SEBLLEO@GMAIL.COM

CQF Module 2, Session 5: Martingales I
Advanced Stochastic Calculus and Martingales

Module 1.3

CQF

→ BM

→ Itô Calculus

SDE

determ.

random.

In this lecture...

$$dY(t) = \underbrace{f(Y(t),t)\,dt}_{\text{determ.}} + \underbrace{g(Y(t),t)\,dX(t)}_{\text{random.}}$$

$X(t)$: BM

... we will expand on the stochastic calculus lecture (Lecture 1.3) to introduce further probabilistic methods:

- the probabilistic universe;
- sample space, filtration and probability measures;
- conditional and unconditional ~~measures~~; expectations
- change of measure and the Radon Nicodým derivative;
- definition and properties of martingales.

Introduction

In the quantitative finance literature, most articles written in the past 10 years will 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.

In this lecture, we will go beyond Itô calculus to present a more advanced view of probabilities based largely on a class of stochastic processes called martingales.

Section 1: The Probabilistic Universe

measure.

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

measure space

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

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

Example: Obtaining a "Head" in a coin toss experiment is an event.

1 dice

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.

Events can then be regarded as subsets of Ω.

1.1.1. Reminder: Elementary Set Operations

A **set** S is a collection of distinct numbers or other mathematical objects called **elements**.

In probability theory, the outcomes of an experiment are the elements of the sample space associated with this experiment.

events

Example: In our dice experiment, the outcomes 1, 2, etc. are elements of the sample space Ω_d.

Mathematically, if e is an element then we write $e \in S$ to indicate that e belongs to S.

event "1" $\in \{1, 2, 3, 4, 5, 6\}$

belong to

Example: $1 \in \{1, 2, 3, 4, 5, 6\}$.

Note in the above that 1 is the element "1" representing the event "the draw produces a 1".

$A \subset S$

A **subset** A of the set S is a set of elements belonging to S.

Example: In the dice experiment, define the subset E corresponding to the event "the throw produces an even number." E is defined as $E = \{2, 4, 6\}$ and is indeed a subset of Ω_d.

Mathematically, we write $A \subset S$ to indicate that A is a subset of S.

The **complement** of a subset A in Ω, denoted by A^c is the subset of all the elements which belong to Omega but not to A.

Example: In the dice experiment,

$$E^c = \{1, 3, 5\} =: O$$

i.e. $O = E^c$ represents the event "the throw produces an odd number".

Now, on to elementary set operations. elementary set operations are the means throw which we can define new sets based on existing ones by combining or breaking them apart.

The **union** of two sets A and B corresponds to the logical "or" meaning that the union of A and B is comprised of all the elements present in *either* A and B. Mathematically, the union of two sets A and B is denoted by $A \cup B$, so...

Example: In the dice experiment, take P to be the event "the throw produces a prime number", so $P = \{2, 3, 5\}$ and E to represent the event "the throw produces an even number". Then $E \cup P = \{2, 3, 4, 5, 6\}$.

The **intersection** of two sets A and B corresponds to the logical "and" meaning that the union of A and B is comprised of all the elements present in *both* A and B. Mathematically, the union of two sets A and B is denoted by $A \cap B$, so...

Example: In the dice experiment, still take P to be the event "the throw produces a prime number" and E to represent the event "the throw produces an even number". Then $E \cap P = \{2\}$.

Finally, we define the **set difference**, denoted by $A \backslash B$ as the set of all elements belonging to A but not B.

Example: In the dice experiment, still take P to be the event "the throw produces a prime number" and E to represent the event "the throw produces an even number". Then $E \backslash P = \{4, 6\}$.

One last notation needs to be introduced: the **null set** or empty set, denoted by $\{\emptyset\}$.

Example: In the dice experiment, take O to be the event "the throw produces an odd number"and E to represent the event "the throw produces an even number". Then $O \cap E = \{\emptyset\}$.

In probability theory, the null set is used to account for the null event, i.e. nothing possible occurs.

The properties of the null set broadly correspond to those of the number 0 in algebra. For example, if we let A be a set:

1. $A \cup \{\emptyset\} = A$;
2. $A \cap \{\emptyset\} = \{\emptyset\}$;
3. $A \backslash \{\emptyset\} = A$.

Using set notation and set operations, we are now able to define a wider class of events. We have just seen that using sets we can easily define complex events such as "event ω is even" or "event ω is odd but is not a prime."

