MATH 627: Probability Theory

Felix Zhou [1]

June 1, 2023

[1]From Professor Elena Rudo's Lectures at Yale University in Fall 2022

Contents

Part I

Introduction

Chapter 1

Introduction

1.1 Mid-17th Century: Pascal, de Fermat

1.1.1 Classical Probability Model

All (finite) outcomes of a random experiment are equally probable, say $E = \{e_1, \dots, e_n\}$. We can identity these events with a *probability model* $\Omega = \{\omega_1, \dots, \omega_n\}$. Here each ω_i is an *elementary event*. An event is some $A \subseteq \Omega$ and the probability

$$\mathbb{P}\{A\} = \frac{|A|}{|\Omega|}.$$

Note that

1) $\mathbb{P}(A) \geq 0$
2) $\mathbb{P}(\Omega) = 1, \mathbb{P}(\varnothing) = 0$
3) $\forall A_1, A_2$ with $A_1 \cap A_2 = \varnothing$, $\mathbb{P}(A_1) + \mathbb{P}(A_2) = \frac{|A_1|+|A_2|}{|\Omega|} = \mathbb{P}(A_1 + A_2)$

1.1.2 18th Century: Bernoulli, de Moivre, Laplace

The concept of Bernoulli trials were introduced during this time.

1.1.3 19th Century: Generalizations

Our elementary event space now generalizes to countable cardinality $E = \{e_1, e_2, \dots\}$ with corresponding probability model $\Omega = \{\omega_1, \omega_2, \dots\}$. An event is again a subset $A \subseteq \Omega$.

However, we no longer enforce equiprobable events.

1) $\mathbb{P}(\omega_i) = p_i \geq 0$
2) $\sum_i p_i = 1$
3) $\mathbb{P}(A) = \sum_j \mathbb{P}(\omega_{i_j}) = \sum_j p_{i_j}$ where $A = \bigcup_j \{w_{i_j}\}$

Shortcomings

We are still limited to countably finite events, when bigger cardinalities were already known.

There was no concept of "geometric probability". Thus for $F \subseteq G \subseteq \mathbb{R}^N$, what is the probability that a random point of G belongs to F?

Independently, Brownian motion was already discovered. Moreover, Chebyshev, Lyapenov, and Markov introduced random real numbers, expectations, variance, and limit theorems.

Borel and Lebesgue introduced measure theory, which was evidently useful for probability.

Finally, Hilbert left the axiomatization of probability as his 6th problem.

1.2 1930: Kolmogorov

Kolmogorov came up with a measure theoretic axiomatization of probability and more importantly, applied this theory to study stochastic processes, showing the broad applicableness of this model.

Chapter 2

Axiomatic Approach to Probability

2.1 Probability Space

Definition 2.1.1 (Algebra)
An $\mathcal{F}$-system of subsets of Ω is an *algebra* if

a) $\Omega \in \mathcal{F}$

b) $A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$ ($A \in \mathcal{F}$ are called events or measurable sets)

c) $A_1, A_2 \in \mathcal{F}$ implies that $A_1 \cup A_2, A_1 \cap A_2 \in \mathcal{F}$

Definition 2.1.2 (σ-algebra)
An $\mathcal{F}$-system of subsets of Ω is a σ-*algebra* if

a) $\Omega \in \mathcal{F}$

b) $A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$ ($A \in \mathcal{F}$ are called events or measurable sets)

c) $A_1, A_2, \dots \in \mathcal{F}$ implies that $\bigcup_{n=1}^{\infty} A_i, \bigcap_{n=1}^{\infty} A_i \in \mathcal{F}$ (note that only one of the two is needed)

Definition 2.1.3 (Measurable Space)
$(\Omega, \mathcal{F})$ where $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra is a *measurable space.*

Definition 2.1.4 (Probability Measure)
Let $(\Omega, \mathcal{F})$ be a measurable space. Then $\mathbb{P} : \mathcal{F} \to \mathbb{R}_+$ is a *probabilistic measure* on $(\Omega, \mathcal{F})$ if

a) $\mathbb{P}(A) \geq 0$ (non-negativity)

b) $\mathbb{P}(\Omega) = 1$ (normalization)

c) $A_1, A_2, \ldots, \in \mathcal{F}$ with $A_i \cap A_j = \varnothing$ implies that $P(\sqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Definition 2.1.5 (Probability Space)
Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathbb{P}$ a probability measure. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space.*

Note that our new model strictly includes all previous models.

2.2 Experiments

We would like to differentiate the event space from our universal probability space. For instance, our probability space might consist of the outcomes of a dice roll. However, we can derive multiple event spaces such as the value of the roll or the parity of the roll from a single probability space.

Definition 2.2.1 ($\mathcal{F}/\mathcal{E}$-measurable)
Let $(E, \mathcal{E})$ be another measurable space. Then $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ is $\mathcal{F}/\mathcal{E}$-*measurable* if for all $B \in \mathcal{E}$, $X^{-1}(B) \in \mathcal{F}$.

Proposition 2.2.1
For all $B \in \mathcal{E}$, define $\mathbb{P}_x(B) := \mathbb{P}(X^{-1}(B))$. Then $\mathbb{P}_x$ is a probability measure on $(E, \mathcal{E})$ called the *distribution / law / law of distribution of* X.

Proof
Check definitions.

2.3 σ-Algebra & Algebras

Remark that a σ-algebra is an algebra. Moreover, a finite algebra is a σ-algebra.

Example 2.3.1
The *trivial* σ-algebra $\mathcal{F}_* := \{\varnothing, \Omega\} \subseteq \Omega$ is a σ-algebra.

$\mathcal{F}^* := 2^\Omega$ is the "richest" σ-algebra on Ω. If $|\Omega| < \infty$ like in classical probability, we always consider $\mathcal{F}^*$.

Let $A \subseteq \Omega$, then $\mathcal{F}_A := \{A, \bar{A}, \Omega, \varnothing\}$ is the *σ-algebra generated by A*.

Lemma 2.3.2
Let $\mathcal{B} \subseteq 2^\Omega$, there exists a smallest algebra and σ-algebra, denoted $\alpha(\mathcal{B}), \sigma(\mathcal{B})$, respectively, containing $\mathcal{B}$.

We say $\alpha(\mathcal{B}), \sigma(\mathcal{B})$ is the (σ-)algebra generated by $\mathcal{B}$.

Proof
Take $\sigma(\mathcal{B}) := \bigcap\{G : G \text{ is } \sigma\text{-algebra} \wedge \mathcal{B} \subseteq G\}$ and similarly for $\alpha(\mathcal{B})$.

2.3.1 Borel Sets & σ-Algebras

Let (Ω, ρ) be a metric space.

Definition 2.3.1 (Borel σ-Algebra)
Consider the metric topology on (Ω, ρ). The Borel σ-algebra is the one generated by open sets.

2.3.2 Borel σ-Algebra on $\mathbb{R}$

Note that similar things can be said for $[0, 1]$ and $\mathbb{R}^N$.

Theorem 2.3.3
We have

$$\begin{aligned}\mathcal{B}_\mathbb{R} &= \sigma\{(a, b) : -\infty \leq a < b \leq \infty\}\\ &= \sigma\{[a, b]\}\\ &= \sigma\{[a, b)\}\\ &= \sigma\{(a, b]\}\\ &= \sigma\{(-\infty, a)\}\\ &= \dots\end{aligned}$$

Proof
Any open set on $\mathbb{R}$ is a countable union of open intervals.

Remark that any finite union, intersection of intervals of the type $[a, b)$ are again intervals of the same type. Thus they form an algebra.

2.3.3 Cylinder σ-Algebras

Let Ω be an infinite-dimensional space, say $\Omega = \mathbb{R}^{\mathbb{N}}$.

Definition 2.3.2 (Cylinder Set)
Let $A_1, \dots, A_k \in \mathcal{B}_{\mathbb{R}}$ with $k \in \mathbb{N}$. A *cylinder set*

$$C_{i_1 i_2 \dots i_k}(A_1, A_2, A_3) := \{x \in \mathbb{R}^{\mathbb{N}} : x_{i_j} \in A_j\}$$

is contrained at finitely many coordinates.

Note that cylinder sets form an algebra.

Definition 2.3.3 (Cylinder σ-Algebra)
The σ-algebra generated by all cylinder sets.

2.3.4 σ-Algebra vs Algebras

Lemma 2.3.4
An algebra $\mathscr{A}$ is a σ-algebra if and only if $\mathscr{A}$ is a monotone class. That is, for all $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{A}$, then $A := \bigcup A_n \in \mathscr{A}$. Similarity if $A_1 \supseteq A_2 \supseteq \dots \in \mathscr{A}$, then $\bigcap A_n \in \mathscr{A}$.

Proof
($\implies$) Easy.

($\impliedby$) Check definitions. The only non-trivial aspect is σ-additivity. For $A_1, A_2, \dots \in \mathscr{A}$, consider $B_n = \bigcup_{i=1}^n A_i$ and observe that $\bigcup_n A_n = \bigcup_n B_n$.

2.4 (Probability) Measures

Let $(\Omega, \mathcal{F})$ be a measurable space.

Definition 2.4.1 (Measure)
A *measure* is a set function $\mu : \mathcal{F} \to \mathbb{R}$ such that

1) $\mu(A) \geq 0$ for all $A \in \mathcal{F}$
2) $\mu(\varnothing) = 0$
3) $A_1, A_2, \cdots \in \mathcal{F}$ with $A_i \cap A_j = \varnothing$ implies that $\mu(\bigsqcup_n A_n) = \sum_n \mu(A_n)$ (σ-additivity)

We say a measure is *finite* if $\mu(\Omega) < \infty$ and σ*-finite* if $\Omega = \bigcup_n \Omega_n$ with each $\mu(\Omega_n) < \infty$.

We say μ is *normalized* if $\mu(\Omega) = 1$ and also refer to it as a *probability measure*.

Note that the definition of a probability measure no longer requires $\mu(\varnothing) = 0$ since that can be derived using $\mu(\Omega) = 1$ and σ-additivity.

2.4.1 Basic Properties of Probability Measures

Proposition 2.4.1

1) $\mu(\bar{A}) = 1 - \mu(A)$
2) $0 \leq \mu(A) \leq 1$
3) $B \subseteq A \implies \mu(B) \leq \mu(A)$ (monotonicity)
4) For $A, B \in \mathcal{F}$,
$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \leq \mu(A) + \mu(B)$$
(semiadditivity)
5) $\mu(\bigcup_{i=1}^n A_i)$?
6) $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$ (semiadditivity)

Theorem 2.4.2 (Continuity of Measure)
Let $(\Omega, \mathcal{F})$ be a measurable space equipped with $\mu : \mathcal{F} \to \mathbb{R}_+$, a finitely additive set function. The following are equivalent:

1) μ is σ-additive (probability measure)
2) μ is continuous from below, thus for all $A_1 \subseteq A_2 \subseteq \dots \in \mathcal{F}$, $\lim_n \mu(A_n) = \mu(\bigcup_n A_n)$
3) μ is continuous from above
4) μ is continuous at $\varnothing$, thus if $A_1 \supseteq A_2 \supseteq \dots$ with $\bigcap_n A_n = \varnothing$, then $\lim_n \mu(A_n) = 0$

Proof
($1 \implies 2$) Observe that

$$\bigcup_n A_n = A_1 \sqcup (A_2 \setminus A_1) \sqcup (A_3 \setminus A_2) \sqcup \dots$$

Thus

$$\begin{aligned}\mu(\bigcup_n A_n) &= \mu(A_1) + \mu(A_2 \setminus A_1) + \mu(A_3 \setminus A_2) + \dots \\ &= \lim_n \mu(A_n).\end{aligned}$$

($2 \implies 3$) Consider the increasing sequence of complements.

($3 \implies 4$) Easy.

($4 \implies 1$) Let $A_1, A_2, \dots, \in \mathcal{F}$ be disjoint.

We claim that $\mu(\bigsqcup_{i=n+1}^\infty A_i) \to 0$ as $n \to \infty$. Indeed, let $B_n := \bigsqcup_{i=n+1}^\infty A_i$. Notice that $B_n \downarrow \bigcap_{n=0}^\infty B_n = \varnothing$. By assumption $\mu(B_n) = \mu(\bigsqcup_{i=n+1}^\infty A_i) \to 0$.

But then

$$\begin{aligned}\mu(\bigsqcup_{i=1}^\infty A_i) &= \mu(\bigsqcup_{i=1}^n A_i) + \mu(\bigsqcup_{i=n+1}^\infty A_i) \\ &\to \lim_n \mu(\bigsqcup_{i=1}^n A_i) + 0 && n \to \infty \\ &= \lim_n \sum_{i=1}^n \mu(A_i) \\ &= \sum_{i=1}^\infty \mu(A_i)\end{aligned}$$

2.5 Deterministic & Null Sets

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

Definition 2.5.1 (Null Set)
$M \in \mathcal{F}$ is a *null set*if $\mathbb{P}(M) = 0$.

Definition 2.5.2 (Negligible Set)
$N \subseteq \Omega$ is *negligible* if there is some null set $M \in \mathcal{F}$ such that $N \subseteq M$.

Definition 2.5.3 (Complete Space)
$(\Omega, \mathcal{F}, \mathbb{P})$ is *complete* if any negligible set is also a member of $\mathcal{F}$.

Definition 2.5.4 (Almost Surely)
$A \in \mathcal{F}$ occurs *almost surely* (a.s.) OR with probability 1 if $\mathbb{P}(A) = 1$.

Proposition 2.5.1

1) A countable union of null sets is null
2) A countable union of negligible sets is negligible
3) A countable intersection of a.s. events happens a.s.

Lemma 2.5.2 (Borel-Cantelli)
Let $A_1, A_2, dots \in \mathcal{F}$ be events and

$$A := \{\omega \in \Omega : \omega \text{ is contained in infinitely many } A_i\text{'s}\}$$

(a) If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A) = 0$

(b) If $\sum_n \mathbb{P}(A_n) = \infty$ and the A_n's are independent, then $\mathbb{P}(A) = 1$.

2.6 Random Elements (Measurable Functions)

Definition 2.6.1 (Measurable Function)
Let $(F, \mathcal{F}), (E, \mathcal{E})$ be measurable spaces.
$X : F \to E$ is *measurable* if for every $B \in \mathcal{E}$, $X^{-1}(B) \in \mathcal{F}$.

Definition 2.6.2 (Random Element)
$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ is a *random element*.

Example 2.6.1
$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is a *random variable*.

$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N})$ is a *random vector*.

$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^T, \sigma_{\text{cylinder}})$ is a *random process*. Here we take $T = \mathbb{N}$ or $T = [0, T]$.

Let $X : (\Omega, \mathcal{F}, \mathbb{R}) \to (E, \mathcal{E})$ be a random element. Recall the *pushforward measure on* $(E, \mathcal{E})$ */ distribution of* X given by $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$.

Definition 2.6.3
The *pullback σ-algebra* is given by

$$\sigma_X = X^{-1}(\mathcal{E}).$$

Note that $\sigma_X \subseteq \mathcal{F}$ and that different pullback σ-algebras can occur with different random elements.

Example 2.6.2
Suppose $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (E, \mathcal{E}, \mathbb{P}_X) \xrightarrow{f} (\mathbb{R}, \mathcal{B})$. Then $f \circ X : \Omega \to \mathbb{R}$ is a *statistic*.

For instance, the mean or variance are all statistics.

2.6.1 Measures on $\mathbb{R}$

Recall the classical proability spaces $(\Omega, 2^\Omega, \mathbb{P})$ where Ω is finite and $\mathbb{P}$ comes from the $p(\omega_i)$'s. This is the discrete measure.

We also have the discrete measure on $(\mathbb{R}, \mathcal{B})$ where we assign some real numbers $a_1, a_2, \ldots$ with a probability $p(a_i)$ such that $\sum_i p(a_i) = 1$. Then for $A \in \mathcal{B}$, $\mathbb{P}(A) = \sum_{a_i \in A} p(a_i)$.

We wish to enrich our zoo of measures.

Theorem 2.6.3 (Carathéodory)
Let $\mathscr{A}$ be an algebra and $\mathcal{F} = \sigma(\mathscr{A})$. Suppose $\mathbb{P}_{\mathscr{A}} : \mathscr{A} \to \mathbb{R}$ is nonnegative, normalized, and σ-additive.
Then there exists a unique probability measure $\mathbb{P}_{\mathcal{F}} : (\Omega, \mathcal{F}) \to \mathbb{R}$ which is an extension of $\mathbb{P}_{\mathscr{A}}$.

2.7 Borel & Lebesgue Measure on $\mathbb{R}$

Let $\mathscr{A}$ be the algebra on $(0, 1]$ be generated by intervals of the form $(a, b]$ for $0 \le a \le b \le 1$. Recall $\mathcal{B}_{(0,1]} = \sigma(\mathscr{A})$.

Now for $A = \bigsqcup_{i=1}^n (a_i, b_i]$, we define

$$\lambda(A) = \sum_{i=1}^n (b_i - a_i).$$

Note that $\lambda \ge 0$ and $\lambda((0, 1]) = 1$. We also claim that λ is σ-additive. If we show this then λ extends uniquely to a probability measure on $((0, 1], \mathcal{B}_{(0,1]})$ called the *Borel Measure*.

Proposition 2.7.1
λ is σ-additive.

Proof
It suffices to show continuity at $\varnothing$.

Suppose that $A_1 \supseteq A_2 \supseteq \ldots \downarrow \varnothing$. By definition, each A_i is a finite disjoint union of intervals. But then the maximum and minimum endpoints exist and their differences must converge to 0. But this is an upperbound on $\lambda(A_i)$ so that $\lambda(A_i) \to 0$ as desired.

Note that the *Lebesgue measure* coincides with the Borel measure on $\mathcal{B}_{(0,1]}$. However, the number of measurable sets under the Borel measure has cardinality $2^{\mathbb{N}}$ while the number of measurable sets under the Lebesgue measure has cardinality $2^{\mathbb{R}}$. In fact, the Lebesgue measure $(\mathbb{R}, \mathcal{L}_{\mathbb{R}}, \lambda)$ is the completion of the Borel measure $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$ and contains a richer set of null events.

2.7.1 More Measures

We now include the Borel/Lebesgue measure in our zoo. We can also construct measures by combining them. For instance, in the universe $(0, 1]$, we assign $(0, 1/2]$ the Borel measure and some discrete measure on $(1/2, 1]$.

2.8 Lebesgue Integration

2.8.1 Construction of Lebesgue Integration

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

Indicator Functions

Let $S \in \mathcal{F}$. We write $\mathbb{1}_S : \Omega \to \mathbb{R}$ denote the indicator function of S. Then we define

$$\int_\Omega \mathbb{1}_S(\omega) d\mu(\omega) := \mu(S).$$

Simple Functions

Recall a simple function is a linear combination of indicator functions. We define

$$\int_\Omega \sum_{k=1}^n a_k \mathbb{1}_{S_k}(\omega) d\mu(\omega) := \sum_{k=1}^n a_k \mu(S_k).$$

Note the rigorous construction requires careful treatment of negative coefficients.

Non-negative Functions

Let $f : \Omega \to \mathbb{R}_+$ be measurable. We define

$$\int_\Omega f(\omega) d\mu(\omega) := \sup\left\{ \int_\Omega s(\omega) d\mu(\omega) : s \text{ simple} \wedge 0 \leq s \leq f \right\}.$$

Measurable Functions

Finally, consider $f : \Omega \to \mathbb{R}$ measurable. we decompose $f = f^+ - f^-$ and define

$$\int_\Omega f(\omega) d\mu(\omega) := \int_\Omega f^+(\omega) d\mu(\omega) - \int_\Omega f^-(\omega) d\mu(\omega).$$

Note that we ask at least one of the integrals be finite.

2.8.2 Properties

Recall an L^p-space for a Lebesgue measurable space $(\Omega, \mathcal{F}, \mu)$ is given by

$$L^p(\Omega, \mathcal{F}, \mu) := \left\{ f : \Omega \to \mathbb{R} : \int_\Omega |f(\omega)|^p d\mu(\omega) < \infty \right\}.$$

1) $\int f d\mu = \int g d\mu$ implies that $f = g$ a.e. in μ
2) $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$ (linearity)
3) $f \le g$ implies that $\int f d\mu \le \int g d\mu$ (monotonicity)
4) Suppose $f_{n+1} \ge f_n$, $f_n \to f$ a.e., and $f, f_n \ge \eta$ satisfying $\int \eta d\mu > -\infty$, then $\int f_n d\mu \to \int f d\mu$ (monotone convergence theorem)
5) Supppose $f_n \to f$ a.e. and there is some $g \in L^p(\Omega, \mathcal{F}, \mu)$ with $|f_n|, |f| \le g$. Then $\int f_n d\mu \to \int f d\mu$ and $f \in L^p(\Omega, \mathcal{F}, \mu)$ (dominated convergence theorem)

2.8.3 Examples

Consider $\Omega = \{0, 1\}$ with the measure $p(0) = \frac{1}{3}, p(1) = \frac{2}{3}$. Consider $f : 0 \mapsto 1, 1 \mapsto 10$. Then

$$\int_\Omega f(\omega) d\mu(\omega) = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 10.$$

2.9 Absolute Continuity

Let $p : \mathbb{R} \to \mathbb{R}_+$ be $\mathcal{B}_\mathbb{R}$-measurable such that

$$\int_\mathbb{R} p(x) d\lambda(x) = 1.$$

We can define a probability measure on $\mathbb{R}$ where for $A \in \mathcal{B}$,

$$\mathbb{P}_p(A) = \int_\mathbb{R} p(x) \mathbb{1}_A(x) d\lambda(x).$$

Proposition 2.9.1
$\mathbb{P}_p$ is a probability measure.

Proof
Non-negativity and normalization is easy to show. For σ-additivity, let $A_i \in \mathcal{B}_\mathbb{R}$ for $i \in \mathbb{N}$.

Then

$$
\begin{aligned}
\mathbb{P}_p(\bigsqcup_i A_i) &:= \int_{\bigsqcup_i A_i} p(x) d\lambda(x) \\
&= \int_{\mathbb{R}} p(x) \mathbb{1}_{\bigsqcup_i A_i} d\lambda(x) \\
&= \int \lim_n p(x) \sum_{i=1}^{n} \mathbb{1}_{A_i} d\lambda(x) \\
&= \lim_n \int p(x) \sum_{i=1}^{n} \mathbb{1}_{A_i} d\lambda(x) && \text{monotone convergence} \\
&= \sum_{i=1}^{\infty} \int_{\mathbb{R}} p(x) \mathbb{1}_{A_i}(x) d\lambda(x) \\
&= \sum_{i=1}^{\infty} \mathbb{P}_p(A_i).
\end{aligned}
$$

We can generalize this idea.

Definition 2.9.1 (Absolutely Continuous)
Let μ, ν be measures on some measurable space $(\Omega, \mathcal{F})$. μ is *absolutely continuous with respect to* ν if $\nu(A) = 0$ implies that $\mu(A) = 0$ and write $\mu << \nu$.

Recall that a σ-finite measure on $(\Omega, \mathcal{F})$ means that Ω is a countable union of finite measure subsets.

Theorem 2.9.2 (Radon-Nikodym)
Let $\mu << \nu$ be σ-finite measures on $(\Omega, \mathcal{F})$. There exists a unique measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}_+, \mathcal{B})$ such that for all $A \in \mathbb{R}$,

$$\mu(A) = \int_A f(x) d\nu(x).$$

Note that this theorem can be reversed.

Definition 2.9.2 (Radon-Nikodym Derivative)
We say that f is the *Radon-Nikodym derivative* and write

$$f(x) = \frac{d\mu}{d\nu}(x).$$

2.10 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be probability spaces.

Definition 2.10.1 (Independent)
$A, B \in \mathcal{F}$ are independent with respect to $\mathbb{P}$ if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Definition 2.10.2 (Mutually Independent)
We say $A_i \in \mathcal{F}$ for $i \in [n]$ are independent if for all sub-indices $I \subseteq [n]$,

$$\mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

Note that this is stronger than pairwise independence!

Definition 2.10.3 (Mutually Independent)
We say $A_t \in \mathcal{F}$ for $t \in T$ are independent if for all finite sub-indices $I \subseteq T$, $\{A_i : i \in I\}$ is independent.

Definition 2.10.4 (Independent Variables)
Consider $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (E_i, \mathbb{E}_i, \mathbb{P}_{X_i})$ for $i \in [n]$. The X_i's are mutually independent with respect to $\mathbb{P}$ if for all $B_i \in \mathbb{E}_i$, $\{X^{-1}(B_i) : i \in [n]\}$ is mutually independent.

Definition 2.10.5 (Independent σ-Algebra)
Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be σ-algebras on Ω. We say there are independent if any $\{A_i \in \mathcal{F}_i : i \in [n]\}$ are independent.

Note that random variables are independent if and only if their pullback σ-algebras are independent.

2.11 Direct Product of (Finite) Measure Spaces

Suppose $(\Omega_i, \mathcal{F}_i, \mu_i)$ are finite measure spaces for $i \in [n]$. We construct the direct product space as follows:

1) $\Omega = \times_{i=1}^{n} \Omega_i$

2) Let $\mathscr{A} := \{\times_{i=1}^n A_i : A_i \in \Omega_i\}$ be an algebra and take $\mathcal{F} := \sigma(\mathscr{A})$.
3) Define μ on $\mathscr{A}$ as $\mu(\times_{i=1}^n A_i) = \prod_{i=1}^n \mu(A_i)$. Then we extend using Carathéodory's theorem.

Proposition 2.11.1
$\mu : \mathscr{A} \to \mathbb{R}$ is nonnegative, normalized, and σ-additive.

This justifies the use of Carathéodory's theorem.

Proof
Nonnegativity and normalization is easy. We show σ-additivity by showing continuity at $\varnothing$.

Suppose some $\times_{i=1}^n A_i \downarrow \varnothing$. It must be that some $A_i \downarrow \varnothing$. But then $\mu(A_i) \to 0$ and by finiteness, $\mu(\times_{i=1}^n A_i) \to 0$.

Definition 2.11.1 (Direct Product of Measure Spaces)
$(\Omega, \mathcal{F}, \mu)$ from above.

Proposition 2.11.2
The coordinate projections from the direct product are independent random elements.

The proof will involve showing that they are random elements and then showing independence.

Example 2.11.3
An example of a direct product is Bernoulli trials.

Part II

Random Variables & Vectors

Chapter 3

Random Variables & Vectors

3.1 Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Recall a random variable is a measurable function

$$\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}).$$

We consider random variables equivalent if they agree a.e., or $\mathbb{P}\{\omega : \xi(\omega) \neq \eta(\omega)\} = 0$.

Recall $\mathbb{P}_\xi$ is a probability measure on $(\mathbb{R}, \mathcal{B})$ induced by ξ with $\mathbb{P}_\xi(A) := \mathbb{P}(\xi^{-1}(A))$.

Proposition 3.1.1

1) Let $g : \mathbb{R} \to \mathbb{R}$ be measurable. Then $g \circ \xi : \Omega \to \mathbb{R}$ is a random variable.
2) If $\xi_i, i \in \mathbb{N}$ are random variables, then
   a) $\max(\xi_1, \xi_2), \min(\xi_1, \xi_2)$ are random variables.
   b) $\sup_n \xi_n, \inf_n \xi_n$ are random variables.
   c) If $\xi_n \to \xi$ pointwise, then ξ is a random variable.
3) Let ξ, η be random variables. Then so are the following:
   a) $\xi + \eta$
   b) $a\xi + b\eta$
   c) $\xi \cdot \eta$

Proof (sketch)
2) We have

$$(\max(\xi_1, \xi_2))^{-1}(-\infty, x] = \xi_i^{-1}(-\infty, x] \cap \xi_2^{-2}(-\infty, x].$$

3) We have

$$(\xi+\eta)^{-1}(-\infty,x] = \bigcup_{q\in\mathbb{Q}}\{\omega : \xi(\omega) < q, \eta(\omega) < x - q\}.$$

3.2 Distribution Function

Definition 3.2.1 (Distribution Function)
Let $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$. Define the *distribution function* $F_\xi : \mathbb{R} \to \mathbb{R}$ be

$$F_\xi(x) := \mathbb{P}_\xi(-\infty, x].$$

Proposition 3.2.1

1) $F_\xi(b) - F_\xi(a) = \xi^{-1}(a, b]$
2) F_ξ is non-decreasing
3) $F_\xi(-\infty) := \lim_{x\to-\infty} F(x) = 0$ and $F_\xi(\infty) := \lim_{x\to\infty} F(x) = 1$
4) F_ξ is right-continuous
5) F_ξ has left limits but is not necessarily left continuous

Proof (sketch)
3) Continuity of measure.

Definition 3.2.2 (CADLAG)
The CADLAG function class $\mathcal{D}$ is right continuous and has left limits but is not necessarily left-continuous.

Theorem 3.2.2

1) The class of all distribution functions coincides with CADLAG functions which are non-decreasing and whose limits at $\pm\infty$ are $0, 1$, respectively.
2) Each distribution $\mathbb{P}_\xi$ and distribution function F_ξ determine each other uniquely.

Proof (sketch)
2) Given $F_\xi(x)$, define $\mathbb{P}_\xi(a, b] := F_\xi(b) - F_\xi(a)$. Then apply Carathéodory's theorem to extend to $\sigma\{(a, b] : a \leq b]\}$.

3.3 Discrete Random Variables (and Distributions)

Definition 3.3.1 (Discrete Random Variable)
A random variable $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ is discrete if there is some $A = \{a_i : i \in \mathbb{N}\} \subseteq \mathbb{R}$ countable such that $\mathbb{P}_\xi(A) = 1$.

We can write $\mathbb{P}_\xi(B) = \sum_i p_i \mathbb{1}_{a_i \in B}$.

3.3.1 Delta Measure

$\xi(\Omega) = 1$ so

$$\mathbb{P}_\xi(0) = 1.$$

3.3.2 Bernoulli Variable

We write $\mathrm{Be}(p)$ to denote

$$\mathbb{P}_\xi(0) = 1 - p, \mathbb{P}_\xi(1) = p.$$

Thus $\mathbb{P}_\xi(A) = (1-p)\mathbb{1}_{0 \in A} + p\mathbb{1}_{1 \in A}$.

