Qualifying Exam Notes

Emily Saunders

November 2020

Contents

1 General Probability

Resources for this section are my own lecture notes from Professor Corwin's Fall 2019 Analysis & Probability I course and *Probability Theory and Examples, 5th ed.* By Rick Durrett.

1.1 Basic Measure Theory

1.1.1 Definition of a Measure and Basic Properties

Definition 1 (σ-algebra)**.** Let Ω be a set. We say $\mathcal{F} \subset 2^\Omega$ is a σ-algebra if,

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
3. For every countable collection $\{A_i\} \subset \mathcal{F}$, $\bigcup A_i \subset \mathcal{F}$

Note that given 2 we could equivalently require that 1 read $\phi \in \mathcal{F}$ and that 3 read that $\mathcal{F}$ is closed to countable intersections.

Definition 2 (Measurable Space)**.** A pair $(\Omega, \mathcal{F})$ is called a measurable space.

Definition 3 (Measure and Probability Measure)**.** A measure μ on $(\Omega, \mathcal{F})$ is any countably additive, non-negative function, $\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ i.e.

1. $\mu(A) \geq \mu(\phi) = 0$
2. $\mu(\bigcup_i A_i) \leq \sum_i \mu(A_i)$ with equality holding if the A_i are disjoint (or if their intersections have measure 0)

If $\mu(\Omega) = 1$, then we call μ a probability measure and often denote it by $\mathbb{P}$.

Definition 4. A measure μ on $(\Omega, \mathcal{F})$ is σ-finite if $\Omega = \bigcup_i A_i$ for $A_i \in \mathcal{F}$ and $\mu(A_i) < \infty$ for all i.

Example 1. All finite measures are σ-finite.

Example 2. The Lebesgue measure on $\mathbb{R}$ is $\sigma - finite$ with $A_i = [i, i+1]$ for $i \in \mathbb{Z}$

Definition 5 (Measure Space and Probability Space)**.** We say that a triple $(\Omega, \mathcal{F}, \mu)$ is a measure space. If μ is a probability measure, we call it a probability space.

Proposition 1. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the following hold,*

1. *Monotonicity: if $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$*
2. *Continuity from above: if $A_1 \subset A_2 \subset ...$ such that $\bigcup_i A_i = A$ then $\mathbb{P}(A_i) \to \mathbb{P}(A)$*
3. *Continuity from below: if $A_1 \supset A_2 \supset ...$ and $\bigcap_i A_i = A$ then $\mathbb{P}(A_i) \to \mathbb{P}(A)$*

The proof of these properties follows pretty immediately from the definition.

Proposition 2. *For any collection $\mathcal{F} \subset 2^\Omega$, there exists a unique σ-algebra $\sigma(\mathcal{F})$ such that,*

1. *$\mathcal{F} \subset \sigma(\mathcal{F})$*
2. *for all σ-algebras $\mathcal{B}$ such that $\mathcal{F} \subset \mathcal{B}$, $\sigma(\mathcal{F}) \subset \mathcal{B}$*

Remark 1. This construction requires the axiom of choice why?.

Definition 6 (Borel Sets)**.** Let X be a topological space with topology $\mathcal{T} \subset 2^X$. We say that $\mathcal{B} = \sigma(\mathcal{T})$ is the Borel σ-algebra.

1.1.2 $\pi - \lambda$ Systems

DO THIS SECTION

1.1.3 Carathéodory Extension Theorem

We can use these minimally generated σ-algebras to define measures on spaces. First, we say that $\mathcal{A} \subset 2^{\Omega}$ is an algebra if it contains Ω and is closed to compliments and finite unions.

Theorem 1 (Carathéodory Extension Theorem)**.** *Given a function,*

$$\mu_0 : \mathcal{A} \to \mathbb{R}_{\geq 0}$$

that is countably additive (known as a premeasure), there exists a measure,

$$\mu : \sigma(\mathcal{A}) \to \mathbb{R}_{\geq 0}$$

extending μ_0, and if $\mu_0(\Omega) < \infty$ (known as being finite), then μ is unique.

Proof. We begin by defining an outer measure,$\mu^* : 2^{\Omega} \to \mathbb{R} \cup \{+\infty\}$ via the following formula:

$$\mu^*(A) = \inf\{\sum_i \mu(A_i) | A_i \in \mathcal{A}, \bigcup_i A_i \supset A\}$$

Note that we can assume that the A_i are disjoint by the countably additive assumption on μ. We prove the following properties of this outer measure:

1. Monotonicity
2. Countably Subadditive
3. $\mu^* = \mu$ on $\mathcal{A}$

Monotonicity is clear from the definition. For countably subadditve, we need to show,

$$\mu^*(\bigcup_i A_i) \leq \sum_i \mu^*(A_i)$$

Fix $\varepsilon > 0$. There exists a $B_{i,j} \in \mathcal{A}$ such that $\bigcup_j B_{i,j} \supset A_i$ and $\sum_j \mu(B_{i,j}) \leq \mu^*(A_i) + \varepsilon/2^i$ by definition of infimum. Then we have that,

$$\sum_i \mu^*(A_i) + \varepsilon \geq \sum_{i,j} \mu(B_{i,j}) \geq \mu(\bigcup_{i,j} B_{i,j}) \geq \mu^*(\bigcup_i A_i)$$

where the second inequality follows from countable additivity on $\mathcal{A}$ and the last follows by monotonicity.

To see that $\mu^* = \mu$ on $\mathcal{A}$, note that by the countably additive property of μ the quantity $\sum_i \mu(A_i)$ for $A \subset \bigcup_i A_i$ is minimized when $A = \bigcup_i A_i$ and $\sum_i \mu(A_i) = \mu(\bigcup_i A_i) = \mu(A)$.

We now define a subset $E \subset \Omega$ to be *measurable* if for all other $A \subset \Omega$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

Note that by the properties we've proven $\leq$ follows immediately, so we really only need to check $\geq$. We define $\mathcal{M}$ to be the collection of all measurable sets. First we show that $\mathcal{M}$ is an algebra.

FINISH

□

The standard σ-algebra on $\mathbb{R}$ is the Borel σ-algebra $\mathcal{B}$ defined as $\sigma(\mathcal{A})$ for $\mathcal{A} = \{(a_1, b_1] \cup .... \cup (a_k, b_k] | a_1 < b_1 < ... < a_k < b_k\}$. We define measures on $(\mathbb{R}, \mathcal{B})$ via distribution functions:

Definition 7. A Lebesgue-Steiljes distribution function $F : \mathbb{R} \to \mathbb{R}$ is a function that is non-decreasing and right continuous. Additionally, if $F(-\inf) = 0$ and $F(+\inf) = 1$, we call it a probability distribution function.

We use these distribution functions to define measures on $\mathbb{R}$ by defining them on the generating algebra $\mathcal{A}$ of $\mathcal{B}$ to be,

$$\mu(\bigcup_{i=1}^{k}(a_i, b_i]) = \sum_{i=1}^{k}(F(b_i) - F(a_i))$$

In probability language, F is the *Cumulative Distribution Function* or CDF. When $F(x) = x$, we call the resulting measure the *Lebesgue Measure* on $\mathbb{R}$.

Include some comments about the space of measurable sets $\mathcal{M}$ and how we define them via the axiom of choice

1.1.4 Measurable Functions and Random Variables

We now move on to discussing the morphisms in the category of measurable spaces, measurable functions.

Definition 8 (Measurable Function)**.** A function $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ is *measurable* if for every $E \in \mathcal{F}'$, $X^{-1}(E) \in \mathcal{F}$.

Definition 9 (Random Variable and Random Vector)**.** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B})$ is measurable, then X is a random vector. In the case $d = 1$, X is a random variable

Theorem 2. *If $\mathcal{A}$ is a collection of subsets in Ω' such that $\sigma(\mathcal{A}) = \mathcal{F}'$, then X is measurable if and only if for all $E \in \mathcal{A}$, $X^{-1}(E) \in \mathcal{B}$. That is, it is sufficient to check measurability on a generating set of your σ-algebra.*

Proof. Proof needed □

Definition 10. For a function $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$, define $\sigma(X)$ to be the smallest σ-algebra on Ω such that $X : (\Omega, \sigma(X)) \to (\Omega', \mathcal{F}')$ is measurable.

Remark 2. Note that this construction also depends on $\mathcal{F}$.

Proposition 3. *A function $Y : (\Omega, \sigma(X)) \to (\Omega'', \mathcal{F}'')$ is measurable if and only if $Y = f \circ X$ for some $f : (\Omega', \mathcal{F}') \to (\Omega'', \mathcal{F}'')$ measurable.*

Proof. Proof needed QUESTION: IS THIS ONLY TRUE IN BOREL? Possible counter example: $\mathcal{F}$ is the trivial σ-algebra □

Proposition 4. *If X, Y are measurable functions $(\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, then so are $X + Y$, cX, XY, $\max(X, Y)$, $\min(X, Y)$, $\inf_n X_n$, $sup_n X_n$, $\liminf_x X_n$, $\limsup_n X_n$. For $\{X_n\}_{n \geq 0}$, the set $\Omega = \{\omega \in \Omega | \lim(X_i(\omega))$ exists$\}$ is a measurable set. meaning it is in $\mathcal{F}$?*

Theorem 3 (Lusin's)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then for all $\varepsilon > 0$ there exists a continuous $g : \mathbb{R} \to \mathbb{R}$ such that the set $\{x|f(x) = g(x)\}$ is closed and the compliment has Lebesgue measure $\leq \varepsilon$*

Check the statement of this because it was wonky in my notes In other words, any measurable function $f : \mathbb{R} \to \mathbb{R}$ can be closely approximated by a continuous function g.

