## Contents

1 Brownian Motion ........................................... 3
   1.1 Axioms of Probability ..................................... 3
   1.2 Random Variables ......................................... 11
   1.3 Random Vectors ........................................... 20
       1.3.1 Random Vectors Generalia ............................. 20
       1.3.2 Marginal Distributions ............................... 21
       1.3.3 Expectations, etc. .................................... 22
       1.3.4 The Variance-Covariance Matrix ...................... 23
       1.3.5 Independence .......................................... 24
       1.3.6 Functions of Random Vectors ......................... 25
   1.4 The Multivariate Normal Distribution .................... 30
       1.4.1 The Definition ........................................ 30
       1.4.2 The Bivariate Case .................................... 31
   1.5 Brownian Motion Defined .................................. 40
       1.5.1 Stochastic Processes .................................. 40
       1.5.2 The Distribution of a Stochastic Process ............ 41
       1.5.3 Brownian Motion ...................................... 41
       1.5.4 A Little History ...................................... 44
       1.5.5 Gaussian Processes ................................... 44
       1.5.6 Examples of Gaussian Processes ....................... 45
       1.5.7 Brownian Motion as a Limit of Random Walks ........ 46
   1.6 Simulation .................................................. 57
       1.6.1 Random Number Generators ............................ 57
       1.6.2 Simulation of Random Variables ....................... 60
       1.6.3 Simulation of Normal Random Vectors ................. 63
       1.6.4 Simulation of the Brownian Motion and Gaussian Processes 64
CONTENTS

1.6.5 Monte Carlo Integration .............................................. 65
1.7 Conditioning ............................................................ 70
  1.7.1 Conditional Densities .............................................. 70
  1.7.2 Other Conditional Quantities ................................. 72
  1.7.3 $\sigma$-algebra = Amount of Information ................... 74
  1.7.4 Filtrations .......................................................... 76
  1.7.5 Conditional Expectation in Discrete Time .................. 77
  1.7.6 A Formula for Computing Conditional Expectations ....... 81
  1.7.7 Further properties of Conditional Expectation .......... 82
  1.7.8 What happens in continuous time? ......................... 84
  1.7.9 Martingales ........................................................ 87
  1.7.10 Martingales as Unwinnable Gambles ....................... 89

2 Stochastic Calculus ...................................................... 105
  2.1 Stochastic Integration ............................................... 106
    2.1.1 What do we really want? .................................. 106
    2.1.2 Stochastic Integration with respect to Brownian Motion .. 107
    2.1.3 The Itô Formula .............................................. 109
    2.1.4 Some properties of stochastic integrals ................. 111
    2.1.5 Itô processes ............................................... 115
  2.2 Applications to Finance ........................................... 122
    2.2.1 Modelling Trading and Wealth ............................ 122
    2.2.2 Derivatives .................................................. 123
    2.2.3 Arbitrage Pricing ........................................... 126
    2.2.4 Black and Scholes via PDEs ............................... 127
    2.2.5 The “Martingale method” for pricing options .......... 130

A Odds and Ends .......................................................... 133
  A.1 No $\mathbb{P}$ for all Subsets of $[0, 1]$ ....................... 134
  A.2 Marginals: Normal, Joint: Not ................................ 135
  A.3 Mersenne Twister .................................................. 136
  A.4 Conditional Expectation Cheat-Sheet ......................... 141
Brownian Motion

The probable is what usually happens.
—Aristotle

It is a truth very certain that when it is not in our power to determine what is true we ought to follow what is most probable
—Descartes - “Discourse on Method”

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.
—Pierre Simon Laplace - “Thorie Analytique des Probabilits, 1812 ”

Anyone who considers arithmetic methods of producing random digits is, of course, in a state of sin.
—John von Neumann - quote in “Conic Sections” by D. MacHale

I say unto you: a man must have chaos yet within him to be able to give birth to a dancing star: I say unto you: ye have chaos yet within you...
—Friedrich Nietzsche - “Thus Spake Zarathustra”

1.1 Axioms of Probability

All of probability starts with a set, the state space, which we usually denote by Ω. Every element ω ∈ Ω stands for a state of the world, i.e. represents a possible development of the events of interest.

Example 1.1.1. The universe of Derek the Daisy is very simple. All that really matters to him is whether it rains or shines. Therefore, his Ω has two elements: Ω = {R, S},
where $R$ stands for rain and $S$ for sun. But even daisies live more than one day, so he might be interested in what happens tomorrow. In that case, we should have $\Omega = \{(R, R), (R, S), (S, R), (S, S)\}$ - there are four possible states of the world now, and each $\omega \in \Omega$ is an ordered pair.

From the previous example, it is clear how to picture Derek as an even wiser daisy, looking 3, 4, or even 100 days in advance. It is also clear how to add snow, presence of other daisies or even solar eclipses into the model; you would just list all the variables of interest and account for all the possibilities. There will of course exist an $\omega \in \Omega$ describing the situation (the world, the future, the parallel universe, . . . ) in which there is a solar eclipse every day from today until the 100th day. One might object that such a display of events does not seem too likely to happen. The same person should immediately realize that he or she has just discovered the notion of probability: while the clerk $\Omega$ blindly keeps track of every conceivable contingency, the probability brings reason (and inequality) to the game. Its job is to assign a number to each $\omega \in \Omega$ - higher numbers to likelier states of the world, and smaller numbers to unusual coincidences. It can also take groups of $\omega$’s (subsets of $\Omega$) and assign numbers to those by simply adding the probabilities. The subsets of $\Omega$ are usually referred to as events. For an event $A \subseteq \Omega$ we write $\mathbb{P}[A]$ for the probability assigned to $A$. With these new concepts in mind, we venture into another example.

**Example 1.1.2.** Derek is getting sophisticated. He is not interested in meteorology anymore. He wants to know how his stocks are doing (this is a course in mathematical finance after all !). He has bought $100 worth of shares of the Meadow Mutual Fund a year ago, and is wondering how much his portfolio is worth today. He reads his Continuous-Time Finance Lecture Notes and decides that the state space $\Omega$ will consist of all positive numbers (nobody said state spaces should be finite!!!!!). So far so good - all the contingencies are accounted for and all we need to do is define the probability. You have to agree that it is very silly to expect his portfolio to be worth $1,000,000,000,000. It is much more likely it will be around $100. With that in mind he starts:

“hm... let’s see... what is the probability that my portfolio will be worth exactly $101.22377 (in Derek’s world there are coins worth 0.1 cents, 0.01 cents, etc.) Well, ... I have to say that the probability must be 0. Same for $95.23213526763, and $88.23868726345345345384762, . . . and $144.34322423345345343423. . . I must be doing something wrong.”
But then he realizes:

“my broker Mr. Mole has told me that it is extremely unlikely that I will lose more than 10%. He also said that these days no fund can hope to deliver a return larger than 15%. That means that I should be quite confident that the value of my portfolio lies in the interval \[90\$, 115\$\] say, ... 90% confident. That must be it!”

And Derek is right on the money here. When \(\Omega\) is infinite, it usually makes no sense to ask questions like *What is the probability that \(\omega \in \Omega\) will turn out to be the true state of the world?*. Not that you cannot answer it. You can. The answer is usually trivial and completely uninformative - zero. The only meaningful questions are the ones concerning the probabilities of events: *What is the probability that the true state of the world \(\omega\) will turn out to be an element of the set \(A\)?* In Derek’s case, it made much more sense to assign probabilities to subsets of \(\Omega\) and it cannot be done by summing over all elements as in the finite case. If you try to do that you will end up adding 0 to itself an uncountable number of times. Strange ...

Where did this whole discussion lead us? We are almost tempted to define the probability as a function assigning a number to each subset of \(\Omega\), and complying with a few other well-know rules (additive, between 0 and 1, etc.), namely

**Definition 1.1.3** (Tentative). Let \(\Omega\) be a set. A **probability** is a function \(\mathbb{P}\) from the family \(\mathcal{P}(\Omega)\) of all subsets of \(\Omega\) to the set of reals between 0 and 1 such that

- \(\mathbb{P}[\Omega] = 1\), and
- \(\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]\), whenever \(A, B\) are subsets of \(\Omega\), and \(A \cap B = \emptyset\).

... but the story is not so simple. In 1924, two Polish mathematicians Stefan Banach and Alfred Tarski proved the following statement:

**Banach-Tarski Paradox:** *It is possible to take a solid ball in 3-dimensional space, cut it up into finitely many pieces and, moving them using only rotation and translation, reassemble the pieces into two balls the same size as the original.*
...and showed that the concept of measurement cannot be applied to all subsets of the space. The only way out of the paradox is to forget about the universal notion of volume and restrict the class of sets under consideration. And then, if you think about it, probability is a weird kind of a volume. A volume on events, not bodies in space, but still a volume. If you are unhappy with an argument based on the premise that probability is a freaky kind of volume, wait until you read about \( \sigma \)-additivity and then go to the Appendix to realize that it is impossible to define a \( \sigma \)-additive probability \( P \) on all subsets of \([0, 1]\), where any subset \( A \) has the same probability as its translated copy \( x + A \) (as long as \( x + A \subset [0, 1] \)).

The rescue from the apparently doomed situation came from Russian mathematician Andrei Nikolae-vich Kolmogorov who, by realizing two important things, set foundations for the modern probability theory. So, what are the two important ideas? First of all, Kolmogorov supported the idea that the probability function can be useful even if it is not defined on all subsets of \( \Omega \). If we restrict the class of events which allow a probability to be assigned to them, we will still end up with a useful and applicable theory. The second idea of Kolmogorov was to require that \( P \) be countably-additive (or \( \sigma \)-additive), instead of only finitely-additive. What does that mean? Instead of postulating that \( P(A) + P(B) = P(A \cup B) \), for disjoint \( A, B \subset \Omega \), he required more: for any sequence of pairwise disjoint sets \( (A_n)_{n \in \mathbb{N}} \), Kolmogorov kindly asked \( P \) to satisfy

\[
P[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n]
\]  

(countable additivity).

If you set \( A_1 = A, A_2 = B, A_3 = A_4 = \ldots = \emptyset \), you get immediately the weaker statement of finite additivity.

So, in order to define probability, Kolmogorov used the notion of \( \sigma \)-algebra as the third
and the least intuitive ingredient in the axioms of probability theory:

**Definition 1.1.4.** A collection $\mathcal{F}$ of subsets of $\Omega$ is called a $\sigma$-algebra (or a $\sigma$-field) if the following three conditions hold

- $\Omega \in \mathcal{F}$.
- for each $A \in \mathcal{F}$, we must also have $A^c \in \mathcal{F}$, where $A^c \triangleq \{ \omega \in \Omega : \omega \notin A \}$.
- if $(A_n)_{n \in \mathbb{N}}$ is a sequence of elements of $\mathcal{F}$, then their union $\bigcup_{n=1}^{\infty} A_n$ must belong to $\mathcal{F}$ as well.

Typically $\mathcal{F}$ will contain all the events that allow probability to be assigned to them, and we call elements of $\mathcal{F}$ the measurable sets. As we have seen, these will not always be all subsets of $\Omega$. There is no need to worry, though: **In practice, you will never encounter a set so ugly that it is impossible to assign a probability to it.**

**Remark 1.1.1.** The notion of $\sigma$-algebra will however prove to be very useful very soon in a role quite different from the present one. Different $\sigma$-algebras on the same state-space will be used to model different amounts of information. After all, randomness is just the lack of information.

We finish the story about foundations of probability with the axioms of probability theory

**Definition 1.1.5.** A triple $(\Omega, \mathcal{F}, P)$ is called a probability space if $\Omega$ is a non-empty set, $\mathcal{F}$ is a $\sigma$-algebra of subsets of $\Omega$ and $P$ is a probability on $\mathcal{F}$, i.e.

- $P : \mathcal{F} \to [0, 1]$
- $P[\Omega] = 1$
- $P[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n]$, for any sequence $(A_n)_{n \in \mathbb{N}}$ of sets in $\mathcal{F}$, such that $A_n \cap A_m = \emptyset$, for $n \neq m$.  

Last Updated: Spring 2005
Exercise 1.1.6. Let $\mathcal{F}$ and $\mathcal{G}$ be $\sigma$-algebras on $\Omega$. Prove the following

1. $\mathcal{F} \cap \mathcal{G}$ is a $\sigma$-algebra.

2. $\mathcal{F} \cup \mathcal{G}$ is not necessarily a $\sigma$-algebra.

3. There exists the smallest $\sigma$-algebra $\sigma(\mathcal{F} \cup \mathcal{G})$ containing both $\mathcal{F}$ and $\mathcal{G}$. (Hint: prove and use the fact that $\mathcal{P}(\Omega) = \{ A : A \subseteq \Omega \}$ is a $\sigma$-algebra containing both $\mathcal{F}$ and $\mathcal{G}$.)

Exercise 1.1.7. Which of the following are $\sigma$-algebras?

1. $\mathcal{F} = \{ A \subseteq \mathbb{R} : 0 \in A \}$.

2. $\mathcal{F} = \{ A \subseteq \mathbb{R} : A$ is finite $\}$.

3. $\mathcal{F} = \{ A \subseteq \mathbb{R} : A$ is finite, or $A^c$ is finite $\}$.

4. $\mathcal{F} = \{ A \subseteq \mathbb{R} : A$ is open $\}$. (A subset of $\mathbb{R}$ is called open if it can be written as a union of open intervals $(a, b)$.)

5. $\mathcal{F} = \{ A \subset \mathbb{R} : A$ is open or $A$ is closed $\}$. (Note: $A$ is closed if and only if $A^c$ is open).

Exercise 1.1.8. Let $\mathcal{F}$ be a $\sigma$-algebra and $\{ A_n \}_{n \in \mathbb{N}}$ a sequence of elements of $\mathcal{F}$ (not necessarily disjoint). Prove that

1. There exists a pairwise disjoint sequence $\{ B_n \}_{n \in \mathbb{N}}$ of elements in $\mathcal{F}$ such that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. (A sequence $\{ B_n \}_{n \in \mathbb{N}}$ is said to be pairwise disjoint if $B_n \cap B_m = \emptyset$ whenever $n \neq m$.)

2. $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$. 
CHAPTER 1. BROWNIAN MOTION 1.1. AXIOMS OF PROBABILITY

Solutions to Exercises in Section 1.1

Solution to Exercise 1.1.6:

1. We need to prove that the collection \( \mathcal{H} \triangleq \mathcal{F} \cap \mathcal{G} \) of subsets of \( \Omega \) satisfies the three axioms of a \( \sigma \)-algebra:

   - It is obvious that \( \Omega \in \mathcal{H} \) since \( \Omega \in \mathcal{F} \) and \( \Omega \in \mathcal{G} \) by assumption.
   - Secondly, suppose that \( A \) is an element of \( \mathcal{H} \). Then \( A \in \mathcal{F} \) and \( A \in \mathcal{G} \) by assumption, and thus \( A^c \in \mathcal{F} \) and \( A^c \in \mathcal{G} \) since both are \( \sigma \)-algebras. We conclude that \( A^c \in \mathcal{H} \).
   - Then, let \( \{A_n\}_{n \in \mathbb{N}} \) be a sequence of subsets of \( \Omega \) with \( A_n \in \mathcal{H} \) for each \( n \). Then, by assumption \( A_n \in \mathcal{F} \) and \( A_n \in \mathcal{G} \) for any \( n \), and so \( \cup_n A_n \in \mathcal{F} \) and \( \cup_n A_n \in \mathcal{G} \) because \( \mathcal{F} \) and \( \mathcal{G} \) are \( \sigma \)-algebras. Hence, \( \cup_n A_n \in \mathcal{H} \).

2. Take \( \Omega = \{1, 2, 3, 4\} \) and take

   \[
   \mathcal{F} = \{\emptyset, \{1, 2, 3, 4\}, \{1, 2\}, \{3, 4\}\},
   \]

   and

   \[
   \mathcal{G} = \{\emptyset, \{1, 2, 3, 4\}, \{1, 3\}, \{2, 4\}\}.
   \]

   It is easy to show that both \( \mathcal{F} \) and \( \mathcal{G} \) are \( \sigma \)-algebras. The union \( \mathcal{F} \cup \mathcal{G} \) is not a \( \sigma \)-algebra since \( \{1, 2\} \in \mathcal{F} \cup \mathcal{G} \) and \( \{1, 3\} \in \mathcal{F} \cup \mathcal{G} \), but \( \{1, 2\} \cup \{1, 3\} = \{1, 2, 3\} \notin \mathcal{F} \cup \mathcal{G} \).

3. Just take the intersection of all \( \sigma \)-algebras containing \( \mathcal{F} \) and \( \mathcal{G} \). The intersection is nonempty because, at least, \( \mathcal{P} \) (the set containing all subsets of \( \Omega \)) is there.

Solution to Exercise 1.1.7:

1. By definition \( A = \{0\} \in \mathcal{F} \), but \( A^c = (-\infty, 0) \cup (0, \infty) \notin \mathcal{F} \) since \( 0 \notin A^c \). Therefore, \( \mathcal{F} \) is not a \( \sigma \)-algebra.

2. Each element of the sequence \( A_n \triangleq \{n\} \) is in \( \mathcal{F} \). However, \( \cup_n A_n = \mathbb{N} \) is not an element of \( \mathcal{F} \) because it is obviously not finite. We conclude that \( \mathcal{F} \) is not a \( \sigma \)-algebra.

3. The same counterexample as above proves that \( \mathcal{F} \) is not a \( \sigma \)-algebra since neither \( \mathbb{N} \) nor \( \mathbb{R} \setminus \mathbb{N} \) are finite.

4. Remember that a set \( A \) is said to be \textbf{open} if for each \( x \in A \) there exists \( \varepsilon > 0 \) such that \( (x - \varepsilon, x + \varepsilon) \subseteq A \). It is quite evident that \( A = (-\infty, 0) \) is an open set (prove it rigorously if you feel like it: just take \( \varepsilon = |x| \)). However, \( A^c = [0, \infty) \) is not an open set since for \( x = 0 \) no \( \varepsilon > 0 \) will do the trick; no matter how small \( \varepsilon > 0 \) I take, the interval \( (-\varepsilon, \varepsilon) \) will always contain negative numbers, and so it will never be a subset of \( A^c \).
5. Suppose that $\mathcal{F}$ is a $\sigma$-algebra. Define $A = [1, 2]$, $B = (0, 1)$. The set $A$ is closed and $B$ is open, and so they both belong to $\mathcal{F}$. Therefore, the union $A \cup B = (0, 2]$ is an element of $\mathcal{F}$. This is a contradiction with the definition of $\mathcal{F}$ since neither $(0, 2]$ nor $(-\infty, 0] \cup (2, \infty)$ are open (argument is very similar to that from 4.) By contraposition, $\mathcal{F}$ is not a $\sigma$-algebra.

Solution to Exercise 1.1.8:

1. Define $B_1 = A_1$, $B_2 = A_2 - B_1$, $B_3 = A_3 - (B_1 \cup B_2)$, \ldots In general $B_n = A_n - (B_1 \cup B_2 \cup \ldots \cup B_{n-1})$. It is easy to verify that $B_n$’s are disjoint and that, for each $n$, \[ \bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k. \] Inductively, it is also quite obvious that $B_n \in \mathcal{F}$.

2. By De Morgan’s rules, we have \[ \cap_{n \in \mathbb{N}} A_n = \left( \cup_{n \in \mathbb{N}} A_n^c \right)^c. \]
1.2 Random Variables

Look at the following example

**Example 1.2.1.** Derek’s widened interest in the world around him makes the structure of the state space $\Omega$ more and more complicated. Now he has to take the value of his portfolio in US$, the weather, the occurrence of the solar eclipse and the exchange rate between the American Dollar and MCU (Meadow Currency Unit) into account. In symbols, we have

$$\Omega = \mathbb{R}_+ \times \{R, S\} \times \{\text{Eclipse, No eclipse}\} \times \mathbb{R}_+,$$

and the typical element of $\Omega$ is a quadruplet very much like $(114.223, R, \text{No eclipse}, 2.34 \text{ MCU/\$}),$ or $(91.12, R, \text{Eclipse}, 1.99 \text{ MCU/\$}).$

If Derek were interested in the value of his portfolio expressed in MCUs, he would just multiply the first and the fourth component of $\omega,$ namely the value of the portfolio in $\$$ multiplied by the exchange rate. Of course, this can be done for any $\omega \in \Omega$ and the result could be different in different states of the world. In this way, Derek is extracting information from what can be read off $\omega$: the value of the portfolio in MCUs is a function of $\omega.$ It is variable and it is random so we can give the following definition:

**Definition 1.2.2 (tentative).** A function $X : \Omega \rightarrow \mathbb{R}$ is called a random variable.

Why tentative? It has to do with those guys from Poland - Banach and Tarski: for the very same reason that we have introduced a $\sigma$-algebra as a constituent in the probability space. The true (non-tentative) definition of a random variable will respect the existence of the $\sigma$-algebra $\mathcal{F}$:

**Definition 1.2.3.** We say that the function $X : \Omega \rightarrow \mathbb{R}$ is $\mathcal{F}$-measurable, if for any real numbers $a < b$ the set

$$X^{-1}((a, b)) \triangleq \{\omega \in \Omega : X(\omega) \in (a, b)\},$$

is an element of $\mathcal{F}.$ A random variable is any $\mathcal{F}$-measurable function $X : \Omega \rightarrow \mathbb{R}.$

The notion of $\mathcal{F}$-measurability makes sense once you realize that in practice you are going to be interested in the probability of events of the form $X^{-1}((a, b))$ for some real numbers $a, b.$ For example, “what is the probability that my portfolio return will be between 10% and 20% at the end of the next year?” If you want to be able to assign probabilities to such events - and you do - your random variables had better be $\mathcal{F}$-measurable. However, you do not need to worry about measurability at all: all random variables encountered in practice are measurable, and in this course we will never make an issue of it.

So, in practice any function going from $\Omega$ to $\mathbb{R}$ is a random variable, and tells us something about the state of the world. In general, $\Omega$ is going to be a huge set in which we can put everything relevant (and irrelevant), but random variables are going to be the manageable pieces of it. Look at the following example:
Example 1.2.4. Consider the world in which Derek wants to keep track of the value of his portfolio not only a year from now, but also on every instant from now until a year from now. In that case each \( \omega \) would have to describe the evolution of the portfolio value as time \( t \) (in years) goes from 0 (today) to 1 (a year from today). In other words each \( \omega \) is a real-valued function \( \omega : [0, 1] \rightarrow \mathbb{R} \). By a leap of imagination you could write \( \omega \) as an uncountable product of \( \mathbb{R} \) by itself, each copy of \( \mathbb{R} \) corresponding to the value of the portfolio at a different time-instant. So, \( \Omega \) is going to be a collection of all functions from \([0, 1]\) to \( \mathbb{R} \). A huge set, indeed.

In the previous example, the value of the portfolio at a particular time point (say \( t_0 = 0.5 \), i.e. 6 months from now) is a random variable. For any state of the world \( \omega \), it returns the number \( \omega(t_0) \). For the \( \omega \) from the picture, that will be approx. $75. You can take - (a) the net gain of the portfolio over the year, (b) the log-return in the second month, or (c) the maximum value of the portfolio - and express them as functions of \( \omega \). They will all be random variables (under the reasonable assumptions, always verified in practice, that is):

(a) \( \omega(1) - \omega(0) \), \hspace{1cm} (b) \( \log(\omega(\frac{2}{12})/\omega(\frac{1}{12})) \), \hspace{1cm} (c) \( \max_{0 \leq t \leq 1} \omega(t) \).

Our discussion of random variables so far never mentioned the probability. Suppose that there is a probability \( \mathbb{P} \) defined on \( \Omega \), and that we are given a random variable \( X \). We might be interested in probabilities of various events related to \( X \). For example, we might want to know the probability \( \mathbb{P} \{ \omega \in \Omega : a < X(\omega) < b \} \) (\( \mathbb{P}[a < X < b] \) in the standard shorthand).

This is where the notion of the distribution of the random variable comes in handy. Even though the probability spaces differ hugely from one application to another, it will turn out that there is a handful of distributions of random variables that come up over and over again. So, let us give a definition first

**Definition 1.2.5.** Let \( X : \Omega \rightarrow \mathbb{R} \) be a random variable on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). The function \( F : \mathbb{R} \rightarrow [0, 1] \) defined by \( F(x) \triangleq \mathbb{P}[X \leq x] \) is called the distribution function (or just the distribution) of the random variable \( X \).

In a sense, the distribution function tells us all about \( X \) when we take it out of the context, and sometimes that is enough. For the purposes of applications, it is useful to single out two important classes of random variables discrete and continuous\(^1\).

**Definition 1.2.6.** Let \( X \) be a random variable on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

(a) we say \( X \) is a **discrete random variable** if there exists a sequence \((x_n)_{n \in \mathbb{N}}\) such that

\[
\mathbb{P}[X \in \{x_1, x_2, \ldots\}] = \sum_{n=1}^{\infty} \mathbb{P}[X = x_n] = 1.
\]

\(^1\)Caution: there are random variables which are neither discrete nor continuous, but we will have no use for them in this course.
In other words, $X$ always has some $x_n$ as its value.

(b) we say $X$ is a **continuous random variable** if there exists a integrable function $f_X : \mathbb{R} \to \mathbb{R}_+$ such that

$$P[a < X < b] = \int_a^b f_X(x) \, dx, \text{ for all } a < b. \tag{1.2.1}$$

When $X$ is a discrete random variable, the probabilities of all events related to $X$ can be expressed in terms of the probabilities $p_n = P[X = x_n]$. For example, $P[a \leq X \leq b] = \sum\{n : a \leq x_n \leq b\} p_n$. In particular, $F(x) = \sum\{n : x_n \leq x\} p_n$. In the case when the sequence $(x_n)_{n \in \mathbb{N}}$ contains only finitely many terms (or, equivalently, when there exists $k \in \mathbb{N}$ such that $p_n = 0$, for $n > k$), the distribution of $X$ can be represented in the form of the **distribution table**

$$X \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_k \\ p_1 & p_2 & \ldots & p_k \end{pmatrix}$$

For a continuous random variable $X$, the function $f_X$ from the definition 1.2.6.(b), is called the **density function** of the random variable $X$, or simply the **density** of $X$. For a random variable $X$ with the density $f_X$, the relation (1.2.1) holds true even when $a = -\infty$ or $b = \infty$. It follows that

$$F(x) = \int_{-\infty}^x f_X(x) \, dx, \text{ and that } \int_{-\infty}^\infty f_X(x) = 1,$$

since $P[X \in \mathbb{R}] = 1$.

Given all these definitions, it would be fair to give an example or two before we proceed.

**Example 1.2.7.** This is the list of Derek’s favorite distributions. He is sort of a fanatic when it comes to collecting different famous distributions. He also likes the trivia that comes with them.

1. **Discrete distributions**

   (a) **(Bernoulli: $B(p)$)** The Bernoulli distribution is the simplest discrete distribution. It takes only 2 different values: 1 and 0, with probabilities $p$ and $1 - p$, respectively. Sometimes 1 is called *success* and 0 *failure*. If $X$ is the outcome of tossing of an unfair coin, i.e. a coin with unequal probabilities for head and tails, then $X$ will have the Bernoulli distribution. Even though it might not seem like a very interesting distribution, it is ubiquitous, and it is very useful as a building block for more complicated models. Just think about binomial trees in discrete-time finance. Or winning a lottery, or anything that has exactly two outcomes.
(b) (Binomial: $B(n, p)$) Adding the outcomes of $n$ unrelated (I should say independent here) but identical Bernoulli random variables gives you a binomial random variable. If you adopt the success-failure interpretation of a Bernoulli random variable, the Binomial says how many successes you have had in $n$ independent trials. The values a Binomial random variable can take are $x_0 = 0$, $x_1 = 1$, $\ldots$, $x_n = n$, with probabilities

$$p_m = \mathbb{P}[X = m] = \binom{n}{m} p^m (1 - p)^{n-m},$$

for $m = 0, 1, \ldots, n$.

(c) (Poisson, $P(\lambda)$) The Poisson distribution is a limiting case of a Binomial distribution when $n$ is very large, $p$ is very small and $np \sim \lambda$. The probability distribution is given by $p_n = \mathbb{P}[X = n] = e^{-\lambda} \frac{\lambda^n}{n!}$ for $n = 0, 1, \ldots$.

(d) (Geometric, $G(p)$) If you are tossing an unfair coin ($\mathbb{P}[\text{Tails}] = p$), the Geometric distribution will be the distribution of the number of tosses it takes to get your first Tails. The values it can take are $n = 1, 2, \ldots$ and the corresponding probabilities are $p_n = p(1 - p)^{n-1}$. 
2. Continuous distributions

(a) \( f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x < a \text{ or } x > b \end{cases} \)

The Uniform distribution models complete ignorance of an unknown quantity as long as it is constrained to take values in a bounded interval. The parameters \( a \) and \( b \) denote the end-points of that interval, and the density function is given by

\[ f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x < a \text{ or } x > b \end{cases} \]

(b) \( f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right), \quad x \in \mathbb{R} \)

The normal distribution was originally studied by DeMoivre (1667-1754), who was curious about its use in predicting the probabilities in gambling! The first person to apply the normal distribution to social data was Adolph Quetelet (1796-1874). He collected data on the chest measurements of Scottish soldiers, and the heights of French soldiers, and found that they were normally distributed. His conclusion was that the mean was nature’s ideal, and data on either side of the mean were a deviation from nature’s ideal. Although his conclusion is arguable, he nonetheless represented normal distribution in a real-life setting. The density function of the normal distribution with parameters \( \mu \) (mean), \( \sigma \) (standard deviation) is

\[ f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right), \quad x \in \mathbb{R} \]
1.2. RANDOM VARIABLES

CHAPTER 1. BROWNIAN MOTION

(c) **Exponential** distribution is the continuous-time analogue of the Geometric distribution. It usually models waiting times and has a nice property of no memory. Its density is

\[ f(x) = \begin{cases} \lambda \exp(-\lambda x), & x > 0 \\ 0, & x \leq 0, \end{cases} \]

where the parameter \( \lambda \) stands for the rate - i.e. the reciprocal of the expected waiting time.

(d) **Double Exponential** distribution arises as a modification of the Exponential distribution when the direction of the uncertainty is unknown - like an upward or a downward movement of the (log-) stock price. The density function is given by

\[ f(x) = \frac{1}{2\lambda} \exp(-\lambda|x|), \ x \in \mathbb{R} \]

where \( \lambda \) has the same meaning as in the previous example.

(e) **Arcsin** distribution. Consider a baseball team that has a 0.5 probability of winning each game. What percentage of the season would we expect it to have a losing record (more games lost than won so far)? A winning record? The rather unexpected answer is supplied by the Arcsin distribution. Its density function is

\[ f(x) = \begin{cases} \frac{1}{\pi \sqrt{x(1-x)}}, & x \in (0,1) \\ 0, & x \leq 0 \text{ or } x \geq 1. \end{cases} \]
CHAPTER 1. BROWNIAN MOTION 1.2. RANDOM VARIABLES

Exercises

Exercise 1.2.8. Let $X$ and $Y$ be two continuous random variables on the same probability space. It is true that their sum is a continuous random variable? What about the sum of two discrete random variables - is their sum always discrete?

Exercise 1.2.9. Let $X$ be a continuous random variable with the density function $f(x)$ such that $f(x) > 0$ for all $x$. Furthermore, let $F(x) = \int_{-\infty}^{x} f(x) \, dx$ be the distribution function of $X$: $F(x) = P[X \leq x]$. We define the random variable $Y$ by $Y(\omega) = F(X(\omega))$. Prove that $Y$ is a uniformly distributed continuous random variable with parameters $a = 0$ and $b = 1$. In other words, $Y \sim U(0, 1)$.

Exercise 1.2.10. Let $X$ be a standard normal random variable, i.e. $X \sim N(0,1)$. Prove that for $n \in \mathbb{N} \cup \{0\}$, we have

$$
E[X^n] = \begin{cases} 
1 \cdot 3 \cdot 5 \cdots (n-1), & \text{if } n \text{ is even}, \\
0, & \text{if } n \text{ is odd}.
\end{cases}
$$

Exercise 1.2.11. Let $X$ be an exponentially distributed random variable with parameter $\lambda > 0$, i.e. $X \sim \text{Exp}(\lambda)$. Compute the density $f_Y(y)$ of the random variable $Y = \log(X)$.

(Hint: compute the distribution function $F_Y(y)$ of $Y$ first.)
Solutions to Exercises in Section 1.2

Solution to Exercise 1.2.8: The first statement is not true. Take $X$ to be any continuous random variable and take $Y = -X$. Then $Z = X + Y$ is the constant random variable $0$, i.e. $\mathbb{P}[Z = 0] = 1$. Suppose there exists a density $f_Z$ for the random variable $Z$. The function $f_Z$ should have the property that

- $f_Z(z) \geq 0$, for all $z > 0$. 
- $\int_{-\infty}^{\infty} f_Z(z) \, dz = 1$
- $\int_a^b f_Z(z) \, dz = 0$ for every interval $(a,b)$ which does not contain $0$.

It is quite obvious that such a function cannot exist. For those of you who have heard about “Dirac’s delta function”, I have to mention the fact that it is not a function even though physicists like to call it that way.

The second statement is true. Let $\{x_n\}_{n \in \mathbb{N}}$ be (the countable) sequence of values that $X$ can take, and let $\{y_n\}_{n \in \mathbb{N}}$ be the sequence of values the random variable $Y$ can take. The sum $Z = X + Y$ will always be equal to some number of the form $x_n + y_m$, and it is thus enough to prove that there are countably many numbers which can be written in the form $x_n + y_m$. In order to prove that statement, let us first note that the number of numbers of the form $x_n + y_m$ is “smaller” than the number of ordered pairs $(n, m)$. This follows from the fact that the pair $(n, m)$ determines uniquely the sum $x_n + y_m$, while the same sum might correspond to many different pairs. But we know that the sets $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$ both have countably many elements, which concludes the proof.

Solution to Exercise 1.2.9: Let us compute the distribution function of the random variable $Y$: (note that $F_X^{-1}$ exists on $(0,1)$ since $F_X$ is a strictly increasing function due to the strict positivity of $f_X$)

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[F_X(X) \leq y] = \begin{cases} 1, & y > 1 \\ \mathbb{P}[X \leq F_X^{-1}(y)] = F_X(F_X^{-1}(y)) = y, & y \in (0,1) \\ 0, & y < 0. \end{cases}$$

It is clear now that $F_Y(y) = \int_{-\infty}^{y} f_{U(0,1)}(z) \, dz$, where

$$f_{U(0,1)} = \begin{cases} 0, & y < 0 \text{ or } y > 1 \\ 1, & 0 < y < 1, \end{cases}$$

is the density of the uniform random variable with parameters $a = 0$ and $b = 1$. 

---

Last Updated: Spring 2005
Solution to Exercise 1.2.10: Let $f(x)$ denote the density of the standard normal: $f(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$. Then

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) \, dx.$$ 

Letting $a_n = E[X^n]$ we immediately note that $a_n = 0$ whenever $n$ is odd, because of the symmetry of the function $f$. For $n = 0$, we have $a_0 = 1$ since $f$ is a density function of a random variable. Finally, we use partial integration, L’Hospital rule and the fact that $f'(x) = -xf(x)$ (where exactly are they used?) to obtain

$$a_{2n} = \int_{-\infty}^{\infty} x^{2n} f(x) \, dx = \frac{x^{2n+1} f(x)}{2n+1} \bigg|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{x^{2n+1}}{2n+1} (-xf(x)) \, dx = \frac{a_{2n+2}}{2n+1}.$$ 

Having obtained the recursive relation $a_{2n+2} = (2n+1)a_{2n}$ the claim now follows by induction.

Solution to Exercise 1.2.11: The distribution function $F_X$ of the exponentially distributed random variable $X$ with parameter $\lambda$ is given by

$$F_X(x) = \begin{cases} \int_0^x \lambda e^{-\lambda y} \, dy, & x \geq 0 \\ 0, & x < 0 \end{cases} = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Therefore, for $y \in \mathbb{R}$ we have

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P} [\log(X) \leq y] = \mathbb{P}[X \leq e^y] = F_X(e^y) = 1 - e^{-\lambda e^y}.$$ 

Differentiation with respect to $y$ gives

$$f_Y(y) = \lambda e^{y-\lambda e^y}.$$
1.3 Random Vectors

1.3.1 Random Vectors Generalia

Random variables describe numerical aspects of the phenomenon under observation that take values in the set of real numbers \( \mathbb{R} \). Sometimes we might be interested in a piece of information consisting of more than one number - say the height and the weight of a person, or the prices of several stocks in our portfolio. Therefore, we are forced to introduce **random vectors** - (measurable functions) from \( \Omega \) to \( \mathbb{R}^n \).

There is a number of concepts related to random vectors that are direct analogies of the corresponding notions about random variables. Let us make a list. In what follows, let \( \mathbf{X} = (X_1, X_2, \ldots, X_n) \) be an \( n \)-dimensional random vector.

1. The function \( F_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R} \), defined by
   \[
   F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n],
   \]
   is called the **distribution function** of the random vector \( \mathbf{X} \).

2. \( \mathbf{X} \) is said to be a **continuous random vector** if there exists a nonnegative integrable function \( f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R} \) such that for \( a_1 < b_1, a_2 < b_2, \ldots, a_n < b_n \) we have
   \[
   \mathbb{P}[a_1 < X_1 < b_1, a_2 < X_2 < b_2, \ldots, a_n < X_n < b_n] =
   \int_{a_n}^{b_n} \ldots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{\mathbf{X}}(y_1, y_2, \ldots, y_n) \, dy_1 \, dy_2 \ldots \, dy_n
   \]
   Subject to some regularity conditions (you should not worry about them) we have
   \[
   f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\mathbf{X}}(x_1, x_2, \ldots, x_n).
   \]

**Example 1.3.1.** Let \( \mathbf{X} = (X_1, X_2, X_3) \) be a random vector whose density function \( f_{\mathbf{X}} \) is given by
   \[
   f_{\mathbf{X}}(x_1, x_2, x_3) \triangleq \begin{cases} 
   \frac{2}{3}(x_1 + x_2 + x_3), & 0 \leq x_1, x_2, x_3 \leq 1 \\
   0, & \text{otherwise.}
   \end{cases}
   \]
   It is easy to show that \( f_{\mathbf{X}} \) indeed possesses all the properties of a density function (positive, integrates to one). Note that the random vector \( \mathbf{X} \) takes values in the unit cube \([0, 1]^3\) with probability one. As for the distribution function \( F_{\mathbf{X}} \), it is not impossible to compute it explicitly, even though the final result will be quite messy (there are quite a few cases to consider). Let us just mention that \( F_{\mathbf{X}}(x_1, x_2, x_3) = 0 \) for \( x_1 < 0, x_2 < 0 \) or \( x_3 < 0 \) and \( F_{\mathbf{X}}(x_1, x_2, x_3) = 1 \) for all \( x_1, x_2, x_3 \) are greater than 1.

Even though the definition of a density teaches us only how to compute probabilities of the from \( \mathbb{P}[a_1 < X_1 < b_1, a_2 < X_2 < b_2, a_3 < X_3 < b_3] \), we can compute more complicated...
CHAPTER 1. BROWNIAN MOTION

1.3. RANDOM VECTORS

probabilities. For nice enough set \( A \subseteq \mathbb{R}^3 \) we have

\[
P[X \in A] = \iiint_A f_X(y_1, y_2, y_3) \, dy_1 \, dy_2 \, dy_3. \tag{1.3.1}
\]

In words, the probability that \( X \) lies in the region \( A \) can be obtained by integrating \( f_X \) over that region. Of course, there is nothing special about \( n = 3 \) here. The same procedure will work in any dimension.

**Example 1.3.2.** Let \( X = (X_1, X_2, X_3) \) be the random vector from Example 1.3.1. Suppose we are interested in the probability \( P[X_1 - 2X_2 \geq X_3] \). The first step is to cast the inequality \( X_1 - 2X_2 \geq X_3 \) into the form \( X \in A \) for a suitably chosen set \( A \subseteq \mathbb{R}^3 \). In our case \( A = \{ (x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - 2x_2 - x_3 \geq 0 \} \), and so the required probability can be obtained by using formula (1.3.1):

\[
P[X_1 - 2X_2 \geq X_3] = \iiint_{(x_1,x_2,x_3)\in\mathbb{R}^3:x_1-2x_2-x_3\geq0} f_X(x_1, x_2, x_3) \, dx_1 \, dx_2 \, dx_2.
\]

Furthermore, since \( f_X \) is equal to 0 outside the cube \( 0 < x_1, x_2, x_3 < 1 \), we have reduced the problem of finding the probability \( P[X_1 - 2X_2 \geq X_3] \) to the computation of the integral

\[
\iiint_{0<x_1,x_2,x_3<1,x_1-2x_2-x_3\geq0} \frac{2}{3}(x_1 + x_2 + x_3) \, dx_1 \, dx_2 \, dx_3.
\]

The Maple command

\[
> \text{int(int(int(2/3*(x+y+z)*piecewise(x-2*y-z>0, 1,0),x=0..1),y=0..1),z=0..1)};
\]

does the trick, and we get \( P[X_1 - 2X_2 \geq X_3] = \frac{1}{16} \).

1.3.2 Marginal Distributions

For a component random variable \( X_k \) of the random vector \( X \) there is a simple way of obtaining its distribution function \( F_{X_k} \) from the distribution of the random vector (why is that so?)

\[
F_{X_k}(x) = \lim_{x_1 \to \infty, \ldots, x_{k-1} \to \infty, x_{k+1} \to \infty, \ldots, x_n \to \infty} F(x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n).
\]

You would get the distribution function \( F_{X_k,X_l}(x, x') \) in a similar way by holding the \( k^{th} \) and the \( l^{th} \) coordinate fixed and taking the limit of \( F_X \) when all the other coordinates tend to \( \infty \), etc.