3.3.3 Binomial

We write $\mathrm{Bin}(N, p)$ to denote

$$\mathbb{P}_\xi(k) := \binom{N}{l} p^k (1-p)^{N-k}.$$

3.3.4 Geometric

We write $\mathrm{Geo}(p)$ to denote

$$\mathbb{P}_x i(k) = p(1-p)^k.$$

3.3.5 Poisson

We write $\mathrm{Po}(\lambda)$ to denote

$$\mathbb{P}_\xi(k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

This is used to model events happening independently in some interval with fixed rate.

Theorem 3.3.1 (Poisson (informal))
As $N \to \infty$ and $p_N \to 0$

$$|\mathbb{P}\{\mathrm{Bin}(N, p_N) = k\} - \mathbb{P}\{\mathrm{Po}(Np_N) = k\}| \to 0.$$

3.4 Absolutely Continuous Distributions

Recall that if we have a density $p : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}_+, \mathcal{B})$ measurable such that $\int_{\mathbb{R}} p(x) d\lambda(x) = 1$, we can define a distribution

$$\mathbb{P}_\xi(A) := \int_A p(x) d\lambda(x).$$

3.4.1 Uniform

We write $\xi \sim U[a, b]$ to denote that it comes from the density given by $p(x) = \frac{1}{b-1} \mathbb{1}_{[a,b]}$.

3.4.2 Normal / Gaussian

We write $\xi \sim N(a, \sigma^2)$ to denote that it came from the density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

3.4.3 Exponential

We write $\xi \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$ to denote that it came from the density

$$p(x) = \lambda e^{-\lambda x}.$$

Recall that the exponential distribution is *memoryless*

$$\mathbb{P}(\xi(\omega) > x + y : \xi(\omega) > y) = \mathbb{P}\{\xi(\omega) > x\}.$$

3.5 Continuous Singular Distributions

Definition 3.5.1 (Continuous Singular)
A random variable ξ is *continuous singular* if the corresponding distribution function F_ξ is continous with $F'_\xi(x) = 0$ a.e. with respect to λ.

An example is the Cantor distribution.

Theorem 3.5.1

1) Any distribution function is a convex combination of discrete, absolutely continuous, and continous singular distributions functions.
2) The corresponding distribution measure is also a convex combination of discrete, absolutely continuous, and continuous singular measures.

3.6 Random Vectors

Recall that a random vector is a measurable function

$$\bar{\xi} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)).$$

Note that we can think of $\mathcal{B}(\mathbb{R}^d) = \sigma\{\times_{i=1}^d (a_i, b_i]\}$ or $\sigma\{\times_{i=1}^d B_i : B_i \in \mathcal{B}(\mathbb{R})\}$. Alternatively, it is a fact that $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\times d}$. The proof is not obvious at an initial glance.

Definition 3.6.1 (Distribution)
$\mathbb{P}_{\bar{\xi}}$ is a distribution on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given by

$$\mathbb{P}_{\bar{\xi}}(A) = \mathbb{P}(\bar{\xi}^{-1}(A)).$$

Definition 3.6.2 (Distribution Function)
$F_{\bar{\xi}}(x)$ given by

$$\bar{x} \mapsto \mathbb{P}_{\bar{\xi}}\{\times_{i=1}^d (-\infty, x_i]\}.$$

Note that we can show there is a bijection between distributions and distribution functions, analogously to the 1-dimensional case.

3.7 Discrete Random Vectors

3.7.1 Discrete Random Vectors

Delta Measure

$\bar{\xi}(\Omega) = x$ for some $x \in \mathbb{R}^d$.

Multinomial Distribution

We write $\bar{\xi} \sim \text{Mult}(p_1, \ldots, p_d)$ to denote

$$\mathbb{P}(\bar{\xi} = x) = \frac{N!}{x_i! \ldots x_d!} p_1^{x_1} \ldots p_d^{x_d}$$

where $\sum_i p_i = 1$.

Random Walk

We can let $\bar{\xi}$ be the position of a lattice random walk in $\mathbb{R}^d$.

3.7.2 Absolutely Continuous Distributions in $\mathbb{R}^d$

Recall absolutely continuous distribution have an associate density $p : \mathbb{R}^d \to \mathbb{R}_+$ where

$$\mathbb{P}_{\bar{\xi}}(A) = \int_A p(x) d\lambda(x).$$

Note that the distribution function is simply

$$F_{\bar{\xi}}(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} p(x) d\lambda(x).$$

Uniform Distribution

We write $\bar{\xi} \sim U(D)$ for some $Dd \in \mathcal{B}(\mathbb{R}^d)$ to denote

$$\mathbb{P}_{\bar{\xi}}(x) = \frac{1}{\lambda(D)} \mathbb{1}_D.$$

Gaussian Distribution

We write $\bar{\xi} \sim N(A, \Sigma)$ for some PSD $\Sigma \in \mathbb{R}^{d\times d}$ to denote

$$\mathbb{P}_{\bar{\xi}}(x) = \frac{1}{(\sqrt{2\pi})^2 |\det \Sigma|} \exp\left[-\frac{1}{2}(x-\bar{A})^T\Sigma^{-1}(x-\bar{A})\right].$$

3.8 Joint & Marginal Distributions

Let $\bar{\xi}(\omega) = (\xi_1(\omega), \dots, \xi_d(\omega))$ be a random vector. Then the i-th projection $\xi_i(\omega) = \pi_i(\bar{\xi}(\omega))$ is a random variable.

Definition 3.8.1 (Joint Distribution)
We say $\mathbb{P}_{\bar{\xi}}$ is the *joint distribution* of the ξ_i's.

Definition 3.8.2 (Marginal Distribution)
We say $\mathbb{P}_{\bar{\eta}}$ where $\bar{\eta}$ is any subvector of ξ, is a *marginal distribution* of $\bar{\xi}$.

Note that marginals are uniquely defined given the joint, but the converse is not necessarily true!

Example 3.8.1
Let $\xi_1, \xi_2 \sim U[0,1]i$ and $\bar{\xi} = (\xi_1, \xi_2)$.

If ξ_1, ξ_2 are independent, then $\bar{\xi} = \lambda^{[0,1]} \times \lambda^{[0,1]}$.

Now suppose $\xi_1 = \xi_2$. Then $\bar{\xi} = (\xi_1, \xi_1)$ is different!

Thus we cannot reconstruct a random vector from marginals unless they are independent.

Recall that $\xi_1, \dots, \xi_d$ are mutually independent if $\{\xi_i^{-1}(A_i) : i \in [d]\}$ is independent as events for all $A_i \in \mathcal{B}(\mathbb{R})$.

Proposition 3.8.2
1) $\bar{\xi} = (\xi_1, \dots, \xi_d)$ is uniquely defined given
   a) $\mathbb{P}_{\xi_i}$ for $i \in [d]$
   b) ξ_i's are mutually independent
2) $\mathbb{P}_{\bar{\xi}} = \times_{i=1}^d \mathbb{P}_{\xi_i}$
3) $F_{\bar{\xi}}(x) := \mathbb{P}_{\bar{\xi}}(I_{-\infty}(x)) = \prod_{i=1}^d \xi_i(x_i)$

Proof
It suffices to show that

$$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_{\bar{\xi}}) = \prod_{i=1}^d (\mathbb{R}, \mathcal{B}, \mathbb{P}_{\xi_i}).$$

Moreover, by Carathéodory's extension theorem, it suffices to show the distributions coincide on boxes.

$$\begin{aligned}
\mathbb{P}_{\bar{\xi}}(\times_{i=1}^d I_i) &:= \mathbb{P}(\bar{\xi}^{-1}(\times_{i=1}^d I_i)) \\
&= \mathbb{P}(\bigcap_{i=1}^d \xi_i^{-1}(I_i)) \\
&= \prod_{i=1}^d \mathbb{P}(\xi_i^{-1}(I_i)) && \text{independence} \\
&=: \prod_{i=1}^d \mathbb{P}_{\xi_i}(I_i).
\end{aligned}$$

Let ξ_i's be independent, absolutely continouos distributions with densities $p_i(x)$. Remark that by Fubini's theorem

$$\begin{aligned}
F_\xi(x_i) &= \int_{-\infty}^{x_i} p_i(t)dt \\
F_{\bar{\xi}}(x) &= \prod_{i=1}^d F_{\xi_i}(x_i) \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} p_1(x_1)\dots p_d(x_d)dx_d \dots dx_1.
\end{aligned}$$

Also, if the ξ_i's are discrete and we let $\bar{\xi} = (\xi_i)$, then $\bar{\xi}$ take values in some countable set A. We know $\mathbb{P}_{\bar{\xi}}$ if and only if we know all $\mathbb{P}\{\xi_i = a_i : i \in [d]\}$ for all $a \in A$.

3.9 Moments of Random Variables & Random Vectors

Let $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P}_\xi)$.

3.9.1 Expectation

Definition 3.9.1 (Expectation)
The *expectation* of ξ is

$$\mathbb{E}\xi := \int_\Omega \xi(\omega) d\mathbb{P}(\omega).$$

Note that we require either of $\mathbb{E}\xi^+, \mathbb{E}\xi^-$ to be finite.

We say that the expectation si finite if $\xi \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, that is, $\int |\xi| d\mathbb{P} < \infty$.

Proposition 3.9.1

1) If $\xi(\omega)$ is bounded, then $\mathbb{E}\xi$ exists.
2) If $\xi = a$ is constant, then $\mathbb{E}\xi = a$
3) If ξ is discrete, taking on values a_i with probability p_i, then

$$\mathbb{E}\xi = \sum_i \int_{\xi^{-1}(a_i)} \xi(\omega) d\mathbb{P}(\omega) = \sum_i a_i p_i.$$

Properties

prop $\mathbb{E}(a\xi + b\eta) = a\mathbb{E}\xi + b\mathbb{E}\eta$

prop $\xi \geq \eta$ implies that $\mathbb{E}\xi \geq \mathbb{E}\eta$

prop $|\mathbb{E}\xi| \leq \mathbb{E}|\xi|$

Recall that proofs about Lebesgue integrals always follows the same path: start from indicator functions, proceed to simple functions, then considering approximations. Let $A_{k,n} := g^{-1}\left[\frac{k}{n}, \frac{k+1}{n}\right)$ for $k \in \mathbb{Z}$. Then a simple approximation is

$$f_n(\omega) = \sum_k \frac{k}{n} \mathbb{1}_{A_{k,n}}.$$

Corollary 3.9.1.1
Let $\xi_i(\omega), k \in \mathbb{N}$ be non-negative. By the monotone convergence theorem,

$$\mathbb{E}\left(\sum_{i=1}^\infty \xi_k\right) = \sum_{k=1}^\infty \mathbb{E}\xi_k.$$

Theorem 3.9.2 (Change of Variables)
Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E}, \mathbb{P}_X)$ be a random element, and $g : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B})$ be measurable. For any $A \in \mathcal{E}$,

$$\int_A g(e) d\mathbb{P}_X(e) = \int_{X^{-1}(A)} g(X(\omega)) d\mathbb{P}(\omega).$$

Corollary 3.9.2.1

1) $\int_A g(x) d\mathbb{P}_\xi(x) = \int_{\xi^{-1}(A)} g(\xi(\omega)) d\mathbb{P}(\omega)$
2) $\mathbb{E} g(\xi) = \int_{\mathbb{R}} g(x) d\mathbb{P}_\xi(x)$
3) If ξ is and absolutely continuous distribution with denity $p(x) = \frac{d\mathbb{P}_\xi(x)}{d\lambda(x)}$, then

$$\mathbb{E} f(\xi) \underbrace{=}_{2)} \int_{\mathbb{R}} f(x) d\mathbb{P}_\xi(x) = \int_{\mathbb{R}} f(x) p(x) dx.$$

(needs proof)

Note that for discrete distributions, we can directly compute $\mathbb{E}\xi = \sum_i a_i p_i$ witht the LHS. On the other hand, for absolutely continuous distributions, we use the RHS $\mathbb{E}\xi = \int_{\mathbb{R}} x p(x) dx$.

Proposition 3.9.3
Let ξ, η be independent random variables with finite expectation. Then $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$.

Proof
Let the joint be $\bar{\xi} = (\xi, \eta)$, let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $(x, y) \mapsto xy$. Then

$$\begin{aligned}
\mathbb{E}(\xi\eta) &= \mathbb{E} f(\xi, \eta) \\
&= \int_{\mathbb{R}^2} x_1 x_2 d\mathbb{P}_{\bar{\eta}}(x_1, x_2) \\
&= \int_{\mathbb{R}^2} x_1 x_2 d\mathbb{P}_\xi(x_1) \mathbb{P}_\eta(x_2) && \text{independence} \\
&= \int_{\mathbb{R}} x_1 d\mathbb{P}_\xi(x_1) \cdot \int_{\mathbb{R}} x_2 d\mathbb{P}_\eta(x_2) && \text{Fubini}
\end{aligned}$$

Examples

For $\xi \sim \text{Be}(p)$,

$$\mathbb{E}\xi = 0(1-p) + 1(p) = p.$$

For $\xi_N \sim \text{Bin}(N, p)$,

$$\mathbb{E}\xi_N = \sum_{k=1}^{N} k \binom{N}{k} p^k (1-p)^{N-k}.$$

Instead of computing this, we can instead realize it is the sum of IID bernoulli variables so

$$\mathbb{E}\left(\sum_{i=1}^{N} \xi^{(i)}\right) = Np.$$

For $\xi \sim U[a, b]$,

$$\mathbb{E}\xi = \int_{\mathbb{R}} x \frac{1}{b-a} \mathbb{1}_{[a,b]} = \int_a^b \frac{x}{b-a} = \frac{a+b}{2}.$$

3.9.2 Median

This is not a true moment. However, it is some measure of centrality.

Definition 3.9.2 (Median)
A *median* M_ξ is any $x \in \mathbb{R}$ such that

$$\mathbb{P}_\xi(-\infty, x] \geq \frac{1}{2}$$
$$\mathbb{P}_\xi[x, \infty) \geq \frac{1}{2}$$

for symmetric distributions, the median is the mean but not always! It is a fact that $|M_\xi - \mathbb{E}\xi| \leq \sqrt{\text{Var}\,\xi}$.

3.9.3 Variance

Definition 3.9.3 (Variance)
Let ξ be a random variable with $|\mathbb{E}\xi| < \infty$. Then its variance is

$$\begin{aligned}\text{Var}\,\xi &:= \mathbb{E}(\xi - \mathbb{E}\xi)^2 \\ &= \int_\omega (\xi(\omega) - \mathbb{E}\xi)^2 d\mathbb{P}(\omega) \\ &= \mathbb{E}\xi^2 - 2\mathbb{E}\xi \cdot \mathbb{E}\xi + (\mathbb{E}\xi)^2 \\ &= \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2.\end{aligned}$$

Proposition 3.9.4

1) The variance is non-negative with equality if and only if the variable is constant a.e.
2) $\operatorname{Var}(\xi + a) = \operatorname{Var}\xi$
3) $\operatorname{Var}(a\xi) = a^2 \operatorname{Var}\xi$

Note that the variance of a discrete variable can be computed as

$$\sum_{i=1}^{\infty} (a_i - \mathbb{E}\xi)^2 p_i.$$

which the absolutely continuous ones can be computed as

$$\int_{\mathbb{R}} (x - \mathbb{E}\xi)^2 p(x) dx.$$

Examples

For $\xi \sim \operatorname{Be}(p)$,

$$\operatorname{Var}\xi = p - p^2.$$

For $\xi \sim \operatorname{Bin}(N, p)$,

$$\operatorname{Var}\xi = N(p - p^2)$$

where we use the fact that independent variances are summable.

For $\xi \sim U[a, b]$,

$$\operatorname{Var}\xi = \int_{\mathbb{R}} \left(x - \frac{a+b}{2}\right)^2 \frac{\mathbb{1}_{[a,b]}}{b-a} dx.$$

3.9.4 Covariance

Now, this is not a moment but it is a useful notion concerning the variance.

Definition 3.9.4 (Covariance)
Let ξ, η be random variables with finite expectation. Then

$$\begin{aligned}\operatorname{Cov}(\xi, \eta) &:= \mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta) \\ &= \int_{\mathbb{R}^2} (x_1 - \mathbb{E}\xi)(x_2 - \mathbb{E}\eta) dP_{(\xi,\eta)}(x_1, x_2).\end{aligned}$$

Note that if ξ, η are independent, we can then factorize the joint as a product and apply Fubini's theorem to realize that $\operatorname{Cov}(\xi, \eta) = 0$.

Now, by Hölder's inequality,

$$|\operatorname{Cov}(\xi, \eta)| \leq \sqrt{\operatorname{Var}\xi \cdot \operatorname{Var}\eta}.$$

Definition 3.9.5 (Correlation Coefficient)
We define

$$\rho(\xi, \eta) = \frac{\operatorname{Cov}(\xi, \eta)}{\sqrt{\operatorname{Var}\xi \cdot \operatorname{Var}\eta}} \in [-1, 1].$$

We say ξ, η are *uncorrelated* if they have correlation 0.

Proposition 3.9.5
We have

$$\operatorname{Var}(\xi + \eta) = \operatorname{Var}\xi + \operatorname{Var}\eta + 2\operatorname{Cov}(\xi, \eta).$$

Corollary 3.9.5.1

1) If $\xi_i, i \in [n]$ are pairwise uncorrelated, then $\operatorname{Var}(\sum_i \xi_i) = \sum_i \operatorname{Var}\xi_i$.
2) If ξ_i are independent, then their variances are additive.

3.9.5 Higher Order Moments

Definition 3.9.6 (Absolute Moment of Order p)
We define

$$\mathbb{E}|\xi|^p = \int_\Omega |\xi(\omega)|^p d\mathbb{P}(\omega).$$

Note that the p-th absolute exists if and onky if $\xi \in L^p(\Omega, \mathcal{F}, \mathbb{P})$.

Similarly, we have

Definition 3.9.7 (Moment of Order p)
$\mathbb{E}\xi^p$.

Definition 3.9.8 (Absolute Central Moment of Order p)
$\mathbb{E}|\xi - \mathbb{E}\xi|^p$.

Definition 3.9.9 (Central Moment of Order p)
$\mathbb{E}(\xi - \mathbb{E}\xi)^p$.

Theorem 3.9.6 (Jensen's Inequality)
For all $q \leq p$,

$$(\mathbb{E}|\xi|^q)^{\frac{1}{q}} \leq (\mathbb{E}|\xi|^p)^{\frac{1}{p}} .$$

Theorem 3.9.7 (Hölder's Inequality)
For $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}|\xi\eta| \leq (\mathbb{E}|\xi|^p)^{\frac{1}{p}} (\mathbb{E}|\xi|^q)^{\frac{1}{q}} .$$

3.9.6 The Moment Problem

Problem 1 (Moment Problem)
Let $a_1, a_2, \cdots \in \mathbb{R}$.

(1) When does this form a sequence of moments for some random variable?

(2) If so, does this sequence uniquely determine a distribution?

The answer to (1) is the following theorem, while the answer to (2) is not fully fleshed out.

Theorem 3.9.8 (Hamburger)
$a_n = \mathbb{E}\xi^n, n \in \mathbb{N}$ for some $\xi, \mathbb{P}_\xi$ if and only if the matrix $A := [a_{i+j}]_{i,j} \succeq 0$.

Note that

$$\sum_{i,j=1}^{n} \mathbb{E}\xi^i \xi^j x_i x_j = \mathbb{E}\left(\sum_{i=1}^{n} \xi^i x_i\right)^2 \geq 0$$

so one directly is needed.

3.9.7 Moment Inequalities

Lemma 3.9.9 (Chebyshev)
Let ξ be a non-negative ranodm variable. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be non-decreasing. Then for all $\epsilon > 0$,

$$\mathbb{P}\{\xi \geq \epsilon\} \leq \frac{\mathbb{E}\varphi(\xi)}{\varphi(\epsilon)}.$$

Proof
Let

$$A_\epsilon := \{\omega : \xi(\omega) \geq \epsilon\}$$

and note that it is a subset of

$$B_\epsilon := \{\omega : \varphi(\xi(\omega)) \geq \varphi(\epsilon)\}.$$

We have

$$\begin{aligned}\mathbb{P}\{\xi \geq \epsilon\} &= \int_{A_\epsilon} d\mathbb{P}(\omega) \\ &\leq \int_{B_\epsilon} d\mathbb{P}(\omega) \\ &\leq \int_{B_\epsilon} \frac{\varphi(\xi(\omega))}{\varphi(\epsilon)} d\mathbb{P}(\omega) \\ &= \frac{1}{\varphi(\epsilon)} \mathbb{E}\varphi(\epsilon).\end{aligned}$$

Corollary 3.9.9.1
1) $\mathbb{P}\{|\eta| \geq \epsilon\} \leq \frac{\mathbb{E}\varphi(|\eta|)}{\varphi(\epsilon)}$
2) $\mathbb{P}\{|\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbb{E}\xi_i| \geq \epsilon\} \leq \frac{\mathrm{Var}\sum_{i=1}^n \xi_i}{\epsilon^2}$
3) $\mathbb{P}\left\{\frac{|\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbb{E}\xi_i|}{n} \geq \epsilon\right\} \leq \frac{\mathrm{Var}\left(\sum_{i=1}^n \xi_i\right)}{n^2\epsilon^2}$

Thus if $\mathrm{Var}\left(\sum_{i=1}^n \xi_i\right) \in o(n^2)$, the LHS of 3) tends to 0 as $n \to \infty$. This is also known as Markov's Law of Large Numbers.

Chapter 4

Characteristic Functions

Characteristic functions will be our main tool to attain limit theorems.

4.1 Characteristic Functions

Let ξ be a random variable.

Definition 4.1.1 (Characteristic Function)
The characteristic function of ξ is given by

$$\varphi_\xi(t) := \mathbb{E}e^{it\xi} = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_\xi(x).$$

Note that we can equivalently express this as

$$\mathbb{E}(\cos t\xi) + i\mathbb{E}(\sin t\xi).$$

If ξ is discrete, we can compute

$$\varphi_\xi(t) = \sum_{k=1}^{\infty} p_k e^{ita_k}$$

If ξ is absolutely continuous,

$$\varphi_\xi(t) = \int_{\mathbb{R}} e^{itx} p(x) dx.$$

Note that this is essentially the inverse Fourier transform!

Proposition 4.1.1

1) $\varphi_\xi(0) = 1$
2) $|\varphi_\xi(t)| = \mathbb{E}|e^{it\xi}| = 1$
3) $\varphi_\xi(t)$ is uniformly continuous in t
4) $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$
5) $\varphi_{a\xi+b}(t) = \mathbb{E}e^{it(a\xi+b)} = e^{ibt}\varphi_\xi(at)$
6) $\forall t \in \mathbb{R}, \varphi_\xi(t) \in \mathbb{R}$ if and only if $\mathbb{P}_\xi$ is symmetric, ie $\mathbb{P}_\xi(A) = \mathbb{P}_\xi(-A)$

Proof

3) We have

$$\begin{aligned} |\varphi_\xi(t+h) - \varphi_\xi(t)| &= |\mathbb{E}e^{it\xi}(e^{ih\xi} - 1)| \\ &\leq \mathbb{E}|e^{ih\xi} - 1| \\ &\to 0 \qquad\qquad h \to 0 \end{aligned}$$

by the dominated convergence theorem.

6) Suppose $\mathbb{P}_\xi$ is symmetic. It must be that $\mathbb{E}(\sin t\xi) = 0$ as sin is an odd function. But then $\operatorname{Im}\mathbb{P}_\xi(t) = 0$ as desired.

Now suppose φ_ξ is real-valued. Then $\varphi_{-\xi}(t) = \overline{\varphi_\xi(t)} = \varphi_\xi(t)$.

Example 4.1.2

a) If ξ follows a δ-distribution, then $\varphi_\xi(t) = e^{ita}$ where $\mathbb{P}\{\xi = a\} = 1$
b) $\xi \sim \operatorname{Be}(p)$, then $\varphi_\xi(t) = (1-p)\cdot 1 + p \cdot e^{it}$
c) $\xi \sim \operatorname{Po}(\lambda)$, then $\varphi_\xi(t) = \sum_{k=0}^\infty e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}e^{\lambda eit}$
d) $\xi \sim \exp(\lambda)$, then $\varphi_\xi(t) = \int_{\mathbb{R}_+} e^{itx}(\lambda e^{-\lambda x})dx = -\frac{\lambda}{it-\lambda}$
e) $\xi \sim N(0,1)$.

$$\begin{aligned} \varphi_\xi(t) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itx}e^{-\frac{x^2}{2}}dx \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int_{\mathbb{R}} e^{-\frac{1}{2}(x-it)^2}dx \\ &= \frac{e - \frac{t^2}{2}}{\sqrt{2\pi}}\int_{\mathbb{R}-it} e^{-\frac{x^2}{2}}dx \\ &= e^{-\frac{t^2}{2}}. \end{aligned}$$

Theorem 4.1.3 (Bochner-Khintchin)
Let $\varphi(t) : \mathbb{R} \to \mathbb{C}$. Then φ is the characteristic function of some probability distribution if and only if

1) $\varphi(0) = 1$
2) $\varphi(t)$ is continuous
3) $\sum_{k,\ell=1}^n \varphi(t_k - t_\ell) z_k \bar{z}_\ell \geq 0$ for all $t_i \in \mathbb{R}, z_i \in \mathbb{C}, i \in [n]$

The forward implication is rather straightforwards. Indeed,

$$0 \leq \mathbb{E}\left|\sum_{k=1}^n e^{it_k\xi} z_k\right|^2 = \sum_{k,\ell=1}^n e^{it_k\xi} e^{-t_\ell \xi} z_k \bar{z}_\ell.$$

Proposition 4.1.4
$E^{-|t|^\alpha}$ is a characteristic function if and only if $\alpha \in (0, 2]$.

Theorem 4.1.5 (Levy)
Let ξ_1, ξ_2 be random variables such that $\varphi_{\xi_1}(t) = \varphi_{\xi_2}(t)$ for all $t \in \mathbb{R}$. Then $\xi \equiv \xi_2, \mathbb{P}_{\xi_1} \equiv \mathbb{P}_{\xi_2}$ as distributions.

Theorem 4.1.6 (Inversion Formula)
Let $\xi, \mathbb{P}_\xi, \varphi_\xi$ be a random variable, its pushforward measure, and characteristic function.

1) $\mathbb{P}_\xi(a, b) + \frac{1}{2}\mathbb{P}_\xi\{a\} + \frac{1}{2}\mathbb{P}_\xi\{b\}$ is equal to

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\xi(t) dt.$$

2) If $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$, then there is some density $p(x)$ of $\mathbb{P}_\xi$ given by

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_\xi(t) dt.$$

Proof (2)
Let $a < b$.

$$\begin{aligned}
\int_a^b p(x)dx &= \frac{1}{2\pi}\int_a^b \int_{\mathbb{R}} e^{-itx}\varphi_\xi(t)dtdx \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_\xi(t)\left(\int_a^b e^{-itx}dx\right)dt && \text{Fubini} \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(t)\frac{e^{-ita}-e^{-itb}}{it}dt \\
&= T_{\to\infty}\frac{1}{2\pi}\int_{-T}^{T} \varphi(t)\frac{e^{-ita}-e^{-itb}}{it}dt \\
&= \mathbb{P}_\xi(a,b) + \frac{1}{2}\mathbb{P}_\xi\{a,b\}.
\end{aligned}$$

Since p follows the requirements of being a density on intervals, it must do so on all Borel sets as desired.

4.2 Characteristic Functions & Moments

Theorem 4.2.1
Let $\xi \in L^k(\Omega)$. Then the j-th derivative

$$\varphi_\xi^{(j)}(t) = \mathbb{E}\{(i\xi)^j e^{it\xi}\} = \int_{\mathbb{R}} (ix)^j e^{itx} d\mathbb{P}_\xi(x)$$

exists for $j \in [k]$.

Proof
We argue by induction.

$$\begin{aligned}
\frac{\varphi^{(j)}(t+h)-\varphi^{(j)}(t)}{h} &= \mathbb{E}\frac{1}{h}\left[(i\xi)^j e^{it\xi}(e^{ih\xi}-1)\right] \\
&\to \mathbb{E}\left\{\lim_{h\to 0}(i\xi)^j\frac{e^{i(t+h)\xi}-e^{it\xi}}{h}\right\} && h \to 0 \\
&= \mathbb{E}\left\{(i\xi)^{j+1}e^{it\xi}\right\}
\end{aligned}$$

Swapping the limit and the integral requires an application of the dominated convergence theorem.

Note that

$$\begin{aligned} |x^{ix} - 1| &= \left| i \int_0^t e_{iy} dy \right| \\ &\leq \left| \int_0^x 1 dy \right| \\ &= |x|. \end{aligned}$$

It follows that

$$\begin{aligned} \left| \frac{1}{h} \left[(i\xi)^j e^{it\xi} (e^{ih\xi} - 1) \right] \right| &\leq \frac{1}{|h|} \cdot 1 \cdot |\xi^{j+1} h| \\ &= |\xi|^{j+1}. \end{aligned}$$

Thus our application of the dominated convergence theorem above is justified.

Corollary 4.2.1.1
We have

1) $\mathbb{E}\xi^j = \frac{\varphi_\xi^{(j)}(0)}{i^j}$
2) $\varphi_\xi(t) = \sum_{i=0}^k \frac{(it)^j}{j!} \mathbb{E}\xi^j + o(t^k)$

Theorem 4.2.2
Suppose that $\varphi_\xi^{(2k)}(0)$ exists and is finite. Then $\mathbb{E}\xi^{2k} < \infty$.

Proof (Sketch)
We consider $k = 1$, as the rest is similar.