Now we consider how a measurable map pushes forward a measure on $(\Omega, \mathcal{F})$ to a measure on $(\Omega', \mathcal{F}')$.

Definition 11 (Distribution Function)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a random variable. The distribution function of X, $F_X : \mathbb{R} \to [0, 1]$ is defined by,

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega|X(\omega) \leq x\})$$

Furthermore, this function defines a probability measure on $(\mathbb{R}, \mathcal{B})$.

Remark 3. Random variables are *not* defined by their distribution functions.

Proposition 5. *All distribution functions have the following properties:*

1. *Non-decreasing*
2. $\lim_{x\to+\infty} F(X) = 1, \lim_{x\to-\infty} F(X) = 0$
3. *Right continuous*
4. *Left limits exist and are equal to $\mathbb{P}(X < x)$ (we denote these limits as $F(x-)$)*
5. $F(x) - F(x-) = \mathbb{P}(X = x)$

1.1.5 Integration

The goal of this section is to define integration with respect to a σ-finite measure μ on a measurable space $(\Omega, \mathcal{F})$. We construct $\int_\Omega f d\mu$ in 4 steps, at each stage we can verify the following properties:

1. If $\varphi \geq 0$ almost everywhere, then $\int \varphi d\mu \geq 0$
2. For all $a \in \mathbb{R}$, $\int a\varphi d\mu = a \int \varphi d\mu$.
3. $\int(\varphi + \psi)d\mu = \int \varphi d\mu + \int \psi d\mu$
4. If $\varphi \leq \psi$ almost everywhere, $\int \varphi d\mu \leq \int \psi d\mu$
5. $|\int \varphi d\mu| \leq \int |\varphi|d\mu$

Step 1 For φ a simple function $\varphi = \sum_{i=1}^n a_i 1_{A_i}$, with A_i disjoint measurable sets with $\mu(A_i) < \infty$, we define,

$$\int \varphi d\mu = \sum_{i=1}^n a_i \mu(A_i)$$

Step 2 For f bounded and $f = 0$ outside some E such that $\mu(E) < \infty$, we define,

$$\int f d\mu = \sup_{\varphi \leq f} \int \varphi d\mu = \inf_{\psi \geq f} \int \psi d\mu$$

To prove that this equality is true, we can approximate our function f from above and below by simple functions via the following sequences:

$$\psi_n = \sum_{k=-n}^{n} \frac{kM}{n} 1_{f^{-1}([\frac{(k-1)M}{n}, \frac{kM}{n}])} \qquad \varphi_n = \sum_{k=-n}^{n} \frac{(k-1)M}{n} 1_{f^{-1}([\frac{(k-1)M}{n}, \frac{kM}{n}])}$$

Remark 4. Note that, as opposed to Riemannian integration, this allows us to subdivide based on our range instead of our domain. This allows us to integrate more functions, for instance,

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \text{else} \end{cases}$$

can now be integrated on any bounded subset of $\mathbb{R}$.

Step 3 For $f \geq 0$, we define,

$$\int f d\mu = \sup\{\int h d\mu | 0 \leq h \leq f, ||h||_\infty < \infty, \mu(\{x | h(x) \neq 0\}) < \infty\}$$

Step 4 Let f be such that, as defined above, $\int |f| d\mu < \infty$ (we say f is *integrable*). Define,

$$f^-(x) = \begin{cases} -f(x) & f(x) \leq 0 \\ 0 & \text{else} \end{cases} \qquad f^+(x) = \begin{cases} f(x) & f(x) \geq 0 \\ 0 & \text{else} \end{cases}$$

Then we define,

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

Remark 5. For intuition as to why this is the correct notion of "integrable", note that in the countable case we want to define the integral to be $\sum_{i \in \Omega} f(i)\mu(i)$, however if we do not have that this sum converges absolutely, that is, if $\sum_{i \in \Omega} |f(i)|\mu(i) = \infty$, then the sum $\sum_{i \in \Omega} f(i)\mu(i)$ can be rearranged to equal whatever value we want. This is clearly not a well defined notion of integration.

The following is common notation for specific types of integrals,

1. $(\mathbb{R}^d, \mathcal{B}, \text{Lebesgue})$: $\int_A f(x) dx$
2. Ω countable: $\int f d\mu = \sum_{i \in \Omega} f(i)\mu(i)$
3. $(\mathbb{R}, \mathcal{B}, \mu)$ with $\mu([a, b]) = G(b) - G(a)$: $\int f d\mu = \int f(x) dG$

Definition 12 (Expectation)**.** Let $\mathbb{P}$ be a probability measure and X a positive random variable. Then we define $\mathbb{E}[X] =\in X d\mathbb{P}$. If X is not necessarily ≥ 0, we define $\mathbb{E}[X] = \int x^+ d\mathbb{P} - \int X^- d\mathbb{P}$ provided both X^+ and X^- are integrable. A random variable X is integrable if $\mathbb{E}[X] < \infty$.

Now we can prove the first of the Borel-Cantelli lemmas:

Definition 13. For $\{A_n\}_n$ a sequence of subsets in Ω,

$$\limsup_{n\to\infty} A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \limsup_{n\to\infty} 1_{A_n}$$

(i.e. the set of all $\omega \in \Omega$ such that $\omega \in A_k$ for infinite k)

$$\liminf_{n\to\infty} A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \liminf_{n\to\infty} 1_{A_n}$$

(i.e. the set of all elements $\omega \in \Omega$ such that $\omega \in A_k$ for all but finitely many k.

Lemma 1 (Borel-Cantelli Lemma (1)). *If* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, *then* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$

Proof. Let $N(\omega) = \sum_{k=1}^{\infty} 1_{A_k}(\omega)$. Then $\mathbb{E}[|N|] = \mathbb{E}[N] = \sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, therefore N must be $< \infty$ almost surely. □

Remark 6. The converse to this (that if the probabilities of the sets are not summable the lim sup occurs with positive probability) is not true.

Proposition 6 (Jensen's Inequality). *For* $\varphi : \mathbb{R} \to \mathbb{R}$ *convex,*

$$\varphi(\int f d\mu) \leq \int \varphi(f) d\mu$$

Proof. Let $L(x) = ax + b$ be a linear function such that $L(\int f d\mu) = \varphi(\int f d\mu)$ and $L(x) \leq \varphi(x)$ for all x (such a function exists if φ is convex). Then we have,

$$\int \varphi(f) d\mu \geq \int af + bd\mu = a \int f d\mu + b = L(\int f d\mu) = \varphi(\int f d\mu)$$

□

Proposition 7 (Holder's Inequality). *For* $p, q \in (1, \infty)$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int |fg| d\mu = ||fg||_1 \leq ||f||_p ||g||_q$$

where $||f||_p = (\int |f|^p d\mu)^{1/p}$

Remark 7. When $p = q = 2$, this is Cauchy-Schwartz inquality.

Proof. If $||f||_p$ or $||g||_p = 0$, then $fg = 0$ almost everywhere. Therefore it is sufficient to prove the result when $||f||_p, ||g||_p > 0$. By deviding both sides by $||f||_p ||g||_p$, we can assume that $||f||_p ||g||_p = 1$. Fix $y \geq 0$ and for all $x \geq 0$, let,

$$\varphi(x) = x^p/p + y^q/q - xy$$

Then we have that,

$$\varphi'(x) = x^{p-1} - y \qquad \varphi''(x) = (p-1)x^{p-2}$$

Therefore x has a minimum at $x_0 = y^{1/(p-1)}$. Then we have that $q = p/(p-1)$, so $x_0^p = y^q$. Thus we have,

$$\varphi(x_0) = y^q(1/p + 1/q) - y^{1/(p-1)}y = 0$$

Since x_0 was the minimum it follows that $xy \leq x^p/p + y^q/q$, so if we let $x = |f|, y = |g|$, we get,

$$\int |fg| d\mu \leq \frac{1}{p} + \frac{1}{q} = 1 = ||f||_p ||g||_p$$

□

Proposition 8 (Markov's Inequality)**.** *Suppose* $\varphi : \mathbb{R} \to \mathbb{R}$ *has* $\varphi \geq 0$*. Let* $A \in \mathcal{B}$ *and let* $i_A = \inf\{\varphi(a) | a \in A\}$*. Then,*

$$i_A \mathbb{P}(X \in A) \leq \mathbb{E}[\varphi(X) 1_{X \in A}] \leq \mathbb{E}[\varphi(X)]$$

Proof. By definition of i_A, we have,

$$i_A 1_{X \in A} \leq \varphi(X) 1_{X \in A} \leq \varphi(X)$$

Since $\mathbb{E}[1_{X \in A}] = \mathbb{P}(X \in A)$, taking expectation gives us,

$$i_A \mathbb{P}(X \in A) \leq \mathbb{E}[\varphi(X) 1_{X \in A}] \leq \mathbb{E}[\varphi(X)]$$

□

Example 3 (Chebyshev's Inequality)**.** When $\varphi(x) = x^2$, we have that,

$$\mathbb{P}(|X - \mathbb{E}[X]| > \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Definition 14 (Product Measure)**.** Let $(X, \mathcal{A}, \mu_1), (Y, \mathcal{B}, \mu_2)$ be σ-finite measure spaces. Let $\Omega = X \times Y$, $S = \{A \times B | A \in \mathcal{A}, B \in \mathcal{B}\}, \sigma(S) = \mathcal{F}$. Then the product measure $\mu = \mu_1 \times \mu_2$ on $(\Omega, \mathcal{F})$ is defined to be the Cáratheodory extension of the measure defined on S by $\mu(A \times B) = \mu_1(A)\mu_2(B)$.

Remark 8. Visually, we can think of this measure in $\mathbb{R}^2$ as being defined on rectangles by length $\times$ width and defined on other sets as the limit of approximating the set by a covering of smaller and smaller rectangles and taking their measures.

