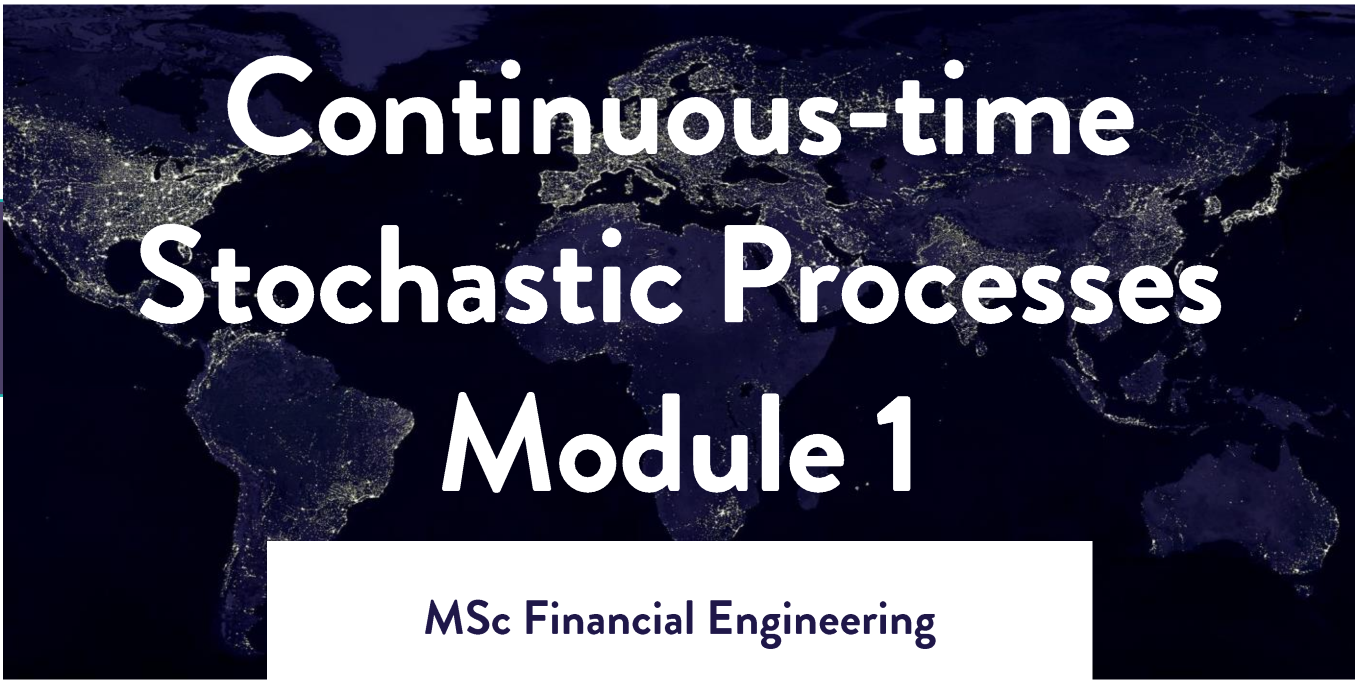

Continuous-time Stochastic Processes Module 1

MSc Financial Engineering

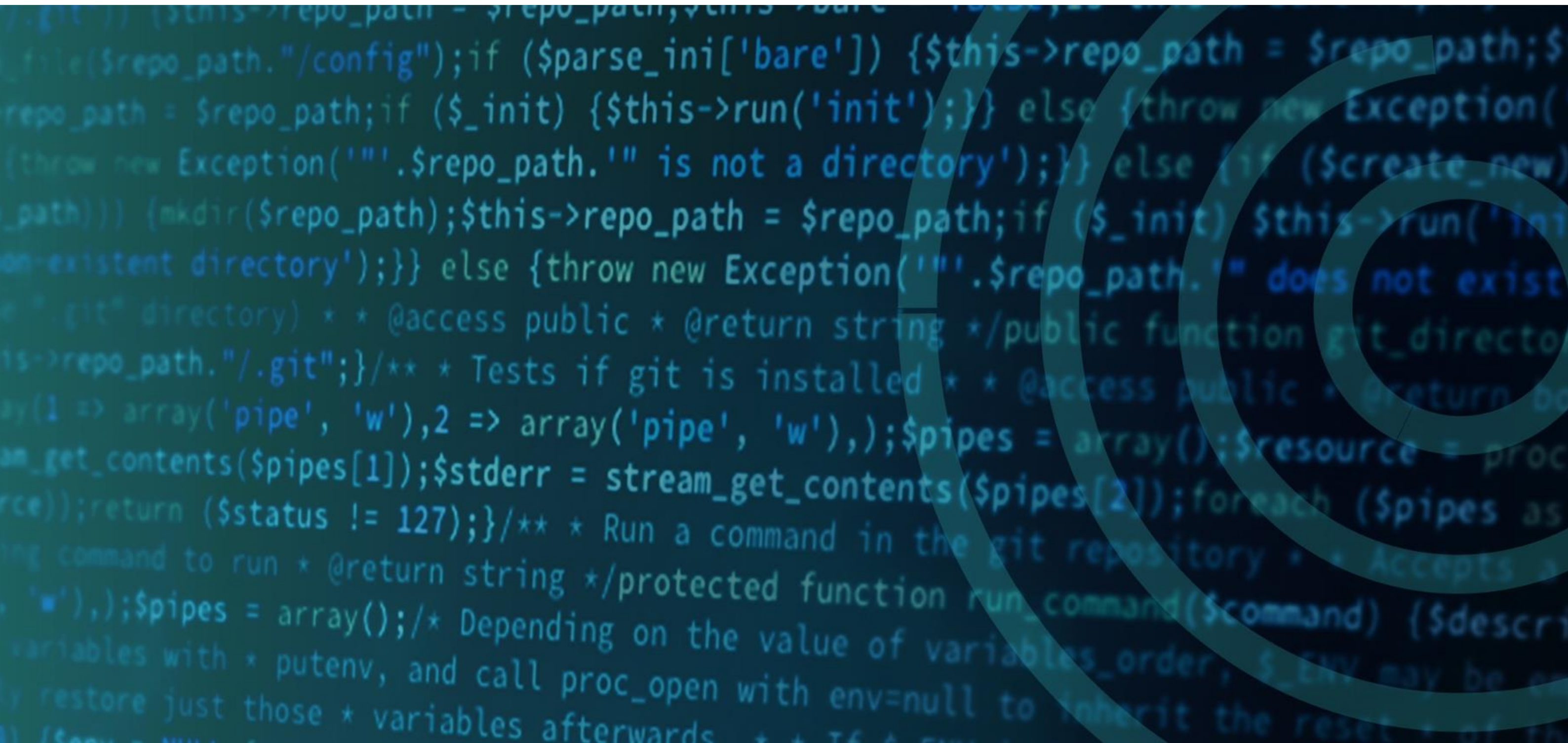

Table of Contents

1. Brief

This document contains the core content for Module 1 of Continuous-time Stochastic Processes, entitled Brownian Motion and Continuous-time Martingales. It consists of five sets of notes and three video lecture scripts.

