18.675: Theory of Probability
Cheat Sheet

Brian Lee

12/15/25

Rough Overview

Roughly we divide things into things to know how to prove and things we simply stated:

(a) **Things we stated but didn't prove:** Caratheodory extension theorem, Construction of Lesbegue Integral, Properties of Lesbegue integral, Monotone Convergence Theorem, Fubini's Theorem, Jensen and Holder inequalities, Gaussian approximation of L^p function through convolution, integral preservation under ergodic functions, Birkhoff's Ergodic theorem, Example of martingale that is L^1 bounded and converges a.s. but not in L^1

(b) **Things we proved:** Dynkin's $\pi - \lambda$ lemma, Unique extension of a measure on a π-system for finite measures, Existence of the Lesbegue measure, Proof of translation equivariance of Lesbegue, Equivalence of indpenedence of σ-algebras generated by π-systems to independence of the π-systems, First and Second Borel-Cantelli lemmas, neccesary and sufficient conditions for measurability of a function, Monotone class theorem, conditions for independence of collection of random variables, Convergence a.e. implies convergence in measure for finite measure spaces, convergence in measure implies convergece a.e. along a subsequence, Kolmogorov's $0 - 1$ law, Properties of Integrable Functions, Swapping integral with summation, Fatou's Lemma, Strictness of the inequality in Fatou's Lemma, Dominated Convergence, existence of product measure, Markov's inequality, orthogonal decomposition in L^2, bounded convergence theorem, convergence in probability does not imply convergence in L^1, uniform integrability of L^p-boudned family of random variables, uniform integrability characterization as tail probabilities, L^1 convergence equivalent to uniform integrability, Convolution is L^p, Convolution theorem for L^1 variables, Fourier inversion formula for general L^1 functions, Plancherel's identity for L^1 functions, characteristic function determines the law, equivalence of characteristic function convergence to weak convergence of random variable, Ergodicity of shift-map in a Bernoulli scheme, von-Neumann's ergodic theorem, WLLN, SLLN on fourth moment bound, SLLN for iid variables, CLT, Existence of conditional expectation, Conditional Monotone Convergence, Conditional Fatou's Lemma, Conditional Dominated Convergence, Conditional Jensen's Inequality, "Tower property" of conditional expectation, optional stopping theorem, simple random walk hits $-a$ before b with probability $b/(a+b)$, non-negative supermartingale converges a.s. towards an a.s. finite limit, Doob's upcrossing inequality, a.s. marginale convergence theorem, Doob's maximal inequality, Doob's L^p-inequality, martingale convergence in L^p theorem, proof that $\mathbb{E}[X|\mathcal{G}]$

is UI, martingale convergence in L^1-theorem, UI OST, a.s. backwards L^p martingale convergence theory, SLLN and Kolomogorov $0-1$ law with martingales, Kakutani's product martingale theorem, Radon-Nikodyn Theorem for countably generated σ-algebras

1. L1-5: Measure Theory

1.1 Measure Theory Basics

We start with the formal definitions of π and d-systems.

Definition 1.1 (*σ-algebras*)
For a set E, a **σ-algebra** $\mathcal{E}$ on E is a set of subsets of E such that

i. **(Empty set)** $\emptyset \in \mathcal{E}$

ii. **(Complements)** $\forall A \in \mathcal{E} \implies A^c \in \mathcal{E}$

iii. **(Countable Intersection)** $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E} \implies \cup_{n=1}^{\infty} A_n \in \mathcal{E}$

Definition 1.2 (*Measurable Spaces*)
A pair $(E, \mathcal{E})$ is called a **measurable space**.

Given a measurable space, one can induce a measure onto the space through a countably additive function as follows;

Definition 1.3 (*Measure on a Measurable Space*)
A function $\mu : \mathcal{E} \to [0, \infty)$ is called a **measure** on the measurable space $(E, \mathcal{E})$ if the following conditions hold:

i. (Zero-Measure) $\mu(\emptyset) = 0$

ii. (Countable Additivity) $\forall (A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ disjoint, we have

$$\mu(\cup_n A_n) = \sum_n \mu(A_n)$$

Equipped with a measure, a measurable space turns into a "measure space":

Definition 1.4 (*Measure Space*)
A pair $(E, \mathcal{E}, \mu)$ where $(E, \mathcal{E})$ is a measurable space and μ is a measure on $(E, \mathcal{E})$ is called a **measure space**.

Definition 1.5

In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we call Ω the **sample space** and $\mathcal{F}$ the **events**. For $A \in \mathcal{F}$, we call $\mathbb{P}[A]$ the **probability** of A.

We can naturally break down σ-algebras into two "smaller" components, namely the π and d-systems:

DEFINITION 1.6

For countable E, let $\mathcal{E} = 2^E$ define a mass function $m : E \to [0, \infty)$. If μ is a measure on $(E, \mathcal{E})$ we have

$$\mu(A) = \sum_{x \in A} \mu(\{x\}) \tag{1.1}$$

Defining $\mu(\{x\}) = m(x)$ gives the **discrete measure**.

We now break down the notion of a σ-algebra into two notions, a notion of "intersectingness" in the π-system and a notion of "quotiontness" in the d-system:

DEFINITION 1.7 (*π and d-systems*)

Let $A \subseteq 2^E$. $\mathcal{A}$ is a **π-system** if

(a) $\emptyset \in \mathcal{A}$

(b) $\forall A, B \in \mathcal{A}, A \cap B \in \mathcal{A}$

$\mathcal{A}$ is called a **d-system** or a **λ-system** if

(a) $E \in A$

(b) $\forall A, B \in \mathcal{A}, A \subseteq B, B \backslash A \in \mathcal{A}$

(c) $\forall\ A_j$ increasing, $\cup_j A_j \in \mathcal{A}$

THEOREM 1.1

A system that is both a π and a d system is a σ-algebra.

LEMMA 1.2 (*Dynkin's $\pi - \lambda$ lemma*)

If $\mathcal{A}$ is a π-system, then any d-system that contains $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.

Proof. Let $\mathcal{D}$ be the intersection of all d-systems. This thing is a σ-algebra and hence we are done. □

The assumption that $\mathcal{D}$ is a d-system is required: a simple counterexample is the set of integers and the integers greater than points in your set.

Definition 1.8
For $\mathcal{A} \subseteq \mathcal{P}(E)$, we define the **$\sigma$-algebra generated by $\mathcal{A}$** as the intersection of all σ-algebras containing A.

1.2 Construction of the Lesbegue Measure

We now focus our attention to constructing a "canonical" measure on $\mathbb{R}$, namely the one that measures $\mathbb{R}$ in the obvious way.

Definition 1.9 (*Rings*)
We have $\mathcal{A} \subseteq 2^E$ is called a **ring** if

1. $\emptyset \in \mathcal{A}$
2. $\forall A, B \in \mathcal{A}, B \backslash A \in \mathcal{A}, A \cup B \in \mathcal{A}$

Definition 1.10 (*Algebras*)
We have $\mathcal{A} \subseteq 2^E$ is called an **algebra** if

1. $\emptyset \in \mathcal{A}$
2. $\forall A, B \in \mathcal{A}, B \backslash A \in \mathcal{A}, A \cup B \in \mathcal{A}$

The following theorem will allow us to define a unique Borel measure with the Lesbegue property:

Theorem 1.3 (*Caratheodory Extension Theorem*)
Let $\mathcal{A}$ be a ring on E and μ is a countably additive set functions on $\mathcal{A}$. Then μ extends uniquely to a measure on $\sigma(\mathcal{A})$.

Definition 1.11 (*Borel Measures*)
If E is Hausdorff Topological Space. We define the **Borel σ-algebra** of E to be the σ-algebra generated by the set of open sets U, denoted $\mathcal{B}(E)$. A measure μ on $\mathcal{B}(E)$ is called a **Borel measure** and if for all compact K we have $\mu(K) < \infty$ then μ is said to be a **Radon measure**.

Theorem 1.4 (*Existence and Uniqueness of the Lesbegue Measure*)
There exists a unique Borel measure μ on $\mathcal{B}(\mathbb{R})$ with $\mu([a, b]) = b - a$. The unique such measure is called the **Lesbegue Measure** on $\mathbb{R}$.

Proof. Uniqueness follows from defining $\mu_n = \mu((n, n+1] \cap A)$ and similar for λ_n and using Dynkin's $\pi - \lambda$ lemma on the π-system $\{(n, n+1]\}$ which implies $\mu_n = \lambda_n$ and thus $\mu = \lambda$

globally. Existence follows from defining he ring of finite unions of disjoint intervals and then using Carahtheodory to extend to a measure on the σ-algebra. This follows from Bolzano Weierstrass. □

PROPOSITION 1.5

The Lesbegue measure is **translation invariant**, ie. if $B \in \mathcal{B}$, $x \in \mathbb{R}$,

$$\mu(B + x) = \mu(B) \tag{1.2}$$

PROPOSITION 1.6 (*Translational Converse*)

Let $\tilde{\mu}$ be a Borel measure on $\mathbb{R}$ that is translation invariant and $\tilde{\mu}([0,1]) = 1$. Then $\tilde{\mu}$ is Lesbegue.

Proof. Chunk the space into rational intervals. □

DEFINITION 1.12

Let $(E, \mathcal{E})$ be a measure space and μ a measure on $\mathcal{E}$. We say μ is **finite** if $\mu(E) < \infty$ and σ-finite if $E = \cup E_n$ for some $\mu(E_n) < \infty$.

1.3 Construction of the Lesbegue-Stieltjes Measure

DEFINITION 1.13

For $(E, \mathcal{E})$ and $(G, \mathcal{G})$ we say $f : E \to G$ is **measurable** if $\forall A \in \mathcal{G}$, $f^{-1}(A) \in \mathcal{E}$. If $G = \mathbb{R}$ (Borel) we say f is a **measurable function on** E while if $G = [0, \infty)$ we have a **non-negative measurable function on** E.

THEOREM 1.7

If $f : E \to G$ with $\mathcal{G} = \sigma(\mathcal{A})$ and f is measurable on $\mathcal{A}$ then f is measurable.

Continuous functions on a topological space are always measurable.

THEOREM 1.8

We have for (f_n) nonnegative measurable on E (or real-valued), $f_1 + f_2$, $f_1 f_2$, $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$, and $\limsup_n f_n$ are all measurable.

The following theorem essentially says that for vector spaces of bounded functions that are closed under monotone limits contains every bounded measurable function.

Theorem 1.9

Let $(E, \mathcal{E})$ be a measurable space and $\mathcal{A}$ be a π-system generting $\mathcal{E}$ and let $\mathcal{V}$ be a vector space of bounded functions $f : E \to \mathbb{R}$ such that

1. $1 \in \mathcal{V}$ and $1_A \in \mathcal{V}$ for all $A \in \mathcal{V}$
2. If $f_n \in \mathcal{V}$ for all n and f is bounded with $0 \leq f_n \uparrow f$ then $f \in \mathcal{V}$

Then $\mathcal{V}$ contains every bounded function.

We now define a notion of "pushing forward" on one space onto another.

Definition 1.14

Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable functions and let μ be a measure on $\mathcal{E}$. For any measurable function $f : E \to G$, we define the **image measure** or **pushforward** $\nu := \mu \circ f^{-1}$ on $\mathcal{G}$ given by $\nu(A) = \mu(f^{-1}(A))$ for $A \in \mathcal{G}$.

We now have that the threshold functions for any non-decreasing right-continuous function is also left-continuous and non-decreasing.

