Stochastic Processes for Finance

Patrick Roger

bookboon
The eBook company

Stochastic Processes for Finance

Patrick Roger

Strasbourg University, EM Strasbourg Business School

June 2010

Stochastic Processes for Finance

ISBN 978-87-7681-666-7

Contents

Introduction

In "Probability for Finance" (Roger, 2014) we presented the essential tools from probability theory which are useful in one-period financial models[1]. We assume here that the reader is comfortable with these elements of probability theory. This second book is an extension to multi-period models, either in the discrete or continuous-time framework.

When dealing with multi-period models, one of the key problems is to take into account the revelation of information over time, especially the information transmitted by the observation of economic variables like prices, interest rates or exchange rates. We already referred to this problem when conditional expectations were developed in "Probability for Finance". This technical tool will be used extensively in the chapters of this book.

As mentioned before, there are several approaches to study multi-period models, depending on the way time is measured. Roughly speaking, financial models can be divided in two families. In discrete-time models, markets are open on a finite or countable set of dates, denoted $0, 1, .., T$. In continuous-time models, markets are always open and the set of dates is an interval $[0; T]$.

These two categories have their own advantages and drawbacks. Discrete-time models are easier to understand and sometimes allow to solve valuation problems that cannot be easily managed in continuous-time. This is the case for the valuation of American options. However, continuous-time models often provide simple analytical solutions (also called closed-form solutions) when discrete-time models only provide untractable solutions and/or bulky formulas.

It is also clear that discrete-time models use a less sophisticated mathematical machinery and make easier economic interpretations. It is the reason why chapter 1 starts with the presentation of discrete-time stochastic processes. A section is devoted to Markov chains which are common tools, especially in credit risks models. A particular subset of discrete-time processes, namely martingales, is especially important in finance, leading to devote a large part of this chapter to these processes (in their discrete-time version).

[1] For example in portfolio choice models like the one developed by Harry Markowitz (1952, 1959).

The second chapter addresses continuous-time processes. Financial models started to use these tools at the end of the sixties[2]. On a mathematical point of view, they are more demanding and more difficult to understand. In some cases, they reveal "strange" objects like continuous and nowhere differentiable functions. One more time, the subset of continuous martingales is of primary importance in valuation models but one particular process, called "Brownian motion" or "Wiener process" is the main building block of a broad category of processes used in financial models. Consequently, a non negligible part of this second chapter is devoted to the study of the Wiener process.

Chapter 3 introduces stochastic calculus. When managing a portfolio in continuous-time, one has to combine quantities of assets with price variations to calculate the return of the portfolio. In discrete-time, it is simply written as a sum (over the set of dates) of products (quantities times price variations) and aggregated over the different stocks in the portfolio. In continuous-time, it is technically more difficult to perform these calculations. This problem is solved by using stochastic integrals.

Moreover, a usual problem in finance is to determine the dynamics of prices of derivative securities (like futures contracts or options), knowing the dynamics of the underlying asset of the contract. It can be achieved by using Itô's lemma, which can be seen as a Taylor series expansion for stochastic processes. It is constantly used in continuous-time valuation models of derivative securities.

Finally, it can be shown (in arbitrage-free pricing models), that the today price of a security is its expected tomorrow price[3] discounted at the risk-free rate, the expectation being calculated under a so called risk-neutral probability measure. It means that in the risk-neutral world, risky assets do not deliver a risk premium. But they deliver such a premium in the real world. Therefore, a mathematical tool is needed to express the dynamics of a security price in the risk-neutral world, knowing the corresponding dynamics in the real world. Girsanov theorem is the right tool to perform this transformation; it is presented at the end of chapter 3.

[2] See Merton (1969, 1971)

[3] We assume that no dividends are paid between today and tomorrow.

Chapter 1

Discrete-time stochastic processes

1.1 Introduction

In one-period models, the only uncertainty is about the state of nature occuring at date T, trading being realized only at date 0. There is neither learning by investors nor information disclosure at intermediate dates for the simple reason that there are no intermediate dates.

In this chapter, we introduce a multi-period market in a simple framework. It is not the most general context since the set of relevant dates is assumed finite. Continuous markets will be described in the next chapter. In section 1.2 we introduce stochastic processes in this constrained framework. Section 1.3 addresses the question of information revelation over time. The good understanding of this section is essential, not because it is technically difficult, but because it describes the formal model of information revelation and the assumptions commonly used in multi-period financial models. In section 1.5 we enter the core of the subject by presenting martingales and their use in financial models.

Three questions are addressed.

1) The Doob decomposition of a martingale, useful to decompose the return on a financial asset in two components, a predictable one and a zero-mean "surprise"[1].

[1] This term has a precise mathematical translation but for now the reader can keep in

2) We show that in a market without arbitrage opportunities, there doesn't exist a dynamic strategy allowing to "beat the market".

3) We prove that even if one chooses the date of entry and the liquidation date in a smart way, one cannot beat the market. The financial gain/loss of the strategy has the same characteristics as the one obtained using the standard *buy and hold* strategy.

Obviously, the results obtained in (2) and (3) are based on disputable assumptions, like market perfection, which are not always satisfied on real markets. However, this approach provides solid foundations to the theory of financial valuation. Deviations occur temporarily but, in most cases, the activity of arbitrageurs bring back prices toward the theoretical values.

1.2 The general framework

As mentioned in the introduction, we consider a discrete-time economy with a relevant set of dates denoted as $\mathcal{T} = \{0, 1, ..., T\}$ where $T < +\infty$. The latter assumption is not mandatory but finite-horizon models cover the major part of the financial literature. The financial market is open at dates in the set $\mathcal{T}_- = \{0, 1, ..., T-1\}$; the last transactions are realized at date $T-1$ and the securities pay a liquidating dividend or a terminal payoff by date T. It is equivalent to assume that agents consume their remaining wealth at date T and die just after!

Uncertainty is described by a probability space $(\Omega, \mathcal{A}, P)$ where Ω is the set of states of nature, $\mathcal{A}$ is a tribe on Ω and P is a probability measure on $\mathcal{A}$.As the market is open at different dates, prices and returns are sequences of random variables indexed by time. Such a sequence is called a stochastic process.

Definition 1 *A discrete-time stochastic process is a sequence of random variables $X = (X_0, ..., X_T)$ defined on $(\Omega, \mathcal{A})$ and taking values in* [2] $\mathbb{R}$.

This definition may seem restrictive because variables X_t are real random variables. We could also assume that they are n-dimensional random vectors

mind the usual meaning of this word.

[2] The set $\mathbb{R}$ is always equipped with the Borel tribe $\mathcal{B}_{\mathbb{R}}$.

but it will not be necessary in the present chapter. In the same way, indexation could be done by any ordered set but it is of little interest in financial models where time is the natural index.

The variables X_t can be either discrete or continuous. For example, when the binomial distribution of stock prices was presented in the first book, it was referred to a stochastic process of stock prices X written as:

$$X_t = X_{t-1} \times Y_t$$

where the variables Y_t were taking values u and d with probabilities p and $1 - p$. When a positive initial value X_0 is defined arbitrarily, X is a stochastic process and the X_t are discrete random variables (with finite support in this case). On the contrary, if the X_t were assumed to be lognormally distributed, the variables would be continuous because they could take any value in $\mathbb{R}^+$. So it is important to distinguish what is discrete. There exist discrete(continuous)-time stochastic processes with discrete (continuous) variables. The time dimension and the state dimension are clearly different.

State	$X_1(\omega)$	$X_2(\omega)$
ω_1	100	95
ω_2	100	106
ω_3	105	102
ω_4	105	109

Table 1.1: Definition of X_1 et X_2

1.3 Information revelation over time

1.3.1 Filtration on a probability space

The definition of a stochastic process shows that the time-dimension must be explicitly included in the description of a financial market and in the sequence of prices. Specifying the distributions of the X_t is not sufficient to describe what really happens on the market. In particular, even if X_1 and X_2 (a stock price at two successive dates) follow the same probability distribution, the value of X_1 is known at date 1 when the value of X_2 is not yet revealed. Moreover, it is important to note that all the variables in the process (whatever their time-index is) are defined on the same space Ω.

To illustrate the point in a simple framework, consider $\Omega = \{\omega_i, i = 1, ..., 4\}$ and $\mathcal{A} = \mathcal{P}(\Omega)$; assume that X_1 and X_2 are defined by table 1.1.

Observing X_1 at date 1 provides some information[3]. For example, $X_1 = 100$ reveals that the event $\{\omega_1, \omega_2\}$ occurs. On the contrary, observing $X_1 = 105$ means that the pair $\{\omega_3, \omega_4\}$ occurs. Two remarks can be done at this stage. First, the conditional probabilities related to the possible values of X_2 are changed after the observation of X_1. If $\{\omega_1, \omega_2\}$ is true, only prices 95 and 106 remain possible at date 2. Second, even if observing X_1 reveals information, uncertainty remains because the terminal date ($T = 2$) has not yet been reached. We do not know exactly at date 1 what will be the value of X_2 at date 2.

Obviously this example is overly simplified. But the definitions of X_1 and X_2 seems to possess reasonable properties when comparing to real markets. Agents accumulate information over time but uncertainty remains about future prices (at least for risky assets). We let the reader check that the

[3] This type of example was already studied in the chapter of "Probability for Finance" dealing with conditional expectations.

description of X_1 and X_2 is relevant because the σ-algebra generated by X_1 is strictly included in the σ-algebra generated by the pair (X_1, X_2).

To build a consistent financial model, it is then necessary to choose a relevant definition of information revelation over time. In particular, at a given date t, agents observe date-t prices which are no more random at future dates. Moreover, it seems reasonable to assume that the quantity of information held by investors increases over time (corresponding to the idea that investors don't forget anything!).

In technical terms, these points can be summarized as follows. The list of events you know to be true or false at a given date s is included in the corresponding list at a future date t. It also means that if an event is true at a given date it will also be true at any future date. Filtrations are the right mathematical tool to translate these ideas.

Definition 2 *A filtration on a probability space* $(\Omega, \mathcal{A}, P)$ *is an increasing sequence* $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, ..., \mathcal{F}_T\}$ *of sub-tribes of* $\mathcal{A}$. *The quadruple* $(\Omega, \mathcal{A}, P, \mathcal{F})$ *is called a filtered probability space.*

Increasingness of tribes $\mathcal{F}_t$ is understood here as:

$$\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$$

for any $t \geq 1$.

In a model ending at date T, it is reasonable to assume that no uncertainty remains at the terminal date T. Consequently, it is often assumed that $\mathcal{F}_T = \mathcal{A}$. In what follows, we always consider this assumption as satisfied without recalling it systematically.

In the same way, we assume that nothing is known at date 0, that is $\mathcal{F}_0 = \{\emptyset; \Omega\}$. Despite the fact you know nothing, you are able to say that the impossible event doesn't occur and that the sure event occurs.

In the example of table 1.1, the relevant filtration is:

$$\begin{aligned}
\mathcal{F}_0 &= \{\emptyset, \Omega\} \\
\mathcal{F}_1 &= \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\} \\
\mathcal{F}_2 &= \mathcal{P}(\Omega)
\end{aligned}$$

At the intermediate date $t = 1$ you are able to say if $\{\omega_1, \omega_2\}$ or $\{\omega_3, \omega_4\}$ occurs because you observe X_1 which brings the necessary information. Consequently, at date 0, you can describe the events you will know to be true or false at future dates. The reader has to be conscious that it is a strong assumption. In the real life (and especially on financial markets), many events occur which were even not imagined by investors before they occured[4].

1.3.2 Adapted and predictable processes

The information carried by a stochastic process X is increasing over time and we need to specify what is meant by "agents do not forget" and "X_t is not random after date t". Consider that the future price of IBM stock is a stochastic process, that is a sequence of random variables. At date t, investors observe the stock price X_t and X_t will not change at date $t+1$ (it will be X_{t+1} at this future date). It means that X_t is random up to date t and becomes a number (or a constant random variable) after date t.

If $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, ..., \mathcal{F}_T\}$ describes the evolution of information over time, $\mathcal{F}_t$ is the date-t list of events you know as true or false. In particular, you know if, for example, $\{X_t \leq 100\}$ is true or not, where X_t is the IBM stock price. It simply means that X_t is measurable with respect to $\mathcal{F}_t$ (the same remark applies to any date t).

Definition 3 *1) A stochastic process $X = (X_0, ..., X_T)$ is adapted to the filtration $\mathcal{F}$ if, for any t, X_t is $\mathcal{F}_t$-measurable.*

2) The natural filtration of a stochastic process X is the sequence $\mathcal{F}^X$ of sub-tribes of $\mathcal{A}$ such that $\mathcal{F}_t^X$ is the tribe generated by the variables $X_s, s \leq t$.

Consider one-more time the Cox-Ross-Rubinstein model (1979) with $X_t = X_{t-1}Y_t$, the Y_t being independent and taking values u and d with probabilities p and $1-p$. If $T = 2$, there are four possible paths for X between $t = 0$ and $T = 2$. Then, 4 states are necessary to describe the evolution of prices. We note $\Omega = \{uu, ud, du, dd\}$; at date 1, X_1 is observed and we know if the price lies on a path starting by an *up* move or a *down* move. As mentioned before, the events $\{uu, ud\}$ and $\{du, dd\}$ are known to be true or false by that date.

[4] Probably, the number of investors who were considering the bankruptcy of Lehman Brothers as possible was very low, one year before it occured in September 2008.

$\mathcal{F}_1 = \{\emptyset, \{uu, ud\}, \{du, dd\}, \Omega\}$ is then the relevant tribe to describe the information known at date 1. X_1 is defined by:

$$\begin{aligned} X_1(uu) &= X_1(ud) = uX_0 \\ X_1(du) &= X_1(dd) = dX_0 \end{aligned}$$

This variable is $\mathcal{F}_1$-measurable and in fact $\mathcal{F} = \mathcal{F}^X$. This is a general remark; when only one risky asset is traded on the market, the natural filtration of the price process is sufficient to modelize the information revelation in the economy.

State	X_0	$X_1(\omega)$	$X_2(\omega)$
ω_1	1	u	u^2
ω_2	1	u	ud
ω_3	1	d	du
ω_4	1	d	d^2

Table 1.2: Definition of the price process X

When Ω is finite with $Card(\Omega) = n$, any random variable can be identified to a vector with n components. The space $L^2(\Omega, \mathcal{A}, P)$ is itself identified to $\mathbb{R}^n$. Moreover, $L^2(\Omega, \mathcal{F}_t, P)$, the space of $\mathcal{F}_t$-measurable random variables is a vector subspace of $L^2(\Omega, \mathcal{A}, P)$ (see the geometric interpretation of conditional expectations in the first book).

Obviously, the dimension of this subspace increases with t. In the former example, $L^2(\Omega, \mathcal{F}_1, P)$ was two-dimensional, characterized by:

$$L^2(\Omega, \mathcal{F}_1, P) = \left\{(x, y, z, t) \in \mathbb{R}^4 \text{ such that } x = y \text{ and } z = t\right\}$$

In other words, the tribe $\mathcal{F}_1$ separates neither the first pair of states, nor the second pair. Table 1.2 shows the entire price process. To simplify, it is assumed that $X_0 = 1$ in table 1.2.

Consider now an investor taking a position on a market where K assets are traded, whose price processes are denoted as $X^k, k = 1, ..., K$.

He takes a position at date t and rebalances his portfolio at date $t+1$. He still holds the date-t portfolio at date $t+1$ just before the rebalancing. We then need to choose notations, either $\theta'_t = \left(\theta^1_t, ...\theta^K_t\right)$ or $\theta'_{t+1} = \left(\theta^1_{t+1}, ...\theta^K_{t+1}\right)$ for the portfolio held between t et $t+1$.

In discrete-time, the two choices are equivalent but, to keep the vocabulary consistent with the choices to be made in continuous-time in next chapters, it is preferable to select the second notation. However, one has to remark that θ_{t+1} is revealed at date t. This corresponds to what is called a predictable process.

Definition 4 *A stochastic process $X = (X_1..., X_T)$ is predictable if, for any $t \geq 1$, X_t is $\mathcal{F}_{t-1}$-measurable.*

The essential interest of predictable processes will appear in the section devoted to martingales. The process in definition 4 starts with an index $t = 1$

since X_1 is $\mathcal{F}_0$-measurable. When a process θ denotes quantities of stocks, one can argue that the agent may already possess some stocks at date 0, as an initial endowment. In this case, it is convenient to denote θ_0 this endowment, knowing that the first transactions realized at date 0 will transform θ_0 in θ_1.

Example 1 *Other stochastic processes than quantities are naturally predictable. Consider a market on which it is possible to invest at each date and for one period in a locally risk-free asset (savings account); let r_t be the rate of return on this account on the time-period $[t; t+1]$ and B_t the amount obtained at date t by a rolling investment of one dollar in the savings account since date 0.*

Recall that r_t is known at date t and the process r is adapted to the filtration $\mathcal{F}$, explaining the expression "locally risk-free" (only risk-free one period ahead). The process B is then predictable. In fact, it can be written:

$$B_t = \prod_{s=0}^{t-1}(1 + r_s)$$

B_t is known at date $t-1$, then B_t is $\mathcal{F}_{t-1}$-measurable.

1.4 Markov chains

1.4.1 Introduction

Rating agencies like Moodys, Standard & Poor's or Fitch, regularly publish statistics about the evolution of the ratings of a number of financial instruments. From time to time they publish tables like the one on figure 1.1. The first column with elements identified from AAA (the best grade) to CCC (the worst grade before default) provides possible grades given by the rating agency, on a given year p, to given financial instruments like corporate bonds.

The first line gives the possible states of the rating process on year $p+1$. There are two more elements in this line. "D" means default and "NR" means "not rated". The numbers in the matrix provide the proportion of firms moving from a rating to another one. For example, 88.46% on the top left means that 88.46% of bonds rated AAA on year p are still rated AAA on year $p+1$. The number just on the right is 8.05. It signifies that 8.05% of

	AAA	AA	A	BBB	BB	B	CCC	D	NR
AAA	88.46	8.05	0.72	0.06	0.11	0.00	0.00	0.00	2.61
AA	0.63	88.27	7.47	0.56	0.05	0.13	0.02	0.00	2.87
A	0.08	2.32	87.64	5.02	0.65	0.22	0.01	0.05	4.01
BBB	0.03	0.29	5.54	82.49	4.68	1.02	0.11	0.17	5.67
BB	0.02	0.11	0.58	7.01	73.83	7.64	0.89	0.98	8.93
B	0.00	0.09	0.21	0.39	5.98	72.76	3.42	4.92	12.23
CCC	0.17	0.00	0.34	1.02	2.20	9.64	53.13	19.29	14.21

Figure 1.1: One-year ratings migration

the bonds rated AAA on year p have been downgraded to AA on year $p+1$. This kind of matrix is called a transition matrix or a migration matrix.

On a mathematical point of view, these transition matrices are linked to a category of stochastic processes called Markov processes. They are defined as follows.

1.4.2 Definition and transition probabilities

Definition 5 *A process X defined on $(\Omega, \mathcal{A}, P, \mathcal{F})$ is a **Markov process** if for any $(B_1, ..., B_n) \in \mathcal{B}_{\mathbb{R}}^n$ and any $(t_1, ..., t_n) \in \mathcal{T}^n$ such that $t_1 < .... < t_n$:*

$$P\left(X_{t_n} \in B_n \left| X_{t_j} \in B_j, j = 1, .., n-1\right.\right) = P\left(X_{t_n} \in B_n \left| X_{t_{n-1}} \in B_{n-1}\right.\right)$$

The Markov property says that the history of the process up to date t_{n-1} does not matter to determine what happens between t_{n-1} and t_n. The only important information is to know where is located the process at date t_{n-1}.

If the state space is finite, the set of values taken by the variables $(X_p, p \in \mathbb{N})$ is denoted $(x_1, ..., x_n)$ and Markov processes have specific properties. They are called **finite Markov chains**.

Definition 6 *Let $(X_p, p \in \mathbb{N})$ a finite Markov chain taking values in $(x_1, ..., x_n)$;*
1) Transition probabilities of order 1 are the quantities:

$$\pi(x_i, p-1, p, x_j) = P\left(X_p = x_j \left| X_{p-1} = x_i\right.\right) \tag{1.1}$$

2) The matrix $\Pi_{p-1} = (\pi(x_i, p-1, p, x_j), i, j = 1, ..., n)$ is the transition matrix of the Markov chain at date $p-1$.
2) When the transition probabilities do not depend on p, the Markov chain is said homogeneous (or stationary) and notations are simplified by writing $\pi(x_i, p-1, p, x_j) = \pi_{ij}$

$\pi(x_i, p-1, p, x_j)$ is the probability for the process X to be in state x_j at date p knowing it is in state x_i at date $p-1$. These probabilities are a specific case of the probabilities appearing in definition 5.

If two lines (corresponding to "D" and "NR") are added to the transition matrix in figure 1.1 (with ones as diagonal terms and zero otherwise), the resulting square matrix is the transition matrix of a Markov chain with 9 states.