Consider Taylor's theorem

$$\varphi(h) = \varphi(0) + h\varphi'(0) + \frac{h^2}{2}\varphi''(0) + o(h^2).$$

On one hand,

$$\begin{aligned} \frac{\varphi(h) + \varphi(-h) - 2\varphi(0)}{h^2} &= \frac{h^2 \varphi''(0)}{2h^2} \\ &\to \frac{\varphi''(0)}{2} && h \to 0 \\ &\in [-M, M]. \end{aligned}$$

On the other hand,

$$\begin{aligned}\frac{\varphi(h)+\varphi(-h)-2\varphi(0)}{h^2} &= \frac{1}{h^2}\mathbb{E}\left[e^{ih\xi}+e^{-h\xi}-2\right] \\ &= \frac{1}{h^2}\mathbb{E}\left[\left(\sin\frac{\xi h}{2}\right)^2\right] \\ &\to \mathbb{E}\xi^2. && h\to 0\end{aligned}$$

It remains to argue that the two limits do not differ much. Note that the convergence of the second limit requires an application of the dominated convergence theorem.

Corollary 4.2.2.1

1) If $\varphi_\xi \in C^\infty$, then ξ has all moments $\mathbb{E}\xi^k < \infty$
2) If ξ is such that $\mathbb{E}|\xi|^k < \infty$ for all k, then $\varphi_\xi \in C^\infty$
3) $\varphi_\xi(t) = \sum_{k=0}^\infty \frac{(it)^k}{k!}\mathbb{E}\xi^k$ for all

$$|t| < \frac{1}{\lim_n \sup\left(\frac{\left(\mathbb{E}|\xi|^k\right)^{\frac{1}{k}}}{k}\right)}$$

(assuming the limit is finite).

4.3 Distributions & Characteristic Functions for Sums of Random Variables

Theorem 4.3.1

Let ξ_1, ξ_2 be independent random variables with distributions $\mathbb{P}_{\xi_1}, \mathbb{P}_{xi_2}$ and characteristic functions $\varphi_{\xi_1}, \varphi_{\xi_2}$.

1) We have the following *convolutional distribution*

$$\begin{aligned}\mathbb{P}_{\xi_1+\xi_2}(A) &= \int_A \mathbb{P}_{\xi_1}(A-x)d\mathbb{P}_{\xi_2}(x) \\ &= \int_A \mathbb{P}_{\xi_2}(A-x)d\mathbb{P}_{\xi_2}(x)\end{aligned}$$

2) $\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t)\cdot\varphi_{\xi_2}(t)$

Proof
1) Consider $\eta := (\xi_1, \xi_2)$. Then $\mathbb{P}_\eta = \mathbb{P}_{\xi_1} \times \mathbb{P}_{\xi_2}$.

Define
$$B_A := \{x \in \mathbb{R}^2 : x_1 + x_2 \in A\}.$$

Then

$$\begin{aligned}
\mathbb{P}_{\xi_1+\xi_2}(A) &= \mathbb{P}\{\xi_1 + \xi_2 \in A\} \\
&= \mathbb{P}\{\eta \in B_A\} \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{B_A}(x) d\mathbb{P}_\eta(x) \\
&= \int_{\mathbb{R}} d\mathbb{P}_{\xi_1}(x_1) \left[\int_{\mathbb{R}} \mathbb{1}_{B_A}(x) d\mathbb{P}_{\xi_2}(x_2) \right].
\end{aligned}$$

Now,

$$\begin{aligned}
\int_{\mathbb{R}} \mathbb{1}_{B_A}(x) d\mathbb{P}_{\xi_2}(x_2) &= \int_{\mathbb{R}} \mathbb{1}_{A-x_1}(x_2) d\mathbb{P}_{\xi_2}(x_2) \\
&= \mathbb{P}_{\xi_2}(A - x_1).
\end{aligned}$$

This concludes the proof.

2) We have

$$\begin{aligned}
\varphi_{\xi_1+\xi_2}(t) &= \mathbb{E}e^{it(\xi_1+\xi_2)} \\
&= \left(\mathbb{E}e^{it\xi_1}\right)\left(\mathbb{E}e^{it\xi_2}\right) && \text{independence} \\
&= \varphi_{\xi_1}(t)\varphi_{\xi_2}(t).
\end{aligned}$$

Corollary 4.3.1.1
1) Let F_{ξ_1}, F_{ξ_2} be distribution functions. Then $F_{\xi_1+\xi_2}(x) = \int_{\mathbb{R}} F_{\xi_1}(x-y) dF_{\xi_2(y)}$. (Some more general integral)
2) Let p_{ξ_1}, p_{ξ_2} be densities. Then $p_{\xi_1+\xi_2}(x) = \int_{\mathbb{R}} p_{\xi_1}(x-y)p_{\xi_2}(y)dy$.

Corollary 4.3.1.2
1) Let $\xi_1, \dots, \xi_n$ be independent, then $\varphi_{\sum_i \xi_i} = \prod_{i=1}^n \varphi_{\xi_i}(t)$
2) If $\xi_1, \dots, \xi_n$ are iid, then $\varphi_{\sum_i \xi_i}(t) = \varphi_{\xi_1}(t)^n$

4.4 Gaussian Random Variables

Let $\xi \sim N(a, \sigma)$ and recall it has density

$$p(x) = p_a(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

for $a \in \mathbb{R}$ and $\sigma > 0$. Note that $\xi \equiv \sigma\xi_{0,1} + a$ where $\xi_{0,1} \sim N(0, 1)$.

Let us define

$$N(a, \sigma = 0) := \delta_a$$

where $\mathbb{P}\{\delta_a = a\} = 1$ is a point mass. Then $p_a(x, \sigma)$ satisfies

$$\begin{aligned} \frac{\partial p}{\partial \sigma} &= \frac{\partial^2 p}{\partial x^2} && \text{heat equation} \\ p(x, 0) &= \delta_a(x). \end{aligned}$$

Here we consider some more general derivative at $p(x, 0)$.

Proposition 4.4.1
If $\xi \sim N(0, \sigma)$,

1) $\mathbb{E}\xi = a$
2) $\operatorname{Var}\xi = \sigma^2$
3) $\mathbb{E}(\xi - a)^k = \sigma^k(k-1)!!$ for k even, otherwise it is 0
4) $\varphi_\xi(t) = \exp\left(ita - \frac{\sigma^2 t^2}{2}\right)$

Corollary 4.4.1.1

1) If $\xi_1 \sim N(a_1, \sigma_1)$, $\xi_2 \sim N(a_2, \sigma_2)$ are independent, then $\xi_1 + \xi_2 \sim N(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
2) Suppose $\xi_1, \ldots, \xi_n \sim N(0, \sigma)$ are independent, then $\sum_{i=1}^n \xi_i \sim N(0, \sqrt{\sum_i \sigma^2}) = N(0, \sqrt{n} \cdot \sigma)$
3) Suppose $\xi_1, \ldots, \xi_n \sim N(0, \sigma)$ are independent, then $\sum_{i=1}^n \xi_i \equiv \sqrt{n} \cdot \xi_1$

Example 4.4.2
Suppose $\xi_1, \xi_2 \sim \mathrm{Exp}(\lambda = 1)$ are iid. Recall then $p_{\xi_i}(x) = e^{-x}$ for $x > 0$. We thus have

$$\begin{aligned}
\varphi_{\xi_1-\xi_2}(t) &= \varphi_{\xi_1}(t) \cdot \varphi_{-\xi_2}(t) \\
&= \varphi_{\xi_1}(t) \cdot \varphi_{\xi_2}(-t) \\
&= \frac{1}{1-it} \cdot \frac{1}{1+it} \\
&= \frac{1}{1+t^2}.
\end{aligned}$$

This is the characteristic function of the *Laplacian distribution* with a density $p_\xi(x) = \frac{1}{2}e^{-|x|}$. To verify this,

$$\varphi_\xi(t) = \int_{\mathbb{R}} \frac{e^{itx} \cdot e^{-|x|}}{2} dx = \frac{1}{1+t^2}.$$

Coincidentally, the Laplacian distribution's characteristic function is proportional to the density of the *Cauchy distribution* with density $p(x) = \frac{1}{\pi(1+x^2)}$.

Indeed,

$$\begin{aligned}
p_L(x) &= \frac{1}{2}e^{-|x|} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{1}{1+t^2}dt \\
&= 2\mathbb{E}e^{it\xi_{\text{Cauchy}}}
\end{aligned}$$

Note that if $\xi_1, \dots, \xi_n$ are iid Cauchy, then $\sum_i \xi_i \equiv n\xi_1$. This can be shown through characteristic functions.

4.5 Stable Distributions

Definition 4.5.1 (Stable)
We say a distribution ξ is stable if for all $n \geq 1$, there is some $a_n > 0$ and $b_n \in \mathbb{R}$ such that the following holds: If $\xi_1, \dots, \xi_n \sim \xi$ iid, then $\sum_i \xi_i \setminus a_n\xi + b_n$.

Equivalently,

$$\varphi_\xi(t)^n = \varphi_{a_n\xi+b_n(t)} = e^{ib_nt}\varphi_\xi(a_nt).$$

Theorem 4.5.1 (Lévy-Khintchine)
ξ is stable if and only if the characteristic function is of the form

$$\varphi_\xi(t) = \exp\left[it\beta - c|t|^\alpha(1 + i\theta \operatorname{sgn}(t) G(t,\alpha))\right].$$

Here $0 < \alpha \leq 2, \beta \in \mathbb{R}, c \geq 0$, and $|\theta| < 1$ are hyperparameters and

$$G(t,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2}, & \alpha \neq 1 \\ -\frac{2\log|t|}{\pi}, & \alpha = 1 \end{cases}$$

Corollary 4.5.1.1
All symmetric stable distributions have characteristic functions of the form

$$\varphi_\xi(t) = e^{-c|t|^\alpha}$$

for $\alpha \in (0,2]$.

First note the inverse implication is straightforwards. Indeed, we can just multiply the characteristic functions.

On the other hand, consier $\alpha \geq 2$. Then

$$\varphi^{(2)}(0) \begin{cases} \neq 0, & \alpha = 2 \\ 0, & \alpha > 0 \end{cases}$$

But from a previous theorem $\mathbb{E}\xi^2 = 0$ so that $\xi \equiv 0$ is a constant.

Proposition 4.5.2
Suppose $\varphi_\xi(t) = e^{-c|t|^\alpha}$. For $r < \alpha$,

$$\mathbb{E}|\xi|^r < \infty$$

and for $r \geq \alpha$,

$$\mathbb{E}|\xi|^r = \infty.$$

Depending on the qualities of iid distributions, they tend to their family of "preferred stable distributions".

Chapter 5

Limit Theorems

Let η_k be a sequence of random variations. We wish to pose the question of convergence. For example,

$$\xi_k := \frac{1}{k}\sum_{i=1}^{k} \eta_k \to ?$$

$$\zeta_k := \frac{1}{\sqrt{k}}\left[\sum_{i=1}^{k} \eta_i - \sum_{i=1}^{k} \mathbb{E}\eta_i\right] \to ?$$

5.1 Convergence of Random Variables

Before we explore limits of random variables, we must first pin down precise notions of convergence.

Let $\xi_n, \xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (X, \rho)$ where (X, ρ) is an arbitrary metric space, ie $(\mathbb{R}, |\cdot|)$.

Definition 5.1.1 (Convergence Almost Everywhere)
We say $\xi_n \to \xi$ a.e., converges in $\mathbb{P}$ a.e., or with probability 1 if

$$\mathbb{P}\{\omega : \xi_n(\omega) \to \xi(\omega)\} = 1.$$

First, let us check that the set of pointwise convergence is measurable. Define

$$A_k^\epsilon := \{\omega : \rho(\xi_k(\omega), \xi(\omega)) > \epsilon\}$$

$$A^\epsilon := \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k^\epsilon =: \limsup_k A_k^\epsilon.$$

This is the set of points that that for all $n \geq 1$, there exists $k \geq n$ such that $\rho(\xi_k(\omega), \xi(\omega)) > \epsilon$. Thus the set of points of non-convergence is precisely

$$\bigcup_{m \in \mathbb{N}} A^{\frac{1}{m}} = \{\omega : \xi_n(\omega) \not\to \xi(\omega)\}.$$

We have constructed the set of pointwise (non-)convergence with countable operations on countably many measurable sets, which is thus measurable.

Proposition 5.1.1
$\xi_n \to \xi$ almost everywhere if and only if for all $\epsilon > 0$,

$$\mathbb{P}\left\{\omega : \sup_{k \geq n} \rho(\xi_k(\omega), \xi(\omega)) > \epsilon\right\} \to 0$$

as $n \to \infty$.

Proof
Remark that $A^{\frac{1}{n}} \supseteq A^{\frac{1}{m}}$ for $n \geq m$. Now, $\xi_n \to \xi$ almost everywhere if and only if

$$\begin{aligned}
\mathbb{P}\bigcup_m A^{\frac{1}{m}} = 0 &\iff \forall m, \mathbb{P}A^{\frac{1}{m}} = 0 \\
&\iff \forall \epsilon > 0, \mathbb{P}(A^{\epsilon}) = 0 \\
&\iff \forall \epsilon > 0, \mathbb{P}\bigcup_{k \geq n} A_k^{\epsilon} \to 0 && n \to \infty \\
&\iff \forall \epsilon > 0, \mathbb{P}\left\{\omega : \sup_{k \geq n} \rho(\xi_k(\omega), \xi(\omega)) > \epsilon\right\} \to 0. && n \to \infty
\end{aligned}$$

Proposition 5.1.2
<u>Borel-Cantelli Lemma:</u> If for all $\epsilon > 0$,

$$\sum_{k=1}^{\infty} \mathbb{P}\{\omega : \rho(\xi_k(\omega), \xi(\omega)) > \epsilon\} < \infty,$$

then $\xi_n \to \xi$ almost surely.
<u>Second Borel-Cantelli Lemma:</u> Also, if the ξ_k's are independent and

$$\sum_{k=1}^{\infty} \mathbb{P}\{\omega : \rho(\xi_k(\omega), \xi(\omega)) > \epsilon\} = \infty,$$

then $\mathbb{P}\{\omega : \xi_k(\omega) \not\to \xi(\omega)\} = 1$.

Definition 5.1.2 (Convergence in Probability)
Suppose ξ_n, ξ are random variables. We say $\xi_n \to \xi$ in probability if for every $\epsilon > 0$,

$$\mathbb{P}\{\omega : |\xi_n(\omega) - \xi(\omega)| > \epsilon\} \to 0.$$

Definition 5.1.3 (Convergence in Distribution)
We say $\xi_n \to \xi$ in distribution if for every bounded continuous $f : X \to \mathbb{R}$,

$$\mathbb{E}f(\xi_n) \to \mathbb{E}f(\xi).$$

Definition 5.1.4 (Convergence in Mean of Order p)
Suppose ξ_n, ξ are random variables. We say $\xi_n \to \xi$ in mean of order p if

$$\mathbb{E}|\xi_n - \xi|^p \to 0$$

for $\rho > 0$.

Let $([0,1], \mathcal{B}[0,1], \lambda)$ be our probability space.

Example 5.1.3
Consider $\xi \equiv 0$ and

$$\xi_n(\omega) := \begin{cases} e^n, & \omega \in \left[0, \frac{1}{n}\right] \\ 0, & \omega > \frac{1}{n} \end{cases}$$

$\xi_n \to \xi$ almost everywhere.

Also, $\lambda\{\omega : |\xi(\omega)| > \epsilon\} \leq \frac{1}{n}$. Thus we have convergence in probability.

Now, for any bounded continuous f,

$$\mathbb{E}f(\xi_n) = \frac{1}{n}f(e^n) + \frac{n-1}{n}f(0) \to f(0)$$

since f is bounded. Thus we have convergence in distribution.

However, $\mathbb{E}|\xi_n|^p = (e^n)^p \cdot \frac{1}{n} \not\to \xi$ as $n \to \infty$. Thus we do not have convergence in mean of order p for any $p \geq 1$.

Example 5.1.4
Take

$$\xi_n^{(i)} := \mathbb{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right]}$$

and flatten that to a 1-dimensional sequence.

We see that $\xi_n^{(i)}(\omega) \not\to 0$ a.e. since it takes on the value 1 for infinite many times in any tail. However,

$$\mathbb{P}\{\omega : |\xi_n^{(i)}(\omega) - 0| \geq \epsilon\} = \frac{1}{n} \to 0.$$

Thus we do have convergence in probability.

Theorem 5.1.5

1) convergence almost everywhere implies convergence in probability
2) convergene in L_p implies convergence in probability
3) convergence in probability implies convergene in distribution

Proof

1) We have

$$\begin{aligned}
&\mathbb{P}\{\omega : \rho(\xi_n(\omega), \xi(\omega)) > \epsilon\} \\
&\leq \mathbb{P}\left\{\omega : \sup_{k \geq n} \rho(\xi_n(\omega), \xi(\omega)) > \epsilon\right\} \\
&\to 0. && \text{convergence almost everywhere}
\end{aligned}$$

2) By Chebyshev's inequality, and L_p convergence,

$$\begin{aligned}
&\mathbb{P}\{\omega : \rho(\xi_n(\omega), \xi(\omega)) > \epsilon\} \\
&\leq \frac{1}{\epsilon^p} \mathbb{E}|\xi_n - \xi|^p \\
&\to 0.
\end{aligned}$$

3) Let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Thus $|f| \leq c$ for some $c > 0$.

Fix some $\epsilon > 0$, there is some $N \in \mathbb{N}$ such that $\mathbb{P}\{|\xi| > N\} < \frac{\epsilon}{4c}$.

Now, f is continuous, and thus uniformly continuous on $[-N, N]$. There is some $\delta > 0$ such that $|x - y| < \delta$ implies that $|f(x) - f(y)| < \epsilon$ for all $x, y \in [-N, N]$.

Let A, B, C be the following partition of Ω:

$$\begin{aligned}
A &:= \{\omega : |\xi_n - \xi| < \delta, |\xi| < N\} \\
B &:= \{\omega : |\xi_n - \xi| < \delta, |\xi| \geq N\} \\
C &:= \{\omega : |\xi_n - \xi| \geq \delta\}.
\end{aligned}$$

Then

$$
\begin{aligned}
&|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi)| \\
&\leq \mathbb{E}|f(\xi_n) - f(\xi)| \\
&= \int_\Omega |f(\xi_n) - f(\xi)| d\mathbb{P}(\omega) \\
&= \int_A |f(\xi_n) - f(\xi)| d\mathbb{P}(\omega) + \int_B |f(\xi_n) - f(\xi)| d\mathbb{P}(\omega) + \int_C |f(\xi_n) - f(\xi)| d\mathbb{P}(\omega) \\
&< \epsilon \cdot \mathbb{P}(A) + 2c \cdot \frac{\epsilon}{4c} + 2c \cdot \mathbb{P}\{|\xi_n - \xi| \geq \delta\} \\
&\to 0 \qquad n \to \infty
\end{aligned}
$$

Proposition 5.1.6
Let ξ_n be a random variable and $\xi \equiv a$. Then $\xi_n \to \xi$ in probability if and only if $\xi_n \to \xi$ in distribution.

5.2 Convergence of Probabilty Measures

Let $(E, \mathcal{E}, \rho)$ be a metric space. Let $\mathbb{P}_n$ be a sequence of probability measures on E.

Definition 5.2.1 (Weak Convergence)
We say $\mathbb{P}_n \to \mathbb{P}$ weakly if for every $f : E \to \mathbb{R}$ that is bounded and continuous,

$$\int_E f(x) d\mathbb{P}_n(x) \to \int_E f(x) d\mathbb{P}(x).$$

We note that we can think of distributions abstractly as a black box which computes expectations. By definition, $\xi_n \to \xi$ in distribution if and only if $\mathbb{P}_{\xi_n} \to \xi$ weakly.

Definition 5.2.2 (Convergence in General)
We say $\mathbb{P}_n \to \mathbb{P}$ in general if for every $A \in \mathcal{E}$ such that $\mathbb{P}(\partial A) = 0$,

$$\mathbb{P}_n(A) \to \mathbb{P}(A).$$

Note that the condition that the boundary has measure zero is to avoid point masses on boundaries.

Theorem 5.2.1
The following are equivalent.

1) $\mathbb{P}_n \to \mathbb{P}$ weakly
2) $\lim_n \sup \mathbb{P}_n(A) \leq \mathbb{P}(A)$ for all A closed
3) $\lim_n \inf \mathbb{P}_n(A) \geq \mathbb{P}(A)$ for all A open
4) $\mathbb{P}_n \to \mathbb{P}$ in general

Proof
<u>1) $\implies$ 2)</u> Let A be closed. We can the define

$$\begin{aligned}
\rho(x, A) &:= \inf_{y\in A} \rho(x, y) \\
A^\epsilon &:= \{x : \rho(x, A) < \epsilon\} \\
f_A^\epsilon(x) &:= \left(1 - \frac{1}{\epsilon}\rho(x, A)\right)_+ \\
&= \begin{cases} 1, & x \in A \\ 0, & x \in \overline{A^\epsilon} \\ \text{continuous}, & \text{else} \end{cases}
\end{aligned}$$

Remark that $A^\epsilon \downarrow A$ as $\epsilon \to 0$. We can think of f_A^ϵ as a continuous bounded approximation of $\mathbb{1}_A$.

Then for every $\epsilon > 0$,

$$\begin{aligned}
\lim_n \sup \mathbb{P}_n(A) &= \int_E \mathbb{1}_A d\mathbb{P}_n(x) \\
&= \int_E f_A^\epsilon(x) d\mathbb{P}_n(x) \\
&\to \int_E f_A^\epsilon(x) d\mathbb{P}(x) && \text{weak convergence} \\
&\leq \mathbb{P}(A^\epsilon).
\end{aligned}$$

But as $\epsilon \to 0$, $\mathbb{P}(A^\epsilon) \downarrow \mathbb{P}(A)$ by the continuity of measure.

<u>2) $\iff$ 3)</u> This is not hard to see by considering complements.

<u>2), 3) $\implies$ 4)</u> Let $B \in \mathcal{E}$ be such that $\mathbb{P}(\partial B) = 0$. Remark that $\mathcal{B} \cup \partial B$ is closed and contains B, while $B \setminus \partial B$ is open and contained in B.

We have

$$\begin{aligned}
\limsup_n \mathbb{P}_n(B) &\leq \limsup_n \mathbb{P}_n(B \cup \partial B) \\
&\leq \mathbb{P}(B \cup \partial B) \\
&= \mathbb{P}(B) \\
\liminf_n \mathbb{P}_n(B) &\geq \liminf_n \mathbb{P}_n(B \setminus \partial B) \\
&\geq \mathbb{P}_n(B \setminus \partial B) \\
&= \mathbb{P}(B)
\end{aligned}$$

Thus the limit exists and $\mathbb{P}_n(B) \to \mathbb{P}(B)$.

4) $\iff$ 1) Let $f : E \to \mathbb{R}$ be bounded and continuous. Then there is some $C > 0$ such that $|f| \leq C$.

Consider $\mathbb{P}_f$, the push-forward measure on $\mathbb{R}$. We claim that it has at most countably many atoms. This is not hard to see by consider sets of the form $\{x \in \mathbb{R} : \mathbb{P}_f(x) \geq \frac{1}{n}\}$ which has at most n elements.

Fix $\epsilon > 0$.

Let $-C = t_0 < t_1 < \cdots < t_k = C$ be a partition of $[-C, C]$. From the claim above, we can always choose t_i such that

(i) $|t_i - t_{i+1}| < \epsilon$

(ii) $\left|\sum_{i=0}^{k-1} t_i \mathbb{P}(B_i) - \int f(x) d\mathbb{P}(x)\right| < \epsilon$

(iii) $\mathbb{P}\{x : f(x) = t_i\} = \mathbb{P}_f(t_i) = 0$

Then

$$B_i := f^{-1}[t_i, t_{i+1}]$$

for $0 \leq i < k$ forms a partition of E such that $\mathbb{P}(\partial B_i) = 0$. By assumption,

$$\sum_{i=0}^{k-1} t_i \mathbb{P}_n(B_i) \to \sum_{i=0}^{k-1} t_i \mathbb{P}(B_i).$$

Now to show weak convergence,

$$
\begin{aligned}
&\left|\int f(x)d\mathbb{P}_n(x) - \int f(x)d\mathbb{P}(x)\right| \\
&\leq \left|\int f(x)d\mathbb{P}_n(x) - \sum_{i=0}^{k-1} t_i\mathbb{P}_n(B_i)\right| + \left|\sum_{i=0}^{k-1} t_i\mathbb{P}_n(B_i) - \sum_{i=0}^{k-1} t_i\mathbb{P}(B_i)\right| \\
&\quad + \left|\sum_{i=0}^{k-1} t_i\mathbb{P}(B_i) - \int f(x)d\mathbb{P}(x)\right| \\
&\leq \max_{0\leq i<k}|t_i - t_{i+1}| + \left|\sum_{i=0}^{k-1} t_i\mathbb{P}_n(B_i) - \sum_{i=0}^{k-1} t_i\mathbb{P}(B_i)\right| \\
&\quad + \left|\sum_{i=0}^{k-1} t_i\mathbb{P}(B_i) - \int f(x)d\mathbb{P}(x)\right| \\
&< \epsilon + \left|\sum_{i=0}^{k-1} t_i\mathbb{P}_n(B_i) - \sum_{i=0}^{k-1} t_i\mathbb{P}(B_i)\right| + \epsilon \\
&\to 0. && n\to\infty
\end{aligned}
$$

In particular, this shows that weak limits are unique. Indeed, if a sequence had two weak limits, they must agree on all open intervals. But then by Carath'eodory's extension theorem, they are in fact the same measures.

Consider a family of probability measures $\{\mathbb{P}_\alpha : \alpha \in A\}$ defined on a measurable space $(E, \mathcal{E})$ equipped additionally with a metric ρ.

Definition 5.2.3 (Relative Compactness)
$\{P_\alpha\}$ is *relatively compact* if any sequence $\mathbb{P}_n$ contains a weakly converging subsequence whose limit is a probability measure.

Definition 5.2.4 (Tight)
$\{P_\alpha\}$ is *tight* if for all $\epsilon > 0$, there is some $K_\epsilon \subseteq E$ compact such that

$$\mathbb{P}_\alpha(K_\epsilon) > 1 - \epsilon$$

for all $\alpha \in A$.

Theorem 5.2.2 (Prokhorov)
Suppose $(E, \mathcal{E}, \rho)$ is complete and separable. Then $\{\mathbb{P}_\alpha\}$ is relatively compact if and only if it is tight.

Remark that in the setting of Prokhorov's theorem, tighting is equivalent to the statement that any converging subsequence converges to a probability measure.

Lemma 5.2.3
There is some $K > 0$ such that for all $a > 0$, if $\mathbb{P}$ is any probability measure,

$$\int_{|x|>\frac{1}{a}} d\mathbb{P}(x) \leq \frac{K}{a} \int_0^a [1 - \operatorname{Re}\varphi(t)]\, dt.$$

Proof
By computation,

$$\begin{aligned}
\frac{1}{a}\int_0^a [1 - \operatorname{Re}\varphi(t)]\, dt &= \frac{1}{a}\int_0^a \int_{-\infty}^{\infty} [1 - \cos tx]\, d\mathbb{P}(x)dt \\
&= \int_{-\infty}^{\infty} \frac{1}{a}\int_0^a [1 - \cos tx]\, dtd\mathbb{P}(x) \\
&= \int_{-\infty}^{\infty} \left[1 - \frac{\sin ax}{ax}\right] d\mathbb{P}(x) \\
&\geq \inf_{|y|\geq 1}\left[1 - \frac{\sin y}{y}\right] \cdot \mathbb{P}\left\{|x| > \frac{1}{a}\right\}.
\end{aligned}$$

By taking $\frac{1}{K} = \inf_{|y|\geq 1}\left[1 - \frac{\sin y}{y}\right]$, we are done.

Theorem 5.2.4 (Lévy's Continuity Theorem for $\mathbb{R}^d$)
Let $\mathbb{P}_n, n \in \mathbb{N}$ be a sequence of probability measures on $\mathbb{R}^d$ with characteristic functions $\varphi_n(t) = \mathbb{E}\left[\exp(-i\langle t, \xi_n\rangle)\right]$, where $\mathbb{P}_n = \mathbb{P}_{\xi_n}$.

1) If $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ to a probability measure with characteristic function $\varphi(t)$, then $\varphi_n(t) \to \varphi(t)$ for every $t \in \mathbb{R}$ as $n \to \infty$.
2) If the limit $\varphi(t) = \lim_{n\to\infty} \varphi_n(t)$ exists for all t, then $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ to a probability measure and $\varphi(t)$ is its characteristic function. Moreover, $\varphi(t)$ is continuous.

Proof ($d = 1$)
<u>1)</u> Suppose $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$. For every $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous,

$$\int_{\mathbb{R}} f(x) d\mathbb{P}_n(x) \to \int_{\mathbb{R}} f(x) f\mathbb{P}(x).$$

Let $f(x) := \operatorname{Re}\exp(itx)$. Then $\int_{\mathbb{R}} \operatorname{Re}\exp(itx) d\mathbb{P}_n(x) \to \int_{\mathbb{R}} \operatorname{Re}\exp(itx) d\mathbb{P}(x)$ and similariy for $\operatorname{Im}\exp(itx)$. Thus $\varphi_n(t) \to \varphi(t)$ as desired.

2) First, we show that $\{\mathbb{P}_n, \mathbb{P} : n \in \mathbb{N}\}$ is tight. Note that it suffices to show the claim for all $n \geq N$ where $N \in \mathbb{N}$ is fixed. This is because finite unions preserve compactness. We have

$$
\begin{aligned}
&\mathbb{P}_n \left[-\frac{1}{a}, \frac{1}{a}\right]^c \\
&= \int_{|x|>\frac{1}{a}} d\mathbb{P}_n(x) \\
&\leq \frac{K}{a} \int_0^a [1 - \operatorname{Re}\varphi_n(t)]\, dt && \text{lemma} \\
&< \epsilon + \frac{K}{a} \int_0^a [1 - \operatorname{Re}\varphi(t)]\, dt. && \text{CDT, large } n \\
&< \epsilon + K\epsilon'
\end{aligned}
$$

An appropriate choice of ϵ' yields the result.