2. Course Context

Continuous-time Stochastic Processes is the fourth course presented in the WorldQuant University (WQU) Master of Science in Financial Engineering (MScFE) program. The aim of the course is to introduce derivative pricing when trading happens in continuous time. It begins by focusing on the stochastic calculus of Brownian motion and its generalization to continuous semimartingales. These ideas are then applied to continuous trading and the pricing of financial derivatives.

2.1 Course-level Learning Outcomes

Upon completion of the Continuous-time Stochastic Processes course, you will be able to:

1. Define and identify Brownian motion processes in multiple dimensions.
2. Solve stochastic differential equations.
3. Apply Ito's Lemma for continuous semimartingales.
4. Apply Girsanov's theorem to construct equivalent local martingale measures.
5. Price and hedge derivatives in various asset price models.
6. Derive the Black-Scholes partial differential equation.
7. Construct asset prices models based on Levy processes.
8. Price interest rate derivatives.

2.2 Module Breakdown

The Continuous-time Stochastic Processes course consists of the following one-week modules:

1. Brownian Motion and Continuous-time Martingales
2. Stochastic Calculus I: Ito Process
3. Stochastic Calculus II: Semimartingales
4. Continuous Trading
5. The Black-Scholes Model
6. An Introduction to the Levy Processes
7. An Introduction to Interest Rate Modeling

3. Module 1:

Brownian Motion and Continuous-time Martingales

Welcome to the first module of the Continuous-time Stochastic Processes course. In this module, we will review some probability concepts in order to provide you with the correct tools to understand the rest of the course. Moreover, in this module you will also see the most important continuous time stochastic process: the Brownian motion. Although we will study Brownian motions in more detail during the second module, it is important to have a first contact with this key concept. However, first we will take a look at some useful constructs in probability theory.

3.1 Module-level Learning Outcomes

After completing this module, you will be able to:

1. Construct the product of two or more probability spaces.
2. Interpret the convergence of a sequence of random variables.
3. Define and identify a Brownian motion in multiple dimensions.
4. Calculate the quadratic variation of a martingale.

3.2 Transcripts and Notes

3.2.1 Notes: Introduction to Probability Theory

In this section we will introduce the concepts of σ-algebra, probability space, and filtration.

σ- algebra

Let Ω be a nonempty set and let F be a collection of sub-sets of Ω . We say that F is a σ-algebra if:

a) the empty set ∅ belongs to F;
b) whenever a set A belongs to F, its complement A^c also belongs to F; and
c) whenever a sequence of sets $A_1; A_2;$ ... belongs to F, their union also belongs to F.

For example, if Ω = (1, 2, 3), then F is the following collection of 2^3 = 8 sets:

$$F = \{(1), (1,2), (1,2,3), (2), (1,3), \emptyset, (3), (2,3)\}.$$

It easy to check that the three above conditions hold in this simple example.

Probability space

Let Ω be a nonempty set and let F be a σ-algebra of subsets of Ω. A probability measure P is a function that, to every set $A \in F$, assigns a number in $[0,1]$, called the probability of A and written $P(A)$. We require:

a) $P(\Omega) = 1$; and

b) (countable additivity) whenever $A_1, A_2, \ldots$ is a sequence of disjoint sets
c) in F, then

$$P\left(\bigcup_{n-1}^{\infty} A_n\right) = \sum_{n-1}^{\infty} P\left(A_n\right).$$

The triple (Ω, F, P) is called a probability space.

From the above definition, we can say that a probability measure must satisfy the following conditions:

1 If A and B are disjoint sets in F,

$$P(A \cup B) = P(A) + P(B).$$

2 (finite additivity) If $A_1, A_2, \ldots, A_N$ are finitely many disjoint sets in F, then

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n).$$

3 In the special case that $N = 2$ and $A_1 = A, A_2 = B = A^c$, we get

$$P(A^c) = 1 - P(A).$$

Filtration

Let Ω be a nonempty set. Let T be a fixed positive number and assume that for each $t \in [0, T]$ there is a σ-algebra $F(t)$. Assume further that if $s \leq t$, then every set in $F(s)$ is also in $F(t)$. Then we call the collection of σ -algebras $F(t), 0 \leq t \leq T$, a filtration.

A filtration tells us about the information that will be available in the future. In other words, when we get to time t, we will know for each set in $F(t)$ that the true ω

(event) lies in that set.

Adapted stochastic process

Let Ω be a nonempty sample space equipped with the filtration $F(t), 0 \leq t \leq T$. Let $X(t)$ be a collection of random variables indexed by $t \in [0, T]$. We say this collection of random variables is an adapted stochastic process if, for each t, the random variable $X(t)$ is $F(t)$-measurable.

$\boldsymbol{F(t)}$-measurable

Remember that, generally speaking, a random variable X is $F(t)$- measurable if and only if the information in $F(t)$ is sufficient to determine the value of X. For instance, a portfolio position $\Delta(t)$ taken at t must be $F(t)$-measurable. In other words, it must depend only on information available to the investor at time t.

In most practical cases (along this course), asset prices, portfolio processes, and wealth processes will be adapted.

3.2.2 Transcript: Products

Hi, in this video we introduce the product of two measure spaces.

Let (Ω_1, F_1, μ_1) and (Ω_2, F_2, μ_2) be two measure spaces. We want to introduce a σ −algebra on the product space of Ω_1 and Ω_2 which we can draw on a graph as follows. The x-axis is Ω_1 and the y −axis is Ω_2. The product is just a set of all points on the plane. The first thing that we ought to do is introduce a σ −algebra in $\Omega_1 \times \Omega_2$. Then, once we have a σ −algebra, we're going to introduce a measure on the σ −algebra that is going to be compatible with the two measures, μ_1 and μ_2.

An obvious candidate for the σ −algebra would be the product $F_1 \times F_2$, which is just a set of all sets of the following form: $\{A_1 \times A_2 : A_1 \in F_1 \; and \; A_2 \in F_2\}$. This will therefore be an obvious candidate for a σ −algebra on the product. Unfortunately, however, this is not a σ −algebra as it is not closed under all the properties of a σ −algebra. So, what we do is define the product σ −algebra of $F_1 \otimes F_2$ as the smallest σ −algebra containing these collections. It is a σ −algebra generated by the product of $F_1 \times F_2$ and we call it the product σ −algebra. And, finally, we define a measure, which we call the product measure, by denoting it as follows: $\mu_1 \otimes \mu_2$. We first define it on this collection: $\{A_1 \times A_2 : A_1 \in F_1$ and $A_2 \in F_2\}$.

