Risk Analysis in Finance and Insurance, Second Edition presents an accessible yet comprehensive introduction to the main concepts and methods that transform risk management into a quantitative science. Taking into account the interdisciplinary nature of risk analysis, the author discusses many important ideas from mathematics, finance, and actuarial science in a simplified manner. He explores the interconnections among these disciplines and encourages readers toward further study of the subject. This edition continues to study risks associated with financial and insurance contracts, using an approach that estimates the value of future payments based on current financial, insurance, and other information.

**New to the Second Edition**
- Expanded section on the foundations of probability and stochastic analysis
- Coverage of new topics, including financial markets with stochastic volatility, risk measures, risk-adjusted performance measures, and equity-linked insurance
- More worked examples and problems

Reorganized and expanded, this updated book illustrates how to use quantitative methods of stochastic analysis in modern financial mathematics. These methods can be naturally extended and applied in actuarial science, thus leading to unified methods of risk analysis and management.

**Features**
- Introduces the main ideas, techniques, and stochastic models of financial mathematics
- Focuses on the foundations and key concepts of the modern methodology of quantitative financial analysis
- Explores the problems of managing insurance risks
- Examines the multiple intrinsic connections between insurance risks and financial risks
- Shows how stochastic analysis is a powerful tool for modeling financial and insurance risk processes
Risk Analysis in Finance and Insurance
Second Edition
Aims and scope:
The field of financial mathematics forms an ever-expanding slice of the financial sector. This series aims to capture new developments and summarize what is known over the whole spectrum of this field. It will include a broad range of textbooks, reference works and handbooks that are meant to appeal to both academics and practitioners. The inclusion of numerical code and concrete real-world examples is highly encouraged.

Series Editors
M.A.H. Dempster
Centre for Financial Research
Department of Pure Mathematics and Statistics
University of Cambridge

Dilip B. Madan
Robert H. Smith School of Business
University of Maryland

Rama Cont
Center for Financial Engineering
Columbia University
New York

Published Titles
American-Style Derivatives; Valuation and Computation, Jerome Detemple
Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing, Pierre Henry-Labordère
Credit Risk: Models, Derivatives, and Management, Niklas Wagner
Engineering BGM, Alan Brace
Financial Modelling with Jump Processes, Rama Cont and Peter Tankov
Interest Rate Modeling: Theory and Practice, Lixin Wu
Introduction to Credit Risk Modeling, Second Edition, Christian Bluhm, Ludger Overbeck, and Christoph Wagner
Introduction to Stochastic Calculus Applied to Finance, Second Edition, Damien Lamberton and Bernard Lapeyre
Monte Carlo Methods and Models in Finance and Insurance, Ralf Korn, Elke Korn, and Gerald Kroisandt
Portfolio Optimization and Performance Analysis, Jean-Luc Prigent
Quantitative Fund Management, M. A. H. Dempster, Georg Pflug, and Gautam Mitra
Robust Libor Modelling and Pricing of Derivative Products, John Schoenmakers
Stochastic Finance: A Numeraire Approach, Jan Vecer
Stochastic Financial Models, Douglas Kennedy
Structured Credit Portfolio Analysis, Baskets & CDOs, Christian Bluhm and Ludger Overbeck
Understanding Risk: The Theory and Practice of Financial Risk Management, David Murphy
Unravelling the Credit Crunch, David Murphy

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
4th, Floor, Albert House
1-4 Singer Street
London EC2A 4BQ
UK
Risk Analysis in Finance and Insurance
Second Edition

Alexander Melnikov
Preface to 2nd edition

This book deals with the notion of “risk” and is devoted to analysis of risks in finance and insurance. We will study risks associated with financial and insurance contracts, by which we understand risks to be uncertainties that may result in financial loss and affect the ability to make payments associated with the corresponding contracts. Our approach to this analysis is based on the development of a methodology for estimating the present value of the future payments given current financial, insurance, and other information. Using this approach, one can appropriately define notions of price for a financial contract, of premium for insurance policy, and of reserve for an insurance company. Historically, financial risks were subject to elementary mathematics of finance and were treated separately from insurance risks, which were analyzed in actuarial science. The development of quantitative methods based on stochastic analysis is a key achievement of modern financial mathematics. These methods can be naturally extended and applied in the area of actuarial science, thus leading to unified methods of risk analysis and management.

The aim of this book is to give an accessible comprehensive introduction to the main ideas, methods, and techniques that transform risk management into a quantitative science. Because of the interdisciplinary nature of this book, many important notions and facts from mathematics, finance, and actuarial science are discussed in an appropriately simplified manner. Our goal is to present interconnections among these disciplines and to encourage our reader toward further study of the subject. We indicate some initial directions in the Bibliographic Remarks.

This edition is reorganized in a way that allows a natural flow of topics covered in the first edition to be combined together with new additions such as: financial markets with stochastic volatility, risk measures, risk-adjusted performance measures, equity-linked insurance, and so forth. The substantial extension of the section regarding the foundations of Probability and Stochastic Analysis makes this book self-contained. Furthermore, an increased number of worked examples and a collection of some 140 problems, which is accompanied by the Instructor’s Solutions Manual, make this edition more attractive both from a research and a pedagogical perspective. This book can be readily used as a textbook for a Mathematical Finance course, both at introductory undergraduate and advanced graduate levels. It has been used for teaching Mathematical Finance at both levels at the University of Alberta, and many
student comments and recommendations are taken into account in this edition.

The author thanks his graduate students Anna Evstafyeva, Hao Li, and Henry Heung for their help in introducing some new worked examples and problems in this edition. The author is also very grateful to Dr. Alexei Filinkov for translating, editing, and preparing the manuscript.

Alexander Melnikov

University of Alberta, Edmonton, Canada

October 2010
Introduction

Financial and insurance markets always operate under various types of uncertainties that can affect the financial position of companies and individuals. In financial and insurance theories, these uncertainties are usually referred to as risks. Given certain states of the market, and the economy in general, one can talk about risk exposure. It is expected that individuals, companies, and public establishments that aim to accumulate wealth should examine their risk exposure. The process of risk management consists of a sequence of corresponding actions over a period of time that are designed to mitigate the level of risk exposure. Some of the main principles and ingredients of risk management are qualitative identification of risk, estimation of possible losses, choosing the appropriate strategies for avoiding losses and for shifting the risk to other parts of the financial system, including analysis of the involved costs, and using feedback for developing adequate controls.

The first six chapters of this book are devoted to the financial market risks. We aim to give an elementary and yet comprehensive introduction to the main ideas, methods, and stochastic models of financial mathematics. The probabilistic approach appears to be one of the most efficient ways of modeling uncertainties in financial markets. Risks (or uncertainties of financial market operations) are described in terms of statistically stable stochastic experiments, and therefore estimation of risks is reduced to the construction of financial forecasts adapted to these experiments. Using conditional expectations, one can quantitatively describe these forecasts given the observable market prices and events. Thus, it can be possible to construct dynamic hedging strategies and those for optimal investment. The foundations and key concepts of the modern methodology of quantitative financial analysis are the main focus of Chapters 1–6.

Insurance against possible financial losses is one of the key ingredients of risk management. However, the insurance business is an integral part of the financial system. Chapters 7–8 focus on the problems of managing insurance risks. Multiple intrinsic connections between insurance risks and financial risks are also considered.

Our treatment of insurance risk management demonstrates that methods of risk evaluation and management in insurance and finance are interrelated and can be treated using a single integrated approach. Estimations of future payments and their corresponding risks are key operational tasks of financial and insurance companies. Management of these risks requires an accurate
evaluation of the present values of future payments, and therefore the adequate modeling of (financial and insurance) risk processes. Stochastic analysis is one of the most powerful tools for such modeling, and it is the fundamental basis of our presentation.

Finally, we note that probabilistic methods were used in finance and insurance since the early 1950s. They were developed extensively over the past decades, especially after the seminal papers by F. Black and M. Scholes and R. C. Merton, published in 1973.
## Contents

1 Introductory concepts of Financial Risk Management and Related Mathematical Tools  
   1.1 Introductory concepts of the securities market  
   1.2 Probabilistic foundations of financial modeling and pricing of contingent claims  
   1.3 Elements of probability theory and stochastic analysis  

2 Financial Risk Management in the Binomial Model  
   2.1 The binomial model of a financial market. Absence of arbitrage, uniqueness of a risk-neutral probability measure, martingale representation.  
   2.2 Hedging contingent claims in the binomial market model. The Cox-Ross-Rubinstein formula  
   2.3 Pricing and hedging American options  
   2.4 Utility functions and St. Petersburg’s paradox. The problem of optimal investment.  
   2.5 The term structure of prices, hedging, and investment strategies in the Ho-Lee model  
   2.6 The transition from the binomial model of a financial market to a continuous model. The Black-Scholes formula and equation.  

3 Advanced Analysis of Financial Risks: Discrete Time Models  
   3.1 Fundamental theorems on arbitrage and completeness. Pricing and hedging contingent claims in complete and incomplete markets.  
   3.2 The structure of options prices in incomplete markets and in markets with constraints.  
   3.3 Hedging contingent claims in mean square  
   3.4 Gaussian model of a financial market in discrete time. Insurance appreciation and discrete version of the Black-Scholes formula.  

4 Analysis of Risks: Continuous Time Models  
   4.1 The Black-Scholes model. “Greek” parameters in risk management, hedging, and optimal investment.  
   4.2 Beyond of the Black-Scholes model  
   4.3 Imperfect hedging and risk measures
## Contents

5 Fixed Income Securities: Modeling and Pricing 139  
5.1 Elements of deterministic theory of fixed income instruments 139  
5.2 Stochastic modeling and pricing bonds and their derivatives 156  

6 Implementations of Risk Analysis in Various Areas of Financial Industry 165  
6.1 Real options: pricing long-term investment projects 165  
6.2 Technical analysis in risk management 173  
6.3 Performance measures and their applications 183  

7 Insurance and Reinsurance Risks 189  
7.1 Modeling risk in insurance and methodologies of premium calculations 189  
7.2 Risks transfers via reinsurance 199  
7.3 Elements of traditional life insurance 208  
7.4 Risk modeling and pricing in innovative life insurance 215  

8 Solvency Problem for an Insurance Company: Discrete and Continuous Time Models 225  
8.1 Ruin probability as a measure of solvency of an insurance company 225  
8.2 Solvency of an insurance company and investment portfolios 241  
8.3 Solvency problem in a generalized Cramér-Lundberg model 254  

Appendix A Problems 265  
A.1 Probability theory and elements of stochastic analysis 265  
A.2 General questions on financial markets 270  
A.3 Binomial model 274  
A.4 The Black-Scholes model 281  
A.5 Bond market 284  
A.6 Risk and performance measurement 287  
A.7 Elements of insurance and actuarial science 293  

Appendix B Bibliographic Remarks 299  

Bibliography 303  

Glossary of Notation 311  

Index 313
Chapter 1


1.1 Introductory concepts of the securities market

The notion of an asset (anything of value) is one of the fundamental notions in mathematical finance. Assets can be risky and non-risky. Here risk is understood to be an uncertainty that can cause loss (e.g., of wealth). The most typical representatives of such assets are the basic securities: stocks $S$ and bonds (bank accounts) $B$. These securities constitute the basis of a financial market that can be understood as a space of assets equipped with a structure for their trading.

Stocks are share securities that accumulate capital required for a company’s successful operation. The stockholder holds the right to participate in the control of the company and to receive dividends.

Bonds are debt securities issued by a government or a company for accumulating capital, restructuring debts, and so forth. In contrast to stocks, bonds are issued for a specified period of time. The essential characteristics of a bond include their exercise and maturity time, face value (principal or nominal), coupons (payments up to maturity), and yield (return up to maturity). The zero-coupon bond is similar to a bank account, and its yield corresponds to a bank interest rate.

An interest rate $r \geq 0$ is typically quoted by banks as an annual percentage. Suppose that a client opens an account with a deposit of $B_0$, then at the end of a 1-year period the client’s non-risky profit is $\Delta B_1 = B_1 - B_0 = rB_0$. After $n$ years, the balance of this account will be $B_n = B_{n-1} + rB_0$, given that only the initial deposit $B_0$ is reinvested every year. In this case, $r$ is referred to as a simple interest.