1.2 Convergence

1.2.1 Convergence Theorems for Integrals

Definition 15 (Convergence in Measure)**.** A sequence of functions $f_n : \Omega \to \mathbb{R}$ converges in measure to $f : \Omega \to \mathbb{R}$ if for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \mu(|f_n - f| > \varepsilon) = 0$$

Theorem 4 (Bounded Convergence Theorem)**.** *Assume there exists some* $E \subset \Omega$ *such that* $\mu(E) > 0$ *and there exists some* $M > 0$ *such that for all* n*,* $f_n = 0$ *on* E^c *and* $||f_n||_\infty < M$*, and assume that* $f_n \to f$ *in measure. Then,*

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu$$

Proof. Let $\varepsilon > 0$ and $G_n = \{x : |f_n(x) - f(x)| < \varepsilon\}$. Define $B_n = E - G_n$. Then we have,

$$|\int f d\mu - \int f_n d\mu| = |\int (f_n - f) d\mu| \leq \int |f - f_n| d\mu$$

$$= \int_{G_n} |f - f_n| d\mu + \int_{B_n} |f - f_n| d\mu \leq \varepsilon\mu(E) + 2M\mu(B_n)$$

Since $f_n \to f$ in measure, $\mu(B_n) \to 0$. Therefore since ε was arbitrary and $\mu(E) < \infty$, $\int f_n d\mu \to \int f d\mu$. □

The following is an example to illustrate the necessity of the bounded condition.

Example 4. Consider the function $f_n : \mathbb{R} \to \mathbb{R}$ that is $1/n$ on $[0, n]$ and 0 otherwise. Then $\int f_n(x)dx = 1$ for all n, but $\lim_{n\to\infty} f_n = 0$, therefore $\lim_{n\to\infty} \int f_n(x)dx \neq \int \lim_{n\to\infty} f_n(x)dx$

Remark 9. If μ is finite (for instance, a probability measure), then we have that $\mu(\Omega) < \infty$, therefore all that we need to check to apply the theorem is that the sequence $\{f_n\}_n$ is uniformly bounded.

Theorem 5 (Fatou's Inequality)**.** *If $f_n \geq 0$, then*

$$\liminf_{n\to\infty}(\int f_n d\mu) \geq \int (\liminf_{n\to\infty} f_n) d\mu$$

Proof. Let $g_n(x) = \inf_{m\geq n} f_m(x)$. Then we have, $f_n(x) \geq g_n(x)$ and $g_n \xrightarrow{n\to\infty} \liminf_{n\to\infty} f_n(x)$. Since we have that $\int f_n d\mu \geq \int g_n d\mu$, it is sufficient to show,

$$\liminf_{n\to\infty} \int g_n d\mu \geq \int \lim_{n\to\infty} g_n d\mu$$

Let $E_1 \subset E_2 \subset ...$ be sets of finite measure such that $\bigcup_{n=1}^{\infty} E_n = \Omega$. Since $g_n \geq 0$, $g_n \geq g_{n+1}$ and for fixed m we have that $\min(g_n, m)1_{E_m} \to \min(\liminf_{n\to\infty} f_n, m)1_{E_m}$, the bounded convergence theorem tells us that,

$$\liminf_{n\to\infty} \int g_n d\mu \geq \int g_n d\mu \geq \int_{E_m} \min(g_n, m) d\mu \to \int_{E_m} \min(\liminf_{n\to\infty} f_n, m) d\mu$$

Taking the sup over all m concludes the proof. □

Theorem 6 (Monotone Convergence Theorem)**.** *If $f_n \geq 0$ and $f_n \nearrow f$ monotonically, then $\int f_n d\mu \nearrow \int f d\mu$*

Proof. Fatou's inequality tells us that $\liminf_{n\to\infty} \int f_n d\mu \geq \int f d\mu$. On the other hand, $f_n \leq f$ implies that $\limsup_{n\to\infty} \int f_n d\mu \leq \int f d\mu$. □

Theorem 7 (Dominated Convergence Theorem)**.** *If $f_n \xrightarrow{a.e.} f$, and $|f_n| \leq g$ for g integrable, then $\lim_{n\to\infty} f_n d\mu \to \int f d\mu$*

Proof. We have that $f_n + g \geq 0$, so Fatou's inequality tells us that,

$$\liminf_{n\to\infty} \int f_n + g d\mu \geq \int f + g d\mu$$

Subtracting $\int g d\mu$ from both sides tells us that,

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int f d\mu$$

We can apply the same logic to $-f_n$ to obtain,

$$\limsup_{n\to\infty} \int f_n d\mu \leq \int f d\mu$$

which tells us that $lim_{n\to\infty} \int f_n d\mu \to \int f d\mu$. □

Theorem 8 (Fubini's Theorem)**.** *Let $(X, \mathcal{A}, \mu_1), (Y, \mathcal{B}, \mu_2)$ be measure spaces, and let $(X \times Y, \sigma(\mathcal{A} \times \mathcal{B}), \mu_1 \times \mu_2)$ be the product space as defined in definiton 13. If $f : X \times Y \to \mathbb{R}$ is a measureable function such that $f \geq 0$ or $\int |f| d\mu < \infty$, then,*

$$\int_{X\times Y} f d\mu = \int_X \int_Y f(x, y)\mu_2(dy)\mu_1(dx)$$

Proof. We will prove the first equality, which implies the second by symmetry. To prove that the integral on the left makes sense, we need to check that (i) for a fixed x, the function $f(x, -)$ is $\mathcal{B}$-measureable, and (ii) for a fixed y, the function $x \mapsto \int_Y f(x, y)\mu_2(dx)$ is $\mathcal{A}$-measurable.

We begin with the case of f an indicator function 1_E. Let $E_x = \{y \in Y : (x, y) \in E\}$ be the cross-section of E at x.

Lemma 2. *If $E \in \sigma(\mathcal{A} \times \mathcal{B})$, then $E_x \in \mathcal{B}$.*

Proof. We can easily verify that $(E^c)_x = E_x^c$ and $(\bigcup_{i=0}^{\infty} E_i)_x = \bigcup_{i=0}^{\infty} E_{i,x}$, therefore if $\mathcal{E}$ is the collection of sets in $\sigma(\mathcal{A} \times \mathcal{B})$ then $\mathcal{E}$ is a σ-algebra. Observing that $\mathcal{A} \times \mathcal{B} \subset \mathcal{E}$ concludes the proof. □

Lemma 3. *If $E \in \sigma(\mathcal{A} \times \mathcal{B})$, then $g(x) = \mu_2(E_x)$ is $\mathcal{A}$-measurable and,*

$$\int_X g d\mu_1 = \mu_1 \times \mu_2(E)$$

Proof. Requires $\pi - \lambda$ stuff, so come back to this (maybe make an appendix?) □

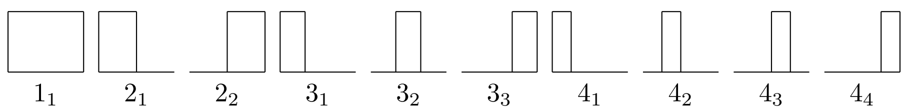
□

1.2.2 Convergence of Random Variables

Definition 16 (Almost Sure Convergence)**.** A sequence of random variables $\{X_n\}_n$ converges almost surely to X (denoted $X_n \xrightarrow{a.s.} X$) if $\mathbb{P}(\{\omega \in \Omega | X_n(\omega) \text{ does not converge to } X(\omega)\}) = 0$

Definition 17 (Convergence in Probability)**.** A sequence of random variables $\{X_n\}_n$ defined on the same probability space converges in probability to a random variable X (denoted $X_n \xrightarrow{p} X$) if they converge in measure with respect to $\mathbb{P}$.

Remark 10. This definition is weaker than almost sure convergence. For example, consider the following sequence of random variables on $[0,1]$ with uniform measure: For each $k \in \mathbb{N}$ we subdivide $[0,1]$ into k segments of length $\frac{1}{k}$ and label the i-th segment of length $\frac{1}{k}$ k_i. Then we get the following sequence of segments $1_1, 2_1, 2_2, 3_1, 3_2, 3_3, 4_1, ...$ Define X_n to be 1 on the n-th segment in this sequence and 0 otherwise. Then we have that for any $0 < \varepsilon < 1$, if $X_n = k_i, \mathbb{P}(|X_n - X| > \varepsilon) = \frac{1}{k} \to 0$, however for every $x \in [0,1]$ the sequence $X_n(x)$ does not converge, so this sequence does not converge almost surely (or even anywhere). The first few instances elements of this sequence are illustrated below:



Proposition 9. *A sequence of random variables $\{X_n\}_n$ converges in probability to X if and only if for every subsequence $n(k)$ there exists a further subsequence $\{X_{n(k_j)}\}_j$ that converges almost surely to X.*

Proof. To prove convergence in probability implies every subsequence has a almost surely convergent further subsequence it is sufficent to check this just for the entire sequence (since any subsequence converges in probability as well).This direction of the proposition is an application of the first Borel-Cantelli Lemma. For all $k \in \mathbb{N}$, there exists an $n_k > n_{k-1}$ such that,

$$\mathbb{P}(|X_{n_k} - X| > \frac{1}{k}) < \frac{1}{2^k}$$

Thus since $\sum_{k=1}^{\infty} \mathbb{P}(|X_{n_k} - X_n| > \frac{1}{k}) < \infty$, it happens infinitely often with probability 0, therefore converges almost surely.

