THEORY OF FINANCIAL RISK: 
FROM DATA ANALYSIS TO RISK 
MANAGEMENT

Jean-Philippe Bouchaud 
and 
Marc Potters

DRAFT
May 24, 1999

Contents

Foreword ix

1 Probability theory: basic notions 1
1.1 Introduction 1
1.2 Probabilities 4
    1.2.1 Probability distributions 4
    1.2.2 Typical values and deviations 5
    1.2.3 Moments and characteristic function 7
    1.2.4 Divergence of moments – Asymptotic behaviour 8
1.3 Some useful distributions 9
    1.3.1 Gaussian distribution 9
    1.3.2 Log-normal distribution 10
    1.3.3 Lévy distributions and Paretian tails 11
    1.3.4 Other distributions (*) 15
1.4 Maximum of random variables – Statistics of extremes 17
1.5 Sums of random variables 22
    1.5.1 Convolutions 22
    1.5.2 Additivity of cumulants and of tail amplitudes 23
    1.5.3 Stable distributions and self-similarity 24
1.6 Central limit theorem 25
    1.6.1 Convergence to a Gaussian 25
    1.6.2 Convergence to a Lévy distribution 30
    1.6.3 Large deviations 30
    1.6.4 The CLT at work on a simple case 33
    1.6.5 Truncated Lévy distributions 37
    1.6.6 Conclusion: survival and vanishing of tails 38
1.7 Correlations, dependence and non-stationary models (*) 39
    1.7.1 Correlations 39
    1.7.2 Non-stationary models and dependence 40
1.8 Central limit theorem for random matrices (*) 42
1.9 Appendix A: non-stationarity and anomalous kurtosis 46
1.10 References 47
Contents

2 Statistics of real prices 49
  2.1 Aim of the chapter 49
  2.2 Second order statistics 53
    2.2.1 Variance, volatility and the additive-multiplicative crossover 53
    2.2.2 Autocorrelation and power spectrum 56
  2.3 Temporal evolution of fluctuations 60
    2.3.1 Temporal evolution of probability distributions 60
    2.3.2 Multiscaling – Hurst exponent (*) 68
  2.4 Anomalous kurtosis and scale fluctuations 70
  2.5 Volatile markets and volatility markets 76
  2.6 Statistical analysis of the Forward Rate Curve (*) 79
    2.6.1 Presentation of the data and notations 80
    2.6.2 Quantities of interest and data analysis 81
    2.6.3 Comparison with the Vasicek model 84
    2.6.4 Risk-premium and the $\sqrt{\theta}$ law 87
  2.7 Correlation matrices (*) 90
  2.8 A simple mechanism for anomalous price statistics (*) 93
  2.9 A simple model with volatility correlations and tails (*) 95
  2.10 Conclusion 96
  2.11 References 97

3 Extreme risks and optimal portfolios 101
  3.1 Risk measurement and diversification 101
    3.1.1 Risk and volatility 101
    3.1.2 Risk of loss and ‘Value at Risk’ (VaR) 104
    3.1.3 Temporal aspects: drawdown and cumulated loss 109
    3.1.4 Diversification and utility – Satisfaction thresholds 115
    3.1.5 Conclusion 119
  3.2 Portfolios of uncorrelated assets 119
    3.2.1 Uncorrelated Gaussian assets 121
    3.2.2 Uncorrelated ‘power law’ assets 125
    3.2.3 ‘Exponential’ assets 127
    3.2.4 General case: optimal portfolio and VaR (*) 129
  3.3 Portfolios of correlated assets 130
    3.3.1 Correlated Gaussian fluctuations 130
    3.3.2 ‘Power law’ fluctuations (*) 133
  3.4 Optimised trading (*) 137
  3.5 Conclusion of the chapter 139
  3.6 Appendix B: some useful results 140
  3.7 References 141

4 Futures and options: fundamental concepts 143
  4.1 Introduction 143
    4.1.1 Aim of the chapter 143
    4.1.2 Trading strategies and efficient markets 144
  4.2 Futures and Forwards 147
    4.2.1 Setting the stage 147
    4.2.2 Global financial balance 148
    4.2.3 Riskless hedge 150
    4.2.4 Conclusion: global balance and arbitrage 151
  4.3 Options: definition and valuation 152
    4.3.1 Setting the stage 152
    4.3.2 Orders of magnitude 155
    4.3.3 Quantitative analysis – Option price 156
    4.3.4 Real option prices, volatility smile and ‘implied’ kurtosis 161
  4.4 Optimal strategy and residual risk 170
    4.4.1 Introduction 170
    4.4.2 A simple case 171
    4.4.3 General case: ‘$\Delta$’ hedging 175
    4.4.4 Global hedging/instantaneous hedging 180
    4.4.5 Residual risk: the Black-Scholes miracle 180
    4.4.6 Other measures of risk – Hedging and VaR (*) 185
    4.4.7 Hedging errors 187
    4.4.8 Summary 188
  4.5 Does the price of an option depend on the mean return? 188
    4.5.1 The case of non-zero excess return 188
    4.5.2 The Gaussian case and the Black-Scholes limit 193
    4.5.3 Conclusion. Is the price of an option unique? 196
  4.6 Conclusion of the chapter: the pitfalls of zero-risk 197
    4.7 Appendix C: computation of the conditional mean 198
    4.8 Appendix D: binomial model 200
    4.9 Appendix E: option price for (suboptimal) $\Delta$-hedging 201
  4.10 References 202

5 Options: some more specific problems 205
  5.1 Other elements of the balance sheet 205
    5.1.1 Interest rate and continuous dividends 205
    5.1.2 Interest rates corrections to the hedging strategy 208
    5.1.3 Discrete dividends 209
    5.1.4 Transaction costs 209
  5.2 Other types of options: ‘Puts’ and ‘exotic options’ 211
    5.2.1 ‘Put-Call’ parity 211

FOREWORD

Finance is a rapidly expanding field of science, with a rather unique link to applications. Correspondingly, the recent years have witnessed the growing role of Financial Engineering in market rooms. The possibility of accessing and processing rather easily huge quantities of data on financial markets opens the path to new methodologies, where systematic comparison between theories and real data becomes not only possible, but required. This perspective has spurred the interest of the statistical physics community, with the hope that methods and ideas developed in the past decades to deal with complex systems could also be relevant in Finance. Correspondingly, many PhD in physics are now taking jobs in banks or other financial institutions.

However, the existing literature roughly falls into two categories: either rather abstract books from the mathematical finance community, which are very difficult to read for people trained in natural science, or more professional books, where the scientific level is usually poor.1 In particular, there is in this context no book discussing the physicists’ way of approaching scientific problems, in particular a systematic comparison between ‘theory’ and ‘experiments’ (i.e. empirical results), the art of approximations and the use of intuition.2 Moreover, even in excellent books on the subject, such as the one by J.C. Hull, the point of view on derivatives is the traditional one of Black and Scholes, where the whole pricing methodology is based on the construction of riskless strategies. The idea of zero risk is counter-intuitive and the reason for the existence of these riskless strategies in the Black-Scholes theory is buried in the premises of Ito’s stochastic differential rules.

It is our belief that a more intuitive understanding of these theories is needed for a better control of financial risks all together. The models discussed in Theory of Financial Risk are devised to account for real markets’ statistics where the construction of riskless hedges is impossible. The mathematical framework required to deal with these cases is however not more complicated, and has the advantage of making issues at stake,

1There are notable exceptions, such as the remarkable book by J. C. Hull, Futures and options and other derivatives, Prentice Hall, 1997.
in particular the problem of risk, more transparent.

Finally, commercial softwares are being developed to measure and control financial risks (some following the ideas developed in this book).\footnote{For example, the software Profiler, commercialised by the company ATSM, heavily relies on the concepts introduced in Chapter 3.} We hope that this book can be useful to all people concerned with financial risk control, by discussing at length the advantages and limitations of various statistical models.

Despite our efforts to remain simple, certain sections are still quite technical. We have used a smaller font to develop more advanced ideas, which are not crucial to understand the main ideas. Whole sections, marked by a star (*), contain rather specialised material and can be skipped at first reading. We have tried to be as precise as possible, but have sometimes been somewhat sloppy and non-rigourous. For example, the idea of probability is not axiomatised: its intuitive meaning is more than enough for the purpose of this book. The notation $P(.)$ means the probability distribution for the variable which appears between the parenthesis, and not a well determined function of a dummy variable. The notation $x \to \infty$ does not necessarily mean that $x$ tends to infinity in a mathematical sense, but rather that $x$ is large. Instead of trying to derive results which hold true in any circumstances, we often compare order of magnitudes of the different effects: small effects are neglected, or included perturbatively.\footnote{$a \simeq b$ means that $a$ is of order $b$, $a \ll b$ means that $a$ is smaller than – say – $b/10$. A computation neglecting terms of order $(a/b)^2$ is therefore accurate to one percent. Such a precision is often enough in the financial context, where the uncertainty on the value of the parameters (such as the average return, the volatility, etc., can be larger than 1%).}

Finally, we have not tried to be comprehensive, and have left out a number of important aspects of theoretical finance. For example, the problem of interest rate derivatives (swaps, caps, swaptions...) is not addressed – we feel that the present models of interest rate dynamics are not satisfactory (see the discussion in Section 2.7). Correspondingly, we have not tried to give an exhaustive list of references, but rather to present our own way of understanding the subject. A certain number of important references are given at the end of each chapter, while more specialised papers are given as footnotes where we have found it necessary.

This book is divided into five chapters. Chapter 1 deals with important results in probability theory (the Central Limit Theorem and its limitations, the extreme value statistics, etc.). The statistical analysis of real data, and the empirical determination of the statistical laws, are discussed in Chapter 2. Chapter 3 is concerned with the definition of risk, Value-at-Risk, and the theory of optimal portfolio, in particular in the case where the probability of extreme risks has to be minimised. The problem of forward contracts and options, their optimal hedge and the residual risk is discussed in detail in Chapter 4. Finally, some more advanced topics on options are introduced in Chapter 5 (such as exotic options, or the role of transaction costs). Finally, an index and a list of symbols are given in the end of the book, allowing one to find easily where each symbol or word was used and defined for the first time.

This book appeared in its first edition in French, under the title: Théorie des Risques Financiers, Aléa-Saclay-Eyrolles, Paris (1997). Compared to this first edition, the present version has been substantially improved and augmented. For example, we discuss the theory of random matrices and the problem of the interest rate curve, which were absent from the first edition. Several points were furthermore corrected or clarified.

Acknowledgements

This book owes a lot to discussions that we had with Rama Cont, Didier Sornette (who participated to the initial version of Chapter 3), and to the entire team of Science and Finance: Pierre Cizeau, Laurent Laloux, Andrew Matacz and Martin Meyer. We want to thank in particular Jean-Pierre Aguilar, who introduced us to the reality of financial markets, suggested many improvements, and supported us during the many years that this project took to complete. We also thank the companies ATSM and CFM, for providing financial data and for keeping us close to the real world. We also had many fruitful exchanges with Jeff Miller, and also Alain Arnould, Aubry Miens,\footnote{with whom we discussed Eq. (1.24), which appears in his Diplomarbeit.} Erik Aurell, Martin Baxter, Jean-Francois Chauwin, Nicole El Karoui, Stefano Galluccio, Gaëlle Gego, David Jeammet, Imre Kondor, Jean-Michel Lasry, Rosario Mantegna, Jean-François Muzy, Nicolas Sagna, Gene Stanley, Christian Walter, Mark Wexler and Karol Zyczkowski. We thank Claude Godrèche, who edited the French version of this book, for his friendly advices and support. Finally, JPB wants to thank Elisabeth Bouchaud for sharing so many far more important things.

This book is dedicated to our families, and more particularly to the memory of Paul Potters.
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 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.

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

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

1.2.1 Probability distributions

Contrarily to the throw of a dice, which can only return an integer between 1 and 6, the variation of price of a financial asset\(^3\) can be arbitrary (we disregard 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 \(x = \xi\) is equal to \(P(x)dx\). In the following, we shall denote as \(P(\cdot)\) the probability density for the variable appearing as the argument of the function. This is a potentially ambiguous, but very useful notation.

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

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

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

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

An equivalent way of describing the distribution of \(X\) is to consider its cumulative distribution \(P_\langle(x)\), defined as:

\[
P_\langle(x) = P(X < x) = \int_{-\infty}^x P(x')dx'. \tag{1.3}
\]

\(P_\langle(x)\) takes values between zero and one, and is monotonously increasing with \(x\). Obviously, \(P_\langle(-\infty) = 0\) and \(P_\langle(+\infty) = 1\). Similarly, one defines \(P_\rangle(x) = 1 - P_\langle(x)\).

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 \(x_{\text{med}}\) is such that the probabilities that \(X\) lies between two specific values \(x\) is of course equal to one. In other words, \(P_\langle(x_{\text{med}}) = P_\rangle(x_{\text{med}}) = \frac{1}{2}\). The mean, or expected value of \(X\), which we shall note as \(m\) or \(\langle x \rangle\) in the following, is the average of all possible values of \(X\), weighted by their corresponding probability:

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

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) \(\triangle\), by the root mean square (RMS) \(\sigma\) (or, in financial terms, the volatility ), or by the ‘full width at half maximum’ \(\Delta W/2\).

The mean absolute deviation from a given reference value is the average of the distance between the possible values of \(X\) and this reference

---

\(^3\)Asset is the generic name for a financial instrument which can be bought or sold, like stocks, currencies, gold, bonds, etc.
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 $X$:

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

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. 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:

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

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

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

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 $z = 0$,

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

---

4One 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.

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

---

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.

---

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 \( \mu \)). The derivatives of order larger or equal to \( \mu \) of the characteristic function thus do not exist at the origin \( (z = 0) \).

### 1.3 Some useful distributions

#### 1.3.1 Gaussian distribution

The most commonly encountered distributions are the ‘normal’ laws of Laplace and Gauss, which we shall simply call 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 \( \sigma \) is defined as:

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

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 \( m_{2n} = (2n-1)(2n-3)...(2n-1)!! \sigma^{2n} \).

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

\[
P_G(z) = \exp\left(-\frac{\sigma^2 z^2}{2} + imz\right). \tag{1.16}\]

\(^7\)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.
Its logarithm is a second order polynomial, for which all derivatives of order larger than two are zero. In particular, the kurtosis of a Gaussian variable is zero. As mentioned above, the kurtosis is often taken as a measure of the distance from a Gaussian distribution. When \( \kappa > 0 \) (leptokurtic distributions), the corresponding distribution density has a marked peak around the mean, and rather ‘thick’ tails. Conversely, when \( \kappa < 0 \), the distribution density has a flat top and very thin tails. For example, the uniform distribution over a certain interval (for which tails are absent) has a kurtosis \( \kappa = -\frac{6}{5} \).

A Gaussian variable is peculiar because ‘large deviations’ are extremely rare. The quantity \( \exp(-x^2/2\sigma^2) \) decays so fast for large \( x \) that deviations of a few times \( \sigma \) are nearly impossible. For example, a Gaussian variable departs from its most probable value by more than \( 2\sigma \) only 5\% of the times, of more than \( 3\sigma \) in 0.2\% of the times, while a fluctuation of \( 10\sigma \) has a probability of less than \( 2 \times 10^{-23} \); 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:

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

(1.17)

the moments of which being: \( m_n = x_0^n \exp(n^2\sigma^2/2) \).

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-normal distributions account for the following symmetry in the problem of exchange rates: 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 principle in favour of log-normals: since the price of an asset cannot be negative, its statistics cannot be Gaussian since the latter admits in principle negative values, 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, the corresponding distribution density has a flat top and very thin tails. For \( n \) 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 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 Pareto tails

Lévy distributions (noted \( L_a(x) \) 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 ‘Pareto tails’:

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

where \( 0 < \mu < 2 \) is a certain exponent (often called \( \alpha \)), 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 \( P_\mu(x) = (A_+ / x)^\mu \) 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 general, 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’):

\[ L_1(x) = \frac{A}{x^2 + \pi^2 A^2}. \quad (1.19) \]

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

\[ \hat{L}_\mu(z) = \exp(-\mu|z|^\mu), \quad (1.20) \]

where \( a_\mu \) is a certain constant, proportional to the tail parameter \( A_\mu \).\(^{12}\) It is thus clear that in the limit \( \mu = 2 \), one recovers the definition of a Gaussian. When \( \mu \) 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.

\[ \text{In the case } \beta \neq 0, \text{ one would have:} \]

\[ L_\mu^\beta(z) = \exp \left[ -\mu|z|^\mu \left( 1 + i\beta \tan(\mu\pi/2) \frac{z}{|z|} \right) \right] \quad (\mu \neq 1). \quad (1.21) \]

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

\(^{11}\) 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 \( 1 < \mu < 2 \), \( A_\mu = \mu \Gamma((\mu - 1)\pi/2) a_\mu / \pi \).
cut-off for large arguments is to set:\[ \mu = 0.8, 1.2, 1.6 \text{ and } 2 \text{ (this last value actually corresponds to a Gaussian). The smaller } \mu, \text{ the sharper the ‘body’ of the distribution, and the fatter the tails, as illustrated in the inset.} \]

important for finite \( x \). The full asymptotic series actually reads:

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

The presence of the subleading terms may lead to a bad empirical estimate of the exponent \( \mu \) based on a fit of the tail of the distribution. In particular, the ‘apparent’ exponent which describes the function \( L_\mu \) for finite \( x \) is larger than \( \mu \), and decreases towards \( \mu \) for \( x \to \infty \), but more and more slowly as \( \mu \) gets nearer to the Gaussian value \( \mu = 2 \), for which the power-law tails no longer exist. Note however that one also often observes empirically the opposite behaviour, i.e. an apparent Pareto exponent which grows with \( x \). This arises when the Pareto distribution (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 \( \mu \) 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

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

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

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

**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

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

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

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

**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 \( \omega \) (e.g. the times when the price of an asset changes). The number of points \( n \) in an arbitrary interval of length \( \ell \) is distributed according to the Poisson distribution:

\[
P(n) = \frac{(\omega \ell)^n}{n!} \exp(-\omega \ell). \tag{1.27}
\]

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:

$$P_E(x) = \frac{\alpha}{2} \exp(-\alpha|x|),$$  \hspace{1cm} (1.32)

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

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

$$P_S(x) \equiv \frac{1}{\sqrt{\pi}} \frac{\Gamma((1+\mu)/2)}{\Gamma(\mu/2)} \frac{\alpha^\mu}{(a^2 + x^2)^{(1+\mu)/2}},$$  \hspace{1cm} (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 $\mu$. 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:

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

$$P(x_{\text{max}} < \Lambda) = [P_{\leq}(\Lambda)]^N.$$  \hfill (1.35)

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

$$P(x_{\text{max}} < \Lambda) = [1 - P_{>}(\Lambda)]^N \simeq e^{-NP_{>}(\Lambda)}.$$  \hfill (1.36)

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

$$P_{>}(\Lambda_{\text{med}}) = 1 - \left(\frac{1}{2}\right)^{1/N} \simeq \frac{\log 2}{N}.$$  \hfill (1.37)

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

$$P_{>}(\Lambda_p) \simeq -\frac{\log p}{N}.$$  \hfill (1.38)

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

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 $\Lambda$ expected next year, with 95% confidence, is defined such that $P(x_{\text{max}} < \Lambda_{\text{med}}) = 1/2$:

$$P_{>}(\Lambda_{\text{med}}) = 1 - \left(\frac{1}{2}\right)^{1/N} \simeq \frac{\log 2}{N}.$$  \hfill (1.39)

Interestingly, the distribution of $x_{\text{max}}$ 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 $P_{>}(x) \sim \exp(-ax)$, one finds:

$$\Lambda_{\text{med}} = \frac{\log N}{a},$$  \hfill (1.40)

which grows very slowly with $N$.$^{14}$ Setting $x_{\text{max}} = \Lambda_{\text{med}} + \frac{u}{\Lambda_{\text{med}}}$, one finds that the deviation $u$ around $\Lambda_{\text{med}}$ is distributed according to the Gumbel distribution:

$$P(u) = e^{-e^{-u}}.$$  \hfill (1.41)

The most probable value of this distribution is $u = 0.15$. This shows that $\Lambda_{\text{med}}$ is the most probable value of $x_{\text{max}}$. The result (1.40) is actually much more general, and is valid as soon as $P(x)$ decreases more rapidly than any power-law for $x \to \infty$: the deviation between $\Lambda_{\text{med}}$ (defined as (1.34)) and $x_{\text{max}}$ 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,

$$P_{>}(x) \simeq \frac{\Lambda^u}{x^u}.$$  \hfill (1.41)

$^{14}$For example, for a symmetric exponential distribution $P(x) = \exp(-|x|)/2$, the median value of the maximum of $N = 10000$ variables is only 6.3.
and the typical value of the maximum is given by:

$$\Lambda_{\text{max}} = A_+ N^{1/\mu}. \quad (1.42)$$

Numerically, for a distribution with $\mu = 3/2$ and a scale factor $A_+ = 1$, the largest of $N = 10000$ 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:

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

Its asymptotic behaviour for $u \to \infty$ is still a power law of exponent $1 + \mu$. Said differently, both power-law tails and exponential tails are stable with respect to the ‘max’ operation.\(^{16}\) The most probable value $x_{\text{max}}$ is now equal to $(\mu/1 + \mu)^{1/\mu} \Lambda_{\text{max}}$. As mentioned above, the limit $\mu \to \infty$ formally corresponds to an exponential distribution. In this limit, one indeed recovers $\Lambda_{\text{max}}$ 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 $N\mu$, where $\mu$ 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 $\text{rms}$ of the sum is equal to $\sigma \sqrt{N}$. But for $\mu < 2$, $x_{\text{max}}$ 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 $N[n]$ below. (In particular, $N[1] = x_{\text{max}}$.) The distribution $P_n$ of $N[n]$ can be obtained in full generality as:

$$P_n(N[n]) = C^N_n \left( P(x = N[n]) \right)^{(N-n)} \left( P(x < N[n]) \right)^{n-1}. \quad (1.44)$$

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 $N[n]$ and $N - n$ are smaller than $N[n]$. One can study the position $N^*\[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 $N^*\[n]$. The calculation simplifies in the limit where

\(^{16}\)A third class of laws stable under ‘max’ concerns random variables which are bounded from above – i.e. such that $P(x) = 0$ for $x > x_M$, with $x_M$ finite. This leads to the Weibull distributions, which we will not consider further in this book.
then, if \( P(x) \) decreases as in (1.14), one finds, for \( \Lambda \to \infty \),

\[
\langle x \rangle_\Lambda = \frac{\mu}{\mu - 1} \Lambda,
\]

independently of the tail amplitude \( A_x' \).\(^{17}\) The average \( \langle x \rangle_\Lambda \) is thus always of order of \( \Lambda \) itself, with a proportionality factor which diverges as \( \mu \to 1 \).

### 1.5 Sums of random variables

In order to describe the statistics of future prices of a financial asset, one \textit{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

\[
P(x) = \int dx_1 P_1(x_1) P_2(x-x_1) = \int dx_2 P_1(x_2) P_2(x-x_2).
\]

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

\[
P(x, N = 2) = \int dx' P_1(x') P_2(x-x').
\]

One thus understands how powerful is the hypothesis that the increments are IID, i.e. that \( P_1 = P_2 = \ldots = P_N \). 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 \( P(x, N = 2) = |P_1 * P_2(x)| \), reads in Fourier space:

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

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.

\(^{17}\)This means that \( \mu \) can be determined by a one parameter fit only.
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 $\kappa$, 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 $\mu$. The cumulants of order higher than $\mu$ 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 $A_\mu$: 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 $\mu$, 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:

$$A_{\mu,N}^\mu \equiv N A_{\mu}^\mu. \quad (1.54)$$

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 $P(x,N) = [P_1(x_1)]^N$. 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:

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

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).

The family of all possible stable laws coincide (for continuous variables) with the Lévy distributions defined above,\(^{18}\) 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 $A_N = N^{\frac{\mu}{2}}$, and one finds, for $\mu < 2$:

$$\langle |x|^q \rangle \propto AN^{\frac{\mu}{2}} \quad q < \mu \quad (1.56)$$

where $A = A_\mu = A_{1\mu}$. In words, the above equation means that the order of magnitude of the fluctuations on ‘time’ scale $N$ is a factor $N^{\frac{\mu}{2}}$ 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 $\mu$, 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 $\sigma^2$ towards a Gaussian. In a more precise

---

\(^{18}\) For discrete variables, one should also add the Poisson distribution (1.27).

\(^{19}\) The case $\mu = 1$ is special and involves extra logarithmic factors.
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 $a_N = N^{1/2}$.

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 $x_1$. 
way, the result is then the following:

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

(1.57)

for all finite $u_1$, $u_2$. Note however that for finite $N$, the distribution of the sum $X = X_1 + \ldots + X_N$ 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 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 $X = p_1 X_1 + p_2 X_2 + \ldots + p_N X_N$, 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 $P_i(x)$ 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 $P_i(x)$. 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 $P_i(x)$ has power-law tails with $\mu > 2$ (such that $\sigma$ is finite), the Gaussian ‘realm’ grows barely faster than $\sqrt{N}$ (as $\sim N \log N$).

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 to the Gaussian basin of attraction. More precisely, the necessary and sufficient condition for $P_i(x)$ to belong to this basin is that:

$$\lim_{u \to \infty} \int_{|u|<u} du' u^2 P_i(u') = 0.$$  (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 $I$ (or entropy) associated with a probability distribution $P$ is defined as:

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

(1.59)

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

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

(1.60)

where $\zeta$ is fixed by the condition $\int dx' x'^2 P(x') = \sigma^2$ and $\zeta'$ 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:

$$I_G = \frac{3}{2} + \frac{1}{2} \log (2\pi) + \log(\sigma) \simeq 2.419 + \log(\sigma).$$

(1.61)
For comparison, one can compute the entropy of the symmetric exponential distribution, which is:

$$I_E = 2 + \frac{\log 2}{2} + \log(\sigma) \simeq 2.346 + \log(\sigma).$$  (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 $A^n = A^n_+ = A^n_-$ (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 $\mu$ and with a tail amplitude $N A^n$. If the positive and negative tails of the elementary distribution $P_1(x_1)$ are characterised by different amplitudes ($A^n_+$ and $A^n_-$) one then obtains an asymmetric Lévy distribution with parameter $\beta = (A^n_+ - A^n_-)/(A^n_+ + A^n_-)$. 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:

$$\lim_{u \to \pm\infty} \frac{P_{<}(u)}{P_{>}(u)} = \frac{1 - \beta}{1 + \beta},$$  (1.63)

and for all $r$,

$$\lim_{u \to \pm\infty} \frac{P_{<}(u + ru)}{P_{>}(u + ru)} = r^\mu.$$  (1.64)

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

$$P_{<}(u) \simeq \frac{A^n_-}{|u|^\mu} \quad \text{and} \quad P_{>}(u) \simeq \frac{A^n_+}{u^\mu},$$  (1.65)

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

### 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 $\sigma^2$, one defines a ‘rescaled variable’ $U$ as:

$$U = \frac{X - Nm}{\sigma \sqrt{N}},$$  (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:

$$\lim_{N \to \infty} P_{>}(u) = P_{G>}(u),$$  (1.67)

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

$$P_{G>}(u) = \int_u^\infty \frac{du'}{\sqrt{2\pi}} \exp(-u'^2/2) = \frac{1}{2} \text{erfc}\left(\frac{u}{\sqrt{2}}\right).$$  (1.68)

However, the above convergence is not uniform. The value of $N$ such that the approximation $P_{>}(u) \simeq P_{G>}(u)$ 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 $\Delta P_{>}(u) \equiv P_{>}(u) - P_{G>}(u)$,

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

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

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

and

$$Q_2(u) = \frac{1}{12} \lambda_5 u^5 + \frac{1}{8} (\frac{7}{9} \lambda_4 - \frac{10}{7} \lambda_2^2) u^3 + (\frac{5}{27} \lambda_3^2 - \frac{1}{8} \lambda_4) u.$$  (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 \( P_1 \), \( \lambda_3 \) 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 \( P_{\mathcal{G} > 0}(u) \) \( \simeq \exp(-u^2/2)/(u\sqrt{2\pi}) \). One then obtains the following condition:

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

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

---

Cramér function

More generally, when \( N \) is large, one can write the distribution of the sum of \( N \) IID random variables as:

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

where \( S \) is the so-called Cramér 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:

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

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:

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

with \( S(u) \) given by:

\[
\frac{d}{dz} \log[\hat{P}_1(z)] \bigg|_{z=z^*} = iz^* u + \log[\hat{P}_1(z^*)],
\] (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:

\[
P_1(x) = \Theta(x_1) \alpha e^{-\alpha x_1},
\] (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

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

---

20The above arguments can actually be made fully rigorous, see [Feller].

21We assume that their mean is zero, which can always be achieved through a suitable shift of \( x_1 \).
\( \sigma^2 = \alpha^{-2} \). 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:\(^{22}\)

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

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 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:

\[
\frac{d}{dx} x^{N-1} e^{-\alpha x} \bigg|_{x^*} = 0, \tag{1.80}
\]

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

\[
\log P(x, N) = -K(N-1) - \log m - \frac{\alpha^2(x-x^*)^2}{2(N-1)} + \frac{\alpha^3(x-x^*)^3}{3(N-1)^2} + O(x-x^*)^4, \tag{1.81}
\]

where

\[
K(N) \equiv \log N! + N - N \log N \simeq \frac{1}{2} \log(2\pi N). \tag{1.82}
\]

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 \( \hat{P}_1(x_1) \) 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:

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

which, expanded around \( z = 0 \), gives:

\[
\hat{P}_N^1(k) \simeq 1 - \frac{Nz^2a^2}{2} + \frac{|z|^3a^3}{3} + O(z^4). \tag{1.87}
\]

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 \( N\sigma^2 \):

\[
P(x, N) \simeq \frac{1}{\sqrt{2\pi N\theta}} \exp \left( -\frac{x^2}{2N\theta^2} \right). \tag{1.88}
\]
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:

$$P(x, N) \simeq \frac{2N a^3}{\pi x^4}. \quad (1.89)$$

The above two expressions (1.88) and (1.89) are not incompatible, since they 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:

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

One thus find,

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

(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:

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

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:

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

which tends to zero for large $N$. However, one should notice that as $\mu$ 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 $\mu$.

### 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 $1 < \mu < 2$:

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

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

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

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 $\mu < 2$):

$$N \gg N^* = \lambda_4 \Rightarrow (N a_\mu)^{1/2} \gg \alpha^{-1}. \quad (1.96)$$

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 $\mu$ and tail amplitude $A^\mu \propto a_\mu$ in the region where $x$ is large but still much smaller than $\alpha^{-1}$ (we thus also assume that $\alpha$ 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 $x_{\text{max}} \simeq AN^{1/2}$ (cf. (1.42)). If $x_{\text{max}}$ is very small compared to $\alpha^{-1}$, 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 $N^{1/2}$, as if $\alpha$ was zero. However, as illustrated in Fig. 1.9, this regime ends when $x_{\text{max}}$
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:

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

(1.97)

where \( \sigma \equiv C(0) \). From this expression, it is readily seen that if \( C(\ell) \) decays faster than \( 1/\ell \) for large \( \ell \), the sum over \( \ell \) 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 \( \ell \) 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 Gaussian variables, it is easy to show that the resulting sum is also Gaussian, whatever the value of \( \nu \). Another solvable case is when the \( X_i \)

\[25\] We again assume in the following, without loss of generality that the mean \( m \) is zero.

Figure 1.9: Behaviour of the typical value of \( X \) as a function of \( N \) for TLD variables. When \( N \ll N^* \), \( x \) grows \( N^{2/3} \) (dotted line). When \( N \sim N^* \), \( x \) reaches the value \( \alpha^{-1} \) 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).

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 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,
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 $P_1(x_1), P_2(x_2), \ldots, P_N(x_N)$ 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 $\sigma_k^2$, which is itself a random variable. We shall denote by $\langle \ldots \rangle_k$ the average over the random variable $\sigma_k$, to distinguish it from the notation $\langle \ldots \rangle$ which we have used to describe the average over the probability distribution $P_k$. If $\sigma_k$ varies rapidly, it is impossible to separate the two sources of uncertainty. Thus, the empirical histogram constructed from the series $\{x_1, x_2, \ldots, x_N\}$ leads to an ‘apparent’ distribution $\overline{P}$ which is non-Gaussian even if each individual $P_k$ is Gaussian. Indeed, from:

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

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

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

Since for any random variable one has $\kappa \geq (\overline{\sigma^2})^2$ (the equality being reached only if $\sigma^2$ 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 $\sigma$ 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:

$$\log |\overline{P}(x)| \propto -x \frac{2}{\kappa}.$$  \hfill (1.100)

Since $c < 2 + c$, this asymptotic decay is always much slower than in the Gaussian case, which corresponds to $c \rightarrow \infty$. 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 $\sigma^2$ is distributed as an completely asymmetric Lévy distribution ($\beta = 1$) of exponent equal to $2\mu$. 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 $2\mu$.

If the fluctuations of $\sigma_k$ are themselves correlated, one observes an interesting case of dependence. For example, if $\sigma_k$ is large, $\sigma_{k+1}$ 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 $\epsilon_k$ are IID random variables of zero mean and unit variance, and $\sigma_k$ 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:

$$\overline{\langle x_i x_j \rangle} = \overline{\sigma_i \sigma_j} \langle \epsilon_i \epsilon_j \rangle = \delta_{i,j} \sigma^2.$$  \hfill (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 $X_k$:

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

which indeed has an interesting temporal behaviour: see Section 2.4.\textsuperscript{26} However, even if the correlation function $\sigma_i^2 \sigma_j^2 - \sigma^2 \overline{\sigma}^2$ decreases very slowly 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

\textsuperscript{26}Note that for $i \neq j$ this correlation function can be zero either because $\sigma$ is identically equal to a certain value $\sigma_0$, or because the fluctuations of $\sigma$ are completely uncorrelated from one time to the next.
variable. A way to see this is to compute the average kurtosis of the sum, \( \kappa_N \). As shown in Appendix A, one finds the following result:

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

(1.103)

where \( \kappa_0 \) is the kurtosis of the variable \( \epsilon \), and \( g(\ell) \) the correlation function of the variance, defined as:

\[
\sigma_i^2 \sigma_j^2 - \sigma^2 = \sigma^2 g(|i - j|)
\]

(1.104)

It is interesting to see that for \( N = 1 \), the above formula gives \( \kappa_1 = \kappa_0 + (3 + \kappa_0)g(0) \) with \( \kappa_0 \), 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 \( \ell \), the kurtosis \( \kappa_N \) 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 \( \ell \), one finds that for large \( N \):

\[
\kappa_N \propto \frac{1}{N} \quad \text{for} \quad \nu > 1; \quad \kappa_N \propto \frac{1}{N^\nu} \quad \text{for} \quad \nu < 1.
\]

(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 all elements of the matrix \( H \) 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 \( H \) is a square, \( N \times N \) symmetric matrix. Its eigenvalues are \( \lambda_\alpha \), with \( \alpha = 1, ..., N \). The density of eigenvalues is defined as:

\[
\rho(\lambda) = \frac{1}{N} \sum_{\alpha=1}^{N} \delta(\lambda - \lambda_\alpha),
\]

(1.106)

where \( \delta \) is the Dirac function. We shall also need the so-called ‘resolvent’ \( G(\lambda) \) of the matrix \( H \), defined as:

\[
G_{ij}(\lambda) \equiv \left( \frac{1}{\lambda I - H} \right)_{ij},
\]

(1.107)

where \( I \) is the identity matrix. The trace of \( G(\lambda) \) can be expressed using the eigenvalues of \( H \) as:

\[
\text{Tr} G(\lambda) = \sum_{\alpha=1}^{N} \frac{1}{\lambda - \lambda_\alpha}.
\]

(1.108)

The ‘trick’ that allows one to calculate \( \rho(\lambda) \) in the large \( N \) limit is the following representation of the \( \delta \) function:

\[
\frac{1}{x - i\epsilon} = P P \frac{1}{x} + i\pi \delta(x) \quad (\epsilon \to 0),
\]

(1.109)

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

\[
\rho(\lambda) = \lim_{\epsilon \to 0} \frac{1}{\pi} \text{Im} \left( \text{Tr} G(\lambda - i\epsilon) \right).
\]

(1.110)

Our task is therefore to obtain an expression for the resolvent \( G(\lambda) \). This can be done by establishing a recursion relation, allowing one to compute \( G(\lambda) \) for a matrix \( H \) 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 \( H \)) using the standard formula for matrix inversion:

\[
G_{00}^{N+1}(\lambda) = \frac{\text{minor}(\lambda I - H)_{00}}{\det(\lambda I - H)}. \quad (1.111)
\]

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 \( G_{00}^{N+1}(\lambda) \):

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

(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 $(H_{ij}^2) = \sigma^2/N$. 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 $G_{i0}^{N+1}$, and establish self-consistently that $G_{ij} \sim 1/\sqrt{N}$ for $i \neq j$. On the other hand, due to the diagonal term $\lambda$, $G_{ii}$ 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 $H_{i0} \sim 1/\sqrt{N}$, this term can be neglected compared to $\lambda$. This finally leads to a simplified recursion relation, valid in the limit $N \to \infty$:

$$
\frac{1}{G_{i0}^{N+1}(\lambda)} \simeq \lambda - \sum_{i=1}^{N} H_{ii}^2 G_{i0}^{N}(\lambda). \quad (1.113)
$$

Now, using the CLT, we know that the last sum converges, for large $N$, towards $\sigma^2/N \sum_{i=1}^{N} G_{i0}^{N}(\lambda)$. This result is independent of the precise statistics of the $H_{ii}$, provided their variance is finite. This shows that $G_{i0}$ converges for large $N$ towards a well defined limit $G_\infty$, which obeys the following limit equation:

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

The solution to this second order equation reads:

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

(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 $\lambda$ 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:

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

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 $C$, which have the special property of being positive definite. $C$ can be written as $C = HH^\dagger$, where $H^\dagger$ is the matrix transpose of $H$. In general, $H$ 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 $C$ are simply obtained from those of $H$ by squaring them:

$$
\lambda_C = \lambda_H^2. \quad (1.117)
$$

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

$$
\rho(\lambda_C)d\lambda_C = \rho(\lambda_H)d\lambda_H, \quad (1.118)
$$

which leads to:

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

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:

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

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 $\sim 1/\sqrt{\lambda}$. 

\footnote{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).}

The density of eigenvalues also vanishes above a certain upper edge \( \lambda_{\text{max}} \).

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 \( \lambda_{\text{max}} \) and below \( \lambda_{\text{min}} \), which goes to zero when \( N \) becomes large.\(^{29}\)

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 \( \delta x_i \)'s can be written as \( \sigma_i \epsilon_i \). The \( \sigma_i \)'s are correlated as:

\[
(D_k - D)(D_l - D) = D^2 g(|k - l|) \quad D_k \propto \sigma_k^2. \tag{1.121}
\]

Let us first compute \( \left\langle \left( \sum_{i=1}^{N} \delta x_i \right)^4 \right\rangle \), where \( \langle \ldots \rangle \) means an average over the \( \epsilon_i \)'s and the overline means an average over the \( \sigma_i \)'s. If \( \langle \epsilon_i \rangle = 0 \), and\(^{29}\) one finds:

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

where we have used the definition of \( \kappa_0 \) (the kurtosis of \( \epsilon \)). On the other hand, one must estimate \( \left\langle \sum_{i=1}^{N} \delta x_i^2 \right\rangle^2 \). One finds:

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

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

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

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:**

- **Extreme value statistics:**
2

STATISTICS OF REAL PRICES

Le marché, à son insu, obéit à une loi qui le domine : la loi de la probabilité.¹

(Bachelier, Théorie de la spéculation.)

2.1 Aim of the chapter

The easy access to enormous financial databases, containing thousands of asset time series, sampled at a frequency of minutes or sometimes seconds, allows to investigate in detail the statistical features of the time evolution of financial assets. The description of any kind of data, be it of physical, biological, or financial origin, requires however an interpretation canvas, which is needed to order and give a meaning to the observations. To describe necessarily means to simplify, and even sometimes betray: the aim of any empirical science is to approach reality progressively, through successive improved approximations.

The goal of the present chapter is to present in this spirit the statistical properties of financial time series. We will propose some mathematical modelling, as faithful as possible (though imperfect) of the observed properties of these time series. It is interesting to note that the word ‘modelling’ has two rather different meanings within the scientific community. The first one, often used in applied mathematics, engineering sciences and financial mathematics, means that one represents reality using appropriate mathematical formulas. This is the scope of the present chapter. The second, more common in the physical sciences, is perhaps more ambitious: it aims at finding a set of plausible causes which are needed to explain the observed phenomena, and therefore, ultimately, to justify the chosen mathematical description. We will however only discuss in a cursory way the ‘microscopic’ mechanisms of price formation and evolution, of adaptive traders’ strategies, herding behaviour between

¹The market, without knowing it, obeys to a law which overwhelms it: the law of probability.
traders, feedback of price variations onto themselves, etc., which are certainly at the origin of the interesting statistics that we shall report below. We feel that this aspect of the problem is still in its infancy, and will evolve rapidly in the coming years. We briefly mention, at the end of this chapter, two simple models of herding and feedback, and give references of several very recent articles.

We shall describe several types of markets:

- Very liquid, ‘mature’ markets (of which we take three examples: a U.S. stock index (S&P 500), an exchange rate (DEM/$), and a long term interest rate index (the German BUND);
- Very volatile markets, such as emerging markets like the Mexican peso;
- Volatility markets: through option markets, the volatility of an asset (which is empirically found to be time dependent) can be seen as a price which is quoted on markets (see Chapter 4);
- Interest rate markets, which give fluctuating prices to loans of different maturities, between which special type of correlations must however exist.

We will voluntarily limit our study to fluctuations taking place on rather short time scales (typically from minutes to months). For longer time scales, the available data is in general too short to be meaningful. From a fundamental point of view, the influence of the average return is negligible for short time scales, but become crucial on long time scales. Typically, a stock varies by several % within a day, but its average return is – say – 10% per year, or 0.04% per day. Now, the ‘average return’ of a financial asset appears to be unstable in time: the past return of a stock is seldom a good indicator of future returns. Financial time series are intrinsically non-stationary: new financial products appear and influence the markets, trading techniques evolve with time, as does the number of participants and their access to the markets, etc. This means that taking very long historical data set to describe the long term statistics of markets is a priori not justified. We will thus avoid this difficult (albeit important) subject of long time scales.

The simplified model that we will present in this chapter, and that will be the starting point of the theory of portfolios and options discussed in later chapters, can be summarised as follows. The variation of price of the asset $X$ between time $t = 0$ and $t = T$ can be decomposed as:

$$x(T) = x_0 + \sum_{k=0}^{N-1} \delta x_k, \quad N = \frac{T}{\tau},$$

(2.1)

where,

- In a first approximation, and for $T$ not too large, the price increments $\delta x_k$ are random variables which are (i) independent as soon as $\tau$ is larger than a few tens of minutes (on liquid markets) and (ii) identically distributed, according to a TLD (TLD) $P_{1}(\delta x) = L_{1}(\delta x)$ with a parameter $\mu$ approximatively equal to $3/2$, for all markets.\footnote{Alternatively, a description in terms of Student distributions is often found to be of comparable quality, with a tail exponent $\mu \sim 3 - 5$ for the SP500, for example.}

The results of Chapter 1 concerning sums of random variables, and the convergence towards the Gaussian distribution, allows one to understand the observed ‘narrowing’ of the tails as the time interval $T$ increases.

- A refined analysis however reveals important systematic deviations from this simple model. In particular, the kurtosis of the distribution of $x(T) - x_0$ decreases more slowly than $1/N$, as it should if the increments $\delta x_k$ were IID random variables. This suggests a certain form of temporal dependence, of the type discussed in Section 1.7.2. The volatility (or the variance) of the price increments $\delta x$ is actually itself time dependent: this is the so-called ‘heteroskedasticity’ phenomenon. As we shall see below, periods of high volatility tend to persist over time, thereby creating long-range higher order correlations in the price increments. On long time scales, one also observes a systematic dependence of the variance of the price increments on the price $x$ itself. In the case where the RMS of the variables $\delta x$ grows linearly with $x$, the model becomes multiplicative, in the sense that one can write:

$$x(T) = x_0(1 + \prod_{k=0}^{N-1} \eta_k), \quad N = \frac{T}{\tau},$$

(2.2)

where the returns $\eta_k$ have a fixed variance. This model is actually more commonly used in the financial literature. We will show that reality must be described by an intermediate model, which interpolates between a purely additive model (2.1), and a multiplicative model (2.2).

**Studied assets**

The chosen stock index is the futures contract on the Standard and Poor’s 500 (S&P 500) U.S. stock index, traded on the Chicago Mercantile Exchange.
tile Exchange (CME). During the time period chosen (from November 1991 to February 1995), the index rose from 375 to 480 points (Fig. 2.1 (a)). Qualitatively, all the conclusions reached on this period of time are more generally valid, although the value of some parameters (such as the volatility) can change significantly from one period to the next.

The exchange rate is the U.S. dollar ($) against the German mark (DEM), which is the most active exchange rate market in the world. During the analysed period, the mark varied between 58 and 75 cents (Fig. 2.1 (b)). Since the interbank settlement prices are not available, we have defined the price as the average between the bid and the ask prices.3

Finally, the chosen interest rate index is the futures contract on long term German bonds (BUND), quoted on the London International Financial Futures and Options Exchange (LIFFE). It is typically varying between 85 and 100 points (Fig. 2.1 (c)).

The indices S&P 500 and BUND that we have studied are thus actually futures contracts (cf. Section 4.2). The fluctuations of futures prices follow in general those of the underlying contract and it is reasonable to identify the statistical properties of these two objects. Futures contracts exist with several fixed maturity dates. We have always chosen the most liquid maturity and suppressed the artificial difference of prices when one changes from one maturity to the next (roll). We have also neglected the weak dependence of the futures contracts on the short time interest rate (see Section 4.2): this trend is completely masked by the fluctuations of the underlying contract itself.

2.2 Second order statistics

2.2.1 Variance, volatility and the additive-multiplicative crossover

In all that follows, the notation $\delta x$ represents the difference of value of the asset $X$ between two instants separated by a time interval $\tau$:

$$\delta x_k = x(t + \tau) - x(t) \quad t = k\tau. \quad (2.3)$$

In the whole modern financial literature, it is postulated that the relevant variable is not the increment $\delta x$ itself, but rather the return $\eta = \delta x/x$. It is therefore interesting to study empirically the variance of $\delta x$, conditioned to a certain value of the price $x$ itself, which we shall denote $\langle \delta x^2 \rangle_x$. If the return $\eta$ is the natural random variable, one should observe that $\sqrt{\langle \delta x^2 \rangle_x} = \sigma_1 x$, where $\sigma_1$ is constant (and equal to the RMS

3There is, on all financial markets, a difference between the bid price and the ask price for a certain asset at a given instant of time. The difference between the two is called the ‘bid/ask spread’. The more liquid a market, the smaller the average spread.
2.2 Second order statistics

Figure 2.2: RMS of the increments $\delta x$, conditioned to a certain value of the price $x$, as a function of $x$, for the three chosen assets. For the chosen period, only the exchange rate DEM/$ conform to the idea of a multiplicative model: the straight line corresponds to the best fit $\langle (\delta x^2) \rangle_x^{1/2} = \sigma x$. The adequacy of the multiplicative in this case is related to the symmetry $$/dem \rightarrow dem/$.

