Theory of Financial Risk and Derivative Pricing

From Statistical Physics to Risk Management

SECOND EDITION

Jean-Philippe Bouchaud and Marc Potters

CAMBRIDGE
UNIVERSITY PRESS

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

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Jean-Philippe Bouchaud and Marc Potters 2000, 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000
This edition published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typefaces Times 10/13 pt. and Helvetica *System* $\LaTeX\,2_\varepsilon$ [TB]

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

Library of Congress Cataloguing in Publication data
Bouchaud, Jean-Philippe, 1962–
Theory of financial risk and derivative pricing : from statistical physics to risk
management / Jean-Philippe Bouchaud and Marc Potters.–2nd edn
p. cm.
Rev. edn of: Theory of financial risks. 2000.
Includes bibliographical references and index.
ISBN 0 521 81916 4 (hardback)
1. Finance. 2. Financial engineering. 3. Risk assessment. 4. Risk management.
I. Potters, Marc, 1969– II. Bouchaud, Jean-Philippe, 1962– Theory of financial risks.
III. Title.
HG101.B68 2003
658.15′5 – dc21 2003044037

ISBN 0 521 81916 4 hardback

Contents

Foreword *page* xiii
Preface xv

1 Probability theory: basic notions 1
  1.1 Introduction 1
  1.2 Probability distributions 3
  1.3 Typical values and deviations 4
  1.4 Moments and characteristic function 6
  1.5 Divergence of moments – asymptotic behaviour 7
  1.6 Gaussian distribution 7
  1.7 Log-normal distribution 8
  1.8 Lévy distributions and Paretian tails 10
  1.9 Other distributions (*) 14
  1.10 Summary 16

2 Maximum and addition of random variables 17
  2.1 Maximum of random variables 17
  2.2 Sums of random variables 21
    2.2.1 Convolutions 21
    2.2.2 Additivity of cumulants and of tail amplitudes 22
    2.2.3 Stable distributions and self-similarity 23
  2.3 Central limit theorem 24
    2.3.1 Convergence to a Gaussian 25
    2.3.2 Convergence to a Lévy distribution 27
    2.3.3 Large deviations 28
    2.3.4 Steepest descent method and Cramèr function (*) 30
    2.3.5 The CLT at work on simple cases 32
    2.3.6 Truncated Lévy distributions 35
    2.3.7 Conclusion: survival and vanishing of tails 36
  2.4 From sum to max: progressive dominance of extremes (*) 37
  2.5 Linear correlations and fractional Brownian motion 38
  2.6 Summary 40

3 Continuous time limit, Ito calculus and path integrals 43
- 3.1 Divisibility and the continuous time limit 43
  - 3.1.1 Divisibility 43
  - 3.1.2 Infinite divisibility 44
  - 3.1.3 Poisson jump processes 45
- 3.2 Functions of the Brownian motion and Ito calculus 47
  - 3.2.1 Ito's lemma 47
  - 3.2.2 Novikov's formula 49
  - 3.2.3 Stratonovich's prescription 50
- 3.3 Other techniques 51
  - 3.3.1 Path integrals 51
  - 3.3.2 Girsanov's formula and the Martin–Siggia–Rose trick (*) 53
- 3.4 Summary 54

4 Analysis of empirical data 55
- 4.1 Estimating probability distributions 55
  - 4.1.1 Cumulative distribution and densities – rank histogram 55
  - 4.1.2 Kolmogorov–Smirnov test 56
  - 4.1.3 Maximum likelihood 57
  - 4.1.4 Relative likelihood 59
  - 4.1.5 A general caveat 60
- 4.2 Empirical moments: estimation and error 60
  - 4.2.1 Empirical mean 60
  - 4.2.2 Empirical variance and MAD 61
  - 4.2.3 Empirical kurtosis 61
  - 4.2.4 Error on the volatility 61
- 4.3 Correlograms and variograms 62
  - 4.3.1 Variogram 62
  - 4.3.2 Correlogram 63
  - 4.3.3 Hurst exponent 64
  - 4.3.4 Correlations across different time zones 64
- 4.4 Data with heterogeneous volatilities 66
- 4.5 Summary 67

5 Financial products and financial markets 69
- 5.1 Introduction 69
- 5.2 Financial products 69
  - 5.2.1 Cash (Interbank market) 69
  - 5.2.2 Stocks 71
  - 5.2.3 Stock indices 72
  - 5.2.4 Bonds 75
  - 5.2.5 Commodities 77
  - 5.2.6 Derivatives 77

