Stochastic Calculus Toolbox, Part I

Stochastic Calculus and Probability

CQF

In this lecture...

... we expand on the stochastic calculus lectures to introduce further probabilistic concepts:

- The probabilistic ecosystem:
  - sample space;
  - filtration;
  - measures;
- Conditional and unconditional expectation;
- Change of measure and the Radon-Nikodym derivative;
- Further Itô calculus: tips & tricks

Introduction

In the quantitative finance literature, most articles written in the past 15 years start with the words:

"*Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space...*"

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space, and its inclusion in quantitative finance articles and books reflects the increasing influence of probability theory and probabilists over the subject.

1 The Probabilistic Ecosystem

In this section, we discuss the triple $(\Omega, \mathcal{F}, \mathbb{P})$, constituted of

1. the sample space Ω;
2. the filtration $\mathcal{F}$;
3. the probability measure $\mathbb{P}$.

This triple is called a *probability space* and it represents the foundations of the probabilistic universe. The probability space gives their structure to the concepts of Brownian motion, Itô calculus and stochastic differential equations that you are already familiar with.

Whether it is explicitly mentioned or not in the articles or books you read, the probability space is always there somewhere in the background.

1.1 The Sample Space Ω

1.1.1 Events in the Sample Space

An **event**, or **outcome** is the result of a given experiment.

Example

Obtaining a “Head” in a coin toss experiment is an event.

Usually, events are denoted by the Greek lower cap letter omega, ω.

The **sample space** is the set of all possible events. Generally, the sample space is denoted by the Greek capital letter Omega, Ω.

Example

1 if our experiment is the toss of one coin, the sample space Ω_c will be defined as:

$$\Omega_c = \{\{H\}, \{T\}\}$$

2 if our experiment is the throw of one dice, the sample space Ω_d will be defined as:

$$\Omega_d = \{1, 2, 3, 4, 5, 6\}$$

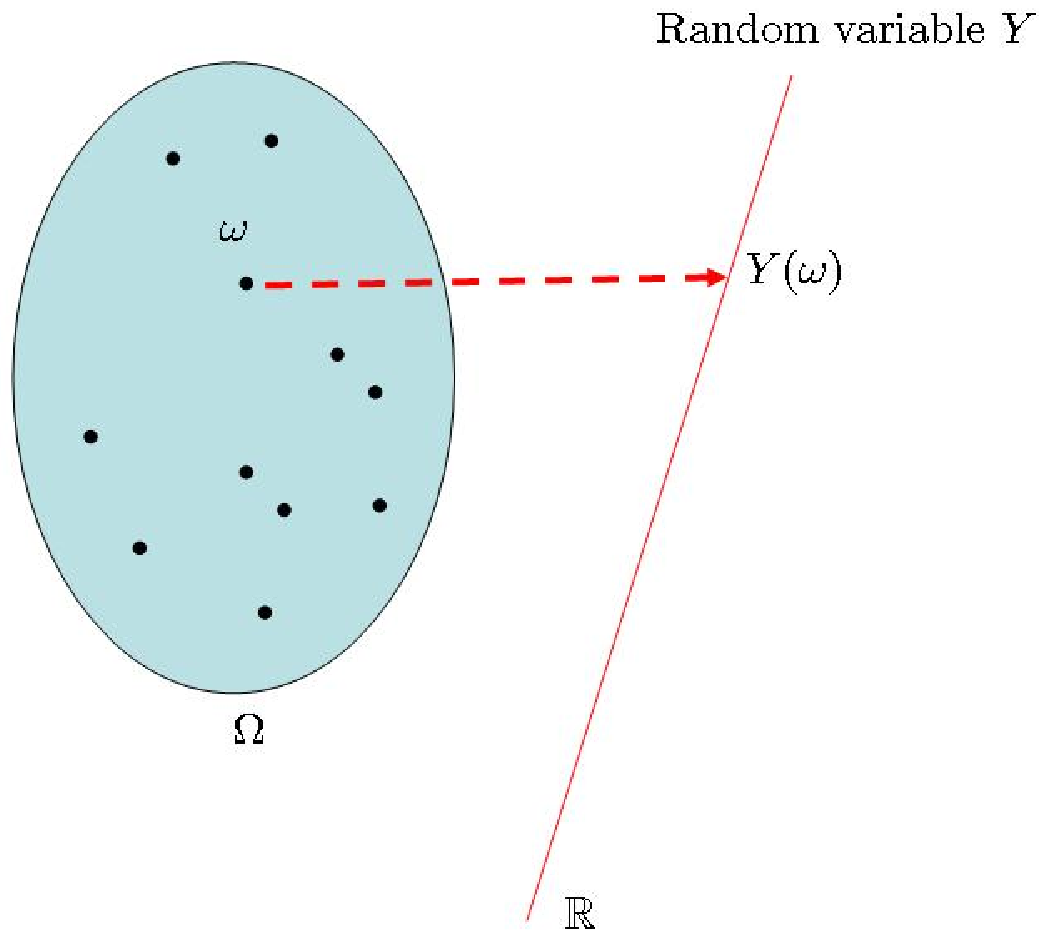
This framework is built on **set theory**. Elementary events or outcomes are elements of the sample space Ω, defined as the set of all outcomes as well as their unions and intersections.

1.1.2 Events, Random Variables and Stochastic Processes

We can view a **random variable** Y as a *function* assigning a numerical value to each individual event $\omega \in \Omega$.

To reflect this idea, we write Y more precisely as $Y(\omega)$ where $\omega \in \Omega$ is an event.

Figure : Sample space, events and random variables



Example

Take the coin toss game in which you gain 1 if the toss produces a Head and lose 1 if the toss produces a Tail.

The sample space Ω has two events:

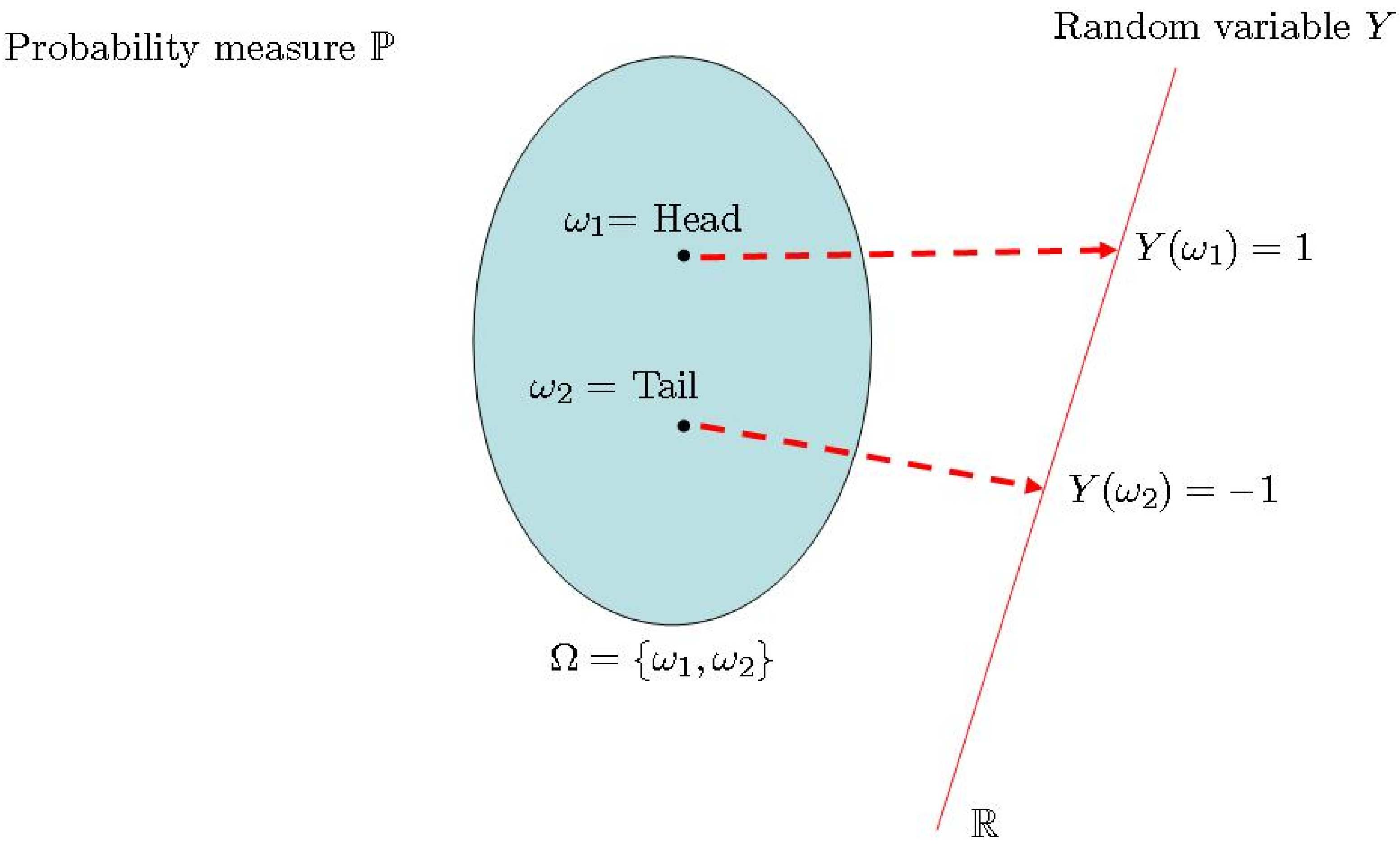
- ω_1 = Head;
- ω_2 = Tail.

so $\Omega = \{\omega_1, \omega_2\}$.