Figure 2.3: RMS of the increments $\delta x$, conditioned to a certain value of the price $x$, as a function of $x$, for the S&P 500 for the 1985-1998 time period.

Figure 2.4: RMS of the increments $\delta x$, conditioned to a certain value of the price $x$, as a function of $x$, for the CAC 40 index for the 1991-1995 period; it is quite clear that during that time period $\langle (\delta x^2) \rangle_x$ was almost independent of $x$. 
of $\eta$). Now, in many instances (Figs. 2.2 and 2.4), one rather finds that $\sqrt{\langle \delta x^2 \rangle} |x|$ is independent of $x$, apart from the case of exchange rates between comparable currencies. The case of the CAC40 is particularly interesting, since during the period 1991-1995, the index went from 1400 to 2100, leaving the absolute volatility nearly constant (if anything, it is seen to decrease with $x$!)

On longer time scales, however, or when the price $x$ rises substantially, the $\text{rms}$ of $\delta x$ increases significantly, as to become proportional to $x$ (Fig. 2.3). A way to model this crossover from an additive to a multiplicative behaviour is to postulate that the $\text{rms}$ of the increments progressively (over a time scale $T_{\sigma}$) adapt to the changes of price of $x$. Schematically, for $T < T_{\sigma}$, the prices behave additively, while for $T > T_{\sigma}$, multiplicative effects start playing a significant role:\footnote{In the additive regime, where the variance of the increments can be taken as a constant, we shall write $\langle \delta x^2 \rangle = \sigma^2 x_0^2 = D T$.}

\begin{equation}
\langle (x(T) - x_0)^2 \rangle = D T \quad (T \ll T_{\sigma});
\end{equation}
\begin{equation}
\left\langle \log^2 \left( \frac{x(T)}{x_0} \right) \right\rangle = \sigma^2 T \quad (T \gg T_{\sigma}). \tag{2.4}
\end{equation}

On liquid markets, this time scale is on the order of months.

2.2.2 Autocorrelation and power spectrum

The simplest quantity, commonly used to measure the correlations between price increments, is the temporal two-point correlation function $C_{k\ell}^\tau$, defined as:\footnote{In principle, one should subtract the average value $\langle \delta x \rangle = m_{\tau} = m_1$ from $\delta x$. However, if $\tau$ is small (for example equal to a day), $m_{\tau}$ is completely negligible compared to $\sqrt{D \tau}$.}

\begin{equation}
C_{k\ell}^\tau = \frac{1}{D \tau} \langle \delta x_k \delta x_\ell \rangle. \tag{2.5}
\end{equation}

Figure 2.5 shows this correlation function for the three chosen assets, and for $\tau = 5$ minutes. If the increments are uncorrelated, the correlation function $C_{k\ell}^\tau$ should be equal to zero for $k \neq l$, with a $\text{rms}$ equal to $\sigma = 1/\sqrt{N}$, where $N$ is the number of independent points used in the computation. Figure 2.5 also shows the 3$\sigma$ error bars. We conclude that beyond 30 minutes, the two-point correlation function cannot be distinguished from zero. On less liquid markets, however, this correlation time is longer. On the U.S. stock market, for example, this correlation time has significantly decreased between the 60's and the 90's.

On very short time scales, however, weak but significant correlations do exist. These correlations are however small to allow profit making:

![Figure 2.5: Correlation function $C_{k\ell}^\tau$ for the three chosen assets, as a function of the time difference $|k - l|\tau$, and for $\tau = 5$ minutes. Up to 30 minutes, some weak but significant correlations do exist (of amplitude $\sim 0.05$). Beyond 30 minutes, however, the two-point correlations are not statistically significant.](image)
2.2 Second order statistics

Let us briefly mention another equivalent way of presenting the same results, using the so-called power spectrum, defined as:

\[ S(\omega) = \frac{1}{N} \left( \sum_{k,\ell=1}^{N} \delta x_k \delta x_\ell e^{i\omega(k-\ell)} \right). \]  

(2.6)

The case of uncorrelated increments leads to a flat power spectrum, \( S(\omega) = S_0 \). Figure 2.7 shows the power spectrum of the $/DEM time series, where no significant structure appears.

the potential return is smaller than the transaction costs involved for such a high frequency trading strategy, even for the operators having direct access to the markets (cf. 4.1.2). Conversely, if the transaction costs are high, one may expect significant correlations to exist on longer time scales.

We have performed the same analysis for the daily increments of the three chosen assets (\( \tau = 1 \text{ day} \)). Figure 2.6 reveals that the correlation function is always within 3\( \sigma \) of zero, confirming that the daily increments are not significantly correlated.

Power spectrum

Figure 2.7: Power spectrum \( S(\omega) \) of the time series $/DEM, as a function of the frequency \( \omega \). The spectrum is flat: for this reason one often speak of white noise, where all the frequencies are represented with equal weights. This corresponds to uncorrelated increments.
2.3 Temporal evolution of fluctuations

2.3.1 Temporal evolution of probability distributions

The results of the previous section are compatible with the simplest scenario where the price increments $\delta x_k$ are, beyond a certain correlation time, independent random variables. A much finer test of this assumption consists in studying directly the probability distributions of the price increments $x_N - x_0 = \sum_{k=0}^{N-1} \delta x_k$ on different time scales $N = T/\tau$. If the increments are independent, then the distributions on different time scales can be obtained from the one pertaining to the elementary time scale $\tau$ (chosen to be larger than the correlation time). More precisely (see Section 1.5.1), one should have $P(x, N) = \left[ P_1(\delta x_1) \right]^{*N}$.

The elementary distribution $P_1$

The elementary cumulative probability distribution $P_{1>}(\delta x)$ is represented in Figs. 2.8, 2.9 and 2.10. One should notice that the tail of the distribution is broad, in any case much broader than a Gaussian. A fit using a truncated Lévy distribution of index $\mu = 3/2$, as given by Eq. (1.23), is quite satisfying.\(^6\) The corresponding parameters $A$ and $\alpha$ are given in Tab. 2.1 (For $\mu = 3/2$, the relation between $A$ and $a_{3/2}$ reads: $a_{3/2} = 2\sqrt{2}\pi A^{3/2}/3$). Alternatively, as shown in Fig. 1.5, a fit using a Student distribution would also be acceptable.

We have chosen to fix the value of $\mu$ to $3/2$. This reduces the number of adjustable parameters, and is guided by the following observations:

- A large number of empirical studies on the use of Lévy distributions to fit the financial market fluctuations report values of $\mu$ in the range $1.6 - 1.8$. However, in the absence of truncation (i.e. with $\alpha = 0$), the fit overestimates the tails of the distribution. Choosing a higher value of $\mu$ partly corrects for this effect, since it leads to a thinner tail.

- If the exponent $\mu$ is left as a free parameter, it is in many cases found to be in the range $1.4 - 1.6$, although sometimes smaller, as in the case of the $$/DEM ($\mu \simeq 1.2$).

- The particular value $\mu = 3/2$ has a simple theoretical interpretation, which we will briefly present in Section 2.8.

\(^6\)A more refined study of the tails actually reveals the existence of a small asymmetry, which we neglect here. Therefore, the skewness $\lambda_3$ is taken to be zero.

Figure 2.8: Elementary cumulative distribution $P_{1>}(\delta x)$ (for $\delta x > 0$) and $P_{1<}(\delta x)$ (for $\delta x < 0$), for the S&P 500, with $\tau = 15$ minutes. The thick line corresponds to the best fit using a symmetric TLD $L_\mu^{(0)}$, of index $\mu = 3/2$. We have also shown on the same graph the values of the parameters $A$ and $\alpha^{-1}$ as obtained by the fit.
Figure 2.9: Elementary cumulative distribution for the DEM/$, for $\tau = 15$ minutes, and best fit using a symmetric TLD $L^{(i)}_\mu$, of index $\mu = 3/2$. In this case, it is rather $100\delta x/x$ that has been considered. The fit is not very good, and would have been better with a smaller value of $\mu \sim 1.2$. This increases the weight of very small variations.

Figure 2.10: Elementary cumulative distribution for the BUND, for $\tau = 15$ minutes, and best fit using a symmetric TLD $L^{(i)}_\mu$, of index $\mu = 3/2$. 
Table 2.1: Value of the parameters $A$ and $\alpha^{-1}$, as obtained by fitting the data with a symmetric tld $L^{(i)}_{\mu}$, of index $\mu = 3/2$. Note that both $A$ and $\alpha^{-1}$ have the dimension of a price variation $\delta x$, and therefore directly characterise the nature of the statistical fluctuations. The other columns compare the RMS and the kurtosis of the fluctuations, as directly measured on the data, or via the formulas (1.94, 1.95).

<table>
<thead>
<tr>
<th>Asset</th>
<th>Variance $\sigma^2_i$</th>
<th>Kurtosis $\kappa_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A$</td>
<td>$\alpha^{-1}$</td>
</tr>
<tr>
<td>SP500</td>
<td>0.22</td>
<td>2.21</td>
</tr>
<tr>
<td>BUND</td>
<td>0.0091</td>
<td>0.275</td>
</tr>
<tr>
<td>$$/DEM$$</td>
<td>0.0447</td>
<td>0.96</td>
</tr>
</tbody>
</table>

In order to characterise a probability distribution using empirical data, it is always better to work with the cumulative distribution function rather than the distribution density. To obtain the latter, one indeed has to choose a certain width for the bins in order to construct frequency histograms, or to smooth the data using, for example, a Gaussian with a certain width. Even when this width is carefully chosen, part of the information is lost. It is furthermore difficult to characterise the tails of the distribution, corresponding to rare events, since most bins in this region are empty. On the other hand, the construction of the cumulative distribution does not require to choose a bin width. The trick is to order the observed data according to their rank, for example in decreasing order. The value $x_k$ of the $k^{th}$ variable (out of $N$) is then such that:

$$P_>(x_k) = \frac{k}{N+1}. \quad (2.7)$$

This result comes from the following observation: if one draws an $N + 1$ random variable from the same distribution, there is an a priori equal probability $1/N + 1$ that if falls within any of the $N + 1$ intervals defined by the previously drawn variables. The probability that if falls above the $k^{th}$ one, $x_k$ is therefore equal to the number of intervals beyond $x_k$, which is equal to $k$, times $1/N + 1$. This is also equal, by definition, to $P_>(x_k)$. (See also the discussion in Section 1.4, and Eq. (1.45)). Since the rare events part of the distribution is a particular interest, it is convenient to choose a logarithmic scale for the probabilities. Furthermore, in order to check visually the symmetry of the probability distributions, we have systematically used $P_<(\delta x)$ for the negative increments, and $P_>(\delta x)$ for positive $\delta x$.

Maximum likelihood

Suppose that one observes a series of $N$ realisations of the random iid variable $X$, $\{x_1, x_2, \ldots, x_N\}$, drawn with an unknown distribution that one would like to parameterise, for simplicity, by a single parameter $\mu$. If $P_\mu(x)$ denotes the corresponding probability distribution, the a priori probability to observe the particular series $\{x_1, x_2, \ldots, x_N\}$ is proportional to:

$$P_\mu(x_1)P_\mu(x_2)\ldots P_\mu(x_N). \quad (2.8)$$

The most likely value $\mu^*$ of $\mu$ is such that this a priori probability is maximised. Taking for example $P_\mu(x)$ to be a power-law distribution:

$$P_\mu(x) = \frac{\mu^{\mu^*}}{x^{1+\mu}} \quad x > x_0, \quad (2.9)$$

(with $x_0$ known), one has:

$$P_\mu(x_1)P_\mu(x_2)\ldots P_\mu(x_N) \propto e^{N \log \mu + N \mu \log x_0 - (1+\mu) \sum_{i=1}^N \log x_i}. \quad (2.10)$$

The equation fixing $\mu^*$ is thus, in this case:

$$N \mu^* + N \log x_0 - \sum_{i=1}^N \log x_i = 0 \Rightarrow \mu^* = \frac{N}{\sum_{i=1}^N \log (x_i^x_0)}. \quad (2.11)$$

This method can be generalised to several parameters. In the above example, if $x_0$ is unknown, its most likely value is simply given by:

$$x_0 = \text{min}\{x_1, x_2, \ldots, x_N\}.$$

Convolutions

The parameterisation of $P_\mu(\delta x)$ as a tld allows one to reconstruct the distribution of price increments for all time intervals $T = N\tau$, if one assumes that the increments are iid random variables. As discussed in Chapter 1, one then has $P(\delta x, N) = [P_\mu(\delta x_1)]^N$. Figure 2.11 shows the cumulative distribution for $T = 1$ hour, 1 day and 5 days, reconstructed from the one at 15 minutes, according to the simple iid hypothesis. The symbols show empirical data corresponding to the same time intervals. The agreement is acceptable; one notices in particular the progressive deformation of $P(\delta x, N)$ towards a Gaussian for large $N$. The evolution of the variance and of the kurtosis as a function of $N$ is given in Tab. 2.2, and compared to the results that one would observe if the simple convolution rule was obeyed, i.e. $\sigma^2_N = N\sigma^2_1$ and $\kappa_N = \kappa_1/N$. For these liquid assets, the time scale $T^* = \kappa_1\tau$ which sets the convergence towards the Gaussian is on the order of days. However, it is clear from Tab. 2.2 that this convergence is slower than it ought to be: $\kappa_N$ decreases much more slowly than the $1/N$ behaviour predicted by an iid hypothesis. A
Table 2.2: Variance and kurtosis of the distributions $P(\delta x, N)$ measured or computed from the variance and kurtosis at time scale $\tau$ by assuming a simple convolution rule, leading to $\sigma^2_N = N\sigma^2_1$ and $\kappa_N = \kappa_1/N$. The kurtosis at scale $N$ is systematically too large – cf. Section 2.4. We have used $N = 4$ for $T = 1$ hour, $N = 28$ for $T = 1$ day and $N = 140$ for $T = 5$ days.

<table>
<thead>
<tr>
<th>Asset</th>
<th>Variance $\sigma^2_N$</th>
<th>Kurtosis $\kappa_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP500 (T=1 hour)</td>
<td>1.06</td>
<td>1.12</td>
</tr>
<tr>
<td>BUND (T=1 hour)</td>
<td>$9.49 \times 10^{-3}$</td>
<td>$9.68 \times 10^{-3}$</td>
</tr>
<tr>
<td>$/$DEM (T=1 hour)</td>
<td>$6.03 \times 10^{-2}$</td>
<td>$6.56 \times 10^{-2}$</td>
</tr>
<tr>
<td>SP500 (T=1 day)</td>
<td>7.97</td>
<td>7.84</td>
</tr>
<tr>
<td>BUND (T=1 day)</td>
<td>$6.80 \times 10^{-2}$</td>
<td>$6.76 \times 10^{-2}$</td>
</tr>
<tr>
<td>$/$DEM (T=1 day)</td>
<td>0.477</td>
<td>0.459</td>
</tr>
<tr>
<td>SP500 (T=5 days)</td>
<td>38.6</td>
<td>39.20</td>
</tr>
<tr>
<td>BUND (T=5 days)</td>
<td>0.341</td>
<td>0.338</td>
</tr>
<tr>
<td>$/$DEM (T=5 days)</td>
<td>2.52</td>
<td>2.30</td>
</tr>
</tbody>
</table>

Closer look at Fig. 2.11 also reveals systematic deviations: for example the tails at 5 days are distinctively fatter than they should be.

Tails, what tails?

The asymptotic tails of the distributions $P(\delta x, N)$ are approximately exponential for all $N$. This is particularly clear for $T = N \tau = 1$ day, as illustrated in Fig. 2.12 in a semi-logarithmic plot. However, as mentioned in Section 1.3.4 and in the above paragraph, the distribution of price changes can also be satisfactorily fitted using Student distributions (which have power-law tails) with rather high exponents. In some cases, for example the distribution of losses of the S&P 500 (Fig. 2.12), one sees a slight upward bend in the plot of $P(x)$ versus $x$ in a linear-log plot. This indeed suggests that the decay could be slower than exponential. Many authors have proposed that the tails of the distribution of price changes is a power-law with an exponent $\mu$ in the range 3 to 5, or a stretched exponential $\exp[-|\delta x|^c] c < 1$.\(^7\) For example, the most

---

likely value of $\mu$ using a Student distribution to fit the daily variations of the S&P in the period 1991-95 is $\mu = 5$. Even if it is rather hard to distinguish empirically between an exponential and a high power-law, this question is very important theoretically. In particular, the existence of a finite kurtosis requires $\mu$ to be larger than 4. As far as applications to risk control, for example, are concerned, the difference between the extrapolated values of the risk using an exponential or a high power-law fit of the tails of the distribution is significant, but not dramatic. For example, fitting the tail of an exponential distribution by a power-law, using 1000 days, leads to an effective exponent $\mu \approx 4$. An extrapolation to the most probable drop in 10000 days overestimates the true figure by 30%. In any case, the amplitude of very large crashes observed in the century are beyond any reasonable extrapolation of the tails, whether one uses an exponential or a high power-law. The a priori probability of observing a -22% drop in a single day, as happened on the New-York Stock Exchange in October 1987, is found in any case to be much smaller than $10^{-4}$ per day, that is, once every 40 years. This suggests that major crashes are governed by a specific amplification mechanism, which drives these events outside the scope of a purely statistical analysis, and require a specific theoretical description.

2.3.2 Multiscaling – Hurst exponent (*)

The fact that the autocorrelation function is zero beyond a certain time scale $\tau$ implies that the quantity $\langle [x_N - x_0]^2 \rangle$ grows as $D\tau N$. However, measuring the temporal fluctuations using solely this quantity can be misleading. Take for example the case where the price increments $\delta x_k$ are independent, but distributed according to a TLD of index $\mu < 2$. As we have explained in the Section 1.6.5, the sum $x_N - x_0 = \sum_{k=0}^{N-1} \delta x_k$ behaves, as long as $N \ll N^* = \kappa$, as a pure ‘Lévy’ sum, for which the truncation is inessential. Its order of magnitude is therefore $x_N - x_0 \sim A N^{\frac{1}{\mu}}$, where $A^\mu$ is the tail parameter of the Lévy distribution. However, the second moment $\langle [x_N - x_0]^2 \rangle = D\tau N$, is dominated by extreme fluctuations, and is therefore governed by the existence of the exponential truncation which gives a finite value to $D\tau$ proportional to $A^\mu \alpha^{\nu-2}$. One can check that as long as $N \ll N^*$, one has $\sqrt{D\tau N} \gg A N^{\frac{1}{2}}$. This means that in this case, the second moment overestimates the amplitude of probable fluctuations. One can generalise the above result to the $q^{th}$ moment of the price difference, $\langle [x_N - x_0]^q \rangle$. If $q > \mu$, one


*On this point, see [Bouchaud-Cont].
finds that all moments grow like $N$ in the regime $N \ll N^*$, and like $N^{q/\mu}$ if $q < \mu$. This is to be contrasted with the sum of Gaussian variables, where $\langle [x_N - x_0]^q \rangle$ grows as $N^{q/2}$ for all $q > 0$. More generally, one can define an exponent $\zeta_q$ as $\langle [x_N - x_0]^q \rangle \propto N^{\zeta_q \rho}$ if $q < \mu$ is not constant with $q$, one speaks of multiscaling. It is not always easy to distinguish true multiscaling from apparent multiscaling, induced by crossover or finite size effects. For example, in the case where one sums uncorrelated random variables with a long range correlation in the variance, one finds that the kurtosis decays slowly, as $\kappa_N \propto N^{-\nu}$, where $\nu < 1$ is the exponent governing the decay of the correlations. This means that the fourth moment of the difference $x_N - x_0$ behaves as:

$$\langle [x_N - x_0]^4 \rangle = (D\tau N)^2 [3 + \kappa_N] \sim N^2 + N^{2-\nu}.$$

If $\nu$ is small, one can fit the above expression, over a finite range of $N$, using an effective exponent $\zeta_4 < 2$, suggesting multiscaling. This is certainly a possibility that one should keep in mind, in particular analysing financial time series (see [Mandelbrot, 1998]).

Another interesting way to characterise the temporal development of the fluctuations is to study, as suggested by Hurst, the average distance between the ‘high’ and the ‘low’ in a window of size $t = n\tau$:

$$H(n) = \langle \max_{k=\ell+1}^{\ell+n}(x_k) - \min_{k=\ell+1}^{\ell+n}(x_k) \rangle.$$

The Hurst exponent $H$ is defined from $H(n) \propto n^H$. In the case where the increments $\delta x_k$ are Gaussian, one finds that $H \equiv 1/2$ (for large $n$). In the case of a TLD of index $1 < \mu < 2$, one finds:

$$H(n) \propto A n^{\frac{\mu}{2}} \quad (n \ll N^*) \quad H(n) \propto \sqrt{D\tau n} \quad (n \gg N^*).$$

The Hurst exponent therefore evolves from an anomalously high value $H = 1/\mu$ to the ‘normal’ value $H = 1/2$ as $n$ increases. Figure 2.13 shows the Hurst function $H(n)$ for the three liquid assets studied here. One clearly sees that the ‘local’ exponent $H$ slowly decreases from a high value ($\sim 0.7$, quite close to $1/\mu = 2/3$) at small times, to $H \approx 1/2$ at long times.

### 2.4 Anomalous kurtosis and scale fluctuations

*For in a minute there are many days*  
(Shakespeare, Romeo and Juliet.)

As mentioned above, one sees in Fig. 2.11 that $P(\delta x, N)$ systematically deviates from $[P_1(\delta x_0)]^N$. In particular, the tails of $P(\delta x, N)$ are...
2.4 Anomalous kurtosis and scale fluctuations

Anomalously 'fat'. Equivalently, the kurtosis $\kappa_N$ of $P(\delta x, N)$ is higher than $\kappa_1/N$, as one can see from Fig. 2.14, where $\kappa_N$ is plotted as a function of $N$ in log-log coordinates.

Correspondingly, more complex correlations functions, such as that of the squares of $\delta x_k$, reveal a non trivial behaviour. An interesting quantity to consider is the amplitude of the fluctuations, averaged over one day, defined as:

$$\gamma = \frac{1}{N_d} \sum_{k=1}^{N_d} |\delta x_k|,$$

where $\delta x_k$ is the five minutes increments, and $N_d$ is the number of five minutes interval within a day. This quantity is clearly strongly correlated in time (Figs. 2.15 and 2.16): the periods of strong volatility persist far beyond the day time scale.

A simple way to account for these effects is to assume that the elementary distribution $P_1$ also depends of time. One actually observes that the level of activity on a market (measured by the volume of transactions) on a given time interval can vary rather strongly with time. It is reasonable to think that the scale of the fluctuations $\gamma$ of the price depends directly on the frequency and volume of the transactions. A simple hypothesis is that apart from a change of this level of activity, the mechanisms leading to a change of price are the same, and therefore that the fluctuations have the same distributions, up to a change of scale. More precisely, we shall assume that the distribution of price changes is such that:

$$P_{1k}(\delta x_k) = \frac{1}{\gamma_k} P_{10} \left( \frac{\delta x_k}{\gamma_k} \right),$$

where $P_{10}(u)$ is a certain distribution normalised to 1 and independent of $k$. The factor $\gamma_k$ represents the scale of the fluctuations: one can define $\gamma_k$ and $P_{10}$ such that $\int du |u| P_{10}(u) \equiv 1$. The variance $D_k \tau$ is then proportional to $\gamma_k^2$.

In the case where $P_{10}$ is Gaussian, the model defined by (2.16) is known in the literature as a ‘stochastic volatility’ model. Model (2.16) is however more general since $P_{10}$ is a priori arbitrary.

If one assumes that the random variables $\delta x_k/\gamma_k$ are independent and of zero mean, one can show (see Section 1.7.2 and Appendix A) that the average kurtosis of the distribution $P(\delta x, N)$ is given, for $N \geq 1$, by:

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

where $\kappa_0$ is the kurtosis of the distribution $P_{10}$ defined above (cf. (2.16)).

Figure 2.14: Kurtosis $\kappa_N$ at scale $N$ as a function of $N$, for the BUND. In this case, the elementary time scale $\tau$ is 30 minutes. If the iid assumption was true, one should find $\kappa_N = \kappa_1/N$. The straight line has a slope of $-0.43$, which means that the decay of the kurtosis $\kappa_N$ is much smaller, as $\simeq 20/N^{0.43}$. 

and \( g \) the correlation function of the variance \( D_k \) (or of the \( \gamma_k^2 \)):

\[
(D_k - \overline{D})(D_\ell - \overline{D}) = \overline{D}^2 g(|\ell - k|).
\]

(2.18)

The overline means that one should average over the fluctuations of the \( D_k \). It is interesting to notice that even in the absence of ‘bare’ kurtosis (\( \kappa_0 = 0 \)), volatility fluctuations are enough to induce a non zero kurtosis \( \kappa_1 = \kappa_0 + (3 + \kappa_0)g(0) \).

The empirical data on the kurtosis are well accounted for using the above formula, with the choice \( g(\ell) \propto \ell^{-\nu} \), with \( \nu = 0.43 \) in the case of the Bund. This choice for \( g(\ell) \) is also in qualitative agreement with the decay of the correlations of the \( \gamma \)'s (Fig. 2.16). However, a fit of the data using for \( g(\ell) \) the sum of two exponentials \( e^{-\ell/\tau_1} + e^{-\ell/\tau_2} \) is also acceptable. One finds two rather different time scales: \( \tau_1 \) is shorter than a day, and a long correlation time \( \tau_2 \) of a few tens of days.

One can thus quite clearly see that the scale of the fluctuations, that the market calls the volatility, changes with time, with a rather long persistence time scale. This slow evolution of the volatility in turn leads to an anomalous decay of the kurtosis \( \kappa_N \) as a function of \( N \). As we shall see in Section 4.3.4, this has direct consequences for the dynamics of the volatility smile observed on option markets.
2.5 Volatile markets and volatility markets

We have considered, up to now, very liquid markets, where extreme price fluctuations are rather rare. On less liquid/less mature markets, the probability of extreme moves is much larger. The case of short term interest rates is also interesting, since the evolution of – say – the three month rate is directly affected by the decision of central banks to increase or to decrease the day to day rate. As discussed further in Section 2.6 below, this leads to a rather high kurtosis, related to the fact that the short rate often does not change at all, but sometimes changes a lot. The kurtosis of the U.S. three month rate is on the order of 10 for daily moves (Fig. 2.17). Emerging markets (such as South America or Eastern Europe markets) are obviously even wilder. The example of the Mexican peso is interesting, because the cumulative distribution of the daily changes of the rate peso/USD reveals a power-law tails, with no obvious truncation, with an exponent \( \mu = 1.5 \) (Fig. 2.18). This data set corresponds to the years 92-94, just before the crash of the peso (December 1994). A similar value of \( \mu \) has also been observed, for example, on the fluctuations of the Budapest Stock Exchange.\(^9\)

Another interesting quantity is the volatility itself that varies with time, as emphasized above. The price of options reflect quite accurately the value of the historical volatility in a recent past (see Section 4.3.4). Therefore, the volatility can be considered as a special type of asset, which one can study as such. We shall define as above the volatility \( \gamma_k \) as the average over a day of the absolute value of the five minutes price increments. The autocorrelation function of the \( \gamma \)'s is shown in Fig. 2.16; it is found to decrease slowly, perhaps as a power-law with an exponent \( \nu \) in the range 0.1 to 0.5 (Fig. 2.16).\(^{10}\) The distribution of the measured volatility \( \gamma \) is shown in Fig. 2.19 for the S&P, but other assets lead to similar curves. This distribution decreases slowly for large \( \gamma \)'s, again as an exponential or a high power-law. Several functional forms have been suggested, such as a log-normal distribution, or an inverse Gamma distribution (see Section 2.9 for a specific model for this behaviour). However, one must keep in mind that the quantity \( \gamma \) is only an approximation for the ‘true’ volatility. The distribution shown in Fig. 2.19 is therefore the convolution of the true distribution with a measurement error distribution.

\(^{10}\)On this point, see, e.g. [Arneodo], and Y. Liu et al., “The statistical properties of volatility of prices fluctuations”, available on http://xxx.lanl.gov/cond-mat/9903369.
2.6 Statistical analysis of the Forward Rate Curve (*)

The case of the interest rate curve is particularly complex and interesting, since it is not the random motion of a point, but rather the consistent history of a whole curve (corresponding to different loan maturities) which is at stake. The need for a consistent description of the whole interest rate curve is furthermore enhanced by the rapid development of interest rate derivatives (options, swaps, options on swaps, etc.) [Hull].

Present models of the interest rate curve fall into two categories: the first one is the Vasicek model and its variants, which focuses on the dynamics of the short term interest rate, from which the whole curve is reconstructed.\(^\text{11}\) The second one, initiated by Heath, Jarrow and Morton takes the full forward rate curve as dynamical variables, driven by (one or several) continuous-time Brownian motion, multiplied by a maturity dependent scale factor. Most models are however primarily motivated by their mathematical tractability rather than by their ability to describe the data. For example, the fluctuations are often assumed to be Gaussian, thereby neglecting ‘fat tail’ effects.

\(^{11}\)For a compilation of the most important theoretical papers on the interest rate curve, see: [Hughston].
Our aim in this section is not to discuss these models in any detail, but rather to present an empirical study of the forward rate curve (FRC), where we isolate several important qualitative features that a good model should be asked to reproduce.\footnote{This section is based on the following papers: J.-P. Bouchaud, N. Sagna, R. Cont, N. ElKaroui, M. Potters, “Phenomenology of the interest rate curve,” \textit{RISK}, June 1998, and to appear in \textit{Applied Mathematical Finance}.} Some intuitive arguments are proposed to relate the series of observations reported below.

### 2.6.1 Presentation of the data and notations

The forward interest rate curve (FRC) at time $t$ is fully specified by the collection of all forward rates $f(t, \theta)$, for different maturities $\theta$. It allows for example to calculate the price $B(t, \theta)$ at time $t$ of a (so-called ‘zero-coupon’) bond, which by definition pays 1 at time $t + \theta$. The forward rates are such that they compound to give $B(t, \theta)$:

$$B(t, \theta) = \exp\left(- \int_0^\theta du f(t, u)\right),$$  

(2.19)

$r(t) = f(t, \theta = 0)$ is called the ‘spot rate’. Note that in the following $\theta$ is always a time difference; the maturity date $T$ is $t + \theta$.

Our study is based on a data set of daily prices of Eurodollar futures contracts on interest rates.\footnote{In principle forward contracts and future contracts are not strictly identical – they have different margin requirements – and one may expect slight differences, which we shall neglect in the following.} The interest rate underlying the Eurodollar futures contract is a 90-day rate, earned on dollars deposited in a bank outside the U.S. by another bank. The interest of studying forward rates rather than yield curves is that one has a direct access to a ‘derivative’ (in the mathematical sense – see Eq. (2.19)), which obviously contains more precise informations than the yield curve (defined from the logarithm of $B(t, \theta)$ itself).

In practice, the futures markets price three months forward rates for fixed expiration dates, separated by three month intervals. Identifying three months futures rates to instantaneous forward rates, we have available time series on forward rates $f(t, T_i - t)$, where $T_i$ are fixed dates (March, June, September and December of each year), which one can convert into fixed maturity (multiples of three months) forward rates by a simple linear interpolation between the two nearest points such that $T_i - t \leq \theta \leq T_{i+1} - t$. Between 1990 and 1996, one has at least 15 different Eurodollar maturities for each market date. Between 1994 and 1996, the number of available maturities rises to 30 (as time grows, longer and longer maturity forward rates are being traded on future markets); we shall thus often use this restricted data set. Since we only have daily data, our reference time scale will be $\tau = 1$ day. The variation of $f(t, \theta)$ between $t$ and $t + \tau$ will be denoted as $\delta f(t, \theta)$:

$$\delta f(t, \theta) = f(t + \tau, \theta) - f(t, \theta).$$  

(2.20)

### 2.6.2 Quantities of interest and data analysis

The description of the FRC has two, possibly interrelated, aspects:

1. What is, at a given instant of time, the shape of the FRC as a function of the maturity $\theta$?
2. What are the statistical properties of the increments $\delta f(t, \theta)$ between time $t$ and time $t + \tau$, and how are they correlated with the shape of the FRC at time $t$?

The two basic quantities describing the FRC at time $t$ is the value of the short term interest rate $f(t, \theta_{\text{min}})$ (where $\theta_{\text{min}}$ is the shortest available maturity), and that of the short term/long term spread $s(t) = \ldots$
2.6 Statistical analysis of the Forward Rate Curve (*)

Figure 2.21: The average FRC in the period 94-96, as a function of the maturity $\theta$. We have shown for comparison a one parameter fit with a square-root law, $A(\sqrt{\theta} - \sqrt{\theta_{\text{min}}})$. The same $\sqrt{\theta}$ behaviour actually extends up to $\theta_{\text{max}} = 10$ years, which is available in the second half of the time period.

$$f(t, \theta_{\text{max}}) - f(t, \theta_{\text{min}}),$$

where $\theta_{\text{max}}$ is the longest available maturity. The two quantities $r(t) \simeq f(t, \theta_{\text{min}})$, $s(t)$ are plotted versus time in Fig. 2.20; note that:

- The volatility $\sigma$ of the spot rate $r(t)$ is equal to $0.8%/\sqrt{\text{year}}$. This is obtained by averaging over the whole period.
- The spread $s(t)$ has varied between 0.53% and 4.34%. Contrarily to some European interest rates on the same period, $s(t)$ has always remained positive. (This however does not mean that the FRC is increasing monotonously, see below).

Figure 2.21 shows the average shape of the FRC, determined by averaging the difference $f(t, \theta) - r(t)$ over time. Interestingly, this function is rather well fitted by a simple square-root law. This means that on average, the difference between the forward rate with maturity $\theta$ and the spot rate is equal to $A\sqrt{\theta}$, with a proportionality constant $A = 0.85%/\sqrt{\text{year}}$ which turns out to be nearly identical to the spot rate volatility. We shall propose a simple interpretation of this fact below.

Let us now turn to an analysis of the fluctuations around the average shape. These fluctuations are actually similar to that of a vibrated elastic string. The average deviation $\Delta(\theta)$ can be defined as:

$$\Delta(\theta) \equiv \sqrt{\langle [f(t, \theta) - r(t) - s(t)\sqrt{\theta/\theta_{\text{max}}}]^2 \rangle},$$

(2.21)

and is plotted in Fig. 2.22, for the period 94-96. The maximum of $\Delta$ is reached for a maturity of $\theta^* = 1$ year.

We now turn to the statistics of the daily increments $\delta f(t, \theta)$ of the forward rates, by calculating their volatility $\sigma(\theta) = \sqrt{\langle \delta f(t, \theta)^2 \rangle}$, their excess kurtosis

$$\kappa(\theta) = \frac{\langle (\delta f(t, \theta))^4 \rangle}{\sigma^4(\theta)} - 3,$$

(2.22)
and the following ‘spread’ correlation function:

\[ C(\theta) = \frac{\langle \delta f(t, \theta_{\text{min}}) (\delta f(t, \theta) - \delta f(t, \theta_{\text{min}})) \rangle}{\sigma^2(\theta_{\text{min}})}, \tag{2.23} \]

which measures the influence of the short term interest fluctuations on the other modes of motion of the FRC, subtracting a trivial overall translation of the FRC.

Figure 2.23 shows \( \sigma(\theta) \) and \( \kappa(\theta) \). Somewhat surprisingly, \( \sigma(\theta) \), much like \( \Delta(\theta) \) has a maximum around \( \theta^* = 1 \) year. The order of magnitude of \( \sigma(\theta) \) is 0.05%/√day, or 0.8%/√year. The kurtosis \( \kappa(\theta) \) is rather high (on the order of 5), and only weakly decreasing with \( \theta \).

Finally, \( C(\theta) \) is shown in Fig. 2.22; its shape is again very similar to those of \( \Delta(\theta) \) and \( \sigma(\theta) \), with a pronounced maximum around \( \theta^* = 1 \) year. This means that the fluctuations of the short term rate are amplified for maturities around one year. We shall come back to this important point below.

### 2.6.3 Comparison with the Vasicek model

The simplest FRC model is a one factor model due to Vasicek, where the whole term structure can be ascribed to the short term interest rate, which is assumed to follow a so-called ‘Ornstein-Uhlenbeck’ (or mean reverting) process defined as:

\[ \frac{dr(t)}{dt} = \Omega(r_0 - r(t)) + \sigma \xi(t), \tag{2.24} \]

where \( r_0 \) is an ‘equilibrium’ reference rate, \( \Omega \) describes the strength of the reversion towards \( r_0 \) (and is the inverse of the mean reversion time), and \( \xi(t) \) is a Gaussian noise, of volatility 1. In its simplest version, the Vasicek model prices a bond maturing at \( T \) as the following average:

\[ B(t, T) = \left\langle \exp - \int_t^T du \ r(u) \right\rangle, \tag{2.25} \]

where the averaging is over the possible histories of the spot rate between now and the maturity, where the uncertainty is modelled by the noise \( \xi \). The computation of the above average is straightforward and leads to (using Eq. (2.19)):

\[ f(t, \theta) = r(t) + (r_0 - r(t))(1 - e^{-\Omega \theta}) - \frac{\sigma^2}{2 \Omega^2} (1 - e^{-\Omega \theta})^2 \tag{2.26} \]

The basic results of this model are as follows:

![Figure 2.23: The daily volatility and kurtosis as a function of maturity. Note the maximum of the volatility for \( \theta = \theta^* \), while the kurtosis is rather high, and only very slowly decreasing with \( \theta \). The two curves correspond to the periods 90-96 and 94-96, the latter period extending to longer maturities.](image)
2.6 Statistical analysis of the Forward Rate Curve (*)

Since \( \langle r_0 - r(t) \rangle = 0 \), the average of \( f(t, \theta) - r(t) \) is given by

\[
\langle f(t, \theta) - r(t) \rangle = -\sigma^2 / 2\Omega^2 (1 - e^{-t\Omega})^2,
\]

and should thus be negative, at variance with empirical data. Note that in the limit \( \Omega \theta \ll 1 \), the order of magnitude of this (negative) term is very small: taking \( \sigma = 1\% / \sqrt{\text{year}} \) and \( \theta = 1 \text{ year} \), it is found to be equal to 0.005%, much smaller than the typical differences actually observed on forward rates.

- The volatility \( \sigma(\theta) \) is monotonously decreasing as \( \exp(-\Omega \theta) \), while the kurtosis \( \kappa(\theta) \) is identically zero (because \( \xi \) is Gaussian).
- The correlation function \( C(\theta) \) is negative and is a monotonous decreasing function of its argument, in total disagreement with observations (Fig. 2.22).
- The variation of the spread \( s(t) \) and of the spot rate should be perfectly correlated, which is not the case (Fig. 2.22): more than one factor is in any case needed to account for the deformation of the FRC.

An interesting extension of Vasicek’s model designed to fit exactly the ‘initial’ FRC \( f(t = 0, \theta) \) was proposed by Hull and White [Hull]. It amounts to replacing the above constants \( \Omega \) and \( r_0 \) by time dependent functions. For example, \( r_0(t) \) represents the anticipated evolution of the ‘reference’ short term rate itself with time. These functions can be adjusted to fit \( f(t = 0, \theta) \) exactly. Interestingly, one can then derive the following relation:

\[
\frac{\partial f(t)}{\partial t} = \left( \frac{\partial f(t, 0)}{\partial t} \right),
\]

up to a term of order \( \sigma^2 \) which turns out to be negligible, exactly for the same reason as explained above. On average, the second term (estimated by taking a finite difference estimate of the partial derivative using the first two points of the FRC) is definitely found to be positive, and equal to 0.8%/year. On the same period (90-96), however, the spot rate has decreased from 8.1% to 5.9%, instead of growing by 5.6%.

In simple terms, both the Vasicek and the Hull-White model mean the following: the FRC should basically reflect the market’s expectation of the average evolution of the spot rate (up to a correction on the order of \( \sigma^2 \), but which turns out to be very small – see above). However, since the FRC is on average increasing with the maturity (situations when the FRC is ‘inverted’ are comparatively much rarer), this would mean that the market systematically expects the spot rate to rise, which it does not. It is hard to believe that the market persists in error for such a long time. Hence, the upward slope of the FRC is not only related to what the market expects on average, but that a systematic risk premium is needed to account for this increase.

2.6.4 Risk-premium and the \( \sqrt{\theta} \) law

The average FRC and Value-at-Risk pricing

The observation that on average the FRC follows a simple \( \sqrt{\theta} \) law (i.e. \( \langle f(t, \theta) - r(t) \rangle \propto \sqrt{\theta} \)) suggests an intuitive, direct interpretation. At any time \( t \), the market anticipates either a future rise, or a decrease of the spot rate. However, the average anticipated trend is, in the long run, zero, since the spot rate has bounded fluctuations. Hence, the average market’s expectation is that the future spot rate \( r(t) \) will be close to its present value \( r(t = 0) \). In this sense, the average FRC should thus be flat. However, even in the absence of any trend in the spot rate, its probable change between now and \( t = \theta \) is (assuming the simplest random walk behaviour) of the order of \( \sigma \sqrt{\theta} \), where \( \sigma \) is the volatility of the spot rate. Hence, money lenders are tempted to protect themselves against this potential rise by adding to their estimate of the average future rate a risk premium on the order of \( \sigma \sqrt{\theta} \) to set the forward rate at a satisfactory value. In other words, money lenders take a bet on the future value of the spot rate and want to be sure not to lose their bet more frequently than – say – once out of five. Thus their price for the forward rate is such that the probability that the spot rate at time \( t + \theta \), \( r(t + \theta) \) actually exceeds \( f(t, \theta) \) is equal to a certain number \( p \):

\[
\int_{f(t, \theta)}^{\infty} dr' P(r', t + \theta | r, t) = p,
\]

where \( P(r', t' | r, t) \) is the probability that the spot rate is equal to \( r' \) at time \( t' \) knowing that it is \( r \) now (at time \( t \)). Assuming that \( r' \) follows a simple random walk centred around \( r(t) \) then leads to:

\[
f(t, \theta) = r(t) + A r(0) \sqrt{\theta}, \quad A = \sqrt{2} \operatorname{erfc}^{-1}(2p),
\]

which indeed matches the empirical data, with \( p \sim 0.16 \).

Hence, the shape of today’s FRC can be thought of as an envelope for the probable future evolutions of the spot rate. The market appears to

---

[16] This assumption is certainly inadequate for small times, where large kurtosis effects are present. However, one the scale of months, these non Gaussian effects can be considered as small.
price future rates through a Value at Risk procedure (Eqs. 2.29, 2.30 – see Chapter 3 below) rather than through an averaging procedure.

The anticipated trend and the volatility hump

Let us now discuss, along the same lines, the shape of the FRC at a given instant of time, which of course deviates from the average square root law. For a given instant of time \( t \), the market actually expects the spot rate to perform a biased random walk. We shall argue that a consistent interpretation is that the market estimates the trend \( m(t) \) by extrapolating the past behaviour of the spot rate itself. Hence, the probability distribution \( P(r', t+\theta | r, t) \) used by the market is not centred around \( r(t) \) but rather around:

\[
r(t) + \int_{t}^{t+\theta} du \, m(t, t+u),
\]

(2.31)

where \( m(t, t') \) can be called the anticipated bias at time \( t' \), seen from time \( t \).

It is reasonable to think that the market estimates \( m \) by extrapolating the recent past to the nearby future. Mathematically, this reads:

\[
m(t, t+u) = m_1(t) Z(u) \quad \text{where} \quad m_1(t) \equiv \int_{\theta}^{\infty} dv \, K(v) \delta(r(t-v)),
\]

(2.32)

and where \( K(v) \) is an averaging kernel of the past variations of the spot rate. One may call \( Z(u) \) the trend persistence function; it is normalised such that \( Z(0) = 1 \), and describes how the present trend is expected to persist in the future. Equation (2.29) then gives:

\[
f(t, \theta) = r(t) + A \sigma \sqrt{\theta} + m_1(t) \int_{0}^{\theta} du \, Z(u).
\]

(2.33)

This mechanism is a possible explanation of why the three functions introduced above, namely \( \Delta(\theta) \), \( \sigma(\theta) \) and the correlation function \( C(\theta) \) have similar shapes. Indeed, taking for simplicity an exponential averaging kernel \( K(v) \) of the form \( \exp[-v\epsilon] \), one finds:

\[
\frac{dm_1(t)}{dt} = -c m_1 + \epsilon \frac{dr(t)}{dt} + \epsilon \xi(t),
\]

(2.34)

where \( \xi(t) \) is an independent noise of strength \( \sigma_\xi^2 \), added to introduce some extra noise in the determination of the anticipated bias. In the absence of temporal correlations, one can compute from the above equation the average value of \( m_1^2 \). It is given by:

\[
\langle m_1^2 \rangle = \frac{\epsilon}{2} (\sigma^2(0) + \sigma^2_\xi).
\]

(2.35)

Figure 2.24: Comparison between the theoretical prediction and the observed daily volatility of the forward rate at maturity \( \theta \), in the period 94-96. The dotted line corresponds to Eq. (2.38) with \( \sigma_\xi = \sigma(0) \), while the full line is obtained by adding the effect of the variation of the coefficient \( A \sigma(0) \) in Eq. (2.33), which adds a contribution proportional to \( \theta \).

In the simple model defined by Eq. (2.33) above, one finds that the correlation function \( C(\theta) \) is given by:\(^{17}\)

\[
C(\theta) = \epsilon \int_{0}^{\theta} du \, Z(u).
\]

(2.36)

Using the above result for \( \langle m_1^2 \rangle \), one also finds:

\[
\Delta(\theta) = \frac{\sigma^2(0) + \sigma^2_\xi}{2\epsilon} C(\theta),
\]

(2.37)

thus showing that \( \Delta(\theta) \) and \( C(\theta) \) are indeed proportional.