Alternatively, the earned interest also can be reinvested (compounded), and then, at the end of $n$ years, the balance will be $B_n = B_{n-1}(1 + r) = B_0(1 + r)^n$. Note that here the ratio $\Delta B_n/B_{n-1}$ reflects the profitability or return of the investment as it is equal to $r$, the compound interest.
Now suppose that interest is compounded \( m \) times per year, then
\[
B_n = B_{n-1} \left( 1 + \frac{r(m)}{m} \right)^m = B_0 \left( 1 + \frac{r(m)}{m} \right)^{mn}.
\]
Such rate \( r(m) \) is quoted as a **nominal** (annual) interest rate, and the equivalent **effective** (annual) interest rate is equal to
\[
r = \left( 1 + \frac{r(m)}{m} \right)^m - 1.
\]

Let \( t \geq 0 \), and consider the ratio
\[
\frac{B_{t+\frac{1}{m}} - B_t}{B_t} = \frac{r(m)}{m},
\]
where \( r(m) \) is a nominal annual interest rate compounded \( m \) times per year. Then another rate
\[
r = \lim_{m \to \infty} \frac{B_{t+\frac{1}{m}} - B_t}{\frac{1}{m} B_t} = \lim_{m \to \infty} r(m) = \frac{1}{B_t} \frac{dB_t}{dt}
\]
is called the nominal annual interest rate **compounded continuously**. Clearly, \( B_t = B_0 e^{rt} \).

Thus, the concept of interest is one of the essential components in the description of money value time evolution. Now consider a series of periodic payments (deposits) \( f_0, f_1, \ldots, f_n \) (**annuity**). It follows from the formula for compound interest that the present value of \( k \)-th payment is equal to \( f_k (1+r)^{-k} \), and therefore the present value of the annuity is \( \sum_{k=0}^{n} f_k (1+r)^{-k} \).

**Worked Example 1.1** Let an initial deposit into a bank account be $10,000. Given that \( r(m) = 0.1 \), find the account balance at the end of 2 years for \( m = 1, 3, \) and 6. Also find the balance at the end of each of years 1 and 2 if the interest is compounded continuously at the rate \( r = 0.1 \).

**Solution** Using the notion of compound interest, we have
\[
B_2^{(1)} = 10,000 \left( 1 + 0.1 \right)^2 = 12,100
\]
for interest compounded once per year;
\[
B_2^{(3)} = 10,000 \left( 1 + \frac{0.1}{3} \right)^{2 \times 3} \approx 12,174
\]
for interest compounded three times per year;
\[
B_2^{(6)} = 10,000 \left( 1 + \frac{0.1}{6} \right)^{2 \times 6} \approx 12,194
\]
for interest compounded six times per year.
For interest compounded continuously we obtain

\[ B_1^{(∞)} = 10,000 e^{0.1} ≈ 11,052, \quad B_2^{(∞)} = 10,000 e^{2×0.1} ≈ 12,214. \]

Stocks are significantly more volatile than bonds, and therefore they are characterized as **risky assets**. Similarly to bonds, one can define their **profitability** or **return** \( \rho_n = \Delta S_n / S_{n-1} \), \( n = 1, 2, \ldots \), where \( S_n \) is the price of a stock at time \( n \). Then we have the following discrete equation for stock prices:

\[ S_n = S_{n-1}(1 + \rho_n), \quad S_0 > 0. \]

The mathematical model of a financial market formed by a bank account \( B \) (with an interest rate \( r \)) and a stock \( S \) (with profitability \( \rho_n \)) is referred to as a \((B, S)\)-market.

The volatility of prices \( S_n \) is caused by a great variety of sources, some of which may not be easily observed. In this case, the notion of **randomness** appears to be appropriate, so that \( S_n \), and therefore \( \rho_n \), can be considered as **random variables**. Since at every time step \( n \) the price of a stock goes either up or down, then it is natural to assume that profitabilities \( \rho_n \) form a sequence of independent random variables \((\rho_n)_{n=1,2,\ldots}\) that take values \( b \) and \( a \) \((b > a)\) with probabilities \( p \) and \( q \), respectively \((p + q = 1)\). Next, we can write \( \rho_n \) as a sum of its mean \( \mu = bp + aq \) and a random variable \( w_n = \rho_n - \mu \), which has the expectation equal to zero. Thus, profitability \( \rho_n \) can be described in terms of an independent random deviation \( w_n \) from the mean profitability \( \mu \).

When the time steps become smaller, the oscillations of profitability become more chaotic. Formally, the limit continuous model can be written as

\[ \frac{dS_t}{S_t} \equiv \frac{dS_t}{dt} = \mu + \sigma \dot{W}_t, \]

where \( \mu \) is the mean profitability, \( \sigma \) is the volatility of the market, and \( \dot{W}_t \) is the Gaussian white noise.

The formulas for compound and continuous interest rates together with the corresponding equation for stock prices define the binomial (Cox-Ross-Rubinstein) and the diffusion (Black-Scholes) models of the market, respectively.

A participant in a financial market usually invests free capital in various available assets that then form an **investment portfolio**. The effective management of the capital is realized through a process of building and managing the portfolio. The redistribution of a portfolio with the goal of limiting or minimizing the risk in various financial transaction is usually referred to as **hedging**. The corresponding portfolio is then called a **hedging portfolio**. An investment strategy (portfolio) that may give a profit even with zero initial investment is called an **arbitrage** strategy. The presence of arbitrage reflects the instability of a financial market.

The development of a financial market offers the participants **derivative securities**, that is, securities that are formed on the basis of the basic securities – stocks and bonds. The derivative securities (forwards, futures, options,
etc.) require smaller initial investment and play the role of insurance against possible losses. They also increase the liquidity of the market.

For example, suppose company $A$ plans to purchase shares of company $B$ at the end of the year. To protect itself from a possible increase in shares prices, company $A$ reaches an agreement with company $B$ to buy the shares at the end of the year for a fixed (forward) price $F$. Such an agreement between the two companies is called a forward contract (or simply, forward).

Now suppose that company $A$ plans to sell some shares to company $B$ at the end of the year. To protect itself from a possible fall in price of those shares, company $A$ buys a put option (seller’s option), which confers the right to sell the shares at the end of the year at the fixed strike price $K$. Note that, in contrast to the forwards case, a holder of an option must pay a premium to its issuer.

Futures contract is an agreement similar to the forward contract, but the trading takes place on a stock exchange, a special organization that manages the trading of various goods, financial instruments, and services.

Finally, we reiterate here that mathematical models of financial markets, methodologies for pricing various financial instruments and for constructing optimal (minimizing risk) investment strategies, are all subject to modern mathematical finance.

1.2 Probabilistic foundations of financial modeling and pricing of contingent claims

Suppose that a non-risky asset $B$ and a risky asset $S$ are completely described at any time $n = 0, 1, 2, \ldots$ by their prices. Therefore, it is natural to assume that the price dynamics of these securities are the essential component of a financial market. These dynamics are represented by the following equations:

\[
\begin{align*}
\Delta B_n &= rB_{n-1}, \quad B_0 = 1, \\
\Delta S_n &= \rho_n S_{n-1}, \quad S_0 > 0,
\end{align*}
\]

where $\Delta B_n = B_n - B_{n-1}$, $\Delta S_n = S_n - S_{n-1}$, $n = 1, 2, \ldots$; $r \geq 0$ is a constant rate of interest and $\rho_n$ represent the level of risk. Quantities $\rho_n$ will be specified later in this section.

Another important component of a financial market is the set of admissible actions or strategies that are allowed in dealing with assets $B$ and $S$. A sequence $\pi = (\pi_n)_{n=1,2,\ldots} = (\beta_n, \gamma_n)_{n=1,2,\ldots}$ is called an investment strategy (portfolio) if for any $n = 1, 2, \ldots$, the quantities $\beta_n$ and $\gamma_n$ are determined by prices $S_1, \ldots, S_{n-1}$. In other words, $\beta_n = \beta_n(S_1, \ldots, S_{n-1})$ and $\gamma_n = \gamma_n(S_1, \ldots, S_{n-1})$ are functions of $S_1, \ldots, S_{n-1}$, and they are interpreted
as the amounts of assets $B$ and $S$, respectively, at time $n$. The value of a portfolio $\pi$ is

$$X_\pi^n = \beta_n B_n + \gamma_n S_n,$$

where $\beta_n B_n$ represents the part of the capital deposited in a bank account and $\gamma_n S_n$ represents the investment in shares. If the value of a portfolio can change only because of changes in assets prices $\Delta X_\pi^n = X_\pi^n - X_\pi^{n-1} = \beta_n \Delta B_n + \gamma_n \Delta S_n$, then $\pi$ is said to be a self-financing portfolio. The class of all such portfolios is denoted $SF$.

A common feature of all derivative securities in a $(B, S)$-market is their potential liability (payoff) $f_N$ at a future time $N$. For example, for forwards, we have $f_N = S_N - F$ and for call options $f_N = (S_N - K)^+ = \max\{S_N - K, 0\}$. Such liabilities inherent in derivative securities are called contingent claims. One of the most important problems in the theory of contingent claims is their pricing at any time before the expiry date $N$. This problem is related to the problem of hedging contingent claims. A self-financing portfolio is called a hedge for a contingent claim $f_N$ if $X_\pi^N \geq f_N$ for any behavior of the market. If a hedging portfolio is not unique, then it is important to find a hedge $\pi^*$ with the minimum value $X_\pi^* \leq X_\pi^N$ for any other hedge $\pi$. Hedge $\pi^*$ is called the minimal hedge. The minimal hedge gives an obvious solution to the problem of pricing a contingent claim: the fair price of the claim is equal to the value of the minimal hedging portfolio. Furthermore, the minimal hedge manages the risk inherent in a contingent claim.

Next, we introduce some basic notions from probability theory and stochastic analysis that are helpful in studying risky assets. We start with the fundamental notion of an “experiment” when the set of possible outcomes of the experiment is known, but it is not known a priori which of those outcomes will take place (this constitutes the randomness of the experiment).

**Example 1.1 (Trading on a stock exchange)**

A set of possible exchange rates between the dollar and the euro is always known before the beginning of trading but not the exact value. □

Let $\Omega$ be the set of all elementary outcomes $\omega$ and let $\mathcal{F}$ be the set of all events (non-elementary outcomes), which contains the impossible event $\emptyset$ and the certain event $\Omega$.

Next, suppose that after repeating an experiment $n$ times, an event $A \in \mathcal{F}$ occurred $n_A$ times. Let us consider random experiments that have the following property of statistical stability: for any event $A$, there is a number $P(A) \in [0, 1]$ such that $n_A/n \to P(A)$ as $n \to \infty$. This number $P(A)$ is called the probability of event $A$. Probability $P : \mathcal{F} \to [0, 1]$ is a function with the following properties:

1. $P(\Omega) = 1$ and $P(\emptyset) = 0$;
2. $P(\cup_k A_k) = \sum_k P(A_k)$ for $A_i \cap A_j = \emptyset$. 

The triple \((\Omega, \mathcal{F}, P)\) is called a *probability space*. For the rest of this section, we assume that the set \(\Omega\) is countable. In this case, \((\Omega, \mathcal{F}, P)\) is referred to as a *discrete probability space*.