To prove the other direction, we first prove the following lemma,

Lemma 4. *Let y_n be a sequence of elements in a topological space. If every subsequence of y_n has a further subsequence that converges to y, then $y_n \to y$.*

Proof. Suppose $y_n \nrightarrow y$. Then there exists an open neighborhood $y \in U$ that does not contain infinitely many y_n. From these we can then construct a subset that does not converge to y. □

Thus applying this lemma to the sequence $y_n = \mathbb{P}(|X_n - X| > \varepsilon)$ for all y gives the desired result. □

Corollary 1. *If f is continuous, $X_n \xrightarrow{p} X$, then $f(X_n) \xrightarrow{p} f(X)$.*

Definition 18 (Convergence in Distribution/Weak Convergence)**.** A sequence of distributions functions $\{F_n\}_n$ converges weakly to a function F (denoted $F \Rightarrow F$) if $F_n(y) \to F(y)$ for all y that are continuity points of F. A sequence of random variables $\{X_n\}_n$ converges in distribution to a random variable X (denoted $X_n \Rightarrow X$) if their distribution functions converge weakly to the distribution function of X.

Remark 11. This definition is far weaker than the others. In particular, it is not even required that the random variables all be defined on the same probability space.

Theorem 9. *If $F_n \Rightarrow F$ then there are random variables $Y_n, n \geq 1$ and Y such that Y_n has distribution function F_n, Y has distribution function F and $Y_n \xrightarrow{a.s.} Y$.*

Proof. Page 103 in Durrett □

Theorem 10. *$X_n \Rightarrow X$ if and only if for every bounded continuous function g, we have that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$.*

Proof. Let Y_n have the same distribution as X_n and be such that Y_n converges almost surely to $Y \overset{d}{\sim} X$. Since g is continuous we have that $g(Y_n) \to g(Y)$ almost surely and the bounded convergence theorem implies that,

$$\mathbb{E}[g(x_n)] = \mathbb{E}[g(Y_n)] \to \mathbb{E}[g(Y)] = \mathbb{E}(g(X)]$$

To prove the converse, let,

$$g_{x,\varepsilon}(y) = \begin{cases} 1 & y \leq x \\ 0 & y \geq x + \varepsilon \\ \text{linear interpolation} & x \leq y \leq x + \varepsilon \end{cases}$$

I.e. g is a continuous interpolation between $1_{y \leq x}$ and $1_{y \leq x+\varepsilon}$. Then we have that,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \limsup_{n\to\infty} \mathbb{E}[g_{x,\varepsilon}(X_n)] = \mathbb{E}[g_{x,\varepsilon}(X)] \leq \mathbb{P}(X \leq x + \varepsilon)$$

Letting $\varepsilon \to 0$, we get $\limsup_{n\to\infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x)$. In the other direction we have that,

$$\liminf_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \liminf_{n\to\infty} \mathbb{E}[g_{x-\varepsilon,\varepsilon}(X_n)] = \mathbb{E}[g_{x-\varepsilon.\varepsilon}(X)] \geq \mathbb{P}(X \leq x - \varepsilon)$$

So again we take $\varepsilon \to 0$ and get that $\liminf_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \mathbb{P}(X \leq x)$ if x is a continuity point. □

Theorem 11 (Helly's Selection Theorem)**.** *For every sequence F_n of distribution functions, there is a subsequence F_{n_k} and a right continuous nondecreasing function F so that $\lim_{k\to\infty} F_{n_k}(y) = F(y)$ for all continuity points y of F.*

Proof. We begin by making a diagonal argument on the rationals. Letting $q_1, q_2, ...$ be an enumeration of the rationals, we notice that for every i, $\{F_n(q_i)\}_n$ is a sequence bounded by $[0, 1]$. Therefore we can inductively construct a subsequence $n_{k(i)}$ such that $n_{k(i)}$ is a subsequence of $n_{k(i-1)}$ and $\{F_{n_{k(i)}}(q_i)\}_{k(i)}$ converges. Then we diagonally construct a subsequence $n_i = n_{i(i)}$ so that for all

$q \in \mathbb{Q}$, $\{F_{n_i}(q)\}_i$ converges to some b_q. Furthermore, we have that for $p, q \in \mathbb{Q}, p < q$, $b_p < b_q$. We now define,

$$F(x) = \inf\{b_q | q \in \mathbb{Q}, q > x\}$$

It is easily verifiable by the monotone property mentioned above that $F(q) = b_q$ and F is right continuous and monotone. Thus all that remains to show is that $F_{n_i}(y) \to F(y)$ for y a continuity point of F. By continuity, for every $\varepsilon > 0$ we can select rationals p, q such that $p < y < q$ and,

$$F(y) - \varepsilon < F(p) \leq F(y) \leq F(q) < F(y) + \varepsilon$$

Since our subsequence converges on the rationals to F, and F_{n_i} is monotone for all i, we see that for sufficiently large i,

$$F(y) - \varepsilon < F_{n_i}(p) \leq F_{n_i}(y) \leq F_{n_i}(q) < F(y) + \varepsilon$$

which concludes the proof. □

Remark 12. Note that this is NOT equivalent to the corresponding measures of the subsequence converging weakly to a measure μ as F is not necessarily a distribution function (i.e. it may not have the appropriate limits at $\pm\infty$). For a simple example, let F_n be such that $F(x) = 0$ for $x < n$ and $F_n(x) = 1$ for $x \geq n$. Then the resulting pointwise limit F is 0, which is clearly not a probability measure as $lim_{x \to +\infty} F(x) \neq 1$. This example illustrates that generally in the limit mass is able to escape to $\pm\infty$ (we could just as easily compute a sequence where mass escapes to $-\infty$ by letting $G_n(x) = F_{-n}(x)$). However, the following definition and theorem gives necessary and sufficient criteria to keep this from happening.

Definition 19 (Tightness)**.** A sequence of probaility measures μ_n on $\mathbb{R}$ is tight if for every ε, there exists a M_ε such that,

$$\limsup_{n \to \infty} 1 - \mu_n([-M_\varepsilon, M_\varepsilon]) \leq \varepsilon$$

Equivalently, a sequence of distribution functions F_n is tight if for every ε there exists a M_ε such that,

$$\limsup_{n \to \infty} 1 - F(M_\varepsilon) + F(-M_\varepsilon) \leq \varepsilon$$

Theorem 12. *A squence of distribution functions F_n is tight if and only if every subsequential limit is the distribution function of a probability measure.*

Proof. Suppose the sequence F_n is tight and $F_{n_k}(y) \to F(y)$ for all continuity points of F. Fix some $\varepsilon > 0$ and let $r < -M_\varepsilon, s > M_\varepsilon$. Then we have that,

$$1 - F(s) + F(r) = \lim_{k \to \infty} 1 - F_{n_k}(s) + F_{n_k}(r) \leq \limsup_{n \to \infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \leq \varepsilon$$

This implies that $\limsup_{x \to \infty} 1 - F(x) + F(-x) \leq \varepsilon$. Since ε was arbitrary and $0 \leq F(-x) \leq F(x) \leq 1$, this implies that F is a probability distribution function.

To prove the converse, suppose that F_n is not tight. Then there is an $\varepsilon > 0$ and a subsequence such that,

$$1 - F_{n_k}(k) + F_{n_k}(-k) \geq \varepsilon$$

for all k. By choosing a further subsequence, we may assume that there exists a monotone, right continuous F such that $F_{n_k}(y) \to F$ for all continuity points of F. Let $r < 0 < s$ be continuity points of F. Then we have,

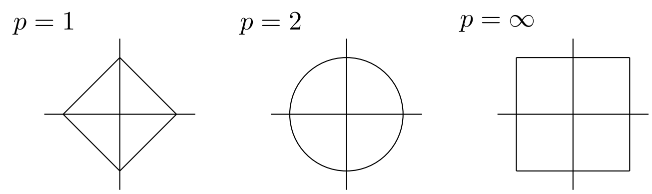
$$1 - F(s) + F(r) = \lim_{k\to\infty} 1 - F_{n_k}(s) + F)n_k(r) \geq \liminf_{k\to\infty} 1 - F_{n_k}(k) + F_{n_k}(-k) \geq \varepsilon$$

Therefore F cannot be a probability distribution function. □

1.3 L^p Space

Definition 20 (L^p Space)**.** $L^p(\Omega, \mathcal{F}, \mu)$ is the space of all functions $f : \Omega \to \mathbb{R}$ such that $||f||_p = (\int |f|^p d\mu)^{1/p}$ is finite. $||f||_\infty = \inf\{M \geq 0 | \mu(\{x||f(x)| > M\}) > 0\}$ (i.e. M such that $|f|$ is bounded by M almost surely).

To illustrate the difference between these norms, note that when $|\Omega| = 2$, these are all norms on $\mathbb{R}^2$. The following illustrate the circle $\{x \in \mathbb{R}^2 || |x||_p = 1\}$ for various p:



Theorem 13 (Minkowski's Theorem)**.** $||f + g||_p \leq ||f||_p + ||g||_p$.

Proof. Proof needed □

Theorem 14 (Riesz-Fischer)**.** *L^p is complete for all $p \in [1, \infty]$*

Proof. Proof needed □

Corollary 2. *$|| - ||_p$ is a seminorm (fails the 0 only for 0 condition). If we mod out by the relation $f \sim g$ if $f = g$ almost surely, then $|| - ||_p$ is a Banach space (complete normed vector space).*

The following are some other useful properties of L^p space:

1. L^2 is a Hilbert Space (a real or complex inner product space that is also a complete metric space).
2. Embeddings: for $1 \leq p < q \leq \infty$ the following hold:
   - $L^q \subset L^p$ if and only if Ω does not contain sets of finite but arbitrarily large measure (for instance if μ is a probability measure).
   - $L^p \subset L^q$ if and only if Ω does not contain sets of nonzero but arbitrarily small measure.
   - If $\Omega = \{1, 2, ..., n\}$, then $L^p \cong L^q \cong \mathbb{R}^n$.
3. Dual Spaces: For $p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, L^p and L^q are dual to each other as Banach spaces, with identification given by $T_g f = \int fg d\mu$. If μ is σ-finite, then $(L^1)^* \cong L^\infty$ but we only have that $(L^\infty)^* \subset L^1$.

