THEORY OF FINANCIAL RISK:

Basic notions in probability

Jean-Philippe Bouchaud
and
Marc Potters

DRAFT
November 10, 2000

This text corresponds to Chapter 1 of
Theory of Financial Risk
Evaluation copy available for free at
http://www.science-finance.fr/

send comments, typos, etc. to
book@science-finance.fr

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 et 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, and in assuming that these frequencies reflect some intimate mechanism of the markets themselves, implying 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 thirty years ago: many factors contribute to the evolution of the way markets behave (development of derivative markets, worldwide 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 sudden turmoil of 1997 – see Fig. 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] 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, 1967).

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

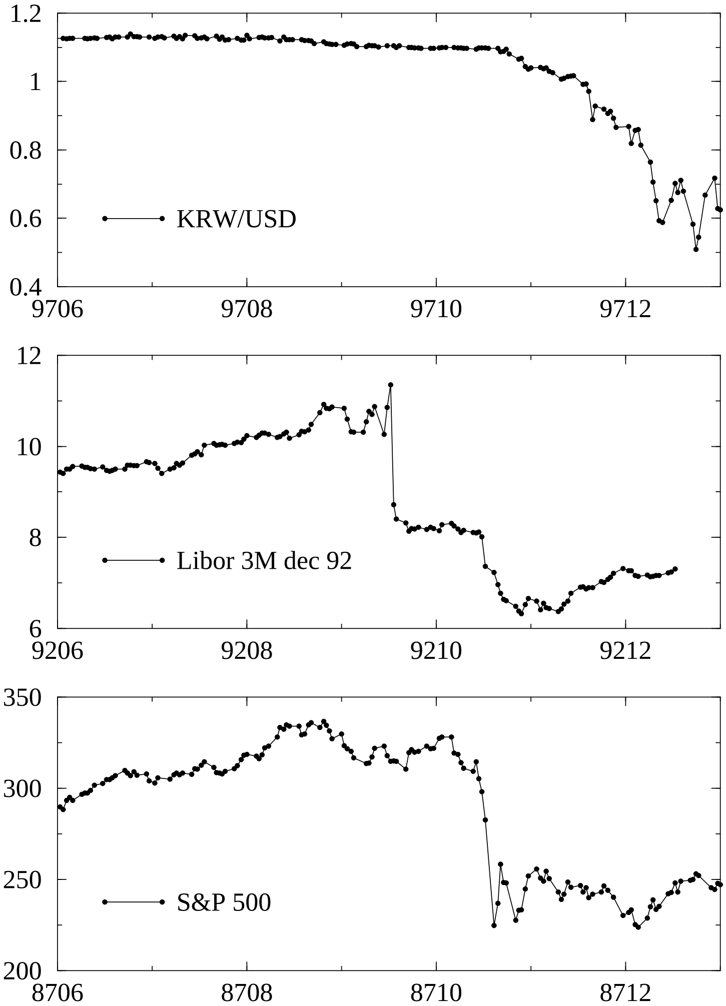

Figure 1.1: Three examples of statistically unforseen crashes: the Korean won against the dollar in 1997 (top), the British 3 month short term interest rates futures in 1992 (middle), and the S&P 500 in 1987 (bottom). In the example of the Korean Won, it is particularly clear that the distribution of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

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 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 dx around $X = x$ is equal to $P(x)dx$. In the following, we shall denote as $P(.)$ 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,

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

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

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. $P(x)dx$ is thus invariant upon a change of unit, i.e. under the change of variable $x \to \gamma x$. More generally, $P(x)dx$ is invariant upon any (monotonous) change of variable $x \to y(x)$: in this case, one has $P(x)dx = P(y)dy$.

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 normalised, that is that the integral of $P(x)$ over the whole range of possible values for X must be equal to one:

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

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 $\int_{-\infty}^{+\infty}$.

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

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

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

$\mathcal{P}_<(x)$ takes values between zero and one, and is monotonously 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 rather 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. The median $x_{\rm med}$ is such that the probabilities that X be greater or less than this particular value are equal. In other words, $\mathcal{P}_<(x_{\rm med}) = \mathcal{P}_>(x_{\rm med}) = \frac{1}{2}$. 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:

$$m \equiv \langle x \rangle = \int xP(x)dx. \tag{1.4}$$

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) $E_{\rm abs}$, 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]

$$E_{\rm abs} \equiv \int |x - x_{\rm med}|P(x)dx. \tag{1.5}$$

[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 minimise, respectively, the MAD and the RMS.

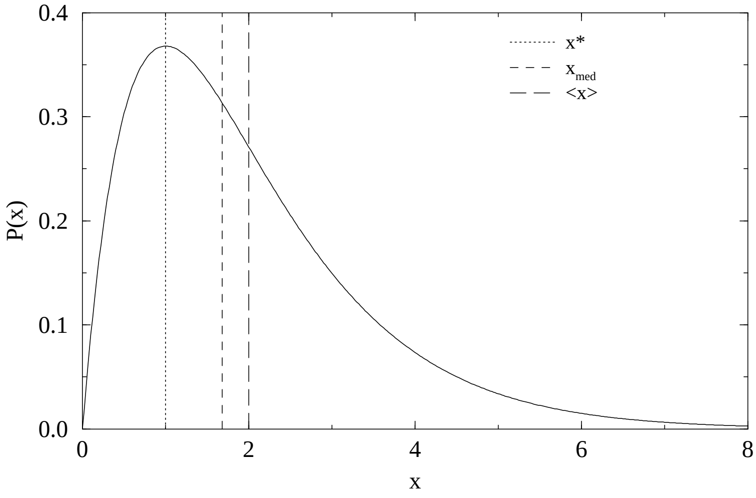

Figure 1.2: The 'typical value' of a random variable X drawn according to a distribution density $P(x)$ can be defined in at least three different ways: through its mean value $\langle x \rangle$, its most probable value x^* or its median x_{med}. In the general case these three values are distinct.

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

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

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 $P(x^* \pm \frac{w_{1/2}}{2}) = \frac{P(x^*)}{2}$, which corresponds to the points where the probability density has dropped by a factor 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 $[x^* - \frac{w}{2}, x^* + \frac{w}{2}]$ 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, 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:

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

Accordingly, the mean m is the first moment ($n = 1$), while the variance is related to the second moment ($\sigma^2 = m_2 - m^2$). The above definition (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 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 normalised, 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 define the *cumulants* c_n of a distribution as the successive derivatives of the logarithm of its characteristic function:

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

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

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 normalise the cumulants by an appropriate power of the variance, such that the resulting quantity is dimensionless. One thus define the *normalised cumulants* λ_n,

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

One often uses the third and fourth normalised 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 characterised by the fact that all cumulants of order larger than two are identically zero. Hence the 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 n^{th} 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 (1.7) would diverge for $|x|$ large. If one restricts to distribution densities 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 normalisable probability distribution, the integral (1.2) must converge, which requires $\mu > 0$.]

The characteristic function of a distribution having an asymptotic power law behaviour given by (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 ($z = 0$).

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

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 in the following Gaussians. 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 next chapter, the price fluctuations of most financial assets.

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

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

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

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

$$\hat{P}_G(z) = \exp\left(-\frac{\sigma^2 z^2}{2} + imz\right). \tag{1.16}$$

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

[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.

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, while 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]

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

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

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

[8] A log-normal distribution has the remarkable property that the knowledge of all its moments is not sufficient to characterise the corresponding distribution. It is indeed easy to show that the following distribution: $\frac{1}{\sqrt{2\pi}} x^{-1} e^{-\frac{1}{2}(\log x)^2}[1 + a\sin(2\pi \log x)]$, 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.

the price of an asset cannot be negative, its statistics cannot be Gaussian since the latter admits in principle negative values, while 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 be 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, specially in the 'body' of the distribution (Fig. 1.3). As for the tails, we shall see below that Gaussians substantially underestimate their weight, while the log-normal predicts that large 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 fifties and sixties 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

[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 three million

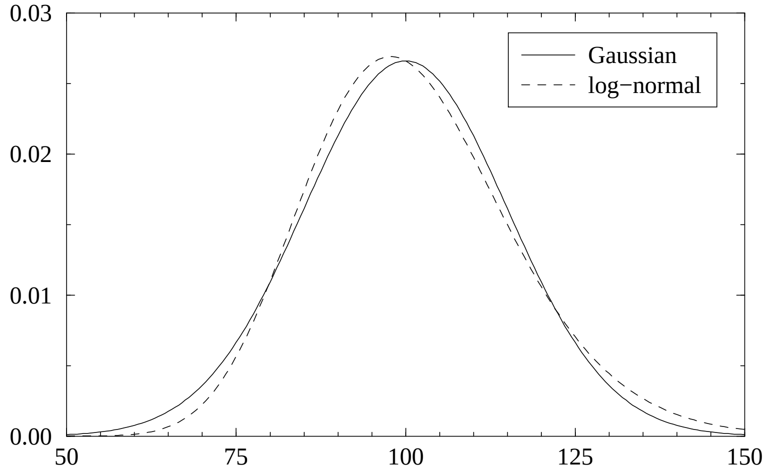

Figure 1.3: Comparison between a Gaussian (thick line) and a log-normal (dashed line), with $m = x_0 = 100$ and σ equal to 15 and 15% respectively. The difference between the two curves shows up in the tails.

'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 $\mu \geq 2$, however, those tails do not correspond to the asymptotic behaviour of a Lévy distribution.

In full generality, Lévy distributions are characterised 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 (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 (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 loose 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:

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

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

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

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

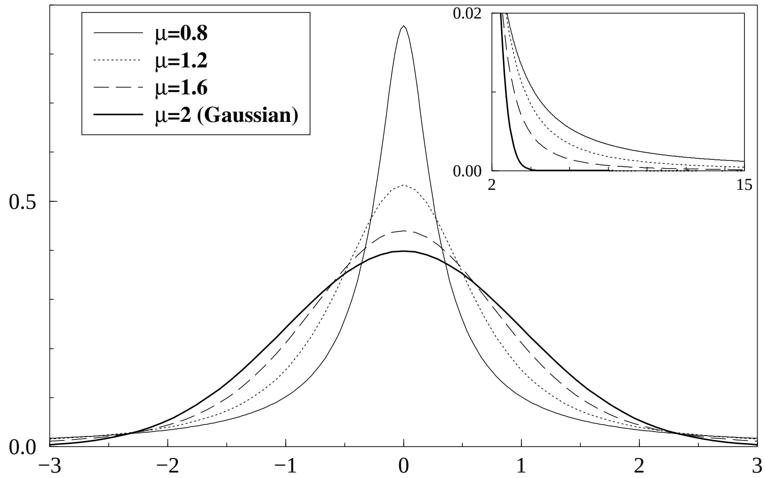

Figure 1.4: Shape of the symmetric Lévy distributions with $\mu = 0.8,\ 1.2,\ 1.6$ and 2 (this last value actually corresponds to a Gaussian). The smaller μ, the sharper the 'body' of the distribution, and the fatter the tails, as illustrated in the inset.