5.3 Financial markets 79
  5.3.1 Market participants 79
  5.3.2 Market mechanisms 80
  5.3.3 Discreteness 81
  5.3.4 The order book 81
  5.3.5 The bid-ask spread 83
  5.3.6 Transaction costs 84
  5.3.7 Time zones, overnight, seasonalities 85
5.4 Summary 85

6 Statistics of real prices: basic results 87
6.1 Aim of the chapter 87
6.2 Second-order statistics 90
  6.2.1 Price increments vs. returns 90
  6.2.2 Autocorrelation and power spectrum 91
6.3 Distribution of returns over different time scales 94
  6.3.1 Presentation of the data 95
  6.3.2 The distribution of returns 96
  6.3.3 Convolutions 101
6.4 Tails, what tails? 102
6.5 Extreme markets 103
6.6 Discussion 104
6.7 Summary 105

7 Non-linear correlations and volatility fluctuations 107
7.1 Non-linear correlations and dependence 107
  7.1.1 Non identical variables 107
  7.1.2 A stochastic volatility model 109
  7.1.3 GARCH(1,1) 110
  7.1.4 Anomalous kurtosis 111
  7.1.5 The case of infinite kurtosis 113
7.2 Non-linear correlations in financial markets: empirical results 114
  7.2.1 Anomalous decay of the cumulants 114
  7.2.2 Volatility correlations and variogram 117
7.3 Models and mechanisms 123
  7.3.1 Multifractality and multifractal models (*) 123
  7.3.2 The microstructure of volatility 125
7.4 Summary 127

8 Skewness and price-volatility correlations 130
8.1 Theoretical considerations 130
  8.1.1 Anomalous skewness of sums of random variables 130
  8.1.2 Absolute vs. relative price changes 132
  8.1.3 The additive-multiplicative crossover and the q-transformation 134

8.2 A retarded model 135
8.2.1 Definition and basic properties 135
8.2.2 Skewness in the retarded model 136
8.3 Price-volatility correlations: empirical evidence 137
8.3.1 Leverage effect for stocks and the retarded model 139
8.3.2 Leverage effect for indices 140
8.3.3 Return-volume correlations 141
8.4 The Heston model: a model with volatility fluctuations and skew 141
8.5 Summary 144

9 Cross-correlations 145
9.1 Correlation matrices and principal component analysis 145
9.1.1 Introduction 145
9.1.2 Gaussian correlated variables 147
9.1.3 Empirical correlation matrices 147
9.2 Non-Gaussian correlated variables 149
9.2.1 Sums of non Gaussian variables 149
9.2.2 Non-linear transformation of correlated Gaussian variables 150
9.2.3 Copulas 150
9.2.4 Comparison of the two models 151
9.2.5 Multivariate Student distributions 153
9.2.6 Multivariate Lévy variables (*) 154
9.2.7 Weakly non Gaussian correlated variables (*) 155
9.3 Factors and clusters 156
9.3.1 One factor models 156
9.3.2 Multi-factor models 157
9.3.3 Partition around medoids 158
9.3.4 Eigenvector clustering 159
9.3.5 Maximum spanning tree 159
9.4 Summary 160
9.5 Appendix A: central limit theorem for random matrices 161
9.6 Appendix B: density of eigenvalues for random correlation matrices 164

10 Risk measures 168
10.1 Risk measurement and diversification 168
10.2 Risk and volatility 168
10.3 Risk of loss, 'value at risk' (VaR) and expected shortfall 171
10.3.1 Introduction 171
10.3.2 Value-at-risk 172
10.3.3 Expected shortfall 175
10.4 Temporal aspects: drawdown and cumulated loss 176
10.5 Diversification and utility – satisfaction thresholds 181
10.6 Summary 184

11 Extreme correlations and variety 186
11.1 Extreme event correlations 187
11.1.1 Correlations conditioned on large market moves 187
11.1.2 Real data and surrogate data 188
11.1.3 Conditioning on large individual stock returns: exceedance correlations 189
11.1.4 Tail dependence 191
11.1.5 Tail covariance (*) 194
11.2 Variety and conditional statistics of the residuals 195
11.2.1 The variety 195
11.2.2 The variety in the one-factor model 196
11.2.3 Conditional variety of the residuals 197
11.2.4 Conditional skewness of the residuals 198
11.3 Summary 199
11.4 Appendix C: some useful results on power-law variables 200

