FE5222 Advanced Derivative Pricing

Wu Lei

Risk Management Institute, National University of Singapore

rmiwul@nus.edu.sg

August 15, 2019

Overview

Probability Space

Definition

A probability space is a triplet of $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is the sample space,
- $\mathcal{F}$ is a σ-algebra, and
- $\mathbb{P}$ is a probability measure.

Sample Space

Definition

A sample space Ω is the set of all possible outcomes. Each $\omega \in \Omega$ represents an outcome.

Example

- Toss a coin once, $\Omega = \{H, T\}$

Sample Space

Definition

A sample space Ω is the set of all possible outcomes. Each $\omega \in \Omega$ represents an outcome.

Example

- Toss a coin once, $\Omega = \{H, T\}$
- Toss a coin twice, $\Omega = \{HH, HT, TH, TT\}$

Sample Space

Definition

A sample space Ω is the set of all possible outcomes. Each $\omega \in \Omega$ represents an outcome.

Example

- Toss a coin once, $\Omega = \{H, T\}$
- Toss a coin twice, $\Omega = \{HH, HT, TH, TT\}$
- Toss a coin infinite times? $\Omega = [0, 1)$

σ-algebra

Definition

A σ-algebra (σ-field) $\mathcal{F}$ is a family of subsets in Ω such that

- $\Omega \in \mathcal{F}$,
- If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$, and
- If $E_n \in \mathcal{F}$, $n = 1, 2, \ldots$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{F}$.

σ-algebra

Definition

A σ-algebra (σ-field) $\mathcal{F}$ is a family of subsets in Ω such that

- $\Omega \in \mathcal{F}$,
- If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$, and
- If $E_n \in \mathcal{F}, n = 1, 2, \ldots$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{F}$.

Definition

A set $E \in \mathcal{F}$ is called an event.

Borel σ-algebra on $\mathbf{R^n}$

Definition

The Borel σ-algebra $\mathcal{B}(\mathbf{R^n})$ the smallest σ-algebra containing all open sets in $\mathbf{R^n}$.

Borel σ-algebra on $\mathbf{R^n}$

The σ-algebra $\mathcal{B}(\mathbf{R^n})$ includes all subsets in $\mathbf{R^n}$ of our interest. For example, the Borel σ-algebra $\mathcal{B}(\mathbf{R})$ includes

- Open set
- Close set
- Singleton: $\{a\}$
- Half-open (closed) sets: $(a, b], [a, b)$

σ-algebra

Since the power set $2^{\Omega} = \text{all subsets of } \Omega$ is a σ-algebra, why do we need other σ-algebras?

σ-algebra

Since the power set $2^\Omega = \text{all subsets of } \Omega$ is a σ-algebra, why do we need other σ-algebras?

1. Overcome technical difficulties
   The σ-algebra 2^Ω is too big to define a meaningful probability measure in some cases.

σ-algebra

Since the power set $2^\Omega = \text{all subsets of } \Omega$ is a σ-algebra, why do we need other σ-algebras?

1. Overcome technical difficulties
   The σ-algebra 2^Ω is too big to define a meaningful probability measure in some cases.
2. Model information
   More importantly we can use σ-algebra to model information.

σ-algebra

The outcome $\omega \in \Omega$ is what has really happened. This may not be observable by a person. Each $E \in \mathcal{F}$ is a set of ω's in Ω. If our information is $\mathcal{F}$, then we know whether a particular E has happened or not. We may know one of the $\omega' s$ in E has happened, but don't know exactly which ω.

σ-algebra

Example

In the example of tossing a coin twice, the sample space is $\Omega = \{HH, HT, TH, TT\}$, we can have the following two different σ-algebras.

- $\mathcal{F}_1 = \{\emptyset, \{HT, HH\}, \{TH, TT\}, \Omega\}$
- $\mathcal{F}_2 = \{\emptyset, \{HT, HH, TH\}, \{TT\}, \Omega\}$

σ-algebra

If $\omega = HT$, a person possessing the information $\mathcal{F}_1$ knows that the event $\{HT, HH\}$ (i.e., the first toss is head) has happened. But he or she does not know whether it is HT or HH. This can happen if this person only observes the first coin toss.