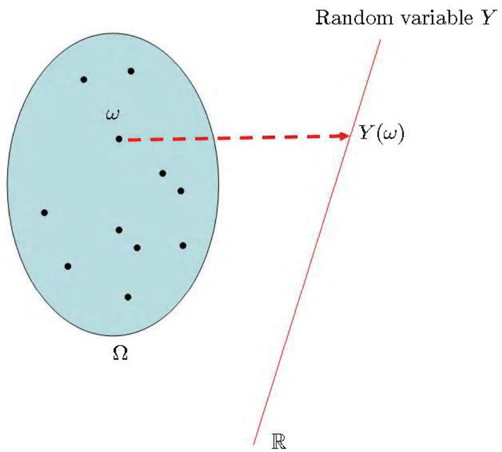
These concepts are at the very heart of Probability theory and will therefore form the base of most of our analysis.

1.1.2. Events, Random Variables and Stochastic Processes

A **random variable** Y is a *function* which assign to each individual event $\omega \in \Omega$ a numerical value.

To reflect its definition as a function Y can be more precisely written as $Y(\omega)$ where $\omega \in \Omega$ is an individual 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:

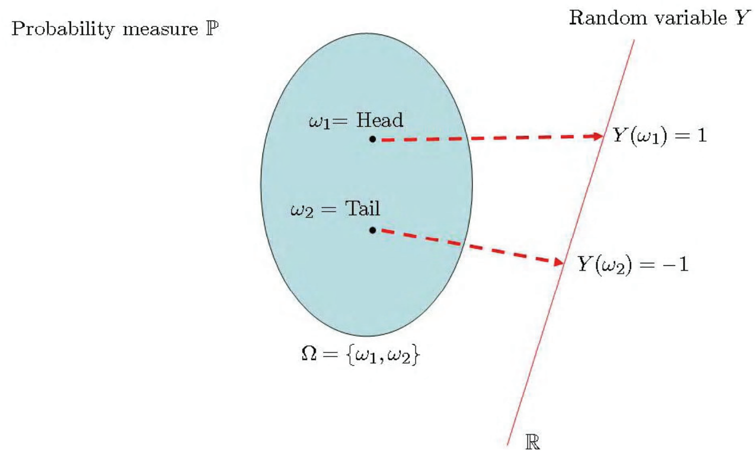
- $\omega_1 =$ Head;
- $\omega_2 =$ Tail.

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



$S(t)(\omega)$ $\longmapsto$ $S(\omega)$

for each $t \in [0,T] \to$ RV

A **stochastic process** $S(t)$ can be viewed as 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 can express this idea more precisely by denoting the process by $S(\omega)(t)$.

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, the **sample space** for a stock can be written as

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

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

A trajectory, or **sample path** for the stock would then be an elementary trajectory, such as DU.

Note also that our formulation of the sample space is not unique. Indeed, 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 would happen after the end of the third time period?

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

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

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 period 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.

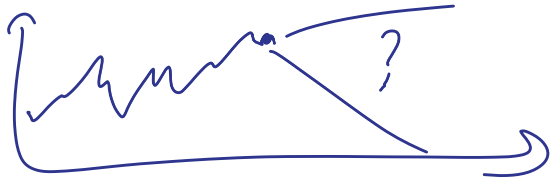
To keep track of the increasing flow of information we need to define a new mathematical object: the **filtration**.

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).



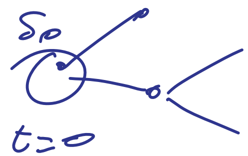
Concretely, this definition implies 2 rules:

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

Note: as a corollary to these three rules we see that:

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

$S \to S_0$



How can we see the filtration?

In most case it is very difficult to describe explicitly the filtration. But in the case of the binomial model, this can be done.

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 in turns conditions our prediction of how the stock could behave.

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 we 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 what the first two moves of the stock trajectory, the filtration starts to take shape and to become more complicated.

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

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 will 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 will 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.

$\{\mathcal{F}\}_{t \in [0, T]}$ ← notation

Since we will be working in continuous time, we will denote our filtration slightly differently.

For an experiment starting at time 0 and ending at time T, we will 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 $0 \leq t_1 \leq t_2 \leq T$,

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

final filtration

t_1 t_2 final time

This property is quite intuitive. Since we consider that information gets constantly recorded and accumulates 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.

$\mathcal{F}_t \rightarrow$ have information up to time t.

$\rightarrow$ you will know the exact value of S_t.

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 $\mathcal{F}_t$.

$(\Omega, \mathcal{F}, \mathbb{P})$ ← probability measure

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).