The profit and loss (P&L) from the game is a random variable Y defined by

- $Y(\omega_1) = 1$;
- $Y(\omega_2) = $ -1.

Figure : Sample space, events and random variables in a coin toss game



A **stochastic process** $S(t)$ is a sequence of random variables indexed by time t

Hence, a stochastic process $S(t)$ is a function of both the individual events ω and time t. We express this idea more precisely by denoting the process by $S(\omega)(t)$.

Remark

The word "stochastic" comes from the Greek $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa\acute{o}\varsigma$ meaning "to aim at, to guess at." It was first used in German ("stochastik") in the 1930s with the current meaning of "random".

The term "process" comes from the old French *proces*, itself derived from the Latin *procedo*. It means "journey" and emphasises the chronological dimension of the stochastic process.

1.1.3 Sample Paths and Sample Sets for Stochastic Processes

Let's take the simple example of the binomial model in order to articulate the intuition behind the meaning of **sample space** and **sample path** of stochastic processes.

After two time periods, we can write the **sample space** for the stock price trajectories as

$$\Omega_2 = \{UU, UD, DU, DD\}$$

where U and D respectively represent the event "up-move" and "down-move"

Then, a **sample path** for the stock is an individual trajectory, such as DU.

Remark

Our formulation for the sample space is not unique.

We could have built our sample space based on the stock price rather than its trajectory. If the tree is recombining, we would therefore lose information with this formulation as the outcome UD becomes indistinguishable from the outcome DU.

As a general rule, we will formulate the sample space in order to generate as much information as possible on the experiment.

Now what happens after the end of the third time period?

After three time periods, the sample space for the stock becomes:

$$\begin{aligned}\Omega_3 \quad = \quad & \{UUU, UUD, UDU, UDD, DUU, \\ & DDU, DUD, DDD\}\end{aligned}$$

Clearly, the sample space has expanded to reflect additional information related to the third movement of the stock.

Underneath, we can still see the possibilities for the first two move, i.e. UD, DU, DD, but now each of these possibilities has itself branched out to give additional outcomes, resulting in an expanding sample space.

As the number of time periods becomes larger and larger, it becomes increasingly difficult to track all of the possible outcomes and all of the sample space generated through time (i.e $\Omega_1, \Omega_2, \Omega_3, \ldots, \Omega_t, \ldots$).

So, how can we keep track of an ever expanding sample space?

The answer is simple and elegant: keep track of the increasing flow of information separately from the sample space.

To keep track of the increasing flow of information we need to define a new mathematical object: the **filtration**.

1.2 The Filtration $\mathcal{F}$

The filtration, $\mathcal{F}$, is an indication of how information about the experiment builds up over time as more results become available. It can be thought of as an increasing family of events.

More than just a family of events, the filtration $\mathcal{F}$ is a set formed of all possible combinations of events $A \subset \Omega$, their unions and complements.

Technically, the filtration $\mathcal{F}$ is a well-defined object called a σ-field (a concept from Measure theory).

Concretely, this definition implies 2 rules:

1. $A \subset \mathcal{F} \Rightarrow A^c \subset \mathcal{F}$;
2. $A_i \subset \mathcal{F} \quad \forall i = 1, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i \subset \mathcal{F}$.

Remark

As a corollary to these two rules we see that:

3. 2. implies that $\Omega \subset \mathcal{F}$;
4. 1. and 3. imply that $\emptyset \subset \mathcal{F}$;
5. 1. and 2. imply that $\bigcap_{i=1}^{\infty} A_i \subset \mathcal{F}$.

How can we see the filtration?

In most cases, it is very difficult to describe explicitly the filtration. The binomial model is a notable exception.

Consider a 3-period binomial model. At the end of each period, new information becomes available to help us predict the true trajectory of the stock.

At time 0, before the start of trading, we only have the trivial filtration

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

since we do not have any information regarding the trajectory of the stock.

After the first period, we are in a position to tell whether the stock started by moving up or down. This conditions our prediction of how the stock could behave in future periods.

After the first time period, the filtration is given by:

$$\mathcal{F}_1 = \{\Omega, \emptyset, u, d\}$$

where

$$u = \{UUU, UUD, UDU, UDD\}$$

represents the possible paths the stock can take if the first move was an up-move, and

$$d = \{DDD, DDU, DUD, DUU\}$$

represents the possible paths the stock can take if the first move we a down-move.

After the second time period we know the first two moves of the stock trajectory.

The filtration starts to take shape and becomes more complicated.

$$\mathcal{F}_2 = \{\Omega, \emptyset, ud, du, dd, uu \cup ud \dots\}$$

where $uu = \{UUU, UUD\}$, $ud = \{UDU, UDD\}$, $du = \{DUU, DUD\}$, $dd = \{DDU, DDD\}$ represent the possible path conditioned by the information we have available at the end of the second period.

At the end of our experiment, after the third period, we know with certainty what the true path of the stock was.

Because of the inclusion of all intersections and unions, as the number of period increases, the filtration quickly become enormous.

However, we will not need to worry about it: the mathematical object $\mathcal{F}$ is keeping track of all this for us.

When we work in continuous time, we denote the filtration slightly differently.

For an experiment starting at time 0 and ending at time T, we define the **filtration** as the set $\{\mathcal{F}\}_{t\in[0,T]}$.

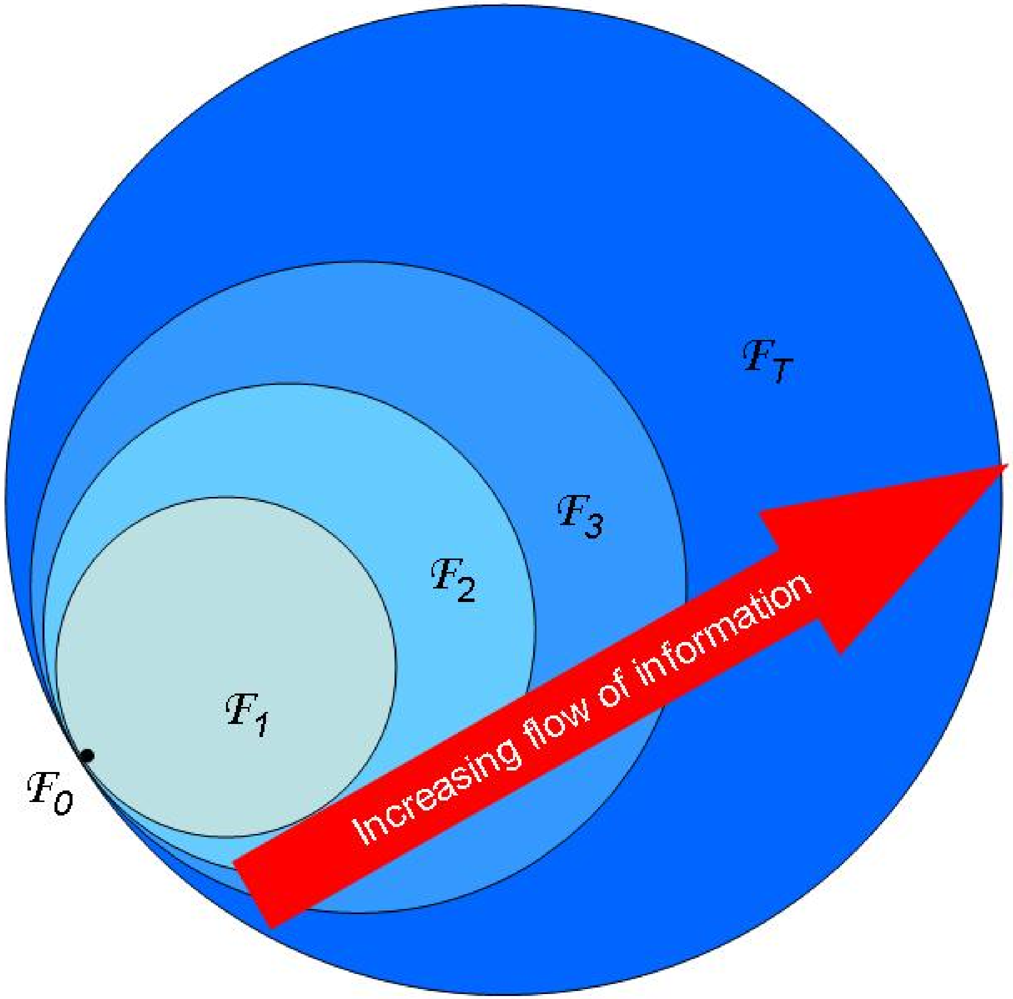
It is important to keep in mind a very important, but quite intuitive fact:

Key Fact

For $0 \leq t_1 \leq t_2 \leq T$,

$$\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \subseteq \mathcal{F}_T \equiv \mathcal{F}$$

Figure : With the passage of time information accumulates and the filtration increases



This property is quite intuitive. Since we consider that information gets constantly recorded and accumulated without ever getting lost or forgotten, it is only logical that as time goes by we have an increasing amount of information.

So information accumulates up until the end of the experiment at time T.

The filtration at time T therefore represent all the information we will ever have on this specific experiment.

Definition (Adapted (Measurable) Process)

A stochastic process S_t is said to be **adapted to the filtration** $\mathcal{F}_t$ (or **measurable with respect to** $\mathcal{F}_t$, or $\mathcal{F}_t$-**adapted**) if the value of S at time t is known given the information set $\mathcal{F}_t$.

1.3 The Probability Measure $\mathbb{P}$

$\mathbb{P}$ is the probability measure, a special type of “function”, called a measure, assigning probabilities to subsets (i.e. the outcomes).

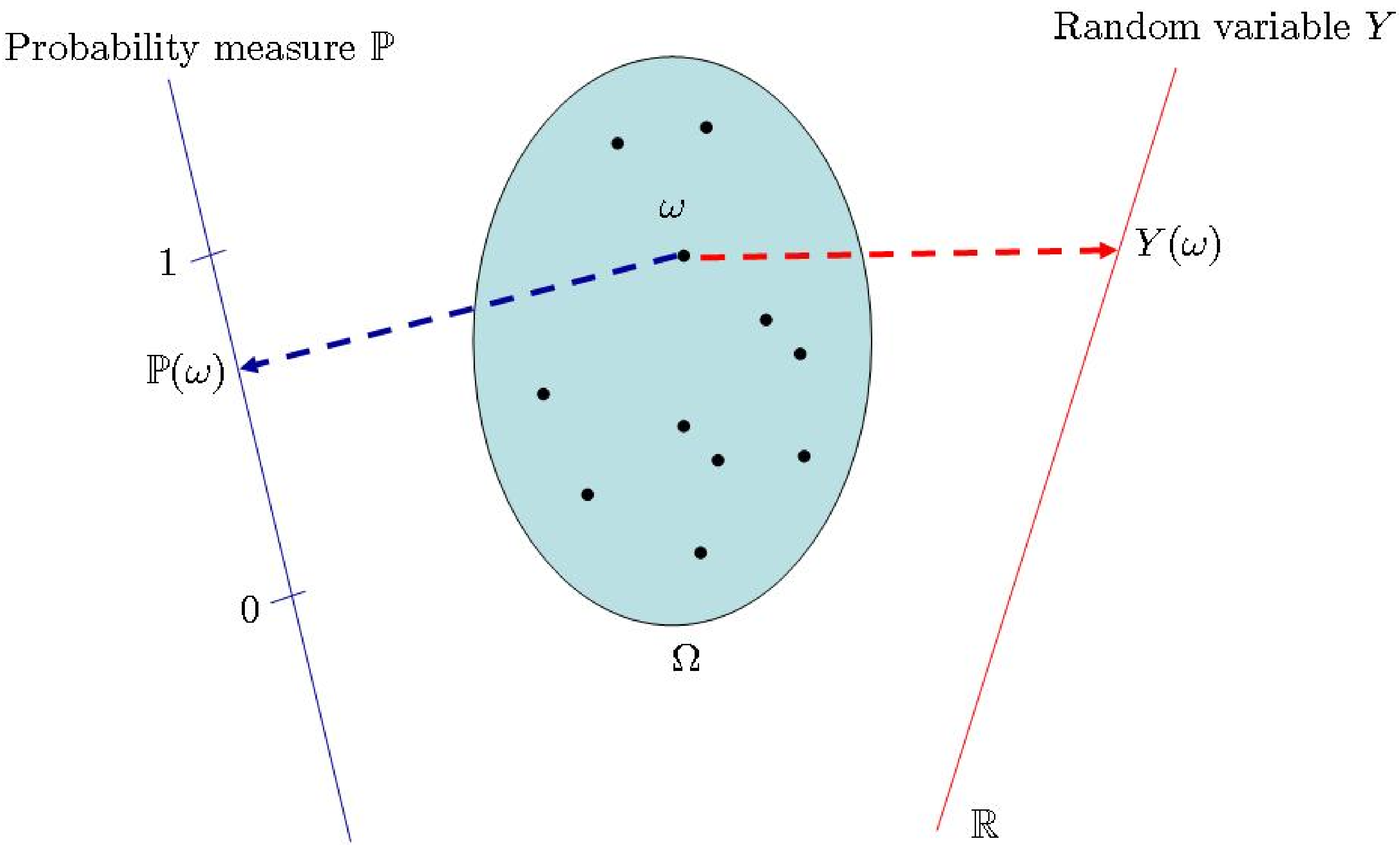
Unsurprisingly, the mathematics of probability measures come from the field of Measure Theory.

Probability measures are similar to the cumulative density functions (CDF) we are used to manipulating. The main difference is that probability measures are defined on general sets while CDFs are defined on intervals (such as $\mathbb{R}$).

1.3.1 Probability Measure, Events and Random Variable

A **probability measure** is a *function* which assigns to every individual events $\omega \in \Omega$ a number in the interval $[0, 1]$, known as the probability of event ω.

Figure : Probability measure $\mathbb{P}$



Example

Take the coin toss game in which you gain 1 if the toss produces a Head and lose 1 if the toss produces a Tail.

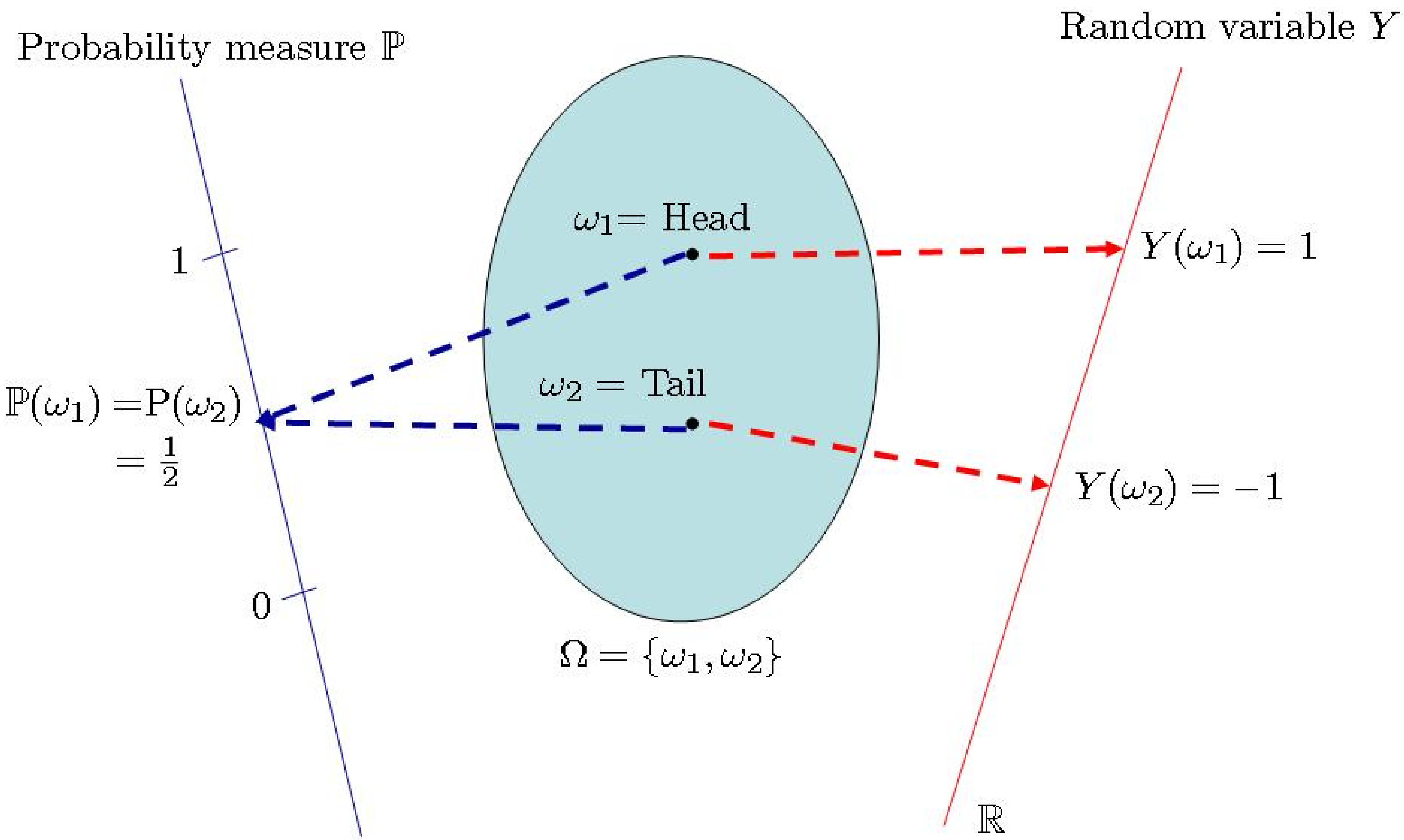
The sample space Ω has two events:

- $\omega_1 =$ Head;
- $\omega_2 =$ Tail.

so $\Omega = \{\omega_1, \omega_2\}$.

If the coin is fair, the two events will be equiprobable and the probability measure $\mathbb{P}$ is defined as $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \frac{1}{2}$.

Figure : Probability measure $\mathbb{P}$ in a coin toss game



1.3.2 Axioms of Probability Theory

Practically, a probability measure $\mathbb{P}$ satisfies the three defining properties, or axioms, of probabilities:

Axioms

1. $0 \leq \mathbb{P}(A) \leq 1$;
2. $\mathbb{P}(\Omega) = 1$;
3. For any mutually exclusive events A_i, $i = 1, \ldots n \leq +\infty$,

$$\mathbb{P}(\cup_{i=1}^{n} A_i) = \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

1.3.3 Probabilities and Expectations Revisited

The definition of expectation (and therefore variance, covariance and all other moments) is intimately linked with the concept of integration. Indeed, taking the expectation of a random variable X is equivalent to integrating this variable with respect to the differential of a CDF.

We are used to the definition of the mathematical expectation for real-valued random variables:

$$\begin{aligned} \mathbb{E}\left[h(X)\right] &= \int_{\mathbb{R}} h(x)p(x)dx \\ &= \int_{\mathbb{R}} h(x)d(P(x)) \end{aligned}$$

where p is the PDF and P is the CDF.

Example

To compute the expectation of a normal random variable $X \sim N(\mu, \sigma^2)$ above some level K, we start by writing

$$h(x) = [x - K]^+ = \max[x - K, 0]$$

$$\begin{aligned} P(x) &= \Phi(x) \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\} dz \end{aligned}$$

So,

$$\begin{aligned} \mathbb{E}\left[[X-K]^{+}\right] &= \int_{\mathbb{R}} [x-K]^{+} dP(x) \\ &= \int_{K}^{+\infty} (x-K) dP(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{K}^{+\infty} (x-K) \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right\} dx \end{aligned}$$

Now, all we have to do is perform the relatively innocuous task of evaluating this integral.

Remark

This example is one of the main ideas in the derivation of the option pricing formula from a probabilistic perspective (more on this in Module 3).

To see this, set $x := \log S$ and $K := \log E$. Because we assume that the share price S are lognormally distributed, the log return on the share, defined as $\log S$ is normally distributed.

In our probabilistic universe $(\Omega, \mathcal{F}, \mathbb{P})$, the mathematical expectation is formulated as a (Lebesgue) integral with respect to the measure $\mathbb{P}$:

$$\mathbb{E}\left[h(X)\right] \;=\; \int_\Omega h(x) d\mathbb{P}$$

In addition, there is a very useful *relation between expectations and probabilities*.

In our formula for the expectation, choose h(x) to be the **indicator function** $\mathbf{1}_{x \in A}$ for a subset $A \subset \Omega$.

The indicator function $\mathbf{1}_{x \in A}$ is a function returning 1 if $x \in A$ and 0 if $x \notin A$.

The expectation of the indicator function of an event is the probability associated with this event:

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{X\in A\}}\right] &= \int_\Omega \mathbf{1}_{\{x\in A\}} d\mathbb{P} \\ &= \int_A d\mathbb{P} \\ &= \mathbb{P}(A)\end{aligned}$$

which is simply the probability that $X \in A$.

This "trick" is very useful to convert a complicated-looking expectation into a nicer-looking probability. In particular, we use this technique at some key points in the probabilistic derivation of the Black-Scholes formula (more on this in Module 3).

2 Conditional And Unconditional Expectation

So far we have only discussed very general *unconditional* expectations.

Now we introduce *conditional* expectations, which are expectation conditional on some information.

2.1 Example

The difference between conditional and unconditional expectations can be viewed through a simple card game.

We start by assigning a numerical value to each card:

- Cards 1 to 10 are worth 1 point each;
- The Jake is worth 2 points;
- The Queen is worth 4 points;
- The King is worth 5 points.

Example (Unconditional expectation)

Pick a card. What is the expected point value of the card?

Example (Conditional expectation)

Assume that we have already picked two cards: a King and a Jake, and that we have not returned these cards to the deck. Pick a card. What is the expected point value of the card?

Hence, the difference between conditional and unconditional expectation is **information**.

In our probabilistic universe $(\Omega, \mathcal{F}, \mathbb{P})$, information is represented by the filtration $\mathcal{F}$. It therefore seems natural to consider *conditional expectation with respect to the filtration* $\mathcal{F}$.

2.2 Definition

Now that we know about filtration and measures, we can extend the definition of conditional expectations to include expectations conditional on a filtration.

Mathematically, we define:

$$\mathbb{E}\left[X|\mathcal{F}\right]$$

for a random variable X and a filtration $\mathcal{F}$.

A couple of notes...

- In general, $Y = \mathbb{E}\left[X|\mathcal{F}\right]$ will be a random variable;
- $Y = \mathbb{E}\left[X|\mathcal{F}\right]$ is $\mathcal{F}$-adapted.

Condition expectations will come in handy when we define martingales later in the next lecture.

2.3 Properties of Conditional Expectations

Conditional expectations have the following useful properties:

1. *Linearity*:

$$\mathbb{E}\left[aX + bY|\mathcal{F}\right] = a\mathbb{E}\left[X|\mathcal{F}\right] + b\mathbb{E}\left[Y|\mathcal{F}\right]$$

2. *Tower Property (i.e. Iterated Expectations)*: if $\mathcal{F} \subset \mathcal{G}$,

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

In plain English, when you take iterated expectations with respect to several levels of information, you could as well take a single expectation with respect to the smallest set of information available.

3. As a special case of the Tower property, we have

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}\right]\right] = \mathbb{E}\left[X\right]$$

since "no filtration" is always a smaller information set than any filtration

4. *Taking Out What Is Known*: if X is $\mathcal{F}$-measurable, then the value of X is known once we know $\mathcal{F}$. Therefore,

$$\mathbb{E}\left[X|\mathcal{F}\right] = X$$

5. *Taking Out What Is Known (2)*: by extension, if X is $\mathcal{F}$-measurable but not Y,

$$\mathbb{E}\left[XY|\mathcal{F}\right] = X\mathbb{E}\left[Y|\mathcal{F}\right]$$

6. *Independence*: if X is independent from $\mathcal{F}$, then knowing $\mathcal{F}$ is useless to predict the value of X. Hence,

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

7. *Positivity*: if $X \geq 0$ then $\mathbb{E}[X|\mathcal{F}] \geq 0$.

8. *Jensen's Inequality*: let f be a convex function, then

$$f(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[f(X)|\mathcal{F}]$$

In the next lecture, we will see an example of how to use some of these rules to prove that a discrete time stochastic process is a martingale.

3 Changing Probability Measure

You have seen in the Binomial Model lecture that there is more than just one probability measure.

Indeed, the lecture introduced you to the distinction between the "real" or "physical" probability measure, which we encounter every day on our Bloomberg or Reuters screen, and the so-called "risk-neutral" measure, which is used for pricing.