Every event \(A \in \mathcal{F}\) can be associated with its *indicator*:

\[
I_A(\omega) = \begin{cases} 
1, & \text{if } \omega \in A \\
0, & \text{if } \omega \in \Omega \setminus A 
\end{cases}
\]

Any function \(X : \Omega \to \mathbb{R}\) is called a *random variable*. An indicator is the simplest example of a random variable. Any random variable \(X\) on a discrete probability space is discrete since the range of function \(X(\cdot)\) is countable: \((x_k)_{k=1,2,...}\). In this case, we have the following representation:

\[
X(\omega) = \sum_{k=1}^{\infty} x_k I_{A_k}(\omega),
\]

where \(A_k \in \mathcal{F}\) and \(\bigcup_k A_k = \Omega\). A discrete random variable \(X\) is called *simple* if the corresponding sum is finite. The function

\[
F_X(x) := P(\{\omega : X \leq x\}), \quad x \in \mathbb{R}
\]

is called the *distribution function* (or *cumulative distribution function*) of \(X\). For a discrete \(X\), we have

\[
F_X(x) = \sum_{k : x_k \leq x} P(\{\omega : X = x_k\}) = \sum_{k : x_k \leq x} p_k.
\]

The sequence \((p_k)_{k=1,2,...}\) is called the *probability distribution* of a discrete random variable \(X\), and we have \(\sum_k p_k = 1\).

Note the following properties of the distribution function:

1. \(F_X(x)\) are non-decreasing functions of \(x\);
2. \(F_X(x) \searrow 0\) as \(x \to -\infty\) and \(F_X(x) \nearrow 1\) as \(x \to +\infty\).

The *expectation* (expected value or mean value) of \(X\) is

\[
E(X) = \sum_{k \geq 1} x_k p_k.
\]

Given a random variable \(X\), for most functions \(g : \mathbb{R} \to \mathbb{R}\) it is possible to define a random variable \(Y = g(X)\) with expectation

\[
E(Y) = \sum_{k \geq 1} g(x_k) p_k.
\]

In particular, for any \(k = 1, 2, \ldots\), the quantity \(E(X^k)\) is called the \(k\)-th moment of \(X\), and the quantity

\[
Var(X) = E\left[(X - E(X))^2\right]
\]

is called the *variance* of \(X\).

Note the following straightforward properties:
1. **Linearity of the expectation:** for any random variables \(X_1, \ldots, X_n\) and any constants \(c_1, \ldots, c_n\), we have
\[
E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i);
\]

2. For any random variable \(X\) and any constant \(c\),
\[
\text{Var}(cX) = c^2 \text{Var}(X).
\]

**Example 1.2 (Examples of discrete probability distributions)**

1. **Bernoulli:**
\[
p_0 = P(\{\omega : X = a\}) = p, \quad p_1 = P(\{\omega : X = b\}) = 1 - p,
\]
where \(p \in [0, 1]\) and \(a, b \in \mathbb{R}\).

2. **Binomial:**
\[
p_m = P(\{\omega : X = m\}) = \binom{n}{k} p^m (1-p)^{n-m},
\]
where \(p \in [0, 1]\), \(n \geq 1\), and \(m = 0, 1, \ldots, n\).

3. **Poisson (with parameter \(\lambda > 0\):**
\[
p_m = P(\{\omega : X = m\}) = e^{-\lambda} \frac{\lambda^m}{m!}
\]
for \(m = 0, 1, \ldots\).

Consider a positive random variable \(\tilde{Z}\) on a probability space \((\Omega, \mathcal{F}, P)\). Suppose that \(E(\tilde{Z}) = 1\), then, for any event \(A \in \mathcal{F}\), define its new probability:
\[
\tilde{P}(A) = E(\tilde{Z}\mathbb{I}_A). \tag{1.3}
\]

The expectation of a random variable \(X\) with respect to this new probability is \(\tilde{E}(X) = E(\tilde{Z}X)\), and this rule is referred to as *change of the probability measure under the expectation sign*. Random variable \(\tilde{Z}\) is called the *density* of the probability \(\tilde{P}\) with respect to \(P\). The proof of this formula is based on the linearity of the expectation:
\[
\tilde{E}(X) = \sum_k x_k \tilde{P}(\{\omega : X = x_k\}) = \sum_k x_k E(\tilde{Z}\mathbb{I}_{\{\omega : X = x_k\}})
\]
\[
= \sum_k E(\tilde{Z} x_k \mathbb{I}_{\{\omega : X = x_k\}}) = E(\tilde{Z} \sum_k x_k \mathbb{I}_{\{\omega : X = x_k\}})
\]
\[
= E(\tilde{Z}X).
\]
For discrete random variables $X$ and $Y$ with values $(x_i)_{i=1,2,...}$ and $(y_i)_{i=1,2,...}$, respectively, consider the probabilities

$$P\left(\{\omega \colon X = x_i, Y = y_i\}\right) = p_{ij}, \quad p_{ij} \geq 0, \quad \sum_{i,j} p_{ij} = 1.$$ 

These probabilities form the joint distribution of $X$ and $Y$. Denote $p_{i\bullet} = \sum_j p_{ij}$ and $p_{\bullet j} = \sum_i p_{ij}$, then random variables $X$ and $Y$ are called independent if $p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$, which implies that $E(XY) = E(X)E(Y)$.

The quantity

$$E(X|Y = y_i) := \sum_i x_ip_{ij}$$

is called the conditional expectation of $X$ with respect to the event $\{Y = y_i\}$. The random variable $E(X|Y)$ is called the conditional expectation of $X$ with respect to $Y$ if $E(X|Y)$ is equal to $E(X|Y = y_i)$ on every set $\{\omega : Y = y_i\}$. In particular, for indicators $X = I_A$ and $Y = I_B$, we obtain

$$E(X|Y) = P(A|B) = \frac{P(AB)}{P(B)},$$

which is called the conditional expectation of the event $A$ given the event $B$.

We mention some properties of conditional expectations:

1. $E(X) = E(E(X|Y))$, in particular, for $X = I_A$ and $Y = I_B$, we have $P(A) = P(B)P(A|B) + P(\Omega \setminus B)P(A|\Omega \setminus B)$;

2. If $X$ and $Y$ are independent, then $E(X|Y) = E(X)$;

3. Since by definition $E(X|Y)$ is a function of $Y$, then conditional expectation can be interpreted as a prediction of $X$ given the information from the “observed” random variable $Y$.

Finally, for a random variable $X$ with values in $\{0, 1, 2, \ldots\}$ we introduce the notion of a generating function

$$\phi_X(x) = E(x^X) = \sum_i x^ip_i.$$ 

We have

$$\phi(1) = 1, \quad \frac{d^k\phi}{dx^k}\bigg|_{x=0} = k!p_k$$

and

$$\phi_{X_1 + \ldots + X_k}(x) = \prod_{i=1}^k \phi_{X_i}(x)$$

for independent random variables $X_1, \ldots, X_k$. 

Example 1.1 (Trading on a stock exchange: Revisited)

Consider the following time scale: $n = 0$ (present time), \ldots, $n = N$ (can be one month, quarter, year, etc.).

An elementary outcome can be written in the form of a sequence $\omega = (\omega_1, \ldots, \omega_N)$, where $\omega_i$ is an elementary outcome representing the results of trading at time step $i = 1, \ldots, N$. Now we consider a probability space $(\Omega, \mathcal{F}_N, P)$ that contains all trading results up to time $N$. For any $n \leq N$, we also introduce the corresponding probability space $(\Omega, \mathcal{F}_n, P)$ with elementary outcomes $(\omega_1, \ldots, \omega_n) \in \mathcal{F}_n \subseteq \mathcal{F}_N$.

Thus, to describe the evolution of trading on a stock exchange, we need a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ called a stochastic basis, where $\mathcal{F} = (\mathcal{F}_n)_{n \leq N}$ is called a filtration (or information flow):

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_N.$$

By definition, the set $\mathcal{F}_n$ contains all trading information up to time $n$; therefore, if $A \in \mathcal{F}_n \subset \mathbb{P}$, then $\mathcal{F}_n$ also contains the complement of $A$ and it is closed under taking countable unions and intersections. The sets of events that satisfy such properties are $\sigma$-algebras, and we will discuss them in detail later in this chapter.

Now consider a $(B, S)$-market. Since asset $B$ is non-risky, we can assume that $B(\omega) \equiv B_n$ for all $\omega \in \Omega$. For a risky asset $S$, it is natural to assume that prices $S_1, \ldots, S_N$ are random variables on stochastic basis $(\Omega, \mathcal{F}_N, \mathbb{F}, P)$. Each of $S_n$ is completely determined by the trading results up to time $n \leq N$ or, in other words, by the set of events $\mathcal{F}_n$. We also assume that the sources of trading randomness are exhausted by the stock prices; that is, $\mathcal{F}_n = \sigma(S_1, \ldots, S_n)$ is a set generated by random variables $S_1, \ldots, S_n$.

Let us consider a specific example of a $(B, S)$-market. Let $\rho_1, \ldots, \rho_N$ be independent random variables taking values $a$ and $b$ ($a < b$) with probabilities $P(\{\omega: \rho_k = b\}) = p$ and $P(\{\omega: \rho_k = a\}) = 1 - p \equiv q$, respectively. Define the stochastic basis: $\Omega = \{a, b\}^N$ is the space of sequences of length $N$ whose elements are equal to either $a$ or $b$; $\mathcal{F} = 2^\Omega$ is the set of all subsets of $\Omega$. The filtration $\mathbb{F}$ is generated by the prices $(S_n)$ or equivalently by the sequence $(\rho_n)$:

$$\mathcal{F}_n = \sigma(S_1, \ldots, S_n) = \sigma(\rho_1, \ldots, \rho_n),$$

which means that every random variable on the probability space $(\Omega, \mathcal{F}_n, P)$ is a function of $S_1, \ldots, S_n$ or, equivalently, of $\rho_1, \ldots, \rho_n$ due to relations

$$\frac{\Delta S_k}{S_{k-1}} = \rho_k, \quad k = 1, 2, \ldots.$$

A financial $(B, S)$-market defined on this stochastic basis is called binomial.

Consider a contingent claim $f_N$. Since its maturity date is $N$, then, in general, $f_N = f(S_1, \ldots, S_N)$ is a function of all “history” $S_1, \ldots, S_N$. The key problem now is to estimate (or predict) $f_N$ at any time $n \leq N$.
given the available market information $\mathcal{F}_n$. We would like these predictions $E(f_N|\mathcal{F}_n)$, $n = 0, 1, \ldots, N$, to have the following intuitively natural properties:

1. $E(f_N|\mathcal{F}_n)$ is a function of $S_1, \ldots, S_n$, but not of future prices $S_{n+1}, \ldots, S_N$;

2. A prediction based on the trivial information $\mathcal{F}_0 = \{\emptyset, \Omega\}$ should coincide with the mean value of a contingent claim: $E(f_N|\mathcal{F}_0) = E(f_N)$;

3. Predictions must be compatible:

   $E\left(E(f_N|\mathcal{F}_n)\right) = E\left(E(f_N|\mathcal{F}_{n+1})\right|\mathcal{F}_n)$,

   in particular

   $E\left(E(f_N|\mathcal{F}_n)\right) = E\left(E(f_N|\mathcal{F}_{n})\right|\mathcal{F}_0) = E(f_N)$;

4. A prediction based on all possible information $\mathcal{F}_N$ should coincide with the contingent claim: $E(f_N|\mathcal{F}_N) = f_N$;

5. Linearity:

   $E(\phi f_N + \psi g_N|\mathcal{F}_n) = \phi E(f_N|\mathcal{F}_n) + \psi E(g_N|\mathcal{F}_n)$

   for $\phi$ and $\psi$ defined by the information in $\mathcal{F}_n$;

6. If $f_N$ does not depend on the information in $\mathcal{F}_n$, then a prediction based on this information should coincide with the mean value

   $E(f_N|\mathcal{F}_n) = E(f_N)$;

7. Denote $f_n = E(f_N|\mathcal{F}_n)$; then from Property 3, we obtain

   $E(f_{n+1}|\mathcal{F}_n) = E\left(E(f_N|\mathcal{F}_{n+1})\right|\mathcal{F}_n) = E(f_N|\mathcal{F}_n) = f_n$

   for all $n \leq N$. Such stochastic sequences are called martingales.

Now we consider how to calculate predictions. Comparing the notions of a conditional expectation and a prediction, we see that a prediction of $f_N$ based on $\mathcal{F}_n = \sigma(S_1, \ldots, S_n)$ is equal to the conditional expectation of a random variable $f_N$ with respect to random variables $S_1, \ldots, S_n$.