For the densities of continuous random vectors, similar conclusions are true. Suppose we are interested in the density function of the random vector \( X' \triangleq (X_2, X_3) \), given that the
density of $X = (X_1, X_2, \ldots, X_n)$ is $f_X(x_1, x_2, \ldots, x_n)$. The way to go is to \textit{integrate out} the variables $X_1, X_4, \ldots, X_n$:

$$f_{X'}(x_2, x_3) = f_{(X_2, X_3)}(x_2, x_3) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(y_1, x_2, x_3, y_4, \ldots, y_n) \, dy_1 \, dy_4 \cdots dy_n.$$  

You would, of course, do the analogous thing for any $k$-dimensional sub-vector $X'$ of $X$. When $k = 1$, you get distributions (densities) of random variables $X_1, X_2, \ldots, X_n$. The distribution (density) of a sub-vector $X'$ of $X$ is called the \textit{marginal distribution (density)}.

\textbf{Example 1.3.3}. Continuing with the random vector from Example 1.3.1, the random variable $X_2$ has a density $f_{X_2}$ given by

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(y_1, x_2, y_3) \, dy_1 \, dy_3.$$  

The fact that $f_X$ is 0 when any of its arguments is outside $[0, 1]^3$, we get

$$f_{X_2}(x_2) = \int_0^{1} \int_0^{1} \frac{2}{3}(y_1 + x_2 + y_3) \, dy_1 \, dy_3.$$  

By simple integration (or Maple) we have:

$$f_{X_2}(x_2) = \begin{cases} \frac{2}{3}(1 + x_2), & 0 < x_2 < 1 \\ 0, & \text{otherwise}. \end{cases}$$  

By symmetry $f_{X_1} = f_{X_2} = f_{X_3}$. Note that the obtained function $f_{X_2}$ is positive and integrates to 1 - as it should. Similarly,

$$f_{(X_2, X_3)}(x_2, x_3) = \int_{-\infty}^{\infty} f_X(y_1, x_2, x_3) \, dy_1 = \begin{cases} \frac{2}{3}(\frac{1}{2} + x_2 + x_3), & 0 < x_2, x_3 < 1 \\ 0, & \text{otherwise}. \end{cases}$$  

\subsection*{1.3.3 Expectations, etc.}

Given a random vector $X$ we define its \textbf{expectation vector (or mean vector)} $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ by the formula:

$$\mu_k = \int_{-\infty}^{\infty} x_k f_X(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \ldots \, dx_n,$$

provided that all the integrals converge absolutely, i.e. $\int_{-\infty}^{\infty} |x_k| \, f_X(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \ldots \, dx_n < \infty$, for all $k$. Otherwise, we say that the expectation \textbf{does not exist}. In order not to repeat this discussion every time when dealing with integrals, we shall adopt the following convention and use it tacitely throughout the book:

\textbf{Last Updated: Spring 2005} 22  Continuous-Time Finance: Lecture Notes
Any quantity defined by an integral will be said not to exist if its defining integral does not converge absolutely.

It is customary in probability theory to use the notation $\mathbb{E}[X]$ for the expectation $\mu$ of $X$. It can be shown that $\mathbb{E}$ acts linearly on random vectors, i.e. if $X$ and $X'$ are two random vectors defined on the same probability space $\Omega$, and of the same dimension $n$, then for $\alpha, \beta \in \mathbb{R}$,

$$\mathbb{E}[\alpha X + \beta X'] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[X'].$$

See Example 1.3.9 for a proof of this statement.

In general, for any function $g : \mathbb{R}^n \to \mathbb{R}$ we define

$$\mathbb{E}[g(X)] \triangleq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \, \ldots \, dx_n.$$  

When $g$ takes values in $\mathbb{R}^m$, $\mathbb{E}[g(X)]$ is defined componentwise, i.e.

$$\mathbb{E}[g(X)] = \left( \mathbb{E}[g_1(X)], \mathbb{E}[g_2(X)], \ldots, \mathbb{E}[g_m(X)] \right)$$

where $g = (g_1, g_2, \ldots, g_m)$. The definition of the expectation is just the special case with $g(x_1, x_2, \ldots, x_n) = (x_1, x_2, \ldots, x_n)$.

### 1.3.4 The Variance-Covariance Matrix

Just as the expectation is used as a measure of a center of a random vector, the measure of spread is provided by the variance-covariance matrix $\Sigma$, i.e. the matrix whose $(i,j)^{th}$ entry is given by

$$\Sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j)f_X(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \, \ldots \, dx_n,$$

where $\mu_i = \mathbb{E}[X_i]$ and $\mu_j = \mathbb{E}[X_j]$. When $n = 1$, we retrieve the familiar concept of variance ($\mu = \mathbb{E}[X_1]$):

$$\Sigma = (\Sigma_{11}) = \text{Var}[X_1] = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) \, dx.$$  

One easily realizes (try to prove it!) that for general $n$, we have

$$\Sigma_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$

This quantity is called the covariance between random variables $X_i$ and $X_j$, and is usually denoted by $\text{Cov}(X_i, X_j)$. We will give examples of the meaning and the computation of the variance-covariance matrix in the section devoted to the multivariate normal distribution.
1.3. RANDOM VECTORS

Using the agreement that the expectation \( E[A] \) of a matrix \( A \) (whose entries are random variables), is just the matrix of expectations of entries of \( A \), we can define the variance-covariance matrix of the random vector \( X \) simply by putting \((T\)-denotes the transposition of the vector\)

\[
\Sigma \triangleq E[(X - \mu)^T(X - \mu)].
\]  
(1.3.2)

(Note: by convention \( X \) is a row-vector, so that \( X^T \) is a column vector and \((X - \mu)^T(X - \mu)\) is a \( n \times n \)-matrix.) Using this representation one can easily show the following statement.

**Proposition 1.3.4.** Let \( X \) be a random vector and let \( B \) be a \( m \times n \) matrix of real numbers. Then the variance-covariance matrix \( \Sigma' \) of the random vector \( X' = XB \) is given by

\[
\Sigma' = B^T \Sigma B,
\]  
(1.3.3)

where \( \Sigma \) is the variance-covariance matrix of the random vector \( X \).

1.3.5 Independence

It is important to notice that the distribution of a random vector carries more information than the collection of the distribution functions of the component random variables. This extra information is sometimes referred to as the dependence structure of the random vector. It is quite obvious that height and weight may both be normally distributed, but without the knowledge of their joint distribution - i.e. the distribution of the random vector (height, weight) - we cannot tell whether taller people tend to be heavier than the shorter people, and if so, whether this tendency is strong or not. There is, however, one case when the distributions of the component random variables determine exactly the distribution of the vector:

**Definition 1.3.5.** Random variables \( X_1, X_2, \ldots, X_n \) are said to be independent if

\[
F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_n}(x_n) = F_X(x_1, x_2, \ldots, x_n),
\]

where \( X \triangleq (X_1, X_2, \ldots, X_n) \), and \( F_{X_1}, F_{X_2}, \ldots F_{X_n} \) are the distribution functions of the component random variables \( X_1, X_2, \ldots, X_n \).

The definition implies (well, it needs a proof, of course) that for independent continuous random variables, the density function \( f_X \) factorizes in a nice way:

\[
f_X(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n),
\]  
(1.3.4)

and, furthermore, the converse is also true: if the densities of the random variables \( X_1, X_2, \ldots, X_n \) satisfy (1.3.4), then they are independent.
CHAPTER 1. BROWNIAN MOTION

1.3. RANDOM VECTORS

**Example 1.3.6.** For the random vector $X$ from Example 1.3.1, the component random variables $X_1$, $X_2$ and $X_3$ are not independent. If they were, we would have (look at the previous example)

$$
\frac{2}{3}(x_1 + x_2 + x_3) = f_X(x_1, x_2, x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) = \frac{8}{27}(1 + x_1)(1 + x_2)(1 + x_3),
$$

for all $0 \leq x_1, x_2, x_3 \leq 1$. This is obviously a contradiction.

Finally, let me mention that (1.3.4) implies easily the validity of the following theorem, and the corresponding rule: *independence means multiply!*

**Theorem 1.3.7.** Let random variables $X_1, X_2, \ldots, X_n$ be independent. For all (sufficiently nice) real functions $h_1 : \mathbb{R} \to \mathbb{R}$, $h_2 : \mathbb{R} \to \mathbb{R}$, ..., $h_n : \mathbb{R} \to \mathbb{R}$ we have

$$
\mathbb{E}[h_1(X_1)h_2(X_2)\ldots h_n(X_n)] = \mathbb{E}[h_1(X_1)]\mathbb{E}[h_2(X_2)]\ldots\mathbb{E}[h_n(X_n)],
$$

subject to existence of all expectations involved. In particular, if $X$ and $Y$ are independent,

$$
\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \text{ and } \text{Cov}(X, Y) = 0.
$$

How does the last formula in the theorem follow from the rest?

1.3.6 Functions of Random Vectors

Let $X$ be a random vector with density $f_X$ and let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a function. The composition of the function $h$ and the random vector $X$ (viewed as a function $\Omega \to \mathbb{R}^n$) is a random vector again. The natural question to ask is whether $X' \triangleq h(X)$ admits a density, and if it does, how can we compute it? The answer is positive, and we will state it in a theorem which we give without the proof. Hopefully, everything will be much clearer after an example.

**Theorem 1.3.8.** Let $h : \mathbb{R}^n \to \mathbb{R}^n$, $h = (h_1, h_2, \ldots, h_n)$ be a one-to-one mapping such that $h$ and its inverse $h^{-1}$ are continuously differentiable. (The inverse being defined on a subset $B \subseteq \mathbb{R}^n$). Then $X' \triangleq h(X)$ admits a density $f_{X'}$, and it is given by the formula

$$
f_{X'}(x'_1, x'_2, \ldots, x'_n) = f_X\left(h^{-1}(x'_1, x'_2, \ldots, x'_n)\right)\left|\text{det } Jh^{-1}(x'_1, x'_2, \ldots, x'_n)\right|,
$$

for $(x'_1, x'_2, \ldots, x'_n) \in B$ and $f_X(x'_1, x'_2, \ldots, x'_n) = 0$ outside $B$. Here, for a differentiable function $g$, $|\text{det } Jg(x'_1, x'_2, \ldots, x'_n)|$ denotes the absolute value of the determinant of the Jacobian matrix $Jg$ (matrix whose $ij$-the entry is the partial derivative of the $i$th component function $g_i$ with respect to $j$th variable $x'_j$.)

**Example 1.3.9.**
1. When \( n = 1 \), the formula in Theorem 1.3.8 is particularly simple. Let us illustrate by an example. Suppose that \( X \) has a standard normal distribution, and define \( h(x) \triangleq \exp(x) \). The density of the resulting random variable \( X' = \exp(X) \) is then given by

\[
f_{X'}(x') = \frac{1}{x'\sqrt{2\pi}} \exp \left( -\frac{1}{2} \log^2(x') \right) \left( = f_X(\log(x')) \left| \frac{d}{dx'} \log(x') \right| \right),
\]

for \( x' > 0 \) and \( f_{X'}(x') = 0 \) for negative \( x' \). The random variable \( X' \) is usually called log-normal and is used very often to model the distribution of stock returns.

2. Let \( \mathbf{X} \) be a 2-dimensional random vector \( \mathbf{X} = (X_1, X_2) \) with density \( f_{\mathbf{X}}(x_1, x_2) \). We are interested in the density function of the sum \( X_1 + X_2 \). In order to apply Theorem 1.3.8, we need an invertible, differentiable function \( h : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \), whose first component \( h_1 \) will be given by \( h_1(x_1, x_2) = x_1 + x_2 \). A natural candidate will be the function \( h : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) given by \( h(x_1, x_2) = (x_1 + x_2, x_1 - x_2) \).

Moreover, both \( h \) and \( h^{-1} \) are linear functions and we have

\[
h^{-1}(x'_1, x'_2) = \left( \frac{x'_1 + x'_2}{2}, \frac{x'_1 - x'_2}{2} \right).
\]

The Jacobian matrix of the function \( h^{-1} \) is given by

\[
Jh^{-1}(x'_1, x'_2) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix},
\]

so \( |\det Jh^{-1}(x'_1, x'_2)| = \frac{1}{2} \). Therefore

\[
f_{\mathbf{X}'}(x'_1, x'_2) = \frac{1}{2} f_{\mathbf{X}} \left( \frac{x'_1 + x'_2}{2}, \frac{x'_1 - x'_2}{2} \right).
\]

We are only interested in the density of the random variable \( X_1 + X_2 \) which is the first component of the random vector \( \mathbf{X}' = h(\mathbf{X}) \). Remembering from subsection 1.3.2 how to obtain the marginal from the joint density, we have

\[
f_{X_1+X_2}(x) = \frac{1}{2} \int_{-\infty}^{\infty} f_{\mathbf{X}} \left( \frac{x + x'_2}{2}, \frac{x - x'_2}{2} \right) dx'_2. \tag{1.3.5}
\]

By introducing a simple change of variables \( y = \frac{x + x'_2}{2} \), the expression above simplifies to

\[
f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(y, x - y) dy. \tag{1.3.6}
\]
In the special case when $X_1$ and $X_2$ are independent, we have $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ and the formula for $f_{X_1+X_2}(x)$ becomes

$$f_{X_1+X_2}(x) = (f_{X_1} \ast f_{X_2})(x) = \int_{-\infty}^{\infty} f_{X_1}(y)f_{X_2}(x-y) \, dy.$$  

The operation $f_{X_1} \ast f_{X_2}$ is called the convolution of functions $f_{X_1}$ and $f_{X_2}$. Note that we have just proved that the convolution of two positive integrable functions is a positive integrable function, and that

$$\int_{-\infty}^{\infty} (f \ast g)(y) \, dy = \left( \int_{-\infty}^{\infty} f(x) \, dx \right) \left( \int_{-\infty}^{\infty} g(x') \, dx' \right).$$

How exactly does the last statement follow from the discussion before?

3. Let us use the result we have just obtained to prove the linearity of mathematical expectation. Let $X_1$ and $X_2$ be random variables on the same probability space. We can stack one on top of the other to obtain the random vector $X = (X_1, X_2)$. By (1), the density of the sum $X_1 + X_2$ is then given by (1.3.6), so (assuming everything is defined)

$$E[X_1 + X_2] = \int_{-\infty}^{\infty} x f_{X_1+X_2}(x) \, dx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X_1}(y)f_{X_2}(x-y) \, dy \, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X_1}(y)f_{X_2}(x-y) \, dy \, dx.$$  

After change of variables $\xi = x - y$, we get

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X_1}(y)f_{X_2}(x-y) \, dy \, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\xi + y) f_{X_1}(y)f_{X_2}(\xi) \, dy \, d\xi =$$

$$= \int_{-\infty}^{\infty} y \left( \int_{-\infty}^{\infty} f_{X_1}(\xi) \, d\xi \right) f_{X_2}(y) \, dy + \int_{-\infty}^{\infty} \xi \left( \int_{-\infty}^{\infty} f_{X_1}(y) \, dy \right) f_{X_2}(\xi) \, d\xi =$$

$$= \int_{-\infty}^{\infty} y f_{X_2}(y) \, dy + \int_{-\infty}^{\infty} \xi f_{X_1}(\xi) \, d\xi = E[X_2] + E[X_1].$$

You prove that $E[\alpha X] = \alpha E[X]$ (you do not need the strength of Theorem 1.3.8 to do that. Just compute the density of $\alpha X$ from the first principles).
1.3. RANDOM VECTORS

CHAPTER 1. BROWNIAN MOTION

Exercises
Solutions to Exercises in Section 1.3
1.4 The Multivariate Normal Distribution

1.4.1 The Definition

The multivariate normal distribution (MND) is one of the most important examples of multivariate distributions. It is a direct generalization of the univariate (standard) normal and shares many of its properties. Besides from being analytically tractable as well as very applicable in modelling a host of every-day phenomena, the MND often arises as the limiting distribution in many multidimensional Central Limit Theorems.

Definition 1.4.1. A random vector \( X = (X_1, X_2, \ldots, X_n) \) is said to have a multivariate normal distribution if for any vector \( a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n \) the distribution of the random variable \( Xa^T = a_1X_1 + a_2X_2 + \cdots + a_nX_n \) has a univariate normal distribution (with some mean and variance depending on \( a \)).

Putting \( a_i = 1, a_k = 0, \) for \( k \neq i \), the definition above states that each component random variable \( X_i \) has a normal distribution, but the converse is not true: there are random vectors whose component distributions are normal, but the random vector itself is not multivariate normal (See Appendix, Section A.2). The following theorem reveals completely the structure of a multivariate normal random vector:

Theorem 1.4.2. Let the random vector \( X = (X_1, X_2, \ldots, X_n) \) have a multivariate normal distribution, let \( \mu = (\mu_1, \mu_2, \ldots, \mu_n) \) be its expectation (mean) vector, and let \( \Sigma \) be its variance-covariance matrix. Then the distribution of \( X \) is completely determined by \( \mu \) and \( \Sigma \). Also, when \( \Sigma \) is non-singular, its density \( f_X \) is given by

\[
f_X(x_1, x_2, \ldots, x_n) = (2\pi)^{-\frac{n}{2}}(\det \Sigma)^{-\frac{1}{2}} \exp \left( -\frac{1}{2}Q_{\mu, \Sigma}(x_1, x_2, \ldots, x_n) \right), \tag{1.4.1}
\]

where the function \( Q_{\mu, \Sigma} \) is a quadratic polynomial in \( x = (x_1, x_2, \ldots, x_n) \) given by

\[
Q_{\mu, \Sigma}(x) = (x - \mu)\Sigma^{-1}(x - \mu)^T = \sum_{i=1}^{n} \sum_{j=1}^{n} \tau_{ij}(x_i - \mu_i)(x_j - \mu_j),
\]

and \( \tau_{ij} \) is the \((i, j)\)th element of the inverse matrix \( \Sigma^{-1} \).

The converse is also true: if a random vector \( X \) admits a density of the form (1.4.1), then \( X \) is multivariate normal.

Note that when \( n = 1 \), the mean vector becomes a real number \( \mu \), and the variance-covariance matrix \( \Sigma \) becomes \( \sigma^2 \in \mathbb{R} \). It is easy to check (do it!) that the formula above gives the familiar density of the univariate normal distribution.

---

2A distribution is said to be multivariate if it pertains to a (multidimensional) random vector. The various distributions of random variables are often referred to as univariate.
1.4.2 The Bivariate Case

The case when \( n = 2 \) is often referred to as the \textit{bivariate normal distribution}. Before we give our next example, we need to remember the notion of \textit{correlation} \( \text{corr}(X,Y) \) between two random variables \( X \) and \( Y \):

\[
\text{corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.
\]

The number \( \text{corr}(X,Y) \) takes values in the interval \([-1, 1]\) and provides a numerical measure of the relation between the random variables \( X \) and \( Y \). When \( X \) and \( Y \) are independent, then \( \text{corr}(X,Y) = 0 \) (because \( \text{Cov}(X,Y) = 0 \), as we have seen in Section 1.3), but the converse is not true (T. Mikosch provides a nice example (Example 1.1.12) on the page 21).

\textbf{Example 1.4.3.} Let \( X = (X,Y) \) have a bivariate normal distribution. If we let \( \mu_X = \mathbb{E}[X] \), \( \mu_Y = \mathbb{E}[Y] \), \( \sigma_X = \sqrt{\text{Var}(X)} \), \( \sigma_Y = \sqrt{\text{Var}(Y)} \), and \( \rho = \text{corr}(X,Y) \), then the mean vector \( \mu \) and the variance-covariance matrix \( \Sigma \) are given by

\[
\mu = (\mu_X, \mu_Y), \quad \Sigma = \begin{pmatrix}
\sigma_X^2 & \rho \sigma_X \sigma_Y \\
\rho \sigma_X \sigma_Y & \sigma_Y^2
\end{pmatrix}.
\]

The inverse \( \Sigma^{-1} \) is not hard to compute when \( \rho^2 < 1 \), and so the quadratic \( Q_{\mu,\Sigma} \) simplifies to

\[
Q_{\mu,\Sigma}(x,y) = \frac{1}{1-\rho^2} \left( \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho \left( \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} \right) \right).
\]

Therefore, the density of the bivariate normal is given by

\[
f_X(x,y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left( \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho \left( \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} \right) \right) \right)
\]

In order to visualize the bivariate normal densities, try the following Maple commands (here we have set \( \sigma_X = \sigma_Y = 1, \mu_X = \mu_Y = 0 \)):

\[
\text{plot3d}(f(x,y,\rho) \rightarrow (1/(2*Pi*sqrt(1-\rho^2))) \cdot \exp((-1/(2*(1-\rho^2)))) \cdot f(x,y,\rho), x=-4..4, y=-4..4, grid=[40,40]); \quad \text{with(plots)}; \quad \text{contourplot}(f(x,y,\rho), x=-4..4, y=-4..4, rho=-0.8..0.8, frames=20, grid=[30,30]); \quad \text{animate3d}(f(x,y,\rho), x=-4..4, y=-4..4, rho=-0.8..0.8, frames=20, grid=[30,30]);
\]

One notices right away in the preceding example that when \( X \) and \( Y \) are uncorrelated (i.e. \( \rho = 0 \)), then the density \( f_X \) can be written as

\[
f_X(x,y) = \left( \frac{1}{\sqrt{2\pi \sigma_X^2}} \exp\left( -\frac{(x-\mu_X)^2}{2\sigma_X^2} \right) \right) \cdot \left( \frac{1}{\sqrt{2\pi \sigma_Y^2}} \exp\left( -\frac{(y-\mu_Y)^2}{2\sigma_Y^2} \right) \right),
\]

so that \( f_X \) factorizes into a product of two univariate normal density functions with parameters \( \mu_X, \sigma_X \) and \( \mu_Y, \sigma_Y \). Remember the fact that if the joint density can be written as the product of the densities of its component random variables, then the component random variables are independent. We have just discovered the essence of the following proposition:
Proposition 1.4.4. Let $X = (X, Y)$ be a bivariate normal random vector. Random variables $X$ and $Y$ are independent if and only if they are uncorrelated i.e. $\text{corr}(X, Y) = 0$.

For general multivariate normal random vectors we have the following nice result:

Proposition 1.4.5. Let $X = (X_1, X_2, \ldots, X_n)$ be a random vector whose marginal distributions (distributions of component vectors) are normal. If the collection $(X_1, X_2, \ldots, X_n)$ is independent, then $X$ has a multivariate normal distribution.

This result will enable us to construct arbitrary multivariate normal vectors from $n$ independent standard normals (but I will have to wait until the section about simulation to see give you more details). Also, we can prove a well-known robustness property of normal distributions in the following example:

Example 1.4.6. Statement: Let $X$ and $Y$ be two independent normally distributed random variables. Then, their sum $X + Y$ is again a normally distributed random variable.

How do we prove that? The idea is to put $X$ and $Y$ into a random vector $X = (X, Y)$ and conclude that $X$ has a bivariate normal distribution by the Proposition 1.4.5. But then, by the definition of multivariate normality, any linear combination $\alpha X + \beta Y$ must be normal. Take $\alpha = \beta = 1$, and you are done.

To recapitulate, here are the most important facts about multivariate normals:

- linear combinations of the components of a multivariate normal are normal
- the multivariate normal distribution is completely determined by its mean and the variance-covariance matrix: once you know $\mu$ and $\Sigma$, you know the density (if $\Sigma$ is invertible) and everything else there is to know about $X$
- bivariate normals have (relatively) simple densities and are completely determined by $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ and $\rho$
- for bivariate normals independence is equivalent to $\rho = 0$.
- random vectors with independent and normally distributed components are multivariate normal

Last Updated: Spring 2005
CHAPTER 1. BROWNIAN MOTION

THE MULTIVARIATE NORMAL DISTRIBUTION

Exercises

Exercise 1.4.7.

1. Let $X = (X_1, X_2, \ldots, X_n)$ have a multivariate normal distribution with mean $\mu$ and the variance-covariance matrix $\Sigma$, and let $B$ be an $n \times m$ matrix. Show that the $m$-dimensional random vector $XB$ has a multivariate normal distribution (Hint: use the definition) and compute its mean and the variance-covariance matrix in terms of $\mu$, $\Sigma$ and $B$.

2. Let $X = (X_1, X_2, X_3)$ be multivariate normal with mean $\mu = (1, 2, 3)$ and the variance-covariance matrix

$$
\Sigma = \begin{pmatrix}
5 & 4 & -9 \\
4 & 6 & -7 \\
-9 & -7 & 22
\end{pmatrix}.
$$

Further, let $B$ be a $3 \times 2$ matrix

$$
B = \begin{pmatrix}
1 & 3 \\
4 & 3 \\
2 & 7
\end{pmatrix}.
$$

Find the distribution of the bivariate random vector $XB$. Use software!

Exercise 1.4.8. Derek the Daisy owns a portfolio containing two stocks MHC (Meadow Honey Company) and BII (Bee Industries Incorporated). The statistical analyst working for Mr. Mole says that the values of the two stocks in a month from today can be modelled using a bivariate normal random vector with means $\mu_{MHC} = 101$, $\mu_{BII} = 87$, variances $\sigma_{MHC} = 7$, $\sigma_{BII} = 10$ and $\text{corr}(\text{MHC, BII}) = 0.8$ (well, bees do make honey!). Help Derek find the probability that both stocks will perform better than their means, i.e.

$$
P[MHC \geq 101, BII \geq 87].
$$

(Hint: of course you can always find the appropriate region $A$ in $\mathbb{R}^2$ and integrate the joint density over $A$, but there is a better way. Try to transform the variables so as to achieve independence, and then just use symmetry.)

Exercise 1.4.9. Let $X = (X_1, X_2)$ be a bivariate normal vector with mean $\mu = (2, 3)$ and variance-covariance matrix $\Sigma = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$.
1. Find the distributions of the following random variables:
   
   (a) \( X_1 \)
   
   (b) \( X_1 + X_2 \)
   
   (c) \( aX_1 + bX_2 \)

2. What is the correlation (coefficient) between \( X_1 \) and \( X_2 \)?

**Exercise 1.4.10.** Let \( X \) have a multivariate normal distribution with the mean (vector) \( \mu \in \mathbb{R}^n \) and the positive-definite symmetric variance-covariance matrix \( \Sigma \in \mathbb{R}^{n \times n} \). Let \( a \) be an arbitrary vector in \( \mathbb{R}^n \). Prove that the random variable

\[
Y = \frac{(X - \mu) a^T}{\sqrt{a^T \Sigma a}}
\]

has the standard normal distribution \( N(0, 1) \).

**Exercise 1.4.11.** Let \( X = (X_1, X_2, X_3) \) be a multivariate-normal vector such that \( \mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = 0 \) and \( \text{Var}[X_1] = 1 \), \( \text{Var}[X_2] = 2 \) and \( \text{Var}[X_3] = 3 \). Suppose further that \( X_1 \) is independent of \( X_2 - X_1 \), \( X_1 \) is independent of \( X_3 - X_2 \), and \( X_2 \) is independent of \( X_3 - X_2 \).

1. Find the variance-covariance matrix of \( X \).

2. Find \( P[X_1 + X_2 + X_3 \leq 1] \) in terms of the distribution function \( \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi \) of the standard unit normal.

**Exercise 1.4.12.** The prices \((S_1, S_2)\) of two stocks at a certain date are modelled in the following way: let \( X = (X_1, X_2) \) be a normal random vector

with mean \( \mu = (\mu_1, \mu_2) \) and the variance-covariance matrix \( \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \).

The stock-prices are then given by \( S_1 = \exp(X_1) \) and \( S_2 = \exp(X_2) \).

(a) Find \( \text{Cov}(S_1, S_2) \) in terms of \( \mu \) and \( \Sigma \).

(b) Find the probability that at the price of at least one of the stocks is in the interval \([s_1, s_2]\), for some constants \(0 < s_1 < s_2\). You can leave your answer in terms of the following:

i) the cdf of the unit normal

\[
\Phi(x) = \int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi, \quad \text{and}
\]

ii) the joint cdf of a bivariate normal with zero mean, unit marginal variances, and correlation \( \rho \):

\[
\Psi(x_1, x_2, \rho) = \int_{-\infty}^{\xi_1} \int_{-\infty}^{\xi_2} \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2(1 - \rho^2)} (\xi_1^2 + \xi_2^2 - 2\rho \xi_1 \xi_2) \right) d\xi_1 d\xi_2
\]
Solutions to Exercises in Section 1.4

Solution to Exercise 1.4.7:

1. First we prove that the random vector $X'$ has a bivariate normal distribution. It will be enough (by Definition 1.4.1) to prove that $X'a^T$ is a normal random variable for any vector $a \in \mathbb{R}^m$.

$$X'a^T = (XB)a^T = Xc^T, \text{ where } c = Ba^T.$$ Using the multivariate normality of $X$, the random variable $Xc^T$ is normal for any $c \in \mathbb{R}^n$, and therefore so is $X'a^T$.

It remains to identify the mean and the variance of $X'$:

$$\mu' = \mathbb{E}[X'] = \mathbb{E}[XB] = (\text{by componentwise linearity of expectation}) = \mathbb{E}[X]B = \mu B.$$ As for the variance-covariance matrix, we use the linearity of expectation again and representation (1.3.2) to obtain

$$\Sigma' = \mathbb{E}[(X'-\mu')^T(X'-\mu')] = \mathbb{E}[B^T(X-\mu)(X-\mu)B] = B^T\mathbb{E}[(X-\mu)(X-\mu)]B = B^T\Sigma B.$$ We conclude that $X' \sim N(\mu B, B^T\Sigma B)$.

2. Note that here we have a special case of part 1., so we know that we are dealing with a normal random vector. The mean and the variance can be obtained using the results just obtained:

$$\Sigma' = B^T\Sigma B = \begin{pmatrix} 7 & 3 & 3 \\ 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} 5 & 4 & -9 \\ 6 & 7 & 22 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 4 & 3 \\ 2 & 7 \end{pmatrix} = \begin{pmatrix} 73 & 100 \\ 100 & 577 \end{pmatrix},$$ and

$$\mu' = \mu B = \begin{pmatrix} 1, & 2, & 3 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 4 & 3 \\ 2 & 7 \end{pmatrix} = \begin{pmatrix} 15, & 30 \end{pmatrix}.$$ Therefore the distribution of $X'$ is multivariate normal:

$$N\left(\begin{pmatrix} 15, & 30 \end{pmatrix}, \begin{pmatrix} 73 & 100 \\ 100 & 577 \end{pmatrix}\right)$$

Here are the Maple commands I used to get the new mean and the new variance-covariance matrix:
1.4. THE MULTIVARIATE NORMAL DISTRIBUTION

> with(linalg):
> B:=matrix(3,2,[1,3,4,3,2,7]);
> Sigma:=matrix(3,3,[5,4,-9,4,6,-7,-9,-7,22]);
> mu:=matrix(1,3,[1,2,3]);
> mu1:=evalm(mu &* B);
> Sigma1:=evalm(transpose(B) &* Sigma &* B);

...and some more that will produce the picture of the density of $X'$:

> Q:=evalm(transpose([x1,x2]-mu1) &* inverse(Sigma1) &* ([x1,x2]-mu1));
> f:=(x1,x2)->1/(2*Pi*sqrt(det(Sigma1)))*exp(-1/2*Q);
> plot3d(f(x1,x2),x1=-30..40,x2=-70..70);

Solution to Exercise 1.5.13: Let $X = (X_1, X_2)$ be the random vector, where $X_1$ is the price of MHC, and $X_2$ of BII:

$$\mu = (110, 87), \Sigma = \begin{pmatrix} 49 & 56 \\ 56 & 100 \end{pmatrix}.$$ 

We are interested in the probability $p = \mathbb{P}[X_1 \geq 101, X_2 \geq 87]$, and we can rewrite this as $p = \mathbb{P}[X \in A]$, where $A = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 101, x_2 \geq 87\}$, and compute

$$p = \int_{A} \int f_X(x_1, x_2) \, dx_1 \, dx_2 = \int_{110}^{\infty} \int_{87}^{\infty} f_X(x_1, x_2) \, dx_1 \, dx_2,$$

where $f_X$ is the density of the multivariate normal random vector $X$:

$$f_X(x_1, x_2) = \frac{\exp\left(-\frac{1}{2(1-0.8^2)} \left(\frac{(x_1-110)^2}{7^2} + \frac{(x_2-87)^2}{10^2} - 2 \cdot 0.8 \frac{(x_1-110)(x_2-87)}{7 \cdot 10}\right)\right)}{2 \pi 7 \cdot 10 \sqrt{1-0.8^2}}.$$ 

And while this integral can be evaluated explicitly, it is quite cumbersome and long.

A different approach would be to try and modify (transform) the vector $X = (X_1, X_2)$ in order to simplify the computations. First of all, to get rid of the mean, we note that $p = \mathbb{P}[X_1 \geq 110, X_2 \geq 87] = \mathbb{P}[X_1 - 110 \geq 0, X_2 - 87 \geq 0] = \mathbb{P}[X_1 - \mu_1 \geq 0, X_2 - \mu_2 \geq 0],$

and realize that $X' \triangleq (X_1 - \mu_1, X_2 - \mu_2)$ is still a normal random vector with the same variance-covariance matrix and mean $\mu' = (0, 0)$, so that

$$p = \mathbb{P}[X'_1 \geq 0, X'_2 \geq 0].$$

Last Updated: Spring 2005
CHAPTER 1. BROWNIAN MOTION

THE MULTIVARIATE NORMAL DISTRIBUTION

If \(X'_1\) and \(X'_2\) were independent, the probability \(p\) would be very easy to compute: \(\mathbb{P}[X'_1 \geq 0] \mathbb{P}[X'_2 \geq 0] = 1/4\) by symmetry. However, \(\rho \neq 0\), so we cannot do quite that. We can, however, try to form linear combinations \(Y_1\) and \(Y_2\) (or ‘mutual funds’ in the language of finance) so as to make \(Y_1\) and \(Y_2\) into independent standard normal variables \((N(0, 1))\), and see if the computations simplify. So, let us put

\[
Y = (Y_1, Y_2) = XA = (X_1, X_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},
\]

and try to determine the coefficients \(a_{ij}\) so that the vector \(Y\) has the identity matrix as its variance-covariance matrix. From previous exercise we know how to compute the variance-covariance matrix of \(Y\), so that we get the following matrix equation; we are looking for the matrix \(A\) so that

\[
A^T \begin{pmatrix} 49 & 56 \\ 56 & 100 \end{pmatrix} A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.4.2}
\]

However, we do not need to rush and solve the equation, before we see what information about \(A\) we actually need. Let us suppose that we have the sought-for matrix \(A\), and that it is invertible (it will be, do not worry). Note that \(p = \mathbb{P}[X \in (0, \infty) \times (0, \infty)] = \mathbb{P}[AX \in A\left([0, \infty) \times [0, \infty)\right)] = \mathbb{P}[Y \in A\left([0, \infty) \times [0, \infty)\right)],\)

so all we need from \(A\) is the shape of the region \(D = A\left([0, \infty) \times [0, \infty)\right)\). Since \(A\) is a linear transformation, the region \(D\) will be the infinite wedge between \((a_{11}, a_{12}) = (1, 0)A\) and \((a_{21}, a_{22}) = (0, 1)A\). Actually, we need even less information form \(A\): we only need the angle \(\angle D\) of the wedge \(D\). Why? Because the probability we are looking for is obtained by integrating the density of \(Y\) over \(D\), and the density of \(Y\) is a function of \(x_1^2 + x_2^2\) (why?), so everything is rotationally symmetric around \((0, 0)\). It follows that \(p = \angle D/2\pi\).

To calculate \(\angle D\), we remember our analytic geometry:

\[
\cos(\angle D) = \cos\left(\angle((1, 0)A, (0, 1)A)\right) = \frac{a_{11}a_{21} + a_{12}a_{22}}{\sqrt{a_{11}^2 + a_{12}^2} \sqrt{a_{21}^2 + a_{22}^2}},
\]

and we also compute

\[
B \triangleq AA^T = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{11}a_{21} + a_{12}a_{22} & a_{21}^2 + a_{22}^2 \end{pmatrix},
\]

and realize that \(\cos(\angle D) = \frac{B_{12}}{\sqrt{B_{11}B_{22}}}\), so that the only thing we need to know about \(A\) is the product \(B = A^T A\). To compute \(B\), we go back to equation (1.4.2) and multiply it by \(A\) on the left and \(A^T\) on the right to obtain

\[
(AA^T) \Sigma (AA^T) = (AA^T),
\]
and since $A$ and $AA^T$ will be invertible, we can multiply both sides by $(AA^T)^{-1}\Sigma^{-1}$ from the right to obtain

$$B = AA^T = \Sigma^{-1},$$

and so $\cos(\angle D) = -\rho$.

by simply plugging the elements of $\Sigma^{-1}$ into the expression for $\cos(\angle D)$.

We can now calculate $p$:

$$p = \frac{\arccos(-\rho)}{2\pi} \approx 0.398.$$

Solution to Exercise 1.4.9:

1. All three random variables are linear combinations of elements of a multivariate normal and therefore have normal distributions themselves. Thus, it is enough to identify their means and variances:

   (a) $E[X_1] = \mu_1 = 2$, $\text{Var}[X_1] = \Sigma_{11} = 1$, and so $X_1 \sim N(2, 1)$.

   (b) $E[X_1 + X_2] = \mu_1 + \mu_2 = 5$, $\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}[X_1, X_2] = 1$, and so $X_1 + X_2 \sim N(5, 1)$.

   (c) $E[aX_1 + bX_2] = 2a + 3b$, $\text{Var}[aX_1 + bX_2] = a^2 + 4b^2 - 4ab$, and so $aX_1 + bX_2 \sim N(2a + 3b, a^2 + 4b^2 - 4ab)$.

2. The formula says

$$\rho(X_1, X_2) = \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1]\text{Var}[X_2]}} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}} = -1.$$

Solution to Exercise 1.4.10: Realize first that $Y$ can be written as a linear combination of the components of $X$ and is therefore normally distributed. It is thus enough to prove that $E[Y] = 0$ and $\text{Var}[Y] = 1$:

$$E[Y] = \frac{1}{\sqrt{a\Sigma a^T}}E[(X - \mu)a^T] = \frac{1}{\sqrt{a\Sigma a^T}} \sum_{k=1}^{n} a_k E[X_k - \mu_k] = 0.$$

$$\text{Var}[Y] = \frac{1}{a\Sigma a^T} \text{Var}[(X - \mu)a^T] = \frac{1}{a\Sigma a^T} E[(X - \mu)^2] = \frac{1}{a\Sigma a^T} E[(X - \mu)(X - \mu)^T] = \frac{aE[(X - \mu)^T(X - \mu)]a^T}{a\Sigma a^T} = 1,$$

since $\Sigma = E[(X - \mu)^T(X - \mu)]$ by definition.

Solution to Exercise 1.4.11:

Solution to Exercise 1.4.12:
CHAPTER 1. BROWNIAN MOTION

THE MULTIVARIATE NORMAL DISTRIBUTION

(a) First of all, let us derive an expression for \( \mathbb{E}[e^Z] \) for a normal random variable \( Z \) with mean \( \mu \) and variance \( \sigma^2 \). To do that, we write \( Z = \mu + \sigma Z_0 \), where \( Z_0 \) is a standard unit normal. Then

\[
\mathbb{E}[e^Z] = e^{\mu} \mathbb{E}[e^{\sigma Z_0}] = e^{\mu} \int_{-\infty}^{\infty} e^{\sigma x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\mu + \frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} dx = e^{\mu + \frac{\sigma^2}{2}}.
\]

We know that \( X_1 \sim N(\mu_1, \Sigma_{11}) \) and \( X_2 \sim N(\mu_2, \Sigma_{22}) \), and so

\[
\mathbb{E}[e^{X_1}] = e^{\mu_1 + \frac{\sigma_{11}^2}{2}}, \text{ and } \mathbb{E}[e^{X_2}] = e^{\mu_2 + \frac{\sigma_{22}^2}{2}}.
\]

Further, \( X_1 + X_2 \sim N(\mu_1 + \mu_2, \Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}) \) because \( \text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}[X_1, X_2] \), so that

\[
\mathbb{E}[e^{X_1}e^{X_2}] = \mathbb{E}[e^{X_1 + X_2}] = e^{\mu_1 + \mu_2 + \frac{\sigma_{11}^2 + \sigma_{22}^2 + 2\sigma_{12}}{2}}.
\]

Therefore

\[
\text{Cov}(X_1, X_2) = \mathbb{E}[e^{X_1}e^{X_2}] - \mathbb{E}[e^{X_1}]\mathbb{E}[e^{X_2}] = e^{\mu_1 + \mu_2 + \frac{\sigma_{11}^2 + \sigma_{22}^2 + 2\sigma_{12}}{2}} - e^{\mu_1 + \mu_2 + \frac{\sigma_{11}^2}{2}} e^{\mu_2 + \frac{\sigma_{22}^2}{2}} = e^{\mu_1 + \mu_2 + \frac{\sigma_{11}^2 + \sigma_{22}^2}{2}} \left( \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} \right).
\]

(b) At least one of the prices \( S_1 = e^{X_1}, S_2 = e^{X_2} \) will take value in \( [s_1, s_2] \) if and only if at least one of the components \( X_1, X_2 \) of the bivariate normal vector \( \mathbf{X} \) takes values in the interval \( [\log(s_1), \log(s_2)] \). This probability is equal to the probability that the random vector \( \mathbf{X}' = (X_1', X_2') = (\frac{X_1 - \mu_1}{\sqrt{\Sigma_{11}}}, \frac{X_2 - \mu_2}{\sqrt{\Sigma_{22}}}) \) falls into the set

\[
A = \{(y_1, y_2) \in \mathbb{R} : y_1 \in [l_1, r_1] \text{ or } y_2 \in [l_2, r_2]\}
\]

where

\[
l_1 = \frac{\log(s_1) - \mu_1}{\sqrt{\Sigma_{11}}}, \quad r_1 = \frac{\log(s_2) - \mu_1}{\sqrt{\Sigma_{11}}}, \quad l_2 = \frac{\log(s_1) - \mu_2}{\sqrt{\Sigma_{22}}}, \quad r_2 = \frac{\log(s_2) - \mu_2}{\sqrt{\Sigma_{22}}}.
\]

We have transformed \( \mathbf{X} \) into \( \mathbf{X}' \) because \( \mathbf{X}' \) is now a bivariate normal vector with correlation coefficient \( \rho' = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}} \), whose marginals \( X_1 \) and \( X_2 \) are standard normal (mean 0, and variance 1). To calculate the probability \( \mathbb{P}[\mathbf{X}' \in A] \), we write

\[
\mathbb{P}[\mathbf{X}' \in A] = 1 - \mathbb{P}[\mathbf{X}' \in A^c]
\]

\[
= 1 - \mathbb{P}[X_1' < l_1, X_2' < l_2] - \mathbb{P}[X_1' < l_1, X_2' > r_2] - \mathbb{P}[X_1' > r_1, X_2' < l_2] - \mathbb{P}[X_1' > r_1, X_2' > r_2]
\]

\[
= 1 - \Phi(l_1, l_2, \rho') - \left( \Phi(l_1) - \Psi(l_1, r_2, \rho') \right) - \left( \Phi(l_2) - \Psi(l_2, r_1, \rho') \right) - \Psi(-r_1, -r_2, \rho').
\]

We have used the fact that the distribution of \( \mathbf{X}' \) is symmetric (equal to that of \( -\mathbf{X}' \)).
1.5 Brownian Motion Defined

1.5.1 Stochastic Processes

After random variables and random vectors come stochastic processes. You can think of them as of very large random vectors (where \( n = \infty \)), or as random elements having functions (instead of numbers or vectors) as values. Formally, we have the following definition

Definition 1.5.1. A stochastic process is a collection

\[ \{X_t : t \in T\}, \]

of random variables \( X_t \), defined on the same probability space. \( T \) denotes the index set of the stochastic process and is usually taken to be

- \( T = \mathbb{N} \) (discrete-time processes),
- \( T = [0, T] \), for some \( T \in \mathbb{R} \) (finite-horizon continuous-time processes), or
- \( T = [0, \infty) \) (infinite-horizon continuous-time processes).