By Prokhorov's theorem, tightness implies the existence of a subsequence converging to a probability measure. But any subsequence converges to the same limit as the original sequence since weak limits are unique by an earlier remark. Thus $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

By 1), $\varphi_n(t) \to \psi(t)$ for all $t \in \mathbb{R}$ where ψ is the characteristic function of $\mathbb{P}$. Since limits in $\mathbb{C}$ are unique, it must be that $\varphi_n(t) \to \varphi(t) = \psi(t)$ for all $t \in \mathbb{R}$.

We now show continuity. This is not hard since

$$
\begin{aligned}
|\mathbb{E}[\exp(itx)] - E[\exp(it(x+h))]| &\leq \mathbb{E}|\exp(itx)[1 - \exp(ith)]| \\
&\to 0. && h \to 0
\end{aligned}
$$

Corollary 5.2.4.1
If $\mathbb{P}_n, \mathbb{P}$ are probability distributions with characteristic functions φ_n, φ, then $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ if and only if $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$ for every $t \in \mathbb{R}$.

5.3 Laws of Large Numbers

5.3.1 Definitions

Let $\xi_1, \xi_2, \dots$ be a sequence of random variables and

$$S_n(\omega) := \sum_{i=1}^n \xi_i(\omega).$$

Definition 5.3.1 (Law of Large Numbers)
We say $\{\xi_i\}$ obeys the *law of large numbers* if there is some $L \in \mathbb{R}$ such that

$$\frac{1}{n} S_n \xrightarrow{p/d} L.$$

Equivalently, we can require $\mathbb{P}_{\frac{1}{n}S_n} \xrightarrow{w} \delta_L$ where δ_L is the delta distribution at L.

Suppose $\{\xi_i\}$ obeys the LLN and $\mathbb{E}\xi_i$ exists. Then

$$\mathbb{E}\frac{1}{n} S_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\xi_i \xrightarrow{p/d} L.$$

In other words, $\frac{1}{n}S_n - \mathbb{E}\frac{1}{n}S_n \xrightarrow{p/d} 0$.

Definition 5.3.2 (Strong Law of Large Numbers)
We say $\{\xi_i\}$ obeys the SLLN if there is some $L \in \mathbb{R}$ such that

$$\frac{1}{n} S_n \xrightarrow{a.s.} L.$$

Example 5.3.1 (Bernoulli's Golden Theorem)
Suppose $\xi_i \sim \text{Be}(p)$ iid. Then $\frac{1}{n}S_n(\omega) \to p$. This convergence is in probability, distribution, and almost everywhere.

5.3.2 Laws of Large Numbers

Theorem 5.3.2 (Markov's LLN)
Let $\xi_i, i \in \mathbb{N}$ be random variables with $\mathbb{E}\xi_i, \operatorname{Var}\xi_i < \infty$. If $\frac{\operatorname{Var}\left(\sum_{i=1}^{n}\xi_i\right)}{n^2} \to 0$, then the LLN holds:

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\xi_i \xrightarrow{p/d} 0.$$

Proof
By Chebyshev's inequality,

$$\begin{aligned}
&\mathbb{P}\left\{\omega : \left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\xi_i\right| \geq \epsilon\right\} \\
&\leq \frac{\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)}{\epsilon^2} \\
&\to 0. \qquad n \to \infty
\end{aligned}$$

Corollary 5.3.2.1 (Chebyshev's LLN)
Let $\xi_i, i \in \mathbb{N}$ be independent random variables such that there is some $C \in \mathbb{R}$ for which $\operatorname{Var}\xi_i < C$. Then LLN holds.

Theorem 5.3.3 (Khinchin's LLN)
Let $\xi_i, i \in \mathbb{N}$ be iid random variables with characteristic function $\varphi(t)$. If $\varphi'(0)$ exists, then LLN holds:

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i(\omega) \xrightarrow{p/d} a =: \frac{\varphi'(0)}{i}.$$

Note here that we do not require ξ_i has finite expectation. However, if it exists, then $\mathbb{E}\xi_i = \frac{\varphi'(0)}{i}$.

Proof
Define $\eta_i := \xi_i - a$, which are also iid random variables. We show the equivalent statement

$$\frac{1}{n}S_n := \frac{1}{n}\sum_{i=1}^{n}\eta_i \xrightarrow{p/d} 0.$$

Remark that

$$\begin{aligned}
\varphi_\eta(0) &= e^{-ita}\varphi_\xi(t) \\
\varphi'_\eta(0) &= -itae^{-ita}\underbrace{\varphi_\xi(0)}_{=1} + e^{-ita}\underbrace{\varphi'_\xi(0)}_{=ai} \\
&= 0
\end{aligned}$$

Indeed, our plan is to apply Lévy's continuity theorem. Now, $\varphi_{\delta_0}(t) = \mathbb{E}e^{it0} = 1$ for every

$t \in \mathbb{R}$. Thus we show that $\varphi_{\frac{1}{n}S_n}(t) \to 1$ for every $t \in \mathbb{R}$.

$$\begin{aligned}\varphi_{\frac{1}{n}S_n}(t) &= \prod_{i=1}^n \varphi_{\frac{1}{n}\eta_i}(t) \\ &= \prod_{i=1}^n \varphi_{\eta_i}\left(\frac{t}{n}\right) \\ &= \varphi_\eta\left(\frac{t}{n}\right)^n.\end{aligned}$$

Taking logarithms,

$$\begin{aligned}\log \varphi_{\frac{1}{n}S_n}(t) &= n \log \varphi_\eta\left(\frac{t}{n}\right) \\ &= n \log\left[\varphi_\eta(0) + \frac{t}{n}\varphi_\eta'(0) + o\left(\frac{1}{n}\right)\right] \\ &= n \log\left(1 + o\left(\frac{1}{n}\right)\right) \\ &= n \cdot o\left(\frac{1}{n}\right) \\ &\to 0. \qquad n \to \infty\end{aligned}$$

5.3.3 Strong Law of Large Numbers (Almost Surely Convergence of Series of Random Variables)

Let $\xi_i, i \in \mathbb{N}$ be independent random variables and define $S_n := \sum_{i=1}^n \xi_i$.

Lemma 5.3.4 (Kolmogorov's Inequality)
Let $\xi_i, i \in \mathbb{N}$ be independent with $\mathbb{E}\xi_i = 0$ and $\mathbb{E}\xi_i^2 < \infty$. Then for every $\epsilon > 0$,

$$\mathbb{P}\left\{\max_{k\in[n]} S_k > \epsilon\right\} \leq \frac{\mathbb{E}S_n^2}{\epsilon^2}.$$

Theorem 5.3.5 (Kolmogorov, Khinchin)
Let $\xi_i, i \in \mathbb{N}$ be random variables with $\mathbb{E}\xi_i = 0$ and $\sum_{i=1}^n \mathbb{E}\xi_i^2 < \infty$. Then $\sum_{i=1}^\infty \xi_i$ converges with probability 1.

Proof
We show that the set of ω where $S_n(\omega)$ is a Cauchy sequence has measure 1. Recall that

$\mathbb{R}$ is complete and thus Cauchy sequences always converge.

Indeed, the statement above is true if and only if

$$\begin{aligned}
&\mathbb{P}\left\{\omega : \sup_{k\geq n}|S_{n+k} - S_n| \geq \epsilon\right\} \to 0 && n\to\infty \\
&\iff \mathbb{P}\left\{\omega : \lim_{N\to\infty} \sup_{k\in[N]} |S_{n+k} - S_n| \geq \epsilon\right\} \to 0 && n\to\infty \\
&\iff \lim_{N\to\infty} \mathbb{P}\left\{\max_{k\in[N]}|S_{n+k} - S_n| \geq \epsilon\right\} \to 0 && n\to\infty
\end{aligned}$$

The last if and only if holds due to the downward continuity of measure.

Observe that

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}\left\{\max_{k\in[N]}|S_{n+k} - S_n| \geq \epsilon\right\} \\
&\leq \lim_{N\to\infty} \left(\sum_{k=n+1}^{n+N} \mathbb{E}\xi_k^2\right)/\epsilon^2 && \text{Kolmogorov's Inequality} \\
&\to 0. && n\to\infty
\end{aligned}$$

The last limit is justified since the tail series of a convergent series must tend towards 0.

Lemma 5.3.6 (Kronecker)
Let $0 < b_n \uparrow \infty$ and $x_n \in \mathbb{R}$ be such that $\sum_{n=1}^{\infty} x_n = L \in \mathbb{R}$. Then

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j \to 0.$$

Theorem 5.3.7 (Kolmogorov)
Let $\xi_i, i \in \mathbb{N}$ be independent random variables such that $\operatorname{Var}\xi_i < \infty$. Let $0 < b_n \uparrow \infty$ be such that $\sum_{n=1}^{\infty} \frac{\operatorname{Var}\xi_n}{b_n^2} < \infty$. Then

$$\frac{S_n(\omega) - \mathbb{E}S_n}{b_n} \xrightarrow{a.s.} 0.$$

Proof
Observe that

$$\frac{S_n(\omega) - \mathbb{E}S_n(\omega)}{b_n} = \frac{1}{b_n} \sum_{k=1}^{n} b_k \left(\frac{\xi_k(\omega) - \mathbb{E}\xi_k}{b_k} \right).$$

Let $x_k(\omega) := \frac{\xi_k(\omega) - \mathbb{E}\xi_k}{b_k}$. If we show that $\sum_{k=1}^{\infty} x_k \in \mathbb{R}$ a.s. Then we can apply Kronecker's lemma to conclude the proof.

Note that $\mathbb{E}x_k = 0$ and $\operatorname{Var} x_k = \frac{\operatorname{Var} \xi_k}{b_k^2}$. By assumption, $\sum_{k=1}^{\infty} \operatorname{Var} x_k < \infty$. Thus by the Kolmogorov-Khinchin Theorem, $\sum_{k=1}^{\infty} x_k \to 0$ converges a.s.

From our remark above, this terminates the proof.

Corollary 5.3.7.1
Let $\xi_i, i \in \mathbb{N}$ be independent random variable with finite variance. If $\sum_{n=1}^{\infty} \frac{\operatorname{Var} \xi_n}{n^2} < \infty$, then SLLN holds.

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i(\omega) - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\xi_i \xrightarrow{a.s.} 0.$$

Example 5.3.8
Let $\xi_i \sim \operatorname{Be}(p), i \in \mathbb{N}$ such that $\operatorname{Var} \xi_i = p(1-p)$. Choose $b_n := \frac{1}{n}$. Then $\sum_{n=1}^{\infty} \frac{p(1-p)}{b_n^2} < \infty$. It follows that

$$\frac{S_n(\omega) - np}{n} \xrightarrow{a.s.} 0.$$

In fact, we can do even better and show that

$$\frac{S_n(\omega) - np}{\sqrt{n} \log n} \xrightarrow{a.s.} 0.$$

Theorem 5.3.9 (Kolmogorov's SLLN)
Let $\xi_i, i \in \mathbb{N}$ be iid random variables such that $\mathbb{E}|\xi_i| < \infty$. Then SLLN holds:

$$\frac{1}{n} S_n \xrightarrow{a.s.} \mathbb{E}\xi.$$

5.3.4 Monte Carlo Methods

Let $f : [0,1] \to [0,1]$ be continuous. Suppose we wish to compute $\int_0^1 f(x)dx$. Let $(\xi_i, \eta_i) \sim U[0,1]^2$ iid. Define

$$\rho_i := \begin{cases} 1, & \eta_i \leq f(\xi_i) \\ 0, & \eta_i > f(\xi_i) \end{cases}$$

Then ρ is the characteristic function of the complement of the graph of f and we have

$$\int_0^1 f(x)dx = \mathbb{E}\rho_i \xleftarrow{w} \frac{1}{n}\sum_{i=1}^n \rho_i.$$

Thus we can use stochastic approximations of deterministics quantities.

5.4 Central Limit Theorems

The essence of LLN is that demeaning and rescaling a series leads to convergence to 0. Central limit theorems (CLTs) ask if we apply a different scaling factor, can we get to some non-trivial limits?

Lévy's theorem is the most useful and is seen the most often.

Theorem 5.4.1 (Lévy)
Let $\xi_i, i \in \mathbb{N}$ be iid random variables such that $\operatorname{Var}\xi_i = \sigma^2 < \infty$. Define $S_n := \sum_{i=1}^n \xi_i$.

1) $\frac{S_n - n\mathbb{E}\xi_1}{\sqrt{n \operatorname{Var}\xi_1}} \xrightarrow{d} N(0,1)$
2) For every $x \in \mathbb{R}$,

$$\mathbb{P}\left\{\frac{X_n - n\mathbb{E}\xi_1}{\sqrt{n\operatorname{Var}\xi_1}} \le x\right\} \to \Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

Proof
Recall that convergence in distribution is equivalent to weak convergence which is equivalent to convergence in general. Since a distribution function uniquely determines a distribution, we need only show 1) and 2) automatically follows.

Define $\eta_i(\omega) := \xi_i(\omega) - \mathbb{E}\xi_i$. Then η_i are iid random varaibles with mean 0 and variance σ^2. It suffices then to show that

$$\frac{\sum_{i=1}^n \eta_i}{\sqrt{n}\sigma} \xrightarrow{d} Z$$

where $Z \sim N(0,1)$. But by the corollary to Lévy's continuity theorem, it suffices to show that for every $t \in \mathbb{R}$,

$$\begin{aligned}\varphi_{\frac{\Sigma_n}{\sigma\sqrt{n}}}(t) &= \varphi_{\frac{\eta_1}{\sigma\sqrt{n}}}(t)^n \\ &= \varphi_{\eta_1}\left(\frac{t}{\sigma\sqrt{n}}\right)^n \\ &\to \varphi_{N(0,1)}(t) \\ &= \exp\left(-\frac{t^2}{2}\right).\end{aligned}$$

Since $\operatorname{Var}\eta_1 < \infty$ we know that $\eta \in L^2(\Omega)$ and thus by a previous theorem, $\varphi_\eta(t)$ is twice differentiable. Thus we can write

$$\begin{aligned}\varphi_\eta(t) &= \varphi_\eta(0) + \varphi_\eta'(0)t + \varphi_\eta''(0)\frac{t^2}{2} + o(t^2) \\ &= \varphi_\eta(0) + 0 \cdot t + \operatorname{Var}\eta_1 \frac{t^2}{2} + o(t^2) \\ &= 1 - \frac{\sigma^2 t^2}{2} + o(t^2).\end{aligned}$$

Since we are holding t, σ fixed,

$$\begin{aligned}\varphi_{\frac{\Sigma_n}{\sigma\sqrt{n}}}(t) &= \varphi_\eta\left(\frac{t}{\sigma\sqrt{n}}\right)^n \\ &= \left[1 - \frac{\sigma^2}{2} \cdot \frac{t^2}{n\sigma^2} + o\left(\frac{1}{n}\right)\right]^n \\ &= \left(1 - \frac{t^2}{2n}\right)^n + \underbrace{o\left(\frac{1}{n}\right)\dots}_{\in o(1)} \\ &\to \exp\left(-\frac{t^2}{2}\right). && n \to \infty\end{aligned}$$

We can alternatively take the log of $\varphi_{\frac{\Sigma_n}{\sqrt{n}\sigma}}$.

The following is an even stronger result which requires more assumptions of moments.

Theorem 5.4.2
Let $\xi_i, i \in \mathbb{N}$ be iid random variables with $\operatorname{Var}\xi_i = \sigma^2$ and $\mathbb{E}|\xi_1|^3 < \infty$. Then as $n \to \infty$,

$$\sup_x \left| \mathbb{P}\left\{ \frac{\sum_{i=1}^n \xi_i(\omega) - \mathbb{E}\xi_1 \cdot n}{\sigma\sqrt{n}} \le x \right\} - \Phi(x) \right| \le \frac{C\mathbb{E}|\xi_1|^3}{\sigma^3\sqrt{n}} \to 0$$

where C is some universal constant.

Example 5.4.3
Let $\xi_i \sim \operatorname{Be}(p), i \in \mathbb{N}$ be iid and consider $S_n := \sum_{i=1}^n \xi_i$. Then as $n \to \infty$,

$$\sup_x \left| \mathbb{P}\left\{ \frac{S_n(\omega) - np}{\sqrt{np(1-p)}} \le x \right\} - \Phi(x) \right| \to 0$$

Theorem 5.4.4 (Lindeberg's CLT)
Let $\xi_i, i \in \mathbb{N}$ be independent random variables with $\operatorname{Var}\xi_i < \infty$. Define $B_n^2 := \sum_{i=1}^n \operatorname{Var}\xi_i$.
If *Lindeberg's condition* holds: For every $\epsilon > 0$,

$$T_{i,\epsilon} := \{\vec{x} \in \mathbb{R} : |x - \mathbb{E}\xi_i| > \epsilon B_n\}$$

$$\frac{1}{B_n^2} \sum_{i=1}^n \int_{T_{i,\epsilon}} |x - \mathbb{E}\xi_i|^2 d\mathbb{P}_{\xi_i}(x) \to 0. \qquad n \to \infty$$

Then

$$\frac{\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbb{E}\xi_i}{B_n} \xrightarrow{d} N(0,1).$$

Corollary 5.4.4.1 (Lindeberge's CLT Implies Lévy's CLT)
Let $B_n^2 := n \operatorname{Var}\xi_1$. For every i and ϵ, $T_{i,\epsilon} \downarrow \varnothing$ as $n \to \infty$. Thus

$$\begin{aligned} \frac{1}{B_n^2} \sum_{i=1}^n \int_{T_{i,\epsilon}} |x - \mathbb{E}\xi_i|^2 d\mathbb{P}_{\xi_i}(x) &= \frac{n}{n \operatorname{Var}\xi_1} \int_{T_{1,\epsilon}} \dots d\mathbb{P}_{\xi_1}(x) \\ &\to 0. \qquad n \to \infty \end{aligned}$$

Corollary 5.4.4.2 (Lyapunov)
Let $\xi_i, i \in \mathbb{N}$ be independent random variables with finite variance. Define $B_n^2 = \operatorname{Var} \sum_{i=1}^n \xi_i$. If *Lyapunov's condition* holds: For every $\delta > 0$,

$$\frac{1}{B_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|\xi_i - \mathbb{E}\xi_i|^{2+\delta} \to 0. \qquad n \to \infty$$

Then CLT holds.

Proof
We have

$$\begin{aligned}\mathbb{E}|\xi_i - \mathbb{E}\xi_i|^{2+\delta} &\geq \int_{T_{i,\epsilon}} |x - \mathbb{E}\xi_i|^{2+\delta} d\mathbb{P}_{\xi_i}(x) \\ &\geq \epsilon^\delta B_n^\delta \int_{T_{i,\epsilon}} |x - \mathbb{E}\xi_i|^2 d\mathbb{P}_{\xi_i}(x).\end{aligned}$$

So

$$\begin{aligned}\frac{1}{B_n^2} \sum_{i=1}^n \int_{T_{i,\epsilon}} |x - \mathbb{E}\xi_i|^2 d\mathbb{P}_{\xi_i}(x) &\leq \frac{1}{\epsilon^\delta B_n^{2+\delta}} \mathbb{E}|\xi_i - \mathbb{E}\xi_i|^{2+\delta} \\ &\to 0. \qquad n \to \infty\end{aligned}$$

5.5 More on Limits

There are more results on convergence. For instance, the laws of iterated logarithms and the laws of large deviation.

5.5.1 Violating CLT

Let $\xi_i, i \in \mathbb{N}$ be iid random variables. We know that if $\operatorname{Var} \xi_i < \infty$, then the central limit theorem holds. Yet we know that some stable distributions do not have expectation, yet there are CLT-like results for these distributions.

Theorem 5.5.1
Let η be a random variable. Then it is the limit of some

$$\frac{\sum_{i=1}^{n} \xi_i - A_n}{B_n} \xrightarrow{d} \eta$$

if and only if η is stable.

We can prove this theorem using the idea of grouping partial sums into a higher-level series.

5.5.2 Fundamental Theorem of Statistics

Let $\xi_i, \dots, \xi_N$ be iid random variables and $F(x)$ their common distribution function.

Definition 5.5.1 (Empirical Distribution Function)
For $N \geq 1$,

$$F_N(x, \omega) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}\{\xi_k(\omega) \leq x\}.$$

Notice that $F_N(x, \cdot)$ is a random variable and F_N is in fact a stochastic process.

By Kolmogorov's SLLN, for every $x \in \mathbb{R}$,

$$F_N(x, \omega) - \mathbb{E}F_N(x, \omega) = F(x) \xrightarrow{a.s.} 0.$$

Theorem 5.5.2 (Glivenko-Cantelli)
Let $\xi_i, i \in \mathbb{N}$ be iid random variables. Then

$$D_N(\omega) := \sup_{x \in \mathbb{R}} |F_N(x, \omega) - F(x)| \xrightarrow{a.s.} 0.$$

Proof
First we note that

$$D_N(\omega) = \sup_{q \in \mathbb{Q}} |F_N(q, \omega) - F(q)|$$

since F is right continuous. A countable supremum of measurable functions remain measurable.

Now, we determine some bounds. Fix $M \geq 2$. Define

$$\begin{aligned} X_{M,K} &= \min\left\{x \in \mathbb{R} : \frac{K}{M} \leq F(x)\right\} && 1 \leq K < M \\ X_{M,0} &= -\infty \\ X_{M,M} &= \infty. \end{aligned}$$

Fix $\omega \in \Omega$. We write $F_N(x-0,\omega) := \lim_{\epsilon\to 0^+} F_N(x-\epsilon,\omega)$. For any $x \in [X_{M,K}, X_{M,K+1}) \neq \varnothing$,

$$\begin{aligned} F_N(x,\omega) - F(x) &\leq F_N(X_{M,K+1} - 0,\omega) - F(X_{M,K}) \\ &= F_N(X_{M,K+1} - 0,\omega) - F(X_{M,K+1} - 0) + \underbrace{F(X_{M,K+1} - 0) - F(X_{M,K})}_{\leq \frac{1}{M}}. \end{aligned}$$

Similarly,

$$F_N(x,\omega) - F(x) \geq F_N(X_{M,K},\omega) - F(X_{M,K}) - \frac{1}{M}.$$

Thus we can compare F_N, F at only finitely many points. It follows that for every $\omega \in \Omega$,

$$\begin{aligned} &\sup_{x\in\mathbb{R}} |F_N(x,\omega) - F(x)| \\ &\leq \max_{0\leq k,\ell\leq M} \{|F_N(x_{M,K},\omega) - F(X_{M,K})|, |F_N(X_{M,\ell} - 0,\omega) - F(X_{M,\ell} - 0)|\} + \frac{1}{M} \\ &\xrightarrow{a.s.} \frac{1}{M} && N \to \infty \end{aligned}$$

Note that $F_N(x,\omega) \to F(x)$ a.s. in L^∞. Thus for any φ that is continuous under the sup norm,

$$\varphi(F_N(x,\omega)) \to \varphi(F(x))$$

ω-a.s.

Theorem 5.5.3 (Kolmogorov)
We have

$$\mathbb{P}\{\sqrt{N} \cdot \sup_x |F_N(x,\omega) - F(x)| \leq y\} \to K(y)$$

where $K(y) := \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2y^2)$.

Chapter 6

Theory of Random Vectors

Most of the theory we developed for random variables extend to random vectors. Let $\bar{\xi} = (\xi_1, \ldots, \xi_d)$ denote a random vector.

6.1 Moments

Definition 6.1.1 (Expectation)
We define
$$\mathbb{E}\bar{\xi} := (\mathbb{E}\xi_1, \ldots \mathbb{E}\xi_d).$$

Definition 6.1.2 (Mixed Moments)
We define the *mixed moment of order* $\bar{k}$ as

$$\begin{aligned}\mathbb{E}(\xi_1^{k_1}, \cdots \xi_d^{k_d}) &= \int_\Omega \prod_{i=1}^d \xi_i(\omega)^{k_i} d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}^d} \prod_{i=1}^d \xi_i(\omega)^{k_i} d\mathbb{P}_{\bar{\xi}}(\bar{x}).\end{aligned}$$

Definition 6.1.3 (Covariance Matrix)
We define the covariance matrix $R_{\bar{\xi}} := [r_{ij}]_{i,j=1}^d$ with entries
$$r_{ij} = \text{Cov}(\xi_i, \xi_j).$$

We say $R_{\bar{\xi}}^{-1}$ is the *concentration matrix* if it exists.

Proposition 6.1.1

1) We have

$$\mathbb{R}_{\bar{\xi}} = \mathbb{E}\left[(\bar{\xi} - \mathbb{E}\bar{\xi})(\bar{\xi} - \mathbb{E}\bar{\xi})^T\right]$$
$$= \mathbb{E}(\bar{\xi}\bar{\xi})^T - \mathbb{E}\bar{\xi}\mathbb{E}\bar{\xi}^T.$$

2) $R_{\bar{\xi}}$ is symmetric, positive semidefinite.

3) $\det R_{\bar{\xi}} = 0.$ if and only if $\bar{\xi}$ lives in some hyperplane, ie $x^T R_{\bar{\xi}} x = 0$ for some $x \neq 0$.

4) Any symmetric positive definite matrix is the covariance matrix of some random vector.

5) If $A : \mathbb{R}^d \to \mathbb{R}^k$ is a matrix, then $R_{A\bar{\xi}} = AR_{\bar{\xi}}A^T$.

6.2 Characteristic Functions

Definition 6.2.1 (Characteristic Function)
We define

$$\varphi_{\bar{\xi}}(\bar{t}) = \int_{\mathbb{R}^d} \exp(i\langle \bar{t}, \bar{x}\rangle) d\mathbb{P}_{\bar{\xi}}(\bar{x})$$
$$= \mathbb{E}\exp\big(i\langle \bar{t}, \bar{\xi}\rangle\big).$$

Note that

$$\varphi_{\bar{\xi}}(\bar{t}) = \int_{\mathbb{R}^d} \exp(i\langle \bar{t}, \bar{x}\rangle) p_{\bar{\xi}}(\bar{x}) d\bar{x}$$

if the density exists.

We remark that the same theory from random variables can be developed for random vectors.

There is a unique correspondance between distributions and their characteristic function.

For the sum of two independent random vectors, the characteristic function is the product of characteristic functions.

If A is a $d \times d$ matrix,

$$\varphi_{A\bar{\xi}+\bar{b}}(\bar{t}) = \exp\big(i\langle \bar{b}, \bar{t}\rangle\big)\varphi_{\bar{\xi}}(A^T\bar{t}).$$

If it exists,

$$\mathbb{E}(\xi_i^{k_i} : i \in d) = i^{\sum_{i=1}^d k_i} \frac{\partial^{\sum_{i=1}^d k_i} \varphi_{\bar{\xi}}}{\partial t_1^{k_1} \dots \partial t_d^{k_d}}(0)$$

The coordinates ξ_i are independent random variables if and only if

$$\varphi_{\bar{\xi}}(t) = \prod_{i=1}^{d} \varphi_{\xi_i}(t_i).$$

6.3 Limit Theorems

These results are constructed analogously.

Theorem 6.3.1 (Lévy)
Let $\mathbb{P}_n, \mathbb{P}$ be probability measures in $\mathbb{R}^d$. Then

$$\mathbb{P}_n \xrightarrow{w} \mathbb{P} \iff \forall t \in \mathbb{R}^d, \lim_n \varphi_n(t) = \varphi(t).$$

6.3.1 Law of Large Numbers

Theorem 6.3.2
Let $\bar{\xi}_i, i \in \mathbb{N}$ be iid random vectors with characteristic function $\varphi(\bar{t})$ such that

$$\frac{\partial \varphi}{\partial t_k}(0)$$

exists and is finite for all $k \in [d]$. Then

$$\frac{1}{n}\sum_{i=1}^{n} \bar{\xi}_i(\omega) \xrightarrow{p/d} \left(\frac{1}{i} \cdot \frac{\partial \varphi}{\partial t_k}(0)\right)_{k\in[d]}.$$

The proof is analogous to the 1-dimensional case. We use Taylor expansion and Lévy's theorem.

6.3.2 Central Limit Theorem

Theorem 6.3.3
Let $\bar{\xi}_i, i \in \mathbb{N}$ be iid random vectors with finite $\mathbb{E}\bar{\xi}$ and covariance matrix R. Then

$$\frac{\sum_{i=1}^{n} \bar{\xi}_i(\omega) - n\mathbb{E}\bar{\xi}}{\sqrt{n}} \xrightarrow{d} N(0, R).$$

Part III

Stochastic Processes

Chapter 7

Conditioning

7.1 Conditional Probability of Events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$ be with that $\mathbb{P}(B) > 0$.

Definition 7.1.1 (Conditional Probability of *A* given *B*)
We define

$$\mathbb{P}\{A \mid B\} := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Alternatively, we can write $\mathbb{P}_B(A)$.

Proposition 7.1.1
$\mathbb{P}_B$ is a probability measure on both $(\Omega, \mathcal{F})$ and $(B, F\big|_B)$.

Proposition 7.1.2

1) $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$
2) A, B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ or $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
3) $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i \mid A_1 \cap \ldots A_{i-1})$ (Multiplication Formula)

Proposition 7.1.3 (Law of Total Probability)
Let $A_1, A_2, \cdots \in \mathcal{F}$ be a countable partition of Ω. For every $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i)\mathbb{P}(B \mid A_i).$$

Proposition 7.1.4 (Baye's Formula)
Let $A_i \in \mathcal{F}, i \in \mathbb{N}$ be a countable partition of Ω. For $B \in \mathcal{F}$,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

7.2 Conditional Probability of Random Variables

Let ξ, η be random variables with joint distribution $\mathbb{P}_{\xi,\eta}$. How do we quantify the probability $\{\omega : \xi(\omega) = x \mid \eta = \dots\}$?