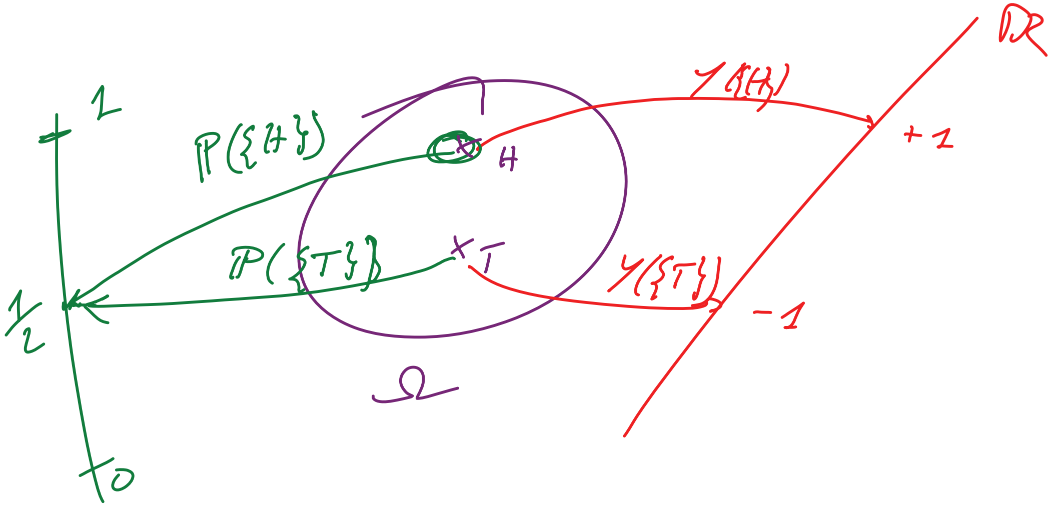
Unsurprisingly, the mathematics of probability measures stem from the field of Measure Theory.

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

Coin Toss Experiment :

$\to H = +£1$

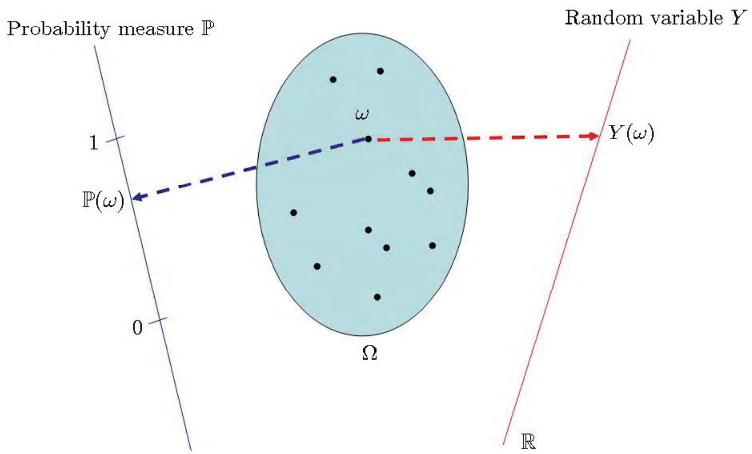
$\to T = -£1$



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.

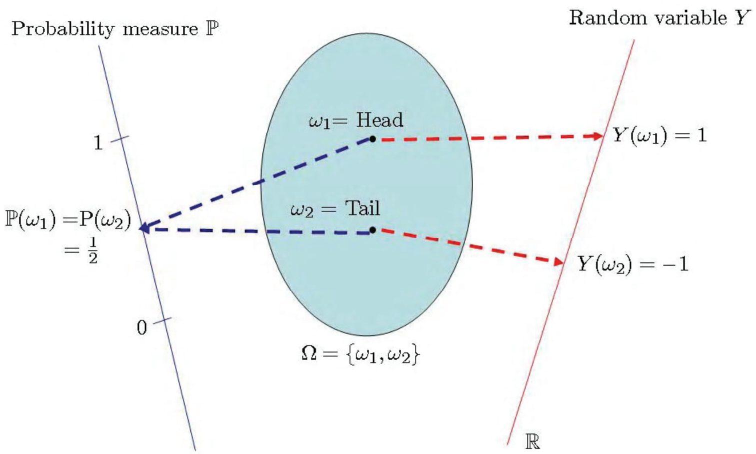
The sample space Ω has two events:

- $\omega_1 =$ 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 $\mathbb{P}$ satisfy 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, $i = 1, \ldots n \leq +\infty$,

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

$$\mathbb{P}(\underbrace{\{H\} \cup \{T\}}_{\Omega}) = \mathbb{P}(H) + \mathbb{P}(T) = 1$$

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:

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

where p is the PDF and P is the CDF.