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 (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 generalisation 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 (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 (1.20) for $\alpha = 0$. Note that

[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).

the argument in the exponential can also be written as:

$$\frac{a_\mu}{2\cos(\pi\mu/2)}\left[(\alpha+iz)^\mu+(\alpha-iz)^\mu-2\alpha^\mu\right]. \tag{1.24}$$

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:

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

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

$$\mathcal{P}_>(x) = \frac{1}{\left(1+\frac{\alpha x}{\mu}\right)^\mu} \underset{\mu\to\infty}{\longrightarrow} \exp(-\alpha x). \tag{1.26}$$

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:

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

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

$$P_H(x) \equiv \frac{1}{2x_0K_1(\alpha x_0)}\exp-[\alpha\sqrt{x_0^2+x^2}], \tag{1.28}$$

  where the normalisation $K_1(\alpha x_0)$ is a modified Bessel function of the second kind. For x small compared to x_0, $P_H(x)$ behaves as a Gaussian while its asymptotic behaviour for $x \gg x_0$ is fatter and reads $\exp -\alpha|x|$.

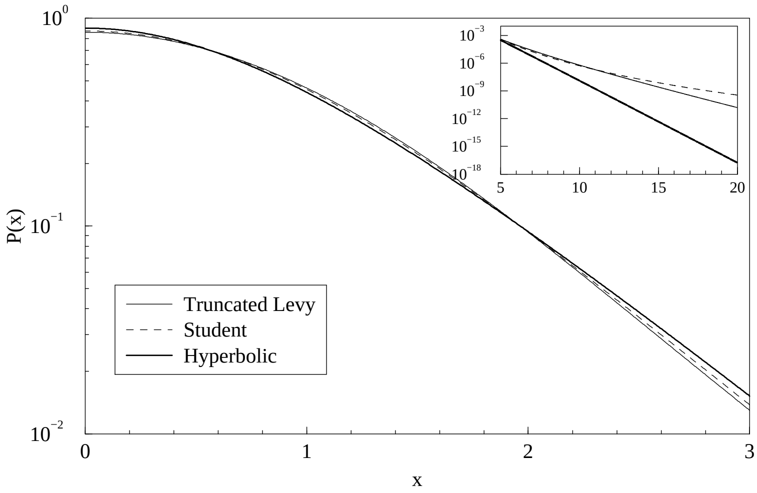

Figure 1.5: Probability density for the truncated Lévy ($\mu = 3/2$), Student and hyperbolic distributions. All three have two free parameters which were fixed to have unit variance and kurtosis. The inset shows a blow-up of the tails where one can see that the Student distribution has tails similar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

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).

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 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 characterise; this will justify the above simple criterion (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 $\mathcal{P}(x_{\max} < \Lambda_{\text{med}}) = 1/2$:

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

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 $p = 1/e \simeq 0.37$. 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 $x_{\max} = \Lambda_{\max} + \frac{u}{\alpha}$, one finds that the deviation u around $\Lambda_{\max}$ is distributed according to the Gumbel distribution:

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

The most probable value of this distribution is $u = 0$. This shows that $\Lambda_{\max}$ is the most probable value of $x_{\max}$. The result (1.40) is actually much more general, and is valid as soon as $P(x)$ decreases more rapidly

[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.

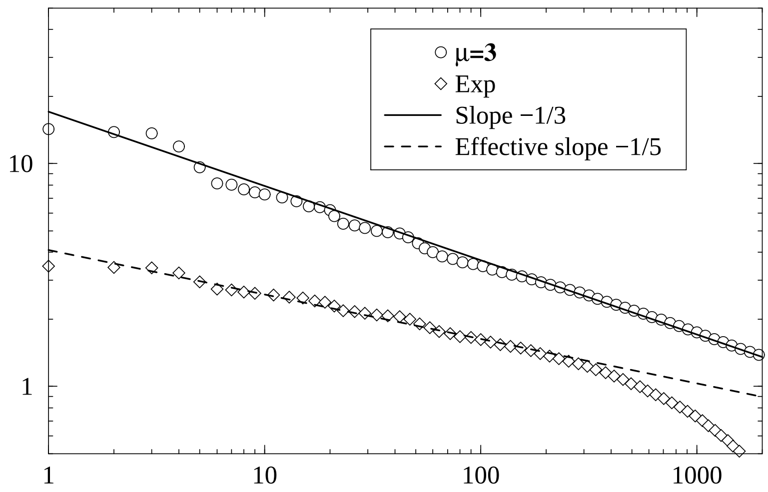

Figure 1.6: Amplitude versus rank plots. One plots the value of the n^{th} variable $\Lambda[n]$ as a function of its rank n. If $P(x)$ behaves asymptotically as a power-law, one obtains a straight line in log-log coordinates, with a slope equal to $-1/\mu$. For an exponential distribution, one observes an effective slope which is smaller and smaller as N/n tends to infinity. The points correspond to synthetic time series of length 5 000, drawn according to a power law with $\mu = 3$, or according to an exponential. Note that if the axis x and y are interchanged, then according to Eq. (1.45), one obtains an estimate of the cumulative distribution, $\mathcal{P}_>$.

than any power-law for $x \to \infty$: the deviation between $\Lambda_{\max}$ (defined as (1.34)) and $x_{\max}$ is always distributed according to the Gumbel law (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. (1.14). In this case,

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

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 $\mu = 3/2$ and a scale factor $A_+ = 1$, the largest of $N = 10\,000$ variables is on the order of 450, while for $\mu = 1/2$ 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.*[15] 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 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 n^{th} 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:

$$P_n(\Lambda[n]) = C_N^n \; P(x = \Lambda[n]) \; (\mathcal{P}(x > \Lambda[n])^{n-1}(\mathcal{P}(x < \Lambda[n])^{N-n}. \quad (1.44)$$

The previous expression means that one has to choose n variables among N as the n largest ones, 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 generalises (1.34):

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

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

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

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

[15] 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.

In the case of an exponential tail, one finds that $\Lambda[n]^* \simeq \log(\frac{N}{n})/\alpha$; while 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 organised: for example, the ratio of the largest value $x_{\max} \equiv \Lambda[1]$ to the second largest $\Lambda[2]$ is on 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 (1.47) is very useful to identify empirically the nature of the tails of a probability distribution. One sorts in decreasing order the set of observed values $\{x_1, x_2, .., x_N\}$ 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 (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 axis x and y of Fig. 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:

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

then, if $P(x)$ decreases as in (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.[16] *The average $\langle x \rangle_\Lambda$ is thus always of order of Λ itself, with a proportionality factor which diverges as $\mu \to 1$.*

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

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 five minutes 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 not always justified, see 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.

1.5.1 Convolutions

What is the distribution of the sum of two independent random variable? This sum can for example represent the variation of price of an asset between today and the day after tomorrow (X), which is the sum of the increment between today and tomorrow (X_1) and between tomorrow and the day after tomorrow (X_2), both assumed to be random and independent.

Let us thus consider $X = X_1 + X_2$ where X_1 and X_2 are two random variables, independent, and distributed according to $P_1(x_1)$ and $P_2(x_2)$, respectively. The probability that X is equal to x (within dx) is given by the sum over all possibilities of obtaining $X = x$ (that is all combinations of $X_1 = x_1$ and $X_2 = x_2$ such that $x_1 + x_2 = x$), weighted by their respective probabilities. The variables X_1 and X_2 being independent, the joint probability that $X_1 = x_1$ and $X_2 = x - x_1$ is equal to $P_1(x_1)P_2(x - x_1)$, from which one obtains:

$$P(x, N = 2) = \int dx' P_1(x') P_2(x - x'). \tag{1.50}$$

This equation defines the convolution between $P_1(x)$ and $P_2(x)$, which we shall write $P = P_1 \star P_2$. The generalisation to the sum of N independent random variables is immediate. If $X = X_1 + X_2 + ... + X_N$ with X_i distributed according to $P_i(x_i)$, the distribution of X is obtained as:

$$P(x, N) = \int \prod_{i=1}^{N-1} dx'_i P_1(x'_1)...P_{N-1}(x'_{N-1}) P_N(x - x'_1 - ... - x'_{N-1}). \quad (1.51)$$

One thus understands how powerful is the hypothesis that the increments are IID, i.e., that $P_1 = P_2 = .. = P_N$. Indeed, according to this hypothesis, one only needs to know the distribution of increments over a unit time interval to reconstruct that of increments over an interval of length N: it is simply obtained by convoluting the elementary distribution N times with itself.

The analytical or numerical manipulations of Eqs. (1.50) and (1.51) are much eased by the use of Fourier transforms, for which convolutions become simple products. The equation $P(x, N = 2) = [P_1 \star P_2](x)$, reads in Fourier space:

$$\hat{P}(z, N = 2) = \int dx e^{iz(x - x' + x')} \int dx' P_1(x') P_2(x - x') \equiv \hat{P}_1(z) \hat{P}_2(z). \quad (1.52)$$

In order to obtain the N^{th} convolution of a function with itself, one should raise its characteristic function to the power N, and then take its inverse Fourier transform.

1.5.2 *Additivity of cumulants and of tail amplitudes*

It is clear that the mean of the sum of two random variables (independent or not) is equal to the sum of the individual means. The mean is thus additive under convolution. Similarly, if the random variables are independent, one can show that their variances (if they are well defined) also add simply. More generally, all the cumulants (c_n) of two independent distributions simply add. This follows from the fact that since the characteristic functions multiply, their logarithm add. The additivity of cumulants is then a simple consequence of the linearity of derivation.