7.2.1 Discrete Random Variables

Suppose (ξ, η) takes on values $\{(x_i, y_i) : i \in \mathbb{N}\}$

Definition 7.2.1 (Conditional Distribution of Discrete Random Variables)
Let x_i be such that $\mathbb{P}\{\xi(\omega) = x_i\} > 0$. We define

$$\mathbb{P}_{\eta|\xi=x_i}(y) := \mathbb{P}\{\eta(\omega) = y \mid \xi(\omega) = x_i\}.$$

The conditional distribution function is

$$F_{\eta|\xi=x_i}(t) = \sum_{y_j \le t} \mathbb{P}_{\eta|\xi=x_i}(y_j).$$

Note that

$$\mathbb{P}_{\eta|\xi=x_i}(A) = \sum_j \mathbb{P}_{\eta|\xi=x_i}(y_j)\mathbb{1}\{y_j \in A\}$$

is a probability measure on $\mathbb{R}$.

Example 7.2.1
Consider N iid Bernoulli trials and let $\xi(\omega)$ be the result of the 1st trial and $\eta(\omega)$ the number of successes.

Then

$$\begin{aligned}\mathbb{P}_{\eta|\xi=1}(k) &= \mathbb{P}\{\eta(\omega) = k \mid \xi = 1\} \\ &= \frac{\binom{N-1}{k-1} p^k (1-p)^{N-k}}{p} \\ &= \binom{N-1}{k-1} p^{k-1}(1-p)^{(N-1)-(k-1)}\end{aligned}$$

which we realize is just another binomial distribution!

7.2.2 Absolutely Continuous Random Variables

Let $p_\xi(x), p_\eta(y)$ be marginal densities

$$\begin{aligned}p_\xi(x) &= F'(x) && \text{a.e.} \\ &= \frac{\partial}{\partial x}\int_{-\infty}^{x}\int_{-\infty}^{\infty} p_{\xi,\eta}(u,y)dydu \\ &= \int_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dy.\end{aligned}$$

Definition 7.2.2 (Conditional Density of Absolutely Continuous Random Variables)
Let $x \in \mathbb{R}$ be such that $p_\xi(x) > 0$. We define

$$p_{\eta|\xi=x}(y) = \frac{p_{\xi,\eta}(x,y)}{p_\xi(x)}.$$

The conditional distribution function is thus

$$F_{\eta|\xi=x}(t) := \int_{-\infty}^{\infty} p_{\eta|\xi=x}(y)dy.$$

Proposition 7.2.2
The conditional distribution of $\eta \mid \xi = x$ is a probability measure on $\mathbb{R}$.

$$\mathbb{P}_{\eta|\xi=x}(A) = \int_A p_{\eta|\xi=x}(y)dy.$$

Proof
Non-negativity is clear.

For normalization,

$$\begin{aligned}\mathbb{P}_{\eta|\xi=x}(\mathbb{R}) &= \int_{\mathbb{R}} \frac{p_{\xi,\eta}(x,y)}{p_\xi(x)} dy \\ &= \frac{1}{p_\xi(x)} \int_{\mathbb{R}} p_{\xi,\eta}(x,y) dy \\ &= \frac{1}{p_\xi(x)} p_\xi(x) \\ &= 1.\end{aligned}$$

Finally, for σ-additivity,

$$\begin{aligned}\mathbb{P}_{\eta|\xi=x}(\bigsqcup_i A_i) &= \int_{\bigsqcup_i A_i} p_{\eta|\xi=x}(y) dy \\ &= \sum_i \int_{A_i} p_{\eta|\xi=x}(y) dy \\ &= \sum_i \mathbb{P}(A_i).\end{aligned}$$

We remark here that we could have alternatively defined

$$\mathbb{P}_{\eta|\xi=x}(A) = \lim_{\epsilon \to 0^+} \mathbb{P}\{\eta \in A : |\xi(\omega) - x| < \epsilon\}.$$

Example 7.2.3
Let $(\xi, \eta) \sim U(B_1^2)$ be uniformly distributed across the unit sphere in $\mathbb{R}^2$. Then

$$\begin{aligned}p_{\xi,\eta}(x,y) &= \frac{1}{\pi} \mathbb{1}\{x^2 + y^2 \leq 1\} \\ p_\xi(x) &= \int_{-\sqrt{1-x^2}} \sqrt{1-x^2} \frac{1}{\pi} dy \\ &= \frac{2}{\pi} \sqrt{1-x^2} \\ p_{\eta|\xi=x}(y) &= \frac{p_{\xi,\eta}(x,y)}{p_\xi(x)} \\ &= \frac{\mathbb{1}\{y^2 \leq 1 - x^2\}}{2\sqrt{1-x^2}}.\end{aligned}$$

Note that $y \sim U\left[-\sqrt{1-x^2}, \sqrt{1-x^2}\right]$.

7.3 Conditional Expectation

Recall that we defined conditional probability for events but conditional probability of random variables for two special cases. We will see how to define conditional probability for random variables in general. It turns out that approaching this from the perspective of conditional expectation is quite natural.

7.3.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\xi : \Omega \to \mathbb{R}$ a random variable, and $\mathcal{G} \subseteq \mathcal{F}$ a σ-subalgebra of $\mathcal{F}$.

Definition 7.3.1 (Conditional Expectation)
The *conditional expectation of* ξ *with respect to* $\mathcal{G}$ is any random variable $\mathbb{E}\{\xi \mid \mathcal{G}\} : \Omega \to \mathbb{R}$ such that

1) $\mathbb{E}\{\xi \mid \mathcal{G}\}$ is $\mathcal{G}$-measurable.
2) $\int_A \xi(\omega) d\mathbb{P}(\omega) = \int_A \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega) d\mathbb{P}(\omega)$ for any $A \in \mathcal{G}$.

We should think of the conditional expectation as "an average over a coarser σ-subalgebra while preserving the value of the integral".

Example 7.3.1
Consider the probability space $([0,1], \mathcal{B}[0,1], \lambda)$ with the σ-subalgebra

$$\mathcal{G}_n := \sigma\left\{\left[\frac{k}{n}, \frac{k+1}{n}\right] : 0 \leq k < n\right\}.$$

Then

$$\mathbb{E}\{\xi \mid \mathcal{G}_n\} = \sum_{k=0}^{n-1} \left(\frac{k + (k+1)}{2n}\right) \mathbb{1}_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}.$$

If we did not require $\mathbb{E}\{\xi \mid \mathcal{G}\}$ to be $\mathcal{G}$-measurable, then we can simply take $\mathbb{E}\{\xi \mid \mathcal{G}\} = \xi$! This condition forces us "collapse the random variable" so that it "contains only as much information as permitted by $\mathcal{G}$".

Example 7.3.2
If $\mathcal{G} = \{\Omega, \varnothing\}$ is trivial, then $\mathbb{E}\{\xi \mid \mathcal{G}\}(\omega) = \mathbb{E}\xi$.

If $\mathcal{G} \supseteq \sigma(\xi)$, then we can take $\mathbb{E}\{\xi \mid \mathcal{G}\} = \xi(\omega)$ since that clearly satisfies the defining property.

7.3.2 Construction

We first remark that $\mathbb{E}\{\xi \mid \mathcal{G}\}$ is in fact an equivalence class of random variables satisfying the definiting conditions. In fact, it can be show that it is unique up to null sets of $\mathbb{P}$.

In order to rigorously construct the conditional expectation, we proceed similarly to the construction of the Lebesgue integral. We begin with $\xi \geq 0$ and define a general conditional expectation as

$$\mathbb{E}\{\xi \mid \mathcal{G}\} = \mathbb{E}\{\xi^+ \mid \mathcal{G}\} - \mathbb{E}\{\xi^- \mid \mathcal{G}\},$$

assuming that at least one of the two values are finite.

Theorem 7.3.3
For a non-negative random variable $\xi \geq 0$, $\mathbb{E}\{\xi \mid \mathcal{G}\}$ exists and is non-negative.

The outline of the proof is to define a measure Q on $(\Omega, \mathcal{G})$ which is absolutely continuous with respect to $\mathbb{P}\big|_{\mathcal{G}}$. Then, we show that we can take $\mathbb{E}\{\xi \mid \mathcal{G}\}$ to be the Radon-Nikodym derivative $\frac{dQ}{d\mathbb{P}}$.

Proof
Define

$$Q(A) := \int_A \xi(\omega) d\mathbb{P}(\omega)$$

for all $A \in \mathcal{G}$. We claim that Q is a measure on $(\Omega, \mathcal{G})$.

Indeed, it is clearly non-negative and 0 at $\varnothing$. We need only demonstrate σ-additivity.

$$\begin{aligned} Q(\bigsqcup_i A_i) &= \int_{\bigsqcup_i A_i} \xi(\omega) d\mathbb{P}(\omega) \\ &= \sum_i \int_{A_i} \xi(\omega) d\mathbb{P}(\omega) \qquad (\star) \\ &= \sum_i Q(A_i) \end{aligned}$$

$(\star)$ This equality is intuitive but requires proof, which we omit.

Furthermore, $\mathbb{P}\big|_{\mathcal{G}}(A) = 0$ implies that $Q(A) = 0$, hence we see that $Q << \mathbb{P}\big|_{\mathcal{G}}$.

Since the restriction a finite measure, we conclude by the Radon-Nikodym theorem that there exists some $\frac{dQ}{d\mathbb{P}}(\omega) \geq 0$ that is $\mathcal{G}$-measurable.

We can take $\mathbb{E}\{\xi \mid \mathcal{G}\} = \frac{dQ}{d\mathbb{P}}(\omega)$. Indeed,

$$\begin{aligned}\int_A \frac{dQ}{d\mathbb{P}}(\omega) d\mathbb{P}(\omega) &= \int_A dQ(\omega) \\ &= Q(A) \\ &= \int_A \xi(\omega) d\mathbb{P}(\omega).\end{aligned}$$

7.3.3 Conditional Probability

Consider the special case where $\xi = \mathbb{1}_B$ for some $B \in \mathcal{F}$.

Definition 7.3.2 (Conditional Probability)
The *conditional probability* $\mathbb{P}\{B \mid \mathcal{G}\} : \Omega \to [0,1]$ is given by

$$\mathbb{P}\{B \mid \mathcal{G}\} := \mathbb{E}\{\mathbb{1}_B \mid \mathcal{G}\}.$$

The following is a characterization of conditional property which we could have used as an alternative definition.

Proposition 7.3.4
$\mathbb{P}\{B \mid \mathcal{G}\} : \Omega \to [0,1]$ is any random variable that is

1) $\mathcal{G}$-measurable
2) $\mathbb{P}(A \cap B) = \int_A \mathbb{P}\{B \mid \mathcal{G}\} d\mathbb{P}(\omega)$ for any $A \in \mathcal{G}$.

Example 7.3.5
If $\mathcal{G}$ is trivial, then $\mathbb{P}\{A \mid \mathcal{G}\} = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$.

If $G = \sigma(B)$ for some $B \in \mathcal{F}$, then

$$\mathbb{P}(A \mid \sigma(B))(\omega) = \begin{cases} \mathbb{P}(A \mid B), & \omega \in B \\ \mathbb{P}(A \mid B^c), & \omega \notin B \end{cases}$$

If $B_1, \ldots, B_n$ is a partition of Ω,

$$\mathbb{P}\{A \mid \sigma(B_1, \ldots, B_n)\} = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i) \mathbb{1}_{B_i}.$$

7.3.4 Properties

Lemma 7.3.6
Let ξ, η be $\mathcal{G}$-measurable random variables. If

$$\int_A \xi(\omega) d\mathbb{P}(\omega) = \int_A \eta(\omega) d\mathbb{P}(\omega)$$

for every $A \in \mathcal{G}$, then $\xi = \eta$ a.e.
Note here we mean that $\{\omega : \xi(\omega) \neq \eta(\omega)\} \in \mathcal{G}$ and has measure zero under $\mathbb{P}\big|_{\mathcal{G}}$.

Most of these properties are proven using the definiting conditions as well as the lemma above.

Proposition 7.3.7

a) $\mathbb{E}\{\xi \equiv c \mid \mathcal{G}\}(\omega) = c$ a.s. for all $c \in \mathbb{R}$

b) $\{a\xi + b\eta \mid \mathcal{G}\}(\omega) = a\mathbb{E}\{\xi \mid \mathcal{G}\} + b\mathbb{E}\{\eta \mid \mathcal{G}\}$ a.s.

c) If $\xi \leq \eta$ a.s., then $\mathbb{E}\{\xi \mid \mathcal{G}\}(\omega) \leq \mathbb{E}\{\xi \mid \mathcal{G}\}$ a.s.

d) $|\mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)| \leq \mathbb{E}\{|\xi| \mid \mathcal{G}\}(\omega)$

e) $\mathbb{E}\{\xi \mid \mathcal{F}\} = \xi$ a.s.

f) $\mathbb{E}\{\xi \mid \{\Omega, \varnothing\}\} = \mathbb{E}\xi$ a.s.

g) $\mathbb{E}\left[\mathbb{E}\{\xi \mid \mathcal{G}\}\right] = \mathbb{E}\xi$

h) If $\sigma(\xi), \mathcal{G}$ are independent, then $\mathbb{E}\{\xi \mid \mathcal{G}\} = \mathbb{E}\xi$ a.s.

Proof (h)
For any $A \in \mathcal{G}$,

$$\begin{aligned}
\int_A \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega) d\mathbb{P}(\omega) &= \int_A \xi(\omega)\mathbb{P}(\omega) \\
&= \mathbb{E}[\xi \mathbb{1}_A] \\
&= \mathbb{E}[\xi] \cdot \mathbb{E}[\mathbb{1}_A] && \text{independence} \\
&= \mathbb{E}\xi \cdot \mathbb{P}(A) \\
&= \int_A \mathbb{E}\xi d\mathbb{P}(\omega).
\end{aligned}$$

Then we apply the useful lemma.

Proposition 7.3.8
If η is $\mathcal{G}$-measurable, then $\mathbb{E}\{\xi\eta \mid \mathcal{G}\}(\omega) = \eta\mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)$.

Proof
We will show that case where $\xi \geq 0$ and $\eta = \mathbb{1}_B$ for some $B \in \mathcal{G}$. The rest proceeds as in the construction of the Lebesgue integral. For every $A \in \mathcal{G}$,

$$\begin{aligned}\int_A \eta(\omega)\mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)d\mathbb{P}(\omega) &= \int_{A\cap B} \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)d\mathbb{P}(\omega) \\ &= \int_{A\cap B} \xi(\omega)d\mathbb{P}(\omega) \\ &= \int_A \eta(\omega)\xi(\omega)d\mathbb{P}(\omega).\end{aligned}$$

Lemma 7.3.9 (Smoothing)
For $\mathcal{G}_1 \subseteq \mathcal{G}_2$,

$$\mathbb{E}\{\mathbb{E}\{\xi \mid \mathcal{G}_1\} \mid \mathcal{G}_2\} = \mathbb{E}\{\xi \mid G_1\} = \mathbb{E}\{\mathbb{E}\{\xi \mid G_2\} \mid \mathcal{G}_1\}$$

a.s.

Proof
The first equality is easy since $\mathbb{E}\{\xi \mid \mathcal{G}_1\}$ is automatically $\mathcal{G}_2$-measurable. To see the second equality, we apply the defining relation multiple times. Indeed, for any $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$,

$$\begin{aligned}\int_A \mathbb{E}\{\mathbb{E}\{\xi \mid \mathcal{G}_2\} \mid \mathcal{G}_1\}(\omega)d\mathbb{P}(\omega) &= \int_A \mathbb{E}\{\xi \mid \mathcal{G}_2\}d\mathbb{P}(\omega) \\ &= \int_A \xi d\mathbb{P}(\omega) \\ &= \int_A \mathbb{E}\{\xi \mid \mathcal{G}_1\}d\mathbb{P}(\omega).\end{aligned}$$

Then, an application our the ever useful lemma terminates the proof.

We now proceed to limiting properties.

Proposition 7.3.10

1) If $\xi_n \uparrow \xi$ as $n \to \infty$, then

$$\mathbb{E}\{\xi_n \mid \mathcal{G}\}(\omega) \uparrow \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)$$

a.s.

2) If $|\xi_n| \leq \eta \in L^1$ and $\xi_n \to \xi$ a.s., then

$$\mathbb{E}\{\xi_n \mid \mathcal{G}\}(\omega) \to \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega)$$

a.s.

Proof (1)

We remark that $\mathbb{E}\{\xi_n \mid \mathcal{G}\}(\omega)$ is monotonically increasing and thus the limit $Z(\omega)$ exists. For every $A \in \mathcal{G}$, we have

$$\begin{aligned}
\int_A \lim_n \mathbb{E}\{\xi_n \mid \mathcal{G}\}(\omega) d\mathbb{P}(\omega) &= \lim_n \int_A \mathbb{E}\{\xi_n \mid \mathcal{G}\}(\omega) d\mathbb{P}(\omega) && \text{MCT} \\
&= \lim_n \int_A \xi_n(\omega) d\mathbb{P}(\omega) \\
&= \int_A \lim_n \xi_n(\omega) d\mathbb{P}(\omega) && \text{MCT} \\
&= \int_A \xi(\omega) d\mathbb{P}(\omega) \\
&= \int_A \mathbb{E}\{\xi \mid \mathcal{G}\}(\omega) d\mathbb{P}(\omega).
\end{aligned}$$

Yet another application of the useful lemma completes the proof.

Chapter 8

Martingales

The fundamental theory of martingales was developed by Doob.

8.1 Introduction

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

Definition 8.1.1 (Filtration / Flow)
An increasing sequence of sub σ-algebras

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$$

is a *filteration* or *flow*.

We sometimes write

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{n\geq 1} \mathcal{F}_n\right).$$

Example 8.1.1 (Natural Filtration)
If $\{\xi_n\}_{n\geq 1}$ is a sequence of random variables, the sequence given by

$$\mathcal{F}_n := \sigma\{\xi_i\}_{i=1}^n$$

is a filteration known as the *natural filtration*.

Definition 8.1.2
A sequence of random variables $X_n, n \geq 1$ is *adapted* to the filtration $\mathcal{F}_n, n \geq 1$, if X_n is $\mathcal{F}_n$-measurable for all $n \geq 1$.

Definition 8.1.3 (Predictable)
We say the sequence X_n is predictable with respect to $\mathcal{F}_n$ if X_{n+1} is $\mathcal{F}_{n-1}$-measurable.

Example 8.1.2
For a measurable function F,

$$F_n(\omega) := F(\xi_1(\omega), \dots, \xi_n(\omega))$$

is $\{\mathcal{F}_n\}$ adapted.

For example, F_n could be the sum or product of its arguments.

Definition 8.1.4 (Martingale)
Let $\{X_n\}$ be adapted with respect to $\{\mathcal{F}_n\}$ such that $\mathbb{E}|X_n| < \infty$. Then $\{X_n\}$ is a *martingale* sequence if

$$\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{=} X_n(\omega)$$

Definition 8.1.5 (Submartingale)
Let $\{X_n\}$ be adapted with respect to $\{\mathcal{F}_n\}$ such that $\mathbb{E}|X_n| < \infty$. Then $\{X_n\}$ is a *submartingale* sequence if

$$\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{\geq} X_n(\omega)$$

Definition 8.1.6 (Submartingale)
Let $\{X_n\}$ be adapted with respect to $\{\mathcal{F}_n\}$ such that $\mathbb{E}|X_n| < \infty$. Then $\{X_n\}$ is a *supermartingale* sequence if

$$\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{\leq} X_n(\omega)$$

Note that the definition of a martingales require an instance of a filtration. If left unspecified, we typically consider the natural filtration.

8.2 Properties

Proposition 8.2.1

1) $\{X_n\}$ is $\{\mathcal{F}_n\}$-martingale if and only if $\mathbb{E}\{X_{n+k} \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{=} X_n(\omega)$ for all $k \geq 1$.
2) $\mathbb{E}X_n = \mathbb{E}X_{n+k}$ for all $k \geq 1$.

Proof

1) ($\Longleftarrow$) is by definition. We argue ($\Longrightarrow$) by induction on k.

The base $k = 1$ is by assumption. Then,

$$\begin{aligned}
\mathbb{E}\{X_{n+k+1} \mid \mathcal{F}_n\}(\omega) &= \mathbb{E}\{\mathbb{E}\{X_{n+k+1} \mid \mathcal{F}_{n+1}\} \mid \mathcal{F}_n\}(\omega) \\
&= \mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}(\omega) && \text{induction hypothesis} \\
&= X_n(\omega).
\end{aligned}$$

2) We have

$$\begin{aligned}
\mathbb{E}X_n &= \mathbb{E}\left(\mathbb{E}\{X_{n+k} \mid \mathcal{F}_n\}\right) && 1) \\
&= \mathbb{E}X_{n+k} && \text{defining relation}
\end{aligned}$$

We can think of a martingale sequence as a "fair game", submartingale sequence as a "favorable game", and supermartingale as an "unfavorable game".

Proposition 8.2.2

Let $\xi_1, \xi_2, \ldots$ be independent random variables such that $\mathbb{E}|\xi_i| < \infty, \mathbb{E}\xi_i = 0$. Then

$$S_n := \sum_{i=1}^{n} \xi_i$$

forms a martingale.

Proof

Clearly S_n is $\mathcal{F}_n$-measurable under the natural filtration. Moreover,

$$\begin{aligned}
\mathbb{E}\{S_{n+1} \mid \mathcal{F}_n\} &= \mathbb{E}\{S_n + \xi_{n+1} \mid \mathcal{F}_n\} \\
&= S_n + \mathbb{E}\{\xi_{n+1} \mid \mathcal{F}_n\} \\
&= S_n + \mathbb{E}\xi_{n+1} && \sigma(\xi_{n+1}), \mathcal{F}_n \text{ are independent} \\
&= S_n.
\end{aligned}$$

Example 8.2.3
If ξ_i are the steps of a symmetric random walk, then it is a martingale.

Proposition 8.2.4
Let $\eta_1, \eta_2, \dots$ be independent random variables such that $\mathbb{E}|\eta_i| < \infty, \mathbb{E}\eta_i = 1$. Then

$$X_n := \prod_{i=1}^{n} \xi_i$$

forms a martingale.

Proof
X_n is clearly adapted to the natural filtration. Moreover,

$$\begin{aligned}
\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\} &= \mathbb{E}\{X_n \eta_{n+1} \mid \mathcal{F}_n\} \\
&= X_n \mathbb{E}\{\eta_{n+1} \mid \mathcal{F}_n\} && X_n \text{ is } \mathcal{F}_n\text{-measurable} \\
&= X_n \mathbb{E}\eta_{n+1} \\
&= X_n
\end{aligned}$$

Example 8.2.5
We begin the game with \$1 and each round, we either double our wealth or lose everything. Then if $\eta_i \sim 2\,\mathrm{Be}\left(\frac{1}{2}\right)$,

$$X_n := \prod_{i=1}^{n} \eta_i$$

is the random variable modeling our wealth after n rounds.

From our work above, X_n is martingale.

Example 8.2.6 (Doob's Martingale)
Let ξ be an arbitrary integrable random variable and $\{\mathcal{F}_n\}$ a given filtration.

Then

$$X_n := \mathbb{E}\{\xi \mid \mathcal{F}_n\}(\omega) \in \mathcal{F}_n$$

is known as *Doob's martingale.*

Proposition 8.2.7
$\{X_n\}$ is submartingale if and only if $\{-X_n\}$ is supermartingale.

Proposition 8.2.8
If $\{X_n\}, \{Y_n\}$ are martingales with respect to $\{\mathcal{F}_n\}$,

a) $aX_n + bY_n$ is martingale

b) $\max(X_n, Y_n)$ is submartingale

c) $\min(X_n, Y_n)$ is supermartingale

Proposition 8.2.9
Suppose $\{X_n\}$ is $\{\mathcal{F}_n\}$-submartingale and $g : \mathbb{R} \to \mathbb{R}$ is convex and increasing. Then $\{g(X_n)\}$ is submartingale.

Proof
We have

$$\begin{aligned} g(X_n) &\leq g(\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}) && \text{submartingale definition} \\ &\leq \mathbb{E}\{g(X_{n+1}) \mid \mathcal{F}_n\}. && \text{Jensen's inequality} \end{aligned}$$

Proposition 8.2.10

1) If $\{X_n\}$ forms a non-negative submartingale, then $\{X_n^\alpha\}$ is submartingale for any $\alpha \geq 1$.

2) If X_n is martingale, then $|X_n|^\alpha$ is submartingale for all $\alpha \geq 1$.

Proof (2))
We have

$$\begin{aligned} \mathbb{E}\{|X_{n+1}| \mid \mathcal{F}_n\} &\geq |\mathbb{E}\{X_{n+1} \mid \mathcal{F}_n\}| \\ &= |X_n|. \end{aligned}$$

So $|X_n|$ is a non-negative martingale and we can apply 1).

Theorem 8.2.11 (Doob's Maximal Inequality)
Let $\{X_k\}_{k=1}^n$ be a non-negative $\{\mathcal{F}_n\}$-submartingale. Then

$$\mathbb{P}\left\{\max_{k\in[n]} X_k \geq t\right\} \leq \frac{1}{t}\mathbb{E}X_n$$

for all $t > 0$.

Proof
Define

$$A_k := \{\omega : X_1(\omega), \dots, X_{k-1}(\omega) < t, X_k(\omega) \geq t\}.$$

Then

$$\begin{aligned} A &:= \left\{\omega : \max_{k\in[n]} X_k(\omega) \geq t\right\} \\ &= \bigsqcup_{k=1}^{n} A_k. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}X_n &\geq \int_A X_n(\omega) d\mathbb{P}(\omega) \\ &= \sum_{k=1}^{n} \int_{A_k} X_n(\omega) d\mathbb{P}(\omega) \\ &= \sum_{k=1}^{n} \int_{A_k} \mathbb{E}\{X_n \mid \mathcal{F}_k\}(\omega) d\mathbb{P}(\omega) && A_k \in \mathcal{F}_k \\ &\geq \sum_{k=1}^{n} \int_{A_k} X_k(\omega) d\mathbb{P}(\omega) && \text{submartingale definition} \\ &\geq \sum_{k=1}^{n} \int_{A_k} t \cdot d\mathbb{P}(\omega) \\ &= t\mathbb{P}(A). \end{aligned}$$

Corollary 8.2.11.1 (Kolmogorov's Inequality)
Let $\xi_1, \xi_2, \dots$ be independent random variables with zero mean and finite variance. Then

$$S_k := \sum_{i=1}^{k} \xi_k$$

is martingale and so S_k^2 is submartingale by the properites of submartingales.
It follows that

$$\begin{aligned} \mathbb{P}\left\{\max_{k\in[n]} |S_k| \geq \epsilon\right\} &= \mathbb{P}\left\{\max_{k\in[n]} |S_k|^2 \geq \epsilon^2\right\} \\ &\leq \frac{\mathbb{E}[S_n^2]}{\epsilon^2}. \end{aligned}$$

8.3 Martingale Differences

Definition 8.3.1 (Martingale Difference)
Let $\{\xi_n\}$ be $\{\mathcal{F}_n\}$-adapted and integrable. Set $\xi_0 \equiv 0$ for convenience.
$\{\xi_n\}$ is a *martingale difference* if

$$\mathbb{E}\{\xi_{n+1} \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{\equiv} 0$$

for all n.

Similarly we define *submartingale difference* if

$$\mathbb{E}\{\xi_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\geq} 0$$

for all n and *supermartingale difference* if

$$\mathbb{E}\{\xi_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\leq} 0$$

for all n.

Proposition 8.3.1

a) If $\{X_n\}$ is $\{\mathcal{F}_n\}$-martingale, then $\xi_n := X_n - X_{n-1}, \xi_0 \equiv 0$ is a martingale difference.

b) If ξ_n is a martingale difference with respect to $\{\mathcal{F}_n\}$, then $X_n := \sum_{i=1}^{n} \xi_n$ is martingale

c) Similar statements hold for submartingale and super martingale differences.

Proposition 8.3.2 (Orthogonality for Martingale Differences)
Let $\{X_n\}$ be $\{\mathcal{F}_n\}$-martingale and $\xi_n := X_n - X_{n-1}$. Then

$$\mathbb{E}\left[\xi_n \xi_m\right] = \begin{cases} \mathbb{E}[\xi_n^2], & m = n \\ 0, & m \neq n \end{cases}$$

Proof
If $m < n$,

$$\begin{aligned} \mathbb{E}\left[\xi_n \xi_m\right] &= \mathbb{E}\left[\mathbb{E}\{\xi_m \xi_n \mid \mathcal{F}_m\}\right] \\ &= \mathbb{E}\left[\xi_m \cdot \mathbb{E}\{\xi_n \mid \mathcal{F}_m\}\right] \\ &= \mathbb{E}\left[\xi_m \cdot 0\right] \\ &= 0. \end{aligned}$$

Theorem 8.3.3 (Doob's Decomposition)
Let $\{X_n\}$ be $\{\mathcal{F}_n\}$-submartingale. Then X_n uniquely decomposes as

$$X_n = M_n + A_n$$

where M_n is martingale and the *compensator* A_n is predictable and increasing a.s. That is, $A_{n+1} \overset{a.s.}{\geq} A_n$ with $A_0 \equiv 0$.

Proof
Let $M_0 := X_0$ and $A_0 \equiv 0$.

For $n \geq 1$, recursively define

$$\begin{aligned} M_n &= M_{n-1} + X_n - \mathbb{E}\{X_n \mid \mathcal{F}_{n-1}\} \\ &= M_0 + \sum_{i=1}^{n} X_i - \mathbb{E}\{X_i \mid \mathcal{F}_{i-1}\} \\ &\in \mathcal{F}_n. \end{aligned}$$

We claim that M_n is martingale. Indeed, M_n is the sum of integrable random variables and is thus integrable. Moreover,

$$\begin{aligned} \mathbb{E}\{M_n \mid \mathcal{F}_{n-1}\} &= M_{n-1} + \mathbb{E}\{X_n \mid \mathcal{F}_{n-1}\} - \mathbb{E}\{\mathbb{E}\{X_i \mid \mathcal{F}_{n-1}\} \mid \mathcal{F}_{n-1}\} \\ &= M_{n-1}. \end{aligned}$$

Next, recursively define the a.s. increasing sequence

$$\begin{aligned} A_n &:= A_{n-1} + \underbrace{\mathbb{E}\{X_n \mid \mathcal{F}_{n-1}\} - X_{n-1}}_{\overset{a.s.}{\geq} 0} \\ &= \sum_{j=1}^{n} \mathbb{E}\{X_j \mid \mathcal{F}_{j-1}\} - X_{j-1} \\ &\in \mathcal{F}_{n-1}. \end{aligned}$$

We have

$$\begin{aligned} M_n + A_n &= X_0 + \sum_{j=1}^{n} (X_j - X_{n-1}) \\ &= X_n \end{aligned}$$

as desired.