So, $\mu_1 \otimes \mu_2 (A_1 \times A_2) := \mu_1(A_1) \times \mu_2(A_2)$.

That defines the product measure only on this collection: $\{A_1 \, x \, A_2 : A_1 \in F_1 \; and \; A_2 \in F_2\}$ but we have to extend it to the σ −algebra generated by that collection. As we know, there is a unique extension from this collection to the whole σ −algebra and we are going to call this measure $(\mu_1 \otimes \mu_2)$ the product measure of μ_1 and μ_2.

Therefore, the product space will look like this: $(\Omega_1 \times \Omega_2, F_1 \otimes F_2, \mu_1 \otimes \mu_2)$ with its sample space being $(\Omega_1 \times \Omega_2)$, its σ −algebra being the product σ −algebra of $(F_1 \otimes F_2)$ and the measure being $(\mu_1 \otimes \mu_2)$. We are always going to assume that the two measures, μ_1 and μ_2, are σ −finite. This product space formulation satisfies many interesting properties:

1. $A \in F_1 \otimes F_2 \Rightarrow \Pi_i(A) \in F_1$
   The first one is that if a set A is product measurable, i.e. if it belongs to this product's σ −algebra, then the projections $\Pi_i(A)$ are measurable in the individual spaces. So, going back to our diagram, if I have a set A that belongs to the product σ −algebra, then its projection on to each individual space is measurable with respect to the σ −algebra on each individual space. The same applies to Ω_2.

2. $F \in m\, F_1 \otimes F_2 \Rightarrow \omega_2 \mapsto F(\omega_1, \omega_2)$
   The second property satisfied by the product measures is that if I have a function F that is measurable with respect to the product σ −algebra, then, if I fix one of the variables, ω_2, and I consider this as a function of ω_1, with ω_2 fixed, then this function: $\omega_2 \mapsto F(\omega_1, \omega_2)$ is also measurable with respect to the σ −algebra ω_1. The same applies when I fix ω_1 and I consider $\omega_1 \mapsto F(\omega_1, \omega_2)$ as a function of ω_2, this is measurable with respect to F_2.

Now, let's move on to integration. Integration with respect to the product measure is summarized in the form of two theorems. These theorems are called Fubini-Tonelli theorems and they say that if you have a product measurable function F to evaluate the integral of F over the product space $\Omega_1 \times \Omega_2$ with respect to the product measure $\mu_1 \otimes \mu_2$, this can be evaluated either in the normal way we evaluate integrals – by doing a simple function approximation – or we can evaluate it as an iterated integral. First, with respect to Ω_2, $F\, d\mu_2$ and then with respect to μ_1. That's the first way we can evaluate it. Or, we can evaluate it as an integral first over Ω_1 with respect to μ_1 and then with respect to μ_2. This is true provided the following holds:

1. Either $F \geq 0$, and in that case the theorem is called Tonelli's theorem, or
2. $F \in L^1\, of\, (\Omega_1 \times \Omega_R\, F_1 \otimes F_2, \mu_1 \otimes \mu_2)$ so it's integrable with respect to the product space.

Those two conditions guarantee that we can evaluate the integral of a function F over the product as an iterated integral without worrying about the order of integration.

Let's look at an example. Consider the space ℕ natural numbers, with the powerset as a σ −algebra and the counting measure, #. Consider the real line with the Borel σ −algebra and the measure μ that is defined as follows:

It is absolutely continuous with respect to the Lebesgue measure and its Raydon-Nikodym derivative, or density with respect to the Lebesgue measure, is given by this quantity here. To evaluate an integral of the following form — so, to evaluate, for instance, an integral over $\mathbb{N} \times \mathbb{R}$ of a function $2^{-n}, x^4$, with respect to the product of these two measures, so with respect to counting measure # and μ, since the integrand is non-negative, we can apply Tonelli's theorem to get that this is equal to an integral over the natural numbers of an integral over the real line of $x^4, 2^{-n}$ with respect to μ and this with respect to the counting measure.

We can further simplify this in many ways: the integral inside here is, ℕ, because ℕ depends only on the counting measure. We can take this out as a constant and get the integral over the natural numbers of 2^n with respect to the counting measure # (that's the first term), times the second term, which is the integral over R of x^4 times — with respect to μ because μ is absolutely continuous — which would just be e^{-x} times the indicator 0 to infinity of x with respect to the Lebesgue measure. These two integrals can be evaluated using the techniques of the previous sections.

Now that we've looked at products, in the next set of notes and video lecture, we're going to move on to convergence.

3.2.3 Notes: Further Probability Convergence

We now move on to convergence of a sequence of random variables. Throughout this subsection we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence of random variables (X_n) on $((\Omega, \mathcal{F}, \mathbb{P})$, and a random variable X. We want to define what it means for the sequence (X_n) to converge to the random variable X as $n \to \infty$.

Sequence convergence

A sequence of real numbers $x_1, x_2, \dots$ converges to a real number x if the distance between x_n and x, $d(x_n, x)$, becomes arbitrarily small as $n \to \infty$. We measure this distance using the absolute value:

$$d(x, y) := |x - y|, x, y \in \mathbb{R}.$$

Thus $x_n \to x$ as $n \to \infty$ if and only if $\mathrm{d}(\mathrm{x_n}, \mathrm{x}) \to 0$ as $n \to \infty$.

Since random variables are real-valued, we might think about using a similar distance when defining convergence of random variables. The only problem is that if X and Y are random variables, then

$$d(X, Y) = |X - Y|$$

is also a random variable. Thus, $d(X_n, X)$ could converge to zero for some outcomes, but not for others.

As an example, consider the random experiment of tossing infinitely many fair coins. A typical outcome of this experiment is an infinite sequence $\omega = \omega_1 \omega_2 \dots$, where $\omega_i = \mathrm{H}$ or T is the outcome of the ith toss. Thus,

$$\Omega = \{\omega = \omega_1\omega_2 \dots : \omega_i = \text{H } or \text{ T}\}.$$

Define the sequence of random variables $\{X_n : n = 1,2, \dots\}$ as

$$X_n(\omega) = \begin{cases} 1 & \text{if the } n\text{th toss is H} \\ 0 & \text{otherwise.} \end{cases}$$

So, for instance, $X_1(\text{HHTTH} \dots) = 1$, while $X_3(\text{HHTTH} \dots) = 0$.