12 Optimal portfolios 202
12.1 Portfolios of uncorrelated assets 202
12.1.1 Uncorrelated Gaussian assets 203
12.1.2 Uncorrelated ‘power-law’ assets 206
12.1.3 ‘Exponential’ assets 208
12.1.4 General case: optimal portfolio and VaR (*) 210
12.2 Portfolios of correlated assets 211
12.2.1 Correlated Gaussian fluctuations 211
12.2.2 Optimal portfolios with non-linear constraints (*) 215
12.2.3 ‘Power-law’ fluctuations – linear model (*) 216
12.2.4 ‘Power-law’ fluctuations – Student model (*) 218
12.3 Optimized trading 218
12.4 Value-at-risk – general non-linear portfolios (*) 220
12.4.1 Outline of the method: identifying worst cases 220
12.4.2 Numerical test of the method 223
12.5 Summary 224

13 Futures and options: fundamental concepts 226
13.1 Introduction 226
13.1.1 Aim of the chapter 226
13.1.2 Strategies in uncertain conditions 226
13.1.3 Trading strategies and efficient markets 228
13.2 Futures and forwards 231
13.2.1 Setting the stage 231
13.2.2 Global financial balance 232
13.2.3 Riskless hedge 233
13.2.4 Conclusion: global balance and arbitrage 235

13.3 Options: definition and valuation 236
  13.3.1 Setting the stage 236
  13.3.2 Orders of magnitude 238
  13.3.3 Quantitative analysis – option price 239
  13.3.4 Real option prices, volatility smile and ‘implied’ kurtosis 242
  13.3.5 The case of an infinite kurtosis 249
13.4 Summary 251

14 Options: hedging and residual risk 254
14.1 Introduction 254
14.2 Optimal hedging strategies 256
  14.2.1 A simple case: static hedging 256
  14.2.2 The general case and ‘Δ’ hedging 257
  14.2.3 Global hedging vs. instantaneous hedging 262
14.3 Residual risk 263
  14.3.1 The Black–Scholes miracle 263
  14.3.2 The ‘stop-loss’ strategy does not work 265
  14.3.3 Instantaneous residual risk and kurtosis risk 266
  14.3.4 Stochastic volatility models 267
14.4 Hedging errors. A variational point of view 268
14.5 Other measures of risk – hedging and VaR (*) 268
14.6 Conclusion of the chapter 271
14.7 Summary 272
14.8 Appendix D 273

15 Options: the role of drift and correlations 276
15.1 Influence of drift on optimally hedged option 276
  15.1.1 A perturbative expansion 276
  15.1.2 ‘Risk neutral’ probability and martingales 278
15.2 Drift risk and delta-hedged options 279
  15.2.1 Hedging the drift risk 279
  15.2.2 The price of delta-hedged options 280
  15.2.3 A general option pricing formula 282
15.3 Pricing and hedging in the presence of temporal correlations (*) 283
  15.3.1 A general model of correlations 283
  15.3.2 Derivative pricing with small correlations 284
  15.3.3 The case of delta-hedging 285
15.4 Conclusion 285
  15.4.1 Is the price of an option unique? 285
  15.4.2 Should one always optimally hedge? 286
15.5 Summary 287
15.6 Appendix E 287

16 Options: the Black and Scholes model 290
16.1 Ito calculus and the Black-Scholes equation 290
16.1.1 The Gaussian Bachelier model 290
16.1.2 Solution and Martingale 291
16.1.3 Time value and the cost of hedging 293
16.1.4 The Log-normal Black–Scholes model 293
16.1.5 General pricing and hedging in a Brownian world 294
16.1.6 The Greeks 295
16.2 Drift and hedge in the Gaussian model (*) 295
16.2.1 Constant drift 295
16.2.2 Price dependent drift and the Ornstein–Uhlenbeck paradox 296
16.3 The binomial model 297
16.4 Summary 298

17 Options: some more specific problems 300
17.1 Other elements of the balance sheet 300
17.1.1 Interest rate and continuous dividends 300
17.1.2 Interest rate corrections to the hedging strategy 303
17.1.3 Discrete dividends 303
17.1.4 Transaction costs 304
17.2 Other types of options 305
17.2.1 ‘Put-call’ parity 305
17.2.2 ‘Digital’ options 305
17.2.3 ‘Asian’ options 306
17.2.4 ‘American’ options 308
17.2.5 ‘Barrier’ options (*) 310
17.2.6 Other types of options 312
17.3 The ‘Greeks’ and risk control 312
17.4 Risk diversification (*) 313
17.5 Summary 316

18 Options: minimum variance Monte–Carlo 317
18.1 Plain Monte-Carlo 317
18.1.1 Motivation and basic principle 317
18.1.2 Pricing the forward exactly 319
18.1.3 Calculating the Greeks 320
18.1.4 Drawbacks of the method 322
18.2 An ‘hedged’ Monte-Carlo method 323
18.2.1 Basic principle of the method 323
18.2.2 A linear parameterization of the price and hedge 324
18.2.3 The Black-Scholes limit 325
18.3 Non Gaussian models and purely historical option pricing 327
18.4 Discussion and extensions. Calibration 329