**Worked Example 1.2** Suppose that the monthly price evolution of stock $S$ is given by

$$S_n = S_{n-1}(1 + \rho_n), \quad n = 1, 2, \ldots,$$

where profitabilities $\rho_n$ are independent random variables taking values $0.2$ and $-0.1$ with probabilities $0.4$ and $0.6$, respectively. Given that the current price $S_0 = 200$ ($\$$), find the predicted mean price of $S$ for the next two months.
Solution Since 
\[
E(\rho_1) = E(\rho_2) = 0.2 \cdot 0.4 - 0.1 \cdot 0.6 = 0.02 ,
\]
then 
\[
E\left(\frac{S_1 + S_2}{2} \bigg| S_0 = 200\right) = E\left(\frac{S_0(1 + \rho_1) + S_0(1 + \rho_1)(1 + \rho_2)}{2} \bigg| S_0 = 200\right) \\
= \frac{S_0}{2} \left[ E(1 + \rho_1) + E(1 + \rho_1)E(1 + \rho_2) \right] \\
= \frac{100}{2} \left[ 1.02 + 1.02 \cdot 1.02 \right] = 206.4 .
\]

We finish this section noting that there are various indications that the use of discrete probability spaces can significantly limit the class of stochastic experiments available for stochastic modeling. Below, we discuss one of most illustrative considerations of this nature.

Let function \( f(x), x \in \mathbb{R}, \) be non-negative with 
\[
\int_{-\infty}^{\infty} f(x) \, dx = 1 .
\]
Then function 
\[
F(x) = \int_{-\infty}^{x} f(y) \, dy
\]
satisfies properties (D1)–(D2) of a distribution function: \( F \) is a non-decreasing function of \( x \) such that \( F(x) \downarrow 0 \) as \( x \to -\infty \) and \( F(x) \uparrow 1 \) as \( x \to +\infty \). In this case, function \( f \) is referred to as density of the distribution function \( F \).

For example, function 
\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]
defines a normal distribution function \( F \) with parameters \( \mu \in \mathbb{R} \) and \( \sigma \in \mathbb{R}^+ \). In particular, a standard normal distribution is a normal distribution with \( \mu = 0 \) and \( \sigma^2 = 1 \). Usually, it is denoted \( \Phi \) or \( \mathcal{N}(0,1) \).

One can pose a natural question: given an arbitrary function \( F \) that satisfies all properties of a distribution function, is there a probability space and a random variable \( X \) such that \( F \) is the distribution function of \( X \)?

The following example gives a positive answer to this question and illustrates that non-discrete probability spaces and non-discrete random variables (that are of particular importance in stochastic modeling) exist.

Consider the Borel algebra \( \mathcal{B}([0,1]) \). It is the set of all Borel subsets of the interval \([0,1] \), which contains all possible subintervals of \([0,1] \), their complements and countable unions and intersections, and therefore it is a \( \sigma \)-algebra. For elements of \( \mathcal{B}([0,1]) \) we can introduce a (unique) Lebesgue measure \( m \), that, for intervals, is equal to their length. Let \( \Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]), \) and \( P = m, \) then \( (\Omega, \mathcal{F}, P) \) is a probability space. We now define a random variable \( X : [0,1] \to \mathbb{R} \) in the following way: \( X(0) = X(1) = 0; \) further for each \( x \in (0,1), \) let \( \omega = F(x) \) and define \( X(\omega) = x \). Then the probability of \( \{ \omega : X(\omega) \leq x \} \) is equal to the length of the interval \([0,F(x)] = F(x) \).

In the next section, we discuss in detail the fundamental general notion of a probability space that facilitates the quantitative description of numerous statistically stable stochastic experiments.
1.3 Elements of probability theory and stochastic analysis

One of the pivotal foundations of modern probability theory is the notion of a *probability space* \((\Omega, \mathcal{F}, P)\), where

- a set \(\Omega\) is the *space of elementary outcomes* \(\omega\);
- the *space of all events* is represented by \(\mathcal{F}\), the set of all subsets of \(\Omega\), including \(\emptyset\) and \(\Omega\), which also contains all their complements and countable unions and intersections (\(\sigma\)-algebra);
- a *probability measure* (or simply, *probability*) \(P\) is a function \(P : \mathcal{F} \rightarrow [0, 1]\) that satisfies the following properties:
  1. \(P(\Omega) = 1, P(\emptyset) = 0\);
  2. \(P(\Omega \setminus A) = 1 - P(A)\) for all \(A \in \mathcal{F}\);
  3. \(P(\bigcup_{k} A_k) = \sum_{k} P(A_k)\) for any sequence of pairwise disjoint sets \(A_k \in \mathcal{F}\).

This construction is often referred to as *Kolmogorov’s axiomatic formulation* of probability theory. For detailed accounts of its various aspects, we refer the reader to standard probability textbooks.

Any pair \((\Omega, \mathcal{F})\) in this construction is called a *measurable space*. In particular, the pair \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) is called the *Borel space*.

We say that function \(X : \Omega \rightarrow \mathbb{R}\) is *measurable* if the inverse image of any Borel set \(B \in \mathcal{B}(\mathbb{R})\) is in \(\mathcal{F}\):

\[ X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}. \]

Such functions are called *random variables*. The *indicator*

\[ X(\omega) = I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}, \quad A \in \mathcal{F}, \]

and linear combinations of indicators

\[ X(\omega) = \sum_{k \geq 1} x_k I_{A_k}(\omega), \quad A_k \in \mathcal{F}, \ x_k \in \mathbb{R} \]

are the simplest examples of random variables.

A random variable \(X\) induces a probability measure on the Borel space \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) defined by

\[ P_X(B) := P\{\omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}). \]
For sets $B = (-\infty, x]$, $x \in \mathbb{R}$, then we can introduce a function
\[ F(x) = F_X(x) = P_X((-\infty, x]), \quad x \in \mathbb{R}, \quad (1.4) \]
which is called the distribution function of random variable $X$. This definition generalizes the notion of the distribution function for a discrete random variable introduced by formula (1.1), and if $X$ is discrete, then function (1.4) coincides with (1.2). Note that function $F$ satisfies properties (D1)–(D2) of a distribution function and that it is right continuous.

A large class of non-discrete random variables can be characterized by absolutely continuous distribution functions of the integral form:
\[ F(x) = \int_{-\infty}^{x} f(y) \, dy, \]
where non-negative function $f$ is called the density of a random variable. Here, we mention some useful examples of such distributions:

1. **Uniform** on $[a, b]$ distribution:
\[ f(y) = \begin{cases} 1/(b-a), & y \in [a,b] \\ 0, & y \notin [a,b] \end{cases}; \]

2. **Normal** (Gaussian) distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$:
\[ f(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}, \quad y \in \mathbb{R}; \]

3. **Gamma** distribution with parameters $\alpha > 0$ and $\beta > 0$:
\[ f(y) = \frac{y^{\alpha-1} e^{-y/\beta}}{\Gamma(\alpha) \beta^\alpha}, \quad y \geq 0, \]
where $\Gamma$ is the gamma function. In particular, if $\alpha = 1$ and $\beta = 1/\lambda$, then this distribution is referred to as exponential distribution with parameter $\lambda$. Its density has the form
\[ f(y) = \lambda e^{-\lambda y}, \quad y \geq 0, \quad \lambda > 0; \]

4. **Student** distribution:
\[ f(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \left( 1 + \frac{y^2}{n} \right)^{-\frac{n+1}{2}}, \quad y \in \mathbb{R}, \]
with degrees of freedom $n = 1, 2, \ldots$;

5. **$\chi^2$** distribution:
\[ f(y) = \frac{y^{n/2-1} e^{-y/2}}{\Gamma(n/2) 2^{n/2}}, \quad y \geq 0, \quad n = 1, 2, \ldots, \]
which is a gamma distribution with parameters $\alpha = n/2$ and $\beta = 2$. 


Given a random variable $X$, one can introduce various random variables associated with it in the following way. Consider an arbitrary measurable (Borel) function $\varphi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$; then the composition of functions $X$ and $\varphi$,

$$Y = \varphi \circ X,$$  

is a random variable, since, for any $B \in \mathcal{B}(\mathbb{R}),$ we have

$$\{\omega : Y(\omega) \in B\} = \{\omega : \varphi(X(\omega)) \in B\} = \{\omega : X(\omega) \in \varphi^{-1}(B)\} \in \mathcal{F}.$$  

In particular, we can introduce random variables $X^+ := \max\{0, X\}, X^- := \max\{0, -X\}, |X| = X^+ + X^-, X^k$, and so forth.

Denote $\mathcal{F}^X$ the set of all random events of the type $\{\omega : X(\omega) \in B\}$ for all possible $B \in \mathcal{B}(\mathbb{R})$. It is not difficult to show that $\mathcal{F}^X$ is a $\sigma$-algebra. This referred to as $\sigma$-algebra generated by random variable $X$. We also note that, if $Y$ as a random variable that is measurable with respect to the $\sigma$-algebra $\mathcal{F}^X$, then there exists a Borel function $\varphi$ such that the representation (1.5) holds for $Y$. Later, we will discuss martingale representations as dynamic versions of this fact.

In the previous section, in the case of discrete random variables on a discrete probability space, we saw that distribution functions and expectations are the key quantitative characteristics of random variables. We now introduce the notion of expectation in the non-discrete case. First, we consider a discrete random variable $X$ on probability space $(\Omega, \mathcal{F}, P)$. Suppose it can attain the finite number of values: $x_1, \ldots, x_n$, then we can write

$$X(\omega) = \sum_{k=1}^{n} x_k I_{A_k}(\omega),$$  

where $A_k = \{\omega : X(\omega) = x_k\}, k = 1, 2, \ldots, n,$ and $\bigcup_{k=1}^{n} A_k = \Omega$.

As in the case of a discrete probability space, we define the expectation of $X$ by

$$E(X) = \sum_{k=1}^{n} x_k P(A_k).$$

Now let $X$ be an arbitrary non-negative random variable on $(\Omega, \mathcal{F}, P)$. We define a sequence $(X_n)_{n=1,2,\ldots}$ of discrete random variables with finite numbers of values by

$$X_n(\omega) = \sum_{k=1}^{2^n} \frac{k - 1}{2^n} I_{\left\{\omega : \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n}\right\}}(\omega) + n I_{\left\{\omega : X(\omega) \geq n\right\}}(\omega).$$

Clearly, $X_n(\omega) \nearrow X(\omega)$ for each $\omega \in \Omega$. Since sequence $(E(X_n))_{n=1,2,\ldots}$ is non-decreasing, then there exists the limit

$$E(X) := \lim_{n \to \infty} E(X_n),$$
which is called the expectation of $X$. Note that this limit is not necessarily finite.

An arbitrary random variable $X$ on $(\Omega, \mathcal{F}, P)$ can be written in the form $X = X^+ - X^-$, where $X^+$ and $X^-$ are non-negative random variables. If at least one of the expectations is finite, we define

$$E(X) = E(X^+) - E(X^-).$$

If both $E(X^+) < \infty$ and $E(X^-) < \infty$, then

$$E(|X|) = E(X^+) + E(X^-) < \infty,$$

and we say that $X$ is a random variable with finite expectation.

This construction of the expectation is identical to the definition of the Lebesgue integral for a measurable function $X$, and the following integral notation is also common in the probability theory:

$$E(X) = \int_{\Omega} X \, dP = \int_{\Omega} X(\omega) P(d\omega).$$

Because of construction, expectations are linear and monotonic. Change of variables in the Lebesgue integral allows the following representation of expectation in terms of the distribution function of $X$:

$$E(\varphi(X)) = \int_{\Omega} \varphi(X) \, dP = \int_{\mathbb{R}} \varphi(x) P_X(dx) = \int_{\mathbb{R}} \varphi(x) dF_X(x),$$

where $\varphi$ is a Borel function that is integrable with respect to $P_X$.