Theorem 1.10 (*Existence of the Lebesgue-Stietjes Measure*)

Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, right-continuous, and non-decreasing. Then, there exists a unique Radon measure dg on $\mathbb{R}$ such that $\forall a, b \in \mathbb{R}$ with $a < b$

$$dg([a, b]) = g(b) - g(a) \tag{1.3}$$

Moreover, all such non-zero Radon measures on $\mathbb{R}$ can be obtained this way. We call dg the **Lebesgue-Stietjes measure** associated with g.

1.4 Borel-Cantelli Lemmas

Definition 1.15 (*limsup and liminf of sets*)

Given a sequence of events (A_n) we define the lim sup and lim inf of the sequence of events:

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m \qquad \liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m \tag{1.4}$$

We call $\limsup A_n = \{A_n \text{ infinitely often (i.o.)}\}$ and $\liminf A_n = \{A_n \text{ eventually (e.v.)}\}$.

Lemma 1.11 (*Borel-Cantelli Lemma*)
Let (A_n) be a sequence of events.

(a) If $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}[A_n \text{ i.o.}] = 0$

(b) If (A_n) are independent and $\sum_n \mathbb{P}[A_n] = \infty$ then $\mathbb{P}[A_n \text{ i.o.}] = 1$.

Proof. a) Union Bound. b) Demorgan's Law, Union bound, independence of A_n^C, $1 - x \leq \exp(-x)$ to finish. □

2. L6-7: Probability Spaces and Random Variables, Independence, Convergence Theorems

Definition 2.1 (*Probability Space*)
A **probability space** is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}(\Omega) = 1$.

Definition 2.2 (*Independence of Events*)
A sequence of events $(A_n) \in \mathcal{F}$ is said to be **independent** if $\mathbb{P}[\cap_{n\in\mathcal{J}} A_n] = \prod_{n\in\mathcal{J}} \mathbb{P}[A_n]$ for $\mathcal{J} \subseteq \mathbb{N}$.

Definition 2.3 (*Independence of σ-algebras*)
A sequence of σ-algebras $(\mathcal{A}_n)$ with $\mathcal{A}_n \subseteq \mathcal{F}$ is said to be **independent** if for any (A_n) with $A_n \in \mathcal{A}_n$ we have independence. The connection comes from the fact that (A_j) is independent iff the σ-algebras generated by them are.

Theorem 2.1 (*Independence of π-systems*)
Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are π-systems, if

$$\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2] \qquad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \tag{2.1}$$

then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2$ are independent.

Proof. Define two measures $\mu(A) = \mathbb{P}[A \cap A_1]$ and $\nu(A) = \mathbb{P}[A]\mathbb{P}[A_1]$ for $A_1 \in \mathcal{A}$ and use Dynkin to finish. □

We can define random variables on probability spaces:

Definition 2.4
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. Then, an ***E*-valued random variable** is a measurable function $X : \Omega \to E$.

Think of E as the "value space" for your random variable. For example, if Ω is the possible results of 10 berounlli coin flips, then a natural $E = \{0, 1, \cdots, 10\}$. Another typical value space is $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

DEFINITION 2.5 (*Distribution of a Random Variable*)
Given a random variable $X : \Omega \to E$, the **distribution** or **law of X**, $\mu_X : E \to \mathbb{R}$ is the image measure:

$$\mu_X = \mathbb{P} \circ X^{-1}$$

denoted $\mathbb{P}[X \in A] = \mu_X(A)$.

DEFINITION 2.6 (*Distribution function of a $\mathbb{R}$-valued random variable*)
If $E = \mathbb{R}$, μ_X is determined by $\mu_X((-\infty, y])$ for $y \in \mathbb{R}$ by countable additivity. We set $F_X(y) := \mu_X((-\infty, y])$. F_X is called the **distribution function of X** or equivalently the **cumulative distribution function (cdf) of X**.

Later (final lecture) we will define the probability density function for $\mathbb{R}$-valued random variables in terms of the Radon-Nikodym derivative.

PROPOSITION 2.2
We have for any $\mathbb{R}$-valued random variable, the distribution function $F_X(y)$ has $\lim_{y \to -\infty} F_X(y) = 0$ and $\lim_{y \to \infty} F_X(Y) = 1$. Furthermore, F_X is non-decreasing and right-continuous.

For non $\mathbb{R}$-valued random variables, we can generalize the notion of a distribution function:

DEFINITION 2.7
A non-decreasing, right-continuous function $f : \mathbb{R} \to [0, 1]$ such that $\lim_{x \to = \infty} f(x) = 0$, $\lim_{x \to \infty} f(x) = 1$ is called a **distribution function**.

Indeed, for any such distribution function there is a corresponding random variable with that distribution function as its law.

2.1 Existence of Random Variables

We now show that for any distribution function F that we can create random variables with the given law:

THEOREM 2.3
Let F be any distribution function. THen $\exists$ some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X on Ω such that $\mathcal{F}_X = \mathcal{F}$.

Proof. Take $B((0, 1))$ with Lesbegue measure. Define

$$X(\omega) := \inf\{x : \omega \leq F(x)\} \tag{2.2}$$

Then, $X(\omega) \leq x \Leftrightarrow \omega \leq F(x)$ so $F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[(0, F(x)]] = F(x)$ $\square$

Definition 2.8 (*Independence of a sequence of random variables*)
A sequence of random variables (X_n) is said to be **independent** if their σ-algebras are independent.

Proposition 2.4
Two real-valued random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent iff

$$\mathbb{P}[X \le a, Y \le b] = \mathbb{P}[X \le a]\mathbb{P}[Y \le b] \tag{2.3}$$

Proof. Definition of independence of σ-algebras. □

Definition 2.9 (*Rademacher Functions*)
There exists

TODO: rademachar functions, existence of a Bernoulli sequence

2.2 Convergence

We now define three notions of convergence in (rough) order of strongest to weakest: a.e. convergence, convergence in measure, and convergence in distribution (or weak convergence).

Definition 2.10 (*Convergence Almost Everywhere/Surely*)
Let $(E, \mathcal{E}, \mu)$ be a measure space and suppose $(f_n), f : E \to \mathbb{R}$ are measurable functions. We say that $f_n \to f$ converges μ-**almost everywhere (a.e.)** if $\mu(\{x \in E : f_n(x) \not\to f\}) = 0$.

In probability spaces $\mu = \mathbb{P}$, we say it converges almost surely (a.s.).

Definition 2.11 (*Convergence in Measure/Probability*)
Suppose that $(E, \mathcal{E}, \mu)$ is a measure space. Suppose that $(f_n), f : E \to \mathbb{R}$ are measurable. We say $f_n \to f$ convergences **in measure** if $\forall \varepsilon > 0$,

$$\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0 \tag{2.4}$$

In probability space, we say that $f_n \to f$ converges **in probability**.
The last notion of convergence concerns real-valued random variables.

Definition 2.12 (*Convergence in Distribution*)
Suppose we have a sequence (X_n), X of real-valued random variables with distribution functions F_{X_n}, F_X. We say $X_n \to X$ **in distribution** if we have $F_{X_n} \to F_x$ poiuntwise at all points at which F_X is continuous.

The main connection between the first two types of convergence are as follows: for finite measure spaces, a.e. convergence is stronger, while for the infinite measure case only a smaller guarantee can be made:

THEOREM 2.5 (*Relationship between a.e. convergence and in measure*)
Let $(E, \mathcal{E}, \mu)$ be a measure space.

(a) If the measure space $\mu(E) < \infty$, we have a.e. convergence implies $f_n \to f$ in measure

(b) If $f_n \to f$ in measure, $\exists$ a subsequence n_k such that $f_{n_k} \to f$ a.e..

Proof. (a) Fix $\varepsilon > 0$. Then it is a simple Fatou's Lemma.

(b) This is just the First Borel-Cantelli Lemma.

□

THEOREM 2.6
Let $X, (X_n)$ be real-valued random variables.

1. If X_n and X are defined on the same probability space and $X_n \to X$ in probability, then $X_n \to X$ in distribution.

2. If $X_n \to X$ in distribution, then there are random vriables $\tilde{X}$ and $(\tilde{X}_n)$ defined on a common probability space and $\tilde{X}_n \to \tilde{X}_n$ almost surely.

2.3 Kolomogorov's 0-1 Law

We now show, a certain type of event, known as a *tail- event*, happens either almost surely or doesn't happen almost surely. The intuition is essentially as follows: If you ignore any finite beginning and only look "infinitely far out," independence makes the remaining randomness so homogeneous that events in the far-future must either 0% or 100%.

DEFINITION 2.13 (*Tail σ-algebra*)
Let (X_n) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We let $\tau_n = \sigma(X_{n+1}, X_{n+2}, \cdots)$. Then, we define the **tail σ-algebra**:

$$\tau = \cap_{n \in \mathbb{N}} \tau_n$$

THEOREM 2.7 (*Kolomogorov's 0-1 Law*)
Let (X_n) be a real-valued sequence of independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the tail σ-algebra τ is trivial, ie. $\forall A \in \tau$, $\mathbb{P}[A] \in \{0, 1\}$. Moreover, if X is a τ-measurable random variable, then $\exists c \in \mathbb{R}$ such that $\mathbb{P}[X = c] = 1$, ie. it is a dirac-delta function.

3. L8-9: Integration

We now use our tools from measure theory to define a more general notion of integration known as *Lesbegue Integration*.

3.1 Basics of Lesbegue Integration

We now generalize the notion of Riemann integration using measure theory to the notion of the Lesbegue integral, allowing us to integrate more generally. We do this by defining it as a supremum of simple functions which we now describe.

Definition 3.1 (*Simple Functions*)
A function $f : E \to \mathbb{R}$ is called a **simple function** if it is of the form

$$f = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k} \tag{3.1}$$

where $0 \le a_k < \infty$ and $A_k \in \mathcal{E}$

We now define the integral first for simple functions in the natural way:

Definition 3.2 (*Lesbegue Integral for Simple Functions*)
The **Lesbegue Integral** of a simple function $f = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k}$ is defined:

$$\mu(f) = \int f d\mu := \sum_{k=1}^{m} a_k \mu(A_k) \tag{3.2}$$

We define the integral $\mu(f)$ of general real-valued functions as a supremum over simple functions:

Definition 3.3
We define the **Lesbegue integral** $\mu(f)$ of a non-negative measurable functions f by:

$$\mu(f) := \int f d\mu = \sup\{\mu(g) : g \text{ simple}, g \le f\} \tag{3.3}$$

For general f, we define $\mu(f) = \mu(f^+) - \mu(f^-)$.

Definition 3.4 (*Integrable Function*)
We say f is integrable if $\mu(|f|) < \infty$, ie. it is L^1.

Proposition 3.1 (*Properties of the Lebesgue Integral*)
We have for integrable functions f, g and constants $\alpha, \beta \ge 0$, we have

(a) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$

(b) $f \le g$ implies $\mu(f) \le \mu(g)$

(c) If f is non-negative, $f = 0$ a.e. iff $\mu(f) = 0$

3.2 Integrals and Limits

The main convergence theorems we have to go from limits of functions to the limits of integrals are the Monotone Convergence Theorem and the Dominated Convergence Theorem. Roughly, the statements are as follows:

Theorem 3.2 (*Montone Convergence Theorem*)
Let f be a non-negative measurable function and let (f_n) be a sequence of such functions. Suppose that $f_n \uparrow f$. Then $\mu(f_n) \uparrow \mu(f)$.

The key result in all of probability is the following:

Lemma 3.3 (*Fatou's Lemma*)
Let (f_n) be a sequence of non-negative measurable functions. Then

$$\mu(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty} \mu(f_n)$$

Specifically, for integrals we have

$$\int_E f d\mu \leq \liminf_{n\to\infty} \int_E f_n d\mu$$

Proof. Follows from MCT and some definitions. □

Theorem 3.4 (*Dominated Convergence Theorem*)
Let f be a measurable function and let (f_n) be a sequence of such functions. Suppose that $f_n(x) \to f(x)$ for all $x \in E$ and that $|f_n| \leq g$ for all n, for some integrable function g. Then f and f_n are integrable for all n, and $\mu(f_n) \to \mu(f)$.