1.4.3 Chapman-Kolmogorov equations

Figure 1.1 showed the one-year transition probabilities. But it is also important to estimate the transition probabilities for 2 years, 3 years and, more generally m years. For example, a bank lending money to a firm for 5 years is interested in the 5-year default probability.

For a homogeneous Markov chain, the Chapman-Kolmogorov equations allow to deduce the m-year transition probabilities from the corresponding one-year transition probabilities.

Proposition 1 *Let $X = (X_p, p \in \mathbb{N})$ denote a N-state homogeneous Markov chain with one-period transition probability matrix $\pi = (\pi_{ij}, i, j = 1, ..., N)$. Let $\pi_{ij}^{(m+n)} = P(X_{m+n} = x_j \,|X_0 = x_i)$ denote the $m + n$-period transition probability from state i to state j.*

$$\pi_{ij}^{(m+n)} = \sum\nolimits_{k=1}^{n} \pi_{ik}^{(m)} \pi_{kj}^{(n)} \tag{1.2}$$

To keep things simple, suppose $m = n = 1$ and $N = 3$. For example, $\pi_{12}^{(2)}$ is the probability for the process to be in state x_2 at date 2, knowing the date-0 state was x_1. The graph below shows the 3 possible paths with the corresponding transition probabilities.

$$\begin{array}{ccccc} & \overset{\pi_{11}}{\nearrow} & x_1 & \overset{\pi_{12}}{\searrow} & \\ x_1 & \overset{\pi_{12}}{\rightarrow} & x_2 & \overset{\pi_{22}}{\rightarrow} & x_2 \\ & \underset{\pi_{13}}{\searrow} & x_3 & \underset{\pi_{32}}{\nearrow} & \end{array}$$

We then deduce:

$$\pi_{12}^{(2)} = \pi_{11}\pi_{12} + \pi_{12}\pi_{22} + \pi_{13}\pi_{32} \tag{1.3}$$

More generally, the Chapman-Kolmogorov equations "count" the possible paths from x_i to x_j in $m + n$ steps and simply cumulate the probabilities of all these paths. However, the formulation is very interesting because $\pi_{ij}^{(m+n)}$ is the inner product of the i-th row of the m-period transition matrix $\left(\pi_{ij}^{(m)}, i, j = 1, ..., n\right)$ with the j-column of the n-period transition matrix $\left(\pi_{ij}^{(n)}, i, j = 1, ..., n\right)$. We then have the useful property.

Corollary 1 *Let $X = (X_p, p \in \mathbb{N})$ denote a N-state homogeneous Markov chain with one-period transition probability matrix $\pi = (\pi_{ij}, i, j = 1, ..., N)$. Let $\pi_{ij}^{(m)} = P(X_m = x_j \,|X_0 = x_i)$ denote the m-period transition probability from state i to state j and $\pi^{(m)}$ the m-period transition matrix containing the $\pi_{ij}^{(m)}$.*

$$\pi^{(m)} = \pi^m \tag{1.4}$$

where π^m is the m-th power of the one-period transition matrix π.

1.4.4 Classification of states

Accessibility and communication

Definition 7 *1) A state x_j of a Markov chain X is* ***accessible*** *from a state x_i if there exists an integer $k > 0$ such that:*

$$\pi_{ij}^{(k)} > 0 \tag{1.5}$$

2) Two states x_i and x_j of a Markov chain ***communicate*** *if there exist integers k and k' such that:*

$$\pi_{ij}^{(k)} > 0 \text{ and } \pi_{ji}^{(k')} > 0 \tag{1.6}$$

Example 2 *Consider the following matrix π.*

$$\pi = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.7 & 0.3 \end{bmatrix} \tag{1.7}$$

It can be observed that when the process starts in state x_1 or x_2, states x_3 and x_4 are never reached; they are not accessible. We can also see that x_1 and x_2 communicate. The same is true for x_3 and x_4. When starting from one of these two states, x_1 or x_2 are not accessible. In fact, for this specific chain, there exists 2 "closed" classes $\{x_1, x_2\}$ and $\{x_3, x_4\}$.

The following proposition generalizes the above remark.

Proposition 2 *Let $X = (X_p, p \in \mathbb{N})$ denote a N-state homogeneous Markov chain with one-period transition probability matrix $\pi = (\pi_{ij}, i, j = 1, ..., N)$. There exists a partition $(C_1, C_2, ...C_M)$ of the N states such that two states communicate if and only if they are in the same class[5]. The sets C_k are called* ***communicating classes.***

When looking at figure 1.1, we easily see that the set of ratings AAA to CCC communicate even if some transition probabilities are equal to 0. But the state D corresponding to default does not communicate with the others.

[5] The relation $\mathcal{R}$, defined by $x_i \; \mathcal{R} \; x_j$ iff x_i and x_j communicate, is an equivalence relation (reflexive, symmetric and transitive).

The implicit assumption is that a defaulted instrument never recovers. It does not mean that D is not accessible from the other states. In fact, the column "D" on figure 1.1 contains non zero entries. For example, a corporate bond rated CCC has a probability 19.29% to default on the next year. D is not in the same communicating class as the other states because reaching D prevents to come back to an other state.

The state NR is particular because we don't know if a non rated financial instrument will be rated in the future. However, a simplifying assumption is to consider that NR instruments at a given date will stay NR in the future.

Definition 8 *A Markov chain is said* ***irreducible*** *if all states communicate, in other words there is only one communicating class.*

Periodicity

Consider a binomial model for the evolution of a stock price S defined for any date $t+1$ by

$$S_{t+1} = \begin{cases} uS_t \text{ with probability } p \\ dS_t \text{ with probability } 1-p \end{cases} \tag{1.8}$$

A usual calibration for this model is to assume $d = 1/u$. In this case, if the initial price S_0 is \$100, you are sure that the price cannot come back to \$100 in less than two periods.

Definition 9 *Let* $X = (X_p, p \in \mathbb{N})$ *denote a* N*-state homogeneous Markov chain with one-period transition probability matrix* $\pi = (\pi_{ij}, i, j = 1, ..., N)$.

1) The ***periodicity*** *of a state* x_i*, denoted* $t(i)$ *is the greatest common divisor of numbers* m *such that* $\pi^{(m)}(i,i) > 0$*. By convention,* $t(i) = 0$ *if for any* m*,* $\pi^{(m)}(i,i) = 0$*.*

2) A Markov chain X *is* ***aperiodic*** *if* $t(i) = 1$ *for any* i*.*

In the binomial model viewed as a Markov chain, the periodicity of states is equal to 2.

Definition 10 *Let* $X = (X_p, p \in \mathbb{N})$ *denote a homogeneous Markov chain with one-period transition probability matrix* π*.*

*1) Let $p_i(n)$ denote the probability of coming back to state i after n periods. A state i is **recurrent** if*

$$\sum_{i=1}^{+\infty} p_i(n) = 1 \tag{1.9}$$

*2) A state which is not recurrent is said **transient**.*
*3) A state i is said **positively recurrent** if*

$$\lim_{m \to +\infty} \pi_{ii}^{(m)} = q_i > 0 \tag{1.10}$$

The chain comes back almost surely on any recurrent state after a first passage. When the state i is positively recurrent, the transition probability of coming back to state i in n states never vanishes, even when the length of the transition period tends to infinity. When all states are positively recurrent, the chain is called positively recurrent.

1.4.5 Stationary distribution of a Markov chain

Proposition 3 *Let $X = (X_p, p \in \mathbb{N})$ denote an aperiodic, positively recurrent, homogeneous Markov chain. We get:*

$$q_i = \sum_{j=1}^{+\infty} q_j \pi_{ij} \tag{1.11}$$

with $\sum_{i=1}^{+\infty} q_i = 1$. Moreover, the probabilities q_i are uniquely determined by the following relationships:

$$\begin{aligned} \forall i, q_i &\geq 0 && (1.12) \\ \sum_{i=1}^{+\infty} q_i &= 1 && (1.13) \\ q_i &= \sum_{i=1}^{+\infty} q_j \pi_{ij} && (1.14) \end{aligned}$$

The probability measure by the q_i on the states of the chain is called the ***stationary distribution*** *of the Markov chain.*

In fact, for a finite Markov chain, looking for the stationary distribution of the chain consists in calculating the powers π^m of the one-period transition probability matrix. When m tends to infinity, the lines of π^m become identical. Each line represents the probability measure $(q_i, i = 1, ..., n)$.

Example 3 *Consider the very simple case of a two-state chain characterized by the following transition matrix.*

$$\pi = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \tag{1.15}$$

This chain obviously satisfies the assumptions of proposition 3. The successive powers of π give:

$$\pi^2 = \begin{bmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{bmatrix} \quad \pi^3 = \begin{bmatrix} 0.504 & 0.496 \\ 0.496 & 0.504 \end{bmatrix} \text{ and } \lim_{m \to +\infty} \pi^m = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \tag{1.16}$$

The stationary distribution of the chain means that the probability of being in one state on the long run does not depend on the initial state.

1.5 Martingales

We first recall the essential properties of conditional expectations, already presented in chapter 4 of Roger (2010).

Proposition 4 *Let (Y, Z) be two random variables in $L^2(\Omega, \mathcal{A}, P)$ and $\mathcal{B}, \mathcal{B}'$ two sub-tribes of $\mathcal{A}$ satisfying $\mathcal{B} \subset \mathcal{B}'$:*

1) If Z is a constant $c \in \mathbb{R}$, $E(Z|\mathcal{B}) = c$

2) $\forall (a,b) \in \mathbb{R}^2$, $E(aZ + bY|\mathcal{B}) = aE(Z|\mathcal{B}) + bE(Y|\mathcal{B})$

3) If $Z \leq Y$, $E(Z|\mathcal{B}) \leq E(Y|\mathcal{B})$

4) $E(E(Z|\mathcal{B}')|\mathcal{B}) = E(Z|\mathcal{B})$ (law of iterated expectations)

5) If Z is $\mathcal{B}$-measurable $E(ZY|\mathcal{B}) = Z\,E(Y|\mathcal{B})$

6) If Z is independent of $\mathcal{B}$, $E(Z|\mathcal{B}) = E(Z)$

We can now define the stochastic processes known as martingales.

Definition 11 *a) Let $(\Omega, \mathcal{A}, \mathcal{F}, P)$ a filtered probability space; a stochastic process $X = (X_0, ..., X_T)$ is a $(\mathcal{F}, P)$-**martingale** if:*
i) X is adapted to $\mathcal{F}$
ii) $\forall t \in \mathcal{T}$, $X_t \in L^1(\Omega, \mathcal{A}, P)$
iii) $\forall t \in \mathcal{T}^*$, $X_{t-1} = E[X_t|\mathcal{F}_{t-1}]$
*b) X is a $(\mathcal{F}, P)$-**supermartingale** if (iii) is replaced by $X_{t-1} \geq E[X_t|\mathcal{F}_{t-1}]$*
*c) X is a $(\mathcal{F}, P)$-**submartingale** if (iii) is replaced by $X_{t-1} \leq E[X_t|\mathcal{F}_{t-1}]$*

In this definition, we specify $(\mathcal{F}, P)$-martingale because being a martingale depends simultaneously on the filtration and the probability measure as shown in (i) to (iii).

In the following, we will simply write "martingale" when no confusion is possible or P-martingale when we want to specify the probability-measure.

Point iii) of the definition may also be written $E[X_t - X_{t-1}|\mathcal{F}_{t-1}] = 0$ because X is $\mathcal{F}$-adapted and X_{t-1} is $\mathcal{F}_{t-1}$-measurable (see (4) of proposition 4).

Remark 1 *Using the law of iterated expectations, it is obvious to see that if X is a martingale (supermartingale, submartingale) we have, for any pair (s,t), $s \leq t$:*

$$E\left[X_t \left|\mathcal{F}_s\right.\right] = (\leq, \geq)\ X_s$$

This relationship could also be used equivalently in point (iii) of definition 11.

Just look at $E\left[X_{t+1} \left|\mathcal{F}_{t-1}\right.\right]$. Using definition 11, we know that $X_{t-1} = E\left[X_t \left|\mathcal{F}_{t-1}\right.\right]$ and $X_t = E\left[X_{t+1} \left|\mathcal{F}_t\right.\right]$. It implies

$$X_{t-1} = E\left[E\left[X_{t+1} \left|\mathcal{F}_t\right.\right] \left|\mathcal{F}_{t-1}\right.\right] \tag{1.17}$$

Using now point (4) of proposition 4 leads to

$$X_{t-1} = E\left[X_{t+1} \left|\mathcal{F}_{t-1}\right.\right] \tag{1.18}$$

because $\mathcal{F}_{t-1} \subset \mathcal{F}_{t+1}$.

The properties of conditional expectations show that, if X is a square integrable martingale, $E\left[X_t \left|\mathcal{F}_{t-1}\right.\right]$ is the best $\mathcal{F}_{t-1}$-measurable approximation of X_t (in the OLS sense). Consequently, for a martingale X, X_{t-1} is the best approximation of X_t conditioned on the information known at date $t-1$. In geometrical terms, X_{t-1} is the orthogonal projection of X_t on the subspace of $\mathcal{F}_{t-1}$-measurable variables.

This kind of process is then very well fitted to modelize "fair games" because $E(X_t)$ is constant[6].

The most standard examples of martingales are the random walk (for example the gambler's wealth in a repeated fair game) and the Doob martingale. They are described hereafter.

Definition 12 *A random walk is a stochastic process X such that $X_0 = c \in \mathbb{R}$ and:*

$$X_t = X_{t-1} + Y_t$$

where the variables Y_t are independent.

[6] Properties of conditional expectations lead to

$$E\left(E\left[X_t \left|\mathcal{F}_{t-1}\right.\right]\right) = E\left[X_t\right] = E\left[X_{t-1}\right]$$

Example 4 *A gambler starts a fair game at date 0 with a wealth X_0; his date-t wealth X_t is defined by:*

$$X_t = X_{t-1} + Y_t$$

where Y_t is the gain/loss of the t-th draw. The Y_s are independent and zero-mean (insuring the fairness of the game) random variables.

Equivalently, the date-t wealth can be written:

$$X_t = X_0 + \sum_{s=1}^{t} Y_s$$

Assume that $\mathcal{F}$ is the natural filtration of the process Y, that is $\mathcal{F}_s$ is the tribe generated by $Y_u, u \leq s$. As Y_s is the result of the s-th draw and X_s the gambler's wealth after s draws, X is $\mathcal{F}$-adapted and we get:

$$\begin{aligned} E\left[X_t \left| \mathcal{F}_{t-1} \right.\right] &= E\left[X_{t-1} + Y_t \left| \mathcal{F}_{t-1} \right.\right] \\ &= E\left[X_{t-1} \left| \mathcal{F}_{t-1} \right.\right] + E\left[Y_t \left| \mathcal{F}_{t-1} \right.\right] \end{aligned}$$

X being $\mathcal{F}$-adapted, the first term on the RHS is equal to X_{t-1}; moreover, the variables Y_t are independent of each other, implying that Y_t is independent of $\mathcal{F}_{t-1}$. It follows that $E\left[Y_t \left| \mathcal{F}_{t-1} \right.\right] = E\left[Y_t\right] = 0 \Rightarrow E\left[X_t \left| \mathcal{F}_{t-1} \right.\right] = X_{t-1}$, then proving that X is a martingale.

Example 5 ***Doob's martingale***

Let $Y \in L^1(\Omega, \mathcal{A}, P)$*; define the stochastic process* X *as:*

$$X_t = E[Y|\mathcal{F}_t]$$

X *is obviously a martingale. Using the law of iterated expectations, it follows immediately*

$$\begin{aligned} E[X_t|\mathcal{F}_{t-1}] &= E[E[Y|\mathcal{F}_t]|\mathcal{F}_{t-1}] \\ &= E[Y|\mathcal{F}_{t-1}] = X_{t-1} \end{aligned}$$

However, it is worth to mention that a process with a constant expectation ($E[X_t]$ independent of t) is not always a martingale. The following example illustrates this point

Example 6 *Consider an urn containing an even number of balls* T*, half of them being white, the other half being black. At each date* $t \leq T$ *a ball is randomly drawn without replacement. Let* Y *(resp.* Z*) the stochastic process counting the number of white (black) balls having been drawn. The relevant filtration is the natural filtration of* Y *(or equivalently* Z*).*

Let $X_t = Y_t - Z_t$ *the difference between the number of white and black balls after* t *draws. The stochastic process satisfies* $X_0 = 0$ *and we can write:*

$$X_t = X_{t-1} + \delta_t$$

with $\delta_t = 1$ *if the* t*-th ball is white and* $\delta_t = -1$ *if it is black. Consequently:*

$$E[X_t|\mathcal{F}_{t-1}] = X_{t-1} + E[\delta_t|\mathcal{F}_{t-1}]$$

The essential difference between this process and the random walk is that δ_t *is no more independent of* $\mathcal{F}_{t-1}$*. In fact, assume that* $Y_t = s$ *and* $Z_t = t - s$*. We then get:*

$$P(\delta_{t+1} = 1|Y_t = s) = \frac{\frac{T}{2} - s}{T - t}$$

This quantity is different from $\frac{1}{2}$ *as soon as* $s \neq \frac{t}{2}$*, and in this case,* $E[\delta_{t+1}|Y_t = s] \neq 0$*.*

It shows that X *is not a martingale. However, for any* t*,* $E(Y_t) = E(Z_t) = \frac{t}{2}$ *leading to* $E[X_t] = 0$*.* X *is then an example of a constant mean stochastic process which is not a martingale. Obviously, this feature comes from* *the "without replacement" characteristic of the draws. After each draw, the probability of drawing a white (black) ball in the next draw is changed. But seen from date 0, it is a fair game since the mean number of white(black) balls drawn between* 0 *and* t *is equal to* $\frac{t}{2}$*.*

It is worth to note that the "problem" comes from the fact that the terminal value of Y *and* Z *is perfectly known at date 0. In fact, before the first draw, one already knows that* $Y_T = Z_T = \frac{T}{2}$ *implying* $X_T = 0$*. In technical terms, these three variables are* $\mathcal{F}_0$*-measurable.*

1.5.1 Doob decomposition of an adapted process

The result stated hereafter shows that martingales come naturally in the description of any adapted process.

Proposition 5 *Let* X *be a stochastic process adapted to* $\mathcal{F}$*, each* X_t *being integrable.* X *may be decomposed in the following way:*

$$\forall t \in \mathcal{T}, \ \ X_t = X_0 + M_t + A_t \tag{1.19}$$

where M *is a martingale satisfying* $M_0 = 0$ *and* A *is a predictable process such that* $A_0 = 0$*. If* X *is a sub-martingale,* A *is increasing* $(A_t \leq A_{t+1})$

Proof. If $X_t = X_0 + M_t + A_t$, we can write:

$$E\left[X_t - X_{t-1} \left| \mathcal{F}_{t-1}\right.\right] = E\left[M_t - M_{t-1} \left| \mathcal{F}_{t-1}\right.\right] + E\left[A_t - A_{t-1} \left| \mathcal{F}_{t-1}\right.\right]$$

As M is a martingale, the first term on the RHS is 0. A being predictable, the second term is equal to $A_t - A_{t-1}$. Summing over the time-index leads to:

$$A_t = \sum_{s=1}^{t} E\left[X_s - X_{s-1} \left| \mathcal{F}_{s-1}\right.\right]$$

(We check on this equation that if X is a sub-martingale, A is increasing).

Defining M as follows:

$$M_t = X_t - X_0 - \sum_{s=1}^{t} E\left[X_s - X_{s-1} \left| \mathcal{F}_{s-1}\right.\right]$$

gives the desired result. ∎

The Doob decomposition 2.6 has a simple economic interpretation. Let X_t be the cumulative return on a financial asset between 0 and t. The decomposition means that in each period $[s; s+1]$, the return has two components: a predictible one and a zero-mean "surprise" (the martingale part). In fact:

$$X_t - X_{t-1} = M_t - M_{t-1} + A_t - A_{t-1}$$

It leads to:

$$E\left[X_t - X_{t-1} \left| \mathcal{F}_{t-1}\right.\right] = E\left[A_t - A_{t-1} \left| \mathcal{F}_{t-1}\right.\right] = A_t - A_{t-1}$$

In an economy populated by risk-neutral agents, $A_t - A_{t-1}$ would be the variation linked to the (locally) risk-free rate.

1.5.2 Martingales and self-financing strategies

Consider a market where K stocks are traded. Their date-t prices are denoted as $X_t = (X_t^1, ..., X_t^K)$. They are measured in units of a **numéraire**. Assume that the price processes X^k are martingales and denote $\theta_t = (\theta_t^1, ..., \theta_t^K)$ the portfolio held by an agent between $t-1$ and t. The date-t portfolio value is:

$$V_t(\theta) = \sum_{k=1}^{K} \theta_t^k X_t^k$$

To be able to compare $V_t(\theta)$ and $V_0(\theta)$ in a meaningful way, it is necessary that no additional funds are invested and no funds withdrawn at intermediate dates s, $0 < s < t$. Moreover, prices at different dates have to be measured in the same unit (called the numéraire[7]). This remark allows us to introduce the notion of self-financing (or self-financed) strategy. But first consider the example described in table 1.3. There are three dates and two stocks. The initial endowment of the investor is 5 units of stock 1 and 10 units of stock 2. The initial value of the position is then $5 \times \$100 + 10 \times \$60 = \$1100$.