The cumulants of a given law convoluted N times with itself thus follow the simple rule $c_{n,N} = N c_{n,1}$ where the $\{c_{n,1}\}$ are the cumulants of the elementary distribution P_1. Since the cumulant c_n has the dimension of X to the power n, its relative importance is best measured in terms of the normalised cumulants:

$$\lambda_n^N \equiv \frac{c_{n,N}}{(c_{2,N})^{\frac{n}{2}}} = \frac{c_{n,1}}{c_{2,1}} N^{1-n/2}. \quad (1.53)$$

The normalised cumulants thus decay with N for $n > 2$; the higher the cumulant, the faster the decay: $\lambda_n^N \propto N^{1-n/2}$. The kurtosis κ, defined above as the fourth normalised cumulant, thus decreases as $1/N$. This is basically the content of the CLT: when N is very large, the cumulants of order > 2 become negligible. Therefore, the distribution of the sum is only characterised by its first two cumulants (mean and variance): it is a Gaussian.

Let us now turn to the case where the elementary distribution $P_1(x_1)$ decreases as a power law for large arguments x_1 (cf. (1.14)), with a certain exponent μ. The cumulants of order higher than μ are thus divergent. By studying the small z singular expansion of the Fourier transform of $P(x, N)$, one finds that the above additivity property of cumulants is bequeathed to the tail amplitudes $A_{\pm}^{\mu}$: the asymptotic behaviour of the distribution of the sum $P(x, N)$ *still behaves as a power-law* (which is thus conserved by addition for all values of μ, provided one takes the limit $x \to \infty$ *before* $N \to \infty$ – see the discussion in 1.6.3), with a tail amplitude given by:

$$A_{\pm,N}^{\mu} \equiv N A_{\pm}^{\mu}. \tag{1.54}$$

The tail parameter thus play the role, for power-law variables, of a generalised cumulant.

1.5.3 Stable distributions and self-similarity

If one adds random variables distributed according to an arbitrary law $P_1(x_1)$, one constructs a random variable which has, in general, a different probability distribution ($P(x, N) = [P_1(x_1)]^{\star N}$). However, for certain special distributions, the law of the sum has exactly the same shape as the elementary distribution – these are called *stable* laws. The fact that two distributions have the 'same shape' means that one can find a (N dependent) translation and dilation of x such that the two laws coincide:

$$P(x, N)dx = P_1(x_1)dx_1 \quad \text{where } x = a_N x_1 + b_N. \tag{1.55}$$

The distribution of increments on a certain time scale (week, month, year) is thus *scale invariant*, provided the variable X is properly rescaled. In this case, the chart giving the evolution of the price of a financial asset as a function of time has the same statistical structure, independently of the chosen elementary time scale – only the average slope and the amplitude of the fluctuations are different. These charts are then called *self-similar*, or, using a better terminology introduced by Mandelbrot, *self-affine* (Figs. 1.7 and 1.8).

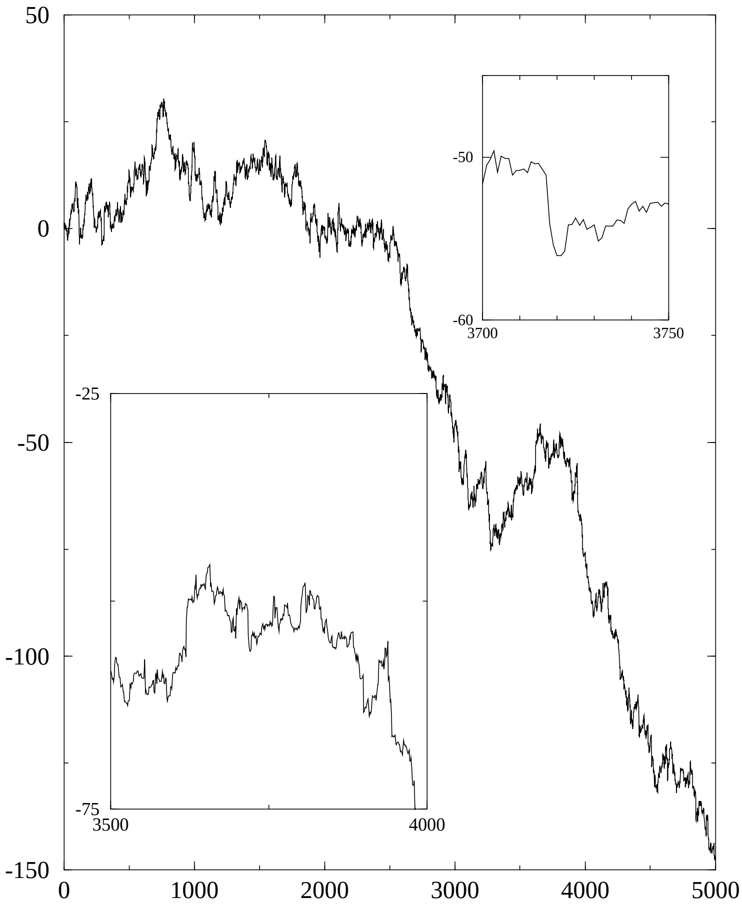

Figure 1.7: Example of a self-affine function, obtained by summing random variables. One plots the sum x as a function of the number of terms N in the sum, for a Gaussian elementary distribution $P_1(x_1)$. Several successive 'zooms' reveal the self similar nature of the function, here with $a_N = N^{1/2}$.

The family of all possible stable laws coincide (for continuous variables) with the Lévy distributions defined above,[17] which include Gaussians as the special case $\mu = 2$. This is easily seen in Fourier space, using the explicit shape of the characteristic function of the Lévy distributions. We shall specialise here for simplicity to the case of symmetric distributions $P_1(x_1) = P_1(-x_1)$, for which the translation factor is zero ($b_N \equiv 0$). The scale parameter is then given by $a_N = N^{\frac{1}{\mu}}$,[18] and one finds, for $\mu < 2$:

$$\langle |x|^q \rangle^{\frac{1}{q}} \propto A N^{\frac{1}{\mu}} \qquad q < \mu \tag{1.56}$$

where $A = A_+ = A_-$. In words, the above equation means that the order of magnitude of the fluctuations on 'time' scale N is a factor $N^{\frac{1}{\mu}}$ larger than the fluctuations on the elementary time scale. However, once this factor is taken into account, the probability distributions are identical. One should notice the smaller the value of μ, the faster the growth of fluctuations with time.

1.6 Central limit theorem

We have thus seen that the stable laws (Gaussian and Lévy distributions) are 'fixed points' of the convolution operation. These fixed points are actually also *attractors*, in the sense that any distribution convoluted with itself a large number of times finally converges towards a stable law (apart from some very pathological cases). Said differently, the limit distribution of the sum of a large number of random variables is a stable law. The precise formulation of this result is known as the *central limit theorem* (CLT).

1.6.1 Convergence to a Gaussian

The classical formulation of the CLT deals with sums of IID random variables of *finite variance* σ^2 towards a Gaussian. In a more precise way, the result is then the following:

$$\lim_{N\to\infty} \mathcal{P}\left(u_1 \le \frac{x - mN}{\sigma\sqrt{N}} \le u_2\right) = \int_{u_1}^{u_2} \frac{du}{\sqrt{2\pi}} e^{-u^2/2}, \tag{1.57}$$

for all *finite* u_1, u_2. Note however that for finite N, the distribution of the sum $X = X_1 + ... + X_N$ in the tails (corresponding to extreme events) can be very different from the Gaussian prediction; but the weight of these non-Gaussian regions tends to zero when N goes to infinity. The

[17] For discrete variables, one should also add the Poisson distribution (1.27).
[18] The case $\mu = 1$ is special and involves extra logarithmic factors.

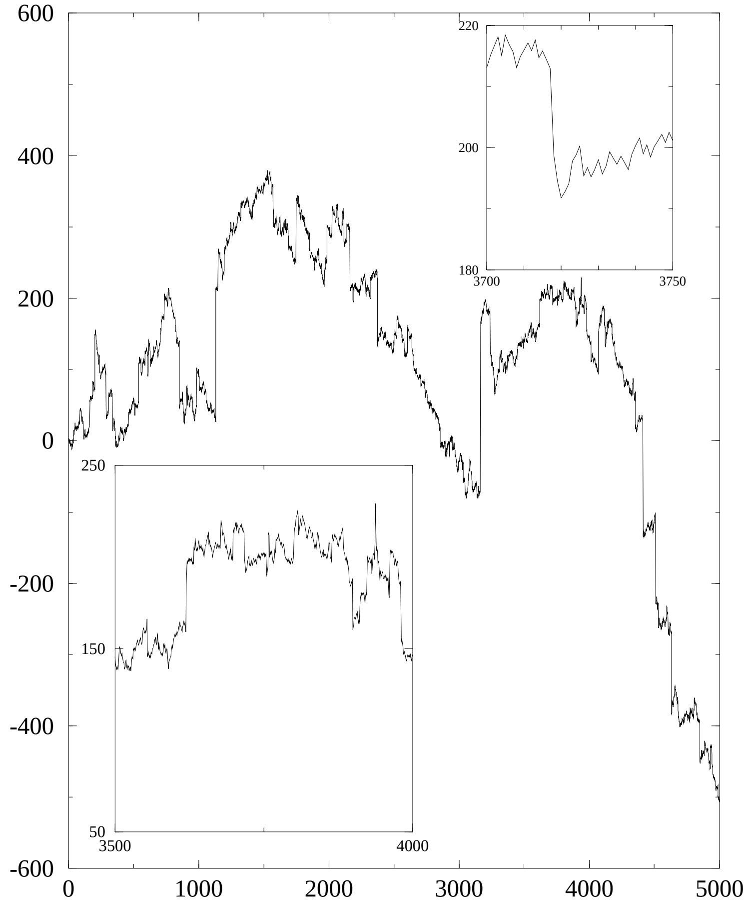

Figure 1.8: In this case, the elementary distribution $P_1(x_1)$ decreases as a power-law with an exponent $\mu = 1.5$. The scale factor is now given by $a_N = N^{2/3}$. Note that, contrarily to the previous graph, one clearly observes the presence of sudden 'jumps', which reflect the existence of very large values of the elementary increment x_1.