Proof. Simple Fatou's Lemma but on the $g \pm f_n$s. □

We also have the following:

Theorem 3.5 (*Bounded Convergence Theorem*)
Let (X_n) be a sequence of random variables with $X_n \to X$ in probability and $|X_n| \leq C$ for some $C < \infty$. Then $X_n \to X$ in L^1.

3.3 Transformation of Integrals

Restriction of the integral to measurable subsets (definite integral).

Proposition 3.6 (*Push-forward of the Lesbegue Integral*)
Let $(E, \mathcal{E}, \mu)$ be a measure space, $(G, \mathcal{G})$ a measurable space, and $f : E \to G$ a measurable function. Define $\nu = \mu \circ f^{-1}$ to be the image measure on $(G, \mathcal{G})$. Then, for all non-negative measurable functions g on G,

$$\nu(g) = \mu(g \circ f)$$

Note, here we only have that the push-forward of the Lesbegue integral actually induces a measure on the new space, not that (if the target space is $\mathbb{R}^d$) it is the Lesbegue space. One can show that in fact this is generally the case (see Pset2 Problem 3: the main idea is to show they are equivalent on a π-system and then use Dynkin's $\pi - \lambda$ lemma) and then induce the Jacobian formula. In probability space, we can get a measure on the sigma algebra of the transformed space (such as Borel σ-algebra):

Corollary 3.7 (*Expectation of a transformed random variable*)
In probability terms, if X is a G-valued random variable, ie. $X : \Omega \to G$ and $g : G \to \mathbb{R}$, we have (if $\mathbb{P}$ is the probability measure on μ_X and $\mu_X := \mathbb{P} \circ X^{-1}$):

$$\mathbb{E}[g(X)] = \int_\Omega g(X) d\mathbb{P} = \int_\mathcal{G} g d\mu_X$$

Given a non-negative function on the *same* space, we can also *change* the integral measure on E by re-weighing, ie. the integral induces a valid measure on the space:

Theorem 3.8 (*ReweigChing of density: Integral Induces a New Measure*)
Let $(E, \mathcal{E}, \mu)$ be a measure space and let $f : E \to [0, \infty)$ be a non-negative measurable function on E. Define a *reweighing of the density* $\nu(A) = \mu(f \mathbb{1}_A), \forall A \in \mathcal{E}$. Then, ν is a measure on E, and for all non-negative measurable functions g on E,

$$\nu(g) = \mu(fg)$$

That is, the integral of functions on the space is precisely given by the integral of the product with f on the original space.

3.4 Product Measures

Definition 3.5 (*Product σ-algebra*)
Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be *finite* measure spaces. The set

$$\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$

is a π-system of subsets of $E := E_1 \times E_2$. The **product σ-algebra** is then given by

$$\mathcal{E}_1 \otimes \mathcal{E}_2 := \sigma(\mathcal{A})$$

We have piecewise, for any product measurable function that each component is measurable:

Lemma 3.9
If $f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable, then for all $x_1 \in E_1$, the function $x_2 \mapsto f(x_1, x_2) : \mathcal{E}_2 \to \mathbb{R}$ is $\mathcal{E}_2$-measurable.

Proof. Monotone Class Theorem. □

A similar result holds for piecewise integration:

Lemma 3.10
Let f be a bounded, non-negative measurable function on E. Define for $x_1 \in E_1$,

$$f_1(x_1) = \int_E f(x_1, x_2) d\mu_2$$

If f is bounded then $f_1 : E_1 \to \mathbb{R}$ is a bounded, $\mathcal{E}_1$-measurable function. If f is non-negative, then $f_1 : E_1 \to [0, \infty)$ is also an $\mathcal{E}_1$-measurable function.

Theorem 3.11
There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}_1 \otimes \mathcal{E}_2$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

$\forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$.

Trivially, switching the order/indices of the measures also preserves the measure:

Lemma 3.12
Let $\hat{\mathcal{E}} = \mathcal{E}_2 \otimes \mathcal{E}_1$ and $\hat{\mu} := \mu_2 \otimes \mu_1$. For $f : E_1 \times E_2 \to [0, \infty)$, define $\hat{f} : E_2 \times E_1 \to [0, \infty)$ by $\hat{f}(x_2, x_1) = f(x_1, x_2)$. Then, $\hat{f}$ is a non-negative $\hat{\mathcal{E}}$-measurable function and $\hat{\mu}(\hat{f}) = \mu(f)$.

The main theorem for switching the order of integration is as follows: the first part allows integration on the product measure by integrating each individual σ-algebra iteratively. The second part, we get that the iterative "slices" are integrable a.e. if the integrand is non-negative or μ-integrable. This allows us with the previous lemma to switch order of integration.

THEOREM 3.13 (*Fubini's theorem*)
Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be σ-finite measure spaces and let $\mu = \mu_1 \otimes \mu_2$ be the product measure on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$. Let $f : E_1 \times E_2 \to \mathbb{R}$ be $\mathcal{E}_1 \otimes \mathcal{E}_2$-measurable.

i. If $f \geq 0$, then
$$\int_{E_1 \times E_2} f \, d\mu = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \, d\mu_2 \right) d\mu_1.$$

ii. If $f \in L^1(\mu)$, define
$$A_1 := \left\{ x_1 \in E_1 : \int_{E_2} |f(x_1, x_2)| \, \mu_2(dx_2) < \infty \right\}$$
and $f_1 : E_1 \to \mathbb{R}$ by
$$f_1(x_1) := \begin{cases} \int_{E_2} f(x_1, x_2) \, \mu_2(dx_2), & x_1 \in A_1, \\ 0, & x_1 \notin A_1. \end{cases}$$
Then $\mu_1(E_1 \setminus A_1) = 0$, the function f_1 is μ_1-integrable, and
$$\int_{E_1} f_1(x_1) \, \mu_1(dx_1) = \int_{E_1 \times E_2} f \, d\mu.$$

From the product measure, we can build iteratively the Lesbegue measure on $\mathbb{R}^n$ (this ends up being equivalent to the inner/outer measure construction of the Lesbegue measure on $\mathbb{R}^n$, see page 7 of Durret).
TODO: product measures of independent random variables

4. L10-12: Functional Analysis (L^p Theory)

4.1 Basics of L^p-theory

TODO: definition of the L^p norm, convergence in L^p.

4.2 L^p-inequalities

L^p-theory gives us a set of nice inequalities to work with that allows us to do a lot of the heavy lifting later on. We have four main ones: Markov, Jensen, Hölder, and Minkowski:

THEOREM 4.1 (*Markov's Inequality*)
Let f be a non-negative measurable function and let $\lambda \geq 0$. Define $\{f \geq \lambda\}$ to be the set of points $\{x \in E | f(x) \geq \lambda\}$. Then,

$$\lambda\mu(f \geq \lambda) \leq \mu(f)$$

or for a non-negative random-variable:

$$\lambda\mathbb{P}[X \geq \lambda] \leq \mathbb{E}[X]$$

Jensen allows us to show inequalities for convex functions and is perhaps our most important inequality.

DEFINITION 4.1
Let $I \subseteq \mathbb{R}$ be an interval. A function $c : I \to \mathbb{R}$ is **convex** if $\forall x, y \in I$ and $t \in [0, 1]$, we have

$$c(tx + (1-t)y) \leq tc(x) + (1-t)c(y)$$

LEMMA 4.2 (*Jensen's Inequality*)
Let X be an integrable random-variable with values in I and let $c : I \to \mathbb{R}$ be convex. Then, $\mathbb{E}[c(X)]$ is well-defined and

$$\mathbb{E}[c(X)] \geq c(\mathbb{E}[X])$$

COROLLARY 4.3 (*Monotonocity of L^p-norms*)
We have for a random variables X and $1 \leq p < q < \infty$ that

$$\|X\|_p \leq \|X\|_q$$

and $L^q(\mathbb{P}) \subseteq L^p(\mathbb{P})$

The following results hold on conjugate indices:

DEFINITION 4.2 (*Conjugate Indices*)
For $p, q \in [1, \infty]$, we say that p and q are **conjugate indices** if

$$1/p + 1/q = 1$$

THEOREM 4.4 (*Hölder's inequality*)
Let $p, q \in (1, \infty)$ be conjugate indices. Then, for all measurable functions f and g, we have

$$\mu(|fg|) \leq \|f\|_p \|g\|_q$$

THEOREM 4.5 (*Minkowski's Inequality*)
For $p \in [1, \infty)$ and measurable functions f and g, we have

$$\|f + g\|_p \leq \|f\|_p + \|q\|_p$$

TODOs: approximation in L^p

4.3 $\mathcal{L}^p$-spaces, Completeness, and Hilbert Space Properties of $\mathcal{L}^2$

DEFINITION 4.3 ($\mathcal{L}^p$*-spaces*)
Define an equivalence relation $\sim$ on L^p, where $f \sim g$ if $f = g$ a.e.. Let $[f]$ be the set of equivalence classes of f induced by this equivalence relation and define the **$\mathcal{L}^p$-space**:

$$\mathcal{L}^p = \{[f] : f \in L^p\}$$

DEFINITION 4.4 (*Banach and Hilbert Spaces*)
A normed vector space V is said to be **complete** if every Cauchy sequence in V converges. A complete normed vector space is called a **Banach space** and a complete inner product space is called a **Hilber Space**.

THEOREM 4.6 (*Completeness of* $\mathcal{L}^p$)
Let $p \in [1, \infty]$. Let (f_n) be a sequence in L^p such that

$$\|f_n - f_m\|_p \to 0 \qquad \text{as } n, m \to \infty$$

Then, there exists $f \in L^p$ such that

$$\|f_n - f\|_p \to 0 \qquad \text{as } n \to \infty$$

Namely, $\mathcal{L}^p$ is complete.

COROLLARY 4.7
We have

(a) $\mathcal{L}^p$ is a Banach space for all $1 \leq p \leq \infty$.

(b) $\mathcal{L}^2$ is a Hilbert space.

We have the following properties that follow as $\mathcal{L}^2$ is a Hilbert Space:

THEOREM 4.8 (*Hilbert Space Properties of $\mathcal{L}^2$*)
We have

(a) **(Pythagoras' Rule)** For $f, g \in L^2$,

$$\|f+g\|_2^2 = \|f\|_2^2 + \|g\|_2^2$$

(b) **(Parallelogram Law)** For $f, g \in L^2$,

$$\|f+g\|_2^2 + \|f-g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2)$$

(c) **(Orthogonal Projection)** Let V be a closed subspace of L^2. Then each $f \in L^2$ has a decomposition $f = v + u$ where $v \in V$ and $u \in V^\perp$. Moreover, $\|f - v\|_2 \leq \|f - g\|_2$ for all $g \in V$ with equality iff $g = v$ a.e..

4.4 Convergence in $L^1(\mathbb{P})$

THEOREM 4.9 (*Bounded Convergence Theorem*)
Let (X_n) be a sequence of random variables with $X_n \to X$ in probablity and $|X_n| \leq C$ for all n for some constant $C < \infty$. Then $X_n \to X$ in L^1.

A way to summarize when to use MCT, DCT, BCT:

- If you have an increasing non-negative sequence, use MCT.
- If you have a non-monotone sequence, but you can find an integrable envelope, use DCT.
- If the space has finite measure and is uniformly bounded, use BCT.
- If you have no domination hypothesis, but have non-negative functions, use Fatou's Lemma.

We can achieve a stronger result for L^1-convergence of random variables with the notion of uniform convergence.