At date 1, before transactions, the investor still holds the same quantities and the value of the portfolio is $5 \times \$130 + 10 \times \$65 = \$1300$. After transactions, we see in the table that he holds 6 units of stock 1 and 8 units of stock 2.

[7] In most common cases, these prices are measured in date-0 or date-T monetary units depending on the choice between discounting and capitalization.

	Quantities		Prices (in $)		Position	
Date	Stock1	Stock2	Stock1	Stock2	Value (in $)	
0	5	10	100	60	1100	
1	6	8	130	65	1300	
2	4	11	120	80	1360	

Table 1.3: Self-financing strategy S

It means that he bought one unit of stock 1 for a cost of 130 and sold 2 units of stock 2, then receiving 130. It turns out that the purchase of stock 1 is exactly financed by the sale of stock 2. Obviously, when calculating the portfolio value at date 1 after transactions, we get $6 \times \$130 + 8 \times \$65 = \$1300$, that is the same value.

This is the intuition of a "self-financing" strategy. If you want to hold more units of stock 1, you have to sell the equivalent amount of stock 2. We let the reader check that, at date 2, the strategy is still self-financing.

Definition 13 *a) A* ***portfolio strategy*** *is a bounded predictable process θ.*
b) A strategy θ is ***self-financing*** *if, at any date t:*

$$\sum_{k=1}^{K} \theta_t^k X_t^k = \sum_{k=1}^{K} \theta_{t+1}^k X_t^k \tag{1.20}$$

The LHS is equal to the date-t liquidation value of the portfolio θ_t. The RHS of equation 1.20 is the cost of the new portfolio θ_{t+1} built at date t. Therefore, the equality means that funds are neither added to, nor withdrawn from, the strategy. On a practical point of view, it means that buying stocks at an intermediate date must be financed by selling other stocks already in the portfolio.

The following proposition shows that if price processes are martingales, the value process of any self-financing strategy is also a martingale.

It means that you cannot beat the market by picking stocks when price processes are martingales, or equivalently, that you cannot transform a fair game into a favorable game.

Proposition 6 *Let $X = \left(X^1, ..., X^K\right)$ be a martingale taking values in $\mathbb{R}^K$ and $\theta = \left(\theta^1, ..., \theta^K\right)$ a self-financing strategy; the process Y defined by:*

$$\begin{aligned} Y_0 &= \sum_{k=1}^{K} \theta_1^k X_0^k \\ Y_t &= \sum_{k=1}^{K} \theta_t^k X_t^k \end{aligned}$$

is a martingale.

Proof. Remark first that Y_t is the date-t value of the strategy θ, Y_0 being the initial cost of the portfolio

$$E\left[Y_t - Y_{t-1} \left| \mathcal{F}_{t-1}\right.\right] = E\left[\sum_{k=1}^{K} \left(\theta_t^k X_t^k - \theta_{t-1}^k X_{t-1}^k\right) \left| \mathcal{F}_{t-1}\right.\right]$$

The self-financing hypothesis allows to write:

$$E\left[Y_t - Y_{t-1} \left| \mathcal{F}_{t-1}\right.\right] = E\left[\sum_{k=1}^{K} \theta_t^k \left(X_t^k - X_{t-1}^k\right) \left| \mathcal{F}_{t-1}\right.\right]$$

because $\sum_{k=1}^{K} \theta_t^k X_t^k = \sum_{k=1}^{K} \theta_{t+1}^k X_t^k$.

As θ is predictable, θ_t can be put outside the conditional expectation in the last term. It is the same for the summation term since the expectation operator is linear. Consequently:

$$E\left[Y_t - Y_{t-1}\left|\mathcal{F}_{t-1}\right.\right] = \sum_{k=1}^{K} \theta_t^k E\left[\left(X_t^k - X_{t-1}^k\right)\left|\mathcal{F}_{t-1}\right.\right]$$

But X is a martingale, so the conditional expectations on the RHS are equal to 0. We finally get:

$$E\left[Y_t - Y_{t-1}\left|\mathcal{F}_{t-1}\right.\right] = 0$$

or equivalently:

$$E\left[Y_t\left|\mathcal{F}_{t-1}\right.\right] = Y_{t-1}$$

■

Proposition 6 has a "financial smell" but it is a consequence of the following mathematical result.

Proposition 7 *Let* $X = \left(X^1, ..., X^K\right)$ *a martingale taking values in* $\mathbb{R}^K$ *and* $\theta = \left(\theta^1, ..., \theta^K\right)$ *a bounded predictable process; the stochastic process* Z *defined by:*

$$\begin{aligned} Z_0 &= 0 \\ Z_t &= \sum_{k=1}^{K} \sum_{s=1}^{t} \theta_s^k \left(X_s^k - X_{s-1}^k\right) \end{aligned}$$

is a martingale.

Proof. It is sufficient to write:

$$E\left[Z_t - Z_{t-1}\left|\mathcal{F}_{t-1}\right.\right] = E\left[\sum_{k=1}^{K} \theta_t^k \left(X_t^k - X_{t-1}^k\right)\left|\mathcal{F}_{t-1}\right.\right]$$

The same arguments as in proposition 6 can be used to get:

$$E\left[Z_t - Z_{t-1}\left|\mathcal{F}_{t-1}\right.\right] = \sum_{k=1}^{K} \theta_t^k E\left[\left(X_t^k - X_{t-1}^k\right)\left|\mathcal{F}_{t-1}\right.\right]$$

and the conditional expectation is 0 because X is a martingale. ■

The definition of Z in proposition 7 gives an idea of what will be the stochastic integral described in chapter 3. To show the analogy with the usual Stieltjes integral, remember the way expectations of continuous variables are defined. Let Y be a random variable with density f_Y and CDF F_Y.

$$E(Y) = \int_{-\infty}^{+\infty} y\ f_Y(y)\ dy = \int_{-\infty}^{+\infty} y\ dF_Y(y)$$

The expectation is in fact the Stieltjes integral of the identity function with respect to the CDF. This formulation is linked to the preceding chapter by writing $E(Y) = E(Y\mathbf{1}_\Omega) = \langle Y, \mathbf{1}_\Omega \rangle_{L^2}$. Consequently, $\mathbf{1}_\Omega$ represents (as in the Riesz theorem) the expectation operator which is linear.

Suppose now that Y is a discrete random variable taking values $y_1 < y_2, ... < y_n$; we then have:

$$E(Y) = \sum y_i \left(F_Y(y_i) - F_Y(y_{i-1})\right) = \int_{-\infty}^{+\infty} y\ dF_Y(y)$$

One can note that this formulation is close to the one used in the definition of Z which could be written:

$$Z_t = \sum_{k=1}^{K} \int_0^t \theta_s^k dX_s^k$$

Each integral $\int_0^t \theta_s^k dX_s^k$ is the gain/loss of strategy θ on asset k between 0 and t. Without entering the details now, there is an essential difference between $E(X)$ and Z_t; $E(X)$ is a real number but Z_t is a random variable. Moreover the CDF of Y is increasing but there is no reason for X^k to be increasing. These differences are "heavy" and impose some precautions in defining the stochastic integral.

1.5.3 Investment strategies and stopping times

A second important question concerning investment strategies is the following: is there a way to select optimally the liquidation date of the portfolio? The intuition is simple. If a position is liquidated as soon as it generates a positive profit, it is surely a gain. The answer to the question is yes if the time-horizon of the investor is infinite and if he can bear any level of intermediate losses! In other cases (that is on real markets populated by real agents), the answer is no.

The most simple example is a heads or tails game where 1$ is won (tails) or lost (heads) each time the (fair) coin is flipped. Successive results are assumed independent.

Let X_t be the gambler's wealth at date t; it is defined by:

$$X_t = X_0 + \sum_{s=1}^{t} Y_s$$

where the Y_s are random variables taking values 1 and -1 with equal probabilities (corresponding to gain or losses), X_0 being the initial wealth of the gambler.

Consider the strategy in which the player stops gambling at the first date t such that $X_t = X_0 + 1$. This strategy is clearly a winning one because the gambler ends the game with $X_0 + 1$ with probability 1. However, to get such a result, any intermediate loss must be acceptable and an infinite time-horizon is necessary. If the gambler knows he has to stop after, say, T games, he cannot be sure to end with a profit.

These two conditions don't seem reasonable (especially on financial markets). A finite horizon and bounded intermediate possible losses prevent to transform a fair game into a favorable game, as we will see now. The notion of stopping time arises naturally to formalize this kind of situation.

Definition 14 *A **stopping time** is a random variable v defined on $(\Omega, \mathcal{A}, P)$ taking values in $\mathcal{T}$ such that:*

$$\forall t \in \mathcal{T} \quad \{v = t\} = \{\omega \in \Omega \text{ such that } v(\omega) = t\} \in \mathcal{F}_t$$

This definition may, at a first glance, seem very abstract but it is easily interpretable if ν is seen as a date at which a decision must be taken (typically, liquidating a portfolio for example). With this view in mind, $\{v = t\} \in \mathcal{F}_t$ simply means that liquidating the portfolio at date t can be decided only using past and present information. In other words, saying that the liquidation date is a stopping time means that you don't read in a crystal ball (revealing the future) before taking the decision.

To give more intuition, assume that you send the following order to your broker: "buy 100 IBM stocks at the minimum price between now and the end of the month". The execution date is a random variable but to know that the price on the 25th day of the month reaches the minimum, you must know the prices on the following days! Consequently, the execution date is not a stopping time (in technical words it is $\mathcal{F}_T$-measurable, not $\mathcal{F}_t$-measurable). Obviously, no broker would accept such an order.

Remark that in discrete-time models with a finite Ω, it is equivalent to define a stopping time by replacing $\{v = t\}$ in definition 14 by $\{v \leq t\}$, or equivalently by $\{v > t\}$. The reason comes from the definition of a filtration. As $\mathcal{F}_{t-1} \subset \mathcal{F}_t$, $\{v = t-1\} \in \mathcal{F}_{t-1}$ and so $\{v = t-1\} \in \mathcal{F}_t$. This reasoning can be done with any date $t-k$ with $k < t$. Finally, as $\mathcal{F}_t$ is a tribe, $\{v \leq t\} \in \mathcal{F}_t \Rightarrow \{v \leq t\}^c = \{v > t\} \in \mathcal{F}_t$.

The Doob's **optional stopping theorem** answers the question asked before concerning the existence of an optimal liquidation date.

Proposition 8 *Let v be a stopping time and X a martingale; if one of the following conditions is satisfied, then X_v is integrable and $E(X_v) = E(X_0)$.*

1) v is bounded

2) X is bounded, that is there exists K such that for every $t \in \mathcal{T}$ and every $\omega \in \Omega$, $|X_t(\omega)| \leq K$.

This proposition is interpreted as follows. Point (1) says that if the number of draws is finite, the expected gain is zero. Point (2) means that, if X_t

denotes the date-t wealth of the agent, resulting from investments in financial assets when prices are martingales, the expected wealth at the liquidation date ν is equal to the initial wealth, as soon as wealth is bounded in absolute value at any date. In other words, if you are not ready to bear arbitrary large intermediate losses, you cannot beat the market, even by choosing intelligently a liquidation date. It is interesting to note that a purely mathematical result has such an intuitive financial interpretation.

Obviously, in the abovementioned heads and tails game, conditions (1) and (2) are not satisfied if the game is pursued as long as a gain of 1 unit is not reached.

To illustrate the point, define the stochastic process Y by:

$$Y_t = \begin{cases} X_t \text{ if } t < v \\ X_v \text{ if } t \geq v \end{cases}$$

Y is called the **stopped process** of X, sometimes denoted X^v where v is a bounded stopping time. The following result is then obtained.

Proposition 9 *If ν is a bounded stopping time and X a stochastic process adapted to $\mathcal{F}$, the stopped process Y is also adapted to $\mathcal{F}$. Moreover, if X is a martingale (super-martingale), it is also the case for Y.*

Proof. Let $\xi_s = \mathbf{1}_{\{v \geq s\}}$; this variable is valued 1 when $\{v \geq s\}$ occurs and 0 otherwise. Y can be decomposed as follows:

$$Y_t = X_0 + \sum_{s=1}^{t} \xi_s \left(X_s - X_{s-1}\right)$$

In fact, if $t < v$, the indicator functions in the sum are equal to 1 and, consequently $Y_t = X_t$. On the contrary, if $t > v$ the indicator functions are equal to 1 for $s \leq v$ and 0 beyond. We then get as expected $Y_t = X_v$. As $\{v \geq t\} = \{v \leq t-1\}^c$ it follows that the process ξ is predictable implying that Y is adapted.

Moreover, if X is a martingale (super-martingale), Y is also a martingale because ξ is predictable.In fact we can directly write:

$$\begin{aligned} E\left[Y_t - Y_{t-1} \left|\mathcal{F}_t\right.\right] &= E\left[\xi_t \left(X_t - X_{t-1}\right) \left|\mathcal{F}_t\right.\right] && (1.21) \\ &= \xi_t E\left[\left(X_t - X_{t-1}\right) \left|\mathcal{F}_t\right.\right] && (1.22) \end{aligned}$$

X is a martingale, therefore

$$E\left[\left(X_t - X_{t-1}\right)|\mathcal{F}_t\right] = 0 \tag{1.23}$$

This proof in fact uses some of the properties of conditional expectations recalled in proposition 4. ■

Example 7 ***A doubling strategy***

A little bit more sphisticated strategy is to double the stake after each losing draw and to stop the game after the first win. Suppose, to keep things simple, that the gambler starts with $X_0 = 0$ and bets one unit on the first draw.

If t unfavorable draws occur in a row, the gambler's wealth is:

$$X_t = -\sum_{s=1}^{t} 2^{s-1} = -\sum_{s=0}^{t-1} 2^s$$

We recognize the sum of the t first terms of a geometric sequence. We then get:

$$X_t = -\frac{2^t - 1}{2 - 1} = -(2^t - 1)$$

The stake for the $t+1$-th draw is 2^t. In case of winning, the final wealth is then $X_{t+1} = 1 = X_0 + 1$.

Observe that the expectation of the stopping time $v = \inf\{t \ / \ X_t = 1\}$ is finite because the mean number of draws is:

$$E(v) = \sum_{s=1}^{+\infty} \left(\frac{1}{2^s}\right) s$$

It can be rewritten as:

$$\sum_{s=1}^{+\infty} \left(\frac{1}{2^s}\right) s = \sum_{s=1}^{+\infty} \sum_{u=s}^{+\infty} \left(\frac{1}{2^u}\right)$$

But $\sum_{u=s}^{+\infty} \left(\frac{1}{2^u}\right) = \frac{1}{2^{s-1}}$ implies:

$$E(v) = \sum_{s=1}^{+\infty} \frac{1}{2^s} = 2$$

However, v does not satisfy conditions of theorem 8 because it is not almost surely bounded. In the same way, X is not almost surely bounded because the doubling strategy can generate incredibly high intermediate losses (if you are especially unlucky!), a situation that cannot be easily dealt with on real financial markets[8].

1.5.4 Stopping times and American options

An American (European) put option with maturity T and strike price K is a contract giving the right to his holder to sell a given underlying asset like a stock, an index or a currency, at a given price K, at any date before T (at date T).

Denote Y_t the payoff received by the holder if he exercises the put at date t.We get $Y_t = \max(K - S_t; 0)$, S_t being the date-t price of the underlying asset. Obviously, we assume that investors never exercise their option at date t when $K \leq S_t$. It would be irrational to sell the underlying asset for K when you can sell it for a higher price on the market.

[8] Think to the cases of LTCM, Lehman Brothers, etc.

Moreover, the decision to exercise early is taken by comparing $K - S_t$ to the value of the option if not exercised. To value this kind of option, the notion of "Snell envelope" of a stochastic process is very useful.

Definition 15 *Let Y be a stochastic process adapted to $\mathcal{F}$; the process X defined by :*

$$\begin{cases} X_T = Y_T \\ X_t = \max\left(Y_t; E\left(X_{t+1} \left| \mathcal{F}_t \right.\right)\right) \;\; si\ t < T \end{cases}$$

is called the Snell envelope of Y.

We then get the following proposition.

Proposition 10 *The Snell envelope X of Y is the smallest super-martingale greater than Y, that is satisfying $X_t \geq Y_t$ for any t.*

Proof. From the definition it follows immediately $X_t \geq E\left(X_{t+1} \left| \mathcal{F}_t \right.\right)$; X is then a super-martingale. Let Z be a super-martingale greater than Y. We are going to show that $\forall t,\ Z_t \geq X_t$. This relationship is true for $t = T$ because of the definition of X. Let us use backward induction by assuming that $Z_s \geq X_s$ for $s \geq t_0$ and proving that $Z_{t_0-1} \geq X_{t_0-1}$.

As Z is a super-martingale, we can write :

$$Z_{t_0-1} \geq E\left(Z_{t_0} \left| \mathcal{F}_{t_0-1} \right.\right) \geq E\left(X_{t_0} \left| \mathcal{F}_{t_0-1} \right.\right) \tag{1.24}$$

The second inequality comes from the recurrence assumption. Moreover Z is greater than Y; consequently:

$$Z_{t_0-1} \geq Y_{t_0-1} \tag{1.25}$$

Inequalities 1.24 and 1.25 imply:

$$Z_{t_0-1} \geq \max\left(Y_{t_0-1}; E\left(X_{t_0} \left| \mathcal{F}_{t_0-1} \right.\right)\right) = X_{t_0-1}$$

Using backward induction, we get the same result for any t.

■

The optimal exercise date of the American put will be the first date t at which $X_t = Y_t$.

Proposition 11 *Let v the random variable defined by:*

$$v = \inf\left(t \ / \ X_t = Y_t\right)$$

v is a stopping time and the stopped process X^v is a martingale.

Proof. The definition of X implies $v \leq T$. We can also write:

$$\{v = s\} = \bigcap_{u=1}^{s-1} \{X_u > Y_u\} \bigcap \{X_s = Y_s\}$$

Each of the events appearing on the RHS is in $\mathcal{F}_s$ since the variables $X_u, Y_u,\ u \leq s$ are $\mathcal{F}_s$-measurable; consequently $\{v = s\} \in \mathcal{F}_s$ and v is a stopping time.

X_t^v may be decomposed using the variables $\xi_s = \mathbf{1}_{\{v \geq s\}}$, defined in proposition 9, as follows :

$$X_t^v = X_0 + \sum_{s=1}^{t} \xi_s \left(X_s - X_{s-1}\right)$$

Using $X_s = \max\left(Y_s; E\left(X_{s+1} \left|\mathcal{F}_s\right.\right)\right)$ we deduce $X_s^v(\omega) = X_s(\omega)$ for $\omega \in \{v \geq s\}$, which leads to:

$$X_s(\omega) = E\left(X_{s+1} \left|\mathcal{F}_s\right.\right)(\omega) \text{ pour } \omega \in \{v \geq s\}$$

It is then sufficient to prove that $X_s^v - X_{s-1}^v = \xi_s \ \left(X_s - E\left(X_s \left|\mathcal{F}_{s-1}\right.\right)\right)$. But if $\{v \geq s\}$, the LHS is equal to $X_s - X_{s-1}$ and the RHS too by definition of X_s. On the event $\{v < s\}$, the LHS is 0 because $X_s^v = X_{s-1}^v = Y_v$. It is also the case for the RHS because the indicator function is 0 on the event $\{v < s\}$.

We finally get:

$$E\left(X_s^v - X_{s-1}^v \left|\mathcal{F}_{s-1}\right.\right) = E\left(\xi_s \left(X_s - E\left(X_s \left|\mathcal{F}_{s-1}\right.\right)\right) \left|\mathcal{F}_{s-1}\right.\right)$$

ξ being predictable, ξ_s goes out the conditional expectation which is then equal to 0. We have shown that:

$$E\left(X_s^v - X_{s-1}^v \left|\mathcal{F}_{s-1}\right.\right) = 0$$

that is X^v is a martingale. ■

To conclude this section, the following proposition shows that the stopping time v is the solution of the early exercise problem for American put options.

Proposition 12 *In*[9] *the set of $\mathcal{T}$-valued stopping times, denoted as $\mathbb{T}$, the stopping time v defined by:*

$$v = \inf \{t \ / \ \ X_t = Y_t\}$$

solves the optimization problem:

$$E[Y_v] = \sup_{u \in \mathbb{T}} E[Y_u]$$

The interpretation of this result for American options is clear. The random variable $Y_t = \max(K - S_t; 0)$ is called the **intrinsic value** of the option at date t. The difference $X_t - Y_t$ is called the speculative value of the put which is always positive or equal to 0. The above proposition means that it is optimal to exercise an American put as soon as its speculative value is 0. When early exercise is done at the optimal date, the price process of the put option (in numéraire units) is a martingale.

