Stochastic Calculus for Mathematical Finance

Frederic Utzet

Departament of Mathematics
Universitat Autònoma de Barcelona

January 2016

Contents

Foreword

This is an introductory course on Mathematical Finance where the emphasis is on probabilistic methods. Advanced stochastic processes are the basis of Mathematical Finance: Martingale theory, Markov processes, Itô formula and stochastic differential equations, etc. are widely used in this discipline. Hence, the more probability and stochastic process the student knows, the more easily and deeply will she/he understand Mathematical Finance. The ideal would be that before studying Mathematical Finance (or at the same time), the student takes a course on stochastic processes; but this is normally unrealistic. For this reason, I tried to cover the jump from elementary to advanced courses of probability introducing the probabilistic tools that are needed, but using as few technicalities as possible. This course could be complemented in two ways. First, the presentation of the financial world here is not comprehensive: I have chosen some elements of Finance that I consider interesting, representative and illustrative. However many topics are missing and I recommend the study of some standard book on Finance such as Hull [8]. On the other hand, many mathematical details (in particular, the majority of proofs) have been skipped over. The interested reader, should look at some more advanced books such as Shreve [19, 20] or Steele [21].

Introduction: Arbitrages and fair games

1. Arbitrages

One of the essential words in finance is **arbitrage**, that colloquially means to earn money without risking anything, and it is also called a *free lunch*. For example, a person looks carefully at the stock exchange market in Barcelona and Frankfurt where shares of Bayer are traded. One day she realizes that to buy a share in Barcelona costs € 52.44 (*ask*) and in Frankfurt the shares are bought for € 52.89(*bid*). So if she has fast access to buying and selling shares, and the transaction costs are low, then she can buy in Barcelona and sell in Frankfurt, and make money. Indeed, currently a private person has no such opportunities, because the banks, the *Hedge Funds*, etc. have powerful computers looking for and executing arbitrage opportunities 24 hours at day; this is part of the so-called algorithmic trading and, as it is well known, it can produce moments of panic in the stock exchanges. The people devoted to executing arbitrages are called *arbs*, and besides their own enrichment, they play an important role since they remove some imperfections from the market, and then the market is a *fair game* for other people.

2. Arbitrages under a random setup

Besides the simple arbitrage that we commented on above, in general an arbitrage depends on random factors. Thus, an arbitrage is a trading strategy that, on the one hand, ensures that the *arb* does not loss anything, and on the other hand, with nonzero probability she can win something. We study a very simple example: the roulette game. The modern version of roulette was designed in the XVII century by the French mathematician and philosopher Blaise Pascal. In Pascal's roulette there were 36 betting spaces, numbered from 1 to 36, and the possibility of different bets. See Figure 1.

For example, we can bet a coin on a certain number, say number 15, and if the result of the wheel is 15, then we recover the bet and gain 35 coins. Denote by C_0 our initial capital and by C_1 the capital after the game. See the possible results in Figure 2.

Another possibility is to bet on two numbers, say numbers 1 and 2; if the result is one of these numbers, then we recover the bet and receive 17 coins. So our capital C_1 is either 18 or 0. In the same way, there are bets on 4 numbers, 6 numbers, etc, but it must be a divisor of 36. The simplest bet is on odd or even, or black or red (18 numbers each), and the gain is one coin (and we recover our bet).

The talent of Pascal is to exhibit the fact that roulette is a **fair game**: if a gambler plays many times, some times she will win, and other times she will lose, but when the afternoon is over (assuming a large enough afternoon, like those of the XVII century) approximately the gambler's capital will have not changed; she will just have passed the afternoon. For example, the bet on 15 has the expectation

$$\mathbf{E}[C_1] = 36 \times \frac{1}{36} + 0 \times \frac{35}{36} = 1 = C_0.$$

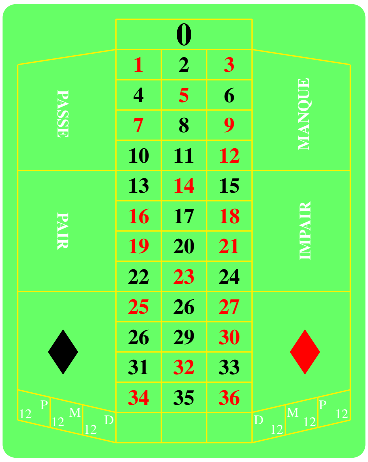


Figure 1. Current roulette table; in the Pascal's one there was no zero pocket

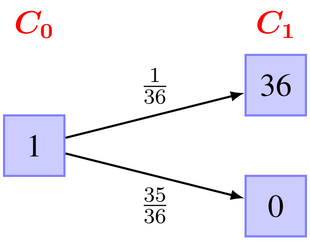


Figure 2. A bet of 1 coin to number 15 in Pascal's roulette

A similar computation can be performed with all other bets.

When roulette was exported to the casinos, they made some modifications in order to ensure that the casino always get some money; casinos are a business, not a philanthropic exercise. At the end of the XIX century, the director of the Hamburg casino, whose name was Mr. Blanc, modified roulette by adding the number 0 but keeping the same remuneration for the bets; for example a bet for a number receives 35 times the bet plus the bet. The inclusion of the number 0 has two effects: the first is that unbalances the bets: the play is unfair; second, the 0 is neither even nor odd, neither red nor black, so that when the result is number 0, many bets are lost. So the odds are now is against the player. For example, the bet of one coin on 15 has expectation

$$\mathbf{E}[C_1] = \frac{36}{37} = 0.973 < C_0.$$

With the inclusion of 0, every bet is favorable to the casino. The legend says that Mr. Blanc walked between the roulette's tables and with a happy face whispered "jouez noire, jouez rouge, Blanc toujours gagne". We should stress that to accelerate the gains, the American casinos added a double 0; it is really remarkable that gamblers should pay an entry ticket to play in an unfair game.

As well as the enrichment of the casino that is guaranteed by the unfair game, a good croupier is interested in the bets on the table being balanced; for example, in the simplest case, that there are the same quantity of bets on odd numbers as on even numbers; in this way the croupier does not need any money to pay the winning bets. But in that situation, the perfect result for the casino is that the ball falls in number 0. Then the casino wins all the money. That is

- If the result is a number between 1 and 36, the casino wins and loses nothing.

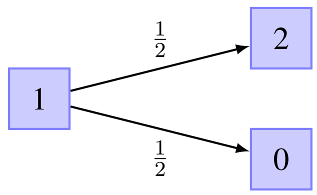


Figure 3. A bet of 1 coin to odd numbers in Pascal's roulette

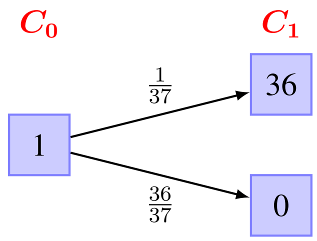


Figure 4. Bet to number 15 with Blanc's roulette

- If the result is 0, the casino wins all bets in the table.

Under this condition, the casino is doing an arbitrage: it never loses, and with some probability it makes profit. As the reader can notice the possibility of arbitrage is due to the existence of the number 0, which makes the game unfair.

3. The stock exchange

The stock exchange can be considered as a game infinitely more complex than roulette. There are many different bets: you can bet on both sides of the game, you can play as a banc, etc. The most essential thing is what is called the Arbitrage Theorem, which says that if we compute the prices of the stocks with a probability such that the dynamics of the prices are fair, then there is no possibility of arbitrage. In this course we will study these kinds of results .

4. Dynamical bets: martingales

We have commented that if C_0 denotes the initial capital of a player and i C_1 her capital after a play, then it is said that the game is **fair** if

$$\mathbf{E}[C_1] = C_0.$$

However, the model is too simple even for a game in a casino, where some strategies like "double the bet if you lost" are usual. Hence we need to consider a stochastic process $C_0,\ C_1,\ C_2, \dots,$ as a model of the players capital, and the definition of of dynamic fair game is the notion of martingale, that says that

$$\mathbf{E}[C_{n+1} \,|\, C_0, C_1, \dots, C_n] = C_n,\ n \geq 0,$$

where the expression of the right hand side is a **conditional expectation**.

In order to study such a dynamical model, we need to start the course by the definition and properties of conditional expectation.

Chapter 1

Conditional probability and expectation. First look

In this chapter we review some basics concepts on conditional probability and conditional expectation, and in Chapter 3 we rewrite them in a general form needed to understand the main topics of Mathematical Finance.

1.1 Introductory example

As a simple example for the behavior of the price of a stock or commodity we can consider the following dynamics. Denote by $S_0 = 20$ € today's price, which is known; the price for tomorrow is unknown, but we assume that it can take just two values: $S_1 = 24$ € with probability $1/4$, or 16 € with probability $3/4$:

$$\mathsf{P}(S_1 = 24) = \frac{1}{4} \quad \text{and} \quad \mathsf{P}(S_1 = 16) = \frac{3}{4}$$

The prices for the day after tomorrow have probabilities that depend on the price of tomorrow, and so they are also random. Assume that they are:

$$\begin{aligned}
\mathsf{P}(S_2 = 28.8 \,|\, S_1 = 24) &= \frac{1}{4}, \\
\mathsf{P}(S_2 = 19.2 \,|\, S_1 = 24) &= \frac{3}{4}, \\
\mathsf{P}(S_2 = 19.2 \,|\, S_1 = 16) &= \frac{1}{4}, \\
\mathsf{P}(S_2 = 12.8 \,|\, S_1 = 16) &= \frac{3}{4}.
\end{aligned}$$

We can sketch the price's dynamics in the tree given in Figure 1.1.

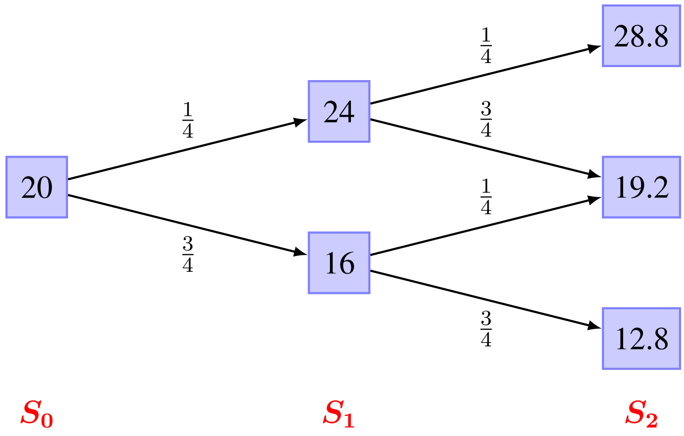


Figure 1.1. Binomial model for a price dynamics

Remark. Although S_0 is deterministic, that is, always $S_0 = 20$, some times it is convenient to write

$$\mathbf{P}(S_1 = 24) = \mathbf{P}(S_1 = 24 \mid S_0 = 20) = \frac{1}{4},$$

or

$$\mathbf{P}(S_2 = 19.2 \mid S_1 = 24) = \mathbf{P}(S_2 = 19.2 \mid S_0 = 20,\ S_1 = 24) = \frac{1}{4},$$

in order to handle more symmetric notations.

The information about the probabilities related to S_1 can be summarized in Table 1.1.

S_1	$\mathbf{P}(S_1 = \cdot)$
24	1/4
16	3/4

Table 1.1. Distribution of S_1

From table Table 1.1 we can compute the expectation of S_1:

$$\mathbf{E}[S_1] = 24 \times \frac{1}{4} + 16 \times \frac{3}{4} = 18.$$

Or its variance

$$\mathrm{Var}(S_1) = \mathbf{E}\left[\left(S_1 - 18\right)^2\right] = (24 - 18)^2 \times \frac{1}{4} + (16 - 18)^2 \times \frac{3}{4} = 12.$$

Or, in general, the expectation of an arbitrary function of S_1, $h(S_1)$:

$$\mathbf{E}[h(S_1)] = h(24)\,\frac{1}{4} + h(18)\,\frac{3}{4}. \tag{1.1}$$

For example, for $h(x) = x^4$ we have

$$\mathbf{E}[S_1^4] = 24^4 \times \frac{1}{4} + 18^4 \times \frac{3}{4} = 161676.$$

1.1.1 Conditional probabilities

For S_2 there are two tables of conditional probabilities, depending on the value of S_1. See Table 1.2

S_2	$\mathsf{P}(S_2 = \cdot \mid S_1 = 24)$
28.8	1/4
19.2	3/4

S_2	$\mathsf{P}(S_2 = \cdot \mid S_1 = 16)$
19.2	1/4
12.8	3/4

Table 1.2. Conditional probabilities of S_2 depending on S_1

From the conditional probabilities of S_2 in Table 1.2, we can deduce the unconditional probabilities of S_2: By the formula of the total probabilities,

$$\mathsf{P}(S_2 = 28.8) = \mathsf{P}(S_1 = 24)\mathsf{P}(S_2 = 28.8 \,|\, S_1 = 24) + \mathsf{P}(S_1 = 16)\mathsf{P}(S_2 = 28.8 \,|\, S_1 = 16) = 0.0625.$$

In a similar way we can complete the Table 1.3.

S_2	$\mathsf{P}(S_2 = \cdot)$
28.8	0.0625
19.2	0.375
12.8	0.5625

Table 1.3. (Unconditional) Probability table of S_2

Table 1.2 is a direct translation of Figure 1.1, which is a typical way to assign probabilities on Mathematical Finance. However, it is worth to remember that conditional probabilities are linked with the two-dimensional random vector (S_1, S_2) by the formula

$$\mathsf{P}(S_1 = 24,\ S_2 = 28.8) = \mathsf{P}(S_1 = 24)\,\mathsf{P}(S_2 = 28.8 \,|\, S_1 = 24) = \frac{1}{4} \times \frac{1}{4} = 0.0625.$$

Computing the other values we have a two–dimensional Table 1.4.

		S_2		
		28.8	19.2	12.8
S_1	24	0.0625	0.1875	0
	16	0	0.1875	0.5625

Table 1.4. Distribution of the two–dimensional vector (S_1, S_2)

From Table 1.4, adding by rows or columns we can recover the marginal tables (unconditional probabilities) of S_1 and S_2 which are, respectively, in Tables 1.1 and 1.3.

1.1.2 Conditional expectation

From Table1.2, we can compute the conditional expectation of S_2:

$$\mathbf{E}[S_2 \mid S_1 = 24] = 28.8 \times \frac{1}{4} + 19.2 \times \frac{3}{4} = 21.6,$$

and

$$\mathbf{E}[S_2 \mid S_1 = 16] = 19.2 \times \frac{1}{4} + 12.8 \times \frac{3}{4} = 14.4.$$

In a similar way that the expectation $\mathbf{E}[S_1] = 18$ gives us a one–number summary of the random variable S_1, the conditional expectations $\mathbf{E}[S_2 \mid S_1 = 24]$ and $\mathbf{E}[S_2 \mid S_1 = 16]$ gives a summary of S_2 as a function of the random variable S_1. So, we have a new random variable that we will write

$$\mathbf{E}[S_2 \mid S_1] : \Omega \longrightarrow \mathbb{R}$$

that takes the values

$$\boxed{\mathbf{E}[S_2 \mid S_1] = \begin{cases} 21.6, & \text{if } S_1 = 24, \\ 14.4, & \text{if } S_1 = 16. \end{cases}} \tag{1.2}$$

Or in a more compact notation,

$$\boxed{\mathbf{E}[S_2 \mid S_1] = 21.6 \ \mathbf{1}_{\{S_1=24\}} + 14.4 \ \mathbf{1}_{\{S_1=16\}} = 21.6 \ \mathbf{1}_{\{24\}}(S_1) + 14.4 \ \mathbf{1}_{\{16\}}(S_1),} \tag{1.3}$$

where

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

In some circumstances it is convenient to give an explicit expression of the distribution of the conditional random variable $\mathbf{E}[S_2 \mid S_1]$. In this particular case, we can deduce from (1.2) that

$$\boxed{\mathbf{E}[S_2 \mid S_1] = \begin{cases} 21.6, & \text{with probability 1/4}, \\ 14.4, & \text{with probability 3/4}. \end{cases}}$$

or written in a usual table,

$\mathbf{E}[S_2 \mid S_1]$	$\mathbf{P}\big(\mathbf{E}[S_2 \mid S_1] = \cdot\big)$
21.6	1/4
14.4	3/4

Table 1.5. Distribution of $\mathbf{E}[S_2 \mid S_1]$

We can compute, for example, the expectation of $\mathbf{E}[S_2 \mid S_1]$:

$$\mathbf{E}\big[\mathbf{E}[S_2 \mid S_1]\big] = 21.6 \times \frac{1}{4} + 14.4 \times \frac{3}{4} = 18.$$

It is not a casuality that $\mathbf{E}[S_2] = \mathbf{E}\big[\mathbf{E}[S_2 \mid S_1]\big]$.

Comment. (*Important.*) The preceding computations can be done more transparent if we specify a (discrete) probability space coherent with this model. Take

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\},$$

that represents the four possible paths from S_0 to S_2, see Figure 1.2.

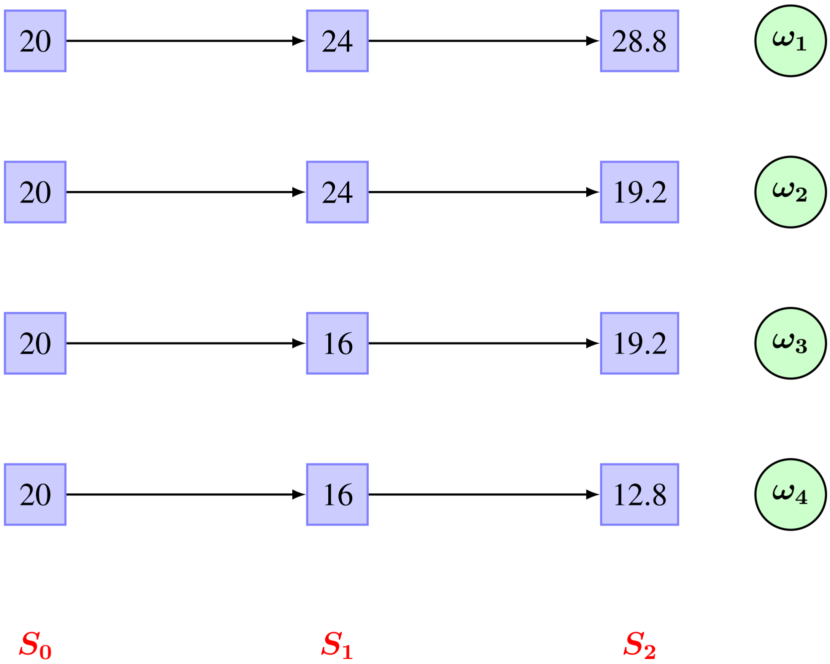


Figure 1.2. The elementary events of Ω

The probabilities of the elementary events are

$$\mathbf{P}(\omega_1) = 0.0625, \quad \mathbf{P}(\omega_2) = \mathbf{P}(\omega_3) = 0.1875 \quad \text{and} \quad \mathbf{P}(\omega_4) = 0.5625.$$

Then

$$S_0(\omega_i) = 20, \ i = 1, \ldots, 4,$$

$$S_1(\omega_1) = S_1(\omega_2) = 24, \quad \text{and } S_1(\omega_3) = S_1(\omega_4) = 16,$$

and so on. Moreover,

$$\mathbf{E}[S_2 \,|\, S_1](\omega_1) = \mathbf{E}[S_2 \,|\, S_1](\omega_2) = 21.6$$

and

$$\mathbf{E}[S_2 \,|\, S_1](\omega_3) = \mathbf{E}[S_2 \,|\, S_1](\omega_4) = 14.4.$$

From Table 1.3 we deduce

$$\mathbf{E}[S_2] = 16.2$$

Alternatively, we could use the formula of total expectations:

$$\mathbf{E}[S_2] = \mathbf{P}(S_1 = 24)\mathbf{E}(S_2 = 28.8 \,|\, S_1 = 24) + \mathbf{P}(S_1 = 16)\mathbf{E}(S_2 = 28.8 \,|\, S_1 = 16) = 16.2.$$

Let us consider the random variable $(S_2 - 19)^+$, that means, we consider the function $h(S_2)$ where

$$h(x) = (x-19)^+ = \begin{cases} x - 19, & \text{if } x \geq 19, \\ 0, & \text{otherwise.} \end{cases}$$

Of course, this corresponds to the payoff of an European call option, but for the moment we do not need any property of options. The expectation of $(S_2 - 19)^+$ is computed using formula (1.1) for S_2 with the unconditional probabilities given in Table 1.3:

$$\mathbf{E}\big[(S_2 - 19)^+\big] = 9.8 \cdot 0.0625 + 0.2 \cdot 0.375 + 0 \cdot 0.5625 = 0.6875.$$

Similarly, from Table 1.2 and a version of formula (1.1) for conditional expectations we obtain

$$\mathbf{E}\big[(S_2 - 19)^+ \,|\, S_1 = 24\big] = 9.8 \cdot 0.25 + 0.2 \cdot 0.75 = 2.6$$

and

$$\mathbf{E}\big[(S_2 - 19)^+ \,|\, S_1 = 16\big] = 0.05.$$

Equivalently,

$$\mathbf{E}\big[(S_2 - 19)^+ \,|\, S_1\big] = 2.6\ \mathbf{1}_{\{S_1 = 24\}} + 0.05\ \mathbf{1}_{\{S_1 = 16\}}$$

1.2 Discrete random variables

Let X and Y two discrete random variables that take values, respectively, in the finite or enumerable sets C_X and C_Y (in general, we will assume $\mathbf{P}\{X = x\} > 0,\ \forall x \in C_X$, and $\mathbf{P}\{Y = y\} > 0,\ \forall y \in C_Y$). In the introductory example,

$$X = S_2 \qquad \text{and} \qquad C_X = \{28.8,\ 19.2,\ 12.8\},$$

and

$$Y = S_1 \qquad \text{and} \qquad C_Y = \{24,\ 16\}.$$

Assume that we know perfectly the link between X and Y. This information can be given in one of the two forms

1. There are given the **joint probabilities**

$$\mathbf{P}(X = x, Y = y), \quad \forall (x, y) \in C_X \times C_Y.$$

Usually, this information is given in a two–dimensional table like Table 1.4.

2. For every $\boldsymbol{y} \in \boldsymbol{C_Y}$, there are given the **conditional probabilities**

$$\mathbf{P}\{X = x \,|\, Y = y\}, \quad x \in C_X.$$

Such information is given through a graphic as Figure 1.1 or a set of tables (one for each value of y) as in Table 1.2.

Recall that the link between joint and conditional probabilities is given by

$$\boxed{\mathbf{P}\{X = x \,|\, Y = y\} = \frac{\mathbf{P}\{X = x, Y = y\}}{\mathbf{P}\{Y = y\}}.}$$

Also remember that the fact that X and Y are independent is equivalent to

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\,\mathbf{P}(Y = y), \quad \forall (x, y) \in C_X \times C_Y,$$

or

$$\mathbf{P}\{X = x \,|\, Y = y\} = \mathbf{P}(X = x), \quad \forall (x, y) \in C_X \times C_Y.$$

Anyway, we can define the conditional expectation of X given $Y = y$ by

$$\mathbf{E}[X \,|\, Y = y] = \sum_{x \in C_X} x\,\mathbf{P}\{X = x \,|\, Y = y\},$$

assuming

$$\sum_{x \in C_X} |x|\,\mathbf{P}\{X = x \,|\, Y = y\} < \infty. \tag{1.4}$$

Assume that X has finite conditional expectation; this implies that (1.4) is true for all $y \in C_Y$. Hence, we can define a new random variable called ***conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{Y}$***, denoted $\mathbf{E}[\boldsymbol{X} \mid \boldsymbol{Y}]$, by

$$\mathbf{E}[X \,|\, Y](\omega) = \mathbf{E}[X \,|\, Y = y], \quad \text{if} \quad Y(\omega) = y.$$

This expression can be written as

$$\mathbf{E}[X \,|\, Y] = \sum_{y \in C_Y} \mathbf{E}[X \,|\, Y = y]\ \mathbf{1}_{\{Y = y\}},$$

or

$$\mathbf{E}[X \,|\, Y] = \sum_{y \in C_Y} \mathbf{E}[X \,|\, Y = y]\ \mathbf{1}_{\{y\}}(Y).$$

It is proved that

The random variable $\mathbf{E}[X \mid Y]$ verifies

(i) It is a function of Y:

$$\mathbf{E}[X \mid Y] = g(Y).$$

(ii) For every subset $B \subset C_Y$,

$$\mathbf{E}\big[\,\mathbf{1}_{\{Y\in B\}}\,X\big] = \mathbf{E}\big[\,\mathbf{1}_{\{Y\in B\}}\,\mathbf{E}[X \mid Y]\big],$$

or equivalently,

$$\int_{\{Y\in B\}} X\,d\mathbf{P} = \int_{\{Y\in B\}} \mathbf{E}[X \mid Y]\,d\mathbf{P}.$$

Property (i) is clear; see, for example, (1.2).

The equality in (ii) is a bit cryptic. The random variable $\mathbf{1}_{\{Y\in B\}}X$ takes the values

$$\mathbf{1}_{\{Y\in B\}}(\omega)X(\omega) = \begin{cases} X(\omega), & \text{if } Y(\omega)\in B \\ 0, & \text{otherwise.} \end{cases}$$

Then (ii) means that given B, the mean value of X restricted to a subset of ω such that $Y(\omega) \in B$ is the same as the mean value of the variable $\mathbf{E}[X \mid Y]$ on the same set of ω. For the introductory example, with $X = S_2$ and $Y = S_1$, in (ii) above take $B = \{24\}$. Then $S_2\ \mathbf{1}_{\{S_1=24\}}$ can take the values 28.8, 19.2 and 0, with probabilities

$$\mathbf{P}\big(S_2\ \mathbf{1}_{\{S_1=24\}} = 28.8\big) = \mathbf{P}\big(S_2 = 28.8, S_1 = 24\big) = 0.0625,$$

where this last value has been obtained from Table 1.4. In a similar way, we can complete the table Hence,

$S_2\ \mathbf{1}_{\{S_1=24\}}$	$\mathbf{P}\big(S_2\ \mathbf{1}_{\{S_1=24\}} = \cdot\big)$
28.8	0.0625
19.2	0.1875
0	0.75

Table 1.6. Distribution of $S_2\ \mathbf{1}_{\{S_1=24\}}$

$$\mathbf{E}\big[\,\mathbf{1}_{\{Y\in B\}}\,X\big] = 28.8\cdot 0.0625 + 19.2\cdot 0.1875 + 0\cdot 0.75 = 5.4$$

To compute the other side of (ii), from (1.3) we deduce that $\mathbf{E}[S_2 \mid S_1]\ \mathbf{1}_{\{S_1=24\}}$ can take the values 21.6 and 0, with probabilities,

$$\mathbf{P}\big(\mathbf{E}[S_2 \mid S_1]\ \mathbf{1}_{\{S_1=24\}} = 21.6\big) = \mathbf{P}(S_1 = 24) = \frac{1}{4},$$

and

$$\mathbf{P}\big(\mathbf{E}[S_2 \mid S_1]\ \mathbf{1}_{\{S_1=24\}} = 0\big) = 1 - \frac{1}{4} = \frac{3}{4}.$$

And (ii) follows.

Another point of view. (*The reader may skip this paragraph.*) We repeat the previous computations with the help of the space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. It is easy to compute

$$\mathbf{E}\big[S_2 \ \mathbf{1}_{\{S_1=24\}}\big] = \sum_{i=1}^{4} \big(S_2 \ \mathbf{1}_{\{S_1=24\}}\big)(\omega_i)\mathbf{P}(\omega_i)$$

and

$$\mathbf{E}\big[\mathbf{E}[S_2 \,|\, S_1] \ \mathbf{1}_{\{S_1=24\}}\big] = \sum_{i=1}^{4} \big(\mathbf{E}[S_2 \,|\, S_1] \ \mathbf{1}_{\{S_1=24\}}\big)(\omega_i)\mathbf{P}(\omega_i).$$

More formulas for the discrete case.

1. **Computation of the expectation of a function of the random variable**

$$\mathbf{E}[h(X)] = \sum_{x \in C_X} h(x)\, \mathbf{P}\{X = x\},$$

   assuming that the series converges absolutely.

2. **Computation of the conditional expectation**

$$\mathbf{E}[h(X) \,|\, Y = y] = \sum_{x \in C_X} h(x)\, \mathbf{P}\{X = x \,|\, Y = y\},$$

   assuming that the series converges absolutely.

3. **Total expectation formula**: Assume that the random variable $h(X)$ ha finite expectation. Then

$$\mathbf{E}[h(X)] = \sum_{y \in C_Y} \mathbf{E}[h(X) \,|\, Y = y]\, \mathbf{P}\{Y = y\}.$$

1.3 Continuous random variables

1.3.1 Example

As in the previous example, take $S_0 = 20$ €, but in this case, the prices of tomorrow are given by a continuous random variable S_1 with probability density function (**pdf** from now on)

$$f(x) = \begin{cases} 0.006(x-15)(25-x), & \text{if } x \in (15, 25), \\ 0, & \text{otherwise.} \end{cases} \tag{1.5}$$

or written with the help of indicator functions,

$$f(x) = 0.006(x-15)(25-x)\, \mathbf{1}_{(15,20)}(x).$$

A plot of such pdf is given in Figure 1.3.

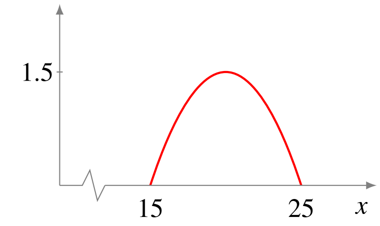


Figure 1.3. Probability density function of S_1

The expectation of S_1 is

$$\mathbf{E}[S_1] = 0.006 \int_{15}^{25} x(x-15)(25-x)\,dx = 20.$$

For the sake of clarity it is convenient to write $f_{S_1}(x)$ instead of $f(x)$. Now we specify the probabilities for the prices the day after tomorrow in function of tomorrow's ones: For $y \in (15, 25)$ (that is, for an acceptable value of S_1), if $S_1 = y$, then S_2 has a pdf given by

$$f_{S_2\,|\,S_1}(x\,|\,y) = 0.006(5+x-y)(5-x+y)\,\mathbf{1}_{(y-5,y+5)}(x).$$

See Figure 1.4.

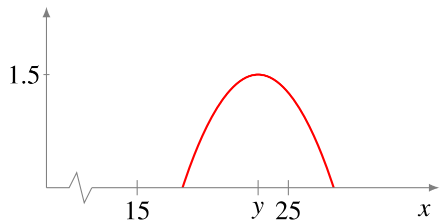


Figure 1.4. Probability density function of S_2 conditionated to $S_1 = y$

The conditional expectation is

$$\mathbf{E}[S_2\,|\,S_1 = y] = 0.006 \int_{y-5}^{y+5} x(5+x-y)(5-x+y)\,dx = y.$$

Remember that this expression is only correct for $y \in (15, 25)$. Then, we can define a new random variable

$$\boxed{\mathbf{E}[S_2\,|\,S_1] = S_1}$$

Remark. From a mathematical point of view it is convenient to have the conditional pdf $f(x|y)$ defined for all $y \in \mathbb{R}$ and not only for $y \in (15, 25)$. This is accomplished defining, for $y \notin (15, 25)$, $f(x|y)$ the value of an arbitrary density (as function of x). Really, the values of such $f(x|y)$ outside the admissible range of y do not appear in any formula. However, it must be careful with automatic computations of conditionals pdf, since it is easy to make mistakes.

We recall that the conditional pdf is related with the **joint pdf** $f(x,y)$ of the random vector (S_1,S_2) by the formula

$$f(x,y) = f_{S_1}(y) f_{S_2|S_1}(x|y),$$

(here $f_{S_1}(y)$ is the pdf of S_1 using the letter y instead of x) that is,

$$\begin{aligned} f(x,y) &= \big(-0.006(y-15)(y-25)\,\mathbf{1}_{(15,20)}(y)\big)\big(-0.006(x-y+5)(x-y-5)\big)\,\mathbf{1}_{(y-5,y+5)}(x) \\ &= 0.006^2(y-15)(y-25)(x-y+5)(x-y-5)\,\mathbf{1}_B(x,y), \end{aligned}$$

where B is the set of points (x,y) such that $y \in (15,25)$, and $x \in (y-5, y+5)$, that is, the parallelogram given in Figure 1.5.

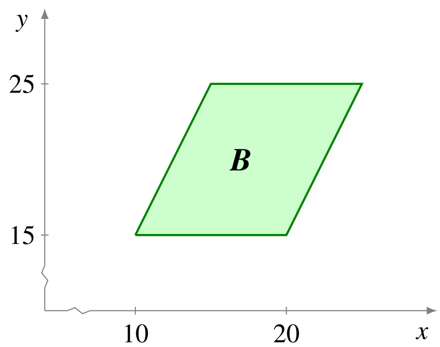


Figure 1.5. The joint pdf of (S_1,S_2) is zero out the set B

1.3.2 The formulas for the continuous case

Let (X,Y) a continuous random vector with marginal densities $f_X(x)$ and $f_Y(y)$. We assume that it is known either

1 The joint pdf $f_{(X,Y)}(x,y)$

or

2 For every y such that $f_Y(y) > 0$, the conditional pdf $f_{X|Y}(x|y)$.

The link between both informations is given by the formula

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)},$$

or

$$f_{(X,Y)}(x,y) = f_Y(y)\, f_{X|Y}(x|y).$$

The conditional expectation of X given $Y=y$ is

$$\mathsf{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx,$$

assuming $\int_{-\infty}^{\infty} |x| f_{X|Y}(x|y)\,dx < \infty$.

It is worth to remark that the expression $\mathbf{E}[X \mid Y = y]$ is a function of y, say

$$\mathbf{E}[X \mid Y = y] = h(y).$$

So we can define a new random variable by

$$\mathbf{E}[X \mid Y] = h(Y).$$

The properties in page 18 are also true in this setup:

The random variable $\mathbf{E}[X \mid Y]$ verifies

(i) It is a function of Y:

$$\mathbf{E}[X \mid Y] = h(Y).$$

(ii) For every Borelian set $B \in \mathcal{B}(\mathbb{R})$,

$$\int_{\{Y \in B\}} X \, d\mathbf{P} = \int_{\{Y \in B\}} \mathbf{E}[X \mid Y] \, d\mathbf{P}.$$

More formulae for the continuous case.

1.

$$\mathbf{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) \, dx,$$

if the integral converges absolutely.

2.

$$\mathbf{E}[h(X) \mid Y = y] = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y) \, dx, \tag{1.6}$$

if the integral converges absolutely.

3. *Total expectation formula*: If $h(X)$ has finite expectation, then,

$$\mathbf{E}[h(X)] = \int_{-\infty}^{\infty} \mathbf{E}[h(X) \mid Y = y] f_Y(y) \, dy.$$

1.4 To practice more

We study a couple of examples.

Exercice 1

A word is selected at random (that is each word with probability 1/7) of the sentence

IT IS TOO GOOD TO BE TRUE

Consider the following two random variables

- X is the number of letters that has the word.
- Y is the number of **O** that has the word.

1. Construct the two-dimensional table of the vector (X, Y).

2. Deduce the marginal distribution of X and Y.

3. Compute the conditional distribution of X given Y.

4. Give the expression of the conditional expectation $\mathsf{E}[X \mid Y]$.

Solution. 1. The two dimensional table is computed in the following way:

$$\mathsf{P}(X = 2, Y = 0) = \mathsf{P}(\textbf{IT}, \ \textbf{IS}, \ \textbf{BE}\ ) = \frac{3}{7}, \text{ etc.}$$

See Table 1.7.

		X		
		2	3	4
	0	3/7	0	1/7
Y	1	1/7	0	0
	2	0	1/7	1/7

Table 1.7. Joint distribution of (X, Y)

2. See table 1.8.

X	2	3	4
$\mathsf{P}\{X = i\}$	4/7	1/7	2/7

Y	0	1	2
$\mathsf{P}\{Y = i\}$	4/7	1/7	2/7

Table 1.8. Marginal distributions of X and Y

3. The conditional probabilities are computed using

$$\mathsf{P}\{X = i \,|\, Y = j\} = \frac{\mathsf{P}\{X = i, Y = j\}}{\mathsf{P}\{Y = j\}},$$

and the result is in Table 1.9.

4.

$$\mathsf{E}[X \mid Y = 0] = 2 \cdot \frac{3}{4} + 4 \cdot \frac{1}{4} = \frac{5}{2},$$

and analogously,

$$\mathsf{E}[X \mid Y = 1] = 2 \quad \text{and} \quad \mathsf{E}[X \mid Y = 2] = 7/2$$

X	2	4
$\mathsf{P}\{X=i \mid Y=0\}$	3/4	1/4

X	2
$\mathsf{P}\{X=i \mid Y=1\}$	1

X	3	4
$\mathsf{P}\{X=i \mid Y=2\}$	1/2	1/2

Table 1.9. Conditional distribution of X given Y

Both expressions are joined:

$$\mathsf{E}[X \mid Y] = \begin{cases} 5/2, & \text{if } Y=0, \\ 2, & \text{if } Y=1 \\ 7/2, & \text{if } Y=2 \end{cases}$$

Or more compactly,

$$\boxed{\mathsf{E}[X \mid Y] = 5/2\ \mathbf{1}_{\{Y=0\}} + 2\ \mathbf{1}_{\{Y=1\}} + 7/2\ \mathbf{1}_{\{Y=2\}}.}$$

Exercise 2

Let (X,Y) a random vector uniform on the triangle $\mathbf{T}$ with vertices (0,0), (1,0) and (1,1) (see Figure 1.6), with pdf

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & \text{if } (x,y) \in \mathbf{T}, \\ 0, & \text{otherwise.} \end{cases}$$

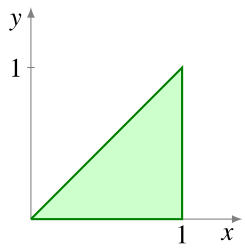


Figure 1.6. Triangle **T**

1. Compute the pdf of Y.

2. Compute the conditional pdf of X given Y.

3. Compute the conditional expectation of X given Y.

Solution. 1. The pdf of Y (also called the marginal pdf of Y, because we deduce from the pdf of a vector) is computed in the following way:

- If $y \notin (0,1)$, then $f_{(X,Y)}(x,y) = 0,\ \forall x$, so $f_Y(y) = 0$.

- If $y \in (0,1)$,

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & \text{if } x \in (y,1), \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\, dx = 2(1-y).$$

To summarize both cases,

$$f_Y(y) = \begin{cases} 2(1-y), & \text{if } 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Or in compact notation,

$$f_Y(y) = 2(1-y)\, \mathbf{1}_{(0,1)}(y).$$

See Figure 1.7.

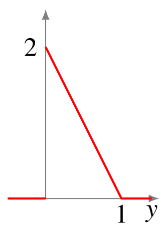


Figure 1.7. Pdf of Y

2. The conditional pdf of X given $Y = y$ is computed for $y \in (0,1)$, and gives

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \begin{cases} \dfrac{1}{1-y}, & \text{if } x \in (y,1), \\ 0, & \text{otherwise.} \end{cases}$$

Or in compact notation,

$$f_{X|Y}(x|y) = \frac{1}{1-y}\, \mathbf{1}_{(y,1)}(x).$$

Therefore, it is a uniform distribution on $(y,1)$. See Figure 1.8.

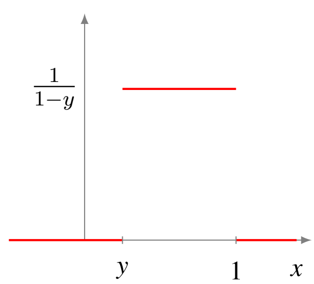


Figure 1.8. Conditional pdf of X given $Y = y \in (0,1)$

For $y \notin (0,1)$, since $f_Y(y) = 0$, we put $f_{X|Y}(x|y),\ x \in \mathbb{R}$ the value of an arbitrary density.

3. The conditional expectation for $y \in (0,1)$, is

$$\mathbf{E}[X \mid Y = y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\, dx = \frac{1}{2}(1-y),$$

Then

$$\mathbf{E}[X \mid Y] = \frac{1}{2}(1-Y).$$

1.5 Exercises

1. A coin is tossed three times. Denote by X the number of Heads, and by Y the difference in absolute value between the number of Heads and the number of Tails.

   (a) Construct the two-dimensional table of the vector (X, Y).

   (b) Deduce the marginal distribution of X and Y.

   (c) Compute the conditional distribution of X given Y.

   (d) Give the expression of the conditional expectation $\mathbf{E}[X \mid Y]$.

2. Consider a random vector (X, Y) with density

$$f(x,y) = \begin{cases} \dfrac{1+xy}{4}, & \text{if } -1 < x < 1 \text{ and } -1 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Compute the pdf of Y.

   (b) Compute the conditional pdf of X given $Y = y$. Note that in the formula that you obtain, the role of x and y is totally different.

   (c) Compute the conditional expectation of X given $Y = y$. The result should be a function of y, not of x.

3. With the data of the introductory example (page 11) consider a digital option with payoff $\mathbf{1}_{\{S_2=28.8\}}$. Compute $\mathbf{E}[\,\mathbf{1}_{\{S_2=28.8\}}]$ and $\mathbf{E}[\,\mathbf{1}_{\{S_2=28.8\}} \mid S_1]$.

4. Following with Example 1.3.1, compute $\mathbf{E}[S^2 \mid S_1]$. *Indication:* Use formula (1.6).

Chapter 2

Information and measurability

In this chapter we study how the notion of the information associated to a random experiment is managed.

2.1 Probability space

We recall that a probability space is a triplet $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is an arbitrary set, $\mathcal{F}$ is a family of subsets of Ω, called the collection of **events**, that has a structure of σ–field:

(i) $\Omega \in \mathcal{F}$.

(ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, where A^c is the event contrary to A, defined by the complementary set $A^c = \{\omega \in \Omega : \omega \notin A\}$.

(iii) If $A_1, A_2, \dots \in \mathcal{F}$, then

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Finally, the third ingredient is a map $\mathbf{P} : \mathcal{F} \to [0,1]$ such that

(i) $\mathbf{P}(\Omega) = 1$.

(ii) If $A_1,\ A_2, \dots, A_n, \dots \in \mathcal{F}$ pairwise disjoints: $A_n \cap A_n = \emptyset$, when $n \neq m$, then

$$\mathbf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

In general, in elementary courses of probability the role of the set of events $\mathcal{F}$ is not considered. This can be done because in discrete models (when Ω is finite or enumerable), normally it is possible to take $\mathcal{F}$ as the collection of all subsets of Ω, which is a σ–field, so it is preferable do not bother the students with unnecessary definitions; when the model is continuous (Ω is $\mathbb{R}$) then it is better to work directly with continuous random variables, and again to skip $\mathcal{F}$. However, in advanced probability theory (and in Mathematical finance) the set $\mathcal{F}$ is important to model the information associated to a random experiment.

2.1.1 Some examples

Example 1. We play with an unfair dice and we do not know the probabilities of the possible elementary results, 1, 2,..., 6. We take $\Omega = \{1,\ 2, \dots, 6\}$. After 1000 plays, we are informed that 200 times the result has been 3. With such information, our estimation of the probabilities are

$$\mathbf{P}(3) = \frac{200}{1000} = \frac{1}{5} \quad \text{and} \quad \mathbf{P}(1,2,4,5,6) = \frac{4}{5}.$$

From that information it is impossible to deduce the probability of 6, or the probability of an odd number. So, associated with the information that we have, the family of events $\mathcal{F}$ is

$$\mathcal{F} = \big\{\emptyset, \Omega, \{3\}, \{1,2,4,5,6\}\big\},$$

because the events of $\mathcal{F}$ are the only ones to which we will give a probability.

Indeed, in a higher level of abstraction, we even can forget the probability, and say that the elements of $\mathcal{F}$ are the *observable sets* in the random experiment.

Example 2. Return to the introductory example of the first chapter, and the set $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ introduced in page 15. After day 1, we will know S_1, that is, if $S_1 = 24$ or $S_1 = 16$; equivalently, we will know what event $\{\omega_1, \omega_2\}$ or $\{\omega_3, \omega_4\}$ has happened, but it is impossible to know if ω_1 has happened until day 2. Write

$$\mathcal{F}_1 = \big\{\emptyset, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\big\}.$$

(Note that it is a σ–field). That family of events is called the information after day 1. Now let

$$\begin{aligned}\mathcal{F}_2 = &\big\{\emptyset, \Omega, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\},\\ &\{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}\big\}.\end{aligned}$$

(Again note that it is a σ–field.) It is called the information after day 2, because after that day we will know of each event in $\mathcal{F}_2$ if it has happened or no. For example, take $\{\omega_1, \omega_2\}$: this has happened if $S_1 = 24$ (we will know this). Or $\{\omega_1\}$, that has happened if $S_1 = 24$ and $S_2 = 28.8$.

To complete the picture, put

$$\mathcal{F}_0 = \big\{\emptyset, \Omega\big\}.$$

We always know if the elements of $\mathcal{F}_0$ has happened of not: $\emptyset$ never happens, Ω always happens. This is the information after day 0: we have $S_0 = 20$, that we already knew. Note that

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2.$$

This is the **flow of information**.

2.1.2 A trick to short the scripture of a σ–field

In the previous example, $\mathcal{F}_2$ is the family of all subsets of Ω. Using a bit of combinatorics we can show that if Ω has n elements, then it has 2^n subsets. That means that written explicitly a σ–field is, even in easy cases, very large, or impossible. The trick to short that scripture is to use the notion of a σ–field generated by a family of subsets:

Given a family (finite or no) of subsets of Ω

$$A_1, A_2, \dots \subset \Omega,$$

the ***σ*–field generated** by that family is the small σ-field that contains the sets $A_1, A_2, \dots$, and it is denoted by

$$\boldsymbol{\sigma(A_1, A_2, \dots)}.$$

For example, returning to Example 2 above, write $A_1 = \{\omega_1, \omega_2\}$ and $A_2 = \{\omega_3, \omega_4\}$. Then

$$\mathcal{F}_1 = \sigma\big(A_1, A_2\big) = \sigma\big(A_1\big).$$

(Please, explain the second equality). And

$$\mathcal{F}_2 = \sigma\big(\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\big).$$

2.2 Measurable functions: discrete functions

We will consider a set Ω, a σ–field $\mathcal{F}$ and a map $X : \Omega \longrightarrow \mathbb{R}$. Remember the shortening

$$\{X \in A\} = \{\omega \in \Omega : \; X(\omega) \in A\},$$

and when $A = \{x\}$, then

$$\big\{X \in \{x\}\big\} = \{X = x\}.$$

Assume that X takes a finite or enumerable number of possible values, and denote by C_X the image set of X.