Example of Mathematical Expectation

If we wanted to compute the expectation of a normal random variable $X \sim N(\mu, \sigma^2)$ to the right of some level K, we would write

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

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

So,

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

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

This example forms the main idea leading to the derivation of option prices from a probabilistic perspective (See Lecture 3.3).

To see this, set $x := \log S$ and $K := \log E$. The intuition behind the log is that since we will assume that the share price S are lognormally distributed, it follows that the log return on the share, defined as $\log S$, will be 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 $\mathbb{P}$:

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

In addition, there is an interesting *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:

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

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 (See Lecture 3.3).

1.3.4. Conditional Expectation With Respect to a Filtration

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

We will now introduce *conditional* expectations, which are expectation conditional on some information.

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.

10
+ 2
+ 4
+ 5

21 / 13

$\mathbb{E}[\text{Points}] = \frac{21}{13}$

Example of unconditional expectation:
Pick a card. What is the expected point value of the card?

Example of 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?

50 cards

$\mathbb{E}[\text{Points} \mid \underbrace{\text{King \& Jake out}}_{\text{information}}]$

Hence, the difference between conditional and conditional 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* $\mathcal{F}$.

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 can define:

filtration

$$\mathbf{E}[X|\mathcal{F}]$$

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

define & prove.

This concept will come in handy when we define martingales later during this session.

1.3.5. Properties Conditional Expectations

Conditional expectations have the following useful properties:

1. *Linearity*:

R.V.

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

constant

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

4PM

8AM

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

4pm 8AM 8.AM

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 Property 2, we have

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

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,

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

if you know
$\mathcal{F}$ you know
X

so
$\mathbb{E}[X$ given
you know
$\mathcal{F}]$
$= X$

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

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

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

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

7. *Positivity*: if $X \geq 0$ then $\mathbf{E}[X|\mathcal{F}] \geq 0$.
8. *Jensen's Inequality*: let f be a convex function, then

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

Later in this session, we will see an example of how some of these rules can be used to prove that a stochastic process is a martingale.

1.3.6. 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 will just outline the rules allowing us to define equivalent measures.

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 $\mathbb{Q} << \mathbb{P}$.

$\mathbb{Q}(A) \Rightarrow \mathbb{P}(A)$

$\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

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 are said to be **equivalent**.

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

The Radon Nikodým Theorem

Key Fact (The Radon Nikodým 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 $A \subset \Omega$ weight function

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

under $\mathbb{Q}$. ← proba under $\mathbb{P}$

$\mathbb{Q}(\omega) = \text{weight} \times \mathbb{P}(\omega)$

where

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

is called the **Radon Nikodým derivative**

The formulation presented in the previous slide is the most general because it applies to continuous distributions.

The Radon Nikodým theorem simplifies considerably when we deal with *discrete distributions*, such as coin toss, throw of a dice or the binomial mode. In particular we have for all subsets $A \subset \Omega$

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

where the Radon Nikodým derivative is given by

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

Example of Change of Measure

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~~ set Ω has two events:

- ω_1 = Head;
- ω_2 = Tail.

Use RN Theorem to assign weights

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

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

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

with $0 < p < 1$.

$\bar{\mathbb{P}}(\omega_1) = \frac{1}{2}$

weight

$W \times \mathbb{P}(\omega_1) = \bar{\mathbb{P}}(\omega_1)$

$W \times p = \frac{1}{2}$

$W = \frac{1}{2p}$

Suppose we want to evaluate our P&L, based on a world where all coins are fair.

We can do that by introducing 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 to travel from the real world $(\Omega, \mathcal{F}, \mathbb{P})$ to the "fair" world $(\Omega, \mathcal{F}, \bar{\mathbb{P}})$, we must have *added extra weight* to each of the two events. Namely,