For $\varphi(x) = x^k$, $k = 1, 2, \ldots$, the expectation $E(X^k)$ is called the $k$-th moment of random variable $X$. Suppose $E(X) = \mu$ and let $\varphi(x) = (x - \mu)^k$, then the corresponding moments are called centered moments. The second centered moment is called the variance of $X$: $\text{Var}(X) = E((X - \mu)^2)$, and it is one of the key measures of the dispersion of values of $X$ about the mean value $\mu$. The common additional measures are skewness:

$$S = \frac{E((X - \mu)^3)}{(\text{Var}(X))^{3/2}},$$

and kurtosis:

$$K = \frac{E((X - \mu)^4)}{(\text{Var}(X))^2}.$$

If a probability space $(\Omega, \mathcal{F}, P)$ is non-discrete, then some further important properties of expectations involve the notion of $P$-almost surely or simply almost surely. We say that a property holds almost surely (we write a.s.) if there is a set $N \in \mathcal{F}$ such that the probability (measure) of $N$ is zero and the property holds for all $\omega \in \Omega \setminus N$. The following properties can be readily derived:
1. If \( X = 0 \) (a.s.), then \( E(X) = 0 \);

2. If \( X = Y \) (a.s.) and \( E|X| < \infty \), then \( E|Y| < \infty \) and \( E(X) = E(Y) \);

3. If \( X \geq 0 \) and \( E(X) = 0 \), then \( X = 0 \) (a.s.);

4. If \( E|X| < \infty \), \( E|Y| < \infty \), and \( E(X I_A) \leq E(Y I_A) \) for all \( A \in \mathcal{F} \), then \( X \leq Y \) (a.s.).

Let \( (X_n)_{n=1,2,...} \) be a sequence of arbitrary random variables on a non-discrete probability space \((\Omega, \mathcal{F}, P)\). Suppose that \( X_n \to X \) (a.s.) as \( n \to \infty \), where \( X \) is some random variable. We are now interested in conditions that can guarantee that \( E(X_n) \) converges to \( E(X) \). The most general condition of this nature is the uniform integrability of the sequence \( (X_n)_{n=1,2,...} \):

\[
\lim_{c \to \infty} \sup_n E|X_n| \cdot I_{\left\{ \omega : |X_n(\omega)| > c \right\}} = 0.
\]

The following result is fundamental.

**Theorem 1.1** Let \( (X_n)_{n=1,2,...} \) be a uniformly integrable sequence of random variables such that \( X_n \to X \) (a.s.) as \( n \to \infty \). Then the limit random variable \( X \) has finite expectation. Furthermore, \( \lim_{n \to \infty} E(X_n) = E(X) \) and \( \lim_{n \to \infty} E|X_n - X| = 0 \) (convergence in \( L^1 \)).

Often there is a need for modeling with a \( n \)-dimensional random variable \( X = (X_1, \ldots, X_n) \). Denote \( B(\mathbb{R}^n) \) the Borel algebra on \( \mathbb{R}^n \), then the random vector \( X \) can be defined as a measurable mapping from a measurable space \((\Omega, \mathcal{F})\) into the Borel space \((\mathbb{R}^n, B(\mathbb{R}^n))\). A natural quantitative probabilistic characteristic of the random vector \( X \) is the \( n \)-dimensional distribution function

\[
F_X(x_1, \ldots, x_n) = P(\omega : X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,
\]

which is also referred to as a joint distribution function of random variables \( X_1, \ldots, X_n \).

Random variables \( X_1, \ldots, X_n \) are said to be independent if \( F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \times \ldots \times F_{X_n}(x_n) \), where \( F_{X_1}, \ldots, F_{X_n} \) are the corresponding one-dimensional distribution functions. Similarly to the one-dimensional case, a \( n \)-dimensional distribution function \( F_X \) generates a probability measure \( P_X \) on \((\mathbb{R}^n, B(\mathbb{R}^n))\) such that

\[
P_X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) = F_X(x_1, \ldots, x_n),
\]

which is called a \( n \)-dimensional distribution. We denote it \( P_n \), and we note its consistency property:

\[
P_n(B) = P_{n+1}(B \times \mathbb{R}), \quad n = 1, 2, \ldots
\]

(1.6)
for all $B \in \mathcal{B}(\mathbb{R}^n)$.

Many important facts in probability theory are related to infinite families of random variables. A sequence of random variables $(X_n)_{n=1,2,...}$ is one of the typical examples of such families. Very often it is natural to assume that finite dimensional distributions $P_k$ of $X_1,\ldots,X_k$ are known for all $k = 1,2,\ldots$. In this case, it is important to consider the question of the existence of a probability space that will accommodate sequence $(X_n)_{n=1,2,...}$. The affirmative answer to this question is given by the celebrated Kolmogorov’s consistency theorem, where the consistency property of finite dimensional distributions becomes the consistency condition on finite dimensional distributions. Note that the example that we discussed at the end of the previous section, which motivates the study of non-discrete probability spaces and non-discrete random variables, is a simple consequence of Kolmogorov’s theorem.

The notion of a joint distribution of random variables $X_1,\ldots,X_n$ allows one to compute quantitative characteristics of random variables that are functions of $X_1,\ldots,X_n$. We illustrate this in the case of two random variables.

**Theorem 1.2** Let $X_1$ and $X_2$ be independent random variables with finite expectations. Then

$$E(X_1X_2) = E(X_1)E(X_2). \quad (1.7)$$

**Proof** We need to prove this statement only for non-negative random variables since the general case then follows from the following relations:

$$X_i = X_i^+ - X_i^-,$$

and

$$X_1X_2 = X_1^+X_2^+ - X_1^-X_2^+ - X_1^+X_2^- + X_1^-X_2^-.$$

Let

$$X_{i,n} := \sum_{k=0}^{\infty} \frac{k}{n^2} I_{\left\{ \omega: \frac{k}{n^2} \leq X_i(\omega) < \frac{k+1}{n^2} \right\}}, \quad i = 1,2,$$

and note that by construction we have

$$X_{i,n} \leq X_i \quad \text{and} \quad |X_{i,n} - X_i| \leq \frac{1}{n}, \quad i = 1,2.$$

Since discrete random variables $X_{1,n}$ and $X_{2,n}$ are independent, we obtain

$$E(X_{1,n}X_{2,n}) = \sum_{k,l=0}^{\infty} \frac{k}{n^2} E\left( I_{\left\{ \omega: \frac{k}{n^2} \leq X_1(\omega) < \frac{k+1}{n^2} \right\}} I_{\left\{ \omega: \frac{l}{n^2} \leq X_2(\omega) < \frac{l+1}{n^2} \right\}} \right)$$

$$= \sum_{k,l=0}^{\infty} \frac{k}{n^2} E\left( I_{\left\{ \omega: \frac{k}{n^2} \leq X_1(\omega) < \frac{k+1}{n^2} \right\}} \right) E\left( I_{\left\{ \omega: \frac{l}{n^2} \leq X_2(\omega) < \frac{l+1}{n^2} \right\}} \right)$$

$$= E(X_{1,n})E(X_{2,n}).$$
By Theorem 1.1, we have
\[
E(X_{i,n}) \to E(X_i) \quad \text{as} \ n \to \infty, \ i = 1, 2,
\]
and therefore
\[
\left| E(X_1 X_2) - E(X_{1,n} X_{2,n}) \right| \leq E(\left| X_1 X_2 - X_{2,n} X_{2,n} \right|)
\leq E(\left| X_1 \right| | X_2 - X_{2,n} |) + E(\left| X_{2,n} \right| | X_1 - X_{1,n} |)
\leq \frac{1}{n} E(\left| X_1 \right|) + \frac{1}{n} E(\left| X_2 + \frac{1}{n} \right|) \to 0 \quad \text{as} \ n \to \infty.
\]
Thus,
\[
E(X_1 X_2) = \lim_{n \to \infty} E(X_{1,n} X_{2,n}) = \lim_{n \to \infty} E(X_{1,n}) \lim_{n \to \infty} E(X_{2,n})
= E(X_1) E(X_2). \quad \square
\]

In the context of this result, it is useful to introduce the notion of covariance of random variables \(X_1\) and \(X_2\) with expectations \(\mu_1\) and \(\mu_2\), respectively:
\[
\text{Cov}(X_1 X_2) = E((X_1 - \mu_1)(X_2 - \mu_2)),
\]
and the notion of the correlation coefficient of \(X_1\) and \(X_2\):
\[
\text{Cor}(X_1 X_2) = \frac{\text{Cov}(X_1 X_2)}{\sqrt{\text{Var}(X_1)} \sqrt{\text{Var}(X_2)}}.
\]

If random variables \(X_1\) and \(X_2\) are independent, then Theorem 1.2 implies that \(\text{Cov}(X_1 X_2) = \text{Cor}(X_1 X_2) = 0\). In this case, we say that \(X_1\) and \(X_2\) are uncorrelated. Note that the inverse is not true: the uncorrelated random variables are not necessarily independent. We also note here that for random variables that are not necessarily independent, equality (1.7) can be replaced with an appropriate inequality. The following Cauchy-Bunyakovskiy inequality is a classical example of such inequalities.

**Theorem 1.3** Let \(X_1\) and \(X_2\) be random variables with finite second moments. Then
\[
E(\left| X_1 X_2 \right|) \leq \sqrt{E(X_1^2) E(X_2^2)}.
\]

Note that the Cauchy-Bunyakovskiy inequality immediately implies that \(|\text{Cor}(X_1 X_2)| \leq 1\).

The next property of expectations is formulated for random variables that are defined via formula (1.5): \(Y = \varphi \circ X\). Let the Borel function \(\varphi\) be convex downward; that is, suppose that for each \(x_0\) there exists a number \(\lambda = \lambda(x_0)\) such that
\[
\varphi(x) \geq \varphi(x_0) + (x - x_0) \lambda(x_0)
\quad \text{(1.8)}
\]
for all \(x \in \mathbb{R}\). We arrive at the following Jensen’s inequality.
Theorem 1.4 Suppose that \( \varphi \) is a convex downward Borel function and a random variable \( X \) is such that the expectations \( E(|X|) \) and \( E(\varphi(X)) \) are finite. Then

\[
\varphi(E(X)) \leq E(\varphi(X)).
\]

Proof Let \( x_0 = E(X) \), then inequality (1.8) with \( x = X \) implies

\[
\varphi(X) \geq \varphi(E(X)) + (X - E(X)) \lambda(E(X)).
\]

Using linearity and monotonicity of expectations, we conclude

\[
E(\varphi(X)) \geq \varphi(E(X)) + \lambda(E(X)) E(X - E(X)) = \varphi(E(X)). \quad \square
\]

The last property of expectations that we wish to mention in this section is called the Chebyshev inequality.

Theorem 1.5 Let \( X \) be a non-negative random variable with finite expectation. Then for any \( \varepsilon > 0 \),

\[
P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}. \quad (1.9)
\]

If \( X \) is an arbitrary random variable with finite variance, then

\[
P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}. \quad (1.10)
\]

Proof Inequality (1.9) holds true since

\[
P(X \geq \varepsilon) = E \left( I \{ \omega : X(\omega) \geq \varepsilon \} \right) \leq \frac{E \left( I \{ \omega : X(\omega) \geq \varepsilon \} \right)}{\varepsilon} \leq \frac{E(X)}{\varepsilon}.
\]

Inequality (1.10) is in fact the inequality (1.9) when applied to the non-negative random variable \( (X - E(X))^2 \) with \( \varepsilon^2 > 0 \). \( \square \)

In order to formulate one of the most important corollaries of inequalities (1.9)–(1.10), we need to introduce the following type of convergence. We say that sequence \( (X_n)_{n=1,2,...} \) converges in probability to a random variable \( X \) if for any \( \varepsilon > 0 \)

\[
P(|X_n - X| \geq \varepsilon) \to 0 \quad \text{as} \quad n \to \infty.
\]

Consider a sequence \( (X_n)_{n=1,2,...} \) of independent identically distributed random variables with expectations \( \mu \) and variances \( \sigma^2 \). Define a new sequence

\[
\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k,
\]

Financial Risk Management and Related Mathematical Tools
where, clearly, \( E(X_n) = \mu \) and \( \text{Var}(X_n) = \sigma^2/n \). Then for any fixed \( \varepsilon > 0 \), inequality (1.10) implies

\[
P(|X_n - \mu| \geq \varepsilon) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \to 0 \quad \text{as } n \to \infty.
\]

Thus, the sequence of the empirical means \( \overline{X}_n \) converges in probability to the theoretical mean \( \mu \). This result is usually referred to as the law of large numbers (LLN).