The notion of a stochastic processes is one of the most important in the modern probability theory and mathematical finance. It used to model a myriad of various phenomena where a quantity of interest varies continuously through time in a non-predictable fashion. In this course we will mainly be interested in continuous-time stochastic processes on finite and infinite time-horizons.

Every stochastic process can be viewed as a function of two variables - \( t \) and \( \omega \). For each fixed \( t, \omega \mapsto X_t(\omega) \) is a random variable, as postulated in the definition. However, if we change our point of view and keep \( \omega \) fixed, we see that the stochastic process is a function mapping \( \omega \) to the real-valued function \( t \mapsto X_t(\omega) \). These functions are called the trajectories of the stochastic process \( X \). Figure on the left shows the graph of the evolution of the Dow Jones Chemical index from January 2002 to December 2002. It is not hard to imagine the stock index pictured here as a realization (a trajectory) of a stochastic process.
CHAPTER 1. BROWNIAN MOTION

1.5. BROWNIAN MOTION DEFINED

1.5.2 The Distribution of a Stochastic Process

In contrast to the case of random vectors or random variables, it will not be very easy to define a notion of a distribution for a stochastic process. Without going into details why exactly this is a problem, let me just mention that the main culprit is infinity again. There is a way out, however, and it is provided by the notion of finite-dimensional distributions:

Definition 1.5.2. The finite-dimensional distributions of a random process \((X_t)_{t \in T}\) are all distribution functions of the form

\[
F(x_1, x_2, \ldots, x_n) \triangleq \mathbb{P}[X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n],
\]

for all \(n \in \mathbb{N}\) and all \(n\)-tuples \((t_1, t_2, \ldots, t_n)\) of indices in \(T\).

For a huge majority of stochastic processes encountered in practice, the finite-dimensional distributions (together with the requirement of regularity of its paths) will be sufficient to describe their full probabilistic structure.

1.5.3 Brownian Motion

The central object of study of this course is Brownian motion. It is one of the fundamental objects of applied mathematics and one of the most symmetric and beautiful things in whole wide world (to quote a mathematician who wanted to remain anonymous). To define the Brownian motion, it will be enough to specify its finite-dimensional distributions and ask for its trajectories to be continuous:

Definition 1.5.3. Brownian motion is a continuous-time, infinite-horizon stochastic process \((B_t)_{t \in [0, \infty)}\) such that

1. \(B_0 = 0\) (Brownian motion starts at 0),

2. for any \(t > s \geq 0\),

   (a) the increment \(B_t - B_s\) is normally distributed with mean \(\mu = 0\) and variance \(\sigma^2 = t - s\) (the increments of the Brownian motion are normally distributed)

   (b) the random variables \(B_{s_m} - B_{s_{m-1}}, B_{s_{m-1}} - B_{s_{m-2}}, \ldots, B_{s_1} - B_0\), are independent for any \(m \in \mathbb{N}\) and any \(s_1 < s_2 < \cdots < s_m\) (the increments of the Brownian motion are independent)

3. the trajectories of a Brownian motion are continuous function (Brownian paths are continuous).

Did I really specify the finite-dimensional distributions of the Brownian motion in the preceding definition? The answer is yes:
Proposition 1.5.4. Let \((B_t)_{t \in [0, \infty)}\) be a Brownian motion. For any \(n\)-tuple of indices \(0 \leq t_1 < t_2 < \cdots < t_n\), the random vector \((B_{t_1}, B_{t_2}, \ldots, B_{t_n})\) has the multivariate normal distribution with mean \(\mu = (0, 0, \ldots, 0)\) and the variance-covariance matrix

\[
\Sigma = \begin{pmatrix}
  t_1 & t_1 & \ldots & t_1 \\
  t_1 & t_2 & \ldots & t_2 \\
  \vdots & \vdots & \ddots & \vdots \\
  t_1 & t_2 & \ldots & t_n
\end{pmatrix}, \quad \text{i.e.} \quad \Sigma_{ij} = \min(t_i, t_j) = t_{\min(i,j)}.
\]

Proof. First of all, let us prove that the random vectors \((B_{t_1}, B_{t_2}, \ldots, B_{t_n})\) have the multivariate normal distribution. By definition (see Definition 1.4.1) it will be enough to prove that \(X \triangleq a_1 B_{t_1} + a_2 B_{t_2} + \cdots + a_n B_{t_n}\) is a normally distributed random variable for each vector \(a = (a_1, a_2, \ldots, a_n)\). If we rewrite \(X\) as

\[
X = a_n(B_{t_n} - B_{t_{n-1}}) + (a_n + a_{n-1})(B_{t_{n-1}} - B_{t_{n-2}}) + (a_n + a_{n-1} + a_{n-2})(B_{t_{n-2}} - B_{t_{n-3}}) + \ldots
\]

we see that \(X\) is a linear combination of the increments \(X_k \triangleq B_{t_k} - B_{t_{k-1}}\). I claim that these increments are independent. To see why, note that the last increment \(X_n\) is independent of \(B_{t_1}, B_{t_2}, \ldots, B_{t_{n-1}}\) by definition, and so it is also independent of \(B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots\). Similarly the increment \(B_{t_{n-1}} - B_{t_{n-2}}\) is independent of ‘everybody’ before it by the same argument. We have therefore proven that \(X\) is a sum of independent normally distributed random variables, and therefore normally distributed itself (see Example 1.4.6). This is true for each vector \(a = (a_1, a_2, \ldots, a_n)\), and therefore the random vector \((B_{t_1}, B_{t_2}, \ldots, B_{t_n})\) is multivariate normal. Let us find the mean vector and variance-covariance matrix of \((B_{t_1}, B_{t_2}, \ldots, B_{t_n})\).

First, for \(m = 1, \ldots, n\), we write the telescoping sum

\[
\mathbb{E}[B_{t_m}] = \mathbb{E}[X_m] + \mathbb{E}[X_{m-1}] + \ldots + \mathbb{E}[X_1], \quad \text{where} \quad X_k \triangleq B_{t_k} - B_{t_{k-1}}.
\]

By Definition 1.5.3-2.(a), the random variables \(X_k, k = 1, \ldots, m\) have mean 0, so \(\mathbb{E}[B_{t_m}] = 0\) and, thus, \(\mu = (0,0,\ldots,0)\). To find the variance-covariance matrix we suppose that \(i < j\) and write

\[
\Sigma_{ij} = \mathbb{E}[B_{t_i} B_{t_j}] = \mathbb{E}[B_{t_i}^2] - \mathbb{E}[B_{t_i} (B_{t_j} - B_{t_i})] = \mathbb{E}[B_{t_i}^2],
\]

since \(B_{t_i}\) and \(B_{t_j} - B_{t_i}\) are independent by Definition 1.5.3-2.(b). Finally, the random variable \(B_{t_i}\) is normally distributed with mean 0, and variance \(t_i\) (just take \(t = t_i\) and \(s = 0\) in Definition 1.5.3, part 2.(a)), and hence \(\mathbb{E}[B_{t_i}^2] = \text{Var}[B_{t_i}] = t_i\). The cases \(i = j\) and \(i \geq j\) can be dealt with analogously. 

\[\square\]
1.5. BROWNIAN MOTION DEFINED

The graph on the right shows a typical trajectory (realization) of a Brownian motion. It is important to note that for a random variable a realization is just one point on the real line, and a $n$-tuple of points for a random vector. For Brownian motion it is a whole function. Note, also, the jaggedness of the depicted trajectory. Later on we will see that no trajectory of the Brownian motion is differentiable at any of its points. . .

Figure 1.13: A Trajectory of Brownian motion

... nevertheless, by the definition, every trajectory of Brownian motion is continuous. That is not the case with the (approximation to) a trajectory of an “independent process” shown on the right. An “independent process” is the process $(X_t)_{t \in [0,1]}$ for which $X_t$ and $X_s$ are independent for $t \neq s$, and normally distributed. This is what you might call - “a completely random function $[0,1] \rightarrow \mathbb{R}$.”

Figure 1.14: A Trajectory of an “Independent” Process
1.5.4 A Little History

Brownian Motion, the physical phenomenon, was named after the English naturalist and botanist Robert Brown (upper-right picture) who discovered it in 1827, is the zig-zagging motion exhibited by a small particle, such as a grain of pollen, immersed in a liquid or a gas. The first explanation of this phenomenon was given by Albert Einstein (lower-left caricature) in 1905. He showed that Brownian motion could be explained by assuming the immersed particle was constantly bombarded by the molecules of the surrounding medium. Since then, the abstracted process has been used beneficially in such areas as analyzing price levels in the stock market and in quantum mechanics. The mathematical definition and abstraction of the physical process as a stochastic process, given by the American mathematician Norbert Wiener (lower-right photo) in a series of papers starting in 1918. Generally, the terms Brownian motion and Wiener process are the same thing, although Brownian motion emphasizes the physical aspects, and Wiener process emphasizes the mathematical aspects. Bachelier process is an uncommonly applied term meaning the same thing as Brownian motion and Wiener process. In 1900, Louis Bachelier (upper-left photo) introduced the limit of random walk as a model for the prices on the Paris stock exchange, and so is the originator of the idea of what is now called Brownian motion. This term is occasionally found in financial literature and European usage.

1.5.5 Gaussian Processes

Back to mathematics. We have seen that both the Brownian motion and the Independent process share the characteristic that their finite-dimensional distributions are multivariate normal. In case of the Independent process, it is because the distributions of random variables $X_t$ are normal and independent of each other. In case of the Brownian motion, it is the content of Proposition 1.5.4. The class of processes which share this property is important enough to merit a name:

**Definition 1.5.5.** A continuous-time stochastic process $(X_t)_{t \in [0, \infty)}$ (or $(X_t)_{t \in [0,T]}$ for some
A Gaussian process \( T \in \mathbb{R} \) is called a \textbf{Gaussian processes} if its finite-dimensional distributions are multivariate normal, i.e. for each \( t_1 < t_2 < \cdots < t_n \) the random vector \((X_{t_1}, X_{t_2}, \ldots, X_{t_n})\) is multivariate normal.

The name - Gaussian process - derives from the fact that the normal distribution is sometimes also called the Gaussian distribution, after Carl Friedrich Gauss, who discovered many of its properties. Gauss, commonly viewed as one of the greatest mathematicians of all time (if not the greatest), is (was) properly honored by Germany on their 10 Deutschmark bill shown in the figure left. With each Gaussian process \( X \) we associate two functions: \( \mu_X : [0, \infty) \to \mathbb{R} \) and \( c_X : [0, \infty) \times [0, \infty) \to \mathbb{R} \). The function \( \mu_X \), defined by, \( \mu_X(t) \triangleq \mathbb{E}[X_t] \) is called the \textbf{expectation function} of \( X \) and can be thought of (loosely) as the trend around which the process \( X \) fluctuates. The function \( c_X \), defined by \( c_X(t,s) = \text{Cov}(X_t, X_s) \), is called the \textbf{covariance function} of the process \( X \) and we can interpret it as the description of dependence structure of the process \( X \). It is worth noting that the function \( c \) is \textbf{positive semi-definite} (see Exercise 1.4.9 for the precise definition and proof). Conversely, it can be proved that given (almost any) function \( \mu \) and any positive-semidefinite function \( c \), we can construct a stochastic process which has exactly \( \mu \) as its expectation function, and \( c \) as its covariance function.

\subsection*{1.5.6 Examples of Gaussian Processes}

Finally, here are several examples of Gaussian processes used extensively in finance and elsewhere.

\textbf{Example 1.5.6. (Independent Process)} The simplest example of a Gaussian process is the Independent process - the process \( (X)_{t \in [0, \infty)} \) such that each \( X_t \) is normally distributed with mean \( \mu \) and variance \( \sigma^2 \), and such that the collection random variables \((X)_{t \in [0, \infty)}\) are independent. The expectation and covariance functions are given by

\[
\mu_X(t) = \mu, \quad c_X(t, s) = \begin{cases} 
\sigma^2, & t = s, \\
0, & t \neq s.
\end{cases}
\]
Example 1.5.7. (Brownian Motion) The Brownian motion is the most important example of a Gaussian process. The expectation and covariance functions are

\[ \mu_X(t) = 0, \quad c_X(t,s) = \min(t,s). \]

Example 1.5.8. (Brownian Motion with Drift) Let \( (B_t)_{t \in [0,\infty)} \) be a Brownian motion and let \( b \) be a constant in \( \mathbb{R} \). We define the process \( (X_t)_{t \in [0,\infty)} \) by

\[ X_t = B_t + bt. \]

The process \( (X_t)_{t \in [0,\infty)} \) is still a Gaussian process with the expectation and covariance functions

\[ \mu_X(t) = bt, \quad c_X(t,s) = \min(t,s). \]

Example 1.5.9. (Brownian Bridge) The Brownian Bridge (or Tied-Down Brownian Motion) is what you get from Brownian Motion \( (B_t)_{t \in [0,1]} \) on the finite interval \([0,1]\), when you require that \( B_1 = 0 \). Formally, \( X_t = B_t - tB_1 \), where \( B \) is some Brownian motion. In the exercises I am asking you to prove that \( X \) is a Gaussian process and to compute its expectation and covariance function.

1.5.7 Brownian Motion as a Limit of Random Walks

Brownian motion describes an idealization of a motion of a particle subjected to independent ‘shoves’. It is therefore not unreasonable to imagine our particle as a closely approximated by a random walker taking each step, independently of its position or previous steps, either to the left or to the right with equal probabilities. Mathematically, we have the following definition:

Definition 1.5.10. Let \( X_1, X_2, \ldots \) be a sequence of independent Bernoulli random variables, i.e.,

\[ X_i \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix} \]
The stochastic process \((S_n)_{n \in \mathbb{N}_0}\) defined by
\[
S_0 = 0, \quad S_n = X_1 + X_2 + \cdots + X_n, \quad n \geq 1,
\]
is called the random walk.

Suppose now that we rescale the random walk in the following way: we decrease the time interval between steps from 1 to some small number \(\Delta t\), and we accordingly decrease the step size to another small number \(\Delta x\). We would like to choose \(\Delta t\) and \(\Delta x\) so that after 1 unit of time, i.e. \(n = 1/\Delta t\) steps, the standard deviation of the resulting random variable \(S_n\) is normalized to 1. By the independence of steps of the random walk - and the fact that for the random variable which takes values \(\Delta x\) and \(-\Delta x\) with equal probabilities, the variance is \((\Delta x)^2\) - we have
\[
\text{Var}[S_n] = \text{Var}[X_1] + \text{Var}[X_1] + \cdots + \text{Var}[X_n] = n(\Delta x)^2 = \frac{(\Delta x)^2}{\Delta t},
\]
so we should choose \(\Delta x = \sqrt{\Delta t}\). Formally, for each \(\Delta t\), we can define an infinite-horizon continuous-time process \((B_{\Delta t})_{t \in [0, \infty)}\) by the following procedure:

- take a random walk \((S_n)_{n \in \mathbb{N}_0}\) and multiply all its increments by \(\Delta x = \sqrt{\Delta t}\) to obtain the new process \((S_{\Delta x}^n)_{n \in \mathbb{N}_0}\)
  \[
  S_{\Delta x}^0 = 0, \quad S_{\Delta x}^n = \sqrt{\Delta t}X_1 + \sqrt{\Delta t}X_2 + \cdots + \sqrt{\Delta t}X_n = \sqrt{\Delta t}S_n, \quad n \geq 1.
  \]
- define the process \((B_{\Delta t})_{t \in [0, \infty)}\) by
  \[
  B_{\Delta t}^t = \begin{cases} 
  S_{\Delta x}^n, & \text{if } t \text{ is of the form } t = n\Delta t, \\
  \text{interpolate linearly}, & \text{otherwise}.
  \end{cases}
  \]

It can be shown that the processes \((B_{\Delta t})_{t \in [0, \infty)}\) converge to a Brownian motion in a mathematically precise sense (the sense of weak convergence), but we will neither elaborate on the exact definition of this form of convergence, nor prove any rigorous statements. Heuristically, weak convergence will allow us to use the numerical characteristics of the paths of the process \(B_{\Delta t}\), which we can simulate on a computer, as approximations of the analogous characteristics of the Brownian Motion. We will cover the details later in the section on simulation.
Exercises

Exercise 1.5.11. Let \((B_t)_{t \in [0, T]}\) be a Brownian motion. Which of the following processes are Brownian motions? Which are Gaussian processes?

- \(X_t \triangleq -B_t, \ t \in [0, \infty)\),
- \(X_t = \sqrt{t}B_1, \ t \in [0, \infty)\),
- \(X_t = |B_t|, \ t \in [0, \infty)\),
- \(X_t = B_{T+t} - B_T, \ t \in [0, \infty), \) where \(T\) is a fixed constant - \(T \in [0, \infty)\)
- \(X_t = B_{et}, \ t \in [0, \infty)\)
- \(X_t = \frac{1}{\sqrt{u}}B_{ut}, \ t \in [0, \infty), \) where \(u\) is a fixed positive constant \(u \in (0, \infty)\)

Exercise 1.5.12. Prove that the Brownian motion with drift is a Gaussian process with the expectation and covariance functions given in (1.5.8).

Exercise 1.5.13. Prove that the Brownian Bridge is a Gaussian process and compute its expectation and covariance functions. Does it have independent increments?

Exercise 1.5.14. A function \(\gamma : [0, \infty) \times [0, \infty) \to \mathbb{R}\) is called **positive semi-definite** if for any \(n \in \mathbb{N}\) and any \(t_1 < t_2 < \cdots < t_n\) the matrix \(\Gamma\) defined by

\[
\Gamma \triangleq \begin{pmatrix}
\gamma(t_1, t_1) & \gamma(t_1, t_2) & \cdots & \gamma(t_1, t_n) \\
\gamma(t_2, t_1) & \gamma(t_2, t_2) & \cdots & \gamma(t_2, t_n) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(t_n, t_1) & \gamma(t_n, t_2) & \cdots & \gamma(t_n, t_n)
\end{pmatrix},
\]

is symmetric positive semi-definite. Let \((X_t)_{t \in [0, \infty)}\) be a Gaussian process. Show that its covariance function \(c_X\) is positive semi-definite.

Exercise 1.5.15. Let \((B_t)_{t \in [0, T]}\) be a Brownian motion, and let \(X = \int_0^1 B_s \, ds\) be the area under the trajectory of \(B\). Compute the mean and variance of \(X\) by following the following procedure

1. Compute the mean and variance of the random variable \(X^\Delta t = \int_0^1 B_s^\Delta t \, ds\).
2. Let \(\Delta t \to 0\) to obtain the mean and variance of \(X\).

Exercise 1.5.16. Let \((B_t)_{t \in [0, 1]}\) be a Brownian motion on \([0, 1]\), and let \((X_t)_{t \in [0, 1]}, X_t = B_t - tB_1\) be the corresponding Brownian Bridge. Define the process \((Y_t)_{t \in [0, \infty)}\) by \(Y_t = (1+t)X_{\frac{t}{1+t}}\). Show that \(Y\) is a Brownian motion, by showing that
CHAPTER 1. BROWNIAN MOTION 1.5. BROWNIAN MOTION DEFINED

- \((Y_t)_{t \in [0, \infty)}\) has continuous paths,
- \((Y_t)_{t \in [0, \infty)}\) is a Gaussian process, and
- the mean and the covariance functions of \((Y_t)_{t \in [0, \infty)}\) coincide with the mean and the covariance functions of the Brownian motion.

Use the result above to show that for any Brownian motion \((W_t)_{t \in [0, \infty)}\) we have

\[
\lim_{t \to \infty} \frac{W_t}{t} = 0.
\]

Exercise 1.5.17. (Brownian Motion in the Plane)

Let \((B^1_t)_{t \in [0, T]}\) and \((B^2_t)_{t \in [0, T]}\) be two independent Brownian motions (independent means that for any \(t\), \(B^1_t\) is independent of the whole process \((B^2_t)_{t \in [0, T]}\), and vice versa). You can think of the trajectory \(t \mapsto (B^1_t(\omega), B^2_t(\omega))\) as a motion of a particle in the plane (see picture).

Let the random variable \(R_t\) denote the distance from the Brownian particle \((B^1_t, B^2_t)\) to the origin, i.e.

\[
R_t = \sqrt{B^1_t^2 + B^2_t^2}.
\]

Find the density \(f_{R_t}\) of \(R_t\), for fixed \(t > 0\). Calculate the mean \(\mu(t)\) and standard deviation \(\sqrt{\sigma^2(t)}\) of \(R_t\) as functions of \(t\), and sketch their graphs for \(t = [0, 5]\). What do they say about the expected displacement of the Brownian particle from the origin as a function of time \(t\)? (Hint: compute first the distribution function \(F_{R_t}\) by integrating a multivariate normal density over the appropriate region, and then get \(f_{R_t}\) from \(F_{R_t}\) by differentiation. To get \(\mu(t)\) and \(\sqrt{\sigma^2(t)}\) you might want to use Maple, or Mathematica, or some other software package capable of symbolic integration, because the integrals involved might get ugly. Once you get the results, you might as well use the same package to draw the graphs for you.)

Exercise 1.5.18. (Exponential Brownian Motion)

The process \((S_t)_{t \in [0, T]}\) defined by

\[
S_t = s_0 \exp(\alpha B_t + \beta t),
\]

for some constants \(s_0, \alpha, \beta \in \mathbb{R}\), with \(\alpha \neq 0\) and \(s_0 > 0\), is called exponential Brownian motion. The exponential Brownian motion is one of the most wide-spread stock-price models today, and the celebrated Black-Scholes formula rests on the assumption that stocks follow exponential Brownian motion. It is the purpose of this exercise to give meanings to parameters \(\alpha\) and \(\beta\).

1. Calculate the expectation \(E[S_t]\) as a function of \(t\). (Do not just copy it from the book. Do the integral yourself.) What must be the relationship between \(\alpha\) and \(\beta\), so that \(X_t\) models a stock with 0 rate of return, i.e. \(E[S_t] = s_0\) for all \(t \geq 0\)?
2. If you had to single out one parameter (can be a function of both \(\alpha\) and \(\beta\)) and call it the \textit{rate of return of the stock} \((S)_{t\in[0,T]}\), what would it be? Why? Why would some people call your conclusion paradoxical?

**Exercise 1.5.19.** (a) Let \(X_1\) and \(X_2\) be two independent normal random variables with mean 0 and variance 1. Find the density function of the random variable \(Y = \frac{X_1}{X_2}\). Do not worry about the fact that \(X_2\) might be equal to 0. It happens with probability 0. (Hint: You might want to use polar coordinates here.)

(b) Let \(B_t, t \in [0, \infty)\) be a Brownian motion. Define the process \(Y_t = \frac{B_{t} - B_{\frac{1}{4}t}}{B_{\frac{1}{4}t} - B_{\frac{1}{2}t}}, t \in (0, \infty)\), \(Y_0 = 0\). Is \(Y_t\) a Gaussian process? If it is, find its mean and covariance functions. If it is not, prove rigorously that it is not.
CHAPTER 1. BROWNIAN MOTION  1.5. BROWNIAN MOTION DEFINED

Solutions to Exercises in Section 1.5

Solution to Exercise 1.5.11 (Note: in what follows we always assume $t > s$).

1. $X_t = -B_t$ is a continuous process, $X_0 = -B_0 = 0$, and its increments $X_t - X_s$ are independent of the past because they are just the negatives of the increments of the Brownian motion. Finally, the distribution of $X_t - X_s = -(B_t - B_s)$ is normal with mean 0 and variance $t - s$, by symmetry of the normal distribution. Therefore, $X_t$ is a Brownian motion.

2. The process $X_t = \sqrt{t}B_1$ is not a Brownian motion because its increments $X_t - X_s = B_1(\sqrt{t} - \sqrt{s})$ are not independent of the past: $X_t - X_s = \frac{\sqrt{t} - \sqrt{s}}{\sqrt{t}}X_r$, for $0 < r < t - s$, so $X_t - X_s$ cannot be independent of $X_r$.

On the other hand, $X_t$ is a Gaussian process. To show that, we take indices $t_1, \ldots, t_n$, and constants $a_1, a_2, \ldots, a_n$ and form the linear combination $Y = a_1X_{t_1} + a_2X_{t_2} + \cdots + a_nX_{t_n}$. By the definition of $X_t$, we have $Y = \gamma B_1$, with $\gamma = (a_1\sqrt{t_1} + a_2\sqrt{t_2} + \cdots + a_n\sqrt{t_n})$. Therefore $Y$ is a normally distributed random variable, and we conclude that $X$ is a Gaussian process.

3. $X_t = |B_t|$ is not a Gaussian process because $X_1 = |B_1|$ is not a normally distributed random variable (it is nontrivial, and yet never takes negative values), as it should be in the case of a Gaussian process. Since Brownian motion is a Gaussian process, we conclude that $X_t$ is not a Brownian motion, either.

4. The process $X_t = B_{T+t} - B_T$ has continuous paths (since the Brownian motion $B_t$ does), and $X_0 = B_{T+0} - B_T = 0$. The increments $X_t - X_s$ have the normal $N(0, t - s)$-distribution since $X_t - X_s = B_{T+t} - B_{T+s}$, and the distribution of $B_{T+t} - B_{T+s}$ is $N(0, T + t - (T + s)) \sim N(0, t - s)$, by the definition of the Brownian motion. Finally, the increment $X_t - X_s = B_{T+t} - B_{T+s}$ is independent of the history up to the time $T + s$, i.e. $X_t - X_s$ is independent of the random vector $(B_{T+s_0}, B_{T+s_1}, \ldots, B_{T+s_n})$ for any collection $s_0, s_1, \ldots, s_n$ in $[0, s]$. We can always take $s_0 = 0$, and conclude that, in particular, $X_t - X_s$ is independent of $B_T$. Therefore, $X_t - X_s$ is independent of the random vector $(B_{T+s_1} - B_T, B_{T+s_2} - B_T, \ldots, B_{T+s_n} - B_T) = (X_{s_1}, X_{s_2}, \ldots, X_{s_n})$, and it follows that $X_t$ is a Brownian motion (and a Gaussian process).

5. Since $B_t$ is a Gaussian process, the random vector $(B_{s_1}, B_{s_2}, \ldots, B_{s_n})$ is multivariate normal for any choice of $s_1, s_2, \ldots, s_n$.

In particular, if we take $s_1 = e^{t_1}, \ldots, s_n = e^{t_n}$, we have that $(B_{e^{t_1}}, B_{e^{t_2}}, \ldots, B_{e^{t_n}})$ is a multivariate normal. Therefore, $X_t$ is a Gaussian process.

$X_t$ is not a Brownian motion since $\operatorname{Var}[X_1] = \operatorname{Var}[B_t] = e \neq 1$. 

Last Updated: Spring 2005

Continuous-Time Finance: Lecture Notes
6. For $u > 0$, the process $X_t = \frac{1}{\sqrt{u}} B_{ut}$ has continuous paths and $X_0 = 0$. The increment $X_t - X_s = \frac{1}{\sqrt{u}} (B_{ut} - B_{us})$ is independent of the history of $B$ before (and including) $u$, which is exactly the history of $X$ before (and including) $s$. Finally, the distribution of the increment $X_t - X_s$ is normal with mean $0$ and variance $\text{Var}[\frac{1}{\sqrt{u}} (B_{tu} - B_{su})] = \frac{1}{u} (tu - ts) = t - s$. Therefore, $X$ is a Gaussian motion (and a Gaussian process).

**Solution to Exercise 1.5.12** Let $(t_1, t_2, \ldots, t_n)$ be an arbitrary $n$-tuple of indices. We have to prove that the random vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is multivariate normal, where $X_t = B_t + bt$ is a Brownian motion with drift. To do that, we take constants $a_1, a_2, \ldots, a_n$ and compute

$$Y = a_1 X_{t_1} + a_2 X_{t_2} + \cdots + a_n X_{t_n} = Z + bt(a_1 + a_2 + \cdots + a_n),$$

where $Z = a_1 B_1 + a_2 B_2 + \cdots + a_n B_n$ is a normal random variable since the Brownian motion is a Gaussian process. The random variable $Y$ is also normal since it is constructed by adding a constant to a normal random variable. Therefore, $X$ is a Gaussian process. To get $\mu_X$ and $c_X$, we write (noting that adding a constant to a random variable does not affect its covariance with another random variable):

$$\mu_X(t) = \mathbb{E}[X_t] = \mathbb{E}[B_t + bt] = bt$$

$$c_X(t, s) = \text{Cov}[X_t, X_s] = \text{Cov}[X_t - bt, X_s - bs] = \text{Cov}[B_t, B_s] = \min(t, s)$$

**Solution to Exercise 1.5.13**

Just like in the previous solution, for an arbitrary $n$-tuple of indices $(t_1, t_2, \ldots, t_n)$ in $[0, 1]$, we are trying to prove that the random vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is multivariate normal. Here $X_t = B_t - B_1 t$ is a Brownian bridge. Again, we take constants $a_1, a_2, \ldots, a_n$ and compute

$$Y = a_1 X_{t_1} + a_2 X_{t_2} + \cdots + a_n X_{t_n} = -(a_1 + a_2 + \cdots + a_n) t B_1 + a_1 B_{t_1} + a_2 B_{t_2} + \cdots + a_n B_{t_n},$$

which is normal because $(B_1, B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ is a multivariate normal random vector. Therefore, $X$ is a Gaussian process. To get $\mu_X$ and $c_X$, we use the fact that $\mathbb{E}[B_t B_s] = st$ and $\mathbb{E}[B_t B_1] = t$, $\mathbb{E}[B_s B_1] = s$, for $t, s \leq 1$, and

$$\mu_X(t) = \mathbb{E}[X_t] = \mathbb{E}[B_t - B_1 t] = 0$$

$$c_X(t, s) = \text{Cov}[X_t, X_s] = \mathbb{E}[(B_t - t B_1)(B_s - s B_1)] = \mathbb{E}[B_t B_s] - s \mathbb{E}[B_t B_1] - t \mathbb{E}[B_s B_1] + ts \mathbb{E}[B_1 B_1]$$

$$= \min(t, s) - ts (= s(1 - t), \text{ when } s < t).$$

**Solution to Exercise 1.5.14** Let $X_t$ be a Gaussian process, and let $c_X$ be its covariance process. We will assume that $\mathbb{E}[X_t] = 0$ for each $t$, since adding a constant to each $X_t$ will not affect the covariance function.
CHAPTER 1. BROWNIAN MOTION 1.5. BROWNIAN MOTION DEFINED

Let \( \mathbf{\Gamma} \) be defined as in the exercise with \( \gamma = c_X \), and some \( t_1, t_2, \ldots, t_n \). To prove that \( \mathbf{\Gamma} \) is non-negative definite, we have to show that \( \mathbf{\Gamma} \) is symmetric, and that for any (row) vector \( \mathbf{a} \) we have \( \mathbf{a} \mathbf{\Gamma} \mathbf{a}^T \geq 0 \).

The first statement is easy since \( \Gamma_{ij} = \text{Cov}[X_{t_i}, X_{t_j}] = \text{Cov}[X_{t_j}, X_{t_i}] = \Gamma_{ji} \). To show the non-negative definiteness, we take the vector \( \mathbf{a} = (a_1, a_2, \ldots, a_n) \) and construct the random variable \( Y = a_1 X_1 + a_2 X_2 + \ldots a_n X_n \). Since variances are always non-negative, we have

\[
0 \leq \text{Var}[Y] = \text{Var}[\mathbf{X} \mathbf{a}^T] = \mathbb{E}[\mathbf{a} \mathbf{X}^T \mathbf{a}^T] = \mathbf{a} \mathbf{\Gamma} \mathbf{a}^T.
\]

**Solution to Exercise 1.5.15** Let us fix \( \Delta t = 1/n \) and compute the integral

\[
A(n) = \int_0^1 B_s^{\Delta t} \, ds,
\]

which can be interpreted as the area under the graph of the linearly-interpolated random walk \( B^{\Delta t} \). From geometry of the trajectories of \( B^{\Delta t} \) we realize that the area \( A(n) \) is the sum of areas of \( n+1 \) trapezia. The \( k^{th} \) trapezium’s altitude is \( \Delta t \) (on the x-axes) and its sides are \( S_{k-1} \) and \( S_k \), the process \( (S_n) \) being the random walk used in the construction of \( B^{\Delta t} \). Therefore, the sum of the areas of the trapezia is

\[
A(n) = \frac{1}{2} \Delta t \left( (S_0+S_1) + (S_1+S_2) + \cdots + (S_{n-1}+S_n) \right) = \frac{1}{2} \Delta t \left( (2n-1)X_1 + (2n-3)X_2 + \cdots + 3X_{n-1} + X_n \right),
\]

where \( X_k \) is the increment \( S_k - S_{k-1} \). Since \( X_k \) takes values \( \pm \Delta x = \pm \sqrt{\Delta t} \) with probabilities 1/2, and since \( X_k \) and \( X_l \) are independent for \( k \neq l \), so \( \mathbb{E}[X_k] = 0 \), \( \text{Var}[X_k] = (\Delta x)^2 = \Delta t \), and \( \text{Cov}[X_k, X_l] = 0 \) for \( k \neq l \). Therefore

\[
\mathbb{E}[A(n)] = 0, \quad \text{Var}[A(n)] = \frac{(\Delta t)^2}{4} \left( (2n-1)^2(\Delta x)^2 + (2n-3)^2(\Delta x)^2 + \cdots + (\Delta x)^2 \right) = \frac{(\Delta t \Delta x)^2}{4} \alpha(n),
\]

where \( \alpha(n) = 1^2 + 3^2 + \cdots + (2n-1)^2 \). It can be shown that \( \alpha(n) = 4/3n^3 - 1/3n \), so, using the fact that \( \Delta t \Delta x = n^{-3/2} \) we get

\[
\text{Var}[A(n)] = n^{-3} \left( \frac{1}{3} n^3 - \frac{1}{12} n \right) = \frac{1}{3} \left( 1 - \frac{1}{12} n^{-2} \right).
\]

Letting \( n \to \infty \) and using the approximating property of \( B^{\Delta t} \) we get

\[
\mathbb{E}\left[ \int_0^1 B_t \, dt \right] = \lim_{n \to \infty} \mathbb{E}[A(n)] = 0, \quad \text{Var}\left[ \int_0^1 B_t \, dt \right] = \lim_{n \to \infty} \text{Var}[A(n)] = \lim_{n \to \infty} \left( \frac{1}{3} - \frac{1}{12} n^{-2} \right) = \frac{1}{3}.
\]

**Solution to Exercise 1.5.16**

Last Updated: Spring 2005 53 Continuous-Time Finance: Lecture Notes
1.5. BROWNIAN MOTION DEFINED

- The trajectories of \((Y_t)_{t \in [0, \infty)}\) are continuous because they are obtained from the trajectories of the Brownian Bridge by composition with the continuous function \(t \mapsto \frac{t}{1+t}\), and multiplication by the continuous function \(t \mapsto (1 + t)\). The trajectories of the Brownian Bridge are continuous because they are obtained from the (continuous) trajectories of the Brownian motion by adding the continuous function \(-t \mapsto tB_1\).

- It can easily be shown that the process \(Y\) is Gaussian by considering linear combinations \(\alpha_1Y_1 + \alpha_2Y_2 + \cdots + \alpha_nY_n\), and rewriting them in terms of the original Brownian motion \(B\) (just like we did in class and in HW 2).

- The expectation function \(\mu_Y\) coincides with the one for Brownian motion \((\mu_B(t) = 0)\) since
  
  \[
  E[Y_t] = E[(1 + t)X_{t/(1+t)}] = (1 + t)E[B_{\frac{t}{1+t}}] - \frac{t}{1+t}B_1 = (1 + t)E[B_{\frac{t}{1+t}}] - tE[B_1] = 0.
  \]

  The same is true for the covariance function, because for \(s \leq t\) we have
  
  \[
  c_Y(s, t) = E[Y_s Y_t] = (1 + t)(1 + s)E[(B_{\frac{t}{1+t}}) - \frac{t}{1+t}B_1)(B_{\frac{s}{1+s}} - \frac{s}{1+s}B_1)]
  \]
  
  \[
  = (1 + t)(1 + s) \left( \frac{s}{1+s} - \frac{s}{1+s} \frac{t}{1+t} \frac{s}{1+s} \frac{t}{1+t} \right)
  \]
  
  \[
  = s = \min(s, t) = c_B(s, t).
  \]

For the Brownian Bridge \(X_t\) we have \(\lim_{t \to 1^-} X_t = 0\), so

\[
\lim_{t \to \infty} \frac{Y_t}{t} = \lim_{t \to \infty} \frac{(1 + t)X_{t/(1+t)}}{t} = \left[ s = \frac{t}{1+t} \right] = \lim_{s \to 1^-} \frac{1}{s} X_s = 0.
\]

We know that \(Y_t\) is a Brownian motion, so we are done.

**Solution to Exercise 1.5.17**

To compute the distribution function \(F_{R_t}(x) = \mathbb{P}[R_t \leq x]\) for \(t > 0\) and \(x > 0\), we write

\[
R_t^2 = (B_t^1)^2 + (B_t^2)^2
\]

so that

\[
F_{R_t}(x) = \mathbb{P}[R_t^2 \leq x_2] = \mathbb{P}[(B_t^1)^2 + (B_t^2)^2 \leq x^2] = \mathbb{P}[(B_t^1, B_t^2) \in D(x)],
\]

where \(D(x)\) is the disk of radius \(x\) centered at the origin. The sought-for probability can be found by integrating the joint density of the random vector \((B_t^1, B_t^2)\) over the region \(D(x)\). Since \(B_t^1\) and \(B_t^2\) are independent, the joint density \(f_{(B_t^1, B_t^2)}(x_1, x_2)\) is given by

\[
f_{(B_t^1, B_t^2)}(x_1, x_2) = \varphi_1(x_1) \varphi_2(x_2),
\]

where \(\varphi_i\) is the density of the normal random variable \(B_t^i\) (and of \(B_t^2\) as well). We are dealing with Brownian motions, so \(B_t^1 \sim N(0, t)\) and

\[
f_{(B_t^1, B_t^2)}(x_1, x_2) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x_1^2}{2t}} \cdot \frac{1}{\sqrt{2\pi t}} e^{-\frac{x_2^2}{2t}} = \frac{1}{2\pi t} e^{-\frac{x_1^2 + x_2^2}{2t}}.
\]
CHAPTER 1. BROWNIAN MOTION 1.5. BROWNIAN MOTION DEFINED

You can, of course, get the same expression for the joint density by calculating the mean and the variance-covariance matrix of the multivariate-normal random vector \((B_1^t, B_2^t)\) and using the formula given in the notes.

We proceed by writing

\[
\mathbb{P}[R_t \leq x] = \int\int_{x_1^2 + x_2^2 \leq x^2} \frac{1}{2\pi t} e^{-\frac{x_1^2 + x_2^2}{2t}} \, dx_1 \, dx_2,
\]

and passing to polar coordinates \((r, \theta)\) (and remembering that \(dx_1 \, dx_2 = r \, dr \, d\theta\) simplifies the integral to

\[
\mu(t), \quad t \in [0, 5].
\]

Now it is very easy to differentiate \(F_{R_t}(x)\) with respect to \(x\) to obtain

\[
f_{R_t}(x) = \frac{x}{te^{-\frac{x^2}{2t}}}, \quad x > 0.
\]

(The distribution with such density is called the Gamma-distribution).