H

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

T

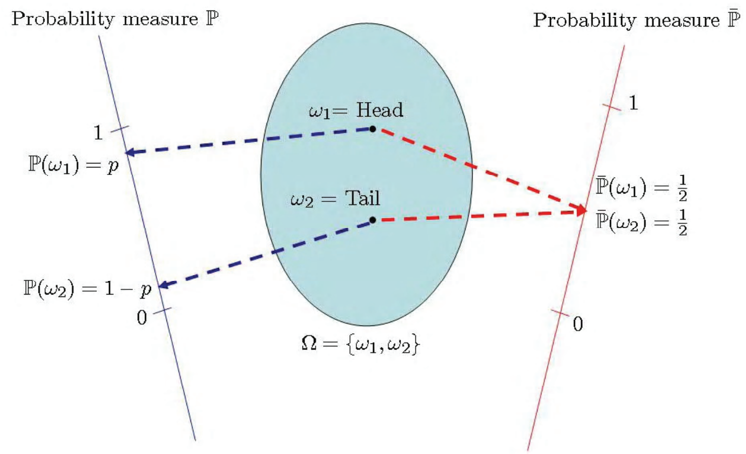
so that

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

RN derivative

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

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

and hence the measure $\mathbb{P}$ and $\bar{\mathbb{P}}$ are indeed equivalent.

$(\Omega, \mathcal{F}, \mathbb{P})$

The main interest of change of measure is to make difficult problems easier to solve. Indeed, 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 Lecture 3.3 and Lecture 3.6).

2. Martingales

The late 1920s and 1930s were a Golden Age for probability theory. The most difficult property of Brownian motion, the continuity of their path , was finally proved by Norbert Wiener in 1924, less than 2 decades after Einstein conjectured that this result may never be proved.

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 up until now.

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.

Our treatment of martingales will be introductory and non-axiomatic. A full treatment of martingales requires a good knowledge of analysis, measure theory and axiomatic probability theory among other things, as well as long study hours.

We will proceed as follows. First, we will present the definitions and results in a discrete time, since discrete time is generally easier and requires fewer technical conditions. We will then "intuit" that these results and definitions still hold as $\delta t \to dt$, subject to appropriate technical conditions and limitations.

$(\Omega, \mathcal{F}, \mathbb{P})$

2.1. Definitions

Definition (Discrete Time Martingale)

A discrete time stochastic process

$$\{M_t : t = 0, \ldots, T\}$$

such that M_t is $\mathcal{F}_t$-measurable for $t = 0, \ldots, T$ is a **martingale** if

technical

$$\mathbf{E}\,|M_t| < \infty$$

integrability condition

and

$$\mathbf{E}\,[M_{t+1}|\mathcal{F}_t] = M_t \tag{1}$$

The first equation represents a standard integrability condition.

The second equation tells you that the expected value of M at time $t+1$ conditional on all the information available up to time t is the value of M at time t. In short, a Martingale is a **driftless process**.

If we take expectation on both sides of (1), then

$$\mathbf{E}[M_{t+1}] = \mathbf{E}[M_t]$$

This is due to the **Tower Property** of conditional expectations we have encountered earlier.

Example: Proving That a Process Is A Martingale

Let Y be a $\mathcal{F}$-measurable random variable with $\mathbf{E}[|Y|] < \infty$ *and define the sequence of random variables*

$$X_n = \mathbf{E}[Y|\mathcal{F}_n], \qquad n = 1, \ldots$$

Prove that X_n is a martingale.

Proving that a process is a martingale involves proving two elements:

1. that the expectation exists, i.e. $\mathbf{E}[|X_n|] < \infty$;
2. that the process is driftless, i.e. $\mathbf{E}[X_n|\mathcal{F}_m] = X_m$ for $m < n$

Proving integrability:
By the definition of X_n,

$$\mathbf{E}[|X_n|] \quad = \quad \mathbf{E}[|\mathbf{E}[Y|\mathcal{F}_n]|]$$

By **Jensen's Inequality**,

$$\mathbf{E}[|\mathbf{E}[Y|\mathcal{F}_n]|] \quad \leq \quad \mathbf{E}[\mathbf{E}[|Y||\mathcal{F}_n]]$$

By the **Tower Property**,

$$\mathbf{E}[\mathbf{E}[|Y||\mathcal{F}_n]] \quad = \quad \mathbf{E}[|Y|]$$

Since Y_n is integrable,

$$\mathbf{E}[|Y|] < \infty$$

To conclude,

$$\mathbf{E}[|X_n|] \quad \leq \quad \mathbf{E}[\mathbf{E}[|Y||\mathcal{F}_n]] = \mathbf{E}[|Y|] < \infty$$

and hence X_n is integrable for $n = 1, \ldots$

Proving driftlessness:
For $m < n$, by the definition of X_n,

$$\mathbf{E}\left[X_n | \mathcal{F}_m\right] = \mathbf{E}\left[\mathbf{E}\left[Y | \mathcal{F}_n\right] | \mathcal{F}_m\right]$$