σ-algebra

Similarly, a person possessing the information $\mathcal{F}_2$ knows that the event $\{HT, HH, TH\}$ (i.e., at least the one toss is head) has happened. But he or she does not know whether it is HT, HH or TH.

Probability Measure

Definition

A probability measure $\mathbb{P}$ is a mapping from $\mathcal{F}$ to $[0,1]$ such that

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$, where $E_i \cap E_j = \emptyset \ \forall i \neq j$.

Probability Measure

Definition

A probability measure $\mathbb{P}$ is a mapping from $\mathcal{F}$ to $[0,1]$ such that

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$, where $E_i \cap E_j = \emptyset \ \forall i \neq j$.

Note that a probability measure is only defined on the σ-algebra. We cannot assign a probability to the set $E \notin \mathcal{F}$.

Almost Surely

Definition

Let $E \in \mathcal{F}$ and $\mathbb{P}(E) = 1$. Then we say the event E happens almost surely (a.s.).

Note that E does not always happen. The event it does not happen is insignificant in probabilistic sense.

Random Variable

Definition

A random variable (r.v.) X is a real-valued[1] function on Ω such that

$$X^{-1}(B) \in \mathcal{F}$$

for all $B \in \mathcal{B}(\mathbf{R})$, where

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

[1] In some cases, we shall allow X to take ∞ or $-\infty$.

Random Variable

Note that

- The requirement for $X^{-1}(B)$ to be $\mathcal{F}$-measurable is to ensure we can assign probability to the events we are interested in. For example, if X is the stock price, we would like to ask what is the probability that X is between 100 and 120 tomorrow. To assign the probability to such an event, we must have

$$\{\omega : 100 \leq X(\omega) \leq 120\} \in \mathcal{F}$$

.

Random Variable

Note that

- The requirement for $X^{-1}(B)$ to be $\mathcal{F}$-measurable is to ensure we can assign probability to the events we are interested in. For example, if X is the stock price, we would like to ask what is the probability that X is between 100 and 120 tomorrow. To assign the probability to such an event, we must have

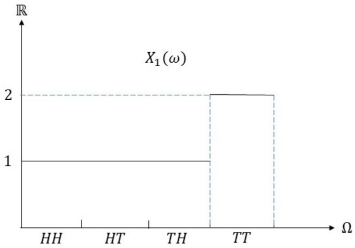
$$\{\omega : 100 \leq X(\omega) \leq 120\} \in \mathcal{F}$$

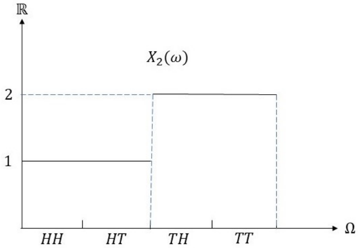
.

- If X is a random variable, f is a Borel function, then $Y = f(X)$ is also a random variable.

Information from a Random Variable

Suppose we don't observe $\omega \in \Omega$, but we know the values of the random variable X_1 or X_2.





Information from a Random Variable

A person who knows X_1 can tell whether $\{HH, HT, TH\}$ or $\{TT\}$ happens by observing the values of X_1. Similarly a person who knows X_2 can tell between the event $\{HH, HT\}$ and $\{TH, TT\}$. Hence X_1 and X_2 convey different information!

Information from a Random Variable

Definition

Let X be a random variable, the σ-algebra generated by X is defined as

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$$

which is the smallest σ-algebra such that X is $\sigma(X)$-measurable.

Information from a Random Variable

Note that

- The σ-algebra $\sigma(X)$ is the information contained in the random variable X.

Information from a Random Variable

Note that

- The σ-algebra $\sigma(X)$ is the information contained in the random variable X.
- In the above example, a person knows the values of both X_1 and X_2 will have more information. For example, if he/she observes $X_1(\omega) = 1$ and $X_2(\omega) = 2$, he/she can immediately deduce that $\omega = TH$.