18.5 Summary 331
18.6 Appendix F: generating some random variables 331

19 The yield curve 334
19.1 Introduction 334
19.2 The bond market 335
19.3 Hedging bonds with other bonds 335
19.3.1 The general problem 335
19.3.2 The continuous time Gaussian limit 336
19.4 The equation for bond pricing 337
19.4.1 A general solution 339
19.4.2 The Vasicek model 340
19.4.3 Forward rates 341
19.4.4 More general models 341
19.5 Empirical study of the forward rate curve 343
19.5.1 Data and notations 343
19.5.2 Quantities of interest and data analysis 343
19.6 Theoretical considerations (*) 346
19.6.1 Comparison with the Vasicek model 346
19.6.2 Market price of risk 348
19.6.3 Risk-premium and the $\sqrt{\theta}$ law 349
19.7 Summary 351
19.8 Appendix G: optimal portfolio of bonds 352

20 Simple mechanisms for anomalous price statistics 355
20.1 Introduction 355
20.2 Simple models for herding and mimicry 356
20.2.1 Herding and percolation 356
20.2.2 Avalanches of opinion changes 357
20.3 Models of feedback effects on price fluctuations 359
20.3.1 Risk-aversion induced crashes 359
20.3.2 A simple model with volatility correlations and tails 363
20.3.3 Mechanisms for long ranged volatility correlations 364
20.4 The Minority Game 366
20.5 Summary 368

Index of most important symbols 372
Index 377

1

Probability theory: basic notions

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

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

1.1 Introduction

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

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

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

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

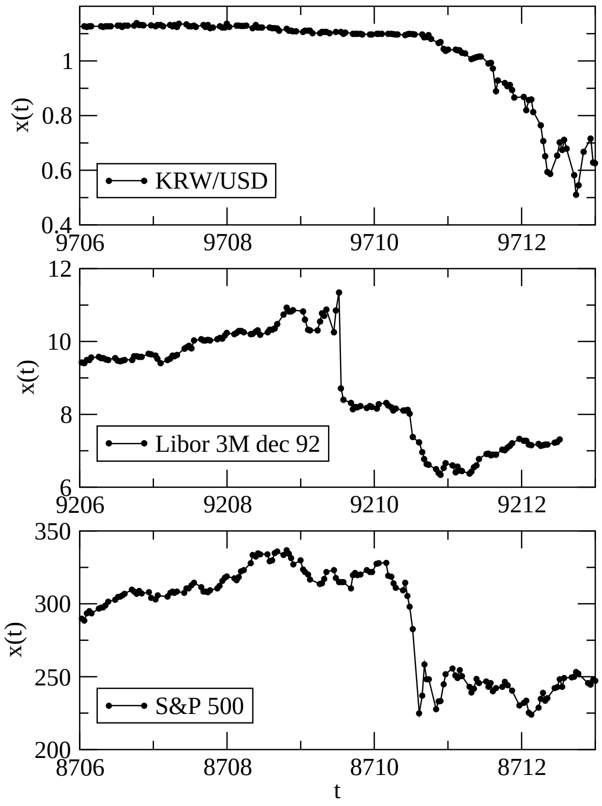

Fig. 1.1. Three examples of statistically unforeseen crashes: the Korean won against the dollar in 1997 (top), the British 3-month short-term interest rates futures in 1992 (middle), and the S&P 500 in 1987 (bottom). In the example of the Korean won, it is particularly clear that the distribution of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

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*.[†]

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

1.2 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[‡] 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 $\mathrm{d}x$ around $X = x$ is equal to $P(x)\,\mathrm{d}x$. In the following, we shall denote as $P(\cdot)$ the probability density for the variable appearing as the argument of the function. This is a potentially ambiguous, but very useful notation.