We also mention here the following statistical application of the law of large numbers. Suppose that parameters \( \mu \) and \( \sigma^2 \) of sequence \((X_k)_{k=1,2,...}\) are unknown and we wish to estimate them from a finite subsequence \( X_1, \ldots, X_n \). It turns out that \( X_n \) and

\[
s^2_n = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2
\]

are adequate approximations for \( \mu \) and \( \sigma^2 \), respectively. These approximations are not biased since \( E(\overline{X}_n) = \mu \) and \( E(s^2_n) = \sigma^2 \), and the law of large numbers guarantees that \( \overline{X}_n \) and \( s^2_n \) converge in probability to \( \mu \) and \( \sigma^2 \), respectively.

Recall, that so far we have discussed three types of convergence of a sequence \((X_n)_{n=1,2,...}\):

1. convergence almost surely;
2. convergence in mean; and
3. convergence in probability.

Note that these three types of convergence do not involve distributions of random variables \( X_n \). Is it possible to introduce a type of convergence that would involve only distributions of random variables \( X_n \)? The answer to this question is affirmative, and it reflects one of the key essential ideas of probability theory.

We say that a sequence of random variables \((Y_n)_{n=1,2,...}\) with distribution functions \((F_{Y_n})_{n=1,2,...}\) converges in distribution to a random variable \( Y \) with a continuous distribution function \( F_Y \), if for all \( x \in \mathbb{R} \)

\[
F_{Y_n}(x) \to F_Y(x) \quad \text{as } n \to \infty.
\]

We now consider again a sequence \((X_n)_{n=1,2,...}\) of independent identically distributed random variables with expectations \( \mu \) and variances \( \sigma^2 \). Define

\[
Y_n = \frac{\sum_{k=1}^{n} X_k - \mu n}{\sigma \sqrt{n}}, \quad n = 1, 2, \ldots.
\]
Sequence \((Y_n)_{n=1,2,...}\) converges in distribution to a standard normal random variable \(Y = \mathcal{N}(0,1)\); that is, for all \(x \in \mathbb{R}\),

\[
F_{Y_n}(x) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy \quad \text{as} \quad n \to \infty.
\]

Because of the extreme importance of this result in probability theory and statistics, it (and its various modifications) is referred to as the central limit theorem (CLT).

Independent Bernoulli random variables

\[
X_n = \begin{cases} 
1 & \text{with probability } p \in [0,1] \\
0 & \text{with probability } q = 1 - p , \, n = 1, 2, \ldots 
\end{cases},
\]

are often used in our discussions throughout this book. Denote \(S_n = X_1 + \ldots + X_n\), then since for all \(k = 1, 2, \ldots, \), \(E(X_k) = p\) and \(Var(X_k) = p \cdot q\) for such variables, the central limit theorem takes the following form

\[
\frac{S_n - np}{\sqrt{npq}} \to \mathcal{N}(0,1) \quad \text{in distribution}.
\]

This result is usually referred to as integral De Moivre-Laplace theorem and it follows readily from the following local De Moivre-Laplace theorem:

\[
B_n(k) = P(S_n = k) = \binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}
\]

for \(k = 0, 1, \ldots, n \to \infty\), or denoting \(k = np + x\sqrt{npq} = 0, 1, \ldots, n\), we can write\(^1\)

\[
B_n(k) \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}}.
\]

Note that the above approximations with normal distribution and normal density are at least of the order \(n^{-1/2}\).

The general notion of conditional expectation is essential in modern probability theory. Its rigorous introduction involves the use of absolutely continuous measures. Let \((\Omega, \mathcal{F}, P)\) be a probability space and consider a \(\sigma\)-algebra \(\mathcal{G} \subseteq \mathcal{F}\) which is a sub-algebra of the original \(\sigma\)-algebra \(\mathcal{F}\).

Consider a non-negative random variable \(X\). Denote \(E(X|\mathcal{G})\) a random variable on \((\Omega, \mathcal{F}, P)\) that is measurable with respect to the smaller \(\sigma\)-algebra \(\mathcal{G}\) and that

\[
E(X I_G) = E(E(X|\mathcal{G}) \cdot I_G) \quad \text{for any } G \in \mathcal{G}.
\]

Random variable \(E(X|\mathcal{G})\) is called the conditional expectation of random variable \(X\) with respect to the \(\sigma\)-algebra \(\mathcal{G}\). It is well defined in the following sense:

\(^1\)We say that functions \(f\) and \(g\) are equivalent: \(f \sim g\), if the limit of their ratio is equal to one.
if \( Y' \) and \( Y'' \) are two \( \sigma \)-measurable random variables that satisfy equality (1.11), then they are almost surely equal. Indeed, from (1.11) we have

\[
E((Y' - Y'') \cdot I_G) = 0 \quad \text{for any } G \in \mathcal{G}.
\]

Consider sets

\[
G' = \{ \omega : Y' - Y'' > 0 \} \quad \text{and} \quad G'' = \{ \omega : Y' - Y'' < 0 \},
\]

then

\[
E((Y' - Y'') \cdot I_{G'}) = 0 = E((Y' - Y'') \cdot I_{G''})
\]

or \( E|Y' - Y''| = 0 \), and therefore

\[
P\{\{ \omega : Y'(\omega) \neq Y''(\omega) \}\} = 0.
\]

In order to explore the existence of a conditional expectation, we now introduce the notion of absolutely continuous measures. We say that measure \( Q \) is absolutely continuous with respect to measure \( P \) (we write \( Q \ll P \)) if for any set \( A \in \mathcal{F} \), \( P(A) = 0 \) implies that \( Q(A) = 0 \). For example, if \( Z \) is a non-negative random variable on \((\Omega, \mathcal{F}, P)\) with finite expectation, then measure \( Q \), defined by the equality \( Q(A) = E(Z I_A) \), is absolutely continuous with respect to measure \( P \). It turns out the inverse statement holds true, and it is known as the Radon-Nikodým theorem: for any measure \( Q \ll P \), there exists a measurable function \( Z = \frac{dQ}{dP} \) (the density), such that

\[
Q(A) = \int_A Z \, dP. \tag{1.12}
\]

Applying this result to measure \( Q \) with \( Q(G) = E(X I_G) \) for all \( G \in \mathcal{G} \), we conclude that there exists a \( \mathcal{G} \)-measurable density \( Z \), which is indeed the conditional expectation \( E(X|\mathcal{G}) \).

This construction can be extended to an arbitrary random variable \( X \) in the following standard way. We use the decomposition \( X = X^+ - X^- \), and assuming that at least one of the expectations \( E(X^\pm|\mathcal{G}) \) is finite, we define

\[
E(X|\mathcal{G}) = E(X^+|\mathcal{G}) - E(X^-|\mathcal{G}).
\]

The classical notion of conditional expectation of a random variable \( X \) with respect to a random variable \( Y \) also can be obtained from the construction above:

\[
E(X|Y) := E(X|\mathcal{F}^Y),
\]

where \( \mathcal{F}^Y \) is a \( \sigma \)-algebra generated by \( Y \). Furthermore, if we take \( X = I_A \) and \( Y = I_B \), then we arrive at the definition of conditional probability \( P(A \mid B) := E(I_A | I_B) \).

Let us list some useful properties of conditional expectations.

1. If random variable \( X \) is constant, then \( E(X|\mathcal{G}) = X \) (a.s.);
2. If $X \leq Y$ (a.s.), then
\[ E(X|\mathcal{G}) \leq E(Y|\mathcal{G}) \] (a.s.),
and in particular, $|E(X|\mathcal{G})| \leq E(|X||\mathcal{G})$ (a.s.);
3. If $\mathcal{G} = \mathcal{F}$, then $E(X|\mathcal{G}) = X$ (a.s.);
4. $E(E(X|\mathcal{G})) = E(X)$ (a.s.);
5. If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_1)$ (a.s.);
6. We say that random variable $X$ is \textit{independent} of the $\sigma$-algebra $\mathcal{G}$, if all events of the type
\[ \{ \omega : X(\omega) \in B \} \in \mathcal{B}(\mathbb{R}) \]
are independent of any event from the $\sigma$-algebra $\mathcal{G}$. In this case, $E(X|\mathcal{G}) = X$ (a.s.);
7. If random variables $\varphi_1$ and $\varphi_2$ are measurable with respect to $\mathcal{G}$, then
\[ E(\varphi_1 X_1 + \varphi_2 X_2|\mathcal{G}) = \varphi_1 E(X_1|\mathcal{G}) + \varphi_2 E(X_2|\mathcal{G}) \] (a.s.),
where $X_1$ and $X_2$ are given random variables. In particular, if $\varphi_1$ and $\varphi_2$ are constant, then this property generalizes the linearity of expectations;
8. If a sequence $(X_n)_{n=1,2,\ldots}$ of random variables is such that $|X_n| \leq Y$ (a.s.), $E(Y) < \infty$, and $X_n \to X$ (a.s.), then the dominated convergence theorem implies
\[ E(X_n|\mathcal{G}) \to E(X|\mathcal{G}) \] (a.s.) as $n \to \infty$.

Further insights about probabilistic properties of sequence $(X_n)_{n=1,2,\ldots}$ of random variables on probability space $(\Omega, \mathcal{F}, P)$ can be obtained by introducing the corresponding sequence of $\sigma$-algebras $(\mathcal{F}_n^X)_{n=1,2,\ldots}$, where $\sigma$-algebra $\mathcal{F}_n^X$ is generated by the first $n$ random variables: $X_1, \ldots, X_n$. It is customary to interpret each $\mathcal{F}_n^X$ as information associated with the given sequence up to time $n$. Since more information becomes available as $n$ increases, the sequence $(\mathcal{F}_n^X)$ is usually referred to as \textit{natural information flow} or as \textit{natural filtration}. As we mentioned earlier, such an approach to studying sequences of random variables corresponds well to the nature of financial markets.

Thus, when modeling financial markets on probability space $(\Omega, \mathcal{F}, P)$, it is natural to introduce a filtration $\mathcal{F} = (\mathcal{F}_n)_{n=0,1,2,\ldots}$. It is customary to assume that $\mathcal{F}_0 = \emptyset$ and $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n, n = 1, 2, \ldots$. The probability space with filtration $(\Omega, \mathcal{F}, \mathcal{F}, P)$ is referred to as \textit{stochastic basis}. One of the advantages of this general approach is an opportunity to accommodate a situation when the observed information flow is bigger or smaller than the filtration generated by a specific sequence $(X_n)$. Financial interpretations of such situations
include, for example, the case of obtaining some additional information (e.g., insider information) and the case when it is impossible to obtain the complete information (e.g., in the case of non-tradable assets). We say that a sequence \((X_n)_{n=1,2,...}\) of random variables is a \textit{stochastic sequence} if it is adapted to a filtration \(\mathcal{F}^X_n \subseteq \mathcal{F}_n\) for all \(n\). Motivated by financial applications, we focus our attention on studying such sequences. We say that an integrable stochastic sequence \((X_n)_{n=1,2,...}\) defined on a stochastic basis \((\Omega, \mathcal{F}, \mathcal{F}, P)\) is a \textit{martingale} if
\[ E(X_n|\mathcal{F}_{n-1}) = X_{n-1} \quad \text{(a.s.)} \]
for all \(n \geq 1\).

If
\[ E(X_n|\mathcal{F}_{n-1}) \geq X_{n-1} \quad \text{a.s. or} \quad E(X_n|\mathcal{F}_{n-1}) \leq X_{n-1} \quad \text{a.s.} \]
for all \(n \geq 1\), then \(X\) is called a \textit{submartingale} or a \textit{supermartingale}, respectively.