CLT only concerns the *central* region, which keeps a finite weight for N large: we shall come back in detail to this point below.

The main hypotheses insuring the validity of the Gaussian CLT are the following:

- The X_i must be independent random variables, or at least not 'too' correlated (the correlation function $\langle x_i x_j \rangle - m^2$ must decay sufficiently fast when $|i - j|$ becomes large – see 1.7.1 below). For example, in the extreme case where all the X_i are perfectly correlated (i.e. they are all equal), the distribution of X is obviously the same as that of the individual X_i (once the factor N has been properly taken into account).

- The random variables X_i need not necessarily be identically distributed. One must however require that the variance of all these distributions are not too dissimilar, so that no one of the variances dominates over all the others (as would be the case, for example, if the variances were themselves distributed as a power-law with an exponent $\mu < 1$). In this case, the variance of the Gaussian limit distribution is the average of the individual variances. This also allows one to deal with sums of the type $X = p_1 X_1 + p_2 X_2 + ... + p_N X_N$, where the p_i are arbitrary coefficients; this case is relevant in many circumstances, in particular in the Portfolio theory (cf. Chapter 3).

- Formally, the CLT only applies in the limit where N is infinite. In practice, N must be large enough for a Gaussian to be a good approximation of the distribution of the sum. The minimum required value of N (called N^* below) depends on the elementary distribution $P_1(x_1)$ and its distance from a Gaussian. Also, N^* depends on how far in the tails one requires a Gaussian to be a good approximation, which takes us to the next point.

- As mentioned above, the CLT does not tell us anything about the tails of the distribution of X; only the central part of the distribution is well described by a Gaussian. The 'central' region means a region of width at least on the order of $\sqrt{N}\sigma$ around the mean value of X. The actual width of the region where the Gaussian turns out to be a good approximation for large finite N crucially depends on the elementary distribution $P_1(x_1)$. This problem will be explored in Section 1.6.3. Roughly speaking, this region is of width $\sim N^{3/4}\sigma$ for 'narrow' symmetric elementary distributions, such that all even moments are finite. This region is however sometimes of much smaller extension: for example, if $P_1(x_1)$ has power-law tails with $\mu > 2$ (such that σ is finite), the Gaussian 'realm' grows barely faster than $\sqrt{N}$ (as $\sim \sqrt{N \log N}$).

The above formulation of the CLT *requires the existence of a finite variance. This condition can be somewhat weakened to include some 'marginal' distributions such as a power-law with $\mu = 2$. In this case the scale factor is not $a_N = \sqrt{N}$ but rather $a_N = \sqrt{N \ln N}$. However, as we shall discuss in the next section, elementary distributions which decay more slowly than $|x|^{-3}$ do not belong the the Gaussian basin of attraction. More precisely, the necessary and sufficient condition for $P_1(x_1)$ to belong to this basin is that:*

$$\lim_{u\to\infty} u^2 \frac{\mathcal{P}_{1<}(-u) + \mathcal{P}_{1>}(u)}{\int_{|u'|<u} du'\ u'^2 P_1(u')} = 0. \tag{1.58}$$

This condition is always satisfied if the variance is finite, but allows one to include the marginal cases such as a power-law with $\mu = 2$.

The central limit theorem and information theory

It is interesting to notice that the Gaussian is the law of maximum entropy – or minimum information – such that its variance is fixed. The missing information quantity $\mathcal{I}$ (or entropy) associated with a probability distribution P is defined as:

$$\mathcal{I}[P] \equiv -\int dx\ P(x)\ \log\left[\frac{P(x)}{e}\right]. \tag{1.59}$$

The distribution maximising $\mathcal{I}[P]$ for a given value of the variance is obtained by taking a functional derivative with respect to $P(x)$:

$$\frac{\partial}{\partial P(x)}\left[\mathcal{I}[P] - \zeta \int dx'\ x'^2 P(x') - \zeta' \int dx'\ P(x')\right] = 0, \tag{1.60}$$

where ζ is fixed by the condition $\int dx\ x^2 P(x) = \sigma^2$ and ζ' by the normalisation of $P(x)$. It is immediate to show that the solution to (1.60) is indeed the Gaussian. The numerical value of its entropy is:

$$\mathcal{I}_G = \frac{3}{2} + \frac{1}{2}\log(2\pi) + \log(\sigma) \simeq 2.419 + \log(\sigma). \tag{1.61}$$

For comparison, one can compute the entropy of the symmetric exponential distribution, which is:

$$\mathcal{I}_E = 2 + \frac{\log 2}{2} + \log(\sigma) \simeq 2.346 + \log(\sigma). \tag{1.62}$$

It is important to understand that the convolution operation is 'information burning', since all the details of the elementary distribution $P_1(x_1)$ progressively disappear while the Gaussian distribution emerges.

1.6.2 Convergence to a Lévy distribution

Let us now turn to the case of the sum of a large number N of IID random variables, asymptotically distributed as a power-law with $\mu < 2$, and with a tail amplitude $A^\mu = A_+^\mu = A_-^\mu$ (cf. (1.14)). The variance of the distribution is thus infinite. The limit distribution for large N is then a stable Lévy distribution of exponent μ and with a tail amplitude NA^μ. If the positive and negative tails of the elementary distribution $P_1(x_1)$ are characterised by different amplitudes (A_-^μ and A_+^μ) one then obtains an asymmetric Lévy distribution with parameter $\beta = (A_+^\mu - A_-^\mu)/(A_+^\mu + A_-^\mu)$. If the 'left' exponent is different from the 'right' exponent ($\mu_- \neq \mu_+$), then the smallest of the two wins and one finally obtains a totally asymmetric Lévy distribution ($\beta = -1$ or $\beta = 1$) with exponent $\mu = \min(\mu_-, \mu_+)$. The CLT generalised to Lévy distributions applies with the same precautions as in the Gaussian case above.

Technically, a distribution $P_1(x_1)$ belongs to the attraction basin of the Lévy distribution $L_{\mu,\beta}$ if and only if:

$$\lim_{u\to\infty} \frac{\mathcal{P}_{1<}(-u)}{\mathcal{P}_{1>}(u)} = \frac{1-\beta}{1+\beta}; \tag{1.63}$$

and for all r,

$$\lim_{u\to\infty} \frac{\mathcal{P}_{1<}(-u) + \mathcal{P}_{1>}(u)}{\mathcal{P}_{1<}(-ru) + \mathcal{P}_{1>}(ru)} = r^\mu. \tag{1.64}$$

A distribution with an asymptotic tail given by (1.14) is such that,

$$\mathcal{P}_{1<}(u) \underset{u\to-\infty}{\simeq} \frac{A_-^\mu}{|u|^\mu} \text{ and } \mathcal{P}_{1>}(u) \underset{u\to\infty}{\simeq} \frac{A_+^\mu}{u^\mu}, \tag{1.65}$$

and thus belongs to the attraction basin of the Lévy distribution of exponent μ and asymmetry parameter $\beta = (A_+^\mu - A_-^\mu)/(A_+^\mu + A_-^\mu)$.

1.6.3 Large deviations

The CLT teaches us that the Gaussian approximation is justified to describe the 'central' part of the distribution of the sum of a large number of random variables (of finite variance). However, the definition of the *centre* has remained rather vague up to now. The CLT only states that the probability of finding an event in the *tails* goes to zero for large N. In the present section, we characterise more precisely the region where the Gaussian approximation is valid.

If X is the sum of N IID random variables of mean m and variance σ^2, one defines a 'rescaled variable' U as:

$$U = \frac{X - Nm}{\sigma\sqrt{N}}, \tag{1.66}$$

which according to the CLT tends towards a Gaussian variable of zero mean and unit variance. Hence, for any *fixed* u, one has:

$$\lim_{N\to\infty} \mathcal{P}_>(u) = \mathcal{P}_{G>}(u), \tag{1.67}$$

where $\mathcal{P}_{G>}(u)$ is the related to the error function, and describes the weight contained in the tails of the Gaussian:

$$\mathcal{P}_{G>}(u) = \int_u^\infty \frac{du'}{\sqrt{2\pi}} \exp(-u^2/2) = \tfrac{1}{2}\mathrm{erfc}\left(\frac{u}{\sqrt{2}}\right). \tag{1.68}$$

However, the above convergence is *not uniform*. The value of N such that the approximation $\mathcal{P}_>(u) \simeq \mathcal{P}_{G>}(u)$ becomes valid depends on u. Conversely, for fixed N, this approximation is only valid for u not too large: $|u| \ll u_0(N)$.

One can estimate $u_0(N)$ in the case where the elementary distribution $P_1(x_1)$ is 'narrow', that is, decreasing faster than any power-law when $|x_1| \to \infty$, such that all the moments are finite. In this case, all the cumulants of P_1 are finite and one can obtain a systematic expansion in powers of $N^{-1/2}$ of the difference $\Delta\mathcal{P}_>(u) \equiv \mathcal{P}_>(u) - \mathcal{P}_{G>}(u)$,

$$\Delta\mathcal{P}_>(u) \simeq \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \left(\frac{Q_1(u)}{N^{1/2}} + \frac{Q_2(u)}{N} + \ldots + \frac{Q_k(u)}{N^{k/2}} + \ldots\right), \tag{1.69}$$

where the $Q_k(u)$ are polynomials functions which can be expressed in terms of the normalised cumulants λ_n (cf. (1.12)) of the elementary distribution. More explicitely, the first two terms are given by:

$$Q_1(u) = \tfrac{1}{6}\lambda_3(u^2 - 1), \tag{1.70}$$

and

$$Q_2(u) = \tfrac{1}{72}\lambda_3^2 u^5 + \tfrac{1}{8}(\tfrac{1}{3}\lambda_4 - \tfrac{10}{9}\lambda_3^2)u^3 + (\tfrac{5}{24}\lambda_3^2 - \tfrac{1}{8}\lambda_4)u. \tag{1.71}$$

One recovers the fact that if all the cumulants of $P_1(x_1)$ of order larger than two are zero, all the Q_k are also identically zero and so is the difference between $P(x, N)$ and the Gaussian.