Information from Random Variables

Definition

Let $X_\lambda, \lambda \in \Lambda$ be random variables, the σ-algebra generated by X_λ, denoted by $\sigma(X_\lambda, \lambda \in \Lambda)$, is the smallest σ-algebra that contains $\sigma(X_\lambda)$ for all $\lambda \in \Lambda$.

Information from Random Variables

Definition

Let $X_\lambda, \lambda \in \Lambda$ be random variables, the σ-algebra generated by X_λ, denoted by $\sigma(X_\lambda, \lambda \in \Lambda)$, is the smallest σ-algebra that contains $\sigma(X_\lambda)$ for all $\lambda \in \Lambda$.

$\sigma(X_\lambda, \lambda \in \Lambda)$ can be interpreted as the information contained in all X_λ.

Information from Random Variables

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. X is a random variable. If $\sigma(X) \subset \mathcal{G}$, then we say X is $\mathcal{G}$-measurable.

Information from Random Variables

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. X is a random variable. If $\sigma(X) \subset \mathcal{G}$, then we say X is $\mathcal{G}$-measurable.

$\sigma(X)$ is the information contained in X. A person with the knowledge of $\mathcal{G}$ knows whether an event E happens or not for any $E \in \mathcal{G}$. Since $\sigma(X) \subset \mathcal{G}$. The person knows whether $X^{-1}(B)$ happens or not for all $B \in \mathcal{B}(R)$.

Information from Random Variables

Theorem

Let X and Y be two random variables. If X is $\sigma(Y)$-measurable (i.e. $\sigma(X) \subset \sigma(Y)$), then there is a Borel function f such that $X(\omega) = f(Y(\omega))$.

Convergence

Definition

Let $X_n, n = 1, 2, \ldots$ be a sequence of random variables, X be a random variable, X_n is said to converge to X almost surely (a.s.) if

$$\lim_{n \to \infty} X_n(\omega) = X(\omega)$$

for all $\omega \in \Omega$ except on a set $N \subset \Omega$ with $\mathbb{P}(N) = 0$.

Convergence in Probability

Definition

Let $X_n, n = 1, 2, \ldots$ be a sequence of random variables, X be a random variable, X_n is said to converge to X in probability if for any $\delta > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|X_n(\omega) - X(\omega)| > \delta) = 0$$

Convergence of Expectation

Theorem (Monotone Convergence Theorem)

Let $0 \leq X_1 \leq X_n \leq \ldots$ *and* $\lim_{n\to\infty} X_n = X$, *then*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

Note that

1. $X(\omega)$ may take value $+\infty$ at some ω
2. $\mathbb{E}(X)$ may be $+\infty$

Convergence of Expectation

Theorem (Dominated Convergence Theorem)

Let $|X_n| \le Y$ *for all* $n = 1, 2, \ldots$, $\lim X_n = X$ *and* $\mathbb{E}(Y) < \infty$, *then*

$$\lim \mathbb{E}(X_n) = \mathbb{E}(X).$$

Convergence of Expectation

Pay attention to the conditions under which the two convergence theorems hold!

- For the Monotone Convergence Theorem, X_n are assumed to be non-negative.
- For the Dominated Convergence Theorem, X_n are bounded by an integrable random variable Y.

Change of Measure

A probability measure essentially defines a weight for each $\omega \in \Omega$. We can scale the weights provided that they still sum up to 1. This process is called *change of measure*.

Change of Measure

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $Z(\omega)$ be a non-negative random variable such that $\mathbb{E}[Z] = 1$. Define

$$\widetilde{\mathbb{P}}(E) = \int_E Z(\omega) d\mathbb{P}(\omega)$$

for all $E \in \mathcal{F}$. Then $\widetilde{\mathbb{P}}$ is a probability measure.

Change of Measure

Theorem

Let $\widetilde{\mathbb{E}}[\cdot]$ denote the expectation under $\widetilde{\mathbb{P}}$, we have

$$\widetilde{\mathbb{E}}[Y] = \mathbb{E}[YZ]$$

provided that both expectations exist.