Uniquness can be shown by taking the conditional expectation of two candidates.

Let $M_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ be martingale. Then $M_n = M_0 = \sum_{j=1}^n \xi_j$ where ξ_j is a martingale difference. By Doob's decomposition, we have

$$M_n^2 = m_n + \langle M_n \rangle$$

where m_n is martingale and $\langle M_n \rangle$ is known as the *quadratic variation of* M_n.

Proposition 8.3.4
We have

$$\langle M_n \rangle = \sum_{j=1}^n \mathbb{E}\{\xi_j^2 \mid \mathcal{F}_{j-1}\}.$$

Example 8.3.5
Suppose $M_n = \sum_{i=1}^n \xi_i$ for some independent zero-mean random variables ξ_i with finite variance. Then

$$\begin{aligned} \langle M_n \rangle &= \sum_{j=1}^n \mathbb{E}\{\xi_j^2 \mid \mathcal{F}_{j-1}\} \\ &= \sum_{j=1}^n \mathbb{E}\xi_j^2 \\ &= \operatorname{Var} M_n. \end{aligned}$$

Remark that $M_n^2 - \operatorname{Var}[M_n]$ is thus a martingale.

Chapter 9

Stopping

9.1 Stopping

Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$ be a filtration.

Definition 9.1.1 (Stopping Time)
A random variables $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ is a *stopping time* if for every $n \in \mathbb{N}$,

$$\{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n.$$

Proposition 9.1.1
The following are equivalent:

(1) $\forall n, \{\omega : \tau \leq n\} \in \mathcal{F}_n$

(2) $\forall n, \{\omega : \tau = n\} \in \mathcal{F}_n$

(3) $\forall n, \{\omega : \tau > n\} \in \mathcal{F}_n$

Proof
$\underline{(1) \iff (3)}$ Observe that

$$\{\tau \leq n\} \in \mathcal{F}_n \iff \overline{\{\tau \leq n\}} = \{\tau > n\} \in \mathcal{F}_n.$$

$\underline{(1) \iff (2)}$ We have

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$$
$$\{\tau \leq n\} = \bigcup_{k=1}^{n} \{\tau = k\}.$$

Proposition 9.1.2

1) $\tau \equiv N$ for some constant $N \in \mathbb{N}$ is a stopping time.
2) If τ, σ are stopping times, then so are their max and mins.
3) The sum of stopping times is a stopping time.
4) If $\tau_k \downarrow \tau$ is a decreasing sequence of stopping times, then τ is a stopping time.

Proof

1) We have

$$\{\tau \leq n\} = \begin{cases} \Omega, & n \geq N \\ \varnothing, & n < N \end{cases}$$

2) For every $n \in \mathbb{N}$,

$$\{\max(\tau, \sigma) \leq n\} = \{\tau \leq n\} \cap \{\sigma \leq n\}.$$

For the minimum, we take the union.

3)

$$\{\tau + \sigma \leq n\} = \bigcup_{k=1}^{n} \{\tau \leq k\} \cap \{\sigma \leq n - k\}.$$

Proposition 9.1.3 (Pre-τ σ-Algebra)

Let τ be a stopping time. Then the following is a σ-algebra

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty : \forall n, A \cap \{\tau \leq n\} \in \mathcal{F}_n\}.$$

We may also take $\mathcal{F}$ in place of $\mathcal{F}_\infty$.

Proof

Firstly, we have that $\Omega \cap \{\tau \leq n\} = \{\tau \leq n\} \in \mathcal{F}_n$ since τ is a stopping time.

Next, for all $A \in \mathcal{F}_\tau$,

$$\bar{A} \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus (A \cap \{\tau \leq n\}) \in \mathcal{F}_n.$$

Thus $\mathcal{F}_\tau$ is closed under complements.

Finally, let $A_k \in \mathcal{F}_\tau$. We have

$$\left(\bigcup_{k=1}^{\infty} A_k\right) \cap \{\tau \leq n\} = \bigcup_{k=1}^{\infty} (A_k \cap \{\tau \leq n\}) \in \mathcal{F}_n.$$

Hence $\mathcal{F}_\tau$ is closed under countable unions.

Proposition 9.1.4

1) $\tau \equiv n \implies \mathcal{F}_\tau = \mathcal{F}_N$
2) If $\sigma \overset{a.s.}{\leq} \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$
3) τ is $\mathcal{F}_\tau$-measurable

Proof

1) For any $A \in \mathcal{F}$, $A \cap \{\tau \leq n\} = \varnothing$ if $n < N$ and $A \cap \{\tau \leq n\} = A$ if $n \geq N$.

2) Suppose that for every $n \in \mathbb{N}$, $A \cap \{\sigma \leq n\} \in \mathcal{F}_n$. Then

$$\begin{aligned} A \cap \{\tau = n\} &= A \cap \{\sigma \leq \tau\} \cap \{\tau = n\} \\ &= A \cap \{\sigma \leq n\} \\ &\in \mathcal{F}_n. \end{aligned}$$

But then

$$A \cap \{\tau \leq n\} = \bigcup_{k=1}^{\infty} A \cap \{\tau = k\} \in \mathcal{F}_n$$

and $A \in \mathcal{F}_\tau$ as desired.

3) For any $s \in \mathbb{R}$, we need to show that $\tau^{-1}(-\infty, s] \in \mathcal{F}_\tau$. Indeed,

$$\begin{aligned} \{\tau \leq s\} \cap \{\tau \leq n\} &= \{\tau \leq \min(\lfloor s \rfloor, n)\} \\ &\in \mathcal{F}_n. \end{aligned}$$

9.2 Stopping a "Process"

Let $\{X_n\}_{n\geq 0}$ be $\{\mathcal{F}_n\}_{n\geq 0}$-adapted and τ a stopping time.

Example 9.2.1 (First Hitting Time)
τ_B given by

$$\omega \mapsto \min\{n : X_n(\omega) \in B\}$$

is a stopping time.

Indeed,

$$\begin{aligned} \{\tau_B \leq n\} &= \bigcup_{k=1}^{n} \{\omega : \forall j < k, X_j(\omega) \notin B \wedge X_k(\omega) \in B\} \\ &\in \mathcal{F}_n. \end{aligned}$$

In general,

$$\tau_B(\omega) := \max\{0 \le n \le N : X_n \notin B\}$$

is NOT a stopping time. Consider the natural filtration and remark it depends on the "future".

Proposition 9.2.2 (Configuration of X_n at τ)
Define

$$X_\tau(\omega) := \sum_{n=0}^{\infty} X_n \mathbb{1}\{\tau(\omega) = n\}.$$

Then X_τ is a $\mathcal{F}_\tau$-measurable random variable.

Proof
For every $n \in \mathbb{N}$,

$$\{\omega : X_\tau(\omega) \in B\} \cap \{\tau \le n\} = \bigcup_{k=1}^{n} \{\omega : X_k(\omega) \in B, \tau = k\} \in \mathcal{F}_n.$$

Theorem 9.2.3 (Doob Optimal Sampling / Stopping)
Let $\{X_n\}_{n=1}^M$ be a $\{\mathcal{F}_n\}$-martingale (sub, super). Let

$$\tau_1 \overset{a.s.}{\le} \tau_2 \overset{a.s.}{\le} \dots \overset{a.s.}{\le} \tau_p$$

be stopping times taking values from 1 to M.
Then

$$(X_{\tau_1}, F_{\tau_1}), \dots, (X_{\tau_p}, \mathcal{F}_{\tau_p})$$

is martingale (sub, super).
Remark that by taking $\tau_0 \equiv 1, \tau_{p+1} \equiv M$, we equivalently have

$$(X_{\tau_0}, \mathcal{F}_{\tau_0}), (X_{\tau_1}, \mathcal{F}_{\tau_1}), \dots, (X_{\tau_p}, \mathcal{F}_{\tau_p}), (X_{\tau_{p+1}}, \mathcal{F}_{\tau_{p+1}})$$

being martingale (sub, super).
In particular,

$$\mathbb{E}X_{\tau_0} = \mathbb{E}X_{\tau_1} = \cdots = \mathbb{E}X_{\tau_p} = \mathbb{E}X_{\tau_{p+1}}$$

with $\overset{a.s.}{\le}, \overset{a.s.}{\ge}$ holding in the case of sub, supermartingales, respectively.

Proof
We have already shown that $X_{\tau_i} \in \mathcal{F}_{\tau_i}$.

Next, we have that

$$\begin{aligned}
\mathbb{E}|X_\tau| &= \sum_{k=1}^{M} \int_{\tau=k} |X_k| d\mathbb{P} \\
&\leq \sum_{k=1}^{m} \mathbb{E}|X_k| \\
&< \infty. && X_k \text{ is martingale}
\end{aligned}$$

Finally, we need to show that for stopping times $\sigma \leq \tau$, $X_\sigma = \mathbb{E}\{X_\tau \mid \mathcal{F}_\sigma\}$. This holds if and only if for every $A \in \mathcal{F}_\sigma$,

$$\begin{aligned}
\int_A X_\sigma d\mathbb{P} &= \int_A \mathbb{E}\{X_\tau \mid \mathcal{F}_\sigma\} d\mathbb{P} && \text{defining relation} \\
&= \int_A X_\tau d\mathbb{P}. && A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_\tau
\end{aligned}$$

First, consider the simpler case where

$$0 \overset{a.s.}{\leq} \tau - \sigma \overset{a.s.}{\leq} 1.$$

For every $A \in \mathcal{F}_\tau$, we have

$$\begin{aligned}
&\int_A (X_\tau - X_\sigma) d\mathbb{P} \\
&= \sum_{k=1}^{M} \int_{A\cap\{\sigma=k\}\cap\{\tau=k+1\}} X_{k+1} - X_k d\mathbb{P} \\
&= \sum_{k=1}^{M} \int_{A\cap\{\sigma=k\}\cap\{\tau>k\}} X_{k+1} - X_k d\mathbb{P} && 0 \overset{a.s.}{\leq} \tau - \sigma \overset{a.s.}{\leq} 1 \\
&= \sum_{k=1}^{M} \int_{A\cap\{\sigma=k\}\cap\{\tau>k\}} \mathbb{E}\{X_{k+1} - X_k \mid \mathcal{F}_k\} d\mathbb{P} && (A \cap \{\sigma = k\}), \{\tau > k\} \in \mathcal{F}_k \\
&\begin{cases} = 0, & \{X_n\} \text{ is martingale} \\ \geq 0, & \{X_n\} \text{ is submartingale} \\ \leq 0, & \{X_n\} \text{ is supermartingale} \end{cases}
\end{aligned}$$

Now, for the general case of $\sigma \overset{a.s.}{\leq} \tau$, we can inject stopping times

$$\sigma = \rho_0 \overset{a.s.}{\leq} \rho_1 \overset{a.s.}{\leq} \dots \overset{a.s.}{\leq} \rho_M = \tau,$$

satisfying $0 \le \rho_j - \rho_{j-1} \le 1$ a.s. For instance, we can take

$$\rho_k := \min(\sigma + k, \tau)$$

for $k = 0, 1, \ldots, M$, which we know to be stopping times.

Then for every $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_{\rho_1} \subseteq \ldots \subseteq \mathcal{F}_{\rho_\tau}$,

$$\begin{aligned}\int_A X_\sigma d\mathbb{P} &= \int_A X_{\rho_1} d\mathbb{P} \\ &= \int_A X_{\rho_2} d\mathbb{P} \\ &= \ldots \\ &= \int_A X_\tau d\mathbb{P}\end{aligned}$$

with $\le, \ge$ holding for the sub, supermartingale cases, respectively.

We note that there is a converse of Doob's theorem.

Proposition 9.2.4
For a $\{\mathcal{F}_n\}$-adapted sequence $\{X_n\}$, X_n is martingale if and only if for all bounded stopping times τ,

$$\mathbb{E}X_\tau = \mathbb{E}X_1.$$

We also have the same stopping theorem without the bounded stopping time assumption, but now with a restriction on the increments.

Theorem 9.2.5
Let $\{X_n\}$ be $\{\mathcal{F}_n := \sigma(X_1, \ldots, X_n)\}$-martingale and τ a stopping time. If

1) $\mathbb{E}\tau < \infty$

2) there is some $c \in \mathbb{R}$ such that $\mathbb{E}\{|X_{n+1} - X_n| \mid \mathcal{F}_n\}(\omega) \overset{a.s.}{\le} c$ for all $\omega \in \{\tau(\omega) > n\}$

then

$$(X_1, \mathcal{F}_1), (X_\tau, \mathcal{F}_\tau)$$

is a martingale and in particular,

$$\mathbb{E}X_1 = \mathbb{E}X_\tau.$$

Definition 9.2.1 (Stopped Sequence / Process)
We define

$$X_n^\tau(\omega) := X_{\tau(\omega)\wedge n}(\omega) \in \mathcal{F}_n.$$

Proposition 9.2.6
If $(X_n, \mathcal{F}_n)$ is martingale, then $(X_n^\tau, \mathcal{F}_n)$ is also martingale.

Chapter 10

Convergence of Martingales

10.1 Convergence of Martingales

10.1.1 Upcrossing Inequality

Lemma 10.1.1 (Upcrossing Inequality)
Let $(X_n, \mathcal{F}_n)$ be submartingale.
Define $B_n(a,b)$ to be the number of upward crossings of $[a,b]$ by trajectories of $\{X_n\}$.
That is, define

$$\begin{aligned}
\tau_0(\omega) &:= 0 \\
\tau_1(\omega) &:= \min\{k \geq \tau_0(\omega) : X_k(\omega) \leq a\} \\
\tau_2(\omega) &:= \min\{k \geq \tau_1(\omega) : X_k(\omega) \geq b\} \\
&\dots \\
\tau_{2m-1}(\omega) &:= \min\{k \geq \tau_{2m-2}(\omega) : X_k(\omega) \leq a\} \\
\tau_{2m}(\omega) &:= \min\{k \geq \tau_{2m-1}(\omega) : X_k(\omega) \geq b\} \\
B_n(a,b)(\omega) &:= \max\{k : \tau_{2k}(\omega) \leq n\}.
\end{aligned}$$

Then

$$\mathbb{E}B_n(a,b) \leq \frac{\mathbb{E}(X_n - a)^+}{b-a}.$$

Proof
For $k > 0$, define

$$\tilde{X}_k(\omega) := (X_k - a)^+ \geq 0$$

and remark that convex increasing functions preserves submartingales.

For the sake of convenience, define $\tilde{b} := b - a$ and note that it suffices to show that

$$\mathbb{E}B_n(0, \tilde{b}) \le \frac{\mathbb{E}\tilde{X}_n}{\tilde{b}}.$$

Define

$$\varphi_i(\omega) := \begin{cases} 1, & \exists 2 \nmid m, \tau_m(\omega) < i \le \tau_{m+1}(\omega) \\ 0, & \exists 2 \mid m, \tau_m(\omega) < i \le \tau_{m+1}(\omega) \end{cases}$$

to be the indicator variable that i lies on an increasing trajectory.

Observe that for every ω,

$$\tilde{b} \cdot B_n(0, \tilde{b})(\omega) \le \sum_{i=1}^{n} \varphi_i(\omega)(\tilde{X}_i - \tilde{X}_{i-1})(\omega).$$

That is, each crossing must "travel" a distance of at least $\tilde{b}$.

Now, we have

$$\begin{aligned} \{\varphi_i(\omega) = 1\} &= \bigcup_{m: 2 \nmid m} \underbrace{\{\tau_m < i\}}_{\in \mathcal{F}_{i-1}} \setminus \underbrace{\{\tau_{m+1}\} < i}_{\in \mathcal{F}_{i-1}} \\ &\in \mathcal{F}_{i-1}. \end{aligned}$$

It follows that

$$\begin{aligned} \tilde{b}\mathbb{E}B_n(0, \tilde{b}) &\le \mathbb{E} \sum_{i=1}^{n} \varphi_i \cdot (\tilde{X}_i - \tilde{X}_{i-1}) \\ &= \sum_{i=1}^{n} \int_{\{\varphi_i = 1\}} (\tilde{X}_i - \tilde{X}_{i-1}) d\mathbb{P} \\ &= \sum_{i=1}^{n} \int_{\{\varphi_i = 1\}} \mathbb{E}\{\tilde{X}_i - \tilde{X}_{i-1} \mid \mathcal{F}_{i-1}\} d\mathbb{P} \\ &= \sum_{i=1}^{n} \int_{\{\varphi_i = 1\}} \mathbb{E}\{\tilde{X}_i \mid \mathcal{F}_{i-1}\} - \tilde{X}_{i-1} d\mathbb{P} \\ &\le \sum_{i=1}^{n} \mathbb{E}\left[\tilde{X}_i - \tilde{X}_{i-1}\right] && \mathbb{E}\{\tilde{X}_i \mid \mathcal{F}_{i-1}\} \overset{a.s.}{\ge} \tilde{X}_{i-1} \\ &= \mathbb{E}\tilde{X}_n. \end{aligned}$$

10.1.2 Doob's Theorem

Theorem 10.1.2 (Doob)
Let $(X_n, \mathcal{F}_n)$ be submartingale with $\sup_n \mathbb{E}X_n^+ < \infty$. Then with probability 1,

(1) $\exists X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$

(2) $X_\infty \in L^1$

Proof
Define

$$A_{p,q} := \{\omega : \exists p, q \in \mathbb{Q} : \liminf_n X_n \leq p < q \leq \limsup_n X_n\}.$$

We have

$$\begin{aligned} A &:= \{\omega : \nexists \lim_{n\to\infty} X_n\} \\ &= \bigcup_{p<q\in\mathbb{Q}} A_{p,q}. \end{aligned}$$

We first claim that $\mathbb{P}(A) = 0$.

Recall $B_n(p,q)(\omega)$ and remark that it is monotonically increasing. It follows that there is some

$$B_\infty(p,q) = \lim_n B_n(p,q).$$

If $\omega \in A_{p,q}$, then we must have $B_\infty(p,q)(\omega) = \infty$. Hence if there is some $A_{p,q}, \mathbb{P}(A_{p,q}) > 0$, we must have $\mathbb{E}B_\infty(p,q) = \infty$. We will show that this is a contradiction.

Indeed, Recall by our lemma that

$$\begin{aligned} \mathbb{E}B_n(p,q) &\leq \frac{\mathbb{E}(X_n - p)^+}{q-p} \\ &\leq \frac{\mathbb{E}X_n^+ + |p|}{q-p}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}B_\infty(p,q) &= \mathbb{E}\lim_n B_n(p,q) \\ &= \lim_n \mathbb{E}B_n(p,q) && \text{MCT} \\ &\leq \sup_n \mathbb{E}B_n(p,q) \\ &\leq \sup_n \frac{\mathbb{E}X_n^+ + |p|}{q-p} \\ &< \infty. && \text{assumption} \end{aligned}$$

By contradiction, $\mathbb{P}(A) = 0$.

Now, to see that $X_\infty \in L^1$,

$$\begin{aligned}
\mathbb{E}|X_\infty| &= \mathbb{E}\lim_n |X_n| \\
&\le \liminf_n \mathbb{E}|X_n| && \text{Fatou's Lemma} \\
&\le \sup_n \mathbb{E}|X_n| \\
&\le \sup_n (2\mathbb{E}X_n^+ - \mathbb{E}X_n) \\
&\le \underbrace{2\sup_n \mathbb{E}X_n^+}_{<\infty} - \mathbb{E}X_1 && \mathbb{E}X_n = \mathbb{E}X_1 \\
&< \infty.
\end{aligned}$$

Corollary 10.1.2.1
Let $\{X_n\}$ be a sequence of random variables. Suppose any of the below hold:

a) $\{X_n\}$ is submartingale and bounded from above

b) $\{X_n\}$ is supermartingale and bounded from below

Then there is some

$$X_\infty(\omega) \overset{a.s.}{=} \lim_n X_n(\omega) \in L^1.$$

Corollary 10.1.2.2
The pair $X_\infty, \mathcal{F}_\infty := \sigma\{\mathcal{F}_i, i \ge 1\}$, closes the sequence. That is,

$$X_1, X_2, \dots, X_\infty$$

is a sub / super martingale.

Proof
We show that case of submartingales.

We have

$$\begin{aligned}
\mathbb{E}\{X_\infty \mid \mathcal{F}_m\}(\omega) &= \mathbb{E}\Big\{\lim_n X_n \mid \mathcal{F}_m\Big\}(\omega) \\
&\ge \limsup_n \mathbb{E}\{X_n \mid \mathcal{F}_m\}(\omega) && \text{Fatou's Lemma} \\
&\ge X_m(\omega). && n \ge m
\end{aligned}$$

Note that this shows

$$\mathbb{E}\{X_infty \mid \mathcal{F}_n\}(\omega) = X_n(\omega)$$

in the case of martingales.

Theorem 10.1.3
Let $(X_n, \mathcal{F}_n)$ be martingale. The following are equivalent:

1) $X_n \xrightarrow{a.s} X_\infty$ and $(X_\infty, \mathcal{F}_\infty)$ is a closing element
2) $X_n \xrightarrow{L^1} X_\infty$, that is, $\mathbb{E}|X_n - X_\infty| \to \infty$ as $n \to \infty$
3) There is some random variable $\xi(\omega)$ such that $X_n(\omega) = \mathbb{E}\{\xi \mid \mathcal{F}_n\}(\omega)$.

The proof of 1) $\iff$ 2) is not hard. The proofs involving 3) are less trivial.

Chapter 11

Applications of Martingales

11.1 Kolmogorov's 0-1 Law

Let $\xi_1, \xi_2, \dots$ be independent random variables. We have

$$\begin{aligned}
\mathcal{F}_n &:= \sigma\{\xi_i : i \in [n]\} \\
&\uparrow \mathcal{F}_\infty \\
&:= \sigma\{\xi_i : i \in \mathbb{N}\} && \text{past} \\
\mathcal{F}_n^\infty &:= \sigma\{\xi_i : i \geq n\} \\
&\downarrow \mathfrak{X} \\
&:= \cap_{n\geq 1} \mathcal{F}_n^\infty. && \text{future}
\end{aligned}$$

Remark that the *tail σ-algebra* $\mathfrak{X}$ is a σ-algebra. It captures events of the "ultimate" future. For instance

$$\begin{aligned}
\{\omega : \max_n \xi_n(\omega) \in A = \infty\} &\in \mathfrak{X} && A \in \mathcal{F} \\
\{\omega : \exists \lim_n \xi_n(\omega)\} &\in \mathfrak{X}.
\end{aligned}$$

However,

$$\begin{aligned}
\{\omega : \lim_n \frac{1}{n} \sum_{i=1}^n \xi_n < c\} &\notin \mathfrak{X} && c \in \mathbb{R} \\
\{\omega : \forall n, \xi_n = 0\} &\notin \mathfrak{X}.
\end{aligned}$$

Finally, note that $\mathfrak{X} \subseteq \mathcal{F}_\infty$, since

$$\bigcap_n \mathcal{F}_n^\infty = \bigcap_n \underbrace{\sigma\{\xi_i : i \geq n\}}_{\subseteq \mathcal{F}_\infty}.$$

Theorem 11.1.1 (Kolmogorov)
For all $A \in \mathfrak{X}$,

$$\mathbb{P}(A) \in \{0, 1\}.$$

Proof
Let $\eta := \mathbb{1}_A(\omega) \in \mathfrak{X}$. Remark that $\eta \in L_1$ is bounded. Consider the Doob martingale

$$X_n := \mathbb{E}\{\eta \mid \mathcal{F}_n\}$$

where $\mathcal{F}_n$ is the natural filtration. Then X_n is martingale and satisfies the conditions of convergence.

$$X_n(\omega) \xrightarrow{a.s./L_1} X_\infty(\omega) := \mathbb{E}\{\mathbb{1}_A \mid \mathcal{F}_\infty\} = \mathbb{1}_A(\omega).$$

The last equality follows from the remark that $A \in \mathfrak{X} \subseteq \mathcal{F}_\infty$.

Remark that $\mathfrak{X}$ is independent of $\mathcal{F}_n$. Thus $X_n = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$! In particular, $\mathbb{1}_A$ is constant and thus is ones of 0-1.

11.2 Fundamental Theorems of Financial Mathematics

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_n, n \geq 1$ a filtration with $\mathcal{F}_0$ the trivial σ-algebra.

Definition 11.2.1 (Market)
We define a (B, S)-market satisfying the following.
$B = (B_n)_{n\geq 0}$ are random variables with B_0 constant, $B_n \in \mathcal{F}_{n-1}$ (predictable), and

$$B_n = B_0 \prod_{k=1}^{n} (1 + b_k).$$

Here $b_n = \frac{B_n - B_{n-1}}{B_{n-1}}$ is the "interest rate" of a stable instrument (banks, bonds, etc).
$S = (S_n)_{n\geq 0}$ are random variables with S_0 constant, $S_n \in \mathcal{F}_n$ (adapted), and

$$S_n = S_0 \prod k = 1^n (1 + s_k).$$

Here $s_n = \frac{S_n - S_{n-1}}{S_{n-1}}$ is the "return" is some more volatile instrument.

Definition 11.2.2 (Portfolio)
A *portfolio* is a random sequence $(\beta_n, \sigma_n)_{n\geq 1}$ with the "value"

$$X_n^\pi(\omega) = \beta_n(\omega)B_n(\omega) + \sigma_n(\omega)S_n(\omega).$$

We say a portfolio is *self-financing* if

$$\Delta X_n^\pi = \beta_n(\omega)\Delta B_n(\omega) + \sigma_n(\omega)\Delta S_n(\omega).$$

Thus there is no "outside" influence.

Definition 11.2.3 (Arbitrage Opportunity)
We say a portfolio π provides opportunity for arbitrary at time n if $X_0 \equiv 0$ and at time n,

$$X_n^\pi(\omega) \overset{a.s.}{\geq} 0$$
$$\mathbb{P}\{X_n^\pi(\omega) > 0\} > 0.$$

We say the a (B, S)-market is arbitrage free if no self-financing portfolio provides an arbitrage oppertunity.

Theorem 11.2.1
In finite discrete time, a (B, S)-market is arbitrage free if and only if there is some $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ with $\mathbb{P}, \tilde{\mathbb{P}}$ absolutely continuous wwith respect to each other, such that

$$\frac{S}{B} = \frac{S_n}{B_n}$$

is $\tilde{\mathbb{P}}$-martingale.

Definition 11.2.4 (N-Complete)
We say that a (B, S)-market is N-complete if any bounded $\mathcal{F}_n$-measurable $f_N(\omega)$ is *replicable*, meaning there exists a self-financing portfolio π such that

$$X_N(\omega) \overset{a.s.}{=} f_N(\omega).$$

Theorem 11.2.2
An arbitrage free (B, S)-market is N-complete for some $N \in \mathbb{N}$ if there exists a unique $\tilde{\mathbb{P}}$-martingale measure.

Example 11.2.3 (CRR-Model)
Suppose the (B, S) market has constant interest $b_k \equiv b, k \in [N]$. Moreover, assume that S_k takes on only two values A, B

$$\mathbb{P}\{S_k = C\} = \begin{cases} q, C = A \\ p, C = B \end{cases}$$

where $p + q = 1$ and $-1 < A < b < B$.

Thus

$$\begin{aligned} \frac{S_n}{B_n} &= \left(\frac{1+S_n}{1+b}\right)\left(\frac{S_{n-1}}{B_{n-1}}\right) \\ &= \frac{S_0}{B_0}\prod_{k=1}^{n}\left(\frac{1+S_k}{1+b}\right). \end{aligned}$$

If we want the above to be a martingale in some $\tilde{P}$, we can ask that

$$\tilde{E}\left(\frac{1+S_n}{1+b}\right) = 1 \iff \mathbb{E}S_n = b.$$

Then

$$\tilde{P}\{S_n = C\} = \begin{cases} \frac{B-b}{B-A}, & C = A \\ \frac{b-A}{B-A}, & C = B \end{cases}$$

11.3 Random Walks

Let $\xi_1, \xi_2, \ldots$ be iid random variables.

Definition 11.3.1 (Random Walk)
We say

$$S_n = \sum_{i=0}^{n} \xi_i$$

is a *random walk*.

A *simple random walk* satifies $\mathbb{P}\{\xi_i = 1\} = p, \mathbb{P}\{\xi_i = -1\} = 1 - p =: q$. Note that we can easily generalize these notions to $\mathbb{R}^d$.

11.3.1 Revisiting Points

We wish to determine whether we return to ξ_0. If so, how many times do we do so? Recall by Komogorov's 0-1 law,

$$P\{\max_n S_n = 0 = \infty\} \in \{0, 1\}.$$

Now, remark that the walk can only return to 0 on even times.

$$\begin{aligned}
\mathbb{P}(B_{2n}) &:= \mathbb{P}\{\omega : S_{2n}(\omega) = 0\} \\
&= \binom{2n}{n} p^n q^n \\
&\approx \frac{(4pq)^n}{\sqrt{\pi n}}. && \text{Sterling's Formula}
\end{aligned}$$

If $pq < 1$, then $\sum_n \mathbb{P}(B_{2n}) < \infty$ and Borel-Cantelli yields that the probability of visiting 0 infinitely many times is 0.

For $p < \frac{1}{2}$, we can show that

$$\mathbb{E}\left[\sup_n S_n\right] < p.$$

Proposition 11.3.1
If $p = q = \frac{1}{2}$,

$$\mathbb{P}\{S_n = N \text{ infinitely many times}\} = 1$$

for any $N \in \mathbb{Z}$.