\textbf{Example 1.2}

Let \((Y_n)_{n=0,1,...}\) be a sequence of independent random variables with zero expectations \(E(Y_n)\). Then the sequence of the partial sums \(X_n = \sum_{k=0}^{n} Y_k\) is a martingale with respect to the natural filtration \((\mathcal{F}^X_n)_{n=0,1,...}\). If expectations \(E(Y_n)\) are non-negative (non-positive), then \((X_n)\) is a submartingale (supermartingale), respectively. \(\square\)

\textbf{Example 1.3}

Let \((Y_n)_{n=0,1,...}\) be a sequence of independent random variables with expectations \(E(Y_n) = 1\). Then the sequence of the partial products \(X_n = \prod_{k=0}^{n} Y_k\) is a martingale with respect to the natural filtration \((\mathcal{F}^X_n)_{n=0,1,...}\). If expectations \(E(Y_n) \geq 1\) (or \(\leq 1\)), then \((X_n)\) is a submartingale (supermartingale), respectively. \(\square\)

\textbf{Remark}

Given a martingale \((X_n)_{n=0,1,...}\), there is a simple way of constructing submartingales (and therefore supermartingales, due to the symmetric relationship between submartingales and supermartingales). Suppose that \(\varphi\) is a convex downward Borel function such that \(E|\varphi(X_n)| < \infty\) for all \(n = 0,1,\ldots\). Then Jensen’s inequality implies that \((\varphi(X_n))_{n=0,1,...}\) is a submartingale.

The following notion of a \textit{stopping time} (or \textit{Markov time}) is closely related to the introduced notion of a stochastic sequence. A random variable \(\tau : \Omega \rightarrow \mathbb{Z}_+ \equiv \{0,1,\ldots\}\) is a \textit{stopping time} if
\[ \{ \omega : \tau(\omega) \leq n \} \in \mathcal{F}_n \quad \text{for all } n = 0,1,\ldots, \]
or equivalently, \(\{ \omega : \tau(\omega) = n \} \in \mathcal{F}_n \) for all \(n = 0,1,\ldots\). We can interpret stopping times as random times where randomness does not depend on the
future (beyond time $n$). Thus, we arrive at the following definition. Suppose (for technical reasons) that $\sigma$-algebra $\mathcal{F}$ is the minimal $\sigma$-algebra that contains the filtration $(\mathcal{F}_n)_{n=0,1,\ldots}$. Let $\tau$ be a stopping time, then $\sigma$-algebra

$$\mathcal{F}_\tau = \{ A \in \mathcal{F} : A \cap \{ \tau = n \} \in \mathcal{F}_n \text{ for all } n = 0, 1, \ldots \}$$

is referred to as the information that is available up to the stopping time $\tau$. Clearly, if $\tau_1 \leq \tau_2$ (a.s), then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.

Now let us discuss some useful properties of martingales and stopping times.

1. Let $(X_n)_{n=0,1,\ldots}$ be a martingale (submartingale, supermartingale) and suppose that stopping times $\tau_1 \leq \tau_2 \leq N$ (a.s), then

$$E(X_{\tau_2} | \mathcal{F}_{\tau_1}) = X_{\tau_1}$$

(1.13)

$$E(X_{\tau_2} | \mathcal{F}_{\tau_1}) \geq X_{\tau_1}, \ E(X_{\tau_2} | \mathcal{F}_{\tau_1}) \leq X_{\tau_1}, \text{ respectively). In particular, } E(X_{\tau_1}) = E(X_{\tau_2})$$

**Proof** of this property, which generalizes the obvious property of deterministic times, readily follows from the following observations:

$$E(I_{A \cap \{ \tau_1 = n \}} (X_{\tau_2} - X_{\tau_1})) = E(I_{A \cap \{ \tau_1 = n \}} \sum_{k=n}^N (X_{k+1} - X_k) I_{\{\tau_2 > k\}})$$

$$= \sum_{k=n}^N E(I_{A \cap \{ \tau_1 = n \}} I_{\{\tau_2 > k\}} (X_{k+1} - X_k))$$

$$= \sum_{k=n}^N E(E[(X_{k+1} - X_k) | \mathcal{F}_k] I_{A \cap \{ \tau_1 = k \}} I_{\{\tau_2 > k\}}) = 0$$

for $A \in \mathcal{F}_{\tau_1}$ and $n < N$. Similarly, we can show $E(I_{A \cap \{ \tau_1 = n \}} (X_{\tau_2} - X_{\tau_1})) \geq 0$ for submartingales and $E(I_{A \cap \{ \tau_1 = n \}} (X_{\tau_2} - X_{\tau_1})) \leq 0$ for supermartingales.

2. **Kolmogorov-Doob inequalities.** Let $(X_n)_{n=0,1,\ldots,N}$ be a submartingale. Then, for any real $a > 0$, we have

$$P\{ \omega : \max_{0 \leq n \leq N} X_n \geq a \} \leq \frac{E(X_N^+)}{a}.$$  

(1.14)

**Proof** Consider the stopping time

$$\tau = \inf \{ n \leq N : X_n \geq a \}$$

and the set

$$A = \{ \omega : \max_{0 \leq n \leq N} X_n \geq a \} \in \mathcal{F}_\tau.$$  

Note $a \cdot I_A \leq X_\tau \cdot I_A$. Calculating expectations we obtain

$$a \cdot P(A) \leq E(X_\tau \cdot I_A) \leq E(X_N \cdot I_A) \leq E(X_N^+ \cdot I_A) \leq E(X_N^+),$$
which proves (1.14).

If \((X_n)_{n=0,1,...,N}\) is a martingale, then by applying (1.14) to submartingales \(|X_n|\) and \(X_n^2\), we arrive at the following inequalities for martingales:

\[
P\{ \omega : \max_{0 \leq n \leq N} |X_n| \geq a \} \leq \frac{E(|X_N|)}{a} \tag{1.15}
\]

and

\[
P\{ \omega : \max_{0 \leq n \leq N} |X_n| \geq a \} \leq \frac{E(|X_N|^2)}{a^2} \tag{1.16}
\]

under assumption that \(E(|X_N|^2) \leq \infty\). Taking limits in (1.14)–(1.16) as \(N \to \infty\), we obtain the following inequalities for infinite time intervals:

\[
P\{ \omega : \sup_n X_n \geq a \} \leq \sup_n E(X_n^+), \tag{1.17}
\]

\[
P\{ \omega : \sup_n |X_n| \geq a \} \leq \sup_n E(|X_n|), \tag{1.18}
\]

and

\[
P\{ \omega : \sup_n |X_n| \geq a \} \leq \frac{\sup_n E(|X_n|^2)}{a^2} \tag{1.19}
\]

under assumption that all supremums in the right-hand sides are finite.

3. **Convergence of martingales and submartingales.** Let \((X_n)_{n=0,1,...}\) be a submartingale with \(\sup_n E(|X_n|) < \infty\), then there exists

\[
X_\infty = \lim_n X_n \quad \text{such that } E(|X_\infty|) < \infty.
\]

**Proof** Suppose that this limit does not exist on a set of positive probability measure; that is,

\[
P\{ \omega : \limsup_n X_n > \liminf_n X_n \}.
\]

Note that the set \(\{ \omega : \limsup_n X_n > \liminf_n X_n \}\) can be written as a countable union of sets of the form \(\{ \omega : \limsup_n X_n > y > x > \liminf_n X_n \}\), where \(x\) and \(y\) are all possible rational numbers. Thus, (1.20) implies that there exist rational numbers \(x < y\) such that

\[
P\{ \omega : \limsup_n X_n > y > x > \liminf_n X_n \}.
\]

Denote \(\beta(x,y)\) the number of upcrossings of interval \((x,y)\) by the submartingale \((X_n)_{n=0,1,...}\). Kolmogorov-Doob inequalities imply (we omit this rather technical proof) the following Doob estimate for the expected number of upcrossings:

\[
E(\beta(x,y)) \leq \frac{\sup_n E(X_n^+) + |x|}{y - x}.
\]
Since \((X_n)_{n=0,1,...}\) is a submartingale, we have
\[
\sup_n E(|X_n|) < \infty \iff \sup_n E(X_n^+) < \infty
\]
and Doob estimate implies that \(E(\beta(x,y)) < \infty\). Thus, \(\beta(x,y) < \infty\) (a.s.), which contradicts (1.21) and (1.20). \(\square\)

If in addition to almost sure convergence of the submartingale \((X_n)_{n=0,1,...}\), one wishes to establish convergence in \(L^1\), then the condition of finiteness of \(\sup_n E(|X_n|)\) must be replaced by a stronger condition of the uniform integrability. In particular, in the case of martingales we arrive at the following Lévy’s structural characterization theorem: a stochastic sequence \((X_n)_{n=0,1,...}\) is a uniformly integrable martingale if and only if there exists a uniformly integrable random variable \(X_\infty\) such that \(X_n = E(X_\infty|\mathcal{F}_n), \ n = 0,1,\ldots,\) and \(X_n \to X_\infty\) both almost surely and in \(L^1\).

Note that, in particular, a martingale \((X_n)_{n=0,1,...,N}\) is uniformly integrable since it is a finite family of random variables; therefore, by Lévy’s characterization theorem, we have \(X_n = E(X_N|\mathcal{F}_n), \ n = 0,1,\ldots,N\). This property of martingales can be readily obtained from the first principles without Lévy’s theorem, and it explains why the use of martingales is so natural in financial applications of the dynamic programming method.

We can use now the convergence properties of uniformly integrable martingales to extend property (1.13) from the case of bounded stopping times to the case of almost surely finite stopping times. Namely, let \((X_n)_{n=0,1,...}\) be a uniformly integrable martingale and consider stopping times \(\sigma \leq \tau < \infty\) (a.s.). Then
\[
E(X_\tau | \mathcal{F}_\sigma) = X_\sigma \quad \text{(a.s.)}. \tag{1.22}
\]

In particular, for a finite stopping time \(\tau\), we have
\[
E(|X_\tau|) < \infty \quad \text{and} \quad E(X_\tau) = E(X_0). \tag{1.23}
\]

We omit the complete proof of property (1.22); we only explain (1.23). For a fixed \(N\), we introduce a bounded stopping time \(\tau_N := \tau \wedge N = \min(\tau, N)\). Property (1.13) implies that \(E(X_0) = E(X_{\tau_N})\), and we observe
\[
E(|X_{\tau_N}|) = 2E(X_{\tau_N}^+) - E(X_{\tau_N}) \leq 2E(X_{\tau_N}^+) - E(X_0).
\]

Since \((X_n^+)_{n=0,1,...}\) is a submartingale, we obtain
\[
E(X_{\tau_N}^+) = \sum_{k=0}^{N} E(X_k^+ I_{\{\tau_N=k\}}) + E(X_N^+ I_{\{\tau>N\}}) \leq \sum_{k=0}^{N} E(X_k^+ I_{\{\tau_N=k\}}) + E(X_N^+ I_{\{\tau>N\}}) = E(X_N^+) \leq E(|X_N|) \leq \sup_n E(|X_n|). \]
Thus,
\[ E(|X_{\tau_N}|) \leq 3 \sup_n E(|X_n|), \]
and after taking limits as \( N \to \infty \), we arrive at
\[ E(|X_\tau|) \leq \limsup_n E(|X_{\tau_N}|) \leq 3 \sup_n E(|X_n|) < \infty. \]
Further, the uniform integrability of \((X_n)_{n=0,1,...}\) implies
\[ E(|X_n| \cdot I_{\{\tau>n\}}) \to 0 \quad \text{as} \quad n \to \infty. \quad (1.24) \]
Note that \( P(\tau > n) \to 0 \) as \( n \to \infty \) since \( \tau \) is a bounded stopping time and \( P(\tau < \infty) = 1. \)

Consider the following decomposition:
\[ X_\tau = X_{\tau \wedge n} + (X_\tau - X_n) \cdot I_{\{\tau>n\}}. \]
Taking expectations and using the martingale property, we obtain
\[ E(X_\tau) = E(X_{\tau \wedge n}) + E(X_\tau \cdot I_{\{\tau>n\}}) - E(X_n \cdot I_{\{\tau>n\}}) \quad (1.25) \]
Taking into account convergence (1.24) and finiteness of \( X_{\tau} \), we conclude that second and third terms in the right-hand side of this equality vanish as \( n \to \infty. \)
Thus, equality (1.25) implies \( E(X_\tau) = E(X_0). \) \( \square \)

4. Martingales and absolute continuity of probability measures.
Let \((\Omega, \mathcal{F}, \mathcal{F}, P)\) be a stochastic basis, and consider another probability measure \( \bar{P} \ll P \) with density \( d\bar{P}/dP = Z \). Let \( \bar{P}_n \) and \( P_n \) be their restrictions on \( \mathcal{F}_n \), and denote \( Z_n := d\bar{P}_n/dP_n. \)
Sequence \((Z_n)_{n=0,1,...}\) is called the local density of \( \bar{P} \) with respect to \( P \) and it is a martingale with respect to the initial measure \( P \). Indeed, let \( A \in \mathcal{F}_n \), then
\[ E(Z_{n+1} \cdot I_A) = E \left( \frac{d\bar{P}_{n+1}}{dP_{n+1}} \cdot I_A \right) = \bar{P}_{n+1}(A) = \bar{P}_n(A) = E \left( \frac{d\bar{P}_n}{dP_n} \cdot I_A \right) \]
\[ = E(Z_n \cdot I_A) \quad (1.26) \]
for all \( n = 0, 1, \ldots \). Lévy’s structural characterization implies that \( Z_n \to Z \) (a.s.) as \( n \to \infty \), and \( Z_n = E(Z | \mathcal{F}_n) \).