We will say that X is $\mathcal{F}$ **measurable** if

$$\{X = x\} \in \mathcal{F}, \quad \forall x \in C_X.$$

Continuation of Example 1. Consider Example 1, where $\Omega = \{1, 2, \dots, 6\}$ and $\mathcal{F} = \big\{\emptyset, \Omega, \{3\}, \{1,2,4,5,6\}\big\}$. Let X be a random variable that represents to bet 1 € to even numbers: if the result is even, we receive 1 € and recover our bet; if the result is odd, we loose the euro. That map is $X : \{1,2,3,4,5,6\} \longrightarrow \mathbb{R}$,

$$X(2) = X(4) = X(6) = 2, \quad X(1) = X(3) = X(5) = -1.$$

As a map, X is properly defined, but it is not $\mathcal{F}$-measurable, since

$$\{X = 2\} = \{2, 4, 6\} \notin \mathcal{F}.$$

That means, with the information that we have of this random experiment, it will be impossible to compute the probability to win the bet. On the contrary, consider to bet 1 € to number 3; if the result is 3, we receive, say, 5 € and recover the initial euro; if the result is not 3, we loose the euro; this bet is formalized with the map $Y : \{1,2,3,4,5,6\} \longrightarrow \mathbb{R}$,

$$Y(3) = 6, \quad Y(1) = Y(2) = Y(4) = Y(5) = Y(6) = -1.$$

Here, $C_Y = \{-1, 6\}$, and

$$\{Y = -1\} = \{1,2,4,5,6\} \in \mathcal{F} \quad \text{and} \quad \{Y = 6\} = \{3\} \in \mathcal{F}.$$

So Y is $\mathcal{F}$ measurable.

2.2.1 σ–field generated by a function

We continue with the discrete case. Here we invert the order of the ingredients: consider a set Ω and a function $X : \Omega \longrightarrow \mathbb{R}$.

> The small σ–field (on Ω) such that X is measurable is called the σ–field generated by X and is denoted by $\sigma(X)$

Return again to Example 1, with $\Omega = \{1,\, 2, \ldots, 6\}$ and $\mathcal{F} = \big\{\emptyset, \Omega, \{3\}, \{1,2,4,5,6\}\big\}$. We saw that Y is $\mathcal{F}$ measurable. However, let

$$\mathcal{G} = \sigma\big(\{1,2,3\}, \{4,5,6\}\big);$$

it is easy (or can be deduced from exercise) that Y is also $\mathcal{G}$ measurable. However,

$$\sigma(Y) = \mathcal{F},$$

because $\mathcal{F}$ is the smaller σ–field (we need to check all the other σ–fields).

2.3 Measurable function: general case

When the hypothesis that the map $X : \Omega \longrightarrow \mathbb{R}$ is discrete is suppressed, then appear technical difficulties due to the fact that the sets $\{X = x\}$ can all have zero probability, and the definition given in the previous section is not useful at all when we have in mind to work with probability. So we need a new definition.

> We will say that $X : \Omega \longrightarrow \mathbb{R}$ is $\mathcal{F}$ **measurable** if
>
> $$\{X \le t\} \in \mathcal{F}, \quad \forall t \in \mathbb{R}.$$
>
> The σ–field generated by X, $\sigma(X)$, is the small σ–field that contains all the sets $\{X \le t\} \in \mathcal{F}, \quad \forall t \in \mathbb{R}$.

(Why when X is discrete this definition contains the old one?)

2.3.1 Random variables

In a general setup we will start with a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. A random variable X is a map $X : \Omega \longrightarrow \mathbb{R}$ that is $\mathcal{F}$ measurable. Many times, as in Chapter 1, we omit to mention the reference space $(\Omega, \mathcal{F}, \mathbf{P})$, and to write the ω. But you should keep in mind that they are here, at least in a subliminal way, and they can appear when you need to clarify some points.

2.3.2 Random vectors

Consider again the reference space $(\Omega, \mathcal{F}, \mathbf{P})$. A random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a map $\boldsymbol{X} : \Omega \longrightarrow \mathbb{R}^n$ such that every component X_j is $\mathcal{F}$ measurable. Equivalently, for every $t_1, \ldots, t_n \in \mathbb{R}$,

$$\{X_1 \le t_1, \ldots, X_n \le t_n\} \in \mathcal{F},$$

where the comma means intersection; that is,

$$\{X_1 \le t_1, \ldots, X_n \le t_n\} = \{X_1 \le t_1\} \cap \cdots \cap \{X_1 \le t_1\}.$$

2.3.3 The Borel σ–field

To construct a reasonable σ–field over a set is not always an easy task; this is particularly true if we want to define a probability on that σ–field, because if the σ–field has few events, the probabilistic model is poor, and if it has many events, then it is not clear how to build a probability. In particular, when $\Omega = \mathbb{R}$ this is a main problem: the set of all subsets of $\mathbb{R}$ is enormous. The usual σ–field on $\mathbb{R}$ that we take is the small σ–field that contains all the open sets of $\mathbb{R}$; this is called the **Borel**-sigma field on $\mathbb{R}$, and it is denoted by $\mathcal{B}(\mathbb{R})$. With previous notations,

$$\mathcal{B}(\mathbb{R}) = \sigma(\text{open sets of } \mathbb{R}).$$

It can be proved that it is also generated for a class of intervals, for example, the closed intervals, or it is very useful that

$$\mathcal{B}(\mathbb{R}) = \sigma\big((-\infty, t],\ t \in \mathbb{R}\big).$$

2.3.4 Measurable real functions

When $\Omega = \mathbb{R}$, that means, we consider an ordinary map $h : \mathbb{R} \longrightarrow \mathbb{R}$, then normally we say that is measurable when we take $\mathcal{F} = \mathcal{B}(\mathbb{R})$.

2.3.5 A main result in measurability

A main property of measurable function is the following result:

Consider two maps $X, Z : \Omega \longrightarrow \mathbb{R}$. Then X is $\sigma(Z)$ measurable if and only if there is a measurable map $h : \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$X = h(Z).$$

We stress that this result closes the circle about measurable functions: Roughly speaking, a function X is measurable with respect a σ–field $\mathcal{F}$ if it depends on the information given by $\mathcal{F}$; if the information is given by a map Z, then $\mathcal{F} = \sigma(Z)$, and in such case, X should be a function of Z.

2.4 Exercices

1. A coin is tossed three times. Write H or T for head and tail respectively, and let

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

(a) Write the σ–fields $\mathcal{F}_j$, $j = 1, 2, 3$ corresponding to the information after the j throw

(b) Write the σ–field $\mathcal{G}_2$ corresponding to the the information of just the 2n throw (and nothing about the first and three throw). Check that $\mathcal{F}_2 \not\subset \mathcal{G}_2$.

2. In Example 2 (page 28), prove

(a) S_j is $\mathcal{F}_j$ measurable, $j = 0, 1, 2$.

(b) S_1 is $\mathcal{F}_2$ measurable.

(c) S_2 is not $\mathcal{F}_1$ measurable

3. With the function X given in the Continuation of Example 1 (page 29) compute $\sigma(X)$.

4. With the functions S_1, S_2 and S_3 given in Example 2 (page 28) check that $\sigma(S_j) = \mathcal{F}_j,\ j = 0, 1, 2.$

5. With the data of Example 2 of page 28, consider the function

$$X(\omega_1) = X(\omega_2) = 1,\ X(\omega_3) = X(\omega_4) = 5.$$

Prove that X is $\sigma(S_1)$ measurable. You can prove this building the σ-field generated by S_1 and checking the condition of measurability, or using the property displayed above.

6. On a set Ω consider two σ–fields, $\mathcal{F}$ and $\mathcal{G}$, with $\mathcal{F} \subset \mathcal{G}$. $X : \Omega \longrightarrow \mathbb{R}$ be $\mathcal{F}$ measurable. Prove that X is also $\mathcal{G}$ measurable.

7. In Exercise 2 (a) it should be proved that S_0 is $\mathcal{F}_0$ measurable. A generalization of that fact is the following: Consider an arbitrary set Ω and the so–called trivial σ–field $\mathcal{F} = \{\emptyset, \Omega\}$. Let $X : \Omega \longrightarrow \mathbb{R}$ a constant map, that is, there is a number $a \in \mathbb{R}$ such that $X(\Omega) = a,\ \forall \omega \in \Omega.$ Prove that X is $\mathcal{F}$ measurable. Prove also the reciprocal: if $X : \Omega \longrightarrow \mathbb{R}$ is $\mathcal{F}$ measurable, then X is a constant map.

Chapter 3

Conditional expectation: General case

In this chapter we give a general definition of conditional expectation. That definition seems very abstract and cryptic. However, it is also very powerful, and after some practice it allows to work with general properties in a short and elegant way.

3.1 Conditional expectation given a σ-field

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. The following result is both a property and a definition.

Let $X : \Omega \longrightarrow \mathbb{R}$ be a random variable (that means, $\mathcal{F}$ measurable) with $\mathbf{E}[|X|] < \infty$ and $\mathcal{G} \subset \mathcal{F}$ be a σ-field. There is a random variable

$$\mathbf{E}[X \mid \mathcal{G}] : \Omega \longrightarrow \mathbb{R}$$

called **the conditional expectation of X given $\mathcal{G}$**, unique a.s., such that

(i) $\mathbf{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ measurable.

(ii) For every $A \in \mathcal{G}$,

$$\int_A \mathbf{E}[X \mid \mathcal{G}] \, d\mathbf{P} = \int_A X \, d\mathbf{P} \,. \tag{*}$$

3.1.1 Short return to $\mathbf{E}[X \mid Y]$

Given two random variables X, $Y : \Omega \longrightarrow \mathbb{R}$, define

$$\mathbf{E}[X \mid Y] \underset{\text{def}}{=} \mathbf{E}[X \mid \sigma(Y)].$$

3.1.2 Properties.

Certain amount of practice is necessary to fully understand the notion of general conditional expectation. To work in the properties will help. Some properties need advanced use of abstract Lebesgue integral, and we omit. We will assume that all random variables are integrable.

1.

$$\boxed{\mathbf{E}\big[\mathbf{E}[X \mid \mathcal{G}]\big] = \mathbf{E}[X].}$$

Proof. By the property (*) applied to $A = \Omega \in \mathcal{G}$

$$\int_\Omega \mathbf{E}[X \mid \mathcal{G}]\, d\mathbf{P} = \int_\Omega X\, d\mathbf{P} = \mathbf{E}[X].$$

that is what we want.

2. Let $c \in \mathbb{R}$. Then

$$\boxed{\mathbf{E}[c \mid \mathcal{G}] = c.}$$

Proof. The way that we prove this property is very illustrative of the procedure to work with general conditional expectation. Our objective is to compute $\mathbf{E}[c \mid \mathcal{G}]$ and we have a *candidate* (it is always needed to guess a *candidate*): c. Then, we will check that the *candidate* satisfies both properties of the conditional expectation, and by the unicity, the *candidate* will be confirmed as the true value.

(i) Measurability: c is $\mathcal{G}$ measurable.Yes. (Please, check this fact).

(ii) Property (*). We need to check that for every $A \in \mathcal{G}$,

$$\int_A c\, d\mathbf{P} = \int_A c\, d\mathbf{P}$$

(in the left hand side there is the *candidate* c, in the right hand side the random variable $X = c$). Which, of course, is obvious.

3. Linearity.

$$\boxed{\mathbf{E}[aX + bY \mid \mathcal{G}] = a\,\mathbf{E}[X \mid \mathcal{G}] + b\,\mathbf{E}[Y \mid \mathcal{G}]\,.}$$

Proof. Here the *candidate* is $a\,\mathbf{E}[X \mid \mathcal{G}] + b\,\mathbf{E}[X \mid \mathcal{G}]$.

(i) Measurability. By definition, $\mathbf{E}[X \mid \mathcal{G}]$ i $\mathbf{E}[Y \mid \mathcal{G}]$ are both $\mathcal{G}$ measurable; hence, $a\,\mathbf{E}[X \mid \mathcal{G}] + b\,\mathbf{E}[X \mid \mathcal{G}]$ is.

(ii) Property (*). We want to prove that for every $A \in \mathcal{G}$,

$$\int_A \Big(a\,\mathbf{E}[X \mid \mathcal{G}] + b\,\mathbf{E}[Y \mid \mathcal{G}]\Big)\, d\mathbf{P} = \int_A \Big(aX + bY\Big)\, d\mathbf{P}. \tag{3.1}$$

By the definition of conditional expectation,

$$\int_A \mathbf{E}[X \mid \mathcal{G}]\, d\mathbf{P} = \int_A X\, d\mathbf{P}$$
$$\int_A \mathbf{E}[Y \mid \mathcal{G}]\, d\mathbf{P} = \int_A Y\, d\mathbf{P}$$

And (3.1) follows.

4.

$$\boxed{X \le Y \Longrightarrow \mathbf{E}[X \mid \mathcal{G}] \le \mathbf{E}[Y \mid \mathcal{G}]\,.}$$

5.

$$\boxed{Z \quad \mathcal{G}-\text{measurable} \quad \Longrightarrow \quad \mathbf{E}[XZ \mid \mathcal{G}] = Z\,\mathbf{E}[X \mid \mathcal{G}].}$$

In particular, if X is $\mathcal{G}$-measurable, then

$$\mathbf{E}[X \mid \mathcal{G}] = X.$$

Intuitively, the random variable Z is deterministic if it is known $\mathcal{G}$, so it behaves as that in the conditional expectation.

Proof. (You can skip). The *candidate* is $Z\,\mathbf{E}[X \mid \mathcal{G}]$. We need to check

(i) The random variable $Z\mathbf{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.

(ii) For every $A \in \mathcal{G}$,

$$\int_A Z\mathbf{E}[X \mid \mathcal{G}]\, d\mathbf{P} = \int_A ZX\, d\mathbf{P}. \tag{3.2}$$

(i) is clear. To prove (ii), the standard *machine* of integration theory is put to work: first check (3.2) for $Z = \mathbf{1}_B$, for $B \in \mathcal{G}$. By linearity the equality is extended to simple random variables, etc.

6. Tower property.

$$\boxed{\mathcal{G}_1 \subset \mathcal{G}_2 \quad \Longrightarrow \quad \mathbf{E}\big[\mathbf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2\big] = \mathbf{E}\big[\mathbf{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1\big] = \mathbf{E}[X \mid \mathcal{G}_1]\,.}$$

(Usualy, it is written $\mathbf{E}[X \mid \mathcal{G}_1 \mid \mathcal{G}_2]$ instead of $\mathbf{E}\big[\mathbf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2\big]$).

Proof. To prove $\mathbf{E}[X \mid \mathcal{G}_1 \mid \mathcal{G}_2] = \mathbf{E}[X \mid \mathcal{G}_1]$ it is needed to use

$$\left.\begin{array}{c} \mathbf{E}[X \mid \mathcal{G}_1] \text{ és } \mathcal{G}_1 - \text{measurable} \\ \mathcal{G}_1 \subset \mathcal{G}_2 \end{array}\right\} \Longrightarrow \mathbf{E}[X \mid \mathcal{G}_1] \;\; \mathcal{G}_2 - \text{measurable}$$

and apply (5). To prove $\mathbf{E}[X \mid \mathcal{G}_2 \mid \mathcal{G}_1] = \mathbf{E}[X \mid \mathcal{G}_1]$ use the properties (i) and (ii) that characterize the conditional expectation.

7. Conditional expectation and independence

$$\boxed{X \text{ } and \text{ } \mathcal{G} \textbf{ independent} \quad \Longrightarrow \quad \mathsf{E}[X \,|\, \mathcal{G}] = \mathsf{E}[X].}$$

Remember that it is said that X and $\mathcal{G}$ are independent if $\forall A \in \mathcal{G}$, the random variables X and $\mathbf{1}_A$ are independent, or equivalently, if $\forall B \in \mathcal{B}(\mathbb{R})$ and $\forall A \in \mathcal{G}$, the events $\{X \in B\}$ are A independent.

Proof. Exercise.

8. Let $\{X_n,\, n \in \mathbb{N}\}$ be a sequence of integrable random variables.

a. *(Monotone convergence Theorem)* If $Y \le X_n \nearrow X$, $a.s.$, where Y is an integrable random variable, then $\mathsf{E}[X_n \,|\, \mathcal{G}] \nearrow \mathsf{E}[X \,|\, \mathcal{G}]$, $a.s.$

b. *(Dominated convergence Theorem)* If $|X_n| \le Y$, where Y is an integrable random variable, and i $X_n \to X$, $a.s.$, then
$$\mathsf{E}[X_n \,|\, \mathcal{G}] \to \mathsf{E}[X \,|\, \mathcal{G}]\,, \quad a.s.$$

9. $\forall p \ge 1$,
$$|\mathsf{E}[X \,|\, \mathcal{G}]|^p \le \mathsf{E}\big[|X|^p \,|\, \mathcal{G}\big]\,.$$

In particular,
$$\big(\mathsf{E}[X \,|\, \mathcal{G}]\big)^2 \le \mathsf{E}\big[X^2 \,|\, \mathcal{G}\big].$$

3.2 Some useful properties for conditional expectations given a random variable

All properties that we have studied can be translated to $\mathsf{E}[X \,|\, Y]$ and $\mathsf{E}[X \,|\, Y = y]$. Some of them have an interesting version:

Property 5. Let X and Y random variables and consider a map $f : \mathbb{R} \to \mathbb{R}$. Then

$$\boxed{\mathsf{E}[Xf(Y) \,|\, Y] = f(Y)\mathsf{E}[X \,|\, Y],}$$

or equivalently,

$$\boxed{\mathsf{E}[Xf(Y) \,|\, Y = y] = f(y)\mathsf{E}[X \,|\, Y = y]\,.}$$

The following property allows to simplify the computations in some practical situation:

Property. Let X and Y independent random variables and $h : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\boxed{\mathsf{E}[h(X,Y) \,|\, Y = y] = \mathsf{E}[h(X,y)],}$$

where $\mathsf{E}[h(X,y)]$ means that in the expression of $\mathsf{E}[h(X,Y)]$ it has been changed the random variable Y by the value y, and computed the expectation with respect to X.

Example. This example shows the point although it is a bit artificial. We throw two dices; denote the results by X and Y. We want to compute $\mathbf{E}[\sqrt{X+Y} \,|\, Y = 6]$. In agreement with the previous property,

$$\mathbf{E}[\sqrt{X+Y} \,|\, Y = 6] = \mathbf{E}[\sqrt{X+6}] = \frac{1}{6}\sum_{j=1}^{6} \sqrt{j+6} = 3.07.$$

Extension. The previous property attains its full power when it is used for an arbitrary value of Y, that means, we have

$$\mathbf{E}[h(X,Y) \,|\, Y = y] = \mathbf{E}[h(X,y)]$$

as a function of y.

Equivalently, write

$$\varphi(y) = \mathbf{E}[h(X,y)],$$

then

$$\mathbf{E}[h(X,Y) \,|\, Y] = \varphi(Y)\,. \tag{3.3}$$

This expression is also written

$$\mathbf{E}[h(X,y)]|_{y=Y},$$

that means that we first compute $\mathbf{E}[h(X,y)]$, and after we change y by Y.

3.3 Appendix. Conditional expectation and convexity

Here is a good point to work slowly and recall the very important notion of **convex** and **concave** functions, which play an important role in many parts of theoretical and applied mathematics; in particular in Mathematical finance.

Let I an interval of $\mathbb{R}$ (finite or infinite).

A map $f : I \longrightarrow \mathbb{R}$ is said to be **convex** if for every $x, x' \in I$, and for every $a \in [0,1]$,

$$f\big(ax + (1-a)x'\big) \le af(x) + (1-a)f(x').$$

When the above inequality is strict for every $x \neq x'$ and $a \in (0,1)$, then the function is called **strictly convex**. A function $f(x)$ such that $-f(x)$ is convex is called **concave**

As you can see in Figure 3.1, the inequality in the definition means that the curve between x and x' lies below the chord joining $(x, f(x))$ and $(x', f(x'))$.

Exercise. Check that Figure 3.1 is correct showing that the straighline by $(x, f(x))$ and $(x', f(x'))$ maps the point $ax + (1-a)x'$ to $af(x) + (1-a)f(x')$.

Remark. Some authors use the name **concave up** instead of convex, and **concave down** instead of concave.

Two main properties of convex functions. For the proofs, see, for example, the classical book of Roberts and Varberg [14].

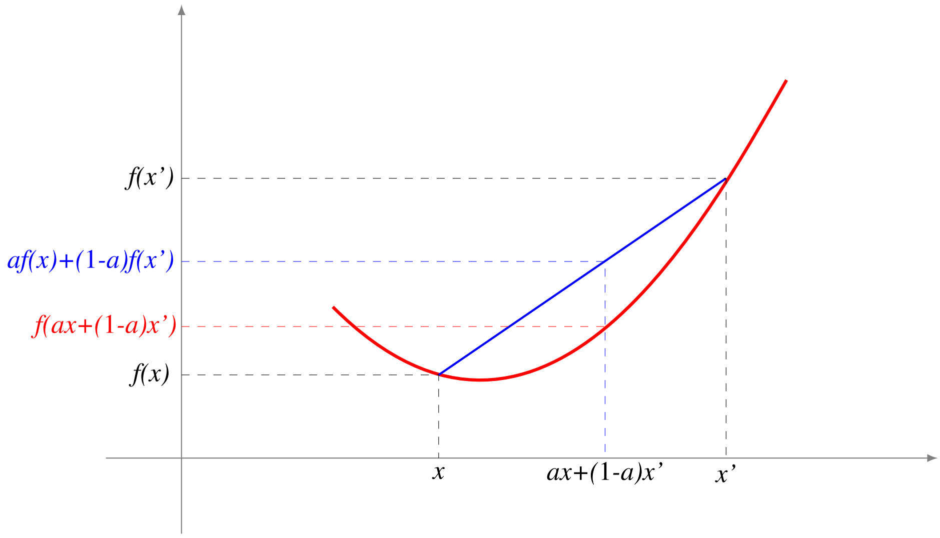


Figure 3.1. A convex function

1. Every convex function is continuous in the interior of I. Moreover it is differentiable at every point of interior of I except, perhaps, in a countable number of points.

2. If f is two times differentiable in an open interval, then f is convex if and only if $f''(x) \geq 0$. If $f''(x) > 0$, then it is strictly convex.

Examples of convex functions

1. Remember that a function $f(x) = ax + b$, which corresponds to a straightline, is called **affine function**. The affine functions are convex. Really, as we will see, they are the building blocks of all convex functions.

2. $f(x) = |x|$, $I = \mathbb{R}$.

3. For $p > 1$, $f(x) = x^p$, $I = [0, \infty]$.

4. For $p > 1$, $f(x) = |x|^p$, $I = \mathbb{R}$.

5. Fixed $c \in \mathbb{R}$, $f(x) = (x - c)^+$, $I = \mathbb{R}$, where

$$z^+ = \begin{cases} z, & \text{if } z \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

   See Figure 3.2. This is an example of a convex function that is differentiable at every point except at point $x = c$.

6. Both $f(x) = e^{-x}$ and $f(x) = e^x$, $I = \mathbb{R}$.

7. $f(x) = -\sqrt{x}$, $I = [0, \infty)$.

8. $f(x) = -\log x$, $I = (0, \infty)$.

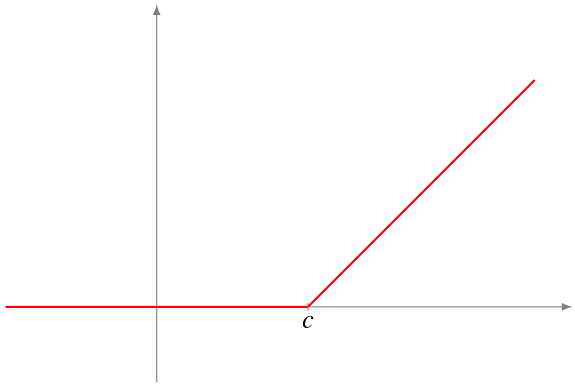


Figure 3.2. The convex function $f(x) = (x - c)^+$

Exercise. Consider a differentiable convex function f on an open interval I. Show that the tangents to f at every point of I are bellow the curve (see Figure 3.3). That is, for every $x_0 \in I$,

$$f(x) \geq f'(x_0)(x - x_0) + f(x_0), \ \forall x \in I.$$

Denote by $\mathcal{A}$ the set of affine functions. Deduce that

$$f = \sup\{h \in \mathcal{A} : \ f(x) \geq h(x), \ \forall x \in I\}.$$

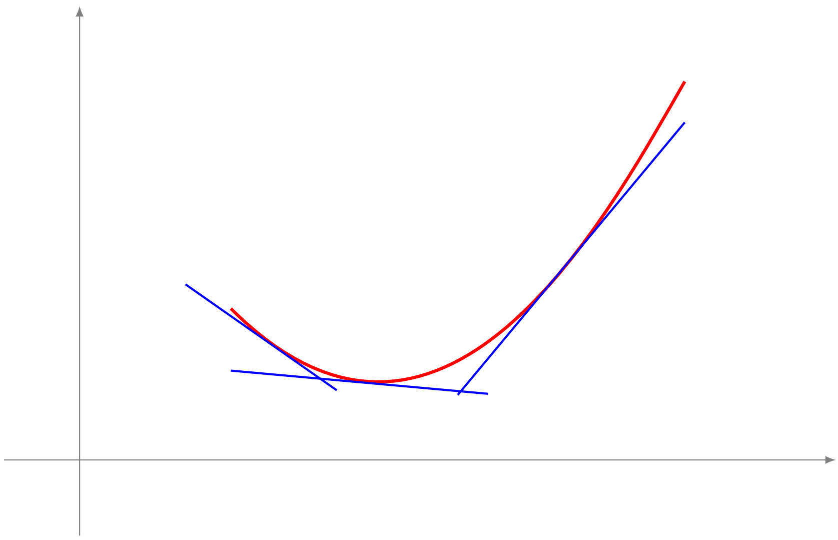


Figure 3.3. Tangents of a convex function

Expectation and convex functions. The main property of convex functions related with probability is the Jensen's inequality:

> Let $f : I \to \mathbb{R}$ be a convex function, and let X be a random variable such that $\mathbf{P}(X \in I) = 1$. Then
>
> $$f\big(E[X]\big) \leq E[f(X)].$$
>
> (Assume all the expectations exists).

Proof. A short proof is based in the fact that the property

$$f = \sup\{h \in \mathcal{A} : \; f(x) \geq h(x), \; \forall x \in I\}$$

given in the previous exercise for differentiable functions, holds for every convex function: in a point where the function is not differentiable the tangent need to be changed by a *support line* (see Roberts and Varberg [14, pag. 12]). Let

$$\mathcal{D} = \{h \in \mathcal{A} : f(x) \geq h(x), \; \forall x \in I\}.$$

For $h(x) = ax + b \in \mathcal{D}$,

$$f(x) \geq ax + b, \; \forall x \in I.$$

Hence, with probability 1,

$$f(X(\omega)) \geq aX(\omega) + b$$

and then

$$\mathbf{E}[f(X)] \geq a\mathbf{E}[X] + b = h([\mathbf{E}[X]). \tag{3.4}$$

On the other hand, since $\mathbf{E}[X] \in I$,

$$f\big(\mathbf{E}[X]\big) = \sup\{h([\mathbf{E}[X]]), \; h \in \mathcal{D}\},$$

and from (3.4) it follows

$$f\big(\mathbf{E}[X]\big) \; \leq \; \mathbf{E}[f(X)].$$

Many well known inequalities can be deduced from Jensen's inequality. For example,

$$\big(\mathbf{E}[X]\big)^2 \leq \mathbf{E}[X^2],$$

since $f(x) = x^2$ is convex.

Jensen's inequality for conditional expectations. Let $f : I \to \mathbb{R}$ be a convex function, and let X be a random variable such that $\mathbf{P}(X \in I) = 1$. Then

$$f\big(\mathbf{E}[X \,|\, \mathcal{G}]\big) \; \leq \; \mathbf{E}[f(X) \,|\, \mathcal{G}].$$

(Assume $\mathbf{E}[|X|] < \infty$ and $\mathbf{E}[|f(X)|] < \infty$.)

The proof is the same as before.

3.4 Exercises

1 Let $\mathcal{G} = \{\emptyset, \Omega\}$. Prove that

$$\mathbf{E}[X \mid \mathcal{G}] = \mathbf{E}[X].$$

(Remember that a random variable $\mathcal{G}$ measurable is constant.)

2 This problem reproduces the computation of an European (*call*) in a simplified version. Consider two independent random variables, $X \sim \mathcal{U}(-2,2)$ (pdf $f_X(x) = \frac{1}{4},\ x \in (-2,2)$) and $Y \sim \mathcal{U}(0.5,1)$ (pdf $f_Y(y) = 2,\ y \in (0.5,1)$). The objective is to compute

$$\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+}\right].$$

A way to proceed is to use that the expectation of the function $h(X,Y)$ of a random vector (X,Y) with joint pdf $f(x,y)$ is

$$\mathbf{E}[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y) dx dy.$$

Then, we need to compute

$$\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+}\right] = \frac{1}{4}\, 2 \iint_{(-2,2)\times(0.5,1)} \left(\frac{x}{y} - 1\right)^{+} dx\, dy.$$

An alternative way (and perhaps short) is to use the properties of conditional expectation: Since

$$\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+}\right] = \mathbf{E}\left[\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+} \mid Y\right]\right],$$

we start computing $\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+} \mid Y\right]$.

a. Show that

$$\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+} \mid Y\right] = \mathbf{E}\left[\left(\frac{X}{y} - 1\right)^{+}\right]\Bigg|_{y=Y}.$$

b. Fix $y \in (0.5, 1)$. Then

$$\mathbf{E}\left[\left(\frac{X}{y} - 1\right)^{+}\right] = \mathbf{E}\left[\left(\frac{X}{y} - 1\right) \mathbf{1}_{\left\{\frac{X}{y} - 1 > 0\right\}}\right] = \mathbf{E}\left[\left(\frac{X}{y} - 1\right) \mathbf{1}_{\{X > y\}}\right] = \frac{1}{4}\int_{y}^{2} \left(\frac{x}{y} - 1\right) dx.$$

c. Compute $\mathbf{E}\left[\left(\frac{X}{Y} - 1\right)^{+}\right]$.

Chapter 4

Martingales with discrete parameter

In this chapter we study a short introduction to martingale theory, which is a crucial notion in Financial Mathematics because it is a dynamic model for a fair game. One of the main principles of our mathematical model of a market is that it is fair for all participants; or more accurately, the market is fair for almost all participants; the fact that some people, the *abitrageurs* or *arbs*, enjoy arbitrage opportunities, and indeed use it, has the consequence that such unfair moments disappear quickly, and the market returns to be fair. Currently, nobody trust is such principle of perfect unfair markets, but it is the starting point of our models.

We will restrict ourselves to the study of a short part of the theory of discrete time martingale. For more general results the interested reader can consult the books of Shiryaev [17] or Durret [6].

4.1 Definitions and first properties

Fix a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and consider an increasing family of sub–σ–fields of $\mathcal{F}$,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$$

that is called a ***filtration***. It is said that a sequence of random variables (also called discrete time stochastic process) $\{X_n,\, n \geq 0\}$ is a martingale relative to $\{\mathcal{F}_n,\, n \geq 0\}$, or that $\{(X_n, \mathcal{F}_n),\, n \geq 0\}$ is a martingale, if

1. $E[|X_n|] < \infty, \;\; \forall n \geq 0.$

2. X_n is $\mathcal{F}_n$ measurable $\forall n \geq 0$.

3.
$$\boxed{\mathbf{E}[X_{n+1} \,|\, \mathcal{F}_n] = X_n, \;\; \forall n \geq 0.} \tag{4.1}$$

Some times it is convenient to start the martingale at time 1: $X_1,\, X_2, \ldots,$ or to consider the martingale for n in a finite rang, or for negative numbers, etc. There is no difficulty in doing the corresponding adaptations.

We will interpret $\mathcal{F}_n$ as the information a time n: at that moment, of a result $\omega \in \Omega$ we only know if every one of the events $A \in \mathcal{F}_n$ contains or not ω, that is, if A has happen or not. A stochastic process $\{X_n,\, n \geq 0\}$ that satisfies that for each n the random variable X_n is $\mathcal{F}_n$ measurable, is called

adapted to the filtration $\{\mathcal{F}_n,\, n \geq 0\}$. Intuitively, this is interpreted as that at time n, the observer know the value of X_n. Many times, but not always, we will take

$$\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}.$$

Indeed, when the filtration is not specified, it is assumed that the filtration is that one. In such a case, the condition (4.1) is written

$$\boxed{\mathbf{E}[X_{n+1} \,|\, X_0, \dots, X_n] = X_n.} \tag{4.2}$$

It is worth to remark that if $\{X_n,\, n \geq 0\}$ is a martingale, then

$$\mathbf{E}[X_0] = \mathbf{E}[X_1] = \cdots .$$

When in the expression (4.1) it is changed the equality for an inequality, then the processes are called sub or supermartingale. Specifically, it is said that $\{X_n,\, n \geq 0\}$ is a ***submartingale*** (respectively ***supermartingale***) relative to $\{\mathcal{F}_n,\, n \geq 0\}$ if

1. $\mathbf{E}[|X_n|] < \infty$.

2. X_n is $\mathcal{F}_n$ measurable.

3.

$$\mathbf{E}[X_{n+1} \,|\, \mathcal{F}_n] \geq X_n.$$

(resp. $\mathbf{E}[X_{n+1} \,|\, \mathcal{F}_n] \leq X_n$.)

If $\{X_n,\, n \geq 0\}$ is a submartingale,

$$\mathbf{E}[X_0] \leq \mathbf{E}[X_1] \leq \cdots .$$

For a supermartingale the reversed inequalities are obtained.

If $\{X_n,\, n \geq 0\}$ is a submartingale (resp. super), then $\{-X_n,\, n \geq 0\}$ is a supermartingale (resp. sub). By this reason, it suffices to consider one of the two cases to study the properties.

Examples

1. Here is the main example of martingale (and sub and super). Let $\{Y_n,\, n \geq 1\}$ a sequence of independent random variables, all with finite expectation. Define

$$X_n = \sum_{j=1}^{n} Y_j,$$

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n) = \sigma(Y_1, \dots, Y_n).$$

Then, from the properties of the conditional expectation,

$$\mathbf{E}[X_{n+1} \,|\, X_1, \dots, X_n] = \sum_{j=1}^{n} Y_j + \mathbf{E}[Y_{n+1} \,|\, Y_1, \dots, Y_n] = X_n + \mathbf{E}[Y_{n+1}].$$

Hence:

- If $\mathsf{E}[Y_n] = 0$, for all n, then X is a martingale.
- If $\mathsf{E}[Y_n] \geq 0$, for all n, then X is a submartingale.
- If $\mathsf{E}[Y_n] \leq 0$, for all n, then X is a supermartingale.

A martingale is a generalization of a sum of independent centered random variables. The main idea is to change the independence between the *increments* of the sum by a weaker property.

2. As a particular case of the previous example, consider a person that plays with a fair coin; if the result is Head she wins 1 €, if Tail, she losses 1 €. The total gain after the first play is

$$Y_1 = \begin{cases} 1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}. \end{cases}$$

So,

$$\mathsf{E}[Y_1] = 0.$$

Remember that when this condition is fulfilled, it is said that the game is ***fair***: If the person plays many times, the mean value of gains will be 0: she just expends time. In the casino games, the expectation of the player is negative: a player can have good or bad luck, but if she plays many times, the result is always a loss.

Now, assume that the player can loss a infinity of euros and she is playing indefinitely. Denote by $Y_2, Y_3, \ldots,$ the gain or loss of the second, third, etc. plays, and by

$$X_n = \sum_{j=1}^{n} Y_j$$

the total gain or loss after the n play. It is a martingale: if the evolution of the total gain until play n is known, $X_1, \ldots, X_n$, then, in the mean, after the play $n+1$ the total gain or loss is X_n:

$$\mathsf{E}[X_{n+1} \,|\, X_1, \ldots, X_n] = X_n.$$

This property wants to capture the fact that the game is fair at every time.

3. Following the previous example, if the game is unfair against the player, that means with probability of Head $p < 1/2$, then

$$\mathsf{E}[Y_n] = p - (1-p) = 2p - 1 < 0.$$

So, $\mathsf{E}[\sum_{j=1}^{n+1} Y_j \,|\, Y_1, \ldots, Y_n] = \sum_{j=1}^{n} Y_j + \mathsf{E}[Y_{n+1}] < \sum_{j=1}^{n} Y_j$. and,

$$X_n = \sum_{j=0}^{n} Y_j$$

is a supermartingale.

4. A typical way to build a martingale is the following: Take a random variable Z with finite expectation and a filtration $\{\mathcal{F}_n,\, n \geq 0\}$. Define

$$X_n = \mathbf{E}[Z \,|\, \mathcal{F}_n].$$

Then $\{(X_n, \mathcal{F}_n),\, n \geq 0\}$ is a martingale. (Please, check).

Properties (easy to prove)

1. Let $\{(X_n, \mathcal{F}_n),\, n \geq 0\}$ be a martingale. From the property (4.1) and the tower property of conditional expectation, it follows

$$\mathbf{E}[X_m \,|\, \mathcal{F}_n] = X_n, \;\; \forall m \geq n.$$

2. Consider two martingales (respectively submartingales)$\{X_n,\, n \geq 0\}$ and $\{Y_n,\, n \geq 0\}$ with respect to the same filtration $\{\mathcal{F}_n,\, n \geq 0\}$ and $a, b \in \mathbb{R}$ (resp. $a, b \geq 0$). Then $aX_n + bY_n$ is a martingale (resp. submartingalae) with respect to the same filtration.

3. Let $\{(X_n, \mathcal{F}_n),\, n \geq 0\}$ be a martingale (respectively a submartingale) and $f : \mathbb{R} \to \mathbb{R}$ a convex function (resp. convex i increasing) such that $\mathbf{E}[|f(X_n)|] < \infty, \forall n$. Then $\{(f(X_n), \mathcal{F}_n),\, n \geq 0\}$ is a submartingale.

4.2 Martingale transforms

We continue with the example of the player, now assuming that she uses a more sophisticate strategy. She beats 1 € the first play and if she losses, then beats 2 € in the next play. In this way, the player duplicates the bet when she losses; when she wins, the game starts freshly, and the player bets 1 €. This procedure is called *to make a martingale* (see the James Bond's adventure *Casino Royal* by Ian Fleming). That system is defined by the sequence of the results of the money that we codify by $\{Y_n,\, n \geq 1\}$

$$Y_n = \begin{cases} \;\;1, & \text{with probability } \frac{1}{2}, \text{(Head at time } n) \\ -1, & \text{with probability } \frac{1}{2}, \end{cases}$$

(note that they are the same variables as in Example 2 of page 45, although their meaning is different) and by the sequence of the random variables $\{H_n,\, n \geq 1\}$ where H_n is the quantity bet at time n (which is decided before the result Y_n):

$$H_1 = 1 \quad \text{(first bet)},$$

and for $n \geq 2$

$$H_n = \begin{cases} 2H_{n-1}, & \text{si } Y_{n-1} = -1 \text{ (losses at the previous play)}, \\ 1, & \text{si } Y_{n-1} = \;\;1 \text{ (wins at the previous play and starts the game)}. \end{cases}$$

The total gain (positive or negative) of the player after play n is

$$M_n = \sum_{i=1}^{n} H_i Y_i.$$

First Head at play number	Result Y_n	Bet H_n	Partial gain H_nY_n	Total gain M_n
1	1	1	1	1
2	-1	1	-1	
	1	2	2	1
3	-1	1	-1	
	-1	2	-2	
	1	4	4	1

Table 4.1. Wins and losses with the *martingale strategy*

Table 4.1 gives the first steps of M_n for different cases.

So, if the player losses the first k plays and winds at $k+1$, the final gain is

$$-1 - 2 - \cdots - 2^{k-1} + 2^k = 1.$$

That means that always that the player wins, she wins 1 €, whereas if the player encounters a bad sequence of results, she can loss much money (the bets increases as powers of 2). That is the reason why that strategy does not work.

The process M_n is a $\mathcal{F}_n$ martingale (please, check!):

$$\mathbf{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n,$$

and the player with that strategy is playing a fair game. It is also worth to remark that that strategy is *honest*, since H_n is predictable (it is $\mathcal{F}_{n-1}$ measurable): To decide the quantity to bet a time n only the information until result $n-1$ is used.

Finally, remember that the martingale that gives the total gains without strategy were

$$X_n = \sum_{j=1}^{n} Y_n,$$

(put also $X_0 = 0$) and then

$$Y_j = X_j - X_{j-1}.$$

Then

$$M_n = \sum_{j=1}^{n} H_j Y_j = \sum_{j=1}^{n} H_j \big(X_j - X_{j-1}\big).$$

Below we will see that if the game is (strictly) unfair with this strategy (or with any honest strategy) it is impossible to convert it in a favorable. First, we take a bit of perspective.

Let $X = \{X_n,\, n \geq 0\}$ and $H = \{H_n,\, n \geq 0\}$ two stochastic process. The process $H \cdot X$ defined by

$$(H \cdot X)_n := H_0 X_0 + \sum_{j=1}^{n} H_j \big(X_j - X_{j-1}\big), \quad n \geq 1.$$

is called the **transformation of X by H**. To get a stochastic process with good properties there are needed some conditions on X and H. Consider a filtration $\{\mathcal{F}_n,\ n \geq 0\}$. We will assume that

- The process X is martingale (or sub or supermartingale) with respect to $\{\mathcal{F}_n,\ n \geq 0\}$.

- The process H satisfies:

$$\begin{array}{lll} H_0 & \text{is} & \mathcal{F}_0 \text{ mesurable} \\ H_n & \text{is} & \mathcal{F}_{n-1} \text{ mesurable, } \forall n \geq 1. \end{array}$$

  A process that satisfies that property is called **predictable**. Intuitively, it means that at time $n-1$ the value of H_n is known.

In the example of the player using a martingale strategy, take $H_0 = X_0 = 0$, and

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n),\ n \geq 0.$$

(note that $\mathcal{F}_0 = \{\emptyset, \Omega\}$) The process $\{H_n,\ n \geq 0\}$ is predictable and the total gain is $(H \cdot X)_n$. Without strategy, the bet is $G_n = 1$, and the total gain is $(G \cdot X)_n = X_n$.

Theorem of martingale transforms. Let $\{(X_n, \mathcal{F}_n),\ n \geq 0\}$ be a martingale (respectively, sub o supermartingale) and $\{H_n,\ n \geq 0\}$ be a bounded predictable process (resp. predictable, bounded and positive). Then

$$H \cdot X = \{\big((H \cdot X)_n, \mathcal{F}_n\big),\ n \geq 0\}$$

is a martingale (resp. sub o supermartingale.)

Proof. We do the proof for martingales; the other cases need minor changes. We use that conditional expectation is linear and H is predictable:

$$\begin{aligned} \mathbf{E}\Big[(H \cdot X)_{n+1} \,|\, \mathcal{F}_n\Big] &= (H \cdot X)_n + \mathbf{E}\big[H_{n+1}(X_{n+1} - X_n) \,|\, \mathcal{F}_n\Big] = \\ &= (H \cdot X)_n + H_{n+1}\, \mathbf{E}[X_{n+1} - X_n \,|\, \mathcal{F}_n\Big] = (H \cdot X)_n. \end{aligned}$$

4.3 Stopping times

This definition is independent of martingale ideas. Later we will mix all things. As we are doing in this chapter, we will restrict ourselves to the discrete case.

A player is looking for a good moment to stop playing and return home. She could decide to stop at the play $n = 20$, or the first time that his capital is greater than 1000 €, or before doing a bad play and drawback. The first form is *deterministic* (there is no relationship with the result of the game), whereas the other two are random: he does not know if that event will happen at play 26, or 50 or never. However, to stop when his capital is greater than 1000 € can be done by an honest player, but to stop before drawback it is necessary to know the result of the next play, so to advance the future in some way.

To provide a good definition of what rules are *honest* was a extraordinary contribution of the great mathematician J. L. Doob.

The previous ideas are very related with the information concept, that implies, with σ-fields. Consider a measurable space $(\Omega, \mathcal{F})$ (here the probability has no role) and a filtration $\{\mathcal{F}_n,\, n \geq 0\}$. We will say that

$$T : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$$

is a **stopping time** if

$$\forall n \in \mathbb{N}, \quad \{T = n\} \in \mathcal{F}_n \,.$$

To study a few examples it is convenient to introduce a stochastic process $X = \{X_n,\, n \in \mathbb{N}\}$, where X_n is the capital of a player after the play n. Let

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n)\,,$$

that is interpreted as the information available after the n play (deduced from the capital of the player). Then, to stop at play 20 is given by the map

$$\begin{aligned} T &: \Omega \longrightarrow \mathbb{N} \\ T(\omega) &= 20 \end{aligned}$$

hence,

$$\{T = n\} = \begin{cases} \emptyset, & \text{if } n \neq 20, \\ \Omega, & \text{if } n = 20, \end{cases}$$

and is a stopping time.

To stop when the capital is greater than 1000€ is:

$$T : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$$

$$T(\omega) = \inf\{n > 0 : \, X_n(\omega) \geq 1000\},$$

(with the convention $\inf(\emptyset) = \infty$) and then

$$\{T = n\} = \{X_1 < 1000, \ldots, X_{n-1} < 1000, X_n \geq 1000\} \in \mathcal{F}_n \,,$$

Finally, to stop before drawback is

$$T = \inf\{n \geq 0 : X_{n+1} \leq 0\}$$

and in the event $\{T = n\}$ there is involved X_{n+1}, and hence T is not $\mathcal{F}_n$-measurable.
Value of a process at a stopping time Given a stochastic process $X = \{X_n,\, n \geq 0\}$ and a finite stopping time (that is, $T = \infty$ has probability zero), then the random variable

$$X_T : \Omega \longrightarrow \mathbb{R},$$

$$X_T(\omega) = X_{T(\omega)}(\omega)$$

is called the **value of the process X at T.**. When T can be infinite, then a bit of precaution is needed.

Main example: Hitting time of a set by a stochastic process. Let $\{X_n,\, n \geq 0\}$ be an adapted stochastic process, that is, X_n is $\mathcal{F}_n$ measurable, $\forall n$. Consider a Borelian set $G \in \mathcal{B}(\mathbb{R})$. Define

$$T_G(\omega) = \inf\{n \geq 0 : \, X_n(\omega) \in G\}\,,$$

(with $\inf(\emptyset)=\infty$). It is a stopping time and we have

$$X_n \notin G \quad \text{on } \{n < T_G\},$$
$$X_{T_G} \in G \quad \text{on } \{T_G < \infty\}.$$

(It has also sense the stopping time $S_G(\omega)=\inf\{n>0:\ X_n(\omega)\in G\}$).

We remark that in the example of the player, the second stopping rule is T_G, where $G=\{1000, 1001,\dots\}$.

Barrier type derivatives. We here anticipate some notions about derivative products. Today is time $n=0$ and denote by $S_0, S_1\dots$ the price of an asset. An European derivative product its a contract signed today between two parts that has a value at a fixed time N depending on the evolution of the prices $S_0,\dots,S_N$. For example, the value for one part could be $(S_N-K)^+$ (an European call with strike K and maduration time N). In barrier type derivative the value depends if at some time $n\in\{0,\dots,N\}$, the price S_n crosses or not some upper or lower barrier. To be more concrete. Assume $S_0=25$, $N=10$, $K=27$ and we consider an upper barrier at $b=30$. This is called an **up-and-out option**. See Figure 4.1 for a possible path of that barrier option.