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

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

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

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

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

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

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

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

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

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

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

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

1.3 Typical values and deviations

It is quite natural to speak about ‘typical’ values of X. There are at least three mathematical definitions of this intuitive notion: the **most probable** value, the **median** and the **mean**. The most probable value x^* corresponds to the maximum of the function $P(x)$; x^* needs not be unique if $P(x)$ has several equivalent maxima. The median x_{med} is such that the probabilities that X be greater or less than this particular value are equal. In other words, $\mathcal{P}_<(x_{\mathrm{med}}) = \mathcal{P}_>(x_{\mathrm{med}}) = \frac{1}{2}$. The mean, or **expected value** of X, which we shall note as m or $\langle x\rangle$ in the following, is the average of all possible values of X, weighted by their corresponding probability:

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

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

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

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

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

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

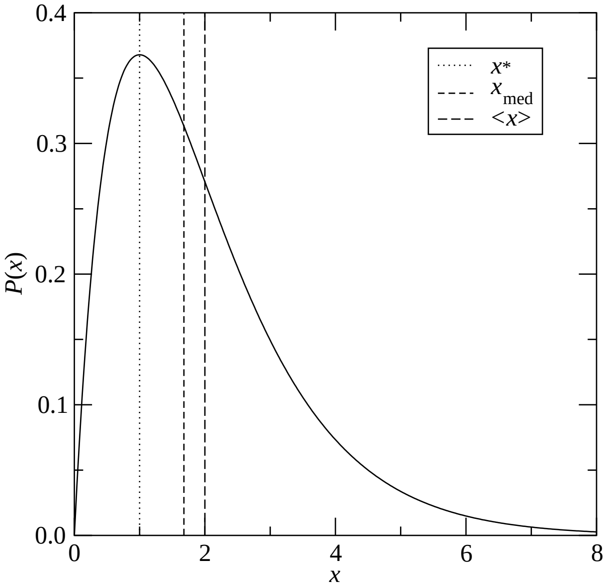

Fig. 1.2. The ‘typical value’ of a random variable X drawn according to a distribution density $P(x)$ can be defined in at least three different ways: through its mean value $\langle x \rangle$, its most probable value x^* or its median x_{med}. In the general case these three values are distinct.

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

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

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

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

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 in the next chapter. However, for the empirical study of fluctuations, it is sometimes preferable to use the MAD; it is more *robust* than the variance, that is, less sensitive to rare extreme events, which may be the source of large statistical errors.

1.4 Moments and characteristic function

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

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

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

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

$$\hat{P}(z) \equiv \int \mathrm{e}^{\mathrm{i}zx} P(x)\,\mathrm{d}x. \tag{1.8}$$

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

$$P(x) = \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}zx} \hat{P}(z)\,\mathrm{d}z. \tag{1.9}$$

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

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

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

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

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

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

[†] This is not rigorously correct, since one can exhibit examples of different distribution densities which possess exactly the same moments, see Section 1.7 below.

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

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

The above definition of cumulants may look arbitrary, but these quantities have remarkable properties. For example, as we shall show in Section 2.2, the cumulants simply add when one sums independent random variables. Moreover a Gaussian distribution (or the normal law of Laplace and Gauss) is characterized by the fact that all cumulants of order larger than two are identically zero. Hence the cumulants, in particular κ, can be interpreted as a measure of the distance between a given distribution $P(x)$ and a Gaussian.

1.5 Divergence of moments – asymptotic behaviour

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

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

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

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

1.6 Gaussian distribution

The most commonly encountered distributions are the ‘normal’ laws of Laplace and Gauss, which we shall simply call **Gaussian** in the following. Gaussians are ubiquitous: for example, the number of *heads* in a sequence of a thousand coin tosses, the exact number of oxygen molecules in the room, the height (in inches) of a randomly selected individual, are all approximately described by a Gaussian distribution.[†] The ubiquity of the Gaussian can be in part traced to the central limit theorem (CLT) discussed at length in Chapter 2, 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 Chapter 6, the price fluctuations of most financial assets.

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

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

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

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

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

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

Its logarithm is a second-order polynomial, for which all derivatives of order larger than two are zero. In particular, the kurtosis of a Gaussian variable is zero. As mentioned above, the kurtosis is often taken as a measure of the distance from a Gaussian distribution. When $\kappa > 0$ (**leptokurtic** distributions), the corresponding distribution density has a marked peak around the mean, and rather 'thick' tails. Conversely, when $\kappa < 0$, the distribution density has a flat top and very thin tails. For example, the uniform distribution over a certain interval (for which tails are absent) has a kurtosis $\kappa = -\frac{6}{5}$. Note that the kurtosis is bounded from below by the value -2, which corresponds to the case where the random variable can only take two values $-a$ and a with equal probability.

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

1.7 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 in Chapter 2. The log-normal distribution density is thus defined as:[†]

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

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

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

From these moments, one deduces the skewness, given by $\varsigma_3 = (e^{3\sigma^2} - 3e^{\sigma^2} + 2)/(e^{\sigma^2} - 1)^{3/2}$, ($\simeq 3\sigma$ for $\sigma \ll 1$), and the kurtosis $\kappa = (e^{6\sigma^2} - 4e^{3\sigma^2} + 6e^{\sigma^2} - 3)/(e^{\sigma^2} - 1)^2 - 3$, ($\simeq 19\sigma^2$ for $\sigma \ll 1$).