By the **Tower Property**,

$$\mathbf{E}\left[\mathbf{E}\left[Y | \mathcal{F}_n\right] | \mathcal{F}_m\right] = \mathbf{E}\left[Y | \mathcal{F}_m\right]$$

By definition of X_m,

$$\mathbf{E}\left[Y | \mathcal{F}_m\right] = X_m$$

Therefore $\mathbf{E}\left[X_n | \mathcal{F}_m\right] = X_m$ for $m < n$ and X_n is a martingale for $n = 1, \ldots$

Definition (Discrete Time Supermartingale (Submartingale))

A discrete time stochastic process

$$\{M_t : t = 0, \ldots, T\}$$

such that M_t is $\mathcal{F}_t$-measurable for $t = 0, \ldots, T$ is a **supermartingale (submartingale)** if

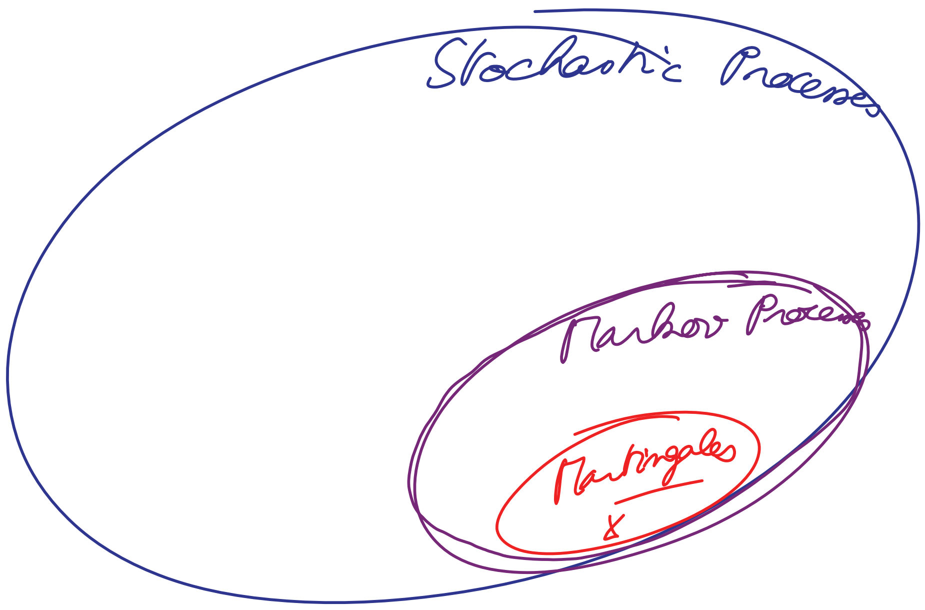
$$\mathbf{E}\,|M_t| < \infty$$

and

$$\mathbf{E}\left[M_{t+1}|\mathcal{F}_t\right] \leq (\geq) M_t$$

As you can see by now, martingales are a very nice mathematical object. Indeed, they achieve our stated objective of “getting rid of the drift” and enable us to focus on what probabilists consider is the most interesting part: the statistical properties of purely random processes.

In addition, Doob and Meyer have developed a powerful theory centered around the definition of sub/super/martingales.


Stochastic Processes
Markov Process
Martingales

2.2. Markov vs. Martingale

Markov processes and martingales are not synonymous.

$\mathbb{E}[M_{t+1} | \mathcal{F}_t] = M(t)$

The martingale property states that the expected value at time $t + 1$ is the value realized at time t. The process is **driftless**.

The Markov property states that that the expected value at time $t + 1$ does not depend on the history (trajectory) up to time t, but it may depend on the value at time t. The process is **memoryless** (no path dependence).

$\mathbb{E}[M_{t+1} | \mathcal{F}_t] = f(M_t)$

The martingale property is therefore a much stronger result than the Markov property.

