \frac{1}{2}$ it is one hundred million! The complete distribution of the maximum, called the Fréchet distribution, is given by:

$$P(u) = \frac{\mu}{u^{1+\mu}} e^{-1/u^{\mu}} \qquad u = \frac{x_{\max}}{A_+ N^{\frac{1}{\mu}}}. \tag{1.43}$$

Its asymptotic behaviour for $u \to \infty$ is still a power-law of exponent $1 + \mu$. Said differently, both power-law tails and exponential tails *are stable with respect to the 'max' operation*.[16] The most probable value $x_{\max}$ is now equal to $(\mu/1+\mu)^{1/\mu} \Lambda_{\max}$. As mentioned above, the limit $\mu \to \infty$ formally corresponds to an exponential distribution. In this limit, one indeed recovers $\Lambda_{\max}$ as the most probable value.

Equation (1.42) allows us to discuss intuitively the divergence of the mean value for $\mu \leq 1$ and of the variance for $\mu \leq 2$. If the mean value exists, the sum of N random variables is typically equal to Nm, where m is the mean (see also below). But when $\mu < 1$, the largest encountered value of X is on the order of $N^{1/\mu} \gg N$, and would thus be larger than the entire sum. Similarly, as discussed below, when the variance exists, the RMS of the sum is equal to $\sigma\sqrt{N}$. But for $\mu < 2$, $x_{\max}$ grows faster than $\sqrt{N}$.

More generally, one can rank the random variables x_i in decreasing order, and ask for an estimate of the nth encountered value, noted $\Lambda[n]$ below. (In particular, $\Lambda[1] = x_{\max}$). The distribution P_n of $\Lambda[n]$ can be obtained in full generality as:

$$P_n(\Lambda[n]) = N C_{N-1}^{n-1} \, P(x = \Lambda[n]) \, (\mathcal{P}(x > \Lambda[n])^{n-1} (\mathcal{P}(x < \Lambda[n])^{N-n}. \tag{1.44}$$

The previous expression means that one has first to choose $\Lambda[n]$ among N variables (N ways), $n-1$ variables among the $N-1$ remaining as the $n-1$ largest ones (C_{N-1}^{n-1} ways), and then assign the corresponding probabilities to the configuration where $n-1$ of them are larger than $\Lambda[n]$ and $N-n$ are smaller than $\Lambda[n]$. One can study the position $\Lambda^*[n]$ of the maximum of P_n, and also the width of P_n, defined from the second derivative of $\log P_n$ calculated at $\Lambda^*[n]$. The calculation simplifies in the limit where $N \to \infty$, $n \to \infty$, with the ratio n/N fixed. In this limit, one finds a relation which generalizes Eq. (1.34):

$$\mathcal{P}_>(\Lambda^*[n]) = n/N. \tag{1.45}$$

[16] A third class of laws, stable under 'max' concerns random variables, which are bounded from above – i.e. such that $P(x) = 0$ for $x > x_M$, with x_M finite. This leads to the Weibull distributions, which we will not consider further in this book.

The width w_n of the distribution is found to be given by:

$$w_n = \frac{1}{\sqrt{N}} \frac{\sqrt{1-(n/N)^2}}{P(x = \Lambda^*[n])}, \tag{1.46}$$

which shows that in the limit $N \to \infty$, the value of the nth variable is more and more sharply peaked around its most probable value $\Lambda^*[n]$, given by Eq. (1.45).

In the case of an exponential tail, one finds that $\Lambda^*[n] \simeq \log(N/n)/\alpha$; whereas in the case of power-law tails, one rather obtains:

$$\Lambda^*[n] \simeq A_+ \left(\frac{N}{n}\right)^{\frac{1}{\mu}}. \tag{1.47}$$

This last equation shows that, for power-law variables, the encountered values are hierarchically organized: for example, the ratio of the largest value $x_{\max} \equiv \Lambda[1]$ to the second largest $\Lambda[2]$ is of the order of $2^{1/\mu}$, which becomes larger and larger as μ decreases, and conversely tends to one when $\mu \to \infty$.