Change of Measure

Theorem

Let $\widetilde{\mathbb{E}}[\cdot]$ denote the expectation under $\widetilde{\mathbb{P}}$, we have

$$\widetilde{\mathbb{E}}[Y] = \mathbb{E}[YZ]$$

provided that both expectations exist.

An intuitive proof.

$$\begin{aligned}
\widetilde{\mathbb{E}}[Y] &= \int Y(\omega) d\widetilde{\mathbb{P}} \\
&= \sum_{\omega} Y(\omega)\widetilde{\mathbb{P}}(\omega) \\
&= \sum_{\omega} Y(\omega)Z(\omega)\mathbb{P}(\omega) \\
&= \int Y(\omega)Z(\omega)d\mathbb{P}(\omega) \\
&= \mathbb{E}[YZ]
\end{aligned}$$

$\square$

Equivalent Probability Measures

Definition

Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two probability measures on $(\Omega, \mathcal{F})$. They are said to be equivalent if

$$\mathbb{P}(E) = 0 \Longleftrightarrow \widetilde{\mathbb{P}}(E) = 0$$

or equivalently

$$\mathbb{P}(E) = 1 \Longleftrightarrow \widetilde{\mathbb{P}}(E) = 1$$

Equivalent Probability Measures

In the risk-neutral pricing approach, we price derivatives under the risk-neutral measure $\widetilde{\mathbb{P}}$. This will require $\widetilde{\mathbb{P}}$ to be equivalent to the real-world measure $\mathbb{P}$ so that no arbitrage in one world implies no arbitrage in the other world.

Equivalent Probability Measures

In the risk-neutral pricing approach, we price derivatives under the risk-neutral measure $\widetilde{\mathbb{P}}$. This will require $\widetilde{\mathbb{P}}$ to be equivalent to the real-world measure $\mathbb{P}$ so that no arbitrage in one world implies no arbitrage in the other world.

$$\begin{aligned}
& \text{No arbitrage in real world} \\
\Longleftrightarrow\quad & \mathbb{P}(\text{arbitrage}) = 0 \\
\Longleftrightarrow\quad & \widetilde{\mathbb{P}}(\text{arbitrage}) = 0 \\
\Longleftrightarrow\quad & \text{No arbitrage in risk-neutral world}
\end{aligned}$$

Radon-Nikodym

Theorem

Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two equivalent probability measures on $(\Omega, \mathcal{F})$, then there exists an a.s. positive random variable Z such that

$$\mathbb{P}(E) = \int Z(\omega) d\widetilde{\mathbb{P}}(\omega)$$

for any $E \in \mathcal{F}$. We usually denote Z as $\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}$.

Independence

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}_i \subset \mathcal{F}$, $i = 1, \ldots, n$, be sub σ-algebras. $\mathcal{G}_i$ are (mutually) independent if

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n)$$

for all $A_i \in \mathcal{G}_i$.

Independence

Note that

- Independence is not equivalent to pairwise independence, which requires

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$$

for all $A_i \in \mathcal{G}_i$ and $A_j \in \mathcal{G}_j, i \neq j$.

Independence

Note that

- Independence is not equivalent to pairwise independence, which requires

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$$

  for all $A_i \in \mathcal{G}_i$ and $A_j \in \mathcal{G}_j, i \neq j$.

- Since we can take $A_i = \Omega$, the definition of independence is equivalent to

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_m}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_m})$$

  for all $A_{i_j} \in \mathcal{G}_{i_j}, j = 1, \ldots, m, m \leq n$.