In order to compute \(\mu(t) = \mathbb{E}[R_t]\) and \(\sigma(t) = \sqrt{\text{Var}[R_t]}\) we simply need to integrate (using Maple, e.g.)

\[
\mu(t) = \int_0^\infty x f_{R_t}(x) \, dx = \int_0^\infty \frac{x^2}{t} e^{-\frac{x^2}{2t}} \, dx = \sqrt{\frac{\pi}{2}} \sqrt{t}.
\]

Similarly,

\[
\mathbb{E}[R_t^2] = \int_0^\infty x^2 f_{R_t}(x) \, dx = \int_0^\infty \frac{x^3}{t} e^{-\frac{x^2}{2t}} \, dx = 2t,
\]

so \(\sigma(t) = \sqrt{\text{Var}[R_t]} = \sqrt{\mathbb{E}[R_t^2] - \mathbb{E}[R_t]^2} = \sqrt{(2 - \frac{t}{2})} \sqrt{t}.\) The graphs of \(\mu(t)\) and \(\sigma(t)\) are given in Figures 2. and 3.

Solution to Exercise 1.5.18

1. To compute the expectation \(\mathbb{E}[S_t]\) observe that \(S_t = g(B_t)\) where \(g(x) = e^{\alpha x + \beta t}\), so that \(\mathbb{E}[S_t] = s_0 \mathbb{E}[g(B_t)]\) and therefore

\[
\mathbb{E}[S_t] = s_0 \int_{-\infty}^{\infty} g(x) \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \, dx = s_0 \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{1}{t}x^2 + \alpha x + \beta t} \, dx,
\]

since \(B_t\) has the normal distribution with mean 0, and variance \(t\). The last integral can be evaluated by completing the square in the exponent, or using a software package, so we have

\[
\mathbb{E}[S_t] = s_0 e^{t(\beta + 1/2\alpha^2)}
\]

and \(\mathbb{E}[S_t] = s_0\) if \(\beta = -\alpha^2/2\).
2. The rate $b(t)$ of (expected) return over the interval $[0, t]$ for the stock $S$ can be defined by

$$b(t) = \frac{\log(E[S_t])}{\log(s_0)} = t(\beta + \alpha^2/2),$$

and the instantaneous rate of return is then $\frac{d}{dt}b(t) = \beta + \alpha^2/2$. The apparent paradox comes from the fact that even for negative $\beta$, the rate of return can be positive. So if we take a Brownian motion with a negative drift (a process that decreases on average) and apply to it a deterministic function (exp) which does not depend on $t$, we get a process which increases on average. Weird . . .

Solution to Exercise 1.5.19

(a) We start by computing the value of the distribution function

$$F(x) = \mathbb{P}[Y \leq x] = \mathbb{P}\left[\frac{X_1}{X_2} \leq x\right] = \mathbb{P}[(X_1, X_2) \in A],$$

where $A = \{(x_1, x_2) \in \mathbb{R}^2 : \frac{x_1}{x_2} \leq x\}$. The joint density function of the random vector $(X_1, X_2)$ is given by

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

Therefore,

$$F(x) = \int_A f_{X_1,X_2}(x_1, x_2) \, dx_1 \, dx_2$$

Without any loss of generality we can assume $x > 0$ because of symmetry of $Y$, so that the region $A$ looks like given in Figure 1., where the slanted line has slope $1/x$. The density function $f_{X_1,X_2}$, and the integration region $A$ are rotationally symmetric, so passing to the polar coordinates makes sense. In polar coordinates the region $A$ is given by

$$A = \{(r, \phi) \in [0, \infty) \times [0, 2\pi) : \phi \in [\arctan(1/x), \pi] \cup [\pi + \arctan(1/x), 2\pi]\},$$

and the integral simplifies to

$$F(x) = \frac{1}{2\pi} \left( \int_{\arctan(1/x)}^{\pi} d\phi + \int_{\pi + \arctan(1/x)}^{2\pi} d\phi \right) = 1 - \frac{1}{\pi} \arctan(1/x),$$

and thus, by differentiation,

$$f_Y(x) = \frac{1}{\pi (1 + x^2)},$$

and we conclude that $Y$ has a Cauchy distribution.

(b) For an arbitrary, but fixed $t > 0$, the random variables $X_1 = \frac{1}{\sqrt{\frac{3}{4}t}} (B_t - B_{\frac{3}{4}t})$ and $X_2 = \frac{1}{\sqrt{\frac{3}{4}t}} (B_{\frac{3}{4}t} - B_{\frac{1}{4}t})$ are independent unit normals (we just have to use the defining properties of Brownian motion). Therefore, part (a) implies that the distribution of $Y_t$ is Cauchy. Cauchy is obviously not normal, so $Y_t$ is not a Gaussian process.
1.6 Simulation

In this section give a brief introduction into the topic of simulation of random variables, random vectors and random processes, and the related Monte Carlo method. There is a vast literature on the topic and numerous web pages dedicated to different facets of the problem.

Simulation of random quantities is one of the most versatile numerical methods in applied probability and finance because of its (apparent) conceptual simplicity and ease of implementation.

1.6.1 Random Number Generators

We start off by introducing the fundamental ingredient of any simulation experiment - the random number generator or RNG. This will usually be a computer function which, when called, produces a number in the range $[0, 1]$. Sometimes (depending of the implementation) the random number generator will produce an integer between 0 and some large number $\text{RAND}_\text{MAX}$, but a simple division (multiplication) by $\text{RAND}_\text{MAX}$ will take care of the transition between the two. We shall therefore talk exclusively about RNGs with the range $[0, 1]$, and the generic RNG function will be denoted by $\text{rand}$, after its MATLAB implementation.

So, far there is nothing that prevents $\text{rand}$ from always returning the same number 0.4, or the sequence 0.5, 0.25, 0.125, ... Such a function will, however, hardly qualify for an RNG since the values it spits out come in a predictable order. We should, therefore, require any candidate for a random number generator to produce a sequence of numbers which is as unpredictable as possible. This is, admittedly, a hard task for a computer having only deterministic functions in its arsenal, and that is why the random generator design is such a difficult field. The state of the affairs is that we speak of good or less good random number generators, based on some statistical properties of the produced sequences of numbers.

One of the most important requirements is that our RNG produce uniformly distributed numbers in $[0, 1]$ - namely - the sequence of numbers produced by $\text{rand}$ will have to cover the interval $[0, 1]$ evenly, and, in the long run, the number of random numbers in each subinterval $[a, b]$ if $[0, 1]$ should be proportional to the length of the interval $b - a$. This requirement if hardly enough, because the sequence

$$0, 0.1, 0.2, \ldots, 0.8, 0.9, 1, 0.05, 0.15, 0.25, \ldots, 0.85, 0.95, 0.025, 0.075, 0.125, 0.175, \ldots$$

will do the trick while being perfectly predictable.

To remedy the inadequacy of the RNGs satisfying only the requirement of uniform distribution, we might require $\text{rand}$ to have the property that the pairs of produced numbers cover the square $[0, 1] \times [0, 1]$ uniformly. That means that, in the long run, every patch $A$ of the square $[0, 1] \times [0, 1]$ will contain the proportion of pairs corresponding to its area. Of course, one could continue with such requirements and ask for triples, quadruples, ... of random
numbers to be uniform in the \([0, 1]^3, [0, 1]^4\ldots\). The highest dimension \(n\) such that the RNG produces uniformly distributed in \([0, 1]^n\) is called the order of the RNG. In the appendix I am including the C source code of Mersenne Twister, an excellent RNG with the order of 623.

Another problem with RNGs is that the numbers produced will start to repeat after a while (this is a fact of life and finiteness of your computer’s memory). The number of calls it takes for a RNG to start repeating its output is called the period of a RNG. You might have wondered how is it that an RNG produces a different number each time it is called, since, after all, it is only a function written in some programming language. Most often, RNGs use a hidden variable called the random seed which stores the last output of \(\text{rand}\) and is used as an (invisible) input to the function \(\text{rand}\) the next time it is called. If we use the same seed twice, the RNG will produce the same number, and so the period of the RNG is limited by the number of different possible seeds. Some operating systems (UNIX, for example) have a system variable called random seed which stores the last used seed and writes it to the hard-disk, so that the next time we use \(\text{rand}\) we do not get the same output. For Windows, there is no such system variable (up to my knowledge) so in order to avoid getting the same random numbers each time you call \(\text{rand}\), you should provide your own seed. To do this you can either have your code do the job of the system and write the seed to a specific file every time you stop using the RNG, or use the system clock to provide you with a seed. The second method is of dubitable quality and you shouldn’t use it for serious applications.

In order to assess the properties of a RNG, a lot of statistical tests have been developed in the last 50 years. We will list only three here (without going much into the details or the theory) to show you the flavor of the procedures used in practice.

They all require a sequence of numbers \((x_n)_{n\in\{1,2,\ldots,N\}}\) to be tested.

- **uniform distribution** Plot the histogram of \(x_1, x_2, \ldots\) by dividing the interval \([0, 1]\) into bins (the good size of the bin is proportional to \(\sqrt{N}\)). All the bins have an approximately equal number of \(x_n\)’s.

- **2,3-dim uniform distribution** Divide your sequence into pairs \((x_1, x_2), (x_3, x_4), (x_5, x_6), \ldots\), and plot the obtained points in the unit square \([0, 1] \times [0, 1]\). There should be no patterns. You can do the same for 3 dimensions, but not really for 4 or more.

- **Maximum of Brownian motion** Use your numbers to produce a sequence \((y_n)_{n\in\mathbb{N}}\) with a (pseudo-) normal distribution (we shall see later how to do that). Divide the obtained sequence into \(M\) subsequences of equal length, and use each to simulate a Brownian motion on the unit interval. For each of the \(M\) runs, record the maximum of the simulated trajectory and draw the histogram of \(M\) the \(M\) maxima. There is
a formula for the density $f_{\text{max}}B$ of the maximum of the trajectory of the Brownian motion, so it is easy to compare the histogram you’ve obtained with $f_{\text{max}}B$.

The simplest RNGs are so-called linear congruential random number generators and a large proportion of the generators implemented in practice belong to this class. They produce random integers on some range $[0, M-1]$, and the idea behind their implementation is simple. Pick a large number $M$, another large number $m$, and yet another number $c$. Start with the seed $x_1 \in [0, M-1]$. You get $x_n$ from $x_{n-1}$ using the formula

$$x_n = mx_{n-1} + c \pmod{M},$$

i.e. multiply $x_{n-1}$ by $m$, add $c$, and take the reminder from the division of the result by $M$. The art is, of course, in the choice of the numbers $M$, $m$, $c$ and $x_1$ and it is easy to come up with examples of linear congruential generators with terrible properties ($m = 1$, $M = 10$, for example).

Finally, we give an example of a (quite bad) random number generator that has been widely used in the sixties and seventies.

**Example 1.6.1.**

RANDU is a linear congruential random number generator with parameters $m = 65539$, $c = 0$ and $M = 2^{31}$. Here is a histogram of the distribution of RANDU. It looks OK.

![Fig 1. Histogram of RANDU output](image1)

Here is a 2-d plot of the pairs of RANDU’s pseudorandom numbers. Still no apparent pattern.

![Fig 2. 2-d plot RANDU output](image2)
Finally, the 3-d plot of the triplets of RANDU’s pseudorandom numbers reveals big problem. See how all the points lie in a dozen-or-so planes in 3D. That wouldn’t happen for truly random numbers. This is called the Marsaglia effect, after M. Marsaglia who published the paper *Random Numbers Fall Mainly in the Planes* in 1968.

![3-d plots RANDU output](image)

**Fig 3.** 3-d plots RANDU output

### 1.6.2 Simulation of Random Variables

Having found a random number generator good enough for our purposes, we might want to use it to simulate random variables with distributions different from the uniform on $[0, 1]$. This is almost always achieved through transformations of the output of a RNG, and we will present several methods for dealing with this problem.

1. **Discrete Random Variables** Let $X$ have a discrete distribution given by

\[
X \sim \begin{pmatrix}
 x_1 & x_2 & \cdots & x_n \\
p_1 & p_2 & \cdots & p_n
\end{pmatrix}.
\]

For discrete distributions taking an infinite number of values we can always truncate at a very large $n$ and approximate it with a distribution similar to the one of $X$. We know that the probabilities $p_1, p_2, \ldots, p_n$ add-up to 1, so we define the numbers $0 = q_0 < q_1 < \cdots < q_n = 1$

\[
q_0 = 0, \ q_1 = p_1, \ q_2 = p_1 + p_2, \ldots, \ q_n = p_1 + p_2 + \cdots + p_n = 1.
\]

To simulate our discrete random variable $X$, we call rand and then return $x_1$ if $0 \leq \text{rand} < q_1$, return $x_2$ if $q_1 \leq \text{rand} < q_2$, and so on. It is quite obvious that this procedure indeed simulates a random variable $X$.  
2. **The Method of Inverse Functions** The basic observation in this method is that, for any continuous random variable \( X \) with the distribution function \( F_X \), the random variable \( Y = F_X(X) \) is uniformly distributed on \([0,1]\). By inverting the distribution function \( F_X \) and applying it to \( Y \), we recover \( X \). Therefore, if we wish to simulate a random variable with an invertible distribution function \( F \), we first simulate a uniform random variable on \([0,1]\) (using `rand`) and then apply the function \( F^{-1} \) to the result. Of course, this method fails if we cannot write \( F^{-1} \) in closed form.

**Example 1.6.2. (Exponential Distribution)** Let us apply the method of inverse functions to the simulation of an exponentially distributed random variable \( X \) with parameter \( \lambda \). Remember that \( f_X(x) = \lambda \exp(-\lambda x), \; x > 0 \), and so \( F_X(x) = 1 - \exp(-\lambda x), \; x > 0 \), and so \( F_X^{-1}(y) = -\frac{1}{\lambda} \log(1 - y) \). Since, \( 1 - \text{rand} \) has the same \( \text{U}[0,1] \)-distribution as \( \text{rand} \), we conclude that
\[
\frac{-\log(\text{rand})}{\lambda}
\]
has the required \( \text{Exp}(\lambda) \)-distribution.

**Example 1.6.3. (Cauchy Distribution)** The Cauchy distribution is defined through its density function
\[
f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.
\]
The distribution function \( F_x \) can be determined explicitly in this example:
\[
F_X(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{1}{1 + x^2} \, dx = \frac{1}{\pi} \left( \frac{\pi}{2} + \arctan(x) \right), \quad \text{and so} \quad F_X^{-1}(y) = \tan \left( \pi \left( y - \frac{1}{2} \right) \right),
\]
yielding that \( \tan(\pi(\text{rand} - 0.5)) \) will simulate a Cauchy random variable for you.

3. **The Box-Muller method** This method is useful for simulating normal random variables, since for them the method of inverse function fails (there is no closed-form expression for the distribution function of a standard normal). It is based on a clever trick:

**Proposition 1.6.4.** Let \( Y_1 \) and \( Y_2 \) be independent \( \text{U}[0,1] \)-distributed random variables. Then the random variables
\[
X_1 = \sqrt{-2 \log(1 - Y_1)} \cos(2\pi Y_2), \quad X_2 = \sqrt{-2 \log(1 - Y_1)} \sin(2\pi Y_2)
\]
are independent and standard normal \( (N(0,1)) \).
The proof of this proposition is quite technical, but not hard, so I will omit it.

Therefore, in order to simulate a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$, we produce call the function `rand` twice to produce two random numbers `rand1` and `rand2`. The numbers

$$X_1 = \sqrt{-2 \log(rand1)} \cos(2\pi \text{rand2}), \quad X_2 = \sqrt{-2 \log(rand1)} \sin(2\pi \text{rand2})$$

will be two independent normals. Note that it is necessary to call the function `rand` twice, but we also get two normal random numbers out of it. It is not hard to write a procedure which will produce 2 normal random numbers in this way on every second call, return one of them and store the other for the next call.

4. **Method of the Central Limit Theorem** The following algorithm is often used to simulate a normal random variable:

(a) Simulate 12 independent uniform random variables (`rands`) - $X_1, X_2, \ldots, X_{12}$.
(b) Set $Y = X_1 + X_2 + \cdots + X_{12} - 6$.

The distribution of $Y$ is very close to the distribution of a unit normal, although not exactly equal (e.g. $P[Y > 6] = 0$. and $P[Z > 6] \neq 0$, for a true normal $Z$). The reason why $Y$ approximates the normal distribution well comes from the following theorem

**Theorem 1.6.5.** Let $X_1, X_2, \ldots$ be a sequence of independent random variables, all having the same distribution. Let $\mu = \mathbb{E}[X_1] (= \mathbb{E}[X_2] = \ldots)$ and $\sigma^2 = \text{Var}[X_1] (= \text{Var}[X_2] = \ldots)$. The sequence of normalized random variables

$$\frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}},$$

converges to the normal random variable (in a mathematically precise sense).

The choice of exactly 12 `rands` (as opposed to 11 or 35) comes from practice: it seems to achieve satisfactory performance with relatively low computational cost. Also, the standard deviation of a $U[0, 1]$ random variable is $1/\sqrt{12}$, so the denominator $\sigma\sqrt{n}$ conveniently becomes 1 for $n = 12$.

The figures below show the densities of random variables obtained as (normalized) sums of $n$ independent uniform random variables for $n = 1, 2, 3, 5, 6, 9, 12$. The plots of their densities are in red, while the density of the standard unit normal is added in ‘dot-dash-dot’ blue.
When choosing between this method and the previously described Box-Muller method, many factors have to be taken into consideration. The Box-Muller method uses only 1 rand per call (on average) as opposed to 12 used by the method based on the central limit theorem. On the other hand, the latter method uses only addition and subtraction, whereas the Box-Muller method utilizes more expensive operations cos, sin, log, and √.

The conclusion of the comparison of the two methods will inevitably rely heavily on the architecture you are running the code on, and the quality of the implementation of the functions cos, sin, . . . .

5. Other methods There is a number of other methods for transforming the output of rand into random numbers with prescribed density (rejection method, Poisson trick, . . . ). You can read about them in the free online copy of Numerical recipes in C at http://www.library.cornell.edu/nr/bookcpdf.html

1.6.3 Simulation of Normal Random Vectors

It is our next task to show how to simulate an n-dimensional normal random vector with mean \( \mu \) and variance-covariance matrix \( \Sigma \). If the matrix \( \Sigma \) is the identity matrix, and \( \mu = (0, 0, \ldots, 0) \), then our problem is simple: we simulate n independent unit normals and organize them in a vector \( \gamma \) (I am assuming that we know how to simulate normal random variables - we can use the Box-Muller method, for example).

When \( \Sigma \) is still identity, but \( \mu \) is a general vector, we are still in luck - just add \( \mu + \gamma \). It is the case of general symmetric positive-semidefinite matrix \( \Sigma \) that is interesting. The idea is to apply a linear transformation \( A \) to the vector of independent unit normals \( \gamma \): \( Y = \gamma A \).
Therefore, we are looking for a matrix $A$ such that
\[ \Sigma = \mathbb{E}[Y^T Y] = \mathbb{E}[A^T \gamma^T \gamma A] = A^T \mathbb{E}[\gamma^T \gamma] A = A^T I A = A^T A. \]

In numerical linear algebra the decomposition $\Sigma = A^T A$ is called the \textit{Cholesky decomposition} and, luckily, there are fast and reliable numerical methods for obtaining $A$ from $\Sigma$. For example, the command \texttt{chol} in MATLAB will produce $A$ from $\Sigma$.

To recapitulate, to simulate an $n$-dimensional normal random vector with mean $\mu$ and variance-covariance matrix $\Sigma$, do the following

1. simulate $n$ independent unit normal random variables and put them in a vector $\gamma$.
2. compute the Cholesky decomposition $A^T A = \Sigma$ of the variance-covariance matrix $\Sigma$.
3. the required vector is $Y = \gamma A + \mu$.

\subsection*{1.6.4 Simulation of the Brownian Motion and Gaussian Processes}

We have already seen that the Brownian motion is a limit of random walks, and that will be the key insight for building algorithms for simulating the paths of Brownian motion. As is inevitably the case with stochastic processes, we will only be able to simulate their paths on a finite interval $[0, T]$, sampled on a finite number of points $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$.

When dealing with continuous processes, we will usually interpolate linearly between these points if necessary.

The independence of increments of Brownian motion will come in very handy because we will be able to construct the whole trajectory from a number of independent steps. Thus, the procedure is the following (I will describe only the case where the points $t_0, t_1, t_2, \ldots, t_n$ are equidistant, i.e. $t_k - t_{k-1} = \Delta t = T/n$.)

1. Choose the horizon $T$ and the number $n$ of time-steps. Take $\Delta t = T/n$.
2. Simulate $n$ independent normal random variables with variance $\Delta t$, and organize them in a vector $\Delta B$.
3. the value of the simulated Brownian motion at $t_i$ will be the sum of the first $i$ terms from $\Delta B$.

The increments do not need to be normal. By the Central Limit Theorem, you can use the Bernoulli random variables with space-steps $\Delta x = \sqrt{\Delta t}$.

The case of a general Gaussian process is computationally more intensive, but conceptually very simple. The idea is to discretize time, i.e. take the points $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ and view the process $(X_t)_{t \in [0,T]}$ you are simulating as being closely approximated by huge multivariate normal vector $X = (X_{t_0}, X_{t_1}, \ldots, X_{t_n})$. So, simulate the normal vector $X$. 

Last Updated: Spring 2005
- its variance-covariance matrix and the mean-vector can be read off the functions \( c_X \) and \( \mu_X \) specified for the Gaussian process \( X \).

Of course, the simulation of a normal vector involves the Cholesky decomposition of the variance-covariance matrix - a relatively expensive operation from the computational point of view. It is therefore, much more efficient to use a special structure of the Gaussian process (when possible) to speed up the procedure - i.e. in the case of the Brownian motion, Brownian motion with drift or the Brownian Bridge.

### 1.6.5 Monte Carlo Integration

Having described some of the procedures and methods used for simulation of various random objects (variables, vectors, processes), we turn to an application in numerical mathematics.

We start off by the following version of the Law of Large Numbers which constitutes the theory behind most of the Monte Carlo applications

**Theorem 1.6.6. (Law of Large Numbers)** Let \( X_1, X_2, \ldots \) be a sequence of identically distributed random variables (works for vectors, too), and let \( g : \mathbb{R} \to \mathbb{R} \) be function such that \( \mu = \mathbb{E}[g(X_1)] (= \mathbb{E}[g(X_2)] = \ldots ) \) exists. Then

\[
\frac{g(X_1) + g(X_2) + \cdots + g(X_n)}{n} \to \mu = \int_{-\infty}^{\infty} g(x) f_{X_1}(x) \, dx, \quad \text{as} \quad n \to \infty.
\]

The key idea of Monte Carlo integration is the following

Suppose that the quantity \( y \) we are interested in can be written as \( y = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \) for some random variable \( X \) with density \( f_X \) and some function \( g \), and that \( x_1, x_2, \ldots \) are random numbers distributed according to the distribution with density \( f_X \). Then the average

\[
\frac{1}{n} (g(x_1) + g(x_2) + \cdots + g(x_n)),
\]

will approximate \( y \).

It can be shown that the accuracy of the approximation behaves like \( 1/\sqrt{n} \), so that you have to quadruple the number of simulations if you want to double the precision of your approximation.

**Example 1.6.7.**

1. (**numerical integration**) Let \( g \) be a function on \([0, 1]\). To approximate the integral \( \int_{0}^{1} g(x) \, dx \) we can take a sequence of \( n \) (\( U[0,1] \)) random numbers \( x_1, x_2, \ldots \),

\[
\int_{0}^{1} g(x) \, dx \approx \frac{g(x_1) + g(x_2) + \cdots + g(x_n)}{n},
\]
because the density of $X \sim U[0, 1]$ is given by

$$f_X(x) = \begin{cases} 
1, & 0 \leq x \leq 1 \\
0, & \text{otherwise}
\end{cases}.$$ 

2. (estimating probabilities) Let $Y$ be a random variable with the density function $f_Y$. If we are interested in the probability $P[Y \in [a, b]]$ for some $a < b$, we simulate $n$ draws $y_1, y_2, \ldots, y_n$ from the distribution $F_Y$ and the required approximation is

$$P[Y \in [a, b]] \approx \frac{\text{number of } y_n \text{’s falling in the interval } [a, b]}{n}.$$ 

One of the nice things about the Monte-Carlo method is that even if the density of the random variable is not available, but you can simulate draws from it, you can still preform the calculation above and get the desired approximation. Of course, everything works in the same way for probabilities involving random vectors in any number of dimensions.

3. (approximating $\pi$)

We can devise a simple procedure for approximating $\pi \approx 3.141592$ by using the Monte-Carlo method. All we have to do is remember that $\pi$ is the area of the unit disk. Therefore, $\pi/4$ equals to the portion of the area of the unit disk lying in the positive quadrant (see Figure), and we can write

$$\frac{\pi}{4} = \int_0^1 \int_0^1 g(x, y) \, dx \, dy,$$

where

$$g(x, y) = \begin{cases} 
1, & x^2 + y^2 \leq 1 \\
0, & \text{otherwise.}
\end{cases}$$

So, simulate $n$ pairs $(x_i, y_i), i = 1 \ldots n$ of uniformly distributed random numbers and count how many of them fall in the upper quarter of the unit circle, i.e. how many satisfy $x_i^2 + y_i^2 \leq 1$, and divide by $n$. Multiply your result by 4, and you should be close to $\pi$. How close? Well, that is another story ... Experiment!

4. (pricing options) We can price many kinds of European- and Asian-type options using Monte Carlo, but all in good time ...
Exercises

Exercise 1.6.8. (Testing RANDU) Implement the linear congruential random number generator RANDU (from the notes) and test its performance by reproducing the 1-d, 2-d and 3-d uniformity-test graphs (just like the ones in the notes). (I suggest using Maple for the 3-d plot since it allows you to rotate 3d graphs interactively.)

Exercise 1.6.9. (Invent your own Random Number Generator)

Use your imagination and produce five different functions that deserve the name random number generator. One of them can be a linear congruential RNG with some parameters of your choice, but the others must be original. Test them and discuss their performance briefly.

Exercise 1.6.10. (Invent your own Random Number Generator Tests) Use more of your imagination to devise 3 new RNG tests. The idea is to simulate some quantity many times and then draw the histogram of your results. If you can compute explicitly the density of the distribution of that quantity, than you can compare your results and draw conclusions. Test you tests on the RNG’s you invented in the previous problem. (Hint: one idea would be to generate \( m \) batches of \( n \) random numbers, and then transform each random number into 0 or 1 depending whether it is \( \geq 0.5 \) or \( < 0.5 \). In this way you obtain \( m \) sequences of \( n \) zeroes and ones. In each of the \( m \) sequences count the longest streak of zeroes. You will get \( m \) natural numbers. Draw their histogram and compare to the theoretical distribution. How do we get the theoretical distribution? Well, you can either look it up in one of the many probability books in the library, search the web, or use a random number generator you know is good to get the same histogram.

For the purposes of this homework assignment, you can assume that the default RNG of your software package is good. )

Exercise 1.6.11. (Trying out Various Methods)

Using the default RNG of your package and the transformation methods we mentiond in class simulate \( n \) draws from each of the following univariate distributions and draw the histograms of your results

1. Exp(\( \lambda \)) with \( \lambda = 1 \).
2. Cauchy
3. Binomial with \( n = 50 \) and \( p = 0.3 \)
4. Poisson with parameter \( \lambda = 1 \) (truncate at a large value of your choice)
5. Normal with \( \mu = 2, \sigma = 3 \) using the Central Limit Theorem Method
6. Normal with $\mu = 2$, $\sigma = 3$ using the Box-Muller Method

**Exercise 1.6.12. (Brownian Bridge)** Simulate and plot 5 trajectories of the Brownian Bridge by

1. simulating trajectories of a Brownian motion first, and then transforming them into the trajectories of the Brownian Bridge
2. using the fact that Brownian Bridge is a Gaussian process and simulate it as a (large) multivariate normal random vector.

**Exercise 1.6.13. (Computing the value of $\pi$)**

Use Monte-Carlo integration to approximate the value of $\pi$ (just like we described it in class). Vary the number of simulations ($n = 10$, $n = 100$, $n = 500$, ...) and draw the graph of $n$ vs. the accuracy of your approximation (absolute value of the difference between the value you obtained and the true value of $\pi$.)

**Exercise 1.6.14. (Probability of 2 stocks going up)** The joint distribution of two stocks $A$ and $B$ is bivariate normal with parameters $\mu_A = 100$, $\mu_B = 120$, $\sigma_A = 5$, $\sigma_B = 15$ and $\rho = 0.65$. Use the Monte-Carlo method to calculate the probability of the event in which both stocks outperform their expectations by at least 10, i.e. $P[A \geq 110, \text{ and } B \geq 120]$. (The trick in the notes we used to solve a similar problem will not work here. You really need a numerical method (such as Monte Carlo) to solve this problem.)
Solutions to Exercises in Section 1.6
1.7 Conditioning

Derek the Daisy’s portfolio consists of two securities - a share of DMF (Daisy’s Mutual Fund) and a share of GI (Grasshopper’s Jumping Industries) whose prices on May, 1st are believed to follow a bivariate normal distribution with coefficients \( \mu_X = 120, \sigma_X = 10 \), \( \mu_Y = 130, \sigma_Y = 5 \) and \( \rho = 0.8 \). Derek’s friend Dennis the Grasshopper serves as the member of the board of GI and has recently heard that GI will enter a merger with CHC (Cricket’s Hopping Corporation) which will drive the price of a share of GJI to the level of 150 on May 1st. This is, of course, insider information, but it is not illegal in the Meadow-world so we might as well exploit it. The question Derek is facing is: what will happen to his other security - one share of DMF? How can he update his beliefs about its distribution, in the light of this new information. For example, what is the probability that DMF will be worth less than 140?

1.7.1 Conditional Densities

Let the random vector \((X_1, X_2)\) stand for the prices of DMF and GI on May, 1st. The elementary probability would try to solve Derek’s problem by computing the conditional probability

\[
P[X_1 \leq 140 | X_2 = 150] = \frac{P\{X_1 \leq 140 \cap X_2 = 150\}}{P\{X_2 = 150\}} = \frac{0}{0},
\]

because \(P[X_2 = 150] = 0\) - the naive approach fails miserably. The rescue comes by replacing the 0-probability-event \(\{X_2 = 150\}\) by an enlargement \(\{X_2 \in (150 - \varepsilon, 150 + \varepsilon)\}\) and letting \(\varepsilon \to 0\). Here are the details . . .

Let \(f_{X_1, X_2}(x_1, x_2)\) be the joint density of \((X_1, X_2)\) and let \(f_{X_2}(x_2)\) be the marginal density of \(X_2\), i.e.

\[
f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1, X_2}(y_1, x_2) \, dy_1.
\]

Then \(P\{X_2 \in (150 - \varepsilon, 150 + \varepsilon)\} = \int_{150-\varepsilon}^{150+\varepsilon} f_{X_2}(y_2) \, dy_2\) and

\[
P\{X_1 \leq 140 \cap X_2 \in (150 - \varepsilon, 150 + \varepsilon)\} = \int_{150-\varepsilon}^{150+\varepsilon} \int_{-\infty}^{140} f_{X_1, X_2}(y_1, y_2) \, dy_1 \, dy_2.
\]

So if we define

\[
P[X_1 \leq 140 | X_2 = 150] \triangleq \lim_{\varepsilon \to 0} \frac{P\{X_1 \leq 140 \cap X_2 \in (150 - \varepsilon, 150 + \varepsilon)\}}{P\{X_2 \in (150 - \varepsilon, 150 + \varepsilon)\}},
\]
we have
\[ \mathbb{P}[X_1 \leq 140|X_2 = 150] = \lim_{\varepsilon \to 0} \frac{\int_{-\infty}^{150-\varepsilon} f(x_1, x_2)(y_1, 150) \, dy_1 \, dy_2}{\int_{-\infty}^{150+\varepsilon} f(x_2) \, dy_2}. \]

To proceed with the argument, let us prove a simple lemma:

**Lemma 1.7.1.** For a sufficiently regular (say continuous) function \( h : \mathbb{R} \to \mathbb{R} \) we have
\[
\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} h(y) \, dy = h(x).
\]

*Proof.* Let’s pick a constant \( a \in \mathbb{R} \) and define the indeterminate integral \( H(x) = \int_a^x h(y) \, dy \). Now
\[
\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} h(y) \, dy = \lim_{\varepsilon \to 0} \frac{H(x+\varepsilon) - H(x-\varepsilon)}{2\varepsilon}
= \frac{1}{2} \lim_{\varepsilon \to 0} \left( \frac{H(x+\varepsilon) - H(x)}{\varepsilon} \right) + \frac{1}{2} \lim_{\varepsilon \to 0} \left( \frac{H(x) - H(x-\varepsilon)}{\varepsilon} \right)
= \frac{1}{2} \left( H'(x) + H'(x) \right) = H'(x) = h(x),
\]
by the Fundamental Theorem of Calculus.

Coming back to our discussion of the conditional expectation, we can use the lemma we have just proved, and write
\[
\mathbb{P}[X_1 \leq 140|X_2 = 150] = \lim_{\varepsilon \to 0} \frac{\int_{-\infty}^{150-\varepsilon} f(x_1, x_2)(y_1, 150) \, dy_1 \, dy_2}{\int_{-\infty}^{150+\varepsilon} f(x_2) \, dy_2} = \frac{\int_{-\infty}^{150} f(x_1, x_2) \, dy_1 \, f(x_2) \, dy_2}{\int_{-\infty}^{150} f(x_2) \, dy_2}
\]

In the case of the bivariate normal, we know the forms of the densities involved:
\[
f_{(x_1, x_2)}(x_1, x_2) = \frac{1}{2\pi \sigma_{X_1} \sigma_{X_2} \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left( \frac{(x_1-\mu_{X_1})^2}{\sigma_{X_1}^2} + \frac{(x_2-\mu_{X_2})^2}{\sigma_{X_2}^2} - 2\rho \frac{(x_1-\mu_{X_1})(x_2-\mu_{X_2})}{\sigma_{X_1} \sigma_{X_2}} \right) \right)
\]
and, since \( X_2 \) is normal with mean \( \mu_{X_2} \) and variance \( \sigma_{X_2}^2 \),
\[
f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi \sigma_{X_2}^2}} \exp \left( -\frac{(x_2-\mu_{X_2})^2}{2\sigma_{X_2}^2} \right).
\]

Thus, when we do the algebra and complete the square we get
\[
\frac{f_{(x_1, x_2)}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{\exp \left( -\frac{1}{2\sigma_{X_1}^2(1-\rho^2)} \left( x_1 - (\mu_{X_1} + \rho \frac{\sigma_{X_1}}{\sigma_{X_2}}(x_2-\mu_{X_2})) \right)^2 \right)}{\sqrt{2\pi \sigma_{X_1}^2(1-\rho^2)}}. \tag{1.7.1}
\]
Plugging \( x_2 = 150 \) into (1.7.1) and integrating from \(-\infty\) to 140 we get

\[
\mathbb{P}[X_1 \leq 140|X_2 = 150] = 0.022, \quad \text{while} \quad \mathbb{P}[X_1 \leq 140] = 0.976,
\]

and we can see how the information that \( X_2 = 150 \) has decreased this probability from 0.98 to 0.02.

We could have repeated the above discussion with any constant \( \xi \) other than 150 and computed the conditional probability of any set \( \{a \leq X_1 \leq b\} \) other than \( \{X_1 \leq 140\} \). We would have obtained that

\[
\mathbb{P}[X_1 \in A|X_2 = \xi] = \int_a^b f_{X_1|X_2}(x_1|X_2 = \xi) \, dx_1,
\]

where

\[
f_{X_1|X_2}(x_1|X_2 = \xi) = \frac{f(x_1, x_2)(x_1, \xi)}{f_{X_2}(\xi)},
\]

and observed that the conditional probabilities can be computed similarly to the ordinary probabilities - we only need to use a different density function. This new density function \( f_{X_1|X_2}(\cdot|X_2 = \xi) \) is called the **conditional density of** \( X_1 \) **given** \( X_2 = \xi \). Observe that for different values of \( \xi \), we get different conditional densities as we should - the updated probabilities depend on the information received.

There is another interesting phenomenon involving multivariate normal distributions - as we have witnessed in (1.7.1) the conditional density of \( X_1 \) given \( X_2 = \xi \) is univariate normal. And the new parameters are

\[
\mu_{X_1}^{\text{new}} = \mu_{X_1} + \rho \frac{\sigma_{X_1}}{\sigma_{X_2}} (\xi - \mu_{X_2}), \quad \sigma_{X_1}^{\text{new}} = \sigma_{X_1} \sqrt{1 - \rho^2}.
\]

This is not an isolated incident - we shall see later that conditional distributions arising from conditioning components of a multivariate normal given (some other) components of that multivariate normal are ... multivariate normal.

### 1.7.2 Other Conditional Quantities

Now when we know how to answer Derek’s question, we can the new concept of conditional expectation to introduce several new (and important) quantities. In what follows we let \((X_1, X_2)\) be a random vector with the joint density function \( f_{X_1,X_2}(x_1, x_2) \).

**Definition 1.7.2.**
1. the **conditional probability** of \( \{X_1 \in A\} \) **given** \( X_2 = \xi \) is defined by

\[
P[X_1 \in A | X_2 = \xi] = \int_A f_{X_1|X_2}(y_1 | X_2 = \xi) \, dy_1.
\]

2. the **conditional expectation** of \( X_1 \) **given** \( X_2 = \xi \) is defined by

\[
E[ X_1 | X_2 = \xi ] = \int_{-\infty}^{\infty} y_1 f_{X_1|X_2}(y_1 | X_2 = \xi) \, dy_1.
\]

3. for a function \( g : \mathbb{R} \to \mathbb{R} \), the **conditional expectation** of \( g(X_1) \) **given** \( X_2 = \xi \) is defined by

\[
E[g( X_1 ) | X_2 = \xi] = \int_{-\infty}^{\infty} g(y_1) f_{X_1|X_2}(y_1 | X_2 = \xi) \, dy_1.
\]

4. the **conditional variance** of \( X_1 \) **given** \( X_2 = \xi \) is defined by

\[
\text{Var}[X_1 | X_2 = \xi] = \int_{-\infty}^{\infty} (y_1 - E[ X_1 | X_2 = \xi])^2 f_{X_1|X_2}(y_1 | X_2 = \xi) \, dy_1.
\]

**Example 1.7.3.** Let \((X_1, X_2)\) have the bivariate normal distribution, just like in the beginning of this section, with parameters \( \mu_{X_1}, \mu_{X_2}, \sigma_{X_1}, \sigma_{X_2}, \) and \( \rho \). Then, using the expression for the conditional density from formula (1.7.1), we have

\[
E[ X_1 | X_2 = \xi] = \mu_{X_1} + \rho \frac{\sigma_{X_1}}{\sigma_{X_2}} (\xi - \mu_{X_2}), \quad \text{Var}[X_1 | X_2 = \xi] = (1 - \rho^2) \sigma_{X_1}^2.
\]

Note how, when you receive the information that \( X_2 = \xi \), the change of your mean depends on \( \xi \), but the decrease in the variance is \( \xi \)-independent.

Of course, conditioning is not reserved only for 2-dimensional random vectors. Let \( \mathbf{X} = (X_1, X_2, \ldots, X_n) \) be a random vector with density \( f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) \). Let us split \( \mathbf{X} \) into 2 sub-vectors \( \mathbf{X}_1 = (X_1, X_2, \ldots, X_k) \) and \( \mathbf{X}_2 = (X_{k+1}, X_{k+2}, \ldots, X_n) \), so that \( \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) \). For \( \xi = (\xi_{k+1}, \xi_{k+2}, \ldots, \xi_n) \), we can mimic the procedure from the beginning of the section, and define the **conditional density** \( f_{\mathbf{X}_1|\mathbf{X}_2}(x_1 | x_2 = \xi_2) \) of the random vector \( \mathbf{X}_1 \) **given** \( \mathbf{X}_2 = \xi \) by

\[
f_{\mathbf{X}_1|\mathbf{X}_2}(x_1 | x_2 = \xi_2) = \frac{f_{\mathbf{X}}(x_1, x_2, \ldots, x_k, \xi_{k+1}, \xi_{k+2}, \ldots, \xi_n)}{f_{\mathbf{X}_2}(\xi_{k+1}, \xi_{k+2}, \ldots, \xi_n)} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, y_2, \ldots, y_k, \xi_{k+1}, \xi_{k+2}, \ldots, \xi_n) \, dy_1 dy_2 \ldots dy_k}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}_2}(\xi_{k+1}, \xi_{k+2}, \ldots, \xi_n) \, dy_1 dy_2 \ldots dy_k}.
\]
1.7.3 σ-algebra = Amount of Information

Remember how I promised to show you another use for a σ-algebra (apart from being a technical nuisance and a topic of a hard homework question)? Well, this is where I'll do it. The last subsection devoted to the conditional densities has shown us how information can substantially change our view of likelihoods of certain events. We have learned how to calculate these new probabilities, in the case when the information supplied reveals the exact value of a random variable, e.g. \( X_2 = \xi \). It this subsection, we would like to extend this notion, and pave the way for the introduction of the concept of conditional expectation with respect to a σ-algebra.