Now, for the outcome corresponding to all heads $\omega = HHH \dots$, the sequence converges to 1, since $X_n(\omega) = 1$ for all n. On the other hand, for the outcome corresponding to all tails $\omega = TTT \dots$, the sequence $X_n(\omega)$ converges to 0, since $X_n(\omega) = 0$ for all n. However, for the outcome of alternative heads and tails $\omega = HTHT \dots$, the sequence $X_n(\omega)$ does not converge, since it alternates between 0 and 1.

Define the sequence of random variables $\{Y_n : n = 1,2,3 \dots\}$ as

$$Y_n := \frac{X_n}{n}, n = 1,2 \dots.$$

Then (Y_n) converges to the constant random variable $Y \equiv 0$ for any outcome $\omega \in \Omega$. Indeed, if $\omega \in \Omega$, then

$$0 \leq |Y_n(\omega) - Y(\omega)| = \left|\frac{X_n(\omega)}{n}\right| \leq \frac{1}{n} \to 0 \text{ as } n \to \infty.$$

So $Y_n(\omega)$ converges to $Y(\omega)$, no matter what the outcome is. With that, we can now define our first mode of convergence.

Almost sure convergence

Let $\{X_n : n = 1,2,3, \ldots\}$ be a sequence of random variables and X be a random variable. We say that X_n converges to X *almost surely* (written as $X_n \xrightarrow{a.s.} X$) if

$$\mathbb{P}\left(\left\{\lim_{n \to \infty} X_n = X\right\}\right) = 1.$$

That is, if the set of outcomes ω for which $X_n(\omega)$ converges to $X(\omega)$ has probability one. In the previous example, the set of outcomes for which $Y_n(\omega)$ converges to $\mathrm{Y}(\omega)$ is equal to the whole sample space, so Y_n converges almost surely to Y. That is,

$$\left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\} = \Omega \Rightarrow \mathbb{P}\left(\left\{\lim_{n\to\infty} Y_n = Y\right\}\right) = \mathbb{P}(\Omega) = 1.$$

This requirement is often too strong. We introduce two other, weaker modes of convergence.

Probability and distribution convergence

If $X_1, X_2, \ldots,$ is a sequence of random variables and X is a random variable, we say that (X_n) converges to X

1. In *probability* ($X_n \xrightarrow{\mathbb{P}} X$) if for any $\epsilon > 0$

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

2. In *distribution* ($X_n \xrightarrow{\mathcal{D}} X$) if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

For all continuity points x of F_X.

Almost sure convergence is the strongest, followed by convergence in probability. That is,

$$X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X \Longrightarrow X_n \xrightarrow{\mathcal{D}} X.$$

Let's look at an example. Let $X \sim N(0,1)$ and define $X_n := \left(1+\frac{1}{n}\right)X^2$ for each $n \geq 1$. Let $Y \sim \chi^2_1$; we show that $X_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$. First note that for $y \geq 0$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_0^y \frac{t^{-\frac{1}{2}}\left(\frac{1}{2}\right)^{\frac{1}{2}} e^{-\frac{1}{2}t}}{\Gamma\left(\frac{1}{2}\right)}\, dt = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

$$= 2\left(\Phi\left(\sqrt{y}\right) - 0.5\right) = 2\Phi\left(\sqrt{y}\right) - 1.$$

Hence,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 2\Phi\left(\sqrt{y}\right) - 1 & y \geq 0. \end{cases}$$

Since F_Y is continuous for all $y \in \mathbb{R}$, we will have to show that

$$\lim_{n\to\infty} F_{X_n}(y) = F_Y(y) \;\; \forall y \in \mathbb{R}.$$

Let $y \geq 0$, then

$$F_{X_n}(y) = \mathbb{P}\left(\left(1+\tfrac{1}{n}\right)X^2 \leqslant y\right) = \mathbb{P}\left(-\sqrt{\tfrac{n}{n+1}y} \leqslant X \leqslant \sqrt{\tfrac{n}{n+1}y}\right) = 2\mathbb{P}\left(X \leqslant \sqrt{\tfrac{n}{n+1}y}\right) = 2\left(\Phi\left(\sqrt{\tfrac{n}{n+1}y}\right) - \Phi(0)\right)$$

The case when $y < 0$ is easy. Thus, $X_n \xrightarrow{\mathcal{D}} Y$.

We claim that $X_n \xrightarrow{\mathbb{P}} X^2$. Pick $\epsilon > 0$. Then

$$\mathbb{P}(|X_n - X^2| > \epsilon) = \mathbb{P}\left(\left|\left(1+\frac{1}{n}\right)X^2 - X^2\right| > \epsilon\right) = \mathbb{P}(|X| > \sqrt{n\epsilon})$$
$$= 2\left(1 - \Phi\sqrt{ne}\,\right)) \to 2\,(1-1) = 0 \text{ as } n \to \infty.$$

Hence, $X_n \xrightarrow{\mathbb{P}} X^2$. Of course this result implies the previous one.

In fact, $X_n \xrightarrow{a.s.} X^2$. Indeed, for every outcome $\omega \in \Omega$,

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty}\,(1+\frac{1}{n})\,X^2(\omega) = X^2(\omega).$$

Hence,

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} X_n = X\right\}\right) = 1.$$

Convergence in $\mathcal{L}^p$

One last mode of convergence that we introduce is convergence in $\mathcal{L}^p$, where $p \geq 1$. We say that (X_n) converges to X in $\mathcal{L}^p$ (written $X_n \xrightarrow{\mathcal{L}^p}$) if

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n - X\,|^{\,p}\right) = 0.$$

Of course, this definition makes sense if all the random variables concerned are in $\mathcal{L}^p$. The case when $p = 1$ is called **convergence in mean**, while $p = 2$ is called **convergence in mean-square**. Convergence in $\mathcal{L}^p$ implies convergence in probability but the converse is not true. For that, we need the notion of *uniform integrability*.

Uniformly integrable

A collection of random variables $\{X_a : \alpha \in I\}$ is said to be **uniformly integrable** (UI) if

$$\lim_{K\to\infty} \sup_{\alpha\in I} \int_{\{|X_a|>K\}} |X_a| d\mathbb{P} = 0.$$

The following theorem gives a converse to the above implication:

Theorem 2.1. *Let* $X = (X_n)$ *be a sequence of random variables in* $\mathcal{L}^1$ *and* X *be a random variable in* $\mathcal{L}^1$*. The following are equivalent:*

1. $X_n \xrightarrow{\mathcal{L}^1} X.$
2. $X_n \xrightarrow{\mathbb{P}} X$ *and* (X_n) *is* UI.

Checking the UI condition is sometimes difficult. Here are some well-known results about UI random variables. Here $\mathcal{C} := \{X_a : \alpha \in I\}$ is a collection of random variables.