Probability measures are by no means unique. We will see in the next lecture that the powerful arsenal of martingale techniques enables us, under certain assumptions, to change measure and transpose our problem subject to the real world measure into an equivalent problem formulated as a martingale under a different measure.

For now, we just outline the rules that allow us to define equivalent measures.

3.1 Equivalent Measure

If two measures $\mathbb{P}$ and $\mathbb{Q}$ share the same sample space Ω and if $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$ for all subset A, we say that $\mathbb{Q}$ is **absolutely continuous** with respect to $\mathbb{P}$ and denote this by $\mathbb{Q} << \mathbb{P}$.

The key point is that all impossible events under $\mathbb{P}$ remain impossible under $\mathbb{Q}$. The probability mass of the possible events will be distributed differently under $\mathbb{P}$ and $\mathbb{Q}$. In short "it is alright to tinker with the probabilities as long as we do not tinker with the (im)possibilities"

If $\mathbb{Q} << \mathbb{P}$ and $\mathbb{P} << \mathbb{Q}$ then the two measures are said to be **equivalent**, denoted by $\mathbb{P} \sim \mathbb{Q}$.

This extremely important result is formalized in the **Radon Nikodym Theorem**.

3.2 The Radon-Nikodym Theorem

Key Fact (The Radon-Nikodym Theorem)

If the measures $\mathbb{P}$ and $\mathbb{Q}$ share the same null sets, then, there exists a random variable Λ such that for all subsets $A \subset \Omega$

$$\mathbb{Q}(A) = \int_A \Lambda d\mathbb{P}$$

where

$$\Lambda = \frac{d\mathbb{Q}}{d\mathbb{P}}$$

is called the **Radon-Nikodym derivative**.

The formulation on the previous slide is the most general: it applies to both continuous and discrete distributions.

The Radon-Nikodym theorem simplifies considerably when we deal specifically with *discrete distributions*, such as coin toss, throw of a dice or the binomial model.

In particular, we have for all subsets $A \subset \Omega$

$$\mathbb{Q}(A) = \Lambda(A)\mathbb{P}(A)$$

where the Radon-Nikodym derivative is given by

$$\Lambda(\cdot) = \frac{\mathbb{Q}}{\mathbb{P}}(\cdot)$$

3.3 Example of Change of Measure

Take the coin toss game in which you gain GBP 1 if the toss produces a Head and lose GBP 1 if the toss produces a Tail.

The sample space Ω has two events:

- $\omega_1 =$ Head;
- $\omega_2 =$ Tail.

so $\Omega = \{\omega_1, \omega_2\}$.

In the general case when the coin is not necessarily "fair," the probability measure $\mathbb{P}$ is

- $\mathbb{P}(\omega_1) = p$;
- $\mathbb{P}(\omega_2) = q = 1 - p$;

with $0 < p < 1$.

Suppose that we want to evaluate our expected P&L in a world where all coins are fair.

To do the, we introduce a new probability measure $\bar{\mathbb{P}}$ such that

- $\bar{\mathbb{P}}(\omega_1) = \frac{1}{2}$;
- $\bar{\mathbb{P}}(\omega_2) = \frac{1}{2}$.

This implies that we must *add extra weight* to each of the two events to travel from the real world $(\Omega, \mathcal{F}, \mathbb{P})$ to the "fair" world $(\Omega, \mathcal{F}, \bar{\mathbb{P}})$. Namely,

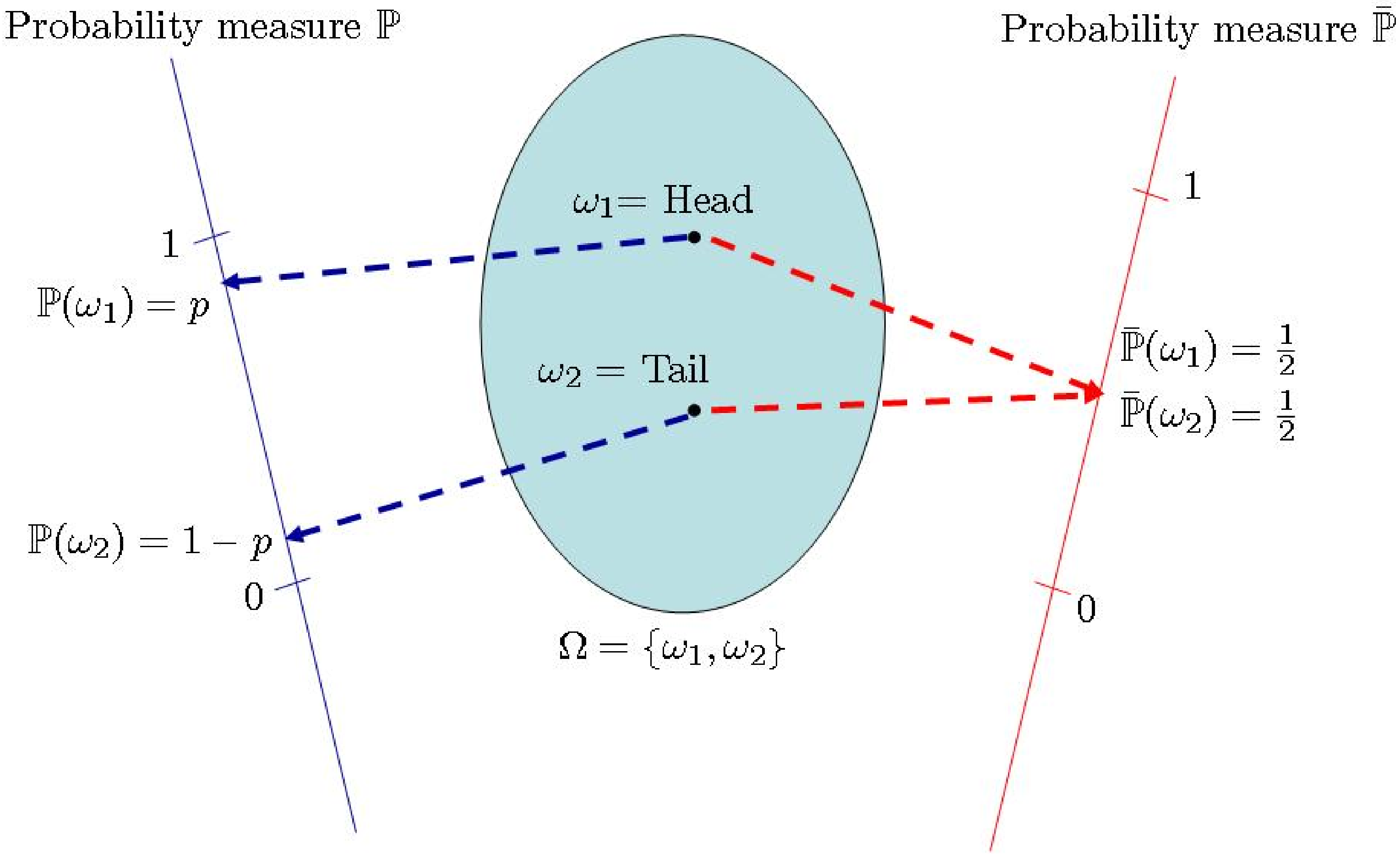
- we reweigth the likelihood of event ω_1 by $\frac{\bar{\mathbb{P}}}{\mathbb{P}}(\omega_1) = \frac{1}{2p}$;
- we reweigth the likelihood of event ω_2 by $\frac{\bar{\mathbb{P}}}{\mathbb{P}}(\omega_2) = \frac{1}{2q}$;

so that

$$\bar{\mathbb{P}}(\cdot) = \frac{\bar{\mathbb{P}}}{\mathbb{P}}(\cdot)\mathbb{P}(\cdot)$$

This last formula is the **Radon-Nikodým formula** for discrete processes!

Figure : Change of probability measure in a coin toss game



Note that since $0 < p < 1$, we have

- $\mathbb{P}(\omega_1) \neq 0$;
- $\mathbb{P}(\omega_2) \neq 0$.

Hence the measure $\mathbb{P}$ and $\bar{\mathbb{P}}$ are **equivalent**.

The main interest of a change of measure is to make difficult problems easier to solve. While some problems might be extremely difficult to tackle under the real-world measure $\mathbb{P}$, it might be possible to find an equivalent measure $\mathbb{Q}$ under which they are much easier to solve.

As a result, the change of measure techniques have become a cornerstone not only of modern probability but also of mathematical finance, where they are widely used in asset pricing (see Lectures on *Black-Scholes via Martingales* and on *PDE vs. Martingales*).

4 Further Itô Clalculus

In this section, we take a look at the Itô product rule, a key application of the Itô formula.