We can now generalize to continuous time:

Definition (Continuous Time Martingale)

A continuous time stochastic process

$$\{M_t : t \in \mathbb{R}^+\}$$

such that M_t is $\mathcal{F}_t$-measurable for $t \in \mathbb{R}^+$ is a **martingale** if

$$\mathbf{E}\,|M_t| < \infty$$

and

$$\mathbf{E}\,[M_t|\mathcal{F}_s] = M_s, \quad 0 \le s \le t$$

③. $(X_t - X_0) \sim \mathcal{N}(0, (t-0))$

driftless

$\Rightarrow$ Martingale property

$$\mathbb{E}[X_t \mid \mathcal{F}_0] = \underbrace{\mathbb{E}[X_t - X_0 \mid \mathcal{F}_0]}_{=0} + \underbrace{\mathbb{E}[X_0 \mid \mathcal{F}_0]}_{= X_0}$$

$$= X_0$$

2.3. The Brownian Motion as a Martingale

In the Stochastic Calculus lecture we saw that Brownian motions are martingales. In fact, not only are Brownian motion martingales, but they can be characterized in terms of martingales.

Key Fact (Levy's Martingale Characterization)

Let $X(t), t > 0$ be a stochastic process and let $\mathcal{F}_t$ be the filtration generated by it. $X(t)$ is a Brownian motion iff all of the following conditions are satisfied:

1. $X(0) = 0$ *a.s.;*
2. *the sample paths* $t \mapsto X(t)$ *are continuous a.s.;*
3. X_t *is a martingale with respect to the filtration* $\mathcal{F}_t$*;*
4. $|X_t|^2 - t$ *is a martingale with respect to the filtration* $\mathcal{F}_t$.

⇒ Quadratic variation of BM

The Lévy characterization can be contrasted with the classical definition of a Brownian motion as a stochastic process $X(t)$ satisfying:

BM

1. $X(0) = 0$ a.s.;
2. ~~the sample paths $t \mapsto X(t)$ are continuous a.s.;~~
3. **independent increments**: for $t_1 < t_2 < t_3 < t_4$ the increments $X_{t_4} - X_{T_3}$ $X_{t_2} - X_{T_1}$ are independent;
4. **normally distributed increments**: $X_t - X_s \sim N(0, |t-s|)$.

In particular, Levy's characterization neither mentions independent increments nor normally distributed increments. Instead, two easily verifiable martingale conditions are introduced.

Definition (Continuous Time Supermartingale (Submartingale))

A discrete time stochastic process

$$\{M_t : t \in \mathbb{R}^+\}$$

such that M_t is $\mathcal{F}_t$-measurable for $t \in \mathbb{R}^+$ is a **supermartingale (submartingale)** if

$$\mathbf{E}\,|M_t| < \infty$$

and

$$\mathbf{E}\left[M_t|\mathcal{F}_s\right] \leq (\geq) M_s \quad \forall 0 \leq s \leq t$$

2.4. Itô Integrals and Martingales

We will now explore the link between Itô integration and martingale.

Consider the stochastic process $Y(t) = X^2(t)$. By Itô, we have

$$X^2(T) = T + \int_0^T 2X(t)dX(t)$$

Taking the expectation, we get

$$E[X^2(T)] = \cancel{T} + \underbrace{E[\int_0^T 2X(t)dX(t)]}_{\text{Itô integral}} = 0$$

$= \cancel{T}$

Now, the quadratic variation property of Brownian motions implies that

$$E[X^2(T)] = T$$

and hence

$$E[\int_0^T 2X(t)dX(t)] = 0$$

Therefore, the Itô integral

$$\int_0^T 2X(t)dX(t)$$

is a martingale.

In fact, this property is shared by all Itô integrals.

Applications in 3.3 & 4.3

Key Fact

Itô Integrals are Martingales.

Let $g(t, X_t)$ be a function on [0,T] and satisfying technical condition. Then

$$\int_0^T g(t, X_t) dX_t$$

is a martingale.

So, Itô integrals are martingales.

But does the converse hold? Can we represent any martingale as an Itô integral?

The answer is yes.

Key Fact (Martingale Representation Theorem)

If M is a martingale, then there exists a function $g(t, X_t)$ satisfying technical condition such that

$$M_T = M_0 + \int_0^T g(t, X_t) dX_t$$

don't usually know what g is.

at time 0

Itô integral

→ 4.3

Example:

We will show that

$$\mathbf{E}\left[X^2(T)\right] = T$$

using only Itô and the fact that Itô integrals are martinagles.