- If $\mathcal{C} = \{X\}$ consists of one element X such that $\mathbb{E}(|X|) < \infty$, then $\mathcal{C}$ is UI *(Hint: this is the most practical)*
- If there exists $Y \in \mathcal{L}^1$ such that $|X| \leq Y$ for all $X \in \mathcal{C}$, then $\mathcal{C}$ is UI
- If there exists $C > 0$ such that $\mathbb{E}(|X|)^p \leq C$ for all $X \in \mathcal{C}$ and some $p > 1$, then $\mathcal{C}$ is UI
- If $\mathcal{C} = \{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}\}$ where $X \in \mathcal{L}^1$, then $\mathcal{C}$ is UI.

3.2.4 Transcript: Convergence

Hi, in this video we study the convergence of sequences of random variables.

Let $(\Omega, F, \mathbb{P})$ be a probability space. Consider $(X_n)_{n=1}^{\infty}$ to be a sequence of random variables and let X be a random variable. We want to define what it means for (X_n) to converge to X. We are going to do it in four modes of convergence.

1 The first one is called almost sure convergence. We say that (X_n) converges to X almost surely, or with probability 1, if and only if the probability of this set $\mathbb{P}\left(\left\{\omega \in \Omega: \lim_{n \to \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$. If this probability is 1, then we say that X_n converges to X almost surely. This is more like pointwise convergence, it's just that we do not insist that it converges pointers everywhere, only almost everywhere.

$$X_n \xrightarrow{a.s.} X \Leftrightarrow \mathbb{P}\left(\left\{\omega \in \Omega: \lim_{n \to \infty} X_n(\omega) = X = (\omega)\right\}\right) = 1$$

2 The second mode of convergence is called convergence in probability. It is slightly weaker than the first mode of convergence. We say that X_n converges to X in probability if and only if for every ϵ positive, the limit as n tends to infinity of the probability of X_n minus X greater than ϵ is equal to 0. So, this is slightly weaker than almost sure convergence in the sense that if X_n converges to X almost surely then it also converges to X in probability.

$$X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow \forall \epsilon > 0, \lim_{n \to \infty} \mathbb{P}(\{|X_n - X| > \epsilon\}) = 0$$

3 The third mode of convergence is called convergence in distribution. We say that X_n converges to X in distribution if and only if the limit as n tends to infinity of CDF of X_n is equal to the CDF of X at X for all points of continuity X of the CDF of X. So, everywhere where F_X of X is continuous, this condition must hold. We can draw it in a diagram like this: so if this is the CDF of X and this is F_X, what it says is that every point where F_X is continuous, which is every point in the diagram, we must have the sequence of CDFs of the X_n's get closer and closer to F_X at X. This is weaker than convergence in probability in the sense that if X_n converges to X in probability then X_n converges to X in distribution.

$$X_n \overset{\mathcal{D}}{\rightarrow} X \Leftrightarrow \lim_{n \rightarrow \infty} F_{Xn}(x) = F_X(x) \text{ for all points for continuity } x \text{ of } F_X$$

4 The last mode of convergence that we look at is what is called convergence in $\mathcal{L}^p$. This says that X_n converges to X in $\mathcal{L}^p$ if and only if the limit as n tends to infinity of the expected value, the absolute value of X_n minus X to the power p, is equal to 0. We study this only for p greater than or equal to 1. This is not weaker than convergence in distribution. There's also no straightforward relationship between convergence almost surely and convergence in $\mathcal{L}^p$. When p is equal to 2 we call it convergence in mean square. When p is equal to 1 we call it convergence in mean.

$$X_n \overset{\mathcal{L}^p}{\rightarrow} X \Leftrightarrow \lim_{n \rightarrow \infty} \in (|X_n - X|^P) = 0$$

Let's look at an example. Let's take the following probability space: $([0,1,], \mathbb{B}([0,1]), \mathbb{P} = \lambda_1)$: Ω is [0,1], with the Borel σ- algebra of [0,1] and $\mathbb{P}$ is just equal to the one-dimensional Lebesgue measure.

Define $X_n = nI_{[0,\frac{1}{2}]}, n = 1, 2, \ldots$

We will define X to be the zero random variable. Then the claim is that X_n converges to X almost surely. Let's show that.

If we fix ω and the sample space $[0,1,]$ then if ω is strictly positive then there exists an N large enough, so N greater than or equal to 1 such that for every n greater than or equal to N we have $\frac{1}{n}$ and this must be less than ω. We can draw this in a diagram like this. This is $0 - 1$. If ω is positive at this point, there exists an n large enough here such that $\frac{1}{N}$ and everything beyond that is less than ω. But if that's the case then $X_n$will be 0 because X_n is n times the indicator of 0 to $\frac{1}{n}$. So, this implies that X_n of this ω will be equal to 0 for n greater than or equal to N which implies that X_n of ω converges to 0 as n tends to infinity. The only potential problem is ω is equal to 0 itself and in that case this does not converge to 0.

But we can ignore that because the point 0 has Lebesgue Measure 0 since $\mathbb{P}$ is equal to the Lebesgue measure. So, therefore, we can safely say that X_n converges to X which is the zero random variable almost surely. On the other hand, we want to show that X_n does not converge to 0 in $L1$. So, if you calculate the expected value of X_n - 0 to the power 1, which is the expected value of X_n, because this is 0 so this is positive, and that will be equal to this case n times $\frac{1}{n}$ which is 1 and this does not converge to 0. So, we have found an example of a sequence of random variables that converges to 0 almost surely but does not converge to 0 in $L1$.

$$E(|X_n - 0|) = E(X_n) = n\frac{1}{n} = 1 \nrightarrow 0$$

3.2.5 Transcript: Brownian Motion

Hi, in this video we study Brownian motion.

A continuous-time stochastic process $W = \{W_t : t \geq 0\}$ is called a Brownian motion if it satisfies the following conditions:

1. $W_0 = 0$. It starts at 0 – in other words, all of its trajectories start at 0.

2. W must have continuous sample paths; meaning that all its trajectories are continuous.

3. W must have independent increments. So, on a diagram what that means is that if we have some time points, $t_1, t_2, t_{3,}$ and t_4, then the random variables – i.e. the increment $W_{t2} - W_{t1}$, between these two time points (t_1 and t_2); and the increment $W_{t4} - W_{t3}$, between these two time points (t_3 and t_4) – are independent. We can extend this to any finite number of increments, as long as they are non-overlapping.

4. Brownian motion has stationary normally distributed increments. So, for s less than t, we have $W_t - W_s$, which has a normal distribution with mean 0 and variance $t - s$. We can also illustrate this on a diagram.

So, here we have time s and time t, and the distribution of the increment is normal, with mean 0, centered at 0, and the variance is dependent on the length of the interval and only on the length of the interval. Of course, this implies, then, that W_t itself is normal with mean 0 and variance t.