First of all, we need to establish the relation between σ-algebras and information. You can picture information as the ability to answer questions (more information gives you a better score on the test . . . ), and the lack of information as ignorance or uncertainty. Clearly, this simplistic metaphor does no justice to the complexity of the concept of information, but it will serve our purposes. In our setting, all the questions can be phrased in terms of the elements of the state-space \( \Omega \). Remember - \( \Omega \) contains all the possible evolutions of our world (and some impossible ones) and the knowledge of the exact \( \omega \in \Omega \) (the “true state of the world”) amounts to the knowledge of everything. So, a typical question would be “What is the true \( \omega \)?”, and the ability to answer that would promote you immediately to the level of a Supreme Being. So, in order for our theory to be of any use to mortals, we have to allow for some ignorance, and consider questions like “Is the true \( \omega \) an element of the event \( A \)?”, where \( A \) could be the event in which the price of the DMF Mutual fund is in the interval \((120, 150)\) on May, 1st. And now comes the core of our discussion . . . the collection of all events \( A \) such that you know how to answer the question “Is the true \( \omega \) an element of the event \( A \)?” is the mathematical description of your current state of information. And, guess what . . . this collection (let us call it \( \mathcal{G} \)) is a σ-algebra. Why?

- First, I always know that the true \( \omega \) is an element of \( \Omega \), so \( \Omega \in \mathcal{G} \).

- If I know how to answer the question Is the true \( \omega \) in \( A \)? \((A \in \mathcal{G})\), I will also know how to answer the question “Is the true \( \omega \) in \( A^c \)?”. The second answer is just the opposite of the first, so \( A^c \in \mathcal{G} \).

- Let \((A_n)_{n \in \mathbb{N}}\) be a sequence of events such that I know how to answer the questions “Is the true \( \omega \) in \( A_n \)?”, i.e. \((A_n) \in \mathcal{G}\) for all \( n \in \mathbb{N} \). Then I know that the answer to the question Is the true \( \omega \) in \( \bigcup_{n \in \mathbb{N}} A_n \) is “No” if I answered “No” to each question “Is the true \( \omega \) in \( A_n \)?”, and it is “Yes” if I answered “Yes” to at least one of them. Consequently \( \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{G} \).
When your state space consists of a finite number of elements, $\sigma$-algebras are equivalent to partitions of the state space (how exactly?). The finer the partition, the more you know about the true $\omega$. The figure above depicts the evolution of a fictitious stock on a two-day time horizon. The price of the stock is 100 on day 0, moves to one of the three possible values on day 1, and branches further on day 2.

On day 0, the information available to us is minimal, the only questions we can answer are the trivial ones “Is the true $\omega$ in $\Omega$?” and “Is the true $\omega$ in $\emptyset$”, and this is encoded in the $\sigma$-algebra $\{\Omega, \emptyset\}$.

On day 1, we already know a little bit more, having observed the value of the stock - let’s denote this value by $S_1$. We can distinguish between $\omega_1$ and $\omega_5$, for example. We still do not know what will happen the day after, so that we cannot tell between $\omega_1$, $\omega_2$ and $\omega_3$, or $\omega_4$ and $\omega_5$. Therefore, our information partition is $\{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5\}, \{\omega_6\}\}$ and the corresponding $\sigma$-algebra $F_1$ is (I am doing this only once!)

$$F_1 = \{\emptyset, \{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5\}, \{\omega_6\}, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}, \{\omega_1, \omega_2, \omega_3, \omega_6\}, \{\omega_4, \omega_5, \omega_6\}, \Omega\}.$$

The $\sigma$-algebra $F_1$ has something to say about the price on day two, but only in the special case when $S_2 = 0$. If that special case occurs - let us call it bankruptcy - then we do not need to wait until day 2 to learn what the stock price at day two is going to be. It is going to remain 0. Finally, when day 2 dawns, we know exactly what $\omega$ occurred and the $\sigma$-algebra $F_2$ consists of all subsets of $\Omega$.

Let us think for a while how we acquired the extra information on each new day. We...
have learned it through the random variables $S_0, S_1$ and $S_2$ as their values gradually revealed themselves to us. We can therefore say that the $\sigma$-algebra $\mathcal{F}_1$ is generated by the random variable $S_1$, in the notation $\mathcal{F}_1 = \sigma(S_1)$. In other words, $\mathcal{F}_1$ consists of exactly those subsets of $\Omega$ that we can describe in terms of $S_1$ only. Mathematically, $\mathcal{F}_1$ is composed of events $\{\omega \in \Omega : S_1(\omega) \in A\}$ where $A$ can be $\{130\}, \{90\}, \{0\}, \{90,130\}, \{0,90\}, \{0,130\}, \{0,90,130\}$ and, of course, $\emptyset$. The $\sigma$-algebra $\mathcal{F}_2$ is generated by $S_1$ and $S_2$. In the same way we can describe $\mathcal{F}_2$ as the set of all subsets of $\Omega$ which can be described only in terms of random variables $S_1$ and $S_2$. In this case the notation is $\mathcal{F}_2 = \sigma(S_1, S_2)$.

Imagine now a trader who slept through day 1 and woke up on day two to observe $S_2 = 110$. If asked about the price of the stock at day 1, she could not be sure whether it was 130 or 90. In other words, the $\sigma$-algebra $\sigma(S_2)$ is strictly smaller (coarser) than $\sigma(S_1, S_2)$, even though $S_2$ is revealed after $S_1$.

1.7.4 Filtrations

We have described the $\sigma$-algebras $\mathcal{F}_0, \mathcal{F}_1$ and $\mathcal{F}_2$ in the previous subsection and interpreted them as the amounts of information available to our agent at day 0, 1 and 2 respectively. The sequence $(\mathcal{F}_n)_{n \in \{0,1,2\}}$ is an instance of a filtration. In mathematical finance and probability theory, a filtration is any family of $\sigma$-algebras which gets finer and finer (an increasing family of $\sigma$-algebras in parlance of probability theory.)

On a state-space in which a stochastic process $S_n$ is defined, it is natural to define the filtration generated by the process $S$, denoted by $\mathcal{F}_n^S$ by $\mathcal{F}_0^S = \sigma(S_0), \mathcal{F}_1^S = \sigma(S_0, S_1), \ldots, \mathcal{F}_n^S = \sigma(S_0, S_1, \ldots, S_n) \ldots$. The filtration $\mathcal{F}_n^S$ will describe the flow of information of an observer who has access to the value of $S_n$ at time $n$ - no more and no less. Of course, there can be more than one filtration on a given asset-price model. Consider the example from the previous subsection and an insider who knows at time 0 whether or not the company will go bankrupt on day 1, in addition to the information contained in the stock price. His filtration, let us call it $(\mathcal{G}_n)_{n \in \{0,1,2\}}$ will contain more information than the filtration $(\mathcal{F}_n)_{n \in \{0,1,2\}}$. The difference will concern only $\mathcal{G}_0 = \{\emptyset, \{\omega_0\}, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}, \Omega\}$ $\supsetneq \{\emptyset, \Omega\} = \mathcal{F}_0$, since the extra information our insider has will be revealed to everyone at time 1, so that $\mathcal{G}_1 = \mathcal{F}_1$ and $\mathcal{G}_2 = \mathcal{F}_2$. Can we still view the insider’s $\sigma$-algebra $\mathcal{G}_0$ as being generated by a random variable? The answer is “yes” if we introduce the random variable $B$ ($B$ is for bankruptcy, not Brownian motion) and define

$$B(\omega) = \begin{cases} 1, & \omega = \omega_0 \\ 0, & \omega \in \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}. \end{cases}$$

---

3 $\mathcal{F}_2$ in our case contains all the information there is, but you can easily imagine the time to go beyond day two, so that there is more to the world than just $S_1$ and $S_2$. 

Last Updated: Spring 2005
CHAPTER 1. BROWNIAN MOTION 1.7. CONDITIONING

Then $\mathcal{G}_0 = \sigma(B) = \sigma(B, S_0)$. Obviously, $\mathcal{G}_1 = \sigma(B, S_1)$, but the knowledge of $S_1$ implies the knowledge of $B$, so $\mathcal{G}_1 = \sigma(S_1)$.

The random variable $B$ has a special name in probability theory. It is called the indicator of the set $\{\omega_0\} \subseteq \Omega$. In general, for an event $A \subseteq \Omega$ we define the indicator of $A$ as the random variable $1_A$ defined by

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \not\in A. \end{cases}$$

Indicators are useful because they can reduce probabilities to expectations: $\mathbb{P}[A] = \mathbb{E}[1_A]$ (prove this in the finite-\Omega case!) For that reason we will develop the theory of conditional expectations only, because we can always understand probabilities as expectations of indicators.

Finally, let us mention the concept of measurability. As we have already concluded, the insider has advantage over the public only at time 0. At time 1, everybody observes $S_1$ and knows whether the company went bankrupt or not, so that the information in $S_1$ contains all the information in $B$. You can also rephrase this as $\sigma(B, S_1) = \sigma(S_1)$, or $\sigma(B) \subseteq \sigma(S_1)$ - the partition generated by the random variable $S_1$ is finer (has more elements) than the partition generated by the random variable $B$. In that case we say that $B$ is measurable with respect to $\sigma(S_1)$,. In general we will say that the random variable $X$ is measurable with respect to $\sigma$-algebra $\mathcal{F}$ if $\sigma(X) \subseteq \mathcal{F}$.

1.7.5 Conditional Expectation in Discrete Time

Let us consider another stock-price model given described in the figure below. The state space $\Omega$ is discrete and $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. There are 3 time periods, and we always assume that the probabilities are assigned so that in any branch-point all the branches have equal probabilities. Stock prices at times $t = 0, 1, 2, 3$ are given as follows: $S_0(\omega) = 100$, for all $\omega \in \Omega$,

$$S_1(\omega) = \begin{cases} 80, & \omega = \omega_1, \omega_2 \\ 120, & \omega = \omega_3, \omega_4 \\ 140, & \omega = \omega_5, \omega_6 \end{cases},$$

$$S_2(\omega) = \begin{cases} 100, & \omega = \omega_1, \omega_2, \omega_3, \omega_4 \\ 130, & \omega = \omega_5, \omega_6 \end{cases},$$

$$S_3(\omega) = \begin{cases} 80, & \omega = \omega_1 \\ 100, & \omega = \omega_2, \omega_3, \omega_4 \\ 120, & \omega = \omega_5, \omega_6 \\ 140, & \omega = \omega_6 \end{cases}.$$

Fig 10. Another Tree
The question we are facing now is how to use the notion of information to give a meaning to the concept of conditional expectation (and then, conditional probability, . . .). We will use the notation

\[ E[X|F] \]

for the **conditional expectation of the random variable** \( X \) **with respect to** (given) the \( \sigma \)-algebra \( F \). Let us look at some examples, and try to figure out what properties the conditional expectation should have.

**Example 1.7.4.** Suppose you are sitting at time \( t = 0 \), and trying to predict the future. What is your best guess of the price at time \( t = 1 \)? Since there is nothing to guide you in your guessing and tell you whether the stock price will go up to 140, 120, or down to 80, and the probability of each of those movements is \( \frac{1}{3} \), the answer should be

\[
\frac{1}{3} \times 140 + \frac{1}{3} \times 120 + \frac{1}{3} \times 80 = 113 \frac{1}{3},
\]

and we can, thus, say that \( E[S_1|F_0^S] = E[S_1] = 113 \frac{1}{3} \), because the expected value of \( S_1 \) given no information (remember, \( F_0^S = \{\emptyset, \Omega\} \)) is just the (ordinary) expectation of \( S_1 \), and that turns out to be \( 113 \frac{1}{3} \).

**Example 1.7.5.** After 24 hours, day \( t = 1 \) arrives and your information contains the value of the stock-price at time \( t = 1 \) (and all stock prices in the past). In other words, you have \( F_1^S \) at your disposal. The task of predicting the future stock prices at \( t = 2 \) is trivial (since there is a deterministic relationship between \( S_1 \) and \( S_2 \) according to the model depicted in the figure above). So, we have

\[
E[S_2|F_1^S](\omega) = \begin{cases} 
130, & S_1(\omega) = 140, \\
100, & S_1(\omega) = 120 \text{ or } 80
\end{cases} = \ldots \text{surprise, surprise!} \ldots = S_2(\omega).
\]

Note that the conditional expectation depends on \( \omega \), but only through the available information.

**Example 1.7.6.** What about day \( t = 2 \)? Suppose first that we have all the information about the past, i.e. we know \( F_2^S \).

If \( S_2 \) is equal to 130 then we know that \( S_3 \) be equal to either 140, or 120, each with probability 1/2, so that \( E[S_3|F_2^S](\omega) = \frac{1}{2} \times 140 + \frac{1}{2} \times 120 = 130 \) for \( \omega \) such that \( S_2(\omega) = 130 \).

On the other hand, when \( S_2 = 100 \) and \( S_1 = 120 \), the value of \( S_3 \) is equal to either 120 or 100, each with probability 1/2. Similarly, when \( S_2 = 100 \) and \( S_1 = 80 \), \( S_3 \) will choose between 100 and 80 with equal probabilities. To summarize,

\[
E[S_3|F_2^S](\omega) = \begin{cases} 
130, & S_2(\omega) = 130, \\
110 = \frac{1}{2} \times 120 + \frac{1}{2} \times 100, & S_2(\omega) = 100 \text{ and } S_1(\omega) = 120, \\
90 = \frac{1}{2} \times 80 + \frac{1}{2} \times 100, & S_2(\omega) = 100 \text{ and } S_1(\omega) = 80
\end{cases}
\]
Let's now turn to the case in which we know the value of $S_2$, but are completely ignorant of the value of $S_1$. Then, our $\sigma$-algebra is $\sigma(S_2)$ and not $F_2^S = \sigma(S_1, S_2)$ anymore. To compute the conditional expectation of $S_3$ we reason as follows:

If $S_2 = 130$, $S_3$ could be either 120 or 140 with equal probabilities, so that $\mathbb{E}[S_3|\sigma(S_2)](\omega) = \frac{1}{2}120 + \frac{1}{2}140 = 130$ for $\omega$ such that $S_2(\omega) = 120$.

When $S_2 = 100$, we do not have the knowledge of the value of $S_1$ at our disposal to tell us whether the price will branch between 100 and 120 (upper 100-dot in the figure), or between 80 and 100 (lower 100-dot in the figure). Since it is equally likely that $S_1 = 120$ and $S_1 = 80$, we conclude that, given our knowledge, $S_3$ will take values 80, 100 and 120, with probabilities $1/4$, $1/2$ and $1/4$ respectively. Therefore,

$$
\mathbb{E}[S_3|\sigma(S_2)](\omega) = \begin{cases} 
130, & S_2(\omega) = 120 \\
100 = \frac{1}{4}80 + \frac{1}{2}100 + \frac{1}{4}120, & S_2(\omega) = 100.
\end{cases}
$$

What have we learned from the previous examples? Here are some properties of the conditional expectation that should be intuitively clear now. Let $X$ be a random variable, and let $F$ be a $\sigma$-algebra. Then the following hold true

(CE1) $\mathbb{E}[X|F]$ is a random variable, and it is measurable with respect to $F$, i.e. $\mathbb{E}[X|F]$ depends on the state of the world, but only through the information contained in $F$.

(CE2) $\mathbb{E}[X|F](\omega) = \mathbb{E}[X]$, for all $\omega \in \Omega$ if $F = \{\emptyset, \Omega\}$, i.e. the conditional expectation reduces to the ordinary expectation when you have no information.

(CE3) $\mathbb{E}[X|F](\omega) = X(\omega)$, if $X$ is measurable in $F$, i.e. there is no need for expectation when you already know the answer (the value of $X$ is known, when you know $F$).
Let us try to picture the conditional expectations from the previous examples from a slightly different point of view. Imagine $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ as a collection of 6 points on the real line, and random variables as real valued functions there. If we superimpose $S_3$ and $\mathbb{E}[S_3|F_2^S]$ on the same plot, we notice that $\mathbb{E}[S_3|F_2^S]$ is constant on every atom\(^4\) of $F_2^S$, as it should be, since there is no information to differentiate between the $\omega$'s in the same atom. This situation is depicted in the figure on the right.

Moreover, the value of $\mathbb{E}[S_3|F_2^S]$ on each $\omega$ in a given atom is the average of the values of $S_3$ on that atom. In our case, each $\omega \in \Omega$ has the same probability, so that the averaging is performed with the same weights. In the general case, one would use the weights proportional to the probabilities, of course.

\(^4\) an atom of a $\sigma$-algebra $\mathcal{F}$ is each element of the partition that generates it. It is the smallest set of $\omega$'s such that $\mathcal{F}$ cannot distinguish between them.

The figure on the left depicts the conditional expectation with respect to the smaller $\sigma$-algebra $\sigma(S_2)$. Again, the value of $\mathbb{E}[S_3|\sigma(S_2)]$ is constant on $\{\omega_1, \ldots, \omega_4\}$, as well as on $\{\omega_5, \omega_6\}$, and is equal to the averages of $S_3$ on these two atoms. By looking at the last two figures, you can easily convince yourself that we would get the identical answer if we computed the conditional expectation of $Y = \mathbb{E}[S_3|F_2^S]$ with respect to $\sigma(S_2)$ instead of the conditional expectation of $S_3$ with respect to $\sigma(S_2)$. There is more at play here than mere coincidence. The following is always true for the conditional expectation, and is usually called the tower property of conditional expectation:

(CE4) When $X$ is a random variable and $\mathcal{F}, \mathcal{G}$ two $\sigma$-algebras such that $\mathcal{F} \subseteq \mathcal{G}$, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}],$$

i.e.

my expectation (given information $\mathcal{F}$) of the value $X$ is equal to the expectation (given information $\mathcal{F}$) of the expectation I would have about the value of $X$ if I had information $\mathcal{G}$ on top of $\mathcal{F}$.
The discussion above points towards a general algorithm for computing the conditional expectations in the discrete-time case.

### 1.7.6 A Formula for Computing Conditional Expectations

Let \( \Omega = \{ \omega_1, \omega_2, \ldots, \omega_n \} \) be a state space and let the probability \( \mathbb{P} \) be assigned so that \( \mathbb{P}[\omega] > 0 \), for each \( \omega \in \Omega \). Further, let \( X : \Omega \to \mathbb{R} \) be a random variable, and let \( \mathcal{F} \) be a \( \sigma \)-algebra on \( \Omega \) which partitions \( \Omega \) into \( m \) atoms \( A^1 = \{ \omega_1^1, \omega_2^1, \ldots, \omega_n^1 \} \), \( A^2 = \{ \omega_1^2, \omega_2^2, \ldots, \omega_n^2 \} \), \ldots, \( A^m = \{ \omega_1^m, \omega_2^m, \ldots, \omega_n^m \} \) (in which case \( A^1 \cup A^2 \cup \cdots \cup A^m = \Omega \), \( A^i \cap A^j = \emptyset \), for \( i \neq j \) and \( n_1 + n_2 + \ldots + n_m = n \)). Then

\[
\mathbb{E}[X|\mathcal{F}](\omega) = \begin{cases} 
\frac{\sum_{i=1}^{n_1} X(\omega_i^1) \mathbb{P}[\omega_i^1]}{\sum_{i=1}^{n_1} \mathbb{P}[\omega_i^1]} = \left( \frac{\sum_{i=1}^{n_1} X(\omega_i^1) \mathbb{P}[\omega_i^1]}{\mathbb{P}[A^1]} \right), & \text{for } \omega \in A^1, \\
\frac{\sum_{i=1}^{n_2} X(\omega_i^2) \mathbb{P}[\omega_i^2]}{\sum_{i=1}^{n_2} \mathbb{P}[\omega_i^2]} = \left( \frac{\sum_{i=1}^{n_2} X(\omega_i^2) \mathbb{P}[\omega_i^2]}{\mathbb{P}[A^2]} \right), & \text{for } \omega \in A^2, \\
\vdots & \\
\frac{\sum_{i=1}^{n_m} X(\omega_i^m) \mathbb{P}[\omega_i^m]}{\sum_{i=1}^{n_m} \mathbb{P}[\omega_i^m]} = \left( \frac{\sum_{i=1}^{n_m} X(\omega_i^m) \mathbb{P}[\omega_i^m]}{\mathbb{P}[A^m]} \right), & \text{for } \omega \in A^m.
\end{cases}
\]

This big expression is nothing but a rigorous mathematical statement of the facts that the conditional expectation is constant on the atoms of \( \mathcal{F} \) and that its value on each atom is calculated by taking the average of \( X \) there using the relative probabilities as weights. Test your understanding of the notation and the formula above by convincing yourself that

\[
\mathbb{E}[X|\mathcal{F}](\omega) = \sum_{i=1}^{m} \mathbb{E}[X|A^i] 1_{A^i}(\omega), \quad \text{where } \mathbb{E}[X|A] = \mathbb{E}[X 1_A]/\mathbb{P}[A]. \tag{1.7.2}
\]

The first simple application of the formula above is to prove the linearity of conditional expectation:

(CE5) The conditional expectation is linear, i.e. for \( X_1, X_2 \) random variables, \( \alpha_1, \alpha_2 \) real constants, and a \( \sigma \)-algebra \( \mathcal{F} \), we have

\[
\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2|\mathcal{F}] = \alpha_1 \mathbb{E}[X_1|\mathcal{F}] + \alpha_2 \mathbb{E}[X_2|\mathcal{F}].
\]
1.7. CONDITIONING

CHAPTER 1. BROWNIAN MOTION

To establish the truth of the above equality, let \( A^1, \ldots, A^n \) be the atoms of the partition generating \( \mathcal{F} \). We can use the formula (1.7.2) to get

\[
E[\alpha_1 X_1 + \alpha_2 X_2 | \mathcal{F}] (\omega) = \sum_{i=1}^{n} \frac{\alpha_1 E[X_1 1_{A^i}] + \alpha_2 E[X_2 1_{A^i}]}{P[A^i]} 1_{A^i}(\omega)
\]

\[
= \alpha_1 E[X_1 | \mathcal{F}] + \alpha_2 E[X_2 | \mathcal{F}].
\]

Let us give another illustration of the usefulness of the formula (1.7.1) by computing a slightly more complicated conditional expectation in our stock-price model.

**Example 1.7.7.** Suppose, for example that all you know is the difference \( S_3 - S_1 \), so that the your information is described by the \( \sigma \)-algebra generated \( S_3 - S_1 \), and given by the partition \( \{ A^1, A^2, A^3 \} \) where \( A^1 = \{ \omega_1, \omega_4, \omega_6 \} \), \( A^2 = \{ \omega_2 \} \) and \( A^3 = \{ \omega_3, \omega_5 \} \), since

\[
S_3(\omega) - S_1(\omega) = \begin{cases} 
0, & \omega = \omega_1, \omega_4, \omega_6 \\
-20, & \omega = \omega_3, \omega_5 \\
20, & \omega = \omega_2 
\end{cases}
\]

In this case \( m = 3, n_1 = 3, n_2 = 1 \) and \( n_3 = 2 \), and

\[
S_3(\omega) - S_0(\omega) = \begin{cases} 
-20, & \omega = \omega_1 \\
0, & \omega = \omega_2, \omega_3 \\
20, & \omega = \omega_4, \omega_5 \\
40, & \omega = \omega_6 
\end{cases}
\]

so that

\[
E[(S_3 - S_0)|\sigma(S_3 - S_1)] = \begin{cases} 
(\frac{1}{3}(-20) + \frac{1}{6}20 + \frac{1}{6}40)/\frac{1}{2} = 10\frac{1}{3} & \omega \in A^1 \\
\frac{1}{2}0/\frac{1}{2} = 0 & \omega \in A^2 \\
(\frac{1}{6}0 + \frac{1}{6}20)/\frac{1}{3} = 10 & \omega \in A^3 
\end{cases}
\]

1.7.7 Further properties of Conditional Expectation

Having established that the conditional expectation is constant on the atoms of the \( \sigma \)-algebra \( \mathcal{F} \), we have used this fact in the derivation of the formula (1.7.2). We can easily generalize this observation (the proof is left as an exercise), and state it in the form of the following very useful property:

(CE6) Let \( \mathcal{F} \) be a \( \sigma \)-algebra, and let \( X \) and \( Y \) be variables such that \( Y \) is measurable with respect to \( \mathcal{F} \). Then

\[
E[XY | \mathcal{F}] (\omega) = Y(\omega)E[X | \mathcal{F}] (\omega),
\]
i.e. the random variable measurable with respect to the available information can be treated as a constant in conditional expectations.

In our final example, we will see how independence affects conditioning, i.e. how conditioning on irrelevant information does not differ from conditioning on no information.

Example 1.7.8. In this example we will leave our stock-price model and consider a gamble where you draw a card from a deck of 52. The rules are such that you win $X = 13$ if the card drawn is a Jack and nothing otherwise. Since there are 4 Jacks in the deck of 52, your expected winnings are $E[X] = \frac{4}{52} \times 13 + \frac{48}{52} \times 0 = 1$. Suppose that you happen to know (by EPS, let’s assume) the suit of the next card you draw. How will that affect you expected winnings? Intuitively, the expectation should stay the same since the proportion of Jacks among spades (say) is the same as the proportion of Jacks among all cards. To show that this intuition is true, we construct a model where

$$
\Omega = \{\omega_1^\spadesuit, \omega_2^\spadesuit, \ldots, \omega_{13}^\spadesuit, \omega_1^\heartsuit, \omega_2^\heartsuit, \ldots, \omega_{13}^\heartsuit, \omega_1^\clubsuit, \omega_2^\clubsuit, \ldots, \omega_{13}^\clubsuit, \omega_1^\diamondsuit, \omega_2^\diamondsuit, \ldots, \omega_{13}^\diamondsuit\},
$$

and, for example, the card drawn in the state of the world $\omega_{12}^\spadesuit$ is the Jack of spades. Every $\omega \in \Omega$ has the probability of $\frac{1}{52}$. The $\sigma$-algebra $\mathcal{F}$ corresponding to your extra information (the suit of the next card) is generated by the partition whose atoms are $A^\spadesuit = \{\omega_1^\spadesuit, \omega_2^\spadesuit, \ldots, \omega_{13}^\spadesuit\}$, $A^\heartsuit = \{\omega_1^\heartsuit, \omega_2^\heartsuit, \ldots, \omega_{13}^\heartsuit\}$, $A^\clubsuit = \{\omega_1^\clubsuit, \omega_2^\clubsuit, \ldots, \omega_{13}^\clubsuit\}$, and $A^\diamondsuit = \{\omega_1^\diamondsuit, \omega_2^\diamondsuit, \ldots, \omega_{13}^\diamondsuit\}$. By the formula for conditional expectation, we have

$$
E[X|\mathcal{F}] = \begin{cases} 
\left(\frac{1}{52} \times 0 + \frac{1}{52} \times 0 + \cdots + \frac{1}{52} \times 13 + \frac{1}{52} \times 0\right) / \frac{1}{4}, & \omega \in A^\spadesuit, \\
\left(\frac{1}{52} \times 0 + \frac{1}{52} \times 0 + \cdots + \frac{1}{52} \times 13 + \frac{1}{52} \times 0\right) / \frac{1}{4}, & \omega \in A^\heartsuit, \\
\left(\frac{1}{52} \times 0 + \frac{1}{52} \times 0 + \cdots + \frac{1}{52} \times 13 + \frac{1}{52} \times 0\right) / \frac{1}{4}, & \omega \in A^\clubsuit, \\
\left(\frac{1}{52} \times 0 + \frac{1}{52} \times 0 + \cdots + \frac{1}{52} \times 13 + \frac{1}{52} \times 0\right) / \frac{1}{4}, & \omega \in A^\diamondsuit,
\end{cases}
$$

just as we expected.

The identity $E[X|\mathcal{F}] = E[X]$ is not a coincidence, either. The information contained in the $\sigma$-algebra $\mathcal{F}$ is, in a sense, independent of value of $X$. Mathematically we have the following definition

Definition 1.7.9.

- We say that the $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$ are independent if for each $A \in \mathcal{F}$ and $B \in \mathcal{G}$ we have $P[A \cap B] = P[A]P[B]$.

- We say that the random variable $X$ and the $\sigma$-algebra $\mathcal{F}$ are independent if the $\sigma$-algebras $\sigma(X)$ and $\mathcal{F}$ are independent.
1.7. CONDITIONING

In general, we have the following property of conditional expectation

\[(CE7) \text{ If the random variable } X \text{ and the } \sigma\text{-algebra } \mathcal{F} \text{ are independent, then}\]

\[E[X|\mathcal{F}](\omega) = E[X], \text{ for all } \omega,\]

i.e. conditioning with respect to independent information is like conditioning on no information at all.

It is quite easy to prove (CE5) by using the formula (1.7.2) and the fact that, because of independence of \(X\) and \(F\), we have \(E[X1_A] = E[X]E[1_A] = E[X]P[A] \text{ for } A \in \mathcal{F}\) (supply all the details yourself!), so

\[E[X|\mathcal{F}] = \sum_{i=1}^{m} \frac{E[X1_{A_i}]}{P[A_i]}1_{A_i} = \sum_{i=1}^{m} E[X]1_{A_i} = E[X] \sum_{i=1}^{m} 1_{A_i} = E[X],\]

because \((A^i)_{i=1,...,m}\) is a partition of \(\Omega\).

1.7.8 What happens in continuous time?

All of the concepts introduced above can be defined in the continuous time, as well. Although technicalities can be overwhelming, the ideas are pretty much the same. The filtration will be defined in exactly the same way - as an increasing family of \(\sigma\)-algebras. The index set might be \([0, \infty)\) (or \([0, T]\) for some time-horizon \(T\)), so that the filtration will look like \((\mathcal{F}_t)_{t \in [0, \infty)}\) (or \((\mathcal{F}_t)_{t \in [0, T]}\)). Naturally, we can also define discrete filtrations \((\mathcal{F}_n)_{n \in \mathbb{N}}\) on infinite \(\Omega\)'s.

The concept of a \(\sigma\)-algebra generated by a random variable \(X\) is considerably harder to define in continuous time, i.e. in the case of an infinite \(\Omega\), but carries the same interpretation. Following the same motivational logic as in the discrete case, the \(\sigma\)-algebra \(\sigma(X)\) should contain all the subsets of \(\Omega\) of the form \(\{\omega \in \Omega : X(\omega) \in (a, b)\}\), for \(a < b\), such that \(a, b \in \mathbb{R}\). These will, unfortunately, not form a \(\sigma\)-algebra, so that (and this is a technicality you can skip if you wish), \(\sigma(X)\) is defined as the smallest \(\sigma\)-algebra containing all the sets of the form \(\{\omega \in \Omega : X(\omega) \in (a, b)\}\), for \(a < b\), such that \(a, b \in \mathbb{R}\). In the same way we can define the \(\sigma\)-algebra generated by 2 random variables \(\sigma(X,Y)\), or infinitely many \(\sigma((X_s)_{s \in [0,t]})\). Think of them as the \(\sigma\)-algebras containing all the sets whose occurrence (or non-occurrence) can be phrased in terms of the random variables \(X, Y, \text{ or } X_s, s \in [0, t]\). It is important to note that the \(\sigma\)-algebras on infinite \(\Omega\)'s do not necessarily carry the atomic structure, i.e. they are not generated by partitions. That is why we do not have a simple way of describing \(\sigma\)-algebras anymore - listing the atoms will just not work.

As an example, consider an asset price modeled by a stochastic process \((X_t)_{t \in [0,\infty)}\) - e.g. a Brownian motion. By the time \(t\) we have observed the values of the random variables \(X_s\) for \(s \in [0, t]\), so that our information dynamics can be described by the filtration \((\mathcal{F}_t)_{t \in [0, \infty)}\).
given by \( \mathcal{F}_t = \sigma((X_s)_{s \in [0,t]}). \) This filtration is called the filtration generated by the process \((X_t)_{t \in [0,\infty]}\) and is usually denoted by \((\mathcal{F}_t^X)_{t \in [0,\infty]}\). In our discrete time example in the previous subsection, the (non-insider’s) filtration is the one generated by the stochastic process \((S_n)_{n \in \{0,1,2\}}\).

Example 1.7.10. Let \( B_t \) be a Brownian motion, and let \( \mathcal{F}_t^B \) be the filtration generated by it. The\( \sigma \)-algebra \( \mathcal{F}_t^B \) will contain information about \( B_1, B_3, B_3 + B_4, -\sin(B_4), 1_{\{B_2 \geq -7\}} \), but not about \( B_8 \), or \( B_9 - B_5 \), because we do not know their values if our information is \( \mathcal{F}_t^B \). We can, thus, say that the random variables \( B_1, B_3, B_3 + B_4, -\sin(B_4), 1_{\{B_2 \geq -7\}} \) are measurable with respect to \( \mathcal{F}_t^B \), but \( B_8 \) and \( B_9 - B_5 \) are not.

The notion of conditional expectation can be defined on general \( \Omega \)'s with respect to general \( \sigma \)-algebras:

**Proposition 1.7.11.** Let \( \Omega \) be a state-space, \( \mathcal{F} \) a \( \sigma \)-algebra on it. For any \( X \) such that \( \mathbb{E}[X] \) exists, we can define the random variable \( \mathbb{E}[X|\mathcal{F}] \), such that the properties (CE1)-(CE7) carry over from the discrete case.

**Remark 1.7.1.** The proviso that \( \mathbb{E}[X] \) exists is of technical nature, and we will always pretend that it is fulfilled. When (and if) it happens that a random variable admits no expectation, you will be warned explicitly.

Calculating the conditional expectation in continuous time is, in general, considerably harder than in discrete time because there is no formula or algorithm analogous to the one we had in discrete time. There are, however, two techniques that will cover all the examples in this course:

- **Use properties (CE1)-(CE7)** You will often be told about certain relationships (independence, measurability, . . .) between random variables and a \( \sigma \)-algebra in the statement of the problem. The you can often reach the answer by a clever manipulation of the expressions using properties (CE1)-(CE7).

- **Use conditional densities** In the special case when the \( \sigma \)-algebra \( \mathcal{F} \) you are conditioning on is generated by a random vector \((X_1, X_2, \ldots, X_n)\) and the joint density of the random vector \((X, X_1, X_2, \ldots, X_n)\) is known you can calculate the conditional expectation

  \[
  \mathbb{E}[g(X)|\mathcal{F}] = \mathbb{E}[g(X)|\sigma(X_1, X_2, \ldots, X_n)]
  \]

by following this recipe:

  - compute the conditional density \( f(x|X_1, X_2, \ldots, X_n) \) (\( x | X_1 = \xi_1, X_2 = \xi_2, \ldots, X_n = \xi_n \)), and use it to arrive at the function

  \[
  h(\xi_1, \xi_2, \ldots, \xi_n) = \mathbb{E}[g(X)|X_1 = \xi_1, X_2 = \xi_2, \ldots, X_n = \xi_n] = \int_{-\infty}^{\infty} g(x) f(x|X_1, X_2, \ldots, X_n)(x|X_1 = \xi_1, X_2 = \xi_2, \ldots, X_n = \xi_n) \, dx.
  \]
1.7. CONDITIONING

CHAPTER 1. BROWNIAN MOTION

- plug $X_1, X_2, \ldots, X_n$ for $\xi_1, \xi_2, \ldots, \xi_n$ in $h(\xi_1, \xi_2, \ldots, \xi_n)$, and you are done. In other words,

$\mathbb{E}[g(X)|\mathcal{F}] = \mathbb{E}[g(X)|\sigma(X_1, X_2, \ldots, X_n)] = h(X_1, X_2, \ldots, X_n)$.

**Example 1.7.12.** Let $(B_t)_{t \in [0, \infty)}$ be a Brownian motion, and let $(\mathcal{F}_t^B)_{t \in [0, \infty)}$ be the filtration it generates. For $s < t$, let us try to compute $\mathbb{E}[B_t|\mathcal{F}_s^B]$ using the rules (CE1)-(CE7). First of all, by the definition of the Brownian motion, the increment $B_t - B_s$ is independent of all that happened before (or at) time $s$, so the random variable $B_t - B_s$ must be independent of the $\sigma$-algebra $\mathcal{F}_s^B$ which contains exactly the past before and up to $s$. Using (CE7) we have $\mathbb{E}[B_t - B_s|\mathcal{F}_s^B] = \mathbb{E}[B_t - B_s] = 0$. The linearity (CE5) of conditional expectation will imply now that $\mathbb{E}[B_t|\mathcal{F}_s^B] = \mathbb{E}[B_s|\mathcal{F}_s^B]$. We know $B_s$ when we know $\mathcal{F}_s^B$, so $\mathbb{E}[B_s|\mathcal{F}_s^B] = B_s$, (CE3). Therefore, $\mathbb{E}[B_t|\mathcal{F}_s^B] = B_s$. Can you compute this conditional expectation using the other method with conditional densities? How about $\mathbb{E}[B_t|\sigma(B_s)]$?

**Example 1.7.13.** Let’s try to compute a slightly more complicated conditional quantity. How much is $\mathbb{P}[B_t > a|\sigma(B_s)]$? Here, we define $\mathbb{P}[A|\sigma(B_s)] \triangleq \mathbb{E}[1_A|\sigma(B_s)]$. We have all the necessary prerequisites for using the conditional-density method:

- we are conditioning with respect to a $\sigma$-algebra generated by a random variable (a random vector of length $n = 1$)
- the joint density of $(B_t, B_s)$ is known
- the required conditional expectation can be expressed as

$$\mathbb{P}[B_t \geq x|\sigma(B_s)] = \mathbb{E}[1_{\{B_t \geq x\}}|\sigma(B_s)] = \mathbb{E}[g(B_t)|\sigma(B_s)], \text{ where } g(x) = \begin{cases} 1, & x \geq a \\ 0, & x < a. \end{cases}$$

We have learned at the beginning of this section that the conditional density of $B_t$ given $B_s = \xi$ is the density of a univariate normal distribution with mean $\xi$ and variance $t - s$, i.e.

$$f_{B_t|B_s}(x|B_s = \xi) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(x - \xi)^2}{2(t-s)}\right),$$

and so $\mathbb{P}[B_t \geq a|\sigma(B_s)] = h(B_s), \text{ where}$

$$h(\xi) = \int_{-\infty}^{\infty} g(x) \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(x - \xi)^2}{2(t-s)}\right) dx = \int_{a}^{\infty} \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(x - \xi)^2}{2(t-s)}\right) dx = 1 - \Phi\left(\frac{a - \xi}{\sqrt{t-s}}\right),$$

i.e.

$$\mathbb{P}[B_t \geq a|\sigma(B_s)] = 1 - \Phi\left(\frac{a - B_s}{\sqrt{t-s}}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard unit normal - $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\xi^2/2) d\xi$. 

Last Updated: Spring 2005
1.7.9 Martingales

Now that we have the conditional expectation at our disposal, we can define the central concept of these lecture notes:

**Definition 1.7.14.** Let \((X_t)_{t \in [0, \infty)}\) stochastic process, and let \((\mathcal{F}_t)_{t \in [0, \infty)}\) be a filtration. \(X\) is called a (continuous-time) \(\mathcal{F}_t\)-martingale if

- \(X_t\) is \(\mathcal{F}_t\)-measurable for each \(t\), and
- \(\mathbb{E}[X_t | \mathcal{F}_s] = X_s\), for all \(s, t \in [0, \infty)\), such that \(s < t\).

Before we give intuition behind the concept of a martingale, here are some remarks:

1. Note that the concept of a martingale makes sense only if you specify the filtration \(\mathcal{F}\). When the filtration is not specified explicitly, we will always assume that the filtration is generated by the process itself, i.e. \(\mathcal{F}_t = \mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)\).

2. In general, if for a process \((X_t)_{t \in [0, \infty)}\), the random variable \(X_t\) is measurable with respect to \(\mathcal{F}_t\), for each \(t \geq 0\), we say that \(X_t\) is **adapted** to the filtration \((\mathcal{F}_t)_{t \in [0, \infty)}\). Of course, analogous definitions apply to the finite horizon case \((X_t)_{t \in [0, T]}\), as well as the discrete case \((X_n)_{n \in \mathbb{N}}\).

3. You can define a discrete time martingale \((X_n)_{n \in \mathbb{N}}\) with respect to a filtration \((\mathcal{F}_n)_{n \in \mathbb{N}}\) by making the (formally) identical requirements as above, replacing \(s\) and \(t\), by \(n, m \in \mathbb{N}\). It is, however, a consequence of the tower property that it is enough to require

\[
\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n, \text{ for all } n \in \mathbb{N},
\]

with each \(X_n\) being \(\mathcal{F}_n\)-measurable.

4. the process that satisfies all the requirements of the definition 1.7.14, except that we have

\[
\mathbb{E}[X_t | \mathcal{F}_s] \geq X_s \text{ instead of } \mathbb{E}[X_t | \mathcal{F}_s] = X_s, \text{ for } s < t,
\]

is called a **submartingale**. Similarly, if we require \(\mathbb{E}[X_t | \mathcal{F}_s] \leq X_s\), for all \(s < t\), the process \(X\) is called a **supermartingale**. Note the apparent lack of logic in these definitions:

- **submartingales** increase on average, and **supermartingales** decrease on average.

5. Martingales have the nice property that \(\mathbb{E}[X_t] = \mathbb{E}[X_0]\) for each \(t\). In other words, the expectation of a martingale is a constant function. It is a direct consequence of the tower property:

\[
\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{F}_0]] = \mathbb{E}[X_0].
\]
The fundamental example of a martingale in continuous time is Brownian motion. To give an example of a martingale in discrete time, let \( X_n \) be a sequence of independent and identically distributed random variables such that the expectation \( \mathbb{E}[X_n] = n \) and \( \mathbb{E}[X_n^2] = n \) are the same, because one can recover the trajectory of the Brownian motion up to time \( t \), from the trajectory of \( X^\mu \) by simply subtracting \( \mu s \) from each \( X^\mu_s \). Similarly, if one knows the trajectory of \( B \) one can get the trajectory of \( X^\mu \) by adding \( \mu s \). Therefore, \( \mathcal{F}_t^B = \mathcal{F}_t^{X^\mu} \) for each \( t \), and we can safely say that \( X^\mu \) is an \( \mathcal{F}^{X^\mu} \) (super-, sub-) martingale.