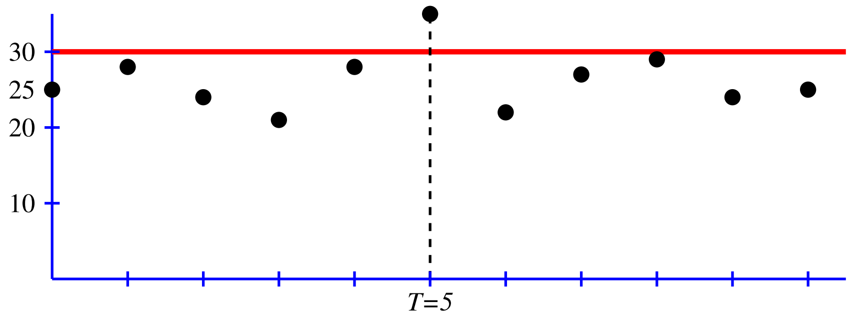


Figure 4.1. Example of barrier option

The time where the barrier is touched (if it is) before the option expires is

$$T=\inf\{0\le n\le 10:\ S_n\ge 30\}.$$

For mathematical coherence, we can put $T=11$ if S does not touch the barrier. The value of the process given in Figure 4.1 at time T is (for this particular path)

$$S_T(\omega)=35,$$

and the value at time 10 of a barrier option with the path given in Figure 4.1 is zero. The formula for that value is

$$V=\begin{cases}(S_{10}-27)^+, & \text{if } \max\{S_0,\dots,S_{10}\}<30,\\ 0, & \text{otherwise,}\end{cases}$$

that can also be written using the stopping time T:

$$V=(S_{10}-27)^+\,\mathbf{1}_{\{T\ge 11\}}.$$

Given a stopping time T it is said that the event $A\in\mathcal{F}$ is **anterior** to T if $\forall n\in\mathbb{N},\ A\cap\{T=n\}\in\mathcal{F}_n$. In order to interpret this cryptic property, assume $\mathcal{F}_n=\sigma(X_0,\dots,X_n)$. Therefore, if the result the experiment is ω and we know $T(\omega) = n$, then from $X_0(\omega), \ldots, X_n(\omega)$ it can be deduced if the event had happened or not. In the barrier option example, $A = \{\min\{S_1, \ldots, S_10\} \geq 20\}$ is not anterior to T, since when $T = 5$ it is impossible to say if A had happen or not. On the contrary, $A = \{\min\{S_1, \ldots, S_T\} \geq 20\}$ is anterior to T.

The family of events anterior to T

$$\mathcal{F}_T = \{A \in \mathcal{F} : \ A \cap \{T = n\} \in \mathcal{F}_n, \, \forall n\}$$

is a σ-field (check!), that naturally is called the σ-field of the events anterior to T, and is interpretated as the information known until time T.

Given an adapted process $\{X_n, \, n \geq 0\}$ and a finite stopping time, it turns out that X_T is $\mathcal{F}_T$-measurable.

Stopped stochastic process

Consider a stochastic process $\{X_n, \, n \geq 0\}$ and a stopping time. The process $\{X_n^T, \, n \geq 0\}$ defined by

$$X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega),$$

where $a \wedge b = \min(a, b)$, is called the **process stopped at** T. See Figure 4.2 for the example of the stopped process S_n^T.

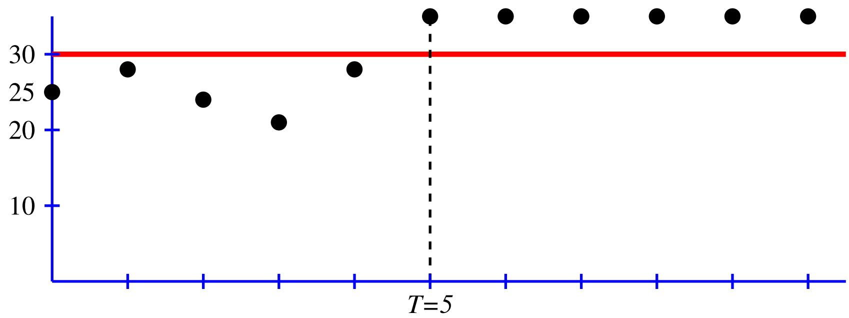


Figure 4.2. A trajectory of the stopped process

4.4 Martingales and stopping times

The first interesting property is that a stopped martingale (or sub or super) is also a martingale (or sub or sup) Specifically, let $\{X_n, \, n \geq 0\}$ be a martingale (sub or sup) with respect to a filtration $\{\mathcal{F}_n, \, n \geq 0\}$ and T a stopping time. Then $\{X_n^T, \, n \geq 0\}$ is a martingale (sub or super). That property is based in the Theorem of martingale transforms and it suffices to observe that putting

$$H_n(\omega) = \begin{cases} 1, & \text{if } T(\omega) \geq n, \\ 0, & \text{otherwise,} \end{cases}$$

in other words,

$$H_n = \mathbf{1}_{\{T \geq n\}},$$

then

$$X_n^T = (H \cdot X)_n.$$

(Check). The process H_n is predictable since

$$\{T \geq n\} = \{T < n\}^c = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}.$$

And the proof is finished applying the above mentioned theorem.

We had commented that for a martingale (sub or super) $\{X_n,\, n \geq 0\}$ (sub o super) we have

$$\mathbf{E}[X_m \,|\, \mathcal{F}_n] = X_n, \;\; m \geq n.$$

(inequalities for sub o super). The **Theorem of optional stopping** states that if we choose the subindex at random, but honestly, that formula is correct. More specifically, if $S \leq T$ are stopping times (with additional conditions), then

$$\mathbf{E}[X_T \,|\, \mathcal{F}_S] = X_S.$$

(respectively, $\mathbf{E}[X_T \,|\, \mathcal{F}_S] \geq X_S$ or $\mathbf{E}[X_T \,|\, \mathcal{F}_S] \leq X_S$.)

4.5 Exercises

1. Consider a sequence of i.i.d. random variables $\{Y_n,\, n \geq 1\}$,

$$\mathbf{P}\{Y_n = 1\} = p \quad \text{and} \quad \mathbf{P}\{Y_n = -1\} = q$$

with $(p + q = 1)$. Write

$$X_n = \sum_{j=1}^{n} Y_j$$

and $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n) = \sigma(X_1, \dots, X_n)$. Finally, put

$$Z_n = \left(\frac{q}{p}\right)^{X_n}.$$

Prove that $\{(Z_n, \mathcal{F}_n),\, n \geq 1\}$ is a martingale.

2. Consider the following sequence of random variables: X_1 is uniform on $(0, 1)$. Given $X_1 = x_1, \dots, X_n = x_n$, X_{n+1} is uniform on $(0, x_n)$. Prove that $\{X_n,\, n \geq 1\}$ is a supermartingale.

3. At time 1, an urn contains one white ball and one black ball. We extract a ball and we replace it by two balls of the same color, so we will have a new composition of the urn; at time 2 we extract a ball from the new urn, and replace, etc. We repeat the procedure indefinitely. Denote by Y_n the number of white balls at time n and by $X_n = Y_n/(n+1)$ the rate of white balls. Check that $\{X_n,\, n \geq 1\}$ is a martingale.

Indication. Note that

$$\mathbf{E}[Y_{n+1} \,|\, Y_1, \dots, Y_n] = \mathbf{E}[Y_{n+1} \,|\, Y_n],$$

compute $\mathbf{E}[Y_{n+1} \,|\, Y_n = j]$, and deduce $\mathbf{E}[Y_{n+1} \,|\, Y_n]$.

4. Fix an integer k. Prove that $T(\omega) = k,\ \forall \omega \in \Omega$, is a stopping time. Prove also that $\mathcal{F}_T = \mathcal{F}_k$.

5. Let S and T two stoping times. Then $S \wedge T = \min(S,T)$ and $S \vee T = \max(S,T)$ are also stopping times.

6. Prove that T is $\mathcal{F}_T$-measurable. (It suffices to show that $\forall j, \{T = j\} \in \mathcal{F}_T$.)

7. Let T a finite stopping time and $\{X_n,\ n \geq 0\}$ a stochastic process. Prove that X_T is a measurable, that is, that for all Borelian set $B \in \mathcal{B}(\mathbb{R})$, we have $\{X_T \in B\}$.

Hint.

$$\{X_T \in B\} = \bigcup_n \{X_T \in B, T = n\} = \bigcup_n \{X_n \in B, T = n\}.$$

Chapter 5

Markov processes

5.1 Introductory example

Compare the two price dynamics shown in Figures 5.1 and 5.2

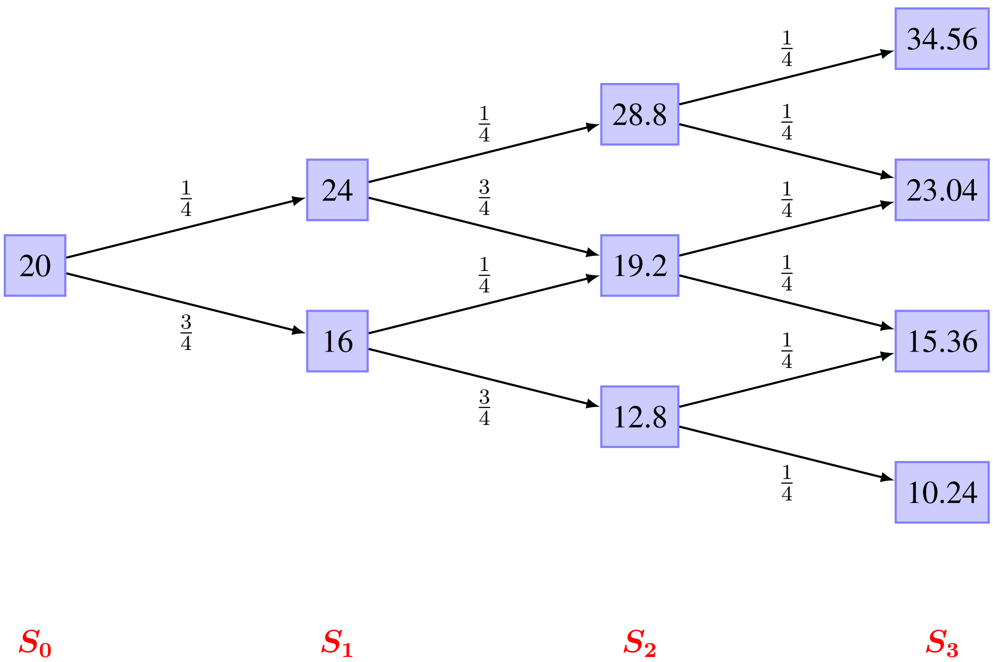


Figure 5.1. A Markov process

The main difference between both stochastic processes is that in the one given in Figure 5.1,

$$\mathbf{P}(S_3 = 23.04 \,|\, S_2 = 19.2, S_1 = 24) = \mathbf{P}(S_3 = 23.04 \,|\, S_2 = 19.2, S_1 = 16) = \frac{1}{4}$$

So if we know that we are at $S_2 = 19.2$, the way as we arrive to that point, through $S_1 = 24$ or $S_1 = 16$, does not matter. On the contrary, in Figure 5.2 we have

$$\mathbf{P}(S_3 = 23.04 \,|\, S_2 = 19.2, S_1 = 24) = \frac{1}{5} \neq \mathbf{P}(S_3 = 23.04 \,|\, S_2 = 19.2, S_1 = 16) = \frac{1}{3},$$

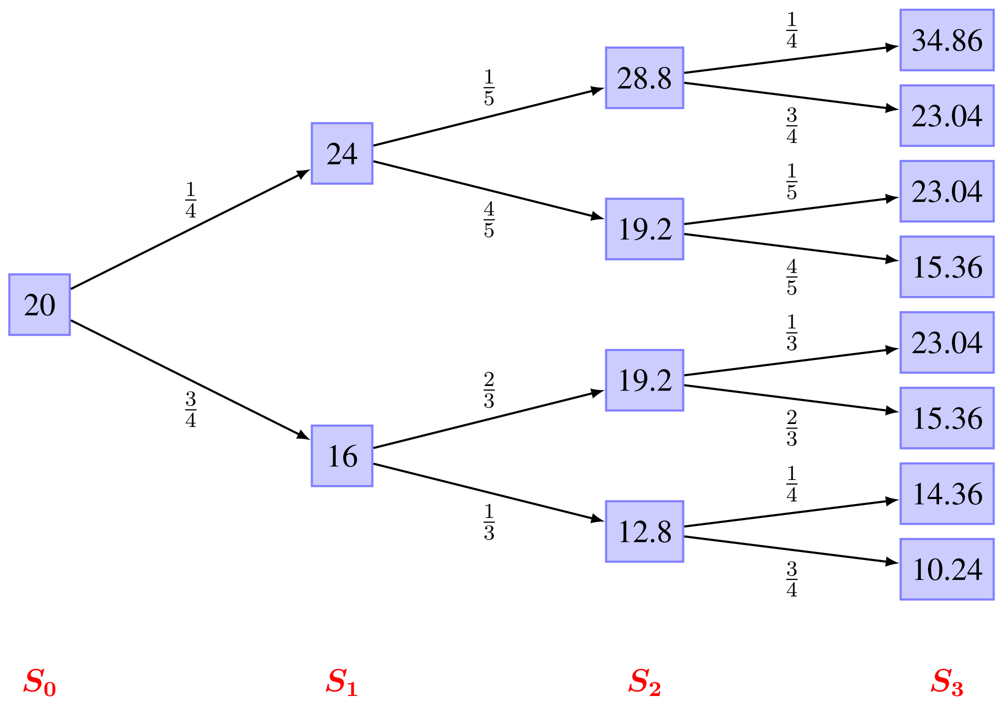


Figure 5.2. A non Markov process

so you need to know all the *trajectory* to $S_2 = 19.2$ in order to compute the probabilities from this point. It is said that the first stochastic process is a Markov (or Markovian) process, or to satisfy the Markov property. Specifically,

A stochastic process $\{X_n,\, n \geq 0\}$ is said to be a Markov process if for every $m > n \geq 1$ and every Borelian set $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{P}(X_m \in B \,|\, X_n, \ldots, X_0) = \mathbf{P}(X_m \in B \,|\, X_n).$$

Remarks.

1. If all the variables involved are discrete, the Markov property is that for every $m > n \geq 1$ and every possible values $a_0, \ldots, a_n, a_m$ of the variables involved,

$$\mathbf{P}(X_m = a_m \,|\, X_n = a_n, \ldots, X_0 = a_0) = \mathbf{P}(X_m = a_m \,|\, X_n = a_n),$$

where all the probabilities can be computed by elementary methods.

2. As in the martingale definition, we can introduce a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$. An adapted process $\{X_n,\, n \geq 0\}$ is said to be Markov with respect to that filtration if

$$\mathbf{P}(X_m \in B \,|\, \mathcal{F}_n) = \mathbf{P}(X_m \in B \,|\, X_n).$$

In the initial definition we take $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$.

An equivalent definition of Markov processes can be given using conditional expectations. We write the most general form:

The (adapted) process $\{X_n,\, n \geq 0\}$ is Markov (with respect to the filtration $\{\mathcal{F}_n,\, n \geq 0\}$) if an only if for every $m > n$ and every function $f : \mathbb{R} \to \mathbb{R}$ positive or such that $\mathsf{E}[|f(X_m)|] < \infty$,

$$\mathsf{E}[f(X_m) \,|\, \mathcal{F}_n] = \mathsf{E}[f(X_m) \,|\, X_n].$$

To check that a process is Markov it suffices to prove the following property

The (adapted) process $\{X_n,\, n \geq 0\}$ is Markov if $\forall n \geq 0$ and every function $f : \mathbb{R} \to \mathbb{R}$ positive

$$\mathsf{E}[f(X_{n+1}) \,|\, \mathcal{F}_n] = g(X_n),$$

where g is a function.

Exercise. Prove that the above condition is sufficient for a Markov process.

5.2 The importance of Markov processes

What you are at this moment contains the whole message of what you were.
A Buddhist sentence[1]

Many phenomena in life can be modelled, at least approximately, or at least, for a period of time, as a Markov process. I used to say that when one is 20 years old, life is a Markov process, because one thinks that life starts again freshly every day, with every new purpose, not matter who you were before; unfortunately, when one is 50 years old, one believes that one's trajectory determines the probabilities of the future, and life becomes non-Markovian.

In a Markov process, computations simplify in a dramatic way; for example, to compute the value of some European derivatives at time n it is needed to compute

$$\mathsf{E}[X \,|\, S_n, \dots, S_1],$$

where (in some cases) X depends only on $S_{n+1}, S_{n+2}, \dots$, where $S_0, S_1, \dots$, is the price of an asset. If the price process is Markov, then it suffices to compute

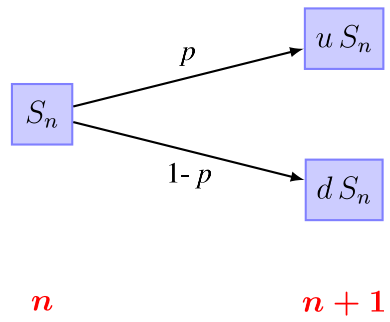
$$\mathsf{E}[X \,|\, S_n].$$

Even for n small, the difference in complexity is enormous. Markov processes with discrete time parameter and discrete random variables are called Markov chains; they are very well studied, and have lot of applications in real life. Markov processes with continuous time parameter are more difficult, and give rise to many practical and theoretical developments. In this course we use an infinitesimal part of the Markov processes results, but it will be very important. In particular, we will study two Markov processes: one in discrete time, the Cox-Ross-Rubinstein (or binomial) model, and the other in continuous time, the Brownian motion.

[1]Brenda Shoshanna, *Zen Miracles*

5.3 The binomial model

A binomial model for a price process $\{S_n,\ n = 0, 1, \dots\}$ starts at a deterministic value $S_0 = s_0 \neq 0$, and at every stage, from n to $n+1$ it can take two values, independently of the way that we arrived to S_n, as the following plot shows:



Where $u \neq d$ are fixed numbers, and we take $u > d$. For example, the Figure 1 we have $u = 1.2$ and $d = 0.8$. In a more formal way,

$$\mathbf{P}\{S_{n+1} = uS_n \,|\, S_n, \dots, S_0\} = p \quad \text{and} \quad \mathbf{P}\{S_{n+1} = dS_n \,|\, S_n, \dots, S_0\} = 1 - p. \tag{5.1}$$

This process is Markov, and is in the core of the so-called Cox-Ross-Rubistein model.

5.4 A two–dimensional Markov process

When we study exotic options in the Cox-Ross-Rubinstein model we need to consider the binomial model of previous section, $\{S_n,\ n = 0, 1, \dots\}$ and the so-called **running maxima** process $\{M_n,\ n = 0, 1, \dots,\}$ defined by

$$M_n = \max\{S_0, \dots, S_n\}.$$

For example, in the chapter of martingales we consider a up-and-out call barrier option with strike price $K = 27$, $N = 10$ and upper barrier $b = 30$, that is, if in some time spot $S_n > 30$, then the value of the option is zero, otherwise, the value at time $N = 10$ is

$$(S_{10} - 27)^+.$$

That condition can be written as

$$(S_{10} - 27)^+ \, \mathbf{1}_{\{M_{10} < 30\}}.$$

So the payoff of the option (the above expression) is a function of the two-dimensional process

$$\{(S_n, M_n),\ n = 0, 1, \dots, N\}.$$

Property. The two-dimensional process $\{(S_n, M_n),\ n = 0, 1, \dots\}$ is Markov with respect to the filtration $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$.

Proof. We will prove (see page 57) that for an arbitrary positive function $f : \mathbb{R}^2 \to \mathbb{R}$, and $n \geq 0$

$$\mathbf{E}[f(S_{n+1}, M_{n+1}) \,|\, \mathcal{F}_n] = g(S_n, M_n), \tag{5.2}$$

for a certain function g. To this end, write

$$R_j = \frac{S_{j+1}}{S_j},\ j = 0, 1, \dots,$$

and note that $R_0, R_1, \dots,$ are i.i.d. random variables with probabilities

$$P(R_j = u) = p \quad \text{and} \quad P(R_j = d) = 1 - p$$

and R_n is independent of $\mathcal{F}_n$.

Denote by $a \vee b$ the maximum between a and b:

$$a \vee b = \max\{a, b\}.$$

We have that

$$M_{n+1} = M_n \vee S_{n+1}.$$

Then

$$f(S_{n+1}, M_{n+1}) = f(S_{n+1}, M_n \vee S_{n+1}) = f\big(S_n R_n, M_n \vee \big(S_n R_n\big)\big).$$

It follows that

$$\mathbf{E}[f(S_{n+1}, M_{n+1}) \,|\, \mathcal{F}_n] = \mathbf{E}[f(S_{n+1}, M_{n+1}) \,|\, S_0, \dots, S_n] = \mathbf{E}\big[f\big(S_n R_n, M_n \vee \big(S_n R_n\big)\big) \,|\, S_0, \dots, S_n\big]$$

and given that S_n and M_n are measurable respect to $\sigma(S_0, \dots, S_n)$, and R_n is independent of that σ-field, we have

$$\mathbf{E}\big[f\big(S_n R_n, M_n \vee \big(S_n R_n\big)\big) \,|\, S_0, \dots, S_n\big] = \mathbf{E}\big[f\big(x R_n, y \vee (x R_n)\big)\big]\big|_{x=S_n, y=M_n},$$

and

$$\mathbf{E}\big[f\big(x R_n, y \vee (x R_n)\big)\big] = ph\big(xu, y \vee (xu)\big) + (1-q)h\big(xd, y \vee (xd)\big).$$

Hence,

$$\mathbf{E}[f(S_{n+1}, M_{n+1}) \,|\, \mathcal{F}_n] = pf\big(S_n u, M_n \vee (S_n u)\big) + (1-q)f\big(S_n d, M_n \vee (S_n d)\big). \tag{5.3}$$

Now take $g(x, y) = pf\big(xu, y \vee (xu)\big) + (1-q)f\big(xd, y \vee (xd)\big)$, and it is deduced (5.2).

5.5 Appendix. Proof that the binomial model is Markov

From (5.1) is is quite intuitive that the process is Markov. For sceptic readers we sketch a proof. Observe that by the tower property, the first part of (5.1) implies

$$\mathbf{P}\{S_{n+1} = uS_n \,|\, S_n\} = \mathbf{P}\{S_{n+1} = uS_n \,|\, S_n, \dots, S_0 \,|\, S_n\} = p,$$

So (1) implies

$$\mathbf{P}\{S_{n+1} = uS_n \,|\, S_n, \dots, S_0\} = \mathbf{P}\{S_{n+1} = uS_n \,|\, S_n\} = p. \tag{5.4}$$

Note that $S_0, \dots, S_n, \dots$ are discrete random variables, and S_n can take only 2^n different values:

$$u^n s_0,\ u^{n-1} d s_0,\ \dots, d^n s_0,$$

which can be written in a generic form as $u^{n-i}d^i s_0$, for $i = 0, \dots, n$. Not all sequences

$$s_0, u^{1-i}d^i s_0, u^{2-i}d^i s_0, \dots, u^{n-i}d^i s_0, \dots$$

have positive probabilities. Take one of this sequences with positive probability and call it $x_1, \dots, x_n$ (it is not needed to write s_0). Then, by the definition of conditional probability,

$$\begin{aligned}
\mathsf{P}(S_{n+1} = uS_n \mid S_n = x_n, \dots, S_1 = x_1) &= \frac{\mathsf{P}(S_{n+1} = uS_n, S_n = x_n, \dots, S_1 = x_1)}{\mathsf{P}(S_n = x_n, \dots, S_1 = x_1)} \\
&= = \frac{\mathsf{P}(S_{n+1} = ux_n, S_n = x_n, \dots, S_1 = x_1)}{\mathsf{P}(S_n = x_n, \dots, S_1 = x_1)} \\
&= \mathsf{P}(S_{n+1} = ux_n \mid S_n = x_n, \dots, S_1 = x_1).
\end{aligned}$$

In a similar way, the second term of (5.4) is

$$\mathsf{P}(S_{n+1} = uS_n \mid S_n = x_n) = \mathsf{P}(S_{n+1} = ux_n \mid S_n = x_n).$$

And the Markov property follows.

Chapter 6

Discrete market models

6.1 Introduction

In a discrete model for a market the time is counted in discrete units like hours, days, weeks, etc., and it is denoted by $n = 0, 1, \dots$ For a sake of concreteness, normally we use today for $n = 0$, and we speak of a *date* n; the adjustments for other units are obvious.

The first ingredient in a market is the so-called **riskless asset** that models the evolution of the money in a trustworthy bank account (there exists?) or in sure government bonds: then, the riskless asset can be interpreted as a bank account that at time 0 has 1 €, and at time n has B_n €. In this chapter we will assume that along the time $n = 0, \dots, N$ the bank gives compound interest with rate $r \geq 0$ by unity of time. Hence the value of one unit of riskless asset at time n is

$$B_n = (1+r)^n, \; n = 0, \dots, N.$$

On the other hand, in the market there are **risky assets** like shares, commodities, foreign currencies, etc. If there are d risky assets we will denote by

$$S_n^{(j)} \geq 0, \; j = 1, \dots, d,$$

the price of the asset j at time n (when $d = 1$ we simply put S_n instead of $S_n^{(1)}$). The initial price $S_0^{(j)}$ is known, however, for $n \geq 1$, $S_n^{(j)}$ is a random variable that will be known at time n. Therefore, in the basis of the model there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots$ As is usual, $\mathcal{F}_n$ is interpreted as the information at time n. Then, the random variable $S_n^{(j)}$ is assumed $\mathcal{F}_n$ measurable; in more technical words, for $j = 1, \dots, d$, the stochastic process $\{S_n^{(j)}, \, n = 0, \dots, N\}$ is adapted to $\{\mathcal{F}_n, \, n = 0, \dots, N\}$. Since the prices at time 0 are known, we will take $\mathcal{F}_0 = \{\emptyset, \, \Omega\}$. Many times we will assume that the filtration has been generated by these processes.

6.2 Portfolios

A **portfolio** is a combination of money, shares, commodities, financial products, etc. that belongs to somebody, that we will call the *owner*. From a mathematical point of view, the simplest portfolio is a vector stochastic process $\mathbf{\Phi} = \{\mathbf{\Phi}_n, \, n = 0, \dots, N\}$, where

$$\mathbf{\Phi}_n = (D_n, H_n^{(1)}, \dots, H_n^{(d)}),$$

with

- D_n is the quantity of riskless assets at time n (not euros: see below). It may be positive or negative; in that last case it means that the owner of the portfolio has borrowed money, and she should return that money adding the corresponding interests; the rate of interest will be the same r. The value in euros of D_n (in this model) is

$$D_n\,B_n = D_n(1+r)^n.$$

- $H_n^{(j)}$ is the quantity of the risky asset j at time n. Also it may be positive or negative. When it is negative it is said that the owner has done **short selling**, that means that she has sold assets that they were not in the portfolio and a bank or someone else has lend to the owner; of course, the owner should return these assets at some time k at price $S_k^{(j)}$. It is worth noting that there is a complete symmetry between risky and riskless assets thanks to the use of the monetary unit B_n. Currently, short selling is allowed under some restrictions.

Finally, we assume that D_0 and $H_0^{(j)}$ are known, and for $n \geq 1$, D_n i $H_n^{(j)}$ are $\mathcal{F}_{n-1}$ measurable (the process $\mathbf{\Phi}$ is predictable), that implies that the quantity of money and assets that there will be in the portfolio at time n is decided (by the owner) with the information available at time $n-1$. We consider some examples.

Example 1. We assume that $r = 0.1$. A person (the owner) has a portfolio with

- 300 €.
- 200 shares of a business with price 3.75 € each (asset 1).
- 300 shares of a business with price 2.83 € each (asset 2).

This is the initial value of the portfolio, that we represent by

n	D_n	B_n	$H_n^{(1)}$	$S_n^{(1)}$	$H_n^{(2)}$	$S_n^{(2)}$
0	300	1	200	3.75	300	2.83

At time 1, the shares 1 down to 3.65 and the shares 2 up to 2.93. Then

n	D_n	B_n	$H_n^{(1)}$	$S_n^{(1)}$	$H_n^{(2)}$	$S_n^{(2)}$
0	300	1	200	3.75	300	2.83
1	300	1.1	200	**3.65**	300	**2.93**

At this moment, the owner buys 100 shares of type 2. This will give the composition of the portfolio at time 2, however, the owner decided that only with the information at time 1: the composition is said to be **predictable**. In this operation the owner expend $100 \times 2.93 = 293$ €. In the owner's account they remain $330 - 293 = 37$ € that corresponds to

$$\frac{37}{1.1} = 33.63 \text{ riskyless assets}$$

We represent it by

n	D_n	B_n	$H_n^{(1)}$	$S_n^{(1)}$	$H_n^{(2)}$	$S_n^{(2)}$
0	300	1	200	3.75	300	2.83
1	300	1.1	200	**3.65**	300	**2.93**
1+	33.63	1.1	200	3.65	**400**	2.93

At time 2 the owner knows the prices of the assets: The shares 1 down to 3.54 and the shares 1 up to 2.95. Then the owner can complete the following table

n	D_n	B_n	$H_n^{(1)}$	$S_n^{(1)}$	$H_n^{(2)}$	$S_n^{(2)}$
0	300	1	200	3.75	300	2.83
1	300	1.1	200	**3.65**	300	**2.93**
1+	33.63	1.1	200	3.65	**400**	2.93
2	33.63	1.21	200	**3.54**	400	**2.95**

The value (in euros) **of a portfolio** $\boldsymbol{\Phi}$ at time n is

$$V_n(\boldsymbol{\Phi}) = D_n B_n + \sum_{j=1}^{d} H_n^{(j)} S_n^{(j)}.$$

(When there is no confusion we write V_n instead $V_n(\boldsymbol{\Phi})$). So, we have another stochastic process $V = \{V_n,\ n = 0, \ldots, N\}$, that is adapted to the filtration $\{\mathcal{F}_n,\ n = 0, \ldots, N\}$.

The quantity $1/B_n = (1+r)^{-n}$ is called the **discount factor** of the prices at time n. The value

$$\widetilde{S}_n^{(j)} = (1+r)^{-n} S_n^{(j)}$$

is called the **discounted value** of the asset j, and

$$\widetilde{\boldsymbol{S}}_n = \left(1, S_n^{(1)}, \ldots, S_n^{(d)}\right)$$

the **vector of discounted prices**, and the **discounted value of the portfolio** is

$$\widetilde{V}_n = (1+r)^{-n} V_n = D_n + \sum_{j=1}^{d} H_n^{(j)} \widetilde{S}_n^{(j)}.$$

Following with the previous example, the value of the portfolio is

n	D_n	B_n	$H_n^{(1)}$	$S_n^{(1)}$	$H_n^{(2)}$	$S_n^{(2)}$	V_n	$\widetilde{V}_n$
0	300	1	200	3.75	300	2.83	1899	1899
1	300	1.1	200	**3.65**	300	**2.93**	1939	1762.73
1+	33.63	1.1	200	3.65	**400**	2.93	1939	1762.73
2	33.63	1.21	200	**3.54**	400	**2.95**	1928.69	1593.96

The final value of the portfolio is 1928.69 €, and the discounted value is 1593.96 €, so the portfolio has decreased its value.

In general, we will restrict ourselves to the called **admissible portfolios**, that are those such that $V_n \geq 0,\ n = 0, \dots, N$.

A portfolio $\mathbf{\Phi}$ is called **self–financing** if the change of the value between two times n and $n+1$ is only due to a recombination of the quantities of assets (including money) and there is no addition or extraction of assets or money; in other words, we can sell, buy, borrow, etc. but always reflecting the operations in the portfolio.

In the above example, we observe here that the $V(1+) = V(1)$: that means that this portfolio is a self-financing.

In order to give a formal expression of self-financing it is convenient to introduce a vectorial notation. Given two vectors $\boldsymbol{x} = (x_1, \dots, x_k)$ and $\boldsymbol{y} = (y_1, \dots, y_k)$, denote by $\boldsymbol{x} \cdot \boldsymbol{y}$ its scalar product:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{k} x_i y_i.$$

Consider the price's vector (including the riskless asset) at time n:

$$\boldsymbol{S}_n = \big(B_n,\, S_n^{(1)}, \dots, S_n^{(d)}\big),$$

and remember that the composition of the portfolio at time n is

$$\mathbf{\Phi}_n = (D_n, H_n^{(1)}, \dots, H_n^{(d)}).$$

Then the value of the portfolio is

$$V_n = \mathbf{\Phi}_n \cdot \boldsymbol{S}_n,$$

A portfolio is **self-financing** if for $n = 0, \dots, N-1$,

$$V_n = \mathbf{\Phi}_n \cdot \boldsymbol{S}_n = \mathbf{\Phi}_{n+1} \cdot \boldsymbol{S}_n. \tag{6.1}$$

That is, the total value of the portfolio at time n has been distributed in a different way in $\mathbf{\Phi}_{n+1}$: if the owner have borrowed money, then she has accounted a negative quantity and the final value have been not modified. However, since the prices change from $\boldsymbol{S}_n$ to $\boldsymbol{S}_{n+1}$, the value will change at $n+1$.

Remark. The current prices at time n are the prices measured in units of B_n. It is said that we are using B_n as a *numeraire*; it is possible to take as a numeraire other riskless or risky asset that have price $S_n > 0,\ \forall n$.

6.2.1 A property of self-financing

First we consider the case of a self–financed portfolio with the riskless asset and only one risky asset. Later we formulate the general case.

Property 6.2.1 *Consider a self-financing portfolio with the riskless asset and only one risky asset, $\mathbf{\Phi}_n = (D_n, H_n)$. Then*

$$\boxed{V_n = V_0 + \sum_{k=1}^{n} D_k\big(B_k - B_{k-1}\big) + \sum_{k=1}^{n} H_k\big(S_k - S_{k-1}\big).} \tag{6.2}$$

and

$$\boxed{\widetilde{V}_n = V_0 + \sum_{k=1}^{n} H_k\big(\widetilde{S}_k - \widetilde{S}_{k-1}\big).} \tag{6.3}$$

Proof. We prove formula (6.3); formula (6.2) is similar. First, since $V_n = D_n B_n + H_n S_n$, we have

$$\widetilde{V}_n = B_n + H_n \widetilde{S}_n. \tag{6.4}$$

On the other hand, from the condition of self-financing (6.1)

$$\widetilde{V}_n = D_n + H_n \widetilde{S}_n = D_{n+1} + H_{n+1}\widetilde{S}_n. \tag{6.5}$$

Staring with (6.4) for $n+1$,

$$\begin{aligned}\widetilde{V}_{n+1} &= D_{n+1} + H_{n+1}\widetilde{S}_{n+1} \\ &= D_{n+1} + H_{n+1}\big(\widetilde{S}_{n+1} - \widetilde{S}_n\big) + H_{n+1}\widetilde{S}_n, \\ &= \widetilde{V}_n + H_{n+1}\big(\widetilde{S}_{n+1} - \widetilde{S}_n\big),\end{aligned}$$

where the last equality follows from (6.5). Then

$$\begin{aligned}&\widetilde{V}_1 = V_0 + H_1\big(\widetilde{S}_1 - \widetilde{S}_0\big), \\ &\widetilde{V}_2 = \widetilde{V}_1 + H_2\big(\widetilde{S}_2 - \widetilde{S}_1\big) = V_0 + H_1\big(\widetilde{S}_1 - \widetilde{S}_0\big) + H_2\big(\widetilde{S}_2 - \widetilde{S}_1\big) \\ &\widetilde{V}_3 = \widetilde{V}_2 + H_3\big(\widetilde{S}_3 - \widetilde{S}_2\big) = \cdots = V_0 + \sum_{k=1}^{3} H_k\big(\widetilde{S}_k - \widetilde{S}_{k-1}\big).\end{aligned}$$

By induction we get formula (6.3).

The formula for a general portfolio with an arbitrary number of risky assets is the following:

Property 6.2.2 *If there are d risky assets in the portfolio,*

$$\boxed{\widetilde{V}_n = V_0 + \sum_{j=1}^{d}\sum_{k=1}^{n} H_k^{(j)}\big(\widetilde{S}_k^{(j)} - \widetilde{S}_{k-1}^{(j)}\big).} \tag{6.6}$$

6.3 Arbitrages

A self-financing portfolio is called an **arbitrage** if

a. $V_0 = 0$.

b. $V_N \geq 0,\ a.s.$

c. $\mathbf{P}(V_N > 0) > 0$. Equivalently, $\mathbf{E}[V_N] > 0$.

So, an arbitrage is a (self-financing) portfolio that starts with 0 value, and that at the end time there are no losses, and with a probability strictly positive there are some gains.

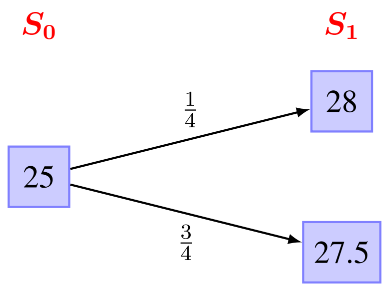


Figure 6.1. Price dynamics of Example 1

6.3.1 Examples.

Example 2. Let $r = 0.1$ and in the market there is a risky asset with initial price $S_0 = 25$ €. The price at time 1 can be 27.5 o 28 € with probabilities 3/4 i 1/4 respectively (see Figure 6.1).

We borrow 25 € and buy one asset.

- If $S_1 = 28$, we sell the asset and return the loan $25 \times 1.1 = 27.5$. We do a profit of $V_1 = 0.5$ €.
- If $S_1 = 27.5$, we sell the share and return the loan: $25 \times 1.1 = 27.5$. Then $V_1 = 0$

So, it is an arbitrage.

Example 3. With the same conditions as before, but now the asset's price at time 1 can be 27 or 27.5 € with probabilities 3/4 and 1/4 respectively (see Figure 6.2).

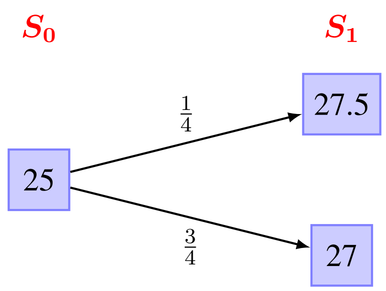


Figure 6.2. Price dynamics of Example 2

In this case, we sell short an asset and put the 25 € in a bank account. At time 1, we take the el 27.5 € from the bank, buy the asset and return it.

6.3.2 The one's price law

Teorema 6.3.1 *Consider a market where there are no arbitrages, and with two risky assets. Then*

$$S_1^{(1)} = S_1^{(2)},\ q.s. \quad \Longrightarrow \quad S_0^{(1)} = S_0^{(2)}.$$

Proof. Assume $S_0^{(1)} < S_0^{(2)}$ (remember that the initial prices are deterministic). At time 0 we sell short asset 2 and buy asset 1. At time 1 we sell asset 1 and return asset 2. This is an arbitrage, that is contradictory with the hypothesis. ■

Comment. This property is called the one's price law and says that the golden rule of business *buy low, sell high* is, in our jargon, an arbitrage.

6.3.3 Equivalent Probabilities

Assume that the today's price of one dollar is 0.75 € and that after 10 days the price will be

$$\begin{cases} 0.85€, & \text{with probability } 0.4 \\ 0.75€, & \text{with probability } 0 \\ 0.7€, & \text{with probability } 0.6, \end{cases}$$

A probabilistic model for such a setup is $\Omega = \{0.85, 0.75, 0.7\},\ \mathcal{F} = \mathcal{P}(\Omega)$ and

$$\mathbf{P}(0.85) = 0.4,\ \mathbf{P}(0.75) = 0,\ \mathbf{P}(0.7) = 0.6.$$

A different person thinks that the prices will be

$$\begin{cases} 0.85€, & \text{with probability } 0.3 \\ 0.75€, & \text{with probability } 0, \\ 0.7€, & \text{with probability } 0.7. \end{cases}$$

Then, Ω and $\mathcal{F}$ will be the same but the probabilities are

$$\mathbf{Q}(0.85) = 0.3,\ \mathbf{Q}(0.75) = 0,\ \mathbf{Q}(0.7) = 0.7.$$

Finally, a third person can think that the probabilities are

$$\mathbf{R}(0.85) = 0.2,\ \mathbf{R}(0.75) = 0.3,\ \mathbf{R}(0.7) = 0.5.$$

It is said that two probabilities $\mathbf{P}_1$ i $\mathbf{P}_2$ on the same space $(\Omega, \mathcal{F})$ are **equivalent** if both have the same sets of zero measure: For $A \in \mathcal{F}$,

$$\mathbf{P}_1(A) = 0 \iff \mathbf{P}_2(A) = 0$$

Hence, in the example, $\mathbf{P}$ are $\mathbf{Q}$ equivalent, but $\mathbf{P}$ and $\mathbf{R}$ no. Intuitively, that two probabilities to be equivalents means that both probabilities agrees in what events can occur although they quantify differently the events.

6.3.4 Arbitrages and martingales

This point explains why the tribe of probabilists were asked to work in Mathematical Finance. Next Theorem is called (a bit pretentiously, we should recognize) the First Theorem of Finance.

Teorema 6.3.2 *In a market there are no arbitrages if and only if there is a probability* $\mathbf{Q}$ *equivalent to* $\mathbf{P}$ *such that the discounted prices of the assets,* $\widetilde{S}^{(j)} = \{\widetilde{S}^{(j)}_n,\ n = 0, \dots, N\}$ *are* $\mathbf{Q}$*-martingales.*

Proof of ($\Longleftarrow$).

We prove that result when $d = 1$; the general case is similar. Assume that there is a probability $\mathbf{Q}$ equivalent to $\mathbf{P}$ such that the discounted price $\{\widetilde{S}_n,\ n \geq 0\}$ is $\mathbf{Q}$-martingale. Then, for any self-financing portfolio, $\mathbf{\Phi}$, by the discounting formula (6.3), changing V_0 by its value,

$$\widetilde{V}_n = D_0 + H_0 S_0 + \sum_{k=1}^{n} H_k \big(\widetilde{S}_k - \widetilde{S}_{k-1}\big).$$

By the property of martingale transforms (see page 47), $\{\widetilde{V}_n,\, n \ge 0\}$ is a $\mathbf{Q}$ martingale and

$$\mathbf{E}_{\mathbf{Q}}[\widetilde{V}_n] = \mathbf{E}_{\mathbf{Q}}[\widetilde{V}_0] = V_0.$$

If $\mathbf{\Phi}$ were an arbitrage, $V_0 = 0$, and then

$$\mathbf{E}_{\mathbf{Q}}[V_N] = 0. \tag{6.7}$$

On the other hand, since $\mathbf{P}$ and $\mathbf{Q}$ are equivalent, it follows that

$$V_N \ge 0,\ \mathbf{P} - \text{a.s.} \quad \Longrightarrow \quad V_N \ge 0,\ \mathbf{Q} - \text{a.s.},$$

and this last property joined with (6.7) implies $V_N = 0$. So $\mathbf{\Phi}$ is not an arbitrage.

The implication ($\Longrightarrow$) is much more difficult and we omit the proof.

> In the previous proof we has seen a remarcable result: Under the probability $\mathbf{Q}$, the discounted value of an arbitrary self-financing portfolio, $\{\widetilde{V}_n,\, n = 0, 1, \dots\}$ is a martingale.

Example 2 (Continuation). A probability $\mathbf{Q}$ equivalent to $\mathbf{P}$ is determined by a number $q \in (0, 1)$ and

$$\mathbf{Q}(S_1 = 28) = q \quad \text{and} \quad \mathbf{Q}(S_1 = 27.5) = 1 - q.$$

The condition $q \ne 0, 1$ is due to the fact that $\mathbf{P}(S_1 = 28) \ne 0$ and $\mathbf{P}(S_1 = 27.5) \ne 0$. So we want to study if there is $q \in (0, 1)$ such that

$$\mathbf{E}_{\mathbf{Q}}[\widetilde{S}_1 \,|\, \mathcal{F}_0] = \widetilde{S}_0.$$

Here, things are very easy: $\widetilde{S}_0 = S_0 = 25$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so $\mathbf{E}_{\mathbf{Q}}[\widetilde{S}_1 \,|\, \mathcal{F}_0] = \mathbf{E}_{\mathbf{Q}}[\widetilde{S}_1]$. Then,

$$\mathbf{E}_{\mathbf{Q}}[\widetilde{S}_1] = q\,\frac{28}{1.1} + (1-q)\frac{27.5}{1.1} = 25.$$

And this equation has only the solution $q = 0$, which is not admissible. This agrees with Theorem 1, since as we have seen in Example 2, in that model there are arbitrages.

Example 4. Consider a model with $r = 0.1$, one risky asset with $S_0 = 25$ and price dynamics

$$\mathbf{P}(S_1 = 24) = \frac{3}{4} \quad \text{and} \quad \mathbf{P}(S_1 = 28) = \frac{1}{4}.$$

Working as in previous example, we need to solve

$$q\,\frac{28}{1.1} + (1-q)\frac{24}{1.1} = 25.$$

The solution (unique) is $q = 0.875$. So in that model there are no arbitrages.

To interpret this result think in a game where a player beats 25 € and receives $28/1.1 = 25.45$ € with probability q or $24/1.1 = 21.82$ € with probability $1 - q$. If we believe that the player is playing a fair game then the probability should be $q = 0.875$. That means, the no arbitrage opportunities property says that the prices in the market reflects that the participants are playing a fair game (a martingale): the prices reflects all the available information and we can deduce the *fair* probabilities of up and down. Of course, nobody believes this, but it is a simplified model that allows us to start computations an obtain an approximation to reality, in the same way as in Newton Physics are used very simplified models.

6.4 Exercices

1. (With EXCEL). Take $r = 0.03$, and a portfolio with a risky asset and three types of actions, with initial composition

$$D_0 = 10000,\ H_0^{(1)} = 100,\ H_0^{(2)} = 200,\ H_0^{(3)} = 250,$$

and initial prices

$$S_0^{(1)} = 22,\ S_0^{(2)} = 16,\ S_0^{(3)} = 28.$$

At time 1 the prices are

$$S_1^{(1)} = 24,\ S_1^{(2)} = 15,\ S_1^{(3)} = 27,$$

and we sell 50 shares of type 1 and buy 30 of type 2. At time 2, the prices are

$$S_2^{(1)} = 26,\ S_2^{(2)} = 19,\ S_2^{(3)} = 30,$$

and we sell 25 shares of type 2 and we buy 100 of type 3. At time 3, the prices are

$$S_3^{(1)} = 25,\ S_3^{(2)} = 17,\ S_3^{(3)} = 32,$$

Compute the final value of the portfolio. At each time, check that the portfolio is self-financing.