Turning now to the volatility \( \sigma(\theta) \), one finds that it is given by:

\[
\sigma^2(\theta) = [1 + C(\theta)]^2 \sigma^2(0) + C(\theta)^2 \sigma^2_\xi.
\]

(2.38)

We thus see that the maximum of \( \sigma(\theta) \) is indeed related to that of \( C(\theta) \). Intuitively, the reason for the volatility maximum is as follows: a variation in the spot rate changes that market anticipation for the trend \( m_1(t) \). But this change of trend obviously has a larger effect

\(^{17}\)In reality, one should also take into account the fact that \( A \sigma(0) \) can vary with time. This brings an extra contribution both to \( C(\theta) \) and to \( \sigma(\theta) \).
2.7 Correlation matrices (*)

As will be detailed in Chapter 3, an important aspect of risk management is the estimation of the correlations between the price movements of different assets. The probability of large losses for a certain portfolio or option book is dominated by correlated moves of its different constituents – for example, a position which is simultaneously long in stocks and short in bonds will be risky because stocks and bonds move in opposite directions in crisis periods. The study of correlation (or covariance) matrices thus has a long history in finance, and is one of the cornerstone of Markowitz’s theory of optimal portfolios (see Section 3.3). However, a reliable empirical determination of a correlation matrix turns out to be difficult: if one considers \( M \) assets, the correlation matrix contains \( M(M - 1)/2 \) entries, which must be determined from \( M \) time series of length \( N \); if \( N \) is not very large compared to \( M \), one should expect that the determination of the covariances is noisy, and therefore that the empirical correlation matrix is to a large extent random, i.e. the structure of the matrix is dominated by ‘measurement’ noise. If this is the case, one should be very careful when using this correlation matrix in applications. From this point of view, it is interesting to compare the properties of an empirical correlation matrix \( C \) to a ‘null hypothesis’ purely random matrix as one could obtain from a finite time series of strictly uncorrelated assets. Deviations from the random matrix case might then suggest the presence of true information. \(^{18}\)

The empirical correlation matrix \( C \) is constructed from the time series of price changes \( \delta xroll_k \) (where \( i \) labels the asset and \( k \) the time) through the equation:

\[
C_{ij} = \frac{1}{N} \sum_{k=1}^{N} \delta x_i^k \delta x_j^k. \tag{2.39}
\]

In the following we assume that the average value of the \( \delta x \)'s has been subtracted off, and that the \( \delta x \)'s are rescaled to have a constant unit volatility. The null hypothesis of independent assets, which we consider now, translates itself in the assumption that the coefficients \( \delta x_k^i \) are independent, identically distributed, random variables. \(^{19}\) The theory of Random Matrices, briefly exposed in Section 1.8, allows one to compute the density of eigenvalues of \( C \), \( p_C(\lambda) \), in the limit of very large matrices: it is given by Eq. (1.119), with \( Q = N/M \).

Now, we want to compare the empirical distribution of the eigenvalues of the correlation matrix of stocks corresponding to different markets with the theoretical prediction given by Eq. (4.17), based on the assumption that the correlation matrix is random. We have studied numerically the density of eigenvalues of the correlation matrix of \( M = 406 \) assets of the S&P 500, based on daily variations during the years 1991-96, for a total of \( N = 1309 \) days (the corresponding value of \( Q \) is 3.22). An immediate observation is that the highest eigenvalue \( \lambda_1 \) is 25 times larger than the predicted \( \lambda_{\text{max}} \) (Fig. 2.25, inset). The corresponding eigenvector is, as expected, the ‘market’ itself, i.e. it has roughly equal components on all the \( M \) stocks. The simplest ‘pure noise’ hypothesis is therefore inconsistent with the value of \( \lambda_1 \). A more reasonable idea is that the components of the correlation matrix which are orthogonal to the ‘market’ is pure noise. This amounts to subtracting the contribution of \( \lambda_1 \) from the nominal value \( \sigma^2 = 1 \), leading to \( \sigma^2 = 1 - \lambda_{\text{max}}/M = 0.85 \). The corresponding fit of the empirical distribution is shown as a dotted line in Fig. 2.25. Several eigenvalues are still above \( \lambda_{\text{max}} \) and might contain some information, thereby reducing the variance of the effectively random part of the correlation matrix. One can therefore treat \( \sigma^2 \) as an adjustable parameter. The best fit is obtained for \( \sigma^2 = 0.74 \), and corresponds to the dark line in Fig. 2.25, which accounts quite satisfactorily for 94% of the spectrum, while the 6% highest eigenvalues still exceed the theoretical upper edge by a substantial amount. These 6% highest eigenvalues are however responsible for 26% of the total volatility.

One can repeat the above analysis on different stock markets (e.g. Paris, London, Zurich), or on volatility correlation matrices, to find very similar results. In a first approximation, the location of the theoretical edge, determined by fitting the part of the density which contains most of the eigenvalues, allows one to distinguish ‘information’ from ‘noise’.

The conclusion of this section is therefore that a large part of the empirical correlation matrices must be considered as ‘noise’, and cannot be trusted for risk management. In the next chapter, we will dwell on Markowitz’ theory of optimal portfolio, which assumes that the correlation matrix is perfectly known. This theory must therefore be taken with a grain of salt, bearing in mind the results of the present section.

\(^{18}\)This section is based on the following paper: L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, “Random matrix theory,” Risk Magazine (March 1999).

\(^{19}\)Note that even if the ‘true’ correlation matrix \( C_{\text{true}} \) is the identity matrix, its empirical determination from a finite time series will generate non trivial eigenvectors and eigenvalues.
2.8 A simple mechanism for anomalous price statistics (*)

We have chosen the family of TLD to represent the distribution of price fluctuations. As mentioned above, Student distributions can also account quite well for the shape of the empirical distributions. Hyperbolic distributions have also been proposed. The choice of TLD’s was motivated by two particular arguments:

- This family of distributions generalises in a natural way the two classical descriptions of price fluctuations, since the Gaussian corresponds to \( \mu = 2 \), and the stable Lévy distributions correspond to \( \alpha = 0 \).

- The idea of TLD allows one to account for the deformation of the distributions as the time horizon \( N \) increases, and the anomalously high value of the Hurst exponent \( H \) at small times, crossing over to \( H = 1/2 \) for longer times.

However, in order to justify the choice of one family of laws over the others, one needs a microscopic model for price fluctuations where a theoretical distribution can be computed. In the next two sections, we propose such ‘models’ (in the physicist sense). These models are not very realistic, but only aim at showing that power-law distributions (possibly with an exponential truncation) appear quite naturally. Furthermore, the model considered in this section leads to a value of \( \mu = 3/2 \), close to the one observed on real data.\(^{20}\)

We assume that the price increment \( \delta x_k \) reflects the instantaneous offset between supply and demand. More precisely, if each operator on the market \( \alpha \) wants to buy or sell a certain fixed quantity \( q \) of the asset \( X \), one has:

\[
\delta x_k \propto q \sum_{\alpha} \varphi_{\alpha},
\]

(2.40)

where \( \varphi_{\alpha} \) can take the values \(-1, 0, +1\), depending on whether the operator \( \alpha \) is selling, inactive, or buying. Suppose now that the operators interact among themselves in an heterogeneous manner: with a small probability \( p/N \) (where \( N \) is the total number of operators on the market), two operators \( \alpha \) and \( \beta \) are ‘connected’, and with probability

1 − p/N, they ignore each other. The factor 1/N means that on average, the number of operator connected to any particular one is equal to p. Suppose finally that if two operators are connected, they come to agree on the strategy they should follow, i.e. ϕa = ϕβ.

It is easy to understand that the population of operators clusters into groups sharing the same opinion. These clusters are defined such that there exists a connection between any two operators belonging to this cluster, although the connection can be indirect and follow a certain ‘path’ between operators. These clusters do not have all the same size, i.e. do not contain the same number of operators. If the size of cluster A is called N(A), one can write:

\[ \delta x_k \propto q \sum_A N(A) \varphi(A), \]

(2.41)

where \( \varphi(A) \) is the common opinion of all operators belonging to A. The statistics of the price increments \( \delta x_k \) therefore reduces to the statistics of the size of clusters, a classical problem in percolation theory [Stauffer]. One finds that as long as \( p < 1 \) (less than one ‘neighbour’ on average with whom one can exchange information), then all \( N(A) \)'s are small compared to the total number of traders \( N \). More precisely, the distribution of cluster sizes takes the following form in the limit where \( 1 − p = \epsilon \ll 1 \):

\[ P(N) \propto N^{3/2} \exp(-\epsilon^2 N) \quad N \ll N. \]

(2.42)

When \( p = 1 \) (percolation threshold), the distribution becomes a pure power-law with an exponent 1 + \( \mu = 5/2 \), and the CLT tells us that in this case, the distribution of the price increments \( \delta x \) is precisely a pure symmetric Lévy distribution of index \( \mu = 3/2 \) (assuming that \( \varphi = \pm 1 \) play identical roles, that is if there is no global bias pushing the price up or down). If \( p < 1 \), on the other hand, one finds that the Lévy distribution is truncated exponentially far in the tail. If \( p > 1 \), a finite fraction of the \( N \) traders have the same opinion: this leads to a crash.

This simple model is interesting but has one major drawback: one has to assume that the parameter \( p \) is smaller than one, but relatively close to one such that Eq. (2.42) is valid, and non trivial statistics follows. One should thus explain why the value of \( p \) spontaneously stabilises in the neighbourhood of the critical value \( p = 1 \). Certain models do actually have this property, of being close to or at a critical point without having to fine tune any of their parameters. These models are called ‘self-organised critical’ [Bak et al.]. In this spirit, let us mention a very recent model of Sethna et al. [Dahmen and Sethna], meant to describe the behaviour of magnets in a time dependent magnetic field. Transposed to the present problem, this model describes the collective behaviour of a set of traders exchanging information, but having all different \textit{a priori} opinions. One trader can however change his mind and take the opinion of his neighbours if the coupling is strong, or if the strength of his \textit{a priori} opinion is weak. All these traders feel an external ‘field’, which represents for example a long term expectation of economy growth or recession, leading to an increased average pessimism or optimism. For a large range of parameters, one finds that the buy orders (or the sell orders) organise as avalanches of various sizes, distributed as a power-law with an exponential cut-off, with \( \mu = 5/4 = 1.25 \). If the anticipation of the traders are too similar, or if the coupling between agents is too strong (strong mimetism), the model again leads to a crash-like behaviour.

### 2.9 A simple model with volatility correlations and tails (*)

In this section, we show that a very simple feedback model where past high values of the volatility influence the present market activity does lead to tails in the probability distribution and, by construction, to volatility correlations. The present model is close in spirit to the ARCH models which have been much discussed in this context. The idea is to write:

\[ x_{k+1} = x_k + \sigma_k \xi_k, \]

(2.43)

where \( \xi_k \) is a random variable of unit variance, and to postulate that the present day volatility \( \sigma_k \) depends on how the market feels the past market volatility. If the past price variations happened to be high, the market interprets this as a reason to be more nervous and increases its activity, thereby increasing \( \sigma_k \). One could therefore consider, as a toy-model:

\[ \sigma_{k+1} - \sigma_0 = (1 - \epsilon)(\sigma_k - \sigma_0) + \epsilon|\sigma_k \xi_k|, \]

(2.44)

which means that the market takes as an indicator of the past day activity the absolute value of the close to close price difference \( x_{k+1} - x_k \). Now, writing

\[ |\sigma_k \xi_k| = |\langle |\sigma \xi| \rangle + \sigma_0 \xi_k|, \]

(2.45)

and going to a continuous time formulation, one finds that the volatility probability distribution \( P(\sigma, t) \) obeys the following ‘Fokker-Planck’ equation:

\[ \frac{\partial P(\sigma, t)}{\partial t} = \epsilon \frac{\partial (\sigma - \bar{\sigma}_0)P(\sigma, t)}{\partial \sigma} + \epsilon^2 \sigma^2 \frac{\partial^2 P(\sigma, t)}{\partial \sigma^2}. \]

(2.46)
where \( \tilde{\sigma}_0 = \sigma_0 - \epsilon(\sigma_\xi) \), and where \( c^2 \) is the variance of the noise \( \delta \xi \).

The equilibrium solution of this equation, \( P_e(\sigma) \), is obtained by setting the left-hand side to zero. One finds:

\[
P_e(\sigma) = \exp\left(-\frac{\tilde{\sigma}_0}{\sigma}\right),
\]

with \( \mu = 1 + (c^2\epsilon)^{-1} > 1 \). Now, for a large class of distributions for the random noise \( \xi_k \), for example Gaussian, it is easy to show, using a saddle-point calculation, that the tails of the distribution of \( \delta \xi \) are power-laws, with the same exponent \( \mu \). Interestingly, a short memory market, corresponding to \( \epsilon \approx 1 \), has much wilder tails than a long-memory market: in the limit \( \epsilon \to 0 \), one indeed has \( \mu \to \infty \). In other words, over-reactions is a potential cause for power-law tails.

### 2.10 Conclusion

The above statistical analysis reveals very important differences between the simple model usually adopted to describe price fluctuations, namely the geometric (continuous time) Brownian motion and the rather involved statistics of real price changes. The Brownian motion description is at the heart of most theoretical work in mathematical finance, and can be summarised as follows:

- One assumes that the relative returns (rather than the absolute price increments) are independent random variables.

- One assumes that the elementary time scale \( \tau \) tends to zero; in other words that the price process is a continuous time process. It is clear that in this limit, the number of independent price changes in an interval of time \( T \) is \( N = T/\tau \to \infty \). One is thus in the limit where the CLT applies whatever the time scale \( T \).

If the variance of the returns is finite, then according to the CLT, the only possibility is that price changes obey a log-normal distribution. The process is also scale invariant, that is that its statistical properties do not depend on the chosen time scale (up to a multiplicative factor – see 1.5.3).

The main difference between this model and real data is not only that the tails of the distributions are very poorly described by a Gaussian law, but also that several important time scales appear in the analysis of price changes:

- A ‘microscopic’ time scale \( \tau \) below which price changes are correlated. This time scale is on the order of a few tens of minutes even on very liquid markets.

- A time scale \( T^* = N^*\tau \), which corresponds to the time where non-Gaussian effects begin to smear out, beyond which the CLT begins to operate. This time scale \( T^* \) depends much on the initial kurtosis on scale \( \tau \). As a first approximation, one has: \( T^* = \kappa_1\tau \), which is equal to several days, even on very liquid markets.

- A time scale corresponding to the correlation time of the volatility fluctuations, which is on the order of ten days to a month or even longer.

- And finally a time scale \( T_\sigma \) governing the crossover from an additive model, where absolute price changes are the relevant random variables, to a multiplicative model, where relative returns become relevant. This time scale is also on the order of a month.

It is clear that the existence of all these time scales is extremely important to take into account in a faithful representation of price changes, and play a crucial role both in the pricing of derivative products, and in risk control. Different assets differ in the value of their kurtosis, and in the value of these different time scales. For this reason, a description where the volatility is the only parameter (as is the case for Gaussian models) are bound to miss a great deal of the reality.

### 2.11 References

- Scaling and Fractals in Financial Markets:
2.11 References

- Percolation, collective models and self organized criticality:


- Other recent market models:


Other descriptions (ARCH, etc.):


The interest rate curve:

L. Hughston (edt.), *Vasicek and beyond*, Risk Publications, 1997

3

EXTREME RISKS AND OPTIMAL PORTFOLIOS

Il n’est plus grande folie que de placer son salut dans l’incertitude.\footnote{Nothing is more foolish than betting on uncertainty for salvation.}

(Madame de Sévigné, Lettres.)

3.1 Risk measurement and diversification

Measuring and controlling risks is now one of the major concern across all modern human activities. The financial markets, which act as highly sensitive economical and political thermometers, are no exception. One of their rôle is actually to allow the different actors in the economic world to trade their risks, to which a price must therefore be given.

The very essence of the financial markets is to fix thousands of prices all day long, thereby generating enormous quantities of data that can be analysed statistically. An objective measure of risk therefore appears to be easier to achieve in finance than in most other human activities, where the definition of risk is vaguer, and the available data often very poor. Even if a purely statistical approach to financial risks is itself a dangerous scientists’ dream (see e.g. Figure 1.1), it is fair to say that this approach has not been fully exploited until the very recent years, and that many improvements can be expected in the future, in particular concerning the control of extreme risks. The aim of this chapter is to introduce some classical ideas on financial risks, to illustrate their weaknesses, and to propose several theoretical ideas devised to handle more adequately the ‘rare events’ where the true financial risk resides.

3.1.1 Risk and volatility

Financial risk has been traditionally associated with the statistical uncertainty on the final outcome. Its traditional measure is the RMS, or,
in financial terms, the *volatility*. We will note by \( R(T) \) the logarithmic return on the time interval \( T \), defined by:

\[
R(T) = \log \left[ \frac{x(T)}{x_0} \right],
\]

where \( x(T) \) is the price of the asset \( X \) at time \( T \), knowing that it is equal to \( x_0 \) today \((t = 0)\). When \(|x(T) - x_0| \ll x_0\), this definition is equivalent to \( R(T) = \log\left[ x(T)/x_0 \right] \).

If \( P(x, T|x_0, 0)dx \) is the conditional probability of finding \( x(T) = x \) within \( dx \), the volatility \( \sigma \) of the investment is the standard deviation of \( R(T) \), defined by:

\[
\sigma^2 = \frac{1}{T} \left[ \int dx P(x, T|x_0, 0)R^2(T) - \left( \int dx P(x, T|x_0, 0)R(T) \right)^2 \right].
\]

The volatility is in general chosen as an adequate measure of risk associated to a given investment. We notice however that this definition includes in a symmetrical way both abnormal gains and abnormal losses. This fact is a priori curious. The theoretical foundations behind this particular definition of risk are numerous:

- First, operational: the computations involving the variance are relatively simple and can be generalised easily to multi-asset portfolios.
- Second, the Central Limit Theorem (CLT) presented in Chapter 1 seems to provide a general and solid justification: by decomposing the motion from \( x_0 \) to \( x(T) \) in \( N = T/\tau \) increments, one can write:

\[
x(T) = x_0 + \sum_{k=0}^{N-1} \delta x_k \quad \text{with} \quad \delta x_k = x_k \eta_k,
\]

where \( x_k = x(t = k\tau) \) and \( \eta_k \) is by definition the instantaneous return. Therefore, we have:

\[
R(T) = \log \left[ \frac{x(T)}{x_0} \right] = \sum_{k=0}^{N-1} \log(1 + \eta_k).
\]

In the classical approach one assumes that the returns \( \eta_k \) are independent variables. From the CLT we learn that *in the limit where \( N \to \infty \), \( R(T) \) becomes a Gaussian random variable centred on a given average return \( \bar{m}T \), with \( \bar{m} = \langle \log(1 + \eta_k) \rangle/\tau \), and whose standard deviation is given by \( \sigma \sqrt{T} \). Therefore, in this limit, the entire probability distribution of \( R(T) \) is parameterised by two quantities only, \( \bar{m} \) and \( \sigma \): any reasonable measure of risk must therefore be based on \( \sigma \).

However, as we have discussed at length in Chapter 1, this is not true for finite \( N \) (which corresponds to the financial reality: there are only roughly \( N \approx 320 \) half-hour intervals in a working month), especially in the ‘tails’ of the distribution, corresponding precisely to the extreme risks. We will discuss this point in detail below.

One can give to \( \sigma \) the following intuitive meaning: after a long enough time \( T \), the price of asset \( X \) is given by:

\[
x(T) = x_0 \exp[\bar{m}T + \sigma \sqrt{T} \xi],
\]

where \( \xi \) is a Gaussian random variable with zero mean and unit variance. The quantity \( \sigma \sqrt{T} \) gives us the order of magnitude of the deviation from the expected return. By comparing the two terms in the exponential, one finds that when \( T \gg \bar{m} \sigma \), the expected return becomes more important than the fluctuations, which means that the probability that \( x(T) \) is smaller than \( x_0 \) (and that the actual rate of return over that period is negative) becomes small. The ‘security horizon’ \( \hat{T} \) increases with \( \sigma \). For a typical individual stock, one has \( \bar{m} \approx 10\% \) per year and \( \sigma = 20\% \) per year, which leads to a \( T \) as long as four years!

The quality of an investment is often measured by its ‘Sharpe ratio’ \( \mathcal{S} \), that is, the ‘signal to noise’ ratio of the mean return \( \bar{m}T \) to the fluctuations \( \sigma \sqrt{T} \):\(^3\)

\[
\mathcal{S} = \frac{\bar{m}T}{\sigma} \equiv \sqrt{\frac{T}{\hat{T}}}.
\]

The Sharpe ratio increases with the investment horizon and is equal to 1 precisely when \( T = \hat{T} \). Practitioners usually define the Sharpe ratio for a 1 year horizon.

Note that the most probable value of \( x(T) \), as given by (3.5), is equal to \( x_0 \exp(\bar{m}T) \), while the mean value of \( x(T) \) is higher: assuming that \( \xi \) is Gaussian, one finds: \( x_0 \exp[(\bar{m}T + \sigma^2 T/2)] \). This difference is due to the fact that the returns \( \eta_k \), rather than the absolute increments \( \delta x_k \), are independent, identically distributed random variables. However, if \( T \) is short (say up to a few months), the difference between the two descriptions is hard to detect. As explained in Section 2.2.1, a purely additive description is actually more adequate at short times. In other words, we shall often in the following write \( x(T) \) as:

\[
x(T) = x_0 \exp[\bar{m}T + \sigma \sqrt{T} \xi] \simeq x_0 + \bar{m}T + \sigma \sqrt{\hat{T}} \xi,
\]

\(^2\)To second order in \( \eta \ll 1 \), we find: \( \sigma^2 = \frac{1}{2} \langle \eta^2 \rangle \) and \( \bar{m} = \frac{1}{2} \langle \eta \rangle + \frac{1}{2} \sigma^2 \).

\(^3\)It is customary to subtract from the mean return \( \bar{m} \) the riskfree rate in the definition of the Sharpe ratio.
where we have introduced the following notations: \( m \equiv \bar{m}x_0 \), \( D \equiv \sigma^2x_0^2 \), which we shall use throughout the following. The non-Gaussian nature of the random variable \( \xi \) is therefore the most important factor determining the probability for extreme risks.

### 3.1.2 Risk of loss and ‘Value at Risk’ (VaR)

The fact that financial risks are often described using the volatility is actually intimately related to the idea that the distribution of price changes is Gaussian. In the extreme case of ‘Lévy fluctuations’, for which the variance is infinite, this definition of risk would obviously be meaningless. Even in a less risky world, this measure of risk has three major drawbacks:

- The financial risk is obviously associated to losses and not to profits. A definition of risk where both events play symmetrical roles is thus not in conformity with the intuitive notion of risk, as perceived by professionals.
- As discussed at length in Chapter 1, a Gaussian model for the price fluctuations is never justified for the extreme events, since the CLT only applies in the centre of the distributions. Now, it is precisely these extreme risks that are of most concern for all financial houses, and thus those which need to be controlled in priority. In recent years, international regulators have tried to impose some rules to limit the exposure of banks to these extreme risks.
- The presence of extreme events in the financial time series can actually lead to a very bad empirical determination of the variance: its value can be substantially changed by a few ‘big days’. A bold solution to this problem is simply to remove the contribution of these so-called aberrant events! This rather absurd solution is actually quite commonly used.

Both from a fundamental point of view, and for a better control of financial risks, another definition of risk is thus needed. An interesting notion that we shall develop now is the probability of extreme losses, or, equivalently, the ‘Value-at-risk’ (VaR).

The probability to loose an amount \( -\delta x \) larger than a certain threshold \( \Lambda \) on a given time horizon \( \tau \) is defined as:

\[
\mathcal{P}[\delta x < -\Lambda] = \mathcal{P}_{\xi}[-\Lambda] = \int_{-\infty}^{-\Lambda} d\delta x \ P_\tau(\delta x),
\]

(3.8)

This definition means that a loss greater than \( \Lambda_{\text{VaR}} \) over a time interval of \( \tau = 1 \) day (for example) happens only every 100 days on average for \( \mathcal{P}_{\text{VaR}} = 1\% \). Let us note that this definition does not take into account the fact that losses can accumulate on consecutive time intervals \( \tau \), leading to an overall loss which might substantially exceed \( \Lambda_{\text{VaR}} \). Similarly, this definition does not take into account the value of the maximal loss ‘inside’ the period \( \tau \). In other words, only the close price over the period \( [k\tau, (k+1)\tau] \) is considered, and not the lowest point reached during this time interval: we shall come back on these temporal aspects of risk in Section 3.1.3.

More precisely, one can discuss the probability distribution \( P(\Lambda, N) \) for the worst daily loss \( \Lambda \) (we choose \( \tau = 1 \) day to be specific) on a temporal horizon \( T_{\text{VaR}} = N\tau = \tau / \mathcal{P}_{\text{VaR}} \). Using the results of Section 1.4, one has:

\[
P(\Lambda, N) = N[\mathcal{P}_{\tau}(-\Lambda)]^{N-1} \mathcal{P}_{\tau}(-\Lambda).
\]

(3.10)

Figure 3.1: Extreme value distribution (the so-called Gumbel distribution) \( P(\Lambda, N) \) when \( P_\tau(\delta x) \) decreases faster than any power-law. The most probable value \( \Lambda_{\max} \), has a probability equal to 0.63 to be exceeded.
3 Extreme risks and optimal portfolios

3.1 Risk measurement and diversification

Most probable worst loss (%), one finds that $\Lambda$ of mean $m_1$ is given by:

$$\mathcal{P}_{\mathcal{G}}< \left( \frac{\Lambda_{\text{VaR}} + m_1}{\sigma_1 \delta x_0} \right) = \mathcal{P}_{\text{VaR}} \rightarrow \Lambda_{\text{VaR}} = \sqrt{2} \sigma_1 \delta x_0 \text{erfc}^{-1}[2 \mathcal{P}_{\text{VaR}}] - m_1,$$

(cf. (1.68)). When $m_1$ is small, minimising $\Lambda_{\text{VaR}}$ is thus equivalent to minimising $\sigma$. It is furthermore important to note the very slow growth of $\Lambda_{\text{VaR}}$ as a function of the time horizon in a Gaussian framework. For example, for $T_{\text{VaR}} = 250 \tau$ (1 market year), corresponding to $\mathcal{P}_{\text{VaR}} = 0.004$, one finds that $\Lambda_{\text{VaR}} \simeq 2.65 \sigma_1 \delta x_0$. Typically, for $\tau = 1$ day, $\sigma_1 = 1\%$, and therefore, the most probable worst day on a market year is equal to $-2.65\%$, and grows only to $-3.35\%$ over a ten year horizon!

In the general case, however, there is no useful link between $\sigma$ and $\Lambda_{\text{VaR}}$. In some cases, decreasing one actually increases the other (see Section 3.2.3 and 3.4). Let us nevertheless mention the Chebyshev inequality, often invoked by the volatility fans, which states that if $\mathcal{P}$:

$$\Lambda_{\text{VaR}}^2 \leq \frac{\sigma^2 \delta x_0^2}{\mathcal{P}_{\text{VaR}}}.$$

This inequality suggests that in general, the knowledge of $\sigma$ is tantamount to that of $\Lambda_{\text{VaR}}$. This is however completely wrong, as illustrated in Fig. 3.2. We have represented $\Lambda_{\text{max}} (= \Lambda_{\text{VaR}}$ with $\mathcal{P}_{\text{VaR}} = 0.63$) as a function of the time horizon $T_{\text{VaR}} = N \tau$ for three distributions $P_\tau(\delta x)$ which all have exactly the same variance, but decay as a Gaussian, as an exponential, or as a power-law with an exponent $\mu = 3$ (cf. (1.83)). Of course, the slower the decay of the distribution, the faster the growth of $\Lambda_{\text{VaR}}$ when $T_{\text{VaR}} \rightarrow \infty$.

Table 3.1 shows, in the case of international bond indices, a comparison between the prediction of the most probable worst day using a Gaussian model, or using the observed exponential character of the tail of the distribution, and the actual worst day observed the following year. It is clear that the Gaussian prediction is systematically over-optimistic. The exponential model leads to a number which is 7 times out of 11 below the observed result, which is indeed the expected result (Fig. 3.1).

Note finally that the measure of risk as a loss probability keeps its meaning even if the variance is infinite, as for a Lévy process. Suppose indeed that $P_\tau(\delta x)$ decays very slowly when $\delta x$ is very large, as:

$$P_\tau(\delta x) \approx \frac{\mu A^\mu}{\delta x^{1+\mu}},$$

with $\mu < 2$, such that $\langle \delta x^2 \rangle = \infty$. $A^\mu$ is the ‘tail amplitude’ of the distribution $P_\tau$; $A$ gives the order of magnitude of the probable values

![Figure 3.2: Growth of $\Lambda_{\text{max}}$ as a function of the number of days $N$, for an asset of daily volatility equal to 1%, with different distribution of price increments: Gaussian, (symmetric) exponential, or power-law with $\mu = 3$ (cf. (1.83)). Note that for intermediate $N$, the exponential distribution leads to a larger VaR than the power-law; this relation is however inverted for $N \rightarrow \infty$.](image-url)
Table 3.1: J.P. Morgan international bond indices (expressed in French Francs), analysed over the period 1989–1993, and worst day observed day in 1994. The predicted numbers correspond to the most probable worst day $\Lambda_{\text{max}}$). The amplitude of the worst day with 95% confidence level is easily obtained, in the case of exponential tails, by multiplying $\Lambda_{\text{max}}$ by a factor $1.53$. The last line corresponds to a portfolio made of the 11 bonds with equal weight. All numbers are in %.

<table>
<thead>
<tr>
<th>Country</th>
<th>Worst day Log-normal</th>
<th>Worst day TLD</th>
<th>Worst day Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>0.92</td>
<td>1.14</td>
<td>1.11</td>
</tr>
<tr>
<td>Canada</td>
<td>2.07</td>
<td>2.78</td>
<td>2.76</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.92</td>
<td>1.08</td>
<td>1.64</td>
</tr>
<tr>
<td>France</td>
<td>0.59</td>
<td>0.74</td>
<td>1.24</td>
</tr>
<tr>
<td>Germany</td>
<td>0.60</td>
<td>0.79</td>
<td>1.44</td>
</tr>
<tr>
<td>Great-Britain</td>
<td>1.59</td>
<td>2.08</td>
<td>2.08</td>
</tr>
<tr>
<td>Italy</td>
<td>1.31</td>
<td>2.60</td>
<td>4.18</td>
</tr>
<tr>
<td>Japan</td>
<td>0.65</td>
<td>0.82</td>
<td>1.08</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.57</td>
<td>0.70</td>
<td>1.10</td>
</tr>
<tr>
<td>Spain</td>
<td>1.22</td>
<td>1.72</td>
<td>1.98</td>
</tr>
<tr>
<td>United-States</td>
<td>1.85</td>
<td>2.31</td>
<td>2.26</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.61</td>
<td>0.80</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Figure 3.3: Cumulative probability distribution of losses (Rank histogram) for the S&P 500 for the period 1989-1998. Shown is the daily loss $(X_{cl} - X_{op})/X_{op}$ (thin line, left axis) and the intraday loss $(X_{lo} - X_{op})/X_{op}$ (thick line, right axis). Note that the right axis is shifted downwards by a factor of two with respect to the left one, so in theory the two lines should fall on top of one another.

3.1.3 Temporal aspects: drawdown and cumulated loss

Worst low

A first problem which arises is the following: we have defined $P$ as the probability for the loss observed at the end of the period $[k\tau,(k+1)\tau]$
Table 3.2: Average value of the absolute value of the open/close daily returns and maximum daily range (high-low) over open for S&P500, DEM/$ and BUND. Note that the ratio of these two quantities is indeed close to 2. Data from 1989 to 1998.

|       | \( A = \frac{\langle |X_{\text{cl}} - X_{\text{op}}| \rangle}{X_{\text{op}}} \) | \( B = \frac{\langle X_{\text{hi}} - X_{\text{lo}} \rangle}{X_{\text{op}}} \) | \( B/A \) |
|-------|-------------------------------------------------|---------------------------------|-------|
| S&P 500 | 0.472% | 0.924% | 1.96 |
| DEM/$  | 0.392% | 0.804% | 2.05 |
| BUND   | 0.250% | 0.496% | 1.98 |

to be at least equal to \( \Lambda \). However, in general, a worse loss still has been reached within this time interval. What is then the probability that the worst point reached within the interval (the ‘low’) is a least equal to \( \Lambda \)? The answer is easy for symmetrically distributed increments (we thus neglect the average return \( m_{\tau} \ll \Lambda \), which is justified for small enough time intervals): this probability is simply equal to \( 2P \),

\[
P[X_{\text{lo}} - X_{\text{op}} < -\Lambda] = 2P[X_{\text{cl}} - X_{\text{op}} < -\Lambda], \tag{3.15}
\]

where \( X_{\text{op}} \) is the value at the beginning of the interval (open), \( X_{\text{cl}} \) at the end (close) and \( X_{\text{lo}} \), the lowest value in the interval (low). The reason for this is that for each trajectory just reaching \( -\Lambda \) between \( k\tau \) and \((k + 1)\tau\), followed by a path which ends up above \( -\Lambda \) at the end of the period, there is a mirror path with precisely the same weight which reaches a ‘close’ value beyond \( -\Lambda \). This factor of 2 in cumulative probability is illustrated in Fig. 3.3. Therefore, if one wants to take into account the possibility of further loss within the time interval of interest in the computation of the VaR, one should simply divide by a factor 2 the probability level \( P_{\text{VaR}} \) appearing in (3.9).

A simple consequence of this ‘factor 2’ rule is the following: the average value of the maximal excursion of the price during time \( \tau \), given by the high minus the low over that period, is equal to twice the average of the absolute value of the variation from the beginning to the end of the period \( \langle |\text{open-close}| \rangle \). This relation can be tested on real data, the corresponding empirical factor is reported in Table 3.2.

\footnote{In fact, this argument – which dates back to Bachelier himself (1900)! – assumes that the moment where the trajectory reaches the point \( -\Lambda \) can be precisely identified. This is not the case for a discontinuous process, for which the doubling of \( P \) is only an approximate result.}

Another very important aspect of the problem is to understand how losses can accumulate over successive time periods. For example, a bad day can be followed by several other bad days, leading to a large overall loss. One would thus like to estimate the most probable value of the worst week, month, etc. In other words, one would like to construct the graph of \( \Lambda_{\text{VaR}}(N\tau) \), for a fixed overall investment horizon \( T \).

The answer is straightforward in the case where the price increments \( \delta P_{\tau} \) are Gaussian, \( P_{\tau} \) is also Gaussian, with a variance multiplied by \( N\tau \). At the same time, the number of different intervals of size \( N\tau \) for a fixed investment horizon \( T \) decreases by a factor \( \sqrt{T} \), one then finds:

\[
\Lambda_{\text{VaR}}(N\tau)|_T \simeq \sigma_{\text{close}} \sqrt{2N \log \left( \frac{T}{\sqrt{2\pi N\tau}} \right)}, \tag{3.16}
\]

where the notation \( \lfloor t \rfloor \) means that the investment period is fixed.\footnote{One can also take into account the average return, \( m_{\tau} = \langle \delta P \rangle \). In this case, one must subtract to \( \Lambda_{\text{VaR}}(N\tau)|_T \) the quantity \( -m_{\tau} N \) (The potential losses are indeed smaller if the average return is positive).} The main effect is then the \( \sqrt{N} \) increase of the volatility, up to a small logarithmic correction.

The case where \( P_{\tau}(\delta P) \) decreases as a power-law has been discussed in Chapter 1: for any \( N \), the far tail remains a power-law (that is progressively ‘eaten up’ by the Gaussian central part of the distribution if the exponent \( \mu \) of the power-law is greater than 2, but keeps its integrity whenever \( \mu < 2 \)). For finite \( N \), the largest moves will always be described by the power-law tail. Its amplitude \( \Lambda_{\text{VaR}} \) is simply multiplied by a factor \( N \) (cf. 1.5.2). Since the number of independent intervals is divided by the same factor \( N \), one finds:\footnote{cf. previous footnote.}

\[
\Lambda_{\text{VaR}}(N\tau)|_T = A \left( \frac{NT}{N\tau} \right)^{\frac{\mu}{2}}, \tag{3.17}
\]

independently of \( N \). Note however that for \( \mu > 2 \), the above result is only valid if \( \Lambda_{\text{VaR}}(N\tau) \) is located outside of the Gaussian central part, the size of which grows as \( \sigma \sqrt{N} \) (Fig. 3.4). In the opposite case, the Gaussian formula (3.16) should be used.

One can of course ask a slightly different question, by fixing not the investment horizon \( T \) but rather the probability of occurrence. This
3.1 Risk measurement and diversification

Extreme risks and optimal portfolios

Figure 3.4: Distribution of cumulated losses for finite \( N \): the central region, of width \( \sim \sigma N^{1/2} \), is Gaussian. The tails, however, remember the specific nature of the elementary distribution \( P_t(\delta x) \), and are in general fatter than Gaussian. The point is then to determine whether the confidence level \( \mathcal{P}_{\text{VaR}} \) puts \( \Lambda_{\text{VaR}}(N\tau) \) within the Gaussian part or far in the tails.

\[
\Lambda_{\text{VaR}}(N\tau) = N^{\frac{1}{2}} \Lambda_{\text{VaR}}(\tau). \tag{3.18}
\]

The growth of the Value-at-Risk with \( N \) is thus faster for small \( \mu \), as expected. The case of Gaussian fluctuations corresponds to \( \mu = 2 \).

**Drawdowns**

One can finally analyse the amplitude of a cumulated loss over a period that is not a priori limited. More precisely, the question is the following: knowing that the price of the asset today is equal to \( x_0 \), what is the probability that the lowest price ever reached in the future will be \( x_{\min} \); and how long such a drawdown will last? This is a classical problem in probability theory [Feller, vol. II, p. 404]. Let us denote as \( \Lambda_{\max} = x_0 - x_{\min} \) the maximum amplitude of the loss. If \( P_t(\delta x) \) decays at least exponentially when \( \delta x \to -\infty \), the result is that the tail of the distribution of \( \Lambda_{\max} \) behaves as an exponential:

\[
P(\Lambda_{\max}) \underset{\Lambda_{\max} \to \infty}{\sim} \exp \left( -\frac{\Lambda_{\max}}{\Lambda_0} \right),
\]

where \( \Lambda_0 > 0 \) is the finite solution of the following equation:

\[
\int d\delta x \exp \left( -\frac{\delta x}{\Lambda_0} \right) P_\tau(\delta x) = 1. \tag{3.20}
\]

Note that for this last equation to be meaningful, \( P_\tau(\delta x) \) should decrease at least as \( \exp(-|\delta x|/\Lambda_0) \) for large negative \( \delta x \)'s. It is clear (see below) that if \( P_t(\delta x) \) decreases as a power-law – say – the distribution of cumulated losses cannot decay exponentially, since it would then decay faster than that of individual losses!

It is interesting to show how this result can be obtained. Let us introduce the cumulative distribution \( P_>(\Lambda) = \int_0^\infty d\Lambda_{\max} P(\Lambda_{\max}) \). The first step of the walk, of size \( \delta x \), can either exceed \( -\Lambda \), or stay above it. In the first case, the level \( -\Lambda \) is immediately reached. In the second case, one recovers the very same problem as initially, but with \( -\Lambda \) shifted to \( -\Lambda - \delta x \). Therefore, \( P_>(\Lambda) \) obeys the following equation:

\[
P_>(\Lambda) = \int_{-\infty}^{-\Lambda} d\delta x P_\tau(\delta x) + \int_{-\Lambda}^{+\infty} d\delta x P_\tau(\delta x) P_>(\Lambda + \delta x). \tag{3.21}
\]

If one can neglect the first term in the right hand side for large \( \Lambda \) (which should be self-consistently checked), then, asymptotically, \( P_>(\Lambda) \) should obey the following equation:

\[
P_>(\Lambda) = \int_{-\Lambda}^{+\infty} d\delta x P_\tau(\delta x) P_>(\Lambda + \delta x). \tag{3.22}
\]

Inserting an exponential shape for \( P_>(\Lambda) \) then leads to Eq. (3.20) for \( \Lambda_0 \), in the limit \( \Lambda \to \infty \). This result is however only valid if \( P_t(\delta x) \) decays sufficiently fast for large negative \( \delta x \)'s, such that (3.20) has a non trivial solution.

Let us study two simple cases:

- For Gaussian fluctuations, one has:

\[
P_\tau(\delta x) = \frac{1}{\sqrt{2\pi D\tau}} \exp \left( -\frac{(\delta x - m)^2}{2D\tau} \right). \tag{3.23}
\]

Equation (3.20) thus becomes:

\[
-\frac{m}{\Lambda_0} + D + 2\Lambda_0^2 = 0 \Rightarrow \Lambda_0 = \frac{D}{2m}, \tag{3.24}
\]

$\Lambda_0$ gives the order of magnitude of the worst drawdown. One can understand the above result as follows: $\Lambda_0$ is the amplitude of the probable fluctuation over the characteristic time scale $\hat{T} = D/m^2$ introduced above. By definition, for times shorter than $\hat{T}$, the average return $m$ is negligible. Therefore, one has: $\Lambda_0 \propto \sqrt{D\hat{T}} = D/m$.

If $m = 0$, the very idea of worst drawdown looses its meaning: if one waits a long enough time, the price can then reach arbitrarily low values. It is intuitively clear that one should not put all his eggs in the same basket. A portfolio, composed of different assets with small mutual correlations, is less risky because the gains of some of the assets more or less compensate the loss of the others. Now, an investment with a small risk and small return must sometimes be prefered to a high yield, but very risky, investment.

The theoretical justification for the idea of diversification comes again from the CLT. A portfolio made up of $M$ uncorrelated assets and of equal volatility, with weight $1/M$, has an overall volatility reduced by a factor $\sqrt{M}$. Correspondingly, the amplitude (and duration) of the worst drawdown is divided by a factor $M$ (cf. Eq. (3.24)), which is obviously satisfying.

This qualitative idea stumbles over several difficulties. First of all, the fluctuations of financial assets are in general strongly correlated; this substantially decreases the possibility of true diversification, and requires a suitable theory to deal with this correlations and construct an ‘optimal’ portfolio: this is Markowitz’s theory, to be detailed below. Furthermore, since price fluctuations can be strongly non-Gaussian, a volatility-based measure of risk might be unadapted: one should rather try to minimise the Value-at-Risk of the portfolio. It is therefore interesting to look for an extension of the classical formalism, allowing one to devise minimum VaR portfolios. This will be presented in the next sections.

Now, one is immediately confronted with the problem of defining properly an ‘optimal’ portfolio. Usually, one invokes the rather abstract concept of ‘utility functions’, on which we shall briefly comment in this section, in particular to show that it does not naturally accommodate the notion of Value-at-Risk.

We will call $W_T$ the wealth of a given operator at time $t = T$. If one argues that the level of satisfaction of this operator is quantified by a certain function of $W_T$ only,7 which one usually calls the ‘utility function’ $U(W_T)$. This function is furthermore taken to be continuous and even twice differentiable. The postulated ‘rational’ behaviour for the operator is then to look for investments which maximise his expected utility, averaged over all possible histories of price changes:

$$\langle U(W_T) \rangle = \int dW_T \ P(W_T)U(W_T).$$

---

7But not of the whole ‘history’ of his wealth between $t = 0$ and $T$. One thus assumes that the operator is insensitive to what can happen between these two dates: this is not very realistic. One could however generalise the concept of utility function and deal with utility functionals $U((W(t))_{0 \leq t \leq T})$. 

The utility function should be non decreasing: a larger profit is clearly always more satisfying. One can furthermore consider the case where the distribution \( P(W_T) \) is sharply peaked around its mean value \( \langle W_T \rangle = W_0 + mT \). Performing a Taylor expansion of \( \langle U(W_T) \rangle \) around \( U(W_0 + mT) \) to second order, one deduces that the utility function must be such that:

\[
\frac{d^2 U}{dW^2} < 0. \tag{3.28}
\]

This property reflects the fact that for the same average return, a less risky investment should always be preferred.

A simple example of utility function compatible with the above constraints is the exponential function \( U(W_T) = -\exp[-W_T/w_0] \). Note that \( w_0 \) has the dimensions of a wealth, thereby fixing a wealth scale in the problem. A natural candidate is the initial wealth of the operator, \( w_0 \propto W_0 \). If \( P(W_T) \) is Gaussian, of mean \( W_0 + mT \) and variance \( DT \), one finds that the expected utility is given by:

\[
\langle U \rangle = -\exp\left[-\frac{W_0}{w_0} - \frac{T}{w_0} (m - D) \right]. \tag{3.29}
\]

One could think of constructing a utility function with no intrinsic wealth scale by choosing a power-law: \( U(W_T) = (W_T/w_0)^\alpha \) with \( \alpha < 1 \) to ensure the correct convexity. Indeed, in this case a change of \( w_0 \) can be reabsorbed in a change of scale of the utility function itself. However, this definition cannot allow for negative final wealths, and is thus problematic.

Despite the fact that these postulates sound reasonable, and despite the very large number of academic studies based on the concept of utility function, this axiomatic approach suffers from a certain number of fundamental flaws. For example, it is not clear that one could ever measure the utility function used by a given agent on the markets. The theoretical results are thus:

- Either relatively weak, because independent of the special form of the utility function, and only based on its general properties.
- Or rather arbitrary, because based on a specific, but unjustified, form for \( U(W_T) \).

On the other hand, the idea that the utility function is regular is probably not always realistic. The satisfaction of an operator is often governed by rather sharp thresholds, separated by regions of indifference (Fig. 3.5). For example, one can be in a situation where a specific project can only be achieved if the profit \( \Delta W = W_T - W_0 \) exceeds a certain amount. Symmetrically, the clients of a fund manager will take their money away as soon as the losses exceed a certain value: this is the strategy of ‘stop-losses’, which fix a level for acceptable losses, beyond which the position is closed. The existence of option markets (which allow one to limit the potential losses below a certain level – see Chapter 4), or of items the price of which is $99 rather than $100, are concrete examples of the existence of these thresholds where ‘satisfaction’ changes abruptly. Therefore, the utility function \( U \) is not necessarily continuous. This remark is actually intimately related to the fact that the Value-at-Risk is often a better measure of risk than the variance. Let us indeed assume that the operator is ‘happy’ if \( \Delta W > -\Lambda \) and ‘unhappy’ whenever \( \Delta W < -\Lambda \). This corresponds formally to the following utility function:

\[
U_\Lambda(\Delta W) = \begin{cases} 
U_1 & (\Delta W > -\Lambda) \\
U_2 & (\Delta W < -\Lambda)
\end{cases}, \tag{3.30}
\]

with \( U_2 - U_1 < 0 \).

The expected utility is then simply related to the loss probability:

\[
\langle U_\Lambda \rangle = U_1 + (U_2 - U_1) \int_{-\infty}^{-\Lambda} d\Delta WP(\Delta W) = U_1 - |U_2 - U_1| P. \tag{3.31}
\]

Therefore, optimising the expected utility is in this case tantamount to minimising the probability of loosing more that \( \Lambda \). Despite this rather appealing property, which certainly corresponds to the behaviour of some market operators, the function \( U_\Lambda(\Delta W) \) does not satisfy the above criteria (continuity and negative curvature).

Confronted to the problem of choosing between risk (as measured by the variance) and return, another very natural strategy (for those not acquainted with utility functions) would be to compare the average return to the potential loss \( \sqrt{DDT} \). This can thus be thought of as defining a risk-corrected, ‘pessimistic’ estimate of the profit, as:

\[
m_\Lambda T = mT - \lambda \sqrt{DDT}, \tag{3.32}
\]

where \( \lambda \) is an arbitrary coefficient that measures the pessimism (or the risk aversion) of the operator. A rather natural procedure would then...
Figure 3.5: Example of a ‘utility function’ with thresholds, where the utility function is non continuous. These thresholds correspond to special values for the profit, or the loss, and are often of purely psychological origin.

be to look for the optimal portfolio which maximises the risk corrected return \( m \lambda \). However, this optimal portfolio cannot be obtained using the standard utility function formalism. For example, Eq. (3.29) shows that the object which should be maximised is \( mT - \lambda \sqrt{DT} \) but rather \( mT - DT/2w_0 \). This amounts to compare the average profit to the square of the potential losses, divided by the reference wealth scale \( w_0 \) – a quantity that depends \( a \ priori \) on the operator.\(^9\) On the other hand, the quantity \( mT - \lambda \sqrt{DT} \) is directly related (at least in a Gaussian world) to the Value-at-Risk \( \Lambda_{\text{VaR}} \), cf. (3.16).