[9] For the proof, see for example Bingham-Kiesel (1998),p 80.

Chapter 2

Continuous-time stochastic processes

2.1 Introduction

The discrete-time processes presented in chapter 1 are easy to understand because they are managed without complex mathematical tools. Moreover, when the number of states of nature is finite, elementary algebraic properties can be used to make easier the understanding. The counterpart is that no "beautiful" analytical formula is obtained, and sometimes it may be difficult to take into account the way real markets work (in continuous time).

The price to pay to enter the continuous-time world seems large to many students who do not have a mathematical background. All vector spaces are infinite dimensional, price paths are functions of time with non intuitive properties. For example, they may be everywhere continuous and nowhere differentiable.

In this chapter, we first present the general definition of a continuous-time process which is not much different of the equivalent notion in discrete-time. We then describe the properties of the paths commonly encountered in the financial literature, and the concepts of filtration, adapted process and predictable process as well.

The next section is devoted to Markov and diffusion/Itô processes. The definition of Markov processes in continuous-time is the same as the one given in discrete-time, except that the set of dates is different. Diffusion and Itô processes introduce more conditions on the paths.

The most important stochastic process in financial models is the Brownian motion (also called Wiener process). We hope that our presentation will look intuitive to the reader. The final section studies the fundamental properties of the Brownian motion, especially those linked to martingales, very useful in finance.

2.2 General framework

The basic notations are the same as in the preceding chapter. $(\Omega, \mathcal{A}, P)$ is a probability space. The set of relevant dates is still denoted $\mathcal{T}$ but now $\mathcal{T} = [0; T]$, $T < +\infty$. It means that markets are open in continuous-time.

Definition 16 *A stochastic process is a family of random variables* $X = (X_t, t \in \mathcal{T})$ *taking values in* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

As mentioned in the introduction, the definition is identical to the one given in discrete-time but the set of indices $\mathcal{T}$ is different. In particular, it is not countable.

Definition 17 *The mapping* $t \to X_t(\omega)$ *for a given* $\omega \in \Omega$ *is called a path (or trajectory) of the process* X.

In discrete-time, a path of a stochastic process is simply a sequence of values taken by the process on the (discrete) set of dates. Here a path is a function defined on an interval $[0; T]$ of the real line.

Three kinds of paths are usually encountered in financial models.

- The description of prices or interest rates dynamics commonly assumes that paths are **continuous** functions. It is the case in the Black-Scholes (1973) option pricing model for the dynamics of the underlying asset of the option contract. It is also the assumption for the short-term rate in the models of Vasicek (1977), Cox-Ingersoll-Ross (1985) or in the model of Heath-Jarrow-Morton (1992) for forward rates.

Figure 2.1 illustrates this continuity assumption. It represents the evolution of daily closing values of the S&P500 index from 1950 to May 2010[1] (more than 12 000 points). We observe that the path of the index is represented as a continuous function, even if it seems to have many points where the function $t \to X_t(\omega)$ is not differentiable.

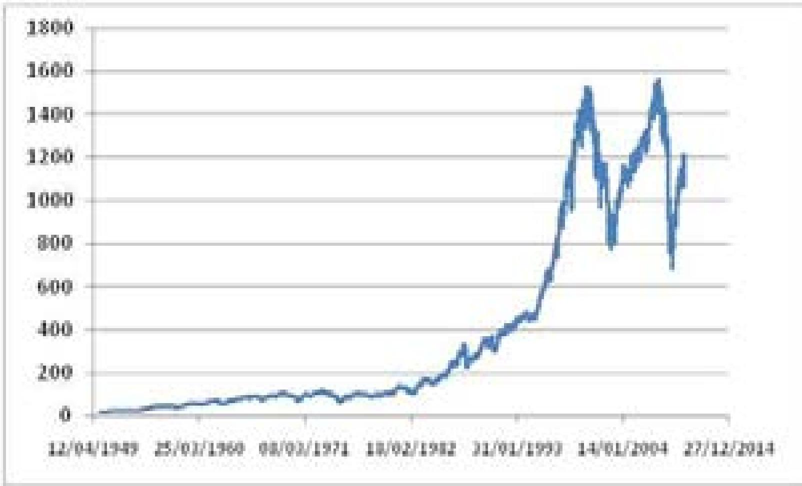

Figure 2.1: S&P500 daily values from 1950 to 2010

- In some cases, it is assumed that paths are **càdlàg**, meaning (in French[2]) "continues à droite et possédant une limite à gauche". The English translation is **right-continuous with left limit** (RCLL). In fact, continuous paths are not always the right tool to describe some economic variables, especially when jumps are possible. For example, it is now well known that stock returns are not normally distributed because extreme returns occur much more frequently than what is predicted by the Gaussian distribution hypothesis.

  Introducing discontinuities in paths may be a way to describe this phenomenon (see Merton (1976), Cox-Ross (1976), etc...), for example by

[1]Dates are in the format JJ/MM/YY on the graph

[2]We provide the French expression because the acronym càdlàg has become standard in many books or papers, even in English ones.

using Poisson processes (see definition 22). Figure 2.2 shows the daily returns of the S&P500 index in October 2008. We observe a continuous curve simply because successive values have been joined, but it is largely artificial. Daily variations are very large (around 10% in absolute value on some days[3]) and it could be more relevant to take jumps into account to represent the evolution of the index.

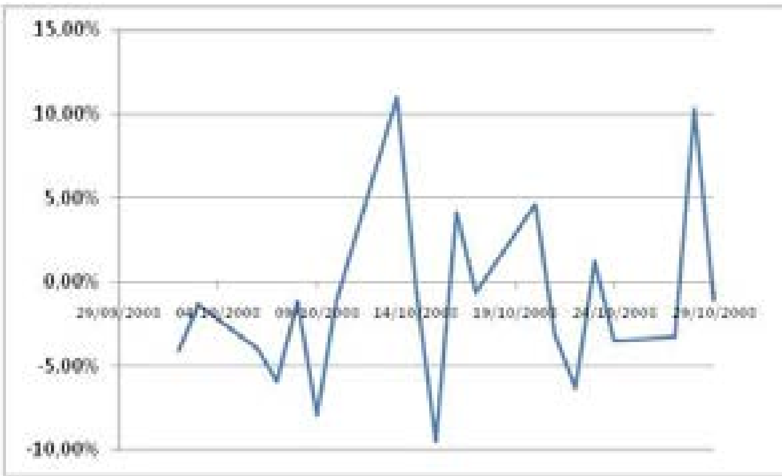

Figure 2.2: S&P500 daily returns in October 2008

[3] October 2008 was a "crazy" month, partly because it followed the bankruptcy of Lehmann Brothers in September. Many other difficult periods were to follow.

- Paths of some stochastic processes may also be **càglàd** ("continues à gauche et possédant une limite à droite"), that is **left-continuous with right limit** (LCRL). This assumption is relevant to describe the quantity of a given asset held by an investor. Obviously, portfolio rebalancing occurs in discrete-time. Choosing left continuity instead of right continuity is based on the following idea. When a path is left-continuous, $X_t(\omega)$ is known as soon as the $X_s(\omega), s < t$ are known. In other words, if you know the composition of your portfolio at any past instant, you also know it now. Remember what we wrote in the preceding chapter about predictable processes for quantities. The idea is the same here and we will come back to this point later on.

Continuous-time stochastic processes are a little bit more complicated when the state-space is not countable. In fact, there may be non-empty but negligible subsets of states of nature and also subsets of dates with Lebesgue measure zero on $\mathcal{T}$. The following definition is then useful to manage this technical problem.

Definition 18 *1. A process Y is called a* ***modification*** *of X if for any $t \in \mathcal{T}$, the set $\Omega_t = \{X_t = Y_t\}$ has probability 1. Y is also called a* ***version*** *of X.*

2. Two processes X and Y are ***indistinguishable*** *if they have in common "almost all" their paths; more precisely:*

$$\exists \Omega^* \in \mathcal{A} \text{ such that } P(\Omega^*) = 1 \text{ and } \forall \omega \in \Omega^*, \forall t \in \mathcal{T}, \quad X_t(\omega) = Y_t(\omega)$$

These two notions, though seemingly close, are different. Let $\Omega = [0;1]\,, \mathcal{T} = [0;1]$ and X defined by $X_t = 0$ for any t and any ω. Let now Y be defined as:

$$\begin{aligned} Y_t(\omega) &= 1 \text{ if } t = \omega \\ &= 0 \text{ elsewhere} \end{aligned}$$

It is obvious that $P(\{X_t = Y_t\}) = 1$ when P is the Lebesgue measure on Ω. Y is then a modification of X. On the contrary, X and Y are not indistinguishable because $\bigcap_t \Omega_t = \emptyset$. We cannot find two identical paths!

We have here an illustration of the technical difficulties evoked in chapter 1 of "Probability for Finance" (Roger, 2010) when defining probability measures and tribes. The problem is that a countable intersection of sets of probability 1 is also a set of probability 1. It is not true when the intersection under consideration is uncountable.

These are only technical difficulties but we have to mention them to be rigorous. However, in financial models, we are rarely concerned with such questions when dealing with real life problems, simply because continuous-time models are an idealization of what is really happening on financial markets.

2.2.1 Filtrations, adapted and predictable processes

The notion of filtration is still relevant in continuous-time models but some more conditions are needed to ensure tractability.

Definition 19

1. *A* ***filtration*** *on Ω is an increasing family $\mathcal{F} = \{\mathcal{F}_t, t \in \mathcal{T}\}$ of sub-tribes of $\mathcal{A}$. The quadruple $(\Omega, \mathcal{A}, \mathcal{F}, P)$ is called a* ***filtered probability space****.*
2. *A filtration $\mathcal{F}$ is* ***right-continuous*** *if for any $t < T$, $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.*
3. *A filtration $\mathcal{F}$ is* ***complete*** *is any tribe $\mathcal{F}_t$ contains all negligible events.*
4. *A process X is* ***adapted*** *to a filtration $\mathcal{F}$ if for any t, X_t is $\mathcal{F}_t$-measurable.*
5. *The* ***natural filtration*** *of a process X, denoted as $\mathcal{F}^X$, is the smallest tribe with respect to which X is adapted. $\mathcal{F}_t^X$ is the tribe generated by the variables $X_s, s \leq t$.*

When Ω is finite, it is generally assumed that $P(\omega) > 0$ for any state ω. When Ω is uncountable, assuming complete tribes is a convenient assumption for purely technical reasons. Without this assumption, we could imagine that some set $B \in \mathcal{F}_t$ satisfies $P(B) = 0$ but a subset A of B is not in $\mathcal{F}_t$. The definition of a probability measure implies that $P(A) \leq P(B)$ when $A \subset B$ but $P(A)$ would not be defined if A was not an event!

In what follows, we always assume that the filtrations we refer to are right-continuous and complete, even if this assumption is not recalled. These two assumptions are sometimes called (with some sense of humor) "usual conditions".

In discrete-time, a predictable process has an intuitive definition. A process is predictable when its date-t value is revealed at date $t-1$ or, more generally, when X_t is $\mathcal{F}_{t-1}$-measurable. It appears immediately that generalizing "X_t is $\mathcal{F}_{t-1}$-measurable" to continuous-time is not natural and may be quite involved.

To avoid the reader technical difficulties, we give a restricted (and then disputable) definition of predictability. It is sufficient for what we deal with in finance and is much more intuitive than the completely general definition.

Definition 20 *X is a **predictable** process if it is adapted to $\mathcal{F}$ and if its paths are left-continuous*[4]*.*

To give an intuition of the definition, consider a function $f : \mathbb{R} \to \mathbb{R}$ and suppose that you know all the values $f(x)$ for $x < x_0$. When is it sufficient to know $f(x_0)$?

Simply when f is left-continuous, because left-continuity means:

$$\lim_{\substack{x \to x_0 \\ x < x_0}} f(x) = f(x_0)$$

The notation combining $x \to x_0$ and $x < x_0$ is sometimes denoted $x \to x_0^-$.

In the same way, left-continuous paths for a stochastic process mean that if you know the values at any date $t < t_0$, you can infer the date-t value of the process. It corresponds pretty well to the intuitive notion of predictability.

Definition 21 *Let X be a process defined on $(\Omega, \mathcal{A}, P)$;*

1. *The increments of X are **independent** if for any $t_1 \leq t_2 \leq ... \leq t_n$, the variables $X_{t_j} - X_{t_{j-1}}$ are independent.*

2. *The increments of X are **stationnary** if for any $t \in \mathcal{T}$ and $h > 0$ such that $t + h \in \mathcal{T}$, the probability distribution of $X_{t+h} - X_t$ depends only on h (and not on t).*

[4] To be completely rigorous, we should consider the tribe $\mathcal{B}_*$ on $\Omega \times \mathcal{T}$ generated by the processes with left-continuous paths. Predictable processes are then the $\mathcal{B}_*$-measurable processes (see Duffie, 1988, p140).

Example 8 *The most simple process satisfying definition 21 in discrete-time is a random walk* X *defined by:*

$$\begin{aligned} X_0 &= 0 \\ X_t &= X_{t-1} + Y_t \end{aligned}$$

where the Y_t *are i.i.d. It is obvious in this case to check that increments are i.i.d because the increments of* X *are exactly the variables* Y_t*. In a next section, we will see that the Brownian motion is the typical example of such a process in continuous-time.*

When X_t *is the logarithm of a stock price (whose process is denoted* S*), the difference* $X_t - X_s$ *is the continuous return of the asset on the interval* $[s;t]$*. In fact,* $X_t = \ln(S_t)$ *and then*

$$X_t - X_s = \ln(S_t) - \ln(S_s) = \ln\left(\frac{S_t}{S_s}\right) \tag{2.1}$$

Assuming that X *has independent increments is then linked to the efficient market hypothesis (EMH). If current prices reflect all past and present information,* $X_t - X_s$ *doesn't depend on past returns. It is not useful to know past prices to infer the distribution of future prices, only the current price is important.*

The stationnarity hypothesis is often used (at least implicitly) in empirical studies, for example when the variance of returns is estimated using a time-series of daily returns.

A second important example of a process with independent increments is the Poisson process.

Definition 22 *A process X is a* ***Poisson process*** *with parameter λ if:*
a) $X_0 = 0$
b) X has independent increments
c) For any pair $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+$ with $s \leq t$, the variable $X_t - X_s$ follows a Poisson distribution with parameter $\lambda(t-s)$.

$$\forall k \in \mathbb{N}, P(X_t - X_s = k) = \frac{(\lambda(t-s))^k}{k!} \exp\left(-\left(\lambda(t-s)\right)\right) \tag{2.2}$$

It is worth to notice that the Poisson process is a continuous-time process but the state-space is discrete. The variables X_t takes values in the set of positive integers $\mathbb{N}$ which is countable.

These processes are commonly used in insurance to represent the arrival of damages; they are then linked to a process Y representing the amount of the claim. For example, at date t, the cumulated amount to be paid by the insurance company is written:

$$Claim_t = \sum_{i=1}^{X_t} Y_i \tag{2.3}$$

where X_t is the number of claims up to date t.

The other field where Poisson processes arise naturally is the microstructure of financial markets. The arrival of buy and sell orders on a stock market is often represented by such a process.

Equation 2.2 means that the probability of an increment greater than one on a very short interval of length h is negligible with respect to h. In fact

$$P(X_{t+h} - X_t = 2) = \frac{(\lambda h)^2}{2} \exp(-\lambda h) \simeq \frac{\lambda^2}{2} h^2 \tag{2.4}$$

At the same time, the probability of a unit increment on a short interval of length h is proportional to h and approximately equal to λh.

2.2.2 Markov and diffusion processes

In this section, we introduce several categories of continuous-time stochastic processes, namely Markov, diffusion and Itô processes. They cover almost all the processes used in financial models.

Markov processes

Definition 23 *A process X defined on $(\Omega, \mathcal{A}, P, \mathcal{F})$ is a **Markov process** if for any $(B_1, ..., B_n) \in \mathcal{B}^n_{\mathbb{R}}$ and any $(t_1, ..., t_n) \in \mathcal{T}^n$ such that $t_1 < .... < t_n$:*

$$P\left(X_{t_n} \in B_n \left| X_{t_j} \in B_j, j = 1, .., n-1\right.\right) = P\left(X_{t_n} \in B_n \left| X_{t_{n-1}} \in B_{n-1}\right.\right)$$

The financial interpretation of this definition is the same as the one given for independent increments. Economic information carried over by X_s for dates $s \leq t_{n-1}$ is the same as the one contained in $X_{t_{n-1}}$.The reader can recognize definition 5 given in chapter 1 but with a different set of dates.

Markov processes are sometimes called "no memory" processes because the path used to reach the set B_{n-1} at t_{n-1} has no influence on the probability distribution of the future variations $X_{t_n} - X_{t_{n-1}}$. Only the state reached at date t_{n-1} matters.

When stock prices are Markov processes, strategies based on technical analysis are useless. There was, and there still is, an intense debate in the financial literature to know if stock prices are Markovian. On one side, proponents of the efficient markets hypothesis believe in the Markovian character of stock prices. One of their arguments is that you cannot become "rich" by trading on past information. On the other side, numerous empirical studies show that some portfolios built on past information provide abnormal returns. To give a simple example, De Bondt and Thaler (1985) show that markets overreact. Buying past losers and short-selling past winners allow to get an abnormal return.

Diffusion processes and Itô processes

Definition 24 *a) A **diffusion process** is a Markov process with continuous paths.*

*b) A process X is an **Itô process** if X is a diffusion process and if there exist two functions μ and σ defined on $\mathbb{R}\times\mathcal{T}$ and taking values respectively in $\mathbb{R}$ and $\mathbb{R}^+$ defined by:*

$$\begin{aligned}
\mu(x,t) &= \lim_{h\to 0} \frac{E\left[X_{t+h} - X_t \left| X_t = x\right.\right]}{h} \\
\sigma^2(x,t) &= \lim_{h\to 0} \frac{V\left[X_{t+h} - X_t \left| X_t = x\right.\right]}{h}
\end{aligned}$$

To interpret μ and σ think of X_t as the logarithm of the date-t price of a stock. The increment $X_{t+h} - X_t$ is the return of the stock on the interval $[t; t+h]$. The function $\mu(x,t)$ is then the instantaneous expected return at date t when the process X is at level x. Dividing by h ensures that μ is expressed per time-unit. The most common time-unit in financial models is the year.

μ measures the direction of evolution of the process. It is generally called the **drift**. In the same way, σ^2 is the instantaneous variance of returns per unit of time. σ is called the **diffusion coefficient** of the process. In the financial literature, authors generally do not make a distinction between diffusion and Itô processes, the latter ones being almost always considered. As we will see later on, the dynamics of such processes may be written as a "stochastic differential" involving μ and σ.

Itô processes are pretty well suited for describing prices, returns or interest rates. Nevertheless, assuming continuous paths is sometimes a restriction, as mentioned before.

2.2.3 Martingales

As in discrete-time, martingales play an important role in continuous-time valuation models. The principles and interpretations are quite close and the definition of continuous-time martingales is almost the same as the definition in discrete-time.

Definition 25 *Let $(\Omega, \mathcal{A}, \mathcal{F}, P)$ a filtered probability space;*

1. *A $(\mathcal{F}, P)$-**martingale** is an integrable process X adapted to $\mathcal{F}$, satisfying:*
$$\forall (s,t) \in \mathcal{T}^2, s \leq t \Rightarrow E\left[X_t \left| \mathcal{F}_s\right.\right] = X_s \tag{2.5}$$

2. *X is a $(\mathcal{F}, P)$-**super-martingale** if equality 2.5 is replaced by $E\left[X_t \left| \mathcal{F}_s\right.\right] \leq X_s$.*

3. *X is a $(\mathcal{F}, P)$-**submartingale** if equality 2.5 is replaced by $E\left[X_t \left| \mathcal{F}_s\right.\right] \geq X_s$.*

We simply write "martingale" instead of $(\mathcal{F}, P)$-martingale when no confusion is possible about the filtration or the probability measure. When X is a martingale, one can always assume that paths are right-continuous, that is X is assumed *càdlàg*. Using the vocabulary describing the different possible paths, we could say that there exists a *càdlàg* version of the martingale X.

Some important results on discrete-time martingales are still valid in continuous time. It is especially the case for the Doob decomposition.

Proposition 13 *Let X be an integrable process adapted to $\mathcal{F}$; X can be decomposed as follows:*

$$\forall t \in \mathcal{T}, \ \ X_t = X_0 + M_t + A_t \tag{2.6}$$

where M is a martingale with $M_0 = 0$ and A is a predictable process such that $A_0 = 0$. If X is a submartingale, A is increasing.

The typical example of martingale in continuous-time is the Brownian motion presented in the next section.

2.3 The Brownian motion

2.3.1 Intuitive presentation

The Brownian motion, also called the Wiener process, is surely the most commonly used stochastic process, in finance and in other sciences as well. It is for the set of stochastic processes what is the Gaussian distribution for the set of probability distributions. It is the reason why we devote a few pages to provide an intuitive feeling of what is a Brownian motion[5].