2. Write the portfolios corresponding to both Examples 2 and 3 in page 66.

Chapter 7

Derivative products

Loosely speaking, a **derivative** or **contingent claim** is a contract which value depends on the price of one (or more) asset, called the *underlying*. We distinguish between different derivatives according if the contract should be finished (or exercised) in a specified date, say N, or could be finished at any time n before a date N. In the former case we speak of **European derivatives**, and in the latter of **American derivatives**. There are still other derivatives, as the **Bermudean** ones that can be exercised at some fixed times $n_1, \ldots, n_r$ (they are between European and American derivatives). We begin with the simplest ones that are the European derivatives. To introduce the concepts that we use, we start considering a **forward contract**

7.1 Introductory example: Forward contracts

The oldest and simple derivative is a **forward contract**.

A **forward contract** is an agreement today (time 0) between a *seller* and a *buyer* that a time N the seller will sell some asset to the buyer at a specified price K, called the **delivery price**.

The buyer is said to held a *long position* and the seller a *short position*.

Denote by $\{S_n,\, n = 0, \ldots, N\}$ the price of the asset, that in some contexts is also called **Spot price**. From the point of view of the buyer, her profit (positive or negative), will be $S_N - K$, because if $S_N > K$ and she buys the asset by K then she could sell the asset and win $S_N - K$; on the contrary, if $S_N < K$ it would be better to buy directly in the market, but as she is obligated to buy at price K, she is losing $K - S_N$ (she is "earning" $-(K - S_N) = S_N - K$). On the other side, the profit or loss for the seller is $K - S_N$. So in each contract there are two different points of view.

Consider the point of view of the buyer. The profit $S_N - K$ can be represented in Figure 7.2 in function of S_n. The buyer can have a maximum loss of $-K$ € if at the expiration date N the price of the underlying is 0, that is $S_N = 0$; on the contrary, she can have an unlimited profit if S_N is much greater than K.

The fair **delivery price** K should be $S_0(1 + r)^T$. Otherwise, there are arbitrage opportunities (please, check).

To enter a forward contract costs nothing; it is said that the value at time 0 of the contract is 0, and write $V_0 = 0$. However, as times goes on, the value of the contract changes. The following simple

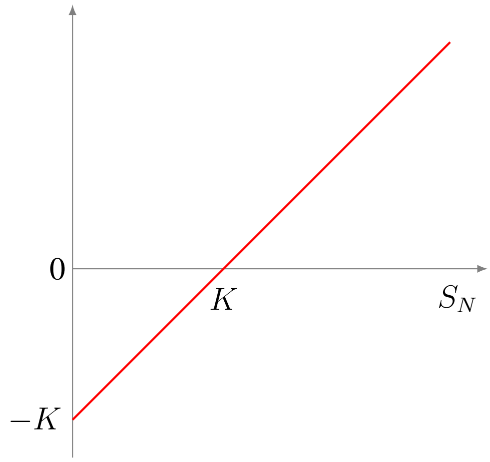


Figure 7.1. Profit of a buyer in a forward contract

example from Dalton [5] (adapted to our notations: we use monthly interest rate) is very illuminating. Today is first of March. Assume that the dynamics of the price of one Kg of frozen peas from First of March to First of September is given in the following table (of course, only the first entry of that table is known on First of March):

Date	1 March	1 April	1 May	1 June	1 July	1 August	1 September
S_n	2.37	2.33	2.29	2.27	2.26	2.27	2.28

For this example, assume that there is a monthly rate of interest $r = 0.5\%$. A farmer, called Farmer A, enters a forward contract to sell the first September 10000 units of frozen peas. The delivery price (for unit) is

$$K = 2.37(1 + 0.005)^6 = 2.442.$$

A second farmer, Farmer B, realizes on first of April that the price of frozen pea was falling, and then he decided to enter a forward contract to sell the first September 10000 units. This second forward contract has a delivery price of

$$K = 2.33(1 + 0.005)^5 = 2.389.$$

However, it is clear that on the first of April, the contract of farmer A is worth than the contract of farmer B, since farmer A will receive the first September 2.442€ per unit, whereas farmer B will receive only 2.389 € per unit. So there is a difference of $d = 2.442 - 2.389 = 0.053$ €. If farmer A would like to sell his contract to farmer B (the first of April), which will be a fair price? This quantity is called the value of the contract A at time 1, V_1. To avoid arbitrages, the quantity V_1 € in a bank account that gives monthly $r = 0.5\%$ from first April to first September should produce exactly $d = 0.053$ €. So

$$V_1(1 + 0.005)^5 = 0.053.$$

Hence $V_1 = 0.052$ €. In a similar way, using a farmer that enters the 1st. May a forward with expiring date on 1st. September we can compute the value of the forward contract the 1st. May. And so on. The following table give the value of the contract of farmer A (per unit) with delivery price $K = 2.442$.

Date	1 March	1 April	1 May	1 June	1 July	1 August	1 September
n	0	1	2	3	4	5	6
V_n	0	0.052	0.104	0.136	0.158	0.160	0.162

In Figure 7.2 there is a plot of the value of the short forward contract of farmer A

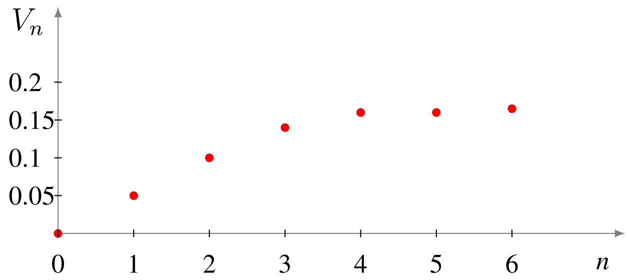


Figure 7.2. Value of the short forward contract of farmer A

Remark. A much more interesting and popular derivative is the Future contract. It is designed to avoid the risk of default in a forward contract. The future contracts are ruled and organized trough an exchange. Using a procedure called **marking to marked** each day the evolution and pricing of the contract is controlled. It is a main tool for speculating and hedging. For present purposes of this course, we do not study the Future markets, and I recommend the excellent cited book of Dalton [5].

Exercise. Plot the payoff of the seller in a forward contract and discuss her possible profits and losses.

7.2 European derivatives

As above commented, an **European derivative** with expiration date N is a contract signed today between two parts related to an operation at time N and that depends of the value of the price of some asset. A major concept is the **payoff** (positive or negative) of a participant in the contract which is the quantity of money obtained by that part, and that depends on the contract and the value of the underlying; For example, in the forward contract, the buyer has a payoff $S_N - K$, and the seller $K - S_N$. In general, the payoff is represented by a random variable X that is $\mathcal{F}_N$ measurable. In the most usual case, X is of the form $f(S_0, \ldots, S_N)$, for an asset S, and the simplest case is when $X = f(S_N)$, as in the (long) forward contract, where $f(x) = x - K$.

7.3 Options

Forward contracts are binding. On the contrary, an **option** gives to one part the right, but no the obligation, to do something.

An **European call option** is a contract signed today (time 0) between a seller and a buyer that gives the right, but not the obligation, to the buyer to buy something at a specified price K. or **exercise price** at some date N. The price K is called the **strike price** and the date N the **expiry date** or **maturity date**.

An **European put option** is a contract signed today (time 0) between a seller and a buyer that gives the right, but not the obligation, to the seller to sell something t the buyer at the exercise price K at the expiry date N.

The jargon increases in complexity:

- The seller in a call option writes the contract, and it is called **the writer of the call** or is is also said that she is **short in a call**.
- The buyer in a call is called the **holder** or the **taker**, or that she **is long in a call**.
- The European Call and Put option are called **Plain vanilla** options, in contrast with other options that are called **exotic**.

7.4 European call options

The payoff of the buyer in an European call option is

$$(S_N - K)^+ = \begin{cases} S_N - K, & \text{if } S_N \geq K, \\ 0, & \text{otherwise.} \end{cases}$$

See Figure 7.3. The rationale is that at day N the buyer will look the price S_N. if $S_N > K$ she will exercise the option, and buy at the price K; if she wants, she can sell the asset at price S_N and win $S_N - K$. On the contrary, if $S_N < K$ she does not exercise the option and buy directly to the market at price K. So there is no lose, in contrast with the forward contract.

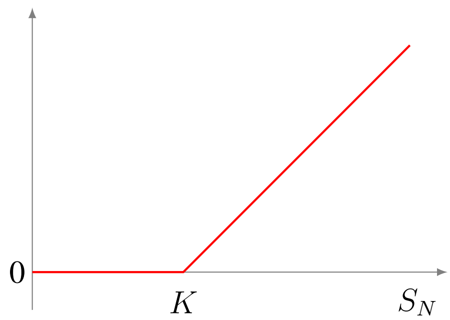


Figure 7.3. Payoff of the buyer in an European call

Remark. In real markets, the exercise of an option is (many times) done symbolically: For example, for a Call, if $S_N > K$, the buyer does not need to buy the asset and sell in the market to get the profit in cash; directly, the bank or financial agent, acting as a seller, gives to the buyer the difference $S_N - K$. This gives rise to a world of more complex options (we will study some of ones later) where have no sense any operation with the underlying.

It is obvious that the buyer has all the advantages in a call option, and the seller the disadvantages (including unlimited losses). So it is clear that in order to sign this contract the buyer should pay some quantity to the seller. This quantity is called the **call price** or the **call premium**.

The key point in modern Mathematical finance is the following question: Can we compute the call price in a way that both buyer and seller agree that this price is correct? The answer, is **YES, WE CAN!** The rest of the course is devoted so such problem, called **the pricing of a derivative**. What is more, how it is avoided the possibility of a big loss of the seller? Indeed, it is possible to assure that the seller will lose nothing; she can use cleverly the money that she receives as a the price to cover the possible losses. This is called **hedging a derivative**, and it is also a must topic in these notes.

Associated with an European call there is also the profit function of each of the parts at the end of the contract. if the price of the derivative is C, then the profit of the buyer is $(S_N - K)^+ - C$. A plot is given in Figure 7.4

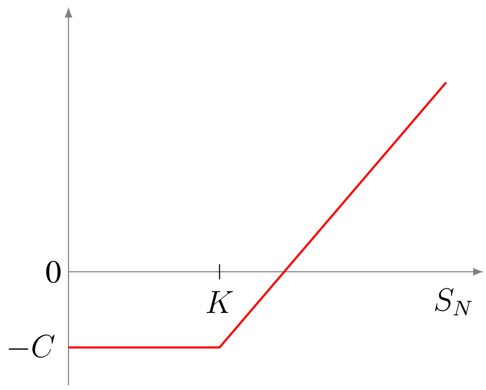


Figure 7.4. Profit of the buyer in an European call

7.4.1 European put options

The payoff of the seller in a European put is $(K - S_N)^+$. Here, the advantages are for the seller in the option, that have to pay a quantity to the buyer, called the **put price** or **put premium**.

7.5 Replicating portfolios

In a general context, the payoff of a European derivative with expiration date N is a random variable X that is $\mathcal{F}_N$ measurable. Observe that in an European call, the payoff of the buyer is

$$X = (S_N - K)^+ \geq 0.$$

The payoff of a seller in an European put is

$$X = (K - S_N)^+ \geq 0.$$

On the contrary, note that in forward contract neither the payoff of the seller or the buyer satisfies that condition of positivity. The study of the derivatives with payoff $X \geq 0$ centers our interest.

A positive $\mathcal{F}_N$ measurable random variable $X \geq 0$ is called **replicable** if there is a self-financing (admissible) portfolio $\mathbf{\Phi}$ such that

$$V_N(\mathbf{\Phi}) = X.$$

The portfolio $\mathbf{\Phi}$ is called the **replicating portfolio** of X.

Remember that a portfolio is called admissible if $V_n \geq 0, \ \forall n$.

Example 1. Consider a market with one risky asset with initial price $S_0 = 25$ €. The price at time 1 can be 28 o 24 € with probabilities 1/4 i 3/4 respectively (see Figure 7.5). Let $r = 0.1$. Consider an European Call with strike price $K = 25$ and maturity $N = 1$. The payoff is

$$X = (S_1 - 25)^+ = \begin{cases} 3, & \text{if } S_1 = 28, \\ 0, & \text{if } S_1 = 24. \end{cases}$$

We are going to prove that this random variable is replicable. It is convenient to specify the space of probability. We can take $\Omega = \{\omega_1, \omega_2\}$, where ω_1 corresponds to the path to $S_1 = 28$ and ω_2 to $S_1 = 24$ (see Figure 7.5).

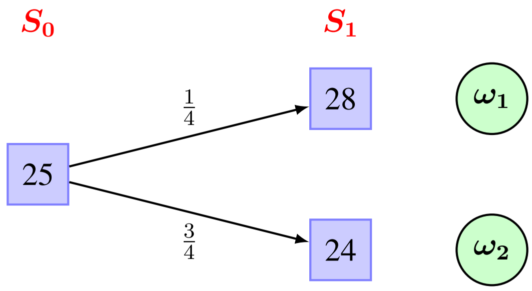


Figure 7.5. Probability space for Example 1

The payoff can be written as

$$X(\omega_i) = (S_1(\omega_i) - 25)^+ = \begin{cases} 3, & \text{if } i = 1, \\ 0, & \text{if } i = 2. \end{cases}$$

In this setup, a portfolio is $\mathbf{\Phi} = \{\mathbf{\Phi}_0, \ \mathbf{\Phi}_1\}$ where

$$\mathbf{\Phi}_n = (D_n, H_n), \ n = 0, \ 1.$$

Since the portfolio is predictable, D_1 and H_1 are $\mathcal{F}_0 = \{\emptyset, \Omega\}$ measurable, and hence, they are deterministic. Write

$$D_1(\omega_1) = D_1(\omega_2) = d_1 \quad \text{and} \quad H_1(\omega_1) = H_1(\omega_2) = h_1$$

We want:

(i) The portfolio is self-financing. This implies that

$$D_0 + H_0 \, 25 = D_1 + H_1 \, 25,$$

or, with the previous notation,

$$D_0 + H_0 \, 25 = d_1 + h_1 \, 25 \tag{7.1}$$

(ii) The portfolio replies X, that is, $V_1 = D_1\,1.1 + H_1\,S_1 = X$, on all possible scenarios; explicitly,

$$D_1(\omega_1)\,1.1 + H_1(\omega_1)\,S_1(\omega_1) = X(\omega_1) = 3$$
$$D_1(\omega_2)\,1.1 + H_1(\omega_2)\,S_1(\omega_2) = X(\omega_2) = 0$$

and then, using again that H_1 and D_1 do not depend on ω,

$$d_1\,1.1 + h_1\,28 = 3$$
$$d_1\,1.1 + h_1\,24 = 0$$

Solving this system we obtain

$$d_1 = -16.36 \quad \text{and} \quad h_1 = 0.75.$$

From equation (7.1) we deduce that the value of the portfolio at time 0 is

$$V_0 = -16.36 + 0.75 \times 25 = 2.39\,€.$$

There are infinite possible values of D_0 and H_0 such that $V_0 = 2.39$ €. We choose the one with $H_0 = 0$ and $D_0 = 2.39$ €.

To check that this portfolio replicates the Call, we write al possible evolution of the market.

- If $S_1 = 28$.

n	D_n	B_n	H_n	S_n	V_n	X
0	2.39	1	0	25	2.39	
0+	-16.36	1	0.75	25	2.39	
1	-16.36	1.1	0.75	28	**3**	**3**

- If $S_1 = 24$.

n	D_n	B_n	H_n	S_n	V_n	X
0	2.39	1	0	25	2.39	
0+	-16.36	1	0.75	25	2.39	
1	-16.36	1.1	0.75	24	**0**	**0**

Note also that de portfolio is admissible.

Now we explicit in words the actions of the buyer and the seller:

Buyer

Time 0: Gives 2.39 € to the seller

Seller

Time 0. Receives 2.39 € and builds the replicating portfolio with

$$D_0 = 2.39, H_0 = 0,$$

and she reorganizes to get at time 1

$$D_1 = -16.36,\ H_1 = 0.75,$$

that means, she borrows 16.36 €, adds to the 2.39 € and buys 0.75 asset.

Time 1

• ***Up:*** $\boldsymbol{S_1 = 28}$. Exercises the option and buys the asset. She paid a total amount of 25+2.39=27.39 €. She saves 0.61 €.

• ***Down:*** $\boldsymbol{S_1 = 24}$. She does not exercise the option and buys the asset in the market. She paid 24+2.39=26.39 €. So she has paid 2.39 € more.

Time 1

• ***Up***: $\boldsymbol{S_1 = 28}$. The buyer exercises the option, so the seller

Receives 25 €.

Sells (in the market) the 0.75 assets: $0.75 \times 28 = 21$ €

Return the $16.36 \times 1.1 = 18$ € loan

The total amount that the seller has

$$25 + 21 - 18 = 28 \text{ €},$$

that is, the same as if the seller has sold the asset in the market.

• ***Down***: $\boldsymbol{S_1 = 24}$. The buyer does not execute the option. Then the seller

Sell the 0'75 asset by per $0.75 \times 24 = 18$ €

Return the loan of 18 €

Retains the asset.

Total,

$$18 - 18 = 0 \text{ €}$$

and retains the asset.

Table 7.1 gives a comparative of buying directly an asset at the market at time 1, or use an European Call. Using a Call reduces the uncertainty of the business.

	With the Call	**Without the Call**
Up	27.39	28
Down	26.39	24
Difference	1	4

Table 7.1. Comparative between buying an asset using or not an European Call

The replicating portfolio showed in that example is called the **hedging portfolio** of the derivative product. It allows to the seller in the European call to hedge or avoid any financial loss.

7.6 Complete markets

In a market there can be an infinity of (European) derivative products; the possibility of construct a portfolio that behaves as the payoff gives us a better understanding of the derivative, and, really, the possibility of hedging that product. So the following definition is really important

> A market is said to be **complete** if every positive $\mathcal{F}_N$ measurable random variable $X \geq 0$ is replicable.

7.6.1 The second main Theorem

The probabilities appear again. This theorem is called *the second main theorem of Mathematical finance*.

Theorem. A market is complete if and only if there is one and only probability, denoted by $\mathbf{P}^*$, equivalent to $\mathbf{P}$, such that the discounted prices of the assets, $\widetilde{S}^{(j)} = \{\widetilde{S}_n^{(j)},\, n = 0, \dots, N\}$ are $\mathbf{P}^*$-martingales.

The probability $\mathbf{P}^*$ is called the **risk neutral probability**. We will discuss later this terminology. We illustrate this theorem with a couple of examples.

Example 1 (continuation). In the previous chapter we computed that the unique probability such that $\{\widetilde{S}_0, , \widetilde{S}_1\}$ is a martingale is $q = 0.875$. So the market is complete and $\mathbf{P}^*$ is given by

$$\mathbf{P}^*\{S_1 = 28\} = 0.875 \quad \text{and} \quad \mathbf{P}^*\{S_1 = 24\} = 0.125$$

Example 2. Consider a *trinomial* model where S_1 can take three values:

$$\mathbf{P}(S_1 = 26) = p_1, \quad \mathbf{P}(S_1 = 25) = p_2 \quad \text{and} \quad \mathbf{P}(S_1 = 24) = p_3,$$

Assume that $\mathbf{P} \leftrightarrow p_1, p_2, p_3 \in (0,1)$ with $p_1 + p_2 + p_3 = 1$ is given. An equivalent probability $\mathbf{Q}$ corresponds to $q_1, q_2, q_3 \in (0,1)$ with $q_1 + q_2 + q_3 = 1$. We are interested in one of such $\mathbf{Q}$ that satisfies

$$\mathbf{E}_{\mathbf{Q}}[\widetilde{S}_1 \,|\, \mathcal{F}_0] = \widetilde{S}_0.$$

Proceeding as in previous Chapter, we need to solve

$$\begin{aligned} 24q_1 + 25q_2 + 26q_3 &= 1.1 \times 25 \\ q_1 + q_2 + q_3 &= 1 \end{aligned}$$

with the restrictions $q_1 > 0$, $q_2 > 0$ and $q_3 > 0$. There are infinite solutions. Hence, there are no arbitrages in this model (1st. main Theorem), but it is not complete (2nd. main Theorem). To convince yourself about the incompleteness, note that a positive random variable $\mathcal{F}_1$ measurable X is determined now by three positive numbers x_1, x_2, x_3 such that

$$X = \begin{cases} x_1, & \text{if } S_1 = 26, \\ x_2, & \text{if } S_1 = 25, \\ x_3, & \text{if } S_1 = 24. \end{cases}$$

With the same notation as in example 1, a portfolio in this model is deterministic, and at time $n = 1$, it is given by the numbers (not random variables) d_1 and h_1. In order that a portfolio replies X it is needed

$$d_1\, 1.1 + h_1 S_1 = X,$$

so

$$\begin{aligned} d_1\, 1.1 + h_1 26 &= x_1 \\ d_1\, 1.1 + h_1 25 &= x_2 \\ d_1\, 1.1 + h_1 24 &= x_3 \end{aligned}$$

Since it is a linear system with 2 unknowns (d_1 and h_1) and 3 equations, there is a solution it is necessary and sufficient that

$$\begin{vmatrix} 1.1 & 26 & x_1 \\ 1.1 & 25 & x_2 \\ 1.1 & 24 & x_3 \end{vmatrix} = 0.$$

This condition is equivalent to

$$x_1 - 2x_2 + x_3 = 0. \tag{7.2}$$

Then in order that a random variable is replicable it is necessary that it satisfies this condition. Since the model is very simple, we can find easily examples of European derivative products that satisfy and that do not satisfy (7.2).

7.7 Pricing and hedging in complete markets

In this section we will consider a complete market with risk neutral probability $\mathbf{P}^*$, and an European derivative with payoff $X \geq 0$, that means, X is $\mathcal{F}_N$–measurable. Denote by $\mathbf{\Phi}$ a replicating portfolio, that is,

$$V_N = X.$$

We saw in the proof of Theorem 3.2 that

$$\widetilde{V} = \{\widetilde{V}_n,\, n = 0, \ldots, N\}$$

is a $\mathbf{P}^*$ martingale, that implies,

$$\mathbf{E}_{\mathbf{P}^*}[\widetilde{V}_N \mid \mathcal{F}_n] = \widetilde{V}_n.$$

Hence, writing the discount factor,

$$\boxed{V_n = (1+r)^{n-N} \mathbf{E}_{\mathbf{P}^*}[X \mid \mathcal{F}_n].}$$

This says that the value of the portfolio $\mathbf{\Phi}$ is determinated at every time n by X, and that random quantity is called the **value of the European derivative at time n**. It is worth remarking that this value only depends on X and not of the particular replicating portfolio.

In particular, since $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

$$V_0 = (1+r)^{-N} \mathbf{E}_{\mathbf{P}^*}[X].$$

This (deterministic) quantity is called the **price** or the **premium** of the derivative. We insist in these definitions, that are really important.

In a complete market, to every European derivative with payoff $X \geq 0$ can be associated a **value** at time n, denoted by $\boldsymbol{V}_n$. When $n = 0$ it is the number

$$V_0 = (1+r)^{-N} \mathbf{E}_{\mathbf{P}^*}[X],$$

that is also called the **price** or the **premium** of the derivative. For $n \geq 1$ the value is the random variable

$$V_n = (1+r)^{n-N} \mathbf{E}_{\mathbf{P}^*}[X \mid \mathcal{F}_n].$$

As a consequence, we can use an European derivative as an ordinary asset, and, for example, incorporate it in a portfolio.

Example 1 (continuation). We compute the price of the European Call with strike $K = 25$ and maduration date $n = 1$, without using the explicit expression of the replicating portfolio. We summarize the model, adding the value of the payoff $X = (S_1 - 25)^+$ to the plot:

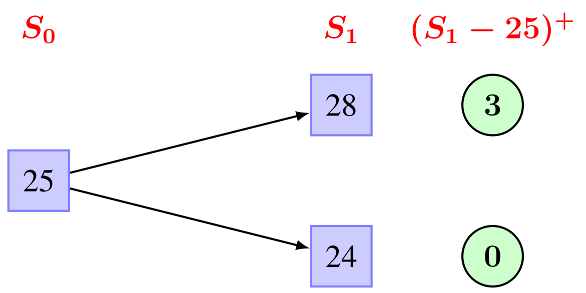


Figure 7.6. Example 1

The market is complete with risk neutral probability given by

$$\mathbf{P}^*\{S_1 = 28\} = 0.875 \quad \text{and} \quad \mathbf{P}^*\{S_1 = 24\} = 0.125$$

so the price of the derivative is

$$V_0 = 1.1^{-1} E_{\mathbf{P}^*}[(X-25)^+] = 1.1^{-1}\big(0.875 \times 3 + 0.125 \times 0\big) = 2.39.$$

Then the initial value of any replicating portfolio should be $V_0 = 2.39$, in agreement with the computations in page 77.

7.8 Call–Put parity

Denote by C_n the price of an European Call with expiry time N and strike price K. As we have seen,

$$C_n = (1+r)^{n-N} \mathbf{E}_{\mathbf{P}^*}[(S_N - K)^+ \mid \mathcal{F}_n].$$

Let P_n the price of an European put with the same strike price.

$$P_n = (1+r)^{n-N} \mathbf{E}_{\mathbf{P}^*}[(K - S_N)^+ \mid \mathcal{F}_n].$$

Observe that

$$a^+ - (-a)^+ = a^+ - a^- = a.$$

Hence, since $\widetilde{S}_n$ is a $\mathbf{P}^*$ martingale,

$$\boxed{C_n - P_n = S_n - K(1+r)^{n-N}.}$$

This is a very nice formula that allows to compute the price of an European Call if we know the price of the Put and viceversa.

7.8.1 Why are options so interesting?

Roughly speaking, in the market there are three type of participants:

- The people that need to go to the market by real economic reasons. For example, a German company that should pay a bill of one million dollar in half a year. Since the change euro/dollar is fluctuating, the company can sign a call with a bank in order to buy one million dollar at some specified price fixed today. This contract will reduce the uncertainty of the total amount that the company should pay in half a year, which is very convenient in order to a rational financial planning.

- The banks and other financial institutions. This is their job. When they are, for example, the seller in an European call, they receive the price of the call and build the hedging portfolio. So, in principle, they win or loss nothing. However, they charge some commission to the companies, or they compute the price of the option and add some quantity.

- The speculators. Under the intuition that some asset will up or down its price, a speculator can act as a buyer or seller in a call or put, not hedging in any case. This is a bet that allows a strong leverage, with a risk of losing all money. For example, assume that someone has 10.000 € to invert, and believes that some shares will increase its value in the next months. The current price of a share is $S_0 = 25$ €, and a Call option that expires in two months with strike $K = 27$ € has a price of 2 €. Then there are two strategies:

1. To buy 400 shares.

2. To buy 5.000 Call options

Now consider two scenarios: After two months, $S_2 = 30$ or $S_2 = 24$. If $S_2 = 30$, then the benefit of the operation (selling the shares, or exercising the option) is

Shares	$30 \times 400 - 10.000 = 2.000$ €
Options	$3 \times 5.000 - 10.000 = 5.000$ €

If $S = 24$, the owner of the shares has the possibility to keep the shares and wait for better times; on the contrary, the owner of the Calls has nothing. Anyway, assuming that the shares are sold, the losses in both scenarios are

Shares	$24 \times 400 - 10.000 = -400$ €
Options	$0 - 10.000 = -10.000$ €

The term **leverage** in the financial world means borrowing some money and reinverting them in order to gain a greater quantity than the interests of the loan; of course, this is very risky. Derivatives allow a kind of leverage without borrowing, at risk of loss all the inversion.

7.9 Exercises

1. Consider a discrete market with $r = 1/9,\ N = 1$, and two risky assets with inicial prices $S_0^1 = 5$ i $S_0^2 = 10$, and

$$(S_1^1, S_1^2) = \begin{cases} (20/3, 40/3) & \text{with probability } p_1, \\ (20/3, 80/9) & \text{with probability } p_2, \\ (40/9, 80/9) & \text{with probability } p_3, \end{cases}$$

where $p_1,\ p_2,\ p_3 \in (0,1)$, with $p_1 + p_2 + p_3 = 1$.

(a) Check that there is an equivalent probability $\mathbf{Q}_1$ such that $\{\widetilde{S}_0^1,\ \widetilde{S}_1^1\}$ is a martingale, and another equivalent probability $\mathbf{Q}_2$ such that $\{\widetilde{S}_0^2,\ \widetilde{S}_1^2\}$ is a martingale, but there is no equivalent probability such that both discounted prices are martingales.

(b) Let $a > 0$, and consider a portfolio $\mathbf{\Phi}$ given by $\mathbf{\Phi}_0 = (0, a/2, -a/4)$. Check that it is an arbitrage.

2. Consider a trinomial model with $r = 1/9,\ N = 1$, $S_0 = 5$ and

$$S_1 = \begin{cases} 20/3 & \text{with probability } p_1, \\ 40/9 & \text{with probability } p_2, \\ 10/3 & \text{with probability } p_3, \end{cases}$$

where $p_1,\ p_2,\ p_3 \in (0,1)$ with $p_1 + p_2 + p_3 = 1$.

(a) Check that the probabilities given by $(p_1, p_2, p_3) = (0.6, 0.2, 0.2)$ and $(p_1, p_2, p_3) = (0.52, 0.44, 0.04)$ are risk neutral. Are these probabilities equivalent?

(b) Check that the set of risk neutral probabilities is

$$\mathcal{M} = \{(p_1, p_2, p_3) = (s, 2 - 3s, -1 + 2s),\ s \in (1/2, 2/3)\}.$$

(c) Are there arbitrages? Is the Europan Call with payoff $(S_1 - 5)^+$ replicable?

(d) Find the conditions in order that a derivative with payoff X of the form

$$X = \begin{cases} X_1 & \text{if } S_1 = 20/3, \\ X_2 & \text{if } S_1 = 40/9, \\ X_3 & \text{if } S_1 = 10/3, \end{cases}$$

(e) Assume X replicable. Prove that for all $\mathbf{Q} \in \mathcal{M}$, the value $\mathbf{E}_{\mathbf{Q}}[X]$ is the same.

3. (Pricing in an incomplete market.) Continue with the previous exercise. Denote by $\mathcal{R}$ the set of derivatives with payoff replicable, and let Y a payoff no replicable. Take $\mathbf{Q} \in \mathcal{M}$ an arbitrary risk neutral probability. We accept the following result:

$$\inf\{\mathbf{E}_{\mathbf{Q}}[X],\ X \geq Y,\ X \in \mathcal{R}\} = \sup\{\mathbf{E}_{\mathbf{Q}}[Y],\ \mathbf{Q} \in \mathcal{M}\},$$

and denote such number by $V_+(Y)$. Note that in the left hand side the probability $\mathbf{Q}$ does not matter due to the part (c) of previous exercise. Analogously,

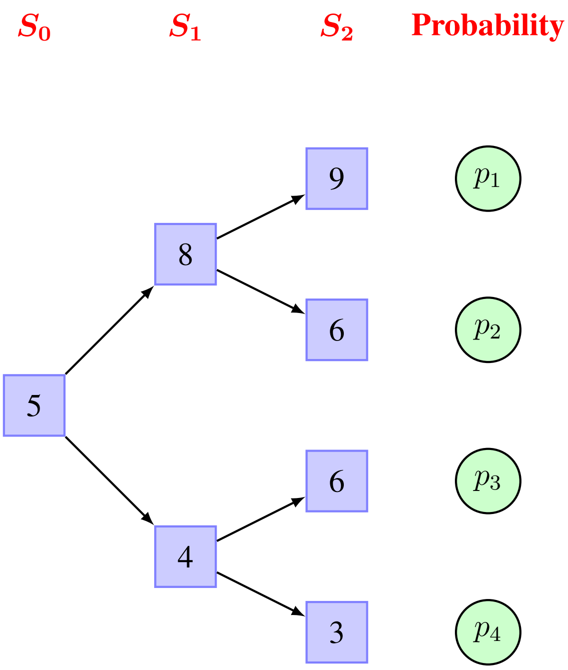
$$V_-(Y) = \sup\{\mathbf{E_Q}[X],\ X \le Y,\ X \in \mathcal{R}\} = \inf\{\mathbf{E_Q}[Y],\ \mathbf{Q} \in \mathcal{M}\}.$$

Now consider a derivative with payoff

$$Y = \begin{cases} 30 & \text{if } S_1 = 20/3, \\ 20 & \text{if } S_1 = 40/9, \\ 10 & \text{if } S_1 = 10/3. \end{cases}$$

Check that it is no replicable. Compute $V_+(Y)$ i $V_-(Y)$.

4. Consider a market with $r = 0$, $N = 2$ and one risky asset with dynamics:



(a) Find the risk neutral probability.

(b) Price the following products:

i. An European Call option with strike price 7.

ii. An European Put option with strike price7.

iii. An Asian option with strike price 7, that is with payoff

$$\big((S_0 + S_1 + S_2)/3 - 7\big)^+.$$

iv. A *chooser option* with strike price 7 and decision time 1; that is, at time 1 the owner has the right to choose between a Call or a Put with strike price 7. Specifically, denote by C_1 and P_1 the value at time 1 of a Call and a Put with expiration time 2 respectively. The chooser choose the Call if $C_1 \ge P_1$. The payoff of the chooser option is

$$(S_2 - 7)^+ \, \mathbf{1}_{\{C_1 \ge P_1\}} + (7 - S_2)^+ \, \mathbf{1}_{\{C_1 < P_1\}}.$$

v. A look-back option with payoff

$$\max(0, S_1 - 7, S_2 - 7).$$

Chapter 8

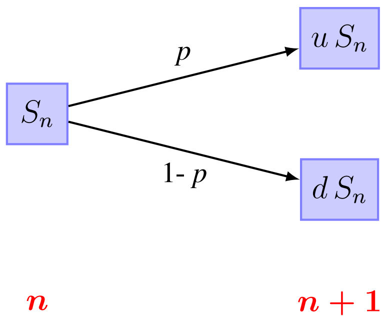
The Cox-Ross-Rubinstein model

8.1 The model

The most important discrete model is the Cox-Ross-Rubinstein model (CRR from now on), introduced in the very nice and clear paper [4] that I strongly recommend . It consists in a riskless asset with dynamics

$$B_n = (1+r)^n$$

and a risky asset with prices given by a binomial model



where it is assumed

- $S_0 = s_0 > 0$.
- $0 < d < u$.
- $r \in (d-1, u-1)$ (and, of course, $r \geq 0$).

The CRR model is complete and the the risk neutral probability is giving by

$$\boxed{p^* = \frac{r+1-d}{u-d}.}$$

We write $p = p^*$ and all computations are done with respect that probability.

Proof. By the Second Theorem in Finance , we need to prove

1. The process $\{\widetilde{S}_n,\ n = 0, 1, \dots\}$ is a martingale with respect the probability given by p above.

2. If $\{\widetilde{S}_n,\ n = 0, 1, \dots\}$ is martingale with respect a probability given by q, then $q = p$.

First note that from the Markov property (since S_n is Markov, then $\widetilde{S}_n$ also is),

$$\mathbf{E}[\widetilde{S}_{n+1} \mid \widetilde{S}_n, \dots, \widetilde{S}_1] = \mathbf{E}[\widetilde{S}_{n+1} \mid \widetilde{S}_n].$$

So we want to prove that

$$\mathbf{E}[\widetilde{S}_{n+1} \mid \widetilde{S}_n = x] = x,$$

for every admissible value x of $\widetilde{S}_n$. Let $y = u^{n-i} d^i s_0$, for some $i = 0, \dots, n$, and $x = (1+r)^{-n} y$. Then

$$\begin{aligned}
\mathbf{E}[\widetilde{S}_{n+1} \mid \widetilde{S}_n = x] &= \mathbf{E}[\widetilde{S}_{n+1} \mid \widetilde{S}_n = (1+r)^{-n} y] = (1+r)^{-n-1} \mathbf{E}[S_{n+1} \mid S_n = y] \\
&= (1+r)^{-n-1} \big(puy + (1-p) dy \big) \\
&= (1+r)^{-n-1} \big(\frac{r-d+1}{u-d}\, uy + \frac{-r+u-1}{u-d}\, dy \big) \\
&= (1+r)^{-n-1} y (r+1) = x.
\end{aligned}$$

Now assume that $\{\widetilde{S}_n,\ n = 0, 1, \dots\}$ is a martingale with respect a probability given by q. In particular,

$$\mathbf{E}_q[\widetilde{S}_1 \mid \widetilde{S}_0] = \mathbf{E}_q[\widetilde{S}_1] = s_0.$$

Hence,

$$(1+r)^{-1} q u s_0 + (1+r)^{-1} (1-q) d s_0 = s_0,$$

and it follows $q = p$.

8.2 Pricing an European derivative with payoff $X = f(S_N)$

We consider the pricing of an European derivative whose payoff depends only on the final value of the asset, for example, an European Call or Put, or digital options. So we consider a payoff of the form $X = f(S_N)$.

We saw in Chapter 7 that the value of the derivative at time n is

$$V_n = (1+r)^{n-N} \mathbf{E}\big[f(S_N) \mid \mathcal{F}_n\big] = (1+r)^{n-N} \mathbf{E}\big[f(S_N) \mid S_n\big],$$

where the last equality is due to the Markov property. Now decompose

$$S_N = S_n F_{n+1,N},$$

where $F_{n+1,N}$ is the multiplicative factor of the price between $n+1$ and N, that is, since there are $N-n$ steps, and at each step it is possible to choose between u and d, then $F_{n+1,N}$ can take the values $u^i d^{N-i}$ for $i = 0, \dots, N-n$, with probabilities

$$\mathbf{P}(F_{n+1,N} = u^i d^{N-n-i}) = \binom{N-n}{i} p^i (1-p)^{N-n-i}.$$

Since S_n is independent of $F_{n+1,N}$, by the properties of the conditional expectations,

$$V_n = v(n, S_n),$$

where the function $v(n,x)$ is given by

$$\begin{aligned} v(n,x) &= (1+r)^{n-N}\mathbf{E}\big[f(xF_{n+1,N})\big] && (8.1)\\ &= (1+r)^{-(N-n)}\sum_{i=0}^{N-n}\binom{N-n}{i}\Big(f(xu^i d^{N-n-i})p^i(1-p)^{N-n-i}\Big). && (8.2)\end{aligned}$$

In particular, the price of the derivative is

$$V_0 = V(0,S_0) = (1+r)^{-N}\sum_{i=0}^{N}\binom{N}{i}\Big(f(S_0u^i d^{N-i})p^i(1-p)^{N-i}\Big).$$

For an European Call with strike price K and maturity date N, the payoff is $X = (S_N - K)^+$. Then the previous formula takes a particularly nice expression. We have

$$V_0 = v(0,S_0) = (1+r)^{-N}\sum_{i=0}^{N}\binom{N}{i}\Big((S_0u^i d^{N-i} - K)^+p^i(1-p)^{N-i}\Big).$$

This expression can be written in closed form as

$$\boxed{V_0 = S_0B(N,p',D) - K\,(1+r)^{-N}\,B(N,p,D),} \tag{8.3}$$

where

$$p' = \frac{up}{1+r}, \quad D = \text{Integer part}\left(\frac{\log\big(K/(S_0d^N)\big)}{\log u - \log d}\right) + 1, \tag{8.4}$$

and

$$B(N,p,D) = \sum_{i=D}^{N}\binom{N}{i}p^i(1-p)^{N-i},$$

that corresponds to $\mathbf{P}\{Z \geq D\}$, where Z is a binomial random variable with parameters N and p (analogously for $B(N,p',D)$).

Exercise. Check the boxed expression. Some miracle happens.

8.2.1 Backward recursion

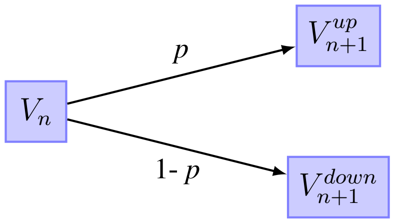
Following with the previous notation, since V_n is the value of a replicating portfolio, $\widetilde{V}_n$ is a martingale. Combining with Markov property and the definition of S_n,

$$\begin{aligned}\widetilde{V}_n &= \mathbf{E}[\widetilde{V}_{n+1}\,|\,\mathcal{F}_n] = \mathbf{E}[\widetilde{V}_{n+1}\,|\,S_n] = (1+r)^{-n-1}\mathbf{E}[v(n+1,S_{n+1})\,|\,S_n]\\ &= (1+r)^{-n-1}\big(v(n+1,uS_n)\,p + v(n+1,dS_n)(1-p)\big).\end{aligned}$$

Then, with a self-explicative notation,

$$\boxed{V_n = \frac{p\,V_{n+1}^{up} + (1-p)V_{n+1}^{down}}{1+r}.} \tag{8.5}$$

Graphically,



So, if f we computed the value of the derivative at all possible states at time $n+1$, we can go backward and compute the value at time n. This procedure is called **backward recursion**, and it is really easy to use, as the next example shows.
Example. The model is given by $r = 0.03$, $S_0 = 25$, $u = 1.09$, $d = 0.85$. The probability of *up* should be

$$p = \frac{r-d+1}{u-d} = 0.75.$$

Consider an European call with $N = 3$ and $K = 25$. The dynamics is given in Figure 8.1.

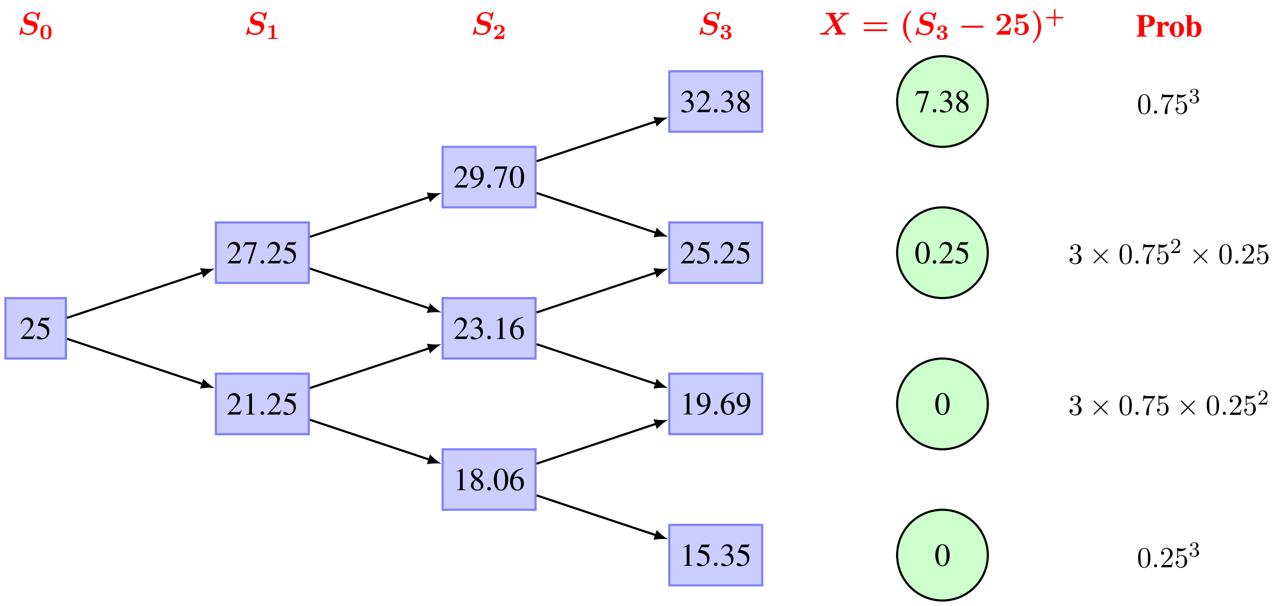


Figure 8.1. Dynamics of the example

The payoff $X = (S_3 - 25)^+$ is given in Figure 8.1. The price of the call is easily computed using the last two columns of that figure:

$$V_0 = 1.03^{-3}E[X] = 1.03^{-3}(7.38 \cdot 0.75^3 + 0.25 \cdot 3 \cdot 0.75^2 0.25 + 0) = 2.94.$$

However, to compute the value of the call at the other nodes is more problematic, since we need to compute conditional expectations. By backward recursion, we start by the values V_3, that is X

because the replicating portfolio takes the same value as X, or directly, because

$$V_N = \mathbf{E}[X \mid \mathcal{F}_N] = X,$$

since X is $\mathcal{F}_N$ measurable. Hence start by Figure 8.2.

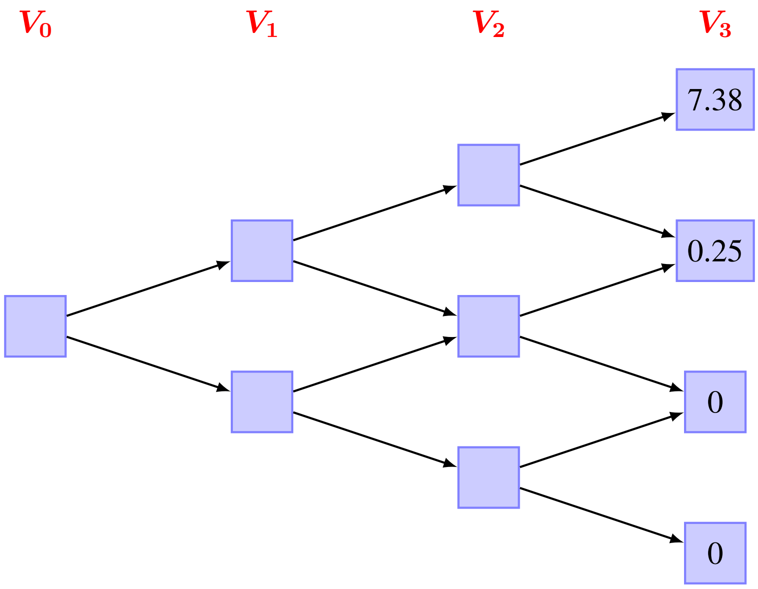


Figure 8.2. Backward recursion

Now we go backward and fill the values: By formula (15.2),

$$V_2 = \frac{p\,V_3^{up} + (1-p)V_3^{down}}{1+r}.$$

That is, for V_2 in the upper position,

$$\frac{0.75 \times 7.38 + 0.25 \times 0.25}{1.03} = 5.43$$

as show in Figure 8.3.