4.1 Reminder: Multidimensional Itô

We can easily extend the Itô formula to functions of several stochastic processes. Such extension will prove useful when valuing bonds using multi-factor interest rate models, option on several underlying assets (basket options) or when we want to model stochastic volatility.

Let start with 2 dimensions.

Let's say we want to price a given financial instrument deriving its value $V(t, S_1, S_2)$ from 2 stochastic processes S_1, S_2, where

$$\begin{aligned} dS_i(t) &= f_i(t, S_k, k = 1, 2)dt + g_i(t, S_k, k = 1, 2)dX_i(t), \\ & \quad i = 1, 2 \end{aligned}$$

and where

$$dX_1(t)dX_2(t) \to \rho dt$$

For convenience, we will simplify the notation as

$$dS_i(t) = f_i dt + g_i dX_i(t), \quad i = 1, 2$$

Writing a naive 2-dimensional Taylor expansion, we see that

$$\begin{aligned} dV(t) &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S_1}dS_1(t) + \frac{\partial V}{\partial S_2}dS_2(t) \\ &+ \frac{1}{2}\frac{\partial^2 V}{\partial S_1^2}dS_1^2(t) + \frac{\partial^2 V}{\partial S_1 \partial S_2}dS_1(t)dS_2(t) + \frac{1}{2}\frac{\partial^2 V}{\partial S_2^2}dS_2^2(t) + \ldots \end{aligned}$$

Since dt is already very small, we will ignore all the terms of order $(dt)^\alpha, \alpha > 1$.

Since $dX_i^2(t) \to dt$ in the mean square limit, we see that

$$dS_i^2(t) \to g_i^2 dt, \quad i = 1, 2$$

in the mean square limit.

Also, since $dX_1(t)dX_2(t) \to \rho dt$, we see that

$$dS_1(t)dS_2(t) \to \rho g_1 g_2 dt$$

in the mean square limit.

Substituting into the Taylor expansion, we get the 2-dimensional Itô formula:

$$\begin{aligned} dV(t) &= \Big(\frac{\partial V}{\partial t} + f_1 \frac{\partial V}{\partial S_1} + f_2 \frac{\partial V}{\partial S_2} + \frac{1}{2} g_1^2 \frac{\partial^2 V}{\partial S_1^2} + \rho g_1 g_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} \\ &\quad + \frac{1}{2} g_2^2 \frac{\partial^2 V}{\partial S_2^2} \Big) dt + g_1 \frac{\partial V}{\partial S_1} dX_1(t) + g_2 \frac{\partial V}{\partial S_2} dX_2(t) \end{aligned}$$

And now for the full n-dimensional formula!

Let's say we want to price a given financial instrument deriving its value $V(t, S_1, \ldots, S_n)$ from n stochastic processes $S_1, \ldots, S_n$, where

$$
\begin{aligned}
dS_i(t) \quad = \quad & f_i(t, S_k, k = 1, \ldots, n)dt + g_i(t, S_k, k = 1, \ldots, n)dX_i(t), \\
& i = 1, \ldots, n
\end{aligned}
$$

We have n different Brownian motions $(X_1, \ldots, X_n)$ at work. We are also given the pairwise correlations of the Brownian increments:

$$dX_i(t)dX_j(t) \to \rho_{ij}dt, \quad i, j = 1, \ldots, n, \ i \neq j$$

Note that $\rho_{ii} = 1$ (i.e. a Brownian motion is perfectly correlated with itself) and we get the usual:

$$(dX_i)^2(t) = dX_i(t)dX_i(t) \to \rho_{ii}dt = dt$$

For convenience, we simplify the notation as

$$dS_i(t) = f_i dt + g_i dX_i(t), \quad i = 1, \ldots, n$$

Now, applying the same logic as in the 2-dimensional case, we obtain the n-dimensional version of Itô's formula

$$dV(t) = \left(\frac{\partial V}{\partial t} + \sum_{i=1}^{n} f_i \frac{\partial V}{\partial S_i} + \frac{1}{2} \sum_{i=1}^{n} g_i^2 \frac{\partial^2 V}{\partial S_i^2} + \sum_{\substack{i=1 \\ j>i}}^{n} \rho_{ij} g_i g_j \frac{\partial^2 V}{\partial S_i S_j} \right) dt + \sum_{i=1}^{n} g_i \frac{\partial V}{\partial S_i} dX_i(t)$$

4.2 Itô Formula for Stochastic Integrals

Given a function $V(t, S(t))$ defined over $[0, T]$ and with $V(0, S(0)) = V_0$, where $S(t)$ is a stochastic process evolving according to $dS(t) = f(t, S(t))dt + g(t, S(t))dX(t)$, with f and g satisfying technical condition, we have:

$$
\begin{aligned}
V(T, S_T) &= V_0 + \int_0^T \left\{ \frac{\partial V}{\partial t} + f(t, S(t)) \frac{\partial V}{\partial S} + \frac{1}{2} g^2(t, S(t)) \frac{\partial^2 V}{\partial S^2} \right\} dt \\
&\quad + \int_0^T g(t, S(t)) \frac{\partial V}{\partial S} dX(t)
\end{aligned}
$$

How did we get to this formula?

Simple. All we did was to define an integration range on which our functional relationship V is valid, namely $[0, T]$, and then to integrate our trusted Itô for SDEs over the integration range.

Implementation Tip

When you need to use Itô for stochastic integral, for all practical purpose:

- *Start from the differential form, which is easier to manipulate, and use Itô or Taylor (substituting $dX^2(t)$ for dt);*
- *Integrate over the required range;*
- *Done!*

We can obtain the integral version of the Itô's formula for n-dimensional processes in the same way:

$$V(T) = V_0 + \int_0^T \left(\frac{\partial V}{\partial t} + \sum_{i=1}^{n} f_i \frac{\partial V}{\partial S_i} + \frac{1}{2} \sum_{i=1}^{n} g_i^2 \frac{\partial^2 V}{\partial S_i^2} + \sum_{\substack{i=1 \\ j>i}}^{n} \rho_{ij} g_i g_j \frac{\partial^2 V}{\partial S_i S_j} \right) dt + \sum_{i=1}^{n} \int_0^T g_i \frac{\partial V}{\partial S_i} dX_i(t)$$

Exercise: What Are We Correlating, Exactly?

Consider a pair of correlated arithmetic Brownian motions :

$$dX(t) = adt + b \cdot dW_1(t), \quad X(0) = 0$$
$$dY(t) = cdt + d \cdot dW_2(t), \quad Y(0) = 0$$

where $W_1(t)$ and $W_2(t)$ are two Brownian motions with $\mathbb{E}\left[dW_1(t)dW_2(t)\right] = \rho dt$.

Using a Euler scheme, simulate the evolution of the two arithmetic Brownian motions using the following sets of parameters:

Short time horizon:

Parameter	Set 1	Set 2	Set 3	Set 4
a	0.1	0.1	0.1	0.1
b	0.2	0.2	0.2	0.2
c	0.1	0.1	-0.1	-0.1
d	0.2	0.2	0.2	0.2
ρ	0.9	-0.9	0.9	-0.9
time interval	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$
discretization δt	0.01	0.01	0.01	0.01

Long time horizon:

Parameter	Set 5	Set 6	Set 7	Set 8
a	0.1	0.1	0.1	0.1
b	0.2	0.2	0.2	0.2
c	0.1	0.1	-0.1	-0.1
d	0.2	0.2	0.2	0.2
ρ	0.9	-0.9	0.9	-0.9
time interval	$[0, 10]$	$[0, 10]$	$[0, 10]$	$[0, 10]$
discretization δt	0.01	0.01	0.01	0.01

What do you observe?

4.3 Itô Product Rule

The **Itô product rule** tells you how to deal with a process $V(t)$ defined as the the product of another two processes $S_1(t)$ and $S_2(t)$, that is:

$$V(t) = S_1(t) \times S_2(t)$$

This is in fact a direct application of a more general result: **Itô's Lemma in 2 dimensions**.

The Itô product rule is an application of Iô's lemma in 2 dimensions with $V(t) = S_1(t) \times S_2(t)$.