Now, from these properties we can derive other properties of a Brownian motion:

1. For $s < t, W_s$ and W_t are jointly normal. Recall that we said earlier that W_t has a normal distribution marginally; therefore, W_s will also have a normal distribution marginally. However, that alone does not imply that W_s and W_t have a joint normal distribution. In the case of a Brownian motion though, we can get that as follows:

$$\begin{pmatrix} W_s \\ W_t \end{pmatrix} = \begin{pmatrix} W_s - W_0 \\ W_t - W_s + W_s - W_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} W_s - W_0 \\ W_t - W_s \end{pmatrix}.$$

We can write this as a matrix times two independents, because these are non-overlapping increments. They are independent and therefore they are jointly normally distributed. We are multiplying them by an invertible matrix and therefore the resulting expression will also have a bivariate normal distribution.

2 The covariance of (W_s, W_t) is the minimum of s and t. In other words, if s is less than t, this will be equal to s.

We get that using the properties as follows:

We write the covariance of W_s and W_t, which is the expected value of $(W_s\, W_t)$ minus the product of the expectations; but the product of the expectation is just 0×0, since each one of them has mean 0. We can rewrite this as:

$$E\big(W_s(W_s + W_t - W_s)\big) = E(W_s^2) + E\big(W_s(W_t - W_s)\big).$$

This is equal to the expected value of W_s^2, (W_s has a normal with mean 0 and variance s), making it equal to s; plus, these two quantities are independent, because they are non-overlapping increments, so this will be $E(W_s)\, E(W_t - W_s) = s, s < t$. Similarly, if t is less than s, the covariance will be equal to t and hence the result. Written in full:

$$Cov(W_s, W_t) = t \wedge s$$

$$Cov(W_s, W_t) = E(W_s\, W_t) - 0 = E(W_s(W_s + W_t - W_s))$$

$$= E\,(W_2^2) + E\,(W_s\,(W_t - W_s))$$

$$= s + E(W_s)\, E(W_t - W_s) = s, s < t.$$

3 Another important property of Brownian motion is that W is a martingale with respect to the natural filtration of W, or any filtration that contains that. So, the main martingale property says that if $s < t$, to calculate the conditional expectation of $(W_t | \mathcal{F}_S^W)$, we use the same trick again of creating an increment. So, this will be $E(W_s + W_t - W_s | \mathcal{F}_S^W)$, which, if you split the expectation, will be W_s, because it's adapted. In addition, this increment is independent of $\mathcal{F}_s$ and therefore, this will be the expected value of the increment, which is 0 and that is equal to W_s. That is the martingale property. Written in full:

$$E(W_t | \mathcal{F}_S^W = E(W_s + W_t - W_s | \mathcal{F}_S^W)$$

$$= W_s + 0 = W_s.$$

4 Finally, W is a Markov process. What that means is that if $\mathcal{F}$ is a function from $\mathbb{R} \rightarrow \mathbb{R}$ that is bounded and Borel-measurable, then the expected value of $(\mathcal{F}(W_t) | \mathcal{F}_S^W)$ is just equal to the expected value of $(\mathcal{F}(W_t) | W_s)$. This is true for $s < t$. It's is also a very useful condition that says that if we have information about Brownian motion, up to and including time s, we can simply replace that information by the value of Brownian motion at time s. Written in full:

$$E(\mathcal{F}(W_t) | \mathcal{F}_S^W) = E(\mathcal{F}(W_t | W_s), \;\; s < t.$$

3.2.6 Notes: Brownian Motion

In this section we introduce what is by far the most important example of a continuous-time stochastic process: *Brownian motion*. A Brownian motion plays a similar role played by a random walk in discrete-time stochastic process. It provides examples of many classes of processes including martingales, Markov processes, and *Levy processes*.

Martingale

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. Recall that a stochastic process $X = \{X_t : t \geq 0\}$ is a martingale if X is $\mathbb{F}$-adapted, integrable and for $s \leq t$,

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s.$$

Brownian motion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process $W = \{W_t : t \geq 0\}$ is called a **Brownian motion** or **standard Brownian motion** if it satisfies the following:

1. $W_0 = 0 \; \mathbb{P} - a.s.$
2. W has continuous sample paths.
3. W has independent increments; i.e. for each

$$0 \leq t_0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5 \ldots \leq t_n - 1 < t_n,$$

the random variables

$$W_{t_1} - W_{t_0}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$$

are independent.

4 W has normally distributed, stationary increments:

$$W_t - W_s \sim N\,(0, t-s) \text{ for } s < t.$$

If $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a filtered probability space and $\mathcal{F}_t^W \subseteq \mathcal{F}_t$ for each $t \geq 0$, then we call $W = \{W_t : t \geq 0\}$ a Brownian motion with respect to $\mathbb{F}$ if it satisfies 1 - 4 and the stronger condition that

5 For $0 \leq s < t, W_t - W_s$ is independent of $\mathcal{F}_s$.

We have "defined" a Brownian motion by specifying certain properties that it must satisfy. One can ask if a stochastic process satisfying these conditions even exists; perhaps these requirements are incompatible. Fortunately, the answer is positive. Indeed, Nobert Weiner proved the following theorem:

Theorem 3.1 (Weiner). *Brownian motion exists. That is, there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a stochastic process W such that W is a Brownian motion* on $(\Omega, \mathcal{F}, \mathbb{P})$.

There are several proofs of this theorem. A popular one constructs a Brownian motion as a limit of a sequence of simple random walks as the step sizes get smaller and smaller.

Let's look at an example. Let W be a Brownian motion. For $\alpha > 0$, define $X_t^\alpha := \frac{1}{\alpha} W_{\alpha^2 t}$ for $t \geq 0$. Then X^α is a Brownian motion. For each $T > 0$, the process $Y_t := W_{t+T} - W_T$ is also a Brownian motion. Check these as an exercise.

Here are some further properties:

Proposition 3.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $W = \{W_t : t \geq 0\}$ *be a Brownian motion. If* $s, t \in (0, \infty)$ *we have*

1 $W_t \sim N(0, t)$

2 $Cov\ (W_t, W_s) = t \wedge s$

3 W_s *and* W_t *have a Bivariate normal distribution.*

Proof.

1 Note that $W_t = W_t - W_0$, so the result follows.

2 If $s < t$, then

$$Cov\ (W_s, W_T) = Cov\ (W_s, W_s + W_t - W_s) = Cov\ (W_s, W_s) + Cov\ (W_s - W_0, W_t - W_s)$$
$$= Var\ (W_s) +\ 0\ =\ s.$$

Similarly, $Cov\ (W_s, W_t) =\ t$ when $t < s$.