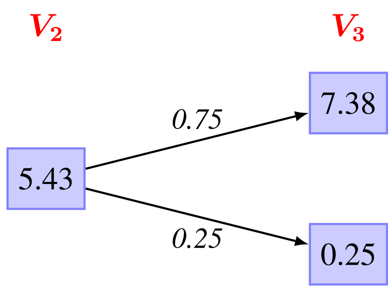


Figure 8.3. Backward recursion, first step

In this way it is obtained the Figure 8.4

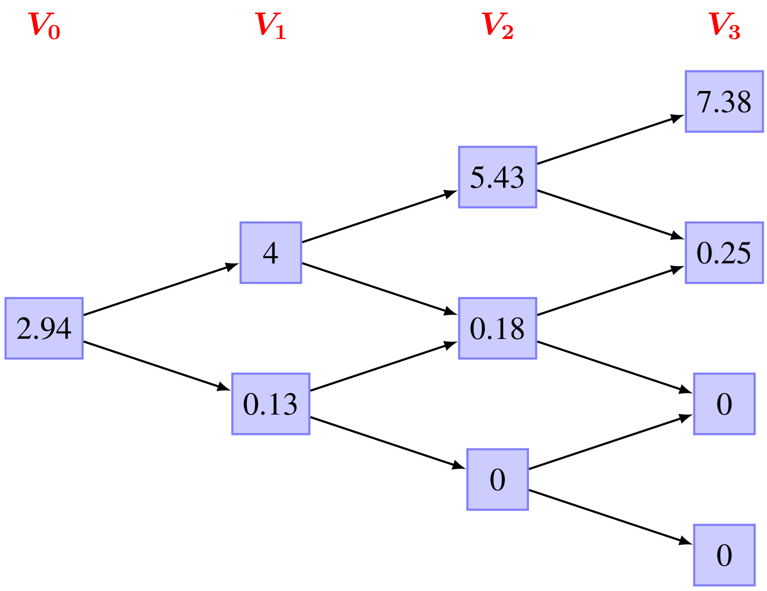


Figure 8.4. Values of the call

8.3 Hedging an European derivative with payoff $X = f(S_N)$

From the previous section we know the values of the derivative at every time, $V_0, \ldots, V_N$. Denote by $\mathbf{\Phi} = \{\mathbf{\Phi}_n,\ n = 0, \ldots, N\}$ the hedging portfolio, with

$$\mathbf{\Phi}_n = (D_n, H_n).$$

Then

$$V_n = v(n, S_n) = D_n(1+r)^n + H_n S_n.$$

As we commented, we start the portfolio with the price of the derivative:

$$D_0 = V_0 \quad \text{and} \quad H_0 = 0.$$

Now, assume that for $n \geq 1$, we observe S_{n-1} (that means, we are considering $S_{n-1}(\omega)$). Since D_n and H_n are $\mathcal{F}_{n-1}$ measurable, they are constants over the two possible paths from that S_{n-1}, that is,

$$\begin{aligned} V_n^{up} &= v(n, uS_{n-1}) = D_n(1+r)^n + H_n u S_{n-1} \\ V_n^{down} &= v(n, dS_{n-1}) = D_n(1+r)^n + H_n d S_{n-1} \end{aligned}$$

Solving that system,

$$\boxed{H_n = \frac{V_n^{up} - V_n^{down}}{(u-d)S_{n-1}},} \tag{8.6}$$

and (puting $S_n^{up} = uS_{n-1}$),

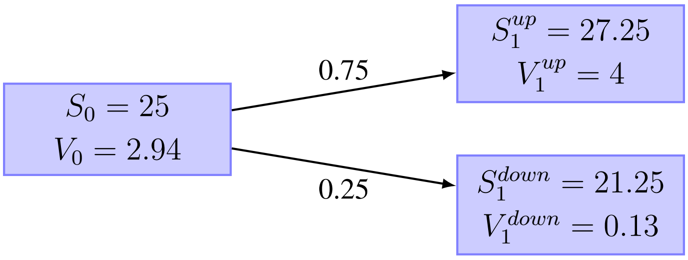
$$\boxed{D_n = \frac{V_n^{up} - H_n S_n^{up}}{(1+r)^n}.} \tag{8.7}$$

Example. We continue the previous data.

Time 0. As we commented,

$$\boxed{D_0 = V_0 = 2.94, \quad \text{and} \quad H_0 = 0.}$$

The next future is given in the following plot:



By formulas (8.6) and (8.7)

$$\boxed{H_1 = \frac{4 - 0.13}{(1.09 - 0.85)25} = 0.64 \quad \text{and} \quad D_1 = \frac{4 - 0.64 \times 27.25}{1.03} = -13.17}$$

The portfolio at this moment is

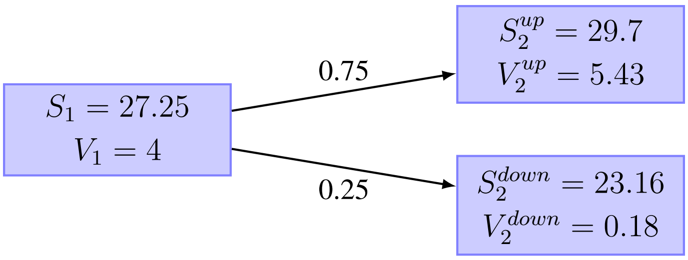
n	D_n	B_n	H_n	S_n	V_n
0	2.94	1	0	25	2.94
0+	-13.17	1	0.64	25	2.94

Time 1. Assume that $S_1 = 27.25$. Then the portfolio is

n	D_n	B_n	H_n	S_n	V_n
0	2.94	1	0	25	2.94
0+	-13.17	1	0.64	25	2.94
1	-13.17	1.03	0.64	27.25	$\boxed{4}$

Note that V_1 at the corresponding node coincides with the value that we have previously computed, see Figure 8.4.

Next step is the following scenario:



Using again formulas (8.6) and (8.7) we compute H_2 and D_2:

$$\boxed{H_2 = 0.8 \quad \text{and} \quad D_2 = -17.36}$$

n	D_n	B_n	H_n	S_n	V_n
0	2.94	1	0	25	2.94
0+	-13.17	1	0.64	25	2.94
1	-13.17	1.03	0.64	27.25	4
1+	-17.36	1.03	0.8	27.25	4

Time 2. Assume that $S_2 = 23.16$.
Then

n	D_n	B_n	H_n	S_n	V_n
0	2.94	1	0	25	2.94
0+	-13.17	1	0.64	25	2.94
1	-13.17	1.03	0.64	27.25	4
1+	-17.36	1.03	0.8	27.25	4
2	-17.36	1.0609	0.8	23.16	$\boxed{0.18}$

Again note that V_2 at this node coincides with the one given in Figure 8.4. This is a self-checking property that it is convenient to use to discover possible mistakes.

With the same operations as before,

$$\boxed{H_3 = 0.04 \quad \text{and} \quad D_3 = -0.8.}$$

Time 3. Assume that $S_3 = 25.25$.The final portfolio is

n	D_n	B_n	H_n	S_n	V_n
0	2.94	1	0	25	2.94
0+	-13.17	1	0.64	25	2.94
1	-13.17	1.03	0.64	27.25	4
1+	-17.36	1.03	0.8	27.25	4
2	-17.36	1.0609	0.8	23.16	0.18
2+	–0.8	1.0609	0.04	23.16	0.18
3	–0.8	1.0927	0.04	25.25	$\boxed{0.25}$

8.4 Pricing and hedging some exotic options

As in the chapter devoted to Markov processes, denote by M_n, $n = 0, \ldots, N$ the running maxima:

$$M_n = \max\{S_0, \ldots, S_n\}.$$

We will extend the backward recursion to exotic options where the payoff is of the form $f(S_N, M_N)$. That options include the Lookback options, where

$$f(S_N, M_N) = (M_N - K)^+ = \max\{0, S_0 - K, \ldots, S_N - K\},$$

or the barrier options type, as

$$f(S_N, M_N) = (S_N - K)^+ \mathbf{1}_{\{M_N < b\}}.$$

The key point is, as we see in the Markov chapter, that the bidimensional process $\{(S_n, M_n),\ n = 0, \ldots, N\}$ is Markov. Hence, the value of the option at time n will be

$$V_n = (1+r)^{n-N}\mathbf{E}\big[f(S_N, M_N)|\mathcal{F}_n\big] = (1+r)^{n-N}\mathbf{E}\big[f(S_N, M_N)|S_n, M_n\big].$$

Write

$$V(n, S_n, M_n) = (1+r)^{n-N}\mathbf{E}\big[f(S_N, M_N)|S_n, M_n\big].$$

Now, using the martingale property as in previous section, we deduce

$$\begin{aligned} V_n &= (1+r)^{-1}\mathbf{E}[v(n+1, S_{n+1}) \mid S_n, M_n] \\ &= (1+r)^{-1}\big(v(n+1, uS_n, (uS_n) \vee M_n)\, p + v(n+1, dS_n, (dS_n) \vee M_n)(1-p)\big). \end{aligned}$$

Then,

$$\boxed{V_n = \frac{v(n+1, uS_n, (uS_n) \vee M_n)\, p + v(n+1, dS_n, (dS_n) \vee M_n)(1-p)}{1+r}.} \tag{8.8}$$

The hedging portfolio is deduced in a similar way as in the previous section.

8.5 Exercises

1. (With Excel). Consider a CRR model with

$$r = 0.04,\ N = 4,\ d = 0.98,\ u = 1.06,\ S_0 = 20\ €$$

and an European Call with maturation time $N = 4$ and strike price $K = 20$ €.

(a) Compute the risk neutral probability.

(b) Construct the binomial tree for S_n.

(c) Compute the probabilities of S_4. Check that they add 1.

(d) Compute $V_4 = (S_4 - 20)^+$, and make a *backward recursion* in order to determine V_3, V_2, V_1 i V_0.

(e) First check of V_0: Compute $(1+r)^{-N}\,\mathbf{E}[(S_N - K)^+]$.

(f) Second check of V_0: Use the formula (8.3)

$$V_0 = S_0 B(N, p', D) - K\,(1+r)^{-N}\,B(N, p, D),$$

where p' and D are given in (8.4) in page 87. Recall that EXCEL uses the cumulative distribution function of a binomial random variable $Z \sim \mathcal{B}in(N, p)$,

$$\mathbf{DISTR.BINOM}(R; N; p) = P\{Z \le R\} = \sum_{j=0}^{R} P\{Z = j\},$$

and we need

$$B(N, p, D) = \sum_{j=D}^{N} P\{Z = j\} = P\{Z \ge D\} = 1 - \mathbf{DISTR.BINOM}(D-1; N; p)$$

(g) Compute the replicating portfolio assuming the following values of S_n:

$$S_0 = 20,\ S_1 = 21.2,\ S_2 = 20.776,\ S_3 = 22.02256 \quad \text{i} \quad S_4 = 21.5821088.$$

At each stage, check the value of the portfolio with the one obtained at point **(d)**. Design a structure of the EXCEL sheet such that these computations become clear and easy.

2. Consider a CRR model with $r = 0$, $S_0 = 20,\ u = 1.1$ and $d = 0.6$

(a) Compute th risk neutral probability and build the binomial tree till $N = 2$.

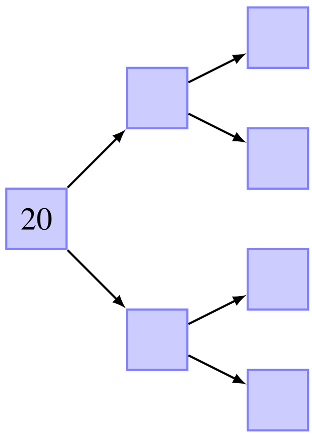
(b) Compute the possible values of a *look–back* with expiring time at $N = 2$ and payoff

$$\max(0, S_1 - 20, S_2 - 20)$$

(c) Compute the price of the option.

(d) Built the replicating portfolio.

Remark. Since the option is path–dependent the binomial tree should be of the form:



Chapter 9

American Options

The majority of options traded in the market are of American type. Hence its study is fundamental, and, in general, much more difficult that of the European one. Merton discovered that there is no benefit in early exercise of an American Call; however, for American Puts, or Calls over an underlying that gives dividends things are different are new formulas are needed. Anyway, the computations are based on the formulas for European options, so that formulas are still very useful.

9.1 Introduction

An American option with maturation time N can be exercised at every time $n = 0, \dots, N$. This fact adds an important difficulty in order to compute the pricing and hedging, and ask for new ideas. Such ideas are a very nice combination of financial intuition and Mathematics. In all the topic we assume that the market is complete.

We study the general case, but to show the rationale, we will have in mind a Call option with strike price K and expiration time N. We consider both the European and the American Call; the European with payoff $(S_N - K)^+$ at time N, and the American one that can be exercised at any time $n = 0, \dots, N$ with payoff $(S_n - K)^+$. Denote by V_n^{eur} the value of the European Call at time n, which is

$$V_n^{eur} = (1+r)^{-(N-n)} \mathbf{E}_{\mathbf{P}^*}[(S_N - K)^+ \mid \mathcal{F}_n],$$

(we omit from now on the subindex $\mathbf{P}^*$), and by V_n^{amer} the value of the American one, that we do not know how to compute. At every time $n < N$, with an European call the owner can only do one thing:

- To keep it or to sell at price V_n^{eur}.

Obviously, to keep or to sell are two different things, but from the point of view of pricing, the only important is that the derivative has a price.

With an American Call the owner can do two things.

- To keep it or to sell at price V_n^{amer}.
- Exercise with a payoff $(S_n - K)^+$.

In general, an American derivative has three ingredients:

1. An adapted stochastic process $X_0, \ldots, X_N$, with $X_n \geq 0$, that is the payoff of exercising the derivative at time n. In the Call case, $X_n = (S_n - K)^+$.

2. A reference European derivative that expires at time N with payoff X_N. We denote the value of that derivative at time n by V_n^{eur}, and it is

$$V_n^{eur} = (1+r)^{-(N-n)} \mathbf{E}[X_N \mid \mathcal{F}_n].$$

For the American Call, obviously we have the European Call with payoff $X_N = (S_N - K)^+$.

3. The value of the American derivative that is a positive adapted stochastic process that we will denote by $V_0^{amer}, \ldots V_N^{amer}$, which for the moment is unknown, and our purpose is to compute it.

To compute the values V_n^{amer} we start by remarking that if in the instant N we have not exercised yet the American derivative, this can only be exercised, and therefore it has the same value that the European one, that is,

$$\boxed{V_N^{amer} = V_N^{eur} = X_N.} \tag{9.1}$$

What is more, at every n it has to be fulfilled

$$\boxed{V_n^{amer} \geq V_n^{eur},}$$

since the American derivative gives to the owner more rights than the European one. In another way, if at time $n < N$, we have

$$V_n^{amer} < V_n^{eur}$$

then there is a possibility of arbitrage selling an European derivative and buying an American one, which will have the same value at time N.

9.2 Value of an American derivative

To calculate the value V_n^{amer} it is used the so-called *dynamic programming* which consists in the following steps

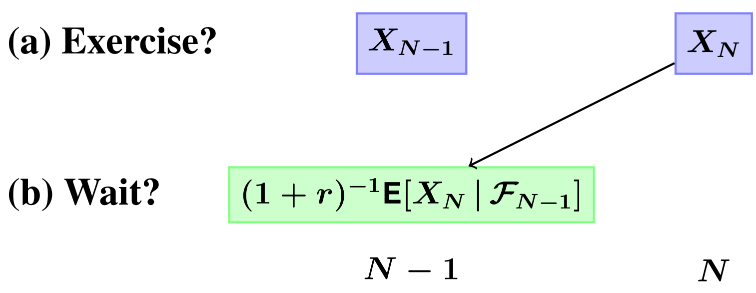
1. As we have commented, at time N the American derivative and the European one have the same value:

$$V_N^{amer} = V_N^{eur}.$$

2. At time $N-1$, with the American derivative the owner can make two things:

(a) To exercise it.

(b) To keep it (or to sell it) so that the owner (or the new owner) may exercise next time,



In case (a) the value is X_{N-1}. In case (b) it is like a European option that matures in one unity of time. Then its value is

$$(1+r)^{-1}\mathbf{E}[X_N \mid \mathcal{F}_{N-1}].$$

Therefore, the value will be the greatest between both possibilities:

$$V_{N-1}^{amer} = \max\{X_{N-1},\ (1+r)^{-1}\mathbf{E}[X_N \mid \mathcal{F}_{N-1}]\}.$$

3. At time $N-2$, dynamic programming says that the value of the derivative is computed with the same ideas: it should be the maximum between the value of exercising it or the value that it has today if the owner wait to the following step to decide,

$$V_{N-2}^{amer} = \max\{X_{N-2},\ (1+r)^{-1}\mathbf{E}[V_{N-1}^{amer} \mid \mathcal{F}_{N-2}]\}.$$

4. With the same arguments, if we have the value at time n, V_n^{amer}, then the value at time $n-1$ is

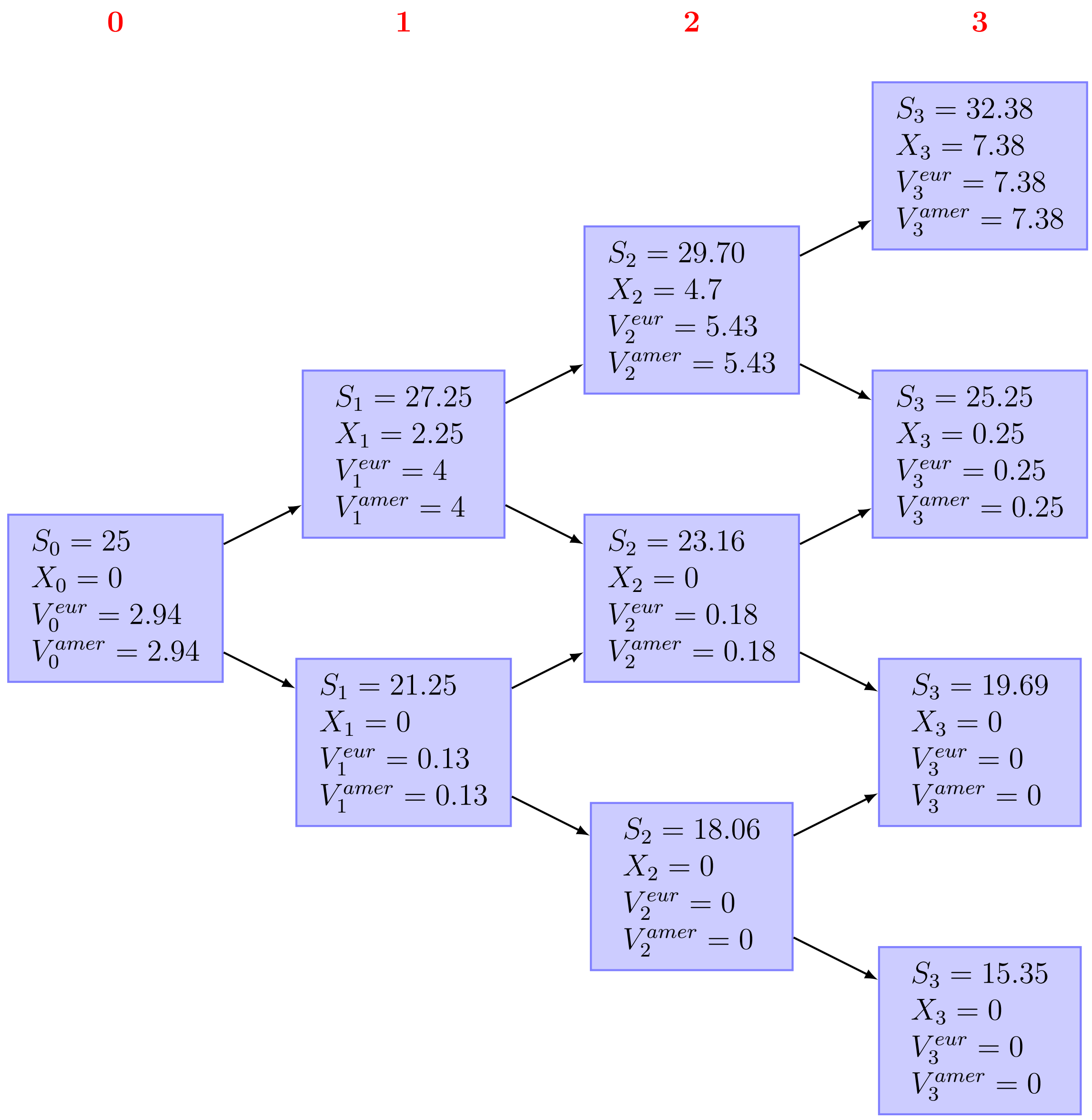
$$\boxed{V_{n-1}^{amer} = \max\{X_{n-1},\ (1+r)^{-1}\mathbf{E}[V_n^{amer} \mid \mathcal{F}_{n-1}]\}.}$$

Note that by construction

$$\boxed{V_n^{amer} \geq X_n} \tag{9.2}$$

Example

Consider a CRR model with $r = 0.03$, $S_0 = 25$, $u = 1.09$, $d = 0.85$, $p = 0.75$ (risk neutral probability) and $N = 3$, and an American Call with strike price $K = 25$ and $N = 3$.



As we see, the value of the American Call is the same as that of the European. This is a property of the American Call over an underlying that does not pay dividends. This is a consequence of the following general result:

Proposition 1. With the above notations, suppose that for every n,

$$V_n^{eur} \geq X_n.$$

Then

$$V_n^{amer} = V_n^{eur},\ n = 0, \ldots, N,$$

and it is optimal not to exercise the American option before the date N. (If the owner needs the money, it is better to sell the derivative at price V_n^{amer}).

Proof. At time N, $V_N^{amer} = V_N^{eur} = X_N$. At time $N-1$,

$$V_{N-1}^{amer} = \max\{X_{N-1},\ (1+r)^{-1}\mathbf{E}[X_N \,|\, \mathcal{F}_{N-1}]\} = \max\{X_{N-1},\ V_{N-1}^{eur}\}.$$

Hence, if $V^{eur}_{N-1} \geq X_{N-1}$,,

$$V^{amer}_{N-1} = V^{eur}_{N-1}.$$

Using the same idea, the property is proved by backward induction, that is, assuming that the property is true for n and proving it for $n-1$.

Corollary 1. Consider a Call. When the underlying does not pay dividends, then $V^{amer}_n = V^{eur}_n$, $n = 0, \ldots, N$, and it is optimal to exercise the Call at time N.

For the demonstration, see the appendix.

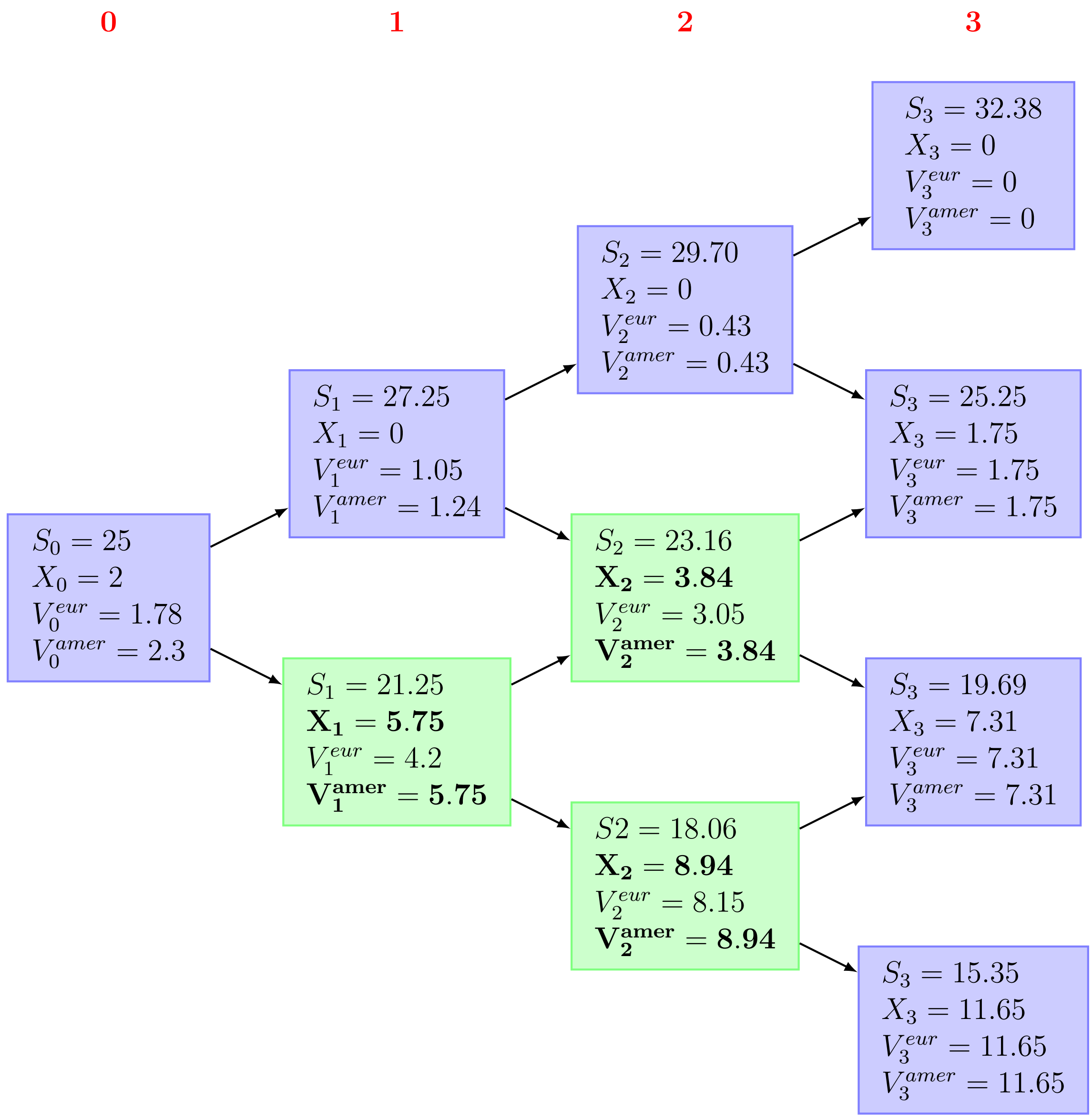
The above results are not true for all American derivatives. In general, European Puts are not worth the same than the American ones even when there are no dividends. The value

$$V^{amer}_0 - V^{eur}_0$$

is named *premium of early exercise*.

On the other hand, the quantity X_n is also called the **intrinsic value** of the American product. As we commented, by construction, $V^{amer}_n \geq X_n$; when on a node $V^{amer}_n = X_n$, it is said that it is a **optimal time** to exercise the derivative. In that point, $X_n \geq V^{eur}_n$. The detection of these times has the maximum interest.

Example. In the same model as before we consider an American Put with strike price $K = 27$, that is, with payoff $X_n = (27 - S_n)^+$.



Then we see that the optimal times are time 1 in lower path, and time 2, in the medium and lower paths.

9.3 Pricing and hedging an American option in a CRR model

The dynamic programming procedure fits perfectly to the backward recursion that we have seen for the CRR model. With the notations that we used there, first the value $V_N^{amer} = V_N^{eur} = X_N$ is computed in all nodes, and then backwarding with the formula

$$\boxed{V_n^{amer} = \max\left\{X_n, (1+r)^{-1}\big(p\,V_{n+1}^{amer,up} + (1-p)V_{n+1}^{amer,down}\big)\right\}}$$

all values V_n^{amer} are obtained.

The hedge is computed forward, starting by

$$D_0 = V_0^{amer} \quad \text{and} \quad H_0 = 0.$$

Assuming that $V_0^{amer} > X_0$, take

$$H_1 = \frac{V_1^{amer,up} - V_1^{amer,down}}{(u-d)S_0},$$

as in the European case, and

$$D_1 = V_0^{amer} - H_1 S_0.$$

In a similar way, if for $n = 1, \dots N-1$, we have $V_n^{amer} > X_n$ then

$$H_{n+1} = \frac{V_{n+1}^{amer,up} - V_{n+1}^{amer,down}}{(u-d)S_n},$$

and

$$D_{n+1} = \frac{V_n^{amer} - H_{n+1} S_n}{(1+r)^n}.$$

However, if $V_n^{amer} = X_n$, that is, (on this node) n is an optimal time to exercise, and the owner of the option does not do, then the seller can consume a quantity

$$C_n = V_n^{amer} - \frac{1}{1+r}\mathbf{E}[V_{n+1}^{amer} \mid \mathcal{F}_n], \tag{9.3}$$

and still maintain the hedge. This is due to the fact that the portfolio is designed to cover the extra premium of the American Put in an optimal time; if the owner does not profit the opportunity, then the seller can do.

Then, in the portfolio the quantity D_n has been reduced to

$$D_n^* = D_n - (1+r)^{-n} C_n,$$

and the value of the portfolio is now

$$V_n^* = D_n^* (1+r)^n + H_n S_n.$$

Note that by substitution of the values we get

$$V_n^* = \frac{1}{1+r}\mathbf{E}[V_{n+1}^{amer} \mid \mathcal{F}_n],$$

as in an European option. In that node, the hedging is

$$H_{n+1} = \frac{V_{n+1}^{amer,up} - V_{n+1}^{amer,down}}{(u-d)S_n},$$

as before, but

$$D_{n+1} = \frac{V_n^* - H_{n+1} S_n}{(1+r)^n}.$$

A bit of algebra shows that

$$(1+r)^{n+1}D_{n+1} + H_{n+1}S_{n+1}^{up} = V_{n+1}^{amer,up}$$

and

$$(1+r)^{n+1}D_{n+1} + H_{n+1}S_{n+1}^{down} = V_{n+1}^{amer,down},$$

so the value of the portfolio is exactly the same of the American Put.

This result seems a bit surprising; certainly, that portfolio is not self–financing, and the discounted value process $\{\widetilde{V}_n^{amer},\ n = 0,\dots,N\}$ is not more a martingale but a supermartingale; The fact that it is a supernartingale is obvious from construction: Since

$$V_{n-1}^{amer} = \max\{X_{n-1},\ (1+r)^{-1}\mathbf{E}[V_n^{amer} \,|\, \mathcal{F}_{n-1}]\},$$

it follows that

$$V_{n-1}^{amer} \ge (1+r)^{-1}\mathbf{E}[V_n^{amer} \,|\, \mathcal{F}_{n-1}],$$

and thus,

$$\widetilde{V}_{n-1}^{amer} \ge \mathbf{E}[\widetilde{V}_n^{amer} \,|\, \mathcal{F}_{n-1}].$$

If there are optimal times $< N$ to exercise, then $\{\widetilde{V}_n^{amer},\ n = 0,\dots,N\}$ is not a martingale (see the example below).

It can be proved that $\{\widetilde{V}_n^{amer},\ n = 0,\dots,N\}$ is the smallest supermartingale such that

$$V_n^{amer} \ge X_n,\ \forall n = 1,\dots,N.$$

It is also rational that a convenient strategy for the owner is to exercise at the first optimal time, that is, at

$$T = \inf\{k = 0,\dots,N:\ X_k = V_k^{amer}\}.$$

Since both X_n and V_n^{amer} are adapted processes, this is a stopping time, that means, it is a honest way to decide when to exercise the option. Moreover (see Appendix 2),

- The stopped process $\widetilde{V}_{n\wedge T}^{amer}$, where $a \wedge b = \min(a,b)$ (see page 51 in Chapter 4) is a martingale.
- The stopping time T satisfies

$$V_0^{amer} = \mathbf{E}[\widetilde{X}_T] = \sup_\tau \mathbf{E}[\widetilde{X}_\tau],$$

  where τ is any stopping time such that $\tau \le N$. That means, between all honest strategies, to exercise at the first optimal times is the best one.

Example. We will do all computations assuming that $S_1 = 21.25$, $S_2 = 18.06$ and $S_3 = 19.69$.

Time 0. The seller starts the portfolio with $D_0 = V_0^{amer} = 2.3$ and $H_0 = 0$. Now,

$$H_1 = \frac{V_1^{amer,up} - V_1^{amer,down}}{(u-d)S_0} = -0.75,$$

and

$$D_1 = V_0^{amer} - H_0 \cdot S_0 = 21.09.$$

So that means, she sells short 0.75 assets, form which she receives $0.75 * 25 = 18.79$ €, that she adds to the 2.3 that there already were in the portfolio. So the portfolio is

n	D_n	B_n	H_n	S_n	V_n
0	2.3	1	0	25	2.3
0+	21.09	1	-0.75	25	2.3

Time 1. Assume that $S_1 = 21.25$ and that the owner exercises the option. The seller has Cash=$21.09 \times 1.03 = 21.72$ € and she should return 0.75 asset. Then

Action	**Cash**	**Assets**
Buy an asset for 27 €	-27	1
Return 0.75 asset		-0.75
Sell 0.25 asset	5.28	-0.25
Cash	21.72	
Total	0	0

Time 1. Assume that $S_1 = 21.25$ and although it is an optimal time, the owner does not exercise the option. Then the seller computes $\mathbf{E}[V_2 \,|\, \mathcal{F}_1] = \mathbf{E}[V_2 \,|\, S_1]$ for this particular node:

$$\mathbf{E}[V_2 \,|\, S_1 = 21.25] = 0.75 \times 3.84 + 0,25 \times 8.94 = 5.11.$$

Then the seller can consume

$$C_1 = V_1 - \frac{1}{1+r}\mathbf{E}[V_2 \,|\, S_1 = 21.25] = 5.75 - \frac{1}{1.03}5.11 = 0.79.$$

So the portfolio is

n	D_n	B_n	H_n	S_n	V_n	C_n	$\widetilde{C}_n$
0	2.3	1	0	25	2.3		
0+	21.09	1	-0.75	25	2.3		
1	21.09	1.03	-0.75	21.25	5.75		
1+	20.32	1.03	-0.75	21.25	4.96	0.79	0.76

The value of the portfolio after the consume is 4.96. Then

$$H_2 = \frac{V_2^{amer,up} - V_2^{amer,down}}{(u-d)S_1} = -1,$$

and

$$D_2 = \frac{4.96 - (-1) \times 21.25}{1.03} = 25.45.$$

Hence,

n	D_n	B_n	H_n	S_n	V_n	C_n	$\widetilde{C}_n$
0	2.3	1	0	25	2.3		
0+	21.09	1	-0.75	25	2.3		
1	21.09	1.03	-0.75	21.25	5.75		
1+	20.32	1.03	-0.75	21.25	4.96	0.79	0.76
1+	25.45	1.03	-1	21.25	4.96		

That is, the seller sells short another 0.25 asset.

Time 2. Assume that $S_2 = 18.06$ and the owner exercises the option. The seller has Cash=$25.45 \times 1.03 = 27$ and she should return an asset. Then

Buy an asset for 27 €	-27	1
Return an asset		-1
Cash	27	
Total	0	0

Time 2. Assume that $S_2 = 18.06$ and the owner does not exercises the option. Then

$$\mathsf{E}[S_3 \,|\, S_2 = 18.06] = 0.75 \times 7.31 + 0.25 \times 11.61 = 8.39.$$

Then the seller consumes

$$C_2 = 8.94 - \frac{8.39}{1.03} = 0.79.$$

The portfolio is

n	D_n	B_n	H_n	S_n	V_n	C_n	$\widetilde{C}_n$
0	2.3	1	0	25	2.3		
0+	21.09	1	-0.75	25	2.3		
1	21.09	1.03	-0.75	21.25	5.75		
1+	20.32	1.03	-0.75	21.25	4.96	0.79	0.76
1++	25.45	1.03	-1	21.25	4.96		
2	25.45	1.0609	-1	18.06	8.94		
2+	24.71	1.0609	-1	18.06	8.94	0.79	0.74

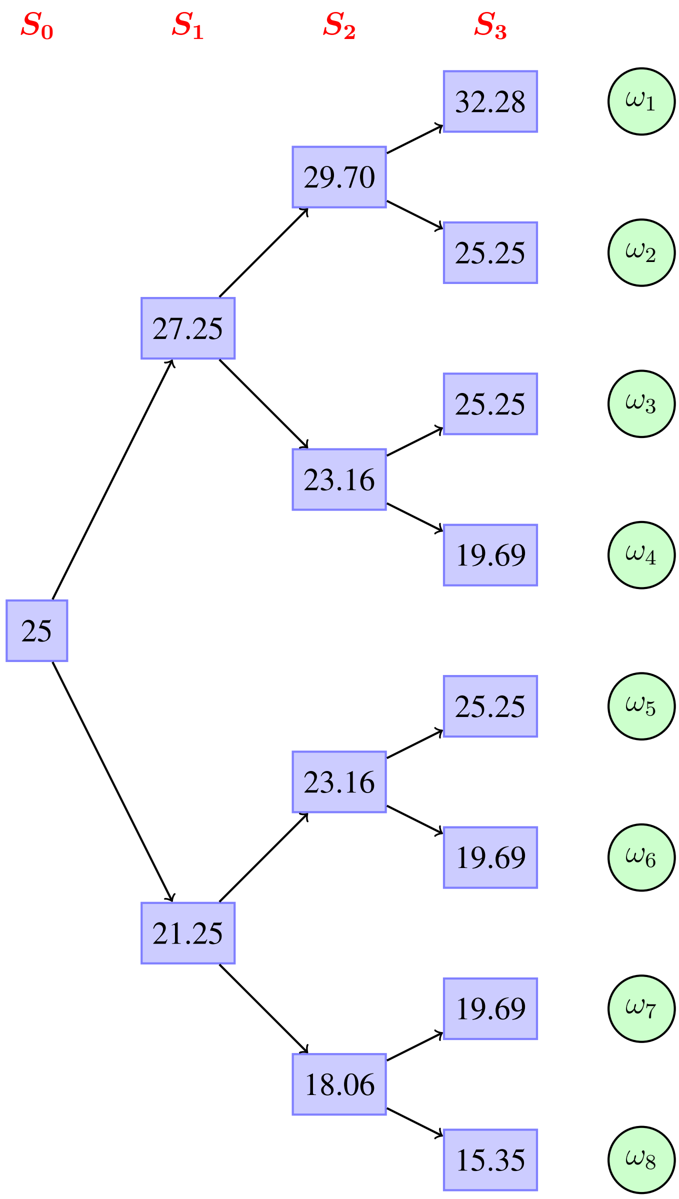
Now compute with the previous formula, $H_3 = -1$ and $B_3 = \frac{V_3^* - H_3 * S_2}{1.03^2} = 24.7$.

Time 3. Assume that $S_3 = 19.69$

n	D_n	B_n	H_n	S_n	V_n	C_n	$\widetilde{C}_n$
0	2.3	1	0	25	2.3		
0+	21.09	1	-0.75	25	2.3		
1	21.09	1.03	-0.75	21.25	5.75		
1+	20.32	1.03	-0.75	21.25	4.96	0.79	0.76
1++	25.45	1.03	-1	21.25	4.96		
2	25.45	1.0609	-1	18.06	8.94		
2+	24.71	1.0609	-1	18.06	8.15	0.79	0.74
2+ +	24.71	1.0609	-1	18.06	8.15		
3	24.71	1.0927	-1	19.68	7.31		

Now the option is exercised: the seller has cash $24.71 \times 1.0927 = 27$ €, so she can buy the asset corresponding to the option, and return it.

We check the supermartingale properties and the stopping time character of the optimal times to exercise. Here it is convenient to desegregate the tree, and write explicitly the space of probability, $\Omega = \{\omega_1, \ldots, \omega_8\}$:



The filtration is generated by S_n, that is,

$$\mathcal{F}_1 = \{\emptyset, \Omega, A_1, A_2\},$$

where

$$A_1 = \{S_1 = 27.25\} = \{\omega_1, \omega_2, \omega_3, \omega_4\},$$

and

$$A_2 = \{S_1 = 21.25\} = A_1^c,$$

and so on. We commented that in general, for all $n = 0, \dots, N-1$,

$$\widetilde{V}_n^{amer}(\omega) \geq \mathbf{E}[\widetilde{V}_{n+1}^{amer} \mid \mathcal{F}_1](\omega), \; \forall \omega \in \Omega, \text{ a.s.}.$$

We will show that the inequality is strict for some n. Compute $\mathbf{E}[V_2^{amer} \mid \mathcal{F}_1] = \mathbf{E}[V_2^{amer} \mid S_1]$ (this is easily computed in the previous diagram):

$$\mathbf{E}[V_2^{amer} \mid S_1 = 27.25] = 0.43 \times 0.75 + 3.84 \times 0.25 = 1.28,$$

and similarly,

$$\mathbf{E}[V_2^{amer} \,|\, S_1 = 21.25] = 5.11.$$

Note that $1.28/1.03 = 1.24$, which is the value of the American Put at this node; on the contrary, $5.11/1.03 = 4.96 \neq 5.75$, due that in this node there is an optimal time to exercise. Then,

$$\mathbf{E}[\widetilde{V}_2^{amer} \,|\, \mathcal{F}_1](\omega) = \begin{cases} 1.21, & \text{if } \omega = \omega_1, \dots, \omega_4, \\ 4.81, & \text{if } \omega = \omega_5, \dots, \omega_8. \end{cases}$$

And

$$\widetilde{V}_1^{amer}(\omega) = \begin{cases} 1.21, & \text{if } \omega = \omega_1, \dots, \omega_4, \\ 5.58, & \text{if } \omega = \omega_5, \dots, \omega_8. \end{cases}$$

Hence, for $\omega \in \{\omega_5, \dots, \omega_8\}$,

$$\widetilde{V}_1^{amer}(\omega) > \mathbf{E}[\widetilde{V}_2^{amer} \,|\, \mathcal{F}_1](\omega).$$

Finally, let

$$T = \inf\{k = 0, \dots, N : \; X_k = V_k^{amer}\}.$$

We have

$$T(\omega) = \begin{cases} 3, & \text{if } \omega = \omega_1, \omega_2, \\ 2, & \text{if } \omega = \omega_3, \omega_4, \\ 1, & \text{if } \omega = \omega_5, \dots, \omega_8. \end{cases}$$

and

$$\{T = 3\} = \{S_1 = 27.25, \; S_2 = 29.70\} \in \mathcal{F}_2 \subset \mathcal{F}_3,$$
$$\{T = 2\} = \{S_1 = 27.25, \; S_2 = 23.16\} \in \mathcal{F}_2,$$

and

$$\{T = 1\} = \{S_1 = 27.25\} \in \mathcal{F}_1,$$

9.4 Appendix

A1. Proof of the corollary 1

First we need a property of the European Calls: denote by V_n^{eur} the value of an European Call that expires at time N and strike price K.

Lemma.

$$V_n^{eur} \geq S_n - K.$$

Proof. Use Jensen's inequality with the function $f(x) = x^+$,

$$\big(\mathbf{E}[S_N - K \,|\, \mathcal{F}_n]\big)^+ \leq \mathbf{E}[(S_N - K)^+ \,|\, \mathcal{F}_n] = (1+r)^{N-n} V_n^{eur}. \tag{9.4}$$

Since $\widetilde{S}_n$ is a martingale, the term within $(...)^+$ is

$$\begin{aligned} \mathbf{E}[S_N \,|\, \mathcal{F}_n] - K &= (1+r)^N \mathbf{E}[\widetilde{S}_N \,|\, \mathcal{F}_n] - K = (1+r)^N \widetilde{S}_n - K = (1+r)^{N-n} S_n - K \\ &= (1+r)^{N-n}\big(S_n - (1+r)^{-(N-n)} K\big). \end{aligned}$$

Joining with (9.4),

$$V_n^{eur} \geq \big(S_n - (1+r)^{-(N-n)}K\big)^+ \geq_{(*)} S_n - (1+r)^{-(N-n)}K \geq_{(**)} S_n - K,$$

where $(*)$ is due that for all $a \in \mathbb{R}$, we have $a^+ \geq a$, and $(**)$ that $(1+r)^{-(N-n)} \leq 1$.

Proof of the corollary 1

We will apply Proposition 1; with this purpose, we will see that for every n,

$$V_n^{eur} \geq (S_n - K)^+.$$

This happens because

- If $S_n \leq K$, then $(S_n - K)^+ = 0$, and since $V_n^{eur} \geq 0$ we have the inequality.
- If $S_n > K$, then $(S_n - K)^+ = S_n - K$, and by the Lemma

  $$V_n^{eur} \geq S_n - K,$$

  and the inequality is also obtained.

A2. Supermartingales and martingales related to American options

The following three properties are true in a general model:

1. $\{\widetilde{V}_n^{amer},\ n = 0, \ldots, N\}$ is the smallest supermartingale such that

$$V_n^{amer} \geq X_n,\ \forall n = 1, \ldots, N.$$

2. Let T be the stopping time

$$T = \inf\{k = 0, \ldots, N :\ X_k = V_k^{amer}\}. \tag{9.5}$$

Then the stopped process $\{\widetilde{V}_{n\wedge T}^{amer},\ n = 0, \ldots, N\}$ is a martingale. Note that since $X_N = V_N^{amer}$ (see (13.7)), the set on the right hand side of (9.5) is nonempty.

3. The stopping time T satisfies

$$V_0^{amer} = \mathbf{E}[\widetilde{X}_T] = \sup_\tau \mathbf{E}[\widetilde{X}_\tau],$$

where τ is any stopping time such that $\tau \leq N$.

Proof of **1.** The proof was given in page 102

Proof of **2.** To simplify the notations write

$$Y_n = V_n^{amer},$$

and set

$$Y_n^T = Y_{T \wedge n}$$

the stopped process. First, for $n = 0, \ldots, N-1$,

$$\widetilde{Y}_{n+1}^T - \widetilde{Y}_n^T = \mathbf{1}_{\{T \geq n+1\}} \big(\widetilde{Y}_{n+1} - \widetilde{Y}_n\big). \tag{9.6}$$

This is checked in the following way: (we can suppress the discounting factor in both sides). if $T(\omega) = k < n+1$, then

$$Y_{n+1}^T(\omega) - Y_n^T(\omega) = Y_k(\omega) - Y_k(\omega) = 0,$$

and also the right hand side of (9.6) is zero. If $T(\omega) \geq n+1$, then the left hand side of (9.6) is

$$Y_{n+1}^T(\omega) - Y_n^T(\omega) = Y_{n+1}(\omega) - Y_n(\omega),$$

which is equal to the right hand side.

By definition of T, on the set $\{T \geq n+1\}$, $Y_n = V_n^{amer} > X_n$, and then

$$Y_n = V_n^{amer} = (1+r)^{-1} \mathbf{E}[V_{n+1}^{amer} \,|\, \mathcal{F}_n],$$

thus

$$\widetilde{Y}_n = \widetilde{V}_n^{amer} = \mathbf{E}[\widetilde{V}_{n+1}^{amer} \,|\, \mathcal{F}_n] = \mathbf{E}[\widetilde{Y}_{n+1} \,|\, \mathcal{F}_n].$$

Returning to (9.6),

$$\widetilde{Y}_{n+1}^T - \widetilde{Y}_n^T = \mathbf{1}_{\{T \geq n+1\}} \big(\widetilde{Y}_{n+1} - \mathbf{E}\big[\widetilde{Y}_{n+1} \,|\, \mathcal{F}_n\big]\big).$$

Hence,

$$\mathbf{E}\big[\widetilde{Y}_{n+1}^T - \widetilde{Y}_n^T \,|\, \mathcal{F}_n\big] = \mathbf{1}_{\{T \geq n+1\}} \mathbf{E}\Big[\widetilde{Y}_{n+1} - \mathbf{E}\big[\widetilde{Y}_{n+1} \,|\, \mathcal{F}_n\big] \,|\, \mathcal{F}_n\Big],$$

because $\{T \geq n+1\} = \{T \leq n\}^c \in \mathcal{F}_n$. It results that

$$\mathbf{E}\big[\widetilde{Y}_{n+1}^T - \widetilde{Y}_n^T \,|\, \mathcal{F}_n\big] = 0,$$

as we wanted to prove.