4. For $|\Omega| = n$, if $p > q$, then $||x||_p \leq ||x||_q \leq n^{1/q-1/p}||x||_p$ (and thus they induce the same topology).

Proposition 10. *For $r > s \geq 1$, convergence in L^r implies convergence in L^s but not vice versa.*

Proposition 11. *Both almost sure convergence and L^p convergence imply convergence in distribution*

1.4 Independence

Definition 21 (Independent Sets)**.** Sets $A_1, ..., A_n$ are indpendant if for all $I \subset \{1, ..., n\}$,

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

Definition 22 (Independent Random Variables)**.** Real valued random variables $X_1, ..., X_n$ are independent if for all $B_1, ..., B_n \in \mathcal{B}$,

$$\mathbb{P}(\bigcap_{i=1}^{n} \{X_i \in B_i\}) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i)$$

An infinite sequence of random variables is said to be independent if every finite subset is independent.

Definition 23 (Independent σ-Algebras)**.** σ-algebras $\mathcal{F}_1, ..., \mathcal{F}_n \subset \mathcal{F}$ are independent if for all $A_1 \in \mathcal{F}_i, ..., A_n \in \mathcal{F}_n$,

$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

Remark 13. Random variables $X_1, ..., X_n$ independent is equivalent to the σ-algebras $\sigma(X_1)...\sigma(X_n)$ independent.

Remark 14. Pairwise independence is not as strong as collective independence. For instance, let X_1, X_2, X_3 be i.i.d *Bernoulli*$(1/2)$ random variables. Then we have that the following sets are pairwise independent, but not collectively independent:

$$A_1 = \{X_2 = X_3\} \quad A_2 = \{X_1 = X_3\} \quad A_3 = \{X_1 = X_2\}$$

each individually has probability $1/2$, and the intersection of any 2 sets is $\{X_1 = X_2 = X_3\}$ which has probability $1/4$, however the intersection of all 3 is still $\{X_1 = X_2 = X_3\}$ which has probability $1/4 \neq (1/2)^3$.

Theorem 15. *For random variables $X_1, ..., X_n$, to prove independence it is sufficient to check that for all $x_1, ..., x_n \in (-\infty, \infty)$,*

$$\mathbb{P}(X_1 < x_1, ..., X_n < x_n) = \prod \mathbb{P}(X_i < x_i)$$

Proof. $\pi - \lambda$ system proof in Durrett □

Proposition 12. *If $(X_1, ..., X_n)$ are absolutely continuous random variables (i.e. they have density $f(x_1, ..., x_n)$) then $X_1, ..., X_n$ are independent if and only if $f(x_1, ..., x_n) = g_1(x_1)...g_n(x_n)$.*

Proof. We can write,

$$\mathbb{P}(X_1 < x_1, ..., X_n < x_n) = \int_{y_i < x_i} f(y_1, ..., y_n) d\vec{y} = \int_{-\infty}^{x_n} ... \int_{-\infty}^{x_1} g_1(y_1)...g_n(y_n) dy_1...dy_n$$

$$= \prod_{i=1}^{n} (\int_{-\infty}^{x_i} g_i(y_i) dy_i) = \prod_{i=1}^{n} \mathbb{P}(X_i < x_i)$$

□

The following are additional properties of independence:

1. Functions of independent random variables are independent.
2. If X, Y are independent with distribution μ, ν respectively, and $h : \mathbb{R}^2 \to \mathbb{R}$ is such that either $h \geq 0$ or $\mathbb{E}[|h(X,Y)|] < \infty$, then $\mathbb{E}[h(X,Y)] = \int \int h(x,y) d\mu(x) d\nu(y)$.
3. If $X_1, ..., X_n$ are independent random variables with $X_i \sim \mu_i$, then $(X_1, ..., X_n) \sim \mu_1 \times ...\mu_n$

Distribution functions for sums of independent random variables can be given by convolutions of their distribution functions. Let X and Y be independent with X having density function f and Y having density function g. Then their sum $X + Y$ has density function,

$$h(z) = (f * g)(z) = \int_{-\infty}^{\infty} f(x) g(z - x) dx = \int_{-\infty}^{\infty} g(y) f(z - y) dy$$

PDFs or CDFs?

Constructions of Infinite Sequences of Independent Random Variables Fill in form page 36 of notes, too tired to do it now

We now can prove a slightly weaker converse to the first Borel-Cantelli Lemma:

Lemma 5 (Borel-Cantelli Lemma (2))**.** *If A_n are independent, then $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ implies that $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.*

Proof. We assume the inequality $1 - x \leq e^{-x}$ for $x \in [0, 1]$. Then we have that,

$$\mathbb{P}((\bigcup_{n=m}^{N} A_n)^c) = \mathbb{P}(\bigcap_{n=m}^{N} A_n^c) = \prod_{n=m}^{N} (1 - \mathbb{P}(A_n)) \leq e^{-\sum_{n=m}^{N} \mathbb{P}(A_n)} \xrightarrow{N\to\infty} 0$$

Which then implies that $\lim_{N\to\infty} \mathbb{P}(\bigcup_{n=m}^{N} A_n) = 1$ for all m. □

1.5 Laws of Large Numbers

1.5.1 L^2 Law of Large Numbers

Definition 24. 2 random variables X, Y are uncorrelated if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ or equivalently if,

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0$$

Remark 15. This is weaker than independence.

Theorem 16 (L^2 Law of Large Numbers)**.** *Let $X_1, ..., X_n \in L^2(\Omega)$ be uncorrolated random variables. Then $Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$.*

Proof. Proof □

Remark 16. Note that this implies that if $X_1, X_2, ...$ are *i.i.d.* with variance $Var(X_i) = \sigma^2$, then we have that,

$$\lim_{n\to\infty} Var(\frac{X_1 + ... + X_n}{n}) = \lim_{n\to\infty} \frac{1}{n^2} Var(X_1 + ... + X_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Then we have that $\frac{\sum_{i=1}^n X_i}{n}$ converges to a constant in L^2. Furthermore, convergence in L^2 implies convergence in L^1, thus $X_n \to \mu$ for $\mu = \mathbb{E}[X_i]$.

1.5.2 Weak Law of Large Numbers

Theorem 17 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, ...$ be i.i.d. with $\mathbb{E}[|X_i|] < \infty$, and set $S_n = X_1 + ... + X_n$, $\mu = \mathbb{E}[X_i]$. Then as $n \nearrow \infty$,*

$$\frac{S_n}{n} \xrightarrow{p} \mu$$

Proof. We begin by truncating our random variables X_i by a cutoff $c > 0$. Let $X_i^c = X_i 1_{|X_i|<c}$, and define $Y_i^c = X_i - X_i^c$. Then we have,

$$\frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\sum_{i=1}^n (X_i^c + Y_i^c) = \frac{1}{n}\sum_{i=1}^n X_i^c + \frac{1}{n}\sum_{i=1}^n Y_i^c$$

Let $\xi_n^c = \frac{1}{n}\sum_{i=1}^n X_i^c$ and $\eta_n^c = \frac{1}{n}\sum_{i=1}^n Y_i^c$, and define, $a_c = \mathbb{E}[X_i^c], b_c = \mathbb{E}[Y_i^c]$. Then we have,

$$\delta_n = \mathbb{E}[|\frac{S_n}{n} - \mu|] = \mathbb{E}[|\xi_n^c + \eta_n^c - \mu|] \le \mathbb{E}[|\xi_n^c - a_c|] + \mathbb{E}[|\eta_n^c - b_c|] \le \mathbb{E}[|\xi_n^c - a_c|] + 2\mathbb{E}[|Y_i^c|]$$

Now note that X_i^c is bounded and therefore has finite second moment. Thus, we can apply the L^2 law of large numbers and see that $\xi_n^c \xrightarrow{p} a_c$, so we get,

$$\limsup_{n\to\infty} \delta_n \le 2\mathbb{E}[|Y_i^c|]$$

$Y_i^c \xrightarrow{a.s.} 0$, so we can apply dominated convergence theorem with X_i as the bounding variable to see that $2\mathbb{E}[|Y_i^c|] \xrightarrow{c\to\infty} 0$, concluding the proof.

In my notes I have at the end the phrase "slightly different version with slightly different hypothesis." What is that referring to? Applying the statement of the L^2 law of large numbers? □

1.5.3 Strong Law of Large Numbers

We now move onto the strong laws of large numbers, considered strong because they prove almost sure convergence rather than convergence in probability (or L^2).

Theorem 18 (L^4 Strong Law of Large Numbers)**.** *Let X_i be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{E}[X_i^4] < \infty$, then $S_n = X_1 + ... + X_n$, $\frac{S_n}{n} \to \mu$ a.s.*

Proof. We can assume $\mu = 0$ by replacing X_i with $X_i - \mu$. Then we have,

$$\mathbb{E}[S_n^4] = \sum_{i=1}^{n} \mathbb{E}[X_i^4] + \binom{4}{2} \sum_{1 \le i < j \le n} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] = n\mathbb{E}[X_i^4] + 3n(n-1)\mathbb{E}[X_i^2] \le cn^2$$

Then we have that

$$\mathbb{P}(|S_n| > n\varepsilon) \le \frac{\mathbb{E}[S_n^4]}{(n\varepsilon)^4} \le \frac{cn^2}{(n\varepsilon)^4} \le \frac{1}{n^2\varepsilon^4}$$

So we have by Borel-Cantelli, $\mathbb{P}(|S_n| > n\varepsilon$ infinitely often$) = 0$ □

Theorem 19 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, ...$ be pairwise independent identically distributed random variables with $\mathbb{E}[|X_i|] < \infty$, $\mathbb{E}[X_i] = \mu$. Then,*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu$$

Additionally, if $\mathbb{E}[X_i^+] = \infty$ and $\mathbb{E}[X_i^-] < \infty$, then $\frac{S_n}{n} \to \infty$ almost surely.