This can be expressed slightly differently: a reasonable objective could be to maximise the value of the ‘probable gain’ \( G_p \), such that the probability of earning more is equal to a certain probability \( p \):

\[
\int_{G_p}^{+\infty} d\Delta W P(\Delta W) = p. \tag{3.33}
\]

In the case where \( P(\Delta W) \) is Gaussian, this amounts to maximise \( mT - \lambda \sqrt{DT} \), where \( \lambda \) is related to \( p \) in a simple manner. Now, one can

\(^9\)This comparison is actually meaningful, since it corresponds to comparing the reference wealth \( w_0 \) to the order of magnitude of the worst drawdown \( D/m \); cf. Eq. (3.24).

\(^{10}\)Maximising \( G_p \) is thus equivalent to minimising \( \Lambda_{\text{VaR}} \) such that \( F_{\text{VaR}} = 1 - p \).

show that it is impossible to construct a utility function such that, in the general case, different strategies can be ordered according to their probable gain \( G_p \). Therefore, the concepts of loss probability, Value-at-Risk or probable gain cannot be accommodated naturally within the framework of utility functions. Still, the idea that the quantity which is of most concern and that should be optimised is the Value-at-Risk sounds perfectly rational. This is at least the conceptual choice that we make in the present monograph.

### 3.1.5 Conclusion

Let us now recapitulate the main points of this section:

- The usual measure of risk through a Gaussian volatility is not always adapted to the real world. The tails of the distributions, where the large events lie, are very badly described by a Gaussian law: this leads to a systematic underestimation of the extreme risks. Sometimes, the measurement of the volatility on historical data is difficult, precisely because of the presence of these large fluctuations.

- The measure of risk through the probability of loss, or the Value-at-Risk, on the other hand, precisely focusses on the tails. Extreme events are considered as the true source of risk, while the small fluctuations contribute to the ‘centre’ of the distributions (and contribute to the volatility) can be seen as a background noise, inherent to the very activity of financial markets, but not relevant for risk assessment.

- From a theoretical point of view, this definition of risk (based on extreme events) does not easily fit into the classical ‘utility function’ framework. The minimisation of a loss probability rather assumes that there exists well defined thresholds (possibly different for each operator) where the ‘utility function’ is discontinuous.\(^{11}\) The concept of ‘Value-at-Risk’, or probable gain, cannot be naturally dealt with using utility functions.

### 3.2 Portfolios of uncorrelated assets

The aim of this section is to explain, in the very simple case where all assets that can be mixed in a portfolio are uncorrelated, how the trade-off between risk and return can be dealt with. (The case where some

\(^{11}\)It is possible that the presence of these thresholds play an important role in the fact that the price fluctuations are strongly non Gaussian.
correlations between the asset fluctuations exist will be considered in the next section). On thus considers a set of $M$ different risky assets $X_i$, $i = 1, ..., M$ and one risk-less asset $X_0$. The number of asset $i$ in the portfolio is $n_i$, and its present value is $x_i^0$. If the total wealth to be invested in the portfolio is $W$, then the $n_i$’s are constrained to be such that $\sum_{i=0}^{M} n_i x_i^0 = W$. We shall rather use the weight of asset $i$ in the portfolio, defined as: $p_i = n_i x_i^0 / W$, which therefore must be normalised to one: $\sum_{i=0}^{M} p_i = 1$. The $p_i$ can be negative (short positions). The value of the portfolio at time $T$ is given by: $S = \sum_{i=0}^{M} n_i x_i(T) = W \sum_{i=0}^{M} p_i x_i(T) / x_i^0$. In the following, we will set the initial wealth $W$ to 1, and redefine each asset $i$ in such a way that all initial prices are equal to $x_i^0 = 1$. (Therefore, the average return $m_i$ and variance $D_i$ that we will consider below must be understood as relative, rather than absolute.)

One furthermore assumes that the average return $m_i$ is known. This hypothesis is actually very strong, since it assumes for example that past returns can be used as estimators of future returns, i.e. that time series are to some extent stationary. However, this is very far from the truth: the life of a company (in particular high-tech ones) is very clearly non stationary; a whole sector of activity can be booming or collapsing, depending upon global factors, not graspable within a purely statistical framework. Furthermore, the markets themselves evolve with time, and it is clear that some statistical parameters do depend on time, and have significantly shifted over the past twenty years. This means that the empirical determination of the average return is difficult: volatilities are such that at least several years are needed to obtain a reasonable signal to noise ratio – this time must indeed be large compared to the ‘security time’ $\hat{T}$. But as discussed above, several years is also the time scale over which the intrinsically non stationary nature of the markets starts being important.

One should thus rather understand $m_i$ as an ‘expected’ (or anticipated) future return, which includes some extra information (or intuition) available to the investor. These $m_i$’s can therefore vary from one investor to the next. The relevant question is then to determine the composition of an optimal portfolio compatible with the information contained in the knowledge of the different $m_i$’s.

The determination of the risk parameters is a priori subject to the same caveat. We have actually seen in Section 2.7 that the empirical determination of the correlation matrices contains a large amount of noise, which blur the true information. However, the statistical nature of the fluctuations seems to be more robust in time than the average returns. The analysis of past price changes distributions appears to be, to a certain extent, predictive for future price fluctuations. It however sometimes happen that the correlations between two assets change abruptly.

### 3.2 Portfolios of uncorrelated assets

#### 3.2.1 Uncorrelated Gaussian assets

Let us suppose that the variation of the value of the $i^{th}$ asset $X_i$ over the time interval $T$ is Gaussian, centred around $m_i T$ and of variance $D_i T$. The portfolio $p = \{p_0, p_1, ..., p_M \}$ as a whole also obeys Gaussian statistics (since the Gaussian is stable). The average return $m_p$ of the portfolio is given by:

$$m_p = \sum_{i=0}^{M} p_i m_i = m_0 + \sum_{i=1}^{M} p_i (m_i - m_0),$$

(3.34)

where we have used the constraint $\sum_{i=0}^{M} p_i = 1$ to introduce the excess return $m_i - m_0$, as compared to the risk-free asset ($i = 0$). If the $X_i$’s are all independent, the total variance of the portfolio is given by:

$$D_p = \sum_{i=1}^{M} p_i^2 D_i,$$

(3.35)

(since $D_0$ is zero by assumption). If one tries to minimise the variance without any constraint on the average return $m_p$, one obviously finds the trivial solution where all the weight is concentrated on the risk-free asset:

$$p_i = 0 \quad (i \neq 0); \quad p_0 = 1.$$  

(3.36)

On the opposite, the maximisation of the return without any risk constraint leads to a full concentration of the portfolio on the asset with the highest return.

More realistically, one can look for a tradeoff between risk and return, by imposing a certain average return $m_p$, and by looking for the less risky portfolio (for Gaussian assets, risk and variance are identical). This can be achieved by introducing a Lagrange multiplier in order to enforce the constraint on the average return:

$$\delta (D_p - \zeta m_p) \bigg|_{p_i = p_i^*} = 0 \quad (i \neq 0),$$

(3.37)

while the weight of the risk-free asset $p_0^*$ is determined via the equation $\sum_i p_i^* = 1$. The value of $\zeta$ is ultimately fixed such that the average value of the return is precisely $m_p$. Therefore:

$$2p_i^* D_i = (m_i - m_0),$$

(3.38)
3 Extreme risks and optimal portfolios

3.2 Portfolios of uncorrelated assets

The least risky portfolio corresponds to the one such that $\zeta = 0$ (no constraint on the average return):

$$p_i^* = \frac{1}{Z D_i} \quad Z = \sum_{j=1}^{M} \frac{1}{D_j}.$$  (3.42)

Its total variance is given by $D_p^* = 1/Z$. If all the $D_i$’s are of the same order of magnitude, one has $Z \sim M/D$; therefore, one finds the result, expected from the CLT, that the variance of the portfolio is $M$ times smaller than the variance of the individual assets.

In practice, one often adds extra constraints to the weights $p_i$ in the form of linear inequalities, such as $p_i^* \geq 0$ (no short positions). The solution is then more involved, but is still unique. Geometrically, this amounts to look for the restriction of a paraboloid to an hyperplane, which remains a paraboloid. The efficient border is then shifted downwards (Fig. 3.6). A much richer case is when the constraint is non-linear.

For example, on futures markets, margin calls require that a certain amount of money is left as a deposit, whether the position is long ($p_i > 0$) or short ($p_i < 0$). One can then impose a leverage constraint, such that $\sum_{i=1}^{M} |p_i| = f$, where $f$ is the fraction of wealth invested as a deposit. This constraint leads to a much more complex problem, similar to the one encountered in hard optimisation problems, where an exponentially large (in $m$) number of quasi-degenerate solutions can be found.\(^{12}\)

**Effective asset number in a portfolio**

It is useful to introduce an objective way to measure the diversification, or the asset concentration, in a given portfolio. Once such an indicator is available, one can actually use it as a constraint to construct portfolios with a minimum degree of diversification. Consider the quantity $Y_2$ defined as:

$$Y_2 = \sum_{i=1}^{M} (p_i^*)^2.$$  (3.43)

If a subset $M' \leq M$ of all $p_i^*$ are equal to $1/M'$, while the others are zero, one finds $Y_2 = 1/M'$. More generally, $Y_2$ represents the average weight of an asset in the portfolio, since it is constructed as the average of $p_i^*$ itself. It is thus natural to define the ‘effective’ number of assets in the portfolio as $M_{\text{eff}} = 1/Y_2$. In order to avoid an overconcentration of the

3.2 Portfolios of uncorrelated assets

3.2.1 Extreme risks and optimal portfolios

The optimal portfolio satisfying $M_{\text{eff}} \geq 2$ is therefore given by the standard portfolio for returns between $r_1$ and $r_2$ and by the $M_{\text{eff}} = 2$ portfolios otherwise.

Figure 3.7: Example of a standard efficient border $\zeta'' = 0$ (thick line) with four risky assets. If one imposes that the effective number of assets is equal to 2, one finds the sub-efficient border drawn in dotted line, which touches the efficient border at $r_1$, $r_2$. The inset shows the effective asset number of the unconstrained optimal portfolio ($\zeta'' = 0$) as a function of average return. The optimal portfolio satisfying $M_{\text{eff}} \geq 2$ is therefore given by the standard portfolio for returns between $r_1$ and $r_2$ and by the $M_{\text{eff}} = 2$ portfolios otherwise.

portfolio on very few assets (a problem often encountered in practice), one can look for the optimal portfolio with a given value for $Y_q$. This amounts to introducing another Lagrange multiplier $\zeta''$, associated to $Y_q$. The equation for $p_i^*$ then becomes:

$$ p_i^* = \frac{\zeta'' m_i + \zeta'}{2(D_i + \zeta'')} \equiv \frac{\zeta'' m_i + \zeta'}{2(D_i + \zeta'')} , \quad (3.44) $$

An example of the modified efficient border is given in Fig. 3.7.

More generally, one could have considered the quantity $Y_q$ defined as:

$$ Y_q = \sum_{i=1}^{M} (p_i^*)^q , \quad (3.45) $$

and used it to define the effective number of assets via $Y_q = M_{\text{eff}}^{1-q}$. It is interesting to note that the quantity of missing information (or entropy) $I$ associated to the very choice of the $p_i^*$’s is related to $Y_q$ when $q \to 1$. Indeed, one has:

$$ I = - \sum_{i=1}^{M} p_i^* \log(\frac{p_i^*}{e}) \equiv 1 - \frac{\partial Y_q}{\partial q} \bigg|_{q=1} . \quad (3.46) $$

Approximating $Y_q$ as a function of $q$ by a straight line thus leads to $I \simeq 1 - Y_2$.

3.2.2 Uncorrelated ‘power law’ assets

As we have already underlined, the tails of the distributions are often non-Gaussian. In this case, the minimisation of the variance is not necessarily equivalent to an optimal control of the large fluctuations. The case where these distribution tails are power-laws is interesting because one can then explicitly solve the problem of the minimisation of the Value-at-Risk of the full portfolio. Let us thus assume that the fluctuations of each asset $X_i$ are described, in the region of large losses, by a probability density that decays as a power-law:

$$ P_T(\delta x_i) \propto \frac{\mu A_i^\mu}{|\delta x_i|^{1+\mu}} , \quad (3.47) $$

with an arbitrary exponent $\mu$, restricted however to be larger than 1, such that the average return is well defined. (The distinction between the cases $\mu < 2$, for which the variance diverges, and $\mu > 2$ will be considered below). The coefficient $A_i$ provides an order of magnitude for the extreme losses associated to the asset $i$ (cf. (3.14)).

As we have mentioned in Section 1.5.2, the power-law tails are interesting because they are stable upon addition: the tail amplitudes $A_i$ (that generalise the variance) simply add to describe the far-tail of the distribution of the sum. Using the results of Appendix B, one can show that if the asset $X_i$ is characterised by a tail amplitude $A_i^\mu$, the quantity $p_i X_i$ has a tail amplitude equal to $p_i^\mu A_i^\mu$. The tail amplitude of the global portfolio $\mathbf{p}$ is thus given by:

$$ A_{p}^\mu = \sum_{i=1}^{M} p_i^\mu A_i^\mu , \quad (3.48) $$

and the probability that the loss exceeds a certain level $\Lambda$ is given by $P = A_{p}^\mu / \Lambda^\mu$. Hence, independently of the chosen loss level $\Lambda$, the minimisation of the loss probability $P$ requires the minimisation of the tail amplitude $A_{p}^\mu$; the optimal portfolio is therefore independent of $\Lambda$. (This
is not true in general: see 3.2.4). The minimisation of \( A_i^\mu \) for a fixed average return \( m_p \) leads to the following equations (valid if \( \mu > 1 \)):

\[
\mu A_i^{\mu-1} A_i^\mu = \zeta (m_i - m_0), 
\]

(3.49)

with an equation to fix \( \zeta \):

\[
\left( \frac{\zeta}{\mu} \right)^{1/\mu} \sum_{i=1}^{M} \frac{(m_i - m_0)^{1/\mu}}{A_i^{1/\mu}} = m_p - m_0. 
\]

(3.50)

The optimal loss probability is then given by:

\[
P^* = 1 \left( \frac{\zeta}{\mu} \right)^{1/\mu} \sum_{i=1}^{M} \frac{(m_i - m_0)^{1/\mu}}{A_i^{1/\mu}}. 
\]

(3.51)

Therefore, the concept of ‘efficient border’ is still valid in this case: in the plane return/probablity of loss, it is similar to the dotted line of Fig. 3.6. Eliminating \( \zeta \) from the above two equations, one finds that the shape of this line is given by \( P^* \propto (m_p - m_0)\mu \). The parabola is recovered in the limit \( \mu = 2 \).

In the case where the risk-free asset cannot be included in the portfolio, the optimal portfolio which minimises extreme risks with no constraint on the average return is given by:

\[
p_i^* = \frac{1}{ZA_i^{1/\mu}} Z = \sum_{j=1}^{M} A_i^{-\mu/\mu}, 
\]

(3.52)

and the corresponding loss probability is equal to:

\[
P^* = 1 \frac{1}{A^\mu} Z^{1-\mu}. 
\]

(3.53)

If all assets have comparable tail amplitudes \( A_i \sim A \), one finds that \( Z \sim MA^{1-\mu} \). Therefore, the probability of large losses for the optimal portfolio is a factor \( M^{\mu-1} \) smaller than the individual probability of loss.

Note again that this result is only valid if \( \mu > 1 \). If \( \mu < 1 \), one finds that the risk increases with the number of assets \( M \). In this case, when the number of assets is increased, the probability of an unfavourable event also increases – indeed, for \( \mu < 1 \) this largest event is so large that it dominates over all the others. The minimisation of risk in this case leads to \( p_{\text{min}} = 1 \), where \( i_{\text{min}} \) is the least risky asset, in the sense that \( A_{i_{\text{min}}} = \min \{A_i^\mu\} \).

3.2 Portfolios of uncorrelated assets

One should now distinguish the cases \( \mu < 2 \) and \( \mu > 2 \). Despite of the fact that the asymptotic power-law behaviour is stable under addition for all values of \( \mu \), the tail is progressively ‘eaten up’ by the centre of the distribution for \( \mu > 2 \), since the CLT applies. Only when \( \mu < 2 \) does this tail remain untouched. We thus again recover the arguments of Section 1.6.4, already encountered when we discussed the time dependence of the VaR. One should therefore distinguish two cases: if \( D_p \) is the variance of the portfolio \( p \) (which is finite if \( \mu > 2 \)), the distribution of the changes \( \delta S \) of the value of the portfolio \( p \) is approximately Gaussian if \( |\delta S| \leq \sqrt{D_p T \log(M)} \), and becomes a power-law with a tail amplitude given by \( A_p^\mu \) beyond this point. The question is thus whether the loss level \( \Lambda \) that one wishes to control is smaller or larger than this crossover value:

- If \( \Lambda \ll \sqrt{D_p T \log(M)} \), the minimisation of the VaR becomes equivalent to the minimisation of the variance, and one recovers the Markowitz procedure explained in the previous paragraph in the case of Gaussian assets.
- If on the contrary \( \Lambda \gg \sqrt{D_p T \log(M)} \), then the formulae established in the present section are valid even when \( \mu > 2 \).

Note that the growth of \( \sqrt{\log(M)} \) with \( M \) is so slow that the Gaussian CLT is not of great help in the present case. The optimal portfolios in terms of the VaR are not those with the minimal variance, and vice-versa.

3.2.3 ‘Exponential’ assets

Suppose now that the distribution of price variations is a symmetric exponential around a zero mean value \( (m_i = 0) \):

\[
P(\delta x_i) = \frac{\alpha_i}{\sqrt{2 \pi}} \exp[-\alpha_i |\delta x_i|], 
\]

(3.54)

where \( \alpha_i^{-1} \) gives the order of magnitude of the fluctuations of \( X_i \) (more precisely, \( \sqrt{2}/\alpha_i \) is the RMS of the fluctuations.). The variations of the full portfolio \( p \), defined as \( \delta S = \sum_{i=1}^{M} p_i \delta x_i \), are distributed according to:

\[
P(\delta S) = \int \frac{dz}{2\pi} \prod_{i=1}^{M} \left[ 1 + (z p_i \alpha_i^{-1})^2 \right]^{1/2} \exp[i z \delta S], 
\]

(3.55)

where we have used (1.50) and the fact that the Fourier transform of the exponential distribution (3.54) is given by:

\[
P(z) = \frac{1}{1 + (z \alpha_i^{-1})^2}. 
\]

(3.56)
Now, using the method of residues, it is simple to establish the following expression for \( P(\delta S) \) (valid whenever the \( \alpha_i / p_i \) are all different):

\[
P(\delta S) = \frac{1}{2} \sum_{i=1}^{M} \frac{\alpha_i}{p_i} \prod_{j \neq i} \left( 1 - \left[ \frac{p_i \alpha_j}{p_j \alpha_i} \right]^2 \right) \exp \left[ -\frac{\alpha_i}{p_i} |\delta S| \right].
\] (3.57)

The probability for extreme losses is thus equal to:

\[
P(\delta S < -\Lambda) \overset{\Lambda \to -\infty}{\simeq} \frac{1}{2} \prod_{j \neq i} \left( 1 - \left[ \frac{p_i \alpha_j}{p_j \alpha_i} \right]^2 \right) \exp[-\alpha^* \Lambda],
\] (3.58)

where \( \alpha^* \) is equal to the smallest of all ratios \( \alpha_j / p_i \), and \( \alpha^* \) the corresponding value of \( i \). The order of magnitude of the extreme losses is therefore given by \( 1 / \alpha^* \). This is then the quantity to be minimised in a Value-at-Risk context. This amounts to choose the \( p_i \)’s such that \( \min_{i} \{ \alpha_i / p_i \} \) is as large as possible.

This minimisation problem can be solved using the following trick. One can write formally that:

\[
\frac{1}{\alpha^*} = \max_i \left\{ \frac{p_i}{\alpha_i} \right\} = \lim_{\mu \to \infty} \left( \sum_{i=1}^{M} \frac{p_i^\mu}{\alpha_i^\mu} \right)^{1/\mu}.
\] (3.59)

This equality is true because in the limit \( \mu \to \infty \), the largest term of the sum dominates over all the others. (The choice of the notation \( \mu \) is on purpose: see below.) For \( \mu \) large but fixed, one can perform the minimisation with respect to the \( p_i \)’s, using a Lagrange multiplier to enforce normalisation. One easily finds that:

\[
p_i^{\mu-1} \propto \alpha_i^\mu.
\] (3.60)

In the limit \( \mu \to \infty \), and imposing \( \sum_{i=1}^{M} p_i = 1 \), one finally obtains:

\[
p_i^* = \frac{\alpha_i}{\sum_{j=1}^{M} \alpha_j},
\] (3.61)

which leads to \( \alpha^* = \sum_{i=1}^{M} \alpha_i \). In this case, however, all \( \alpha_i / p_i^* \) are equal to \( \alpha^* \) and the result (3.57) must be slightly altered. However, the asymptotic exponential fall-off, governed by \( \alpha^* \), is still true (up to polynomial corrections: see cf. 1.6.4). One again finds that if all the \( \alpha_i \)’s are comparable, the potential losses, measured through \( 1 / \alpha^* \), are divided by a factor \( M \).

Note that the optimal weights are such that \( p_i^* \propto \alpha_i \) if one tries to minimise the probability of extreme losses, while one would have found \( p_i^* \propto \alpha_i^2 \) if the goal was to minimise the variance of the portfolio, corresponding to \( \mu = 2 \) (cf. (3.42)). In other words, this is an explicit example where one can see that minimising the variance actually increases the Value-at-Risk.

Formally, as we have noticed in Section 1.3.4, the exponential distribution corresponds to the limit \( \mu \to \infty \) of a power-law distribution: an exponential decay is indeed more rapid than any power-law. Technically, we have indeed established in this section that the minimisation of extreme risks in the exponential case is identical to the one obtained in the previous section in the limit \( \mu \to \infty \) (see (3.52)).

### 3.2.4 General case: optimal portfolio and VaR (*)

In all the cases treated above, the optimal portfolio is found to be independent of the chosen loss level \( \Lambda \). For example, in the case of assets with power-law tails, the minimisation of the loss probability amounts to minimising the tail amplitude \( A^\mu_i \), independently of \( \Lambda \). This property is however not true in general, and the optimal portfolio does indeed depend on the risk level \( \Lambda \), or, equivalently, on the temporal horizon over which risk must be ‘tamed’. Let us for example consider the case where all assets are power-law distributed, but with a tail index \( \mu_i \) that depends on the asset \( X_i \). The probability that the portfolio \( p \) experiences of loss greater than \( \Lambda \) is given, for large values of \( \Lambda \), by:

\[
P(\delta S < -\Lambda) = \sum_{i=1}^{M} p_i^{\mu_i} A^\mu_i / A^\mu_i = \sum_{i=1}^{M} \left( \frac{A^\mu_i}{\mu_i A^\mu_i} \right)^{1/\mu_i}.
\] (3.62)

Looking for the set of \( p_i \)’s which minimises the above expression (without constraint on the average return) then leads to:

\[
p_i^* = \frac{1}{Z} \left( \frac{A^\mu_i}{\mu_i A^\mu_i} \right)^{1/\mu_i},
\] (3.63)

This example shows that in the general case, the weights \( p_i^* \) explicitly depend on the risk level \( \Lambda \). If all the \( \mu_i \)’s are equal, \( A^\mu_i \) factors out and disappears from the \( p_i \)’s.

Another interesting case is that of weakly non-Gaussian assets, such that the first correction to the Gaussian distribution (proportional to the kurtosis \( \kappa_i \) of the asset \( X_i \)) is enough to describe faithfully the
non-Gaussian effects. The variance of the full portfolio is given by
\[ \text{Var}(p) = \sum_{i=1}^{M} p_i^2 \sigma_i^2, \]
while the kurtosis is equal to:
\[ \kappa_p = \sum_{i=1}^{M} p_i^4 \kappa_i / \sigma_i^4. \]
The probability that the portfolio plummets by an amount larger than \( \Lambda \) is therefore given by:
\[ P(\delta S < -\Lambda) \simeq \mathcal{P}_{G^>}(\frac{\Lambda}{\sqrt{D_p T}}) + \kappa_p \frac{\Lambda}{\sqrt{D_p T}}, \]
where \( \mathcal{P}_{G^>} \) is related to the error function (cf. 1.6.3) and
\[ h(u) = \frac{\exp(-u^2)}{\sqrt{2\pi}} (u^3 - 3u). \]
To first order in kurtosis, one thus finds that the optimal weights \( p_i^* \) (without fixing the average return) are given by:
\[ p_i^* = \frac{\varsigma^2}{2D_i} - \frac{\kappa_i \varsigma^3}{D_i^2} \left( \frac{\Lambda}{\sqrt{D_p T}} \right), \]
where \( \varsigma \) is another function, positive for large arguments, and \( \varsigma' \) is fixed by the condition \( \sum_{i=1}^{M} p_i^* = 1 \). Hence, the optimal weights do depend on the risk level \( \Lambda \), via the kurtosis of the distribution. Furthermore, as could be expected, the minimisation of extreme risks leads to a reduction of the weights of the assets with a large kurtosis \( \kappa_i \).

### 3.3 Portfolios of correlated assets

The aim of the previous section was to introduce, in a somewhat simplified context, the most important ideas underlying portfolio optimisation, in a Gaussian world (where the variance is minimised) or in a non-Gaussian world (where the quantity of interest is the Value-at-Risk). In reality, the fluctuations of the different assets are often strongly correlated (or anti-correlated). For example, an increase of short term interest rates often leads to a drop in share prices. All the stocks of the New-York Stock Exchange behave, to a certain extent, similarly. These correlations of course modify completely the composition of the optimal portfolios, and actually make diversification more difficult. In a sense, the number of effectively independent assets is decreased from the true number of assets \( M \).

#### 3.3.1 Correlated Gaussian fluctuations

Let us first consider the case where all the fluctuations \( \delta x_i \) of the assets \( X_i \) are Gaussian, but with arbitrary correlations. These correlations are described in terms of a (symmetric) correlation matrix \( C_{ij} \), defined as:
\[ C_{ij} = \langle \delta x_i \delta x_j \rangle - m_i m_j. \]  
This means that the joint distribution of all the fluctuations \( \delta x_1, \delta x_2, \ldots, \delta x_M \) is given by:
\[ P(\delta x_1, \delta x_2, \ldots, \delta x_M) \propto \exp \left[ -\frac{1}{2} \sum_{ij} (\delta x_j - m_j)(C^{-1})_{ij}(\delta x_j - m_j) \right], \]
where the proportionality factor is fixed by normalisation and is equal to \( 1/\sqrt{(2\pi)^N \det C} \), and \((C^{-1})_{ij}\) denotes the elements of the matrix inverse of \( C \).

An important property of correlated Gaussian variables is that they can be decomposed into a weighted sum of independent Gaussian variables \( e_a \), of mean zero and variance equal to \( D_a \):
\[ \delta x_i = m_i + \sum_{a=1}^{M} O_{ia} e_a \]  
\[ \langle e_a e_b \rangle = \delta_{a,b} D_a. \]
The \( \{e_a\} \) are usually referred to as the ‘explicative factors’ (or principal components) for the asset fluctuations. They sometimes have a simple economic interpretation.

The coefficients \( O_{ia} \) give the weight of the factor \( e_a \) in the evolution of the asset \( X_i \). These can be related to the correlation matrix \( C_{ij} \) by using the fact that the \( \{e_a\} \)'s are independent. This leads to:
\[ C_{ij} = \sum_{a,b=1}^{M} O_{ia} O_{jb} \langle e_a e_b \rangle = \sum_{a=1}^{M} O_{ia} O_{ja} D_a, \]
or, seen as a matrix equality: \( C = O \hat{D} O^\dagger \), where \( O^\dagger \) denotes the matrix transposed of \( O \) and \( \hat{D} \) the diagonal matrix obtained from the \( D_a \)'s. This last expression shows that the \( D_a \)'s are the eigenvalues of the matrix \( C_{ij} \), while \( O \) is the orthogonal matrix allowing to go from the set of assets \( i \)'s to the set of explicative factors \( e_a \).

The fluctuations \( \delta S \) of the global portfolio \( p \) are then also Gaussian (since \( \delta S \) is a weighted sum of the Gaussian variables \( e_a \)), of mean \( m_p = \sum_{i=1}^{M} p_i (m_i - m_0) + m_0 \) and variance:
\[ D_p = \sum_{i,j=1}^{M} p_i p_j C_{ij}. \]
The minimisation of $D_p$ for a fixed value of the average return $m_p$ (and with the possibility of including the risk-free asset $X_0$) leads to an equation generalising Eq. (3.38):

$$2 \sum_{j=1}^{M} C_{ij} p_j^* = \zeta (m_i - m_0), \quad (3.72)$$

which can be inverted as:

$$p_i^* = \frac{\zeta}{2} \sum_{j=1}^{M} C_{ij}^{-1} (m_j - m_0). \quad (3.73)$$

This is Markowitz’s classical result (cf. [Markowitz, Elton & Gruber]).

In the case where the risk-free asset is excluded, the minimum variance portfolio is given by:

$$p_i^* = \frac{1}{2} \sum_{j=1}^{M} C_{ij}^{-1} Z = \sum_{i,j=1}^{M} C_{ij}^{-1}. \quad (3.74)$$

Actually, the decomposition (3.69) shows that, provided one shifts to the basis where all assets are independent (through a linear combination of the original assets), all the results obtained above in the case where the correlations are absent (such as the existence of an efficient border, etc.) are still valid when correlations are present.

In the more general case of non-Gaussian assets of finite variance, the total variance of the portfolio is still given by: $\sum_{i,j=1}^{M} \mu_i \mu_j C_{ij}$, where $C_{ij}$ is the correlation matrix. If the variance is an adequate measure of risk, the composition of the optimal portfolio is still given by Eqs. (3.73, 3.74). Let us however again emphasise that, as discussed in Section 2.7, the empirical determination of the correlation matrix $C_{ij}$ is difficult, in particular when one is concerned with the small eigenvalues of this matrix and their corresponding eigenvectors.

**The CAPM and its limitations**

Within the above framework, all optimal portfolios are proportional to one another, that is, they only differ through the choice of the factor $\zeta$. Since the problem is linear, this means that the linear superposition of optimal portfolios is still optimal. If all the agents on the market choose their portfolio using this optimisation scheme (with the same values for the average return and the correlation coefficients – clearly quite an absurd hypothesis), then the ‘market portfolio’ (i.e. the one obtained by taking all assets in proportion of their market capitalisation) is an optimal portfolio. This remark is at the origin of the ‘CAPM’ (Capital Asset Pricing Model), which aims at relating the average return of an asset with its covariance with the ‘market portfolio’. Actually, for any optimal portfolio $p$, one can express $m_p - m_0$ in terms of the $p^*$, and use Eq. (3.73) to eliminate $\zeta$, to obtain the following equality:

$$m_i - m_0 = \beta_i [m_p - m_0] \quad \beta_i \equiv \frac{(\delta x_i - m_i)(\delta S - m_p)}{\langle (\delta S - m_p)^2 \rangle}. \quad (3.75)$$

The covariance coefficient $\beta_i$ is often called the ‘$\beta$’ of asset $i$ when $p$ is the market portfolio.

This relation is however not true for other definitions of optimal portfolios. Let us define the generalised kurtosis $K_{ijkl}$ that measures the first correction to Gaussian statistics, from the joint distribution of the asset fluctuations:

$$P(\delta x_1, \delta x_2, ..., \delta x_M) = \int \int \cdots \int \prod_{j=1}^{M} \frac{dz_j}{2\pi} \exp \left[ -i \sum_{j} z_j (\delta x_j - m_j) - \frac{1}{2} \sum_{ij} z_i C_{ij} z_j + \frac{1}{4!} \sum_{ijkl} K_{ijkl} z_i z_j z_k z_l + \ldots \right]. \quad (3.76)$$

If one tries to minimise the probability that the loss is greater than a certain $\Lambda$, a generalisation of the calculation presented above (cf. (3.66)) leads to:

$$p_i^* = \frac{\zeta}{2} \sum_{j=1}^{M} C_{ij}^{-1} (m_j - m_0) - \zeta^2 \hat{h} \left( \frac{\Lambda}{\sqrt{D_{p^*}}} \right) \sum_{j,k,l} \sum_{j',k',l'} C_{ijkl} C_{ij'}^{-1} C_{k'l'}^{-1} (m_{j'} - m_0) (m_{k'} - m_0) (m_{l'} - m_0). \quad (3.78)$$

where $\hat{h}$ is a certain positive function. The important point is that the level of risk $\Lambda$ appears explicitly in the choice of the optimal portfolio. If the different operators choose different levels of risk, their optimal portfolio are no longer related by a proportionality factor, and the above CAPM relation does not hold.

### 3.3.2 ‘Power law’ fluctuations (*)

The minimisation of large risks also requires, as in the Gaussian case detailed above, the knowledge of the correlations between large events, and therefore an adapted measure of these correlations. Now, in the
3 Extreme risks and optimal portfolios

3.3 Portfolios of correlated assets

With an extreme case $\mu < 2$, the covariance (as the variance), is infinite. A suitable generalisation of the covariance is therefore necessary. Even in the case $\mu > 2$, where this covariance is a priori finite, the value of this covariance is a mix of the correlations between large negative moves, large positive moves, and all the ‘central’ (i.e. not so large) events. Again, the definition of a ‘tail covariance’, directly sensitive to the large negative events, is needed. The aim of the present section is to define such a quantity, which is a natural generalisation of the covariance for power-law distributions, much as the ‘tail amplitude’ is a generalisation of the variance. In a second part, the minimisation of the Value-at-Risk of a portfolio will be discussed.

‘Tail covariance’

Let us again assume that the $\delta x_i$’s are distributed according to:

$$P(\delta x_i) \propto \mu A_i^\mu |\delta x_i|^{1+\mu}. \quad (3.79)$$

A natural way to describe the correlations between large events is to generalise the decomposition in independent factors used in the Gaussian case, (3.69) and to write:

$$\delta x_i = m_i + \sum_{a=1}^{M} O_{ia} e_a, \quad (3.80)$$

where the $e_a$ are independent power-law random variables, the distribution of which being:

$$P(e_a) \propto \mu A_a^\mu |e_a|^{1+\mu}. \quad (3.81)$$

Since power-law variables are (asymptotically) stable under addition, the decomposition (3.80) indeed leads for all $\mu$ to correlated power-law variables $\delta x_i$.

The usual definition of the covariance is related to the average value of $\delta x_i \delta x_j$, which can in some cases be divergent (i.e. when $\mu < 2$). The idea is then to study directly the characteristic function of the product variable $\pi_{ij} = \delta x_i \delta x_j$. The justification is the following: if $e_a$ and $e_b$ are two independent power-law variables, their product $\pi_{ab}$ is also a power-law variable (up to logarithmic corrections) with the same exponent $\mu$, (cf. Appendix B):

$$P(\pi_{ab}) \propto \mu^2 (A_a^\mu A_b^\mu) \log(|\pi_{ab}|) / |\pi_{ab}|^{1+\mu} \quad (a \neq b). \quad (3.82)$$

On the contrary, the quantity $\pi_{aa}$ is distributed as a power-law with an exponent $\mu/2$ (cf. Appendix B):

$$P(\pi_{aa}) \propto \mu A_a^\mu / 2 |\pi_{aa}|^{1+\mu/2}. \quad (3.83)$$

Hence, the variable $\pi_{ij}$ gets both ‘non-diagonal’ contributions $\pi_{ab}$ and ‘diagonal’ ones $\pi_{aa}$. For $\pi_{ij} \rightarrow \infty$, however, only the latter survive. Therefore $\pi_{ij}$ is a power-law variable of exponent $\mu/2$, with a tail amplitude that we will note $A_{ij}^{\mu/2}$, and an asymmetry coefficient $\beta_{ij}$ (see Section 1.3.3). Using the additivity of the tail amplitudes, and the results of Appendix B, one finds:

$$A_{ij}^{\mu/2} = \sum_{a=1}^{M} |O_{ia} O_{ja}|^{\frac{\mu}{2}} A_a^\mu, \quad (3.84)$$

and

$$C_{ij}^{\mu/2} \equiv \beta_{ij} A_{ij}^{\mu/2} = \sum_{a=1}^{M} \text{sign}(O_{ia} O_{ja}) |O_{ia} O_{ja}|^{\frac{\mu}{2}} A_a^\mu. \quad (3.85)$$

In the limit $\mu = 2$, one sees that the quantity $C_{ij}^{\mu/2}$ reduces to the standard covariance, provided one identifies $A_a^\mu$ with the variance of the explicative factors $D_a$. This suggests that $C_{ij}^{\mu/2}$ is the suitable generalisation of the covariance for power-law variables, which is constructed using extreme events only. The expression (3.85) furthermore shows that the matrix $\hat{O}_{ia} = \text{sign}(O_{ia}) |O_{ia}|^{\mu/2}$ allows one to diagonalise the ‘tail covariance matrix’ $C_{ij}^{\mu/2}$: its eigenvalues are given by the $A_a^\mu$’s.

In summary, the tail covariance matrix $C_{ij}^{\mu/2}$ is obtained by studying the asymptotic behaviour of the product variable $\delta x_i \delta x_j$, which is a power-law variable of exponent $\mu/2$. The product of its tail amplitude and of its asymmetry coefficient is our definition of the tail covariance $C_{ij}^{\mu/2}$.

Optimal portfolio

It is now possible to find the optimal portfolio that minimises the loss probability. The fluctuations of the portfolio $p$ can indeed be written as:

$$\delta S \equiv \sum_{i=1}^{M} p_i \delta x_i = \sum_{a=1}^{M} \left( \sum_{i=1}^{M} p_i O_{ia} \right) e_a. \quad (3.86)$$

\(^{15}\) Note that $\hat{O} = O$ for $\mu = 2$.\(^{14}\) Other generalisations of the covariance have been proposed in the context of Lévy processes, such as the ‘covariation’ [Samorodnitsky-Taqqu].
Due to our assumption that the $e_a$ are symmetric power-law variables, $\delta S$ is also a symmetric power-law variable with a tail amplitude given by:

$$A^\mu_p = \sum_{a=1}^M \left| \sum_{i=1}^M p_i O_{ia} A^\mu_o \right|^\mu. \quad (3.87)$$

In the simple case where one tries to minimise the tail amplitude $A^\mu_p$ without any constraint on the average return, one then finds:

$$\sum_{a=1}^M O_{ia} A^\mu_o V^* = \zeta', \quad (3.88)$$

where the vector $V^*_a = \text{sign}(\sum_{j=1}^M O_{ja} p^*_j)^{\mu-1}$. Once the tail covariance matrix $C^{\mu/2}$ is known, the optimal weights $p^*_i$ are thus determined as follows:

1. The diagonalisation of $C^{\mu/2}$ gives the rotation matrix $\tilde{O}$, and therefore one can construct the matrix $O = \text{sign}(\tilde{O})|\tilde{O}|^{\mu}$ (understood for each element of the matrix), and the diagonal matrix $A$.
2. The matrix $OA$ is then inverted, and applied to the vector $\zeta'/\mu \bar{I}$.

This gives the vector $\tilde{V}^* = \zeta'/\mu (OA)^{-1} \bar{I}$.
3. From the vector $\tilde{V}$ one can then obtain the weights $p^*_i$ by applying the matrix inverse of $O^{\dagger}$ to the vector $\text{sign}(V^*)|V^*|^{\mu-1}$;
4. The Lagrange multiplier $\zeta'$ is then determined such as $\sum_{i=1}^M p^*_i = 1$.

Rather symbolically, one thus has:

$$p = (O^{\dagger})^{-1} \left( \frac{\mu}{\zeta'} (OA)^{-1} \bar{I} \right)^{1/\mu}. \quad (3.89)$$

In the case $\mu = 2$, one recovers the previous prescription, since in that case: $OA = OD = CO$, $(OA)^{-1} = O^{-1}C^{-1}$, and $O^{\dagger} = O^{-1}$, from which one gets:

$$p = \frac{\zeta'}{2} OO^{-1}C^{-1} \bar{I} = \frac{\zeta'}{2} C^{-1} \bar{I}, \quad (3.90)$$

which coincides with Eq. (3.74).

The problem of the minimisation of the Value-at-Risk in a strongly non-Gaussian world, where distributions behave asymptotically as power-laws, and in the presence of correlations between tail events, can thus be solved using a procedure rather similar to the one proposed by Markowitz for Gaussian assets.

---

3.4 Optimised trading (*)

In this section, we will discuss a slightly different problem of portfolio optimisation: how to dynamically optimise, as a function of time, the number of shares of a given stock that one holds in order to minimise the risk for a fixed level of return? We shall actually encounter a similar problem in the next chapter on options when the question of the optimal hedging strategy will be addressed. In fact, much of the notations and techniques of the present section are borrowed from Chapter 4. The optimised strategy found below shows that in order to minimise the variance, the time dependent part of the optimal strategy consists in selling when the price goes up and buying when it goes down. However, this strategy increases the probability of very large losses!

We will suppose that the trader holds a certain number of shares $\phi_n(x_n)$, where $\phi$ depends both on the (discrete) time $t = n \tau$, and on the price of the stock $x_n$. For the time being, we assume that the interest rates are negligible and that the change of wealth is given by:

$$\Delta W_X = \sum_{k=0}^{N-1} \phi_k(x_k) \delta x_k, \quad (3.91)$$

where $\delta x_k = x_{k+1} - x_k$. (See Sections 4.1 and 4.2 for a more detailed discussion of Eq. (3.91)). Let us define the gain $G = \langle \Delta W_X \rangle$ as the average final wealth, and the risk $R^2$ as the variance of the final wealth:

$$R^2 = \langle \Delta W_X^2 \rangle - G^2. \quad (3.92)$$

The question is then to find the optimal trading strategy $\phi^*_k(x_k)$, such that the risk $R$ is minimised, for a given value of $G$. Introducing a Lagrange multiplier $\zeta$, one thus looks for the (functional) solution of the following equation:\(^{17}\)

$$\frac{\delta}{\delta \phi_k(x)} \left[ R^2 - G^2 \right] |_{\phi_k = \phi^*_k} = 0. \quad (3.93)$$

Now, we further assume that the price increments $\delta x_n$ are independent random variables of mean $m \tau$ and variance $D \tau$. Introducing the notation $P(x,k|x_0,0)$ for the price to be equal to $x$ at time $k \tau$, knowing that it is equal to $x_0$ at time $t = 0$, one has:

$$G = \sum_{k=0}^{N-1} \langle \phi_k(x_k) \delta x_k \rangle = m \tau N^{-1} \int dx \, P(x,k|x_0,0) \phi_k(x), \quad (3.94)$$

The following equation results from a functional minimisation of the risk. See Section 4.4.3 for further details.

---

3 Extreme risks and optimal portfolios

where the factorisation of the average values holds because of the absence of correlations between the value of \( x_k \) (and thus that of \( \phi_k \)), and the value of the increment \( \delta x_k \). The calculation of \( \langle \Delta W_X^2 \rangle \) is somewhat more involved, in particular because averages of the type \( \langle \delta x \phi_k (x_k) \rangle \), with \( k > \ell \) do appear: in this case one cannot neglect the correlations between \( \delta x \ell \) and \( x_k \). Using the results of Appendix C, one finally finds:

\[
(\Delta W_X^2) = \frac{D\tau}{N} \sum_{k=0}^{N-1} \int dx \, P(x, k|x_0, 0)\phi_k^2(x)
\]

where the factorisation of the average values holds because of the absence of the correlations between \( \delta x \ell \) and \( x_k \). Using the results of Appendix C, one finally finds:

\[
(\Delta W_X^2) = \frac{D\tau}{N} \sum_{k=0}^{N-1} \int dx \, P(x, k|x_0, 0)\phi_k^2(x)
\]

\[
+ m\tau \sum_{k=0}^{N-1} \int dx dx' P(x', \ell|x_0, 0)P(x, k|x', \ell)\phi_{\ell}(x')\phi_k(x) \frac{x-x'}{k-\ell}.
\]

Taking the derivative with respect to \( \phi_k(x) \), one finds:

\[
2D\tau P(x, k|x_0, 0)\phi_k(x) - 2(1+\zeta)m\tau P(x, k|x_0, 0)\mathcal{G}
\]

\[
+ m\tau P(x, k|x_0, 0) \sum_{\ell=k+1}^{N-1} \int dx' P(x', \ell|x_0, 0)\phi_{\ell}(x') \frac{x-x'}{\ell-k}
\]

\[
+ m\tau \sum_{\ell=k}^{N-1} \int dx^\prime P(x', \ell|x_0, 0)P(x, k|x', \ell)\phi_{\ell}(x') \frac{x-x'}{k-\ell}.
\]

(3.96)

Setting this expression to zero gives an implicit equation for the optimal strategy \( \phi^*_k \). A solution to this equation can be found, for \( m \) small, as a power series in \( m \). Looking for a reasonable return means that \( \mathcal{G} \) should be of order \( m \). Therefore we set: \( \mathcal{G} = \mathcal{G}_0 mT \), with \( T = N\tau \), as expand \( \phi^* \) and \( 1+\zeta \) as:

\[
\phi^*_k = \phi^0_k + m\phi^1_k + \ldots \quad \text{and} \quad 1+\zeta = \frac{\zeta}{m^2} + \frac{\zeta}{m} + \ldots
\]

(3.97)

Inserting these expressions in (3.96) leads to zeroth order, to a time independent strategy:

\[
\phi^0_k(x) = \phi_0 = \frac{\zeta_0 \mathcal{G}_0 T}{D}.
\]

(3.98)

The Lagrange multiplier \( \zeta_0 \) is then fixed by the equation:

\[
\mathcal{G} = \mathcal{G}_0 mT = mT \phi_0 \quad \text{leading to} \quad \zeta_0 = \frac{D}{T} \quad \text{and} \quad \phi_0 = \mathcal{G}_0.
\]

(3.99)

To first order, the equation on \( \phi^1_k \) reads:

\[
D\phi^1_k = \zeta_1 \mathcal{G}_0 - \frac{\phi_0}{2} \sum_{\ell=k+1}^{N-1} \int dx^\prime P(x', \ell|x, k) \frac{x-x'}{\ell-k}
\]

(3.100)

\[
- \frac{\phi_0}{2} \sum_{\ell=0}^{k-1} \int dx' P(x', \ell|x_0, 0)P(x, k|x', \ell) x - x' - \frac{\ell}{k}.
\]

The second term in the right-hand side is of order \( m \), and thus negligible to this order. Interestingly, the last term can be evaluated explicitly without any assumption on the detailed shape for the probability distribution of the price increments. Using the method of Appendix C, one can show that for independent increments:

\[
\int dx' (x'-x_0)P(x', \ell|x_0, 0)P(x, k|x', \ell) = \frac{\ell}{k} (x-x_0) P(x, k|x_0, 0).
\]

(3.101)

Therefore, one finally finds:

\[
\phi^1_k(x) = \zeta_1 \mathcal{G}_0 T \frac{\zeta_0}{D} - \frac{\phi_0}{2D} (x-x_0).
\]

(3.102)

This equation shows that in order to minimise the risk as measured by the variance, a trader should sell stocks when their price increases, and buy more stocks when their price decreases, proportionally to \( m(x-x_0) \). The value of \( \zeta_1 \) is fixed such that:

\[
\sum_{k=0}^{N-1} \int dx P(x, k|x_0, 0)\phi^1_k(x) = 0,
\]

(3.103)

or \( \zeta_1 = m(N-1)/4N \). However, it can be shown that this strategy increases the VaR. For example, if the increments are Gaussian, the left tail of the distribution of \( \Delta W_X \) (corresponding to large losses), using the above strategy, is found to be exponential, and therefore much broader than the Gaussian expected for a time independent strategy:

\[
P_\zeta(\Delta W_X) \approx \Delta W_X \rightarrow -\infty \exp \frac{-2|\Delta W_X|}{\mathcal{G}}.
\]

(3.104)

3.5 Conclusion of the chapter

In a Gaussian world, all measures of risk are equivalent. Minimising the variance or the probability of large losses (or the Value-at-Risk) lead to the same result, which is the family of optimal portfolios obtained by Markowitz. In the general case, however, the VaR minimisation leads to portfolios which are different from the ones of Markowitz. These portfolios depend on the level of risk \( \Lambda \) (or on the time horizon \( T \)), to which the operator is sensitive. An important result is that minimising the variance can actually increase the VaR.
A simple case is provided by power-law distributed assets. In this case, the concept of ‘tail covariance’ (that measure the correlation between extreme events) can precisely be defined and measured. One can then obtain explicit formulae for the controlled VaR optimal portfolios. These formulae boil down, in the appropriate limit, to the classical Markowitz formulae. The minimisation of the VaR (as the relevant optimisation criterion) is however not easy to formulate within the classical framework of ‘utility functions’.