**Example 1.7.16.** To give an example of a martingale in discrete time, let \( Y_1, Y_2, \ldots \) be a sequence of independent and identically distributed random variables such that the expectation \( \mu = \mathbb{E}[Y_1] \) exists. Define

\[
X_0 = 0, \quad X_1 = Y_1, \quad X_2 = Y_1 + Y_2, \quad \ldots, \quad X_n = Y_1 + Y_2 + \cdots + Y_n,
\]

so that \( (X_n)_{n \in \mathbb{N}} \) is a general random walk whose steps are distributed as \( Y_1 \). Let \( \mathcal{F}^X \) be a filtration generated by the process \( (X_n)_{n \in \mathbb{N}} \). Then

\[
\mathbb{E}[X_{n+1} | \mathcal{F}^X_n] = \mathbb{E}[Y_1 + Y_2 + \cdots + Y_{n+1} | \mathcal{F}^X_n] = \mathbb{E}[X_n + Y_{n+1} | \mathcal{F}^X_n] \overset{(CE3)}{=} X_n + \mathbb{E}[Y_{n+1} | \mathcal{F}^X_n]
\]

and we conclude that \( (X_n)_{n \in \mathbb{N}} \) is a supermartingale if \( \mu < 0 \), a martingale if \( \mu = 0 \), and a submartingale if \( \mu > 0 \). In the equation (1.7.3) we could use (CE3) because \( X_n \) is \( \mathcal{F}_n \) adapted, (CE7) because \( Y_{n+1} \) is independent of \( \mathcal{F}_n \), and \( \mathbb{E}[Y_{n+1}] = \mathbb{E}[Y_1] = \mu \), because...
all \( Y_k \) are identically distributed. What can you say about the case where \( X_0 = 1 \), and 
\[
X_n = \prod_{k=1}^{n} Y_k = Y_1 Y_2 \ldots Y_n.
\]

Here is a slightly more convoluted example in continuous time:

**Example 1.7.17.** Let \((B_t)_{t \in [0, \infty)}\) be a Brownian motion, and let \((\mathcal{F}^B)_{t \in [0, \infty)}\) be the filtration generated by \(B\). Let us prove first that \(X_t = e^{\frac{t}{2}} \sin(B_t)\) is a martingale with respect to \((\mathcal{F}^B)_{t \in [0, \infty)}\). As always, we take \(s < t\), and use the facts that \(\sin(B_s)\) and \(\cos(B_s)\) are measurable with respect to \(\mathcal{F}^B_s\), and \(\sin(B_t - B_s)\) and \(\cos(B_t - B_s)\) are independent of \(\mathcal{F}^B_s\):

\[
\mathbb{E}[X_t | \mathcal{F}^B_s] = \mathbb{E}[e^{\frac{t}{2}} \sin(B_t) | \mathcal{F}^B_s] = e^{\frac{t}{2}} \mathbb{E}[\sin(B_s + (B_t - B_s)) | \mathcal{F}^B_s] \\
= e^{\frac{t}{2}} \mathbb{E}[\sin(B_s) \cos(B_t - B_s) + \cos(B_s) \sin(B_t - B_s) | \mathcal{F}^B_s] = \\
= e^{\frac{t}{2}} \left( \sin(B_s) \mathbb{E}[\cos(B_t - B_s)] + \cos(B_s) \mathbb{E}[\sin(B_t - B_s)] \right) = \\
= \sin(B_s) \left( e^{\frac{t}{2}} \mathbb{E}[\cos(B_t - B_s)] \right) = X_s \left( e^{\frac{t-s}{2}} \mathbb{E}[\cos(B_t - B_s)] \right)
\]

We know that \(\sin(B_t - B_s)\) is a symmetric random variable (\(\sin\) is a symmetric function, and \(B_t - B_s \sim N(0, t-s)\)) so \(\mathbb{E}[\sin(B_t - B_s)] = 0\). We are, therefore, left with the task of proving that \(\mathbb{E}[\cos(B_t - B_s)] = e^{-\frac{t-s}{2}}\). We set \(u = t-s\), and remember that \(B_t - B_s \sim N(0, u)\). Therefore,

\[
\mathbb{E}[\cos(B_t - B_s)] = \int_{-\infty}^{\infty} \cos(\xi) \frac{1}{\sqrt{2\pi u}} \exp\left(-\frac{\xi^2}{2u}\right) d\xi = \text{(Maple)} = e^{-\frac{u}{2}} = e^{-\frac{t-s}{2}},
\]

just as required. We can therefore conclude that \(\mathbb{E}[X_t | \mathcal{F}^B_s] = X_s\), for all \(s < t\), proving that \((X_t)_{t \in [0, \infty)}\) is an \((\mathcal{F}^B)_{t \in [0, \infty)}\)-martingale. In order to prove that \((X_t)_{t \in [0, \infty)}\) is an \((\mathcal{F}^X)_{t \in [0, \infty)}\)-martingale, we use the tower property because \(\mathcal{F}^X_t \subseteq \mathcal{F}^B_t\) (why?) to obtain

\[
\mathbb{E}[X_t | \mathcal{F}^X_s] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{F}^B_s] | \mathcal{F}^X_s] = \mathbb{E}[X_s | \mathcal{F}^X_s] = X_s.
\]

**Example 1.7.18.** Finally, here is an example of a process that is neither a martingale, nor a super- or a submartingale, with respect to its natural filtration. Let \(B_t\) be a Brownian motion, and let \(X_t = B_t - tB_1\) be a Brownian Bridge. We pick \(t = 1, s = 1/2\) are write

\[
\mathbb{E}[X_1 | \mathcal{F}^X_{1/2}] = \mathbb{E}[0 | \mathcal{F}^X_{1/2}] = 0.
\]

It is now obvious that for the (normally distributed!) random variable \(X_{1/2}\) we have neither \(X_{1/2} = 0\), \(X_{1/2} \geq 0\), nor \(X_{1/2} \leq 0\).

### 1.7.10 Martingales as Unwinnable Gambles

Let \((X_n)_{n \in \mathbb{N}}\) be a discrete-time stochastic process, and let \((\mathcal{F}_n)_{n \in \mathbb{N}}\) be a filtration. Think of \(X_n\) as a price of a stock on the day \(n\), and think of \(\mathcal{F}_n\) as the publicly available information.
on the day \(n\). Of course, the price of the stock at the day \(n\) is a part of that information so
that \(X_n\) if \(\mathcal{F}_n\)-adapted.

Suppose that we have another process \((H_n)_{n \in \mathbb{N}}\) which stands for the number of stocks in
our portfolio in the evening of day \(n\) (after we have observed \(X_n\), and made the trade for the day). \(H\) will be called the trading strategy for the obvious reasons. We start with \(x\) dollars, credit is available at 0 interest rate, and the stock can be freely shorted, so that our

\[
Y^H = x, \quad Y_2^H = H_1(X_2 - X_1), \quad Y_3^H = Y_2 + H_2(X_3 - X_2), \ldots \quad Y_n^H = \sum_{k=2}^n H_{k-1} \Delta X_k,
\]

with \(\Delta X_k = X_k - X_{k-1}\). The process \(Y^H\) is sometimes denoted by \(Y^H = H \cdot X\). Here is
a theorem which says that you cannot win by betting on a martingale. A more surprising
statement is also true: if \(X\) is a process such that no matter how you try, you cannot win by
betting on it - then, \(X\) must be a martingale.

**Theorem 1.7.19.** \(X_n\) is a martingale if and only if \(E[Y^H_n] = x\) for each \(H, n\). In that case \(Y^H\) is an \(\mathcal{F}_n\)-martingale for each \(H\).

**Proof.**

- **\(X\) martingale \(\implies E[Y^H_n] = x\) for each \(H, n\)**

  Let us compute the conditional expectation \(E[Y^H_{n+1}|\mathcal{F}_n]\). Note first that \(Y^H_{n+1} = Y^H_n + H_n(X_{n+1} - X_n)\), and that \(Y^H_n\) and \(H_n\) are \(\mathcal{F}_n\)-measurable, so that

  \[
  E[Y^H_{n+1}|\mathcal{F}_n] = E[Y^H_n + H_n(X_{n+1} - X_n)|\mathcal{F}_n] = Y^H_n + E[H_n(X_{n+1} - X_n)|\mathcal{F}_n] = Y^H_n + E[H_n|\mathcal{F}_n] - X_n = Y^H_n,
  \]

  so that \(Y^H\) is a martingale, and in particular \(E[Y^H_n] = E[Y^H_1] = x\).

- **\(E[Y^H_n] = x\) for each \(H, n \implies X\) martingale.** (You can skip this part of the proof if you find it too technical and hard!)

  Since we know that \(E[Y^H_n] = x\), for all \(n\), and all \(H\), we will pick a special trading
strategy and use it to prove our claim. So, pick \(n \in \mathbb{N}\), pick any set \(A \in \mathcal{F}_n\), and let \(H\) be the following gambling strategy:

  - do nothing until day \(n\),

\[\text{I am ducking some issues here - existence of expectations, for example. We will therefore cheat a little,}
\]

and assume that all expectations exist. This will not be entirely correct from the mathematical point of view,
but it will save us from ugly technicalities without skipping the main ideas.
– buy 1 share of stock in the evening of the day $n$, if the event $A$ happened (you know whether $A$ happened or not on day $n$, because $A \in \mathcal{F}_n$). If $A$ did not happen, do nothing.

– wait until tomorrow, and liquidate your position,

– retire from the stock market and take up fishing.

Mathematically, $H$ can be expressed as

$$H_k(\omega) = \begin{cases} 1, & \omega \in A \text{ and } k = n \\ 0, & \text{otherwise,} \end{cases}$$

so that for $k \geq n + 1$ we have $Y_k^H = x + (X_{n+1} - X_n)1_A$. By the assumption of the theorem, we conclude that

$$\mathbb{E}[X_{n+1}1_A] = \mathbb{E}[X_n1_A], \text{ for all } n, \text{ and all } A \in \mathcal{F}_n. \quad (1.7.3)$$

If we succeed in proving that (1.7.3) implies that $(X_n)_{n \in \mathbb{N}}$ is a martingale, we are done. We argue by contradiction: suppose (1.7.3) holds, but $X$ is not a martingale. That means that there exists some $n \in \mathbb{N}$ such that $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \neq X_n$. In other words, the random variables $X_n$ and $Y = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$ are not identical. It follows that at least one of the sets

$$B = \{\omega \in \Omega : X_n(\omega) > Y(\omega)\}, \quad C = \{\omega \in \Omega : X_n(\omega) < Y(\omega)\},$$

is nonempty (has positive probability, to be more precise). Without loss of generality, we assume that $\mathbb{P}[B] > 0$. Since both $X_n$ and $Y$ are $\mathcal{F}_n$-measurable, so is the set $B$ and thus $1_B$ is $\mathcal{F}_n$-measurable. Since $Y$ is strictly smaller than $X_n$ on the set $B$, we have

$$\mathbb{E}[X_n1_B] < \mathbb{E}[Y1_B] = \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]1_B] = \mathbb{E}[\mathbb{E}[X_{n+1}1_B|\mathcal{F}_n]] = \mathbb{E}[X_{n+1}1_B],$$

which is in contradiction with (1.7.3) (just take $A = B$). We are done.
Exercises

Exercise 1.7.20. Let \( (B_t)_{t \in [0, \infty)} \) be a Brownian motion. For \( 0 < s < t \), what is the conditional density of \( B_s \), given \( B_t = \xi \). How about the other way around - the conditional density of \( B_t \) given \( B_s = \xi \)?

Exercise 1.7.21. Let \( Y_1 \) and \( Y_2 \) be two independent random variables with exponential distributions and parameters \( \lambda_1 = \lambda_2 = \lambda \). Find the conditional density of \( X_1 = Y_1 \) given \( X_2 = \xi \), for \( X_2 = Y_1 + Y_2 \).

Exercise 1.7.22. In the Cox-Ross-Rubinstein model of stock prices there are \( n \) time periods. The stock price at time \( t = 0 \) is given by a fixed constant \( S_0 = s \). The price at time \( t = k + 1 \) is obtained from the price at time \( t = k \) by multiplying it by a random variable \( X_k \) with the distribution

\[
X_k \sim \left( \frac{1 + a}{p}, \frac{1 - b}{1 - p} \right), \text{ where } a > 0, \ 0 < b < 1, \text{ and } 0 < p < 1.
\]

The return \( X_k \) are assumed to be independent of the past stock prices \( S_0, S_1, \ldots, S_k \). In this problem we will assume that \( n = 3 \), so that the relevant random variables are \( S_0, S_1, S_2 \) and \( S_3 \).

(a) Sketch a tree representation of the above stock-price model, and describe the state-space and the probability measure on it by computing \( \mathbb{P}[\omega] \) for each \( \omega \in \Omega \). Also, write down the values of \( S_0(\omega), S_1(\omega), S_2(\omega), S_3(\omega), X_0(\omega), X_1(\omega) \) and \( X_2(\omega) \) for each \( \omega \in \Omega \).

(Note: you should be able to do it on an \( \Omega \) consisting of only 8 elements).

(b) Find the atoms for each of the following \( \sigma \)-algebras \( \mathcal{F}_k^S \), \( k = 0, 1, 2, 3 \), where \( \mathcal{F}_k^S = \sigma(S_0, S_1, \ldots, S_k) \).

(c) Compute (1) \( \mathbb{E}[S_2 | \mathcal{F}_3^S] \), (2) \( \mathbb{E}[S_2 | \sigma(S_3)] \), (3) \( \mathbb{E} \left[ \frac{S_2}{S_1} | \mathcal{F}_1^S \right] \), (4) \( \mathbb{E}[S_1 + S_2 | \sigma(S_3)] \).

(d) Describe (in terms of the atoms of its \( \sigma \)-algebras) the filtration of an insider who knows whether \( S_3 > S_0 \) or not, on top of the public information filtration \( \mathcal{F}^S \).

Exercise 1.7.23. Let \( \Omega \) be a finite probability space, \( \mathcal{F} \) a \( \sigma \)-algebra on \( \Omega \), and \( X \) a random variable. Show that for all \( A \in \mathcal{F} \),

\[
\mathbb{E} \left[ \mathbb{E}[X | \mathcal{F}] 1_A \right] = \mathbb{E}[X 1_A].
\]

Note: this property is often used as the definition of the conditional expectation in the case of an infinite \( \Omega \).

(Hint: use the fact that each \( A \in \mathcal{F} \) can be written as a finite union of atoms of \( \mathcal{F} \), and exploit the measurability properties of the conditional expectation \( \mathbb{E}[X | \mathcal{F}] \).)
Exercise 1.7.24. Let \((S_n)_{n \in \mathbb{N}}\) be a random walk, i.e.

\[
S_0 = 0, \quad S_n = X_1 + X_2 + \cdots + X_n,
\]

where the steps \(X_i\)'s are independent random variables with distribution

\[
X_i \sim \left( \begin{array}{c} -1 \\ 1/2 \\ 1/2 \end{array} \right).
\]

Compute \(E[S_n|\sigma(X_1)]\) and \(E[X_1|\sigma(S_n)]\). (Hint: Think and use symmetry!)

Exercise 1.7.25. Let \((X,Y)\) be a random vector with the density function given by

\[
f_{(X,Y)}(x,y) = \begin{cases} kxy, & 1 \leq x \leq 2 \text{ and } 0 \leq y \leq x, \\
0, & \text{otherwise.} \end{cases}
\]

(a) Find the constant \(k > 0\), so that \(f\) above defines a probability density function.

(b) Compute the (marginal) densities of \(X\) and \(Y\).

(c) Find the conditional expectation \(E[U|\sigma(X)]\), where \(U = X + Y\).

Exercise 1.7.26. Let \((B_t)_{t \in [0,\infty)}\) be a Brownian motion, and let \(\mathcal{F}_B^t = \sigma((B_s)_{s \in [0,t]})\) be the filtration generated by it.

(a) (4pts) Which of the following are measurable with respect to \(\mathcal{F}_3^B\):

\[
(1) B_3, \quad (2) -B_3, \quad (3) B_2+B_1, \quad (4) B_2+B_{1.5}, \quad (5) B_4-B_2, \quad (6) \int_0^3 B_s \, ds?
\]

(b) (4pts) Which of the following are \(\mathcal{F}_B^t\)-martingales

\[
(1) -B_t, \quad (2) B_t - t, \quad (3) B_t^2, \quad (4) B_t^2 - t, \quad (5) \exp(B_t - \frac{t}{2})?\]

Exercise 1.7.27. The conditional moment-generating function \(M_{X|\mathcal{F}}(\lambda)\) of an \(n\)-dimensional random vector \(X = (X_1, X_2, \ldots, X_n)\) given the \(\sigma\)-algebra \(\mathcal{F}\) is defined by

\[
M_{X|\mathcal{F}}(\lambda) = \mathbb{E}[\exp(-\lambda_1 X_1 - \lambda_2 X_2 - \cdots - \lambda_n X_n)|\mathcal{F}], \text{ for } \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n.
\]

(a) Let \(B_t, \ t \in [0,\infty)\) be a Brownian motion. Calculate \(M_{B_t|\mathcal{F}_s^t}(\lambda)\) for \(t > s \geq 0\) and \(\lambda \in \mathbb{R}\). Here \((\mathcal{F}_t^B)_{t \in [0,\infty)}\) denotes the filtration generated by the Brownian motion \(B\).
1.7. CONDITIONING

CHAPTER 1. BROWNIAN MOTION

(b) Two random variables \(X_1\) and \(X_2\) are said to be **conditionally independent** given the \(\sigma\)-algebra \(\mathcal{F}\), if the random vector \(X = (X_1, X_2)\) satisfies

\[
\mathcal{M}_{X|\mathcal{F}}(\lambda) = \mathcal{M}_{X_1|\mathcal{F}}(\lambda_1) \mathcal{M}_{X_2|\mathcal{F}}(\lambda_2), \quad \text{for all } \lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2.
\]

Show that the random variables \(B_{t_1}\) and \(B_{t_2}\) are conditionally independent given the \(\sigma\)-algebra \(\sigma(B_{t_3})\) for \(0 \leq t_1 < t_3 < t_2\) (note the ordering of indices!!!) In other words, for Brownian motion, past is independent of the future, given the present.

**Exercises**

**Exercise 1.7.28.** Let \((B_t)_{t \in [0,\infty)}\) be a Brownian motion. For \(0 < s < t\), what is the conditional density of \(B_s\), given \(B_t = \xi\). How about the other way around - the conditional density of \(B_t\) given \(B_s = \xi\)?

**Exercise 1.7.29.** Let \(Y_1\) and \(Y_2\) be two independent random variables with exponential distributions and parameters \(\lambda_1 = \lambda_2 = \lambda\). Find the conditional density of \(X_1 = Y_1\) given \(X_2 = \xi\), for \(X_2 = Y_1 + Y_2\).

**Exercise 1.7.30.** In the Cox-Ross-Rubinstein model of stock prices there are \(n\) time periods. The stock price at time \(t = 0\) is given by a fixed constant \(S_0 = s\). The price at time \(t = k + 1\) is obtained from the price at time \(t = k\) by multiplying it by a random variable \(X_k\) with the distribution

\[
X_k \sim \begin{pmatrix} 1 + a & 1 - b \\ p & 1 - p \end{pmatrix}, \quad \text{where } a > 0, \ 0 < b < 1, \ \text{and} \ 0 < p < 1.
\]

The *return* \(X_k\) are assumed to be independent of the past stock prices \(S_0, S_1, \ldots, S_k\). In this problem we will assume that \(n = 3\), so that the relevant random variables are \(S_0, S_1, S_2\) and \(S_3\).

(a) Sketch a tree representation of the above stock-price model, and describe the state-space and the probability measure on it by computing \(P[\omega]\) for each \(\omega \in \Omega\). Also, write down the values of \(S_0(\omega), S_1(\omega), S_2(\omega), S_3(\omega), X_0(\omega), X_1(\omega)\) and \(X_2(\omega)\) for each \(\omega \in \Omega\). (Note: you should be able to do it on an \(\Omega\) consisting of only 8 elements).

(b) Find the atoms for each of the following \(\sigma\)-algebras \(\mathcal{F}_k^S\), \(k = 0, 1, 2, 3\), where \(\mathcal{F}_k^S = \sigma(S_0, S_1, \ldots, S_k)\).

(c) Compute (1) \(E[S_2|\mathcal{F}_3^S]\), (2) \(E[S_2|\sigma(S_3)]\), (3) \(E[S_2|\mathcal{F}_1^S]\), (4) \(E[S_1 + S_2|\sigma(S_3)]\).

(d) Describe (in terms of the atoms of its \(\sigma\)-algebras) the filtration of an insider who knows whether \(S_3 > S_0\) or not, on top of the public information filtration \(\mathcal{F}_3^S\).
Exercise 1.7.31. Let \( \Omega \) be a finite probability space, \( \mathcal{F} \) a \( \sigma \)-algebra on \( \Omega \), and \( X \) a random variable. Show that for all \( A \in \mathcal{F} \),

\[
E[E[X|\mathcal{F}]1_A] = E[X1_A].
\]

Note: this property is often used as the definition of the conditional expectation in the case of an infinite \( \Omega \).

(Hint: use the fact that each \( A \in \mathcal{F} \) can be written as a finite union of atoms of \( \mathcal{F} \), and exploit the measurability properties of the conditional expectation \( E[X|\mathcal{F}] \).)

Exercise 1.7.32. Let \( (S_n)_{n \in \mathbb{N}} \) be a random walk, i.e.

\[
\begin{align*}
(1) & \quad S_0 = 0, \\
(2) & \quad S_n = X_1 + X_2 + \cdots + X_n,
\end{align*}
\]

where the steps \( X_i \)'s are independent random variables with distribution

\[
X_i \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}.
\]

Compute \( E[S_n|\sigma(X_1)] \) and \( E[X_1|\sigma(S_n)] \). (Hint: Think and use symmetry!)

Exercise 1.7.33. Let \( (X,Y) \) be a random vector with the density function given by

\[
f_{(X,Y)}(x,y) = \begin{cases} 
  kxy, & 1 \leq x \leq 2 \text{ and } 0 \leq y \leq x, \\
  0, & \text{otherwise}.
\end{cases}
\]

(a) Find the constant \( k > 0 \), so that \( f \) above defines a probability density function.

(b) Compute the (marginal) densities of \( X \) and \( Y \)

(c) Find the conditional expectation \( E[U|\sigma(X)] \), where \( U = X + Y \).

Exercise 1.7.34. Let \( (B_t)_{t \in [0,\infty)} \) be a Brownian motion, and let \( \mathcal{F}^B_t = \sigma((B_s)_{s \in [0,t]}) \) be the filtration generated by it.

(a) (4pts) Which of the following are measurable with respect to \( \mathcal{F}^B_3 \):

\[
(1) B_3, \quad (2) -B_3, \quad (3) B_2+B_1, \quad (4) B_2+B_{1.5}, \quad (5) B_4-B_2, \quad (6) \int_0^3 B_s ds ?
\]

(b) (4pts) Which of the following are \( \mathcal{F}^B \)-martingales

\[
(1) -B_t, \quad (2) B_t - t, \quad (3) B^2_t, \quad (4) B^2_t - t, \quad (5) \exp(B_t - \frac{t}{2}) ?
\]
Exercise 1.7.35. The conditional moment-generating function $M_{X|F}(\lambda)$ of an $n$-dimensional random vector $X = (X_1, X_2, \ldots, X_n)$ given the $\sigma$-algebra $F$ is defined by

$$M_{X|F}(\lambda) = \mathbb{E}[\exp(-\lambda_1 X_1 - \lambda_2 X_2 - \cdots - \lambda_n X_n)|F],$$
for $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$.

(a) Let $B_t$, $t \in [0, \infty)$ be a Brownian motion. Calculate $M_{B_t|F_s}(\lambda)$ for $t > s \geq 0$ and $\lambda \in \mathbb{R}$. Here $(F^B_t)_{t \in [0, \infty)}$ denotes the filtration generated by the Brownian motion $B$.

(b) Two random variables $X_1$ and $X_2$ are said to be conditionally independent given the $\sigma$-algebra $F$, if the random vector $X = (X_1, X_2)$ satisfies

$$M_{X|F}(\lambda) = M_{X_1|F}(\lambda_1)M_{X_2|F}(\lambda_2),$$
for all $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$.

Show that the random variables $B_{t_1}$ and $B_{t_2}$ are conditionally independent given the $\sigma$-algebra $\sigma(B_{t_3})$ for $0 \leq t_1 < t_3 < t_2$ (note the ordering of indices!!!) In other words, for Brownian motion, past is independent of the future, given the present.
Solutions to Exercises in Section 1.7

Solution to Exercise 1.7.28:
The joint density of \((B_s, B_t)\) is bivariate normal with zero mean vector, variances \(\text{Var}[B_s] = s\), \(\text{Var}[B_t] = t\) and correlation
\[
\rho = \frac{\text{Cov}[B_s, B_t]}{\sqrt{\text{Var}[B_s] \text{Var}[B_t]}} = \frac{s}{\sqrt{st}} = \sqrt{\frac{s}{t}}.
\]
Using (1.7.1) we conclude that the conditional density of \(B_s\), given \(B_t = \xi\) is the density of a normal random variable with
\[
\mu_{B_s|B_t=\xi} = 0 + \sqrt{\frac{s}{t}} \sqrt{\frac{s}{t}} (\xi - 0) = \frac{s}{t} \xi,
\]
and
\[
\sigma_{B_s|B_t=\xi} = \sqrt{s} \sqrt{1 - \frac{s}{t}} = \sqrt{\frac{s(t - s)}{t}}.
\]
As for "the other way around" case, we can use the same reasoning to conclude that the conditional density of \(B_t\) given \(B_s = \xi\) is normal with mean \(\xi\) and standard deviation \(\sqrt{t - s}\).

Solution to Exercise 1.7.29: Let us first determine the joint density of the random vector \((X_1, X_2) = (Y_1, Y_1 + Y_2)\), by computing its distribution function and then differentiating. Using the fact that an exponentially distributed random variable with parameter \(\lambda\) admits a density of the form \(\lambda \exp(-\lambda x), x \leq 0\), we have \((0 \leq x_1 \leq x_2)\)
\[
F_{X_1,X_2}(x_1, x_2) = \mathbb{P}[Y_1 \leq x_1, Y_1 + Y_2 \leq x_2] = \int_{0}^{x_1} \int_{0}^{x_2-y_2} \lambda_1 \exp(-\lambda_1 y_1) \lambda_2 \exp(-\lambda_2 y_2) dy_2 dy_1
\]
\[
= \int_{0}^{x_1} \lambda_1 \exp(-\lambda_1 x_1)(1 - \exp(-\lambda_2 (x_2 - y_1))) dy_1,
\]
and we can differentiate with respect to \(x_1\) and \(x_2\) to obtain
\[
f_{X_1,X_2}(x_1, x_2) = \begin{cases} \lambda^2 \exp(-\lambda x_2), & 0 \leq x_1 \leq x_2, \\ 0, & \text{otherwise} \end{cases}
\]
1.7. CONDITIONING

We use now the formula for conditional probability:

\[
f_{X_1 | X_2}(x_1, x_2 = \xi) = \frac{f_{X_1, X_2}(x_1, \xi)}{f_{X_2}(\xi)} = \begin{cases} \frac{\lambda^2 \exp(-\lambda \xi)}{f_{X_2}(\xi)}, & 0 \leq x_1 \leq \xi, \\ 0, & \text{otherwise} \end{cases}
\]

Knowing that, as a function of \(x_1\), \(f_{X_1 | X_2}(x_1, x_2 = \xi)\) is a genuine density function, we do not need to compute \(f_{X_2}\) explicitly - we just need to insure that (note that integration is with respect to \(x_1\))

\[
1 = \int_0^\xi \frac{\lambda^2 \exp(\lambda \xi)}{f_{X_2}(\xi)} \, dx_1,
\]

and so \(f_{X_2}(\xi) = \lambda^2 \xi \exp(-\lambda \xi)\) and

\[
f_{X_1 | X_2}(x_1, x_2 = \xi) = \begin{cases} \frac{\xi}{\lambda}, & 0 \leq x_1 \leq \xi, \\ 0, & \text{otherwise} \end{cases},
\]

so we conclude that, conditionally on \(Y_1 + Y_2 = \xi\), \(Y_1\) is distributed uniformly on \([0, \xi]\).

As a bi-product of this calculation we obtained that the density if \(X_2 = Y_1 + Y_2\) is \(f_{X_2}(x) = \lambda^2 x \exp(-\lambda x), \ x \geq 0\). The distribution of \(X_2\) is called the Gamma-distribution with parameters \((\lambda, 2)\).

**Solution to Exercise 1.7.30:**

(a) The figure on the right shows the tree representation of the CRR model, drawn to scale with \(a = 0.2\) and \(b = 0.5\). We can always take \(\Omega\) to be the set of all paths the process can take. In this case we will encode each path by a triplet of letters - e.g. UDD will denote the path whose first movement was up, and the next two down. With this convention we have

\[
\Omega = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\},
\]

and \(P[UUU] = p^3, \ P[UUD] = P[UDU] = P[DUU] = p^2(1 - p), \ P[DDU] = P[DUD] = P[UDD] = p(1 - p)^2, \) and \(P[DDD] = (1 - p)^3\), since the probability of an upward movement is \(p\) independently of the other increments.

The following table shows the required values of the random variables (the row represents the element of \(\Omega\), and the column the random variable):
(b) The $\sigma$-algebra $F_0^S$ is trivial so its only atom is $\Omega$. The $\sigma$-algebra $F_1^S$ can distinguish between those $\omega \in \Omega$ for which $S_1(\omega) = s(1 + a)$ and those for which $S_1(\omega) = s(1 - b)$, so its atoms are $\{UUU, UUD, UDU, UDD\}$ and $\{DUU, DUD, DDU, DDD\}$. An analogous reasoning leads to the atoms $\{UUU, UUD\}$, $\{UDU, UDD\}$, $\{DUU, DUD\}$ and $\{DDU, DDD\}$ of $F_2^S$. Finally, the atoms of $F_3^S$ are all 1-element subsets of $\Omega$.

(c)

(1) Since $S_2$ is measurable with respect to $F_3^S$ (i.e. we know the exact value of $S_2$ when we know $F_3^S$), the property (CE3) implies that $E[S_2|F_3^S] = S_2$.

(2) There are four possible values for the random variable $S_3$ and so the atoms of the $\sigma$-algebra generated by it are $A^0 = \{UUU\}$, $A^1 = \{UUD, UDU, UDD\}$, $A^2 = \{DDD, DDU, DDD\}$, and $A^3 = \{UDD, DDD\}$. By the algorithm for computing discrete conditional expectations from the notes, and using the table above, we have

$$E[S_2|\sigma(S_3)](\omega) = \begin{cases} 
\frac{S_2(UUU)P[UUU]}{P[UUU]} = s(1 + a)^2, & \omega \in A^0 \\
\frac{S_2(UUD)P[UUD]+S_2(UDU)P[UDU]+S_2(DUU)P[DUU]}{P[UUD]+P[UDU]+P[DUU]} = s(1+a)^2+2(1+a)(1-b)/3, & \omega \in A^1 \\
\frac{S_2(UDU)P[UDU]+S_2(DDU)P[DDU]+S_2(DDD)P[DDU]}{P[UDU]+P[DDU]+P[DDU]} = s(1-a)(1-b)^2/3, & \omega \in A^2 \\
\frac{S_2(DDD)P[DDD]}{P[DDD]} = s(1-b)^2, & \omega \in A^3
\end{cases}$$

Note how the obtained expression does not depend on $p$. Why is that?

(3) $\frac{S_2}{S_1} = X_1$, and this random variable is independent of the past, i.e. independent of the $\sigma$-algebra $F_1^S$. Therefore, by (CE7), we have

$$E[\frac{S_2}{S_1}|F_1^S](\omega) = E[\frac{S_2}{S_1}] = E[X_1] = p(1 + a) + (1 - p)(1 - b), \text{ for all } \omega \in \Omega.$$
1.7. CONDITIONING

CHAPTER 1. BROWNIAN MOTION

(4) Just like in the part (2), we can get that

\[ E[\sigma(S_3)](\omega) = \begin{cases} 
S_1(UUU) & = s(1 + a), \\
S_1(UUD) + S_1(DUD) & = S_1(DU\cup DU\cup DU) = s\left(\frac{1-b+2(1+a)}{3}\right), \\
S_1(UDD) + S_1(DUD) + S_1(DDU) & = S_1(DD\cup DU\cup DD) = s\left(\frac{1-b+2(1+a)}{3}\right), \\
S_1(DDD) & = s(1 - b), \\
\end{cases} \]

and the desired conditional expectation \(E[S_1 + S_2|\sigma(S_3)]\) is obtained by adding the expression above to the expression from (2), by (CE5).

(d) The filtration \(G\) of the insider will depend on the values of \(a\) and \(b\). I will solve only the case in which \((1 + a)^2(1 - b) > 1\) and \((1 + a)(1 - b)^2 < 1\). The other cases are dealt with analogously. By looking at the table above we have

\[\begin{align*}
G_0 &= \sigma\{UUU, UUD, UDU, DUU\}, \{UDD, DUD, DDU, DDD\} \\
G_1 &= \sigma\{UUU, UUD, UDU\}, \{DUU\}, \{UDD\}, \{DUD, DDU, DDD\} \\
G_2 &= \sigma\{UUU, UUD\}, \{UDU\}, \{DUU\}, \{UDD\}, \{DUD\}, \{DDU, DDD\} \\
G_3 &= \sigma\{UUU\}, \{UUD\}, \{UDU\}, \{DUU\}, \{UDD\}, \{DUD\}, \{DDU\}, \{DDD\}
\end{align*}\]

Solution to Exercise 1.7.31: Assume first that \(A\) is an atom of the \(\sigma\)-algebra \(\mathcal{F}\). Then \(E[X|\mathcal{F}]\) is constant on \(A\), and its value there is the weighted average of \(X\) on \(A\) - let us call this number \(\alpha\). Therefore \(E[E[X|\mathcal{F}]1_A] = E[\alpha 1_A] = \alpha P[A]\). On the other hand, \(\alpha = E[X 1_A]P[A]\), so, after multiplication by \(P[A]\), we get

\[E[E[X|\mathcal{F}]1_A] = \alpha P[A] = E[X 1_A].\]

When \(A\) is a general set in \(\mathcal{F}\), we can always find atoms \(A_1, A_2, \ldots, A_n\) such that \(A = A_1 \cup A_2 \cup \cdots \cup A_n\), and \(A_i \cap A_j = \emptyset\) for \(i \neq j\). Thus, \(1_A = 1_{A_1} + 1_{A_2} + \cdots + 1_{A_n}\), and we can use linearity of conditional (and ordinary) expectation, together with the result we have just obtained to finish the proof.

Solution to Exercise 1.7.32:

The first conditional expectation is easier: note that \(X_1\) is measurable with \(\sigma(X_1)\), so by (CE3), \(E[X_1|\sigma(X_1)] = X_1\). On the other hand \(X_2, X_3, \ldots, X_n\) are independent of \(X_1\), and therefore so is \(X_2 + X_3 + \cdots + X_n\). Using (CE7) we get \(E[X_2 + X_3 + \cdots + X_n|\sigma(X_1)] = E[X_2 + X_3 + \cdots + X_n] = 0\), since the expectation of each increment is 0. Finally,

\[E[S_n|\sigma(X_1)] = E[X_1 + X_2 + \cdots + X_n|\sigma(X_1)] = E[X_1|\sigma(X_1)] + E[X_2 + \cdots + X_n|\sigma(X_1)] = X_1 + 0 = X_1.\]

For the second conditional expectation we will need a little more work. The increments \(X_1, X_2, \ldots, X_n\) are independent and identically distributed, so the knowledge of \(X_1\) should
have the same effect on the expectation of $S_n$, as the knowledge of $X_2$ or $X_{17}$. More formally, we have $S_n = X_k + X_1 + X_2 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_n$, so $\mathbb{E}[X_1|\sigma(S_n)] = \mathbb{E}[X_k|\sigma(S_n)]$, for any $k$. We know that $\mathbb{E}[S_n|\sigma(S_n)] = S_n$, so

$$S_n = \mathbb{E}[S_n|\sigma(S_n)] = \mathbb{E}[X_1 + X_2 + \cdots + X_n|\sigma(S_n)] = \mathbb{E}[X_1|\sigma(S_n)] + \mathbb{E}[X_2|\sigma(S_n)] + \cdots + \mathbb{E}[X_n|\sigma(S_n)]$$

and we conclude that $\mathbb{E}[X_1|\sigma(S_n)] = \frac{S_n}{n}$.

**Solution to Exercise 1.7.33:**

(a) To find the value of $k$, we write

$$1 = \int_1^2 \int_0^x kxy \, dy \, dx = k \frac{15}{8},$$

and so $k = \frac{8}{15}$.

(b) $f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) \, dy = \int_0^x \frac{8}{15} xy \, dy = \frac{4}{15} x^3$, for $1 \leq x \leq 2$, and $f_X(x) = 0$ otherwise.

For $y < 0$ or $y > 1$, obviously $f_Y(y) = 0$.

For $0 \leq y \leq 1$,

$$f_Y(y) = \int_1^2 \frac{8}{15} xy \, dx = \frac{4}{15} y.$$

For $1 < y \leq 2$, we have

$$f_Y(y) = \int_y^2 \frac{8}{15} xy \, dx = \frac{4}{15} y(4 - y^2).$$

(c) The conditional density $f_{Y|X}(y,x = \xi)$, for $1 \leq \xi \leq 2$ is given by

$$f_{Y|X}(y,x = \xi) = \frac{f_{(X,Y)}(\xi,y)}{f_X(\xi)} = \begin{cases} \frac{8}{15} \frac{\xi y}{f_X(\xi)} & 0 \leq y \leq x, \\ \frac{2y\xi^{-2}}{f_X(\xi)} & 0 \leq y \leq \xi, \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\mathbb{E}[Y|X = \xi] = \int_0^\xi 2y y\xi^{-2} \, dy = \frac{2}{3} \xi,$$
1.7. CONDITIONING

CHAPTER 1. BROWNIAN MOTION

and consequently $\mathbb{E}[Y|\sigma(X)] = \frac{2}{3}X$. By (CE3) and (CE5) we have then

$$\mathbb{E}[U|\sigma(X)] = \mathbb{E}[X + Y|\sigma(X)] = X + \mathbb{E}[Y|\sigma(X)] = \frac{5}{3}X.$$

Solution to Exercise 1.7.34:

(a) The $\sigma$-algebra $\mathcal{F}_3^B$ “knows” the values of all $B_t$ for $0 \leq t \leq 3$, so that $B_3, -B_3, B_2 + B_1, B_2 + B_{1.5}$, and $\int_0^3 B_s \, ds$ are measurable there. The random variable $B_4 - B_2$ is not measurable with respect to $\mathcal{F}_3^B$, because we would have to peek into the future from $t = 3$ in order to know its value.

(b)

(2),(3) We know that martingales have constant expectation, i.e. the function $t \mapsto \mathbb{E}[X_t]$ is constant for any martingale $X$. This criterion immediately rules out (2) and (3) because $\mathbb{E}[B_t - t] = -t$, and $\mathbb{E}[B_t^2] = t$ - and these are not constant functions.

(1) The process $-B_t$ is a martingale because

$$\mathbb{E}[(-B_t)|\mathcal{F}_s] = -\mathbb{E}[B_t|\mathcal{F}_s] = -B_s,$$

and we know that $B_t$ is a martingale.