Proof
It suffices to show that

$$\mathbb{P}\left\{\limsup_n \frac{S_n}{\sqrt{n}} = \infty, \liminf_n \frac{S_n}{\sqrt{n}} = -\infty\right\} = 1.$$

Moreover, the probability is either 0 or 1, hence we need only show that it is non-zero.

Indeed,

$$\begin{aligned}
\lim_{c \to \infty} \mathbb{P}\left\{\limsup_n \frac{S_n}{\sqrt{n}} \geq c\right\} &\geq \limsup_n \mathbb{P}\left\{\frac{S_n}{\sqrt{n}} \geq c\right\} && \text{Fatou's Lemma} \\
&= \limsup_n \mathbb{P}\{N(0,1) \geq c\} \\
&= \limsup_n 1 - \operatorname{erf}(c) \\
&> 0.
\end{aligned}$$

11.3.2 Stopped Random Walks

suppose we start at some $a \in [0, m]$ and $p = q = \frac{1}{2}$ so the random walk is a martingale. Define

$$\tau := \min\{n : S_n(\omega) \in \{0, m\}\}$$

and remark it is a stopping time. Furthermore, $\mathbb{E}\tau < \infty$ and

$$\begin{aligned} \mathbb{E}\{|S_{n+1} - S_n| \mid \mathcal{F}_n\} &= \mathbb{E}\{|\xi_{n-1}| \mid \mathcal{F}_n\} \\ &= 1 \end{aligned}$$

is bounded.

It follows that we can apply Doob's optional stopping theorem to conclude that

$$(S_0, \mathcal{F}_0), (S_\tau, \mathcal{F}_\tau)$$

is martingale. In particular,

$$\mathbb{E}S_\tau = \mathbb{E}S_0 = a.$$

Since $S_\tau = 0$ or m, the probability of being either is $\frac{m-a}{m}, \frac{a}{m}$.

Definition 11.3.2 (De Moivre's Martingale)
For $p \neq q$ we can employ *De Moivre's Martingale*

$$M_n(\omega) := \left(\frac{q}{p}\right)^{S_n(\omega)}.$$

We can also ask questions about one-side boundaries $\tau = \min\{n : S_n = 1\}$ if we start at the point 0. Now, suppose $\mathbb{E}S_\tau = \mathbb{E}S_0 = 0$. But $S_\tau = 1$ which is a contradiction. We must have contradicted an assumption in Doob's optional stopping theorem and in particular $\mathbb{E}\tau = \infty$.

Chapter 12

Stochastic Processes

12.1 Stochastic Processes

Definition 12.1.1 (Stochastic Process)
A collection of random elements

$$X(\omega) = \{X_t(\omega) : t \in T\}$$

where T is some index set.

For $T = \mathbb{N}$, X is a *discrete time process*. For $T = [0, \infty)$, we say X is a *continuous time process*. For the sake of culture, $T = \mathbb{R}^d$ means we have a *random field*.

Now, for $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we say X is a *d-dimensional process*.

Lastly, we remark that $X_t(\omega) : T \times \Omega \to \mathbb{R}$ is a function on two variables. For each t, X_t is a random variable. For each ω fixed, $t \mapsto X_t(\omega)$ is a *trajectory* or *realization*.

We can think of X as a "random trajectory", but we need to develop some measurable space first.

12.1.1 Equivalences

Let $\{X_t\}, \{Y_t\}$ be two processes on the same space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 12.1.2 (Indistinguishable)
We say X_t, Y_t are *indistinguishable* if

$$\mathbb{P}\{\omega : \forall t, X_t(\omega) = Y_t(\omega)\} = 1.$$

Definition 12.1.3 (Modification / Strict Equivalence)
We say X_t, Y_t are *modifications* of each other, or *strictly equivalent* if $\forall t \in T$,

$$\mathbb{P}\{\omega : X_t(\omega) = Y_t(\omega)\} = 1.$$

Note that even if X_t, Y_t are modifications of each other, they can still have different properties of trajectories!

Example 12.1.1
Suppose $X_t(\omega) \equiv 0$ while

$$Y_t(\omega) = \begin{cases} 1, & t = \xi(\omega) \\ 0, & t \notin \xi(\omega) \end{cases}$$

where $\xi(\omega)$ is some continuous random variable (with no atoms).

Then for all t,

$$\mathbb{P}\{Y_t(\omega) \neq X_t(\omega)\} = \mathbb{P}\{\xi(\omega) = t\} = 0.$$

Proposition 12.1.2
If X_t, Y_t are modifications of each other and have right continuous trajectories a.s., then X_t, Y_t are indistinguishable.

12.1.2 Finite-Dimensional Distributions

Let $\{X_t\}_{t \in T}$ be a 1-dimensional process.

Definition 12.1.4 (System of Finite-Dimensional Distributions)
A system of finite-dimensional distributions $\mathcal{P}_T$ on $\mathbb{R}^T$ is a family of probability distributions containing a probability distribution $\mathbb{P}_{t_1,\ldots,t_n}$ on $\mathbb{R}^n$ for every possible $n \geq 1, t_1, \ldots, t_n \in T$.

The *system of finite-dimensional distributions* induced by X_t are of the form

$$\mathbb{P}^{(X)}_{t_1,\ldots,t_n}(A) := \mathbb{P}\{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in A\}$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$.

Note that in particular, for $A_i \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}^{(X)}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \mathbb{P}\bigcap_{i=1}^{n}\{X_{t_i} \in A_i\}.$$

Definition 12.1.5 (Consistent)
An arbitrary system of finite-dimensionaal distributions is *consistent* if

a) For every $n \geq 1, t_1, \dots, t_n \in T, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \mathbb{P}_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)})$$

for any permutation $\sigma \in S_n$.

b) For all $n \geq 1, t_1, \dots, t_n, t_{n+1} \in T, A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}_{t_1,\dots,t_n,t_{n+1}}(A \times \mathbb{R}) = \mathbb{P}_{t_1,\dots t_n}(A).$$

Proposition 12.1.3
The system of finite-dimensional distributions induced by X_t is consistent.

Definition 12.1.6 (Wide Equivalence)
X_t, Y_t are *widely equivalent* if their systems of finite-dimensional distributions coincide.

Proposition 12.1.4
If X_t, Y_t are strictly equivalent, then they are also widely equivalent.

Proof
For every $t_1, \dots, t_n \in T, A \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned}
&|\mathbb{P}\{(X_{t_1}, \dots, X_{t_n}) \in A\} - \mathbb{P}\{(Y_{t_1}, \dots, Y_{t_n}) \in A\}| \\
&\leq \mathbb{P}\{(X_{t_i}) \in A, (Y_{t_i}) \notin A\} + \mathbb{P}\{(X_{t_i}) \notin A, (Y_{t_i}) \in A\} \\
&\leq \mathbb{P}\{X_{t_1} \neq Y_{t_1}, \dots, X_{t_n} \neq Y_{t_n}\} \\
&\leq \sum_{i=1}^{n} \mathbb{P}\{X_{t_i} \neq Y_{t_i}\} \\
&= 0.
\end{aligned}$$

12.1.3 Processes as Random Elements

Definition 12.1.7 (Cylinder Set)
A cylinder set in $\mathbb{R}^T$ is given by

$$C_{t_1,\dots,t_n}(A_1,\dots,A_n) := \{x \in \mathbb{R}^n : x_{t_1} \in A_1, \dots, x_{t_n} \in A_n\}.$$

Here $n \geq 1$ and $t_1, \dots, t_n \in T$.

We define the σ-algebra generated by cylinder sets as

$$\mathcal{C}_{\mathbb{R}^T} = \mathcal{C}_T := \sigma\{C_{t_1,\dots,t_n}(A_1,\dots,A_n) : n \geq 1, t_i \in T, A_i \in \mathcal{B}(\mathbb{R})\}.$$

Now, stochastic processes are random elements of the measurable space $(\mathbb{R}^T, \mathcal{C}_T)$. Indeed, we think of a stochastic process as $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^T, \mathcal{C}_T)$. It is measurable since

$$X^{-1}(C_{t_1,\dots,t_n}(A_1,\dots,A_n)) = \{\omega : X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}.$$

12.2 Kolmogorov's Extension Theorem

Theorem 12.2.1 (Kolmogorov)
Suppose $\mathcal{P}_T$ is a consistent family of finite-dimensional distributions. Then there is a probability measure $\mathbb{P}_T$ on $(\mathbb{R}^T, \mathcal{C}_T)$ with the given finite-dimensional distributions.

The theorem is proven by leveraging Carathéodory's extension theorem.

We remark that this theorem characterizes stochastic processes with a consistent system of finite-dimensional distributions! We will employ this theorem to construct Brownian motion.

Example 12.2.2 (sketch)
Define $\mathbb{P}_{t_1,\dots,t_n}$ as follows: Sort and relabel $t_1 < t_2 < \cdots < t_n$. Then define

$$\begin{aligned}
&\mathbb{P}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) \\
&= \int_{A_1}\int_{A_2}\cdots\int_{A_n} p_{t_1}(0,x_1)\cdot p_{t_2-t_1}(x_1,x_2)\dots p_{t_n-t_{n-1}}(x_{n-1},x_n)dx_1\dots dx_n \\
&p_t(x,y) \\
&= \frac{1}{\sqrt{2\pi t}}\exp\left[-\frac{(x-y)^2}{2t}\right].
\end{aligned}$$

Example 12.2.3 (Bernoulli Trials)
Let $X_n(\omega)$ take values in 0, 1 independently with equal probability for all $n \in \mathbb{N}$. Then $X_n \in (\mathbb{Z}_2^n, \mathcal{C}_\mathbb{N})$.

We have

$$\begin{aligned}\mathbb{P}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) &= \mathbb{P}\{x_{t_1} \in A_1, \dots, x_{t_n} \in A_n\} \\ &= \prod_{i=1}^n \mathbb{P}_{\text{Be}}(A_i).\end{aligned}$$

By Kolmogorov's extension theorem, there is some $\mathbb{P}_B$ on $(\mathbb{Z}_2^\mathbb{N}, \mathcal{C}_\mathbb{N})$ corresponding to a stochastic process

$$B_{\text{Be}} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{Z}_2^\mathbb{N}, \mathcal{C}_T).$$

We remark that it can be shown

$$(\mathbb{Z}_2^\mathbb{N}, \mathcal{C}_\mathbb{N}, \mathbb{P}_B) \cong ([0,1], \mathcal{B}[0,1], \lambda).$$

The isomorphism is simply given by

$$y = \sum_{n \in \mathbb{N}} 2^{-n} x_n.$$

Moreover,

$$\mathcal{C}_T = \sigma\{x : \exists n, x_i \text{ is fixed for } i \leq n\}.$$

Thus the image σ-algebra is generated by open intervals.

12.3 Kolmogorov's Continuity Theorem

Definition 12.3.1 (Continuity)
We say a stochastic process $X_t(\omega)$ is *continuous* if

$$\mathbb{P}\{\omega : X_t(\omega) \text{ is continuous}\} = 1.$$

Theorem 12.3.1 (Kolmogorov)
Suppose $X_t(\omega)$ is a stochastic process with $t \in [a, b]$ such that there are constants $\epsilon, \beta, c > 0$ such that for every $t, s \in [a, b]$,

$$\mathbb{E}\left[|X_t - X_s|^\beta\right] < c|t-s|^{1+\epsilon}.$$

The following hold:

1) $X_t(\omega)$ has a continuous modification $\tilde{X}_t(\omega)$.
2) $\tilde{X}_t(\omega)$ is *locally Hölder continuous*, ie there is some $K(\omega) \overset{a.s.}{>} 0$ such that

$$|\tilde{X}_t(\omega) - X_s(\omega)| \leq K(\omega)|t-s|^\alpha$$

for every $\alpha \in \left(0, \frac{\epsilon}{\beta}\right)$.

Is $\tilde{X}_t(\omega)$ unique up to indistinguishability?

Proof
1) By Chebychev's inequality,

$$\begin{aligned}\mathbb{P}\{|X_t - X_s| > \delta\} &\leq \frac{\mathbb{E}|X_s - X_t|^\beta}{\delta^\beta} \\ &\leq \frac{c|t-s|^{1+\epsilon}}{\delta^\beta} \\ &\to 0. & |s-t| \to 0\end{aligned}$$

Hence $X_{s_n} \overset{p}{\to} X_t$ for any $s_n \to t$.

Now fix some $m \geq 1$ and define $q := 2^{-\alpha}$ for some $\alpha \in \left(0, \frac{\epsilon}{\beta}\right)$. From our result above,

$$\begin{aligned}\mathbb{P}\left\{\left|X_{\frac{k+1}{2^m}} - X_{\frac{k}{2^m}}\right| > q^m\right\} &\leq \frac{c2^{-m(1+\epsilon)}}{q^{m\beta}} \\ &= c2^{-m}2^{-m(\epsilon-\alpha\beta)}.\end{aligned}$$

We can then take a union bound

$$\mathbb{P}\left\{\max_k\left|X_{\frac{k+1}{2^m}} - X_{\frac{k}{2^m}}\right| > q^m\right\} \leq \frac{(b-a)}{2^{-m}} \cdot c2^{-m}2^{-m(\epsilon-\alpha\beta)}$$

which is a term in a convergent geometric series. It follows by the Borel-Cantelli lemma that these events occur for only finitely many m a.s. In other words, there is some $\Omega^*, \mathbb{P}(\Omega^*) = 1$ and $M_0 : \Omega^* \to \mathbb{N}$ such that for every $\omega \in \Omega^*$ and $m \geq M_0(\omega)$,

$$\max_k\left|X_{\frac{k+1}{2^m}} - X_{\frac{k}{2^m}}\right| \leq q^m = 2^{-\alpha m}.$$

Now fix $\omega^* \in \Omega^*$ and s, t be dyadic rationals. Moreover, suppose

$$2^{-(n+1)} \leq |t-s| \leq 2^{-n} < 2^{-M_0(\omega^*)}.$$

Without loss of generality, we can write

$$\begin{aligned} s &= k2^{-n} - 2^{-v_1} + \dots + 2^{-v_a} \\ t &= k2^{-n} + 2^{-\ell_1} + \dots + 2^{-\ell_b}. \end{aligned}$$

Here $n < v_i < v_{i+1}$ and $n < \ell_i < \ell_{i+1}$. Then

$$\begin{aligned} |X_t(\omega^*) - X_s(\omega^*)| &\leq |X_t - X_{\frac{k}{2^m}}| + |X_s - X_{\frac{k}{2^m}}| \\ &\leq \sum_i q^{v_i} + \sum_j q^{\ell_j} \\ &\leq 2 \sum_{k=n+1}^{\infty} q^k \\ &= \frac{2q^{n+1}}{1-q} \\ &= \frac{2 \cdot 2^{-\alpha(n+1)}}{1-2^{-\alpha}} \\ &\leq \frac{2}{1-2^{-\alpha}} |t-s|^\alpha. && 2^{-(n+1)} \leq |t-s| \end{aligned}$$

We define

$$\begin{aligned} \tilde{X}_t(\omega) &:= X_t(\omega) && t \text{ dyadic} \\ \tilde{X}_t(\omega) &:= \lim_{t_n \to t} X_{t_n}(\omega). && t_n \text{ dyadic, } t \text{ not} \end{aligned}$$

Thus $\tilde{X}_t$ is continuous by construction.

To see that $\tilde{X}_t$ is a modification, first remark that $\tilde{X}_t = X_t$ for dyadic t. For t non-dyadic, we have $X_{s_n} \xrightarrow{p} X_t$ and $\tilde{X}_{s_n} \xrightarrow{a.s.} X_t$. Hence by an unproven lemma, $X_t = \tilde{X}_t$ a.s.

2) By construction, $\tilde{X}_t(\omega)$ is locally Hölder continuous.

Remark that we proved the theorem for a compact interval $[a, b]$. We can expand onto $[0, \infty)$.

Also, suppose that

$$\mathbb{E}\left[|X_t - X_u|^\delta \cdot |X_u - X_s|^\delta\right] \leq c|t-s|^{1+\delta},$$

then there is a unique modification $\overline{X_t(\omega)}$ with CADLAG trajectories.

Consider the space of continuous functions $C[0, \infty)$ on the half-line. There is a metric

$$\rho(f, g) := \sum_{n=1}^{\infty} \max_{1 \leq t \leq n} (1, |f(t) - g(t)|)$$

such that

$$\mathcal{C}_{[0,\infty)}\Big|_{C[0,\infty)} = \mathcal{B}(C[0, \infty)).$$

Chapter 13

Brownian Motion

13.1 First Construction

Define $\mathbb{P}_{t_1,\dots,t_n}$ as follows: Sort and relabel $t_1 < t_2 < \cdots < t_n$. Then define

$$\begin{aligned}
&\mathbb{P}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) \\
&= \int_{A_1}\int_{A_2}\cdots\int_{A_n} p_{t_1}(0,x_1)\cdot p_{t_2-t_1}(x_1,x_2)\dots p_{t_n-t_{n-1}}(x_{n-1},x_n)dx_1\dots dx_n \\
&p_t(x,y) \\
&= \frac{1}{\sqrt{2\pi t}}\exp\left[-\frac{(x-y)^2}{2t}\right].
\end{aligned}$$

It can be checked that this is a consistent system. Hence we can apply Kolmogorov's extension theorem to extract some stochastic process

$$B_t(\omega) : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^{[0,\infty)}, \mathcal{C}_{[0,\infty)}).$$

We remark that

$$\mathbb{P}_{t_1}(A_1) = \int_{A_1} \frac{1}{\sqrt{2\pi t_1}}\exp\left(-\frac{x_1^2}{2t_1}\right)dx_1$$

so $B_{t_1} \sim N(0,t_1)$!

Moreover, $\mathbb{P}_0 = \delta_0$ is the delta measure at 0. If we wish to start a different distribution, we will need to change the construction of the density.

Furthermore, remark that the increments

$$B_{t_1}, B_{t_1} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$$

are independent. Thus we have

$$\begin{bmatrix} B_{t_1} \\ B_{t_2} - B_{t_1} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{bmatrix} \sim N(0, \operatorname{Diag}(t_1, t_2 - t_1, \dots, t_n - t_{n-1})).$$

We can thus deduce that

$$\begin{aligned} \begin{bmatrix} B_{t_1} \\ B_{t_2} \\ \vdots \\ B_{t_n} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} B_{t_1} \\ B_{t_2} - B_{t_1} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{bmatrix} \\ &\sim N\left(0, \begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \ddots & \ddots & \vdots \\ t_1 & t_2 & \dots & t_n \end{bmatrix}\right). \end{aligned}$$

Thus $\operatorname{Var} B_{t_i} = t_i$ and

$$\operatorname{Cov}(B_{t_i}, B_{t_j}) = \min(t_i, t_j).$$

Remark that

$$\mathbb{E}\left[|B_t - B_s|^4\right] = 3|t - s|^2.$$

Hence the assumption of Kolmogorov's continuity theorem is satisfied with parameter $\beta = 4, \epsilon = 1$ and there is some continuous modification $\tilde{B}_t(\omega)$.

Definition 13.1.1 (Brownian Motion)
A stochastic process $B_t(\omega)$ is a *Brownian motion* if its system of finite-dimensional distributions has locally Hölder continuous trajectories with parameter $\alpha < \frac{1}{2}$.

While this construction seems tailored to Gaussians, it is actually a necessity. We needed that the Gaussian integral works out nicely to show consistency. Indeed, we need that the sum of two increments yields the same class of densities.

13.2 Processes with Independent Increments

Definition 13.2.1 (Independent Increment)
A stochastic process $X_t(\omega)$ has *independent increments* if for all $0 = t_0 < t_1 < \cdots < t_n \in T$,

$$X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$$

are independent.

Note that there is a bijection between the finite dimensional distributions of X_t and the distribution of increments.

We wish to define processes using increments. To do so, we need to adapt the notion of consistency for increments. Note that we define increments on the ordered vectors so permutation invariance is free. We need only check that

$$\begin{aligned}
&\mathbb{P}\{(X_{t_0}, \dots, X_{t_n}) \in A_0 \times \dots, A_k = \mathbb{R}, \dots, A_n\} \\
&= \mathbb{P}\{(X_{t_0}, \dots, X_{t_{k-1}}, X_{t_{k+1}}, \dots, X_{t_n}) \in A_0 \times \cdots \times A_{k-1} \times A_{k+1} \times \cdots \times A_n\} \\
&\iff \\
&\mathbb{P}\{D(X_{t_0}, \dots, X_{t_n}) \in D(A_0 \times \dots, A_k = \mathbb{R}, \dots, A_n)\} \\
&= \mathbb{P}\{D(X_{t_0}, \dots, X_{t_{k-1}}, X_{t_{k+1}}, \dots, X_{t_n}) \in D(A_0 \times \cdots \times A_{k-1} \times A_{k+1} \times \cdots \times A_n)\}.
\end{aligned}$$

Here D is the linear operator which sends the vector to its increments.

Equivalently, Since we must check over some generating set of the borel σ-algebra we check that over boxes B_i,

$$\begin{aligned}
&\mathbb{P}\{D(X_{t_0}, \dots, X_{t_n}) \in (B_0 \times \dots, D(\mathbb{R}, B_{k+1}), \dots, B_n)\} \\
&= \mathbb{P}\{D(X_{t_0}, \dots, X_{t_{k-1}}, X_{t_{k+1}}, \dots, X_{t_n}) \in (B_0 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n)\} \\
&\iff \\
&\mathbb{P}\{(X_{t_k} - X_{t_{k-1}}, X_{t_{k+1}} - X_{t_k}) \in D(\mathbb{R} \times B_{k+1})\} \\
&= \mathbb{P}\{X_{t_{k+1}} - X_{t_{k-1}} \in B_{k+1}\} \\
&\iff \\
&(X_{t_{k+1} - X_{t_k}})(\omega) \overset{ind}{+} (X_{t_k} - X_{t_{k-1}})(\omega) \\
&\overset{d}{=} (X_{t_{k+1}} - X_{t_{k-1}})(\omega).
\end{aligned}$$

Proposition 13.2.1
A process $X_t(\omega)$ with independent increments is uniquely defined by the distribution of $X_0(\omega), (X_t - X_s)(\omega)$ for $s < t \in T$, provided that

$$(X_t - X_s)(\omega) \overset{ind}{+} (X_s - X_u)(\omega) \overset{d}{=} (X_t - X_u)(\omega).$$

Proposition 13.2.2
Let $T = \mathbb{Z}_+$ so that $X_m(\omega)$ is a discrete time process. Then $X_m(\omega)$ has independent increments if and only if $X_m = \sum_{k=0}^{m} \xi_k$ for some independent random variables ξ_k.

Proof
If it has independent increments, then $\xi_0 := X_0, \xi_i := X_i - X_{i-1}$ is independent by assumption.

Conversely,

$$(X_0, X_{m_1} - X_0, \dots, X_{m_n} - X_{m_{n-1}}) = \left(\xi_0, \sum_{k=1}^{m_1} \xi_k, \dots, \sum_{k=m_{n-1}}^{m_n} \xi_k \right).$$

In particular, a Bernoulli process does NOT have independent increments.

13.3 The Increments of Brownian Motion

By the first construction of BM,

$$\begin{aligned}
&\mathbb{P}_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) \\
&= \int_{A_1} \int_{A_2} \cdots \int_{A_n} p_{t_1}(0, x_1) \cdot p_{t_2 - t_1}(x_1, x_2) \dots p_{t_n - t_{n-1}}(x_{n-1}, x_n) dx_1 \dots dx_n \\
&p_t(x, y) \\
&= \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{(x-y)^2}{2t} \right].
\end{aligned}$$

In particular, note that $B_t \sim N(0, t)$.

Moreover,

$$B_{\alpha t} \stackrel{d}{=} \sqrt{\alpha} B_t \sim N(0, \alpha t).$$

Proposition 13.3.1
Brownian motion has independent increments.

Proof
We have

$$\begin{aligned}
&\mathbb{P}\{D(B_0, \ldots, B_{t_n}) \in A_0 \cdots \times A_n\} \\
&= \mathbb{P}\{(B_0, \ldots, B_{t_n}) \in D^{-1}A\} \\
&= \int_{D^{-1}A} p(x_1, \ldots, x_n) dx_1 \ldots dx_n \mathbb{1}\{0 \in A_0\} dx_0 \\
&= \int_A p(y_1; 0, t_1^2) \ldots p(y_n; t_{n-1}, (t_n - t_{n-1})^2) dy_1 \ldots dy_n.
\end{aligned}$$

This coincides with an independent distribution as desired.

In fact, we note that

$$B_t - B_s \sim N(0, t-s) \sim B_{t-s}.$$

Definition 13.3.1 (Stationary)
A process has *stationary* increments if the increment from $s \to t$ is a function of $t - s$.

We have just shown that the increments of BM are independent and stationary.

Recall that $\operatorname{Cov}(B_t, B_s) = \min(t, s)$. We can compute this as

$$\mathbb{E}B_t B_t = \mathbb{E}B_s^2 + \mathbb{E}(B_t - B_s)\mathbb{E}B_s = s.$$

Remark that BM can alternatively be defined as the process with independent and stationary increments with distributions $B_0 \equiv 0$ and $B_t - B_s \sim N(0, t-s)$, provided that

$$(B_t - B_s) \overset{ind}{+} (B_s - B_u) \overset{d}{=} (B_t - B_u)$$

for all $t > s > u$.

In other words, the convolutional distribution

$$D(\ldots) * D(\ldots) = D(\ldots)$$

for normal distributions is again a normal. This is quit a rare property! In fact, these distributions are known as Lévy processes to which Poisson and BM belong.

Finally, as a final remark on GM, recall Komogorov's continuity theorem. There is a modification of BM with locally Hölder continuous trajectories. Thus

$$|(B_t - B_s)(\omega)| < K(\omega)|t - s|^\alpha$$

for any $\alpha \in \left(0, \frac{1}{2}\right)$. We also know that the trajectory B_t is nowhere differentiable with probability 1.

Chapter 14

More Stochastic Processes

14.1 Poisson Processes

Definition 14.1.1 (Poisson Process)
A Poisson process $P_t(\omega)$ is the stochastic process such that

1) $P_0(\omega) \equiv 0$
2) P_t has inddependent and stationary increments
3) $(P_t - P_s)(\omega) \sim \text{Po}(\lambda(t-s))$ for all $t > s \in \mathbb{R}_+$

From our work with independent increments, we need only check that

$$(P_t - P_s) \overset{ind}{+} (P_s - P_u) \overset{d}{=} (P_t - P_u)$$

for every $t > s > u$. In other words,

$$\text{Po}(\lambda(t-s)) * \text{Po}(\lambda(s-u)) \overset{d}{=} \text{Po}(\lambda(t-u)).$$

and then P_t exists. Note that while P_t is well-defined, we do not know prior to checking that it actually exists. This can be done simply by computing

$$\sum_{k=0}^{n} \mathbb{P}\{\xi_1 = k\} \cdot \mathbb{P}\{\xi_2 = n-k\}.$$

Proposition 14.1.1
P_t is *stochastically continuous*, that is,

$$P_{t+h} \overset{p}{\to} P_t$$

as $h \to 0$.

Proof

$$
\begin{aligned}
\mathbb{P}\{|P_{t+h} - P_t| > \epsilon\} &= 1 - \mathbb{P}\{P_{t+h} - P_t = 0\} & \epsilon \in (0,1) \\
&= 1 - \exp(-\lambda h) \\
&\to 0.
\end{aligned}
$$

Proposition 14.1.2
There exists a CADLAG modification of P_t.

Proof
We apply the stronger version of Kolmogorov's continuity theorem

$$
\begin{aligned}
\mathbb{E}(P_t - P_s)(P_s - P_u) &= \lambda^2 |t-s| \cdot |s-u| \\
&\leq \lambda^2 |t-u|^{1+1}.
\end{aligned}
$$

Proposition 14.1.3
P_t is non-decreasing a.s. and $\mathbb{Z}_+$-valued.

Proof
Since the trajectory is CADLAG, it suffices to check that this holds on the rational points.

$$
\begin{aligned}
&\mathbb{P}\{\forall q \geq p \in \mathbb{Q}, P_q(\omega) \geq P_p(\omega)\} = 1 \\
&\iff \forall q \geq p \in \mathbb{Q}, \mathbb{P}\{P_q(\omega) \geq P_p(\omega)\} = 1 \\
&\iff \forall q \geq p \in \mathbb{Q}, \mathbb{P}\{P_q(\omega) - P_p(\omega) \geq 0\} = 1.
\end{aligned}
$$

14.1.1 Counting Renewal Processes

We can think of a Poisson as a "counting renewal" proces, which are non-negative integer-value processes where $P_t - P_s$ model the the number of events which occur during the interval $(s, t]$. In particular, $P_t \operatorname{Po}(\lambda t)$ models the number of occurances up until time t for Poisson processes.

14.1.2 Random Point Processes

We construct a "random measure" on $\mathcal{B}(\mathbb{R})$ as follows. First consider an interval $A = [a, b]$. We define

$$\begin{aligned}\#^P_A(\omega) &:= \sum_{k=1}^{\infty} \mathbb{1}\{T_k \in [a, b]\} \\ &= P_b(\omega) - P_a(\omega) \\ &\sim \text{Po}(\lambda(b-a)).\end{aligned}$$

Then for disjoint intervals $A \sqcup B$,

$$\begin{aligned}\#^P_{A\sqcup B}(\omega) &= \text{Po}(\lambda(\ell(A))) * \text{Po}(\lambda\ell(B)) \\ &= \text{Po}(\lambda\mu(A \cup B)) && \text{Lebesgue measure.}\end{aligned}$$

In general for $C \in \mathcal{B}(\mathbb{R})$,

$$\#^P_C(\omega) \sim \text{Po}(\lambda\mu(C)).$$

Note that $\#^P_C, \#^P_D$ are independent for $C \cap D = \varnothing$.

An interesting example of a random point process are the eigenvalues of some random matrix.

We note that we can extend Poisson processes to $\mathbb{R}^d$ using the notion of a random point process.