Proof of **3.** We have seen that $\widetilde{Y}_n^T$ is a martingale. Its value at time 0 is

$$\widetilde{Y}_{T \wedge 0} = \widetilde{Y}_0 = Y_0 = \mathbf{E}[\widetilde{Y}_N^T \,|\, \mathcal{F}_0] = \mathbf{E}[\widetilde{Y}_T] = \mathbf{E}[\widetilde{X}_T],$$

where we have used that since $T \leq N$, then $T \wedge N = T$, that we are assuming that $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so the conditional expectation to that σ–field is the ordinary expectation, and finally, that $Y_T = X_T$ (see (13.7)).

Furthermore, since $\widetilde{Y}_n$ is a supermartingale, for every stopping time $\tau \leq N$, the stopped process $\widetilde{Y}_n^\tau$ is also a supermartingale (see page 51 in chapter 4). Then

$$Y_0 = \widetilde{Y}_0^\tau \geq \mathbf{E}\big[\widetilde{Y}_N^\tau \,|\, \mathcal{F}_0] = \mathbf{E}[\widetilde{Y}_\tau] \geq \mathbf{E}[\widetilde{X}_\tau],$$

because $Y_n \geq X_n$ for all n, and hence $Y_\tau \geq X_\tau$.

9.5 Exercises

Consider a CRR model with

$$r = 0.04,\ u = 1.06,\ d = 0.98,\ S_0 = 20\,€ \quad \text{and} \quad p = 0.75 \text{ (risk neutral probability)}.$$

We will study an American Put with strike $K = 20\,€$ and maturation time $N = 4$.

Note that the data are the same as exercise in page 93, so it can be used the Put-Call parity to compute the values of the Put.

1. Construct the binomial tree for S_n.

2. Compute $V_4^{eur} = (K - S_4)^+$, and by backward recursion deduce V_n^{eur}, $n = 0, 1, 2, 3$:

$$V_n^{eur} = \frac{pV_{n+1}^{eur,up} + (1-p)V_{n+1}^{eur,down}}{1+r}.$$

3. From $V_4^{amer} = V_4^{eur}$, compute V_n^{amer}, $n = 0, 1, 2, 3$ with the recursion

$$V_n^{amer} = \max\left\{(K - S_n)^+, (1+r)^{-1}\left(p\,V_{n+1}^{amer,up} + (1-p)P_{n+1}^{amer,down}\right)\right\}.$$

Note that in that expression instead of $(K - S_n)^+$ you can use $K - S_n$, because the other factor of the maxima is always positive. Remember that in l'EXCEL for $\max\{A, B\}$ it is necessary to write max(A;B).

4. Check that

$$V_n^{amer} \geq V_n^{eur},\ n = 0, 1, \ldots, 4.$$

5. Compute the premium of early exercise.

6. Check that the node $S_1 = 19.6$ determines an optimal time to exercise. Compute the quantity that the seller of the Put can consume if the owner does not exercise.

Chapter 10

Brownian motion and Itô integral

10.1 Stochastic processes with continuous time parameter

From now on, we will deal with stochastic processes with continuous time parameter, that is, a family of random variables $\{X_t,\, t \in \mathbb{T}\}$, where $\mathbb{T}$ is the set of non-negative real numbers $\mathbb{R}_+ = [0, \infty)$, or an interval $[a, b]$. The random variables are defined on a probability space $(\Omega, \mathcal{F}, P)$.

For each $\omega \in \Omega$, the function

$$
\begin{aligned}
X_.(\omega) : \mathbb{T} &\longrightarrow \mathbb{R} \\
t &\mapsto X_t(\omega)
\end{aligned}
$$

is called a trajectory or realization of the stochastic process.

The most important example of stochastic processes is the Brownian motion.

10.2 The Brownian motion

The Brownian motion was discovered by the botanist Robert Brown in 1828 when he was studying the motion of pollen particles in water; the motion was very irregular and the pollen particles diffused in the liquid. That motion was caused by the shocks of water molecules on the particles of pollen, and Einstein in 1905, in a very brilliant paper, deduced the probabilities of the transition of a pollen particle between two points in some period of time. At the same time, starting by a very good work of Bachelier in 1900 (whose importance was not recognized in that moment) models based in a Brownian motion were used for the price dynamics of a stock. The mathematical construction of a model for the Brownian motion was done in 1924 by Norbert Wiener. However, the more intricate properties were proved by Paul Lévy in the years 1930's and 1940's. Lévy regretted all his live the fact that he was not the first to give a rigorous construction of that mathematical model, and in a nostalgic part of his autobiography wrote "I was very near and I am sure that in one or two years I would arrive, but Wiener advanced me." The mathematical model for the physical phenomenon of the Brownian motion is called the Wiener process, or also, confusing the reality and the model, a Brownian motion.

We will use mainly an unidimensional Brownian motion , but Brown observed a tridimensional motion; the first thing done by Einstein in his paper is to divide the tridimensional motion in three independent unidimensional motions. Then, we will consider the motion in one direction, and the trajectories are curves in the time-space plane. The definition is the following:

A Brownian motion is a stochastic process $\{W_t,\, t \in \mathbb{R}_+\}$ such that

1. $W_0 = 0,$ q.s. (the particle starts a 0).

2. For every $0 \leq s < t$, the random variable $W_t - W_s$ is Gaussian with mean 0 and variance $t - s$. We write

$$W_t - W_s \sim \mathcal{N}(0, t - s).$$

3. For every $0 \leq t_1 < \cdots < t_r$, the random variable

$$W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_r} - W_{t_{r-1}}$$

are independent. (It is said that the process has *independent increments*).

4. With probability 1, the trajectories

$$W_\omega(\cdot) : \mathbb{R}_+ \longrightarrow \mathbb{R}$$

are continuous functions.

From properties 2 and 3 it is deduced that the random vectors $(W_{t_1}, \ldots, W_{t_r})$ are multivariate Gaussian.

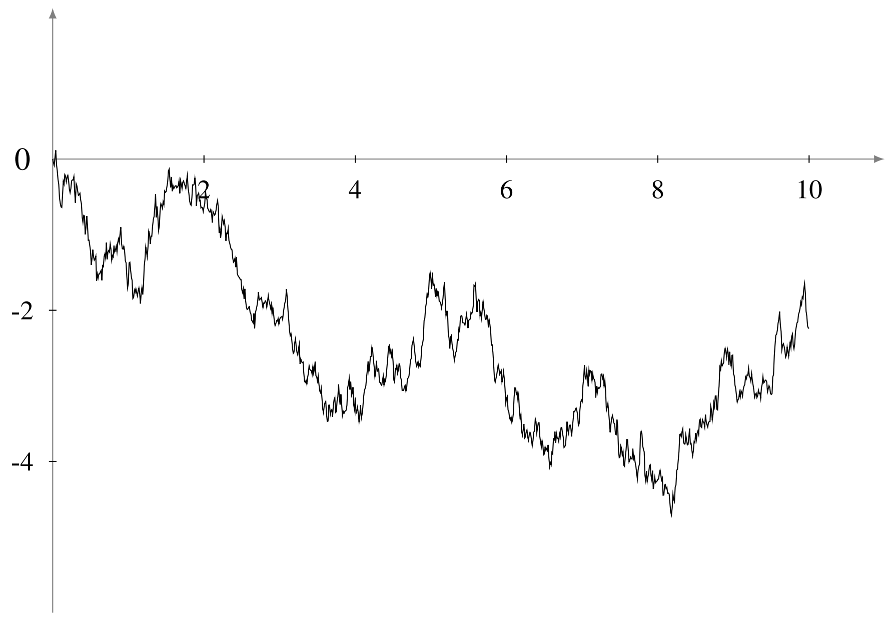


Figure 10.1. Approximate trajectory of a Brownian motion

The continuity of the trajectories is a very natural property, since the particle cannot jump. The trajectories are very irregular; indeed, they are impossible to plot. We will comment later that a trajectory is not differentiable; this is not very realistic, since it means that the particle has no velocity.

10.2.1 Simulation of a Brownian motion

Fix $T > 0$ (for example, take $T = 5$), and take N big (for example, $N = 1000$). Put

$$\delta = \frac{T}{N},$$

which is called the mesh of the approximation. In the example, $\delta = 0.005$. Write

$$t_0 = 0,\ t_1 = \delta,\ t_2 = 2\delta, \dots, t_N = N\delta = T.$$

From the property 2 of the Brownian motion,

$$W_{t_i} - W_{t_{i-1}} \sim \mathcal{N}(0, \delta) = \sqrt{\delta}\, Z,$$

where Z is a standard $\mathcal{N}(0,1)$ random variable. Moreover, by property 3,

$$W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_N - W_{t_{N-1}}$$

are all independent. Hence, the procedure to simulate is

(i) Generate $Z_1, \dots, Z_N$ independent $\mathcal{N}(0,1)$ random variables.

(ii) Put

$$\begin{aligned} &W_0 = 0, \\ &W_{t_i} = W_{t_{i-1}} + \sqrt{\delta} Z_i,\ i = 1, \dots, N. \end{aligned}$$

10.3 Martingales with continuous time parameter

We consider a general stochastic process $M = \{M_t,\, t \in \mathbb{T}\}$, where $\mathbb{T}$ is $[0, \infty)$ or an interval $[a, b]$, and a filtration $\mathbb{F} = \{\mathcal{F}_t,\ t \in \mathbb{T}\}$ so sub-σ-fields of $\mathcal{F}$: that is, if $s < t$, then $\mathcal{F}_s \subset \mathcal{F}_t$. As in the discrete time case, we say that M is a martingale with respect to $\mathbb{F}$ if

1. M_t is $\mathcal{F}_t$ measurable (M is adapted).

2. $\mathsf{E}[|M_t|] < \infty$.

3. $\mathsf{E}[M_t \,|\, \mathcal{F}_s] = M_s,$ for all $s < t$ (with $s, t \in \mathbb{T}$).

The **submartingales** and **supermartingales** are defined in a similar way

Remark. As in the discrete case,

$$E[M_t] = [M_s],\ \forall s, t \in \mathbb{T}.$$

In particular, if $\mathbb{T} = [0, T]$, then

$$E[M_t] = [M_0],\ \forall t \in [0, T].$$

Examples

In the following examples, we consider a a Brownian motion $W = \{W_t,\ t \geq 0\}$ and a filtration $\{\mathcal{F}_t,\ t \geq 0\}$ such that if $s < t$, the variable $W_t - W_s$ is independent of $\mathcal{F}_s$. Remember also that for $s < t$, $W_t - W_s \sim \mathcal{N}(0, t - s)$.

(1) W is a martingale. We only need to check that

$$\mathbf{E}[W_t \mid \mathcal{F}_s] = W_s.$$

This follows from the fact that $W_t - W_s$ is independent of $\mathcal{F}_s$. Hence, by the properties of the conditional expectation,

$$\mathbf{E}[W_t - W_s \mid \mathcal{F}_s] = \mathbf{E}[W_t - W_s] = 0.$$

On the other hand, since W_s is $\mathcal{F}_s$ measurable,

$$\mathbf{E}[W_t - W_s \mid \mathcal{F}_s] = \mathbf{E}[W_t \mid \mathcal{F}_s] - W_s,$$

and the result follows.

(2) $M_t = W_t^2 - t$ is a martingale.

The computation is more tricky. We want to check that

$$\mathbf{E}[W_t^2 - t \mid \mathcal{F}_s] = W_s^2 - s.$$

Equivalently, that

$$\mathbf{E}[W_t^2 \mid \mathcal{F}_s] = t - s + W_s^2.$$

To this end, we write $W_t = (W_t - W_s) + W_s$ and we use that $W_t - W_s$ is independent of $\mathcal{F}_s$ and W_s is $\mathcal{F}_s$ measurable.

$$\begin{aligned}\mathbf{E}[W_t^2 \mid \mathcal{F}_s] &= \mathbf{E}\big[\big((W_t - W_s) + W_s\big)^2 \mid \mathcal{F}_s\big] = \mathbf{E}\big[(W_t - W_s)^2 + 2(W_t - W_s)W_s + W_s^2 \mid \mathcal{F}_s\big] \\ &= \mathbf{E}\big[(W_t - W_s)^2\big] + 2W_s\mathbf{E}[W_t - W_s] + W_s^2 = t - s + W_s^2,\end{aligned}$$

where we have used that $W_t - W_s \sim \mathcal{N}(0, t-s)$.

10.3.1 Properties of martingales

The following properties are easy to prove:

1. Let $\{M_t,\, t \in [0,T]\}$ be a martingale. Then for every $a \in \mathbb{R}$ $N_t := aM_t$ is also a martingale.

2. Let $\{M_t,\, t \in [0,T]\}$ and $\{N_t,\, t \in [0,T]\}$ be martingales. Then $M_t + N_t$ is also a martingale.

10.4 Markov property of the Browian motion

10.4.1 Markov processes in continuous time

Consider a continuous time stochastic process $X = \{X_t,\, t \in \mathbb{T}\}$, and denote by $\mathcal{F}_t^X$ the σ–field generated by the random variables $X_s,\ s \le t$, which is interpreted as the *past* of X until time t. It is said that X is a Markov process if for every Borel set $B \in \mathcal{B}(\mathbb{R})$ and for every $u > t$,

$$\mathbf{P}(X_u \in B \mid \mathcal{F}_t^X) = \mathbf{P}(X_u \in B \mid X_t).$$

(Please, compare with the discrete case, page 56). In many occasions it is convenient that the family of σ–field to be a larger than $\mathcal{F}_t^X$, so we extend the notion of filtration to the continuous case.

Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. A **filtration** is a family $\{\mathcal{F}_t,\ t \in \mathbb{T}\}$ of sub–σ–fields of $\mathcal{F}$ such that if $s \leq t$ then $\mathcal{F}_s \subset \mathcal{F}_t$.
A stochastic process $X = \{X_t,\ t \in \mathbb{T}\}$ is said to be adapted to the filtration $\{\mathcal{F}_t,\ t \in \mathbb{T}\}$ if X_t is $\mathcal{F}_t$ measurable for all $t \in \mathbb{T}$.
If $X = \{X_t,\ t \in \mathbb{T}\}$ is adapted to the filtration $\{\mathcal{F}_t,\ t \in \mathbb{T}\}$ then $\mathcal{F}_t^X \subset \mathcal{F}_t$.

So we can give a more general definition of Markov process in the following way: We say that $X = \{X,\ t \in \mathbb{T}\}$ is a Markov process with respect to the the filtration $\{\mathcal{F}_t,\ t \in \mathbb{T}\}$ if it is adapted and for every Borel set $B \in \mathcal{B}(\mathbb{R})$, for every $u > t$,

$$\mathbf{P}(X_u \in B \,|\, \mathcal{F}_t) = \mathbf{P}(X_u \in B \,|\, X_t).$$

As in the discrete case, we have (see page 57)

The (adapted) process $\{X_t,\ t \in \mathbb{T}\}$ is Markov (with respect to the filtration $\{\mathcal{F}_t,\ t \in \mathbb{T}\}$) if an only if for every $u > t$ and every function $f : \mathbb{R} \to \mathbb{R}$ positive or such that $\mathbf{E}[|f(X_u)|] < \infty$,

$$\mathbf{E}[f(X_u) \,|\, \mathcal{F}_t] = \mathbf{E}[f(X_u) \,|\, X_t]. \tag{10.1}$$

Associate to a Markov process there is a system $P_{tu}(x, B)$, defined for $t < u$, $x \in \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$ such that

1. Fixed t, u, B, the map $P_{tu}(\cdot, B) : \mathbb{R} \to [0, 1]$ is Bórel measurable.

2. Fixed t, u, x, the map $P_{tu}(x, \cdot) : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is a probability measure on $\mathbb{R}$.

3.
$$\mathbf{P}(X_u \in B \,|\, X_t = x) = P_{tu}(x, B). \tag{10.2}$$

These maps are called the **transition probabilities** of the Markov process. That system satisfies the so-called Chapman-Kolmogorov equation:

$$P_{su}(x, B) = \int_{\mathbb{R}} P_{tu}(y, B) P_{st}(x, dy),$$

where this last expressions means that it is an integral with respect to the probability measure $P_{st}(x, \cdot)$. Joining property (10.2) with the defininion of Markov process we have that

$$\mathbf{P}(X_u \in B \,|\, \mathcal{F}_t) = P_{tu}(X_t, B).$$

A Markov process is said to be **homogeneus** if for every $u > t$, and $r > 0$,

$$\mathbf{P}(X_u \in B \,|\, \mathcal{F}_t) = \mathbf{P}(X_{u+r} \in B \,|\, \mathcal{F}_{t+r}),$$

that means, that probability $\mathbf{P}(X_u \in B \,|\, \mathcal{F}_t)$ only depends on the distance b $u - t$. This condition is transferred to the transition probabilities and, in this case, $P_{tu}(\cdot, \cdot)$ only depends on $u - t$.

10.4.2 Brownian motion as a Markov process

A Brownian motion is a Markov process. A short proof is based in an extension (which is not short to prove!) of the property of the conditional expectation (see page 36): If X, Y are independent random variables, then

$$E[h(X,Y)\,|\,Y] = E[h(X,y)]\big|_{y=Y}.$$

The extension is the following: Let X, Y be random variables and $\mathcal{G}$ a σ–field such that X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$ measurable. Then

$$E[h(X,Y)\,|\,\mathcal{G}] = E[h(X,y)]\big|_{y=Y}.$$

On the other hand, for $t < u$, the increment $W_u - W_t$ is independent of the σ–field $\mathcal{F}_t^W$ (please, try to give an intuitive explanation of this fact).

To prove Markov property, we check the property (10.1): consider $t < u$; we have

$$\begin{aligned}\mathbf{E}[f(W_u)\,|\,\mathcal{F}_t^W] &= \mathbf{E}[f(W_u - W_t + W_t)\,|\,\mathcal{F}_t^W] \\ &= \mathbf{E}[f(W_u - W_t + y)]\big|_{y=W_t} = \mathbf{E}[f(W_u)\,|\,W_t]\end{aligned}$$

It is also proved that the probability $P_{tu}(x,\cdot)$ has a density given by

$$p_{tu}(x,z) = \frac{1}{\sqrt{2\pi(u-t)}}\exp\{-\frac{(z-x)^2}{2(u-t)}\},$$

that means, it is a normal $\mathcal{N}(x, u-t)$ distribution. Given that this distribution only depends on $u - t$ it turns out that the Brownian motion is homogeneous.

10.5 Integration with respect to a Brownian motion: the naive approach

Let $W = \{W_t,\, t \geq 0\}$ be a Brownian motion, and $H = \{H_t,\, t \in [0,T]\}$ be another stochastic processes. Our goal is to define and compute

$$\int_0^T H(t)\,dW(t).$$

As we will see later, such kind of integrals are needed to model the financial markets.

It is first convenient to have an insight on the reasons that makes such integral problematic to define. We start with an informal discussion about functions of bounded variation.

10.5.1 Bounded variation function

Roughly speaking, a continuous function $g : [0,T] \to \mathbb{R}$ has bounded variation if the vertical distance travelled by the point $(t, g(t))$ for $t \in [0,T]$ is finite. For exemple, the function $g_1 : [0, 2\pi] \to \mathbb{R}$,

$$g_1(t) = \sin t,$$

has bounded variation. On the contrary, the function on $[0, 2\pi]$

$$g_2(t) = \begin{cases} 0, & \text{t=0}, \\ \sin(1/t), & \text{otherwise}, \end{cases}$$

has no bounded variation, or it is also said that it has infinite variation on $[0, 2\pi]$.

Formally, the definition (valid for continuous and non continuous functions) is the following. Let $g : [a, b] \to \mathbb{R}$ be a function, and consider a partition $\mathcal{P}$ of $[a, b]$:

$$a = t_0 < t_1 < \cdots < t_n = b.$$

The variation of g over the partition $\mathcal{P}$ is defined as

$$V_{\mathcal{P}} g = \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})|.$$

It is said that g has finite variation on $[a, b]$ if there is a constant K such that

$$V_{\mathcal{P}} g \leq K$$

for every partition of $[a, b]$. In that case, the number

$$V_{(a,b)} g = \lim V_{\mathcal{P}} g,$$

where the limit is taken over a sequence of partitions of $[a, b]$ with the mesh tending to 0, is called the variation of order 1 of g on $[a, b]$.

An easy condition that guarantees that a continuous function has bounded variation on $[a, b]$ is that g is differentiable in (a, b) with bounded derivative, $|g'(t)| \leq C$ (C is a constant). This is due that for every partition, by the mean value Theorem applied to each interval $[t_{i-1}, t_i]$

$$g(t_i) - g(t_{i-1}) = g'(s_i)(t_i - t_{i-1}),$$

for some $s_i \in (t_{i-1}, t_i)$. Then

$$\sum_i |g(t_i) - g(t_{i-1})| \leq \sum_i |g'(s_i)(t_i - t_{i-1})| \leq K \sum_i |t_i - t_{i-1}| = K(b - a).$$

The trajectories of a Brownian motion have infinite variation on every interval.

10.5.2 Integral with respect to functions of bounded variation functions

Let $h, g : [a, b] \to \mathbb{R}$ such that h is continuous and g is continuous and has bounded variation (it is possible to define the integral with less restrictive conditions on h and g, but for present introductory purposes that suffices). The Riemann-Stieljes integral is defined by

$$\int_a^b h(t)\, dg(t) = \lim \sum_i h(t_i)\big(g(t_i) - g(t_{i-1})\big),$$

where the limit is taken over a sequence of partitions of $[a, b]$ with the mesh tending to 0. (we can change the point t_i in $h(t_i)$ for any point in the interior of (t_{i-1}, t_i).) In particular, if g is differentiable, we have the familiar formula

$$\int_a^b h(t)\,dg(t) = \int_a^b h(t)g'(t)\,dt.$$

If $g(t) = t$, then we get the ordinary Riemann integral

$$\int_a^b h(t)\,dg(t) = \int_a^b h(t)\,dt.$$

All the main properties of the ordinary Riemann integral are transferred to the Riemann–Stieljes integrals, and we say that Riemann–Stieljes integral with respect to bounded variation function is a satisfactory theory.

Example 1. Let $g : [0, T] \to \mathbb{R}$ a differentiable function of bounded variation. Then,

$$\int_0^T g(t)\,dg(t) = \int_0^T g(t)g'(t)\,dt = \frac{1}{2}\big(g^2(T) - g^2(0)\big). \tag{10.3}$$

The naive point of view for the integral

$$\int_0^T H(t)\,dW(t)$$

is integrate trajectory by trajectory: for each $\omega \in \Omega$, define

$$\int_0^T H_\omega(t)\,dW_\omega(t).$$

However (in general) this approach does not work because the trajectories W_ω are not of bounded variation.

10.5.3 Quadratic variation

A remarkable property of the Brownian motion is that the trajectories have finite quadratic variation. Given a function $g : [a, b] \to \mathbb{R}$ it is said that g has quadratic variation on $[a, b]$ if (with the same notations as before)

$$QV_{\mathcal{P}}g := \sum_{i=1}^{n} \big(g(t_i) - g(t_{i-1})\big)^2 \leq K,$$

for every partition of $[a, b]$. In that case, the number

$$QV_{(a,b)}g = \lim QV_{\mathcal{P}}g,$$

where the limit is taken over a sequence of partitions of $[a, b]$ with the mesh tending to 0, is called the quadratic variation (or variation of order 2) of g on $[a, b]$. In general, the ordinary functions of the Mathematical Analysis have quadratic variation equal to 0. This can look a bit surprising, but the reader can convince herself studiying some simple examples as $g(t) = t$, or $g(t) = t^2$.

Indeed, we have if g is a continous function of bounded variation, then the quadratic variation is 0. The proof is as follows: for an arbitrary partition

$$\sum_{i=1}^{n} \big(g(t_i) - g(t_{i-1})\big)^2 \leq \sup_{i=1,\dots,n} |g(t_i) - g(t_{i-1})| \sum_{i=1}^{n} \big|g(t_i) - g(t_{i-1})\big| \leq \sup_{i=1,\dots,n} |g(t_i) - g(t_{i-1})| \, V_{(a,b)} g,$$

and $\sup_{i=1,\dots,n} |g(t_i) - g(t_{i-1})|$ tends to zero when the mesh of the partition tends to 0.

The trajectories of a Brownian motion have finite quadratic variation on every finite interval and for every $\boldsymbol{\omega}$ (almost sure) and

$$\boldsymbol{QV_{(0,t)} W_\omega = t}.$$

10.6 Itô integral

That property of the quadratic variation of the Brownian motion allows Itô to define the integral. This is a masterpiece of the Mathematics in the XX century. We restrict ourselves to a particular case where the integrant is a process with continuous trajectories and satisfies a condition of integrability. First we need to fix the notations. We consider a Brownian motion $W = \{W_t,\, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and there is a filtration $\{\mathcal{F}_t,\ t \geq 0\}$ such that

(i) W_t is $\mathcal{F}_t$ measurable (W is adapted).

(ii) For $0 \leq s < t$, the random variable $W_t - W_s$ is independent of $\mathcal{F}_s$.

Normally, we take $\mathcal{F}_t = \mathcal{F}_T^W$ the σ–field generated by $\{W_u,\ u \in [0,t]\}$, but, as we commented, it is convenient to have the possibility to use a more general filtration.

Definition. Let $H = \{H_t,\, t \in [0,T]\}$ be a stochastic process such that

(a) H_t is $\mathcal{F}_t$ measurable (H is adapted).

(b) $\int_0^T \mathbf{E}\big[H_t^2\big]\, dt < \infty$,

Then the Itô integral of H is defined as

$$\int_0^T H_t \, dW_t := \lim_n \sum_i H_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}),$$

where the limit in probability is taken over a sequence of partitions $\mathcal{P}_n$, where $\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_k = T\}$ of $[0,T]$ with the mesh tending to 0.

it is worth to remark that in comparison with the definition in Subsection 10.5.2, the main difference is that the limit is in probability, that means, globally as random variables, and not ω by ω.

Example 2. Take $H_t = W_t$. Condition (a) is obviously satisfied, and condition (b) also:

$$\int_0^T \mathbf{E}[W_t^2]\,dt = \int_0^T t\,dt = t^2/2 < \infty,$$

Then

$$\int_0^T W_t\,dW_t = \lim_n \sum_i W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}).$$

We have

$$\sum_i W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) = \frac{1}{2}\underbrace{\sum_i (W_{t_i}^2 - W_{t_{i-1}}^2)}_{(*)} - \frac{1}{2}\underbrace{\sum_i \big(W_{t_i} - W_{t_{i-1}}\big)^2}_{(**)}.$$

Developing the term (*), the factors positive and negative cancels, and the result is W_T^2. The term (**) goes, when the mesh of the partition tends to 0, to the quadratic variation of W on the interval $[0,T]$, which is T. Hence,

$$\int_0^T W_t\,dW_t = \frac{1}{2}\,W_T^2 - \frac{1}{2}\,T.$$

Remember that in (10.3) we obtaned that for a differentiable function of bounded variation g (assume $g(0) = 0$, as in the Brownian motion case)

$$\int_0^T g(t)\,dg(t) = \frac{1}{2}\,g^2(T).$$

That means, in this case the quadratic variation term is missing because it is 0.

10.6.1 Properties of the Itô integral

1.

$$\int_0^T (aH_t + bJ_t)\,dW_t = a\int_0^T H_t\,dW_t + b\int_0^T J_t\,dW_t.$$

2.

$$\mathbf{E}\Big[\int_0^T H_t\,dW_t\Big] = 0 \quad \text{and} \quad \mathbf{E}\Big[\Big(\int_0^T H_t\,dW_t\Big)^2\Big] = \int_0^T \mathbf{E}[H_t^2]\,dt.$$

3. Let $\{h_t,\, t \in [0,T]\}$ be a deterministic differentiable function with continuous derivative. Then we have the *integration by parts formula*

$$\int_0^T h(t)\,dB_t = h(T)B(T) - \int_0^T B_t h'(t)\,dt.$$

Extension. It is possible to define the integral for an adapted process $H = \{H_t,\, t \in [0,T]\}$ such that

$$\int_0^T H_t^2\,dt < \infty, \text{ a.s.}$$

Then the properties about the expectation (and expectation of the square) of $\int_0^T H_t\,dW_t$ are lost. Obviously,

$$\int_0^T \mathbf{E}\big[H_t^2\big]\,dt < \infty \quad \Longrightarrow \quad \int_0^T H_t^2\,dt < \infty, \text{ a.s.}$$

10.6.2 The Itô integral as a martingale

A major result in Itô integration theory is the following: Let $H = \{H_t,\, t \in [0,T]\}$ adapted and such that $\int_0^T \mathbf{E}[H_t^2]\, dt < \infty$. Then, for each $t \in [0,T]$ the process $\{H_s,\, s \in [0,t]\}$ satisfies also the above conditions for the integral, so we can construct a new family of random variables:

$$M_t := \int_0^t H_s \, dW_s.$$

$$\boxed{\text{The process } \{M_t,\, t \in [0,T]\} \text{ is a martingale.}}$$

The idea of the proof is very simple: consider a partition of $[0,T]$, and take $t_j < t_k$.

$$\mathbf{E}\Big[\sum_{i=1}^{k} H_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \,|\, \mathcal{F}_{t_j}\Big] = \underbrace{\mathbf{E}\Big[\sum_{i=1}^{j} H_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \,|\, \mathcal{F}_{t_j}\Big]}_{(*)} + \underbrace{\mathbf{E}\Big[\sum_{i=j+1}^{k} H_{t_i}(W_{t_i} - W_{t_{i-1}}) \,|\, \mathcal{F}_{t_j}\Big]}_{(**)}.$$

By the measurability properties of H and W,

$$(*) = \sum_{i=1}^{j} H_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}).$$

Further, for $i \geq j+1$, since the increment $W_{t_i} - W_{t_{i-1}}$ is independent of $\mathcal{F}_{t_{i-1}}$ and $H_{t_{i-1}}$ is $\mathcal{F}_{t_{i-1}}$ measurable, we can use the tower property of conditional expectations and deduce

$$\begin{aligned} \mathbf{E}\big[H_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \,|\, \mathcal{F}_{t_j}\big] &= \mathbf{E}\big[H_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \,|\, \mathcal{F}_{t_{i-1}} \,|\, \mathcal{F}_{t_j}\big] \\ &= \mathbf{E}\big[H_{t_{i-1}}\, \mathbf{E}[W_{t_i} - W_{t_{i-1}} \,|\, \mathcal{F}_{t_{i-1}}] \,|\, \mathcal{F}_{t_j}\big] = \mathbf{E}\big[H_{t_{i-1}}\, \mathbf{E}[W_{t_i} - W_{t_{i-1}}] \,|\, \mathcal{F}_{t_j}\big] = 0. \end{aligned}$$

Hence, $(**) = 0$. Now, passing to the limit all the expression we get the property.

Remark. Let $\{H_t,\, t \in [0,T]\}$ a stochastic process (or a deterministic function) such that for every ω,

$$\int_0^T |H_t(\omega)|\, dt < \infty.$$

Then

$$D_t = \int_0^t H_s \, ds,\ t \in [0,T]$$

(defined pathwise) is **not** a martingale (unless $H_t = 0, \forall t$.)

10.7 Exercises

1. Fix $a \in \mathbb{R}$. Prove that

$$M_t = \exp\Big\{aW_t - \frac{a^2}{2}\, t\Big\}$$

is a martingale. *Indication:* Use that if $Z \sim \mathcal{N}(0,1)$, then

$$\mathbf{E}\big[e^{aZ}\big] = e^{a^2/2}.$$

2. Prove that $M_t = W_t^3 - 3tW_t$ is a martingale. *Indication:* Write $W_t = (W_t - W_s) + W_s$, as in Example 2 in page 114.

3. We have the following result (that we prove in next chapter): Let $f \in \mathcal{C}^{1,2}(\mathbb{R}_+ \times \mathbb{R})$. Denote by $\mathcal{L}$ the differential operator

$$\mathcal{L}f(t,x) = \frac{\partial}{\partial t}f(t,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x).$$

Then

$$M_t = f(t,B_t) - f(0,0) - \int_0^t \mathcal{L}f(s,B_s)\,ds$$

is a martingale. Use this property (without proving it) to solve previous Exercises 1 and 2.

Chapter 11

The Itô formula

11.1 The chain rule for ordinary calculus

As in the previous chapter it is convenient to remember a few properties of the ordinary differential calculus. Let $x : [0,T] \to \mathbb{R}$ be a differentiable function, and $f : \mathbb{R} \to \mathbb{R}$ be another differentiable function. The chain rule states that

$$\big(f(x(t))\big)' = f'(x(t))\, x'(t). \tag{11.1}$$

For example, if $x(t) = \sin t$ and $f(x) = x^2$, then

$$\big(\sin^2 t\big)' = 2 \sin t\, \cos t.$$

In differential notation, (11.1) looks

$$\boxed{d\big(f(x(t))\big) = f'(x(t))\, dx(t).} \tag{11.2}$$

Also, we can integrate both sides of (11.1) on $[0,t]$ and we get

$$\boxed{f(x(t)) - f(x(0)) = \int_0^t f'(x(s))\, dx(s).}$$

Assume that $x(0) = 0$, and consider again $f(x) = x^2$. We have, in agreement with the results in previous chapter, that

$$x(t)^2 = 2\int_0^t x(s)\, dx(s).$$

However, we also saw that for a Brownian motion, $\int_0^t W_s\, dW_s = W_t^2/2 - t/2$, or equivalently,

$$W_t^2 = 2\int_0^t W_s\, dW_s + \int_0^t ds.$$

So, the chain rule for the Brownian motion should like something as

$$d(W_t^2) = 2W_t dW_t + dt.$$

The apparition of this new term dt is due to the quadratic variation of the Brownian motion.

11.2 The Itô formula.

Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$ function (f is two times differentiable and the derivatives are continuous). Then

$$\boxed{f(W_t) = f(0) + \int_0^t f'(W_s)\,dW_s + \frac{1}{2}\int_0^t f''(W_s)\,ds.} \tag{11.3}$$

In differential notation, we write

$$\boxed{d\big(f(W_t)\big) = f'(W_t)\,dW_t + \frac{1}{2}f''(W_t)\,dt.}$$

The proof is as follows. Remember that the Taylor expansion of f for x in a neighborhood of x_0 is

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x^*)(x - x_0)^2,$$

where x^* is an intermediate point between x and x_0. Start by taking a partition of $[0, t]$, $0 = t_0 < t_1 < \cdots < t_n = t$ and write

$$f(W_t) - f(0) = \sum_{i=1}^{n}\big(f(W_{t_i}) - f(W_{t_{i-1}})\big).$$

Develop each term $f(W_{t_i})$ using the Taylor formula at point $W_{t_{i-1}}$. We have

$$f(W_t) - f(0) = \sum_{i=1}^{n} f'(W_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^{n} f''(W_i^*)(W_{t_i} - W_{t_{i-1}})^2,$$

where W_i^* is a (random) intermediate point between W_{t_i} and $W_{t_{i-1}}$. The first sum goes to $\int_0^t f'(W_s)\,dW_s$ when the mesh of the partition goes to zero. By the quadratic variation property of the Brownian motion (and the continuity properties of all involved processes), the second sum goes to $\int_0^t f''(W_s)\,ds$.

Example 1. Taking $f(x) = x^2$ we get the expression of W_t^2:

$$W_t^2 = 2\int_0^t W_s\,dW_s + \int_0^t ds.$$

Example 2. Taking $f(x) = x^3$,

$$W_t^3 = 3\int_0^t W_s^2\,dW_s + 3\int_0^t W_s\,ds.$$

In differential notation,

$$d(W_t^3) = 3W_t^2\,dW_t + 3W_t\,dt.$$

Exercise 1. Apply Itô formula to $e^{W_t^2}$.

11.3 Extensions of the Itô formula

The Itô formula can be extended to consider other types of functions and the possibility of combining different Brownian motions. It is worth to do an extra effort to study a notation that will do things easier and easy to remember.

11.3.1 Itô processes

Let $\{H_t,\, t \in [0,T]\}$ and $\{G_t,\, t \in [0,T]\}$ two adapted stochastic processes such that

$$\int_0^T H_t^2 dt < \infty, \text{ a.s.} \quad \text{and} \quad \int_0^T |G_t| dt < \infty, \text{ a.s.}$$

A stochastic process of the form

$$\boxed{X_t = X_0 + \int_0^t H_s\, dW_s + \int_0^t G_s\, ds,\ t \in [0,T],}$$

where X_0 is a constant or a random variable independent of the Brownian motion, is called a **Itô process.** Using differential notation, we write (formally!)

$$dX_t = H_t\, dW_t + G_t\, dt.$$

Example 3. Consider $X_t = W_t^2$. By the Itô formula,

$$dX_t = 2W_t\, dW_t + \, dt.$$

Hence, X_t is an Itô process with $H_t = 2W_t$ and $G_t = 1$.

Let $dX_t = H_t\, dW_t + G_t\, dt$ be an Itô process and $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$. function. Then $Y_t = f(X_t)$ is an Itô process and

(a) Integral form (1).

$$\boxed{Y_t = f(X_t) = f(X_0) + \int_0^t f'(X_s) H_s\, dW_s + \int_0^t f'(X_s) G_s\, ds + \frac{1}{2} \int_0^t f''(X_s) H_s^2\, ds.}$$

(b) Differential form (1).

$$\boxed{dY_t = f'(X_t) H_t\, dW_t + f'(X_t) G_t\, dt + \frac{1}{2} f''(X_t) H_t^2\, dt.}$$

(c) Differential form (2). The first and second factor on the right hand side can be grouped in dX_t:

$$\boxed{dY_t = f'(X_t)\, dX_t + \frac{1}{2} f''(X_t) H_t^2\, dt.} \tag{11.4}$$

(d) Integral form (2). For a *good* adapted stochastic process $\{J_t,\, t \in [0,T]\}$ we can define the integral with respect to X by

$$\int_0^t J_s\, dX_s := \int_0^t J_s H_s\, dW_s + \int_0^t J_s G_s\, ds.$$

With such integral, the previous integral expression (1) becomes

$$\boxed{f(X_t) = f(X_0) + \int_0^t f'(X_s)\, dX_s + \frac{1}{2} \int_0^t f''(X_s) H_s^2\, ds.}$$

Exercise 1 (continuation). As in Example 3, consider the Itô process $X_t = W_t^2 = 2\int_0^t W_s\, dW_s + \int_0^t ds$. Compute $e^{W_t^2}$ as the the function $f(x) = e^x$ applied to the Itô process $X_t = W_t^2$.

11.4 The quadratic variation of an Itô process

The concept of the quadratic variation of the Brownian motion can be extended to an Itô process. Again, the formulas are more compact with such notion. Given an Itô process (with some additional conditions)

$$X_t = X_0 + \int_0^t H_s \, dW_s + \int_0^t G_s \, ds,$$

its quadratic variation is

$$\boxed{\langle X \rangle_t := \lim \sum_i \big(X(t_{i+1}) - X(t_i) \big)^2 = \int_0^t H_s^2 ds,}$$

where the limit is in probability, taken over a sequence of partitions with the mesh tending to 0.

For the Brownian case,

$$\langle W \rangle_t = t.$$

It is convenient to remark that the part of X with the integral $\int_0^t G_s \, ds$ (bounded variation part of X) does not have any contribution to the quadratic variation: if we put

$$M_t = \int_0^t H_s \, dW_s,$$

then

$$\langle X \rangle_t = \langle M \rangle_t = \int_0^t H_s^2 ds.$$

Then the Itô formula can be written as

$$\boxed{f(X_t) = f(X_0) + \int_0^t f'(X_s) X_s + \frac{1}{2} \int_0^t f''(X_s) \, d\langle X \rangle_s,}$$

or in differential notation

$$\boxed{d\big(f(X_t)\big) = f'(X_t) \, dX_t + \frac{1}{2} f''(X_t) \, d\langle X \rangle_t.}$$

That formulas are the extension of the basic Itô formula (11.3) to an Itô process with a convenient definition of the stochastic integral with respect to X and the quadratic variation of X.

Exercise 1 (Continuation) The Itô process $X_t = W_t^2 = 2 \int_0^t W_s \, dW_s + \int_0^t ds$, has quadratic variation

$$\langle X \rangle_t = 4 \int_0^t W_s^2 ds.$$

Compute again $e^{W_t^2}$ using that fact.

11.4.1 Itô formula for a function that depends also on time

Let $f(t,x) : [0,T] \times \mathbb{R} \to \mathbb{R}$ a $\mathcal{C}^{1,2}$ function, and

$$X_t = X_0 + \int_0^t H_s\, dW_s + \int_0^t G_s\, ds,$$

be an Itô process. Then, writing $Y_t = f(t, X_t)$ we have

$$\boxed{dY_t = f'_x(t,X_t)\, dX_t + \frac{1}{2} f''_{xx}(t,X_t) H_s^2\, ds + f'_t(t,X_t)\, dt,} \tag{11.5}$$

where

$$f'_x = \frac{\partial f}{\partial x}, \text{ etc.}$$

Equivalently, using the quadratic variation the formula looks simpler:

$$dY_t = f'_x(t,X_t)\, dX_t + \frac{1}{2} f''_{xx}(t,X_t)\, \langle X \rangle_t + f'_t(t,X_t)\, dt,$$

Example 4. Geometric Brownian motion. We want to compute the differential of

$$S_t = s_0 \exp\{\sigma W_t + (\mu - \sigma^2/2)t\},$$

where s_0 is a constant. This process is called a geometric Brownian motion, and is the basis for Black-Scholes model. We will do the computations twice, using different versions of the Itô formula.

(a) Define the Itô process $X_t = \sigma W_t + (\mu - \sigma^2/2)t$ and the function $f(x) = s_0 e^x$, so $S_t = f(X_t)$ We have $H_t = \sigma$ and $G_t = \mu - \sigma^2/2$. By formula (11.4),

$$dS_t = e^{X_t}\, dX_t + \sigma^2 \frac{1}{2}\, e^{X_t}\, dt = S_t\, dX_t + \frac{\sigma^2}{2} S_t\, dt,$$

and by decomposing X_t in its components,

$$dS_t = \sigma\, S_t\, dW_t + \mu S_t\, dt.$$

That expression is also written

$$\frac{dS_t}{S_t} = \sigma\, dW_t + \mu dt.$$

(b) Consider the function

$$g(t,x) = s_0 e^{\sigma^2 x + (\mu - \sigma^2/2)t},$$

and compute the partial derivatives g'_x, etc. Apply the previous Itô formula to $X_t = W_t$.

Exercise 2. Repeat previous applications of the Itô formula but now using the quadratic variation of the Itô process X.

11.5 Multidimensional Itô formulas

We will work from the general case to the most important particular cases. We will write all vectors in column, and given a vector or matrix $\boldsymbol{a}$, then $\boldsymbol{a}^T$ will denote the transposed vector or matrix. Let $\boldsymbol{W}(t) = (W_1(t), \ldots, W_k(t))^T$ be a vector of independent Brownian motions. We will denote by $\{\mathcal{F}_t,\ t \geq 0\}$ the filtration generated by these Brownian motions. Note that each Brownian motion $W_j(t)$ verifies the properties introduced before the definition of Itô integral in previous chapter. Let

$$\boldsymbol{H}(t) = \begin{pmatrix} H_{11}(t) & \cdots & H_{1k}(t) \\ \vdots & & \vdots \\ H_{n1}(t) & \cdots & H_{nk}(t) \end{pmatrix}$$

be a $n \times k$ matrix of adapted stochastic processes and $\boldsymbol{G}(t) = (G_1(t), \ldots, G_n(t))^T$ a vector of of adapted stochastic processes such that

$$\int_0^T H_{ij}^2(t)dt < \infty, \text{ a.s.} \quad \text{and} \quad \int_0^T |G_i(t)|dt < \infty, \text{ a.s.}$$

Consider the n-dimensional Itô process $\boldsymbol{X}_t = (X_1(t), \ldots, X_n(t))^T$

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{H}_s\, d\boldsymbol{W}_s + \int_0^t \boldsymbol{G}_s\, ds,$$

where the product $\boldsymbol{H}_s\, d\boldsymbol{W}_s$ is the product of a matrix by a vector and the integral is componentwise. In differential notation,

$$\begin{aligned} dX_1(t) &= \sum_{j=1}^k H_{1j}\, dW_j(t) + G_1(t)\, dt \\ &\vdots \\ dX_n(t) &= \sum_{j=1}^k H_{nj}\, dW_j(t) + G_n(t)\, dt \end{aligned}$$

We define the quadratic covariance of two processes (assuming that it has sense for these processes) M_i and M_j by *polarization*:

$$\langle M_i,\, M_j\rangle_t = \frac{1}{2}\,\langle M_i + M_j\rangle_t - \frac{1}{2}\,\langle M_i\rangle_t - \frac{1}{2}\,\langle M_j\rangle_t.$$

Note that this is the formula $ab = (a+b)^2/2 - a^2/2 - b^2/2$.

Exercise 3. Prove that if W_1 and W_2 are two independent Brownian motions, then $W_t := (W_1(t) + W_2(t))/\sqrt{2}$ is also a Brownian motion. Deduce that

$$\langle W_1,\, W_2\rangle_t = 0.$$

From previous exercise,

$$\langle W_i,\, W_j\rangle_t = \delta_{ij} t,$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that for processes $C(t)$ and $D(t)$ (under some conditions),

$$\langle \int C(s)dW_i(s), \int D(s)dW_j(s)\rangle_t = \int_0^t C(s)D(s)\, d\langle W_i,\, W_j\rangle_t = \delta_{ij} \int_0^t C(s)D(s)\, dt.$$

Finally, given that the quadratic covariance is bilinear, the quadratic variance of the Itô processes X_i i X_j is (in differential notion)

$$d\langle X_i,\, X_j\rangle_t = \sum_{m=1}^k H_{im}(t)H_{jm}(t)\, dt.$$

Then, the quadratic variation of the vector $\boldsymbol{X}$ is defined by the $n \times n$ matrix

$$\langle \boldsymbol{X}\rangle_t = \int_0^t \boldsymbol{H}(s)\boldsymbol{H}(s)^T\, ds.$$

1. **Multidimensional Itô formula.** Let $f : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^{1,2}$ function. then

$$\boxed{d\big(f(t, \boldsymbol{X}_t)\big) = \sum_{i=1}^n f_i'(t, \boldsymbol{X}_t)dX_i(t) + \frac{1}{2}\sum_{i,j=1}^n f_{i,j}''(t, \boldsymbol{X}_t)\, d\langle X_i, X_j\rangle_t + f_t'(t, \boldsymbol{X}_t)\, dt,}$$

where

$$f_i' = \frac{\partial f}{\partial x_i}, \text{ etc.}$$

Using matrices, it can be written as

$$\boxed{df(t, \boldsymbol{X}_t) = \boldsymbol{f}_{\boldsymbol{x}}'^T d\boldsymbol{X}(t) + \frac{1}{2} tr\big(\boldsymbol{f}_{\boldsymbol{x},\boldsymbol{x}}''\boldsymbol{H}\boldsymbol{H}^T\big)\, dt + f_t'\, dt,}$$

where

$$\boldsymbol{f}_{\boldsymbol{x}}' = \big(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\big)^T, \text{ etc.}$$

and for a matrix $\boldsymbol{A}$,

$$tr(\boldsymbol{A}) = \sum_i A_{ii}.$$

Remark: The expression with the trace is deduced using the symmetry of the matrices $\boldsymbol{f}_{\boldsymbol{x},\boldsymbol{x}}''$ and $\boldsymbol{H}\boldsymbol{H}^T$.