For a general asymmetric elementary distribution P_1, λ_3 is non zero. The leading term in the above expansion when N is large is thus $Q_1(u)$. For the Gaussian approximation to be meaningful, one must at least require that this term is small in the central region where u is of order one, which corresponds to $x - mN \sim \sigma\sqrt{N}$. This thus imposes that $N \gg N^* = \lambda_3^2$. The Gaussian approximation remains valid whenever the relative error is small compared to 1. For large u (which will be justified for large N), the relative error is obtained by dividing Eq. (1.69) by $\mathcal{P}_{G>}(u) \simeq \exp(-u^2/2)/(u\sqrt{2\pi})$. One then obtains the following condition:[19]

$$\lambda_3 u^3 \ll N^{1/2} \text{ i.e. } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/6}. \qquad (1.72)$$

This shows that the central region has an extension growing as $N^{\frac{2}{3}}$.

A symmetric elementary distribution is such that $\lambda_3 \equiv 0$; it is then the kurtosis $\kappa = \lambda_4$ that fixes the first correction to the Gaussian when N is large, and thus the extension of the central region. The conditions now read: $N \gg N^* = \lambda_4$ and

$$\lambda_4 u^4 \ll N \text{ i.e. } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/4}. \qquad (1.73)$$

The central region now extends over a region of width $N^{3/4}$.

The results of the present section do not directly apply if the elementary distribution $P_1(x_1)$ decreases as a power-law ('broad distribution'). In this case, some of the cumulants are infinite and the above cumulant expansion (1.69) is meaningless. In the next section, we shall see that in this case the 'central' region is much more restricted than in the case of 'narrow' distributions. We shall then describe in Section 1.6.5, the case of 'truncated' power-law distributions, where the above conditions become asymptotically relevant. These laws however may have a very large kurtosis, which depends on the point where the truncation becomes noticeable, and the above condition $N \gg \lambda_4$ can be hard to satisfy.

Cramèr function

More generally, when N is large, one can write the distribution of the sum of N IID random variables as:[20]

$$P(x, N) \underset{N\to\infty}{\simeq} \exp\left[-NS\left(\frac{x}{N}\right)\right], \qquad (1.74)$$

where S is the so-called Cramèr function, which gives some information about the probability of X even outside the 'central' region. When the variance is finite, S grows as $S(u) \propto u^2$ for small u's, which again leads to a Gaussian central region. For finite u, S can be computed using Laplace's saddle point method, valid for N large. By definition:

$$P(x, N) = \int \frac{dz}{2\pi} \exp N\left(-iz\frac{x}{N} + \log[\hat{P}_1(z)]\right). \qquad (1.75)$$

[19] The above arguments can actually be made fully rigourous, see [Feller].

[20] We assume that their mean is zero, which can always be achieved through a suitable shift of x_1.

When N is large, the above integral is dominated by the neighbourhood of the point z^* where the term in the exponential is stationary. The results can be written as:

$$P(x,N) \simeq \exp\left[-NS\left(\frac{x}{N}\right)\right], \tag{1.76}$$

with $S(u)$ given by:

$$\left.\frac{d\log[\hat{P}_1(z)]}{dz}\right|_{z=z^*} = iu \qquad S(u) = -iz^*u + \log[\hat{P}_1(z^*)], \tag{1.77}$$

which, in principle, allows one to estimate $P(x,N)$ even outside the central region. Note that if $S(u)$ is finite for finite u, the corresponding probability is exponentially small in N.

1.6.4 The CLT at work on a simple case

It is helpful to give some flesh to the above general statements, by working out explicitly the convergence towards the Gaussian in two exactly soluble cases. On these examples, one clearly sees the domain of validity of the CLT as well as its limitations.

Let us first study the case of positive random variables distributed according to the exponential distribution:

$$P_1(x) = \Theta(x_1)\alpha e^{-\alpha x_1}, \tag{1.78}$$

where $\Theta(x_1)$ is the function equal to 1 for $x_1 \geq 0$ and to 0 otherwise. A simple computation shows that the above distribution is correctly normalised, has a mean given by $m = \alpha^{-1}$ and a variance given by $\sigma^2 = \alpha^{-2}$. Furthermore, the exponential distribution is asymmetrical; its skewness is given by $c_3 = \langle (x-m)^3 \rangle = 2\alpha^{-3}$, or $\lambda_3 = 2$.

The sum of N such variables is distributed according to the N^{th} convolution of the exponential distribution. According to the CLT this distribution should approach a Gaussian of mean mN and of variance $N\sigma^2$. The N^{th} convolution of the exponential distribution can be computed exactly. The result is:[21]

$$P(x,N) = \Theta(x)\alpha^N \frac{x^{N-1}e^{-\alpha x}}{(N-1)!}, \tag{1.79}$$

which is called a 'Gamma' distribution of index N. At first sight, this distribution does not look very much like a Gaussian! For example, its asymptotic behaviour is very far from that of a Gaussian: the 'left' side

[21] This result can be shown by induction using the definition (1.50).

is strictly zero for negative x, while the 'right' tail is exponential, and thus much fatter than the Gaussian. It is thus very clear that the CLT does not apply for values of x too far from the mean value. However, the central region around $Nm = N\alpha^{-1}$ is well described by a Gaussian. The most probable value (x^*) is defined as:

$$\left. \frac{d}{dx} x^{N-1} e^{-\alpha x} \right|_{x^*} = 0, \tag{1.80}$$

or $x^* = (N-1)m$. An expansion in $x - x^*$ of $P(x, N)$ then gives us:

$$\begin{aligned} \log P(x, N) &= -K(N-1) - \log m - \frac{\alpha^2 (x - x^*)^2}{2(N-1)} \\ &\quad + \frac{\alpha^3 (x - x^*)^3}{3(N-1)^2} + O(x - x^*)^4, \end{aligned} \tag{1.81}$$

where

$$K(N) \equiv \log N! + N - N \log N \underset{N\to\infty}{\simeq} \tfrac{1}{2} \log(2\pi N). \tag{1.82}$$

Hence, to second order in $x - x^*$, $P(x, N)$ is given by a Gaussian of mean $(N-1)m$ and variance $(N-1)\sigma^2$. The relative difference between N and $N-1$ goes to zero for large N. Hence, for the Gaussian approximation to be valid, one requires not only that N be large compared to one, but also that the higher order terms in $(x - x^*)$ be negligible. The cubic correction is small compared to 1 as long as $\alpha|x - x^*| \ll N^{2/3}$, in agreement with the above general statement (1.72) for an elementary distribution with a non zero third cumulant. Note also that for $x \to \infty$, the exponential behaviour of the Gamma function coincides (up to subleading terms in x^{N-1}) with the asymptotic behaviour of the elementary distribution $P_1(x_1)$.

Another very instructive example is provided by a distribution which behaves as a power-law for large arguments, but at the same time has a finite variance to ensure the validity of the CLT. Consider the following explicit example of a Student distribution with $\mu = 3$:

$$P_1(x_1) = \frac{2a^3}{\pi(x_1^2 + a^2)^2}, \tag{1.83}$$

where a is a positive constant. This symmetric distribution behaves as a power-law with $\mu = 3$ (cf. (1.14)); all its cumulants of order larger or equal to three are infinite. However, its variance is finite and equal to a^2.

It is useful to compute the characteristic function of this distribution,

$$\hat{P}_1(z) = (1 + a|z|)e^{-a|z|}, \tag{1.84}$$

and the first terms of its small z expansion, which read:

$$\hat{P}_1(z) \simeq 1 - \frac{z^2a^2}{2} + \frac{|z|^3a^3}{3} + O(z^4). \tag{1.85}$$

The first singular term in this expansion is thus $|z|^3$, as expected from the asymptotic behaviour of $P_1(x_1)$ in x_1^{-4}, and the divergence of the moments of order larger than three.

The N^{th} convolution of $P_1(x_1)$ thus has the following characteristic function:

$$\hat{P}_1^{\,N}(z) = (1 + a|z|)^N e^{-aN|z|}, \tag{1.86}$$

which, expanded around $z = 0$, gives:

$$\hat{P}_1^{\,N}(k) \simeq 1 - \frac{Nz^2a^2}{2} + \frac{N|z|^3a^3}{3} + O(z^4). \tag{1.87}$$

Note that the $|z|^3$ singularity (which signals the divergence of the moments m_n for $n \geq 3$) does not disappear under convolution, even if at the same time $P(x, N)$ converges towards the Gaussian. The resolution of this apparent paradox is again that the convergence towards the Gaussian only concerns the centre of the distribution, while the tail in x^{-4} survives for ever (as was mentioned in Section 1.5.3).

As follows from the CLT, the centre of $P(x, N)$ is well approximated, for N large, by a Gaussian of zero mean and variance Na^2:

$$P(x, N) \simeq \frac{1}{\sqrt{2\pi N}a} \exp\left(-\frac{x^2}{2Na^2}\right). \tag{1.88}$$

On the other hand, since the power-law behaviour is conserved upon addition and that the tail amplitudes simply add (cf. (1.14)), one also has, for large x's:

$$P(x, N) \underset{x\to\infty}{\simeq} \frac{2Na^3}{\pi x^4}. \tag{1.89}$$

The above two expressions (1.88) and (1.89) are not incompatible, since these describe two very different regions of the distribution $P(x, N)$. For fixed N, there is a characteristic value $x_0(N)$ beyond which the Gaussian approximation for $P(x, N)$ is no longer accurate, and the distribution is described by its asymptotic power-law regime. The order of magnitude of $x_0(N)$ is fixed by looking at the point where the two regimes match to one another:

$$\frac{1}{\sqrt{2\pi N}a} \exp\left(-\frac{x_0^2}{2Na^2}\right) \simeq \frac{2Na^3}{\pi x_0^4}. \tag{1.90}$$

One thus find,

$$x_0(N) \simeq a\sqrt{N \log N}, \tag{1.91}$$

(neglecting subleading corrections for large N).

This means that the rescaled variable $U = X/(a\sqrt{N})$ becomes for large N a Gaussian variable of unit variance, but this description ceases to be valid as soon as $u \sim \sqrt{\log N}$, which grows very slowly with N. For example, for N equal to a million, the Gaussian approximation is only acceptable for fluctuations of u of less than three or four RMS!