Consider the function $F(t, X_t) = X_T^2$, then by Itô's lemma,

$$
\begin{aligned}
\mathbb{E}\left[X_T^2 \right] &= \underbrace{X_0^2}_{=0} + \frac{1}{2}\int_0^T 2dt + \int_0^T 2X_t dX_t \\
&= \underbrace{\mathbb{E}\left[\int_0^T dt\right]}_{\text{deterministic}} + \underbrace{\mathbb{E}\left[2\int_0^T X_t dX_t\right]}_{\leftarrow\ \text{Itô integral}}
\end{aligned}
$$

since $X_0 = 0$

Taking the expectation,

$$
\mathbf{E}\left[X_T^2\right] = \underbrace{\mathbf{E}\left[\int_0^T dt\right]} + 2\underbrace{\mathbf{E}\left[\int_0^T X_t dX_t\right]}_{=0}
$$

$\int_\Omega \int_0^T \cdot \, dt \, d\mathbb{P}$

Now, $\int_0^T X_t dX_t$ is an Itô integral and as a result $\mathbf{E}\left[\int_0^T X_t dX_t\right] = 0$

Moreover,

deterministic term

$$\int_0^T \mathbb{E}[1]\,dt = \int_0^T dt = T = \mathbf{E}\left[\int_0^T dt\right] = \mathbf{E}\left[T\right] = T$$

We can conclude that

$$\mathbf{E}\left[X^2(T)\right] = T$$

As an aside, we can usually exchange the order of integration between the time integral and the expectation so that

$$\mathbf{E}\left[\int_0^T f(X_t)dt\right] = \int_0^T \mathbf{E}\left[f(X_t)\right] dt$$

This is due to an analysis result known as Fubini's Theorem.

(end of example)

Aside: Properties of Itô Integrals:

Key Fact

1. **Linearity**

$$\int_0^T (\alpha f(t) + \beta g(t))\, dX_t = \int_0^T \alpha f(t) dX_t + \int_0^T \beta g(t) dX_t$$

2. **Itô isometry**:

proofs

$$\mathbb{E}\left[\left|\int_0^T f(t) dX_t\right|^2\right] = \mathbb{E}\left[\int_0^T |f(t)|^2\, dt\right]$$

$0 < s < T$

3. **Martingale**:

Most general version

$$\mathbb{E}\left[\underbrace{\int_0^T f(t) dX_t}_{\text{Itô}}\,\middle|\,\mathcal{F}_s\right] = \int_0^s f(t) \cancel{dX_t}\ dt$$

The **linearity property** is carried over from the general definition of integrals as being the limit of a sum.

The **isometry property** is used to extend the definition of the Itô integral to a very general class of functions. As a result, it is often mentioned as one of the key properties.

(end of aside)

<u>Example:</u> A First Glance at an Exponential Martingale

Consider a SDE for a process $S(t)$:

$$dS_t = \mu dt + \sigma dX_t, \quad S(0) = S$$

Find a deterministic function $f(t)$ such that the process

$$M(t) = \exp\left(S_t + f(t)\right)$$

is a martingale.

Define the function

$$g(u, x) := \exp\left(x + f(u)\right)$$

so that

$$M(t) = g(t, S_t)$$

We will now use Itô's formula. Since

$$
\begin{aligned}
\frac{\partial g}{\partial u}(v,y) &= f'(v)g(v,y) \\
\frac{\partial g}{\partial x}(v,y) &= g(v,y) \\
\frac{\partial^2 g}{\partial x^2}(v,y) &= g(v,y)
\end{aligned}
$$

then, by Itô,

$$
\begin{aligned}
dM(t) &= \frac{\partial g}{\partial u}(t,S_t)dt + \frac{\partial g}{\partial x}(t,S_t)dS + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2}(t,S_t)dt \\
&= \left(f'(t) + \mu + \frac{1}{2}\sigma^2\right) g(t,S_t)dt + \sigma g(t,S_t)dX \\
&= \left(f'(t) + \mu + \frac{1}{2}\sigma^2\right) M(t)dt + \sigma M(t)dX
\end{aligned}
$$

For $M(t)$ to be a martingale, we need it to be driftless. Hence, its drift must be equal to 0 and thus

$$f'(t) + \mu + \frac{1}{2}\sigma^2 = 0$$

Then

$$f(t) = -\left(\mu + \frac{1}{2}\sigma^2\right) t$$

and

$$\exp\left(S_t - \left(\mu + \frac{1}{2}\sigma^2\right) t\right)$$

is a martingale.

In this lecture, we have seen...

- the probability space $(\Omega, \mathcal{F}, \mathbb{P})$;
- various properties of expectations;
- the Radon Nicodým Theorem: a useful result to help us change our setting from a measure $\mathbb{P}$ to a measure $\mathbb{Q}$;
- the definition of martingales as driftless processes;
- the fact that Brownian motion can be defined in terms of martingales;
- the fact that Itô integrals are martingales and that martingales can be represented as Itô integrals.