The property, Eq. (1.47) is very useful in identifying empirically the nature of the tails of a probability distribution. One sorts in decreasing order the set of observed values $\{x_1, x_2, \ldots, x_N\}$ and one simply draws $\Lambda[n]$ as a function of n. If the variables are power-law distributed, this graph should be a straight line in log–log plot, with a slope $-1/\mu$, as given by Eq. (1.47) (Fig. 1.6). On the same figure, we have shown the result obtained for exponentially distributed variables. On this diagram, one observes an approximately straight line, but with an effective slope which varies with the total number of points N: the slope is less and less as N/n grows larger. In this sense, the formal remark made above, that an exponential distribution could be seen as a power-law with $\mu \to \infty$, becomes somewhat more concrete. Note that if the axes x and y of Figure 1.6 are interchanged, then according to Eq. (1.45), one obtains an estimate of the cumulative distribution, $\mathcal{P}_>$.

Let us finally note another property of power-laws, potentially interesting for their empirical determination. If one computes the average value of x conditioned to a certain minimum value Λ:

$$\langle x \rangle_\Lambda = \frac{\int_\Lambda^\infty x P(x)\,\mathrm{d}x}{\int_\Lambda^\infty P(x)\,\mathrm{d}x}, \tag{1.48}$$

then, if $P(x)$ decreases as in Eq. (1.14), one finds, for $\Lambda \to \infty$,

$$\langle x \rangle_\Lambda = \frac{\mu}{\mu - 1}\Lambda, \tag{1.49}$$

independently of the tail amplitude A_+^μ.[17] The average $\langle x \rangle_\Lambda$ is thus always of the same order as Λ itself, with a proportionality factor which diverges as $\mu \to 1$.

[17] This means that μ can be determined by a one parameter fit only.

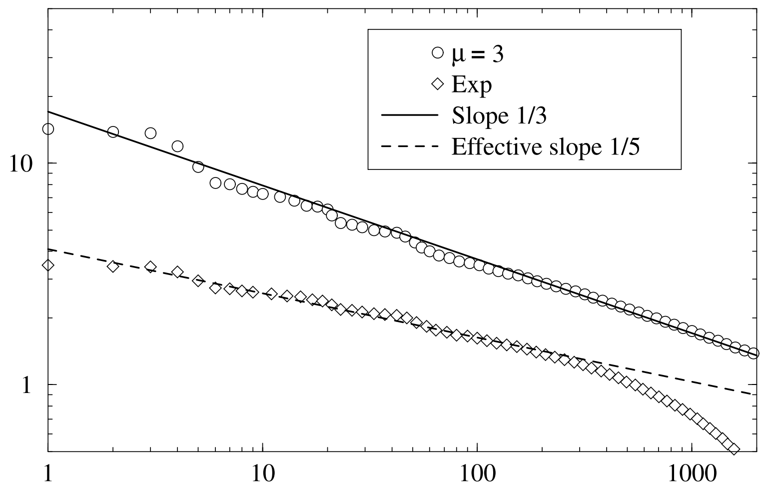

Fig. 1.6. Amplitude versus rank plots. One plots the value of the nth variable $\Lambda[n]$ as a function of its rank n. If $P(x)$ behaves asymptotically as a power-law, one obtains a straight line in log–log coordinates, with a slope equal to $-1/\mu$. For an exponential distribution, one observes an effective slope which is smaller and smaller as N/n tends to infinity. The points correspond to synthetic time series of length 5000, drawn according to a power-law with $\mu = 3$, or according to an exponential. Note that if the axes x and y are interchanged, then according to Eq. (1.45), one obtains an estimate of the cumulative distribution, $\mathcal{P}_>$.

1.5 Sums of random variables

In order to describe the statistics of future prices of a financial asset, one *a priori* needs a distribution density for all possible time intervals, corresponding to different trading time horizons. For example, the distribution of 5-min price fluctuations is different from the one describing daily fluctuations, itself different for the weekly, monthly, etc. variations. But in the case where the fluctuations are independent and identically distributed (iid), an assumption which is, however, usually not justified, see Sections 1.7 and 2.4, it is possible to reconstruct the distributions corresponding to different time scales from the knowledge of that describing short time scales only. In this context, Gaussians and Lévy distributions play a special role, because they are stable: if the short time scale distribution is a stable law, then the fluctuations on all time scales are described by the same stable law – only the parameters of the stable law must be changed (in particular its width). More generally, if one sums iid variables, then, independently of the short time distribution, the law describing long times converges towards one of the stable laws: this is the content of the 'central limit theorem' (CLT). In practice, however, this convergence can be very slow and thus of limited interest, in particular if one is concerned about short time scales.