Independence

Definition

Two random variables X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

Independence

Theorem

Let X and Y be two random variables. Their joint moment generating function is

$$M_{X,Y}(s,t) = \mathbb{E}[e^{sX+tY}]$$

Let

$$M_X(s) = \mathbb{E}[e^{sX}]$$

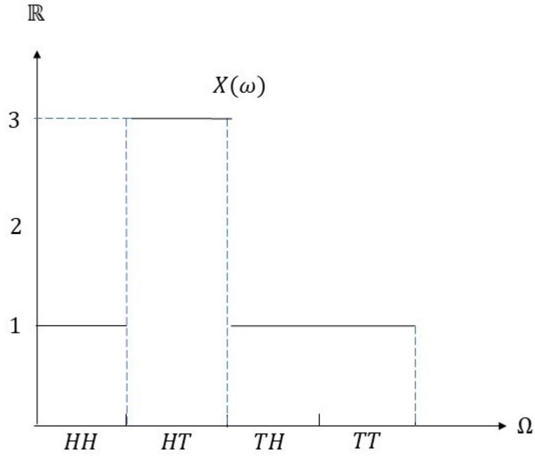
and

$$M_Y(t) = \mathbb{E}[e^{tY}]$$

be the moment generating function for X and Y respectively. Then X and Y are independent if and only if

$$M_{X,Y}(s,t) = M_X(s)M_Y(t)$$

Conditional Expectation



Conditional Expectation

Suppose we have the information $\mathcal{F}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$. Let $\omega = HH$, what will be our estimate of expected values for X?

Conditional Expectation

Suppose we have the information $\mathcal{F}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$. Let $\omega = HH$, what will be our estimate of expected values for X?
Answer: Since our information is $\mathcal{F}_1$, given $\omega = HH$, we can only know that the event $\{HH, HT\}$ has happened. But we don't know whether it is HH or HT. Hence our best estimate of X is $\frac{1+3}{2} = 2$.

Conditional Expectation

And what if $\omega = TH$?

Conditional Expectation

And what if $\omega = TH$?

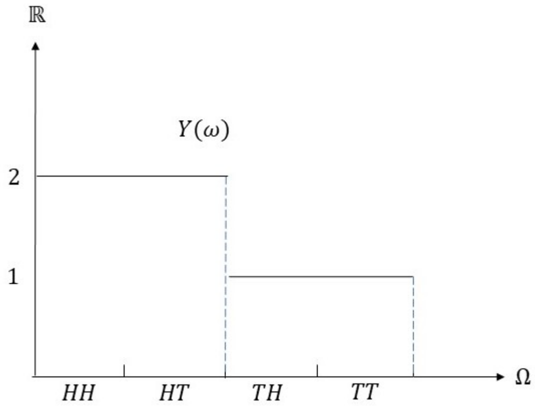
Given our information $\mathcal{F}_1$, we know $\omega = TH$ or TT. In either case, the value of X is 1. Hence our estimate will be 1.

Conditional Expectation

In summary, our estimate of X based on the information $\mathcal{F}_1$ is

$$Y(\omega) = \begin{cases} 2 & \text{if } \omega \in \{HH, HT\} \\ 1 & \text{if } \omega \in \{TH, TT\} \end{cases}$$

Conditional Expectation



Conditional Expectation

Note that

- Y is a random variable
- Y is $\mathcal{F}_1$-measurable
- Y has the same expectation as X on set $\{HH, HT\}$ and $\{TH, TT\}$.

Conditional Expectation

Definition

Let X be an integrable random variable, $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. A random variable Y is the conditional expectation of X given $\mathcal{G}$ if

1. Y is $\mathcal{G}$-measurable
2. $$\int_A Y(\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega)$$
   for all $A \in \mathcal{G}$.

We denote $Y = \mathbb{E}[X|\mathcal{G}]$.

Conditional Expectation

Properties

1. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

Properties

1. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.
2. If Y is $\mathcal{G}$-measurable, then $\mathbb{E}[YX|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$.