DEFINITION 4.5
Let $\mathcal{X}$ be a family of random variables. For $1 \leq p \leq \infty$, we say that $\mathcal{X}$ is **bounded in L^p** if $\sup_{x \in \mathcal{X}} \|X\|_p < \infty$. Define

$$I_{\mathcal{X}}(\delta) = \sup\{\mathbb{E}(|X|\mathbb{1}_A) : X \in \mathcal{X}, A \in \mathcal{F}, \mathbb{P}(A) \leq \delta\}$$

We say that $\mathcal{X}$ is **uniformly integrable (UI)** if $\mathcal{X}$ is bounded in L^1 and $I_X(\delta) \downarrow 0$ as $\delta \downarrow 0$.

LEMMA 4.10
Any finite collection of integrable random variables is UI.

Another characterization of uniform integrability in terms of how "big" $|X|$ is:

LEMMA 4.11
Let $\mathcal{X}$ be a family of random variables. Then $\mathcal{X}$ is UI iff

$$\sup\{\mathbb{E}(|X|\mathbb{1}_{|X|\geq K}) : X \in \mathcal{X}\} \to 0, \quad \text{as } K \to \infty$$

The following is then the definitive result on L^1-convergence of random variables:

THEOREM 4.12 (*L^1 convergence is equivalent to uniform integrability of the sequence*)
Let X be a random variable and (X_n) a sequence of random variables. The following are equivalent:

(a) $X_n \in L^1$ for all n, $X \in L^1$ and $X_n \to X$ in L^1.

(b) $\{X_n : n \in \mathbb{N}\}$ is UI and $X_n \to X$ in probability.

5. L13-15: Fourier Analysis

5.1 Fourier Transform Definitions

Now assume $f : \mathbb{R}^d \to \mathbb{C}$ are complex-valued and $f \in L^1(\mathbb{R}^d)$ implies that f is a *complex-valued* function that is bounded in L^1.

DEFINITION 5.1 (*Fourier Transform of an L^1 Function*)
The **fourier transform** $\hat{f}$ of $f \in L^1(\mathbb{R}^d)$ is defined

$$\hat{f}(u) = \int_{\mathbb{R}^d} f(x) e^{\langle i,x \rangle} dx Fouf \tag{5.1}$$

$\hat{f}$ is continuous and bounded

The Fourier transform can be extended to that of a measure as follows:

DEFINITION 5.2 (*Fourier Transform of a Measure*)
Let μ be a finite Borel measure on $\mathbb{R}^d$. Define the **fourier transform** of μ, denoted $\hat{\mu}$, by

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i\langle u,x \rangle} d\mu \tag{5.2}$$

Note, $\hat{\mu}$ is continuous and bounded by the Dominated Convergence Theorem.

If μ has density f with respect to the Lesbegue measure, $\hat{\mu} = \hat{f}$.

DEFINITION 5.3
The **characteristic function** ϕ_X of a random variable X is the Fourier transform of its law μ_X.

Definition 5.4 (*Convolution of a Function and a Measure*)
For $p \in [1, \infty)$, $f \in L^p(\mathbb{R}^d)$, probability measure ν on $\mathbb{R}^d$, define the **convolution** $f \star \nu \in L^p(\mathbb{R}^d)$ by

$$f \star \nu(x) := \int_{\mathbb{R}^d} f(x-y) d\nu \tag{5.3}$$

when the integral exists and 0 otherwise. Notably, we have $f \star \nu$ is also L^p.

Definition 5.5 (*Convolution of Two Functions*)
Define a measure ν with density g. Then $f \star g := f \star \nu$.

Definition 5.6 (*Convolution of Two Measures*)
For two probability measures μ, ν on $\mathbb{R}^d$, define the **convolution of the two measures** $\mu \star \nu$ to be the distribution of $X + Y$ for two independent random variables X, Y having laws μ and ν respectively. In particular, $\mu \star \nu$ has law $f \star g$.

The main theorem we use in Fourier Analysis is the famous convolution theorem: convolutions become products in Fourier space.

Theorem 5.1 (*Convolution Theorem*)
For $f \in L^1(\mathbb{R}^d)$ and measure ν on $\mathbb{R}^d$, we have

$$\widehat{f \star \nu}(u) = \hat{f}(u)\hat{\nu}(u) \tag{5.4}$$

and for $\mu \in L^1(\mathbb{R}^d)$

$$\widehat{\mu \star \nu}(u) = \hat{\mu}(u)\hat{\nu}(u) \tag{5.5}$$

5.2 Zero-Mean Gaussian Random Variables

Lemma 5.2 (*Characteristic Function of a Gaussian Variable*)
The characteristic function ϕ_Z of a gaussian random variable Z is given by

$$\phi_Z(u) = e^{-u^2/2} \tag{5.6}$$

Lemma 5.3
We have

$$\|f \star g_t\|_1 \leq \|f\|_1, \quad \|f \star g_t\|_\infty \leq (2\pi)^{-d/2} t^{-d/2} \|f\|_1 \tag{5.7}$$

We have also that zero mean Gaussians convolve to not do anything:

Lemma 5.4

Let $f \in L^p(\mathbb{R}^d)$ with $p \in [1, \infty)$. Then $\|f \star g_t - f\|_p \to 0$ as $t \to 0$.

5.3 Fourier Inversion in L^1

Theorem 5.5

Let $f \in L^1(\mathbb{R}^d)$. Define for $t > 0$ and $x \in \mathbb{R}^d$

$$f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-|u|^2 t/2} e^{-i\langle u, x\rangle} du \tag{5.8}$$

where we note also that $f_t = f \star g_t$ Then $\|f_t - f\| \to 0$ as $t \to 0$ and the Fourier inversion formula holds whenever $f \in L^1(\mathbb{R}^d)$ and $\hat{f} \in L^1(\mathbb{R}^d)$

5.4 Fourier Transform in L^2

Theorem 5.6

The Plancherel identity holds for all $f \in L^1 \cap L^2(\mathbb{R}^d)$, ie.

$$\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2 \tag{5.9}$$

There is a "weak" notion of convergence of measures given as follows:

Definition 5.7 (*Weak Convergence of Borel Measures*)

A sequence of Borel probability measures (μ_n) **converges weakly** to μ if for all continuous bounded functions f on $\mathbb{R}^d$, we have $\mu_n(f) \to \mu(f)$ as $n \to \infty$.

Definition 5.8 (*Weak Convergence of Random Variables*)

A sequence of random variables (X_n) converges weakly to X if μ_{X_n} converges weakly to μ_X. Note this doesn't require the random variables have to be defined on the same probability space (and a sequence of random variables can converge weakly to an infinite number of different random variables as long as they share the same law).

We now show the key result of this section, ie. that two random variables have equal law if they have equal characteristic functions. The main reason we built up the machinery on the previous section was to get to here: the main idea is to note that although ϕ_X may not be integrable we can "ameliorate" ϕ_X with a gaussian convolution to get integrable and use the fact that weak limits are unique.

THEOREM 5.7 (*Convergence of Characteristic Functions Implies Weak Convergence*)
Let X be a random variable in $\mathbb{R}^d$. Then μ_X is completely determined by the characteristic function ϕ_X. If ϕ_X is integrable, then μ_X is continuous and bounded with density function

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u) e^{-i\langle u,x\rangle} du \tag{5.10}$$

ie. $\mu_X = \int f_X$. Moreover, if (X_n) is a sequence of random variables in $\mathbb{R}^d$ such that $\phi_{X_n} \to \phi_X$, ie. the characteristic functions converge, then X_n converges to X weakly.

Proof. We first show that equality of the characteristic functions implies equality of the laws. Let Z be a standard normal random variable. Then $X + \sqrt{t}Z$ has density $f_t := \mu_X * g_t$. Now, by the convolution theorem $\hat{f}_t = \hat{\mu}_x \hat{g}_t = \phi_X(u) e^{t\|u\|^2/2}$. Thus, by Fourier Inversion (which is also integrable being a product of integrable random variables), we have

$$f_t = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} \phi_X(u) e^{-t\|u\|^2/2} du \tag{5.11}$$

so for equivalent ϕ_X, we have f_t is uniquely determined. Then, we have for any equivalent ϕ_X we can uniquely determine ϕ_X. Note, f_t is valid since the inner term is integrable since by Jensen,

$$\int_{\mathbb{R}^d} |\phi_X(u) e^{-i\langle u,x\rangle} e^{-t\|u\|^2/2}| du \le \int_{\mathbb{R}^d} \mathbb{E}[|e^{i\langle u,y\rangle}|] |e^{-i\langle u,x\rangle}| e^{-|t|\|u\|^2/2} du = \int_{\mathbb{R}^d} e^{-|t|\|u\|^2/2} < \infty \tag{5.12}$$

(unlike in the non-smoothened case). Thus, we have that for any continuous bounded function $g \in C_b(\mathbb{R}^d)$ that by Bounded Convergence

$$\mathbb{E}[g(X + \sqrt{t}Z)] \to \mathbb{E}[g(X)] \tag{5.13}$$

so $f_t \to \mu_X$ weakly and since the weak limit is unique we have that μ_X is uniquely determined.

If ϕ_X is integrable, we have that since $|\phi_X(u) e^{-i\langle u,x\rangle - t\|u\|^2/2}| \le |\phi_X(u)|$ so by dominated convergence, we have $f_t \to f_X$. We now show that $\mu_X = f_X dx$. To see this, note that the set of continuous compact support sets determines a finite Borel measure, we have

$$\int g\mu_X = \lim_{t\to 0} \int g(x) f_t(x) dx = \int g(x) f(x) dx \tag{5.14}$$

Now if (X_n) is such that $\phi_{X_n} \to \phi_X$ we show that $X_n \to X$ weakly. Suppose g is integrable with derivative bounded. Combine Fourier Inversion, DCT, and a bound on derivative to conclude. □

A stronger version (not proved):

THEOREM 5.8 (*Levy's Continuity Theorem*)
If $\phi_{X_n}(u)$ converges as $n \to \infty$ with limit $\phi(u)$ and if ϕ is continuous in a neighborhood of 0, then ϕ is the characteristic function of some random variable X and $X_n \to X$ in distribution.

5.5 Gaussian Random Variables

Definition 5.9 (*Gaussian Random Variable*)
A $\mathbb{R}$-valued random variable is *Gaussian* if for some $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$, X has density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{5.15}$$

where we take the case $X = \mu$ (constant) to be the degenerate case where $\sigma^2 = 0$.

Lemma 5.9 (*Properties of Univariate Gaussian*)
We have that if $X \sim \mathcal{N}(\mu, \sigma^2)$ that

(a) $\mathbb{E}[X] = \mu$

(b) $\text{var}[X] = \sigma^2$

(c) $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

(d) $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$

Definition 5.10
A random variable X in $\mathbb{R}^n$ is *Gaussian* if $\langle u, X \rangle$ is Gaussian, for all $u \in \mathbb{R}^n$.

Lemma 5.10 (*Properties of Multivate Gaussian*)
We have that if X is a Gaussian random variable in $\mathbb{R}^n$ and A is an $m \times n$ matrix, $b \in \mathbb{R}^m$, that

(a) $AX + b$ is a Gaussian in $\mathbb{R}^m$

(b) $X \in L^2$ and μ_X is determined by $\mu = \mathbb{E}[X]$ and $V = \text{var}(X)$.

(c) $\phi_X(u) = e^{i\langle u, \mu \rangle - \langle u, Vu \rangle/2}$

(d) If V is invertible, then X has a density function on $\mathbb{R}^n$ given by

$$f_X(x) = (2\pi)^{-n/2} (\det V)^{-1/2} \exp\{-\langle x - \mu, V^{-1}(x - \mu) \rangle / 2\} \tag{5.16}$$

(e) Suppose $X = (X_1, X_2)$ where $X_1 \in \mathbb{R}^{n_1}$ and $X_2 \in \mathbb{R}^{n_2}$. Then

$$\text{cov}(X_1, X_2) = 0 \implies X_1 \perp X_2 \tag{5.17}$$

6. L16-17: Ergodic Theory

We now move onto ergodic theory which provides a notion of the statistical properties of deterministic dynamical systems, ie. ergodicity.