For pedagogical reasons, we start with discrete-time random walks. Remember that a random walk is a stochastic process X defined by:

$$\begin{aligned} X_0 &= a \\ X_n &= X_{n-1} + Y_n \end{aligned}$$

where a is a constant and the Y_n are assumed i.i.d.

Let T be a given time horizon; $[0;T]$ is divided in N sub-periods of length $h = T/N$ and we look at what happens when N tends to infinity, T being fixed. The limit process (in a sense to be precised) will be a continuous-time process.

Then, for the moment we assume $\mathcal{T} = \{0, 1, ..., N\}$.

Proposition 14 *Let X be a random walk with the Y_n taking values σ and $-\sigma$ with equal probabilities $p = 1/2$*

[5] This section is largely based on Merton (1982). This paper was reedited in Merton's book (1990) entitled "Continuous-time Finance".

1. *X has independent and stationary increments*

2. $Cov(X_n, X_m) = \sigma^2 \min(n, m)$

Proof. Point 1 is obvious by definition of the Y_n. For point 2, suppose that $m < n$. We can write:

$$\begin{aligned} Cov(X_n, X_m) &= Cov(X_m + \sum_{k=m+1}^{n} Y_k, X_m) \\ &= Cov(X_m, X_m) + \sum_{k=m+1}^{n} cov(Y_k, X_m) \\ &= V(X_m) + \sum_{k=m+1}^{n} Cov\,(Y_k, X_m) \end{aligned}$$

The second equality is obtained through the linearity of the covariance operator. The last equality comes from the equality $Cov(X_m, X_m) = V(X_m)$. Moreover the covariances $Cov\,(Y_k, X_m)$ are equal to zero for $k > m$ due to the independence of the Y_k and $X_m = X_0 + \sum_{k \leq m} Y_k$. Finally, we use the assumption $V(Y_k) = \sigma^2$ to obtain $V(X_m) = \sum_{k=1}^{m} V(Y_k) = m\sigma^2$.

Obviously, if $n < m$, we get $Cov(X_n, X_m) = V(X_n) = n\sigma^2$. ■

Suppose now that X_n is the logarithm of the date-t price of a financial asset where $t = nh$, starting at $X_0 = 0$ (in other words, the initial price of the asset is equal to 1). X_n is then also the cumulated return of the asset on the interval $[0; t]$. The set $\mathcal{T} = \{0, 1, ..., N\}$ is the set of transaction dates occurring in $[0; T]$[6].

Assuming $N \to +\infty$ is equivalent to suppose that the market approaches a continuous-time market. If the cumulated continuous return on $[0; T]$ is denoted Λ_T, we need $\Lambda_T = \lim_{N \to +\infty} X_N$ to get a consistent model.

σ is the standard deviation of returns on a given sub-period. It should depend on N because the cumulated variance is $N\sigma^2$ on $[0; T]$. If T is 3 months, weekly data corresponds to $N = 13$ and $N = 91$ corresponds to daily data. Obviously the daily volatility of returns is different from the corresponding weekly volatility.

[6] For example, T may be the maturity date of an option contract we try to evaluate.

So, the question is to link σ to N in a consistent way. It is also the typical question you need to answer to calibrate the Cox-Ross-Rubinstein model (1979). When the period length changes, u and d must be changed.

Let now $\left(Y_n^N, n = 1, ..., N\right)$ denote the increments of the random walk when $\mathcal{T} = \{0, 1, ..., N\}$. Denote σ_h the standard deviation of Y_n^N. The intuitive idea is to define the date-T logarithm of the price as $X_N = X_0 + \sum_{n=1}^N Y_n^N$.

Note that for any N, $X_N - X_0$ defines the cumulated return of the asset on $[0; T]$, that is Λ_T. Consequently when N tends to infinity and h tends to 0, we should get a consistent description of the continuous-time process of returns.

2.3.2 The assumptions

To ensure the consistency of the approach, some precautions are needed. Careful assumptions are necessary about the evolution of parameters when N changes. We introduce hereafter the three "reasonable" assumptions proposed by Merton (1982).

1. There exists $A_1 > 0$, independent of N, such that:

$$\forall n \in \mathcal{T}, \ \ V(X_n) \geq A_1$$

2. There exists $A_2 > 0$, independent of N, such that:

$$V(X_N) \leq A_2$$

3. There exists $A_3 > 0$, independent of N, such that, for any $n \in \mathcal{T}$:

$$\frac{V(Y_n)}{\max_{j=1}^N V(Y_j)} \geq A_3$$

Comments

- Assumption 1 says that uncertainty never disappears, even if the length of subperiods tends to 0. Consequently, the cumulated return between 0 and t remains stochastic on continuous-time markets.

- Assumption 2 is, in some sense, symmetrical to the first one. It says that variance of returns on $[0;T]$ doesn't explode when time is cut in finer and finer slices. This assumption illustrates what was already said in the Cox-Ross-Rubinstein framework. When the number of sub-periods increases, parameters u and d have to be changed. If they are held constant, the variance of returns would become unbounded.

- Assumption 3 has a specific meaning. It stipulates that the variance of returns remains greater than a given proportion of the maximum variance. This assumption rejects cases in which uncertainty is concentrated in some sub-periods. An example allows to understand why this assumption is reasonable to deal with risky assets. Assume you buy a lotto ticket on Tuesday when the official draw is on the next Saturday. Obviously, you buy a risky asset but it doesn't satisfy assumption 3. The problem is that nothing happens between Tuesday and Saturday. Consequently, the value of the ticket remains the same (no uncertainty about the return on intermediate periods). The value of the ticket changes only during the draw, depending on whether the numbers you chose show up in the official draw. In other words, the lotto ticket is locally risk-free!

- An alternative justification (possibly more serious!) of assumption 3 is that on a continuous market, new information is generated in a continuous flow, implying possible price variations at any moment. For the lotto ticket, no new information becomes available before the Saturday draw starts. We except the case where the sponsor of the game announces bankruptcy on Wednesday!

Remark that, if the process under consideration has stationnary increments, assumption 3 is satisfied because the variance is constant on subperiods of a given length. For the random walk considered at the beginning of the section we could choose $A_3 = 1$.

This set of three assumptions, intuitive at the economic level, allows to specify which processes are good candidates to represent prices and returns of financial assets. We first recall the Landau notations O and o.

Definition 26 *Let f and g be two functions defined on $\mathbb{R}$ and taking values in $\mathbb{R}$. We write $f(h) = O(g(h))$ if $\lim_{h\to 0}\left|\frac{f(h)}{g(h)}\right| < +\infty$ and $f(h) = o(g(h))$ if $\lim_{h\to 0}\left|\frac{f(h)}{g(h)}\right| = 0$.*

$f(h) = O(g(h))$ means that the values of f and g have the same magnitude when h tends to 0. In the same way, $f(h) = o(g(h))$ means that f is infinitely small with respect to g when h tends to 0.

The following proposition specifies the behaviour of the variance σ_h^2 when the length of the interval between two trading dates shrinks to 0.

Proposition 15

$$\sigma_h^2 = O(h) \text{ and } \sigma_h^2 \neq o(h) \tag{2.7}$$

where $O(h)$ and $o(h)$ denote the Landau notations.

Proof.

$$V(X_N) = \sum_{n=1}^{N} V(Y_n^N) = N\sigma_h^2$$

This equality is a direct consequence of the independence Y_n^N. Assumption 2 implies $N\sigma_h^2 \leq A_2$. Replacing N by $\frac{T}{h}$ leads to:

$$\sigma_h^2 \leq \frac{A_2}{T}h \tag{2.8}$$

It means that $\sigma_h^2 = O(h)$.

In the same way, assumption 1 implies:

$$N\sigma_h^2 \geq A_1 \tag{2.9}$$

or, equivalently, $\sigma_h^2 \geq \frac{A_1}{T}h$. it shows that $\sigma_h^2 \neq o(h)$. ■

If the variance of Y_n is $O(h)$, the values taken by Y_n can be written as $\sigma_h = c\sqrt{h}$ and $-c\sqrt{h}$ with c a positive constant. In other words, the magnitude of the return is $O(\sqrt{h})$ on a period of length h.

Consider now the special case $c = 1$; the process X may be written:

$$X_n = X_{n-1} + \delta_n \sqrt{h}$$

where the δ_n are zero-mean independent variables taking values $+1$ and -1.

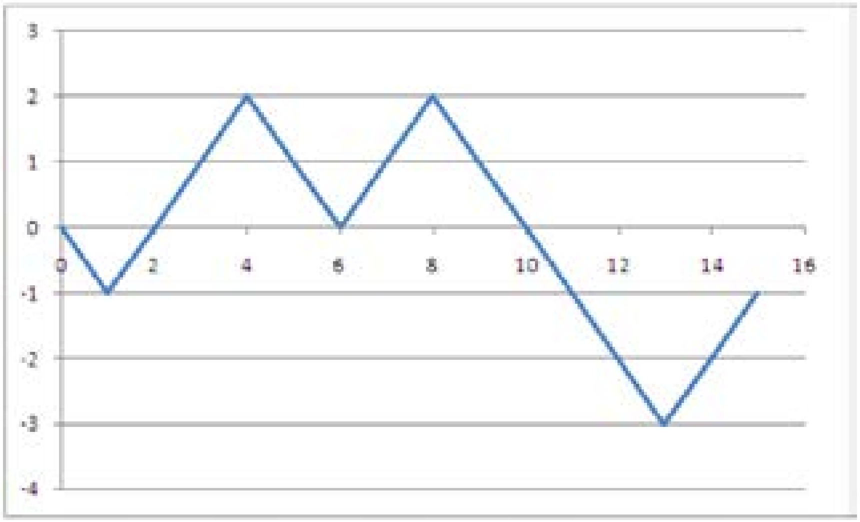

Figure 2.3: A random walk with $h = 1$

Figure 2.3 represents a path with $h = 1$ and figure 2.4 a path with $h = 0.2$. The points for successive dates have been joined by straight lines.

To understand how trajectories of the Brownian motion look like, it is enough to remark that the slope of the successive segment on the two figures is $\pm\frac{\sqrt{h}}{h}$, or equivalently $\pm\frac{1}{\sqrt{h}}$.

We observe that when h decreases, the slope increases. We advice the reader to simulate such paths for different values of h to observe the phenomenon (on an Excel sheet for example). When h tends to 0 the path looks like the price paths provided by any financial website or economic newspaper (or by figures 2.1 and 2.2 at the beginning of this chapter).

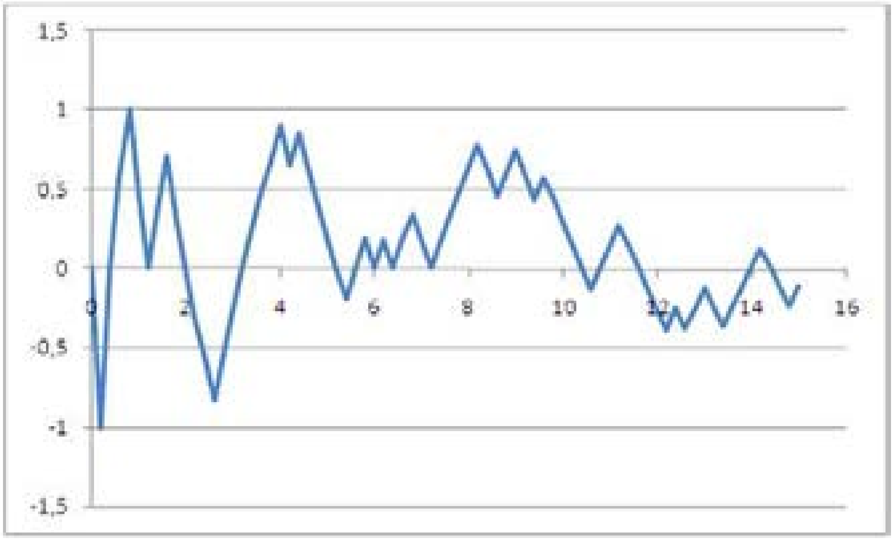

Figure 2.4: A random walk with $h = 0.2$

Define now $\Lambda_T = \lim_{N\to+\infty}\sqrt{\frac{T}{N}}\sum_{n=1}^{N}\delta_n$ where the δ_n are defined as before. The central limit theorem allows to conclude that Λ_T follows a Gaussian distribution with standard deviation $\sqrt{T}$. Here, weak convergence is used.

If two dates T and T' are considered, with $T < T'$, it is clear that $\Lambda_{T'} - \Lambda_T$ is Gaussian with standard deviation $\sqrt{T' - T}$ and that $\Lambda_{T'} - \Lambda_T$ is independent of $\Lambda_T - \Lambda_0$.

Following this intuitive presentation, we can now define more formally the Brownian motion which is the limit process of the random walks developed before when the duration between two transaction dates shrinks to 0. .

2.3.3 Definition and general properties

Definition 27 *Let $(\Omega, \mathcal{A}, P)$ be a probability space; a standard Brownian motion (or Wiener process) is a stochastic process Z satisfying:*

1. $Z_0 = 0$ *P.a.s*[7]
2. *The increments of Z are independent and stationnary*

[7] Remember that a.s means "almost surely", that is "with probability 1".

3. $\forall (s,t) \in \mathbb{R}^+ \times \mathbb{R}^+,\ s < t,\ Z_t - Z_s \sim \mathcal{N}(0, \sqrt{t-s})$

4. *Z has continuous paths.*

Without technical words, the idea behind the brownian motion is the same as the one behind the Gaussian distribution. When a phenomenon is the addition of a large number of independent components, the result should be driven by a Gaussian distribution. Moreover, combining points 1 and 3 of the definition shows that Z_t is a zero-mean Gaussian variable with standard deviation $\sqrt{t}$.

In finance, price variations come from a continuous flow of information leading to demands and supplies of assets by investors. Therefore, if information is immediately reflected in prices (efficient market hypothesis), successive returns should be independent. If it was not the case, investors would take into account this dependency...and it would be ruled out.

The preceding section based on random walks illustrated that variations of Z between two dates t and $t+h$ are of order $\sqrt{h}$, infinitely large w.r.t h (see Landau notations). It explains why Brownian motion paths are very irregular. It is in fact impossible to draw a trajectory of a Brownian motion, because of the following proposition.

Proposition 16 *Brownian motion paths are continuous but nowhere differentiable.*

Most properties of random walks can be transferred witout difficulties to the Brownian motion.

Proposition 17 *1. The Brownian motion Z is a Markov process.*

2. Z is a martingale with respect to its natural filtration.

3. $cov(Z_t, Z_s) = \min(t, s)$

Proof. Point (1) is a direct consequence of the independent increments.

When you deal with Brownian motion, the independence of increments is used in most proofs. The idea is always to write something like $Z_t = Z_s + (Z_t - Z_s)$ (with $s < t$) and use the fact that Z_s and $(Z_t - Z_s)$ are independent. Moreover, $(Z_t - Z_s)$ is also independent of $\mathcal{F}_s^Z$, the natural filtration of Z.

To prove point (2), let s and t denote two dates such that $s < t$. We can write:

$$\begin{aligned} E\left[Z_t \middle| \mathcal{F}_s^Z\right] &= E\left[Z_s + (Z_t - Z_s) \middle| \mathcal{F}_s^Z\right] \\ &= E\left[Z_s \middle| \mathcal{F}_s^Z\right] + E\left[Z_t - Z_s \middle| \mathcal{F}_s^Z\right] \end{aligned}$$

The second term in the RHS is equal to 0 because $Z_t - Z_s$ is independent of $\mathcal{F}_s^Z$:

$$E\left[Z_t - Z_s \middle| \mathcal{F}_s^Z\right] = E\left[Z_t - Z_s\right] = 0 \tag{2.10}$$

As Z is adapted to $\mathcal{F}^Z$, Z_s is $\mathcal{F}_s^Z$-measurable. The properties of conditional expectations imply:

$$E\left[Z_s \middle| \mathcal{F}_s^Z\right] = Z_s \tag{2.11}$$

We finally get the martingale relationship

$$E\left[Z_t \middle| \mathcal{F}_s^Z\right] = Z_s \tag{2.12}$$

To prove point (3), the same "trick" is used (assume that $s < t$) :

$$\begin{aligned} cov(Z_t, Z_s) &= cov(Z_t - Z_s + Z_s, Z_s) \\ &= cov(Z_t - Z_s, Z_s) + V(Z_s) \\ &= cov(Z_t - Z_s, Z_s - Z_0) + s \\ &= s \end{aligned}$$

One more time, independent increments imply that the covariance between $Z_t - Z_s$ and $Z_s - Z_0$ is zero.

Obviously, if $t < s$, Z_s is decomposed in $Z_s - Z_t + Z_t$ and the same reasoning applies. ■

The definition of diffusion and Itô processes in section 2.2.2 introduced the following functions:

$$\begin{aligned}\mu(x,t) &= \lim_{h\to 0} \frac{E\left[X_{t+h} - X_t \left| X_t = x\right.\right]}{h} \\ \sigma^2(x,t) &= \lim_{h\to 0} \frac{V\left[X_{t+h} - X_t \left| X_t = x\right.\right]}{h}\end{aligned}$$

It appears that the standard Wiener process is an Itô process with zero drift ($\mu(x,t) = 0$) and a diffusion coefficient $\sigma(x,t)$ equal to 1.

2.3.4 Usual transformations of the Wiener process

Some functions of the Brownian motion are also martingales; the two most simple examples are summarized in the following proposition.

Proposition 18 *Let X and Y be defined by:*

$$\begin{aligned}X_t &= Z_t^2 - t \\ Y_t &= \exp\left(\gamma Z_t - \frac{\gamma^2 t}{2}\right)\end{aligned}$$

where γ is a real number. X and Y are martingales w.r.t. $\mathcal{F}^Z$.

Proof. Z is square-integrable so X_t is integrable; Z_t is Gaussian so Y_t is log-normal. The two processes are then integrable.

To prove the martingale relationship, we use one more time the decomposition $Z_t = Z_s + (Z_t - Z_s)$ with $s < t$. The conditional expectation $E\left(X_t \left| \mathcal{F}_s^Z\right.\right)$ is transformed in the following way:

$$\begin{aligned}E\left(X_t \left| \mathcal{F}_s^Z\right.\right) &= E\left(Z_t^2 \left| \mathcal{F}_s^Z\right.\right) - t \\ &= E\left((Z_t - Z_s)^2 + 2Z_t Z_s - Z_s^2 \left| \mathcal{F}_s^Z\right.\right) - t\end{aligned}$$

We use now:

1) the linearity of conditional expectations
2) the fact that Z_s is $\mathcal{F}_s^Z-$measurable
3) the fact that Z is a martingale
4) the independence of $Z_t - Z_s$ w.r.t. $\mathcal{F}_s^Z$.
We then get:

$$\begin{aligned}
E\left(X_t \left|\mathcal{F}_s^Z\right.\right) &= E\left((Z_t - Z_s)^2 \left|\mathcal{F}_s^Z\right.\right) + 2E\left(Z_t Z_s \left|\mathcal{F}_s^Z\right.\right) - E\left(Z_s^2 \left|\mathcal{F}_s^Z\right.\right) - t \\
&= E\left((Z_t - Z_s)^2 \left|\mathcal{F}_s^Z\right.\right) + 2E\left(Z_t Z_s \left|\mathcal{F}_s^Z\right.\right) - Z_s^2 - t \\
&= E\left((Z_t - Z_s)^2 \left|\mathcal{F}_s^Z\right.\right) + 2Z_s E\left(Z_t \left|\mathcal{F}_s^Z\right.\right) - Z_s^2 - t \\
&= E\left((Z_t - Z_s)^2 \left|\mathcal{F}_s^Z\right.\right) + 2Z_s^2 - Z_s^2 - t \\
& \quad E\left((Z_t - Z_s)^2\right) + Z_s^2 - t \\
& \quad (t-s) + Z_s^2 - t \\
&= Z_s^2 - s = X_s
\end{aligned}$$

We now show that Y is a martingale, using the same properties of conditional expectations.

$$\begin{aligned}
E\left(Y_t \left|\mathcal{F}_s^Z\right.\right) &= \exp\left(-\frac{\gamma^2 t}{2}\right) E\left(\exp\left(\gamma Z_t\right) \left|\mathcal{F}_s^Z\right.\right) \\
&= \exp\left(-\frac{\gamma^2 t}{2}\right) E\left(\exp\left(\gamma (Z_t - Z_s)\right) \exp(\gamma Z_s) \left|\mathcal{F}_s^Z\right.\right) \\
&= \exp\left(-\frac{\gamma^2 t}{2}\right) E\left(\exp\left(\gamma (Z_t - Z_s)\right)\right) \exp(\gamma Z_s)
\end{aligned}$$

Remark now that $\exp\left(\gamma(Z_t - Z_s)\right)$ follows a log-normal distribution with parameters 0 and $\gamma\sqrt{t-s}$. Consequently $E\left(\exp\left(\gamma(Z_t - Z_s)\right)\right) = \frac{\gamma^2(t-s)}{2}$. Replacing in the last equality leads to:

$$E\left(Y_t \left|\mathcal{F}_s^Z\right.\right) = \exp\left(\gamma Z_s - \frac{\gamma^2 s}{2}\right) = Y_s$$

and ends the proof. ■

The processes X and Y in proposition 18 are in fact specific transformations of the Brownian motion by the following functions:

$$\begin{aligned}
(x,t) &\rightarrow f(x,t) = x^2 - t \\
(x,t) &\rightarrow g(x,t) = \exp\left(\gamma x - \frac{\gamma^2 t}{2}\right)
\end{aligned}$$

Let $z(x,t)$ stands for the density of Z_t defined by:

$$z(x,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right)$$

The functions f and g are solutions of the equation:

$$h(0,s) = \int_{-\infty}^{+\infty} h(u,t+s) z(u,t) du$$

In fact, for f, we have $f(0,s) = -s$; the RHS is written:

$$\begin{aligned}\int_{-\infty}^{+\infty} f(u,t+s) z(u,t) du &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} (u^2 - (t+s)) \exp\left(-\frac{u^2}{2t}\right) du \\ &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} u^2 \exp\left(-\frac{u^2}{2t}\right) du - (t+s)\end{aligned}$$

$\frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} u^2 \exp\left(-\frac{u^2}{2t}\right) du$ is the variance of Z_t, equal to t. For g, we have:

$$\begin{aligned}\int_{-\infty}^{+\infty} g(u,t+s) z(u,t) du &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} \exp\left(\gamma u - \frac{\gamma^2(t+s)}{2} - \frac{u^2}{2t}\right) du \\ &= \exp\left(-\frac{\gamma^2(t+s)}{2}\right) \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} \exp\left(\gamma u - \frac{u^2}{2t}\right) du\end{aligned}$$

The integral is transformed by the usual method:

$$\exp\left(\gamma u - \frac{u^2}{2t}\right) = \exp\left(-\frac{1}{2t}(u-\gamma t)^2 + \frac{\gamma^2 t}{2}\right)$$

which leads to:

$$\begin{aligned}\int_{-\infty}^{+\infty} g(u,t+s) z(u,t) du &= \exp\left(-\frac{\gamma^2 s}{2}\right) \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2t}(u-\gamma t)^2\right) du \\ &= \exp\left(-\frac{\gamma^2 s}{2}\right) = g(0,s)\end{aligned}$$

2.3.5 The general Wiener process

Using a standard Brownian motion as a model for stock returns has at least two serious drawbacks. The first one is that $E(Z_t) = 0$ for any t. Therefore, the expected stock returns are always 0. The second problem comes from the variance $V(Z_t) = t$ meaning that a dotcom stock and an industrial firm (pharmaceutical, automobile, etc.) have the same variance. It does not correspond to empirical observations.