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

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

---

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

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

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

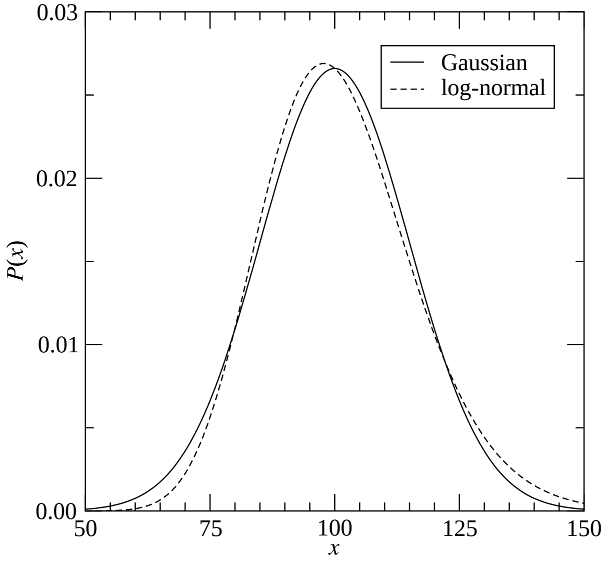

Fig. 1.3. Comparison between a Gaussian (thick line) and a log-normal (dashed line), with $m = x_0 = 100$ and σ equal to 15 and 15% respectively. The difference between the two curves shows up in the tails.

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

1.8 Lévy distributions and Paretian tails

Lévy distributions (noted $L_\mu(x)$ below) appear naturally in the context of the CLT (see Chapter 2), because of their stability property under addition (a property shared by Gaussians). The tails of Lévy distributions are however much 'fatter' than those of Gaussians, and are thus useful to describe multiscale phenomena (i.e. when both very large and very small values of a quantity can commonly be observed – such as personal income, size of pension funds, amplitude of earthquakes or other natural catastrophes, etc.). These distributions were introduced in the 1950s and 1960s by Mandelbrot (following Pareto) to describe personal income and the price changes of some financial assets, in particular the price of cotton. An important constitutive property of these Lévy distributions is their power-law behaviour for large arguments, often called **Pareto tails**:

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

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

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

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

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

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

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

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

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

where a_μ is a constant proportional to the tail parameter A^μ:

$$A^\mu = \mu\Gamma(\mu - 1)\frac{\sin(\pi\mu/2)}{\pi} a_\mu \qquad 1 < \mu < 2, \tag{1.21}$$

and

$$A^\mu = (1 - \mu)\Gamma(\mu)\frac{\sin(\pi\mu/2)}{\pi\mu} a_\mu \qquad \mu < 1. \tag{1.22}$$

It is clear, from (1.20), that in the limit $\mu = 2$, one recovers the definition of a Gaussian. When μ decreases from 2, the distribution becomes more and more sharply peaked around the origin and fatter in its tails, while ‘intermediate’ events lose weight (Fig. 1.4). These distributions thus describe ‘intermittent’ phenomena, very often small, sometimes gigantic.

The moments of the symmetric Lévy distribution can be computed, when they exist. One finds:

$$\langle |x|^\nu \rangle = (a_\mu)^{\nu/\mu}\, \frac{\Gamma(-\nu/\mu)}{\mu\Gamma(-\nu)\, \cos(\pi\nu/2)}, \qquad -1 < \nu < \mu. \tag{1.23}$$

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

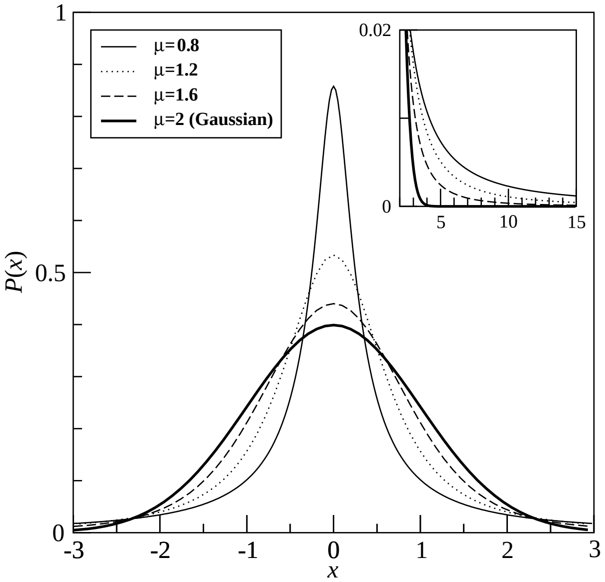

Fig. 1.4. Shape of the symmetric Lévy distributions with $\mu = 0.8,\ 1.2,\ 1.6$ and 2 (this last value actually corresponds to a Gaussian). The smaller μ, the sharper the ‘body’ of the distribution, and the fatter the tails, as illustrated in the inset.