6.1 Basic Definitions

Definition 6.1

Let $(E, \mathcal{E}, \mu)$ be a measure space. A measurable function $\theta : E \to E$ is called a **measure-preserving transformation** if $\mu(\theta^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{E}$.

Intuitively, the function doesn't change the measure of any set in the sigma algebra. Those that map identically under the measure-preserving transform are defined as follows:

Definition 6.2

A set $A \in \mathcal{E}$ is **invariant** if $\theta^{-1}(A) = A$.

We extend this definition to a set of invariant measurable *functions* as follows:

Definition 6.3 (*Invariant Functions*)

A measurable function f is **invariant** with respect to θ if $f = f \circ \theta$

The set of all invariant sets forms a σ-algebra, which we denote by $\mathcal{E}_\theta$.

Lemma 6.1

f is invariant with respect to θ iff f is $\mathcal{E}_\theta$-measurable.

Definition 6.4

A measure preserving function θ is said to be **ergodic** if $\mathcal{E}_\theta$ contains only sets of either measure zero or sets with complements who have measure 0.

Theorem 6.2 (*Measure Preserving Functions Preserve the Integral*)

If f is integrable and θ is measure-preserving then $f \circ \theta$ is integrable and

$$\int_E f d\mu = \int_E f \circ \theta d\mu \tag{6.1}$$

Theorem 6.3

If θ is ergodic and f is invariant then $f = c$ a.e. for some constant c.

6.2 Bernoulli Shifts

The main idea from Ergodic Theory that we will use is the idea of Bernoulli Shifts to rewrite the averages in the Strong Law of Large numbers and then use Birkhoff's a.e.-Ergodic Theorem to show convergence.

We first define a notion of a σ-algebra for a sequenece of random variables having common law:

Definition 6.5 (*Canonical Model of a Sequence of Random Variables*)
Let m be a probability measure on $\mathbb{R}$. There exists an iid sequence of random variables (Y_n) with law m. Let $E = \mathbb{R}^N$ be the set of sequences of real numbers. For any $n \in \mathbb{N}$, set $X_n : E \to \mathbb{R}$ be the **coordinate-map** $(Y_m)_{m\in\mathbb{N}} \mapsto Y_n$. Let $\mathcal{E}$ be the σ-algebra generated by $(X_n)_{n\in\mathbb{N}}$, or equivalently

$$\mathcal{E} = \sigma\left(\prod_{n\in\mathbb{N}} A_n : A_n \in \mathcal{B}(\mathbb{R}), \exists n_0 \in \mathbb{N} \text{ st. } A_n = \mathbb{R} \quad \forall n \geq n_0\right) \tag{6.2}$$

Define $Y : \Omega \to E$ by $Y(\omega) = (Y_1(\omega), Y_2(\omega), \cdots)$ for $\omega \in \Omega$. Then Y is measurable and the image measure μ has $\mu(A) = \prod m(A_n)$ for $A \in \mathcal{A}$ where $\mathcal{A} = \prod A_n$ is the generating π-system. By uniqueness of extension, μ is the unique measure on $\mathcal{E}$ having this property. We define the probability space $(E, \mathcal{E}, \mu)$ to be the **canonical model** of sequences having the law m.

Intuitively, you just define a probability space (in the obvious way) for a sequence of random variables having some law.

Definition 6.6 (*Bernoulli Shifts*)
Given a law m on $\mathbb{R}$ and a canonical model $(E, \mathcal{E}, \mu)$, define the **shift operator** $\theta : E \to E$ by

$$\theta(x_1, x_2, \cdots) = (x_2, x_3, \cdots) \tag{6.3}$$

Then, the quadruplet $(E, \mathcal{E}, \mu, \theta)$ is called a **Bernoulli Shift** or a **Bernoulli Scheme**.

Theorem 6.4
We have that the shift map θ is an ergodic, measure preserving function.

Proof. We have that if $A := A_1 \times \cdots \times A_{n_0} \times \mathbb{R} \times \cdots \in \mathcal{A}$, that $\theta^{-1}(A) = \mathbb{R} \times A$ which is also in $\mathcal{E}$ so it is measurable and the measure-preserving part is straightforward since $m(\mathbb{R}) = 1$. Ergodicity is a straightforward consequence of Kolmogorov's $0 - 1$ law (TODO). ☐

Definition 6.7
Given a measurable function f and a measure-preserving transformation θ, define the sum

$$S_n(f) := f + f \circ \theta + \cdots + f \circ \theta^{n-1} \tag{6.4}$$

6.3 Ergodic Theorems

The main theorem that will be doing the heavy lifting for us in the Strong Law of Large numbers will be the following:

Theorem 6.5 (*Birkhoff's almost-everywhere ergodic theorem*)
Assume that $(E, \mathcal{E}, \mu)$ is a σ-finite measure space, θ is a measure preserving map, and that f is an integrable function on E. Then, there exists an invariant function $\bar{f}$, with $\mu(|\bar{f}|) \leq \mu(|f|)$ such that $S_n(f)/n \to \bar{f}$ a.e. as $n \to \infty$.

The next theorem builds on Birkhoff and shows that for finite measures that almost everywhere convergence of $S_n(f)/n$ can be turned into L^p convergence:

Theorem 6.6 (*von-Neumann's L^p ergodic theorem*)
Assume that $\mu(E) < \infty$ and $p \in [1, \infty)$. Then for all $f \in L^p(\mu)$, $S_n(f)/n \to \bar{f}$ in L^p. w

7. L17-18: Law of Large Numbers, Central Limit Theorems

7.1 Law of Large Numbers

We have three versions of the law of large numbers that we proved in class (The Weak Law of Large Numbers, the Strong Law with Finite Fourth Moments, and the general SLLN).

Theorem 7.1 (*Weak Law of Large Numbers*)
Suppose we have iid (X_i) and define $\mu = \mathbb{E}[X_i]$. We have that $S_n/n \to \mu$ in probability provided the second moment is finite, ie. $\mathbb{E}[X_i^2] < \infty$.

Proof Sketch: Consider the tail $S_n^* = S_n/n - \mu$. Use Markov's inequality on the absolute value to conclude.

Proof. Let $S_n^* = S_n/n - \mu$. Apply Markov's inequality to S_n^* which has $\mathbb{E}[S_n^*] = \mu - \mu = 0$ and $\text{Var}[S_n^*] = \mathbb{E}[(S_n/n - \mu)^2] = \text{Var}[X_i]/n$. By Markov's inequality, we get

$$\mathbb{P}[|S_n^*| \geq \epsilon| \leq \frac{\text{Var}[S_n^*]}{\epsilon^2} = \frac{\text{Var}[X_i]}{n\epsilon^2} \to 0$$

since we have finite second moments we have $S_n/n \to \mu$ a.e. in probability. □

Theorem 7.2 (*SLLN with finite 4th moments*)
Let (X_n) be a sequence of independent random variables (not neccesarily identically distributed) such that for some constants $\mu \in \mathbb{R}$ and $M < \infty$ we have

$$\mathbb{E}[X_n] = \mu, \quad \mathbb{E}[X_n^4] \leq M \quad \forall n \in \mathbb{N} \tag{7.1}$$

Letting $S_n = X_1 + \cdots + X_n$, we have

$$S_n/n \to \mu \text{ a.s. as } n \to \infty \tag{7.2}$$

Proof Sketch: do it on the normalized $Y_n = X_n - \mu$. Then the product of Y_j and any powers is 0. Show that the sum of the expected $(S_n/n)^4$ is thus finite and so $S_n/n \to 0$.

Proof. Set $Y_n = X_n - \mu$ and note $\mathbb{E}[Y_n] = 0$ while $\mathbb{E}[Y_n^4] \leq 16(X_n^4 + \mu^4) \leq 16(M^4 + \mu^4) < \infty$ and so we show just the case where $\mu = 0$. Note the X_i and X_js are independent, so $\mathbb{E}[X_i X_j^3] = \mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i X_j X_k X_l] = 0$ so

$$\mathbb{E}[S_n^4] = \mathbb{E}[\sum X_k^4 + 6\sum X_i^2 X_j^2] \leq nM + 3n(n-1)M \leq 3n^2 M \tag{7.3}$$

where we use Cauchy-Schwarz for the second term. This allows us to bound:

$$\mathbb{E}\sum (S_n/n)^4 \leq 3\sum \frac{M}{n^2} < \infty \tag{7.4}$$

so $\sum (S_n/n)^4$ a.s. and thus $S_n/n \to 0$ a.s.. Note, we required fourth moments in getting the $1/n^2$ term instead of a $1/n$ term which would diverge in order to get finite sum. □

THEOREM 7.3 (*Strong Law of Large Numbers*)
Let (X_n) be a sequence of iid random variables with mean μ. Set $S_n = X_1 + \cdots + X_n$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.

Proof Sketch: Use the Bernoulli scheme on the law, apply Birkhoff and von-Neumann to show convergence to the mean.

Proof. Suppose each $X_i : \Omega \to \mathbb{R}$. Let μ_X be the law of the X_is. Now note by assumption we have that since X_i is a random variable, it is both normalized and has finite first moment, ie. $X_i \in L^1$. Now, let $(E, \mathcal{E}, \mu, \theta)$ be the Bernoulli scheme for random variables with law μ_X. Now, we showed the shift map θ to be measure preserving and ergodic. Then, let $f : E \to \mathbb{R}$ be the coordinate function $f(x_1, \cdots,) = x_1$. Now, we have that with respect to this θ, that $S_n(f) : E \to \mathbb{R}$ satisfies:

$$S_n(f) = \sum_{j=1}^{n} f \circ \theta^{j-1} = x_1 + \cdots + x_n \tag{7.5}$$

Thus, by Birkhoff's Ergodic Lemma, since θ is measure preserving, we have $S_n(f)/n \to \bar{f}$ for some invariant $\bar{f}$. Further, since θ is ergodic, $\bar{f} = c$ a.e.. Note also by von-Neumann, since we have $\mu(|\theta|) = \int |x| d\mu_X < \infty$ we have $S_n(f)/n \to \bar{f}$ in L^1. Thus, $c = \int \bar{f} d\mu = \lim_{n\to\infty} \mu(S_n(f)/n)$. Note that $\mu(S_n(f)/n) = \int (x_1 + \cdots + x_n)/n d\mu = \sum \int x_j/n d\mu_X = \sum(\mu/n) = \mu$. So $S_n(f)/n \to \mu$ a.s.. □

7.2 Central Limit Theorem

The sample mean goes to the mean while the scaled sample mean goes to a Gaussian:

THEOREM 7.4 (*Central Limit Theorem*)

Let (X_n) be a sequence of iid random variables with mean 0 and variance 1. Set $S_n = X_1 + \cdots + X_n$. Then, for all $x \in \mathbb{R}$, as $n \to \infty$,

$$\mathbb{P}[S_n/\sqrt{n} \leq x] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{7.6}$$

That is, $\mu_{S_n/\sqrt{n}} \to \mathcal{N}(0,1)$ pointwise.

Proof Sketch: use the characteristic function, a taylor expansion, and the expansion of the log.