14.2 Lévy Processes

Definition 14.2.1 (Lévy Process)
A stochastic process satisfying:

1) $L_0(\omega) \equiv 0$
2) Stationary and independent increments
3) Stochastically continuous trajectories

Lévy processes are a generalization of both Brownian motion as well as Poisson processes and capture all the processes defined using independent increments.

Proposition 14.2.1
L_t has CADLAG modifications.

Example 14.2.2
Brownian motion with drift

$$B_t + at$$

is a Lévy process.

The gamma process with density

$$p_t(x) := \frac{\lambda^{\gamma t}}{\Gamma(\gamma t)} x^{\gamma t - 1} \exp(-\lambda x)$$

is a Lévy process.

We remark Lévy processes are infinitely divisible distributions. In other words,

$$\begin{aligned} L_t(\omega) &= \sum_{k=1}^{n} \left(L_{\frac{kt}{n}} - L_{\frac{(k-1)t}{n}} \right)(\omega) \\ &= \sum_{k=1}^{n} L^{(k)}_{\frac{t}{n}}. \end{aligned}$$

Another way to express this relation is through the characteristic function of the process

$$\varphi_t(x) = \varphi^n_{\frac{t}{n}}(x).$$

Example 14.2.3 (Infinitely Divisible Distributions)
Any stable distribution is infinitely divisible. Examples include Normal, Cauchy, Gamma, Poisson, χ^2, and Student distributions.

14.3 Gaussian Processes

Definition 14.3.1 (Gaussian process)
A stochastic process is Gaussian if all finite-dimensional distributions are Gaussian vectors

$$(G_{t_1}, \ldots, G_{t_n}) \sim N(A_{t_1,\ldots,t_n}, \Sigma_{t_1,\ldots,t_n}).$$

Example 14.3.1
Brownian motion is a Gaussian process.

Proposition 14.3.2
A random vector $\bar{\xi}$ is a Gaussian vector if and only if for all coefficients $\alpha_i \in \mathbb{R}$,

$$\sum_i \alpha_i \xi_i$$

is a Gaussian random variable.

Proposition 14.3.3
G_t is Gaussian if and only if for every $t_i \in T, \alpha_i \in \mathbb{R}$,

$$\left(\sum_i \alpha_i T_{t_i}\right)(\omega)$$

is a Gaussian random variable.

Example 14.3.4 (Discrete Brownian Motion)
Consider discrete Brownian motion, with $T = \mathbb{Z}_+$, which is a process with independent stationary increments and satisfies

$$\begin{aligned} \mathrm{DB}_0 &\equiv 0 \\ \mathrm{DB}_n(\omega) &= \sum_{k=1}^{n} \xi_k(\omega). & \xi_k \sim N(a, \sigma) \end{aligned}$$

Example 14.3.5 (Ornstein-Uhlenbeck)
This is the Gaussian process given by

$$\mathrm{OU}_t(\omega) = \exp(-\theta t) B_{\exp(2\theta t)}(\omega)$$

for $\theta > 0$.

Note that for every $t \geq 0$,

$$\begin{aligned} \mathrm{OU}_t(\omega) &\overset{d}{=} \exp(-\theta t) \cdot N(0, \exp(2\theta t)) \\ &= N(0, 1). \end{aligned}$$

Example 14.3.6 (Brownian Bridge)
Consider $T = [0, 1]$. A brownian bridge is the process given by

$$\mathrm{BB}_t(\omega) = B_t(\omega) - tB_1(\omega).$$

We have

$$\begin{aligned} \mathrm{BB}_0 &\equiv 0 \\ \mathrm{BB}_1 &\equiv 0 \\ \mathrm{BB}_t &\sim N(0, t(1-t)) \\ \operatorname{Var} \mathrm{BB}_t &= \operatorname{Var}\left[(1-t)B_t - t(B_1 - B_t)\right] \\ &= (1-t)^2 t + t^2(1-t). \end{aligned}$$

We remark that this is a Gaussian process since

$$\sum_i \alpha_i \mathrm{BB}_{t_1} = \left(\sum_i \alpha_i B_{t_i}\right) - \left(\sum_i \alpha_i t_i\right) B_1.$$

14.4 Markov Chains & Processes

Informally speaking, a process is Markov if the probability of the "future" conditioned on the "past" and "present" is just the probability of the probability of the "future" conditioned on the "present". It captures a different notion of "memoryless" compared to the exponential distribution as well as martingales.

14.4.1 Markov Chains

Let S be a countable *phase* (*state*) space.

Definition 14.4.1 (Markov Chain)
A discrete time process $X_n, n \in \mathbb{Z}_+$ is a Markov chain if

1) $X_n \in S$ for all $n \in \mathbb{Z}_+$
2) $\mathbb{P}\{X_n = i \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\} = \mathbb{P}\{X_n = i \mid X_{n-1} = i_{n-1}\}$

Proposition 14.4.1
The second condition is equivalent to any of the following:

(i) $\mathbb{P}\{X_n = i \mid X_{k_j} = i_{k_j}, j \in [\ell]\} = \mathbb{P}\{X_n = i \mid X_{k_\ell} = i_{k_\ell}\}$ for $k_j < k_{j+1}$

(ii) $\mathbb{P}\{X_n \in A_n \mid X_{k_j} \in A_{k_j}, j \in [\ell]\} = \mathbb{P}\{X_n \in A_n \mid X_{k_\ell} \in A_{k_\ell}\}$ for $k_j < k_{j+1}$

(iii)

$$\begin{aligned} &\mathbb{P}\{X_n = i_n, X_{k+1} = i_{k+1}, X_{k-1} = i_{k-1}, \dots, X_0 = i_0 \mid X_k = i_k\} \\ &= \mathbb{P}\{X_n = i_n, \dots, X_{k+1} = i_{k+1} \mid X_k = i_k\} \cdot \mathbb{P}\{X_{k-1} = i_{k-1}, \dots, X_{i_0} = i_0 \mid X_k = i_k\} \end{aligned}$$

Let

$$p_{ij}(n \to n+1) = \mathbb{P}\{X_{n+1} = j \mid X_n = i\}$$

and define the matrix $P(n \to n+1) = [p_{ij}]$.

Proposition 14.4.2
$P(n \to n+1)$ is a *stochastic matrix.* That is,

1) $p_{ij}(n \to n+1) \in [0, 1]$
2) $\sum_j p_{ij}(n \to n+1) = 1$ for all i

Theorem 14.4.3
Let $X_n, n \in \mathbb{Z}_+$ be a Markov chain. Then X_n is completely defined by

1) The distribution of X_0, say $p_k := \mathbb{P}\{X_0 = k\}$ for every $k \in S$.
2) The sequence of transition matrices $P(n \to n+1)$ provided that they are stochastic.

Proof (Sketch)
For the forward direction, we can decompose $P\{X_j = i_j, 0 \leq j \leq n\}$ as a product of transitions. Conversely, we can check that the two conditions define a permutation invariant, consistent, and Markov system of finite dimensional distributions.

Definition 14.4.2 (Homogeneous Markov Chain)
We say a Markov chain is *homogeneous* or *stationary*, if the transition matrix is the same across all steps.

Example 14.4.4 (Homogeneous Markov Chains)

1) Sequence of independent events
2) Random walk
3) Random walks with absorbing / reflecting barriers

Theorem 14.4.5 (Chapman-Kolmogorov Equations)

1) For a stationary Markov chain, the n-step transition matrix is just P^n.
2) For a not-necessarily homogenous Markov chain,

$$P(k \to m) = P(k \to u)P(u \to m)$$

for all $k < u < m$.

Corollary 14.4.5.1
We have

$$\begin{aligned}
&\mathbb{P}\{X_0 \in A_0, X_{k_1} \in A_{k_1}, \ldots, X_{k_\ell} \in A_\ell\} \\
&= \sum_{i_0 \in A_0, i_{k_j} \in A_k, j \in [\ell]} p_{i_0} p_{i_0 \to i_{k_1}}(0 \to k_1) \cdot \cdots \cdot p_{i_{k_{\ell-1}} \to i_{k_\ell}}(k_{\ell-1} \to k_\ell).
\end{aligned}$$

More on Markov Chains

There is further theory on Markov chains such as recurrency and transience. For instance, random walks in dimensions 1, 2 are recurrent (returns almost surely), while for dimensions 3 or higher, they are transient (never return). We can also study ergodicity and asymptotic properties.

14.4.2 Markov Processes

Consider now $T = \mathbb{R}_+$ and $S = \mathbb{R}$. We write

$$\mathcal{F}_s^t := \sigma\{X_u(\omega) : s < u < t\}.$$

Thus $\mathcal{F}_0^s$ is the σ-algebra of the past and $\mathcal{F}_s^\infty$ is the σ-algebra of the future.

Definition 14.4.3 (Markov Process)
A continuous time process $X_t, t \in \mathbb{R}_+$ is *Markov* if for every $s < t$ and $A \in \mathcal{B}$,

$$\mathbb{P}\{X_t \in A \mid \mathcal{F}_0^s\}(\omega) = \mathbb{P}\{X_t \in A \mid \mathcal{F}_s^s\}(\omega).$$

The definition above is difficult to work with. Thus we wish to consider an analogue of the transition matrix.

Definition 14.4.4 (Transition Probability)
Suppose there is a function $P(s, x, t, A)$ such that for every s, t and $A \in \mathcal{B}$,

$$\mathbb{P}\{X_t \in A \mid \mathcal{F}_s^s\}(\omega) = P(x, X_t(\omega), t, A),$$

then P is the *transition probability*.

Note that P should be measurable in x. Moreover, for every s, t, x, it should be a measure in A. Finally, it should satisfy the Chapman-Kolmogorov equation

$$P(s, x, u, A) = \int_{-\infty}^{\infty} p(s, x, t, dy) p(t, y, u, A).$$

For instance, if there exists a transition density $p(s,t,x,y)$ such that

$$P(s,X,t,A) = \int_A p(s,t,x,y)dy,$$

then the Chapman-Kolmogorov equations is as follows.

$$P(s,x,u,z) = \int_{-\infty}^{\infty} p(s,x,t,y)p(t,y,u,z)dy.$$

Note that Chapman-Kolmogorov guarantees the consistency of finite-dimensional distributions.

If transition densities exist,

$$\begin{aligned}
&\mathbb{P}\{X_0 \in A_0, X_{t_j} \in A_j, j \in [\ell]\} \\
&= \int_{A_0} \cdots \int_{A_\ell} f_0(x_0)p(0,x_0,t_1,x_1)\dots p(t_{\ell-1},x_{\ell-1},t_\ell,x_\ell)dx_0dx_1\dots dx_\ell.
\end{aligned}$$

Here f_0 is the density of X_0.

Proposition 14.4.6
Markov processes with transition densities ar uniquely defined by transitions probabilities given the Chapman-Kolmogorov equations.

Example 14.4.7
Any process with independent increments is Markov.

Brownian motion is a Markov process. More generally, any Lévy process is Markov.

14.5 Markov Diffusion

Markov Diffusion Processes are Markov processes with particularly well-behaved sample paths. We omit the precise definition but give some sufficient conditions to be a diffusion proces.

Proposition 14.5.1
A markov process X_t with transition densities is a diffusion process if

1) The limit (drift coefficient)

$$a(s,x) = \lim_{t \to s^-} \int_{-\infty}^{\infty} \frac{(y-x)p(s,x,t,y)}{(t-s)} dy$$

exists.

2) The limit (drift coefficient)

$$b(s,x) = \lim_{t \to s^-} \int_{-\infty}^{\infty} \frac{(y-x)^2 p(s,x,t,y)}{(t-s)} dy$$

exists.

3) We have continuous trajectories, ie

$$\int_{-\infty}^{\infty} |y-x|^{2+\delta} p(s,x,t,y) dy = o(|t-s|).$$

14.5.1 Kolmogorov Forward-Backward Equations

Consider the operators

$$\begin{aligned} \mathscr{D}_{a,b}(\cdot) &= \frac{1}{2} b(s,x) \frac{\partial^2}{\partial x^2}(\cdot) + a(s,x)\frac{\partial}{\partial x}(\cdot) \\ \tilde{\mathscr{D}}_{a,b}(\cdot) &= \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b(t,y)\cdot\right) - \frac{\partial}{\partial y}\left(a(t,y)\cdot\right). \end{aligned}$$

The idea is that with s, x fixed,

$$\frac{\partial}{\partial t} p(s,x,t,y) = \mathscr{D}_{a,b} p(s,x,t,y).$$

Example 14.5.2
Brownian motion has drift coefficient $a(s,x) = 0$ and diffusion coefficient $b(s,x) = 1$. It follows that

$$\mathscr{D}_{a,b} = \frac{1}{2}\Delta.$$

14.6 Convergence of Stochastic Processes

14.6.1 The Space $C[0,\infty)$ & Convergence

The σ-algebra generated by cylinder sets is not "nice" to work with when discussing notions of convergence. Instead, we restrict ourselves to the processes with continuous trajectories and take the Borel σ-algebra under some metric.

We write

$$C[0,\infty)$$

to denote the set of continuous functions $f:[0,\infty)\to\mathbb{R}$ and we equip it with the metric

$$\rho(f,g) := \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0\le t\le n} \{|f(t)-g(t)| \wedge 1\}.$$

Remark that $\rho(f,g)\in[0,1]$.

Proposition 14.6.1
We have

$$\mathcal{B}_{C[0,\infty)} = \mathscr{C}_{\mathbb{R}^{[0,\infty)}}\Big|_{C[0,\infty)}.$$

Thus random processes with continuous trajectories are random elements of $(C[0,\infty),\mathcal{B}_C)$.

We wish to study the weak convergence of measures, or in other words, the convergence in distribution of random processes $X_t^{(n)} \xrightarrow{p} X_t$. Recall that convergence in distribution is defined as the convergence

$$\mathbb{E}F(X_t^{(n)}) \to \mathbb{E}F(X_t)$$

for every bounded continuous $F: C[0,\infty)\to\mathbb{R}$, ie

$$\int_{C[0,\infty)} F(f)d\mathbb{P}_n(f) \to \int_{C[0,\infty)} F(f)\mathbb{P}(f).$$

Recall that sequence of measures $\mathbb{P}_n$ is tight if every subsequence contains a convergene subsequence converging to a probability measure. It is equivalent to relative compactness, which states that there exists some K_ϵ for which $\mathbb{P}_n(K_\epsilon) > 1-\epsilon$.

14.6.2 Convergene of Processes

Proposition 14.6.2
If $X_t^{(n)} \xrightarrow{w} X_t$ and $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ in $(C[0,\infty), \mathcal{B}_C)$,

1) $\mathbb{P}_n$ is relatively compact and tight
2) For every $t_1, \ldots, t_d \in [0, \infty)$,

$$\left(X_{t_1}^{(n)}, \ldots, X_{t_d}^{(n)}\right)(\omega) \xrightarrow{d} (X_{t_1}, \ldots, X_{t_d})(\omega).$$

Proof
To see 2), let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous and bounded. Define $F : C[0,\infty) \to \mathbb{R}$ given by

$$F(g) := f(g(t_1), \ldots, g(t_d)).$$

Remark that $F(X_t^{(n)}(\omega)) = f(X_{t_1}^{(n)}(\omega), \ldots, X_{t_d}^{(n)}(\omega))$.

Theorem 14.6.3
Let $X_t^{(n)}$ be a sequence of random elements of $(C[0,\infty), \mathcal{B})$ that is

1) tight (or relatively compact)
2) convergent in terms of finite dimensional distributions

Then $X_t^{(n)} \xrightarrow{d} X_t$ and $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ where X_t is the process given by the limiting finite dimensional distributions.

We note that in general, tightness is a difficulty property to show. Luckily, there are criterions for the tightness of processes.

14.6.3 Convergene of Random Walks to Brownian Motion

Let $\xi_i, i \in \mathbb{N}$ be iid with zero-mean and variance σ^2. Define $S_0 := 0$ and

$$S_k(\omega) := \sum_{i=1}^{k} \xi_i(\omega).$$

Consider the process

$$Y_t(\omega) := S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1}$$

and rescale it so to

$$X_t^{(n)}(\omega) := \frac{1}{\sigma\sqrt{n}} Y_{\lfloor tn \rfloor}$$

for $t \geq 0$.

Theorem 14.6.4 (Donsker)
The scaled random walk $X_t^{(n)}$ converges in distribution to Brownian motion B_t and $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_B$ on $(C[0,\infty), \mathcal{B}_C)$ a random elements.

We note that $\mathbb{P}_B$ on $(C, \mathcal{B}_C)$ is at times known as the *Wiener measure*.

Donsker's theorem is also known as the functional central limit theorem.

Chapter 15

Stochastic Calculus

15.1 Motivation

Recall that a Brownian motion $B_t(\omega)$ is a stochastic process with independent stationary increments and a.s. ($\alpha < \frac{1}{2}$ Hölder) continuous trajectories. Moreover, it is a square-integrable martingale. Indeed, for $s < t$,

$$\begin{aligned}
\mathbb{E}\{B_t \mid \mathcal{F}_s\} &= \mathbb{E}\{B_s + (B_t - B_s) \mid \mathcal{F}_s\} \\
&= \mathbb{E}B_s + \mathbb{E}[B_t - B_s] && B_s \in \mathcal{F}_s, \text{independence} \\
&= \mathbb{E}B_s.
\end{aligned}$$

Finally, there exists some $\langle B_t \rangle$ such that $B_t^2 - \langle B_t \rangle$ is martingale.

Proposition 15.1.1
For $S = t_0 < t_1 < \cdots < t_n = T$,

$$\lim_{\Delta t \to 0} \sum_{k=0}^{n-1} (B_{t_{k+1}}(\omega) - B_{t_k}(\omega))^2 \xrightarrow{p} T - S.$$

Proof
We have equality under expectation and the variance tends to 0. Thus we can apply Chebyshev's inequality.

Note that we can actually show convergence a.s. More importantly however, the total variation

$$\lim_{\Delta t \to 0} \sum_{k=0}^{n-1} |B_{t_{k+1}}(\omega) - B_{t_k}(\omega)| \geq T - S > 0.$$

Hence B_t is a.s. nowhere differentiable and we have no hope of applying for example the

Riemann–Stieltjes integral.

Still, we would like some notion of a derivative, some definition of differential equations, as well as calculus.

15.2 Construction of the Îto Integral

15.2.1 The Space

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Our space carries a Brownian motion B_t that is adapted to some filtration $\mathcal{F}_t, t \in [0, T]$.

Remark that B_t is a $\mathcal{F}_t$-martingale.

15.2.2 "Integrable" Processes

Consider an arbitrary $f : [0, T] \times \Omega \to \mathbb{R}$. We would like f to be adapted, ie $f(t, \cdot) \in \mathcal{F}_t$ for every t, and square integrable, ie

$$\int_0^T \int_\Omega f(t, \omega)^2 d\mathbb{P} dt = \|f\|_2 < \infty.$$

Write $H_2[0, T]$ to be the set of adapted functions from $L^2([0, T] \times \Omega)$.

15.2.3 Desired Properties

For our desired integral

$$\int_0^T f(t, \omega) dB_t,$$

we wish for it to be a random variable that is $\mathcal{F}_T$ measurable.

Moreover, we would like the following to be satisfied:

(i) Linearity

(ii) $\int_0^T \mathbb{1}_{[s,u]}(t, \omega) dB_t = (B_u - B_s)(\omega)$

(iii) $\mathbb{E}\left[\int_0^T f(t, \omega) dB_t\right] = 0$

(iv)

$$\mathbb{E}\left[\int_0^T f(t, \omega) dB_t\right]^2 = \|f\|_{H_2}^2 := \int_0^T \int_\Omega f(t, \omega)^2 d\mathbb{P}(\omega) dt < \infty$$

First we note that $\int_0^T f(t,\omega)dB_t$ depends on both T, ω thus it is a process. Moreover, The integral is an isometry between $H_2[0,T]$ and $L^2(\Omega)$. Hence many properties can be divined simply by looking at $H_2[0,T]$.

15.2.4 Simple Processes

Similar to the construction of other integrals, we begin with easily integrable functions.

Definition 15.2.1 (Simple Process)
For $f \in H_2[0,T]$, we say that f is *simple* if there are $0 = t_0 < t_1 < \cdots < t_n = T$ such that

$$f(t,\omega) = \sum_{k=0}^{n-1} f(t_k,\omega)\mathbb{1}\{t \in [t_k, t_{k+1})\}.$$

Here $f(t_k, \cdot) \in \mathcal{F}_{t_k}$.

It is clear that f is square-integrable

$$\|f\|_{H_2[0,T]} = \sum_{k=0}^{n-1}(t_{k+1} - t_k)\mathbb{E}\left[f(t_k,\omega)^2\right] < \infty.$$

Moreover, it is adapted by definition.

For such a simple function, we define

$$\int_0^T f(t,\omega)dB_t := \sum_{k=0}^{n-1} f(t_k,\omega)[B_{t_{k+1}}(\omega) - B_{t_k}(\omega)].$$

Note that this definition allows refinement of partitions as it does not chance the value of our random variable. This is due to the fact that B_t has independent increments.

Proposition 15.2.1
The integral satisfies all 4 desirable properties.

Proof
The proof of (i), (ii) is straightforwards. For (iii), we can directly compute

$$\mathbb{E}\left[\int_0^T f(t,\omega)dB_t\right] = \sum_k \mathbb{E}[f(t_k,\omega)] \cdot E[B_{t_{k+1}} - B_{t_k}] = 0$$

by independence.

The isometric property is again proven by computation using independence.

15.2.5 Extending to all $H_2[0,T]$

Proposition 15.2.2
For any $F \in H_2[0,T]$, there is a sequence of simple processes $f_n \uparrow F$ such that

$$\|f_n - F\|_{H_2}^2 \to 0$$

as $n \to \infty$.

Now, Recall that our integral is an isometry. Hence the image $\int_0^T f_n dB_t$ is Cauchy in L^2 since f_n is Cauchy in H_2. By the completeness of $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, there exists a unique limit $\int_0^T F dB_t \in L^2$. We take this element to be the integral of F.

Proposition 15.2.3
We have

$$\int_0^T B_t dB_t = \frac{B_T^2 - T}{2}.$$

Proposition 15.2.4
Our integral satisfies all 4 desirable properties for every $f \in H_2[0,T]$.

Note that with more work, we can extend out integral to $T = \infty$.

For sub-intervals $[A,B) \subseteq [0,T]$, we have for every $F \in \mathcal{F}_B$,

$$\int_A^B F dB_t := \int_0^T \mathbb{1}_{[A,B)} F dB_t.$$

It follows that

$$\int_A^B + \int_B^C = \int_A^C .$$

Now, we have

$$\mathbb{E}\left\{\int_A^B F dB_t \mid \mathcal{F}_A\right\} = 0.$$

We can see this since it certainly holds for simple functions, and L^2 convergence implies L^1 convergence in finite measures

$$\|g\|_{L^1} = \|\mathbb{1}_\Omega g\|_{L^1} \le \mathbb{P}(\Omega) \cdot \|g\|_{L^2}.$$

But then

$$\mathbb{E}\left\{\int_0^T F dB_t \mid \mathcal{F}_s\right\} = \int_0^s F dB_t$$

and the stochastic integral is a martingale.

15.3 Stochastic Differentials

Definition 15.3.1 (Stochastic Differential)
Let $Z_t(\omega)$ for $t \in [0,T]$. Suppose that there are $a(s,\omega) \in L^1[0,T]$ and $\sigma(s,\omega) \in H_2[0,T]$ such that

$$Z_t(\omega) - Z_0(\omega) = \int_0^t a(s,\omega)ds + \int_0^t \sigma(s,\omega)dB_s$$

for every $t \in [0,T]$.
Then we define

$$dZ_t(\omega) = a(t,\omega)dt + \sigma(t,\omega)B_t.$$

Example 15.3.1
$dB_t = dB_t$. Indeed,

$$B_t(\omega) - B_0(\omega) = \int_0^t \mathbb{1}[0,t]dB_t = B_t(\omega) - B_0(\omega).$$

Example 15.3.2
$dB_t^2 = 2B_t dB_t + 1 \cdot dt$. We can see this since $\int_0^t B_s(\omega)dB_s = \frac{B_t^2 - t}{2}$. Hence

$$B_2^2 - B_0^2 = 2\int_0^t B_s dB_s + \int_0^1 1dt.$$

Lemma 15.3.3 (Îto)
Let $f \in C^2(\mathbb{R},\mathbb{R})$ be deterministic. Then

$$df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt.$$

In other words,

$$f(B_t) - f(B_0) = \int_0^t f'(B_t(\omega))dB_t + \frac{1}{2}\int_0^t f''(B_t(\omega))dt.$$

Proof (Idea)
Let us consider the special case of $f \in C^2$ with bounded derivatives $|f'|, |f''|, |f'''| < C \in \mathbb{R}_+$.

Consider the partition $0 = t_0 < t_1 < \cdots < t_n = t$ as $n \to \infty$. By Taylor's theorem,

$$
\begin{aligned}
f(B_t) - f(B_0) &= \sum_{k=0}^{n-1} [f(B_{t_{k+1}}) - f(B_{t_k})] \\
&= \sum_{k=0}^{n-1} f'(B_{t_k})[B_{t_{k+1}} - B_{t_k}] && (i) \\
&\quad + \frac{1}{2} \sum_{k=0}^{n-1} f''(B_{t_k})[(B_{t_{k+1}} - B_{t_k})^2 - (t_{k+1} - t_k)] && (ii) \\
&\quad + \frac{1}{2} \sum_{k=0}^{n-1} f''(B_{t_k})(t_{k+1} - t_k) && (iii) \\
&\quad + \frac{1}{2} \sum_{k=0}^{n-1} f'''(B_{\tilde{t}_k(\omega)})[B_{t_{k+1}} - B_{t_k}]. && (iv)
\end{aligned}
$$

For some $\tilde{t}_k(\omega) \in (t_k, t_{k+1})$.

By construction,

$$(i) = \int_0^T f_n'(B_t) dB_t \xrightarrow{L_2} \int_0^T f'(B_t) dB_t.$$

In addition,

we have shown that

$$(ii) \xrightarrow{p} C \cdot 0$$

and in fact, a.s. convergence holds. Recall this is due to the fact that it is 0 in expectation and the variation tends to 0.

Now,

$$(iii) \to \frac{1}{2} \int_0^t f''(B_t) dt$$

as it is just the normal Riemann sum.

Finally,

$$(iv) \le \frac{1}{6} \max_k [B_{t_{k+1}} - B_{t_k}] C \le \sum_k [B_{t_{k+1}} - B_{t_k}]^2 \xrightarrow{L^2} 0.$$

With Îto's lemma in hand, we can prove some standard theorems.

Theorem 15.3.4
Suppose

$$dX_t = a(t,\omega)dt + \sigma(t,\omega)dB_t.$$

Then for all $f(t,x) \in C^2$,

$$df(t,X_t) = f_t'(t,X_t(\omega))dt + f_x'(t,X_t)dX_t + \frac{1}{2}f_{xx}'(t,X_t)\sigma^2(t,\omega)dB_t.$$

we can also derive formulas for $df(t,X_t,Y_t)$, and so on.

15.4 Stochastic Differential Equations

15.4.1 Motivation

Consider an ODE

$$dX(t) = a(t,X(t))dt.$$

This models bond prices or assets prices.

Suppose now that there is some white noise perturbations

$$dX_t(\omega) = a(t,X_t)X_t(\omega)dt + \sigma(t,X_t(\omega))dB_t.$$

We can think of stock prices.

As a side note, we would like to make sense of *Gaussian White Noise (GWN)* as a stationary Gaussian process with covariance being the delta function. However, this does not exist a stochastic process. Instead, the so called "mean-squared derivative" of Brownian motion is GWN. Informally speaking,

$$\mathbb{E}\left\{\frac{B_{t+s}-B_t}{2} - \mathrm{GWN}_t\right\}^2 \to 0.$$

In applications, we like to heuristically reason about dB_t as GWN.

15.4.2 Theory

A *Stochastic Differential Equation* is written as

$$dX_t(\omega) = a(t,X_t(\omega))dt + \sigma(t,X_t(\omega))dB_t.$$

Equivalently, recall that this means by definition

$$X_t(\omega) = X_0(\omega) + \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

for every $t \in [0, T]$. Thus a SDE is actually an integral equation.

Definition 15.4.1 (Solving a SDE)
We say that $X_t(\omega)$ *solves* the SDE with initial conditions $X_0(\omega)$ if the integral condition holds. In other words,

$$X_t(\omega) = X_0(\omega) + \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

Theorem 15.4.1 (Existence of a Unique Solution)
Suppose the there is a constant $K > 0$ such that the following hold.

(i) $|a(t,x) - a(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le K|x - y|$

(ii) $a^2(t,x) + \sigma^2(t,x) \le K(1+x)^2$

(iii) $X_0 \in \mathcal{F}_0$ and has finite variance

Then there exists a solution to the SDE that is continuous and unique up to indistinguishability.

We remark the proof follows the same lines as for an ODE. We define some iterates

$$\begin{aligned} X_t^{(0)} &:= X_0(\omega) \\ &\dots \\ X_t^{(n)}(\omega) &:= X_0(\omega) + \int_0^t a(s, X_s^{(n-1)}(\omega))ds + \int_0^t \sigma(s, X_s^{(n-1)}(\omega))dB_s. \end{aligned}$$

We can show that $X_t^{(n)}(\omega)$ is Cauchy in $C[0,T]$ a.s. and that the limit is a solution.

Theorem 15.4.2
Under the conditions of the previous theorem, the solution $X_t(\omega)$ is a Markov process. Moreover, it is a Markov diffusion with drift and diffusion coefficients

$$\begin{aligned} a(s,x) &= \mathbb{E}\{X_t - X_s \mid X_s = x\}\Big|_{t=s} \\ \sigma^2(s,x) &= \mathbb{E}\big\{(X_t - X_s)^2 \mid X_s = x\big\}\Big|_{t=s}. \end{aligned}$$

We conclude by remarking that many interesting processes are implicitly described by SDEs and hence are diffusions. Although we may not be able to find a closed-form solution for these SDEs, we can study their distribution and other properties based on the SDE that they solve.