3 Notice that

$$\begin{pmatrix} W_t \\ W_s \end{pmatrix} = \begin{pmatrix} W_t - W_s + W_s - W_0 \\ W_s - W_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} W_t - W_s \\ W_s - W_0 \end{pmatrix}.$$

Since W_t – W_s and $W_s - W_0$ are independent normal random variables, they are jointly normal, which implies that W_t and W_s are also jointly normal.

Theorem 3.2. $W = \{W_t\!: t\ \geq 0\}$ and $M = W_t^2 - t : t \geq 0\}$ *are martingales with respect to* $\mathbb{F}^W$.

Proof.

The first two properties are obvious. Let $s < t$ ($s =\ t$ is clear). Then,

$$\mathbb{E}\ (W_t | \mathcal{F}_s^W) = \mathbb{E}\ (W_s + W_t - W_s | \mathcal{F}_s^W) = W_s + 0 = W_s.$$

Showing that M is a martingale is left as an exercise.

It turns out that there is a converse to this result.

Theorem 3.3 (Levy). *Let M be a continuous martingale such that $M_0 = 0$ and $\left\{M_t^2 - t : t \geq 0\right\}$ are martingales. Then M is a Brownian motion.*

Even though the paths of Brownian motion are continuous, it turns out that they are not differentiable.

3.2.7 Notes: Continuous-time Martingales

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. Recall that a stochastic process $X = \{X_t : t \geq 0\}$ is a martingale if X is $\mathbb{F}$-adapted, integrable and for $s \leq t$,

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s.$$

For a submartingale or a supermartingale, the appropriate inequality replaces the above equality.

A classic example of a continuous-time martingale is Brownian motion discussed in the previous section.

We will always assume that all martingales, submartingales, and supermartingales have cadlag sample paths and that filtrations satisfy the usual conditions. This is not restrictive since it can be shown that under mild conditions, these processes will have cadlag modifications.

A martingale X is uniformly integrable (UI) if the collection $\{X_t : t \geq 0\}$ is uniformly integrable. In this case, X_t converges almost surely as $t \to \infty$ to a random variable X_∞ and

$$X_t = \mathbb{E}(X_\infty | \mathcal{F}_t) \quad \text{for every } t \geq 0.$$

More generally, the martingale convergence theorem states that if $sup_{t \geq 0}\, \mathbb{E}(|X_t|) < \infty$, then X_t converges almost surely to a random variable X_∞. When this is the case, we will define $X_\tau(\omega) \coloneqq X_{\tau(\omega)}(\omega)$even when $\tau(\omega) = \infty$.

With all these concepts we can now state the celebrated *optional sampling theorem:*

Theorem 4.1 (Optional Sampling). *Let X be a UI martingale and τ and σ be stopping times with $\tau \leq \sigma$. Then,*

$$E(X_\sigma | F_\tau) = X_\tau.$$

We have analogous inequalities when X is a sub or super martingale.

Let us look at an example. Consider a Brownian motion W and let $a < 0 < b$ be real numbers. Define $\tau := \inf\{t \geq 0: W_t \notin (a, b)\}$. The martingale W^τ is bounded between a and b, hence it is UI (check this!). We can then apply the optional sampling theorem to get

$$0 = W_0 = W_0^\tau = \mathbb{E}(W_\tau^\tau | \mathcal{F}_0) = \mathbb{E}(W_\tau) = pa + (1-p)b,$$

where $p = \mathbb{P}(W_\tau = a)$. This allows us to solve for p.

We now try to generalize the Doob decomposition of a stochastic process to continuous-time. Unlike the discrete-time case, not every stochastic process X can be written as $X = X_0 + M + A$, where M is a martingale and A is predictable. To get such a decomposition, we will have to restrict our attention to a special class of processes.

A stochastic process X is of *class* D if the collection $\{X_\tau : \tau$ is a finite stopping time$\}$ is UI. We say that X is of *class* DL if for every $a > 0$, the collection $\{X_\tau : \tau$ is a stopping time with $\tau \leq \alpha\}$ is UI.

We can now state the Doob-Meyer decomposition theorem:

Theorem 4.2 (Doob-Meyer Decomposition). *Let X be a submartingale of class DL. Then there exists a unique martingale M and a unique predictable increasing process A, both null at zero, such that*

$$X = X_0 + A + M.$$

Furthermore, M is UI if and only if X is of class D. The pair (A, M) is called the Doob-Meyer decomposition *of X.*

Every non-negative submartingale is of class DL. Indeed, let $a > 0$, then for any stopping time τ with $\tau \leq \alpha$, we have (by the OST above)

$$X_\tau \leq \mathbb{E}\left(X_\alpha | \mathcal{F}_\tau\right).$$

Thus,

$$\{|X_\tau| = X_\tau \colon \tau \text{ is a stopping time with } \tau \leq \alpha\}$$

is dominated by

$$\{\mathbb{E}\left(X_\alpha | \mathcal{F}_\tau\right) : \ \tau \text{ is a stopping time with } \tau \leq \alpha\},$$

implying that X is of class DL.

Let N be a Poisson process with constant rate $\lambda > 0$. Then N is a non-negative submartingale, hence it is of class DL. The Doob decomposition of N is

$$N_t = 0 + (N_t - \lambda t) + \lambda t.$$

Now let X be a square integrable martingale; i.e. $\mathbb{E}(X_t^2) < \infty$ for all $t \geq 0$. From Jensen's inequality, the process X^2 is a non-negative submartingale, hence it is of class DL. Therefore, we can find its Doob-Meyer decomposition (A, M) such that

$$X^2 = X_0^2 + M + A.$$

The process A is called the **predictable quadratic variation** of the square integrable martingale X and is denoted by $\langle X \rangle$. That is, $\langle X \rangle$ is the unique predictable increasing process null at zero such that $X^2 - \langle X \rangle$ is a martingale.

Consider a Brownian motion W. We know that $W_t^2 - t$ is a martingale and the process $(t, \omega) \mapsto t$ is predictable. Hence $\langle W \rangle_t = t$.

Now let X and Y be square integrable martingales. We define the *predictable covariation process* of X and Y via polarization:

$$\langle X, Y \rangle := \frac{1}{4}(\langle X + Y \rangle - \langle X - Y \rangle).$$

The covariation process behaves like an inner product:

1. $\langle X, X \rangle \geq 0$
2. $\langle X, Y \rangle = \langle Y, X \rangle$
3. $\langle aX, Y \rangle = a\langle X, Y \rangle$
4. $\langle X, Y + Z \rangle = \langle X, Y \rangle + \langle X, Z \rangle$

Note that $\langle X, X \rangle = \langle X \rangle$.

It is easy to see that $\langle X, Y \rangle$ is the unique predictable process null at zero such that $XY - \langle X, Y \rangle$ is a martingale. Also, XY is a martingale if and only if $\langle X, Y \rangle = 0$, and in that case, we say that X and Y are *strongly orthogonal*.