Proof. We have that the characteristic function of X_1, $\phi_{X_1}(u) = \mathbb{E}[e^{iuX_1}]$. We have since X_1 has finite variance, we can differentiate twice under the integral, which gives $\phi_{X_1}(0) = 1$ and

$$\phi'_{X_1}(0) = \mathbb{E}[iX_1] = 0, \qquad \phi''_{X_1}(0) = -\mathbb{E}[X_1^2] = -1 \tag{7.7}$$

so by Taylor's Theorem, we have

$$\phi_{X_1}(u) = 1 - u^2/2 + o(u^2) \tag{7.8}$$

Now,

$$\phi_{S_n/\sqrt{n}}(u) = \mathbb{E}[e^{iu(X_1+\cdots+X_n)/\sqrt{n}}] = (\mathbb{E}[e^{iu(X_1/\sqrt{n})}])^n = (1 - u^2/(2n) + o(u^2/2n))^n \tag{7.9}$$

Thus,

$$\log \phi_{S_n/\sqrt{n}}(u) = n \log(1 - u^2/(2n) + o(u^2/n)) \tag{7.10}$$

Recalling $\log(1+z) = z + o(|z|)$, we have that as $n \to \infty$

$$\log \phi_{S_n/\sqrt{n}}(u) = -u^2/2 \implies \phi_{S_n/\sqrt{n}} \to e^{-u^2/2} \tag{7.11}$$

Since this is precisely the characteristic function of the Gaussian cdf $\mathcal{N}(0,1)$, we have the result. □

8. L19: Conditional Expectations

The conditional expectation gives us a notion of "partial averages" for random variables. The conditional expectation is defined weakly, as a random variable $\mathbb{E}[X|\mathcal{G}]$ on a sub-sigma algebra $\mathcal{G} \subseteq \mathcal{F}$.

THEOREM 8.1

Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. There exists a random variable Y such that:

(a) Y is $\mathcal{G}$-measurable.

(b) Y is integrable and $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ for all $A \in \mathcal{G}$.

Definition 8.1
We call such a random variable an **conditional expectation of X given $\mathcal{G}$**

Some properties of the conditional expectations:

Theorem 8.2 (*Properties of Conditional Expectation*)
If X is an integrable random variable and $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-algebra we have

(a) **Law of Total Expectation**: $\mathbb{E}[\mathbb{E}[X \mid G]] = \mathbb{E}[X]$

(b) **Dependence Property**: if X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ almost surely.

(c) **Independence**: if X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ almost surely.

(d) **Nonnegativity**: if $X \geq 0$ almost surely then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely.

(e) **Linearity**: $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\,\mathbb{E}[X|\mathcal{G}] + \beta\,\mathbb{E}[Y|\mathcal{G}]$ almost surely.

(f) **Conditional Monotone Convergence**: if $0 \leq X_n \uparrow X$ almost surely then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ almost surely.

(g) **Conditional Fatou's Lemma**: if $X_n \geq 0$ then $\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$

(h) **Conditional Dominated Convergence**: if $X_n \to X$ and $|X_n| \leq Y$ a.s. for some integrable random variable Y, then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ a.s.

(i) **Conditional Jensen's Inequality**: if c is convex then $\mathbb{E}[c(X)|\mathcal{G}] \geq c\,\mathbb{E}[X|\mathcal{G}]$ almost surely.

(j) L^p **Property**: $\|\,\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p$ for all $1 \leq p \leq < \infty$.

(k) **Tower Property**: if $\mathcal{H} \subseteq \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ almost surely.

(l) **Take Out What's Known Property**: if Y is bounded and $\mathcal{G}$-measurable then $\mathbb{E}[YX|\mathcal{G}] = Y\,\mathbb{E}[X|\mathcal{G}]$

(m) **Independent Conditioning Property**: if $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$ then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$ almost surely.

Last property is not true if $\sigma(X, \mathcal{G})$ is just $\mathcal{G}$ even if $X \perp \mathcal{H}$ (think two Bernoulli variables). Notably, also, for L^1 random variables, the conditional expectations are also UI:

Theorem 8.3
If $X \in L^1$, is a sub σ-algebra, then $\{\mathbb{E}[X|\mathcal{G}] | G \subseteq \mathcal{E}$ is a sub σ-algebra$\}$ is UI.

9. L20-25: Discrete-Time Martingales

We now make rigorous the notion of a random process, along with the notion of a martingale which intuitively describes a notion of a "fair game" where the present is the best indicator of the future.

9.1 Random Processes

Definition 9.1 (*Filtration*)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Define a **filtration** to be a sequence $(\mathcal{F}_n)_{n \geq 0}$ of sub σ-algebras such that

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots \subseteq F \tag{9.1}$$

for all n. Set $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$. Then $\mathcal{F}_\infty \subseteq \mathcal{F}$. Intuitively this is the notion of our knowledge at time n.

Definition 9.2 (*Random Processes*)
A **random process** is a sequence of random variables (X_n) where $X_n : \Omega \to E$. Define the **natural filtration** $\mathcal{F}_n^X = \sigma(X_0, \cdots, X_n)$. We say that (X_n) is **adapted** to a filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable. Note by definition it is adapted to $\mathcal{F}_n^X$. It is thus equivalent to say that $\mathcal{F}_n^X \subseteq \mathcal{F}_n$. We say that (X_n) is **integrable** if each X_n is integrable with respect to each $\mathcal{F}_n$.

9.2 Martingales and Optional Stopping

Definition 9.3 (*Martingales*)
A **martingale** is an adapted integrable random process (X_n) such that $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ a.s. for all $n \geq 0$. If we have $\leq$ we have a **supermartingale** and if we have $\geq$ it is called a **submartingale**.

Intuitively, they can be thought of as having no conditional drift, ie. a game where the present is the best indicator of the future and you can not beat the game systematically.
We now define a *stopping time*, which is essentially a random variable that tells you when you (for each event) should make a bet. The definition essentially formalizes the fact the following idea: for each time n, you should be able to tell that you ave already stopped.

Definition 9.4 (*Stopping Times*)
A random variable $T : \Omega \to \{0, 1, \cdots\} \cup \{\infty\}$ is called a **stopping time** if $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$. We define

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \forall n \geq 0\} \tag{9.2}$$

and define X_T by $X_T(\omega) = X_{T(\omega)}(\omega)$ and X^T a **stopped process** by $X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega)$.

Some examples of stopping times include for any $A \in \mathcal{B}(\mathbb{R})$ the first time $n \geq 0$ such that $X_n \in A$. On the other hand, $T_A = \sup\{n \leq 10 \mid X_n \in A\}$ is not a stopping time since you need to look ahead into the future. Some facts about stopping times:

The main theorem about martingales is essentially the following: in a martingale (a fair game), we have for two stopping times $S \leq T$ that if you stop at any point or make a bet then your expectation doesn't change.

THEOREM 9.2 (*Optional Stopping Theorem*)
Let $X = (X_n)$ be a martingale and let $S \leq T$ be bounded stopping times. Then $\mathbb{E}[X_T] = \mathbb{E}[X_S]$

Proof. Let $n \geq 0$ such that $T \leq n$ and then note

$$X_T = X_S + \sum_{S \leq k < T} (X_{k+1} - X_k) \tag{9.3}$$

so

$$\mathbb{E}[X_T] = \mathbb{E}[X_S] + \sum_{k=0}^{n} \mathbb{E}[(X_{k+1} - X_k)\mathbb{1}_{S \leq k < T}] \tag{9.4}$$

Noting $\{S \leq k < T\} \in \mathcal{F}_k$ we have that

$$\begin{aligned}
\mathbb{E}[(X_{k+1} - X_k)\mathbb{1}_{S \leq k < T}] &= \mathbb{E}[\mathbb{E}[(X_{k+1} - X_k)\mathbb{1}_{S \leq k < T} | \mathcal{F}_k]] && (9.5)\\
&= \mathbb{E}[\mathbb{1}_{S \leq k < T}\, \mathbb{E}[(X_{k+1} - X_k) | \mathcal{F}_k]] && (9.6)\\
&= 0 && (9.7)
\end{aligned}$$

where we use the martingal property. This gives the conclusion. □

Some additional properties of martingales:

THEOREM 9.3
Let X be an adapted integrable process. Then the following are equivalent:

(a) X is a martingale.

(b) For all bounded stopping times $S \leq T$,

$$\mathbb{E}[X_T | \mathcal{F}_S] = X_S \tag{9.8}$$

almost surely.

(c) If T is a stopping time, the stopped process X^T is a martingale.

(d) If $|X_n| \leq Y$ for all $n \in \mathbb{N}$ for some $Y \in L^1$ and T is a stopping time such that $T < \infty$ a.s., then $\mathbb{E}[X_T] = \mathbb{E}[X_S]$.

(e) If $\exists M > 0$ such that $|X_{n+1} - X_n| \leq M$ for all $n \in \mathbb{N}$ a.s. and T is a stopping time such that $\mathbb{E}[T] < \infty$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

A supermartingale version:

THEOREM 9.4 (*Optional Stopping Theorem for Supermartingales*)
Suppose that $X = (X_n)$ is a non-negative supermartingale. Let T be a stopping time that is finite a.s.. Then,
$$\mathbb{E}[X_T] \leq \mathbb{E}[X_0] \tag{9.9}$$

A nice result that follows:

THEOREM 9.5
If $\mathcal{G} = (\mathcal{G}_n)$ is a simple random walk and $T_c := \inf\{n \geq 0 : x_n = c\}$ then
$$\mathbb{P}[T_{-a} \leq T_b] = \frac{b}{a+b} \tag{9.10}$$

Proof. We have $|X_{n+1} - X_n| \leq 1$ and we can show $\mathbb{E}[T] < \infty$ by first defining
$$S : \inf\{k \in \mathbb{N} : \mathcal{G}_{k(a+b)+1} = \cdots = \mathcal{G}_{(k+1)(a+b)} \tag{9.11}$$
Then $\mathbb{E}[S] = \frac{1}{1-p} < \infty$ where $p = 2^{-(a+b)}$. Then, $T \leq (S+1)(a+b)$ so $\mathbb{E}[T] < \infty$. Then by the supermartingale OST, we have
$$\mathbb{E}[X_T] = -a\mathbb{P}[T_{-a} \leq T_b] + b\mathbb{P}[T_{-a} > T_b] \implies \mathbb{P}[T_{-a} \leq T_b] = \frac{b}{a+b} \tag{9.12}$$
□

9.3 Upcrossings and a.s. Martingale Convergence Theorem

DEFINITION 9.5 (*Upcrossing*)
Let $(X_n) \to X$ be a converging sequence in $\mathbb{R}$. Fix $a < b \in \mathbb{R}$. Define $T_0(x) = 0$ and inductively define for $k \geq 0$ where $(X_n) \to X$ is a convergence sequence. Then, define
$$S_{k+1}(x) := \inf\{n \geq T_k(x) : X_n \leq a\} \tag{9.13}$$
$$T_{k+1}(x) := \inf\{n \geq S_{k+1}(x) : X_n \geq b\} \tag{9.14}$$
Now we define the number of disjoint upcrossings contained in $\{0, \cdots, n\}$ by
$$N_n([a,b]) = \sup\{k \geq 0 : T_k(x) \leq n\} \tag{9.15}$$
We then define the number of upcrossings $N([a,b])$ by $N_n([a,b]) \uparrow N([a,b])$.

LEMMA 9.6
A sequence of real numbers (X_n) converges to $X \in \mathbb{R} \cup \{\pm\infty\}$ iff the total number of upcrossings in $[a,b]$ for any $a, b \in \mathbb{Q}$ is finite.