3.6 Appendix B: some useful results

Let us assume that the random variable $X$ is distributed as a power-law, with an exponent $\mu$ and a tail amplitude $A_X^\mu$. The random variable $\lambda X$ is then also distributed as a power-law, with the same exponent $\mu$ and a tail amplitude equal to $\lambda^\mu A_X^\mu$. This can easily be found using the rule for changing variables in probability densities, recalled in Section 1.2.1:

$$P(\lambda x) = \frac{1}{\lambda} \frac{\mu A_X^\mu}{(\lambda x)^{1+\mu}}. \quad (3.105)$$

For the same reason, the random variable $X^2$ is distributed as a power-law, with an exponent $\mu/2$ and a tail amplitude $(A_X^\mu)^{1/2}$. Indeed:

$$P(x^2) = \frac{P(x)}{2x} \approx \frac{\mu A_X^\mu}{2(x^2)^{1+\mu/2}}. \quad (3.106)$$

On the other hand, if $X$ and $Y$ are two independent power-law random variables with the same exponent $\mu$, the variable $Z = X \times Y$ is distributed according to:

$$P(z) = \int dx \int dy P(x) P(y) \delta(xy - z) = \int dx \frac{dx}{x} P(x) P(y = \frac{z}{x}). \quad (3.107)$$

When $z$ is large, one can replace $P(y = z/x)$ by $\mu A_Y^\mu x^{1+\mu}/z^{1+\mu}$ as long as $x \ll z$. Hence, in the limit of large $z$’s:

$$P(z) \approx \int_a^z dx x^\mu P(x) \frac{\mu A_Y^\mu}{z^{1+\mu}} \quad (3.108)$$

where $a$ is a certain constant. Now, since the integral $x$ diverges logarithmically for large $z$’s (since $P(x) \propto x^{-1-\mu}$), one can, to leading order, replace $P(x)$ in the integral by its asymptotic behaviour, and find:

$$P(z) \approx \frac{\mu^2 A_X^\mu A_Y^\mu}{z^{1+\mu}} \log z. \quad (3.109)$$

3.7 References

- **Statistics of drawdowns and extremes:**
  


- **Portfolio theory and CAPM:**
  


- **Optimal portfolios in a Lévy world:**
  


- **Generalisation of the covariance to Lévy variables:**
  
Les personnes non averties sont sujettes à se laisser induire en erreur.\textsuperscript{1}

(Lord Raglan, ‘Le tabou de l’inceste’, quoted by Boris Vian in ‘L’automne à Pékin’.)

4.1 Introduction

4.1.1 Aim of the chapter

The aim of this chapter is to introduce the general theory of derivative pricing in a simple and intuitive – but rather unconventional – way. The usual presentation, which can be found in all the available books on the subject\textsuperscript{2}, relies on particular models where it is possible to construct riskless hedging strategies, which replicate exactly the corresponding derivative product.\textsuperscript{3} Since the risk is strictly zero, there is no ambiguity in the price of the derivative: it is equal to the cost of the hedging strategy. In the general case, however, these ‘perfect’ strategies do not exist. Not surprisingly for the layman, zero risk is the exception rather than the rule. Correspondingly, a suitable theory must include risk as an essential feature, which one would like to minimise. The present chapter thus aims at developing simple methods to obtain optimal strategies, residual risks, and prices of derivative products, which takes into account in an adequate way the peculiar statistical nature of financial markets – as described in Chapter 2.

\textsuperscript{1}Unwarned people may easily be fooled.

\textsuperscript{2}see e.g. [Hull,Wilmott,Baxter].

\textsuperscript{3}A hedging strategy is a trading strategy allowing one to reduce, and sometimes eliminate, the risk.
4.1.2 Trading strategies and efficient markets

In the previous chapters, we have insisted on the fact that if the detailed prediction of future market moves is probably impossible, its statistical description is a reasonable and useful idea, at least as a first approximation. This approach only relies a certain degree of stability (in time) in the way markets behave and the prices evolve.\(^4\) Let us thus assume that one can determine (using a statistical analysis of past time series) the probability density \(P(x, t|x_0, t_0)\), which gives the probability that the price of the asset \(X\) is equal to \(x\) (to within \(dx\)) at time \(t\), knowing that at a previous time \(t_0\), the price was equal to \(x_0\). As in previous chapters, we shall denote as \(\langle O \rangle\) the average (over the ‘historical’ probability distribution) of a certain observable \(O\):

\[
\langle O(x, t) \rangle \equiv \int dx \ P(x, t|x_0, 0)O(x, t). \tag{4.1}
\]

As we have shown in Chapter 2, the price fluctuations are somewhat correlated for small time intervals (a few minutes), but become rapidly uncorrelated (but not necessarily independent!) on longer time scales. In the following, we shall choose as our elementary time scale \(\tau\) an interval a few times larger than the correlation time – say \(\tau = 30\) minutes on liquid markets. We shall thus assume that the correlations of price increments on two different intervals of size \(\tau\) are negligible.\(^5\) When correlations are small, the information on future movements based on the study of past fluctuations is weak. In this case, no systematic trading strategy can be more profitable (on the long run) than holding the market index – of course, one can temporarily ‘beat’ the market through sheer luck. This property corresponds to the efficient market hypothesis.\(^6\)

It is interesting to translate in more formal terms this property. Let us suppose that at time \(t_n = n\tau\), an investor has a portfolio containing, in particular, a quantity \(\phi_n(x_n)\) of the asset \(X\), quantity which can depend on the price of the asset \(x_n = x(t_n)\) at time \(t_n\) (this strategy could actually depend on the price of the asset for all previous times: \(\phi_n(x_n, x_{n-1}, x_{n-2}, \ldots)\)). Between \(t_n\) and \(t_{n+1}\), the price of \(X\) varies by \(\delta x_n\). This generates a profit (or a loss) for the investor equal to \(\phi_n(x_n)\delta x_n\). Note that the change of wealth is not equal to \(\delta(G) = \phi dx + x\delta\phi\), since the second term only corresponds to converting some stocks in cash, or vice versa, but not to a real change of wealth. The wealth difference between time \(t = 0\) and time \(t_N = T = N\tau\), due to the trading of asset \(X\) is thus equal to:

\[
\Delta W_X = \sum_{n=0}^{N-1} \phi_n(x_n)\delta x_n. \tag{4.2}
\]

Since the trading strategy \(\phi_n(x_n)\) can only be chosen before the price actually changes, the absence of correlations means that the average value of \(\Delta W_X\) (with respect to the historical probability distribution) reads:

\[
\langle \Delta W_X \rangle = \sum_{n=0}^{N-1} \langle \phi_n(x_n) \rangle \langle \delta x_n \rangle = \bar{m}\tau \sum_{n=0}^{N-1} \langle x_n \phi_n(x_n) \rangle, \tag{4.3}
\]

where we have introduced the average return \(\bar{m}\) of the asset \(X\), defined as:

\[
\delta x_n \equiv \eta_n x_n \quad \bar{m}\tau = \langle \eta_n \rangle. \tag{4.4}
\]

The above equation (4.3) thus means that the average value of the profit is fixed by the average return of the asset, weighted by the level of investment across the considered period. We shall often use, in the following, an additive model (more adapted on short time scales – cf. 2.2.1) where \(\delta x_n\) is rather written as \(\delta x_n = \eta_n x_0\). Correspondingly, the average return over the time interval \(\tau\) reads: \(m_1 = \langle \delta x \rangle = \bar{m}\tau x_0\). This approximation is usually justified as long as the total time interval \(T\) corresponds to a small (average) relative increase of price: \(\bar{m}T \ll 1\). We will often denote as \(m\) the average return per unit time: \(m = m_1/\tau\).

\(\text{Trading in the presence of temporal correlations}\)

It is interesting to investigate the case where correlations are not zero. For simplicity, we shall assume that the fluctuations \(\delta x_n\) are stationary Gaussian variables of zero mean (\(m_1 = 0\)). The correlation function is then given by \(\langle \delta x_n \delta x_k \rangle = C_{nk}\). The \(C_{nk}^{-1}\)'s are the elements of the matrix inverse of \(C\). If one knows the sequence of past increments \(\delta x_0, \ldots, \delta x_{n-1}\), the distribution of the next \(\delta x_n\) conditioned to such an observation is simply given by:

\[
P(\delta x_n) = \mathcal{N} \exp -\frac{1}{2} \bar{m}_n^{-1} \langle \delta x_n - m_n \rangle^2, \tag{4.5}
\]
existence of riskless assets such as bonds, which yield a known return, and the emergence of more complex financial products such as futures contracts or options, demands an adapted theory for pricing and hedging. This theory turns out to be predictive, even in the complete absence of temporal correlations.

4.2 Futures and Forwards

4.2.1 Setting the stage

Before turning to the rather complex case of options, we shall first focus on the very simple case of forward contracts, which allows us to define a certain number of notions (such as arbitrage and hedging) and notations. A forward contract $F$ amounts to buy or sell today an asset $X$ (the ‘underlying’) with a delivery date $T = N\tau$ in the future.\(^9\) What is the price $F$ of this contract, knowing that it must be paid at the date of expiry?\(^10\)

The naive answer that first comes to mind is a ‘fair game’ condition: the price must be adjusted such that, on average, the two parties involved in the contract fall even on the day of expiry. For example, taking the point of view of the writer of the contract (who sells the forward), the wealth balance associated with the forward reads:

$$\langle \Delta W_F \rangle = F - x(T).$$

This actually assumes that the writer has not considered the possibility of simultaneously trading the underlying stock $X$ to reduce his risk, which he of course should do: see below.

Under this assumption, the fair game condition amounts to set $\langle \Delta W_F \rangle = 0$, which gives for the forward price:

$$F_B = \langle x(T) \rangle \equiv \int dx \: xP(x, T|x_0, 0),$$

if the price of $X$ is $x_0$ at time $t = 0$. This price, that can we shall call the ‘Bachelier’ price, is not satisfactory since the seller takes the risk that the price at expiry $x(T)$ ends up far above $\langle x(T) \rangle$, which could prompt him to increase his price above $F_B$.\(^10\)

\(^7\)The notation $m_n$ has already been used in Chapter 1 with a different meaning. Note also that a general formula exists for the distribution of $\delta x_{n+k}$ for all $k \geq 0$, and can be found in books on optimal filter theory – see references.

\(^8\)A strategy that allows one to generate consistent abnormal returns with zero or minimal risk is called an ‘arbitrage opportunity.’

\(^9\)In practice ‘futures’ contracts are more common that forwards. While forwards are over-the-counter contracts, futures are traded on organised markets. For forwards there are typically no payments from either side before the expiration date while futures are marked-to-market and compensated every day, meaning that payments are made daily by one side or the other to bring back the value of the contract to zero.

\(^10\)Note that if the contract was to be paid now, its price would be exactly equal to that of the underlying asset (barring the risk of delivery default).
Actually, the Bachelier price $\mathcal{F}_B$ is not related to the market price for a simple reason: the seller of the contract can suppress his risk completely if he buys now the underlying asset $X$ at price $x_0$ and waits for the expiry date. However, this strategy is costly: the amount of cash frozen during that period does not yield the riskless interest rate. The cost of the strategy is thus $x_0e^{rT}$, where $r$ is the interest rate per unit time. From the viewpoint of the buyer, it would certainly be absurd to pay more than $x_0e^{rT}$, which is the cost of borrowing the cash needed to pay the asset right away. The only viable price for the forward contract is thus $\mathcal{F} = x_0e^{rT} \neq \mathcal{F}_B$, and is, in this particular case, completely unrelated to the potential moves of the asset $X$!

An elementary argument thus allows one to know the price of the forward contract and to follow a perfect hedging strategy: buy a quantity $\phi = 1$ of the underlying asset during the whole life of the contract. The aim of next paragraph is to establish this rather trivial result, which has not required many maths, in a much more sophisticated way. The importance of this procedure is that one needs to learn how to write down a proper wealth balance in order to price more complex derivative products such as options, on which we shall focus in the next sections.

### 4.2.2 Global financial balance

Let us write a general financial balance which takes into account the trading strategy of the the underlying asset $X$. The difficulty lies in the fact that the amount $\phi_n x_n / n$ which is invested in the asset $X$ rather than in bonds is effectively costly: one ‘misses’ the risk-free interest rate. Furthermore, this loss cumulates in time. It is not a priori obvious to write down the correct balance. Suppose that only two assets have to be considered: the risky asset $X$, and a bond $B$. The whole capital $W$ at time $t_n = n\tau$ is shared between these two assets:

$$W_n = \phi_n x_n + B_n.$$  \hfill (4.11)

The time evolution of $W_n$ is due both to the change of price of the asset $X$, and to the fact that the bond yields a known return through the interest rate $\rho$:

$$W_{n+1} - W_n = \phi_n(x_{n+1} - x_n) + B_n\rho, \quad \rho = r\tau. \hfill (4.12)$$

On the other hand, the amount of capital in bonds evolves both mechanically, through the effect of the interest rate (+$B_n\rho$), but also because the quantity of stock does change in time ($\phi_n \rightarrow \phi_{n+1}$), and causes a flow of money from bonds to stock or vice versa. Hence:

$$B_{n+1} - B_n = B_n\rho - x_{n+1}(\phi_{n+1} - \phi_n). \hfill (4.13)$$

Note that Eq. (4.11) is obviously compatible with the following two equations. The solution of the second equation reads:

$$B_n = (1 + \rho)^n B_0 - \sum_{k=1}^{n} x_k(\phi_k - \phi_{k-1})(1 + \rho)^{n-k}, \hfill (4.14)$$

Plugging this result in (4.11), and renaming $k - 1 \rightarrow k$ in the second part of the sum, one finally ends up with the following central result for $W_n$:

$$W_n = W_0(1 + \rho)^n + \sum_{k=0}^{n-1} \psi_k^n(x_{k+1} - x_k - \rho x_k), \hfill (4.15)$$

with $\psi_k^n \equiv \phi_k(1 + \rho)^{n-k}$. This last expression has an intuitive meaning: the gain or loss incurred between time $k$ and $k + 1$ must include the cost for the forward contract $-\rho x_k$; furthermore, this gain or loss must be forwarded up to time $n$ through interest rate effects, hence the extra factor $(1 + \rho)^{n-k}$. Another useful way to write and understand (4.15) is by introducing the discounted prices $\tilde{x}_k \equiv x_k(1 + \rho)^{-k}$. One then has:

$$W_n = (1 + \rho)^n \left(W_0 + \sum_{k=0}^{n-1} \phi_k(\tilde{x}_{k+1} - \tilde{x}_k)\right). \hfill (4.16)$$

The effect of interest rates can be thought of as an erosion of the value of the money itself. The discounted prices $\tilde{x}_k$ are therefore the ‘true’ change of wealth. The overall factor $(1 + \rho)^n$ then converts this true wealth into the current value of the money.

The global balance associated to the forward contract contains two further terms: the price of the forward $\mathcal{F}$ that one wishes to determine, and the price of the underlying asset at the delivery date. Hence, finally:

$$W_N = \mathcal{F} = x_N + (1 + \rho)^N \left(W_0 + \sum_{k=0}^{N-1} \phi_k(\tilde{x}_{k+1} - \tilde{x}_k)\right). \hfill (4.17)$$

Since by identity $\tilde{x}_N = \sum_{k=0}^{N-1}(\tilde{x}_{k+1} - \tilde{x}_k) + x_0$, this last expression can also be written as:

$$W_N = \mathcal{F} + (1 + \rho)^N \left(W_0 - x_0 + \sum_{k=0}^{N-1} (\phi_k - 1)(\tilde{x}_{k+1} - \tilde{x}_k)\right). \hfill (4.18)$$
4.2.3 Riskless hedge

In this last formula, all the randomness, the uncertainty on the future evolution of the prices, only appears in the last term. But if one chooses \( \phi_k \) to be identically equal to one, then the global balance is completely independent of the evolution of the stock price. The final result is not random, and reads:

\[
W_N = F + (1 + \rho)^N (W_0 - x_0). \tag{4.19}
\]

Now, the wealth of the writer of the forward contract at time \( T = N \tau \) cannot be, with certitude, greater (or less) than the one he would have had if the contract had not been signed, i.e. \( W_0 (1 + \rho)^N \). If this was the case, one of the two parties would be loosing money in a totally predictable fashion (since Eq. (4.19) does not contain any unknown term). Since any reasonable participant is reluctant to give away his money with no hope of return, one concludes that the forward price is indeed given by:

\[
F = x_0 (1 + \rho)^N \simeq x_0 e^{r T} \neq F_B, \tag{4.20}
\]

which does not rely on any statistical information on the price of \( X! \)

**Dividends**

In the case of a stock that pays a constant dividend rate \( \delta = d \tau \) per interval of time \( \tau \), the global wealth balance is obviously changed into:

\[
W_N = F - x_N + W_0 (1 + \rho)^N + \sum_{k=0}^{N-1} \psi_k \left( x_{k+1} - x_k + (\delta - \rho) x_k \right). \tag{4.21}
\]

It is easy to redo the above calculations in this case. One finds that the riskless strategy is now to hold:

\[
\phi_k \equiv \frac{(1 + \rho - \delta)^{N-k-1}}{(1 + \rho)^{N-k}} \tag{4.22}
\]

stocks. The wealth balance breaks even if the price of the forward is set to:

\[
F = x_0 (1 + \rho - \delta)^N \simeq x_0 e^{(r - \delta) T}, \tag{4.23}
\]

which again could have been obtained using a simple no arbitrage argument of the type presented below.

**Variable interest rates**

In reality, the interest rate is not constant in time but rather also varies randomly. More precisely, as explained in 2.6, at any instant of time the whole interest rate curve for different maturities is known, but evolves with time. The generalisation of the global balance, as given by formula (4.17), depends on the maturity of the bonds included in the portfolio, and thus on the whole interest rate curve. Assuming that only short term bonds are included, yielding the (time dependent) 'spot' rate \( \rho_k \), one has:

\[
W_N = F - x_N + W_0 \prod_{k=0}^{N-1} (1 + \rho_k) + \sum_{k=0}^{N-1} \psi_N \left( x_{k+1} - x_k - \rho_k x_k \right), \tag{4.24}
\]

with: \( \psi_N \equiv \phi_k \prod_{l=k+1}^{N-1} (1 + \rho_l) \). It is again quite obvious that holding a quantity \( \phi_k \) of the underlying asset leads to zero risk in the sense that the fluctuations of \( X \) disappear. However, the above strategy is not immune to interest rate risk. The interesting complexity of interest rate problems (such as the pricing and hedging of interest rate derivatives) comes from the fact that one may choose bonds of arbitrary maturity to construct the hedging strategy. In the present case, one may take a bond of maturity equal to that of the forward. In this case, risk disappears entirely, and the price of the forward reads:

\[
F = \frac{x_0}{B(0, N)}, \tag{4.25}
\]

where \( B(0, N) \) stands for the value, at time 0, of the bond maturing at time \( N \).

4.2.4 Conclusion: global balance and arbitrage

From the simple example of forward contracts, one should bear in mind the following points, which are the key concepts underlying the derivative pricing theory as presented in this book. After writing down the complete financial balance, taking into account the trading of all assets used to cover the risk, it is quite natural (at least from the viewpoint of the writer of the contract) to determine the trading strategy for all the hedging assets so as to minimise the risk associated to the contract. After doing so, a reference price is obtained by demanding that the global balance is zero on average, corresponding to a fair price from the point of view of both parties. In certain cases (such as the forward contracts described above), the minimum risk is zero and the true market price cannot differ from the fair price, or else arbitrage would be possible. On the example of forward contracts, the price (4.20) indeed corresponds to the absence of arbitrage opportunities (AAO), that is, of riskless profit. Suppose for example that the price of the forward
is below $F = x_0(1 + \rho)^N$. Now, one can sell the underlying asset now at price $x_0$ and simultaneously buy the forward at a price $F' < F$, that must be paid for on the delivery date. The cash $x_0$ is used to buy bonds with a yield rate $\rho$. On the expiry date, the forward contract is used to buy back the stock and close the position. The wealth of the trader is then $x_0(1 + \rho)^N - F'$, which is positive under our assumption – and furthermore fully determined at time zero: there is profit, but no risk. Similarly, a price $F' > F$ would also lead to an opportunity of arbitrage. More generally, if the hedging strategy is perfect, this means that the final wealth is known in advance. Thus, increasing the price as compared to the fair price leads to a riskless profit for the seller of the contract, and vice-versa. This AAO principle is at the heart of most derivative pricing theories currently used. Unfortunately, this principle cannot be used in the general case, where the minimal risk is non zero, or when transaction costs are present (and absorb the potential profit – see the discussion in Section 4.1.2 above). When the risk is non zero, there exists a fundamental ambiguity in the price, since one should expect that a risk premium is added to the fair price (for example, as a bid-ask spread). This risk premium depends both on the risk-averseness of the market maker, but also on the liquidity of the derivative market: if the price asked by one market maker is too high, less greedy market makers will make the deal. This mechanism does however not operate for over the counter operations (OTC, that is between two individual parties, as opposed to through an organised market). We shall come back to this important discussion in Section 4.6 below.

Let us emphasise that the proper accounting of all financial elements in the wealth balance is crucial to obtain the correct fair price. For example, we have seen above that if one forgets the term corresponding to the trading of the underlying stock, one ends up with the intuitive, but wrong, Bachelier price (4.10).

### 4.3 Options: definition and valuation

#### 4.3.1 Setting the stage

A buy option (or ‘call’ option) is nothing but an insurance policy, protecting the owner against the potential increase of the price of a given asset, which he will need to buy in the future. The call option provides to its owner the certainty of not paying the asset more than a certain price. Symmetrically, a ‘put’ option protects against drawdowns, by insuring to the owner a minimum value for his stock.

More precisely, in the case of a so-called ‘European’ option, the contract is such that at a given date in the future (the ‘expiry date’ or ‘maturity’) $t = T$, the owner of the option will not pay the asset more than $x_s$ (the ‘exercise price’, or ‘strike’ price): the possible difference between the market price at time $T$, $x(T)$ and $x_s$ is taken care of by the writer of the option. Knowing that the price of the underlying asset is $x_0$ now (i.e. at $t = 0$), what is the price (or ‘premium’) $C$ of the call? What is the optimal hedging strategy that the writer of the option should follow in order to minimise his risk?

The very first scientific theory of option pricing dates back to Bachelier in 1900. His proposal was, following a fair game argument, that the option price should equal the average value of the pay-off of the contract at expiry. Bachelier calculated this average by assuming the price increments $\delta x_n$ to be independent Gaussian random variables, which leads to the formula (4.43) below. However, Bachelier did not discuss the possibility of hedging, and therefore did not include in his wealth balance the term corresponding to the trading strategy that we have discussed in the previous section. As we have seen, this is precisely the term which is responsible for the difference between the forward ‘Bachelier price’ $F_B$ (cf. (4.10)) and the true price (4.20). The problem of the optimal trading strategy must thus, in principle, be solved before one can fix the price of the option. This is the problem solved by Black and Scholes in 1973, when they showed that for a continuous time Gaussian process, there exists a perfect strategy, in the sense that the risk associated to writing an option is strictly zero, as is the case for forward contracts. The determination of this perfect hedging strategy allows one to fix completely the price of the option using an AAO argument. Unfortunately, as repeatedly discussed below, this ideal strategy only exists in a continuous time, Gaussian world, that is, if the market correlation time was infinitely short, and if no discontinuities in the market price were allowed – both assumptions rather remote from reality. The hedging strategy cannot, in general, be perfect. However, an optimal hedging strategy can always be found, for which the risk is minimal (cf. Section 4.4). This optimal strategy thus allows one to calculate the fair price of the option and the associated residual risk. One should nevertheless bear in mind that, as emphasised in 4.2.4 and 4.6, there is no such thing as a unique option price.
price whenever the risk is non zero.

Let us now discuss these ideas more precisely. Following the method and notations introduced in the previous section, we write the global wealth balance for the writer of an option between time $t = 0$ and $t = T$ as:

$$W_N = [W_0 + C](1 + \rho)^N - \max(x_N - x_s, 0) + \sum_{k=0}^{N-1} \psi^N_k(x_{k+1} - x_k - \rho x_k).$$  \hspace{1cm} (4.26)

which reflects the fact that:

- The premium $C$ is paid immediately (i.e. at time $t = 0$).
- The writer of the option incurs a loss $x_N - x_s$ only if the option is exercised ($x_N > x_s$).
- The hedging strategy requires to convert a certain amount of bonds into the underlying asset, as was discussed before Eq. (4.17).

A crucial difference with forward contracts comes from the non linear nature of the pay-off, which is equal, in the case of a European option, to $\mathcal{Y}(x_N) = \max(x_N - x_s, 0)$. This must be contrasted with the forward pay-off, which is linear (and equal to $x_N$). It is ultimately the non-linearity of the pay-off which, combined with the non-Gaussian nature of the fluctuations, leads to a non zero residual risk.

An equation for the call price $C$ is obtained by requiring that the excess return due to writing the option, $\Delta W = W_N - W_0(1 + \rho)^N$, is zero on average:

$$(1 + \rho)^N C = \left[\max(x_N - x_s, 0) - \sum_{k=0}^{N-1} \psi^N_k(x_{k+1} - x_k - \rho x_k)\right].$$  \hspace{1cm} (4.27)

This price therefore depends, in principle, on the strategy $\psi^N_k = \phi^N_k(1 + \rho)^{N-k-1}$. This price corresponds to the fair price, to which a risk premium will in general be added (for example in the form of a bid-ask spread).

In the rather common case where the underlying asset is not a stock but a forward on the stock, the hedging strategy is less costly since only a small fraction $f$ of the value of the stock is required as a deposit. In the case where $f = 0$, the wealth balance appears to take a simpler form, since the interest rate is not lost while trading the underlying forward contract:

$$C = (1 + \rho)^N \left[\max(F_N - x_s, 0) - \sum_{k=0}^{N-1} \phi_k(F_{k+1} - F_k)\right].$$  \hspace{1cm} (4.28)

However, one can check that if one expresses the price of the forward in terms of the underlying stock, Eqs. (4.27) and (4.28) are actually identical. (Note in particular that $F_N = x_N$.)

4.3.2 Orders of magnitude

Let us suppose that the maturity of the option is such that non Gaussian ‘tail’ effects can be neglected ($T > T^*$, cf. 1.6.3, 1.6.5, 2.3), so that the distribution of the terminal price $x(T)$ can be approximated by a Gaussian of mean $mT$ and variance $DT = \sigma^2 x_s^2 T$.

If the average trend is small compared to the rms $\sqrt{DT}$, a direct calculation for ‘at the money’ options gives:

$$\langle \max(x(T) - x_s, 0) \rangle = \int_{x_s}^{\infty} dx \frac{x - x_s}{\sqrt{2\pi DT}} \exp\left(-\frac{(x - x_0 - mT)^2}{2DT}\right)$$

$$\simeq \sqrt{\frac{DT}{2\pi}} + \frac{mT}{2} + O\left(\frac{m^2T^3}{D}\right).$$  \hspace{1cm} (4.29)

Taking $T = 100$ days, a daily volatility of $\sigma = 1\%$, an annual return of $m = 5\%$, and a stock such that $x_0 = 100$ points, one gets:

$$\sqrt{\frac{DT}{2\pi}} \simeq 4 \text{ points} \quad \frac{mT}{2} \simeq 0.67 \text{ points.}$$  \hspace{1cm} (4.30)

In fact, the effect of a non zero average return of the stock is much less than what the above estimation would suggest. The reason is that one must also take into account the last term of the balance equation (4.27), which comes from the hedging strategy. This term corrects the price by an amount $-\langle \phi \rangle mT$. Now, as we shall see later, the optimal strategy for an at the money option is to hold, on average $\langle \phi \rangle = \frac{1}{2}$ stocks per option. Hence, this term precisely compensates the increase (equal to $mT/2$) of the average pay-off, $\langle \max(x(T) - x_s, 0) \rangle$. This strange compensation is actually exact in the Black-Scholes model (where the

14We again neglect the difference between Gaussian and log-normal statistics in the following order of magnitude estimate. See 1.3.2, 2.2.1, Eq. (4.43) and Fig. 4.1 below.

15An option is called ‘at-the-money’ if the strike price is equal to the current price of the underlying stock ($x_s = x_0$), ‘out-of-the-money’ if $x_s > x_0$ and ‘in-the-money’ if $x_s < x_0$. 

The interest rate appears in two different places in the balance equation (4.27): in front of the call premium, and in the cost of the trading turns. For a numerical example, it corresponds to a price increase of 2% of the stock price. We shall thus temporarily set \( r = 1\) to simplify the following discussion, and come back to the corrections brought about by a non zero value of \( m \) in Section 4.5.

Since the hedging strategy \( \psi_k \) is obviously determined before the next random change of price \( \delta x_k \), this two quantities are uncorrelated, and one has:

\[
\langle \psi_k \delta x_k \rangle = \langle \delta x_k \rangle = 0 \quad (m = 0).
\]

In this case, the hedging strategy disappears from the price of the option, which is given by the following 'Bachelier' like formula, generalised to the case where the increments are not necessarily Gaussian:

\[
\begin{align*}
C &= (1 + \rho)^{-N} \left( \max(x_N - x_s, 0) \right) \\
&= (1 + \rho)^{-N} \int_{x_s}^{\infty} dx \, (x - x_s) P(x, N|x_0, 0). \tag{4.33}
\end{align*}
\]

In order to use Eq. (4.33), concretely, one needs to specify a model for price increments. In order to recover the classical model of Black and Scholes, let us first assume that the relative returns are IID random variables and write \( \delta x_k \equiv \eta_k x_k \), with \( \eta_k \ll 1 \). If one knows (from empirical observation) the distribution \( P_1(\eta_k) \) of returns over the elementary time scale \( \tau \), one can easily reconstruct (using the independence of the returns) the distribution \( P(x, N|x_0, 0) \) needed to compute \( C \). After changing variables to \( x \rightarrow x_0(1 + \rho)^N e^y \), the formula (4.33) is transformed into:

\[
C = x_0 \int_{y_0}^{\infty} dy \, (e^y - e^{y'}) P_N(y),
\]

where \( y_k \equiv \log(x_k/x_0(1 + \rho)^N) \) and \( P_N(y) \equiv P(y, N|x_0, 0) \). Note that \( y_k \) involves the ratio of the strike price to the forward price of the underlying asset, \( x_0(1 + \rho)^N \).

Setting \( x_k = x_0(1 + \rho)^k e^{y_0} \), the evolution of the \( y_k \)'s is given by:

\[
y_{k+1} - y_k = \left( \frac{\eta_k}{1 + \rho} - \frac{\eta_k^2}{2} \right) y_0 = 0,
\]

where third order terms \((\eta^3, \eta^4, \eta^5, \ldots)\) have been neglected. The distribution of the quantity \( y_N = \sum_{k=0}^{N-1} \frac{\eta_k}{1 + \rho} - \frac{\eta_k^2}{2} \) is then obtained, in Fourier space, as:

\[
\tilde{P}_N(z) = [\tilde{P}_1(z)]^N,
\]

where we have defined, in the right hand side, a modified Fourier transformed:

\[
\tilde{P}_1(z) = \int dy \, P_1(y) \exp \left( iz \frac{\eta}{1 + \rho} - \frac{\eta^2}{2} \right),
\]

The Black and Scholes limit

We can now examine the Black-Scholes limit, where \( P_1(\eta) \) is a Gaussian of zero mean and \( \text{rms} \) equal to \( \sigma_1 = \sigma \sqrt{T} \). Using the above equations (4.36, 4.37) one finds, for \( N \) large: \(^{16}\)

\[
P_N(y) = \frac{1}{\sqrt{2\pi N \sigma_1^2}} \exp \left( -\frac{(y + N \sigma_1^2/2)^2}{2N \sigma_1^2} \right). \tag{4.38}
\]

The Black-Scholes model also corresponds to the limit where the elementary time interval \( \tau \) goes to zero, in which case one of course has \( N = T/\tau \rightarrow \infty \). As discussed in Chapter 2, this limit is not very realistic since any transaction takes at least a few seconds and, more significantly, that correlations persist over at least several minutes. Notwithstanding, if one takes the mathematical limit where \( \tau \rightarrow 0 \), with \( N = T/\tau \rightarrow \infty \) but keeping the product \( N \sigma_1^2 = T \sigma^2 \) finite, one constructs the continuous time log-normal process, or 'geometrical Brownian motion'. In this

\(^{16}\)In fact, the variance of \( P_N(y) \) is equal to \( N \sigma_1^2 / (1 + \rho)^2 \), but we neglect this small correction which disappears in the limit \( \tau \rightarrow 0 \).
limit, using the above form for $P_N(y)$ and $(1 + \rho)^N \to e^{rT}$, one obtains the celebrated Black-Scholes formula:

$$C_{BS}(x_0, x_s, T) = x_0 \int_{y_s}^{\infty} dy \frac{(e^y - e^{y_s})}{\sqrt{2\pi \sigma^2 T}} \exp \left( - \frac{(y + \sigma^2 T/2)^2}{2\sigma^2 T} \right)$$

$$= x_0 \mathcal{P}_G\left( \frac{y-y_s}{\sigma \sqrt{T}} \right) - x_s e^{-rT} \mathcal{P}_G\left( \frac{y-y_s}{\sigma \sqrt{T}} \right),$$  \hspace{1cm} (4.39)

where $y_k = \log(x_k/x_0) - rT \pm \sigma^2 T/2$ and $\mathcal{P}_G(u)$ is the cumulative normal distribution defined by Eq. (1.68). The way $\mathcal{P}_G$ varies with the four parameters $x_0, x_s, T$ and $\sigma$ is quite intuitive, and discussed at length in all the books which deal with options. The derivatives of the price with respect to these parameters are now called the ‘Greeks’, because they are denoted by Greek letters. For example, the larger the price of the stock $x_0$, the more probable it is to reach the strike price at expiry, and the more expensive the option. Hence the so-called ‘Delta’ of the option ($\Delta = \frac{\partial C}{\partial x}$) is positive. As we shall show below, this quantity is actually directly related to the optimal hedging strategy. The variation of $\Delta$ with $x_0$ is noted $\Gamma$ (Gamma) and is defined as: $\Gamma = \frac{\partial \Delta}{\partial x_0}$. Similarly, the dependence in maturity $T$ and volatility $\sigma$ (which in fact only appear through the combination $\sigma \sqrt{T}$ if $rT$ is small) leads to the definition of two further ‘Greeks’: $\Theta = -\frac{\partial C}{\partial T} < 0$ and ‘Vega’ $\nu = \frac{\partial C}{\partial \sigma} > 0$ – the higher the volatility through:

$$D \equiv \sigma^2 x_0^2.$$  \hspace{1cm} (4.40)

The price formula written down by Bachelier corresponds to the limit of short maturity options, where all interest rate effects can be neglected. In this limit, the above equation simplifies to:

$$C_G(x_0, x_s, T) = e^{-rT} \int_{x_s}^{\infty} dx \frac{1}{\sqrt{2\pi c^2(T)}} \exp \left( - \frac{(x-x_0 e^{rT})^2}{2c^2(T)} \right).$$  \hspace{1cm} (4.42)

This equation can also be derived directly from the Black-Scholes price (4.39) in the small maturity limit, where relative price variations are small: $xN/x0 -1 \ll 1$, allowing one to write $y = \log(x/x_0) \simeq (x-x_0)/x_0$. As emphasised in 1.3.2, this is the limit where the Gaussian and log-normal distributions become very similar.

The dependence of $C_G(x_0, x_s, T)$ as a function of $x_s$ is shown in Fig. 4.1, where the numerical value of the relative difference between the Black-Scholes and Bachelier price is also plotted.

$$\frac{xN = x_0(1 + \rho)^N + \sum_{k=0}^{N-1} \delta x_k(1 + \rho)^{N-k-1}}{x_s = x_0(1 + \rho)^N}.$$  \hspace{1cm} (4.41)

When $N$ is large, the difference $xN - x_0(1 + \rho)^N$ becomes, according to the clt, a Gaussian variable of mean zero (if $m = 0$) and of variance equal to:

$$\sigma^2(T) = D\tau \sum_{k=0}^{N-1} (1 + \rho)^{2k} \simeq DT \left[ 1 + \rho(N-1) + O(\rho^2N^2) \right].$$  \hspace{1cm} (4.42)

Dynamical equation for the option price

It is easy to show directly that the Gaussian distribution:

$$P_G(x, T|x_0, 0) = \frac{1}{\sqrt{2\pi DT}} \exp \left( - \frac{(x-x_0)^2}{2DT} \right).$$  \hspace{1cm} (4.44)

obeys the diffusion equation (or heat equation):

$$\frac{\partial P_G(x, T|x_0, 0)}{\partial T} = \frac{D}{2} \frac{\delta^2 P_G(x, T|x_0, 0)}{\delta x^2},$$  \hspace{1cm} (4.45)
with boundary conditions:

\[ P_G(x,0|x_0,0) = \delta(x-x_0). \]  

(4.46)

On the other hand, since \( P_G(x,T|x_0,0) \) only depends on the difference \( x-x_0 \), one has:

\[ \frac{\partial^2 P_G(x,T|x_0,0)}{\partial x^2} = \frac{\partial^2 P_G(x,T|x_0,0)}{\partial x_0^2}. \]  

(4.47)

Taking the derivative of Eq. (4.43) with respect to the maturity \( T \), one finds

\[ \frac{\partial C_G(x_0,x_s,T)}{\partial T} = \frac{D}{2} \frac{\partial^2 C_G(x_0,x_s,T)}{\partial x_0^2}, \]  

(4.48)

with boundary conditions, for a zero maturity option:

\[ C_G(x_0,x_s,0) = \max(x_0-x_s,0). \]  

(4.49)

The option price thus also satisfies the diffusion equation. We shall come back on this point later (4.5.2): it is indeed essentially this equation that Black and Scholes have derived using stochastic calculus in 1973.

### 4.3.4 Real option prices, volatility smile and ‘implied’ kurtosis

#### Stationary distributions and the smile curve

We shall now come back to the case where the distribution of the price increments \( \delta x_k \) is arbitrary, for example a TLD. For simplicity, we set the interest rate \( \rho \) to zero. (A non zero interest rate can readily be taken into account by discounting the price of the call on the forward contract). The price difference is thus the sum of \( N = T/\tau \) iid random variables, to which the discussion of the CLT presented in Chapter 1 applies directly. For large \( N \), the terminal price distribution \( P(x,N|x_0,0) \) becomes Gaussian, while for \( N \) finite, ‘fat tail’ effects are important and the deviations from the CLT are noticeable. In particular, short maturity options or out of the money options, or options on very ‘jagged’ assets, are not adequately priced by the Black-Scholes formula.

In practice, the market corrects for this difference empirically, by introducing in the Black-Scholes formula an ad-hoc ‘implied’ volatility \( \Sigma \), different from the ‘true’, historical volatility of the underlying asset. Furthermore, the value of the implied volatility needed to price properly options of different strike prices \( x_s \) and/or maturities \( T \) is not constant, but rather depends both on \( T \) and \( x_s \). This defines a ‘volatility surface’ \( \Sigma(x_s,T) \). It is usually observed that the larger the difference between \( x_s \) and \( x_0 \), the larger the implied volatility: this is the so-called ‘smile
A simple calculation allows one to understand the origin and the shape of the volatility smile. Let us assume that the maturity $T$ is sufficiently large so that only the kurtosis $\kappa_1$ of $P_1(\delta x)$ must be taken into account to measure the difference with a Gaussian distribution. Using the results of Section 1.6.3, the formula (4.34) leads to:

$$\Delta C_\kappa(x_0, x_s, T) = \frac{\kappa_1 T}{24T} \sqrt{\frac{2T}{2\pi}} \exp\left[-\frac{(x_s - x_0)^2}{2DT}\right] \left(\frac{(x_s - x_0)^2}{DT} - 1\right),$$

(4.50)

where $D \equiv \sigma^2 x_0^2$ and $\Delta C_\kappa = C_\kappa - C_{\kappa=0}$.

One can indeed transform the formula

$$C = \int_{x_s}^{\infty} dx' (x' - x_s) P'(x', x_0, 0),$$

(4.51)

through an integration by parts

$$C = \int_{x_s}^{\infty} dx' P_\geq(x', x_0, 0).$$

(4.52)

After changing variables $x' \rightarrow x' - x_0$, and using Eq. (1.69) of Section 1.6.3, one gets:

$$C = C_G + \frac{\sqrt{DT}}{\sqrt{2\pi}} \int_{u_s}^{\infty} \frac{du}{\sqrt{2\pi}} Q_1(u) e^{-u^2/2}$$

$$+ \frac{\sqrt{DT}}{N \sqrt{2\pi}} \int_{u_s}^{\infty} \frac{du}{\sqrt{2\pi}} Q_2(u) e^{-u^2/2} + ...$$

(4.53)

where $u_s \equiv (x_s - x_0)/\sqrt{DT}$. Now, noticing that:

$$Q_1(u) e^{-u^2/2} = \frac{\lambda_3}{6} \frac{d^3}{du^3} e^{-u^2/2},$$

(4.54)

and

$$Q_2(u) e^{-u^2/2} = -\frac{\lambda_4}{24} \frac{d^4}{du^4} e^{-u^2/2} - \frac{\lambda_3^2}{72} \frac{d^6}{du^6} e^{-u^2/2},$$

(4.55)

We also assume here that the distribution is symmetrical ($P_1(\delta x) = P_1(-\delta x)$), which is usually justified on short time scales. If this assumption is not adequate, one must include the skewness $\lambda_3$, leading to an asymmetrical smile – cf. Eq. (4.56).
the integrations over \( u \) are readily performed, yielding:

\[
C = C_0 + \sqrt{\Delta T} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \left( \frac{\lambda_3}{6\sqrt{N}} u_s + \frac{\lambda_4}{24N} (u_s^2 - 1) \right. \\
\left. + \frac{\lambda_3^2}{72N} (u_s^4 - 6u_s^2 + 3) + ... \right),
\]

(4.56)

which indeed coincides with (4.50) for \( \lambda_3 = 0 \). In general, one has \( \lambda_3^2 \ll \lambda_4 \); in this case, the smile remains a parabola, but shifted and centred around \( s_0(1 - 2\sigma T \lambda_3 / \lambda_4) \).

Note that we refer here to additive (rather than multiplicative) price increments: the Black-Scholes volatility smile is often asymmetrical merely because of the use of a skewed log-normal distribution to describe a nearly symmetrical process.

On the other hand, a simple calculation shows that the variation of \( C_{\kappa=0}(x_0, x_s, T) \) [as given by Eq. (4.43)] when the volatility changes by a small quantity \( \delta \) is given by:

\[
\delta C_{\kappa=0}(x_0, x_s, T) = \delta \sigma x_0 \sqrt{\frac{T}{2\pi}} \exp \left( -\frac{(x_s - x_0)^2}{2DT} \right).
\]

(4.57)

The effect of a non zero kurtosis \( \kappa_1 \) can thus be reproduced (to first order) by a Gaussian pricing formula, but at the expense of using an effective volatility \( \Sigma(x_s, T) = \sigma + \delta \sigma \) given by:

\[
\Sigma(x_s, T) = \sigma \left[ 1 + \frac{\kappa(T)}{24} \left( \frac{(x_s - x_0)^2}{DT} - 1 \right) \right],
\]

(4.58)

with \( \kappa(T) = \kappa_1 / N \). This very simple formula, represented in Fig. 4.2, allows one to understand intuitively the shape and amplitude of the smile. For example, for a daily kurtosis of \( \kappa_1 = 10 \), the volatility correction is on the order of \( \delta \sigma / \sigma \simeq 17\% \) for out of the money options such that \( x_s - x_0 = 3\sqrt{DT} \), and for \( T = 20 \) days. Note that the effect of the kurtosis is to reduce the implied at-the-money volatility as compared to its true value.

Figure 4.3 shows some ‘experimental’ data, concerning options on the BUND (futures) contract – for which a weakly non-Gaussian model is satisfactory (cf. 2.3). The associated option market is furthermore very liquid; this tends to reduce the width of the bid-ask spread, and thus, probably to decrease the difference between the market price and a theoretical ‘fair’ price. This is not the case for OTC options, when the overhead charged by the financial institution writing the option can in some case be very large – cf. 4.4.1 below. In Fig. 4.3, each point corresponds to an option of a certain maturity and strike price. The coordinates of these points are the theoretical price (4.34) along the \( x \) axis (calculated using the historical distribution of the BUND), and the observed market price along the \( y \) axis. If the theory is good, one should observe a cloud of points concentrated around the line \( y = x \). Figure 4.3 includes points from the first half of 1995, which has been rather ‘calm’, in the sense that the volatility has remained roughly constant during that period (see Fig. 4.5). Correspondingly, the assumption that the process is stationary is reasonable. On the other hand, the comparison with Black-Scholes theoretical prices (using the historical value of the volatility) is shown in Fig. 4.4, and reveals systematic differences with the market prices.

The agreement between theoretical and observed prices is however much less convincing if one uses data from 1994, when the volatility of interest rate markets has been high. The following subsection aims at describing these discrepancies in greater details.