Then,

$$\begin{aligned} \frac{\partial V}{\partial S_1} &= S_2 \\ \frac{\partial V}{\partial S_2} &= S_1 \\ \frac{\partial^2 V}{\partial S_1^2} &= 0 \\ \frac{\partial^2 V}{\partial S_2^2} &= 0 \\ \frac{\partial^2 V}{\partial S_1 \partial S_2} &= 1 \end{aligned}$$

and

$$\begin{aligned} dV(t) &= (f_1 S_2(t) + f_2 S_1(t) + \rho g_1 g_2)\, dt \\ &\quad + g_1 S_2(t) dX_1(t) + g_2 S_1(t) dX_2(t) \end{aligned}$$

In integral form,

$$\begin{aligned} V(T) &= V(0) + \int_0^T \left(f_1 S_2(t) + f_2 S_1(t) + \rho g_1 g_2\right) dt \\ &\quad + \int_0^T g_1 S_2(t) dX_1(t) + \int_0^T g_2 S_1(t) dX_2(t) \end{aligned}$$

In fact, the Itô product rule is very close to the product rule from ordinary calculus.

To see this, notice that we can express the Itô product rule as:

$$
\begin{aligned}
& d\left(S_1(t) \times S_2(t)\right) \\
= \; & \underbrace{S_1(t) \times dS_2(t) + dS_1(t) \times S_2(t)}_{\text{ordinary product rule}} + \underbrace{dS_1(t) \times dS_2(t)}_{\text{cross variation adjustment}}
\end{aligned}
$$

This expression is actually the most useful formulation of the Itô product rule.

In financial mathematics, we often need to use the Itô product rule to deal with discounting and change of measures.

Next, we look at two simple and very useful applications of the Itô product rule. We will need these two applications when we price derivatives in Modules 3 and 4!

Application 1

Apply the Itô product rule to compute $V(t) = S_1(t) \times S_2(t)$, where

- $S_1(t)$ is stochastic:

$$dS_1(t) = f_1(t, S_1)dt + g_1(t, S_1)dX(t)$$

- $S_2(t)$ is deterministic:

$$dS_2(t) = f_2(t, S_2)dt$$

What do we need this for?

We need this result to price a stock options in Module 3. Take

- $S_1(t)$ as the stock price;
- $S_2(t)$ as the (deterministic) discount factor.

Applying the Itô product rule (or deriving the result from scratch), we get:

$$dV(t) \quad = \quad (f_1 S_2(t) + f_2 S_1(t))\, dt + g_1 S_2(t) dX(t)$$

The integral form is:

$$V(T) \quad = \quad V(0) + \int_0^T (f_1 S_2(t) + f_2 S_1(t))\, dt + \int_0^T g_1 S_2(t) dX(t)$$

Application 2

Apply the Itô product rule to compute $V(S_1, S_2) = S_1(t) \times S_2(t)$ where

- $S_1(t)$ and $S_2(t)$ are both stochastic;
- but they depend on <u>the same</u> Brownian motion $X(t)$.

Here,

$$dS_i(t) = f_i(t, S_k, k = 1, 2)dt + g_i(t, S_k, k = 1, 2)dX(t),\quad i = 1, 2$$

What do we need this for?

We need this result to price a Zero Coupon Bond in Module 4. We will need to manipulate the product of two processes:

- the bond price, and;
- the Radon-Nikodym derivative of a change of measures.

Applying the Itô product rule (or deriving the result from scratch), we get:

$$\begin{aligned} dV(t) &= (f_1 S_2(t) + f_2 S_1(t) + g_1 g_2)\, dt \\ &\quad + (g_1 S_2(t) + g_2 S_1(t))\, dX(t) \end{aligned}$$

The integral version is:

$$
\begin{aligned}
V(T) &= V(0) + \int_0^T \left(f_1 S_2(t) + f_2 S_1(t) + g_1 g_2\right) dt \\
&\quad + \int_0^T \left(g_1 S_2(t) + g_2 S_1(t)\right) dX(t)
\end{aligned}
$$

In this lecture, we have seen...

- The probabilistic ecosystem:
  - sample space;
  - filtration;
  - measures;
- Conditional expectations with respect to a filtration;
- the Radon-Nikodym Theorem: a useful result to help us change our setting from a measure $\mathbb{P}$ to a measure $\mathbb{Q}$;
- The Itô product rule;

To go a bit further...

There is a plethora of stochastic calculus and stochastic analysis books. Most are good, some are better than others, but the great majority share a common feature: they are rather arcane!

Both of Steven Shreve's books ([6] and [7]) are invaluable. They are clearly written and contain all you need to know on stochastic calculus applied to finance... and then some more.

Baxter and Rennie [1] provides a good overview of the key techniques. The book is intuitive: Baxter and Rennie tend to explain the important concepts in plain old English and then show how this translates in math.

Chin, Nel and Ólafsson [2] contains a large number of worked out exercises on probability and stochastic calculus. It is a very good companion book.

Hull [3] is a *tour de force*. It has an overview of pretty much every topic, but the stochastics are covered more from a finance perspective than a quantitative finance perspective. majority share a common feature: they are rather arcane!

Øksendal [5] is the next step. It gives you all of the results you will ever need to manipulate Itô processes, and their proofs.

Neftci [4] is OK, but a bit too much of a bird's eye view and not quite rigorous enough.

Appendix A: A Short History of Modern Probability

The story of Brownian motions is intertwined with the lives and work of four men: Robert Brown, Louis Bachelier, Albert Einstein and Norbert Wiener.

In 1827, Scottish botanist Robert Brown (1773-1858) observed the seemingly random movement of pollen particles in a fluid. He correctly deduced that this motion was actually a result of the pollen colliding with molecules of water. The concept of Brownian motion was born.

In 1880, Thorvald N. Thiele provided the first mathematical description of the Brownian motion in a paper on least square.

In 1900, Louis Bachelier (1870-1946) used the Brownian motion as a mathematical object to model the dynamics of asset prices. Using as much heuristics as mathematics, Bachelier derived a formula to price options.

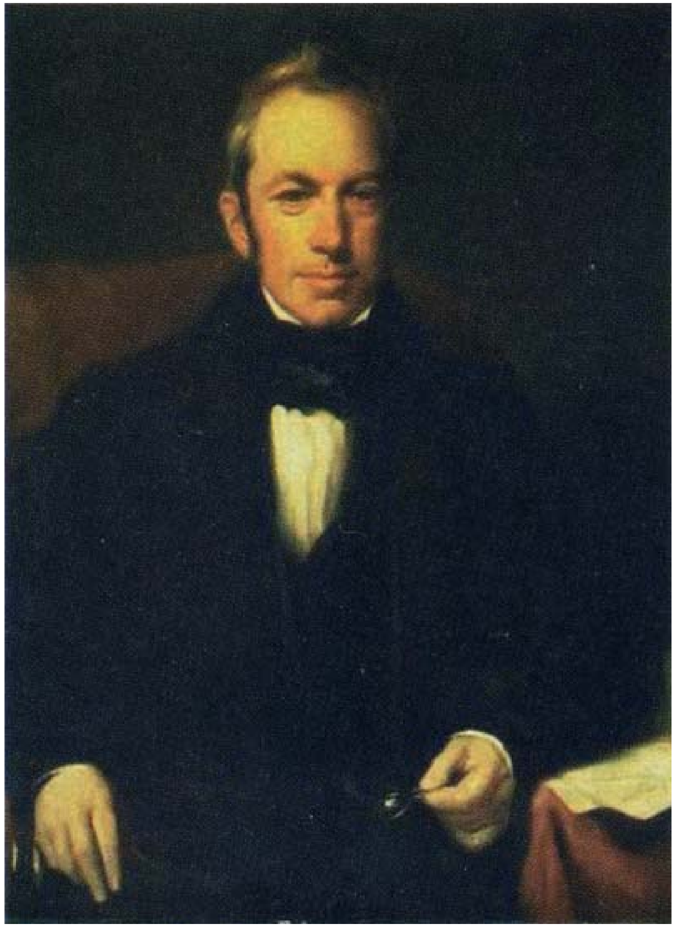


Figure : R. Brown (1773-1858)

Figure : L. Bachelier (1870-1946)

1905 was Albert Einstein's (1879-1955) *annus mirabilis*. One of the four groundbreaking papers he published on that year focused on Brownian motions and drew the attention of the physicists community to stochastic processes.

Einstein did not solve all the mystery of Brownian motions. In fact he even predicted that it might not be possible to prove that Brownian motions are continuous. But he cracked some of their difficulties and applied them to solve interesting problems in physics. As a result, Brownian motions gained a quick recognition. In fact, stochastic processes are a key component of quantum mechanics.

The late 1920s and 1930s were a Golden Age for probability theory.

Lifting the last corner of the veil did not take too long. American mathematician Norbert Wiener (1894-1964) proved that the Brownian motion has a continuous path in 1924, less than 2 decades after Einstein conjectured that this result may never be proved. Almost 90 years later, Wiener's proof remains challenging even for trained mathematician.