2. **Itô formula for two Itô processes respect to the same Brownian motion.**

$$\begin{aligned} dX(t) &= H_1(t)\, dW(t) + G_1(t)\, dt \\ dY(t) &= H_2(t)\, dW(t) + G_2(t)\, dt \end{aligned}$$

Write $\boldsymbol{X}_t = (X_t, Y_t)^T$, then

$$d\langle \boldsymbol{X} \rangle_t = \begin{pmatrix} H_1^2(t)\,dt & H_1(t)H_2(t)\,dt \\ H_1(t)H_2(t)\,dt & H_1^2(t)\,dt \end{pmatrix}$$

and the formula $f(t, x, y)$ gives

$$\begin{aligned} df(t, X_t, Y_t) = {} & f'_x(t, X_t, Y_t) dX(t) + f'_y(t, X_t, Y_t) dY(t) \\ & + \frac{1}{2} f''_{xx}(t, X_t, Y_t)\, H_1^2(t)\, dt + \frac{1}{2} f''_{yy}(t, X_t, Y_t)\, H_2^2(t)\, dt + f''_{xy}(t, X_t, Y_t)\, H_1(t)H_2(t)\, dt \\ & + f'_t(t, X_t, Y_t)\, dt. \end{aligned}$$

3. **Integration by parts formula for two Itô processes respect to the same Brownian motion.** An interesting particular case of previous formula is when $f(t, x, y) = xy$. Then

$$\boxed{X_t Y_t = X_0 Y_0 + \int_0^t Y_s\, dX_s + \int_0^t X_s\, dY_s + \int_0^t H_1(s) H_2(s)\, ds,}$$

which is called *integration by parts formula* .

11.6 Exercises

1. Write in the form

$$dX_t = u(t)\, dW_t + v(t)\, dt, \quad X_0 = x_0$$

the differential of the following processes:

(a) $X_t = W_t^2$.

(b) $X_t = 2 + t + e^{W_t}$.

(c) $X_t = \exp\{W_t - t/2\}$.

(d) $X_t = \exp\{2W_t - t\}$.

2. Assume that the US Dollar/Euro exchange rate is giving by a geometric Brownian process with differential

$$dr_t = \mu r_t\, dt + \sigma r_t\, dW_t.$$

Denote by $r'_t = 1/r_t$ the Euro/Dollar US exchange rate. Compute the differential expression for r'_t.

3. An asset follows a geometric Brownian motion with differential equation

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t,$$

and initial condition $S_0 = s_0 > 0$, where s_0 is a constant. Give the differential expression of $U_t = S_t^2$. Deduce $\mathbf{E}[S_t^2]$. *Indication:* It may be useful the following formula. Consider a linear integral equation (all functions are deterministic and enjoy good properties); the functions $a(t)$ and $b(t)$ are known:

$$h(t) = h(0) + \int_0^t \big(a(s)h(s) + b(s)\big)\, ds.$$

Its solution is

$$h(t) = e^{\int_0^t a(s)\, ds}\Big[h(0) + \int_0^t b(s) e^{-\int_0^s a(u)\, du} ds\Big].$$

4. Fix a number $a > 0$, and define

$$X_t = \big(a^{1/3} + \frac{1}{3} W_t\big)^3.$$

Check that

$$dX_t = X_t^{2/3} dW_t + \frac{1}{3} X_t^{1/3} dt, \quad X_0 = a.$$

5. Let $f(t)$ and $g(t)$ two *good* ordinary functions (for example, continuous functions), and

$$X_t = x_0 \exp\Big\{ \int_0^t f(s)\, dW_s + \int_0^t (g(s) - f^2(s)/2)\, ds\Big\}.$$

Check that

$$dX_t = f(t) X_t\, dW_t + g(t) X_t\, dt.$$

6. In the following cases, represent the random variable F in the form

$$F = a + \int_0^T f(t)\, dW_t,$$

where a is a number that should be computed.

(a) $F = W_T$.

(b) $F = \int_0^T W_t\, dt$. (*Indication:* Use the integration by parts formula of previous chapter (page 120) and $TW_T = T \int_0^T dW_t$.)

(c) $F = W_T^3$

7. Let

$$X_t = e^{at} \cos(3W_t),$$

where a is a number. Write X as an Itô process and compute the value of a such that X is a martingale (that is, the integral with respect to dt disappears).

8. Let

$$X_t = x_0 e^{-ct} + \sigma e^{-ct} \int_0^t e^{cs} dW_s,$$

where $\sigma, c > 0$ are constants. Check that

$$X_t = x_0 - c \int_0^t X_s\, ds + \sigma W_t.$$

Compute $\mathbf{E}[X_t]$ and $\mathbf{E}[X_t^2]$. *Indication:* Consider the Itô process $Y_t := e^{ct} X_t$, and deduce the expression of $e^{-ct} Y_t$.

9. Let
$$X_t = a + \int_0^t e^{2W_s}\, dW_s.$$
Obtain an integral expression of X_t^2 and compute $E[X_t^2]$.

10. Apply the Itô formula to compute
$$\mathsf{E}\Big[\Big(\int_0^t s\, dW_s\Big)^4\Big].$$
Indication: Consider the Itô process $X_t = \int_0^t s\, dW_s$.

11. Use the Itô formula (11.5) to prove the property given in Exercise 3 in Chapter 10

Chapter 12

Stochastic differential equations

12.1 Integral equations

The simplest ordinary differential equations are the ones of order 1:

$$\begin{aligned} x'(t) &= a(t, x(t)) \\ x(0) &= x_0 \end{aligned} \tag{12.1}$$

where $a(t, x)$ is a known function and x_0 is a given number. For example, take $a(t, x) = 5x$ and $x_0 = 2$. The solution of

$$\begin{aligned} x'(t) &= 5x(t) \\ x(0) &= 2 \end{aligned}$$

is $x(t) = 2e^{5t}$. The equation (12.1) can be written

$$\frac{dx(t)}{dt} = a(t, x(t))$$

and hence, we also put

$$dx(t) = a(t, x(t))\, dt.$$

Alternatively, equation (12.1) can be written as an integral equation

$$x(t) = x_0 + \int_0^t a(s, x(s))\, ds.$$

12.2 Stochastic differential equations

In mathematical modelling there is sometimes a differential equation in the form

$$dx(t) = a(t, x(t))\, dt$$

that is necessary to perturb by a random noise; the most convenient way to to this is to consider an expression

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t, \tag{12.2}$$

where $\{W_t,\, t \geq 0\}$ is a Brownian motion. Intuitively, the behaviour of X_t on $[t,\, dt]$ depens on $a(t, X_t)\, dt$ and a random increment $b(t, X_t)\,(W_{t+dt} - W_t)$, where $(W_{t+dt} - W_t)$ is Gaussian $\mathcal{N}(0, dt)$. However, which is the right interpretation of dW_t appearing in (12.2)? The answer is that we want an stochastic process $\{X_t,\, t \in [0,T]\}$ such that

$$X_t = Z + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s)\, dW_s, \tag{12.3}$$

where Z, the initial condition, is a random variable (or a number), $a(s,x)$ is called the drift coefficient and $b(t,x)$ the diffusion term. Usually, the equation is written in differential form

$$\begin{aligned} dX_t &= a(t, X_t)\, dt + b(t, X_t)\, dW_t \\ X_0 &= Z \end{aligned}$$

There is the following result of existence and unicity of solutions: Assume that

1. Z is independent of the Brownian motion and $\mathbf{E}[Z^2] < \infty$.

2. There is a constant C such that

$$|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| < C|x-y|,\ \forall x, y \in \mathbb{R},\ \forall t \in [0,T].$$

It is said that $b(t,x)$ and $a(t,x)$ are uniformly Lipschitz in $[0,T]$.

3. There is a constant C' such that

$$|a(t,x)| + |b(t,x)| < C'(1+|x|),\ \forall x \in \mathbb{R},\ \forall t \in [0,T].$$

It is said that $a(t,x)$ and $b(t,x)$ have linear growth in x.

Then there is one and only one stochastic process $X = \{X_t,\ t \in [0,T]\}$ that satisfies the equation (12.3).

There are mainly three types of stochastic differential equations:

(i) Equations that have an explicit solution. Only the simplest equations as the linear equations belong to this class.

(ii) Equations that it is not known the explicit solution, but it is known the probability distribution of the solution X_t. Very few, but important. An example of such equation is the Cox–Ingersol-Ross model for the interest rates

(iii) Equations that can be solved numerically.

12.3 Some examples of solvable equations

12.3.1 Black-Scholes equation

We saw in previous chapter that the differential of

$$S_t = \exp\{\sigma W_t + (\mu - \sigma^2/2)t\}$$

is

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$$

Hence, given a number $s_0 > 0$, the stochastic differential equation

$$\begin{aligned} dX_t &= \mu X_t\, dt + \sigma X_t\, dW_t \\ X_0 &= s_0 \end{aligned}$$

has a solution given by $X_t = S_t$ Since $b(t,x) = \mu x$ and $a(t,x) = \sigma x$ satisfies the conditions for the unicity, that solutions is unique.

12.3.2 Vasicek model

Consider $c > 0$, $m \geq 0$, $\sigma > 0$ and $x_0 > 0$. The equation

$$\begin{aligned} dX_t &= c(m - X_t)\, dt + \sigma\, dW_t \\ X_0 &= x_0 \end{aligned}$$

has the unique solution

$$X_t = m + (x_0 - m)e^{-ct} + \sigma e^{-ct} \int_0^t e^{cs}\, dW_s.$$

This is proved applying Itô formula to X_t. That equation is used to model the instantaneous interest rate r_t. Here are some properties:

1. As we commented in the chapter devoted to the Itô integral (see the subsection Wiener integral), $\int_0^t e^{cs}\, dW_s$ is a centered Gaussian random variable with variance $\int_0^t e^{2cs}\, ds$. Hence, X_t is Gaussian with

$$\mathbf{E}[X_t] = m + (x_0 - m)e^{-ct}.$$

and

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{2c}(1 - e^{-2ct}).$$

2. Since X_t is Gaussian it can take negative values. Theoretically, this is a drawback. In practice, the parameters can be chosen in such a way that the probability that X_t achieve negative values is negligible.

3. As a consequence of the point **1.**, when $t \to \infty$, X_t converges is distribution to a Gaussian random variable $\mathcal{N}(m, \frac{\sigma^2}{2c})$, independently of the initial value x_0.

It is said that X_t has a mean reversion.

When $m = 0$, the process X_t is called a Ornstein-Uhlenbeck process.

12.3.3 Linear stochastic differential equations

Both previous examples are linear equations. As in the non-stochastic case, the solution of a linear stochastic differential equations is known in closed form. A general linear differential equation has coefficients

$$a(t,x) = a_1(t) + a_2(t)x \quad \text{and} \quad b(t,x) = b_1(t) + b_2(t)x,$$

where $a_1(t), a_2(t), b_1(t), b_2(t)$ are continuous functions. These coefficients are uniformly Lipschitz and have linear growth. So the equation has one and only one solution. First consider the *homogeneous* linear equation

$$\begin{aligned} dY_t &= a_2(t)Y_t dt + b_2(t)Y_t\, dW_t \\ Y_0 &= 1 \end{aligned} \tag{12.4}$$

Applying Itô formula it is checked that the solution is

$$Y_t = \exp\Big\{ \int_0^t \big(a_2(s) - \frac{1}{2} b_2^2(t)\big)\, ds + \int_0^t b_2(s)\, dW_s \Big\}.$$

Now consider the inhomogeneuos equation

$$\begin{aligned} dX_t &= \big(a_1(t) + a_2(t)\big)X_t dt + \big(b_1(t) + b_2(t)\big)X_t\, dW_t \\ X_0 &= Z \end{aligned} \tag{12.5}$$

where $Z > 0$ is independent of the Brownian motion and satisfies $\mathbf{E}[Z^2] < \infty$. By the integration by parts formula of previous chapter it is proved that

$$X_t = Y_t \bigg(Z + \int_0^t Y_s^{-1} \big(a_1(s) - b_1(s) b_2(s)\big)\, ds + \int_0^t Y_s^{-1} b_1(s)\, dW_s \bigg)$$

is the solution of (12.5).

12.4 Stochastic differential equation with known probability distribution of the solution

Strange as it may seem, there are some stochastic differential equations that it is known that they have a unique solution and the probability distribution of the solution, but the solution has no explicit expression. A major example is the square-root process, also called Cox-Ingersoll-Ross model (CIR) for instantaneous interest rate. The equation is

$$\begin{aligned} dX_t &= \alpha\big(\beta - X_t\big)dt + \sigma\sqrt{X_t}\, dW_t \\ X_0 &= x_0 \end{aligned} \tag{12.6}$$

where $\alpha, \beta, \gamma, x > 0$. The coefficient $b(t,x) = \sqrt{x}$ is not Lipschitz, and hence the equation (12.6) does not fit in the standard results of stochastic differential equations. However, using advanced results, it can be proved that it has one and only one solution $X_t \geq 0$. Moreover, the probability distribution of X_t is known to be a non central χ^2 distribution.

12.5 Numerical solution

12.5.1 Euler scheme

Consider a stochastic differential equation

$$dX_t = a(t, X_t)dt + b(t, X_t)\, dW_t \tag{12.7}$$

$$X_0 = x_0 \tag{12.8}$$

for $t \in [0,T]$. Euler scheme is the translation to stochastic differential equations of the classical Euler method to solve numerically an ordinary differential equation. Consider a sequence $\{\mathcal{P}_n,\, n \geq 1\}$ of partitions of $[0,T]$: $\mathcal{P}_n$ is given by the points $0 = t_0^{(n)} < \cdots < t_{m_n}^{(n)} = T$. To simplify the notation, we suppress the superindex and write t_j instead $t_j^{(n)}$ Write

$$\Delta t_j = t_j - t_{j-1} \quad \text{and} \quad \Delta W_{t_j} = W_{t_j} - W_{t_{j-1}}.$$

Define recursively the values of the process at points $t_0, \ldots, t_{m_n}$

$$\begin{aligned} X_0^{(n)} &= x_0 \\ X_{t_1}^{(n)} &= X_0^{(n)} + a(t_1, X_0^{(n)})\Delta t_1 + b(t_1, X_0^{(n)})\Delta W_{t_1}, \end{aligned}$$

and so on, and interpolate linearly between to points $X_{t_i}^{(n)}$ and $X_{t_{i+1}}^{(n)}$. Under some hypothesis, it is proved that when the mesh of the partition goes to 0,

$$\lim_n \mathbf{E}\big[\big(X_t - X_t^{(n)}\big)^2\big] = 0,$$

uniformly in $t \in [0,T]$.

12.5.2 Milstein scheme

Consider again equation (12.7), now with the integral expression, and compute the difference $X_{t_j} - X_{t_{j-1}}$:

$$X_{t_j} - X_{t_{j-1}} = \int_{t_{j-1}}^{t_j} a(s, X_s)\, ds + \int_{t_{j-1}}^{t_j} b(s, X_s)\, dW_s.$$

Assume that both a and b are smooth functions (they are two times differentiable with respect to x variable, etc.). Then we can apply Itô formula to $a(s, X_s)$ and $b(s, X_s)$ and after some computions we can construct the Milstein approximation to X_t:

$$\begin{aligned} X_0^{(n)} &= x_0 \\ X_{t_1}^{(n)} &= X_0^{(n)} + a(t_1, X_0^{(n)})\Delta t_1 + b(t_1, X_0^{(n)})\Delta W_{t_1} + \frac{1}{2} b(t_1, X_0^{(n)}) \frac{\partial b}{\partial x}(t_1, X_0^{(n)})\big((\Delta W_{t_1})^2 - \Delta t_1\big) \end{aligned}$$

and so on. Again under some hypothesis on the coefficients a and b, and X_0 it can be proved that Milstein approximation converges to the solution X_t, and indeed, that approximation produces better results that Euler's one.

For an overview of that point we refer to Mikosch [12], and for a complete study to Kloeden, P. E. and Platen [9].

12.6 Exercises

1. Consider the linear stochastic differential equation

$$dX_t = 2X_t\,dt + 1.5X_t\,dW_t,\ X_0 = 1.$$

Write its solution. Now we will restrict to the time interval $[0,1]$. Generate on that interval a Brownian path with mesh $\delta = 0.05$. Using that path, plot a trajectory of the solution of the solution of the stochastic differential equation.

2. Using the Brownian path generated in previous exercise, plot an Euler approximation of the equation, with mesh 1/10. Repeat with mesh 1/100. Compare with the plot obtained with the exact solution obtained in previous exercise.

3. Repeat the previous exercise with a Milstein scheme.

Chapter 13

Black–Scholes model

13.1 The model

We consider a market in continuous time $t \in [0, T]$. There is a riskless asset B_t that starts with $B_0 = 1$ € and is continuously compounding at instantaneous interest rate $r \geq 0$,

$$B_t = e^{rt};$$

in differential notation,

$$dB_t = re^{rt}\,dt.$$

Equivalently,

$$\boxed{\frac{dB_t}{B_t} = r\,dt, \quad B_0 = 1.}$$

There is also a risky asset with price S_t at time t given by the stochastic differential equation

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t, \tag{13.1}$$

where W_t is a Brownian motion. Equivalently,

$$\boxed{\frac{dS_t}{S_t} = \mu\,dt + \sigma\,dW_t, \quad S(0) = S_0.}$$

As we saw in the previous chapter, we have that

$$\boxed{S_t = S_0 \exp\{(\mu - \sigma^2/2)t + \sigma W_t\}.} \tag{13.2}$$

We take S_0 to be a constant. The filtration is the one generated by the Brownian motion $\mathcal{F}_t = \sigma(W_s,\ s \in [0, t])$.

Properties

1. From the relation

$$\mathbf{E}[e^{aZ}] = e^{a^2/2},$$

where $Z \sim \mathcal{N}(0, 1)$, it follows that

$$\mathbf{E}[S_t] = S_0 e^{\mu t}.$$

So the mean of S_t is given by the exponential curve $S_0 \exp(\mu t)$.

2. S_t has continuous trajectories.

3. For $0 \leq s \leq t$,
$$\frac{S_t}{S_s} = \exp\{(t-s)(\mu - \sigma^2/2) + \sigma(W_t - W_s)\}, \tag{13.3}$$
is independent of $\mathcal{F}_s$. From the Markov property of the Brownian motion, it follows that S_t is a Markov process.

4. From the previous point, it is deduced that the relative increments
$$\frac{S_t - S_s}{S_s} = \frac{S_t}{S_s} - 1$$
over non overlapping intervals are independent and stationary:
$$\frac{S_t - S_s}{S_s} \underset{\text{Law}}{=} \frac{S_{t-s} - S_0}{S_0}$$

5.
$$\log \frac{S_t}{S_s} \underset{\text{Law}}{=} (t-s)(\mu - \sigma^2/2) + \sigma\sqrt{t-s}Z,$$
where $Z \sim \mathcal{N}(0,1)$. Hence,
$$\log \frac{S_t}{S_s} \underset{\text{Law}}{=} \mathcal{N}\big((t-s)(\mu - \sigma^2/2), \sigma^2(t-s)\big).$$
It is said that S_t/S_s has a log-normal distribution.

13.2 Risk neutral probability

Changes of probability for continuous random variables are more difficult that for discrete random variables, and for continuous time stochastic processes are still more difficult. We will use the following simplified version of a deep theorem due to Girsanov:

Girsanov Theorem. Fix $T > 0$. Let $W = \{W_t,\ t \in [0,T]\}$ be a Brownian motion on the probability space $(\Omega.\mathcal{F}, \mathbf{P})$, and fix a number $\lambda \in \mathbb{R}$. There is one and only one probability $\mathbf{Q}$ equivalent to $\mathbf{P}$ such that
$$W_t^* := \lambda t + W_t,\ t \in [0,T]$$
is a $\mathbf{Q}$-Brownian motion. Moreover, if X is a positive or $\mathbf{P}$–integrable random variable, $\mathcal{F}_T$ measurable,
$$\mathbf{E}_{\mathbf{Q}}[X] = \mathbf{E}_{\mathbf{P}}\big[X\ \exp\{-\lambda W_T - \lambda^2 T/2\}\big].$$

$* * *$

Consider the discounted price process,
$$\widetilde{S}_t = e^{-rt} S_t.$$

Then, by the formula of (13.2) S_t applied to μ changed by $\mu - r$ (or by Itô formula), and by (13.1) we have that

$$d\widetilde{S}_t = (\mu - r)\widetilde{S}_t\,dt + \sigma\widetilde{S}_t\,dW_t. \tag{13.4}$$

Since we want $\widetilde{S}_t$ to be a martingale, we need that the term $(\mu - r)\widetilde{S}_t\,dt$ disappears. Then we change the probability to another one $\mathbf{P}^*$ in such a way that

$$W_t^* = \frac{\mu - r}{\sigma}\,t + W_t$$

is a $\mathbf{P}^*$-Brownian motion; that is, we take $\lambda = (\mu - r)/\sigma$ in Girsanov Theorem. Changing W_t by $W_t^* - (\mu - r)/\sigma$ in (13.4),

$$d\widetilde{S}_t = \sigma\widetilde{S}_t\,dW_t^*.$$

It follows that

$$\widetilde{S}_t = S_0 + \sigma\int_0^t \widetilde{S}_t\,dW_t^*$$

is a $\mathbf{P}^*$-martingale.

Equivalently, note that

$$\widetilde{S}_t = S_0\exp\{-\frac{1}{2}\,\sigma^2 t + \sigma W_t^*\},$$

which is a $\mathbf{P}^*$ martingale by Example 3.1 in page 121.

13.3 Continuous time portfolio

A continuous time portfolio with the riskless asset and the risky asset is a two-dimensional stochastic process

$$\mathbf{\Phi} = \{\mathbf{\Phi}_t = (D_t, H_t),\ t \in [0, T]\},$$

where D_t is the quantity of riskless asset and H_t the quantity of risky asset at time t. It is assumed that both are continuous adapted processes, such that

$$\int_0^T |D_t|\,dt < \infty,\ \text{a.s.} \quad \text{and} \quad \int_0^T H_t^2\,dt < \infty,\ \text{a.s.}$$

The value at time t of the portfolio is

$$V_t = D_t e^{rt} + H_t S_t.$$

If $V_t \geq 0$ for all t, we say that the portfolio is **admissible**. Also we say that the portfolio is **self-financing** if for every $t \in [0, T]$,

$$V_t = V_0 + \int_0^t D_s\,dB_s + \int_0^t H_s\,dS_s, \tag{13.5}$$

which is exactly the continuous time version of the non-consumption in discrete time (see (6.2) in Chapter 6). Developing a bit (13.5) is it clear that V_t is an Itô process. From Itô formula, the discounted value of the portfolio

$$\widetilde{V}_t = e^{-rt}V_t$$

satisfies

$$\widetilde{V}_t = V_0 + \int_0^t H_s d\widetilde{S}_s = V_0 + \sigma \int_0^t H_s \widetilde{S}_s \, dW_s^*.$$

(Note that the first equality of this expression is the continuous time equivalent to formula (6.3) in Chapter 6). Hence, under $\mathbf{P}^*$, $\widetilde{V}_t$ is a martingale.

We will accept the following fundamental result, whose proof needs advanced results in stochastic calculus. Anyway, in the next section we will prove that result for a random variable of the form $X = f(S_T)$.

Theorem. The Black-Scholes model is complete: under the probability $\mathbf{P}^*$ for every positive random variable X $\mathcal{F}_T$ measurable there is an admissible self-financing portfolio $\mathbf{\Phi}$ such that it value a time T is

$$V_T = X.$$

* * *

As in the discrete case, the value at time t of an European derivative with payoff X is given by the value V_t of the replicating portfolio at that time. Since $\{\widetilde{V}_t, \ t \in [0,T]\}$ is a $\mathbf{P}^*$–martingale, and $V_T = X$, we have the valuation formula

$$\boxed{V_t = e^{-r(T-t)} \mathbf{E}_{\mathbf{P}^*}[X \mid \mathcal{F}_t].} \tag{13.6}$$

13.4 Black-Scholes formula for a derivative with payoff $f(S_T)$

We are going to give explicitly the replicating portfolio of a derivative with payoff $X = f(S_T)$, that is, we compute $\mathbf{\Phi}_t = (D_t, H_t)$ such that

$$\boxed{V_T = X = f(S_T)} \tag{13.7}$$

We will use the valuation formula (13.6) and the fact that

$$\boxed{V_t = D_t e^{rt} + H_t S_t.} \tag{13.8}$$

Step 1. From (13.7), (13.6) and the Markov property

$$V_t = e^{-r(T-t)} \, \mathbf{E}_{\mathbf{P}^*}[f(S_T) \mid \mathcal{F}_t] = e^{-r(T-t)} \, \mathbf{E}_{\mathbf{P}^*}[f(S_t \, \frac{S_T}{S_t}) \mid S_t].$$

Hence, due to the fact that S_T/S_t is independent of $\mathcal{F}_t$, and property (3.3) of conditional expectation (Chapter 3), we have

$$\boxed{V_t = F(t, S_t),}$$

where

$$\boxed{F(t,x) = e^{-r(T-t)} \mathbf{E}_{\mathbf{P}^*} \big[f(x \, \frac{S_T}{S_t}) \big].}$$

It is typical to write S instead of x in such expressions; and changing S_T/S_t by its expression, we arrive at

$$\boxed{F(t,S) = e^{-r(T-t)} \mathbf{E}_{\mathbf{P}^*} \big[f(S e^{(T-t)(r-\sigma^2/2)+\sigma(W_T^* - W_t^*)}) \big].} \tag{13.9}$$

Now observe that

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t^*, \tag{13.10}$$

so it is an Itô process. Assuming f of class $\mathcal{C}^2$, we can apply Itô formula to $F(t, S_t)$ and we get

$$dV_t = \sigma \frac{\partial F(t,S_t)}{\partial S} S_t\,dW_t^* + r\frac{\partial F(t,S_t)}{\partial S} S_t\,dt + \frac{\partial F(t,S_t)}{\partial t}\,dt + \frac{1}{2}\sigma \frac{\partial^2 F(t,S_t)}{\partial^2 S} S_t^2\,dt \tag{13.11}$$

Step 2. Since the portfolio is self-financing, by (13.5),

$$dV_t = rD_t e^{rt}\,dt + H_t\,dS_t,$$

that joining with (13.8) and (13.10)

$$dV_t = r(V_t - H_tS_t)dt + H_t\,dS_t = rV_t\,dt - rH_tS_t\,dt + rH_tS_t\,dt + \sigma H_tS_t\,dW_t^* = rV_t\,dt + \sigma H_tS_t\,dW_t^*.$$

Now, using that $V_t = F(t, S_t)$ we have the alternative expression for dV_t

$$dV_t = rF(t,S_t)\,dt + \sigma H_tS_t\,dW_t^*. \tag{13.12}$$

Comparing (13.11) and (13.12) we deduce

$$\begin{aligned} H_t &= \frac{\partial F(t,S_t)}{\partial S} \\ rF(t,S_t) &= rS_t\,\frac{\partial F(t,S_t)}{\partial S} + \frac{\partial F(t,S_t)}{\partial t} + \frac{1}{2}\sigma S_t^2\,\frac{\partial^2 F(t,S_t)}{\partial^2 S} \end{aligned}$$

The second equality is true for every number $S_t(\omega)$, and since S_t has a log-normal distribution, this implies that is true for every positive number. Writing S instead S_t in the second equation, and adding the terminal condition, we get the celebrated Black–Scholes *pde* for $F(t,S)$:

$$\boxed{\begin{aligned} &\frac{\partial F}{\partial t} + rS\,\frac{\partial F}{\partial S} + \frac{1}{2}\sigma S^2\,\frac{\partial^2 F}{\partial^2 S} = rF, \\ &F(T,S) = f(S). \end{aligned}} \tag{13.13}$$

The hedging is given by

$$\boxed{\begin{aligned} H_t &= \frac{\partial F(t,S_t)}{\partial S} \\ D_t &= e^{-rt}\big(F(t,S_t) - H_tS_t\big) \end{aligned}} \tag{13.14}$$

13.5 Pricing and hedging an European Call

The particular case of an European Call is of utmost importance. The payoff is $f(S_T) = (S_T - K)^+$, and we write $C(t,S) := F(t,S)$. Then, from (13.9), changing $W_T^* - W_t^*$ by $\sqrt{T-t}\,Z$, where $Z \sim \mathcal{N}(0,1)$,

$$\begin{aligned} C(t,S) &= e^{-r(T-t)}\mathbf{E}\Big[\big(Se^{(T-t)(r-\sigma^2/2)+\sigma\sqrt{T-t}\,Z} - K\big)^+\Big] \\ &= \frac{1}{\sqrt{2\pi}}e^{-r(T-t)}\int_{-\infty}^{\infty}\big(Se^{(T-t)(r-\sigma^2/2)+\sigma\sqrt{T-t}\,z} - K\big)^+ e^{-z^2/2}\,dz \end{aligned}$$

Since

$$\left(Se^{(T-t)(r-\sigma^2/2)+\sigma\sqrt{T-t}\,z}-K\right)^+=0$$

unless

$$Se^{(T-t)(r-\sigma^2/2)+\sigma\sqrt{T-t}\,z}>K,$$

and this inequality is equivalent to

$$z>\frac{\log(K/S)-(r-\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

or writing

$$\boxed{d_-:=\frac{\log(S/K)+(r-\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},}$$

to the inequality

$$z>-d_-.$$

Hence,

$$C(t,S)=\underbrace{\frac{1}{\sqrt{2\pi}}e^{-r(T-t)}\int_{-d_-}^{\infty}Se^{(T-t)(r-\sigma^2/2)+\sigma\sqrt{T-t}\,z}e^{-z^2/2}\,dz}_{(*)}-\underbrace{\frac{1}{\sqrt{2\pi}}e^{-r(T-t)}K\int_{-d_-}^{\infty}e^{-z^2/2}\,dz}_{(**)}.$$

The term (**), given the symmetry of the standard Gaussian density, is

$$(**)=Ke^{-r(T-t)}\Phi(d_-),$$

where

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-z^2/2}\,dz$$

is the cumulative probability distribution function of a $\mathcal{N}(0,1)$ random variable. The factor (*) is

$$(*)=S\frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty}e^{-(T-t)\sigma^2/2+\sigma\sqrt{T-t}\,z-z^2/2}\,dz=S\frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty}e^{-\frac{1}{2}(z-\sigma\sqrt{T-t})^2}\,dz=S\Phi(d_+),$$

where

$$\boxed{d_+=\frac{\log(S/K)+(r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}=d_-+\sigma\sqrt{T-t}.}$$

Then

$$\boxed{C(t,S)=S\Phi(d_+)-Ke^{-r(T-t)}\Phi(d_-).}$$

By (13.14), the value of H_t in the hedging portfolio is $\partial C(t,S)/\partial S$. That partial derivative is not totally trivial, since d_- and d_+ depend on S. Doing the computations we get

$$\boxed{H_t=\Phi(d_+)}.$$

This in a number between 0 and 1, and give the proportion of one asset that is necessary to have in the portfolio to hedge the call.

That results guarantee the exact hedging of an European Call option when the seller does a continuous hedging; of course, in practice that is impossible for two reasons: first, because it is physically impossible to continuously rebalance the portfolio: each change needs some time. The second reason is that we had not accounting the transaction costs; that simplifies the theory and allows for approximate but good formulas; however, a continuous hedging will imply infinite cost, and that cannot be ignored. How often the portfolio should be rebalanced? There is no absolute answer to that important question: it depends of the importance of the portfolio, the volatility, and many other matters.

13.5.1 An example

First, all the quantities in Black-Scholes formula need to be measured in coherent units. Normally, time is measured in years and hence r is a yearly interest rate. Also the volatility is measured as yearly percent and a volatility $\sigma = 0.2$ means $20\,\%$. We will price an European Call of an asset that today has spot $S_0 = 29$ and that matures is 4 weeks from now, with strike price $K = 30$. We take $r = 0.04$ and $\sigma = 0.2$. We will also compute the hedging portfolio assuming that the seller of the Call hedges once a week. There are some conventions on how to count the dates (trading days or natural days, etc). We will use the simplest version, and today is $t = 0$, and the maturation date is $4/52$. Please, check the following computations.

Week	t (years)	S_t	H_t	D_t	B_t	C_t
0	0	29	0.299	-8.368	1	0.296
1	0.019	28	0.086	-2.361	1.0008	0.052
2	0.038	30	0.523	-15.188	1.0015	0.494
3	0.058	28	0.007	-0.20	1.0023	0.002
4	0.077	27	0	0	1.0031	0

Here the hedging in the week four it is assumed to be done an infinitesimal time before the maturation time.

Now we do the hedging for an evolution of the asset price such that the Call finish in the money

Week	t (years)	S_t	H_t	D_t	B_t	C_t
0	0	29	0.299	-8.368	1	0.296
1	0.019	28	0.086	-2.361	1.0008	0.052
2	0.038	30	0.523	-15.188	1.0015	0.494
3	0.058	32	0.991	-29.620	1.0023	2.026
4	0.077	34	1	-29.908	1.0031	4

It is interesting to remark that if the option finish out-the-money, $S_T(\omega) < K$, as in the first table, then the seller of the option has to have nothing in her portfolio. That is due to the fact that in that case, by the continuity of the trajectories of S_t, we have $S_t(\omega) > K$ for t in a neighborhood of T. Hence, $\log(S_t(\omega)/K) > 0$ and

$$\lim_{t \nearrow T} d_+(t) = +\infty,$$

and thus

$$\lim_{t \nearrow T} H_t(\omega) = \lim_{t \nearrow T} \Phi(d_+(t)) = 0.$$

On the contrary, if the option finish in-the-money, $S_T(\omega) > K$, then

$$\lim_{t \nearrow T} H_t(\omega) = 1,$$

that means, the seller of the option needs to have the asset in order to cover the exercise of the option.

In the light of such results, one can understand the stress of the seller of a Call when the expiration date approaches and S_t is moving near the value of K.

13.5.2 The Greeks

The price of a call at a fixed time t, C, is a function of the variables S, r, σ, T and K, that is,

$$C = C(S, r, \sigma, T, K).$$

The study of the dependence of C on the values of such variables and parameters is called **sensibility analysis**, and it is mainly done through the derivatives (in the ordinary Calculus meaning) of C; such derivatives are called **the Greeks**.

The delta

The delta measures the variation of the price of the call in function of the price of the underlying, and it is defined as

$$\Delta = \frac{\partial C}{\partial S} = \Phi(d_+).$$

Since $\Phi(d_+) > 0$, it follows that the price of a Call is an increasing function of the price of the underlying. That means, with the same conditions of expiry time and strike, a Call over an asset with price, say $S = 25$ € is more expensive that over an asset with $S = 20$ €.

The Gamma

$$\Gamma := \frac{\partial \Delta}{\partial S} = \frac{\partial^2 C}{\partial S^2} = \frac{1}{S\sigma\sqrt{T-t}}\Phi'(d_+).$$

That quantity measures the velocity of the change of the price of the option in function of the price of the underlying. Since

$$\Phi'(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} > 0,$$

if follows that $\Gamma > 0$. This implies that C is a convex function of spot. Joining the information from the delta and the gamma we have the following remarkable property:

Property. The price of an European Call option is an increasing and convex function of the underlying. A typical shape of such function is given in Figure 13.1

The vega

$$\mathcal{V} = \frac{\partial C}{\partial \sigma} = S\sqrt{T}\Phi'(d_+) > 0.$$

This also implies that C increases with σ.

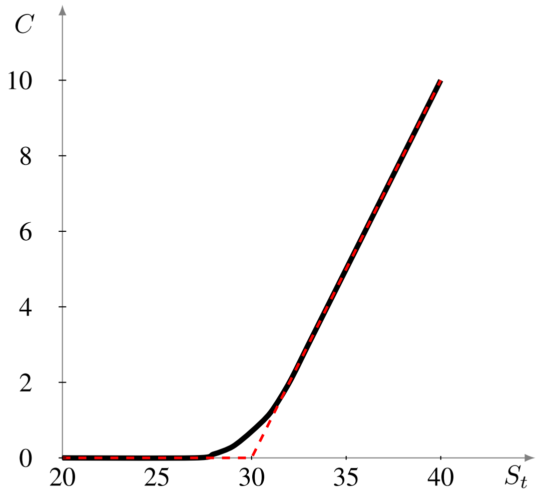


Figure 13.1. Solid line: Value of a Call before the expiring time: $C(S_t, t)$. Dashed line: Payoff of the Call. $K = 30$

The Lambda

$$\Lambda = \frac{\partial C}{\partial K} = -e^{-rT}\Phi(d_-) < 0.$$

Hence, the price of the option decreases as K increases.

The Theta

$$\Theta = \frac{\partial C}{\partial T} = S\sigma\Phi'(d_-)/(2\sqrt{T}) + Ke^{-rT}r\Phi(d_-).$$

It can be proved that $\Theta > 0$. So C increases with T.

The rho

$$\rho = \frac{\partial C}{\partial r} = TKe^{-rT}\Phi(d_-) > 0.$$

13.5.3 Implied volatility

In the formula of the price of an European Call, $C(S, t, T, K, \sigma, r)$, all quantities are directly known except the volatility σ. Moreover, today's market price of the Call is also known. As we see with the vega, the price C is strictly increasing in the volatility, and hence there is a bijection between prices and the volatilities, and we can deduce the value of σ that match that market price. We study an example: Since C depends only on $T - t$, we will take $t = 0$. Consider a Call with the following characteristics: $T = 0.25,\ K = 34,\ r = 0.05$. The value of the underlying today is $S = 30$, and the market price of the the Call is 0.14. Then we can solve (numerically)

$$C(30, 0, 0.25, 34, \sigma, 0.05) = 0.14,$$

and obtain that the volatility that gives that price is

$$\sigma = 0.18.$$

That computation can be done easily with the solver of Excel. That quantity is called **implied volatility** and it is very important. It is written

$$\boxed{\sigma_I = 0.18.}$$

On the other hand, for a liquid underlying, today may be in the market different options on that underlying with the same maduration date and different strike prices. For example, assume that for $T = 0.25$, there are the following options with the corresponding market prices (today's stock is $S = 30$):

$\boldsymbol{K}$	24	26	28	30	32	34	36	38
Market price	6.34	4.45	2.73	1.38	0.54	0.14	0.04	0.02

Now we compute the implied probabilities of that prices and they are

$\boldsymbol{K}$	24	26	28	30	32	34	36	38
$\boldsymbol{\sigma_I}$	0.25	0.23	0.21	0.20	0.19	0.18	0.185	0,21

A plot of this table is given in Figure 13.2. Such figure is called **the smile**. which is contradictory with Black–Scholes model. Explanations of that phenomenon are interesting and controversial, and we refer the reader to Joshi [7].

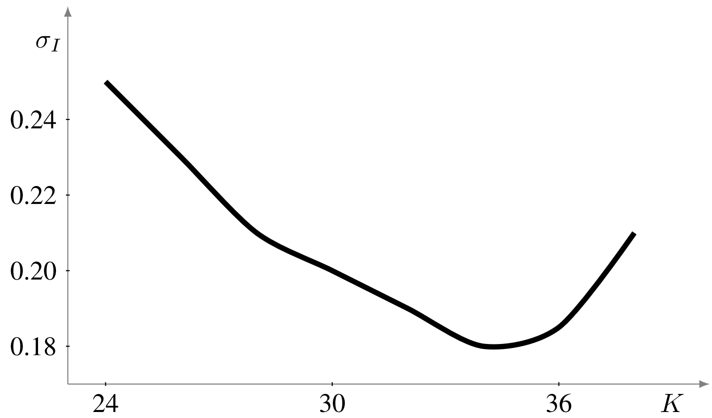


Figure 13.2. Volatility smile

13.6 Historical volatility

Volatility is not an observable quantity: we observe the prices and its behaviour, and from that we can estimate the volatility. As we saw at the beginning of this chapter, under Black-Scholes model,

$$\log \frac{S_{t+u}}{S_t} \sim \mathcal{N}\big((r - \sigma^2/2)u, \sigma^2 u\big).$$

Assume that we take observations

$$S_{t_1}, \dots, S_{t_{n+1}},$$

with a regular time step: $t_{i+1} - t_i = \tau$. Write $L_i = \log \frac{S_{t_{i+1}}}{S_{t_i}}$. Then $L_1, \dots, L_n$ are independent and identically distributed random variables $\mathcal{N}\big((r - \sigma^2/2)\tau, \sigma^2\tau\big)$. Denote by $\overline{L}$ the sample mean:

$$\overline{L} = \frac{1}{n}\sum_{j=1}^{n} L_j.$$

Then

$$\frac{1}{n-1}\sum_{j=1}^{n} (L_j - \overline{L})^2$$

is an unbiased estimator of $\sigma^2\tau$, and it is defined the **historical volatility** as

$$\sigma_H = \sqrt{\frac{1}{(n-1)\tau}\sum_{j=1}^{n} (L_j - \overline{L})^2}.$$

In practice, since

$$\overline{L} = \frac{1}{n}\sum_{j=1}^{n} \log \frac{S_{t_{i+1}}}{S_{t_i}} = \frac{1}{n}\big(\log(S_{t_{n+1}}) - \log(S_{t_1})\big)$$

and

$$\frac{1}{\tau(n-1)}\sum_j (L_j - \overline{L})^2 = \frac{1}{\tau(n-1)}\sum_j L_j^2 - \frac{n}{\tau(n-1)}(\overline{L})^2$$
$$\frac{1}{\tau(n-1)}\sum_j L_j^2 - \frac{n}{\tau(n-1)n^2}\big(\log(S_{t_{n+1}}) - \log(S_{t_1})\big)^2 \approx \frac{1}{\tau(n-1)}\sum_{j=1}^{n} L_j^2.$$

Then, the estimator used is

$$\boxed{\sigma_H = \sqrt{\frac{1}{(n-1)\tau}\sum_{j=1}^{n} \big(\log(S_{t_{i+1}}) - \log(S_{t_i})\big)^2}.}$$

Normally it is taken a period of 90 or 180 trading days: $\tau = 252$.

13.7 Put–call parity

In continuous time, put-call parity formula is the following: Denote by C_t (respectively P_t) the price of an European Call (resp. Put) of maturation time T and strike price K. Then

$$C_t + Ke^{-r(T-t)} = P_t + S_t.$$

13.8 From Cox–Ross–Rubinstein to Black–Scholes

In this section we are going to discretize the Black–Scholes model $\{S_t,\, t \in [0,T]\}$. Assume that the parameters of the model are T, r and σ (the mean μ does not matter, since we are going to work with the risk neutral probability). Take a natural number N (the number of discrete steps), and consider a discrete time unit T/N; define

$$r_N = \frac{T}{N} r$$

the discrete interest rate. Note that

$$\lim_{N\to\infty} (1 + r_N)^N = e^{rT}.$$

Also, with the notation of chapter 3, write

$$u_N = e^{\sigma\sqrt{T/N}} \quad \text{and} \quad d_N = e^{-\sigma\sqrt{T/N}}.$$

Note that $u_N = 1/d_N$, and this will simplify the computations. The risk neutral probability in this model is

$$p_N^* = \frac{r_N - d_N + 1}{u_N - d_N}.$$

Finally, consider the spot prices in this model,

$$S_0^{(N)}, \ldots, S_N^{(N)}.$$

It can be proved that

$$\lim_{N\to\infty} S_N^{(N)} = S_T, \text{ in distribution.}$$

It follows that if we denote by $C_{CRR}^{(N)}$ the price of an European derivative with expiry time $N = T$ and strike price K given by CRR formulas, and C_{BS} the price given by BS formulas,

$$\lim_{N\to\infty} C_{CRR}^{(N)} = C_{BS}.$$

Hence, the BS formula to price an European option can be approximated by the CRR formula.

What is more, it can be proved the convergence in distribution of the law of the CRR process to the BS process, and that implies that also the prices that depend on all the trajectory, as an American option, in BS model can be approximated by the corresponding price in CRR model. This is an important and very useful property.

13.9 Exercises

1. In a Black–Scholes model with $\mu = 5$, $\sigma = 0.2$, $S_0 = 40$, and $r = 0.05$, consider an European Call with maturation time 3 months and strike 50 €. Compute under the risk neutral probability the probability that the option finish in-the-money.

2. Compute the implied volatility of an asset with spot 40 €, if a Call with expiry time in one month with strike 39 € has a premium of 1.5 €. Use an interest rate of 4%.

3. Compute the weekly hedging of an European Call with $T = 30$ days, $r = 0.05$, $\sigma = 0.3$, $K = 30$, and spot given by

Day	0	7	14	21	28
S_t	35	32	39	37	28

4. Consider a digital option that pays 1 € at time T if $S_T > K$, and 0 otherwise. Prove that the price is

$$V_0 = e^{-rT}\Phi(d),$$

where

$$d = \frac{\ln\big(S_0/K\big) + \big(r - \sigma^2/2\big)T}{\sigma\sqrt{T}}.$$

5. In a Black–Scholes model consider a derivative with payoff S_T^3 at time T.

(a) Check that the value of that derivative at time t can be written as

$$V_t = g(t,T)S_t^3,$$

where $g(t,T)$ is a function that you should determine.

(b) Compute the price V_0 of that derivative assuming $T = 3, \sigma = 0.2, r = 0.05$ and $S_0 = 30$ €.

Note: Part (a) can be done in two ways: using the general valuation formulas (13.6) or (13.9), or solving Black-Scholes equation (13.13) with the suitable final condition. It is convenient that the reader works both ways.

Chapter 14

Models with jumps

14.1 Introduction

Abrupt changes in the evolution of the price of an asset are often observed, changes that are not compatible with a geometric Brownian motion. On the other hand, the so called *stylized empirical facts*, that is, some statistical properties of the financial data like the heavy tails of log returns, the volatility clusters, the excess of kurtosis, etc. are also against Black–Scholes model; see for example Cont and Tankov [3]. An extension relatively simple of Black–Scholes model is to add some jumps at random times. As Merton [11] explains, the prices will have a twofold dynamics: on one side, the "normal" one, that makes that the prices follow a geometric Brownian motion (Black-Scholes) and some "abnormals" changes that are traduced in jumps. Such model uses a process called ***jump-diffusion***, that unfortunately, induces an incomplete market: there are infinity neutral probabilities. Therefore, Merton chooses a particular neutral probability and a hedging that assumes that the risk due to jumps is *diversifiable*. Another solution for the hedging is to use other strategies, as the *minimal quadratic hedging*, that we will also study.

The Poisson process or the de jump-diffusion process are both de Lévy processes, and in the last years many other Lévy processes has been used to modeling financial data: Gamma process, inverse Gaussian, etc.; see Cont and Tankov [3] or Schoutens [16].