(4) To show that $B_t^2 - t$ is a martingale, we write $B_t^2 = (B_s + (B_t - B_s))^2 = B_s^2 + (B_t - B_s)^2 + 2B_s(B_t - B_s)$. The random variable $(B_t - B_s)^2$ is independent of $\mathcal{F}_s^B$, so by (CE7), we have

$$\mathbb{E}[(B_t - B_s)^2|\mathcal{F}_s^B] = \mathbb{E}[(B_t - B_s)^2] = (t - s).$$  \hfill (one)

Further, $B_s$ is measurable with respect to $\mathcal{F}_s^B$, so

$$\mathbb{E}[2B_s(B_t - B_s)|\mathcal{F}_s^B] = 2B_s\mathbb{E}[B_t - B_s|\mathcal{F}_s^B] = 2B_s\mathbb{E}[B_t - B_s] = 0,$$  \hfill (two)

by (CE6) and (CE7). Finally $B_s^2$ is $\mathcal{F}_s^B$-measurable, so

$$\mathbb{E}[B_s^2|\mathcal{F}_s^B] = B_s^2.$$  \hfill (three)

Adding (one), (two) and (three) we get

$$\mathbb{E}[B_t^2|\mathcal{F}_s^B] = (t - s) + B_s^2,$$

and so $\mathbb{E}[B_t^2 - t|\mathcal{F}_s^B] = B_s^2 - s$,

proving that $B_t^2 - t$ is a martingale.
Finally, let us prove that $X_t = \exp(B_t - \frac{1}{2}t)$ is a martingale. Write

$$X_t = X_s \exp(-\frac{1}{2}(t-s)) \exp(B_t - B_s),$$

so that $X_s = \exp(B_s - \frac{1}{2}s)$ is $\mathcal{F}_s^B$-measurable, and $\exp(B_t - B_s)$ is independent of $\mathcal{F}_s^B$ (because the increment $B_t - B_s$ is). Therefore, using (CE6) and (CE7) we have

$$\mathbb{E}[X_t|\mathcal{F}_s^B] = X_s \exp(-\frac{1}{2}(t-s))\mathbb{E}[\exp(B_t - B_s)|\mathcal{F}_s^B] = X_s \exp(-\frac{1}{2}(t-s))\mathbb{E}[\exp(B_t - B_s)],$$

and it will be enough to prove that $\mathbb{E}[\exp(B_t - B_s)] = \exp(\frac{1}{2}(t-s))$. This will follow from the fact that $B_t - B_s$ has a $N(0, t-s)$ distribution, and so (using Maple for example)

$$\mathbb{E}[\exp(B_t - B_s)] = \int_{-\infty}^{\infty} \exp(\xi) \frac{1}{\sqrt{2\pi(t-s)}} \exp(-\frac{\xi^2}{2(t-s)}) d\xi = \exp(\frac{1}{2}(t-s)).$$

**Solution to Exercise 1.7.35:**

(a) Using the properties of conditional expectation and the fact that $B_t - B_s$ (and therefore also $e^{-\lambda(B_t - B_s)}$) is independent of $\mathcal{F}_s^B$, we get

$$\mathcal{M}_{B_t|\mathcal{F}_s^B}(\lambda) = \mathbb{E}[\exp(-\lambda B_t)|\mathcal{F}_s^B] = \exp(-\lambda B_s)\mathbb{E}[\exp(-\lambda (B_t - B_s))] = \exp(-\lambda B_s + \frac{\lambda^2(t-s)}{2}).$$

(b) Let us calculate first $Y = \mathbb{E}[\exp(-\lambda_1 B_{t_1} - \lambda_2 B_{t_2})|\mathcal{F}_{t_3}^B]$. By using the properties of conditional expectation and Brownian motion we get:

$$Y = \exp(-\lambda_1 B_{t_1}) \exp(-\lambda_2 B_{t_2}) \mathbb{E}[\exp(-\lambda_2 (B_{t_2} - B_{t_3}))],$$

and thus, by the tower property of conditional expectation, we have

$$\mathcal{M}_{(B_{t_1}, B_{t_2})|\sigma(B_{t_3})}(\lambda_1, \lambda_2) = \mathbb{E}[Y|\sigma(B_{t_3})] = \exp(-\lambda_2 B_{t_3})\mathbb{E}[\exp(-\lambda_2 (B_{t_2} - B_{t_3}))]\mathbb{E}[\exp(-\lambda_1 B_{t_1})|\sigma(B_{t_3})].$$

On the other hand,

$$\mathcal{M}_{B_{t_1}|\sigma(B_{t_3})}(\lambda_2) = \mathbb{E}[\exp(-\lambda_1 B_{t_1})|\sigma(B_{t_3})],$$

and

$$\mathcal{M}_{B_{t_2}|\sigma(B_{t_3})}(\lambda_2) = \mathbb{E}[\exp(-\lambda_2 (B_{t_2} - B_{t_3}) - \lambda_2 B_{t_3})|\sigma(B_{t_3})] = \exp(-\lambda_2 B_{t_3})\mathbb{E}[\exp(-\lambda_2 (B_{t_2} - B_{t_3}))],$$

and the claim of the problem follows.
Chapter 2

Stochastic Calculus
2.1 Stochastic Integration

2.1.1 What do we really want?

In the last section of Chapter 1, we have seen how discrete-time martingales are exactly those processes for which you cannot make money (on average) by betting on. It is the task of this section to define what it means to bet (or trade) on a continuous process and to study the properties of the resulting "wealth"-processes. We start by reviewing what it means to bet on a discrete-time process. Suppose that a security price follows a stochastic process $(X_n)_{n\in\mathbb{N}_0}$, and that the public information is given by the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. The investor (say Derek the Daisy for a change) starts with $x$ dollars at time 0, and an inexhaustible credit line. After observing the value $X_0$, but before knowing $X_1$ - i.e. using the information $\mathcal{F}_0$ only - Derek decide how much to invest in the asset $X$ by buying $H_0$ shares. If $x$ dollars is not enough, Derek can always borrow additional funds from the bank at 0 interest rate\(^1\).

In the morning of day 1, and after observing the price $X_1$, Derek decides to rebalance the portfolio. Since there are no transaction costs, we might as well imagine Derek first selling his $H_0$ shares of the asset $X$, after breakfast and at the price $X_1$ (making his wealth equal to $Y_1 = x + H_0(X_1 - X_0)$ after paying his debts) and then buying $H_1$ shares after lunch. Derek is no clairvoyant, so the random variable $H_1$ will have to be measurable with respect to the $\sigma$-algebra $\mathcal{F}_1$. On day 2 the new price $X_2$ is formed, Derek rebalances his portfolio, pays the debts, and so on ... Therefore, Derek’s wealth at day $n$ (after liquidation of his position) will be

$$Y_n^H = x + H_0(X_1 - X_0) + \cdots + H_{n-1}(X_n - X_{n-1}). \quad (2.1.1)$$

The main question we are facing is How do we define the wealth process $Y^H$ when $X$ is a continuous process and the portfolio gets rebalanced at every instant? Let us start by rebalancing the portfolio not exactly at every instant, but every $\Delta t$ units of time, and then letting $\Delta t$ get smaller and smaller. We pick a time-horizon $T$, and subdivide the interval $[0,T]$ into $n$ equal segments $0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = T$, where $t_k = k\Delta t$ and $\Delta t = T/n$. By (2.1.1) the total wealth after $T = n\Delta t$ units of time will be given by

$$Y_T^H = x + \sum_{k=1}^{n} H_{t_{k-1}}(X_{t_k} - X_{t_{k-1}}) = \sum_{k=1}^{n} H_{t_{k-1}} \Delta X_{t_k}, \quad (2.1.2)$$

Where $\Delta X_{t_k} = X_{t_k} - X_{t_{k-1}}$. To make a conceptual headway with the expression (2.1.4), let us consider a few simple (and completely unrealistic) cases of the structure of process $X$. We

\(^1\)For the sake of compliance with the existing standards, we will always start our discrete and continuous processes from $n = 0$.

\(^2\)Later on we shall add non-zero interest rates into our models, but for now they would only obscure the analysis without adding anything interesting.
will also assume that \( X \) is a deterministic (non-random) process since the randomness will not play any role in what follows.

**Example 2.1.1.** Let \( X \) be non random and let \( X_t = t \). In that case

\[
Y^H_T = x + \sum_{k=1}^{n} H_{t_{k-1}} (X_{t_k} - X_{t_{k-1}}) = x + \sum_{k=1}^{n} H_{t_{k-1}} \Delta t \rightarrow x + \int_0^T H_t \, dt, \tag{2.1.3}
\]

as \( n \to \infty \), since the Riemann sums (and that is exactly what we have in the expression for \( Y^H_T \)) converge towards the integral. Of course, we have to assume that \( H \) is not too wild, so that the integral \( \int_0^T H_t \, dt \) exists.

**Example 2.1.2.** Let \( X \) be non-random again, suppose \( X_t \) is differentiable and let \( \frac{dX_t}{dt} = \mu(t) \).

Then the Fundamental Theorem of Calculus gives

\[
Y^H_T = x + \sum_{k=1}^{n} H_{t_{k-1}} (X_{t_k} - X_{t_{k-1}}) \approx x + \sum_{k=1}^{n} H_{t_{k-1}} \mu(t_{k-1}) \Delta t \stackrel{(n \to \infty)}{\rightarrow} x + \int_0^T H_t \mu(t) \, dt = x + \int_0^T H_t \frac{dX_t}{dt} \, dt. \tag{2.1.4}
\]

From the past two examples it is tempting to define the wealth-process \( Y^H_T \) by \( Y^H_T = \int_0^T H_u \, dB_u \), but there is a big problem. Take \( X_t = B_t \) - a Brownian motion - and remember that the paths \( t \mapsto B_t(\omega) \) are not differentiable functions. The trajectories of the Brownian motion are too irregular (think of the simulations) to admit a derivative. A different approach is needed, and we will try to go through some of its ideas in the following subsection.

### 2.1.2 Stochastic Integration with respect to Brownian Motion

Let us take the the price process \( X \) to be a Brownian motion, \( \mathcal{F}_t^B \) the filtration generated by it (no extra information here), and \( H_t \) a process adapted to the filtration \( \mathcal{F}_t^B \). If we trade every \( \Delta t = T/n \) time units, the wealth equation should look something like this:

\[
Y_T^{(n)} = x + \sum_{k=1}^{n} H_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}),
\]

and we would very much like to define

\[
x + \int_0^T H_t \, dB_t = \lim_{n \to \infty} Y_T^{(n)},
\]

and call it the wealth at time \( T \). The quotes around the limit point to the fact that we still do not know what kind of limit we are talking about, and, in fact, the exact description of it would require mathematics that is far outside the scope of these notes. We can, however, try to work out an example.
Example 2.1.3. Take $H_t = B_t$, so that the expression in question become
\[
\lim_{n \to \infty} \sum_{k=1}^{n} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}).
\]

We would like to rearrange this limit a little in order to make the analysis easier. We start from the identity
\[
B_T^2 = B_T^2 - B_0^2 = \sum_{k=1}^{n} (B_{t_k}^2 - B_{t_{k-1}}^2) = \sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})(B_{t_k} + B_{t_{k-1}})
\]
\[
= \sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}} + 2B_{t_{k-1}}) = \sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})^2 + 2 \sum_{k=1}^{n} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}),
\]
So that
\[
\lim_{n \to \infty} \sum_{k=1}^{n} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) = \frac{1}{2} B_T^2 - \frac{1}{2} \lim_{n \to \infty} \sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})^2.
\]

We have reduced the problem to finding the limit of the sum $\sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})^2$ as $n \to \infty$. Taking the expectation we get (using the fact that $B_{t_k} - B_{t_{k-1}}$ is a normal random variable with variance $t_k - t_{k-1}$)
\[
\mathbb{E}[\sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})^2] = \sum_{k=1}^{n} \mathbb{E}[(B_{t_k} - B_{t_{k-1}})^2] = \sum_{k=1}^{n} (t_k - t_{k-1}) = T.
\]

This suggests that
\[
\lim_{n \to \infty} \sum_{k=1}^{n} (B_{t_k} - B_{t_{k-1}})^2 = T,
\]
a fact which can be proved rigorously, but we shall not attempt to do it here. Therefore
\[
\lim_{n \to \infty} \sum_{k=1}^{n} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) = \frac{1}{2} (B_T^2 - T).
\]

The manipulations in this example can be modified so that they work in the general case and the integral $\int_{0}^{T} H_u \, dB_u$ can be defined for a large class (but not all) “portfolio” processes. Everything will be OK if $H$ is adapted to the filtration $\mathcal{F}_t^B$ and conforms to some other (and less important) regularity conditions. There is, however, more to observe in the example.

---

The trading strategy where the number of shares of an asset ($H_t$) in your portfolio is equal to the price of the asset ($B_t$) is highly unrealistic, but it is the simplest case where we can illustrate the underlying ideas.
above. Suppose for a second that $B_t$ does have a derivative $dB_t/\,dt = \dot{B}_t$. Suppose also that all the rules of classical calculus apply to the stochastic integration, so that

$$\int_0^T B_t \, dB_t = \int_0^T B_t \dot{B}_t \, dt = \int_0^T \frac{d(B^2_t)}{\,dt} \, dt = \frac{1}{2} B_T^2.$$ 

So what happened to $-\frac{1}{2} T$? The explanation lies in the fact that $\lim_{n \to \infty} \sum_{k=1}^n (B_{t_k} - B_{t_{k-1}})^2 = T$, a phenomenon which does not occur in classical calculus. In fact, for any differentiable function $f : [0, T] \to \mathbb{R}$ with continuous derivative we always have

$$\lim_{n \to \infty} \sum_{k=1}^n (f(t_{k+1}) - f(t_k))^2 = \lim_{n \to \infty} \sum_{k=1}^n (f'(\tilde{t}_k)\Delta t)^2 \leq M^2 \sum_{k=1}^n \frac{1}{n^2} = M^2 \frac{1}{n} \to 0,$$

where $M = \sup_{t \in [0, T]} |f'(t)|$, because the Mean Value Theorem implies that $f(t_k) - f(t_{k+1}) = f'(\tilde{t}_k)(t_k - t_{k-1}) = f'(\tilde{t}_k)\Delta t$, for some $\tilde{t}_k \in (t_{k-1}, t_k)$, and $M$ is finite by the continuity of $f'(t)$. In stochastic calculus the quantity

$$\lim_{n \to \infty} \sum_{k=1}^n (X_{t_{k+1}} - X_{t_k})^2,$$

is called the \textbf{quadratic variation} of the process $X$ and is denoted by $\langle X, X \rangle_T$. If the paths of $X$ are differentiable, then the previous calculation shows that $\langle X, X \rangle_T = 0$, and for the Brownian motion $B$ we have $\langle B, B \rangle_T = T$, for each $T$.

### 2.1.3 The Itô Formula

Trying to convince the reader that it is possible to \textit{define} the stochastic integral $\int_0^T H_u \, dB_u$ with respect to Brownian motion, is one thing. Teaching him or her how to \textit{compute} those integrals is quite a different task, especially in the light of the mysterious $-\frac{1}{2} T$ term that appears in the examples given above. We would like to provide the reader with a simple tool (formula?) that will do the trick. Let us start by examining what happens in classical calculus.

\textbf{Example 2.1.4.} Suppose that we are asked to compute the integral

$$\int_{-\pi/2}^{\pi/2} \cos(x) \, dx.$$ 

We could, of course, cut the interval $[-\pi/2, \pi/2]$ into $n$ pieces, form the Riemann-sums and try to find the limit of those, as $n \to \infty$, but very early in the calculus sequences, we are told that there is a much faster way - The Newton-Leibnitz formula (or, The Fundamental Theorem of Calculus). The recipe is the following: to integrate the function $f(x)$ over the interval $[a, b]$, ...
• Find a function $F$ such that $F'(x) = f(x)$ for all $x$.
• The result is $F(a) - F(b)$.

Our problem above can then be solved by noting that $\sin'(x) = \cos(x)$, and so $\int_{-\pi/2}^{\pi/2} \cos(x) \, dx = \sin(\pi/2) - \sin(-\pi/2) = 2$.

Can we do something like that for stochastic integration? The answer is yes, but the rules will be slightly different (remember $-\frac{1}{2}T$). Suppose now that we are given a function $f$, and we are interested in the integral

$$
\int_0^T f(B_t) \, dB_t
$$

Will its value be $F(B_T) - F(0)$, for a function $F$ such that $F'(x) = f(x)$? Taking $f(x) = x$, and remembering the example 2.1.3, we realize that it cannot be the case. We need an extra term $\gamma$ to rectify the situation:

$$
F(B_T) - F(B_0) = \int_0^T F'(B_t) \, dB_t + \gamma.
$$

In the late 1940’s, K. Itô has shown that this extra term is given by $\frac{1}{2} \int_0^T F''(B_t) \, dt$, introducing the famous Itô-formula:

$$
F(B_T) = F(B_0) + \int_0^T F'(B_t) \, dB_t + \frac{1}{2} \int_0^T F''(B_t) \, dt,
$$

for any function $F$ whose second derivative $F''$ is continuous. Heuristically, the derivation of the Itô-formula would spell as follows: use the second order Taylor expansion for the function $F$ -

$$
F(x) \approx F(x_0) + F'(x_0)(x - x_0) + \frac{1}{2} F''(x_0)(x - x_0)^2.
$$

Substitute $x = B_{tk}$ and $x_0 = B_{tk-1}$ to get

$$
F(B_{tk}) - F(B_{tk-1}) \approx F'(B_{tk-1})(B_{tk} - B_{tk-1}) + \frac{1}{2} F''(B_{tk-1})(B_{tk} - B_{tk-1})^2.
$$

Sum the above expression over $k = 1$ to $n$, so that the telescoping sum of the left-hand side becomes $F(B_T) - F(B_0)$:

$$
F(B_T) - F(B_0) \approx \sum_{k=1}^n F'(B_{tk-1})(B_{tk} - B_{tk-1}) + \frac{1}{2} \sum_{k=1}^n F''(B_{tk-1})(B_{tk} - B_{tk-1})^2,
$$

and upon letting $n \to \infty$ we get the Itô-formula. Of course, this derivation is non-rigorous and a number of criticisms can be made. The rigorous proof is, nevertheless, based on the ideas presented above. Before we move on, let us give a simple example illustrating the use of Itô’s formula.
Example 2.1.5. Let us compute \( \int_0^T B_t^2 dB_t \). Applying the Itô-formula with \( F(x) = \frac{1}{3} x^3 \) we get
\[
\frac{1}{3} B_T^3 - \frac{1}{3} B_0^3 = \int_0^T B_t^2 dB_t + \frac{1}{2} \int_0^T 2B_t dt,
\]
so that \( \int_0^T B_t^2 dB_t = \frac{1}{3} B_T^3 - \int_0^T B_t dt \).

Having introduced the Itô-formula it will not be hard to extend it to deal with the integrals of the form \( \int_0^t f(s, B_s) dB_s \). We follow an analogous approximating procedure as above:
\[
F(t + \Delta t, B_t + \Delta B_t) - F(t, B_t) \approx \Delta t \frac{\partial}{\partial t} F(t, B_t) + \Delta B_t \frac{\partial}{\partial x} F(t, B_t) + \frac{1}{2} \left( \left( \Delta B_t \right)^2 \frac{\partial^2}{\partial t^2} F(t, B_t) + 2\Delta t \Delta B_t \frac{\partial^2}{\partial x \partial t} F(t, B_t) + \left( \Delta B_t \right)^2 \frac{\partial^2}{\partial x^2} F(t, B_t) \right),
\]
only using the multidimensional version of the Taylor formula, truncated after the second-order terms. By comparison with the terms \( (\Delta B_t)^2 = (B_{k_i} - B_{k_{i-1}})^2 \) appearing the the quadratic variation, we conclude that \( (\Delta t)^2 \) and \( (\Delta t \Delta B_t) \) are of smaller order, and can thus be safely excluded. Telescoping now gives the, so called, inhomogeneous form of the Itô-formula:
\[
F(t, B_t) = F(0, B_0) + \int_0^t \left( \frac{\partial}{\partial t} F(t, B_t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} F(t, B_t) \right) dt + \int_0^t \frac{\partial}{\partial x} F(t, B_t) dB_t,
\]
for a function \( F \) of two arguments \( t \) and \( x \), continuously differentiable in the first, and twice continuously differentiable in the second.

2.1.4 Some properties of stochastic integrals

The Martingale Property: We have seen at the very end of Chapter 1, that betting on martingales produces no expected gains (or losses), and we even characterized discrete-time martingales as the class of processes possessing that property. It does not require an enormous leap of imagination to transfer the same property to the continuous-time case. Before we formally state this property, it is worth emphasizing that the true strength of stochastic analysis comes from viewing stochastic integrals as processes indexed by the upper limit of integration. From now on, when we talk about the stochastic integral of the process \( H \) with respect to the Brownian motion \( B \), we will have the process \( X_t = \int_0^t H_s dB_s \) in mind (and then \( X_0 = 0 \)). The first properties of the process \( X_t \) is given below:

**(SI1)** For a process \( H \), adapted with respect to the filtration \( F_t^B \) (generated by the Brownian motion \( B \)), the stochastic integral process \( X_t = \int_0^t H_s dB_s \) is a martingale, and its paths are continuous functions.
2.1. STOCHASTIC INTEGRATION

CHAPTER 2. STOCHASTIC CALCULUS

We have to remark right away that the statement above is plain wrong unless we require some regularity on the process $H$. If it is too wild, various things can go wrong. Fortunately, we will have no real problems with such phenomena, so we will tacitly assume that all integrands we mention behave well. For example, $X$ is a martingale if the function $h(s) = \mathbb{E}[H_s^2]$ satisfies $\int_0^t h(s) \, ds < \infty$, for all $t > 0$. As for the continuity of the paths, let us just remark that this property is quite deep and follows indirectly from the continuity of the paths of the Brownian motion. Here is an example that illustrates the coupling between (SI1) and the Itô formula.

Example 2.1.6. By the Itô formula applied to the function $F(x) = x^2$ we get

$$B_t^2 = \int_0^t 2B_s \, dB_s + \frac{1}{2} \int_0^T 2 \, ds = \int_0^t 2B_s \, dB_s + t.$$  

The property (SI1) implies that $B_t^2 - t = \int_0^t 2B_s \, dB_s$ is a martingale, a fact that we have derived before using the properties (CE1)-(CE7) of conditional expectation.

The Itô Isometry: Having seen that the stochastic integrals are martingales, we immediately realize that $\mathbb{E}[X_t] = \mathbb{E}[X_0] = 0$, so we know how to calculate expectations of stochastic integrals. The analogous question can be posed about the variance $\text{Var}[X_t] = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 = \mathbb{E}[X_t^2]$ (because $X_0 = 0$). The answer comes from the following important property of the stochastic integral, sometimes known as the Itô isometry.

(SI2) For a process $H$, adapted with respect to the filtration $\mathcal{F}_t^B$ (generated by the Brownian motion $B$), such that $\int_0^t \mathbb{E}[H_s^2] \, ds < \infty$ we have

$$\mathbb{E}\left[\left(\int_0^t H_s \, dB_s\right)^2\right] = \int_0^t \mathbb{E}[H_s^2] \, ds.$$  

Let us try to understand this property by deriving it for the discrete-time analogue of the stochastic integral $\sum_{k=1}^n H_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})$. The continuous-time version (SI2) will follow by taking the appropriate limit. We start by exhibiting two simple facts we shall use later. The proofs follow from the properties of conditional expectations (CE1)-(CE7).

- For $0 < k < l \leq n$, we have

$$\mathbb{E}\left[H_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})H_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})\right] = \mathbb{E}\left[H_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})H_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})|\mathcal{F}_{t_{l-1}}^B\right]$$  

$$= \mathbb{E}\left[H_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})H_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})|\mathcal{F}_{t_{l-1}}^B\right]$$  

$$= \mathbb{E}\left[H_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})H_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})0\right] = 0.$$  

Last Updated: Spring 2005

112 Continuous-Time Finance: Lecture Notes
• For $0 < k \leq n$ we have
\[
\mathbb{E} \left[ H_{t_{k-1}}^2 (B_{t_k} - B_{t_{k-1}}) \right]^2 = \mathbb{E} \left[ \mathbb{E}[H_{t_{k-1}}^2 (B_{t_k} - B_{t_{k-1}})^2 | \mathcal{F}^B_{t_{k-1}}] \right] = \mathbb{E} \left[ H_{t_{k-1}}^2 \mathbb{E}[(B_{t_k} - B_{t_{k-1}})^2 | \mathcal{F}^B_{t_{k-1}}] \right] = \mathbb{E}[H_{t_{k-1}}^2](t_k - t_{k-1}).
\]

By expanding the square and using the properties above we get
\[
\mathbb{E} \left[ \sum_{k=1}^{n} H_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) \right]^2 = \sum_{0<k,l\leq n} \mathbb{E} \left[ H_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) H_{t_{l-1}} (B_{t_l} - B_{t_{l-1}}) \right] = \sum_{k=1}^{n} \mathbb{E} \left[ H_{t_{k-1}}^2 (B_{t_k} - B_{t_{k-1}})^2 \right] + 2 \sum_{0<k<l\leq n} \mathbb{E} \left[ H_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) H_{t_{l-1}} (B_{t_l} - B_{t_{l-1}}) \right] = \sum_{k=1}^{n} \mathbb{E}[H_{t_{k-1}}^2](t_k - t_{k-1}) \rightarrow \int_0^t \mathbb{E}[H_s^2] ds.
\]

The Itô isometry has an immediate and useful corollary, the proof of which is delightfully simple. Let $H_s$ and $K_s$ be two processes adapted to the filtration $\mathcal{F}^B_t$ generated by the Brownian motion. Then
\[
\mathbb{E} \left[ \left( \int_0^t H_s dB_s \right) \cdot \left( \int_0^t K_s dB_s \right) \right] = \int_0^t \mathbb{E}[H_s K_s] ds.
\]

To prove this, let $L_s = H_s + K_s$ and apply the Itô isometry to the process $X_t = \int_0^t L_s dB_s$, as well as to the processes $\int_0^t H_s dB_s$ and $\int_0^t K_s dB_s$ to obtain
\[
2\mathbb{E} \left[ \int_0^t H_s dB_s \cdot \int_0^t K_s dB_s \right] = \mathbb{E} \left[ \left( \int_0^t (K_s + H_s) dB_s \right)^2 \right] - \mathbb{E} \left[ \left( \int_0^t K_s dB_s \right)^2 \right] - \mathbb{E} \left[ \left( \int_0^t L_s dB_s \right)^2 \right] = \int_0^t \mathbb{E}[(K_s + H_s)^2] ds - \int_0^t \mathbb{E}[K_s^2] ds - \int_0^t \mathbb{E}[H_s^2] ds = 2 \int_0^t \mathbb{E}[K_s H_s] ds.
\]

We are in position now to give several examples, illustrative of the martingale and the Itô isometry properties of the stochastic integration, as well as the use of the Itô formula.

**Example 2.1.7.** The purpose of this example is to calculate the moments (expectations of $k^{th}$ powers) of the normal distribution using stochastic calculus. Remember that we did this by classical integration in Section 1 of Chapter 1. Define $X_t = \int_0^t B_s^k dB_s$, for some $k \in \mathbb{N}$. By Itô’s formula applied to function $F(x) = \frac{x^{k+1}}{k+1}$ we have
\[
X_t = \frac{1}{k+1} B_{t}^{k+1} - \frac{1}{2} \int_0^t k B_s^{k-1} ds.
\]
We know that $E[X_t] = 0$ by (SI1), so putting $t = 1$,

$$E[B_{k+1}^1] = \frac{k(k+1)}{2}E[\int_0^1 B_s^{k-1} ds].$$

(2.1.6)

Define $M(k) = E[B_k^1]$ - which is nothing but the expectation of the unit normal raised to the $k^{th}$ power, and note that $s^{k/2}M(k) = E[B_k^k]$, by the normality of $B_s$, and the fact that its variance is $s$.

Fubini Theorem applied to equation (2.1.6) gives

$$M(k+1) = \frac{k(k+1)}{2} \int_0^1 E[B_s^{k-1}] ds = \frac{k(k+1)}{2} \int_0^1 s^{(k-1)/2}M(k-1) ds = kM(k-1).$$

Trivially, $M(0) = 1$, and $M(1) = 0$, so we get $M(2k-1) = 0$, for $k \in \mathbb{N}$ and $M(2k) = (2k-1)M(2k-2) = (2k-1) \cdot (2k-3)M(2k-4) = \cdots = (2k-1) \cdot (2k-3) \cdots 3 \cdot 1$.

**Example 2.1.8.** In Chapter 1 we used approximation by random walks to calculate $E[\left(\int_0^1 B_s ds\right)^2]$. In this example we will do it using stochastic calculus. Let us start with the function $F(t, B_t) = tB_t$ and apply the inhomogeneous Itô-formula to it

$$tB_t = 0 + \int_0^t B_s ds + \int_0^t s dB_s,$$

and therefore

$$E\left[\left(\int_0^t B_s ds\right)^2\right] = E\left[\left(tB_t - \int_0^t s dB_s\right)^2\right]$$

$$= t^2E[B_t^2] + E\left[\left(\int_0^t s dB_s\right)^2\right] - 2tE[B_t] \int_0^t s dB_s.$$  

(2.1.7)

The first term $t^2E[B_t^2]$ equals $t^3$ by normality of $B_t$. Applying the Itô isometry to the second term we get

$$E\left[\left(\int_0^t s dB_s\right)^2\right] = \int_0^t E[s^2] ds = \int_0^t s^2 ds = \frac{t^3}{3}.$$  

Finally, we apply the equation (2.1.6) (the corollary of Itô isometry) to the third term in (2.1.6). The trick here is to write $B_t = \int_0^t 1 dB_s$:

$$2tE[B_t \int_0^t s dB_s] = 2tE[\int_0^t 1 dB_s \cdot \int_0^t s dB_s] = 2tE[\int_0^t 1 \cdot s ds] = 2t \frac{t^2}{2} = t^3.$$

---

4the classical Fubini Theorem deals with switching integrals in the double integration. There is a version which says that

$$E[\int_0^t H_s ds] = \int_0^t E[H_s] ds,$$

whenever $H$ is not "too large". It is a plausible result once we realize that the expectation is a sort of an integral itself.
CHAPTER 2. STOCHASTIC CALCULUS  2.1. STOCHASTIC INTEGRATION

Putting all together we get

$$E \left( \left( \int_0^t B_s \, ds \right)^2 \right) = \frac{t^3}{3}.$$  

2.1.5 Itô processes

In this subsection we will try to view the Itô formula in a slightly different light. Just like

the Fundamental theorem of calculus allows us to write any differentiable function $F(t)$ as an

indefinite integral (plus a constant)

$$F(t) = F(0) + \int_0^t F'(s) \, ds$$

of the function $F'$ (which happens to be the derivative of $F$), the Itô formula states that if

we apply a function $F$ to the Brownian motion $B_t$ the result will be expressible as a sum of

constant and a pair of integrals - one classical (with respect to $ds$) and the other stochastic

(with respect to $dB_s$):

$$F(B_t) = F(B_0) + \frac{1}{2} \int_0^t F''(B_s) \, ds + \int_0^t F'(B_s) \, dB_s,$$

or, in the inhomogeneous case, when $F(\cdot, \cdot)$ is a function of two arguments $t$ and $x$

$$F(t, B_t) = F(0, B_0) + \int_0^t \left( \frac{\partial}{\partial t} F(s, B_s) + \frac{1}{2} \frac{\partial^2}{\partial x^2} F(s, B_s) \right) \, ds + \int_0^t \frac{\partial}{\partial x} F(s, B_s) \, dB_s.$$  

The processes $X_t$ which can be written as

$$X_t = X_0 + \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dB_s,$$  \hspace{1cm} (2.1.8)

for some adapted processes $\mu_s$ and $\sigma_s$ are called Itô processes and constitute a class of

stochastic processes structured enough so that their properties can be studied using stochastic

calculus, yet large enough to cover most of the financial applications. The process $\mu_s$ is called

the drift process and the process $\sigma_s$ the volatility process$^5$ of the Itô process $X_t$. The two

parts $\int_0^t \mu_s \, ds$ and $\int_0^t \sigma_s \, dB_s$ can be thought of as signal and noise, respectively. The signal

component $\int_0^t \mu_s \, ds$ describes some global trend around which the process $X_s$ fluctuates, and

the volatility can be interpreted as the intensity of that fluctuation. The Itô formula states

that any process $X_t$ obtained by applying a function $F(t, x)$ to a Brownian motion is an Itô

process with

$$\mu_s = \frac{\partial}{\partial t} F(s, B_s) + \frac{1}{2} \frac{\partial^2}{\partial x^2} F(s, B_s) \quad \text{and} \quad \sigma_s = \frac{\partial}{\partial x} F(s, B_s).$$

The question whether an Itô process $X_t$ is a martingale (sub-, super-) is very easy to answer:

$^5$An important subclass of the class of Itô processes are the ones where $\mu_s$ and $\sigma_s$ are functions of the time

$s$ and the value of the process $X_t$ only. Such processes are called diffusions.
2.1. STOCHASTIC INTEGRATION

• \( X_t \) is a martingale if \( \mu_t = 0 \), for all \( t \),
• \( X_t \) is a supermartingale if \( \mu_t \leq 0 \) for all \( t \), and
• \( X_t \) is a submartingale if \( \mu_t \geq 0 \) for all \( t \).

Of course, if \( \mu_t \) is sometimes positive and sometimes negative, \( X_t \) is neither a supermartingale nor a submartingale.

There is a useful, if not completely correct, notation for Itô processes, sometimes called the differential notation. For an Itô process given by (2.1.8), we write

\[
dX_t = \mu_t \, dt + \sigma_t \, dB_t, \quad X_0 = x_0. \tag{2.1.9}
\]

The initial value specification \((X_0 = x_0)\) is usually omitted because it is often not important where the process starts. For a real function \( F \), the representations \( F(t) = F(0) + \int_0^t F'(s) \, ds \) and \( dF(t) = F'(t) \, dt \) are not only formally equivalent because \( \frac{dF(t)}{dt} \) equals \( F'(t) \) in a very precise way described in the fundamental theorem of calculus. On the other hand, it is not correct to say that

\[
\frac{dX_t}{dt} = \mu_t + \sigma_t \frac{dB_t}{dt},
\]

since the paths of the Brownian motion are nowhere differentiable, and the expression \( \frac{dB_t}{dt} \) has no meaning. Therefore, anytime you see (2.1.9), it really means (2.1.8). The differential notation has a number of desirable properties. First, it suggests that the Itô processes are built of increments (a view very popular in finance) and that the increment \( \Delta X_t = X_{t+\Delta t} - X_t \) can be approximated by the sum \( \mu_t \Delta t + \sigma_t \Delta B_t \) - making it thus “normally distributed” with mean \( \mu_t \Delta t \) and variance \( \sigma_t^2 \Delta t \). These approximations have to be understood in the conditional sense, since \( \mu_t \) and \( \sigma_t \) are random variables measurable with respect to \( \mathcal{F}_t \). Another useful feature of the differential notation is the ease with which the Itô formula can be applied to Itô processes. One only has to keep in mind the following simple multiplication table:

\[
\begin{array}{c|cc}
\times & dB_t & dt \\
\hline
dB_t & dt & 0 \\
dt & 0 & 0
\end{array}
\]

Now, the Itô formula for Itô processes can be stated as

\[
dF(t, X_t) = \frac{\partial}{\partial x} F(t, X_t) \, dX_t + \frac{\partial}{\partial t} F(t, X_t) \, dt + \frac{1}{2} \frac{\partial^2}{\partial x^2} F(t, X_t)(dX_t)^2, \tag{2.1.10}
\]

where, using the multiplication table,

\[
(dX_t)^2 = (\mu_t \, dt + \sigma_t \, dB_t)(\mu_t \, dt + \sigma_t \, dB_t) = \mu_t^2 \, (dt)^2 + 2\mu_t\sigma_t \, dtdB_t + \sigma_t^2 \, (dB_t)^2 = \sigma_t^2 \, dt.
\]
Therefore,
\[
dF(t, X_t) = \left( \frac{\partial}{\partial x} F(t, X_t) \mu_t + \frac{\partial}{\partial t} F(t, X_t) + \frac{1}{2} \sigma^2_t \frac{\partial^2}{\partial x^2} F(t, X_t) \right) dt + \frac{\partial}{\partial x} F(t, X_t) \sigma_t dB_t
\]
(2.1.11)

Of course, the formula (2.1.10) is much easier to remember and more intuitive than (2.1.11). Formula (2.1.11), however, reveals the fact that \(F(t, X_t)\) is an Itô process if \(X_t\) is. We have already mentioned a special case of this result when \(X_t = B_t\). A question that can be asked is whether a product of two Itô processes is an Itô process again, and if it is, what are its drift and volatility? To answer the question, let \(X_t\) and \(Y_t\) be two Itô processes

\[
dX_t = \mu_t^X dt + \sigma_t^X dB_t, \quad dY_t = \mu_t^Y dt + \sigma_t^Y dB_t.
\]

In a manner entirely analogous to the one in which we heuristically derived the Itô formula by expanding alla Taylor and using telescoping sums, we can get the following formal expression (in differential notation)

\[
d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t.
\]

This formula looks very much like the classical integration-by-parts formula, except for the extra term \(dX_t dY_t\). To figure out what this term should be, we resort to the differential notation:

\[
dx_t dY_t = (\mu_t^X dt + \sigma_t^X dB_t)(\mu_t^Y dt + \sigma_t^Y dB_t) = \sigma_t^X \sigma_t^Y dt,
\]

so that we can write

\[
d(X_t Y_t) = (X_t \mu_t^Y + Y_t \mu_t^X + \sigma_t^X \sigma_t^Y) dt + (X_t \sigma_t^Y + Y_t \sigma_t^X) dB_t,
\]

showing that the product \(X_t Y_t\) is an Itô process and exhibiting its drift and volatility. Another strength of the differential notation (and the last one in this list) is that it formally reduces stochastic integration with respect to an Itô process to stochastic integration with respect to a Brownian motion. For example, if we wish to define the integral

\[
\int_0^t H_s dX_s, \text{ with respect to the process } dX_t = \mu_t dt + \sigma_t dB_t,
\]

we will formally substitute the expression for \(dX_t\) into the integral obtaining

\[
\int_0^t H_s dX_s = \int_0^t H_s \mu_s ds + \int_0^t H_s \sigma_s dB_s.
\]

A closer look at this formula reveals another stability property of Itô processes - they are closed under stochastic integration. In other words, if \(X_t\) is an Itô process and \(H_s\) is an adapted process, then the process \(Y_t = \int_0^t H_s dX_s\) is an Itô process.
Examples. After all the theory, we give a number of examples. Almost anything you can think of is an Itô process.

Example 2.1.9. The fundamental example of an Itô process is the Brownian motion $B_t$. Its drift and volatility are $\mu_t = 0$, $\sigma_t = 1$, as can be seen from the representation

$$B_t = 0 + B_t = \int_0^t 0 \, ds + \int_0^t 1 \, dB_s.$$ 

An adapted process whose paths are differentiable functions is an Itô process:

$$X_t = \int_0^t \mu_s \, ds, \text{ where } \mu_t = \frac{dX_t}{dt}.$$ 

In particular, the deterministic process $X_t = t$ is an Itô process with $\mu_t = 1$, $\sigma_t = 0$. Then, Brownian motion with drift $X_t = B_t + \mu t$ is an Itô process with $\mu_t = \mu$ and $\sigma_t = 1$. The processes with jumps, or the processes that look into the future are not Itô processes.

Example 2.1.10. [Samuelson’s model of the stock prices] Paul Samuelson proposed the Itô process

$$S_t = \exp\left(\sigma B_t + (\mu - \frac{1}{2} \sigma^2) t\right), \quad S_0 = s_0,$$

as the model for the time-evolution of the price of a common stock. This process is sometimes referred to as the geometric Brownian motion with drift. We shall talk more about this model in the context of derivative pricing in the following sections. Let us now just describe its structure by finding its drift and volatility processes. The Itô formula yields the following expression for the process $S$ in the differential notation:

$$dS_t = \exp\left(\sigma B_t + (\mu - \frac{1}{2} \sigma^2) t\right) \left(\mu - \frac{1}{2} \sigma^2 + \frac{1}{2} \sigma^2\right) dt + \exp\left(\sigma B_t + (\mu - \frac{1}{2} \sigma^2) t\right) \sigma dB_t$$

$$= S_t \mu dt + S_t \sigma dB_t.$$

Apart from being important in finance, this example was a good introduction to the concept of a stochastic differential equation. We say that an Itô process $X_t$ is a solution to the stochastic differential equation (SDE)

$$dX_t = F(t, X_t) \, dt + G(t, X_t) \, dB_t, \quad X_0 = x_0,$$

if, of course,

$$X_t = x_0 + \int_0^t F(s, X_s) \, ds + \int_0^t G(s, X_s) \, dB_s, \text{ for all } t \geq 0.$$
CHAPTER 2. STOCHASTIC CALCULUS

2.1. STOCHASTIC INTEGRATION

The process \( S_t \) in the Samuelson’s stock-price model is thus a solution to the stochastic differential equation

\[
dS_t = S_t \mu \, dt + S_t \sigma \, dB_t, \quad S_0 = s_0, \tag{2.1.12}
\]

and \( F(t, x) = \mu x, \; G(t, x) = \sigma x \). This particular SDE has a closed-form solution in terms of \( t \) and \( B_t \),

\[ S_t = S_0 \exp(\sigma B_t + t(\mu - \frac{1}{2} \sigma^2)), \]

as we have checked above using the Itô formula. It is important to notice that the concept of an SDE is a generalization of the concept of an ordinary differential equation (ODE). Just set \( G(t, x) = 0 \), and you will end-up with the familiar expression

\[ dX_t = F(t, X_t) \, dt, \quad \text{i.e.} \quad \frac{dX_t}{dt} = F(t, X_t), \]

for a generic ODE. With this in mind, an SDE can be thought of as a perturbation or a noisy version of an ODE. Let us try to apply these ideas to the simple case of Samuelson’s model. The noise-free version of the SDE (2.1.12) will be (let us call the solution process \( \hat{S} \)):

\[
d\hat{S}_t = \hat{S}_t \, dt, \quad \text{i.e.} \quad \frac{d\hat{S}_t}{dt} = \mu \hat{S}_t. \tag{2.1.13}
\]

This is an example of a linear homogeneous ODE with constant coefficients and its solution describes the exponential growth:

\[ \hat{S}_t = s_0 \exp(\mu t). \]

If we view \( \mu \) as an interest-rate, \( \hat{S} \) will model the amount of money we will have in the bank after \( t \) units of time, starting from \( s_0 \) and with the continuously compounded interest at the constant rate of \( \mu \). The full SDE (2.1.12) models stock-price evolution, so that we might say that the stock-prices, according to Samuelson, are bank accounts + noise.

Fig 18. Historical stock-price of a stock traded on NYSE

Fig 19. Simulated path of a geometric Brownian motion (Samuelson’s model)
2.1. STOCHASTIC INTEGRATION

Of course, there is much more than that to finance, and Paul Samuelson received a Nobel Prize in economics in 1970 for a number of contributions including his work on financial markets. In terms of the “drift and volatility” interpretation of Itô processes, we may put forward the following non-rigorous formula

$$\frac{\Delta S_{t_i}}{S_{t_i}} = \frac{S_{t_{i+1}} - S_{t_i}}{S_{t_i}} \approx \mu \Delta t + \sigma \Delta B_t,$$

stating that in a small time-interval $\Delta t$, the change in the stock-price, relative to the stock-price at the beginning of the interval, will move deterministically by $\mu \Delta t$ coupled with the noise-movement of $\sigma \Delta B_t$. In particular, the Samuelson model predicts the relative stock returns over small intervals (and over large intervals, too) to be normally distributed. We will have very little to say about the rich theory of stochastic differential equations in this course, mainly due to the lack of time and tools, as well as the fact that we shall work almost exclusively with the Samuelson’s model (2.1.12) whose solution can be obtained explicitly. However, here are some examples of famous SDE’s and their solutions in the closed form (whenever they exist).