Non-stationarity and ‘implied’ kurtosis

A more precise analysis reveals that the scale of the fluctuations of the underlying asset (here the BUND contract) itself varies noticeably around its mean value; these ‘scale fluctuations’ are furthermore correlated with the option prices. More precisely, one postulates that the distribution of price increments \( \delta x \) has a constant shape, but a width \( \gamma_k \) which is time dependent (cf. Section 2.4), i.e.:\(^{18}\)

\[
P_k(\delta x) \equiv \frac{1}{\gamma_k} P_{10} \left( \frac{\delta x_k}{\gamma_k} \right)^k.
\]

(4.59)

where \( P_{10} \) is a distribution independent of \( k \), and of width (for example measured by the MAD) normalised to one. The absolute volatility of the asset is thus proportional to \( \gamma_k \). Note that if \( P_{10} \) is Gaussian, this model is known as the stochastic volatility Brownian motion.\(^{19}\) However, this assumption is not needed, and one can keep \( P_{10} \) arbitrary.

Figure 4.5 shows the evolution of the scale \( \gamma \) (filtered over five days in the past) as a function of time and compares it to the implied volatility \( \Sigma(x_s = x_0) \) extracted from at the money options during the same period. One thus observes that the option prices are rather well tracked by adjusting the factor \( \gamma_k \) through a short term estimate of the volatility of the underlying asset. This means that the option prices primarily

\(^{18}\)This hypothesis can be justified by assuming that the amplitude of the market moves is subordinated to the volume of transactions, which itself is obviously time dependent.

\(^{19}\)In this context, the fact that the volatility is time dependent is called ‘heteroskedasticity’. ARCH models (Auto Regressive Conditional Heteroskedasticity) and their relatives have been invented as a framework to model such effects.
Figure 4.3: Comparison between theoretical and observed option prices. Each point corresponds to an option on the BUND (traded on the LIFFE in London), with different strike prices and maturities. The $x$ coordinate of each point is the theoretical price of the option (determined using the empirical terminal price distribution), while the $y$ coordinate is the market price of the option. If the theory is good, one should observe a cloud of points concentrated around the line $y = x$. The dotted line corresponds to a linear regression, and gives $y = 0.999(3)x + 0.02$ (in basis points units). The inset shows more precisely the difference between market price and theoretical prices as a function of the market price.

Figure 4.4: Comparison between Black-Scholes prices (using the historical volatility) and market prices. This figure is constructed as the previous one, except that the theoretical price is now the Black-Scholes price. The agreement is now only fair; the deviations are furthermore systematic, as shown in the inset (note the scale of the vertical axis compared to the previous figure). The linear regression now gives $y = 1.02(0)x + 6$ (in basis points). The systematic character of the deviations is responsible for the rather large value of the intercept.
fluctuations of the underlying asset (here the BUND contract), or from the implied volatility of at the money options. More precisely, the historical determination of \( \gamma \) comes from the daily average of the absolute value of the five minutes price increments. The two curves are then smoothed over five days. These two determinations are strongly correlated, showing that the option price mostly reflects the instantaneous volatility of the underlying asset itself.

It is interesting to notice that the mere existence of volatility fluctuations leads to a non zero kurtosis \( \kappa_N \) of the asset fluctuations (see 2.4), which has an anomalous time dependence (cf. (2.17)), in agreement with the direct observation of the historical kurtosis (Fig. 4.6). On the other hand, an ‘implied kurtosis’ can be extracted from the market price of options, using Eq. (4.58) as a fit to the empirical smile. Remarkably enough, this implied kurtosis is in close correspondence with the historical kurtosis – note that Fig. 4.6 does not require any further adjustable parameter.

As a conclusion of this section, it seems that market operators have empirically corrected the Black-Scholes formula to account for two distinct, but related effects:

- The presence of market jumps, implying fat tailed distributions \( (\kappa > 0) \) of short term price increments. This effect is responsible for the volatility smile (and also, as we shall discuss next, for the

![Figure 4.5: Time dependence of the scale parameter \( \gamma \), obtained from the analysis of the intra-day fluctuations of the underlying asset (here the BUND), or from the implied volatility of at the money options. More precisely, the historical determination of \( \gamma \) comes from the daily average of the absolute value of the five minutes price increments. The two curves are then smoothed over five days. These two determinations are strongly correlated, showing that the option price mostly reflects the instantaneous volatility of the underlying asset itself.](image)

![Figure 4.6: Plot (in log-log coordinates) of the average implied kurtosis \( \kappa_{\text{imp}} \) (determined by fitting the implied volatility for a fixed maturity by a parabola) and of the empirical kurtosis \( \kappa_N \) (determined directly from the historical movements of the BUND contract), as a function of the reduced time scale \( N = T/\tau \), \( \tau = 30 \) minutes. All transactions of options on the BUND future from 1993 to 1995 were analysed along with 5 minute tick data of the BUND future for the same period. We show for comparison a fit with \( \kappa_N \approx N^{-0.42} \) (dark line), as suggested by the results of Section 2.4. A fit with an exponentially decaying volatility correlation function is however also acceptable (dash line).](image)
fact that options are risky).

- The fact that the volatility is not constant, but fluctuates in time, leading to an anomalously slow decay of the kurtosis (slower than $1/N$) and, correspondingly, to a non trivial deformation of the smile with maturity.

It is interesting to note that, through trial and errors, the market as a whole has evolved to allow for such non trivial statistical features – at least on most actively traded markets. This might be called 'market efficiency'; but contrarily to stock markets where it is difficult to judge whether the stock price is or not the 'true' price, option markets offer a remarkable testing ground for this idea. It is also a nice example of adaptation of a population (the traders) to a complex and hostile environment, which has taken place in a few decades!

In summary, part of the information contained in the implied volatility surface $\Sigma(x_s, T)$ used by market participants can be explained by an adequate statistical model of the underlying asset fluctuations. In particular, in weakly non-Gaussian markets, an important parameter is the time dependent kurtosis – see Eq. (4.58). The anomalous maturity dependence of this kurtosis encodes the fact that the volatility is itself time dependent.

### 4.4 Optimal strategy and residual risk

#### 4.4.1 Introduction

In the above discussion, we have chosen a model of price increments such that the cost of the hedging strategy (i.e. the term $\langle \psi_k \delta x_k \rangle$) could be neglected, which is justified if the excess return $m$ is zero, or else for short maturity options. Beside the fact that the correction to the option price induced by a non zero return is important to assess (this will be done in the next section), the determination of an 'optimal' hedging strategy and the corresponding minimal residual risk is crucial for the following reason. Suppose that an adequate measure of the risk taken by the writer of the option is given by the variance of the global wealth balance associated to the operation, i.e.,

$$ R = \sqrt{\Delta W^2 \langle \phi \rangle}. \quad (4.60) $$

As we shall find below, there is a special strategy $\phi^*$ such that the above quantity is reaches a minimum value. Within the hypotheses of

[20] The case where a better measure of risk is the loss probability or the Value-at-Risk will be discussed below.

the Black-Scholes model, this minimum risk is even, rather surprisingly, strictly zero. Under less restrictive and more realistic hypotheses, however, the residual risk $R^* \equiv \sqrt{\langle \Delta W^2 [\phi^*] \rangle}$ actually amounts to a substantial fraction of the option price itself. It is thus rather natural for the writer of the option to try to reduce further the risk by overpricing the option, adding to the ‘fair price’ a risk premium proportional to $R^*$ – in such a way that the probability of eventually loosing money is reduced. Stated otherwise, option writing being an essentially risky operation, it should also be, on average, profitable.

Therefore, a market maker on option markets will offer to buy and to sell options at slightly different prices (the ‘bid-ask’ spread), centred around the fair price $C$. The amplitude of the spread is presumably governed by the residual risk, and is thus $\pm \lambda R^*$, where $\lambda$ is a certain numerical factor, which measures the price of risk. The search for minimum risk corresponds to the need of keeping the bid-ask spread as small as possible, because several competing market makers are present. One therefore expects the coefficient $\lambda$ to be smaller on liquid markets. On the contrary, the writer of an OTC option usually claims a rather high risk premium $\lambda$.

Let us illustrate this idea by the two Figs. 4.7 and 4.8, generated using real market data. We show the histogram of the global wealth balance $\Delta W$, corresponding to an at-the-money option, of maturity equal to three months, in the case of a bare position ($\phi \equiv 0$), and in the case where one follows the optimal strategy ($\phi = \phi^*$) prescribed below. The fair price of the option is such that $\langle \Delta W \rangle = 0$ (vertical thick line). It is clear that without hedging, option writing is a very risky operation. The optimal hedge substantially reduces the risk, though the residual risk remains quite high. Increasing the price of the option by an amount $\lambda R^*$ corresponds to a shift of the vertical thick line to the left of Fig. 4.8, thereby reducing the weight of unfavourable events.

Another way of expressing this idea is to use $R^*$ as a scale to measure the difference between the market price of an option $C_M$ and its theoretical price:

$$ \lambda = \frac{C_M - C}{R^*}. \quad (4.61) $$

An option with a large value of $\lambda$ is an expensive option, which includes a large risk premium.

#### 4.4.2 A simple case

Let us now discuss how an optimal hedging strategy can be constructed, by focussing first on the simple case where the amount of the underlying
Figure 4.7: Histogram of the global wealth balance $\Delta W$ associated to the writing of an at-the-money option of maturity equal to 60 trading days. The price is fixed such that on average $\langle \Delta W \rangle$ is zero (vertical line). This figure corresponds to the case where the option is not hedged ($\phi \equiv 0$). The RMS of the distribution is 1.04, to be compared with the price of the option $C = 0.79$. The 'peak' at 0.79 thus corresponds to the cases where the option is not exercised, which happens with a probability close to 1/2 for at-the-money options.

Figure 4.8: Histogram of the global wealth balance $\Delta W$ associated to the writing of the same option, with a price still fixed such that $\langle \Delta W \rangle = 0$ (vertical line), with the same horizontal scale as in the previous figure. This figure shows the effect of adopting an optimal hedge ($\phi = \phi^*$), recalculated every half-hour, and in the absence of transaction costs. The RMS $R^*$ is clearly smaller ($= 0.28$), but non zero. Note that the distribution is skewed towards $\Delta W < 0$. Increasing the price of the option by $\lambda R^*$ amounts to diminish the probability of loosing money (for the writer of the option), $P(\Delta W < 0)$. 
asset held in the portfolio is fixed once and for all when the option is written, i.e. at \( t = 0 \). This extreme case would correspond to very high transaction costs, so that changing one’s position on the market is very costly. We shall furthermore assume, for simplicity, that interest rate effects are negligible, i.e. \( \rho = 0 \). The global wealth balance (4.28) then reads:

\[
\Delta W = C - \max(x_N - x_s, 0) + \phi \sum_{k=0}^{N-1} \delta x_k.
\]  

(4.62)

In the case where the average return is zero (\( \langle \delta x_k \rangle = 0 \)) and where the increments are uncorrelated (i.e. \( \delta x_k \delta x_l = D \tau \delta_{k,l} \)), the variance of the final wealth, \( R^2 = \langle \Delta W^2 \rangle - \langle \Delta W \rangle^2 \), reads:

\[
R^2 = N D \tau \phi^2 - 2 \phi \langle (x_N - x_0) \max(x_N - x_s, 0) \rangle + R_0^2,
\]

(4.63)

where \( R_0^2 \) is the intrinsic risk, associated to the ‘bare’, unhedged option (\( \phi = 0 \)):

\[
R_0^2 = \langle \max(x_N - x_s, 0)^2 \rangle - \langle \max(x_N - x_s, 0) \rangle^2.
\]

(4.64)

The dependence of the risk on \( \phi \) is shown in Fig. 4.9, and indeed reveals the existence of an optimal value \( \phi = \phi^* \) for which \( R \) takes a minimum value. Taking the derivative of (4.63) with respect to \( \phi \),

\[
\left. \frac{dR}{d\phi} \right|_{\phi=\phi^*} = 0,
\]

(4.65)

one finds:

\[
\phi^* = \frac{1}{N} \int_{x_s}^{x_N} dx(x - x_s)(x - x_0)P(x, N|x_0, 0),
\]

(4.66)

thereby fixing the optimal hedging strategy within this simplified framework. Interestingly, if \( P(x, N|x_0, 0) \) is Gaussian (which becomes a better approximation as \( N = T/\tau \) increases), one can write:

\[
\frac{1}{D \tau} \int_{x_s}^{x_N} dx(x - x_s)(x - x_0) \exp -\frac{(x - x_0)^2}{2D \tau} - \int_{x_s}^{x_N} dx(x - x_s) \frac{\partial}{\partial x} \exp -\frac{(x - x_0)^2}{2D \tau},
\]

(4.67)

giving, after an integration by parts: \( \phi^* = P \), or else the probability, calculated from \( t = 0 \), that the option is exercised at maturity: \( P \equiv \int_{x_s}^{x_N} dxP(x, N|x_0, 0) \).

Hence, buying a certain fraction of the underlying allows one to reduce the risk associated to the option. As we have formulated it, risk minimisation appears as a variational problem, the result of which thus depending on the family of ‘trial’ strategies \( \phi \). A natural idea is thus to generalise the above procedure to the case where the strategy \( \phi \) is allowed to vary both with time, and with the price of the underlying asset. In other words, one can certainly do better than holding a certain fixed quantity of the underlying asset, by adequately readjusting this quantity in the course of time.

\[ \text{Figure 4.9: Dependence of the residual risk } R \text{ as a function of the strategy } \phi, \]

in the simple case where this strategy does not vary with time. Note that \( R \) is minimum for a well defined value of \( \phi \).

\[ \text{4.4.3 General case: ‘} \Delta \text{’ hedging} \]

If one writes again the complete wealth balance (4.26) as:

\[
\Delta W = C(1 + \rho)^N - \max(x_N - x_s, 0) + \sum_{k=0}^{N-1} \psi^N_k(x_k)\delta x_k,
\]

(4.68)

the calculation of \( \langle \Delta W^2 \rangle \) involves, as in the simple case treated above, three types of terms: quadratic in \( \psi \), linear in \( \psi \), and independent of \( \psi \). The last family of terms can thus be forgotten in the minimisation procedure. The first two terms of \( R^2 \) are:

\[
\sum_{k=0}^{N-1} \langle (\psi^N_k)^2 \rangle \langle \delta x_k^2 \rangle - 2 \sum_{k=0}^{N-1} \langle \psi^N_k \delta x_k \max(x_N - x_s, 0) \rangle,
\]

(4.69)
where we have used $\langle \delta x_k \rangle = 0$ and assumed the increments to be of finite variance and uncorrelated (but not necessarily stationary nor independent):\(^{21}\)

$$
\langle \delta x_k \delta x_l \rangle = \langle \delta x_k^2 \rangle \delta_{k,l}. \tag{4.70}
$$

We introduce again the distribution $P(x, k| x_0, 0)$ for arbitrary intermediate times $k$. The strategy $\psi^N_k$ depends now on the value of the price $x_k$. One can express the terms appearing in Eq. (4.69) in the following form:

$$
\langle (\psi^N_k)^2 \rangle \langle \delta x_k^2 \rangle = \int dx \ [\psi^N_k(x)]^2 P(x, k| x_0, 0) \langle \delta x_k^2 \rangle, \tag{4.71}
$$

and

$$
\langle \psi^N_k \delta x_k \max(x_N - x_s, 0) \rangle = \int dx \ \psi^N_k(x) P(x, k| x_0, 0) \times \int_{x_s}^{+\infty} dx' \ \langle \delta x_k(x, k)\rightarrow(x', N) (x' - x_s) P(x', N|x, k). \tag{4.72}
$$

where the notation $\langle \delta x_k(x, k)\rightarrow(x', N) \rangle$ means that the average of $\delta x_k$ is restricted to those trajectories which start from point $x$ at time $k$ and end at point $x'$ at time $N$. Without this restriction, the average would of course be (within the present hypothesis) zero.

The functions $\psi^N_k(x)$, for $k = 1, 2, ..., N$, must be chosen so that the risk $\mathcal{R}$ is as small as possible. Technically, one must ‘functionally’ minimise (4.69). We shall not try to justify mathematically this procedure; one can simply imagine that the integrals over $x$ are discrete sums (which is actually true, since prices are expressed in cents and therefore are not continuous variables). The function $\psi^N_k(x)$ is thus determined by the discrete set of values of $\psi^N_i(i)$. One can then take usual derivatives of (4.69) with respect to these $\psi^N_i(i)$. The continuous limit, where the points $i$ become infinitely close to one another, allows one to define the functional derivative $\frac{\partial}{\partial \psi^N(x)}$, but it is useful to keep in mind its discrete interpretation. Using (4.71) and (4.73), one thus find the following fundamental equation:

$$
\frac{\partial \mathcal{R}}{\partial \psi^N_k(x)} = 2 \psi^N_k(x) P(x, k| x_0, 0) \langle \delta x_k^2 \rangle - 2 P(x, k| x_0, 0) \int_{x_s}^{+\infty} dx' \langle \delta x_k(x, k)\rightarrow(x', N) (x' - x_s) P(x', N|x, k). \tag{4.73}
$$

Setting this functional derivative to zero then provides a general and rather explicit expression for the optimal strategy $\psi^N_k^*(x)$, where the only assumption is that the increments $\delta x_k$ are uncorrelated (cf. (4.70)):

$$
\psi^N_k^*(x) = \frac{1}{\langle \delta x_k \rangle} \int_{x_s}^{+\infty} dx' \langle \delta x_k(x, k)\rightarrow(x', N) (x' - x_s) P(x', N|x, k). \tag{4.74}
$$

This formula can be simplified in the case where the increments $\delta x_k$ are identically distributed (of variance $\sigma^2$) and when the interest rate is zero. As shown in Appendix C, one then has:

$$
\langle \delta x_k(x, k)\rightarrow(x', N) \rangle = \frac{x' - x}{N - k} \tag{4.75}
$$

The intuitive interpretation of this result is that all the $N - k$ increments $\delta x_l$ along the trajectory $(x, k) \rightarrow (x', N)$ contribute on average equally to the overall price change $x' - x$. The optimal strategy then finally reads:

$$
\phi^N_k^*(x) = \int_{x_s}^{+\infty} dx' \frac{x' - x}{D\tau(N - k)} \langle x' - x_s \rangle P(x', N|x, k). \tag{4.76}
$$

It is easy to see that the above expression varies monotonously from $\phi^N_k^*(x_s = +\infty) = 0$ to $\phi^N_k^*(x_s = -\infty) = 1$ as a function of $x_s$; this result holds without any further assumption on $P(x', N|x, k)$. If $P(x', N|x, k)$ is well approximated by a Gaussian, the following identity then holds true:

$$
\frac{x' - x}{D\tau(N - k)} P_G(x', N|x, k) = \frac{\partial P_G(x', N|x, k)}{\partial x}. \tag{4.77}
$$

The comparison between (4.76) and (4.33) then leads to the so-called ‘Delta’ hedging of Black and Scholes:

$$
\phi^N_k^*(x = x_k) = \frac{\partial c_{BS}[x, x_s, N - k]}{\partial x} \bigg|_{x = x_k} \tag{4.78}
$$

One can actually show that this result is true even if the interest rate is non zero (see Appendix C), as well as if the relative returns (rather than the absolute returns) are independent Gaussian variables, as assumed in the Black-Scholes model.

Equation (4.78) has a very simple (perhaps sometimes misleading) interpretation. If between time $k$ and $k + 1$ the price of the underlying asset varies by a small amount $dx_k$, then to first order in $dx_k$, the variation of the option price is given by $\Delta[x, x_s, N - k] dx_k$ (using the
very definition of \( \Delta \)). This variation is therefore exactly compensated by the gain or loss of the strategy, i.e. \( \phi^N(x = x_k)dx_k \). In other words, \( \phi^N = \Delta(x_k, N - k) \) appears to be a perfect hedge, which ensures that the portfolio of the writer of the option does not vary at all with time (cf. below, when we will discuss the differential approach of Black and Scholes). However, as we discuss now, the relation between the optimal hedge and \( \Delta \) is not general, and does not hold for non-Gaussian statistics.

**Cumulant corrections to \( \Delta \) hedging**

More generally, using Fourier transforms, one can express (4.76) as an cumulant expansion. Using the definition of the cumulants of \( P \), one has:

\[
(x' - x)P(x', N|x, 0) \equiv \int \frac{dz}{2\pi} \hat{P}_N(z) \frac{\partial}{\partial(-iz)} e^{-iz(x' - x)} = -\sum_{n=2}^{\infty} (\frac{z^2}{n})^n c_{n, N} \frac{\partial^{n-1}}{x^{n-1}} P(x', N|x, 0),
\]

(4.79)

where \( \log \hat{P}_N(z) = \sum_{n=2}^{\infty} (iz)^n c_{n, N}/n! \). Assuming that the increments are independent, one can use the additivity property of the cumulants discussed in Chapter 1, i.e.: \( c_{n, N} \equiv N c_{n, 1} \), where the \( c_{n, 1} \) are the cumulants at the elementary time scale \( \tau \). One finds the following general formula for the optimal strategy:

\[
\phi^N(x) = \frac{1}{\delta \tau} \sum_{n=2}^{\infty} c_{n, 1} \frac{\partial^{n-1} C[x, x, N]}{x^{n-1}},
\]

(4.80)

where \( c_{2,1} \equiv \delta \tau, c_{4,1} \equiv \kappa [c_{2,2}]^2 \), etc. In the case of a Gaussian distribution, \( c_{n,1} \equiv 0 \) for all \( n \geq 3 \), and one thus recovers the previous ‘Delta’ hedging. If the distribution of the elementary increments \( \delta x_k \) is symmetrical, \( c_{3,1} = 0 \), and the first correction is therefore of order \( c_{4,1}/c_{2,1} \times (1/\sqrt{\delta \tau N})^2 \equiv \kappa/N \), where we have used the fact that \( C[x, x, N] \) typically varies on the scale \( x \approx \sqrt{\delta \tau N} \), which allows us to estimate the order of magnitude of its derivatives with respect to \( x \). Equation (4.80) shows that in general, the optimal strategy is not simply given by the derivative of the price of the option with respect to the price of the underlying asset.

It is interesting to compute the difference between \( \phi^* \) and the Black-Scholes strategy as used by the traders, \( \phi^*_M \), which takes into account

\[\text{Note that the market does not compute the total derivative of the option price with respect to the underlying price, which would also include the derivative of the implied volatility.}\]

---

Fig. (4.10). As will be discussed below, the use of such an approximate hedging induces a rather small increase of the residual risk.

4.4.4 Global hedging/instantaneous hedging

It is important to stress that the above procedure, where the global risk associated to the option (calculated as the variance of the final wealth balance) is minimised is equivalent to the minimisation of an ‘instantaneous’ hedging error – at least when the increments are uncorrelated. The latter consists in minimising the difference of value of a position which is short one option and $\phi$ long in the underlying stock, between two consecutive times $k\tau$ and $(k+1)\tau$. This is closer to the concern of many option traders, since the value of their option books is calculated every day: the risk is estimated in a ‘marked to market’ way. If the price increments are uncorrelated, it is easy to show that the optimal strategy is identical to the ‘global’ one discussed above.

4.4.5 Residual risk: the Black-Scholes miracle

We shall now compute the residual risk $R^*$ obtained by substituting $\phi$ by $\phi^*$ in Eq. (4.69).24 One finds:

$$R^* = R_0^* - DT \sum_{k=0}^{N-1} \int dx \, P(x, k|x_0, 0)|\phi_k^N(x)|^2,$$

where $R_0$ is the unhedged risk. The Black-Scholes ‘miracle’ is the following: in the case where $P$ is Gaussian, the two terms in the above equation exactly compensate in the limit where $\tau \to 0$, with $N \tau = T$ fixed.

This ‘miraculous’ compensation is due to a very peculiar property of the Gaussian distribution, for which:

$$P_G(x_1, T|x_0, 0)\delta(x_1 - x_2) - P_G(x_1, T|x_0, 0)P_G(x_2, T|x_0, 0) =$$

$$= D \int_0^T dt \int dx \, P_G(x, t|x_0, 0) \frac{\partial P_G(x_1, T|x, t)}{\partial x} \frac{\partial P_G(x_2, T|x, t)}{\partial x}. \quad (4.85)$$

Integrating this equation with respect to $x_1$ and $x_2$, the left hand side gives $R_0^*$. As for the right hand side, one recognises the limit of $\sum_{k=0}^{N-1} \int dx \, P(x, k|x_0, 0)|\phi_k^N(x)|^2$ when $\tau = 0$, where the sum becomes an integral.

24The following ‘zero-risk’ property is true both within an ‘additive’ Gaussian model and the Black-Scholes multiplicative model, or for more general continuous time Brownian motions. This property is more easily obtained using Ito’s stochastic differential calculus – see below. The latter formalism is however only valid in the continuous time limit ($\tau = 0$).

Therefore, in the continuous time Gaussian case, the $\Delta$-hedge of Black and Scholes allows one to eliminate completely the risk associated to an option, as was the case for the forward contract. However, contrarily to the forward contract where the elimination of risk is possible whatever the statistical nature of the fluctuations, the result of Black and Scholes is only valid in a very special limiting case. For example, as soon as the elementary time scale $\tau$ is finite (which is the case in reality), the residual risk $R^*$ is non zero even if the increments are Gaussian. The calculation is easily done in the limit where $\tau$ is small compared to $T$: the residual risk then comes from the difference between a continuous integral $D \int dt \int dx \, P_G(x, t'|x_0, 0)\phi^2(x, t')$ (which is equal to $R_0^*$) and the corresponding discrete sum appearing in (4.84). This difference is given by the Euler-McLaurin formula and is equal to:

$$R^* = \sqrt{\frac{D\tau}{2}} P(1 - P) + O(\tau^2), \quad (4.86)$$

where $P$ is the probability (at $t = 0$) that the option is exercised at expiry ($t = T$). In the limit $\tau \to 0$, one indeed recovers $R^* = 0$, which also holds if $P \to 0$ or 1, since the outcome of the option then becomes certain. However, (4.86) already shows that in reality, the residual risk is not small. Take for example an at-the-money option, for which $P \approx \frac{1}{2}$. The comparison between the residual risk and the price of the option allows one to define a ‘quality’ ratio $Q$ for the hedging strategy:

$$Q \equiv \frac{R^*}{\hat{C}} \simeq \sqrt{\frac{\pi}{4N}}, \quad (4.87)$$

with $N = T/\tau$. For an option of maturity one month, rehedged daily, $N \approx 25$. Assuming Gaussian fluctuations, the quality ratio is then equal $Q \approx 0.2$. In other words, the residual risk is a fifth of the price of the option itself. Even if one rehedges every 30 minutes in a Gaussian world, the quality ratio is already $Q \approx 0.05$. If the increments are not Gaussian, then $R^*$ can never reach zero. This is actually rather intuitive, the presence of unpredictable price ‘jumps’ jeopardises the differential strategy of Black-Scholes. The miraculous compensation of the two terms in (4.84) no longer takes place. Figure 4.11 gives the residual risk as a function of $\tau$ for an option on the Bund contract, calculated using the optimal strategy $\phi^*$, and assuming independent increments. As expected, $R^*$ increases with $\tau$, but does not tend to zero when $\tau$ decreases: the theoretical quality ratio saturates around $Q = 0.17$. In fact, the real risk is even larger than the theoretical estimate since the model is imperfect. In particular, it ignores the volatility fluctuations, i.e. of the
scale factor $\gamma_k$. (The importance of these volatility fluctuations for determining the correct price was discussed above in 4.3.4). In other words, the theoretical curve shown in Fig. 4.11 neglects what is usually called the ‘volatility’ risk. A Monte-Carlo simulation of the profit and loss associated to at-the-money option, hedged using the optimal strategy $\phi^*$ determined above leads to the histogram shown in Fig. 4.8. The empirical variance of this histogram corresponds to $Q \simeq 0.28$, substantially larger than the theoretical estimate.

The ‘stop-loss’ strategy does not work

There is a very simple strategy that leads, at first sight, to a perfect hedge. This strategy is to hold $\phi = 1$ of the underlying as soon as the price $x$ exceeds the strike price $x_s$, and to sell everything ($\phi = 0$) as soon as the price falls below $x_s$. For zero interest rates, this ‘stop-loss’ strategy would obviously lead to zero risk, since either the option is exercised at time $T$, but the writer of the option has bought the underlying when its value was $x_s$, or the option is not exercised, but the writer does not possess any stock. If this were true, the price of the option would actually be zero, since in the global wealth balance, the term related to the hedge perfectly matches the option pay-off!

In fact, this strategy does not work at all. A way to see this is to realise that when the strategy takes place in discrete time, the price of the underlying is never exactly at $x_s$, but slightly above, or slightly below. If the trading time is $\tau$, the difference between $x_s$ and $x_k$ (where $k$ is the time is discrete units) is, for a random walk, on the order of $\sqrt{D\tau}$. The difference between the ideal strategy, where the buy or sell order is always executed precisely at $x_s$, and the real strategy is thus of the order of $N_k\sqrt{D\tau}$, where $N_k$ is the number of times the price crosses the value $x_s$ during the lifetime $T$ of the option. For an at the money option, this is equal to the number of times a random walk returns to its starting point in a time $T$. The result is well known, and is $N_k \propto \sqrt{T/\tau}$ for $T \gg \tau$. Therefore, the uncertainty in the final result due to the accumulation of the above small errors is found to be of order $\sqrt{DT}$, independently of $\tau$, and therefore does not vanish in the continuous time limit $\tau \to 0$. This uncertainty is of the order of the option price itself, even for the case of a Gaussian process. Hence, the ‘stop-loss’ strategy is not the optimal strategy, and was therefore not found as a solution of the functional minimisation of the risk presented above.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4.11}
\caption{Residual risk $R^*$ as a function of the time $\tau$ (in days) between adjustments of the hedging strategy $\phi^*$, for at-the-money options of maturity equal to 60 trading days. The risk $R^*$ decreases when the trading frequency increases, but only rather slowly: the risk only falls by 10% when the trading time drops from one day to thirty minutes. Furthermore, $R^*$ does not tend to zero when $\tau \to 0$, at variance with the Gaussian case (dotted line). Finally, the present model neglects the volatility risk, which leads to an even larger residual risk (marked by the symbol ‘M.C.’), corresponding to a Monte-Carlo simulation using real price changes.}
\end{figure}
Residual risk to first order in kurtosis

It is interesting to compute the first non-Gaussian correction to the residual risk, which is induced by a non zero kurtosis $\kappa_1$ of the distribution at scale $\tau$. The expression of $\frac{\partial R^*}{\partial \kappa_1}$ can be obtained from Eq. (4.84), where two types of terms appear: those proportional to $\frac{\partial P}{\partial \kappa_1}$, and those proportional $\frac{\partial x}{\partial \kappa_1}$. The latter terms are zero since by definition the derivative of $R^*$ with respect to $\phi^*$ is nil. We thus find:

$$\frac{\partial R^*}{\partial \kappa_1} = \int_{x_0}^\infty dx(x-x_0)( \max(N(x)-x_0,0)) \times \frac{\partial P(x,N|x_0,0)}{\partial \kappa_1} - D\tau \sum_k \int dx \frac{\partial P(x',k|x_0,0)}{\partial \kappa_1} \left[ \phi_k^N \right]^2. \quad (4.88)$$

Now, to first order in kurtosis, one has (cf. Chapter 1):

$$\frac{\partial P(x',k|x_0,0)}{\partial \kappa_1} \equiv k \left( \frac{D\tau}{4!} \right)^2 \left[ \phi_k^N \right]^4, \quad (4.89)$$

which allows one to estimate numerically the extra risk, using (4.88). The order of magnitude of all the above terms is given by $\frac{\partial R^*}{\partial \kappa_1} \approx D\tau/4!$. The relative correction to the Gaussian result (4.86) is then simply given, very roughly, by the kurtosis $\kappa_1$. Therefore, tail effects, as measured by the kurtosis, is the major source of the residual risk when the trading time $\tau$ is small.

Stochastic volatility models

It is instructive to consider the case where the fluctuations are purely Gaussian, but where the volatility itself is randomly varying. In other words, the instantaneous variance is given by $D_k = \bar{D} + \delta D_k$, where $\delta D_k$ is itself a random variable of variance $\langle \delta D^2 \rangle$. If the different $\delta D_k$'s are uncorrelated, this model leads to a non zero kurtosis (cf. (2.17) and Section 1.7.2) equal to $\kappa_1 = 3\langle \delta D^2 \rangle^2/D^2$.

Suppose then that the option trader follows the Black-Scholes strategy corresponding to the average volatility $\bar{D}$. This strategy is close, but not identical to, the optimal strategy which he would follow if he knew all the future values of $\delta D_k$ (this strategy is given in Appendix C). If $\delta D_k \ll \bar{D}$, one can perform a perturbative calculation of the residual risk associated to the uncertainty on the volatility. This excess risk is of order $\langle \delta D^2 \rangle$ since one is close to the absolute minimum of the risk, which is reached for $D \equiv 0$. The calculation indeed yields to a relative increase of the risk given, in order of magnitude, by:

$$\frac{\delta R^*}{R^*} \approx \frac{\langle \delta D^2 \rangle}{D^2}. \quad (4.90)$$

4.4.6 Other measures of risk – Hedging and VaR (\textsuperscript{*})

If the observed kurtosis is entirely due to this stochastic volatility effect, one has $\kappa^* \approx \kappa_1$. One thus recovers the result of the previous section. Again, this volatility risk can represent an appreciable fraction of the real risk, especially for frequently hedged options. Figure 4.11 actually suggests that the fraction of the risk due to the fluctuating volatility is comparable to that induced by the intrinsic kurtosis of the distribution.

It is conceptually illuminating to consider the model where the price increments $\delta x_k$ are Lévy variables of index $\mu < 2$, for which the variance is infinite. Such a model is also useful to describe extremely volatile markets, such as emergent countries (like Russia!), or very speculative assets (cf. Chapter 2). In this case, the variance of the global wealth balance $\Delta W$ is meaningless and cannot be used as a reliable measure of the risk associated to the option. Indeed, one expects that the distribution of $\Delta W$ behaves, for large negative $\Delta W$, as a power law of exponent $\mu$.

In order to simplify the discussion, let us come back to the case where the hedging strategy $\phi$ is independent of time, and suppose interest rate effects negligible. The wealth balance (4.28) then clearly shows that the catastrophic losses occur in two complementary cases:

- Either the price of the underlying soars rapidly, far beyond both the strike price $x_s$ and $x_0$. The option is exercised, and the loss incurred by the writer is then:

$$|\Delta W| = x_N(1-\phi) - x_s + \phi x_0 - C \approx x_N (1-\phi). \quad (4.91)$$

The hedging strategy $\phi$, in this case, limits the loss, and the writer of the option would have been better off holding $\phi = 1$ stock per written option.

- Or, on the contrary, the price plummets far below $x_0$. The option is then not exercised, but the strategy leads to important losses since:

$$|\Delta W| = \phi(x_0 - x_N) - C \approx -\phi x_N. \quad (4.92)$$

In this case, the writer of the option should not have held any stock at all ($\phi=0$).

However, both dangers are \textit{a priori} possible. Which strategy should one follow? Thanks to the above argument, it is easy to obtain the tail of the distribution of $\Delta W$ when $\Delta W \to -\infty$ (large losses). Since we
have assumed that the distribution of $x_N - x_0$ decreases as a power-law for large arguments,

$$P(x_N, N| x_0, 0) \sim \frac{\mu A_{\pm}^\mu}{|x_N - x_0|^{1+p}}, \quad (4.93)$$

it is easy to show, using the results of Appendix B, that:

$$P(\Delta W) \sim \frac{\mu \Delta W_0^\mu}{|\Delta W|^{1+p}} \quad \Delta W_0^\mu \equiv A_{\pm}^\mu(1 - \phi)^\mu + A_{\pm}^\phi \frac{\mu}{\mu - 1}. \quad (4.94)$$

The probability that the loss $|\Delta W|$ is larger than a certain value is then proportional to $\Delta W_0^\mu$ (cf. Chapter 3). The minimisation of this probability with respect to $\phi$ then leads to an optimal ‘Value-at-Risk’ strategy:

$$\phi^* = \frac{A_{\pm}^\mu}{A_{\pm}^\mu + A_{\pm}^\phi} \equiv \frac{\mu}{\mu - 1}, \quad (4.95)$$

for $1 < \mu \leq 2$. For $\mu < 1$, $\phi^*$ is equal to 0 if $A_{\pm} > A_{\pm}^\phi$ or to 1 in the opposite case. Several points are worth emphasising:

- The hedge ratio $\phi^*$ is independent of moneyness $(x_s - x_0)$. Because we are interested in minimising extreme risks, only the far tail of the wealth distribution matters. We have implicitly assumed that we are interested in moves of the stock price far greater than $|x_s - x_0|$, i.e., that moneyness only influences the centre of the distribution of $\Delta W$.

- It can be shown that within this Value-at-Risk perspective, the strategy $\phi^*$ is actually time independent, and also corresponds to the optimal instantaneous hedge, where the VaR between times $k$ and $k+1$ is minimum.

- Even if the tail amplitude $\Delta W_0$ is minimum, the variance of the final wealth is still infinite for $\mu < 2$. However, $\Delta W_0$ sets the order of magnitude of probable losses, for example with 95% confidence. As in the example of the optimal portfolio discussed in Chapter 3, infinite variance does not necessarily mean that the risk cannot be diversified. The option price is fixed for $\mu > 1$ by Eq. (4.33), ought to be corrected by a risk premium proportional to $\Delta W_0^\phi$. Note also that with such violent fluctuations, the smile becomes a spike!

\footnote{Note that the above strategy is still valid for $\mu > 2$, and corresponds to the optimal VaR hedge for power-law tailed assets: see below.}

- Finally, it is instructive to note that the histogram of $\Delta W$ is completely asymmetrical, since extreme events only contribute to the loss side of the distribution. As for gains, they are limited, since the distribution decreases very fast for positive $\Delta W$. In this case, the analogy between an option and an insurance contract is most clear, and shows that buying or selling an option are not at all equivalent operations, as they appear to be in a Black-Scholes world. Note that the asymmetry of the histogram of $\Delta W$ is visible even in the case of weakly non-Gaussian assets (Fig. (4.8).

As we have just discussed, the losses of the writer of an option can be very large in the case of a Lévy process. Even in the case of a ‘truncated’ Lévy process (in the sense defined in Chapters 1 and 2), the distribution of the wealth balance $\Delta W$ remains skewed towards the loss side. It then proportional to $\Delta W^\phi$ probability with respect to $\mu$ then leads to an optimal ‘Value-at-Risk’ strategy, minimising of this probability $|\Delta W|$ can therefore be justified to consider other measures of risk, not based on the variance but rather on higher moments of the distribution, such as the Black-Scholes hedging strategy with the value of the implied volatility, discussed in 4.4.3. Denoting the difference between the actual strategy and the optimal one by $\delta \phi(x, t)$, it is easy to show that for small $\delta \phi$, the increase in residual risk is given by:

$$\delta R^2 = D\tau \sum_{k=0}^{N-1} \int dx [\delta \phi(x, t_k)]^2 P(x, k| x_0, 0), \quad (4.96)$$

\footnote{For at the money options, one can actually show that $\Delta W \leq C$}
which is quadratic in the hedging error $\delta \phi$, and thus, in general, rather small. For example, we have estimated in Section 4.4.3 that $\delta \phi$ is at most equal to $0.02\kappa_N$, where $\kappa_N$ is the kurtosis corresponding to the terminal distribution. (See also Fig. (4.10)). Therefore, one has:

$$\delta R^2 < 4 \times 10^{-4} \kappa_N^2 DT. \quad (4.97)$$

For at-the-money options, this corresponds to a relative increase of the residual risk given by:

$$\frac{\delta R}{R} < 1.2 \times 10^{-3} \frac{\kappa_N^2}{Q^2}. \quad (4.98)$$

For a quality ratio $Q = 0.25$ and $\kappa_N = 1$, this represents a rather small relative increase equal to 2% at most. In reality, as numerical simulations show, the increase in risk induced by the use of the Black-Scholes $\Delta$-hedge rather than the optimal hedge is indeed only of a few percents for one month maturity options. This difference however increases as the maturity of the options decreases.

27In the following, we shall again stick to an additive model and discard interest rate effects, in order to focus on the main concepts.

4.5 Does the price of an option depend on the mean return?

4.5.1 The case of non-zero excess return

We should now come back to the case where the excess return $m_1 \equiv \langle \delta x_k \rangle = mT$ is non zero. This case is very important conceptually: indeed, one of the most striking result of Black and Scholes (besides the zero risk property) is that the price of the option and the hedging strategy are totally independent of the value of $m$. This may sound at first rather strange, since one could think that if $m$ is very large and positive, the price of the underlying asset one average increases fast, thereby increasing the average pay-off of the option. On the contrary, if $m$ is large and negative, the option should be worthless.

This argument actually does not take into account the impact of the hedging strategy on the global wealth balance, which is proportional to $m$. In other words, the term $\max(x(N) - x_s, 0)$, averaged with the historical distribution $P_m(x, N|x_0, 0)$, such that:

$$P_m(x, N|x_0, 0) = P_{m=0}(x - Nm_1, N|x_0, 0), \quad (4.99)$$

is indeed strongly dependent on the value of $m$. However, this dependence is partly compensated when one includes the trading strategy, and even vanishes in the Black-Scholes model.

Let us first present a perturbative calculation, assuming that $m$ is small, or more precisely that $(mT)^2/DT \ll 1$. Typically, for $T = 100$ days, $mT = 5\%100/365 = 0.014$ and $\sqrt{DT} \approx 1\%\sqrt{100} \approx 0.1$. The term of order $m^2$ that we neglect corresponds to a relative error of $(0.14)^2 \approx 0.02$.

The average gain (or loss) induced by the hedge is equal to:

$$\langle \Delta W_S \rangle = \int dx P_m(x, k|x_0, 0)\phi^*_N(x). \quad (4.100)$$

To order $m$, one can consistently use the general result (4.80) for the optimal hedge $\phi^*$, established above for $m = 0$, and also replace $P_m$ by $P_{m=0}$:

$$\langle \Delta W_S \rangle = \int dx P_0(x, k|x_0, 0) \quad (4.101)$$

$$\int_{x_s}^{+\infty} dx' (x' - x_s) \sum_{n=2}^{\infty} \frac{(-)^n c_{n1}}{D_T(n-1)!} \partial^{n-1} P_0(x', N|x, k),$$

where $P_0$ is the unbiased distribution ($m = 0$).

Now, using the Chapman-Kolmogorov equation for conditional probabilities:

$$\int dx P_0(x', N|x, k)P_0(x, k|x_0, 0) = P_0(x', N|x_0, 0), \quad (4.102)$$

one easily derives, after an integration by parts, and using the fact that
\( P_0(x', N|x_0, 0) \) only depends on \( x' - x_0 \), the following expression:
\[
\langle \Delta W_S \rangle = m_1 N \left[ \int_{x_s}^{+\infty} dx^t P_0(x^t, N|x_0, 0) \right. \\
+ \left. \sum_{n=3}^{\infty} \frac{c_{n,1}}{D^r(n-1)!} \frac{\partial^{n-3}}{\partial x_0^{n-3}} P_0(x_s, N|x_0, 0) \right]. \tag{4.103}
\]

On the other hand, the increase of the average pay-off, due to a non zero mean value of the increments, is calculated as:
\[
\langle \max(x(N) - x_s, 0) \rangle_m \equiv \int_{x_s}^{+\infty} dx^t \langle x^t - x_s \rangle P_m(x^t, N|x_0, 0) \tag{4.104}
= \int_{x_s}^{+\infty} dx^t (x^t + m_1 N - x_s) P_0(x^t, N|x_0, 0),
\]
where in the second integral, the shift \( x_k \to x_k + m_1 k \) allowed us to remove the bias \( m \), and therefore to substitute \( P_m \) by \( P_0 \). To first order in \( m \), one then finds:
\[
\langle \max(x(N) - x_s, 0) \rangle_m \approx \langle \max(x(N) - x_s, 0) \rangle_0 + m_1 N \int_{x_s}^{+\infty} dx^t P_0(x^t, N|x_0, 0). \tag{4.105}
\]

Hence, grouping together (4.103) and (4.105), one finally obtains the price of the option in the presence of a non zero average return as:
\[
C_m = C_0 - m T \sum_{n=3}^{\infty} \frac{c_{n,1}}{(n-1)!} \frac{\partial^{n-3}}{\partial x_0^{n-3}} P_0(x_s, N|x_0, 0). \tag{4.106}
\]

Quite remarkably, the correction terms are zero in the Gaussian case, since all the cumulants \( c_{n,1} \) are zero for \( n \geq 3 \). In fact, in the Gaussian case, this property holds to all orders in \( m \) (cf. below). However, for non-Gaussian fluctuations, one finds that a non zero return should in principle affect the price of the options. Using again Eq. (4.79), one can rewrite Eq. (4.106) in a simpler form as:
\[
C_m = C_0 + m T \left[ \mathcal{P} - \phi^* \right] \tag{4.107}
\]
where \( \mathcal{P} \) is the probability that the option is exercised, and \( \phi^* \) the optimal strategy, both calculated at \( t = 0 \). From Fig. (4.10) one sees that in the presence of ‘fat tails’, a positive average return makes out-of-the-money options less expensive (\( \mathcal{P} < \phi^* \)), while in-the-money options should be more expensive (\( \mathcal{P} > \phi^* \)).

Again, the Gaussian model (for which \( \mathcal{P} = \phi^* \)) is misleading: the independence of the option price with respect to the market ‘trend’ only holds for Gaussian processes, and is no longer valid in the presence of ‘jumps’. Note however that the correction is usually numerically rather small: for \( x_0 = 100 \), \( m = 10\% \) per year, \( T = 100 \) days, and \( |\mathcal{P} - \phi^*| \sim 0.02 \), one finds that the price change is on the order of 0.05 points, while \( \mathcal{C} \simeq 4 \) points.

\section*{Risk neutral probability}

It is interesting to notice that the result (4.106) can alternatively be rewritten as:
\[
\mathcal{C}_m = \int_{x_s}^{+\infty} dx \ (x - x_s) Q(x, N|x_0, 0), \tag{4.108}
\]
with an ‘effective probability’ (called ‘risk neutral probability’, or ‘pricing kernel’ in the mathematical literature) \( Q \) defined as:
\[
Q(x, N|x_0, 0) = P_0(x, N|x_0, 0) - \frac{m_1}{D^r} \sum_{n=3}^{\infty} \frac{c_{n,N}}{(n-1)!} \frac{\partial^{n-1}}{\partial x_0^{n-1}} P_0(x^t, N|x_0, 0), \tag{4.109}
\]
which satisfies the following equations:
\[
\int dx \ Q(x, N|x_0, 0) = 1, \tag{4.110}
\]
\[
\int dx \ (x - x_0) Q(x, N|x_0, 0) = 0. \tag{4.111}
\]

Using (4.79), Eq. (4.109) can also be written in a more compact way as:
\[
Q(x, N|x_0, 0) = P_0(x, N|x_0, 0) - \frac{m_1}{D^r} (x - x_0) P_0(x, N|x_0, 0) + m_1 \frac{\partial P_0(x, N|x_0, 0)}{\partial x_0}. \tag{4.112}
\]

Note however that Eq. (4.108) is rather formal, since nothing insures that \( Q(x, N|x_0, 0) \) is everywhere positive, except in the Gaussian case where \( Q(x, N|x_0, 0) = P_0(x, N|x_0, 0) \).