We will give two proofs of the strong law of large numbers, the first of which is due to Etemadi in 1981:

Proof. need proof of the second statement, should be an application of Borel-Cantelli? Just as in the proof of the weak law of large numbers, we begin by truncating. Let $Y_k = X_k 1_{|X_k| \le k}$. We first show that it is sufficient to prove the convergence of the truncated variables.

Lemma 6. *Let $T_n = Y_1 + ... + Y_n$. It is sufficient to show that $\frac{T_n}{n} \xrightarrow{a.s.} \mu$.*

Proof. $\sum_{k=1}^{\infty} \mathbb{P}(|X_k| > k) \le \int_0^\infty \mathbb{P}(|X_1| > t)dt = \mathbb{E}[|X_i|] < \infty$ so by Borel-Cantelli we have that $\mathbb{P}(X_k \ne Y_k$ infinitely often$) = 0$ Therefore we have that for every $\omega \in \Omega$ except on a measure 0 set, there exists a $R(\omega) < \infty$ such that $|S_n(\omega) - T_n(\omega)| < R(\omega)$, so $|\frac{S_n(\omega)}{n} - \frac{T_n(\omega)}{n}| < \frac{R(\omega)}{n} \xrightarrow{n\to\infty} 0$. □

Lemma 7. $\sum_{k=0}^{\infty} \frac{Var(Y_k)}{k^2} \le 4\mathbb{E}[X_1] < \infty$

Proof. $Var(Y_k) \le \mathbb{E}[Y_k^2] = \int_0^\infty 2y\mathbb{P}(|Y_k| > y)dy$why?$\le \int_0^k 2y\mathbb{P}(|X_1| > y)$ why not equal?

Since everything is ≥ 0, we can apply Fubini's theorem to see,

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}[Y_k^2]}{k^2} \le \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^\infty 1_{y<k} 2y\mathbb{P}(|X_1| > k)dy = \int_0^\infty (\sum_{k=1}^{\infty} \frac{1_{y<k}}{k^2})2y\mathbb{P}(|X_1| > y)dy$$

So to complete the proof we show that if $y \ge 0$, then $2y\sum_{k>y} \frac{1}{k^2} \le 4$. (For this see Lemma 2.4.4 of Durrett, it is not hard). □

Now our goal is to combine these two lemmas to prove the Strong Law of Large Numbers. The remainder of this proof follows the following outline:

1. Note that we can write $X_n = X_n^+ - X_n^-$, so we can assume that we are working with $X_i > 0$ and combine the results for X^+ and X^- to obtain the general case.
2. Choose a subsequence T_{k_n} and show convergence.
3. Use the monotonicity to conclude that T_n converges almost surely.

To choose the subsequence, For some $\alpha > 1$, define $k_n = [\alpha^n]$ (where $[x]$ is defined to be the closest integer to x). Then we have that k_n grows exponentially. So now we have,

$$\sum_{n=1}^{\infty} \mathbb{P}(|T_{k_n} - \mathbb{E}[T_{k_n}]| > \varepsilon k_n) \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{Var(T_{k_n})}{k_n^2} = \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{m=1}^{k_n} Var(Y_m)$$

$$= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} Var(Y_m) \sum_{k_n \geq m} \frac{1}{k_n^2} \leq 4(1 - \frac{1}{\alpha^2})^{-1} \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{\mathbb{E}[Y_m^2]}{m^2} < \infty$$

Thus by Borel-Cantelli we can conlude that,

$$\frac{T_{n_k} - \mathbb{E}[T_{n_k}]}{n_k} \xrightarrow{a.s.} 0$$

Note that this is sufficient to show this subsequence converges to μ as this implies that,

$$\lim_{k\to\infty} \frac{T_{n_k}}{n_k} = \lim_{k\to\infty} \frac{\mathbb{E}[T_{n_k}]}{n_k} = \lim_{k\to\infty} \frac{\mathbb{E}[S_{n_k}]}{n_k} = \mu$$

Now to handle the intermediate values we have the following inequality for $k_n \leq m \leq k_{n+1}$:

$$\frac{T_{k_n}}{k_{n+1}} \leq \frac{T_m}{m} \leq \frac{T_{k_{n+1}}}{k_n}$$

So then this implies that,

$$(\frac{k_n}{k_{n+1}}) \frac{T_{n_k}}{n_k} \leq \frac{T_m}{m} \leq (\frac{k_{n+1}}{k_n}) \frac{T_{k_{n+1}}}{n_{k+1}}$$

and by our choice of subsequence we have that $\frac{k_{n+1}}{k_n} \to \alpha$ as $n \to \infty$, for arbitrarily chosen $\alpha > 1$, thus,

$$\mu \leq \liminf_{m\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \mu$$

which concludes the first proof. □

To present the second proof we first need to prove the following theorems of Kolmogorov:

Theorem 20 (Kolmogorov's Maximal Inequality)**.** *Let $X_1, X_2, ...$ be independent centered random variables with $Var(X_i) < \infty$. Let $S_n = X_1 + ... + X_n$. Then,*

$$\mathbb{P}(\max_{1\leq k\leq n} |S_k| \geq x) \leq \frac{Var(S_n)}{x^2}$$

Remark 17. Note that this is stronger than the Chebyshev inequality since we take the maximum over all S_k for $k \leq n$.

Proof. Define $A_k = \{|S_k| \geq x, |S_j| < x \text{ for all } j < k\}$. Note that the A_k are disjoint, and that $\bigcup_{k=1}^n A_k = \{\max_{1\leq k\leq n} |S_k| \geq x\}$. Then we have,

$$\mathbb{E}[S+n^2] = \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P}...$$

UNFINISHED IN MY NOTES see Durrett to finish ☐

Theorem 21 (Kolmogorov 0-1 Law)**.** *Define $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, ...)$ the smallest σ-field such that all $X_m, m \geq n$ are measurable. Define the tail σ-field $\mathcal{T} = \bigcap_{n=1}^{\infty}$. If the sequence $X_1, X_2, ...$ are independent, then for an $A \in \mathcal{T}$, $\mathbb{P}(A) = 0$ or 1.*

Proof. The idea of the proof is to show that A is independent from itself and therefore must have probability either 0 or 1. This requires $\pi = \lambda$ systems and can be found in Durrett REVISIT IF YOU EVER GET AROUND TO $\pi - \lambda$ STUFF ☐

We use these two theorems to prove the following theorem that we will the apply to give a second proof of the Strong Law of Large Numbers.

Theorem 22 (Kolmogorov's One Series Theorem)**.** *Suppose $X_1, X_2, ...$ are independent centered random variables. If $\sum_{i=0}^{\infty} Var(X_i) < \infty$, then with probability 1,*

$$\sum_{i=1}^{\infty} X_i(\omega)$$

converges.

Proof. Let S_n be the n-th partial sum of the series. To prove this we will prove

$$\mathbb{P}(\{S_n\}_n \text{ is Cauchy}) = 1$$

By Kolmogorov's maximal inequality we have,

$$\mathbb{P}(\max_{M\leq m\leq N} |S_m - S_M| > \varepsilon) \leq \frac{1}{\varepsilon^2} Var(S_N - S_M) = \frac{1}{\varepsilon^2} \sum_{n=M+1}^{N} Var(X_n) < \infty$$

So we have,

$$\mathbb{P}(\sup_{m,n\geq M} |S_m - S_n| > 2\varepsilon) \leq \mathbb{P}(\sup_{m\geq M} |S_m - S_M| > \varepsilon) \xrightarrow{M\to\infty} 0$$

☐

We also will need the following lemma, which we won't prove here,

Lemma 8 (Kronecker's Lemma)**.** *If $a_n \to \infty$ and $\{x_n\}_n$ is such that $\sum_{i=1}^{\infty} \frac{x_n}{a_n}$ converges then $\frac{1}{a_n} \sum_{i=1}^n x_n \xrightarrow{n\to\infty} 0$.*

Now we give the second proof of the Strong Law of Large Numbers:

Proof. Again, let $Y_k = X_k 1_{|X_k| \leq k}$, and let $Z_k = Y_k - \mathbb{E}[Y_k]$. Recall that by lemma 4 of proof 1 it is sufficient to show to show that $\frac{T_n - \mathbb{E}[T_n]}{n} \to 0$ almost surely. We have, $Var(\frac{Z_k}{k}) = \frac{Var(Y_k)}{k^2} < \infty$ by lemma 5 in the first proof, therefore by Kolmogorov's one series theorem, $\sum_{k=1}^{\infty} \frac{Z_k}{k}$ converges almost surely. By Kronecker's lemma, this implies that $\frac{1}{n}\sum_{k=1}^{n} Z_k \to 0$ almost surely, so,

$$\frac{1}{n}\sum_{k=1}^{n} Z_k = \frac{\sum_{k=1}^{n} Y_k - \mathbb{E}[Y_k]}{n} = \frac{T_n - \mathbb{E}[T_n]}{n} \xrightarrow{a.s.} 0$$

□

1.5.4 Applications

The following are examples of applications of the laws of large numbers chosen from Durrett,

Bernstein Polynomials Let f be continuous on $[0,1]$, and define the *n-th Bernstein polynomial of f* to be,

$$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f(\frac{m}{n})$$

Then $\sup_{x \in I} |f_n(x) - f(x)| \xrightarrow{n \to \infty} 0$.