Finally, the CLT states that the weight of the regions where $P(x, N)$ substantially differs from the Gaussian goes to zero when N becomes large. For our example, one finds that the probability that X falls in the tail region rather than in the central region is given by:

$$\mathcal{P}_<(x_0) + \mathcal{P}_>(x_0) \simeq 2\int_{a\sqrt{N\log N}}^{\infty} \frac{2a^3 N}{\pi x^4} dx \propto \frac{1}{\sqrt{N}\log^{3/2} N}, \tag{1.92}$$

which indeed goes to zero for large N.

The above arguments are not special to the case $\mu = 3$ and in fact apply more generally, as long as $\mu > 2$, i.e. when the variance is finite. In the general case, one finds that the CLT is valid in the region $|x| \ll x_0 \propto \sqrt{N \log N}$, and that the weight of the non Gaussian tails is given by:

$$\mathcal{P}_<(x_0) + \mathcal{P}_>(x_0) \propto \frac{1}{N^{\mu/2-1}\log^{\mu/2} N}, \tag{1.93}$$

which tends to zero for large N. However, one should notice that as μ approaches the 'dangerous' value $\mu = 2$, the weight of the tails becomes more and more important. For $\mu < 2$, the whole argument collapses since the weight of the tails would grow with N. In this case, however, the convergence is no longer towards the Gaussian, but towards the Lévy distribution of exponent μ.

1.6.5 Truncated Lévy distributions

An interesting case is when the elementary distribution $P_1(x_1)$ is a truncated Lévy distribution (TLD) as defined in Section 1.3.3. The first cumulants of the distribution defined by Eq. (1.23) read, for $1 < \mu < 2$:

$$c_2 = \mu(\mu-1)\frac{a_\mu}{|\cos \pi\mu/2|}\alpha^{\mu-2} \qquad c_3 = 0. \tag{1.94}$$

The kurtosis $\kappa = \lambda_4 = c_4/c_2^2$ is given by:

$$\lambda_4 = \frac{(3-\mu)(2-\mu)|\cos \pi\mu/2|}{\mu(\mu-1)a_\mu \alpha^\mu}. \tag{1.95}$$

Note that the case $\mu = 2$ corresponds to the Gaussian, for which $\lambda_4 = 0$ as expected. On the other hand, when $\alpha \to 0$, one recovers a pure Lévy distribution, for which c_2 and c_4 are formally infinite. Finally, if $\alpha \to \infty$ with $a_\mu \alpha^{\mu-2}$ fixed, one also recovers the Gaussian.

If one considers the sum of N random variables distributed according to a TLD, the condition for the CLT to be valid reads (for $\mu < 2$):[22]

$$N \gg N^* = \lambda_4 \Longrightarrow (N a_\mu)^{\frac{1}{\mu}} \gg \alpha^{-1}. \tag{1.96}$$

This condition has a very simple intuitive meaning. A TLD behaves very much like a pure Lévy distribution as long as $x \ll \alpha^{-1}$. In particular, it behaves as a power-law of exponent μ and tail amplitude $A^\mu \propto a_\mu$ in the region where x is large but still much smaller than α^{-1} (we thus also assume that α is very small). If N is not too large, most values of x fall in the Lévy-like region. The largest value of x encountered is thus of order $x_{\max} \simeq AN^{\frac{1}{\mu}}$ (cf. 1.42). If $x_{\max}$ is very small compared to α^{-1}, it is consistent to forget the exponential cut-off and think of the elementary distribution as a pure Lévy distribution. One thus observe a first regime in N where the typical value of X grows as $N^{\frac{1}{\mu}}$, as if α was zero.[23] However, as illustrated in Fig. 1.9, this regime ends when $x_{\max}$ reaches the cut-off value α^{-1}: this happens precisely when N is on the order of N^* defined above. For $N > N^*$, the variable X progressively converges towards a Gaussian variable of width $\sqrt{N}$, at least in the region where $|x| \ll \sigma N^{3/4}/N^{*1/4}$. The typical amplitude of X thus behaves (as a function of N) as sketched in Fig. 1.9. Notice that the asymptotic part of the distribution of X (outside the central region) decays as an exponential for all values of N.

1.6.6 Conclusion: survival and vanishing of tails

The CLT thus teaches us that if the number of terms in a sum is large, the sum becomes (nearly) a Gaussian variable. This sum can represent the temporal aggregation of the daily fluctuations of a financial asset, or the aggregation, in a portfolio, of different stocks. The Gaussian (or non-Gaussian) nature of this sum is thus of crucial importance for risk control, since the extreme tails of the distribution correspond to the most 'dangerous' fluctuations. As we have discussed above, fluctuations

[22] One can see by inspection that the other conditions, concerning higher order cumulants, and which read $N^{k-1}\lambda_{2k} \gg 1$, are actually equivalent to the one written here.

[23] Note however that the variance of X grows like N for all N. However, the variance is dominated by the cut-off and, in the region $N \ll N^*$, grossly overestimates the typical values of X – see Section 2.3.2.

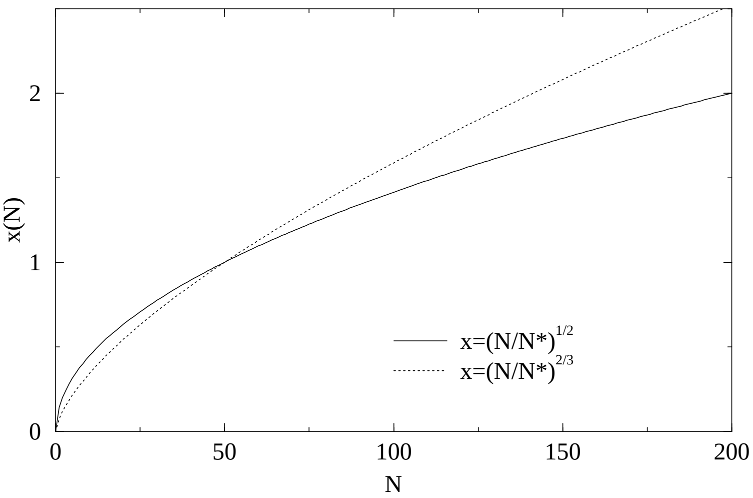

Figure 1.9: Behaviour of the typical value of X as a function of N for TLD variables. When $N \ll N^*$, x grows $N^{\frac{1}{\mu}}$ (dotted line). When $N \sim N^*$, x reaches the value α^{-1} and the exponential cut-off starts being relevant. When $N \gg N^*$, the behaviour predicted by the CLT sets in, and one recovers $x \propto \sqrt{N}$ (plain line).

are never Gaussian in the far-tails: one can explicitly show that if the elementary distribution decays as a power-law (or as an exponential, which formally corresponds to $\mu = \infty$), the distribution of the sum decays in the very same manner outside the central region, i.e. much more slowly than the Gaussian. The CLT simply ensures that these tail regions are expelled more and more towards large values of X when N grows, and their associated probability is smaller and smaller. When confronted to a concrete problem, one must decide whether N is large enough to be satisfied with a Gaussian description of the risks. In particular, if N is less than the characteristic value N^* defined above, the Gaussian approximation is very bad.

1.7 Correlations, dependence and non-stationary models (*)

We have assumed up to now that the random variables where *independent* and *identically distributed*. Although the general case cannot be discussed as thoroughly as the IID case, it is useful to illustrate how the CLT must be modified on a few examples, some of which being particularly relevant in the context of financial time series.

1.7.1 Correlations

Let us assume that the correlation function $C_{i,j}$ (defined as $\langle x_i x_j \rangle - m^2$) of the random variables is non zero for $i \neq j$. We also assume that the process is *stationary*, i.e. that $C_{i,j}$ only depends on $|i-j|$: $C_{i,j} = C(|i-j|)$, with $C(\infty) = 0$. The variance of the sum can be expressed in terms of the matrix C as:[24]

$$\langle x^2 \rangle = \sum_{i,j=1}^{N} C_{i,j} = N\sigma^2 + 2N \sum_{\ell=1}^{N} (1 - \frac{\ell}{N}) C(\ell) \tag{1.97}$$

where $\sigma^2 \equiv C(0)$. From this expression, it is readily seen that if $C(\ell)$ decays faster than $1/\ell$ for large ℓ, the sum over ℓ tends to a constant for large N, and thus the variance of the sum still grows as N, as for the usual CLT. If however $C(\ell)$ decays for large ℓ as a power-law $\ell^{-\nu}$, with $\nu < 1$, then the variance grows faster than N, as $N^{2-\nu}$ – correlations thus enhance fluctuations. Hence, when $\nu < 1$, the standard CLT certainly has to be amended. The problem of the limit distribution in these cases is however not solved in general. For example, if the X_i are correlated

[24] We again assume in the following, without loss of generality that the mean m is zero.

Gaussian variables, it is easy to show that the resulting sum is also Gaussian, whatever the value of ν. Another solvable case is when the X_i are correlated Gaussian variables, but one takes the sum of the *squares* of the X_i's. This sum converges towards a Gaussian of width $\sqrt{N}$ whenever $\nu > 1/2$, but towards a non trivial limit distribution of a new kind (i.e. neither Gaussian nor Lévy stable) when $\nu < 1/2$. In this last case, the proper rescaling factor must be chosen as $N^{1-\nu}$.

One can also construct *anti-correlated* random variables, the sum of which grows slower than $\sqrt{N}$. In the case of power-law correlated or anticorrelated Gaussian random variables, one speaks of 'fractional Brownian motion'. This notion was introduced by Mandelbrot and Van Ness [Mandelbrot].

1.7.2 Non stationary models and dependence

It may happen that the distribution of the elementary random variables $P_1(x_1)$, $P_2(x_2), ..., P_N(x_N)$ are not all identical. This is the case, for example, when the variance of the random process depends upon time – in financial markets, it is a well known fact that the daily volatility is time dependent, taking rather high levels in periods of uncertainty, and reverting back to lower values in calmer periods. For example, the volatility of the bond market has been very high during 1994, and decreased in later years. Similarly, the volatility of stock markets has increased since August 1997.