Conditional Expectation

Properties

1. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.
2. If Y is $\mathcal{G}$-measurable, then $\mathbb{E}[YX|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$.
3. If $\mathcal{H} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$

Conditional Expectation

Properties

1. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.
2. If Y is $\mathcal{G}$-measurable, then $\mathbb{E}[YX|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$.
3. If $\mathcal{H} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$
4. If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$

Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. Suppose the random variables $X_1, \ldots, X_m$ are $\mathcal{G}$-measurable and $Y_1, \ldots, Y_n$ are independent of $\mathcal{G}$. $f(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is a Borel function. Define

$$g(x_1, \ldots, x_m) = \mathbb{E}[f(x_1, \ldots, x_m, Y_1, \ldots, Y_n)]$$

Then

$$\mathbb{E}[f(X_1, \ldots, X_m, Y_1, \ldots, Y_n))|\mathcal{G}] = g(X_1, \ldots, X_m)$$

Stochastic Process

Definition

A stochastic process $X(t, \omega)$ is a function from $[0, \infty) \times \Omega$ to $\mathbf{R^n}$.

[2]

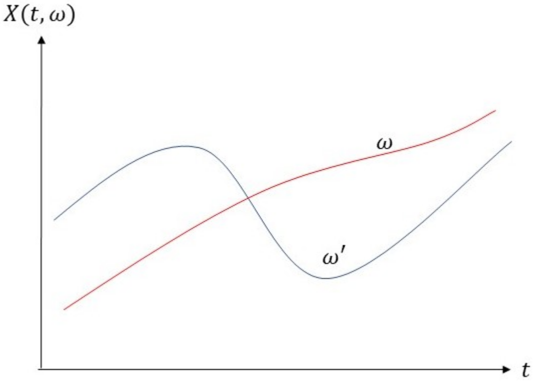
[2] Strictly speaking we need to impose some sort of measurability in the definition, but we ignore it here.

Stochastic Process

In the definition t is interpreted as time. Fix t, $X(t,\omega)$ is a random variable. Hence a stochastic process $X(t,\omega)$ can be interpreted as random variables that evolve with time. For example, it can be stock price.

Stochastic Process

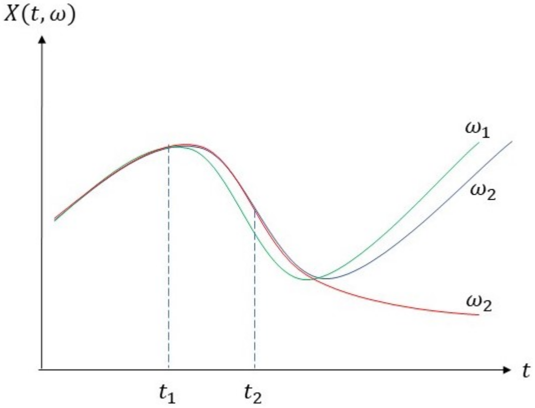
We can also interpret X in another way. Fix ω, $X(\cdot, \omega)$ is a function on $[0, \infty)$. Hence X is a function from Ω to the space of all functions on $[0, \infty)$. This function is called a sample path of X.



Stochastic Process

In the second interpretation, an $\omega \in \Omega$ has been picked. However we don't know exactly ω. Our observation is the value of $X(t, \omega)$ up to time t. There could be many ω's that will have the same sample path up to time t. Based on our observation, we can differentiate ω's whose sample paths up to time t are different from what we have observed, but not the others.

Stochastic Process



Filtration

We model the accumulation of information with filtration.

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{\mathcal{F}_t\}_{t\geq 0}$ be a collection of sub σ-algebra of $\mathcal{F}$. $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration if

$$\mathcal{F}_s \subset \mathcal{F}_t, \qquad \forall\ s \leq t$$

Adapted Process

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration. A stochastic process X is said to be adapted to this filtration if $X(t)$ is $\mathcal{F}_t$-measurable.

Filtration from a Stochastic Process

Definition

Let X be a stochastic process, we can define

$$\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$$

Then

- $\{\mathcal{F}_t^X\}_{t\geq 0}$ is a filtration
- X_t is adapted to $\{\mathcal{F}_t^X\}_{t\geq 0}$

$\{\mathcal{F}_t^X\}_{t\geq 0}$ is called the filtration generated from X.

$\mathcal{F}_t^X$ is the information one has by observing the value of X up to time t.