Proof. Let $X_i \sim Bern(p)$, $S_n = X_1 + ... + X_n$, then we have that, $\mathbb{P}(S_n = m) = \binom{n}{m} p^m (1-p)^{n-m}$. Therefore,

$$f_n(p) = \mathbb{E}[f(\frac{S_n}{n})] \xrightarrow{n \to \infty} f(p)$$

□

High-Dimensional Cubes Concentrate on the Boundary of a Ball Let $X_i \sim Unif[-1,1]$, $Y_i = X_i^2$, Then we can easilyt see that $\mathbb{E}[Y_i] = \frac{1}{3}$ and $Var(Y_i) \leq 1$. Thus by the law of large numbers we have,

$$\frac{X_1^2 + ... + X_n^2}{n} \xrightarrow{n \to \infty} \frac{1}{3}$$

Therefore, by the law of large numbers, for some small $\delta > 0$, the δ-neighborhood of the ball $B(0, \sqrt{\frac{n}{3}})$ contains almost all of the mass of the cube.

Renewal Theorem Let $X_1, X_2, ...$ be i.i.d. random bariables with $0 < X_i < \infty$. Define $T_n = X_1 + ... + X_n$ and let $\mathbb{E}[X_i] = \mu$. Define $N_t = \sup\{n | T_n \leq t\}$ By the law of large numbers we have that $\frac{S_n}{n} \xrightarrow{a.s.} \mu$, and we claim that $\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{\mu}$.

Proof. $T_{N_t} \leq t \leq T_{N_t+1}$, so

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{T_{N_t+1}}{N_t+1}(\frac{N_t+1}{N_t})$$

Both the right hand side and left hand side go to μ as $t \to \infty$ by the law of large numbers so we are done. □

Triangular Arrays Page 51 in Durrett

1.6 Central Limit Theorem

1.6.1 Characteristic Functions

Definition 25 (Characteristic Function)**.** For a random variable X with distribution μ on $\mathbb{R}$ and $t \in \mathbb{R}$, the characteristic function $\varphi_X(t)$ is defined to be,

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itX} d\mu$$

The following are easily verifyable properties of characteristic functions:

1. $\varphi_X(0) = 1$
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$
3. $|\varphi_X(t)| = |\mathbb{E}[e^{itX}]| \leq \mathbb{E}[|e^{itX}|] = \mathbb{E}[1] = 1$
4. φ_X is uniformly continuous on $(-\infty, \infty)$:

$$|\varphi_X(t) - \varphi_X(t+h)| = |\mathbb{E}[e^{itX}(1 - e^{ihX})]| \leq \mathbb{E}[|1 - e^{ihX}|]$$

5. $\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\varphi_X(at)$
6. If $X \in \mathbb{Z}$ almost surely, $\varphi_X(t + 2\pi)\varphi_X(t)$:

$$\varphi_X(t + 2\pi) = \mathbb{E}[e^{i(t+2\pi)X}] = \mathbb{E}[e^{i2\pi X} e^{itX}] = \varphi_X(t)$$

Proposition 13. *If X_1 and X_2 are independent, then $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$*

Proof. We can check this directly,

$$\varphi_{X_1+X_2}(t) = \mathbb{E}[e^{it(X_1+X_2)}] = \mathbb{E}[e^{itX_1}e^{itX_2}] = \mathbb{E}[e^{itX_1}]\mathbb{E}[e^{itX_2}] = \varphi_{X_1}(t)\varphi_{X_2}(t)$$

□

Example 5 (Bernoulli($\frac{1}{2}$))**.** If $X \sim Bern(\frac{1}{2})$, then,

$$\varphi_X(t) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos(t)$$

For general p, $\varphi_X(t) = pe^{it} + (1-p)e^{-it}$.

Example 6 (Binomial($n, \frac{1}{2}$))**.** By proposition 13 and example 3, we have that for $X \sim Bin(n, \frac{1}{2})$,

$$\varphi_X(t) = \cos^n(t)$$

Example 7 (Poisson(λ))**.** If $N \sim Po(\lambda)$, then,

$$\varphi_N(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = e^{\lambda(e^{it}-1)}$$

Example 8 (Normal$(0,\sigma^2)$)**.** For $Z \sim \mathcal{N}(0,\sigma^2)$,

$$\varphi_Z(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2\sigma^2} dx = e^{-\sigma^2 t/2}$$

Example 9 (Uniform)**.** For $U \sim Unif[a,b]$,

$$\varphi_U(t) = \frac{1}{it(b-a)}(e^{itb} - e^{ita})$$

If $a = c$ and $b = -c$, then we have that,

$$\varphi_U(t) = \frac{\sin(ct)}{ct}$$

Example 10 (Triangle Density)**.** For X a random variable with density $f(x) = 1 - |x|$ on $[-1,1]$ (example: $U_1 + U_2$ for U_1, U_2 independent $\sim Unif[-1/2, 1/2]$),

$$\varphi_X(t) = \varphi_{U_1}(t) vp_{U_2}(t) = (\frac{\sin^2(t/2)}{t/2})^2 = \frac{2(1-\cos(t))}{t^2}$$

Example 11 (Exponential)**.** For $X \sim exp(\lambda)$,

$$\varphi_X(t) = \lambda \int_0^{\infty} e^{itx} e^{-\lambda x} dx = \frac{\lambda}{(it - \lambda)}$$

Now the question becomes, how do we reverse this process? That is, how can we recover a distribution from a characteristic function?

Theorem 23 (Inversion Formula for Characteristic Functions)**.** *Let $\varphi(t) = \int_{\mathbb{R}} e^{itx} \mu(dx)$ for μ a probability distribution. Then for $a < b$,*

$$\mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt$$

Proof. Define,

$$I_t = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{-T}^{T} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \mu(dx) dt$$

To show that this integral exists, we observe that,

$$\frac{e^{-ita} - e^{-itb}}{it} = \int_a^b e^{-ity} dy$$

Therefore, $|\frac{e^{-ita} - e^{-itb}}{it}| \le b - a$.

Since μ is a probability measure and $[-T,T]$ is a finite interval we are allowed to apply Fubini's theorem to obtain,

$$I_T = \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt \mu(dx)$$

$$= \int_{\mathbb{R}} (\int_{-T}^{T} \frac{\sin(t(x-a))}{t} dt - \int_{-T}^{T} \frac{\sin(t(x-b))}{t} dt)$$

Define,

$$R(\theta, T) = \int_{-T}^{T} \frac{\sin(t\theta)}{t} dt \qquad S(T) = \int_{0}^{T} \frac{\sin(x)}{x} dx$$

Then we have that

$$I_T = \int_{\mathbb{R}} (R(x-a,T) - R(x-b,T))\mu(dx)$$

Apply a change of variables letting $t = x/\theta$ to get,

$$R(\theta, T) = 2\int_{0}^{T\theta} \frac{\sin(x)}{x} dx = 2S(T\theta)$$

Since sin is an odd function, we have that $R(\theta, T) = sgn(\theta)R(|\theta|, T)$, so we get that $R(\theta, T) = 2sgn(\theta)S(T|\theta|)$ As $T \to \infty$, $S(T) \to \pi/2$ (See exercise 1.7.5 in Durrett). This implies that $\lim_{T\to\infty} R(\theta, T) = \pi sgn(\theta)$, therefore we get that,

$$\lim_{T\to\infty} R(x-a,T) - R(x-b,T) = \begin{cases} 2\pi & a < x < b \\ \pi & x = a \text{ or } x = b \\ 0 & \text{else} \end{cases}$$

We then use dominated convergence to bring the limit inside of the integral to get,

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} (R(x-a,T) - R(x-b,T))\mu(dx) = \frac{1}{2\pi}(\int_{a}^{b} 2\pi\mu(dx) + \pi(\mu(a) + \mu(b))$$

$$= \mu((a,b)) + \frac{1}{2}\mu(\{a,b\})$$

□

The following theorem about inverting point masses won't be proven here but follows a similar structure to the inversion theorem proof above:

Theorem 24. *If $\varphi(t) = \int_{\mathbb{R}} e^{itx}\mu(dx)$, then*

$$\mu(\{a\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita}\varphi(t)dt$$

Corollary 3. *If φ_X is real, then X and $-X$ have the same distribution.*

Corollary 4. *If $X_1 \sim \mathcal{N}(0, \sigma_1^2), X_2 \sim \mathcal{N}(0, \sigma_2^2)$ are independent, then $X_1 + X_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$*

Theorem 25. *If $\int |\varphi(t)|dt < \infty$, then μ has bounded continuous density,*

$$f(y) = \frac{1}{2\pi} \int e^{ity}\varphi(t)dt$$

Proof. By the bound on the integrand mentioned in the inversion theorem and the fact that our integral converges absolutely, we have that,

$$\mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) = \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt \leq \frac{(b-a)}{2\pi}\int_{-\infty}^{\infty} |\varphi(t)|dt \xrightarrow{a\to b} 0$$

which implies that μ has no point masses.