Introducing the general Wiener process allows to solve these two difficulties by introducing a drift (which may be positive or negative) and a per time-unit variance (which may be lower or greater than 1). The natural extension of the standard Brownian motion introduces parameters μ and σ, μ being the drift, and σ^2 the instantaneous variance per time-unit.

Definition 28 *A process W is a general Brownian motion with parameters μ and σ if W_t writes:*

$$\begin{aligned} W_0 &= 0 \\ W_t &= \mu t + \sigma Z_t \end{aligned}$$

where Z is a standard Brownian motion.

Properties of W are easily deduced from those of Z.

Proposition 19 *1. W is a process with independent and stationnary increments*

2. *For $s < t$, $W_t - W_s \sim \mathcal{N}(\mu(t-s), \sigma\sqrt{t-s})$*
3. *$cov(W_t, W_s) = \sigma^2(t-s)$*
4. *W is a sub(super)-martingale w.r.t. its natural filtration if $\mu > (<)0$.*
5. *The paths of W are (a.s) continuous and nowhere differentiable.*

Think of a stock with an initial price of 1. If W_t is the cumulated return of this stock on $[0;t]$, $\mu > 0$ means that the expected return is positive. If there is a risk-free asset traded on the market with a non stochastic return r, risk aversion of agents implies generally that $\mu > r$. Investors require a premium to invest in a risky stock.

Remark 2 *1) Let h be a real number close to 0; the variation of W on $[t;t+h]$ is written as:*

$$W_{t+h} - W_t = \mu h + \sigma(Z_{t+h} - Z_t)$$

When building Z, starting with random walks, we observed that $Z_{t+h} - Z_t$ is $O(\sqrt{h})$, that is infinitely large w.r.t. h when h tends to 0. It is also true for W and the economic consequences are important. Even if the expected return μ is different from zero, it is not possible to predict short-term variations. The term $\sigma(Z_{t+h} - Z_t)$ is much larger (in absolute value) than μh when h is close to 0. It is the reason why observing very short term variations in prices (intraday variations for example) says nothing about the evolution on a longer horizon (yearly variations for example).

2) One has to be careful with intuitions about the Brownian motion. For example, assume $\mu = 0$, corresponding to a process without any drift. It is natural to think that paths will stay "around" zero. However, we will show that paths of such a Brownian motion cross any boundary (starting from 0, the value goes outside any interval $[a;b]$) in a finite time). Consequently, if you observe just one path, you could conclude that there is a drift when it is not the case.

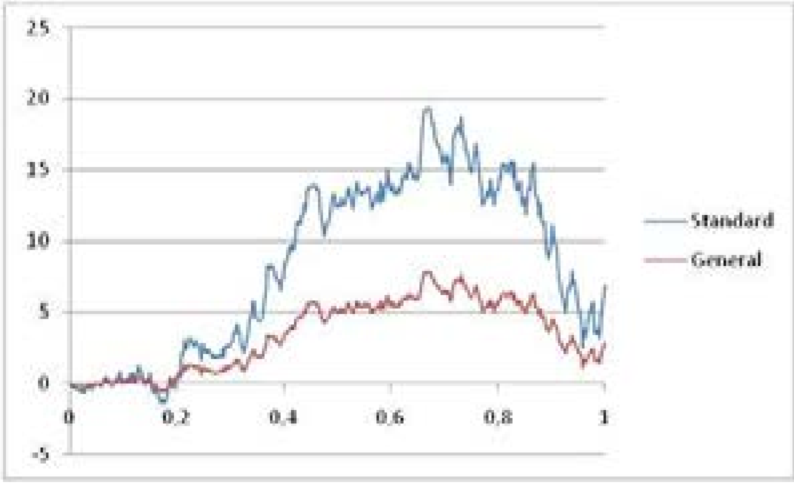

Figure 2.5: Simulation of a standard and a general Brownian motions ($\mu =$ 10% and $\sigma = 20\%$)

Remark 3 *Figure 2.5 shows an example. We simulated a path of a standard Brownian motion on one year (the time-unit) at a daily frequency. It corresponds to the thin blue line. The bold red line is the corresponding general Wiener process with parameters* $\mu = 10\%$ *and* $\sigma = 20\%$*, that are reasonable values for a stock return. Looking at the path on the 9 first months (up to* $t = 0.75$*), the curve shows a "strong positive drift" but in fact the simulation was obtained with a zero drift.*

Before proving the abovementioned "crossing boundary" result, it is necessary to introduce stopping times in continuous-time.

2.3.6 Stopping times

Definition 29 *Let* $(\Omega, \mathcal{A}, \mathcal{F}, P)$ *be a filtered probability space; a stopping time is a random variable* τ *taking values in* $\mathcal{T} \bigcup \{+\infty\}$ *such that* $\{\tau \leq t\} \in \mathcal{F}_t$ *for any* $t \in \mathcal{T}$*.*

This definition is close to the one given in discrete-time. The essential difference is that it uses events $\{\tau \leq t\}$ instead of $\{\tau = t\}$. Time being continuous, it could happen that $P(\{\tau = t\}) = 0$ for any t.

As mentioned before, stopping times come naturally in the valuation of American options for which early exercise is possible. The exercise decision must be based only on past and present information, it cannot be based on future prices which are not yet known.

The value $+\infty$ has been added to the set of possible values of a stopping time. It is used to take into account that, sometimes, it is never optimal to exercise the option, even at the maturity date. In that case, it is conventional to write $\tau = +\infty$.

Consider now a corporate bond with maturity T paying coupons $(C_1, ..., C_T)$. This bond may default at date t meaning that it really pays $F_t < C_t$. The first default date is the stopping-time τ defined as $\tau = \inf\{t \ / \ F_t < C_t\}$. We write $\tau = +\infty$ if default never occurs before the maturity date T.

Proposition 20 *Let τ and τ' be two stopping times w.r.t. a filtration $\mathcal{F}$. $\tau + \tau'$, $\min(\tau, \tau')$, $\max(\tau, \tau')$ are stopping times.*

A common alternative notation for $\min(\tau, \tau')$ is $\tau \wedge \tau'$ and $\tau \vee \tau'$ for $\max(\tau, \tau')$.

Stopping times like $\tau \wedge \tau'$ and $\tau \vee \tau'$ are frequently encountered in life insurance policies. If τ and τ' are the dates of death of two married people, some contracts stipulate that the insurance company will pay a given amount to the survivor at the first death date $(\min(\tau, \tau'))$, some other contracts paying to the heirs at the last death date $(\max(\tau, \tau'))$. A usual joke is to remark that $|\tau - \tau'|$ takes low values when the spouse dies first, but takes very large values when the husband dies first.

More seriously, it must be noted that a sure date t is a stopping-time. Consequently, $t \wedge \tau$ is also a stopping time.

A tribe $\mathcal{F}_\tau$ is linked to any stopping time τ. It contains all the sets $A \bigcap \{\tau \leq t\}$ with $A \in \mathcal{A}$. We then get the following proposition.

Proposition 21 *Let τ be a stopping time and X a stochastic process.*

1. *τ is $\mathcal{F}_\tau$-measurable*
2. *If τ is bounded and X adapted to $\mathcal{F}$, then X_τ is $\mathcal{F}_\tau$-measurable*
3. *If τ and τ' are two stopping times such that $\tau \leq \tau'$, then $\mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'}$*

2.3.7 Properties of the Brownian motion paths

We come now to the "crossing-boundary" property.

Proposition 22 *Let W be a Wiener process with parameters $(0, \sigma)$; let a and b denote two real numbers such that $a < 0 < b$ and τ denote the stopping time defined by:*

$$\tau = \inf \left\{ t \in \mathbb{R}^+ \ / \ W_t = a \quad or\ W_t = b \right\}$$

We then get:

$$\begin{aligned} P\left(W_\tau = a\right) &= \frac{b}{b-a} \\ E(\tau) &= \frac{-ab}{\sigma^2} \end{aligned}$$

Proof. τ is the first date at which the process reaches either a or b. First, remark that the probability of reaching a at date τ (instead of b) doesn't depend on σ, because at date τ the process is in a or b. It is then intuitive that σ only plays a role on the delay to reach one of the bounds but not on the probability of reaching one of these bounds. It follows that $P\left(W_\tau = a\right) = P(Z_\tau = a)$.

Consider X, the martingale defined in proposition 18 :

$$X_t = Z_t^2 - t$$

Let $\tau^* = \tau \wedge s$ with s a positive real number; the stopped process $X_t^{\tau^*}$satisfies:

$$E\left(X_t^{\tau^*}\right) = E(X_0^{\tau^*}) = 0$$

from which we deduce:

$$E\left(Z_{t\wedge\tau^*}^2\right) = E(t \wedge \tau^*) \leq (b-a)^2$$

In fact, the definition of τ implies that before this date, the process Z has reached neither a nor b. But this relationship is true for any s, implying:

$$E(\tau) = \lim_{s\to+\infty} E(\tau^*) \leq (b-a)^2$$

τ is then a bounded stopping time, implying $E(Z_\tau) = E(Z_0) = 0$ by the optional stopping theorem. We can then write:

$$E(Z_\tau) = aP(Z_\tau = a) + b(1 - P(Z_\tau = a)) = 0$$

which leads to:

$$P(Z_\tau = a) = \frac{b}{b-a}$$

The same reasoning applied to X allows to write:

$$E\left(X_t^\tau\right) = 0 = E\left(Z_\tau^2 - \tau\right)$$

or

$$E(Z_\tau^2) = a^2\frac{b}{b-a} + b^2\frac{-a}{b-a} = -ab$$

As $W_t = \sigma Z_t$, we then get the desired result. ■

This result shows that Z crosses one of the boundaries in a finite expected time. It also illustrates why interest rates cannot be easily represented by a Brownian motion. First, they could take negative values, and second they would go higher than any given threshold in a finite expected time...even if the threshold is 100% or 200%. The intuitive idea we can have about the dynamics of interest rates is that the drift should be negative when interest rates are high and positive when they are near zero. We then need more general diffusion processes than the Brownian motion to adequately modelize many economic variables. These processes are presented in the next chapter.

Chapter 3

Stochastic integral and Itô's lemma

3.1 Introduction

In the preceding chapter we presented the elementary properties of the Wiener process. The way we built the Brownian motion was based on random walks with zero-mean binary increments Y_n. Then we deduced a formulation of the general Wiener process by allowing non zero-mean increments and a variance proportional to the delay between two dates. However, the parameters (expectation and variance per time unit) were constant. In particular they depended neither on the date nor on the value already reached by the stochastic process. It may be a restrictive assumption, even for relatively simple economic processes like interest rates or stock prices.

Assume for example that we want to model the stochastic process driving the evolution of a stock price. Using a Wiener process generates several problems but we will focus on just one for the moment. Think to the economing meaning of a constant expected change per time unit. It means for example that a price will have a yearly expected change of $10, whatever the current price is. Obviously a $10 change is not the same when the stock price is $20 or when it is $200. Therefore, the Brownian motion is not a good candidate to model stock prices, even if it seems adequate for returns (except during financial crises!!!).

The Brownian motion assumptions are then restrictive when one wants to describe specific economic variables. As mentioned at the end of the preceding chapter, real data suggest that the probability of an increase (decrease) in interest rates is lower when rates are high (low). A mean-reverting process comes naturally to mind for describing such phenomena. In the same way, it seems intuitive that the volatility of interest rates is higher (lower) when rates are high (low). Figure 3.1 illustrates the point. It represents the so-called U.S Fed rates on a period of 55 years. It seems reasonable to consider a stochastic process with non constant volatility and a varying drift. For example, the Fed rate felt to 0 at the end of the period under consideration due to the financial crisis. The drift cannot be negative after that date.

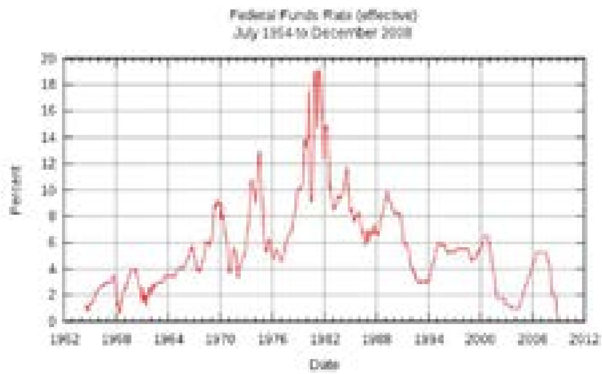

Figure 3.1: Interest rates for Federal Funds (Source: Federal Reserve, http://www.federalreserve.gov)

The definition of Itô processes assumes there exist two functions $\mu(x,t)$ and $\sigma(x,t)$ depending on time and on the level reached by the process under consideration. These processes are good candidates to take into account the abovementioned problems (however, keep in mind that the Brownian motion is a special case of Itô process with constant parameters).

In this chapter we are going to "build" Itô processes using the concept of stochastic integral. We will then present Itô's lemma (also called Fundamental theorem of stochastic calculus) which is the mathematical tool used to price derivative securities in continuous-time. The price of such securities can be written as a regular function of the price of an underlying asset. Itô's lemma helps to find the dynamics of prices of derivative securities, being given the dynamics of the price of the underlying asset.

Finally, we introduce the Girsanov theorem which allows to change the drift of an Itô process without changing its variance. In arbitrage pricing models, this is an important question because derivative contracts are valued under the assumption of no arbitrage. The valuation technique consists in writing the dynamics of prices in a risk-neutral world in which the expected returns are equal to the risk-free rate (which is different from the physical rate[1]).

3.2 The stochastic integral

3.2.1 An intuitive approach

As mentioned before, a brownian motion assumption is not a good way to describe the random evolution of stock prices, even if it may be a good idea to model the logarithm of stock prices. In the following, we continue to refer to the logarithm of a stock price but the stochastic integral we are going to build is a much more general approach than the one used before.

To keep things simple, we nevertheless adopt the same pedagogical approach as in the preceding chapter by starting with random walks. Let a time-interval $[0;T]$ be given, and assume that it is divided in N sub-periods of equal length h with $h = T/N$. For example, $0, h, 2h, ..., Nh$ are the dates at which the market is open and trades occur.

Let X stands for the stochastic process of the logarithm of a stock price defined by:

$$\begin{aligned} X_0 &= 0 \\ X_n &= X_{n-1} + Y_n \end{aligned}$$

for $n \in \mathcal{T} = \{1, ..., N\}$. This formulation means that $X_n(\omega)$ is the value of the process X at date nh in state ω since h is the duration between two successive trading dates. Y_n is still interpreted as the stock return on a period of length h.

[1]We do not use the term "real" rate because it usually refers to rates of return after having taken into account inflation. Here, "physical" refers to the expected return in the real world where agents are risk-averse.

The definition of the random walk Y_n is now a little bit more involved:

$$Y_n = \mu_{n-1} h + \sigma_{n-1}\left(Z_{nh} - Z_{(n-1)h}\right) \tag{3.1}$$

where Z is a standard brownian motion. It is assumed that μ_{n-1} and σ_{n-1} are $\mathcal{F}_{(n-1)h}$-measurable random variables where $\mathcal{F}$ is the natural filtration of Z. This formulation is very general and, for the moment, no other assumptions are made on μ_{n-1}, σ_{n-1}.

μ_{n-1} is the per time-unit conditional expectation of the increment of X on the time-interval $[(n-1)h; nh]$ and σ_{n-1} is the corresponding conditional standard deviation. Conditional moments are defined here with respect to the $\sigma-$algebra $\mathcal{F}_{(n-1)h}$. It also means that σ_{n-1} and $Z_{nh} - Z_{(n-1)h}$ are independent and $V\left(Z_{nh} - Z_{(n-1)h}\right) = h$.

In the preceding chapter, we described Merton's assumptions (1982) about the processes X and Y and we still assume they are satisfied.

Let Λ_T the cumulated return of the stock on $[0; T]$; we want to write Λ_T as the limit (in a sense to be defined) of X_N when N tends to infinity. We then have:

$$X_N = \sum_{n=1}^{N} Y_n = h \sum_{n=1}^{N} \mu_{n-1} + \sum_{n=1}^{N} \sigma_{n-1}\left(Z_{nh} - Z_{(n-1)h}\right)$$

Replacing h by $\frac{T}{N}$, we get:

$$\begin{aligned} X_N &= \sum_{n=1}^{N} Y_n = T\left(\frac{1}{N}\sum_{n=1}^{N} \mu_{n-1}\right) + \sum_{n=1}^{N} \sigma_{n-1}\left(Z_{nh} - Z_{(n-1)h}\right) \\ &= A_N + B_N \end{aligned}$$

with $A_N = T\left(\frac{1}{N}\sum_{n=1}^{N} \mu_{n-1}\right)$ and $B_N = \left(\sum_{n=1}^{N} \sigma_{n-1}\left(Z_{nh} - Z_{(n-1)h}\right)\right)$

To get a consistent continuous-time model, we need to define precisely $A = \lim_{N\to+\infty} A_N$ and $B = \lim_{N\to+\infty} B_N$.

Let α^N a continuous-time process defined by:

$$\alpha_s^N = \mu_{n-1} \text{ if } s \in [(n-1)h; nh[$$

For a given state of nature, $\alpha_s^N(\omega)$ is then constant on each interval $[(n-1)h; nh[$. In other words, $\alpha^N(\omega)$ is then a step function.

For any ω, we can write:

$$A_N(\omega) = \int_0^T \alpha_s^N(\omega)ds$$

where the integral is a usual Riemann-Stieltjes integral. When $N \to +\infty$, we can define, for each state ω, $A(\omega)$ as:

$$A(\omega) = \int_0^T \alpha_s(\omega)ds$$

where $\alpha_s(\omega) = \lim_{N\to+\infty} \alpha_s^N(\omega)$. Here, the limit process α is defined "state-by state".

Unfortunately, we cannot use the same simple approach to define the limit of B_N, simply because the values taken by the random variables $\left(Z_{nh} - Z_{(n-1)h}\right)$ are of order $\sqrt{h}$.

If σ and Z were usual functions and not stochastic processes, we would get the following formulation:

$$\lim_{N\to+\infty} B_N = \int_0^T \sigma_s dZ_s$$

by using a standard Stieltjes integral. However, this approach does not work, as shown in the following example.