Martingale

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on it. A stochastic process $M(t)$ is a martingale if

- $\mathbb{E}[|M(t)|] < \infty$
- $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$ for all $s < t$.

Martingale

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on it. A stochastic process $M(t)$ is a martingale if

- $\mathbb{E}[|M(t)|] < \infty$
- $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$ for all $s < t$.

Note that M is

- A supermartingale if $\mathbb{E}[M(t)|\mathcal{F}_s] \leq M(s)$
- A submartinagle if $\mathbb{E}[M(t)|\mathcal{F}_s] \geq M(s)$

Markov Process

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on it. A stochastic process $X(t)$ is a Markov process if for any Borel function f and $s < t$, there is a function g such that

$$\mathbb{E}[f(X(t))|\mathcal{F}_s] = g(X(s))$$

Markov Process

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on it. A stochastic process $X(t)$ is a Markov process if for any Borel function f and $s < t$, there is a function g such that

$$\mathbb{E}[f(X(t))|\mathcal{F}_s] = g(X(s))$$

The estimate of $f(X(t))$ at time s depends only on the value of X at time s and not on the path of the process before time s.

Markov Process

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on it. A stochastic process $X(t)$ is a Markov process if for any Borel function f and $s < t$, there is a function g such that

$$\mathbb{E}[f(X(t))|\mathcal{F}_s] = g(X(s))$$

The estimate of $f(X(t))$ at time s depends only on the value of X at time s and not on the path of the process before time s. The condition is equivalent to

$$\mathbb{E}[f(X(t))|\mathcal{F}_s] = \mathbb{E}[f(X(t))|X(s)]$$

Brownian Motion

Definition

A stochastic process W is a Brownian motion if

1. For all $\omega \in \Omega$, $W(t, \omega)$ as a function of t is continuous.
2. For $0 = t_0 < t_1 < \ldots < t_n$, the increments $W(t_0), W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ are independent.
3. For any $s < t$, $W(t) - W(s)$ are normally distributed with mean 0 and variance $t - s$.

Filtration for Brownian Motion

Definition

Let $\{\mathcal{F}_t\}$ be a filtration. It is a filtration for the Brownian motion $W(t)$ if

- $W(t)$ is adapted to $\{\mathcal{F}_t\}$
- $\forall t > s$, $W(t) - W(s)$ is independent of $\mathcal{F}_s$

Properties of Brownian Motion

Properties of Brownian Motion

- $W(t)$ is a martingale
- $W(t)$ is a Markov process

Properties of Brownian Motion

Proof of Markov Property.

For any Borel function f and $s < t$, we shall find a function g such that

$$\mathbb{E}[f(W(t))|\mathcal{F}_s] = g(W(s))$$

Note that

$$\mathbb{E}[f(W(t))|\mathcal{F}_s] = \mathbb{E}[f(W(t) - W(s) + W(s))|\mathcal{F}_s]$$

Since $W(t) - W(s)$ is independent of $\mathcal{F}_s$ and $W(s)$ is $\mathcal{F}_s$-measurable, we can compute the conditional expectation on the RHS as

$$\mathbb{E}[f(W(t) - W(s) + W(s))|\mathcal{F}_s] = g(W(s))$$

where $g(x) = \mathbb{E}[f(W(t) - W(s) + x)]$. This proves that $W(t)$ is a Markov process.

Transition Density of Brownian Motion

In the above proof, we can compute the function g explicitly.

$$\begin{aligned} g(x) &= \mathbb{E}[f(W(t) - W(s) + x))] \\ &= \int f(u + x) \frac{1}{\sqrt{2\pi(t-s)}} e^{\frac{u^2}{2(t-s)}} du \\ &= \int f(y) \frac{1}{\sqrt{2\pi(t-s)}} e^{\frac{(y-x)^2}{2(t-s)}} dy \end{aligned}$$

Mathematical Foundation

Wu Lei

Probability

Stochastic Process

Stochastic Calculus

Transition Density of Brownian Motion

Define $p(t-s, x, y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{\frac{(y-x)^2}{2(t-s)}}$, then

$$\mathbb{E}[f(W(t))|\mathcal{F}_s] = \int f(y) p(t-s, W(s), y) dy$$

$p(t-s, x, y)$ is the transition density for $W(t)$ given $W(s) = x$.