If the distribution P_k varies sufficiently 'slowly', one can in principle measure some of its moments (for example its mean and variance) over a time scale which is long enough to allow for a precise determination of these moments, but short compared to the time scale over which P_k is expected to vary. The situation is less clear if P_k varies 'rapidly'. Suppose for example that $P_k(x_k)$ is a Gaussian distribution of variance σ_k^2, which is itself a random variable. We shall denote as $\overline{(...)}$ the average over the random variable σ_k, to distinguish it from the notation $\langle ... \rangle_k$ which we have used to describe the average over the probability distribution P_k. If σ_k varies rapidly, it is impossible to separate the two sources of uncertainty. Thus, the empirical histogram constructed from the series $\{x_1, x_2, ...x_N\}$ leads to an 'apparent' distribution $\overline{P}$ which is non-Gaussian even if each individual P_k is Gaussian. Indeed, from:

$$\overline{P}(x) \equiv \int d\sigma P(\sigma) \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{x^2}{2\sigma^2}, \tag{1.98}$$

one can calculate the kurtosis of $\overline{P}$ as:

$$\overline{\kappa} = \frac{\overline{\langle x^4 \rangle}}{(\overline{\langle x^2 \rangle})^2} - 3 \equiv 3\left(\frac{\overline{\sigma^4}}{(\overline{\sigma^2})^2} - 1\right). \tag{1.99}$$

Since for any random variable one has $\overline{\sigma^4} \geq (\overline{\sigma^2})^2$ (the equality being reached only if σ^2 does not fluctuate at all), one finds that $\overline{\kappa}$ is always positive. The volatility fluctuations can thus lead to 'fat tails'. More precisely, let us assume that the probability distribution of the RMS, $P(\sigma)$, decays itself for large σ as $\exp -\sigma^c$, $c > 0$. Assuming P_k to be Gaussian, it is easy to obtain, using a saddle point method (cf. (1.75)), that for large x one has:

$$\log[\overline{P}(x)] \propto -x^{\frac{2c}{2+c}}. \tag{1.100}$$

Since $c < 2 + c$, this asymptotic decay is always much slower than in the Gaussian case, which corresponds to $c \to \infty$. The case where the volatility itself has a Gaussian tail ($c = 2$) leads to an exponential decay of $\overline{P}(x)$.

Another interesting case is when σ^2 is distributed as an completely asymmetric Lévy distribution ($\beta = 1$) of exponent $\mu < 1$. Using the properties of Lévy distributions, one can then show that $\overline{P}$ is itself a symmetric Lévy distribution ($\beta = 0$), of exponent equal to 2μ.

If the fluctuations of σ_k are themselves correlated, one observes an interesting case of *dependence*. For example, if σ_k is large, σ_{k+1} will probably also be large. The fluctuation X_k thus has a large probability to be large (but of arbitrary sign) twice in a row. We shall often refer, in the following, to a simple model where x_k can be written as a product $\epsilon_k \sigma_k$, where ϵ_k are IID random variables of zero mean and unit variance, and σ_k corresponds to the local 'scale' of the fluctuations, which can be correlated in time. The correlation function of the X_k is thus given by:

$$\overline{\langle x_i x_j \rangle} = \overline{\sigma_i \sigma_j} \langle \epsilon_i \epsilon_j \rangle = \delta_{i,j} \overline{\sigma^2}. \tag{1.101}$$

Hence the X_k are uncorrelated random variables, but they are not independent since a higher order correlation function reveals a richer structure. Let us for example consider the correlation of X_k^2:

$$\overline{\langle x_i^2 x_j^2 \rangle} - \overline{\langle x_i^2 \rangle \langle x_j^2 \rangle} = \overline{\sigma_i^2 \sigma_j^2} - \overline{\sigma_i^2}\,\overline{\sigma_j^2} \qquad (i \neq j), \tag{1.102}$$

which indeed has an interesting temporal behaviour: see Section 2.4.[25] However, even if the correlation function $\overline{\sigma_i^2 \sigma_j^2} - \overline{\sigma}^2$ decreases very slowly

[25] Note that for $i \neq j$ this correlation function can be zero either because σ is identically equal to a certain value σ_0, or because the fluctuations of σ are completely uncorrelated from one time to the next.

with $|i-j|$, one can show that the sum of the X_k, obtained as $\sum_{k=1}^N \epsilon_k \sigma_k$ is still governed by the CLT, and converges for large N towards a Gaussian variable. A way to see this is to compute the average kurtosis of the sum, κ_N. As shown in Appendix A, one finds the following result:

$$\kappa_N = \frac{1}{N}\left[\kappa_0 + (3+\kappa_0)g(0) + 6\sum_{\ell=1}^{N}(1-\frac{\ell}{N})g(\ell)\right], \tag{1.103}$$

where κ_0 is the kurtosis of the variable ϵ, and $g(\ell)$ the correlation function of the variance, defined as:

$$\overline{\sigma_i^2\sigma_j^2} - \overline{\sigma}^2 = \overline{\sigma}^2 g(|i-j|) \tag{1.104}$$

It is interesting to see that for $N = 1$, the above formula gives $\kappa_1 = \kappa_0 + (3+\kappa_0)g(0) > \kappa_0$, which means that even if $\kappa_0 = 0$, a fluctuating volatility is enough to produce some kurtosis. More importantly, one sees that if the variance correlation function $g(\ell)$ decays with ℓ, the kurtosis κ_N tends to zero with N, thus showing that the sum indeed converges towards a Gaussian variable. For example, if $g(\ell)$ decays as a power-law $\ell^{-\nu}$ for large ℓ, one finds that for large N:

$$\kappa_N \propto \frac{1}{N} \quad \text{for} \quad \nu > 1; \qquad \kappa_N \propto \frac{1}{N^\nu} \quad \text{for} \quad \nu < 1. \tag{1.105}$$

Hence, long-range correlation in the variance considerably slows down the convergence towards the Gaussian. This remark will be of importance in the following, since financial time series often reveal long-ranged volatility fluctuations.

1.8 Central limit theorem for random matrices (*)

One interesting application of the CLT concerns the spectral properties of 'random matrices'. The theory of Random Matrices has made enormous progress during the past thirty years, with many applications in physical sciences and elsewhere. More recently, it has been suggested that random matrices might also play an important role in finance: an example is discussed in Section 2.7. It is therefore appropriate to give a cursory discussion of some salient properties of random matrices. The simplest ensemble of random matrices is one where a all elements of the matrix $\mathbf{H}$ are IID random variables, with the only constraint that the matrix be symmetrical ($H_{ij} = H_{ji}$). One interesting result is that in the limit of very large matrices, the distribution of its eigenvalues has universal properties, which are to a large extent independent of the distribution of the elements of the matrix. This is actually the consequence of the CLT, as we will show below. Let us introduce first some notations. The matrix $\mathbf{H}$ is a square, $N \times N$ symmetric matrix. Its eigenvalues are λ_α, with $\alpha = 1, ..., N$. The *density* of eigenvalues is defined as:

$$\rho(\lambda) = \frac{1}{N} \sum_{\alpha=1}^{N} \delta(\lambda - \lambda_\alpha), \tag{1.106}$$

where δ is the Dirac function. We shall also need the so-called 'resolvent' $\mathbf{G}(\lambda)$ of the matrix $\mathbf{H}$, defined as:

$$G_{ij}(\lambda) \equiv \left(\frac{1}{\lambda \mathbf{1} - \mathbf{H}} \right)_{ij}, \tag{1.107}$$

where $\mathbf{1}$ is the identity matrix. The trace of $\mathbf{G}(\lambda)$ can be expressed using the eigenvalues of $\mathbf{H}$ as:

$$\mathrm{Tr}\mathbf{G}(\lambda) = \sum_{\alpha=1}^{N} \frac{1}{\lambda - \lambda_\alpha}. \tag{1.108}$$

The 'trick' that allows one to calculate $\rho(\lambda)$ in the large N limit is the following representation of the δ function:

$$\frac{1}{x - i\epsilon} = PP\frac{1}{x} + i\pi\delta(x) \qquad (\epsilon \to 0), \tag{1.109}$$

where PP means the principal part. Therefore, $\rho(\lambda)$ can be expressed as:

$$\rho(\lambda) = \lim_{\epsilon \to 0} \frac{1}{\pi} \Im \left(\mathrm{Tr}\mathbf{G}(\lambda - i\epsilon) \right). \tag{1.110}$$

Our task is therefore to obtain an expression for the resolvent $\mathbf{G}(\lambda)$. This can be done by establishing a recursion relation, allowing one to compute $\mathbf{G}(\lambda)$ for a matrix $\mathbf{H}$ with one extra row and one extra column, the elements of which being H_{0i}. One then computes $G_{00}^{N+1}(\lambda)$ (the superscript stands for the size of the matrix $\mathbf{H}$) using the standard formula for matrix inversion:

$$G_{00}^{N+1}(\lambda) = \frac{\mathrm{minor}(\lambda \mathbf{1} - \mathbf{H})_{00}}{\det(\lambda \mathbf{1} - \mathbf{H})}. \tag{1.111}$$

Now, one expands the determinant appearing in the denominator in minors along the first row, and then each minor is itself expanded in subminors along their first column. After a little thought, this finally leads to the following expression for $G_{00}^{N+1}(\lambda)$:

$$\frac{1}{G_{00}^{N+1}(\lambda)} = \lambda - H_{00} - \sum_{i,j=1}^{N} H_{0i} H_{0j} G_{ij}^{N}(\lambda). \tag{1.112}$$

This relation is general, without any assumption on the H_{ij}. Now, we assume that the H_{ij}'s are IID random variables, of zero mean and variance equal to $\langle H_{ij}^2 \rangle = \sigma^2/N$. This scaling with N can be understood as follows: when the matrix H acts on a certain vector, each component of the image vector is a sum of N random variables. In order to keep the image vector (and thus the corresponding eigenvalue) finite when $N \to \infty$, one should scale the elements of the matrix with the factor $1/\sqrt{N}$.