3.2.2 Counter-example

Let C_N and D_N defined by :

$$\begin{aligned} C_N &= \sum_{n=1}^{N} Z_{nh}\left(Z_{nh} - Z_{(n-1)h}\right) \\ D_N &= \sum_{n=1}^{N} Z_{(n-1)h}\left(Z_{nh} - Z_{(n-1)h}\right) \end{aligned}$$

with Z a standard Brownian motion. If Z was not a random process but an integrable function, we would get:

$$\lim_{N\to+\infty} C_N = \lim_{N\to+\infty} D_N = \int_0^T Z_s dZ_s = \frac{Z_T^2 - Z_0^2}{2} = \frac{Z_T^2}{2}$$

But we can easily show that the expectations of C_N and D_N are different and do not depend on N. One more time, the "trick" $Z_t = Z_s + (Z_t - Z_s)$ is useful.

$$E(C_N) = E\left(\sum_{n=1}^{N} Z_{nh}\left(Z_{nh} - Z_{(n-1)h}\right)\right) \quad (3.2)$$

$$= \sum_{n=1}^{N} E\left(Z_{nh}\left(Z_{nh} - Z_{(n-1)h}\right)\right) \quad (3.3)$$

Using $Z_{nh} = Z_{(n-1)h} + (Z_{nh} - Z_{(n-1)h})$, we get:

$$E(C_N) = \sum_{n=1}^{N} E\left[\left(Z_{(n-1)h} + (Z_{nh} - Z_{(n-1)h})\right)\left(Z_{nh} - Z_{(n-1)h}\right)\right] \quad (3.4)$$

$$= E\left(\sum_{n=1}^{N} Z_{(n-1)h}\left(Z_{nh} - Z_{(n-1)h}\right)\right) \quad (3.5)$$

$$+E\left(\sum_{n=1}^{N} \left(Z_{nh} - Z_{(n-1)h}\right)^2\right) \quad (3.6)$$

$$= \sum_{n=1}^{N} E\left[\left(Z_{nh} - Z_{(n-1)h}\right)^2\right] + \sum_{n=1}^{N} E\left[Z_{(n-1)h}\left(Z_{nh} - Z_{(n-1)h}\right)\right] \quad (3.7)$$

We use now the properties of the Brownian motion. As $Z_{nh} - Z_{(n-1)h}$ is zero-mean, we can write:

$$E\left(\left(Z_{nh} - Z_{(n-1)h}\right)^2\right) = V\left(Z_{nh} - Z_{(n-1)h}\right) = nh - (n-1)h = h$$

After summation, the first term in expression (3.4) is equal to $Nh = T$.

To evaluate the second term, we use the independence of increments of the Brownian motion.

$$\begin{aligned}\sum_{n=1}^{N} E\left(Z_{(n-1)h}\left(Z_{nh} - Z_{(n-1)h}\right)\right) &= \sum_{n=1}^{N} E\left(Z_{(n-1)h}\right) E\left(Z_{nh} - Z_{(n-1)h}\right) \\ &= 0\end{aligned} \tag{3.8}$$

We then obtain the following result:

$$E(C_N) = T$$

It is important to notice that it does not depend on the way the interval $[0;T]$ is sliced. Equivalently, it does not depend on N.

Remark now that $E(D_N)$ has already been calculated in equation (3.8) and it is equal to 0. Consequently, the two sequences C_N and D_N cannot converge to the same limit since they have different expectations.

This counter-example is interesting because it shows that we can define at least two stochastic integrals (in fact an infinity).

In the following, we choose the one based on D_N. This choice can be easily understood if one refers to the way the gain of a strategy is calculated. Remember that it was defined in discrete-time as:

$$g_n = \sum_{s=1}^{n} \theta_{nh}\left(X_{nh} - X_{(n-1)h}\right)$$

where the coefficients θ_{nh} denote quantities (assumed $\mathcal{F}_{(n-1)h}$-measurable because θ is a predictable process).

It is also the reason why σ_{n-1} was used in the definition of Y_n (see equation (3.1)) and assumed $\mathcal{F}_{(n-1)h}$-measurable.

To get a square-integrable limit, we need to assume that B_N is square-integrables.

$E\left(B_N^2\right)$ may be written as:

$$\begin{aligned} E\left(B_N^2\right) &= E\left(\left(B_N - B_{N-1} + B_{N-1}\right)^2\right) \\ &= E\left(\left(B_N - B_{N-1}\right)^2\right) + E\left(B_{N-1}^2\right) + 2E\left(B_{N-1}\left(B_N - B_{N-1}\right)\right) \\ &= E\left(\sigma_{N-1}^2\left(Z_{Nh} - Z_{(N-1)h}\right)^2\right) + E\left(B_{N-1}^2\right) + 2E\left(B_{N-1}\right)E\left(B_N - B_{N-1}\right) \\ &= hE\left(\sigma_{N-1}^2\right) + E\left(B_{N-1}^2\right) \end{aligned}$$

The last term in the RHS is zero because of the independence of increments of B. Moreover, $\left(Z_{Nh} - Z_{(N-1)h}\right)$ is independent of σ_{N-1}^2 and $E\left(\left(Z_{Nh} - Z_{(N-1)h}\right)^2\right) = h$. We then get, by a recurrence argument:

$$E\left(B_N^2\right) = T\left(\frac{1}{N}\sum_{n=1}^{N} E(\sigma_{n-1}^2)\right)$$

Remark now that if β^N is defined on $[0;T]$ by:

$$\beta_s^N = \sigma_{n-1} \text{ for } s \in [(n-1)h; nh[$$

we finally get:

$$E\left(B_N^2\right) = E\left(\int_0^T \left(\beta_s^N\right)^2 ds\right)$$

The sequence β^N converges to a process which is (naturally) denoted σ. The latter equality shows that $E\left(\int_0^T \left(\sigma_s\right)^2 ds\right) < +\infty$ is a sufficient condition for B_N to be square-integrable.

We can now define more formally the stochastic integral of a stochastic process w.r.t. the Brownian motion.

3.2.3 Definition and properties of the stochastic integral

Definition 30 *Let Z be a standard Brownian motion defined on $\left(\Omega, \mathcal{A}, \mathcal{F}^Z, P\right)$ and σ a process adapted to $\mathcal{F}^Z$. Suppose that σ satisfies:*

$$E\left(\int_0^T \sigma_s^2 ds\right) < +\infty \tag{3.9}$$

The ***Itô stochastic integral*** *of* σ *on* $[0;T]$ *w.r.t.* Z*, denoted as* $\int_0^T \sigma_s dZ_s$*, is the random variable defined by:*

$$\int_0^T \sigma_s dZ_s = \lim_{N \to +\infty} \sum_{n=1}^{N} \sigma_{n-1} \left(Z_{nh} - Z_{(n-1)h} \right)$$

where the limit refers to convergence in quadratic mean.

It is important to remark that the index n in the RHS of the preceding equation is discrete and corresponds to the subdivision of $[0;T]$ in N sub-periods. An equivalent alternative formulation would be to divide $[0;T]$ in sub-intervals $[t_i; t_{i+1}[$ such that $t_N = T$ and to write the limit as:

$$\int_0^T \sigma_s dZ_s = \lim_{N \to +\infty} \sum_{i=1}^{N} \sigma_{t_{i-1}} \left(Z_{t_i} - Z_{t_{i-1}} \right)$$

with $\max(t_i - t_{i-1})$ converging to 0 when $N \to +\infty$. Moreover, this formulation is valid for any horizon up to T. Consequently, the family $\left(\int_0^u \sigma_s dZ_s, u \in \mathcal{T} \right)$ of random variables is a stochastic process sometimes written as $\left(I_u(\sigma), u \in \mathcal{T} \right)$.

It is also worth noting that if σ is predictable, σ_{n-1} can be replaced by σ_n in definition 30. For example, for a stochastic process θ, representing quantities of stocks in a portfolio, the integral $\int_0^T \theta_s dZ_s$ is the gain of the strategy if the price process is Z.

Proposition 23 *Let X and Y be two adapted processes such that $\int_0^t X_s dZ_s$ and $\int_0^t Y_s dZ_s$ exist for any $t \leq T$ and are square-integrable (see definition 30). We then get:*

1. $\forall (a,b) \in \mathbb{R}^2, \int_0^t (aX_s + bY_s) dZ_s = a \int_0^t X_s dZ_s + b \int_0^t bY_s dZ_s$
2. $\forall (t,u) \in \mathcal{T} \times \mathcal{T}, t < u, \int_0^u X_s dZ_s = \int_0^t X_s dZ_s + \int_t^u X_s dZ_s$
3. $\forall (t,u) \in \mathcal{T} \times \mathcal{T}, t < u, \;\; E\left(\int_0^u X_s dZ_s \,|\mathcal{F}_t\right) = \int_0^t X_s dZ_s$
4. $\forall t \in \mathcal{T}, \quad E\left(\left(\int_0^t X_s dZ_s\right)^2\right) = E\left(\int_0^t X_s^2 ds\right)$

Point (1) means that the mapping associating a process X to its stochastic integral is linear. Point (2) is a standard property of integrals. Point (3) is crucial for financial applications because it shows that $\left(\int_0^u X_s dZ_s, u \in \mathcal{T}\right)$ is a martingale. Finally, the last point stipulates that $\int_0^t X_s dZ_s$ belongs to L^2.

Recall that building the stochastic integral in this way aims at looking for a continuous-time limit for the process X with

$$\begin{aligned} X_0 &= 0 \\ X_n &= X_{n-1} + Y_n = X_{n-1} + \mu_{n-1} h + \sigma_{n-1}\left(Z_{nh} - Z_{(n-1)h}\right) \end{aligned}$$

Consequently, the limit process (still denoted X to save some notations) is formulated as:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dZ_s \tag{3.10}$$

When the stochastic integral can be defined, X is an Itô process as introduced in the preceding chapter. The simplified notations $\mu_s = \mu(X_s, s)$ and $\sigma_s = \sigma(X_s, s)$ leads to write equation 3.10 in the general form as a **stochastic differential**:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dZ_t$$

Remark 4 *It is important to keep in mind that it is only a way to translate that X is written as a stochastic integral. Obviously, dZ_t is not a usual differential because the paths of the Brownian motion, though continuous, are nowhere differentiable.*

If the process X is the general Brownian motion W with parameters μ and σ (which are some constants), we write:

$$dW_t = \mu dt + \sigma dZ_t$$

We come back to the initial formulation by writing (remember that $W_0 = 0$):

$$\begin{aligned} \int_0^t dW_s &= W_t = \int_0^t \mu ds + \int_0^t \sigma dZ_s \\ &= \mu \int_0^t ds + \sigma \int_0^t dZ_s \\ &= \mu t + \sigma Z_t \end{aligned}$$

3.2.4 Calculation rules

The following proposition summarizes the calculation rules applied to increments of the Brownian motion (equivalently dZ_t). To avoid technical difficulties, the reader can think of dZ_t as a difference $Z_{t+dt} - Z_t$ where dt is an infinitesimal time-period[2]. The general intuition is that the higher order (>1) terms in dt may be neglected in developments because, on markets working in continuous-time, the higher order terms are infinitely small w.r.t. dt.

Proposition 24 *1. $E(dZ_t) = 0$ et $V(dZ_t) = dt$*

2. $V\left(dZ_t^2\right) = o(dt)$
3. $dZ_t.dt = o(dt)$
4. $E\left(dZ_{t_1} dZ_{t_2}\right) = 0$ *for* $t_1 \neq t_2$
5. *If Z and Z^* are two Wiener processes, we get:*

$$\begin{aligned} E\left(dZ_t dZ_t^*\right) &= r_t dt \\ V\left(dZ_t dZ_t^*\right) &= o(dt) \end{aligned}$$

Point (1) is obvious when dZ_t is identified to $Z_{t+dt} - Z_t$ since the standard Brownian motion is zero-mean and $V(Z_t - Z_s) = t - s$ for $t > s$.

[2] This interpretation is clearly disputable but it allows an intuitive presentation of the results.

Point (2) says that $(dZ_t)^2$ is not random or, more precisely, has a negligible variance. Remember the discrete-time approach with random walks. The increment of Z was $\sqrt{dt}$ or $-\sqrt{dt}$ with equal probabilities. Therefore $(dZ_t)^2$ is a constant equal to dt.

For point (3), $dZ_t = O(\sqrt{dt})$ implies $dZ_t.dt = O(dt^{\frac{3}{2}}) = o(dt)$.

As soon as $t_1 \neq t_2$, dZ_{t_1} and dZ_{t_2} are independent because the increments of a Brownian motion are independent. It proves point (4).

Finally, r_t is the correlation coefficient between increments of two Brownian motions, each being $O(\sqrt{dt})$. The product is then $O(dt)$. In the same way, the variance of the product is $O(dt^2)$, that is $o(dt)$.

These rules are useful to understand Itô's lemma presented hereafter.

3.3 Itô's lemma

In many valuation models, especially for derivative securities, one has to look for the dynamics of a regular function of an Itô process. The most well-known example is the valuation of a European call option with maturity T and strike price K. The final payoff of such an option is written:

$$C_T = \max(X_T - K; 0). \tag{3.11}$$

A valuation model needs the determination of the dynamics of the option price process $C = (C_t, t \in \mathcal{T})$ over its lifetime. The initial value C_0 is especially important for financial purposes. We assume that $C_t = g(X_t, t)$ where g is a function defined on $\mathbb{R}^+ \times \mathcal{T}$ taking values in $\mathbb{R}^+$. If X is an Itô process, the problem is to characterize the process C.

The same kind of problem arises when one wants to express prices (returns) starting from returns (prices). If X is an Itô process representing the cumulated continuous return of a financial security, what is the dynamics of the corresponding price process $Y_t = \exp(X_t)$? Obviously, the question can be asked the other way. If the dynamics of the price process Y is given, what is the dynamics of the continuous rate of return $X_t = \ln(Y_t)$?

An other example is the dynamics of bond prices as functions of the interest rate process. The problem is more complex because a bond price is determined by the term structure of interest rates, not by only one interest rate. To keep things very general, the bond price dynamics is a function of a family of processes driving interest rates. The models used in the financial literature are a simplified version of this complex reality. Some models are based on the assumption that bond prices only depend on the dynamics of the short rate, some others assume that they depend on the short rate and a long-term rate[3]. These simplifying assumptions are generally justified for practical implementation reasons.

The mathematical tool allowing to answer the preceding questions is the so-called Itô's lemma which is a kind of Taylor's series expansion for Itô processes.

[3] See for example Vasicek (1977), Cox, Ingersoll and Ross (1985), Longstaff and Schwartz (1992), Heath, Jarrow and Morton (1992) and Brace, Gatarek and Musiela (1997).

3.3.1 Taylor's formula, an intuitive approach to Itô's lemma

Let f denote a twice differentiable function defined from $\mathbb{R}^2$ to $\mathbb{R}$; the Taylor's series expansion of f at (x_0, t_0) is written as:

$$\begin{aligned} f(x,t) &= f(x_0,t_0) + \frac{\partial f}{\partial x}(x_0,t_0)(x-x_0) + \frac{\partial f}{\partial t}(x_0,t_0)(t-t_0) \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0,t_0)(x-x_0)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(x_0,t_0)(t-t_0)^2 \\ &+\frac{\partial^2 f}{\partial x \partial t}(x_0,t_0)(x-x_0)(t-t_0) + \varepsilon(x_0,t_0) \end{aligned} \quad (3.12)$$

where $\varepsilon(x_0, t_0) \sim o\left((x-x_0)^2 + (t-t_0)^2\right)$. The condition on ε shows that third-order terms are negligible w.r.t. first and second–order terms. In most cases, a first-order development is used in economic applications; it corresponds to the first line of the equation (3.12).

The story is a little bit different when x is an Itô process which is written as a stochastic integral w.r.t. a Brownian motion. The variation of this process on a time-interval of length $t - t_0$ is of order $\sqrt{t-t_0}$. Consequently, the second-order term $\frac{\partial^2 f}{\partial x^2}(x_0,t_0)(x-x_0)^2$ cannot be neglected because it is $O(t-t_0)$. It has the same magnitude as $\frac{\partial f}{\partial t}(x_0,t_0)(t-t_0)$.

Let us now denote $df(x_0,t_0) = f(x,t) - f(x_0,t_0)$; $t - t_0 = dt$ and $x - x_0 = dx$; equation 3.12 becomes :

$$\begin{aligned} df(x_0,t_0) &= \frac{\partial f}{\partial x}(x_0,t_0)dx + \frac{\partial f}{\partial t}(x_0,t_0)dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0,t_0)(dx)^2 \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial t^2}(x_0,t_0)(dt)^2 + \frac{\partial^2 f}{\partial x \partial t}(x_0,t_0)dxdt + \varepsilon(x_0,t_0) \end{aligned} \quad (3.13)$$

Replace x by X_t with X a stochastic process characterized by:

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dZ_t$$

Equation 3.13 (giving up the arguments of partial derivatives to simplify notations) becomes, for (X_t, t) :

$$\begin{aligned} df(X_t,t) &= \frac{\partial f}{\partial x}\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]+\frac{\partial f}{\partial t}dt \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]^2 \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2+\frac{\partial^2 f}{\partial x\partial t}\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]dt+\varepsilon(X_t,t) \end{aligned}$$

Applying now the calculation rules defined in proposition 24 allows to see that the coefficients of $\frac{\partial^2 f}{\partial t^2}$ and $\frac{\partial^2 f}{\partial x \partial t}$ are negigible (they are $o(dt)$). It then follows:

$$\begin{aligned} df(X_t,t) &= \frac{\partial f}{\partial x}\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]+\frac{\partial f}{\partial t}dt \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]^2+\varepsilon'(X_t,t) \end{aligned}$$

with $\varepsilon' = o(dt)$.

When developing the term $\left[\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right]^2$, one negligible term $(O(dt^2))$ appears, one term $dZ_t dt$ is also negligible (because it is $O(dt^{\frac{3}{2}})$) and, finally, one term dZ_t^2 which is $O(dt)$ and then not negligible. After all possible simplifications, it remains:

$$\begin{aligned} df(X_t,t) &= \frac{\partial f}{\partial x}\left(\mu(X_t,t)dt+\sigma(X_t,t)dZ_t\right)+\frac{\partial f}{\partial t}dt \\ &+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2(X_t,t)dt+\varepsilon''(X_t,t) \end{aligned}$$

with $\varepsilon'' = o(dt)$.

We can now write the process $f(X_t, t)$ as a stochastic differential by grouping dt terms on one hand and dZ_t terms on the other hand:

$$df(X_t,t) = \left(\frac{\partial f}{\partial x}\mu(X_t,t)+\frac{\partial f}{\partial t}+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2(X_t,t)\right)dt+\sigma(X_t,t)\frac{\partial f}{\partial x}dZ_t+\varepsilon''(X_t,t)$$

This brief presentation gives the intuition of the result but to prove rigorously Itô's lemma needs more precautions. Especially, the $\varepsilon(X_t, t)$ are stochastic and saying "ε is negligible w.r.t. dt" is not sufficiently precise. Readers can find a complete demonstration of Itô's lemma in specialized books[4].

[4] For example Karatzas-Shreve (2000), chapter 3.

3.3.2 Itô's lemma

We can now write more formally **Itô's lemma** which is nothing else than a Taylor formula in a specific stochastic environment.

Proposition 25 *Let X be an Itô process characterized by*

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dZ_t \tag{3.14}$$

and let $f : \mathbb{R}^2 \to \mathbb{R}$ a function with continuous partial derivatives up to order 2. The process Y defined by $Y_t = f(X_t, t)$ is an Itô process with a stochastic differential given by:

$$dY_t = \left(\frac{\partial f}{\partial x}\mu(X_t, t) + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2(X_t, t) \right) dt + \sigma(X_t, t)\frac{\partial f}{\partial x}dZ_t$$

If we write the stochastic differential of Y in the following form:

$$dY_t = \mu_Y (Y_t, t) \, dt + \sigma_Y (Y_t, t) \, dZ_t$$

we get:

$$\begin{aligned} \mu_Y (Y_t, t) &= \frac{\partial f}{\partial x}\mu(X_t, t) + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2(X_t, t) \\ \sigma_Y (Y_t, t) &= \sigma(X_t, t)\frac{\partial f}{\partial x} \end{aligned}$$

$\mu_Y (Y_t, t)$ is the drift of the process $f(X_t, t)$ and $\sigma_Y (Y_t, t)$ is the corresponding diffusion coefficient.

3.3.3 Applications

From return to price

Let W denote a Wiener process with parameters μ and σ (which are constant for W) and Y the process defined by $Y_t = f(W_t) = \exp(W_t)$. Y is then the transformation of the Brownian motion by the exponential function. Remark that t does not enter the transformation implying $\frac{\partial f}{\partial t} = 0$. The dynamics of Y is obtained by applying Itô's lemma.

$$\begin{aligned} \mu_Y (Y_t, t) &= \frac{\partial f}{\partial x}\mu + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2 = \exp(W_t) \left(\mu + \frac{\sigma^2}{2} \right) = Y_t \left(\mu + \frac{\sigma^2}{2} \right) \\ \sigma_Y (Y_t, t) &= \sigma\frac{\partial f}{\partial x} = \sigma Y_t \end{aligned}$$

or, equivalently:

$$\frac{dY_t}{Y_t} = \left(\mu + \frac{\sigma^2}{2}\right) dt + \sigma dZ_t$$

Y (which may represent the price of a stock) is called a **geometric Brownian motion**.