14.2 The Poisson process

After the Brownian motion, the most important process (in continuous time) is the Poisson process: the trajectories are stepwise (see Figure 14.1) with jumps of size 1 and the number of jumps before time t is given by a Poisson random variable of parameter proportional to t. We start studying the Poisson random variables.

14.2.1 Poisson random variables

A discrete random variable has a Poisson law of parameter $\alpha > 0$ if

$$P\{X = n\} = e^{-\alpha}\,\frac{\alpha^n}{n!},\; n = 0, 1, \dots$$

Its expectation and variance are α:

$$\mathbf{E}[X] = \mathrm{Var}(X) = \alpha.$$

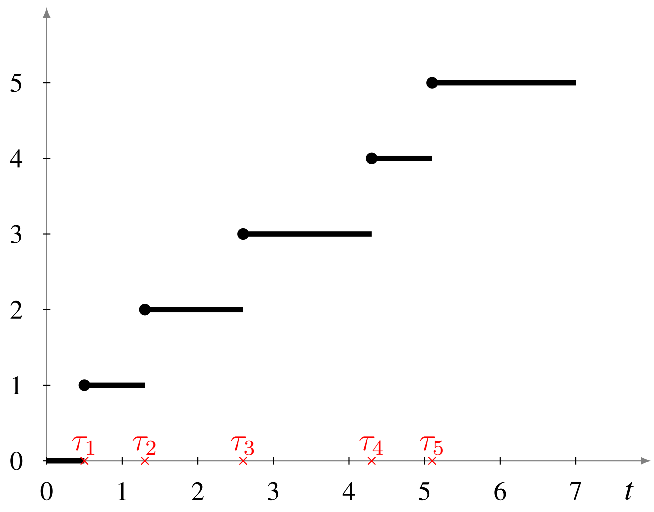


Figure 14.1. Trajectory of a Poisson process of intensity $\lambda = 1$

Moreover, it has exponential moment: for every $u \in \mathbb{R}$,

$$\mathbf{E}[e^{uX}] = \exp\left\{\alpha(e^u - 1)\right\}. \tag{14.1}$$

Exercise. (Expectation of the product of a random number of random variables). Let X a Poisson random variable of parameter α and $\{V_n,\, n \geq 1\}$ a sequence of i.i.d. random variables, independent of X, with expectation $E[V_j] = b$. Prove that

$$\mathbf{E}\Big[\prod_{j=1}^{X} V_j\Big] = \exp\{\alpha(b-1)\}. \tag{14.2}$$

14.2.2 The Poisson process

A Poisson process of intensity $\lambda > 0$ is a stochastic process $N = \{N_t,\, t \geq 0\}$ such that

1. $N_0 = 0$.

2. For $0 \leq s < t$, the random variable $N_t - N_s$ has a Poisson law of parameter $\lambda(t-s)$. Hence, the law of the increment of N between the time instants s and t only depends of the distance between both points. It is said that the process N has **stationary increments**.

3. For $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $N_{t_0}, N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent. It is said that N has **independent increments**.

Next property is important, and among other things, it gives an easy way to simulate a Poisson process without simulate a Poisson random variable (that is not so easy).

Denote by $0 < \tau_1 < \tau_2 < \cdots$ the instant of the jumps of a Poisson process of intensity λ (see Figure 14.1). Then $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ are i.i.d, each with exponential law of parameter λ, that is, $\tau_j - \tau_{j-1}$ has density

$$f(x) = \lambda e^{-\lambda x},\ x > 0.$$

The Poisson process is used as a model for the number of arrivals of people to a place; note that the mean number of jumps in $[0, t]$ is λt, so λ is the mean number of jumps by unity of time, and by that reason is called the intensity.

Two martingales related with the Poisson process

1. The process $N_t - \lambda t$ is a martingale. It is called the compensed Poisson process. See Figure 14.2 for a trajectory of that martingale.

2. From the property of independent increments of the Poisson process and (14.1) it follows that for any $u \in \mathbb{R}$, the process $\exp\{uN_t - \lambda t(e^u - 1)\}$ is a martingale. En particular, $2^{N_t} e^{-\lambda t}$ is a martingale.

Exercise. Prove that both processes above are martingales.

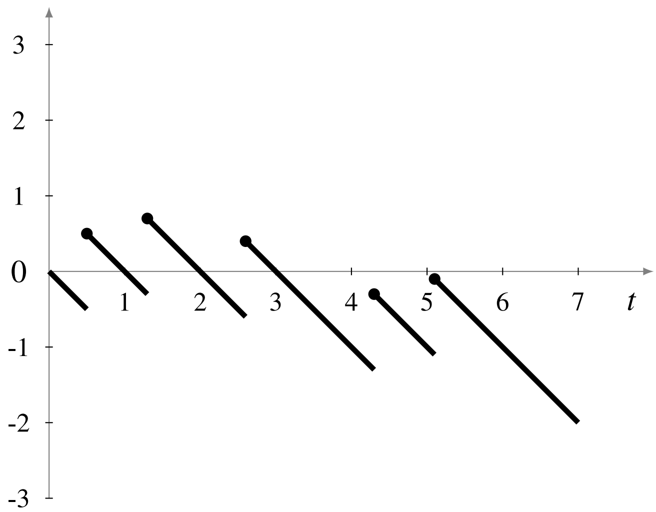


Figure 14.2. Trajectory of a the martingale $N_t - \lambda t$

14.2.3 A generalization: the compound Poisson process

As we commented, the jumps of the Poisson process have always size 1, and it would be more convenient to have more flexibility. With that purpose it is introduced the following process. Let $\{N_t,\, t \geq 0\}$ be a Poisson process of intensity $\lambda > 0$, and $\{A_n,\, n \geq 1\}$ be a sequence of i.i.d. random variables, independent of the Poisson process. Then define

$$Y_t = \sum_{j=1}^{N_t} A_j,$$

(with $Y_t = 0$ if $N_t = 0$). That process is called a **compound Poisson process**. The instant of jumps are given by N_t and the sizes by the random variables A_j.

14.3 The jump–diffusion and a model for the price of an asset

The next process combines a drift $\mu_0 t$, a Brownian part, and jumps given by a compound Poisson process; it is called **jump–diffusion process**:

$$X_t = \mu_0 t + \sigma W_t + \sum_{j=1}^{N_t} A_j,$$

where

- $W = \{W_t,\, t \geq 0\}$ is a Brownian motion.
- $N = \{N_t,\, t \geq 0\}$ is a Poisson process of intensity $\lambda > 0$.
- $A = \{A_n,\, n \geq 1\}$ is a sequence of i.i.d. random variables,
- W, N i A are independents.
- $\sigma \geq 0$ and $\mu_0 \in \mathbb{R}$.

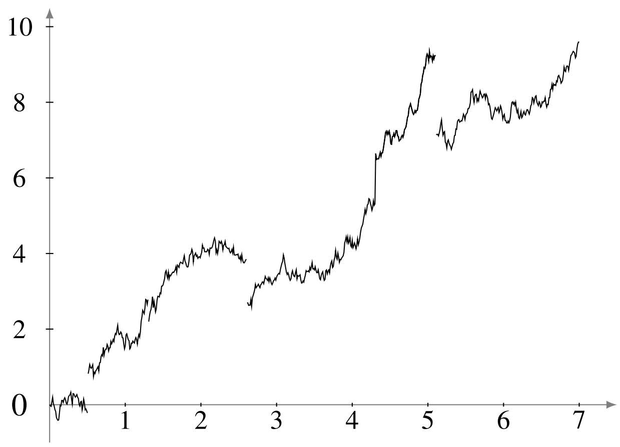


Figure 14.3. Trajectory of a jump–diffusion process

Now consider the price of an asset $S = \{S_t,\, t \in [0,T]\}$ such that its relative change follows a dynamics given by X_t:

$$\boxed{\frac{dS_t}{S_t} = dX_t,}$$

where we assume $A_j > -1$. That is, the behaviour of the relative change between two jumps is given by $\mu dt + \sigma dB_t$, and from time to time there us a jump. The solution of that stochastic differential equation is

$$\boxed{S_t = S_0 \, \exp\left\{(\mu_0 - \sigma^2/2)t + \sigma B_t\right\} \prod_{j=1}^{N_t} \big(A_j + 1\big).}$$

We observe that the condition $A_j > -1$ is necessary in order to have $S_t > 0$. Then the model for S_t is

$$\boxed{S_t = S_0 \exp\big\{(\mu_0 - \sigma^2/2)t + \sigma B_t + \sum_{j=1}^{N_t} \log(A_j + 1)\big\}.} \tag{14.3}$$

Merton, in his computations assumes that A_j+1 is lognormal with parameters $m,\ \epsilon^2$, that is, $A_j+1 = e^{R_j}$, where R_j is $\mathcal{N}(m, \epsilon^2)$.

14.3.1 Exponencial-Lévy models

The Brownian motion, the Poisson process, the jump–diffusion process are examples of Lévy processes, that are processes with independent and stationary increments. The unique Lévy process with continuous trajectories is the Brownian motion. The general theory of Lévy process was very developed from Paul Lévy, but they have no much interest for applications. However, the necessity of models with jumps in mathematical finance give rise to a golden age of Lévy processes.

In agreement with (14.3), S_0 is the exponential of a jump–difussion process. In general, a dynamics of prices such that

$$S_t = S_0 e^{L_t},$$

where L_t is a Lévy process, is called ***exponencial–Lévy*** model.

14.3.2 The jump-diffusion model is not complete

Let r be the interest rate. We are interesting in finding a probability $\mathbf{Q}$ equivalent to the initial one, $\mathbf{P}$ such that the discounted price of the asset,

$$\widetilde{S}_t = e^{-rt} S_t,$$

is a $\mathbf{Q}$-martingale. As we will see, there are infinite solutions. From Girsanov Theorem for jump–diffusion processes it is deduced that given any $\lambda' > 0$, there is a probability $\mathbf{Q}_{\lambda'}$ equivalent to $\mathbf{P}$ such that

- N_t is a $\mathbf{Q}_{\lambda'}$-Poisson process of intensity λ'.
- $B_t + \dfrac{1}{\sigma}(\mu_0 - r + \lambda' k)\, t$ is a $\mathbf{Q}_{\lambda'}$-Brownian motion.
- $\{A_n,\, n \geq 1\}$ are i.i.d. with $\mathbf{E}_{\mathbf{Q}_{\lambda'}}[A_j] = k$.

Then, if we write

$$W_t^* = W_t + \frac{1}{\sigma}(\mu_0 - r - \lambda' k)\, t,$$

we have

$$\widetilde{S}_t = S_0 \, \exp\big\{ - (\lambda' k + \sigma^2/2)t + \sigma W_t^* \big\} \prod_{j=1}^{N_t} \big(A_j + 1\big).$$

which is a $\mathbf{Q}_{\lambda'}$-martingale.

As a consequence,

$\boxed{\text{The market corresponding to a jump–diffusion model is not complete.}}$

The existence of an infinity of neutral probabilities is due to the fact that there are two sources of randomness: one from the Brownian motion, and the other one from the jumps. The Girsanov Theorem that we have used shows that it is possible to change the intensity of jumps in an arbitrary way and at the same time correct the drift in order to keep the martingale property.

Exercise. Prove that $\widetilde{S}_t$ is $\mathbf{Q}_{\lambda'}$-martingale. *Suggestion:* for $0 \leq s < t$, write

$$\widetilde{S}_t = \widetilde{S}_s \exp\big\{ -(\lambda' k + \sigma^2/2)(t-s) + \sigma(W_t^* - W_s^*\big\} \prod_{j=N_s+1}^{N_t} \big(A_j + 1\big),$$

and use the independence property and the exercise of page 154.

14.4 The Merton's solution

Merton [11] proposes to consider the probability $\mathbf{Q}_{\lambda'}$ such that the jumps part is not affected, that is, $\lambda' = \lambda$ and $k = E_P[U_j]$. Denote that probability by $\mathbf{Q}$, and write $\mu = r - \lambda k - \sigma^2/2$, and then S_t is

$$S_t = S_0 \exp\big\{\mu t + \sigma W_t\big\} \prod_{j=1}^{N_t} \big(A_j + 1\big).$$

Merton's argument is that with that election the risk due to the Brownian motion can be hedged using Black-Scholes ideas, and the jumps risk is diversifiable, that implies that a person with different portfolios, the risk of jumps are uncorrelated. Unfortunately, experience shows that this is not the case, and when there are jumps, mainly when they are negative, all portfolios are affected.

* * *

Merton's pricing is the following. Let $\mathcal{F}_t$ be the σ-generated by $\{S_s,\ s \in [0,t]\}$. An European option with payoff $g(S_T)$ has a value at time t

$$V_t = e^{-r(T-t)}\, \mathbf{E}_Q[g(S_T) \,|\, \mathcal{F}_t].$$

Since

$$S_T = S_t \exp\big\{\mu(T-t) + \sigma(W_T - W_t\big\} \prod_{j=N_t+1}^{N_T} \big(A_j + 1\big),$$

and given all the independence that we have,

$$V_t = F(t, S_t),$$

where

$$\begin{aligned} F(t,x) &= e^{-r(T-t)}\, \mathbf{E}_Q[g(S_T) \,|\, S_t = x] \\ &= e^{-r(T-t)}\, \mathbf{E}_Q\Big[g\Big(x \exp\big\{\mu(T-t) + \sigma(W_T - W_t\big\} \prod_{j=N_t+1}^{N_T} (A_j+1)\Big)\Big] \\ &= e^{-r(T-t)}\, \mathbf{E}_Q\Big[g\Big(x \exp\big\{\mu(T-t) + \sigma(W_{T-t}\big\} \prod_{j=1}^{N_{T-t}} (A_j+1)\Big)\Big] \end{aligned}$$

where the last equality is due to the stationary independent increments property of the Brownian motion and Poisson process. Change $A_j + 1$ by e^{R_j}, where $R_j \sim \mathcal{N}(m, \delta^2)$, and then

$$k = \mathbf{E}_Q[A_j] = \exp\{m + \delta^2/2\}.$$

Using that

$$\prod_{j=1}^{n} \big(A_j + 1\big) = \exp\{\sum_{j=1}^{n} R_j\},$$

and $\sum_{j=1}^{n} R_j \sim \mathcal{N}(nm, n\delta^2)$, we have

$$F(t,x) = e^{-r(T-t)} \sum_{n=0}^{\infty} e^{-\lambda(T-t)} \frac{\big(\lambda(T-t)\big)^n}{n!} \mathbf{E}_Q\Big[g\Big(x \, \exp\big\{\mu(T-t) + \sum_{j=1}^{n} R_j + \sigma W_{T-t}\big\}\Big)\Big]$$

Finally, since the R_j are independent of the Brownian motion,

$$\sum_{j=1}^{n} R_j + \sigma W_{T-t} \sim \mathcal{N}(0, n\delta^2 + \sigma^2(T-t)) + nm$$

and then

$$F(t,x) = e^{-r(T-t)} \sum_{n=0}^{\infty} e^{-\lambda(T-t)} \frac{\big(\lambda(T-t)\big)^n}{n!} F^{BS}(T-t, x_n, \sigma_n),$$

where

$$\begin{aligned}
\sigma_n^2 &= \sigma^2 + n\delta^2/(T-t), \\
x_n &= x \exp\{nm + \frac{n\delta^2}{2} - \lambda \exp(m + \frac{\delta^2}{2}) + \lambda(T-t)\}, \\
F^{BS}(t,x,\sigma) &= e^{-rt} \, \mathbf{E}\big[g\big(x \, \exp\big\{(r - \sigma^2/2)t + \sigma\sqrt{t}\, Z\big\}\big)\big],
\end{aligned}$$

with $Z \sim \mathcal{N}(0,1)$. Notice that $F^{BS}(t,x,\sigma)$ is the value of an European option with payoff $g(S_t)$ in a Black-Scholes with volatility σ.

The hedging proposed by Merton is a portfolio $\phi_t = (B_t, H_t)$ given by

$$H_t = \frac{\partial F}{\partial x}(t, S_{t-}) \quad \text{i} \quad H_t^{(0)} = H_t S_t - \int_0^t H_u \, dS_u,$$

that implies that we are hedging the risk due to the Brownian motion.

14.5 Minimal quadratic hedging

Consider a non complet market with an the price of an asset $\{S_t, \ t \in [0,T]\}$, and an arbitrary neutral probability $\mathbf{Q}$. Let X be the payoff of a derivative that we want to hedge. It is called **Minimal quadratic hedging** to a selffinancing portfolio $\phi = \{(B_t, H_t), \ t \in [0,T]\}$ that minimizes

$$\mathbf{E}_{\mathbf{Q}}\big[\big(V_T(\phi) - X\big)^2\big], \tag{14.4}$$

where $V_t(\phi)$ is the value of the portfolio,

$$V_t(\phi) = V_0 + r\int_0^t B_s ds + \int_0^t H_s\, dS_s.$$

In a Merton jump–diffusion model, for $X = g(S_T)$ that portfolio can be computed explicitly, and writing as before

$$F(t,x) = e^{-r(T-t)}\,\mathbf{E}_Q[g(S_T)\,|\,S_t = x],$$

we have

$$H_t = \Delta(t, S_{t-}),$$

where

$$\Delta(t,x) = \frac{1}{\sigma^2 + \lambda \mathbf{E}_Q[A^2]}\Big\{\sigma^2\frac{\partial F(t,x)}{\partial x} + \frac{\lambda}{x}\,\mathbf{E}_Q\Big[\Big(F\big(t,x(1+A)\big) - F(t,x)\Big)A\Big]\Big\}.$$

14.6 Exercise

Our objective is to simulate a Merton model.

1. Start simulating a Poisson process $\{N_t,\ t\in[0,6]\}$ with intensity $\lambda = 3$. You need the jump times $\tau_1, \tau_2, \ldots$, where $\tau_1, \tau_2 - \tau_2, \ldots, \tau_n - \tau_{n-1}, \ldots$ are i.i.d. exponentials of parameter 3.

(a) To simulate an exponential law of parameter λ it is used the inversion method. It s based on the fact that if a random variable X has continuous cumulative distribution function F, and U is uniform in $[0,1]$ (we write $U \sim \mathcal{U}[0,1]$), then $F^{-1}(U)$ has the same law as X, where

$$F^{-1}(u) = \inf\{x:\ F(x) = u\ \},\ u\in(0,1).$$

In the exponential case,

$$F(x) = 1 - \exp\{-\lambda x\},\ x > 0,$$

and then

$$F^{-1}(u) = -\frac{1}{\lambda}\,\log(1-u).$$

So, if $U \sim \mathcal{U}[0,1]$, then $-\frac{1}{\lambda}\,\log(1-U)$ is exponential of paramenter λ. What is more, using that $1-U$ is also uniform in $[0,1]$, then we can simplify the job:

$$U \sim \mathcal{U}[0,1] \quad \Longrightarrow \quad -\frac{1}{\lambda}\,\log U \sim \mathcal{E}xp(\lambda).$$

(b) Simulate $U_1,\ U_2,\ldots,$ uniform in [0,1], independent, and from here the jumping times. You should remember that the number of jumps in $[0,T]$ is random, so it is needed to introduce an instruction to stop the generation when $\tau_n > T$.

(c) Represent a trajectory of the Poisson process.

2. Now we want to simulate

$$S_t = S_0 \exp\big\{(\mu_0 - \sigma^2/2)t + \sigma B_t + \sum_{j=1}^{N_t}\log(A_j+1)\big\},$$

following Merton's model.

(a) Given that $A_j + 1 = e^{R_j}$, with $R_j \sim \mathcal{N}(m, \epsilon^2)$, the sum in the exponent is

$$\sum_{j=1}^{N_t} R_j,$$

so it is compound Poisson process with Gaussian jumps. Take $\mu = 1$ and $\epsilon = 2$ and simulate $R_1, R_2 \dots$.

(b) Simulate Brownian trajectory with mesh $\delta = 0.1$.

(c) Simulate a trajectory of $S_t,\ t \in [0, 6]$, with $\mu_0 = 0.05$, $\sigma = 0.2$, and $S_0 = 25$.

Chapter 15

Interest rate models

15.1 Introduction

In previous chapters we assumed that the interest rate was constant and known. Of course, in reality this is false: first, the instantaneous interest rate is a variable qantity r_t, and second, that number is not observable. In contrast with the price of a share or commodity, there is no official value for the interest rate; some things are similar, as the interest rate given by the European Central Bank, or the EURIBOR, but these numbers are reference indices to compute other indices (as the mortages, for example). The theory and practice of interest rate model is difficult, aggravate by the fact that there is no standard model as Black-Scholes is for assets. In this chapter we will study some models for r_t, and we compute some related quantities. However, this will be a very short and superficial introduction: interest rate model demands a whole course by itself.

15.2 The zero coupon bonds

The most popular fixed income product are the government Bonds issued by a national government. Usually (in the Eurozone) they have a *face value* of 10.000 €, and time of maturity of 6, 12 and 18 month, 3, 5, 10, 15 and 30 years; that is, the bond is bought at a price lower that 10.000 €, and it is repaid by the government at maturity by its face value. Some of them do periodic payments that are called *coupons*. There are many other kind of bonds: soverign bonds, corporative bonds, municipal bonds, but we do not study such casuistics. A **zero coupon bond**, or discounted bond, is an artificial product with face value 1 €, that does not gives coupons, and there are of that bonds for each maturation time every time between today and 30 years, and they mimics the behaviour of the real bonds. The prices of zero coupon bonds are deduced from the prices of the governament's bonds prices.

The curve (done today) with the prices of the zero coupon bonds with maturity each day between tomorrow and 30 years is called the ***prices curve*** of today.

The Figure 15.1 reproduces the price of the zero coupon bond corresponding to April 27, 2009, given by the European Central Bank.

If today (30/04/09) we have a bound that matures, say, on March, 20, 2015, after one year, the 30/5/10 we can sell that bond to someone that the 20/03/15 will receive 1 €. We denote that price by $P(30.04.10, 20.03.15)$.

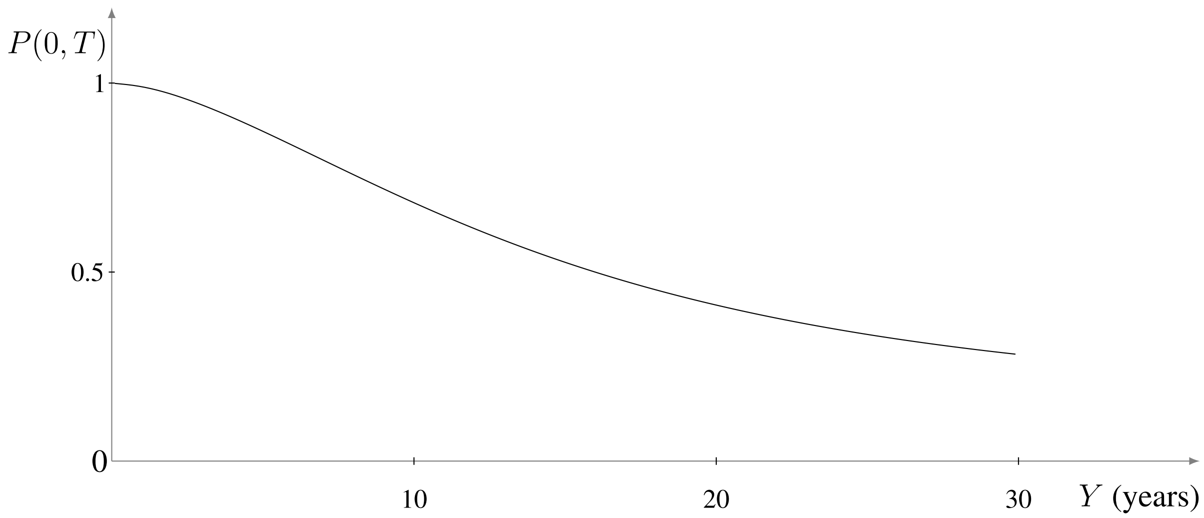


Figure 15.1. Price curve of day 27/04/09

$\boldsymbol{P(t,T)}$ is the value at time t of a zero coupon bond with maturity date T ($t \leq T$).

Hence, we have a family of curves (or a surface) $P(t,T)$ with indices t and T, with $t \leq T$. We have

$$0 < P(t,T) < 1, \text{ for } 0 \leq t < T, \quad \text{and} \quad P(T,T) = 1, \quad \forall T \in \mathbb{R}_+.$$

On the other hand, it is introduced the **Yield** $R(t,T)$ defined

$$R(t,T) = -\frac{\ln P(t,T)}{T-t}. \tag{15.1}$$

The rationale of that formula is that when it is inverted we have

$$P(t,T) = e^{-(T-t)R(t,T)}, \tag{15.2}$$

and then,

$$P(t,T)e^{(T-t)R(t,T)} = 1€,$$

that means, $R(t,T)$ is the rate of interest (instantaneous) that we should apply (continuously) during the period $[t,T]$ in order that $P(t,T)$ gives a quantity of 1 € at time T. For example, if the value of a zero coupon bond in one year is

$$P(30.04.09, 1.05.10) = 0'96 €,$$

then

$$R(30.04.09, 1.05.10) = 3'91\%.$$

For each $\boldsymbol{t}$ fixed, the curve $\boldsymbol{R(t,T),\ T \in (t, t+30]}$ is called the **Yield Curve** or **term structure of the interest rate** corresponding to $\boldsymbol{t}$.

See Figure 15.2 to see the yield curve corresponding to the day April, 27, 2009.

We stress that this curve is not referred to a bond, but to a family of bonds with different maturity date. These curves can have a number of different aspects, see the Figure 15.3 for the Yield curve on 15/09/08. That changes on the yield curve have different interpretations, as the expectative of people about economic or political changes. In Internet the reader can find many yield curves at its interpretations. There are specialists in this field.

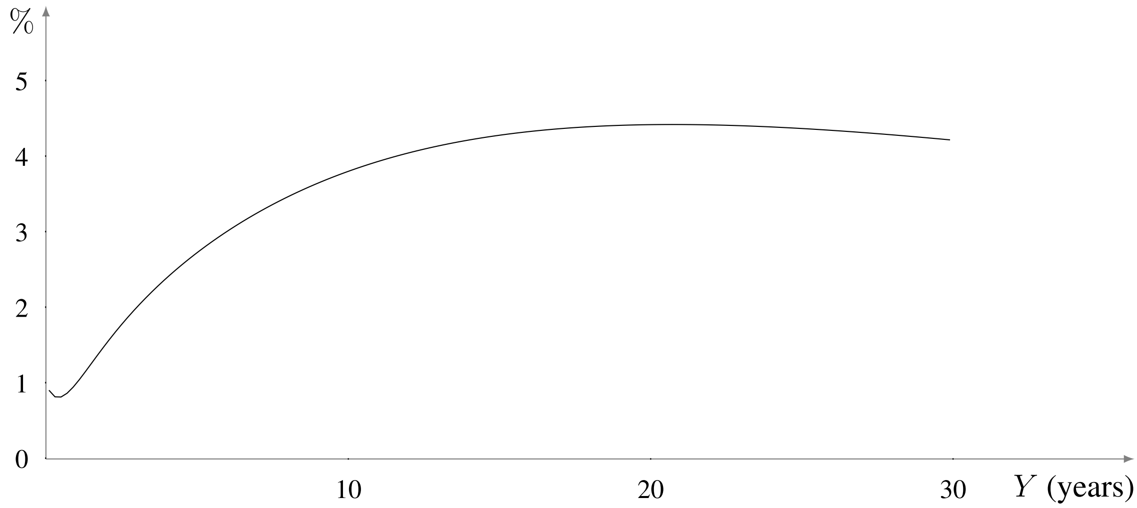


Figure 15.2. Yield curve of day 27/04/09

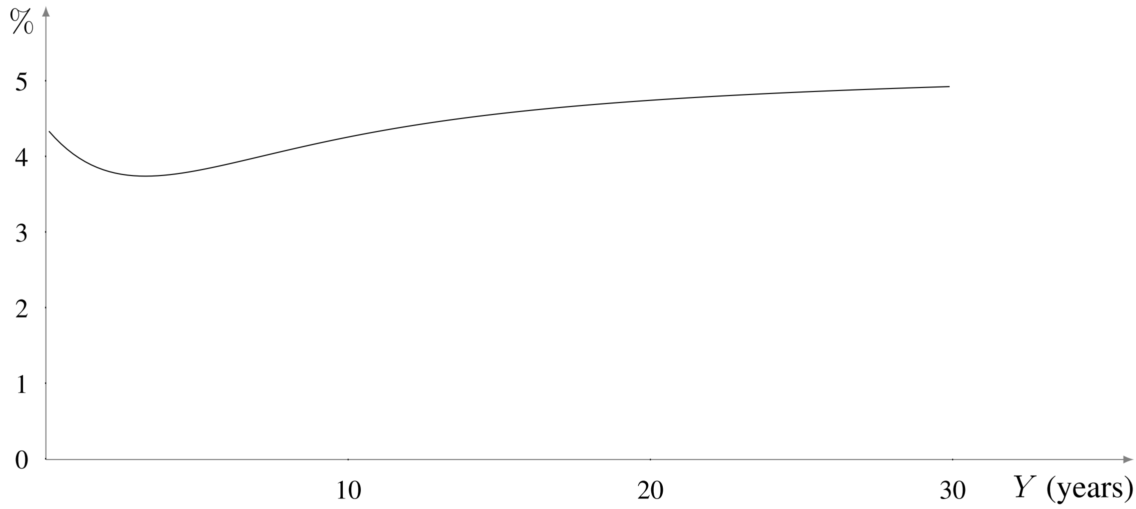


Figure 15.3. Yield curve of day 15/09/08

15.3 Changing the point of view

Fixing t we have the function $P(t,T),\ T \in (t,\tau]$ that, as we said, is builded with a family of bonds and it is a smooth curve. That smoothness is logical, since the price of bonds changes slowly, as we can check looking every day the prices.

However, we can fix T and look the price of a bond in function of the other variable t, that is, we fix T=20/03/15 and look the price of the bonds that matures that day. This is a very irregular curve as we can check if we graph the price the day 21/03/85 of a bond that matures in 30 years, the price 21/03/85 of a bond that matures in 30 years less 1 day, etc., until today (say 30/04/09) where we now the price of a bond that matures in 6 years 11 month. But note that we do not know the prices of tomorrow of a bond that matures the 20/03/15; that is, if today is $t_0 = 30/04/09$, then for $t > t_0$ the price $P(t,T)$ is unknown, and we will model it as a stochastic process.

A similar curve, easier to get, gives the price (until today) of a bond that matures after a fixed quantity of time. In Figure 15.4 there is a plot of the price of a bond with maturity 10 years from

January, 2, 2007 to April, 27, 2009 (data from the European Central Bank).

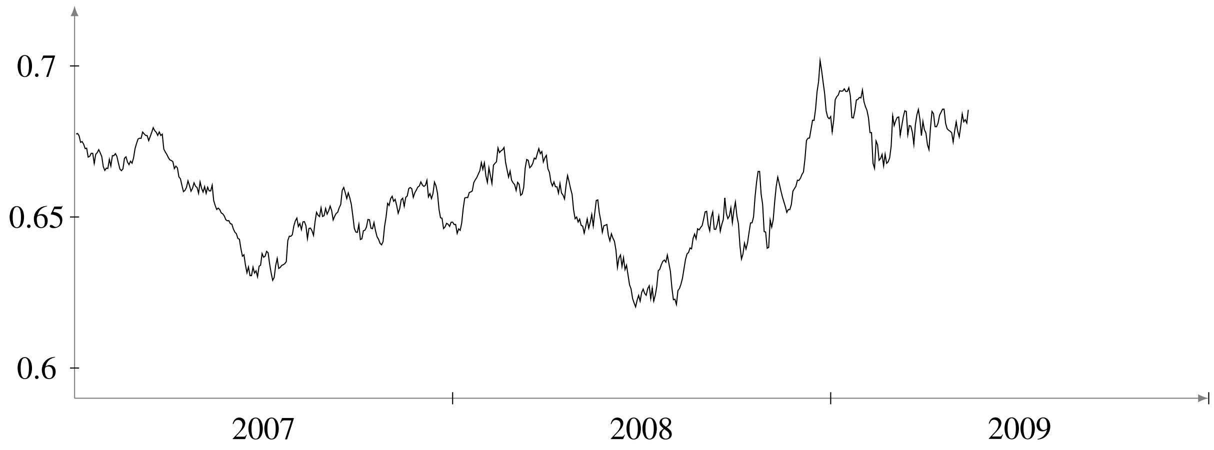


Figure 15.4. Curve $P(t, t+10),\ t = 2/01/07, \dots, 27/04/09$, with the evolution of the price of 10 years coupon bond.

15.4 A look at deterministic world

It the world would be deterministic the current and future interest rate would be known, and for any $t < T$ the quantity $P(t, T)$ would be known. Then it should hold that for any $0 < s < t < u$,

$$R(s, u) = \frac{t-s}{u-s} R(s, t) + \frac{u-t}{u-s} R(t, u),$$

that means, the yield between s and u is a weighted mean of the yields between s and t, and between t and u. Equivalently, by formula (15.1),

$$P(s, u) = P(s, t)P(t, u). \tag{15.3}$$

Exercise. Assume that $\exists t \in (s, u)$ such that

$$P(s, u) > P(s, t)P(t, u).$$

Construct an arbitrage. Analogously, if $P(s, u) < P(s, t)P(t, u)$.

If we assume that fixed t, the price curve

$$\begin{aligned} P(t, \cdot) : [t,\tau] &\longrightarrow [0, 1] \\ T &\mapsto P(t, T) \end{aligned}$$

is smooth, we can compute

$$r_t(u) = \frac{-\partial \log P(t, u)}{\partial u}.$$

From (15.3) it is deduced that the function $r_t(u)$ is the same for all t, and it is written $r(u)$ and

$$P(t,T) = \exp\{-\int_t^T r(u)\,du\}. \tag{15.4}$$

That formula is a generalization of expression $e^{-r(T-t)}$ that we have used from the begining assuming a fixed interest rate $r = r(u)$ for all $u \in [t,T]$.

15.5 The random world: neutral probability

In a random world, formulas (15.3) and (15.4) need to be reconsidered, since, as we commented, if t is in the future, then $P(t,T)$ is random. In general, we write **today** as $t = 0$. We fix a probability space $(\Omega, \mathcal{F}, P)$. The main modelling tool is to assume that there is a stochastic process $\{r_t,\, t \in [0,\tau]\}$ that models the interest rate, that is, $r(t)$ is the interest rate of a bond that matures at time $t + dt$, with $dt \to 0$; equivalently,

$$r(t) = \lim_{T \downarrow t} R(t,T).$$

We also have a filtration $\{\mathcal{F}_t,\, t \in \mathbb{R}_+\}$ that represents, as always, the information at time t, and we assume that $\{r_t,\, t \in [0,\tau]\}$ is adapted, and that fixed T, the process $\{P(t,T),\, t \in [0,T]\}$ is adapted, and consequently, also is $\{R(t,T),\, t \in [0,T]\}$.

From the interest rate r_t we can construct a riskless asset (but random),

$$S_t^{(0)} = \exp\{\int_0^t r(s)\,ds\},$$

that will give the random discount factor: $\exp\{-\int_0^t r(s)\,ds\}$. Specifically, the price at time t of a bond that matures at time T ***discounted to time 0*** is

$$\widetilde{P}(t,T) = \exp\{-\int_0^t r(s)\,ds\}\, P(t,T).$$

The main assumption in mathematical finance that there are no arbitrages, in this setup is guarantied if there is a probability $\mathbf{Q}$ equivalent to $\mathbf{P}$ such that for all T, the process

$$\{\widetilde{P}(t,T),\, t \in [0,T]\}$$

is a martingale. Notice that this condition is really strong because it is referred to an infinity of processes. A way to impose that condition is assuming that r_t is given by a (good) stochastic differential equation.

Under that probability $\mathbf{Q}$ it can be computed the **fair price** of a a bond:

$$\widetilde{P}(t,T) = \mathbf{E_Q}\big[\widetilde{P}(T,T)\,|\,\mathcal{F}_t\big],$$

and given that $P(T,T) = 1$,

$$\widetilde{P}(t,T) = \mathbf{E_Q}\big[\exp\{-\int_0^T r(u)\,du\}\,|\,\mathcal{F}_t\big],$$

Taking off the discount factor we arrive at

$$\boxed{P(t,T) = \mathbf{E_Q}\big[\exp\{-\int_t^T r(u)\,du\}\,|\,\mathcal{F}_t\big].} \tag{15.5}$$

So, (15.5) is a random version of the formula (15.4) in the deterministic world.

15.6 Forward rate

Consider $t < T_1 < T_2$ and assume that today is t. We want an agreement (today) saying that at time T_1 we ask for a loan of 1 €, that we will return with interest at time T_2. Denote by $F(t,T_1,T_2)$ the interest rate that we will pay during the period (T_1,T_2). Since $F(t,T_1,T_2)$ is decided today, that quantity should be $\mathcal{F}_t$ measurable. To compute the price, we can replicate the agreement:

- At time $\boldsymbol{t}$:
  - Buy a bond that matures at time T_1, with price $P(t,T_1)$.
  - Sell x bonds that mature at time T_2, at price $P(t,T_2)$.

  The initial cost of the portfolio should be 0, hence

  $$-P(t,T_1) + xP(t,T_2) = 0,$$

  and then

  $$x = \frac{P(t,T_1)}{P(t,T_2)}.$$

- At time $\boldsymbol{T}_1$: we receive 1 €, from the first bond.
- At time $\boldsymbol{T}_2$: We should pay x €, due to the bonds that we sold.

Then, to avoid arbitrages, the quantity of money that produces 1 € during the period $[T_1,\, T_2]$ should be a. That is,

$$\exp\big\{(T_2 - T_1)F(t,T_1,T_2)\big\} = x = \frac{P(t,T_1)}{P(t,T_2)},$$

thus

$$F(t,T_1,T_2) = -\frac{\log P(t,T_2) - \log P(t,T_1)}{T_2 - T_1}.$$

The following limit is called **instantaneous forward rate**:

$$f(t,T) = \lim_{\Delta T\to 0} F(t,T,T+\Delta T) = -\lim_{\Delta t\to 0}\frac{\log P(t,T+\delta T) - \log P(t,T)}{\Delta T} = -\frac{\partial \log P(t,T)}{\partial T}.$$

$$\boxed{f(t,T) = -\frac{\partial \log P(t,T)}{\partial T}.} \tag{15.6}$$

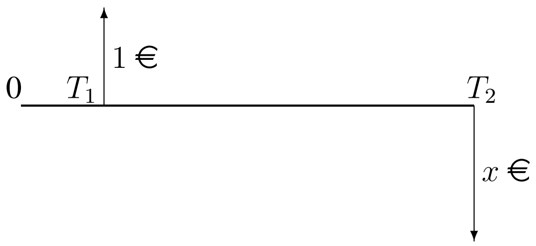


Figure 15.5. Computation of the forward rate

Then

$$\boxed{P(t,T) = \exp\big\{-\int_t^T f(t,u)\,du\big\}.} \tag{15.7}$$

Hence, we have two representations of $P(t,T)$, given by (15.5) and (15.7). It should be noted that in (15.7) does not appear the neutral probability, but we used it implicitly in the non arbitrage argument used to deduce that formula. Equaling both expressions we arrive to

$$\boxed{r(t) = f(t,t), \text{ q.s.}}$$

15.7 The classical modelling, with one factor or indirect

The classical modelling assumes that the interest rate is given by a stochastic differential equation

$$dr(t) = G(t, r(t))\,dt + K(t, r(t))\,dW_t.$$

Examples:

1. Vasicek
$$dr(t) = a(b - r(t))\,dt + \sigma\,dW_t. \tag{15.8}$$

2. Cox–Ingersoll–Ross (CIR)
$$dr(t) = a(b - r(t))\,dt + \sigma\sqrt{r(t)}\,dW_t.$$

3. Dothan
$$dr(t) = ar(t)\,dt + \sigma r(t)\,dW_t.$$

4. Black–Derman–Toy
$$dr(t) = h(t)r(t)\,dt + \sigma(t)r(t)\,dW_t.$$

5. Ho–Lee
$$dr(t) = h(t)\,dt + \sigma\,dW_t. \tag{15.9}$$

6. Hull–White (extended Vasicek)
$$dr(t) = (h(t) - a(t)r(t)\,dt + \sigma(t)\,dW_t.$$

7. Hull–White (extended CIR)

$$dr(t) = (h(t) - a(t)r(t)\,dt + \sigma(t)\sqrt{r(t)}\,dW_t.$$

Since r_t is Markov, from (15.5) it can be assumed that $P(t,T)$ only depends on $r(t)$, that is,

$$P(t,T) = H(t,T,r(t)),$$

where H is a *good* function. Then, by Itô, we can deduce an equation for $P(t,T),\ t \in [0,T]$ and have an insight of a neutral probability. This allows to find closed expressions for $P(t,T)$. For example, for Vasicek model (15.8),

$$P(t,T) = A(t,T)e^{-B(t,T)r_t},$$

where

$$\begin{aligned} B(t,T) &= \frac{1}{a}\big(1 - e^{-a(T-t)}\big) \\ A(t,T) &= \exp\Big\{\frac{1}{a^2}(B(t,T) - T + t)(a^2b - \sigma^2/2) - \frac{1}{4a}\sigma^2 B^2(t,T)\Big\}. \end{aligned}$$

As we see, the formula for $P(t,T)$ only depends of the parameters $a,\ b$ and σ, and the value of $r(t)$. This allows to automatize the computations from r_t.

15.8 The two factors models, or Heath-Jarrow-Morton or direct

The classical modelling depends on the equation of r_t. However, from a economical point of view this is poor (one factor explains all) and from a practical point of view, to fit a price curve it is needed to use complex models. Heath-Jarrow-Morton proposed to model directly $P(t,T)$, or equivalently, thanks to (15.7), $f(t,T)$. Specifically, it is assumed

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t,\ t \in [0,T], \tag{15.10}$$

where $t = 0$ is today, and the initial condition $f(0,T)$ is the forward instantaneous curve that can be deduced from (15.6) for $t = 0$ and $P(0,T)$, o directly from $R(0,T)$.

On the other hand, under the neutral probability , Heath-Jarrow-Morton prove that the coefficients of the equation (15.10) should satisfy

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,du, \tag{15.11}$$

and then it is easy to get simple realistic model fitted to empirical data.

Example. In a (15.10) we take $\sigma(t,T) = \sigma$. Then, by (15.11)

$$\alpha(t,T) = \sigma^2(T-t),$$

the equation for $f(t,T)$ is

$$f(t,T) = f(0,T) + \sigma^2\int_0^t (T-s)\,ds + \sigma\int_0^t dW_s,$$

where

$$f(t,T) = f(0,T) - \sigma^2\big((T-t)^2 - T^2\big)/2 + \sigma dW_t.$$

Since $r(t) = f(t,t)$, a.s., it is deduced the stochastic differential equation

$$r(t) = f(0,t) + \sigma^2 t^2/2 + \sigma\, B_t, \tag{15.12}$$

which is Ho–Lee (15.9) model with

$$h(t) = \frac{\partial f(0,t)}{\partial t} + \sigma^2 t,$$

computed under the neutral probability and fitted to the initial price curve, and we get the closed formula for $P(t,T)$

$$P(t,T) = \frac{P(0,T)}{P(0,t)}\,\exp\Big\{(T-t)f(0,t) - \frac{\sigma^2}{2}\,t(T-t)^2 - (T-t)r(t)\Big\}. \tag{15.13}$$

15.9 Exercises

The objective of this exercise is to clarify the concepts studied in this chapter, avoiding the difficult problems of fitting the model. We will assume that we have fitted a parametric curve to the yield curve

$$R(0,T) = \frac{1}{100}\Big(6 + \frac{T}{30} - e^{-T}\Big),\ T \in [0,30].$$

1. Represent $R(0,T)$. Compute the interest rate of a bond with maturation date 10, 20, 30 years.

2. Deduce the initial curve of the prices of bonds $P(0,T),\ T \in [0,30]$. What is the price of a bond to 10, 20, 30 years?

3. Deduce the instantaneous forward rate $f(0,T),\ T \in [0,30]$. *Indicacition:* From (15.2) and (15.6) deduce the value of $f(0,T)$ from $R(0,T)$.

Consider the HJM model for the instantaneous forward rate with Ho-Lee volatility that we studied in page 170, with $\sigma = 0.001$

4. Simulate three trajectories for $r_t,\ t \in [0,15]$, with mesh $\delta = 1$, corresponding to equation (15.12).

5. From each simulation of r_t, compute $P(t,T)$ in a yearly grid of points:

$$P(t,T),\ t = 0,\ldots,15, T = 0,\ldots,15,\ t \le T,$$

using (15.13).

6. Use the points computed in the previous exercise to plot the three curves $P(5,T),\ T \in [5,15]$, and the three curves $P(t,15),\ t \in [0,15]$. Interpret such curves.

Bibliography

[1] BACK, K., *A Course in Derivative Securities*, Springer 2005.

[2] BAXTER, M. AND RENNIE, A., *Financial Calculus*, Cambridge University Press, 1996.

[3] CONT, R. AND TANKOV, P., *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, Boca Raton, 2004.

[4] COX, J. C., ROSS, S. A. AND RUBINSTEIN, M., Option pricing: A simplified approach, Journal of Financial Economics, **7** (1979) 229-263.

[5] B.DALTON, *Financial Products*, Cambridge, 2008.

[6] DURRET, R., *Probability: Theory and Examples*. 3rd ed., Thomson/Brooks Cole, 2005.

[7] JOSHI, M. S., *The Concepts and Practice of Mathematical Finance*, 2nd edition, Cambridge, 2008.

[8] HULL, J. C., *Options, Futures and other derivatives*, Sixth edition, Prentice Hall, 2005.

[9] KLOEDEN, P. E. AND PLATEN, E., *Numerical Solutions of Stochastic Differential Equations*, Springer, 1992.

[10] LAMBERTON, D. AND LAPEYRE, B., *Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall, 1996.

[11] MERTON, R. C., *Option pricing when underlying stock returns are discontinuous*. Journal of Financial Economics **3** (1976) 125–144.

[12] MIKOSCH, T., *Elementary Stochastic Calculus with Finance in View*, World Scientific, 1999.

[13] PLISKA, S. R., *Introduction to Mathematical Finance*, Blackwell Publishers, Inc, 1997

[14] ROBERTS , A. W. AND VARBERG, D. L., *Convex functions*, Academic Press, 1973.

[15] ROSS, S. M., *An Elementary Introduction to Mathematical Finance.* 2nd. ed.Cambridge University Press, 2003.

[16] SCHOUTENS, W., *Lévy Processes in Finance*, Wiley, 2003.

[17] SHIRYAEV, A. N., *Probability*, Springer, 1984.

[18] SHIRYAEV, A. N., *Essencials of Stochastic Finance: Facts, Models, Theory*, World Scientific, 1999.

[19] SHREVE, S. E., *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer, 2005.

[20] SHREVE, S. E., *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, 2004.

[21] STEELE, J. M., *Stochastic Calculus and Financial Applications*, Springer, 2001.