More general is the following statement which bounds tthe expected number of upcrossings for a supermartingale:

THEOREM 9.7 (*Doob's Upcrossing Inequality*)
Let X be a supermartingale. Then$\forall a < b \in \mathbb{R}$ we have

$$(b-a)\,\mathbb{E}[N_n([a,b])] \leq \mathbb{E}[(X_n - a)^-] \tag{9.16}$$

where $(X_n - a)^- = -\min\{0, X_n - a\}$

Proof. Let $N := N_n([a,b])$. We have that $X_{T_k} - X_{S_k} \geq b - a$ for any $k > 0$. Also,

$$\sum_{k=1}^{n}(X_{T_k \wedge n} - X_{S_k \wedge n}) = \sum_{k=1}^{N} X_{T_k} - X_{S_k} + (X_n - X_{S_{n+1}})\mathbb{1}_{S_{n+1} \leq n} \tag{9.17}$$

Taking expectations, we get

$$0 \geq (b-a)\,\mathbb{E}[N] + \mathbb{E}[(X_n - X_{S_n+1})\mathbb{1}_{S_{n+1} \leq n}] \tag{9.18}$$
$$\geq (b-a)\,\mathbb{E}[N] - \mathbb{E}[(X_n - a)^-] \tag{9.19}$$

where we use the fact that $X_{S_n+1} \leq a$ by definition. We thus get the desired inequality. □

COROLLARY 9.8 (*sup form of Doob's Upcrossing Inequality*)
Let X be a supermartingale. Then

$$(b-a)\,\mathbb{E}[N([a,b])] \leq \sup_{n \geq 0} \mathbb{E}[(X_n - a)^-] \tag{9.20}$$

We use this to show that L^1 martingales converge a.s.:

THEOREM 9.9 (*a.s. Martingale Convergence Theorem*)
Suppose $X = (X_n)$ is a L^1 supermartingale, ie. $\sup_{n \geq 0} \mathbb{E}[|X_n|] < \infty$ then $X_n \to X_\infty$ a.s. for some $X_\infty \in L^1(\mathcal{F}_\infty)$,

Proof. Let $\Omega_0 := \Omega_\infty \cap (\cap_{a<b\in\mathbb{Q}} \Omega_{a,b})$ where

$$\Omega_{a,b} := \{\omega \in \Omega : N([a,b]) < \infty\} \tag{9.21}$$

Now, remark that by Doob's Upcrossing inequality, we have that

$$\mathbb{E}[N([a,b])] \leq \frac{1}{b-a} \sup_{n \geq 0} \mathbb{E}[(X_n - a)^-] \leq \frac{1}{b-a} \sup_{n \geq 0} \mathbb{E}[|X_n|] + \frac{|a|}{b-a} < \infty \tag{9.22}$$

so $\mathbb{P}[N([a,b]) < \infty] = 1$ and thus $\mathbb{P}[\Omega_0] = 1$. We then have that (X_n) converges in Ω_0 a.s. since for any $\omega \in \Omega$, we have that otherwise $\exists a, b \in \mathbb{Q}$ such that $\liminf_{n\to\infty} X_n(\omega) \leq a < b \leq \limsup_{n\to\infty} X_n(\omega)$ which would give an infinite number of upcrossings. Thus, $X_n \to X_\infty$ a.s.. Now, note

$$\mathbb{E}[|X_\infty|] = \mathbb{E}[\lim_{n\to\infty} |X_n|] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|] \leq \limsup_{n\to\infty} \mathbb{E}[X_n] < \infty \tag{9.23}$$

where we use Fatou's Lemma and the definition of (X_n) being in L^1. □

9.4 Doob's L^p martingale inequalities

The next inequality gives us a statement about the supremum of a random variable X_n in a martingale. The first one (the maximal inequality) is analogous to Markov's inequality, while the second one is analogous to Holder.

THEOREM 9.10 (*Doob's Maximal Inequality*)
Let X be a martingale or non-negative submartingale. Define $X_n^* := \sup_{k\le n}|X_k|$ Then, $\forall \lambda \ge 0$,

$$\lambda \mathbb{P}[X_n^* \ge \lambda] \le \mathbb{E}[|X_n| \mathbb{1}_{X_n^* \ge \lambda}] \le \mathbb{E}[|X_n|] \tag{9.24}$$

Proof. Let $T := \inf\{k \ge 0 : X_k \ge \lambda\}$. Then $T \wedge n$ is a bounded stopping time. Then, by OST since $T \wedge n \le n$, we have

$$\mathbb{E}[X_n] \ge \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T]\mathbb{P}[T \le n] + \mathbb{E}[X_n \mathbb{1}_{T>n}] \tag{9.25}$$

$$\ge \lambda \mathbb{P}[X_n^* \ge \lambda] + \mathbb{E}[X_n \mathbb{1}_{T>n}] \tag{9.26}$$

Rearranging gives the desired inequality. □

THEOREM 9.11 (*Doob's L^p-inequality*)
Let X be a martingale or non-negative submartingale. Then, for all $p > 1$, we have

$$\|X_n^*\|_p \le \frac{p}{p-1}\|X_n\|_p \tag{9.27}$$

Proof. WLOG X is nonnegative. Fix $M < \infty$. We have by the Layer-Cake identity, pointwise that

$$(X_n^* \wedge M)^p := \int_0^M p\lambda^{p-1}\mathbb{1}_{X_n^* \ge \lambda} d\lambda \tag{9.28}$$

Thus, by Fubini and Doob's Maximal Inequality,

$$\mathbb{E}[(X_n^* \wedge M)^p] = \int_0^M p\lambda^{p-1}\mathbb{P}[X_n^* \ge \lambda] d\lambda \tag{9.29}$$

$$\le \int_0^M p\lambda^{p-2}\,\mathbb{E}[|X_n|\mathbb{1}_{X_n^* \ge \lambda}] d\lambda \tag{9.30}$$

$$= \mathbb{E}\left[|X_n| \int_0^M p\lambda^{p-2}\mathbb{1}_{X_n^* \ge \lambda} d\lambda\right] \tag{9.31}$$

$$= \frac{p}{p-1}\,\mathbb{E}[X_n(X_n^* \wedge M)^{p-1}] \tag{9.32}$$

$$\le \frac{p}{p-1}\|X_n\|_p \|X_n^* \wedge M\|_p^{p-1} \tag{9.33}$$

where for the last line we use Holder. Dividing by $\|X_n^* \wedge M\|_p^{p-1}$. we get $\|X_n^* \wedge M\|_p \le \frac{p}{p-1}\|X_n\|_p$. Taking $M \to \infty$ and using monotone convergence finishes the proof. c □

Note $p > 1$. The $p = 1$ case requires more work. The next result extends the a.s. martingale convergence to L^p space for $p > 1$.

Definition 9.6
We say that a random process X is **L^p-bounded** if $\sup_{n\geq 0} \|X_n\|_p < \infty$.

Theorem 9.12 (*L^p a.s. martingale convergence theorem*)
Let X be a martingale and $p > 1$. Then the following are equivalent:

(a) $\sup_{n\geq 0} \|X_n\|_p < \infty$

(b) $\exists X_\infty \in L^p$ such that $X_n \to X$ a.s. and $X_n \to X$ in L^p.

(c) $\exists Z \in L^p$ such that $\mathbb{E}[Z|\mathcal{F}_n] = X_n$ a.s.

Proof. First part is just dominated convergence on $|X_n - X_\infty|^p$. More specifically, $X_n \to X_\infty$ a.s. for some $X_\infty \in L^1$ and by Doob's L^p inequality, $\|X^*\|_p \leq \frac{p}{p-1} \sup_{n\geq 0} ||X_n\|_p < \infty$. Then, $|X_n - X_n^*|^p \leq (2X^*)^p < \infty$ so we get the result. For $(ii) \implies (iii)$ just use $Z = X_\infty$. For $(iii) \implies (i)$, we have X_n is a martingale since

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[Z|\mathcal{F}_n] \tag{9.34}$$

by the Tower Property. Also, by conditional Jensen, $\mathbb{E}[Z|\mathcal{F}_n]^p \leq \mathbb{E}[|Z|^p|\mathcal{F}_n]$ so taking expectations we get $\|X_n\|_p \leq \|Z\|_p$ and so X_n is L^p-bounded. □

9.5 Doob's L^1 martingale inequalities

We now focus on uniform integrability in the case of martingales.

Lemma 9.13
Let X be an L^1-bounded martingale. Then, the family

$$\{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \text{ is a sub } \sigma\text{-algebra of } \mathcal{F}\} \tag{9.35}$$

is uniformly integrable.

Proof. Use the definition of UI of a random variable X and then use Markov's inequality along with conditional Jensen. □

This yields a version of the L^1 martingale convergence theorem:

Theorem 9.14 (*L^1 a.s. martingale convergence theorem*)
Let (X_n) be a martingale. Then the following are equivalent:

(a) X is UI.

(b) X_n converges a.s. and in L^1 to some $X_\infty \in L^1$.

(c) $\exists Z \in L^1$ such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for all n a.s.

Proof. $(i) \implies (ii)$ follows from a.s. martingale convergence along with the fact that X is UI. $(ii) \implies (iii)$ follows from taking $Z = X_\infty$. $(iii) \implies (i)$ follows from the Tower Property. ☐

It is possible for non-UI L^1 bounded martingales to converge a.s. but not in L^1. The canonical example is the product of n Ber$(1/2)$ coin flips.

DEFINITION 9.7

If (X_n) is UI, we set $X_T = \sum_{n=0}^{\infty} X_n \mathbb{1}_{T=n} + X_\infty \mathbb{1}_{T=\infty}$ where T is a stopping time and $X_n \to X_\infty$ a.s. and in L^1> So for UI martingales, we allow **stopping at infinity**.

We now extend the OST for UI martingales as follows:

THEOREM 9.15

Let $X = (X_n)$ be a UI martingale and S, T be stopping times such that $S \le T$ a.s.. Then, $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$ a.s.

Note the relaxation of the constraint that S and T are bounded and now that $S \le T$ only a.s. now.

Proof. By L^1-martingale convergence theorem we have $X_n \to X_\infty$ in L^1 and $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$. Also, by conditional Jensen, $\mathbb{E}[|X_T|] \le \mathbb{E}[|X_\infty|] < \infty$. Similarly, $\mathbb{E}[X_T] = \mathbb{E}[X_\infty \mid \mathcal{F}_T]$ with some symbol pushing and the tower property finishes. ☐

9.6 Backwards Martingales

We now define a notion of a backwards martingale which surprisingly gives us easier results foLr the case $p = 1$ unlike in the forward martingale case:

DEFINITION 9.8 (*Backwards Filtration*)

Let $\mathcal{G}_0 \supseteq \mathcal{G}_{-1} \supseteq \cdots$ be a decreasing sequence of sub σ-algebras indexed by $-\mathbb{N} \cup \{0\}$. We call such a decreasing subsequence a **backwards filtration**. Denote by $\mathcal{G}_{-\infty} = \cap_{n\ge 0} \mathcal{G}_{-n}$.

DEFINITION 9.9 (*Backwards Martingale*)

A process $X = (X_n)$ is called a **backwards martingale** if it is **adapted** to the filtration $(\mathcal{G}_n)$ if each X_j is $\mathcal{G}_j$-measurable, $X_0 \in L^1$, and $\mathbb{E}[X_{n+1}|\mathcal{G}_n] = X_n$ for all $n \le -1$.

The main idea behind a backwards martingale (and why you get the $p = 1$ convergence for free unlike in the forward case) is because your backwards martingale is automatically a conditional expectation of a fixed L^1 random variable which means you get uniform integrability for free.

THEOREM 9.16 (*Backwards Martingale Convergence Theorem*)

Let X be a backwards martingale, such that $X_0 \in L^p$ for $p \in [1, \infty)$. Then $X_n \to X_{-\infty} := \mathbb{E}[X_0|\mathcal{G}_{-\infty}]$ a.s. and in L^p.