From price to return

Symetrically, let Y denote a price process characterized by[5] :

$$\begin{aligned} Y_0 &= 1 \\ dY_t &= \mu Y_t dt + \sigma Y_t dZ_t \end{aligned}$$

Let X be defined as $X_t = g(Y_t) = \ln(Y_t)$. In this case, we get:

$$\begin{aligned} \frac{\partial g}{\partial x}\mu(Y_t,t) + \frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\sigma^2(Y_t,t) &= \frac{\partial g}{\partial x}\mu Y_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\sigma^2 Y_t \\ &= \mu - \frac{\sigma^2}{2} \\ \sigma Y_t \frac{\partial g}{\partial x} &= \sigma \end{aligned}$$

These equalities lead to:

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dZ_t$$

Interest rates and the Ornstein-Uhlenbeck process

An Ornstein-Uhlenbeck process is a process X with the following stochastic differential:

$$dX_t = \alpha(\beta - X_t)dt + \sigma dZ_t$$

This stochastic process is often used to describe the dynamics of short-term rates[6]. We observe that the drift is positive when the short rate X_t is

[5] This process is called "geometric Brownian motion". It is the usual process to modelize stock prices, especially in the option valuation model of Black and Scholes (1973).

[6] Vasicek (1977) was the first to propose this process for short term rates.

low (lower than β) and the short rate tends to rise in the short run. On the contrary, when X_t is high, the short rate tends to decrease. So this process is said "mean-reverting". The short rate evolves around its long term mean[7] β. The coefficient α measures the strength of mean reversion.

In discrete-time, we encountered such a process when we described random draws in an urn (with white and black balls) without replacement.

Interest rates and the square root process

In the Ornstein Uhlenbeck process, the diffusion coefficient σ doesn't depend on X_t. As variations of Z during an interval of length dt are of order $\sqrt{dt}$, this process can take negative values because, on the short run, the variation of Z is much larger than the variation coming from the drift.

To solve this problem, Richard (1978) and then Cox-Ingersoll-Ross (1985) proposed the "square root process" defined as follows:

$$dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dZ_t$$

We observe that the variance of the short rate variations is proportional to the level of the short rate. Especially, if the rate reaches 0, only the drift term $\alpha\beta dt$ remains and is positive. Consequently, this process cannot take negative values.

[7] Warning: β is the long-term mean of the short rate, not a long-term rate!

3.4 The Girsanov theorem

3.4.1 Preliminaries

Let W denote a Brownian motion with parameters (μ, σ), defined on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ where $\mathcal{F}$ is the natural filtration of W. It is also the natural filtration of Z, since $W_t = \mu t + \sigma Z_t$. Assume that $W_t = \ln(S_t)$ where S_t is a stock price. $W_t - W_s$ is then the logarithmic return of the stock on the time period $[s; t]$. Denote r the risk-free rate (assumed constant). The no arbitrage assumption implies there exists a risk-neutral probability measure Q under which the drift of W is equal to r. As we deal

with two probability measures P and Q in this section, we denote E_P and E_Q the expectation with respect to P and Q.

Girsanov theorem is the technical tool that allows to transform a process with drift μ in a process with drift $r \neq \mu$.

In order to introduce this result, we first present the method which allows to change the mean of a Gaussian random variable.

Proposition 26 *Let $X \sim \mathcal{N}(0,1)$ defined on $(\Omega, \mathcal{A}, P)$ and Q defined by:*

$$\forall A \in \mathcal{A},\ Q(A) = E_P\left[\mathbf{1}_A \exp\left(\alpha X - \frac{\alpha^2}{2}\right)\right]$$

Q is equivalent to P and $X \sim \mathcal{N}(\alpha, 1)$ under Q.

Proof. As $E_P\left[\exp\left(\alpha X - \frac{\alpha^2}{2}\right)\right] = \exp\left(\frac{-\alpha^2}{2}\right) E_P\left[\exp(\alpha X)\right]$ and $\exp(\alpha X)$ is lognormal with parameters $(0, \alpha)$, the properties of the lognormal distribution imply that $E_P\left[\exp(\alpha X)\right] = \exp\left(\frac{\alpha^2}{2}\right)$.

We then get $Q(\Omega) = E_P\left[\exp\left(\alpha X - \frac{\alpha^2}{2}\right)\right] = 1$. The fact that the probability measures P and Q are equivalent is obvious because of the exponential transformation which takes only strictly positive values.

We can write:

$$\begin{aligned} E_Q[X] &= \int_\Omega X dQ = \int_\Omega X \exp\left(\alpha X - \frac{\alpha^2}{2}\right) dP \\ &= \int_{-\infty}^{+\infty} x \exp\left(\alpha x - \frac{\alpha^2}{2}\right) f_X(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(\alpha x - \frac{\alpha^2}{2}\right) \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{1}{2}(x-\alpha)^2\right) dx = \alpha \end{aligned}$$

where f_X is the density of X under P. We observe that X is, under Q, a Gaussian variable with expectation α and variance 1. Therefore, the variable $\exp\left(\alpha X - \frac{\alpha^2}{2}\right)$ is the Radon-Nikodym derivative of Q w.r.t. P, denoted $\frac{dQ}{dP}$.

■

In a second step, a more sophisticated transformation may be introduced by considering $X_t \sim \mathcal{N}(\mu t, \sigma^2 t)$ with $\mu t = E_P(X_t)$ and $\sigma^2 t = V_P(X_t)$. Suppose you want to transform X_t in a Gaussian variable satisfying $rt = E_Q(X_t)$ and $\sigma^2 t = V_Q(X_t)$. In economic terms, it means that you start in an economy with risk averse agents and you go in another economy with risk neutral agents (the expected return on the risky asset is now the risk-free rate).

Rewrite X as:

$$X_t = \mu t + \sigma Z_t = \mu t + \sigma\sqrt{t} Y_t$$

where Z is a standard Brownian motion and $Y_t \sim \mathcal{N}(0,1)$. To get the desired result, you need to transform Y_t in a gaussian variable following $\mathcal{N}(\alpha, 1)$ with a coefficient α to be defined. However, you look for Q such that:

$$E_Q[X_t] = \mu t + \sigma\sqrt{t}\alpha = rt$$

It implies that α must be chosen as follows:

$$\alpha = \frac{(r-\mu)\sqrt{t}}{\sigma}$$

Proposition 26 says that the following transformation is necessary:

$$\frac{dQ}{dP} = \exp\left[\frac{(r-\mu)\sqrt{t}}{\sigma} Y_t - \frac{1}{2}\left(\frac{(r-\mu)\sqrt{t}}{\sigma}\right)^2\right] = \exp\left[-\frac{\mu - r}{\sigma} Z_t - \frac{t}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right]$$

We can now present Girsanov theorem which generalizes the transformation presented above.

3.4.2 Girsanov theorem

Let $\lambda = (\lambda_t, t \in [0;T])$ denote a process adapted to $\mathcal{F}$ and $L = (L_t, t \in [0;T])$ defined by:

$$L_t = \exp\left(-\int_0^t \lambda_s dZ_s - \frac{1}{2}\int_0^t \lambda_s^2 ds\right)$$

Definition 31 *λ satisfies the Novikov condition if and only if:*

$$E_P\left[\exp\left\{\frac{1}{2}\int_0^T \lambda_s^2 ds\right\}\right] < +\infty$$

Proposition 27 *If λ satisfies the Novikov condition, we get the following properties:*
1) The process L is a $P-$martingale
2) The process Z^* defined by:

$$Z_t^* = Z_t + \int_0^t \lambda_s ds$$

is a Brownian motion on $(\Omega, \mathcal{A}, \mathcal{F}, Q)$ where Q is characterized by:

$$\frac{dQ}{dP} = L_T$$

It is worth noting that when λ is constant, the martingale property for L has already been proved in the preceding chapter. In fact, we have in this case:

$$L_t = \exp\left(-\lambda Z_t - \frac{\lambda^2}{2}t\right)$$

and we proved it is a martingale.

3.4.3 Application

The most common application of Girsanov theorem comes in valuing options by the risk-neutral approach.

In the Black-Scholes model, the dynamics of the stock price is:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t$$

which leads to:

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma Z_t\right)$$

The no arbitrage assuption implies that the discounted price process is a martingale under the risk-neutral probability

But the discounted price is:

$$\exp(-rt)\, S_t = S_0 \exp\left(\left(\mu - r - \frac{\sigma^2}{2}\right)t + \sigma Z_t\right)$$

This process is transformed by Girsanov theorem in:

$$S_t = S_0 \exp\left(-\frac{\sigma^2}{2}t + \sigma Z_t^*\right)$$

with $Z_t^* = Z_t + \left(\frac{\mu - r}{\sigma}\right)t$.

3.5 Stochastic differential equations

3.5.1 Existence and unicity of solutions

When building the stochastic integral, we proved that, under some conditions, a stochastic process is written as a stochastic integral. The symmetrical question is to know which assumptions on μ and σ have to be satisfied for a stochastic differential to define a stochastic process (with nice properties!) written as a stochastic integral.

In this section, it is assumed that the filtration $\mathcal{F}$ on $(\Omega, \mathcal{A}, P)$ is the natural filtration of a standard Brownian motion Z.

Definition 32 *A stochastic differential equation is given by a stochastic differential associated with a boundary condition, that is:*

$$\begin{aligned} X_0 &= c \\ dX_t &= \mu(X_t, t)\, dt + \sigma(X_t, t)\, dZ_t \end{aligned} \tag{3.15}$$

In the general case, c may be a random variable. However, in most financial models, c is a constant, for example the initial price of a financial asset or the initial short-term rate of interest.

Definition 33 *A stochastic process X is a solution of the equation appearing in definition 3.15 on $[0;T]$ if:*

1) X is adapted to $\mathcal{F}$

2) The functions μ and σ satisfy:

$$\int_0^T |\mu(X_t, t)|\, dt < +\infty \text{ and} \int_0^T \sigma^2(X_t, t)\, dt < +\infty \tag{3.16}$$

3) X satisfies:

$$X_t = X_0 + \int_0^t \mu(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dZ_s$$

The following proposition provides conditions on μ and σ for a stochastic differential equation to have a solution.

Proposition 28 *If conditions (a) and (b) hereafter are satisfied, equation 3.15 has a unique solution (P-a.s), which is a stochastic process X adapted to $\mathcal{F}$, with continuous paths, satisfying $E_P\left(\int_0^T X_t^2 dt\right) < +\infty$.*

a) There exists $m > 0$ such that $\forall t \in [0;T], \forall (x,y) \in \mathbb{R}^2$

$$\max\left(|\mu(x,t) - \mu(y,t)|\,;|\sigma(x,t) - \sigma(y,t)|\right) \leq m\,|x-y|$$
$$\mu(x,t)^2 + \sigma(x,t)^2 \leq m\left(1+x^2\right) \text{ for any pair } (x,t)$$

b) X_0 is square integrable, independent of $\mathcal{F}_t$ for any t.

A detailed proof of this result may be found in Oksendal (2000), p66.

The condition appearing in the first part of point (a) is called a Lipschitz condition. It limits the slopes of functions μ and σ which must be finite and bounded by a constant which doesn't depend on t. This condition is standard when solving usual differential equations. The second part of condition (a) puts some restrictions on the growth of μ and σ. As μ is the instantaneous expectation of X variations, the condition means that the LHS must be of order $(1+x^2)^{\frac{1}{2}}$. In other words, we cannot have an "exploding" drift. If the condition were not satisfied, the drift would grow too rapidly with the level reached by the process. For example, it would be the case with $\mu(x,t) = \exp(x)$.

It is worth noting that the solution provided here is called a "strong solution" because the Wiener process Z and the filtration are given. If we were solving the problem for the pair (X, Z), starting from functions μ and σ, we would speak about "weak solutions" (Karatzas-Shreve, 2000).

In the following, we always denote X_t a solution of equation 3.15 and assume that the conditions about existence and unicity of a solution are satisfied.

Proposition 29 *1) The solution of equation 3.15 is a Markov process whose initial distribution is the same as the one of c.*

2) If μ and σ are continuous functions of t, X is an Itô process with parameters μ and σ.

3.5.2 A specific case: linear equations

A stochastic differential equation (SDE in the following) is said linear if it is written as:

$$\begin{aligned} X_0 &= c \\ dX_t &= aX_t dt + \sigma_t dZ_t \end{aligned}$$

where a is a constant.

For this type of equation, the solution is written as:

$$X_t = c\exp(at) + \int_0^t \exp\left[a\left(t-s\right)\right]\sigma_s dZ_s \tag{3.17}$$

In fact, equation 3.17 may be transformed in:

$$X_t \exp(-at) = c + \int_0^t \exp\left(-as\right)\sigma_s dZ_s$$

The process in the RHS, denoted Y, may be written in the following way:

$$\begin{aligned} Y_0 &= c \\ dY_t &= \exp(-at)\sigma_t dZ_t \end{aligned}$$

If we write $\exp(at)Y_t = f\left(Y_t, t\right)$, the partial derivatives of f are given by:

$$\begin{aligned} \frac{\partial f}{\partial t} &= a\exp(at)Y_t \\ \frac{\partial f}{\partial Y_t} &= \exp(at) \\ \frac{\partial^2 f}{\partial Y_t^2} &= 0 \end{aligned}$$

Itô's lemma allows to write:

$$\begin{aligned} df\left(Y_t, t\right) &= a\exp(at)Y_t dt + \exp(at)\exp(-at)\sigma_t dZ_t \\ &= a\exp(at)Y_t dt + \sigma_t dZ_t \end{aligned}$$

Replacing Y_t by $\exp(-at)X_t$ leads to:

$$dX_t = aX_t dt + \sigma_t dZ_t$$

which is the initial stochastic differential.

More generally (see Malliaris-Brock (1982)), the following SDE:

$$\begin{aligned} X_0 &= c \\ dX_t &= a(t)X_t dt + \sigma_t dZ_t \end{aligned}$$

has the following solution X:

$$X_t = \gamma_t \left[c + \int_0^t \gamma_s^{-1} \sigma_s dZ_s \right]$$

where γ is a solution of the differential equation

$$f^{'}(t) = a(t)f(t)$$

Application

Let Y an Ornstein-Uhlenbeck process characterized by:

$$\begin{aligned} Y_0 &= y_0 \\ dY_t &= \alpha\left(\beta - Y_t\right)dt + \sigma dZ_t \end{aligned}$$

Define $X_t = (Y_t - \beta)\exp(\alpha t) = f(Y_t, t)$; the partial derivatives of f are given by:

$$\begin{aligned} \frac{\partial f}{\partial t} &= \alpha \exp(\alpha t)\left(Y_t - \beta\right) \\ \frac{\partial f}{\partial Y_t} &= \exp(\alpha t) \\ \frac{\partial^2 f}{\partial Y_t^2} &= 0 \end{aligned}$$

Itô's lemma leads to:

$$\begin{aligned} dX_t &= \left[\alpha\left(Y_t - \beta\right)\exp(\alpha t) + \exp(\alpha t)\alpha\left(\beta - Y_t\right)\right]dt + \sigma \exp(\alpha t)dZ_t \\ &= \sigma \exp(\alpha t)dZ_t \end{aligned}$$

As $X_0 = y_0 - \beta$, it follows:

$$X_t = y_0 - \beta + \sigma \int_0^t \exp(\alpha s)dZ_s$$

The relationship between X and Y may be written as:

$$Y_t = \beta + X_t \exp(-\alpha t)$$

Therefore:

$$\begin{aligned} Y_t &= \beta + \exp(-\alpha t)\left[y_0 - \beta + \sigma \int_0^t \exp(\alpha s) dZ_s\right] \\ &= \beta\left(1 - \exp(-\alpha t)\right) + y_0 \exp(-\alpha t) + \sigma \int_0^t \exp(-\alpha(t-s)) dZ_s \end{aligned}$$

Y_t follows a Gaussian distribution satisfying:

$$\begin{aligned} E_P\left[Y_t \,|Y_0 = y_0\right] &= \beta\left(1 - \exp(-\alpha t)\right) + y_0 \exp(-\alpha t) \\ V_P\left[Y_t \,|Y_0 = y_0\right] &= \sigma^2 \int_0^t \exp(-2\alpha(t-s)) ds \\ &= \frac{\sigma^2}{2\alpha}\left[1 - \exp\left(-2\alpha t\right)\right] \end{aligned}$$

Bibliography

[1] Biingham. N.H. and Kiesel. R. (2001), *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*, Springer Finance.

[2] Black.F and Scholes.M. (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637-54. D. and M. Musiela (1997)

[3] Brace. A., Gatarek D. and Musiela M. (1997), The Market Model of Interest Rate Dynamics, *Mathematical Finance*, 7, 127-155.

[4] Cox. J. and Ross. S. (1976), The Valuation of Option for Alternative Stochastic Processes, *Journal of Financial Economics* 3**,** 145-166.

[5] Cox. J., Ross. S. and Rubinstein.M (1979), Option Pricing : a Simplified Approach, *Journal of Financial Economics* 7**,** 229-263.

[6] Cox. J., Ingersoll. J. and Ross. S. (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-407.

[7] De Bondt. W. and Thaler. R. (1985), Does the Stock market Overreact?, *Journal of finance,* 40, 3, 793-805.

[8] Dothan. M. (1990), *Prices in Financial Markets*, Oxford university Press

[9] Duffie.D (1988), *Security Markets, Stochastic Models,* Academic Press.

[10] Fama, E. (1970), Efficient markets : a review of theory and empirical work, *Journal of finance,* 25*,* 2, 383-417

[11] Grossman, S.J. (1976), On the Efficiency of Competitive Stock Markets when Traders Have Diverse Information, *Journal of Economic Theory*, 31, 573-585.

[12] Harrison.J.M and Kreps.D.M (1979), Martingales and Arbitrage in Multi-period Securities Markets, *Journal of Economic Theory,* 20, p381-408.

[13] Harrison.J.M and Pliska.S (1981), Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stochastic Processes and their Applications,* 11, 215-60.

[14] Heath. D., Jarrow R. and Morton A. (1992), Bond pricing and the Term Structure of interest Rates: A New methodology, *Econometrica,* 60, 77-105.

[15] Huang.C.F and Litzenberger.R (1988), *Foundations for Financial Economics,* North-Holland.

[16] Hull, J.C. (2009), *Options, Futures, and other Derivatives,* Pearson, 7th ed., Prentice Hall.

[17] Karatzas. I. and Shreve. S.E.(2000), *Brownian Motion and Stochastic Calculus,* 2nd ed.*,* Springer.

[18] Koch-Medina.P. and Merino.S. (2003), *Mathematical Finance and Probability*, Birkhauser Verlag.

[19] Longstaff. F.A. and Schwartz. E.S. (1992), Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model, *Journal of finance*, 47, 1259-1282.

[20] Malkiel, B. (2003), *A Random Walk Down Wall Street*, W.W. Norton & Co. (original edition in 1973).

[21] Malliaris, A.G. and Brock, W.A. (1988), *Stochastic Methods in Economics and Finance*, North Holland.

[22] Markowitz.H (1952), Portfolio Selection, *Journal of Finance,* 7, 77-91.

[23] Markowitz.H (1959), *Portfolio Selection : Efficient Diversification of Investment,* Wiley New York.

[24] Merton, R., C. (1969), Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case, *Review of Economics and Statistics*, 50, 247–257.

[25] Merton, R., C. (1971), Optimal Consumption and Portfolio Rules in a Continuous-Time Model, *Journal of Economic Theory*, 3, 373-413.

[26] Merton, R., C. (1973), Theory of rational option pricing, *Bell Journal of Economics and Management Science,* 4, 141-83.

[27] Merton, R. C. (1976), Option Pricing when Underlying Returns are Discontinuous, *Journal of Financial Economics*, 3, 125-144.

[28] Merton, R. C. (1982), On the Mathematics and Economics Assumptions of Continuous-Time Models, in W.F. Sharpe and C.M. Cootner, eds, *Financial Economics: Essays in the Honor of paul Cootner*, Englewood Cliffs, NJ: Prentice Hall.

[29] Merton, R. C. (1990), *Continuous-Time Finance*, Basil Blackwell.

[30] Mossin.J (1966), Equilibrium in a Capital Asset Market, *Econometrica,* 34, 261-76.

[31] Ő̀ksendal. B. (2000), *Stochastic Differential Equations: An Introduction with Applications*, 5th ed., Springer

[32] Richard. S.F. (1978), An arbitrage model of the term structure of interest rates, *Journal of Financial Economics*, 6, 33-57.

[33] Roger. P. (2010), *Probability for Finance*, Ventus Publishing, www.bookboon.com.

[34] Vasicek O. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188.

Index