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

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

$$\hat{L}_\mu^\beta(z) = \exp\left[-a_\mu |z|^\mu \left(1 + i\beta \tan(\mu\pi/2)\frac{z}{|z|}\right)\right] \quad (\mu \neq 1). \tag{1.24}$$

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

$$L_\mu(x) = \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{\pi n!} \frac{a_\mu^n}{x^{1+n\mu}} \Gamma(1 + n\mu) \sin(\pi\mu n/2). \tag{1.25}$$

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

An interesting generalization of the symmetric Lévy distributions which accounts for this exponential cut-off is given by the **truncated Lévy distributions** (TLD), which will be of much use in the following. A simple way to alter the characteristic function Eq. (1.20) to account for an exponential cut-off for large arguments is to set:

$$\hat{L}_\mu^{(t)}(z) = \exp\left[-a_\mu \frac{(\alpha^2+z^2)^{\frac{\mu}{2}}\cos(\mu \arctan(|z|/\alpha)) - \alpha^\mu}{\cos(\pi\mu/2)}\right], \tag{1.26}$$

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

$$\frac{a_\mu}{2\cos(\pi\mu/2)}[(\alpha+\mathrm{i}z)^\mu + (\alpha-\mathrm{i}z)^\mu - 2\alpha^\mu]. \tag{1.27}$$

The first cumulants of the distribution defined by Eq. (1.26) read, for $1 < \mu < 2$:

$$c_2 = \mu(\mu-1)\frac{a_\mu}{|\cos\pi\mu/2|}\alpha^{\mu-2} \qquad c_3 = 0. \tag{1.28}$$

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

$$\lambda_4 = \frac{(3-\mu)(2-\mu)|\cos\pi\mu/2|}{\mu(\mu-1)a_\mu\alpha^\mu}. \tag{1.29}$$

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.

As explained below in Section 3.1.3, the truncated Lévy distribution has the interesting property of being infinitely divisible for all values of α and μ (this includes the Gaussian distribution and the pure Lévy distributions).

Exponential tail: a limiting case

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

$$\mathcal{P}_>(x) = \frac{A^\mu}{(A+x)^\mu}. \tag{1.30}$$

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

$$\mathcal{P}_>(x) = \frac{1}{[1+(\alpha x/\mu)]^\mu} \underset{\mu\to\infty}{\longrightarrow} \exp(-\alpha x). \tag{1.31}$$

1.9 Other distributions (*)

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

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

$$P(n) \equiv \frac{(\omega\ell)^n}{n!} \exp(-\omega\ell). \tag{1.32}$$

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

$$P_{\mathrm{H}}(x) \equiv \frac{1}{2x_0 K_1(\alpha x_0)} \exp -\left[\alpha\sqrt{x_0^2 + x^2}\right], \tag{1.33}$$

where the normalization $K_1(\alpha x_0)$ is a modified Bessel function of the second kind. For x small compared to x_0, $P_{\mathrm{H}}(x)$ behaves as a Gaussian although its asymptotic behaviour for $x \gg x_0$ is fatter and reads $\exp(-\alpha|x|)$.

From the characteristic function

$$\hat{P}_{\mathrm{H}}(z) = \frac{\alpha x_0 K_1(x_0\sqrt{1+\alpha z})}{K_1(\alpha x_0)\sqrt{1+\alpha z}}, \tag{1.34}$$

we can compute the variance

$$\sigma^2 = \frac{x_0 K_2(\alpha x_0)}{\alpha K_1(\alpha x_0)}, \tag{1.35}$$

and kurtosis

$$\kappa = 3\left(\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)}\right)^2 + \frac{12}{\alpha x_0}\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)} - 3. \tag{1.36}$$

Note that the kurtosis of the hyperbolic distribution is always between zero and three. (The skewness is zero since the distribution is even.)