As we have discussed above, small errors in the hedging strategy do not significantly increase the risk. If the followed strategy is not the

\footnote{The fact that the optimal strategy is equal to the probability \( \mathcal{P} \) of exercising the option also holds in the Black-Scholes model, up to small \( \sigma^2 \) correction terms.}
optimal one but, for example, the Black-Scholes ‘Δ’-hedge (i.e. \( \phi = \Delta = \partial C / \partial \sigma \)), the fair game price is given by Eq. (4.108) with now \( Q(x,N|x_0,0) = P_0(x,N|x_0,0) \), and is now independent of \( m \) and positive everywhere. \(^{29}\) The difference between this ‘suboptimal’ price and the truly optimal one is then, according to (4.107), equal to \( \delta C = mT(\phi^* - \phi) \). As discussed above, the above difference is quite small, and leads to price corrections which are, in most cases, negligible compared to the uncertainty in the volatility.

**Optimal strategy in the presence of a bias**

We now give, without giving the details of the computation, the general equation satisfied by the optimal hedging strategy in the presence of a non zero average return \( m \), when the price fluctuations are arbitrary but uncorrelated. Assuming that \( \langle \delta x_\ell \delta x_{\ell'} \rangle - m^2 = D\delta \delta_{\ell \ell'} \), and introducing the unbiased variables \( \chi_k \equiv x_k - m_k \), one gets for \( \phi_k^*(x) \) the following (involved) integral equation:

\[
\begin{align*}
\left( D\tau \phi_k^*(x) - \int_{\chi_k}^\infty d\chi' (\chi' - \chi_k) \frac{\chi' - S}{N - k} P_0(\chi', N|x_k,k) \right) &= -m_1 \left\{ \int_{\chi_k}^\infty d\chi' (\chi' - \chi_k) [P_0(\chi', N|x_0,0) - P_0(\chi', N|x_k,k)] + \sum_{\ell=1}^{N-1} \int d\chi' \left[ \frac{\chi' - \chi_k}{\ell - k} + m_1 \phi_\ell^*(\chi') P_0(\chi', \ell|x_k,k) \right] \\
&\quad + \sum_{\ell=1}^{N-1} \int d\chi' \left[ \frac{\chi' - \chi_k}{\ell - k} + m_1 \phi_\ell^*(\chi') P_0(\chi', \ell|x_0,0) P_0(\chi_k, k|x_k,k) \right] = -m_1 \phi^* \right\}.
\end{align*}
\]

(4.113)

avec \( \chi = x_k - m_k N \) et

\[
\phi^* \equiv \sum_{k=0}^{N-1} \int d\chi' \phi_k^*(\chi') P_0(\chi', k|x_0,0). \tag{4.114}
\]

In the limit \( m_1 = 0 \), the right hand side vanishes and one recovers the optimal strategy determined above. For small \( m \), Eq. (4.113) is a convenient starting point for a perturbative expansion in \( m_1 \).

Using Eq. (4.113), one can establish a simple relation for \( \phi^* \), which fixes the correction to the ‘Bachelier price’ coming from the hedge:

\(^{29}\)If \( \phi = \Delta \), the result \( Q = P_0 \) is in fact correct (as we show in Appendix E) beyond the first order in \( m \). However, the optimal strategy is not, in general, given by the option \( \Delta \), except in the Gaussian case.
The price of the option in the presence of a non zero return \( m \neq 0 \) is thus given, in the Gaussian case, by:

\[
\mathcal{C}_m(x_0, x_s, T) = \int_0^\infty \frac{udu}{\sqrt{2\pi DT}} \exp \left\{ -\frac{(u - u_0 - mT)^2}{2DT} \right\} - m \bar{\phi}^\prime
\]

\[
= \int_0^\infty \frac{udu}{\sqrt{2\pi DT}} \exp \left\{ -\frac{(u - u_0)^2}{2DT} \right\}
\]

\[
\equiv \mathcal{C}_{m=0}(x_0, x_s, T),
\]

(cf. (4.43)). Hence, as announced above, the option price is indeed independent of \( m \) in the Gaussian case. This is actually a consequence of the fact that the trading strategy is fixed by \( \phi = \frac{\partial f}{\partial x} \), which is indeed correct (in the Gaussian case) even when \( m \neq 0 \). These results, rather painfully obtained here, are immediate within the framework of stochastic differential calculus, as originally used by Black and Scholes. It is thus interesting to pause for a moment and describe how option pricing theory is usually introduced.

**Ito Calculus**

The idea behind Ito's stochastic calculus is the following. Suppose that one has to consider a certain function \( f(x, t) \), where \( x \) is a time dependent variable. If \( x \) was an 'ordinary' variable, the variation \( \Delta f \) of the function \( f \) between time \( t \) and \( t + \tau \) would be given, for small \( \tau \), by:

\[
\Delta f = \frac{\partial f(x, t)}{\partial t} \tau + \frac{\partial f(x, t)}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 f(x, t)}{\partial x^2} \Delta x^2 + \ldots
\]

(4.120)

with \( \Delta x \equiv \frac{x - x_0}{\tau} \). The order \( \tau^2 \) in the above expansion looks negligible in the limit \( \tau \to 0 \). However, if \( x \) is a stochastic variable with independent increments, the order of magnitude of \( x(t) - x(0) \) is fixed by the CLT and is thus given by \( \sigma_1 \sqrt{t/\tau} \xi \), where \( \xi \) is a Gaussian random variable of width equal to 1, and \( \sigma_1 \) is the RMS of \( \Delta x \).

If the limit \( \tau \to 0 \) is to be well defined and non trivial (i.e. such that the random variable \( \xi \) still plays a role), one should thus require that \( \sigma_1 \propto \sqrt{\tau} \). Since \( \sigma_1 \) is the RMS of \( \frac{\Delta x}{\sqrt{\tau}} \), this means that the order of magnitude of \( \frac{\Delta x}{\sqrt{\tau}} \) is proportional to \( 1/\sqrt{\tau} \). Hence the order of magnitude of \( (\frac{\Delta x}{\sqrt{\tau}})^2 \tau^2 \) is not \( \tau^2 \) but \( \tau \): one should therefore keep this term in the expansion of \( \Delta f \) to order \( \tau \).

The crucial point of Ito's differential calculus is that if the stochastic process is a continuous time Gaussian process, then for any small but finite time scale \( \tau \), \( \Delta x \) is already the result of an infinite sum of elementary increments. Therefore, one can rewrite Eq. (4.120), choosing

\[
\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial^2 f}{\partial x^2} \frac{\Delta x^2}{\tau} + \ldots
\]

as an elementary time step \( \tau' \ll \tau \), and sum all these \( \tau'/\tau \) equations to obtain \( \Delta f \) on the scale \( \tau \). Using the fact that for small \( \tau \), \( \partial f/\partial x \) and \( \partial^2 f/\partial x^2 \) do not vary much, one finds:

\[
\Delta x = \sum_{i=1}^{\tau/\tau'} \Delta x_i, \quad \Delta x^2 = \sum_{i=1}^{\tau/\tau'} \Delta x_i^2.
\]

(4.121)

Using again the CLT between scales \( \tau' \ll \tau \) and \( \tau \), one finds that \( \Delta x \) is a Gaussian variable of RMS \( \propto \sqrt{\tau} \). On the other hand, since \( \Delta x_i^2 \) is the sum of positive variables, it is equal to its mean \( \sigma_i^2 \tau'/\tau' \) plus terms of order \( \sqrt{\tau'/\tau'} \), where \( \sigma_i^2 \) is the variance of \( \Delta x_i \). For consistency, \( \sigma_i^2 \) is of order \( \tau' \); we will thus set \( \sigma_i^2 \equiv \sigma_i^2 \equiv \sigma^2 \).

Hence, in the limit \( \tau' \to 0 \), with \( \tau \) fixed, \( \Delta x^2 \) in (4.120) becomes a non random variable equal to \( \sigma^2 \tau \), up to corrections of order \( \sqrt{\tau'/\tau} \). Now, taking the limit \( \tau \to 0 \), one finally finds:

\[
\lim_{\tau \to 0} \frac{\Delta f}{\tau} \equiv \frac{df(x, t)}{dt} = \frac{df(x, t)}{dt} + \frac{df(x, t)}{dx} \frac{dx}{dt} + \frac{D}{2} \frac{\partial^2 f(x, t)}{\partial x^2},
\]

(4.122)

where \( \lim_{\tau \to 0} \frac{\Delta x}{\tau} \equiv \frac{dx}{dt} \). Equation (4.122) means that in the double limit \( \tau \to 0, \tau'/\tau' \to 0 \):

- The second order derivative term does not fluctuate, and thus does not depend on the specific realisation of the random process. This is obviously at the heart of the possibility of finding a riskless hedge in this case.\(^{31}\)
- Higher order derivatives are negligible in the limit \( \tau = 0 \).
- Equation (4.122) remains valid in the presence of a non zero bias \( m \).

Let us now apply the formula (4.122) to the difference \( df/dt \) between \( dC/dt \) and \( \phi dx/dt \). This represents the difference in the variation of the value of the portfolio of the buyer of an option, which is worth \( C \), and that of the writer of the option who holds \( \phi(x, t) \) underlying assets. One finds:

\[
\frac{df}{dt} = \phi \frac{dx}{dt} - \left[ \frac{\partial C(x, x_s, T - t)}{\partial t} + \frac{\partial C(x, x_s, T - t)}{\partial x} \frac{dx}{dt} \right] + \frac{D}{2} \frac{\partial^2 C(x, x_s, T - t)}{\partial x^2}. \quad \quad (4.123)
\]

One thus immediately sees that if \( \phi = \phi^* \equiv \frac{\partial C(x, x_s, T - t)}{\partial x} \), the coefficient of the only random term in the above expression, namely \( \frac{dx}{dt} \), vanishes.

\(^{31}\)It is precisely for the same reason that the risk is also zero for a binomial process, where \( \Delta x \) can only take two values – see Appendix D.
The evolution of the difference of value between the two portfolios is then known with certainty! In this case, no party would agree on the contract unless this difference remains fixed in time (we assume here, as above, that the interest rate is zero). In other words, \( \frac{\partial}{\partial t} \equiv 0 \), leading to a partial differential equation for the price \( C \):

\[
\frac{\partial C(x, x_s, T - t)}{\partial t} = -\frac{D}{2} \frac{\partial^2 C(x, x_s, T - t)}{\partial x^2},
\]

with boundary conditions: \( C(x, x_s, 0) \equiv \max(x - x_s, 0) \), i.e. the value of the option at expiry – see Eq. (4.48). The solution of this equation is the above result (4.43) obtained for \( r = 0 \), and the Black-Scholes strategy is obtained by taking the derivative of the price with respect to \( x \), since this is the condition under which \( \frac{\partial}{\partial x} \) completely disappears from the game. Note also that the fact that the average return \( m \) is zero or non zero does not appear in the above calculation. The price and the hedging strategy are therefore completely independent of the average return in this framework, a result obtained after rather cumbersome considerations above.

**4.5.3 Conclusion. Is the price of an option unique?**

Summarising the above section, the Gaussian, continuous time limit allows one to use very simple differential calculus rules, which only differ from the standard one through the appearance of a second order non fluctuating term — the so-called ‘Ito correction’. The use of this calculus immediately leads to the two main results of Black and Scholes, namely: the existence of a riskless hedging strategy, and the fact that the knowledge of the average trend is not needed. These two results are however not valid as soon as the hypothesis underlying Ito’s stochastic calculus are violated (continuous time, Gaussian statistics). The approach based on the global wealth balance, presented in the above sections, is by far less elegant but more general. It allows one to understand the very peculiar nature of the limit considered by Black and Scholes.

As we have already discussed, the existence of a non zero residual risk (and more precisely of a negatively skewed distribution of the optimised wealth balance) necessarily means that the bid and ask prices of an option will be different, because the market makers will try to compensate for part of this risk. On the other hand, if the average return \( m \) is not zero, the fair price of the option explicitly depends on the optimal strategy \( \phi^* \), and thus of the chosen measure of risk (as was the case for portfolios, the optimal strategy corresponding to a minimal variance of the final result is different from the one corresponding to a minimum Value-at-Risk). The price therefore depends on the operator, of his definition of risk and of his ability to hedge this risk. In the Black-Scholes model, the price is uniquely determined since all definitions of risk are equivalent (and are all zero!). This property is often presented as a major advantage of the Gaussian model. Nevertheless, it is clear that it is precisely the existence of an ambiguity on the price that justifies the very existence of option markets! A market can only exist if some uncertainty remains. In this respect, it is interesting to note that new markets continuously open, where more and more sources of uncertainty become tradable. Option markets correspond to a risk transfer: buying or selling a call are not identical operations, except in the Black-Scholes world where options would actually be useless, since they would be equivalent to holding a certain number of the underlying asset (given by the \( \Delta \)).

**4.6 Conclusion of the chapter: the pitfalls of zero-risk**

The traditional approach to derivative pricing is to find an ideal hedging strategy, which perfectly duplicates the derivative contract. Its price is then, using an arbitrage argument, equal to that of the hedging strategy, and the residual risk is zero. This argument appears as such in nearly all the available books on derivatives and on the Black-Scholes model. For example, the last chapter of J.C. Hull’s remarkable book, called ‘Review of Key Concepts’, starts by the following sentence: The pricing of derivatives involves the construction of riskless hedges from tradable securities. Although there is a rather wide consensus on this point of view, we feel that it is unsatisfactory to base a whole theory on exceptional situations: as explained above, both the continuous time Gaussian model and the binomial model are very special models indeed. We think that it is more appropriate to start from the ingredient which allow the derivative markets to exist in the first place, namely risk. In this respect, it is interesting to compare the above quote from Hull to the following disclaimer, found on the Chicago Board of Exchange document on options: Option trading involves risk!

The idea that zero risk is the exception rather than the rule is important for a better pedagogy of financial risks in general; an adequate estimate of the residual risk – inherent to the trading of derivatives – has actually become one of the major concern of risk management (see also Sections 5.2, 5.3). The idea that the risk is zero is inadequate be-

\[ ^{32} \text{This ambiguity is related to the residual risk, which, as discussed above, comes both from the presence of price ‘jumps’, and from the uncertainty on the parameters describing the distribution of price changes (‘volatility risk’).} \]
cause zero cannot be a good approximation of anything. It furthermore provides a feel of apparent security which can reveal disastrous on some occasions. For example, the Black-Scholes strategy allows one, in principle, to hold an insurance against the fall of one’s portfolio without buying a true Put option, but rather by following the associated hedging strategy. This is called ‘Insurance Portfolio’, and was much used in the eighties, when faith in the Black-Scholes model was at its highest. The idea is simply to sell a certain fraction of the portfolio when the market goes down. This fraction is fixed by the Black-Scholes $\Delta$ of the virtual option, where the strike price is the security level under which the investor does not want to find himself. During the 1987 crash, this strategy has been particularly inefficient: not only because crash situations are the most extremely non-Gaussian events that one can imagine (and thus the zero-risk idea is totally absurd), but also because this strategy feeds back onto the market to make it crash further (a drop of the market mechanically triggers sell orders). According to the Brady commission, this mechanism has indeed significantly contributed to enhance the amplitude of the crash (see the discussion in [Hull]).

### 4.7 Appendix C: computation of the conditional mean

On many occasions in this chapter, we have needed to compute the mean value of the instantaneous increment $\delta x_k$, restricted on trajectories starting from $x_k$ and ending at $x_N$. We assume that the $\delta x_k$’s are identically distributed, up to a scale factor $\gamma_k$. In other words:

$$P_{ik}(\delta x_k) \equiv \frac{1}{\gamma_k} P_{10}(\frac{\delta x_k}{\gamma_k}).$$ (4.125)

The quantity we wish to compute is then:

$$P(x_N, N|x_k, k)(\delta x_k)(x_k, k)\rightarrow(x_N, N) =$$ (4.126)

$$\int \left[ \prod_{j=k}^{N-1} \frac{d\delta x_j}{\gamma_j} P_{10}(\frac{\delta x_j}{\gamma_j}) \right] \delta x_k \delta(x_N - x_k - \sum_{j=k}^{N-1} \delta x_j),$$

where the $\delta$ function insures that the sum of increments is equal to $x_N - x_k$. Using the Fourier representation of the $\delta$ function, the right hand side of this equation reads:

$$\int \frac{dz}{2\pi} e^{iz(x_N-x_k)} i \gamma_k \hat{P}_{10}(\gamma_k z) \left[ \prod_{j=k+1}^{N-1} \hat{P}_{10}(\gamma_j z) \right].$$ (4.127)

- In the case where all the $\gamma_k$’s are equal to $\gamma_0$, one recognises:

$$\int \frac{dz}{2\pi} e^{iz(x_N-x_k)} i \frac{\partial}{N-k \partial z} [\hat{P}_{10}(z\gamma_0)]^{N-k}.$$ (4.128)

Integrating by parts and using the fact that:

$$P(x_N, N|x_k, k) \equiv \int \frac{dz}{2\pi} e^{iz(x_N-x_k)} [\hat{P}_{10}(z\gamma_0)]^{N-k},$$ (4.129)

one finally obtains the expected result:

$$\langle \delta x_k \rangle(x_N, N) \equiv \frac{x_N - x_k}{N - k}.$$ (4.130)

- In the case where the $\gamma_k$’s are different from one another, one can write the result as a cumulant expansion, using the cumulants $c_{n,1}$ of the distribution $P_{10}$. After a simple computation, one finds:

$$P(x_N, N|x_k, k)(\delta x_k)(x_k, k)\rightarrow(x_N, N) =$$

$$\sum_{n=2}^{\infty} \frac{(-\gamma_k)^n c_{n,1}}{(n-1)!} \frac{\partial^n}{\partial x_N^n} P(x_N, N|x_k, k);$$ (4.131)

which allows one to generalise the optimal strategy in the case where the volatility is time dependent. In the Gaussian case, all the $c_n$ for $n \geq 3$ are zero, and the last expression boils down to:

$$\langle \delta x_k \rangle(x_N, N) = \frac{\gamma_k^2 (x_N - x_k)}{2 \sum_{j=k}^{N-1} \gamma_j^2}.$$ (4.132)

This expression can be used to show that the optimal strategy is indeed the Black-Scholes $\Delta$, for Gaussian increments in the presence of a non zero interest rate. Suppose that

$$x_{k+1} - x_k = \rho x_k + \delta x_k, \quad (4.133)$$

where the $\delta x_k$ are identically distributed. One can then write $x_N$ as:

$$x_N = x_0 (1 + \rho)^N + \sum_{k=0}^{N-1} \delta x_k (1 + \rho)^{N-k-1}.$$ (4.134)

The above formula (4.132) then reads:

$$\langle \delta x_k \rangle(x_N, N) = \frac{\gamma_k^2 (x_N - x_k (1 + \rho)^{N-k})}{\sum_{j=k}^{N-1} \gamma_j^2}.$$ (4.135)

with $\gamma_k = (1 + \rho)^{N-k-1}$.
4.8 Appendix D: binomial model

The binomial model for price evolution is due to Cox, Ross et Rubinstein, and shares with the continuous time Gaussian model the zero risk property. This model is very much used in practice [Hull], due to its easy numerical implementation. Furthermore, the zero risk property appears in the clearest fashion. Suppose indeed that between \( t_k = k\tau \) and \( t_{k+1} \), the price difference can only take two values: \( \delta x_k = \delta x_{1,2} \). For this very reason, the option value can only evolve along two paths only, to be worth (at time \( t_{k+1} \)) \( \sum_{i=1}^{k+1} x_i \). Consider now the hedging strategy where one holds a fraction \( \phi \) of the underlying asset, and a quantity \( B \) of bonds, with a risk free interest rate \( \rho \). If one chooses \( \phi \) and \( B \) such that:

\[
\phi_k(x_k + \delta x_1) + B_k(1 + \rho) = C^{k+1}_1, \quad (4.136)
\]

\[
\phi_k(x_k + \delta x_2) + B_k(1 + \rho) = C^{k+1}_2, \quad (4.137)
\]

one sees that in the two possible cases \( (\delta x_k = \delta x_{1,2}) \), the value of the hedge is strictly equal to the option value. The option value at time \( t_k \) is thus equal to \( C^k(x_k) = \phi_k x_k + B_k \), independently of the probability \( p \) \( (1-p) \) that \( \delta x_k = \delta x_1 \) \( | \delta x_2 \). One then determines the option value by iterating this procedure from time \( k+1 = N \), where \( C^N \) is known, and equal to \( \max(x_N - x_s,0) \). It is however easy to see that as soon as \( \delta x_k \) can take three or more values, it is impossible to find a perfect strategy.\(^{33}\)

The Ito process can be obtained as the continuum limit of the binomial tree. But even in its discrete form, the binomial model shares some important properties with the Black-Scholes model. The independence of the premium on the transition probabilities is the analog of the independence of the premium on the excess return in the Black-Scholes model. The magic of zero risk in the binomial model can therefore be understood as follows. Consider the quantity \( s^2 = \langle \delta x_k \rangle^2 \). \( s^2 \) is in principle random, but since changing the probabilities does not modify the option price one can pick \( p = 1/2 \), making \( s^2 \) non-fluctuating \( (s^2 = \langle \delta x_1 \rangle^2 / 4) \). The Ito process shares this property: in the


4.9 Appendix E: option price for (suboptimal) \( \Delta \)-hedging.

If \( \phi = \Delta \), the fact that the ‘risk neutral’ probability \( Q(x,N|x_0,0) \) is equal to \( P_b(x,N|x_0,0) \) is correct, for \( N \) large, beyond the first order in \( m \), as we show now. Taking \( \phi = \partial C / \partial x \) leads to an implicit equation for \( C \):

\[
C(x_0, x_s, N) = \int dx' \mathcal{Y}(x' - x_s) P_m(x', N|x_0, 0) \quad (4.140)
\]

\[
- m_1 \sum_{k=0}^{N-1} \int dx'' \frac{\partial C(x'', x_s, N - k)}{\partial x''} P_m(x'', k|x_0, 0),
\]

with \( \mathcal{Y} \) representing the pay-off of the option. This equation can be solved by making the following ansatz for \( C \):

\[
C(x'', x_s, N - k) = \int dx' \Omega(x' - x_s, N - k) P_m(x', N|x''', k) \quad (4.141)
\]

where \( \Omega \) is an unknown kernel which we try to determine. Using the fact that the option price only depends on the price difference between the strike price and the present price of the underlying, this gives:

\[
\int dx' \Omega(x' - x_s, N) P_m(x', N|x_0, N) = \quad (4.142)
\]

\[
\int dx' \mathcal{Y}(x' - x_s) P_m(x', N|x_0, 0) + m_1 \sum_{k=0}^{N-1} \int dx'' P_m(x'', k|x_0, 0) \times \frac{\partial}{\partial x''} \int dx' \Omega(x' - x_s, N - k) P_m(x', N|x''', k).
\]
Now, using again the Chapman-Kolmogorov equation:
\[
\int dx \ P_m(x', N|x, k) P_m(x, k|x_0, 0) = P_m(x', N|x_0, 0),
\]
(4.143)
one obtains the following equation for \(\Omega\) (after changing \(k \rightarrow N - k\)):
\[
\Omega(x' - x_s, N) + m_1 \sum_{k=1}^{N} \frac{\partial \Omega(x' - x_s, k)}{\partial x'} = \gamma(x' - x_s)
\]
(4.144)
The solution to this equation is simply \(\Omega(x' - x_s) = \gamma(x' - x_s - m_1 N)\). Indeed, if this is the case, one has:
\[
\frac{\partial \Omega(x' - x_s, k)}{\partial x'} \equiv - \frac{1}{m_1} \frac{\partial \Omega(x' - x_s, k)}{\partial k}
\]
(4.145)
and therefore:
\[
\gamma(x' - x_s - m_1 N) - \sum_{k=1}^{N} \frac{\partial \gamma(x' - x_s - m_1 k)}{\partial k} \simeq \gamma(x' - x_s)
\]
(4.146)
where the last equality holds in the large \(N\) limit, when the sum over \(k\) can be approximated by an integral. Therefore, the price of the option is given by:
\[
C = \int dx' \gamma(x' - x_s - m_1 N) P_m(x', N|x_0, 0).
\]
(4.147)
Now, using the fact that \(P_m(x', N|x_0, 0) \equiv P_0(x' - m_1 N, N|x_0, 0)\), and changing variable from \(x' \rightarrow x' - m_1 N\), one finally finds:
\[
C = \int dx' \gamma(x' - x_s) P_0(x', N|x_0, 0).
\]
(4.148)
thereby proving that the pricing kernel \(Q\) is equal to \(P_0\) if the chosen hedge is the \(\Delta\). Note that, interestingly, this is true independently of the particular pay-off function \(\gamma\): \(Q\) thus has a rather general status.

4.10 References

- **Market efficiency:**
  
  

- **Optimal Filters:**
  
  

- **Options and Futures:**
  
  

- **Stochastic differential calculus and derivative pricing:**
  
  

Some classics:


• Option pricing in the presence of residual risk:


• Kurtosis and implied cumulants:


• Stochastic volatility models:


5

OPTIONS: SOME MORE SPECIFIC PROBLEMS

This chapter can be skipped at first reading.
(J.-P. Bouchaud, M. Potters, Theory of financial risk.)

5.1 Other elements of the balance sheet

We have until now considered the simplest possible option problem, and tried to extract the fundamental ideas associated to its pricing and hedging. In reality, several complications appear, either in the very definition of the option contract (see next section), or in the elements that must be included in the wealth balance – for example dividends or transaction costs – that we have neglected up to now.

5.1.1 Interest rate and continuous dividends

The influence of interest rates (and continuous dividends) can be estimated using different models for the statistics of price increments. These models give slightly different answers; however, in a first approximation, they provide the following answer for the option price:

$$C(x_0, x_s, T, r) = e^{-rT}C(x_0e^{rT}, x_s, T, r = 0), \quad (5.1)$$

where $C(x_0, x_s, T, r = 0)$ is simply given by the average of the pay-off over the terminal price distribution. In the presence of a non zero continuous dividend $d$, the quantity $x_0e^{rT}$ should be replaced by $x_0e^{(r-d)T}$, i.e. in all cases the present price of the underlying should be replaced by the corresponding forward price.

Let us present three different models leading to the above approximate formula; but with slightly different corrections to it in general. These three models offer alternative ways to understand the influence of a non zero interest rate (and/or dividend) on the price of the option.
5 Options: some more specific problems

5.1 Other elements of the balance sheet

Writing \( x_k = x_0 + \sum_{j=0}^{k-1} \delta x_j \), we find that \( \langle \Delta W_S \rangle \) is the sum of a term proportional to \( x_0 \) and the remainder. The first contribution therefore reads:

\[
\langle \Delta W_S \rangle_1 = - (\rho - \delta) x_0 \sum_{k=0}^{N-1} \int dx P(x, k|x_0, 0) \phi_k^N(x).
\]
(5.5)

This term thus has exactly the same shape as in the case of a non zero average return treated above, Eq. (4.100), with an effective mean return \( m_{1\text{eff}} = - (\rho - \delta) x_0 \). If we choose for \( \phi^* \) the suboptimal \( \Delta \)-hedge (which we know does only induce minor errors), then, as shown above, this hedging cost can be reabsorbed in a change of the terminal distribution, where \( x_N - x_0 \) becomes \( x_N - x_0 - m_{1\text{eff}} N \), or else: \( x_N - x_0 [1 + (\rho - \delta) N] \).

To order \( (\rho - \delta) N \), this therefore again corresponds to replacing the terminal distribution with the 'dividend' \( k \delta \) is the dividend over a unit time period \( \delta \) is the interest rate of the reference currency). However, the terminal distribution of \( x_N \) must be constructed by independence between price increments and interest rates-dividends

\[
C_k = N \frac{\partial}{\partial x_0} \int dx P(x, k|x_0, 0) \phi_k^N(x_0).
\]
(5.9)

The contribution to the cost of the strategy is then obtained by summing over \( k \) and over \( j < k \), finally leading to:

\[
\langle \Delta W_S \rangle_2 \sim - \frac{N^2}{2} (\rho - \delta) \delta T P(x_s, N|x_0, 0).
\]
(5.10)

This extra cost is maximum for at-the-money options, and leads to a relative increase of the call price given, for Gaussian statistics, by:

\[
\frac{\delta C}{C} = \frac{N}{2} (\rho - \delta) = \frac{T}{2} (r - d),
\]
(5.11)

which is thus quite small for short maturities. (For \( r = 5\% \) annual, \( d = 0 \), and \( T = 100 \) days, one finds \( \delta C/C \leq 1\%).

Independence between price increments and interest rates-dividends

The above model is perhaps not very realistic since it assumes a direct relation between price changes and interest rates/dividends. It might be more appropriate to consider the case where \( x_{k+1} = x_k - \delta x_k \) as an other extreme case; reality is presumably in between these two limiting models. Now the terminal distribution is a function of the difference \( x_N - x_0 \), with no correction to the variance brought about by interest rates. However, in the present case, the cost of the hedging strategy cannot be neglected and reads:

\[
\langle \Delta W_S \rangle = -(\rho - \delta) \sum_{k=0}^{N-1} (x_k \phi_k^N(x_k)).
\]
(5.4)
Multiplicative model

Let us finally assume that \( x_{k+1} - x_k = (\rho - \delta + \eta_k)x_k \), where the \( \eta_k \)'s are independent random variables of zero mean. Therefore, the average cost of the strategy is zero. Furthermore, if the elementary time step \( \tau \) is small, the terminal value \( x_N \) can be written as:

\[
\log \left( \frac{x_N}{x_0} \right) = \sum_{y=0}^{N-1} \log(1+\rho - \delta + \eta_y) \simeq N(\rho - \delta) + y;
\]

or else \( x_N = x_0 e^{(r-d)T+y} \). Introducing the distribution \( P(y,N) \), the price of the option can be written as:

\[
C(x_0, x_s, T, r, d) = e^{-rT} \int dy P(y) \max(x_0 e^{(r-d)T+y} - x_s, 0),
\]

which is obviously of the general form (5.1).

We thus conclude that the simple rule (5.1) above is, for many purposes, sufficient to account for interest rates and (continuous) dividends effects. A more accurate formula, including terms of order \( (r-d)T \) depends somewhat on the chosen model.

5.1.2 Interest rates corrections to the hedging strategy

It is also important to estimate the interest rate corrections to the optimal hedging strategy. Here again, one could consider the above models, that is, the systematic drift model, the independent model or the multiplicative model. In the former case, the result for a general terminal price distribution is rather involved, mainly due to the appearance of the conditional average detailed in Appendix C, Eq. (4.138). In the case where this distribution is Gaussian, the result is simply Black-Scholes’ \( \Delta \) hedge, i.e., the optimal strategy is the derivative of the option price with respect to the price of the underlying contract (see Eq. 4.81). As discussed above, this simple rule leads in general to a suboptimal strategy, but the relative increase of risk is rather small. The \( \Delta \)-hedge procedure, on the other hand, has the advantage that the price is independent of the average return (see Appendix E).

In the ‘independent’ model, where the price change is unrelated to the interest rate and/or dividend, the order \( \rho \) correction to the optimal strategy is easier to estimate in general. From the general formula, Eq. (4.116), one finds:

\[
\phi_k^*(x) = (1+\rho)^{1+k-N} \phi_k^{*0}(x) - \frac{r-d}{D} x \mathbb{C}(x, x_s, N-k) \quad \text{or more accurately,}
\]

\[
\phi_k^*(x) = \frac{(1+\rho)^{1+k-N}}{x \sigma^2(N-k)} \times \int dx' \frac{x' - x}{\ell - k} P(x', \ell, x_k) \phi_k^{*0}(x')
\]

where \( \phi_k^{*0} \) is the optimal strategy for \( r = d = 0 \).

Finally, for the multiplicative model, the optimal strategy is given by a much simpler expression:

\[
\phi_k^*(x) = \frac{(1+\rho)^{1+k-N}}{x \sigma^2(N-k)} \times \int dx' \frac{x' - x}{\ell(1+\rho - \delta)^{N-k}} P(x', N|x, k)
\]

5.1.3 Discrete dividends

More generally, for an arbitrary dividend \( d_k \) (per share) at time \( t_k \), the extra term in the wealth balance reads:

\[
\Delta W_D = \sum_{k=1}^{N} \phi_k^N(x_k) d_k.
\]

Very often, this dividend occurs once a year, at a given date \( d_0 \). In this case, the corresponding share price decreases immediately by the same amount (since the value of the company is decreased by an amount equal to \( d_0 \) times the number of emitted shares): \( x \rightarrow x - d_0 \). Therefore, the net balance \( d_k + \delta x_k \) associated to the dividend is zero. For the same reason, the probability \( P_0(x, N|x_0, 0) \) is given, for \( N > k \), by \( P_{k=0}(x + d_0, N|x_0, 0) \). The option price is then equal to:

\[
C_{d_0}(x, x_s, N) = \mathbb{C}(x, x_s + d_0, N).
\]

If the dividend \( d_0 \) is not known in advance with certainty, this last equation should be averaged over a distribution \( P(d_0) \) describing as well as possible the probable values of \( d_0 \). A possibility is to choose the distribution of least information (maximum entropy) such that the average dividend is fixed to \( \overline{d} \):

\[
P(d_0) = \frac{1}{\overline{d}} \exp(-d_0/\overline{d}).
\]
5.2 Other types of options: ‘Puts’ and ‘exotic options’

The fixed part of the transaction costs is easier to discuss. If these costs are equal to \( \nu \) per transaction, and if the hedging strategy is rebalanced every \( \tau \), the total cost incurred is simply given by:

\[
\Delta W_{tr} = N\nu'.
\]

Comparing the two types of costs leads to:

\[
\frac{\Delta W_{tr}'}{\Delta W_{tr}} \propto \frac{\nu'\sqrt{N}}{\nu},
\]

showing that the latter costs can exceed the former when \( N = \frac{T}{\tau} \) is large, i.e. when the hedging frequency is high.

In summary, the transaction costs are higher when the trading frequency is high. On the other hand, decreasing of \( \tau \) allows one to diminish the residual risk (Fig. 4.11). This analysis suggests that the optimal trading frequency should be chosen such that the transaction costs are comparable to the residual risk.

5.2 Other types of options: ‘Puts’ and ‘exotic options’

5.2.1 ‘Put-Call’ parity

A ‘Put’ contract is a sell option, defined by a pay-off at maturity given by \( \max(x_s - x, 0) \). A Put protects its owner against the risk that the shares he owns drops below the strike price \( x_s \). Puts on stock indices like the S&P 500 are very popular. The price of a European Put will be noted \( C[0,x_s,T] \) (we reserve the notation \( \mathcal{P} \) to a probability). This price can be obtained using a very simple ‘parity’ (or no arbitrage) argument. Suppose that one simultaneously buys a Call with strike price \( x_s \) and a daily volatility of \( \sigma_0 \). One can therefore write that:

\[
\mathcal{C}[x_0,x_s,T] \approx \mathcal{C}[0,x_s,T] + \sigma_0/N.
\]

The transaction costs are in general proportional to \( x_0 \); taking for example \( \nu = 10^{-3}x_0 \), \( \tau = 1 \) day, and a daily volatility of \( \sigma_1 = 1\% \), one finds that the transaction costs represent 1% of the option price. On the other hand, for higher transaction costs (say \( \nu = 10^{-2}x_0 \), a daily rehedging becomes absurd, since the ratio (5.21) is of order 1. The rehedging frequency should then be lowered, such that the volatility on the scale of \( \tau \) increases, to become at least comparable to \( \nu \).
Note in particular that at the money \( x_s = x_0 \), the two contracts have the same value, which is obvious by symmetry (again, in the absence of interest rate effects).

### 5.2.2 ‘Digital’ options

More general option contracts can stipulate that the pay-off is not the difference between the value of the underlying at maturity \( x_N \) and the strike price \( x_s \), but rather an arbitrary function \( \mathcal{Y}(x_N) \) of \( x_N \). For example, a ‘digital’ option is a pure bet, in the sense that it pays a fixed premium \( \mathcal{Y}_0 \) whenever \( x_N \) exceeds \( x_s \). Therefore:

\[
\mathcal{Y}(x_N) = \mathcal{Y}_0 \quad (x_N > x_s); \quad \mathcal{Y}(x_N) = 0 \quad (x_N < x_s).
\] (5.27)

The price of the option in this case can be obtained following the same lines as above. In particular, in the absence of bias (i.e. for \( m = 0 \)) the fair price is given by:

\[
\mathcal{C}_\mathcal{Y}(x_0, N) = \langle \mathcal{Y}(x_N) \rangle = \int dx \ \mathcal{Y}(x) P_0(x, N|x_0, 0),
\] (5.28)

while the optimal strategy is still given by the general formula (4.33). In particular, for Gaussian fluctuations, one always finds the Black-Scholes recipe:

\[
\phi'_\mathcal{Y}(x) \equiv \frac{\partial \mathcal{C}_\mathcal{Y}[x, N]}{\partial x}.
\] (5.29)

The case of a non zero average return can be treated as above, in Section 4.5.1. To first order in \( m \), the price of the option is given by:

\[
\mathcal{C}_\mathcal{Y}, m = \mathcal{C}_\mathcal{Y}, m = 0 - \frac{mT}{\Delta} \sum_{n=1}^{\infty} \frac{c_{n, 1}}{(6 - 1)!} \int dx' \ \mathcal{Y}(x') \frac{\partial^{n-1}}{\partial x^{n-1}} P_0(x', N|x_0, 0),
\] (5.30)

which reveals, again, that in the Gaussian case, the average return disappears from the final formula. In the presence of non zero kurtosis, however, a (small) systematic correction to the fair price appears. Note that \( \mathcal{C}_\mathcal{Y}, m \) can be written as an average of \( \mathcal{Y}(x) \) using the effective, ‘risk neutral’ probability introduced in Section 4.5.1. If the \( \Delta \)-hedge is used, this risk neutral probability is simply \( P_0 \) (see Appendix E).

#### 5.2.3 ‘Asian’ options

The problem is slightly more complicated in the case of the so-called ‘Asian’ options. The pay-off of these options is calculated not on the value of the underlying stock at maturity, but on a certain average of this value over a certain number of days preceeding maturity. This procedure is used to prevent an artificial rise of the stock price precisely on the expiry date, a rise that could be triggered by an operator having an important long position on the corresponding option. The contract is thus constructed on a fictitious asset, the price of which being defined as:

\[
\tilde{x} = \sum_{k=0}^{N} w_k x_i,
\] (5.31)

where the \( \{w_i\} \)’s are some weights, normalised such that \( \sum_{k=0}^{N} w_k = 1 \), which define the averaging procedure. The simplest case corresponds to:

\[
w_N = w_{N-1} = \ldots = w_{N-K+1} = \frac{1}{K}; \quad w_k = 0 \quad (k < N - K + 1),
\] (5.32)

where the average is taken over the last \( K \) days of the option life. One could however consider more complicated situations, for example an exponential average \( (w_k \propto s^{N-k}) \). The wealth balance then contains the modified pay-off: \( \max(\tilde{x} - x_s, 0) \), or more generally \( \mathcal{Y}(\tilde{x}) \). The first problem therefore concerns the statistics of \( \tilde{x} \). As we shall see, this problem is very similar to the case encountered in Chapter 4 where the volatility is time dependent. Indeed, one has:

\[
\sum_{k=0}^{N} w_k x_i = \sum_{k=0}^{N} w_i \left( \sum_{l=0}^{k-1} \delta x_l + x_0 \right) = x_0 + \sum_{k=0}^{N-1} \gamma_k \delta x_k,
\] (5.33)

where

\[
\gamma_k \equiv \sum_{i=k+1}^{N} w_i.
\] (5.34)

Said differently, everything goes as if the price did not vary by an amount \( \delta x_k \), but by an amount \( \delta y_k = \gamma_k \delta x_k \), distributed as:

\[
\frac{1}{\gamma_k} P_1 \left( \frac{\delta y_k}{\gamma_k} \right).
\] (5.35)

In the case of Gaussian fluctuations of variance \( \Delta \tau \), one thus finds:

\[
P(\tilde{x}, N|x_0, 0) = \frac{1}{\sqrt{2\pi D N \tau}} \exp \left[ -\frac{(\tilde{x} - x_0)^2}{2 DN \tau} \right],
\] (5.36)

where

\[
\tilde{D} = \frac{D}{N} \sum_{k=0}^{N-1} \gamma_k^2.
\] (5.37)
More generally, $P(\tilde{x}, N|x_0, 0)$ is the Fourier transform of
\[ \prod_{k=0}^{N-1} \hat{P}_1(\gamma_k z). \] (5.38)

This information is sufficient to fix the option price (in the limit where the average return is very small) through:
\[ C_{\text{asi}}[x_0, x_s, N] = \int_{x_s}^{\infty} \hat{d}(\tilde{x} - x_s) P(\tilde{x}, N|x_0, 0). \] (5.39)

In order to fix the optimal strategy, one must however calculate the following quantity:
\[ P(\tilde{x}, N|x, k)\langle \delta x_k \rangle|(x, k)\rightarrow(\tilde{x}, N), \] (5.40)
conditioned to a certain terminal value for $\tilde{x}$ (cf. (4.74)). The general calculation is given in Appendix C. For a small kurtosis, the optimal strategy reads:
\[ \phi_k |x, N_s| = \frac{\partial C_{\text{asi}}[x, x_s, N - k]}{\partial x} + \frac{\kappa_1 D \tau}{6} \gamma_k^2 \frac{\partial C_{\text{asi}}[x, x_s, N - k]}{\partial x^3}. \] (5.41)

Note that if the instant of time ‘$k$’ is outside the averaging period, one has $\gamma_k = 1$ (since $\sum_{i>k} w_i = 1$), and the formula (4.80) is recovered. If one the contrary $k$ gets closer to maturity, $\gamma_k$ diminishes as does the correction term.

5.2.4 ‘American’ options

We have up to now focussed our attention on ‘European’-type options, which can only be exercised on the day of expiry. In reality, most traded options on organised markets can be exercised at any time between the emission date and the expiry date: by definition, these are called ‘American’ options. It is obvious that the price of American options must greater or equal to the price of a European option with the same maturity and strike price, since the contract is a priori more favourable to the buyer. The pricing problem is therefore more difficult, since the writer of the option must first determine the optimal strategy that the buyer can follow in order to fix a fair price. Now, the optimal strategy for the buyer of a call option is to keep it until the expiry date, thereby converting de facto the option into a European option. Intuitively, this is due to the fact that the average $\langle \text{max}(x_N - x_s, 0) \rangle$ grows with $N$, hence the average pay-off is higher if one waits longer. The argument can be more convincing as follows. Let us define a ‘two-shot’ option, of strike $x_s$, which can only be exercised at times $N_1$ and $N_2 > N_1$ only. At time $N_1$, the buyer of the option may choose to exercise a fraction $f(x_1)$ of the option, which in principle depends on the current price of the underlying $x_1$. The remaining part of the option can then be exercised at time $N_2$. What is the average profit $\langle g \rangle$ of the buyer at time $N_2$?

Considering the two possible cases, one obtains:
\[ \langle g \rangle = \int_{x_s}^{\infty} dx_2 (x_2-x_s) \int dx_1 P(x_2, N_2|x_1, N_1) [1 - f(x_1)] P(x_1, N_1|x_0, 0) \]
\[ + \int_{x_s}^{\infty} dx_1 f(x_1)(x_1-x_s) e^{r \tau(N_2-N_1)} P(x_1, N_1|x_0, 0), \] (5.42)
which can be rewritten as:
\[ \langle g \rangle = C|x_0, x_s, N_2| e^{r \tau N_2} + \int_{x_s}^{\infty} dx_1 f(x_1) \times \]
\[ \times P(x_1, N_1|x_0, 0) (x_1-x_s - C|x_1, x_s, N_2-N_1|) e^{r \tau(N_2-N_1)}. \] (5.43)

The last expression means that if the buyer exercises a fraction $f(x_1)$ of his option, he pockets immediately the difference $x_1 - x_s$, but loses de facto his option, which is worth $C|x_1, x_s, N_2-N_1|$. The optimal strategy, such that $\langle g \rangle$ is maximum, therefore consists in choosing $f(x_1)$ equal to 0 or 1, according to the sign of $x_1-x_s - C|x_1, x_s, N_2-N_1|$. Now, this difference is always negative, whatever $x_1$ and $N_2 - N_1$. This is due to the put-call parity relation (cf. 5.26):
\[ C^\dagger|x_1, x_s, N_2-N_1| = C|x_1, x_s, N_2-N_1| - (x_1-x_s) - x_s (1 - e^{-r \tau(N_2-N_1)}). \] (5.44)

Since $C^\dagger \geq 0$, $C|x_1, x_s, N_2-N_1| - (x_1-x_s)$ is also greater or equal to zero.

The optimal value of $f(x_1)$ is thus zero; said differently the buyer should wait until maturity to exercise his option to maximise his average profit. This argument can be generalised to the case where the option can be exercised at any instant $N_1, N_2, ..., N_n$ with $n$ arbitrary.

Note however that choosing a non zero $f$ increases the total probability of exercising the option, but reduces the average profit! More precisely, the total probability to reach $x_s$ before maturity is twice the probability to exercise the option at expiry (if the distribution of $\delta x$ is even). OTC American options are therefore favourable to the writer of the option, since some buyers might be tempted to exercise before expiry.

---

5 Options: some more specific problems

It is interesting to generalise the problem and consider the case where the two strike prices \( x_{s1} \) and \( x_{s2} \) are different at times \( N_1 \) and \( N_2 \), in particular in the case where \( x_{s1} < x_{s2} \). The average profit (5.43) is then equal to (for \( r = 0 \)):

\[
\langle \mathcal{G} \rangle = C[x_0, x_{s2}, N_2] + \int_{x_s}^{\infty} \int_{x_s}^{\infty} dx_1 f(x_1) P(x_1, N_1|x_0, 0) \times (x_1 - x_s - C[x_1, x_{s2}, N_2 - N_1]).
\] (5.45)

The equation

\[
x^* - x_s - C[x^*, x_{s2}, N_2 - N_1] = 0
\] (5.46)

then has a non trivial solution, leading to \( f(x_1) = 1 \) for \( x_1 > x^* \). The average profit of the buyer therefore increases, in this case, upon an early exercise of the option.