Proof. Fix $a < b \in \mathbb{R}$ and let $N_{-n}([a,b])$ be the number of upcrossings of X on $[a,b]$ by time $-n$. Set $\mathcal{F}_k = (\mathcal{G}_{-n+k}; 0 \leq k \leq n)$ and $X' = (X_{-n+k}; 0 \leq k \leq n)$. Now $\mathcal{F}_k$ is an increasing filtration and X' is a martingale wrt $\mathcal{F}_k$. Since X' is a forward martingale, by Doob's upcrossing inequality, we have that

$$(b-a)\,\mathbb{E}[N_{-n}([a,b])] \leq \mathbb{E}[(X_0' - a)^-] \tag{9.36}$$

$$\leq |a| + \mathbb{E}[|X_0'|] \tag{9.37}$$

We thus have that $N_{-n}([a,b]) < \infty$ and that $X_{-n} \to X_{-\infty}$ a.s. for some $X_{-\infty}$ that is $\mathcal{G}_{-\infty}$-measurable. Also, by Fatou's lemma, we have that

$$\mathbb{E}[|X_{-\infty}|^p] = \mathbb{E}[\lim_{n\to\infty} |X_n|^p] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|^p] \tag{9.38}$$

and by conditional Jensen along with the tower property,

$$|X_{-n}|^p = |\,\mathbb{E}[X_0|\mathcal{G}_{-n}]|^p \leq \mathbb{E}[|X_0|^p|\mathcal{G}_{-n}] \tag{9.39}$$

so

$$\mathbb{E}[|X_{-n}|^p] \leq \mathbb{E}[\mathbb{E}[|X_0|^p|\mathcal{G}_{-n}]] = \mathbb{E}[|X_0|^p] < \infty \tag{9.40}$$

so $X_{-\infty} \in L^p$. Further, since X is a backwards martingale, we have that

$$|X_{-n} - X_{-\infty}|^p = |\,\mathbb{E}[X_0 - X_{-n}|\mathcal{G}_{-n}]|^p \leq \mathbb{E}[|X_0 - X_{-n}|^p|\mathcal{G}_{-n}] \tag{9.41}$$

So $(|X_{-n} - X_{-\infty}|^p)$ is UI (being bounded a UI family) so $X_{-n} \to X_{-\infty}$ in L^p. Also, fixing $A \in \mathcal{G}_{-\infty}$ we have that $\mathbb{E}[X_0 \mathbb{1}_A] = \mathbb{E}[X_{-n}\mathbb{1}_A] =\to \mathbb{E}[X_{-\infty}\mathbb{1}_A]$ so $X_{-n} = -\,\mathbb{E}[X_0|\mathcal{G}_{-\infty}]$ a.s.. □

9.7 Applications, Product Martingales, and Radon-Nikodym Theorem

We can prove the SLLN and Kolomogorov's $0-1$ law using martingales:

Theorem 9.17 (*Kolomogorov's $0-1$ law*)
Let (X_i) be a sequence of independent random variables and $\mathcal{F}_n = \sigma(X_k; k \geq n)$. Then define the tail σ-algebra:

$$\mathcal{F}_\infty := \bigcap_{j=1}^{\infty} \mathcal{F}_j \tag{9.42}$$

Then, for any $A \in \mathcal{F}_\infty$, $\mathbb{P}[A] \in \{0,1\}$.

Proof. Let $\mathcal{G}_n = \sigma(X_1, \cdots, X_n)$. Suppose $A \in \mathcal{F}_\infty$. We show $\mathbb{P}[A] = \mathbb{1}_A$ a.s.. Consider the sequence of random variables $(\mathbb{E}[\mathbb{1}_A|\mathcal{G}_n])_{n\geq 0}$. Now $\mathcal{G}_n$ is a filtration and is UI being the conditional expectation of L^1-random variables. Thus, by the L^1-martingale convergence theorem,

$$\mathbb{E}[\mathbb{1}_A|\mathcal{G}_n] \xrightarrow{L^1} \mathbb{E}[\mathbb{1}_A|\mathcal{G}_\infty] = \mathbb{1}_A \tag{9.43}$$

since $G_\infty \subseteq \mathcal{F}_\infty$. Since $\mathcal{G}_n$ is independent of $\mathcal{F}_\infty$, we have $\mathbb{E}[\mathbb{1}_A \mid \mathcal{G}_n] = \mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$ giving the result. ■ □

THEOREM 9.18 (*Strong Law of Large Numbers*)
Let (X_n) be iid random variables with $\mu = \mathbb{E}[X_j]$. Then $S_n/n \to \mu$ a.s. and in L^1.

Proof. Define $\mathcal{G}_n = \sigma(S_n, \cdots)$. Let $\mu_n = S_n/n$ for all $n \geq 0$. Now, $(\mathcal{G}_n)$ is a backwards filtration. Letting $\mathcal{F}_n = \sigma(X_{n+1}, X_{n+2}, \cdots)$, we also have that since $\sigma(X_1, S_n)$ is independent of $\mathcal{F}_n$ that $\mathbb{E}[X_1|\mathcal{G}_n] = \mathbb{E}[X_1|S_n]$ a.s. for all n. Thus, for all B Borel, we have that

$$\mathbb{E}[X_k \mathbb{1}_{S_n \in B}] = \mathbb{E}[X_1 \mathbb{1}_{S_n \in B}] \tag{9.44}$$

where we use a symmetry argument noting that each of the X_js are iid so they are permutation invariant in the expectation so $\mathbb{E}[X_k|S_n] = \mathbb{E}[X_1|S_n]$ a.s.. Thus,

$$\mathbb{E}[X_1|S_n] + \cdots + \mathbb{E}[X_n|S_n] = \mathbb{E}[S_n|S_n] = S_n = n\,\mathbb{E}[X_1|S_n] \tag{9.45}$$

rearranging gives

$$\mathbb{E}[\mu_1|\mathcal{G}_n] = \mathbb{E}[X_1|S_n] = S_n/n = \mu_n \tag{9.46}$$

a.s. so we have (μ_n) is a backwards martingale.
Thus, by the a.s. backwards martingale convergence theorem, we have that $\exists Y \in L^1$ such that $S_n/n \to Y$ a.s. and in L^1. Now Y is $\cap \mathcal{G}_n$-measurable, ie. in the tail σ-algebra so by Kolomogorov's $0-1$ law, we have that y is constant a.s. so $Y = \mathbb{E}[Y] = \lim_{n\to\infty} \mathbb{E}[S_n/n] = \mu$ a.s.. □

We also have the following theorem about product martingales:

THEOREM 9.19 (*Katakuni's Product Martingale Theorem*)
Let (X_n) be independent, non-negative random variables with mean 1. We set $\mu_0 = 1$, $\mu_n = X_1 \cdots X_n$ for $n \geq 1$. Let $\mu_\infty \in L^1$ be s.t. $\mu_n \to \mu_\infty$ a.s.. We set $a_n = \mathbb{E}[\sqrt{X_n}]$. Then $a_n \in (0, 1]$ and

(a) If $\prod a_n > 0$ then $\mu_n \to \mu_\infty$ in L^1.

(b) If $\prod a_n = 0$ then $\mu_\infty = 0$ a.s.

Proof. We have $\mathbb{E}[\mu_n] = \prod_{k=1}^n \mathbb{E}[X_k] = 1$ by independence so $\mu_n \in L^1$. We also have that

$$\mathbb{E}[\mu_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_{n+1}]\mu_n = \mu_n \tag{9.47}$$

so it is a martingale. By the a.s. martingale convergence theorem, $\exists$ a random variable μ_∞ such that $\mu_n \to \mu_\infty$ a.s.. By Jensen, we have

$$a_n = \mathbb{E}[\sqrt{X_n}] \leq \mathbb{E}[X_n]^{1/2} = 1 \tag{9.48}$$

(a) Assume that $\prod a_n > 0$. Now we show that $\mathbb{E}[\sup_{n\in\mathbb{N}}(\mu_n)] < \infty$ which suffices since then

$$\mathbb{E}[\mu_n \mathbb{1}_A] \leq \mathbb{E}[\sup_{n\in\mathbb{N}} \mu_n \mathbb{1}_A] \leq M\mathbb{P}[A] \downarrow 0 \tag{9.49}$$

as $\mathbb{P}[A] \downarrow 0$ so the (μ_n) is UI and thus by the L^1 a.s.-martingale convergence theorem the convergence in L^1. Define $N_n := \sqrt{X_1 \cdots X_n}/(a_1 \cdots a_n)$ for $n \geq 1$ and $N_0 = 1$. Now, $\mathbb{E}[N_n] = 1$ so $N_n \in L^1$ so (N_n) is an L^1 martingale so $N_n \to N_\infty$ a.s. and

$$\sup_{n\in\mathbb{N}} \mathbb{E}[N_n^2] = \sup_{n\in\mathbb{N}} \frac{1}{a_1^2 \cdots a_n^2} \leq \sup_{n\in\mathbb{N}} \frac{1}{(\prod a_n)^2} < \infty \tag{9.50}$$

and N_n is an L^2 bounded martingale. Now by Doob's L^p-Inequality, we have

$$\mathbb{E}\left[\sup_{1\leq k\leq n}(N_k^2)\right] \leq 2^2\,\mathbb{E}[N_k^2] = 4\,\mathbb{E}[N_k^2] \tag{9.51}$$

so

$$\mu_n = N_n^2\left(\prod_{k=1}^{n} a_k^2\right) \leq N_n^2 \tag{9.52}$$

so

$$\mathbb{E}[\sup_{1\leq k\leq n}\mu_k] \leq \mathbb{E}[\sup_{1\leq k\leq n} N_k^2] \leq 4\,\mathbb{E}[N_k^2] \leq \frac{4}{(\prod_{n=1}^{\infty} a_k^2)} < \infty \tag{9.53}$$

so by Monotone Convergence we get $\mathbb{E}[\sup_{k\geq 1}\mu_k] < \infty$ so (μ_k) is UI. Thus, $\mu_k \to \mu_\infty$ in L^1.

(b) Otherwise, we have that $N_n \to N_\infty$ a.s. and $\prod_{k=1}^{n} a_k \to \prod_{k=1}^{\infty} a_n = 0$ so $\mu_n^2 = N_n^2(\prod_{k=1}^{n} a_k^2) \to N_\infty^2 \cdot 0 = 0$ where we use the fact that $N_\infty \in L^1$ so it is finite a.s. and that $\mu_n \to 0$ a.s..

□

We now define a notion of a "ratio" between probability measures that essentially gives the density of the transform.

THEOREM 9.20 (*Radon-Nikodym Theorem*)
Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$. Assume that $\mathcal{F}$ is countably generated, ie. $\exists(\mathcal{F}_n)$ such that $\mathcal{F}_n \subseteq \mathcal{F}$ and $\mathcal{F} = \sigma(\mathcal{F}_n)$. Then, the following are equivalent:

(a) $\mathbb{P}[A] = 0$ implies $\mathbb{Q}[A] = 0$ for all $A \in \mathcal{F}$

(b) $\forall\varepsilon > 0, \exists\delta > 0$ such that $\forall A \in \mathcal{F}$ with $\mathbb{P}[A] \leq \delta, \mathbb{Q}[A] \leq \varepsilon$.

(c) $\exists$ a non-nnegative random variable X such that

$$\mathbb{Q}[A] = \mathbb{E}_p[X\mathbb{1}_A] \qquad \forall A \in \mathcal{F} \tag{9.54}$$

We call X a **Radon-Nikodym Derivative** of $\mathbb{Q}$ wrt $\mathbb{P}$ and denote it by $d\mathbb{Q}/d\mathbb{P}$.

Proof. $(i) \implies (ii)$: Simply Borel Cantelli. $(ii) \implies (iii)$: L^1-martingale convergence by the obvious definition. $(ii) \implies (iii)$: trivial. □