In the case $x_0 = 0$, one finds the symmetric exponential distribution:

$$P_{\mathrm{E}}(x) = \frac{\alpha}{2} \exp(-\alpha|x|), \tag{1.37}$$

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

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

$$P_{\mathrm{S}}(x) \equiv \frac{1}{\sqrt{\pi}} \frac{\Gamma((1+\mu)/2)}{\Gamma(\mu/2)} \frac{a^\mu}{(a^2 + x^2)^{(1+\mu)/2}}, \tag{1.38}$$

which coincides with the Cauchy distribution for $\mu = 1$, and tends towards a Gaussian in the limit $\mu \to \infty$, provided that a^2 is scaled as μ. This distribution is usually known as

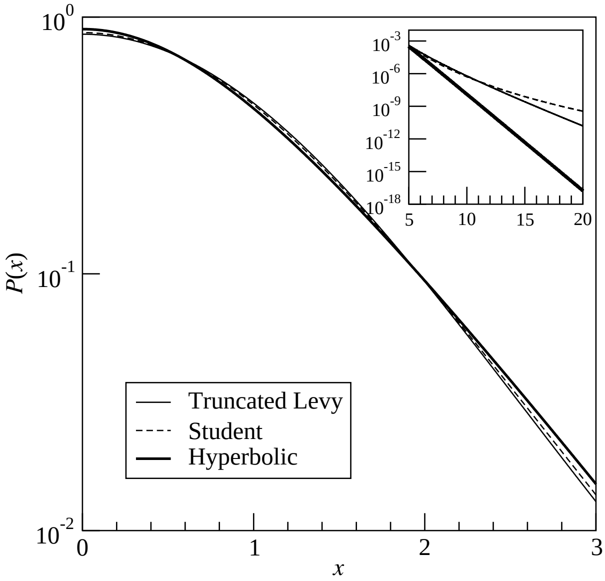

Fig. 1.5. Probability density for the truncated Lévy ($\mu = \frac{3}{2}$), Student and hyperbolic distributions. All three have two free parameters which were fixed to have unit variance and kurtosis. The inset shows a blow-up of the tails where one can see that the Student distribution has tails similar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

Student's t-distribution with μ degrees of freedom, but we shall call it simply the Student distribution.

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: $m_{2n} = (2n-1)!!\,\sigma^{2n}$. The kurtosis of the Student distribution is given by $\kappa = 6/(\mu-4)$. Figure 1.5 shows a plot of the Student distribution with $\kappa = 1$, corresponding to $\mu = 10$.

Note that the characteristic function of Student distributions can also be explicitly computed, and reads:

$$\hat{P}_S(z) = \frac{2^{1-\mu/2}}{\Gamma(\mu/2)}(az)^{\mu/2}K_{\mu/2}(az), \tag{1.39}$$

where $K_{\mu/2}$ is the modified Bessel function. The cumulative distribution in the useful cases $\mu = 3$ and $\mu = 4$ with a chosen such that the variance is unity read:

$$\mathcal{P}_{S,>}(x) = \frac{1}{2} - \frac{1}{\pi}\left[\arctan x + \frac{x}{1+x^2}\right] \qquad (\mu = 3,\ a = 1), \tag{1.40}$$

and

$$\mathcal{P}_{S,>}(x) = \frac{1}{2} - \frac{3}{4}u + \frac{1}{4}u^3, \qquad (\mu = 4,\ a = \sqrt{2}), \tag{1.41}$$

where $u = x/\sqrt{2+x^2}$.

- The **inverse gamma** distribution, for positive quantities (such as, for example, volatilities, or waiting times), also has power-law tails. It is defined as:

  $$P_\Gamma(x) = \frac{x_0^\mu}{\Gamma(\mu)x^{1+\mu}} \exp\left(-\frac{x_0}{x}\right). \tag{1.42}$$

  Its moments of order $n < \mu$ are easily computed to give: $m_n = x_0^n \Gamma(\mu - n)/\Gamma(\mu)$. This distribution falls off very fast when $x \to 0$. As we shall see in Chapter 7, an inverse gamma distribution and a log-normal distribution can sometimes be hard to distinguish empirically. Finally, if the volatility σ^2 of a Gaussian is itself distributed as an inverse gamma distribution, the distribution of x becomes a Student distribution – see Section 9.2.5 for more details.

1.10 Summary

- The most probable value and the mean value are both estimates of the typical values of a random variable. Fluctuations around this value are measured by the root mean square deviation or the mean absolute deviation.
- For some distributions with very fat tails, the mean square deviation (or even the mean value) is infinite, and the typical values must be described using quantiles.
- The Gaussian, the log-normal and the Student distributions are some of the important probability distributions for financial applications.
- The way to generate numerically random variables with a given distribution (Gaussian, Lévy stable, Student, etc.) is discussed in Chapter 18, Appendix F.

- FURTHER READING

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

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

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

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