American puts

Naively, the case of the American puts looks rather similar to that of the calls, and these should therefore also be equivalent to European puts. This is not the case for the following reason. Using the same argument as above, one finds that the average profit associated to a 'two-shot' put option with exercise dates \( N_1, N_2 \) is given by:

\[
\langle \mathcal{G} \rangle = C[x_0, x_{s}, N_2] e^{-r N_2} + \int_{x_s}^{x_0} dx_1 f(x_1) P(x_1, N_1|x_0, 0) \times (x_s - x_1 - C[x_1, x_{s}, N_2 - N_1]) e^{-r (N_2 - N_1)}. \] (5.47)

Now, the difference \( (x_s - x_1 - C[x_1, x_{s}, N_2 - N_1]) \) can be transformed, using the put-call parity, as:

\[
x_s [1 - e^{-r (N_2 - N_1)}] - C[x_1, x_{s}, N_2 - N_1].
\] (5.48)

This quantity may become positive if \( C[x_1, x_{s}, N_2 - N_1] \) is very small, which corresponds to \( x_s > x_1 \) (Puts deep in the money). The smaller the value of \( r \), the larger should be the difference between \( x_1 \) and \( x_s \), and the smaller the probability for this to happen. If \( r = 0 \), the problem of American puts is identical to that of the calls.

In the case where the quantity (5.48) becomes positive, an 'excess' average profit \( \delta \mathcal{G} \) is generated, and represents the extra premium to be added to the price of the European put to account for the possibility of an early exercise. Let us finally note that the price of the American put \( C_{am} \) is necessarily always larger or equal to \( x_s - x_1 \) (since this would be

the immediate profit), and that the price of the 'two-shot' put is a lower bound to \( C_{am} \).

The perturbative calculation of \( \delta \mathcal{G} \) (and thus of the 'two-shot' option) in the limit of small interest rates is not very difficult. As a function of \( N_1, \delta \mathcal{G} \) reaches a maximum between \( \frac{N_2}{2} \) and \( N_2 \). For an at-the-money put such that \( N_2 = 100, r = 5\% \) annual, \( \sigma = 1\% \) per day and \( x_0 = x_1 = 100 \), the maximum is reached for \( N_1 \approx 80 \) and the corresponding \( \delta \mathcal{G} \approx 0.15 \). This must be compared to the price of the European put, which is \( C^1 \approx 4 \). The possibility of an early exercise therefore leads to a 5% increase of the price of the option.

More generally, when the increments are independent and of average zero, one can obtain a numerical value for the price of an American put \( C_{am} \) by iterating backwards the following equation:

\[
C_{am}[x, x_{s}, N + 1] = \max \left( x_s - x, e^{-r T} \int d\delta x P(\delta x; C_{am}[x + \delta x, x_s, N]) \right).
\] (5.49)

This equation means that the put is worth the average value of tomorrow's price if it is not exercised today \( (C_{am} > x_s - x) \), or \( x_s - x \) if it is immediately exercised. Using this procedure, we have calculated the price of a European, American and 'two-shot' option of maturity 100 days, (Fig. 5.1). For the 'two-shot' option, the optimal \( N_1 \) as a function of the strike is shown in the inset.

5.2 Other types of options: ‘Puts’ and ‘exotic options’

Let us now turn to another family of options, called 'barrier' options, which are such that if the price of the underlying \( x_k \) reaches a certain 'barrier' value \( x_b \) during the lifetime of the option, the option is lost. (Conversely, there are options that are only activated if the value \( x_k \) is reached). This clause leads to cheaper options, which can be more attractive to the investor. Also, if \( x_b > x_s \), the writer of the option limits its possible losses to \( x_b - x_s \). What is the probability \( P_b(x, N|x_0, 0) \) for the final value of the underlying to be at \( x \), conditioned to the fact that the price has not reached the barrier value \( x_b \)?

In some cases, it is possible to give an exact answer to this question, using the so-called method of images. Let us suppose that for each time step, the price \( x \) can only change by an discrete amount, \( \pm 1 \) tick. The method of images is explained graphically in Fig. 5.2: one can notice that all the trajectories going through \( x_b \) between \( k = 0 \) and \( k = N \) has a 'mirror' trajectory, with a statistical weight precisely equal (for \( m = 0 \)) to the one of the trajectory one wishes to exclude. It is clear that the conditional probability we are looking for is obtained by subtracting the weight of these image trajectories:

\[
P_b(x, N|x_0, 0) = P(x, N|x_0, 0) - P(x, N|2x_b - x_0, 0).
\] (5.50)
Figure 5.1: Price of a European, American and ‘two-shot’ option as a function of the strike, for a 100 days maturity and a daily volatility of 1%. The top curve is the American price, while the bottom curve is the European price. In the inset is shown the optimal exercise time $N_1$ as a function of the strike for the ‘two-shot’ option.

Figure 5.2: Illustration of the method of images. A trajectory, starting from the point $x_0 = -5$ and reaching the point $x_{20} = -1$ can either touch or avoid the 'barrier' located at $x_b = 0$. For each trajectory touching the barrier, as the one shown in the figure (squares), there exists one single trajectory (circles) starting from the point $x_0 = +5$ and reaching the same final point - only the last section of the trajectory (after the last crossing point) is common to both trajectories. In the absence of bias, these two trajectories have exactly the same statistical weight. The probability of reaching the final point without crossing $x_b = 0$ can thus be obtained by subtracting the weight of the image trajectories. Note that the whole argument is wrong if jump sizes are not constant (for example $\delta x = \pm 1$ or $\pm 2$).

In the general case where the variations of $x$ are not limited to 0, ± 1, the previous argument fails, as one can easily be convinced by considering the case where $\delta x$ takes the values ± 1 and ± 2. However, if the possible variations of the price during the time $\tau$ are small, the error coming from the uncertainty about the exact crossing time is small, and leads to an error on the price $C_b$ of the barrier option on the order of $\langle |\delta x| \rangle$ times the total probability of ever touching the barrier. Discarding this correction, the price of barrier options read:

$$C_b[x_0, x_s, N] = \int_{x_s}^{\infty} dx (x - x_s) [P(x, N|x_0, 0) - P(x, N|2x_b - x_0, 0)]$$

$$\equiv C[x_0, x_s, N] - C[2x_b - x_0, x_s, N] \quad (5.51)$$
5 Options: some more specific problems

\[(x_b < x_0), \text{ or} \]
\[c_b[x_0, x, N] = \int_{x_0}^{x} dx (x - x_0) \left[ P(x, N|x_0, 0) - P(x, N|x_0, 0) \right], \tag{5.52} \]
\[(x_b > x_0); \text{ the option is obviously worthless whenever } x_0 < x < x_b. \]

One can also find ‘double barrier’ options, such that the price is constrained to remain within a certain channel \(x^- \leq x \leq x^+\), or else the option vanishes. One can generalise the method of images to this case. The images are now successive reflections of the starting point \(x_0\) in the two parallel ‘mirrors’ \(x^-\), \(x^+\).

Other types of options

One can find many other types of options, which we shall not discuss further. Some options, for example, are calculated on the maximum value of the price of the underlying reached during a certain period. It is clear that in this case, a Gaussian or log-normal model is particularly inadequate, since the price of the option is governed by extreme events. Only an adequate treatment of the tails of the distribution can allow to price correctly this type of options.

5.3 The ‘Greeks’ and risk control

The ‘Greeks’, which is the traditional name given by professionals to the derivative of the price of an option with respect to the price of the underlying, the volatility, etc., are often used for local risk control purposes. Indeed, if one assumes that the underlying asset does not vary too much between two instant of times \(t\) and \(t + \tau\), one may expand the variation of the option price in series:

\[\delta \mathcal{C} = \Delta \delta x + \frac{1}{2} \Gamma (\delta x)^2 + \nu \delta \sigma - \Theta \tau, \tag{5.53}\]

where \(\delta x\) is the change of price of the underlying. If the option is hedged by simultaneously selling a proportion \(\phi\) of the underlying asset, one finds that the change of the portfolio value is, to this order:

\[\delta \mathcal{W} = (\Delta - \phi) \delta x + \frac{1}{2} \Gamma (\delta x)^2 + \nu \delta \sigma - \Theta \tau. \tag{5.54}\]

Note that the Black-Scholes (or rather, Bachelier) equation is recovered by setting \(\phi^* = \Delta, \delta \sigma \equiv 0\), and by recalling that for a continuous time Gaussian process, \((\delta x)^2 \equiv D \tau\) (see Section 4.5.2). In this case, the portfolio does not change with time \((\delta W = 0)\), provided that \(\Theta = DI/2\), which is precisely Eq. (4.51) in the limit \(\tau \rightarrow 0\).

In reality, due to the non Gaussian nature of \(\delta x\), the large risk corresponds to cases where \(\Gamma (\delta x)^2 \gg \Theta \tau\). Assuming that one chooses to follow the \(\Delta\)-hedge procedure (which is in general suboptimal, see 4.4.3 above), one finds that the fluctuations of the price of the underlying leads to an increase in the value of the portfolio of the buyer of the option (since \(\Gamma > 0\)). Losses can only occur if the implied volatility of the underlying increases. If \(\delta x\) and \(\delta \sigma\) are uncorrelated (which is in general not true), one finds that the ‘instantaneous’ variance of the portfolio is given by:

\[(\delta W)^2 \approx \frac{3 + \kappa_1}{4} \Gamma^2 (\delta x)^2 + \nu^2 (\delta \sigma)^2, \tag{5.55}\]

where \(\kappa_1\) is the kurtosis of \(\delta x\). For an at the money option of maturity \(T\), one has (in order of magnitude):

\[\Gamma \sim \left. \frac{1}{\sigma x_0 \sqrt{T}} \quad \nu \sim x_0 \sqrt{T} \right. \tag{5.56}\]

Typical values are, on the scale of \(\tau = \) one day, \(\kappa_1 = 3\) and \(\delta \sigma \sim \sigma\).

The \(\Gamma\) contribution to risk is therefore on the order of \(\sigma x_0 \sqrt{T}\), and by recalling that for a continuous time Gaussian process, \((\delta x)^2 \equiv D \tau\) (see Section 4.5.2). In this case, the portfolio does not change with time \((\delta W = 0)\), provided that \(\Theta = DI/2\), which is precisely Eq. (4.51) in the limit \(\tau \rightarrow 0\).

In reality, due to the non Gaussian nature of \(\delta x\), the large risk corresponds to cases where \(\Gamma (\delta x)^2 \gg \Theta \tau\). Assuming that one chooses to follow the \(\Delta\)-hedge procedure (which is in general suboptimal, see 4.4.3 above), one finds that the fluctuations of the price of the underlying leads to an increase in the value of the portfolio of the buyer of the option (since \(\Gamma > 0\)). Losses can only occur if the implied volatility of the underlying increases. If \(\delta x\) and \(\delta \sigma\) are uncorrelated (which is in general not true), one finds that the ‘instantaneous’ variance of the portfolio is given by:

\[(\delta W)^2 \approx \frac{3 + \kappa_1}{4} \Gamma^2 (\delta x)^2 + \nu^2 (\delta \sigma)^2, \tag{5.55}\]

where \(\kappa_1\) is the kurtosis of \(\delta x\). For an at the money option of maturity \(T\), one has (in order of magnitude):

\[\Gamma \sim \left. \frac{1}{\sigma x_0 \sqrt{T}} \quad \nu \sim x_0 \sqrt{T} \right. \tag{5.56}\]

Typical values are, on the scale of \(\tau = \) one day, \(\kappa_1 = 3\) and \(\delta \sigma \sim \sigma\).

The \(\Gamma\) contribution to risk is therefore on the order of \(\sigma x_0 \sqrt{T}\), and by recalling that for a continuous time Gaussian process, \((\delta x)^2 \equiv D \tau\) (see Section 4.5.2). In this case, the portfolio does not change with time \((\delta W = 0)\), provided that \(\Theta = DI/2\), which is precisely Eq. (4.51) in the limit \(\tau \rightarrow 0\).
of the portfolio to these ‘explicative variables’ can be measured as the
derivatives of the value of the portfolio with respect to the $e_a$. We shall
therefore introduce the $\Delta$’s and $\Gamma$’s as:

$$
\Delta_a = \frac{\partial f}{\partial e_a}, \quad \Gamma_{a,b} = \frac{\partial^2 f}{\partial e_a \partial e_b}.
$$

(5.57)

We are interested in the probability for a large fluctuation $\delta f^*$ of
the portfolio. We will surmise that this is due to a particularly large
fluctuation of one explicative factor – say $a = 1$ – that we will call the
dominant factor. This is not always true, and depends on the statistics
of the fluctuations of the $e_a$. A condition for this assumption to be
true will be discussed below, and requires in particular that the tail
of the dominant factor should not decrease faster than an exponential.
Fortunately, this is a good assumption in financial markets.

The aim is to compute the Value-at-Risk of a certain portfolio, i.e.,
the value $\delta f^*$ such that the probability that the variation of $f$ exceeds $\delta f^*$
is equal to a certain probability $p$: $P_\text{>}(\delta f^*) = p$. Our assumption about
the existence of a dominant factor means that these events correspond
to a market configuration where the fluctuation $\delta e_1$ is large, while all
other factors are relatively small. Therefore, the large variations of the
portfolio can be approximated as:

$$
\delta f(e_1, e_2, ..., e_M) = \delta f(e_1) + \sum_{a=2}^{M} \Delta_a e_a + \frac{1}{2} \sum_{a,b=2}^{M} \Gamma_{a,b} e_a e_b,
$$

(5.58)

where $\delta f(e_1)$ is a shorthand notation for $\delta f(e_1, 0, ..., 0)$. Now, we use the fact
that:

$$
P_\text{>}(\delta f^*) = \prod_{a=1}^{M} P(e_1, e_2, ..., e_M) \Theta(\delta f(e_1, e_2, ..., e_M) - \delta f^*),
$$

(5.59)

where $\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$. Expanding the $\Theta$ function to
second order leads to:

$$
\Theta(\delta f(e_1) - \delta f^*) + \left[ \sum_{a=2}^{M} \Delta_a e_a + \frac{1}{2} \sum_{a,b=2}^{M} \Gamma_{a,b} e_a e_b \right] (\delta f(e_1) - \delta f^*) + \frac{1}{2} \sum_{a,b=2}^{M} \Delta_a \Delta_b e_a e_b (\delta f(e_1) - \delta f^*),
$$

(5.60)

where $\delta^*$ is the derivative of the $\delta$-function with respect to $\delta f$. In order
to proceed with the integration over the variables $e_a$ in Eq. (5.59), one
should furthermore note the following identity:

$$
\delta(\delta f(e_1) - \delta f^*) = \frac{1}{\Delta_1} \delta(e_1 - e_1^*),
$$

(5.61)

where $e_1^*$ is such that $\delta f(e_1^*) = \delta f^*$, and $\Delta_1$ is computed for $e_1 = e_1^*$,
$a_0 = 1$. Inserting the above expansion of the $\Theta$ function into (5.59)
and performing the integration over the $e_a$ then leads to:

$$
P_\text{>}(\delta f^*) = P_\text{>}(e_1^*) + \sum_{a=2}^{M} \Delta_a \sigma_a^2 P(e_1^*) - \sum_{a=2}^{M} \Delta_a \sigma_a^2 \left( P'(e_1^*) + \frac{\Gamma_{1,1}^a}{\Delta_1} P(e_1^*) \right),$$

(5.62)

where $P(e_1)$ is the probability distribution of the first factor, defined as:

$$
P(e_1) = \prod_{a=1}^{M} d e_a P(e_1, e_2, ..., e_M).
$$

(5.63)

In order to find the Value-at-Risk $\delta f^*$, one should thus solve (5.62) for
$e_1^*$ with $P_\text{>}(\delta f^*) = p$, and then compute $\delta f(e_1^*, 0, ..., 0)$. Note that the
equation is not trivial since the Greeks must be estimated at the solution
point $e_1^*$.

Let us discuss the general result (5.62) in the simple case of a linear
portfolio of assets, such that no convexity is present: the $\Delta$’s are con-
stant and the $\Gamma_{a,b}$’s are all zero. The equation then takes the following
simpler form:

$$
P_\text{>}(e_1^*) - \sum_{a=2}^{M} \Delta_a \sigma_a^2 P'(e_1^*) = p.
$$

(5.64)

Naively, one could have thought that in the dominant factor approxi-
mation, the value of $e_1^*$ would be the Value-at-Risk value of $e_1$ for
the probability $p$, defined as:

$$
P_\text{>}(e_{1,VaR}) = p.
$$

(5.65)

However, the above equation shows that there is a correction term pro-
portional to $P'(e_1^*)$. Since the latter quantity is negative, one sees that
$e_1^*$ is actually larger than $e_{1,VaR}$, and therefore $\delta f^* > \delta f(e_{1,VaR})$. This
reflects the effect of all other factors, which tend to increase the Value-
at-Risk of the portfolio.

The result obtained above relies on a second order expansion; when are higher order corrections negligible? It is easy to see that higher order
terms involve higher order derivatives of $P(e_1)$. A condition for these
terms to be negligible in the limit $p \to 0$, or $e_1^* \to \infty$, is that the suc-
cessive derivatives of $P(e_1)$ become smaller and smaller. This is true

provided that \( P(e_1) \) decays more slowly than exponentially, for example as a power-law. On the contrary, when \( P(e_1) \) decays faster than exponentially (for example in the Gaussian case), then the expansion proposed above completely loses its meaning, since higher and higher corrections become dominant when \( p \to 0 \). This is expected: in a Gaussian world, a large event results from the accidental superposition of many small events, while in a power-law world, large events are associated to one single large fluctuation which dominates over all the others. The case where \( P(e_1) \) decays as an exponential is interesting, since it is often a good approximation for the tail of the fluctuations of financial assets. Taking \( P(e_1) \simeq \alpha_1 \exp -\alpha_1 e_1 \), one finds that \( e^*_1 \) is the solution of:

\[
e^{-\alpha_1 e^*_1} \left[ 1 - \sum_{a=2}^{M} \frac{\Delta_a^2 \sigma_a^2}{2 \Delta_1^2} \right] = p.
\]

Since one has \( \sigma_a^2 \propto \alpha_1^{-2} \), the correction term is small provided that the variance of the portfolio generated by the dominant factor is much larger than the sum of the variance of all other factors.

Coming back to equation (5.62), one expects that if the dominant factor is correctly identified, and if the distribution is such that the above expansion makes sense, an approximate solution is given by \( e^*_1 = e_{1,Var} + \epsilon \), with:

\[
\epsilon \simeq \sum_{a=2}^{M} \frac{\Gamma_{a,Var}^2}{2 \Delta_1^2} = \sum_{a=2}^{M} \frac{\Delta_a^2 \sigma_a^2}{2 \Delta_1^2} \left( \frac{P(e_{1,Var})}{P(e_1)} + \frac{\Gamma_{1,1}}{\Delta_1} \right),
\]

where now all the Greeks at estimated at \( e_{1,Var} \).

### 5.5 Risk diversification (*)

We have put the emphasis on the fact that for real world options, the Black-Scholes divine surprise – i.e. the fact that the risk is zero – does not occur, and a non zero residual risk remains. One can ask whether this residual risk can be reduced further by including other assets in the hedging portfolio. Buying stocks other than the underlying to hedge an option can be called an ‘exogenous’ hedge. A related question concerns the hedging of a ‘basket’ option, the pay-off of which is calculated on a linear superposition of different assets. A rather common example is that of ‘spread’ options, which depend on the difference of the price between two assets (for example the difference between the Nikkei and the S&P 500, or between the French and German rates, etc.). An interesting conclusion is that in the Gaussian case, an exogenous hedge increases the risk. An exogenous hedge is only useful in the presence of non-Gaussian effects. Another possibility is to hedge some options using different options; in other words, one can ask how to optimise a global ‘book’ of options such that the global risk is minimum.

**‘Portfolio’ options and ‘exogenous’ hedging**

Let us suppose that one can buy \( M \) assets \( X^i, i = 1, \ldots, M \), the price of which being \( x_k^i \) at time \( k \). As in Chapter 3, we shall suppose that these assets can be decomposed over a basis of independent factors \( E^a \): 

\[
x_k^i = \sum_{a=1}^{M} O_{a,i} e_k^a.
\]

The \( E^a \) are independent, of unit variance, and of distribution function \( P_a \). The correlation matrix of the fluctuations, \( \langle \delta x^i \delta x^j \rangle \) is equal to \( \sum_a O_{a,i} O_{a,j} = \{O O\}^a_{ij} \).

One considers a general option constructed on a linear combination of all assets, such that the pay-off depends on the value of \( \bar{x} = \sum_i f_i x^i \) and is equal to \( Y(\bar{x}) = \max(\bar{x} - x_a, 0) \). The usual case of an option on the asset \( X^1 \) thus corresponds to \( f_1 = \delta_{1,1} \). A spread option on the difference \( X^1 - X^2 \) corresponds to \( f_1 = \delta_{1,1} - \delta_{1,2} \), etc. The hedging portfolio at time \( k \) is made of all the different assets \( X^i \), with weight \( \phi_k^i \). The question is to determine the optimal composition of the portfolio, \( \phi_k^i \).

Following the general method explained in Section 4.3.3, one finds that the part of the risk which depends on the strategy contains both a linear and a quadratic term in the \( \phi \)’s. Using the fact that the \( E^a \) are independent random variables, one can compute the functional derivative of the risk with respect to all the \( \phi_k^i(x) \). Setting this functional derivative to zero leads to: \(^3\)

\[
\sum_j \{O O\}^i_{ij} \phi_k^i = \int dz Y(z) \left[ \prod_k P_h(z \sum_j f_j O_{ja}) \times \times \sum_a O_{a,i} \frac{\partial}{\partial z} \log P_a \left( z \sum_j f_j O_{ja} \right) \right].
\]

\(^3\)In the following, \( i \) denotes the unit imaginary number, except when it appears as a subscript, in which case it is an asset label.
Using the cumulant expansion of \( P_a \) (assumed to be even), one finds that:

\[
\frac{\partial}{\partial z} \log P_a(z \sum_j f_j O_{ja}) = iz \left( \sum_j f_j O_{ja} \right)^2 - \frac{i^3}{3} \kappa_3 \left( \sum_j f_j O_{ja} \right)^4 + ...
\]

The first term combines with

\[
\sum_a \frac{O_{ja}}{\sum_j f_j O_{ja}}
\]

(5.71)

to yield:

\[
i z \sum_a O_{ja} f_j \equiv iz[O O^\dagger f_{|a|},
\]

(5.72)

which finally leads to the following simple result:

\[
\phi^*_i = f_i \mathcal{P}[\{x_i^k\}, x_s, N - k]
\]

(5.74)

where \( \mathcal{P}[\{x_i^k\}, x_s, N - k] \) is the probability for the option to be exercised, calculated at time \( k \). In other words, in the Gaussian case \( \kappa_3 \equiv 0 \) the optimal portfolio is such that the proportion of asset \( i \) precisely reflects the weight of \( i \) in the basket on which the option is constructed. In particular, in the case of an option on a single asset, the hedging strategy is not improved if one includes other assets, even if these assets are correlated with the former.

However, this conclusion is only correct in the case of Gaussian fluctuations and does not hold if the kurtosis is non zero.\(^4\) In this case, an extra term appears, given by:

\[
\delta \phi^*_i = \frac{1}{6} \left( \sum_j \kappa_3 [O O^\dagger]_{ij} \sum_a f_a O_{ja} \right) \frac{\partial \mathcal{P}(\tilde{x}, x_s, N - k)}{\partial x_s}.
\]

(5.75)

This correction is not, in general, proportional to \( f_i \), and therefore suggests that, in some cases, an exogenous hedge can be useful. However, one should note that this correction is small for at-the-money options \( \tilde{x} = x_s \), since \( \frac{\partial \mathcal{P}(\tilde{x}, x_s, N - k)}{\partial x_s} = 0 \).

Option portfolio

Since the risk associated to a single option is in general non zero, the global risk of a portfolio of options (‘book’) is also non zero. Suppose that the book contains \( p_i \) calls of ‘type’ \( i \) (\( i \) therefore contains the information of the strike \( x_{si} \) and maturity \( T_i \)). The first problem to solve is that of the hedging strategy. In the absence of volatility risk, it is not difficult to show that the optimal hedge for the book is the linear superposition of the optimal strategies for each individual options:

\[
\phi^*_i(x, t) = \sum_i p_i \phi^*_i(x, t).
\]

(5.76)

The residual risk is then given by:

\[
\mathcal{R}^{\ast 2} = \sum_{i,j} p_i p_j C_{ij},
\]

(5.77)

where the ‘correlation matrix’ \( C \) is equal to:

\[
C_{ij} = \langle \max(x(T_i) - x_{si}, 0) \max(x(T_j) - x_{sj}, 0) \rangle - C_i C_j - D \tau \sum_{k=0}^{N-1} \langle \phi^*_i(x, k\tau) \phi^*_j(x, k\tau) \rangle
\]

(5.78)

where \( C_i \) is the price of the option \( i \). If the constraint on the \( p_i \)’s is of the form \( \sum_i p_i = 1 \), the optimum portfolio is given by:

\[
p^*_i = \frac{\sum_{i,j} C^{-1}_{ij} C_{ij}}{\sum_{i,j} C^{-1}_{ij}}
\]

(5.79)

(remember that by assumption the mean return associated to an option is zero).

Let us finally note that we have not considered, in the above calculation, the risk associated to volatility fluctuations, which is rather important in practice. It is a common practice to try to hedge this volatility risk using other types of options (for example, an exotic option can be hedged using a ‘plain vanilla’ option). A generalisation of the Black-Scholes argument (assuming that option prices themselves follow a Gaussian process, which is far from being the case) suggests that the optimal strategy is to hold a fraction

\[
\Delta = \frac{\frac{\partial^2 \mathcal{P}}{\partial x_s^2}}{\frac{\partial \mathcal{P}}{\partial x_s}}
\]

(5.80)

of options of type 2 to hedge the volatility risk associated to an option of type 1.

5.6 References

- More on Options, Exotic Options:

INDEX OF SYMBOLS

A tail amplitude of power laws: \( P(x) \sim \mu A^x / x^{1+\mu} \).............. 8

\( A_i \) tail amplitude of asset \( i \) ................................. 125

\( A_p \) tail amplitude of portfolio \( p \).................................. 125

\( B \) amount invested in the risk-free asset .......................... 148

\( B(t, \theta) \) price, at time \( t \), of a bond that pays 1 at time \( t + \theta \)..... 80

\( c_n \) cumulant of order \( n \) of a distribution. ...................... 8

\( c_{n,1} \) cumulant of order \( n \) of an elementary distribution \( P_1(x) \). ......... 23

\( c_{n,N} \) cumulant of order \( n \) of a distribution at the scale \( N \),
\( P(x, N) \) ..................................................................... 23

\( C \) covariance matrix ................................................................ 45

\( C_{ij} \) element of the covariance matrix. .................................. 90

\( C_{ij}^{\alpha/2} \) ‘tail covariance’ matrix ......................................... 135

\( C \) price of a European call option........................................... 154

\( C^\dagger \) price of a European put option ................................. 211

\( C_G \) price of a European call in the Gaussian Bachelier theory. .... 159

\( C_{BS} \) price of a European call in the Black-Scholes theory. ......... 158

\( C_M \) market price of a European call ....................................... 171

\( C_\kappa \) price of a European call for a non-zero kurtosis \( \kappa \) ......... 163

\( C_m \) price of a European call for a non-zero excess return \( m \) .... 194

\( C_d \) price of a European call with dividends ............................. 209

\( C_{asi.} \) price of a Asian call option........................................ 212

\( C_{am.} \) price of a American call option ................................. 214

\( C_b \) price of a barrier call option .......................................... 217

\( C(\theta) \) yield curve spread correlation function .......................... 217

\( D_{\tau} \) variance of the fluctuations in a time step \( \tau \) in the
additive approximation: \( D = \sigma_1^2 x_0^2 \) .......................... 53

\( D_i \) \( D \) coefficient for asset \( i \) ........................................ 121

\( D_p \tau \) risk associated with portfolio \( p \) ............................... 121
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e_a)</td>
<td>explicative factor (or principal component)</td>
<td>131</td>
</tr>
<tr>
<td>(E_{abs})</td>
<td>mean absolute deviation</td>
<td>5</td>
</tr>
<tr>
<td>(f(t, \theta))</td>
<td>forward value at time (t) of the rate at time (t + \theta)</td>
<td>80</td>
</tr>
<tr>
<td>(\mathcal{F})</td>
<td>forward price</td>
<td>147</td>
</tr>
<tr>
<td>(g(t))</td>
<td>auto-correlation function of (\gamma_k^2)</td>
<td>72</td>
</tr>
<tr>
<td>(G_p)</td>
<td>probable gain</td>
<td>118</td>
</tr>
<tr>
<td>(H)</td>
<td>Hurst exponent</td>
<td>68</td>
</tr>
<tr>
<td>(H)</td>
<td>Hurst function</td>
<td>68</td>
</tr>
<tr>
<td>(I)</td>
<td>missing information</td>
<td>29</td>
</tr>
<tr>
<td>(k)</td>
<td>time index ((t = k\tau)).</td>
<td>50</td>
</tr>
<tr>
<td>(K_n)</td>
<td>modified Bessel function of the second kind of order (n)</td>
<td>16</td>
</tr>
<tr>
<td>(K_{ijkl})</td>
<td>generalised kurtosis</td>
<td>133</td>
</tr>
<tr>
<td>(L_\mu)</td>
<td>Lévy distribution of order (\mu).</td>
<td>12</td>
</tr>
<tr>
<td>(L_\mu(t))</td>
<td>truncated Lévy distribution of order (\mu).</td>
<td>15</td>
</tr>
<tr>
<td>(m)</td>
<td>average return by unit time.</td>
<td>103</td>
</tr>
<tr>
<td>(m(t, t'))</td>
<td>interest rate trend at time (t') as anticipated at time (t).</td>
<td>88</td>
</tr>
<tr>
<td>(m_1)</td>
<td>average return on a unit time scale (\tau): (m_1 = \mu \tau).</td>
<td>113</td>
</tr>
<tr>
<td>(m_i)</td>
<td>return of asset (i).</td>
<td>119</td>
</tr>
<tr>
<td>(m_n)</td>
<td>moment of order (n) of a distribution.</td>
<td>7</td>
</tr>
<tr>
<td>(m_p)</td>
<td>return of portfolio (p).</td>
<td>121</td>
</tr>
<tr>
<td>(M)</td>
<td>number of asset in a portfolio.</td>
<td>115</td>
</tr>
<tr>
<td>(M_{eff.})</td>
<td>effective number of asset in a portfolio.</td>
<td>123</td>
</tr>
<tr>
<td>(N)</td>
<td>number of elementary time step until maturity: (N = T/\tau).</td>
<td>50</td>
</tr>
<tr>
<td>(N^*)</td>
<td>number of elementary time step under which tail effects are important, after the CLT applies progressively.</td>
<td>31</td>
</tr>
<tr>
<td>(O)</td>
<td>coordinate change matrix.</td>
<td>131</td>
</tr>
<tr>
<td>(p_i)</td>
<td>weight of asset (i) in portfolio (p).</td>
<td>119</td>
</tr>
<tr>
<td>(p)</td>
<td>portfolio constructed with the weights ({p_i}).</td>
<td>121</td>
</tr>
<tr>
<td>(P_1(\delta x)) or (P_\delta(\delta x)), elementary return distribution on time scale (\tau).</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>(P_{10})</td>
<td>distribution of rescaled return (\delta x_k/\gamma_k).</td>
<td>72</td>
</tr>
<tr>
<td>(P(x, N))</td>
<td>distribution of the sum of (N) terms.</td>
<td>23</td>
</tr>
<tr>
<td>(P(z))</td>
<td>characteristic function of (P).</td>
<td>7</td>
</tr>
<tr>
<td>(P(x, t</td>
<td>x_0, t_0))</td>
<td>probability that the price of asset (X) be (x) (within (dx)) at time (t) knowing that, at a time (t_0), its price was (x_0).</td>
</tr>
<tr>
<td>(P_0(x, t</td>
<td>x_0, t_0))</td>
<td>probability without bias.</td>
</tr>
<tr>
<td>(P_m(x, t</td>
<td>x_0, t_0))</td>
<td>probability with return (m).</td>
</tr>
<tr>
<td>(P_E)</td>
<td>symmetric exponential distribution.</td>
<td>16</td>
</tr>
<tr>
<td>(P_G)</td>
<td>Gaussian distribution.</td>
<td>9</td>
</tr>
<tr>
<td>(P_H)</td>
<td>hyperbolic distribution.</td>
<td>16</td>
</tr>
<tr>
<td>(P_{LN})</td>
<td>log-normal distribution.</td>
<td>10</td>
</tr>
<tr>
<td>(P_S)</td>
<td>Student distribution.</td>
<td>17</td>
</tr>
<tr>
<td>(P)</td>
<td>probability of a given event (such as an option being exercised).</td>
<td>174</td>
</tr>
<tr>
<td>(P_{&lt;})</td>
<td>cumulative distribution: (P_{&lt;} \equiv P(X &lt; x)).</td>
<td>5</td>
</tr>
<tr>
<td>(P_{&gt;})</td>
<td>cumulative normal distribution, (P_{&gt;}(u) = \text{erf}(u/\sqrt{2})/2).</td>
<td>31</td>
</tr>
<tr>
<td>(Q)</td>
<td>ratio of the number of observation (days) to the number of asset.</td>
<td>45</td>
</tr>
<tr>
<td>(Q(x, t</td>
<td>x_0, t_0))</td>
<td>risk-neutral probability.</td>
</tr>
<tr>
<td>(Q_1(u))</td>
<td>polynomials related to deviations from a Gaussian.</td>
<td>31</td>
</tr>
<tr>
<td>(r)</td>
<td>interest rate by unit time: (r = \rho/\tau).</td>
<td>148</td>
</tr>
<tr>
<td>(r(t))</td>
<td>spot rate: (r(t) = f(t, \theta_{min})).</td>
<td>81</td>
</tr>
<tr>
<td>(\mathcal{R})</td>
<td>risk (rms of the global wealth balance).</td>
<td>170</td>
</tr>
<tr>
<td>(\mathcal{R}_{residual})</td>
<td>residual risk.</td>
<td>170</td>
</tr>
<tr>
<td>(s(t))</td>
<td>interest rate spread: (s(t) = f(t, \theta_{max}) - f(t, \theta_{min})).</td>
<td>81</td>
</tr>
<tr>
<td>(S(u))</td>
<td>Cramér function.</td>
<td>33</td>
</tr>
<tr>
<td>(S)</td>
<td>Sharpe ration.</td>
<td>103</td>
</tr>
<tr>
<td>(T)</td>
<td>time scale, e.g. an option maturity.</td>
<td>50</td>
</tr>
<tr>
<td>(T^*)</td>
<td>time scale for convergence towards a Gaussian.</td>
<td>65</td>
</tr>
<tr>
<td>(T_{\sigma})</td>
<td>crossover time between the additive and multiplicative regimes.</td>
<td>56</td>
</tr>
<tr>
<td>(U)</td>
<td>utility function.</td>
<td>115</td>
</tr>
</tbody>
</table>
\begin{itemize}
  \item \( \mathcal{V} \) \text{ 'Vega', derivative of the option price with respect to volatility.} \hspace{1cm} 158
  \item \( w_{1/2} \) \text{ full-width at half maximum.} \hspace{1cm} 6
  \item \( \Delta W \) \text{ global wealth balance, e.g. global wealth variation between emission and maturity.} \hspace{1cm} 145
  \item \( \Delta W_{S} \) \text{ wealth balance from trading the underlying.} \hspace{1cm} 189
  \item \( x \) \text{ price of an asset.} \hspace{1cm} 50
  \item \( x_{k} \) \text{ price at time } k. \hspace{1cm} 50
  \item \( \bar{x}_{k} = (1 + \rho)^{-k} x_{k} \) \hspace{1cm} 149
  \item \( x_{s} \) \text{ strike price of an option.} \hspace{1cm} 154
  \item \( x^{*} \) \text{ most probable value.} \hspace{1cm} 5
  \item \( x_{\text{med}} \) \text{ median.} \hspace{1cm} 5
  \item \( \delta x_{k} \) \text{ variation of } x \text{ between time } k \text{ and } k+1. \hspace{1cm} 50
  \item \( \delta x_{k}^{i} \) \text{ variation of the price of asset } i \text{ between time } k \text{ and } k+1. \hspace{1cm} 90
  \item \( y_{k} \) \text{ log}(x_{k}/x_{0}) - k \log(1 + \rho). \hspace{1cm} 157
  \item \( \mathcal{Y}(x) \) \text{ pay-off function, e.g. } \mathcal{Y}(x) = \max(x - x_{s}, 0). \hspace{1cm} 154
  \item \( z \) \text{ Fourier variable.} \hspace{1cm} 7
  \item \( Z \) \text{ normalisation.} \hspace{1cm} 123
  \item \( Z(u) \) \text{ persistence function.} \hspace{1cm} 88
  \item \( \alpha \) \text{ exponential decay parameter: } P(x) \sim \exp(-\alpha x). \hspace{1cm} 15
  \item \( \beta \) \text{ asymmetry parameter, \text{ or normalised covariance between an asset and the market portfolio.}} \hspace{1cm} 133
  \item \( \gamma_{k} \) \text{ scale factor of a distribution (potentially } k \text{ dependent).} \hspace{1cm} 72
  \item \( \Gamma \) \text{ derivative of } \Delta \text{ with respect to the underlying: } \partial \Delta / \partial x_{0}. \hspace{1cm} 158
  \item \( \delta_{ij} \) \text{ Kroeneker delta: } \delta_{ij} = 1 \text{ if } i = j, \text{ 0 otherwise.} \hspace{1cm} 41
  \item \( \delta(x) \) \text{ Dirac delta function.} \hspace{1cm} 140
  \item \( \Delta \) \text{ derivative of the option premium with respect to the underlying price, } \Delta = \partial C / \partial x_{0}, \text{ it is the optimal hedge } \phi^{*} \text{ in the Black-Scholes model.} \hspace{1cm} 177
  \item \( \Delta(\theta) \) \text{ RMS static deformation of the yield curve.} \hspace{1cm} 83
  \item \( \zeta, \zeta' \) \text{ Lagrange multiplier.} \hspace{1cm} 29
  \item \( \eta_{k} \) \text{ return between } k \text{ and } k+1: x_{k+1} - x_{k} = \eta_{k} x_{k}. \hspace{1cm} 102
  \item \( \theta \) \text{ maturity of } k \text{ and } k+1 \text{.} \hspace{1cm} 80
  \item \( \Theta \) \text{ derivative of a bond or a forward rate, always a time difference.} \hspace{1cm} 158
  \item \( \kappa \) \text{ kurtosis: } \kappa \equiv \lambda_{4}. \hspace{1cm} 8
  \item \( \kappa_{\text{eff.}} \) \text{ 'effective' kurtosis.} \hspace{1cm} 165
  \item \( \kappa_{\text{imp.}} \) \text{ 'implied' kurtosis.} \hspace{1cm} 165
  \item \( \kappa_{N} \) \text{ kurtosis at scale } N. \hspace{1cm} 65
  \item \( \lambda \) \text{ eigenvalue.} \hspace{1cm} 43
  \item \( \lambda_{n} \) \text{ normalised cumulants: } \lambda_{n} = c_{n}/\sigma^{n}. \hspace{1cm} 8
  \item \( \Delta \) \text{ loss level; } \Delta_{\text{VaR}} \text{ loss level (or Value at Risk) associated to a given probability } P_{\text{VaR}}. \hspace{1cm} 105
  \item \( \mu \) \text{ exponent of a power law, or a Lévy distribution.} \hspace{1cm} 8
  \item \( \pi_{ij} \) \text{ 'product' variable of fluctuations } \delta x_{i} \delta x_{j}. \hspace{1cm} 134
  \item \( \rho \) \text{ interest rate on a unit time interval } \tau. \hspace{1cm} 149
  \item \( \rho^{<n>} \) \text{ interest rate for a maturity } n \text{ rescaled to the elementary time unit: } \rho^{<1>} = \rho. \hspace{1cm} 151
  \item \( \rho^{<n>} \) \text{ interest rate anticipated at } n, \text{ at time } k. \hspace{1cm} 151
  \item \( \rho(\lambda) \) \text{ density of eigenvalues of a large matrix.} \hspace{1cm} 43
  \item \( \sigma \) \text{ volatility.} \hspace{1cm} 6
  \item \( \sigma_{1} \) \text{ volatility on a unit time step: } \sigma_{1} = \sigma / \sqrt{\tau}. \hspace{1cm} 53
  \item \( \Sigma \) \text{ 'implied' volatility.} \hspace{1cm} 161
  \item \( \tau \) \text{ elementary time step.} \hspace{1cm} 50
  \item \( \phi_{k}^{N} \) \text{ quantity of underlying in a portfolio at time } k, \text{ for an option with maturity } N. \hspace{1cm} 148
  \item \( \phi_{k}^{N,*} \) \text{ optimal hedge ratio.} \hspace{1cm} 178
  \item \( \phi_{k}^{*} \) \text{ hedge ratio used by the market.} \hspace{1cm} 171
  \item \( \psi_{k}^{N} \) \text{ hedge ratio corrected for interest rates: } \psi_{k}^{N} = (1 + \rho)^{N-k-1} \phi_{k}^{N}. \hspace{1cm} 149
  \item \( \xi \) \text{ Random variable of unit variance.} \hspace{1cm} 84
  \item \( \equiv \) \text{ equals by definition.} \hspace{1cm} 5
  \item \( \simeq \) \text{ is approximately equal to.} \hspace{1cm} 18
\end{itemize}
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>∞</td>
<td>is proportional to</td>
<td>25</td>
</tr>
<tr>
<td>~</td>
<td>is on the order of, or tends to asymptotically</td>
<td>8</td>
</tr>
<tr>
<td>erfc(x)</td>
<td>complementary error function</td>
<td>31</td>
</tr>
<tr>
<td>log(x)</td>
<td>natural logarithm</td>
<td>8</td>
</tr>
<tr>
<td>Γ(x)</td>
<td>gamma function: $\Gamma(n+1) = n!$</td>
<td>13</td>
</tr>
</tbody>
</table>

**INDEX**

Additive-multiplicative crossover, 53  
Arbitrage opportunity, 146  
absence of (AAO), 151  
ARCH, 95, 165  
Asset, 4  
Basis point, 146  
Bachelier formula, 158  
Bid-ask spread, 53, 146, 171  
Binomial model, 200  
Black and Scholes formula, 157  
Bond, 80, 148  
BUND, 53  
CAPM, 132  
Central limit theorem (CLT), 25  
Characteristic function, 7, 22  
Convolution, 22  
Correlations  
inter-asset, 84, 90, 225  
temporal, 39, 56, 72, 145  
Cramér function, 33  
Cumulants, 8, 23  
Delta, 177, 158, 220  
Distribution  
cumulative, 4, 64  
Fréchet, 20  
Gaussian, 9  
Gumbel, 19, 106  
hyperbolic, 16  
Lévy, 11  
log-normal, 10  
power-law, 8, 133  
Poisson, 15  
stable, 24  
Student, 17, 35, 51  
exponential, 17, 18, 127  
truncated Lévy (TLD), 15, 37, 60  
Diversification, 115, 123  
Dividends, 150, 209  
Drawdown, 113  
Effective number of asset, 123  
Efficiency, 144  
Efficient frontier, 122  
Eigenvalues, 42, 91, 131  
Explicative factors, 131, 134  
Extreme value statistics, 17, 106  
Fair price, 147, 153  
Feedback, 95  
Forward, 147  
rate curve (FRC), 80  
Futures, 53, 53, 147  
Gamma, 158, 220  
German mark (DEM), 53  
Greeks, 158, 220  
Heath-Jarrow-Morton model, 79  
Hedging, 153  
optimal, 170, 175, 192  
Heteroskedasticity, 51, 165  
Hull and White model, 86  
Hurst exponent, 68, 93  
Image method, 217  
Independent identically distributed (IID), 17, 39  
Information, 29  
Interest Rates, 79
<table>
<thead>
<tr>
<th>Index</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ito calculus, 194</td>
<td>Rank histogram, 21</td>
</tr>
<tr>
<td>Kurtosis, 8, 70, 161</td>
<td>Risk residual, 171, 180</td>
</tr>
<tr>
<td>Large deviations, 30</td>
<td>volatility, 184, 227</td>
</tr>
<tr>
<td>Markowitz, H., 132</td>
<td>zero, 150, 181, 197</td>
</tr>
<tr>
<td>Market crash, 3, 198</td>
<td>Risk-neutral probability, 191, 201</td>
</tr>
<tr>
<td>Maturity, 152</td>
<td>Root mean square (RMS), 5</td>
</tr>
<tr>
<td>Mean, 5</td>
<td>Saddle point method, 33, 41</td>
</tr>
<tr>
<td>Mean absolute deviation (MAD), 5</td>
<td>Scale invariance, 25, 96</td>
</tr>
<tr>
<td>Median, 5</td>
<td>Self organisation criticality, 95</td>
</tr>
<tr>
<td>Mimetism, 95</td>
<td>Self-similar, 24</td>
</tr>
<tr>
<td>Moneyness, 155</td>
<td>Semi-circle law, 45</td>
</tr>
<tr>
<td>Non-stationarity, 39, 70, 46, 165</td>
<td>Sharpe ratio, 103</td>
</tr>
<tr>
<td>Option, 152</td>
<td>Skewness, 8</td>
</tr>
<tr>
<td>American, 214</td>
<td>Spot rate, 80</td>
</tr>
<tr>
<td>Asian, 212</td>
<td>Spread, 81</td>
</tr>
<tr>
<td>at-the-money, 155</td>
<td>Stretched exponential, 66</td>
</tr>
<tr>
<td>barrier, 217</td>
<td>Strike price, 153</td>
</tr>
<tr>
<td>Bermudan, 215</td>
<td>S&amp;P 500 index, 3, 51</td>
</tr>
<tr>
<td>call, 152</td>
<td>Tail, 11, 38, 185</td>
</tr>
<tr>
<td>European, 152</td>
<td>covariance, 134</td>
</tr>
<tr>
<td>put, 211</td>
<td>amplitude, 12</td>
</tr>
<tr>
<td>Ornstein-Uhlenbeck process, 84</td>
<td>Theta, 158, 220</td>
</tr>
<tr>
<td>Over the counter (OTC), 171</td>
<td>Tick, 4</td>
</tr>
<tr>
<td>Percolation, 94</td>
<td>Transaction costs, 146, 209</td>
</tr>
<tr>
<td>Portfolio insurance, 198</td>
<td>Utility function, 115</td>
</tr>
<tr>
<td>of options, 226</td>
<td>Vega, 158, 220</td>
</tr>
<tr>
<td>optimal, 121, 129</td>
<td>Value at Risk (VaR), 104, 129, 186, 221</td>
</tr>
<tr>
<td>Power spectrum, 59</td>
<td>Vasicek model, 79, 84</td>
</tr>
<tr>
<td>Premium, 153</td>
<td>Volatility, 5, 53, 76, 101</td>
</tr>
<tr>
<td>Pricing kernel, 191, 201</td>
<td>hump, 88</td>
</tr>
<tr>
<td>Principal components, 131</td>
<td>implied, 161</td>
</tr>
<tr>
<td>Probable gain, 118</td>
<td>smile, 161</td>
</tr>
<tr>
<td>Quality ratio, 114, 181</td>
<td>stochastic, 41, 76, 168, 184</td>
</tr>
<tr>
<td>Random matrices, 42, 90</td>
<td>Wealth balance, 148, 154, 205</td>
</tr>
</tbody>
</table>

Worst low, 109
Zero-coupon, see Bond