Quadratic Variation

Definition

Fix $T > 0$, the quadratic variation of a function f is

$$\lim_{||\Pi|| \to 0} \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^2$$

where $\Pi : 0 = t_0 < t_1 < \ldots < t_n = T$ is a partition of the interval $[0, T]$ and $||\Pi|| = \max_i |t_i - t_{i-1}|$.

Quadratic Variation

If f is smooth enough (e.g., f has continuous first-order derivative), $f(t_i) - f(t_{i-1})$ is small. And $(f(t_i) - f(t_{i-1}))^2$ is even smaller. Hence the summation in the definition will go to 0. However this is not the case for Brownian motion.

Quadratic Variation of Brownian Motion

For each $\omega \in \Omega$, $W(\cdot, \omega)$ can be viewed as a function of t. Hence we can define quadratic variation for $W(\cdot, \omega)$ as

$$\lim_{||\Pi|| \to 0} \sum_{i=1}^{n} (W(t_i, \omega) - W(t_{i-1}, \omega))^2$$

provided that the limit exists.

Quadratic Variation of Brownian Motion

Theorem

The quadratic variation of a Brownian motion W on $[0, T]$ is

$$[W, W](T) = T$$

almost surely.

Quadratic Variation of Brownian Motion

Theorem

The quadratic variation of a Brownian motion W on $[0, T]$ is

$$[W, W](T) = T$$

almost surely.

- The quadratic variation of a stochastic process is defined for each ω. Hence it is usually a random variable.

Quadratic Variation of Brownian Motion

Theorem

The quadratic variation of a Brownian motion W on $[0, T]$ is

$$[W, W](T) = T$$

almost surely.

- The quadratic variation of a stochastic process is defined for each ω. Hence it is usually a random variable.
- In this case, the quadratic variation is a non-zero constant.

Quadratic Variation of Brownian Motion

Brownian motion accumulates variation at the rate of 1 per unit time. Informally we write it as

$$dW(t)dW(t) = dt$$

Quadratic Variation of Brownian Motion

It is easy to see that

$$\lim_{||\Pi|| \to 0} \sum_{i=1}^{n} (W(t_i) - W(t_{i-1}))(t_i - t_{i-1}) = 0$$

and

$$\lim_{||\Pi|| \to 0} \sum_{i=1}^{n} (t_i - t_{i-1})^2 = 0$$

This can be written as

$$dW(t)dt = 0$$

and

$$dtdt = 0$$

Realized Volatility

Suppose $S(t)$ follows a geometric Brownian motion as in the BSM model, then

$$S(t) = S(0)e^{(r-\frac{\sigma^2}{2})t+\sigma W(t)}$$

We often use

$$\frac{1}{n}\sum_{i=1}^{n}\left(\ln(\frac{S(t_i)}{S(t_{i-1})})\right)^2$$

to estimate σ^2. Now we justify it.

Realized Volatility

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left[\ln\left(\frac{S(t_i)}{S(t_{i-1})}\right)\right]^2 \\
= \;& \frac{1}{n}\sum_{i=1}^{n}\left[(r-\tfrac{1}{2}\sigma^2)\Delta t_i + \sigma\Delta W_{t_i}\right]^2 \\
= \;& \frac{1}{n}\sum_{i=1}^{n}\Big[(r-\tfrac{1}{2}\sigma^2)^2(\Delta t_i)^2 + \\
& \qquad \sigma^2(\Delta W_{t_i})^2 + (r-\tfrac{1}{2}\sigma^2)\sigma\Delta t_i\Delta W_{t_i}\Big] \\
\to \;& \frac{T}{n}\sigma^2
\end{aligned}$$

which is daily variance if we take t_i to be daily.

Thank you!