Example 2.1.11 (Ornstein-Uhlenbeck process). The Ornstein-Uhlenbeck process (often abbreviated to OU-process) is an Itô process with a long history and a number of applications. The SDE satisfied by the OU-process is given by

$$dX_t = \alpha (m - X_t) \, dt + \sigma dB_t, \quad X_0 = x_0,$$

for some constants $m, \alpha, \sigma \in \mathbb{R}$, $\alpha, \sigma > 0$. Intuitively, the OU-process models a motion of a particle which drifts toward the level $m$, perturbed with the Brownian noise of intensity $\sigma$. Notice how the drift coefficient is negative when $X_t > m$ and positive when $X_t < m$. Because of that feature, OU-process is sometimes also called a mean-reverting process, and a variant of it is widely used as a model of the interest-rate dynamics. Another of its nice properties is that the OU process is a Gaussian process, making it amenable to empirical and theoretical study.

In the remainder of this example, we will show how this SDE can be solved using some methods similar to the ones you might have seen in your ODE class. The solution we get will not be completely explicit, since it will contain a stochastic integral, but as SDE’s go, even that is
usual more than we can hope for. First of all, let us simplify the notation (without any real loss of generality) by setting $m = 0$, $\sigma = 1$. Let us start by defining the process $Y_t = R_t X_t$, with $R_t = \exp(\alpha t)$, inspired by the technique one would use to solve the corresponding noiseless equation

$$d\hat{X}_t = -\alpha X_t \, dt.$$  

The process $R_t$ is a deterministic Itô process with $dR_t = \alpha R_t \, dt$, so the formula for the product (generalized integration by parts formula) gives

$$dY_t = R_t \, dX_t + X_t \, dR_t = -\alpha X_t R_t \, dt + R_t \, dB_t + X_t \alpha R_t \, dt = R_t \, dB_t = e^{\alpha t} \, dB_t.$$  

We have obtained that $dR_t \, dX_t = 0$ by using the multiplication table, since there is no $dB_t$ term in the expression for $R_t$. It follows now that

$$X_t e^{\alpha t} - x_0 = Y_t - Y_0 = \int_0^t e^{\alpha s} \, dB_s,$$

so that

$$X_t = x_0 e^{-\alpha t} + e^{-\alpha t} \int_0^t e^{\alpha s} \, dB_s.$$  

This expression admits no further significant simplifications.
2.2 Applications to Finance

2.2.1 Modelling Trading and Wealth

In the Meadow Stock Exchange there are only 2 financial instruments. A riskless money-market account \( R_t \) and risky GII (Grasshopper’s Industries Incorporated) common stock \( S_t \). As the only person even remotely interested in investments, Derek the Daisy is allowed to assume any position (long or short) in the stock and in the money-market account (borrowing and lending interest rates are equal). The only requirement is the one of self-financing, i.e. starting with $x, Derek is not allow to inject or withdraw any funds from the two accounts.

He can, however, sell stocks and invest the money in the money-market account, or vice versa. Suppose Derek chooses the following strategy: at time \( t \) his holdings will be \( \pi_t \) shares of stock and \( b_t \) in the money-market account. Of course, \( \pi_t \) and \( b_t \) will not be deterministic in general. The exposure to the stock may depend on Derek’s current wealth at time \( t \), the stock-price, or any number of other factors. Also, not every combination \( (\pi_t, b_t) \) will be allowable - remember, money cannot be created out of thin air. The self-financing condition requires that the total wealth at time \( t \) be composed from the initial wealth $x, the gains from trading, and the deterministic return from the money-market. Equivalently we may say that if Derek’s wealth at time \( t \) is \( W_t \), and he decides to hold exactly \( \pi_t \) shares of stock, the holdings in the money-market account will have to be

\[
b_t = W_t - \pi_t S_t. \tag{2.2.1}
\]

Naturally, if \( W_t \) is small and \( \pi_t \) is large, \( b_t \) might become negative, amounting to borrowing from the money-market account to finance the stock purchase. Assume now that the dynamics of the money-market account and the stock price are given by the following Itô processes:

\[
dS_t = S_t \mu \, dt + S_t \sigma \, dB_t, \quad S_0 = s_0,
\]
\[
dR_t = R_t r \, dt, \quad R_0 = 1,
\]

where \( r \) is the continuously compounded interest-rate, and \( \mu \) and \( \sigma \) are the rate of return and volatility of the stock \( S_t \), respectively. We will assume that these coefficients are known and constant during the planning horizon \([0, T]\). Even though \( S_t \) is defined in terms of a stochastic differential equation, we know from previous section that it admits a closed-form

---

\(^6\)some authors use \( \pi_t \) to denote the total amount of dollars invested in the stock. In our case, this quantity would correspond to \( S_t \pi_t \).

\(^7\)There are efficient statistical methods for estimation of these parameters. The problems and techniques involved in the procedure would require a separate course, so we will simply assume that \( \mu \) and \( \sigma \) are reliable numbers obtained from a third party.

\(^8\)For relatively short planning horizons this assumption is not too bad from the practical point of view. For longer horizons we would have to model explicitly the dynamics of the market coefficients \( \mu \) and \( \sigma \) and the interest-rate \( r \), as well.
CHAPTER 2. STOCHASTIC CALCULUS 2.2. APPLICATIONS TO FINANCE

solution \( S_t = S_0 \exp(\sigma B_t + t(\mu - \frac{1}{2}\sigma^2)) \). However, the SDE representation is more intuitive and will provide us with more information in the sequel. Having specified the models involved, let us look into the dynamics of our wealth and see how it depends on portfolio processes \((\pi_t, b_t)\), subject to the self-financing condition (2.2.1). We start by discretizing time and looking at what happens in a small time-period \( \Delta t \). Using the fact that the change in wealth comes either from the gains from the stock-market, or the interest paid by the money-market, and substituting from equation (2.2.1), we get:

\[
W_{t+\Delta t} \approx W_t + b_t/R_t \Delta R_t + \pi_t \Delta S_t \\
= W_t + (W_t - \pi_t S_t) r \Delta t + \pi_t \Delta S_t \\
\approx W_t + rW_t \Delta t + \pi_t S_t \left( (\mu - r) \Delta t + \sigma \Delta B_t \right)
\]

(2.2.2)

Passing to the limit as \( \Delta t \to 0 \), we get the following SDE, which we call the **wealth equation** for the evolution of the wealth \( W_t = W_{t,x,\pi} \) depending on the portfolio process \( \pi_t \):

\[
dW_{t,x,\pi} = rW_{t,x,\pi} dt + \pi_t S_t \left( (\mu - r) dt + \sigma dB_t \right), \quad W_{0,x,\pi} = x.
\]

(2.2.3)

A quick look at the equation above leads us to give another way of thinking about the growth of wealth. We can think of the wealth \( W_t \) as growing continuously at the riskless rate of \( r \), coupled with the risky growth at the **excess growth rate** \( \mu - r \) and volatility \( \sigma \), multiplied by the factor \( \pi_t S_t \) - our exposure to the stock fluctuations. Moral of the story: any profit excessive to the interest \( r \) comes with a certain amount of risk. The more excess profit you want, the more risk you will have to face.

### 2.2.2 Derivatives

One of the greatest achievements of the modern finance theory is a satisfactory and empirically confirmed derivative-pricing theory. To refresh your memory, here is a “simple” definition of a derivative security from *General Rules and Regulations promulgated under the Securities Exchange Act of 1934*:

The term **derivative securities** shall mean any option, warrant, convertible security, stock appreciation right, or similar right with an exercise or conversion privilege at a price related to an equity security, or similar securities with a value derived from the value of an equity security...

Seriously, a **derivative security** is a financial instrument whose price (value, payoff, ...) depends (sometimes in very complicated ways) on the price (value, payoff, price history, ...) of an **underlying** security. Here is a number of examples. In all these examples we shall assume that \( S_t \) is a stock traded on a securities exchange. That will be our underlying security.
Example 2.2.1 (Call Options). The call option is basically a colored piece of paper that gives the holder the right (but not the obligation, hence the name option) to buy a share of the stock \( S_t \) at time \( T \) for price \( K \).

The underlying security \( S_t \), maturity \( T \) and the strike price \( K \) are a part of the contract. For different \( K \) and/or \( T \) you would get different call options. In the case of the call option, it doesn’t take much to figure out when to exercise the option, i.e. when to use your right to purchase a share of the stock \( S_t \) for the price \( K \). You will do it only when \( S_t > K \). Otherwise you could buy the same share of stock at the market for a price less than \( K \). In that case the option is worthless. In financial slang, we say that the call option is in the money when \( S_t > K \), at the money when \( S_t = K \), and out of the money when \( S_t < K \), even for \( t < T \).

Given that it is easy to figure out when the option should be exercised, and we suppose that the owner of the option will immediately sell the stock at the prevailing market rate if she decides to exercise it, we see that the net payoff of the call option is given by

\[
C = (S_T - K)^+,
\]

where the function \((\cdot)^+\) is defined by

\[
\begin{align*}
(x)^+ &= x, \quad x \geq 0, \\
(x)^+ &= 0, \quad x < 0.
\end{align*}
\]

When the option’s payoff depends on the options price solely through its value \( S_T \) at maturity, i.e. payoff = \( f(S_T) \), it is customary to depict the payoff by plotting the function \( f(\cdot) \), so that the \( x \)-axis stands for the underlying’s price, and the \( y \)-axis for the payoff. Very often, the option’s price at time \( t < T \) will be a function of \( t \) and \( S_t \) only, so that we can plot it at the same graph with \( f \).
Example 2.2.2 (Put Options). The put option is a relative of a call option, with the notable difference that here you are entitled to sell one share of the stock $S_t$ at the price $K$. Just like in the case of the call option, the owner of a put option will choose to exercise it if the stock price is below the level $K$. The notions of in-, at- and out of the money are defined in the same way as in the case of a call option. The payoff of a put option is given by

$$P = (K - S_T)^+,$$

where $T$ is the maturity, and $K$ is the strike price of the put option.

![Fig 22. Payoff of a put option (solid line) with strike $K = 90$, and the price of the underlying stock (dashed line)](image)

Example 2.2.3 (Future and forward contracts). A forward option is like a call option, but now the owner has the obligation to buy a share of the stock $S_t$ at a specified price $K$. A future contract is an institutionalized (and liquid) version of a forward contract traded in the market. The payoff $F$ of a forward contract is given by $F = S_T - K$. Unlike calls and futures, the forward contract can be costly to its holder in case $S_T < K$.

![Fig 23. Payoff of a forward option (solid line) with strike $K = 90$, and the price of the underlying stock (dashed line)](image)

So far we have talked about the most basic and most widespread options, so called vanilla options. As described in the examples above, all of these are European options, meaning that their maturities are fixed and contractually specified in advance. In practice, the majority of traded options are of American type, meaning that the holder (owner) can not only choose whether to exercise the options, but also when to do it. The pricing theory for those options is technically much more involved than the theory of European options, so we shall stick to Europeans in this course. The next few examples are about less traded, but still important, exotic options.
Example 2.2.4 (Asian Options). An Asian option with maturity $T$ and strike price $K$ gives its holder the payoff of

$$A = \left( \frac{1}{T} \int_0^T S_t \, dt - K \right)^{+}.$$ 

In words, Asian options work just like calls, with the difference that the payoff depends on the average stock prices during the entire planning horizon, and not only on the stock-price at the maturity. Stock exchanges like Asian options because they are less amenable to price manipulation by large investors (how?).

Example 2.2.5 (Knock-in options). A Knock-in option is an option whose payoff equals that of a call option, but only if the price of the stock has crossed a predetermined barrier sometime in $[0,T]$. If that doesn’t happen, the option is worthless, even if $S_T > K$. Usually we take a constant barrier at the level $b$, so that the value of a knock-in option is given by

$$K = (S_T - K)^{+} \cdot 1\{\max_{0\leq t \leq T} S_t \geq b\}.$$ 

There are all kinds of variants of knock-in options. Sometimes the option will lose value if the stock-price crosses a barrier. Such options are called Knock-out options.

Example 2.2.6 (Binary (digital) options). A Binary option (also called a Digital option) gives a payoff of $1$ if the stock price at time $T$ is above some level $b$, and nothing if $S_t$ is below $b$. In symbols

$$B = 1\{S_T > b\}.$$ 

Binary options are nothing but gambling instruments because of their extreme sensitivity to the price of the underlying. They are rarely allowed because of the incentive the give to price manipulation. Apparently, George Soros made a large part of his fortune by speculating with foreign-exchange binary options.

2.2.3 Arbitrage Pricing

The derivative securities are either traded on stock-exchanges or offered over-the-counter. In either case, it is important to know what its “fair” price should be. In general, it is difficult to speak of and define the notion of a fair price, since the prices are results of the equilibrium between supply and demand. In financial markets\(^9\), there is a very precise notion of a fair price which leaves very little room for discussion. This price is called the arbitrage price, and is defined by the requirement that it is impossible to make a riskless profit in the market. The economic justification of this requirement is that, if such an opportunity existed, people would exploit it until the prices of the securities are pushed to non-arbitrage levels. Let us look at a simple application of this reasoning to pricing of call options. Suppose that at time

\(^9\)complete financial markets, to be precise
CHAPTER 2. STOCHASTIC CALCULUS

2.2. APPLICATIONS TO FINANCE

$t = 0$, the price of the underlying security is $S_0$. Given that the payoff of the call option is $(S_T - K)^+$, we claim that if such a call option traded at the price $p_C$ higher than $S_0$, there would exist an arbitrage opportunity. To exploit it, at time $t = 0$, an arbitrageur would write (sell) one call option with strike $K$ and maturity $T$, receiving $p_C$ dollars, and then buy 1 share of the stock for $S_0$ dollars, which leaves $p_C - S_0 > 0$ dollars in her pocket. At the maturity, she would sell the share of the stock she bought at $t = 0$, and use the money $S_T$ to pay her obligation $(S_T - K)^+ \leq S_T$ to the holder of the option, locking in a risk-free profit of $(p_C - S_0) + (S_T - (S_T - K)^+) > 0$ dollars. We conclude that the price of a call option should be less then the spot-price of the underlying stock. Otherwise, the clever arbitrageurs would push it down. Suppose now that there exists a self-financing portfolio $\pi$ such that, starting at $x$ and investing according to $\pi$, our total wealth $W_T$ at time $T$ equals $C = (S_T - K)^+$. In that case, there would be no alternative, but to call $x$ the fair price of the call option $C$. Any other price would produce an arbitrage opportunity similar to the one in the paragraph above. Therefore, the problem of option pricing transforms into the problems of finding the capital $x$, from which the payoff of the option can be attained using an appropriate portfolio process. Formally, the arbitrage price of the option $C$ is the (unique) amount of capital $x$ such that there exists a portfolio process $\pi$ such that the wealth process $W_{x,\pi}^x t$ from (2.2.3) satisfies

$$W_{T, x, \pi}^x = C.$$ 

In that case, $\pi$ is called a replicating portfolio for the option $C$. Of course, we can speak of the price of our option at any time $t = T$, and this price at time $0 < t < T$ will be the value of the the wealth $W_{x, \pi}^x t$ under the replicating portfolio $\pi$ at time $t$. In the following two subsection, we will give two derivations for the celebrated Black-and-Scholes formula for the price of a European call. The first one will be based on an explicit solution of a partial differential equation, and the second on a trick which could eventually lead to the powerfull concept of change of measure.

2.2.4 Black and Scholes via PDEs

The PDE approach is based on two assumptions, which can be actually proved, but we shall take them for granted with no proof since it is not hard to grasp their validity intuitively and an a priori proof would require more mathematics than we have developed so far. Also, once we finish the derivation below, we can check that assumptions we put forth a posteriori. So, the assumptions are

- there exists a replicating portfolio $\pi$ and an initial wealth $x$, such that $W_{T, x, \pi}^x = (S_T - K)^+$, and
- the value of the wealth process $W_{x, \pi}^x$ is a smooth function of $S_t$ and $t$ only, i.e. $W_{x, \pi}^x = u(t, S_t)$ for some nice (differentiable enough) function $u(\cdot, \cdot)$. 

Last Updated: Spring 2005 127 Continuous-Time Finance: Lecture Notes
Without the first assumption there would be no arbitrage price of the call option $C$, and
the second one states that the price of $C$ at time $t$ depends only on $t$ and $S_t$. This second
assumption is also quite plausible, since the history of the stock prices is irrelevant to its future
evolution given the stock price today. Having made these two assumptions we immediately
see that all we need now is to determine the value of the function $u(\cdot, \cdot) : [0,t] \times \mathbb{R}^+ \to \mathbb{R},$
at the point $(t, x) = (0, S_0)$, since $u(0, S_0) = W^{x,\pi}_0 = x$. We shall see, however, that in order
to do that we have to determine the values of $u$ for all $t$ and $x$. Let’s start figuring out
how to compute the values of the function $u$.

First of all, since we know that $W^{x,\pi}_t = u(t, S_t)$
we apply the Itô formula to $u(t, S_t)$

\[
\frac{du(t, S_t)}{dt} = \frac{\partial}{\partial t}u(t, S_t) + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2}{\partial x^2}u(t, S_t) dt + \sigma S_t \frac{\partial}{\partial x}u(t, S_t) dB_t.
\]

Remember the expression for the process $W^{x,\pi}_t$

\[
dW^{x,\pi}_t = rW^{x,\pi}_t dt + \pi_t S_t \left( (\mu - r) dt + \sigma dB_t \right)
\]

and equate the terms corresponding to $dt$ and those corresponding to $dB_t$ terms to get

\[
\begin{cases}
\frac{\partial}{\partial t}u(t, S_t) + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2}{\partial x^2}u(t, S_t) + \mu S_t \frac{\partial}{\partial x}u(t, S_t) = ru(t, S_t) + \pi_t S_t (\mu - r), \\
\sigma S_t \frac{\partial}{\partial x}u(t, S_t) = \sigma \pi_t S_t.
\end{cases}
\]

Substituting the expression for $\pi_t S_t$ into the first equation we conclude that $u(\cdot, \cdot)$ should
satisfy the following partial differential equation usually called the Black and Scholes PDE:

\[
\frac{\partial}{\partial t}u(t, x) = -\frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2}u(t, x) - \mu x \frac{\partial}{\partial x}u(t, x) + ru(t, x). 
\tag{2.2.4}
\]

There are infinitely many solutions to this equation, so we need to impose an extra condition
to get a unique solution. Fortunately, we know the exact value of $u(T, x)$, by the requirement that

\[
u(T, S_T) = W^{x,\pi}_T = (S_T - K)^+, \text{ and so } u(T, x) = (x - K)^+.
\]

We have transformed a financial problem of finding a fair price of a call option $C$ to a purely
analytic problem of solving the partial differential equation (2.2.5), subject to the terminal
condition $u(T, x) = (x - K)^+$. To plan an attack on the equation (2.2.5), we start by a
simplifying substitution

\[
u(t, x) = e^{rt}v(T - t, \frac{\log(x)}{\sigma}) = e^{r(T-s)}v(s, y),
\]
with \( y = \frac{\log(x)}{\sigma}, \ s = T - t \) so that the equation for \( v \) becomes

\[
\frac{\partial}{\partial s} v(s, y) = \frac{1}{2} \frac{\partial^2}{\partial y^2} v(s, y) + \left( \frac{r}{\sigma} - \frac{\sigma}{2} \right) \frac{\partial}{\partial y} v(s, y),
\]

(2.2.5)

together with (now initial!) condition \( v(0, y) = (\exp(\sigma y) - K)^+ \). The reason for this specific substitution comes from the desire to get rid of \( x \) and \( x^2 \) in front of \( \frac{\partial}{\partial x} u(t, x) \) and \( \frac{\partial^2}{\partial x^2} u(t, x) \), as well as the term \( ru(t, x) \). By doing this we will transform our partial differential equation (2.2.5) into a standard form of a heat equation. We need another change of variables, though, to accomplish that - we have to get rid of the term involving \( \frac{\partial}{\partial y} v(s, y) \). For that, we set

\[
v(s, y) = w\left(s, y + \left( \frac{r}{\sigma} - \frac{\sigma}{2} \right) s \right) = w(s, z),
\]

with \( z = y + \left( \frac{r}{\sigma} - \frac{\sigma}{2} \right) s \), which finally reduces our PDE to the canonical form of a heat equation

\[
\frac{\partial}{\partial s} w(s, z) = \frac{1}{2} \frac{\partial^2}{\partial z^2} w(s, z),
\]

(2.2.6)

together with the initial condition

\[
w(0, z) = v(0, y + \left( \frac{r}{\sigma} - \frac{\sigma}{2} \right) \cdot 0) = \left( \exp(\sigma z) - K \right)^+.
\]

(2.2.7)

The mathematical theory of partial differential equation has a very precise answer about the solutions of the heat equation with an initial condition assigned. We will cite a general theorem without proof:

**Theorem 2.2.7** (Solution of the heat equation). Given a sufficiently nice function\(^\text{10}\)

\[
\psi : \mathbb{R} \rightarrow \mathbb{R},
\]

there exists a unique function \( w : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R} \) such that

\[
\frac{\partial}{\partial s} w(s, z) = \frac{1}{2} \frac{\partial^2}{\partial z^2} w(s, z), \text{ for } s > 0, \text{ and } z \in \mathbb{R},
\]

and

\[
w(0, z) = \psi(z), \text{ for } z \in \mathbb{R}.
\]

Furthermore, the function \( w \) is given by the following integral:

\[
w(s, z) = \int_{-\infty}^{\infty} e^{-\frac{(z - \xi)^2}{2s}} \psi(\xi) d\xi.
\]

\(^\text{10}\)all the functions appearing in finance are nice enough
Using this result we can immediately write down the solution to our equation (2.2.6) with initial condition (2.2.7):

\[ w(s, z) = \int_{-\infty}^{\infty} e^{-\frac{(s-\xi)^2}{2s}} (\exp(\sigma \xi) - K) \, d\xi = \int_{\log(K)/\sigma}^{\infty} e^{-\frac{(s-\xi)^2}{2s}} \left( \exp(\sigma \xi) - K \right) \, d\xi \]

\[ = -K \Phi\left( \frac{z\sigma - \log(K)}{\sqrt{s}\sigma} \right) + e^{z\sigma + \frac{zs^2}{2s}} \Phi\left( \frac{\sigma (z + s\sigma) - \log(K)}{\sqrt{s}\sigma} \right), \]

where \( \Phi(x) = \int_{-\infty}^{x} e^{-\frac{z^2}{2}} \, dz \). Reversing substitutions we get the following expressions

\[ v(s, y) = -K \Phi\left( \frac{2rs + 2y\sigma - s\sigma^2 - 2\log(K)}{2\sqrt{s}\sigma} \right) + e^{rs + y\sigma} \Phi\left( \frac{2rs + 2y\sigma + s\sigma^2 - 2\log(K)}{2\sqrt{s}\sigma} \right) \]

and finally

\[ u(t, x) = -e^{r(T-t)} K \Phi\left( \frac{(T-t) \left( 2r - \sigma^2 \right) - 2\log(K/x)}{2\sigma \sqrt{T-t}} \right) + x \Phi\left( \frac{(T-t) \left( 2r + \sigma^2 \right) - 2\log(K/x)}{2\sigma \sqrt{T-t}} \right). \]  

(2.2.8)

Setting \( t = 0 \), and \( x = S_0 \) - the stock-price at time \( t = 0 \), we get the famous Black and Scholes formula for the price of a European call option:

\[ u(0, S_0) = -e^{rT} K \Phi\left( \frac{T \left( 2r - \sigma^2 \right) - 2\log(K/S_0)}{2\sigma \sqrt{T}} \right) + S_0 \Phi\left( \frac{T \left( 2r + \sigma^2 \right) - 2\log(K/S_0)}{2\sigma \sqrt{T}} \right). \]

2.2.5 The “Martingale method” for pricing options

In this subsection we would like to present another approach to option pricing in Samuelson’s model. First of all, here is crucial observations about what we did in the previous subsection:

- The moment we wrote down the Black and Scholes PDE the symbol \( \mu \) disappeared without a trace. The function \( u \) and the option price do not depend on \( \mu \) at all !!!!

Since \( \mu \) does not matter, we might as well set it to 0, or to any other value we please, and still obtain the same price. We will use that to derive the Black and Scholes formula in a very elegant and quick manner\(^\text{11}\). Before we do that, let us examine the wealth equation (2.2.3) - in particular let us see what happens to it when we discount it and denominate the wealth

\(^\text{11}\)We couldn’t have arrived to the conclusion that \( \mu \) does not matter directly, without using the PDE approach, so do not dismiss it entirely. There are some other very good reason to keep the PDE approach handy. You will understand which reasons I am talking about when you read through this section.
in real instead of nominal terms. Define the process $V_t$ to be the discounted version of $W^{x,\pi}_t$; i.e. $V_t = e^{-rt}W^{x,\pi}_t$. By Itô’s formula $V_t$ is an Itô process and we have
\[
dV^{x,\pi}_t = -re^{-rt}W^{x,\pi}_t dt + e^{-rt}dW^{x,\pi}_t
\]
\[
= -re^{-rt}W^{x,\pi}_t dt + e^{-rt}\left[rW^{x,\pi}_t dt + \pi_tS_t((\mu - r) dt + \sigma dB_t)\right]
\]
\[
= e^{-rt}\pi_tS_t((\mu - r) dt + \sigma dB_t).
\]

Let us now move to a different market - the one in which $\mu = r$, and denote the stock-price there by $S^r_t$. The stochastic process $S^r_t$ is now the geometric Brownian motion given by the following SDE:
\[
dS^r_t = S^r_t r dt + S^r_t \sigma dB_t, \quad S^r_0 = S_0.
\]

In this market we have
\[
dV^{x,\pi}_t = e^{-rt}\pi_tS^r_t \sigma dB_t
\]
and so $V^{x,\pi}_t$ is a martingale, since its $dt$-term disappears ($\mu - r = 0$). Suppose now that $x$ is the arbitrage price of a contingent claim $X$ (not necessarily a call option), and a let $\pi$ be the replicating portfolio, so that that $W^{x,\pi}_T = X$, and $W^{x,\pi}_0 = x$. The process $V^{x,\pi}_t$ (the discounted wealth $W^{x,\pi}_t$) satisfies now the following
\[
\begin{align*}
& V^{x,\pi}_t \text{ is a martingale}, \\
& V^{x,\pi}_T = e^{-rT}X, \text{ and} \\
& V^{x,\pi}_0 = x,
\end{align*}
\]
so we must have
\[
x = V^{x,\pi}_0 = \mathbb{E}[V^{x,\pi}_T] = \mathbb{E}[e^{-rT}X].
\]

Thus, the arbitrage price of $X$ is given by $x = \mathbb{E}[e^{-rT}X]$ in the market in which $\mu = r$. We know, however, that the arbitrage price does not depend on the parameter $\mu$, so we have the following important conclusion:
\[
\text{the arbitrage price of a European contingent claim } X \text{ is the expectation of } e^{-rT}X \text{ in the market in which } \mu = r.
\]

Let us apply this principle to valuation of the European call option. In this case $X = (S_T - K)^+$, so the Black and Scholes price of $X$ is given by
\[
x = \mathbb{E}[e^{-rT}(S_T - K)^+].
\]
We know that $S_T^r = \exp(\sigma B_T + (r - \frac{1}{2}\sigma^2)T)$, and that $B_T$ has a normal distribution with mean 0 and variance $T$, so we have

$$x = \mathbb{E}[e^{-rT}(S_T^r - K)^+ ] = \mathbb{E}[e^{-rT}\left(\exp(\sigma B_T + (r - \frac{1}{2}\sigma^2)T) - K\right)^+]$$

$$= \int_{-\infty}^{\infty} \frac{e^{-\frac{\xi^2}{2T}}}{\sqrt{2\pi T}} \left(\exp(\sigma \xi + (r - \frac{1}{2}\sigma^2)T) - K\right)^+ d\xi$$

$$= \int_{\frac{1}{T}(\log(K) - T(r - \frac{1}{2}\sigma^2))}^{\infty} \frac{e^{-\frac{\xi^2}{2T}}}{\sqrt{2\pi T}} \left(\exp(\sigma \xi + (r - \frac{1}{2}\sigma^2)T) - K\right) d\xi,$$

and the (standard) evaluation of this integral gives the Black and Scholes formula (2.2.8).

Why is the PDE approach sometimes better? Well, the martingale approach requires you to calculate expectations, which is often faster than solving a PDE. However, not all expectations can be evaluated explicitly (try Asian options!) and we have to resort to numerical techniques and use PDEs.
Appendix A

Odds and Ends
A.1 No \( \mathbb{P} \) for all Subsets of \([0, 1]\)

The following example shows a mind-bending phenomenon - and elucidates the necessity of the nuisance of having probabilities defined on \( \sigma \)-algebras. In order to proceed we have to state an axiom of set theory.

**Axiom of Choice.** Given any collection \( \mathcal{C} \) of nonempty and pairwise disjoint sets, there exists a set containing precisely one element from each set in \( \mathcal{C} \).

**Example A.1.1.** Let us define an equivalence relation \( \sim \) on \( \mathbb{R} \), by stating that \( x \sim y \) if and only if \( y - x \) is a rational number. This equivalence relation will break the set of real numbers \( \mathbb{R} \) into a collection of disjoint nonempty equivalence classes, and by the Axiom of Choice, there exists a set \( A \subseteq \mathbb{R} \) such that \( A \) contains exactly one representative from each class. For each rational number \( q \in \mathbb{Q} \) we can define the translation \( A_q \) of \( A \) by \( A_q \triangleq \{(q+a) \mod 1 : a \in A\} \). Here \( (q+a) \mod 1 \) denotes the fractional part of \( q + a \) (e.g. \( 7.14 \mod 1 = 0.14, -0.92 \mod 1 = 0.08 \)). Convince yourself that \( A_q \) constitutes a countable partition of \([0, 1]\):

\[
\bigcup_{q \in \mathbb{Q}} A_q = \mathbb{R}, \quad A_q \cap A_p = \emptyset, \text{ for } p, q \in \mathbb{Q}, \quad q \neq q.
\]

Suppose there is a probability \( \mathbb{P} \) on all the subsets of \([0, 1]\) such that \( \mathbb{P}[A] = \mathbb{P}[A + x] \), for any \( A \subseteq [0, 1] \) such that \( A + x \subseteq [0, 1] \). We first note that \( \mathbb{P}[A_q] = \mathbb{P}[A_p] \) for any \( p, q \in \mathbb{Q} \).

Can se have the common value of \( \mathbb{P}[A_p] \), \( p \in \mathbb{Q} \) strictly positive, say \( \varepsilon > 0 \)? Surely not! The countable additivity of \( \mathbb{P} \) dictates

\[
1 = \mathbb{P}[0, 1] = \mathbb{P}[\bigcup_{q \in \mathbb{Q}} A_q] = \sum_{q \in \mathbb{Q}} \mathbb{P}[A_q] = \sum_{q \in \mathbb{Q}} \varepsilon = \infty.
\]

So, we must have \( \mathbb{P}[A_p] = 0 \) for all \( p \in \mathbb{Q} \), but we are no better off than before:

\[
1 = \mathbb{P}[0, 1] = \mathbb{P}[\bigcup_{q \in \mathbb{Q}} A_q] = \sum_{q \in \mathbb{Q}} \mathbb{P}[A_q] = \sum_{q \in \mathbb{Q}} 0 = 0.
\]

A contradiction again. Sherlock Holmes said: “When you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth”, so we must agree that there is simply no nice (translationally invariant) probability on \([0, 1]\)
A.2 Marginals: Normal, Joint: Not

In this section we give two examples of bivariate random vectors $X = (X_1, X_2)$ such that $X_1$ and $X_2$ are normally distributed, but the random vector $X$ is not. In the sequel we use the following shorthand:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right),$$

for the density of the standard normal.

**Example A.2.1.** Let $h : \mathbb{R} \to \mathbb{R}$ be any nontrivial function such that

1. $|h(x)| \leq (2\pi e)^{-1/2}$, for all $x \in \mathbb{R}$.
2. $h(x) = 0$, for $x \not\in [-1, 1]$.
3. $h(-x) = -h(x)$, for all $x \in \mathbb{R}$.

Define the random vector $X = (X, Y)$ to have the density

$$f_X(x, y) = \varphi(x)\varphi(y) + h(x)h(y).$$

Conditions 1. and 2. ensure that $f_X \geq 0$, and the integral of $f_X$ over $\mathbb{R} \times \mathbb{R}$ is equal to 1 due to condition 3. Thus, $f_X$ is a true density function. Also, $f_X$ is not a density function of a bivariate normal distribution for no nontrivial $h$ (why?). Let us show that the marginals of $f_X$ are standard normal. We compute

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x, y) \, dy = \int_{-\infty}^{\infty} \varphi(x)\varphi(y) \, dy + \int_{-\infty}^{\infty} h(x)h(y) \, dy$$

$$= \varphi(x) \int_{-\infty}^{\infty} \varphi(y) \, dy + h(x) \int_{-\infty}^{\infty} h(y) \, dy = \varphi(x) \cdot 1 + h(x) \cdot 0 = \varphi(x),$$

and the last two integrals are equal to 1 and 0, respectively due to the facts that $\varphi$ is a density function, and that $h$ is odd ($h(-y) = -h(y)$). By symmetry we would conclude that $f_Y(y) = \varphi(y)$.

**Example A.2.2.** In the following example I will only define the candidate density function and leave it for you as an exercise to prove that it in fact defines a random vector whose marginals are normal, but the joint distribution is not.

$$f_X(x, y) = \begin{cases} 2\varphi(x)\varphi(y), & xy \geq 0 \\ 0, & xy < 0. \end{cases}$$
A.3  Mersenne Twister

In this appendix I am including a C-code for the Mersenne Twister written by M. Matsumoto and T. Nishimura, one of the best free random number generators (its order is 624 and period $2^{19937} - 1$). The interesting thing is that its good properties stem from some deep results in number theory (hence the name). For more info, check the article referenced below. More info (including the source code) is available from ttp://www.mat.keio.ac.jp/matumoto/emt.html+

/* A C-program for MT19937: Real number version([0,1)-interval) */
/* (1999/10/28) */
/* genrand() generates one pseudorandom real number (double) */
/* which is uniformly distributed on [0,1)-interval, for each */
/* call. sgenrand(seed) sets initial values to the working area */
/* of 624 words. Before genrand(), sgenrand(seed) must be */
/* called once. (seed is any 32-bit integer.) */
/* Integer generator is obtained by modifying two lines. */
/* Coded by Takuji Nishimura, considering the suggestions by */
/* Topher Cooper and Marc Rieffel in July-Aug. 1997. */

/* This library is free software; you can redistribute it and/or */
/* modify it under the terms of the GNU Library General Public */
/* License as published by the Free Software Foundation; either */
/* version 2 of the License, or (at your option) any later */
/* version. */
/* This library is distributed in the hope that it will be useful, */
/* but WITHOUT ANY WARRANTY; without even the implied warranty of */
/* MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. */
/* See the GNU Library General Public License for more details. */
/* You should have received a copy of the GNU Library General */
/* Public License along with this library; if not, write to the */
/* Free Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA */
/* 02111-1307 USA */

/* Copyright (C) 1997, 1999 Makoto Matsumoto and Takuji Nishimura. */
/* Any feedback is very welcome. For any question, comments, */
/* see http://www.math.keio.ac.jp/matumoto/emt.html or email */
/* matumoto@math.keio.ac.jp */
#include<stdio.h>

/* Period parameters */
#define N 624
#define M 397
#define MATRIX_A 0x9908b0df /* constant vector a */
#define UPPER_MASK 0x80000000 /* most significant w-r bits */
#define LOWER_MASK 0x7fffffff /* least significant r bits */

/* Tempering parameters */
#define TEMPERING_MASK_B 0x9d2c5680
#define TEMPERING_MASK_C 0xefc60000
#define TEMPERING_SHIFT_U(y) (y >> 11)
#define TEMPERING_SHIFT_S(y) (y << 7)
#define TEMPERING_SHIFT_T(y) (y << 15)
#define TEMPERING_SHIFT_L(y) (y >> 18)

static unsigned long mt[N]; /* the array for the state vector */
static int mti=N+1; /* mti==N+1 means mt[N] is not initialized */

/* Initializing the array with a seed */
void
sgenrand(seed)
    unsigned long seed;
{
    int i;

    for (i=0;i<N;i++) {
        mt[i] = seed & 0xffff0000;
        seed = 69069 * seed + 1;
    }

A.3. MERSENNE TWISTER

```c
mt[i] |= (seed & 0xffffff0000) >> 16;
seed = 69069 * seed + 1;
}
mti = N;
}
/* Initialization by "sgenrand()" is an example. Theoretically, */
/* there are 2^19937-1 possible states as an intial state. */
/* This function allows to choose any of 2^19937-1 ones. */
/* Essential bits in "seed_array[]" is following 19937 bits: */
/* (seed_array[0]&UPPER_MASK), seed_array[1], ..., seed_array[N-1]. */
/* (seed_array[0]&LOWER_MASK) is discarded. */
/* Theoretically, */
/* (seed_array[0]&UPPER_MASK), seed_array[1], ..., seed_array[N-1] */
/* can take any values except all zeros. */
void
lsgenrand(seed_array)
    unsigned long seed_array[];
    /* the length of seed_array[] must be at least N */
{
    int i;

    for (i=0;i<N;i++)
        mt[i] = seed_array[i];
    mti=N;
}

double /* generating reals */
/* unsigned long */ /* for integer generation */
genrand()
{
    unsigned long y;
    static unsigned long mag01[2]={0x0, MATRIX_A};
    /* mag01[x] = x * MATRIX_A for x=0,1 */

    if (mti >= N) { /* generate N words at one time */
        int kk;
    }
```
if (mti == N+1) /* if sgenrand() has not been called, */
    sgenrand(4357); /* a default initial seed is used */
for (kk=0; kk<N-M; kk++) {
    y = (mt[kk]&UPPER_MASK)|(mt[kk+1]&LOWER_MASK);
    mt[kk] = mt[kk+M] ^ (y >> 1) ^ mag01[y & 0x1];
}
for (; kk<N-1; kk++) {
    y = (mt[kk]&UPPER_MASK)|(mt[kk+1]&LOWER_MASK);
    mt[kk] = mt[kk+(M-N)] ^ (y >> 1) ^ mag01[y & 0x1];
}
y = (mt[N-1]&UPPER_MASK)|(mt[0]&LOWER_MASK);
mt[N-1] = mt[M-1] ^ (y >> 1) ^ mag01[y & 0x1];

mti = 0;
}
y = mt[mti++];
y ^= TEMPERING_SHIFT_U(y);
y ^= TEMPERING_SHIFT_S(y) & TEMPERING_MASK_B;
y ^= TEMPERING_SHIFT_T(y) & TEMPERING_MASK_C;
y ^= TEMPERING_SHIFT_L(y);

return ( (double)y * 2.3283064365386963e-10 ); /* reals: [0,1)-interval */
/* return y; */ /* for integer generation */

/* This main() outputs first 1000 generated numbers. */
main()
{
    int i;

    sgenrand(4357);
    for (i=0; i<1000; i++) {
        printf("%10.8f ", genrand());
        if (i%5==4) printf("\n");
    }
}
A.4 Conditional Expectation Cheat-Sheet

Here are the 7 properties of the conditional expectation:

(CE1) $\mathbb{E}[X|\mathcal{F}]$ is a random variable, and it is measurable with respect to $\mathcal{F}$, i.e. $\mathbb{E}[X|\mathcal{F}]$ depends on the state of the world, but only through the information contained in $\mathcal{F}$.

(CE2) $\mathbb{E}[X|\mathcal{F}](\omega) = \mathbb{E}[X]$, for all $\omega \in \Omega$ if $\mathcal{F} = \{\emptyset, \Omega\}$, i.e. the conditional expectation reduces to the ordinary expectation when you have no information.

(CE3) $\mathbb{E}[X|\mathcal{F}](\omega) = X(\omega)$, if $X$ is measurable in $\mathcal{F}$, i.e. there is no need for expectation when you already know the answer (the value of $X$ is known, when you know $\mathcal{F}$.)

(CE4) When $X$ is a random variable and $\mathcal{F}, \mathcal{G}$ two $\sigma$-algebras such that $\mathcal{F} \subseteq \mathcal{G}$, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}],$$

i.e. my expectation (given information $\mathcal{F}$) of the value $X$ is equal to the expectation (given information $\mathcal{F}$) of the expectation I would have about the value of $X$ if I had information $\mathcal{G}$ on top of $\mathcal{F}$.

(CE5) The conditional expectation is linear, i.e. for $X_1, X_2$ random variables, $\alpha_1, \alpha_2$ real constants, and a $\sigma$-algebra $\mathcal{F}$, we have

$$\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2|\mathcal{F}] = \alpha_1 \mathbb{E}[X_1|\mathcal{F}] + \alpha_2 \mathbb{E}[X_2|\mathcal{F}].$$

(CE6) Let $\mathcal{F}$ be a $\sigma$-algebra, and let $X$ and $Y$ be variables such that $Y$ is measurable with respect to $\mathcal{F}$. Then

$$\mathbb{E}[X Y|\mathcal{F}](\omega) = Y(\omega) \mathbb{E}[X|\mathcal{F}](\omega),$$

i.e. the random variable measurable with respect to the available information can be treated as a constant in conditional expectations.

(CE7) If the random variable $X$ and the $\sigma$-algebra $\mathcal{F}$ are independent, then

$$\mathbb{E}[X|\mathcal{F}](\omega) = \mathbb{E}[X],$$

i.e. conditioning with respect to independent information is like conditioning on no information at all.