Problem 1

Compute the following covariance:

$$Cov(7W_3 - W_2, W_1 - W_4).$$

Solution:
First remember that,

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

In our case, we can rewrite as follows:

$$Cov(7W_3 - W_2, W_1 - W_4) = E\,[(7W_3 - W_2)(W_1 - W_4)] - E(7W_3 - W_2)E(W_1 - W_4).$$

It's easy to show that the second term in the above equation is zero. In other words,

$$E(7W_3 - W_2) = \; 7E(W_3) - E(W_2).$$

As $W3$ and $W2$ are Brownian motions, both have zero expected value (check the Brownian motion properties in the lecture notes). The same applies to the other term: $E(W_1 - W_4)$. Thus, we only have,

$$Cov(7W_3 - W_2, W_1 - W_4) = E[(7W_3 - W_2)(W_1 - W_4)].$$

Multiplying,

$$E\,[(7W_3 - W_2)(W_1 - W_4)] = \; 7E(W_3W_1) - 7E(W_3W_4) - E(W_2W_1) + \; E(W_2W_4).$$

Finally, we know that (see in the lecture notes),

$$E(W_s W_t) = t \wedge s = \min(t, s).$$

Thus, we conclude that,

$$E\,[(7W_3 - W_2)(W_1 - W_4)] = \ 7 \times 1 - 7 \times 3 - 1 + 2 = -13.$$

Problem 2

Compute the following variance:

$$Var(W_s + 3W_t), 0 < s < t.$$

Solution:

First remember that,

$$Var(X) = E(X^2) - E(X)^2.$$

In our case, we can rewrite as follows:

$$Var(W_s + 3W_t) = E((W_s + 3W_t)^2) - E(W_s + 3W_t)^2.$$

The second term is zero, $E(W_s \ + \ 3W_t)^2 = \ 0$, and the first term is

$$E\,((\,W_s + 3W_t)^2) = E\,(W_s^2 \ + \ 6W_sW_t \ + \ 9W_t^2).$$

Remember that (see in the lecture notes)

$$E(W_s W_t) = t \wedge s \ = \ min(t, s).$$

Thus, we have,

$$E(W_s^2 + 6W_sW_t + 9W_t^2) = s + 6s + 9t = 7s + 9t.$$

Problem 3

Compute the following expected value:

$$E[W^4].$$

Solution:

Important to remember the n-central moment of a normal distribution, $E\left[\left(X - E(X)\right)^n\right]$ with mean, μ and standard deviation, σ.

$$m_{even} = \sigma^n(n-1)!!, m_{odd} = 0.$$

Taking the above equation into account, we can conclude that

$$E[W^4] = m_4 = 1^4 \times (4-1)!! = 3!! = 3 \times 1 = 3.$$

Final note: If n is odd the n-central moment is always zero, for instance $E[W^3] = 0$.

Problem 4

On the probability space $([0,1), \mathcal{B}([0,1)), \mathbb{P} = \lambda_1)$, define the sequence of random variables $(X_n)_{n=1}^{\infty}$ as follows:

$$X_n(\omega) := \omega^{\frac{2+n}{n^2+5}}, \omega \in [0,1) \text{ and } n \geq 1.$$

For each $\omega \in [0,1)$, find $X(\omega) := \lim_{n \to \infty} X_n(\omega)$.

Solution:

We just have to compute the limit as follows:

$$X_n(\omega) := \lim_{n \to \infty} X_n(\omega) = \lim_{n \to \infty} \omega^{\frac{2+n}{n^2+5}} = \omega^{\lim_{n\to\infty} \frac{2+n}{n^2+5}} = \omega^0 = 1.$$

Remember that almost sure convergence is the strongest, followed by convergence in probability. That is,

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathcal{D}} X.$$

Thus, for this example, we can conclude that converges almost sure, in probability and in distribution to 1.

Problem 5

Let $\{W_t : t \geq 0\}$ be a Brownian motion and let $X_t = W_t^2 + 5t$. If $X = A + M$ is the Doob-Meyer decomposition of X, then, compute A_t.

Solution:

First of all, we need to take into account the following Theorem (see in lecture notes), in other words, and simplifying the theorem, we already know that $M = \{M_t = W_t^2 - t : t \geq 0\}$ is a martingale.

Thus, we can rewrite the process X_t as follows:

$$X_t = W_t^2 + 5t + t - t = W_t^2 - t + 6t,$$

where we can make $M = W_t^2 - t,$ which is a martingale (see lecture notes), and $A = 6t$ is a predictable increasing process.

Problem 6

Let $W = \{W_t : t \geq 0\}$ be a Brownian motion. Then what is $\langle 3W, W + 1\rangle_t$ equal to?

Solution:

First, we are going to solve the example by using the following properties:

1. $\langle X,X\rangle \geq 0$
2. $\langle X,Y\rangle = \langle Y,X\rangle$
3. $\langle \alpha X,Y\rangle = \alpha\langle X,Y\rangle$
4. $\langle X,Y + Z\rangle = \langle X,Y\rangle + \langle X,Z\rangle$

First, we make use of (3) above as,

$$\langle 3W,W + 1\rangle_t = 3\langle W,W + 1\rangle_t.$$

Finally, we have to apply (4) to the above result,

$$\langle 3W,W + 1\rangle_t = 3\langle W,W + 1\rangle_t = 3(\langle W,W\rangle_t + \langle W,1\rangle_t) = 3 * (t + 0) = 3t.$$

We can get the same result by applying

$$\langle X,Y\rangle \coloneqq \frac{1}{4}\left(\langle X+Y\rangle - \langle X-Y\rangle\right) = \frac{1}{4}\left(\langle 4W\rangle - \langle 2W\rangle\right) = \frac{12t}{4} = 3t.$$

Problem 7

On the probability space $([0,1), \mathcal{B}([0,1)), \mathbb{P} = \lambda_1)$, define the sequence of random variables $(X_n)_{n=1}^{\infty}$ as follows:

$$X_n(\omega) \coloneqq \omega^n I_{[0,\frac{1}{2^n})}(\omega), \quad \omega \in [0,1) \text{ and } n \geq 1.$$

Is $X_n(\omega)$ UI?

Solution:

Although it is easy to show that $X_n \xrightarrow{\mathbb{P}} 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|X_n\right| = \lim_{n\to\infty} 2^{-n} = 0,\right.$$

to show that X_n is UI, we need to prove that $E(|X|) < \infty$ (this is just one possible way to proceed),

$$\mathbb{E}\left(|X| = \omega^n * 2^{-n} = (\omega / 2)^n.\right.$$

The above expected vale converges to zero $(n \to \infty)$ if $\omega < 2$. Thus, to conclude, as $\omega \in [0,1)$ we can say that X_n is UI.