In a similar way, one can prove the following formula for change of probability under the conditional expectation sign: for any \( \mathcal{F}_n \)-measurable random variable \( Y \), we have
\[ \bar{E}(Y | \mathcal{F}_{n-1}) = E(Z_nZ_{n-1}^{-1}Y | \mathcal{F}_{n-1}) \quad (a.s.), \quad (1.27) \]
given that the conditional expectation \( E(Z_nZ_{n-1}^{-1}Y | \mathcal{F}_{n-1}) \) is well defined.
The following property that connects the martingale property and absolute continuity of probability measures is related to formula (1.27) and is referred to as the Girsanov theorem. Let \((M_n)_{n=0,1,\ldots}\) (with \(M_0 = 0\)) be a martingale with respect to the original probability \(P\) and suppose \(E(|Y_n Y_{n-1}^{-1} \triangle M_n|) < \infty\) for all \(n = 1, 2, \ldots\). Define \((\widetilde{M}_n)_{n=0,1,\ldots}\) (with \(\widetilde{M}_0 = 0\)) by

\[
\widetilde{M}_n = \triangle \tilde{M}_n = \triangle M_n - E(Y_n Y_{n-1}^{-1} \triangle M_n | \mathcal{F}_{n-1}).
\]

Using (1.27), we calculate

\[
\tilde{E}(\triangle \tilde{M}_n | \mathcal{F}_{n-1}) = \tilde{E}(\triangle M_n - E(Y_n Y_{n-1}^{-1} \triangle M_n | \mathcal{F}_{n-1}) | \mathcal{F}_{n-1})
\]

\[
= \tilde{E}(\triangle M_n | \mathcal{F}_{n-1}) - \tilde{E}(E(Y_n Y_{n-1}^{-1} \triangle M_n | \mathcal{F}_{n-1}) | \mathcal{F}_{n-1}) = 0,
\]

which implies that \((\widetilde{M}_n)_{n=0,1,\ldots}\) is a martingale with respect to \(\widetilde{P} \ll P\).

5. Doob decomposition and predictable characteristics of martingales. The notion of predictability is closely related to the notion of a martingale. We say that a stochastic sequence \((A_n)_{n=0,1,\ldots}\) is predictable if random variables \(A_n\) are \(\mathcal{F}_{n-1}\)-measurable for all \(n\). We also say that a stochastic sequence (not necessarily predictable) \((A_n)_{n=0,1,\ldots}\) is non-decreasing if \(\triangle A_n = A_n - A_{n-1} \geq 0\) (a.s.) for all \(n\).

Let \((X_n)_{n=0,1,\ldots}\) be a submartingale. Then a martingale \((M_n)_{n=0,1,\ldots}\) and a non-decreasing stochastic sequence \((A_n)_{n=0,1,\ldots}\) exist such that the following Doob decomposition

\[
X_n = M_n + A_n, \quad n = 0, 1, \ldots
\]

(1.28) holds.

To prove the existence, we set \(M_0 = X_0, A_0 = 0,\)

\[
M_n = M_0 + \sum_{k=0}^{n-1} \{X_{k+1} - E(X_{k+1} | \mathcal{F}_k)\}, \quad A_n = \sum_{k=0}^{n-1} \{E(X_{k+1} | \mathcal{F}_k) - X_k\}
\]

for \(n = 1, 2, \ldots\). Decomposition (1.28) is unique in the class of predictable stochastic sequences. Indeed, if another martingale \((M'_n)_{n=0,1,\ldots}\) and another non-decreasing stochastic sequence \((A'_n)_{n=0,1,\ldots}\) that satisfy (1.28) exist, then

\[
\triangle A'_{n+1} = A'_{n+1} - A'_n = \triangle A_{n+1} + \triangle M_{n+1} - \triangle M'_{n+1}.
\]

Since \((M_n)_{n=0,1,\ldots}\) and \((M'_n)_{n=0,1,\ldots}\) are martingales and since sequences \((A_n)_{n=0,1,\ldots}\) and \((A'_n)_{n=0,1,\ldots}\) are predictable, we have

\[
\triangle A'_{n+1} = E(\triangle A'_{n+1} | \mathcal{F}_n) = E(\triangle A_{n+1} | \mathcal{F}_n) = \triangle A_{n+1} \text{ (a.s.)},
\]

and therefore \(A_n = A'_n\) and \(M_n = M'_n\) (a.s.), \(n = 0, 1, \ldots\) \(\square\)
If martingale \( (M_n)_{n=0,1,...} \) is such that \( E(M_n) < \infty \) for all \( n = 0, 1, \ldots \), then it is called a square integrable martingale. Applying Doob decomposition (1.28) to the submartingale \( (M^2_n)_{n=0,1,...} \), we conclude that a martingale \( (m_n)_{n=0,1,...} \) and a non-decreasing predictable sequence \( (\langle M \rangle_n)_{n=0,1,...} \) exist such that \( M^2_n = m_n + \langle M \rangle_n \). Sequence \( (\langle M \rangle_n)_{n=0,1,...} \) is called the quadratic characteristic or the compensator of \( M \), and it can be constructed in the following way:

\[
\langle M \rangle_n = \sum_{k=1}^{n} E((\Delta M_k)^2 | \mathcal{F}_{k-1}), \quad \langle M \rangle_0 = 0.
\]

Note that

\[
E((M_k - M_l)^2 | \mathcal{F}_l) = E(M_k^2 - M_l^2 | \mathcal{F}_l) = E(\langle M \rangle_k - \langle M \rangle_l | \mathcal{F}_l), \quad l \leq k,
\]

and \( E(M_k^2) = E(\langle M \rangle_k) \) for \( k = 0, 1, \ldots \).

Recall that one of the measures of association of two random variables with finite second moments is their covariance. A similar measure of association can be introduced for two square integrable martingales \( M \) and \( N \): the sequence \( (\langle M, N \rangle_n)_{n=0,1,...} \) defined by

\[
\langle M, N \rangle_n = \frac{1}{4} \left\{ \langle M + N \rangle_n - \langle M - N \rangle_n \right\}, \quad n = 0, 1, \ldots
\]

is called the mutual quadratic characteristic of \( M \) and \( N \). It is not difficult to show that sequence

\[
(M_n N_n - \langle M, N \rangle_n)_{n=0,1,...}
\]

is a martingale, so the mutual quadratic characteristic of \( M \) and \( N \) is the compensator of their product \( (M_n N_n)_{n=0,1,...} \). Square integrable martingales \( M \) and \( N \) are said to be orthogonal if \( \langle M, N \rangle_n = 0 \) for all \( n = 0, 1, \ldots \).

6. **Discrete stochastic integrals and stochastic exponentials.** Let \( (H_n)_{n=0,1,...} \) be a predictable stochastic sequence and \( (m_n)_{n=0,1,...} \) be a martingale. Stochastic sequence

\[
H \ast m_n = \sum_{k=0}^{n} H_k \Delta m_k
\]

is called a discrete stochastic integral of \( H \) with respect to \( m \). If martingale \( m \) is square integrable, sequence \( H \) is predictable and \( E(H_n^2 \Delta(m)_n) < \infty \) for all \( n = 0, 1, \ldots \), then stochastic integral \( (H \ast m_n)_{n=0,1,...} \) is a square integrable martingale with quadratic characteristic

\[
\langle H \ast m \rangle_n = \sum_{k=0}^{n} H_k^2 \Delta(m)_k.
\]
Further, let \( (M_n)_{n=0,1,...} \) be a fixed square integrable martingale, then one can consider all square integrable martingales \( (N_n)_{n=0,1,...} \) that are orthogonal to \( (M_n)_{n=0,1,...} \) and introduce a family of square integrable martingales of the following form
\[
X_n = M_n + N_n . 
\]
Conversely, any square integrable martingale \( (X_n)_{n=0,1,...} \) can be written in form (1.30), where the orthogonal term \( N \) has the form of the stochastic integral (1.29) with the martingale \( m \) that is orthogonal to the given martingale \( M \). This version of decomposition (1.30) is usually referred to as Kunita-Watanabe decomposition.

Discrete stochastic integrals are naturally related to discrete stochastic differential equations (or stochastic difference equations). Solutions of stochastic difference equations are often used in modeling the dynamics of asset prices in financial markets. Consider a stochastic sequence \( (U_n)_{n=0,1,...} \) with \( U_0 = 0 \). Define new stochastic sequence \( (X_n)_{n=0,1,...} \) with \( X_0 = 1 \) by
\[
\Delta X_n = X_{n-1} \Delta U_n , \quad n = 1, 2, \ldots . 
\]
This simple linear stochastic differential equation has an obvious solution
\[
X_n = \prod_{k=1}^{n} (1 + \Delta U_k) = \varepsilon_n(U) ,
\]
which is called a stochastic exponential. A non-homogeneous version of equation (1.31) has the form
\[
\Delta X_n = \Delta N_n + X_{n-1} \Delta U_n , \quad X_0 = N_0 ,
\]
where \( (N_n)_{n=0,1,...} \) is a given stochastic sequence. A solution of the non-homogeneous equation can be written in terms of solutions of the corresponding homogeneous equation, and it has the form
\[
X_n = \varepsilon_n(U) \left[ N_0 + \sum_{k=1}^{n} \frac{\Delta N_k}{\varepsilon_k(U)} \right] .
\]
Stochastic exponentials have the following useful properties:

(a) \( \frac{1}{\varepsilon_n(U)} = \varepsilon_n(-U^*) \), where \( \Delta U^* = \frac{\Delta U}{1 + \Delta U} \), and \( \Delta U \neq -1 \);

(b) \( (\varepsilon_n(U))_{n=0,1,...} \) is a martingale if and only if \( (U_n)_{n=0,1,...} \) is a martingale;

(c) \( \varepsilon_n(U) = 0 \) for all \( n \geq \tau_0 := \inf \{ k : \varepsilon_k(U) = 0 \} \);

(d) the multiplication rule for stochastic exponentials that correspond to \( (U_n)_{n=0,1,...} \) and \( (V_n)_{n=0,1,...} \):
\[
\varepsilon_n(U)\varepsilon_n(V) = \varepsilon_n(U + V + [U,V]) ,
\]
where
\[
[U, V]_n = \sum_{k=1}^{n} \Delta U_k \Delta V_k \quad \text{and} \quad [U, V]_0 = 0.
\]
Bibliography


This page intentionally left blank