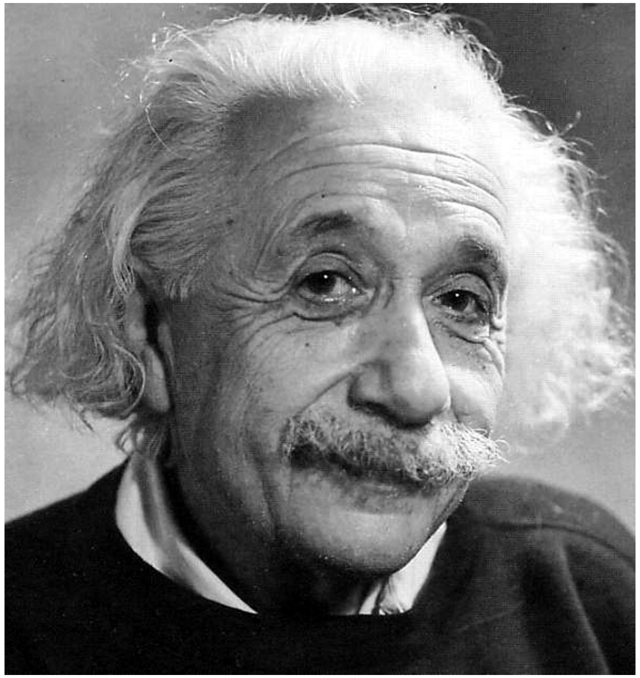


Figure : A. Einstein (1879-1955)

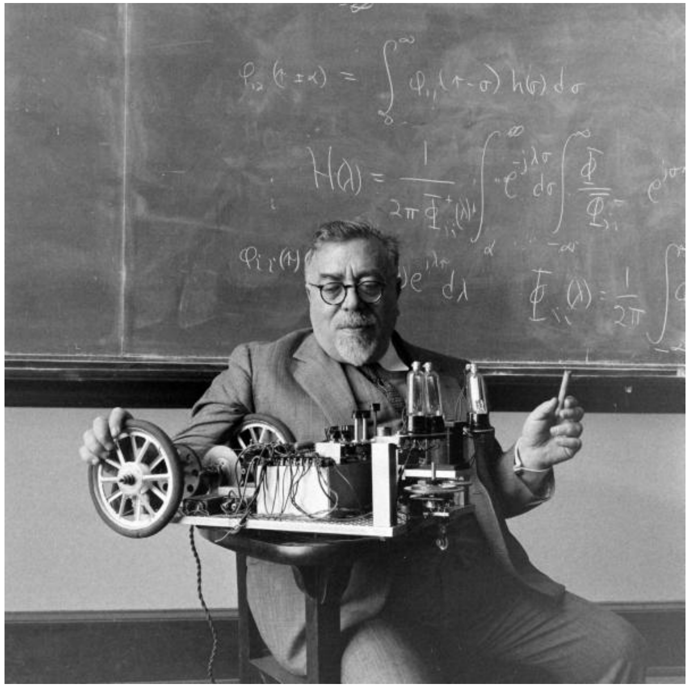


Figure : N. Wiener (1884-1964)

The two towering figures of the era where influential French mathematician Paul Levy (1886-1971) and young Russian mathematician Andrei Kolmogorov (1903-1987).

While Levy worked to extend the current theory of probabilities by introducing a number of new concepts, techniques and methods, Kolmogorov set about to rewrite probability theory using measure theory to give it the most solid mathematical foundations.

Kolmogorov published a short monograph in German detailing his results in 1933 and the Moscow school of probability reached it zenith, remaining influential to this day.

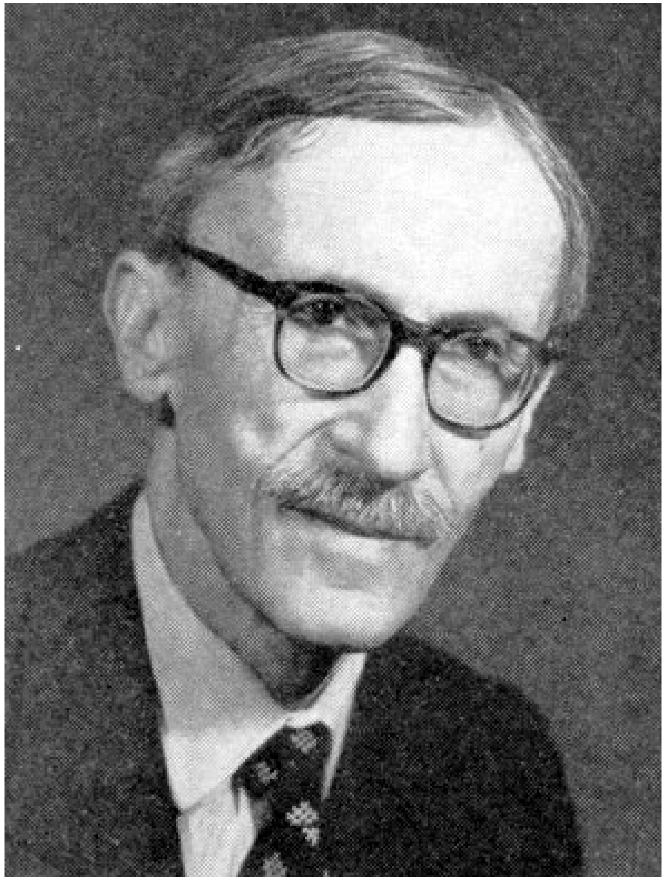


Figure : P. Lévy (1886-1971)

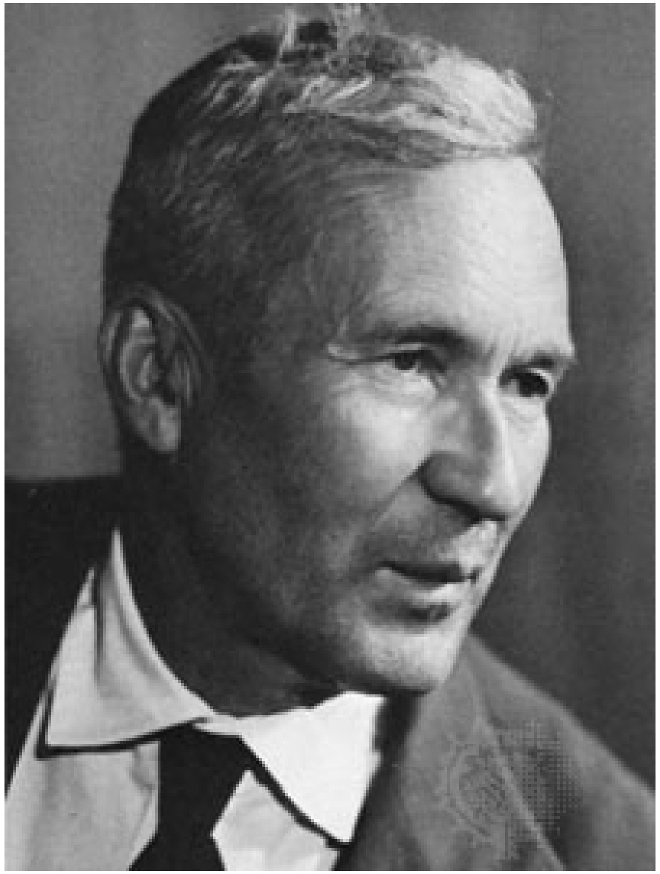


Figure : A. Kolmogorov (1903-1971)

In parallel, in the West, research into advanced topics continued. Starting in the 1950s Joseph Doob (1910-2004) in the USA and later Paul Meyer (1934-2003) in France developed the theory of martingales based on Paul Levy's earlier work.

Although not specifically designed with Kolmogorov's reformulation of Probability Theory in mind, Martingale theory has now been fully integrated into the newly consolidated body of Probability Theory and form an important and often used component.

M. Baxter and A. Rennie.
Financial Calculus: An Introduction to Derivative Pricing.
Cambridge University Press, 1996.

E. Chin, D. Nel, and S. Ólafsson.
Stochastic Calculus, volume 1 of *Problems and Solutions in Mathematical Finance*.
Wiley, 2014.

J. Hull.
Options, Futures, and Other Derivatives.
Pearson Education, 7 edition, 2010.

S. Neftci.
An Introduction to the Mathematics of Financial Derivatives.
Academic Press, 2 edition, 2000.

B. Øksendal.

Stochastic Differential Equations.
Universitext. Springer-Verlag, 6 edition, 2003.

S. Shreve.
Stochastic Calculus for Finance I: The Binomial Asset Pricing Model.
Springer Finance. Springer, 2 edition, 2008.

S. Shreve.
Stochastic Calculus for Finance II: Continuous-Time Models.
Springer Finance. Springer, 2 edition, 2008.