One could also write a recursion relation for G_{0i}^{N+1}, and establish self-consistently that $G_{ij} \sim 1/\sqrt{N}$ for $i \neq j$. On the other hand, due to the diagonal term λ, G_{ii} remains finite for $N \to \infty$. This scaling allows us to discard all the terms with $i \neq j$ in the sum appearing in the right hand side of Eq. (1.112). Furthermore, since $H_{00} \sim 1/\sqrt{N}$, this term can be neglected compared to λ. This finally leads to a simplified recursion relation, valid in the limit $N \to \infty$:

$$\frac{1}{G_{00}^{N+1}(\lambda)} \simeq \lambda - \sum_{i=1}^{N} H_{0i}^2 G_{ii}^N(\lambda). \tag{1.113}$$

Now, using the CLT, we know that the last sum converges, for large N, towards $\sigma^2/N \ \sum_{i=1}^N G_{ii}^N(\lambda)$. This result is independent of the precise statistics of the H_{0i}, provided their variance is finite.[26] This shows that G_{00} converges for large N towards a well defined limit G_∞, which obeys the following limit equation:

$$\frac{1}{G_\infty(\lambda)} = \lambda - \sigma^2 G_\infty(\lambda). \tag{1.114}$$

The solution to this second order equation reads:

$$G_\infty(\lambda) = \frac{1}{2\sigma^2}\left[\lambda - \sqrt{\lambda^2 - 4\sigma^2}\right]. \tag{1.115}$$

(The correct solution is chosen to recover the right limit for $\sigma = 0$.) Now, the only way for this quantity to have a non zero imaginary part when one adds to λ a small imaginary term $i\epsilon$ which tends to zero is that the square root itself is imaginary. The final result for the density of eigenvalues is therefore:

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2} \qquad \text{for } |\lambda| \leq 2\sigma, \tag{1.116}$$

[26] The case of Lévy distributed H_{ij}'s with infinite variance has been investigated in: P. Cizeau, J.-P. Bouchaud, "Theory of Lévy matrices", Phys. Rev. **E 50** 1810 (1994).

and zero elsewhere. This is the well known 'semi-circle' law for the density of states, first derived by Wigner. This result can be obtained by a variety of other methods if the distribution of matrix elements is Gaussian. In finance, one often encounters *correlation matrices* $\mathbf{C}$, which have the special property of being positive definite. $\mathbf{C}$ can be written as $\mathbf{C} = \mathbf{H}\mathbf{H}^\dagger$, where $\mathbf{H}^\dagger$ is the matrix transpose of $\mathbf{H}$. In general, $\mathbf{H}$ is a rectangular matrix of size $M \times N$. In Chapter 2, M will be the number of asset, and N, the number of observation (days). In the particular case where $N = M$, the eigenvalues of $\mathbf{C}$ are simply obtained from those of $\mathbf{H}$ by squaring them:

$$\lambda_C = \lambda_H^2. \tag{1.117}$$

If one assumes that the elements of $\mathbf{H}$ are random variables, the density of eigenvalues of $\mathbf{C}$ can easily be obtained from:

$$\rho(\lambda_C) d\lambda_C = \rho(\lambda_H) d\lambda_H, \tag{1.118}$$

which leads to:

$$\rho(\lambda_C) = \frac{1}{4\pi\sigma^2}\sqrt{\frac{4\sigma^2 - \lambda_C}{\lambda_C}} \qquad \text{for } 0 \leq \lambda_C \leq 4\sigma^2, \tag{1.119}$$

and zero elsewhere. For $N \neq M$, a similar formula exists, which we shall use in the following. In the limit $N, M \to \infty$, with a fixed ratio $Q = N/M \geq 1$, one has:[27]

$$\begin{aligned} \rho(\lambda_C) &= \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{\max} - \lambda_C)(\lambda_C - \lambda_{\min})}}{\lambda}, \\ \lambda_{\min}^{\max} &= \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}), \end{aligned} \tag{1.120}$$

with $\lambda \in [\lambda_{\min}, \lambda_{\max}]$. This form is actually also valid for $Q < 1$, except that there appears a finite fraction of strictly zero eigenvalues, of weight $1 - Q$.

The most important features predicted by Eq. (1.120) are:

- The fact that the lower 'edge' of the spectrum is strictly positive (except for $Q = 1$); there is therefore no eigenvalues between 0 and $\lambda_{\min}$. Near this edge, the density of eigenvalues exhibits a sharp maximum, except in the limit $Q = 1$ ($\lambda_{\min} = 0$) where it diverges as $\sim 1/\sqrt{\lambda}$.

[27] see: A. Edelmann, "Eigenvalues and condition numbers of random matrices," SIAM J. Matrix Anal. Appl. **9**, 543 (1988).

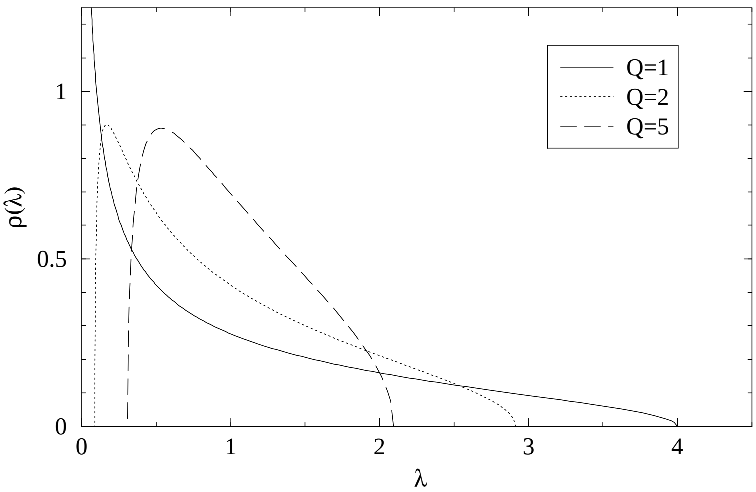

Figure 1.10: Graph of Eq. (1.120) for Q=1, 2 and 5.

- The density of eigenvalues also vanishes above a certain upper edge $\lambda_{\max}$.

Note that all the above results are only valid in the limit $N \to \infty$. For finite N, the singularities present at both edges are smoothed: the edges become somewhat blurred, with a small probability of finding eigenvalues above $\lambda_{\max}$ and below $\lambda_{\min}$, which goes to zero when N becomes large.[28]

In Chapter 2, we will compare the empirical distribution of the eigenvalues of the correlation matrix of stocks corresponding to different markets with the theoretical prediction given by Eq. (1.120).

1.9 Appendix A: non-stationarity and anomalous kurtosis

In this appendix, we calculate the average kurtosis of the sum $\sum_{i=1}^N \delta x_i$, assuming that the δx_i can be written as $\sigma_i \epsilon_i$. The σ_i's are correlated as:

$$\overline{(D_k - \overline{D})(D_\ell - \overline{D})} = \overline{D}^2 g(|\ell - k|) \qquad D_k \propto \sigma_k^2. \tag{1.121}$$

Let us first compute $\overline{\left\langle \left(\sum_{i=1}^N \delta x_i\right)^4 \right\rangle}$, where $\langle ... \rangle$ means an average over the ϵ_i's and the overline means an average over the σ_i's. If $\langle \epsilon_i \rangle = 0$, and

[28] see e.g. M. J. Bowick, E. Brézin, "Universal scaling of the tails of the density of eigenvalues in random matrix models," Phys. Lett **B268**, 21 (1991).

$\langle \epsilon_i \epsilon_j \rangle = 0$ for $i \neq j$, one finds:

$$\overline{\left\langle \left(\sum_{i,j,k,l=1}^{N} \delta x_i \delta x_j \delta x_k \delta x_l \right) \right\rangle} = \sum_{i=1}^{N} \overline{\langle \delta x_i^4 \rangle} + 3 \sum_{i \neq j=1}^{N} \overline{\langle \delta x_i^2 \rangle \langle \delta x_j^2 \rangle}$$

$$= (3 + \kappa_0) \sum_{i=1}^{N} \overline{\langle \delta x_i^2 \rangle^2} + 3 \sum_{i \neq j=1}^{N} \overline{\langle \delta x_i^2 \rangle \langle \delta x_j^2 \rangle}, \tag{1.122}$$

where we have used the definition of κ_0 (the kurtosis of ϵ). On the other hand, one must estimate $\overline{\left\langle \left(\sum_{i=1}^{N} \delta x_i \right)^2 \right\rangle}^2$. One finds:

$$\overline{\left\langle \left(\sum_{i=1}^{N} \delta x_i \right)^2 \right\rangle}^2 = \sum_{i,j=1}^{N} \overline{\langle \delta x_i^2 \rangle}\ \overline{\langle \delta x_j^2 \rangle}. \tag{1.123}$$

Gathering the different terms and using the definition (1.121), one finally establishes the following general relation:

$$\kappa_N = \frac{1}{N^2 \overline{D}^2} \left[N \overline{D}^2 (3 + \kappa_0)(1 + g(0)) - 3N\overline{D}^2 + 3\overline{D}^2 \sum_{i \neq j=1}^{N} g(|i - j|) \right], \tag{1.124}$$

or:

$$\kappa_N = \frac{1}{N} \left[\kappa_0 + (3 + \kappa_0) g(0) + 6 \sum_{\ell=1}^{N} (1 - \frac{\ell}{N}) g(\ell) \right]. \tag{1.125}$$

1.10 References

• Introduction to probabilities:

W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley, New York, 1971.

• Extreme value statistics:

E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, 1958.

• Sum of Random Variables, Lévy distributions:

B. V. Gnedenko & A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison Wesley, 1954.

P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier Villars, 1937–1954.

G. Samorodnitsky & M. S. Taqqu, *Stable Non-gaussian Random Processes*, Chapman & Hall, New York, 1994.

• Broad distributions in Natural Sciences and in Finance:

B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, 1982.

B. B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, 1997.

• Broad distributions in Physics:

E. Montroll, M. Shlesinger, *The Wonderful World of Random Walks*, in Nonequilibrium phenomena II, *From stochastics to hydrodynamics*, Studies in statistical mechanics XI (J. L. Lebowitz & E. W. Montroll, eds.) North Holland, Amsterdam, 1984.

J.-P. Bouchaud, A. Georges, "Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications," *Physics Reports*, **195**, 127 (1990).

G. Zaslavsky, U. Frisch & M. Shlesinger *Lévy Flights and Related Topics in Physics*, in Lecture Notes in Physics **450**, Springer, 1995.

• Theory of Random Matrices:

M. L. Mehta , *Random matrix theory*, Academic Press, New York, 1995.

O. Bohigas, M. J. Giannoni, *Mathematical and Computational Methods in Nuclear Physics*, Lecture Notes in Physics, vol. 209, Springer-Verlag, 1983.