Next, we have that,

$$\mu((x,x+h)) = \frac{1}{2\pi}\int \frac{e^{-itx} - e^{-it(x+h)}}{it}\varphi(t)dt = \frac{1}{2\pi}\int (\int_x^{x+h} e^{-ity}dy)\varphi(t)dt$$

$$= \int_x^{x+h} (\frac{1}{2\pi}\int e^{-ity}\varphi(t)dt)dy$$

Therefore μ has density function,

$$f(y) = \frac{1}{2\pi}\int e^{-ity}\varphi(t)dt$$

□

Proposition 14. *Let X be a random variable with distribution μ. If $\int |x|^n \mu(dx) < \infty$, then,*

1. $\varphi_X^{(n)}(0) = i^n \mathbb{E}[X^n]$
2. $\varphi_X(t) = \sum_{m=0}^{n} \frac{1}{m!}\mathbb{E}[(itX)^m] + g_n(t)$ *for* $g_n(t) = o(t^n)$.
3. $|g_n(t)| \leq \mathbb{E}[\min(|tX|^{n+1}, 2|tX|^n)]$

Proof. 1 follows from the fact that $\varphi_X^{(n)}(t) = \int (ix)^n e^{itx}\mu(dx)$. 2 follows from taylor expanding φ_X.

why is 3 true? □

Theorem 26 (Levy-Cramer Continuity Theorem). *Let $\mu_n, n \geq 1$ be probability measures with characteristic functions φ_n. Then (i) if $\mu_n \Rightarrow \mu$, $\varphi(t) \to \varphi(t)$ for φ the characteristic function of μ. (ii) If $\varphi_n(t)$ converges pointwise to a limit $\varphi(t)$ that is continuous at 0, then the associated sequence of probability measures is tight and converges weakly to the measure μ with characteristic function φ.*

Proof. We will first prove (i). Recall from theorem 10 that $\mu_n \Rightarrow \mu$ if and only if for every bounded continuous function g, $\int g(x)\mu_n(dx) \to \int g(x)\mu(dx)$. We have that,

$$\varphi_n(t) = \int e^{itx}\mu_n(dx) = \int \cos(tx)\mu_n(dx) + i\int \sin(tx)\mu_n(dx)$$

$$\xrightarrow{n\to\infty} \int \cos(tx)\mu(dx) + i\int \sin(tx)\mu(dx) = \varphi(t)$$

Where the limit follows by applying theorem 10 with $g(x) = \cos(tx)$ and $g(x) = \sin(tx)$.

Next we prove (ii). First, we claim that the convergence of characteristic functions implies that μ_n is tight (see page 115 of Durrett for details on the proof).

Next, we apply Helly's selection theorem and theorem 12 to obtain a subsequence μ_{n_k} such that $\mu_{n_k} \Rightarrow \mu$ for μ a probability measure on $\mathbb{R}$. It is clear from (i) that this then implies μ has distribution function φ. This observation and tightness then implies that every sequence has a further subsequence that converges to μ. We now use theorem 10 to prove that this implies $\mu_n \Rightarrow \mu$. Suppose f is continuous and bounded. Then we know that every subsequence of $\{\int f(x)\mu_n(dx)\}_n$ has a further subsequence that converges to $\int f(x)\mu(dx)$, therefore $\int f(x)\mu_n(dx) \to \int f(x)\mu(dx)$, which concludes the proof. □

Higher dimensional characteristic functions and inversion formula/sufficient conditions to check weak convergence

1.6.2 The Central Limit Theorem and Variations

Theorem 27 (The Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Let $Z \sim \mathcal{N}(0,1)$, then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow Z$$

Proof. By centering, we can assume that $\mu = 0$. We will prove this theorem by proving that the characteristic functions $\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t)$ converge pointwise to the characteristic function of Z, $\varphi_Z(t) = e^{-t/2}$.

$$\begin{aligned}
\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) &= \mathbb{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] \\
&= (\mathbb{E}[e^{it\frac{X_i}{\sigma\sqrt{n}}}])^n && \text{By } \{X_k\}_k \text{ i.i.d} \\
&= (1 - \frac{1}{2}(\frac{t}{\sigma\sqrt{n}})^2\sigma^2 + o(\frac{t^2}{\sigma^2 n}))^n && \text{By Proposition 14} \\
&= (1 - \frac{t}{2n} + o(\frac{1}{n}))^n
\end{aligned}$$

We then apply the following lemma to the above formula (which we won't prove here):

Lemma 9. *If $c_j \to 0$, $a_j \to \infty$, and $c_j a_j \to \lambda$, then $(1 + c_j)^{a_j} \to e^\lambda$*

Taking $a_j = j$, and $c_j = -\frac{t}{2j} + o(\frac{1}{j})$, we have that $\lambda = -\frac{t}{2}$, so,

$$\lim_{n\to\infty} \varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \lim_{n\to\infty} (1 - \frac{t}{2n} + o(\frac{1}{n}))^n = e^{-\frac{t}{2}} = \varphi_Z(t)$$

□

COME BACK TO THIS SECTION AND SAY MORE (pg 58 in notes)
Weak convergence in general metric spaces and the Protmanteau theorem

Theorem 28 (Cramér-Wold Device)**.** *Let $\{X_n\}_n$ be a sequence of random vectors in $\mathbb{R}^d$ such that for all $\theta \in \mathbb{R}^d$, $X_n \cdot \theta \Rightarrow X \cdot \theta$, then $X_n \Rightarrow X$*

Proof. Let φ_n and φ be the characteristic functions of X_n and X respectively. Then for all $\theta \in \mathbb{R}^d$ we have,

$$\varphi_n(\theta) = \varphi_{\theta \cdot X_n}(1) \xrightarrow{n\to\infty} \varphi_{\theta \cdot X}(1) = \varphi(\theta)$$

□

Theorem 29 (Central Limit Theorem in $\mathbb{R}^d$). *Let $X_1, X_2, ...$ be i.i.d. random vectors in $\mathbb{R}^d$ with $\mathbb{E}[X_i] = \mu \in \mathbb{R}^d$ and covariance matrix $\Gamma_{i,j} = \mathbb{E}[(X_k^i - \mu_i)(X_k^j - \mu_j)]$. Let X be a multivariate Gaussian random vector in $\mathbb{R}^d$ with $\mathbb{E}[X] = 0$ and covariance matrix Γ, and define $S_n = X_1 + ... + X_n$. Then,*

$$\frac{S_n - \mu n}{\sqrt{n}} \Rightarrow X$$

We now will state a theorem that gives bounds on the rate of convergence of these sequences of random variables to the normal distribution.

Theorem 30. *Let $X_1, X_2, ...$ be i.i.d. random variables in $\mathbb{R}$ with $\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = \sigma^2$, and let F_n be the distribution function for $\frac{S_n}{\sigma\sqrt{n}}$. Denote $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$ the distribution function of $\mathcal{N}(0,1)$, then,*

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{3\mathbb{E}[|X_i|^3]}{\sigma^3\sqrt{n}}$$

Maybe prove this but probably not
Local Limit CLT
The following theorem gives us nice control on the magnitude of fluctuations of a random walk.

Theorem 31 (Law of the Iterated Logarithm). *Let $X_1, X_2, ...$ be i.i.d. random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$. Then,*

$$\limsup_{n\to\infty} |\frac{S_n}{\sqrt{n\log(n)}}| = \sqrt{2} \ a.s.$$

Proof. Proof □

1.7 Conditioning

1.7.1 Condition Expectation

Definition 26 (Conditional Expectation). Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[|X|] < \infty$, Σ a σ-subalgebra of $\mathcal{F}$. Then $\mathbb{E}[X|\Sigma]$ is any Σ-measurable random variable satisfying,

$$\int_A \mathbb{E}[X|\Sigma] d\mathbb{P} = \int_A X d\mathbb{P}$$

for all $A \in \Sigma$.

Remark 18. The existence and uniqueness up to a measure 0 set of this construction will be assumed for now and proven later in the section on Radon-Nikodym derivatives.

Proposition 15. *Conditional expectation satisfies the following properties,*

1. *If X is Σ-measurable, then $\mathbb{E}[X|\Sigma] = X$.*
2. $\mathbb{E}[\mathbb{E}[X|\Sigma]] = \mathbb{E}[X]$
3. *If $X \geq 0$, then $\mathbb{E}[X|\Sigma] \geq 0$ almost surely.*
4. $\mathbb{E}[a_1X_1 + a_2X_2|\Sigma] = a_i\mathbb{E}[X_i|\Sigma] + a_2\mathbb{E}[X_2|\Sigma]$
5. *If Z is Σ-measurable and bounded then $\mathbb{E}[ZX|\Sigma] = Z\mathbb{E}[X|\Sigma]$ almost surely.*
6. *If Z is a bounded Σ-measurable random variable, then $\mathbb{E}[Z\mathbb{E}[X|\Sigma]] = \mathbb{E}[ZX]$.*
7. *If $\Sigma' \subset \Sigma \subset \mathcal{F}$, then $\mathbb{E}[\mathbb{E}[X|\Sigma]|\Sigma'] = \mathbb{E}[X|\Sigma']$*
8. *Conditional Jensen's Inequality: If φ is a convex real valued function, then $\varphi(\mathbb{E}[X|\Sigma]) \leq \mathbb{E}[\varphi(X)|\Sigma]$*

Proof. 1 follows trivially from the definiton. For 2, we take $A = \Omega$ and see that from the definition,

$$\mathbb{E}[\mathbb{E}[X|\Sigma]] = \int_\Omega \mathbb{E}[X|\Sigma]d\mathbb{P} = \int_\Omega Xd\mathbb{P} = \mathbb{E}[X]$$

For 3, we let $A = \mathbb{E}[X|\Sigma]^{-1}(-\infty, 0)$. Notice that since $(-\infty, 0) \in \mathcal{B}$, $A \in \Sigma$. Therefore we know that,

$$0 \geq \int_A \mathbb{E}[X|\Sigma]d\mathbb{P} = \int_A Xd\mathbb{P} \geq 0$$

Thereofore $\int_A \mathbb{E}[X|\Sigma]d\mathbb{P} = 0$ and since $\mathbb{E}[X|\Sigma] < 0$ on A, this implies that $\mathbb{P}(A) = 0$.

To prove 4, first notice that the left hand side is clearly Σ-measurable, so it satisfies the first condition of the definition. To check the second condition, we let $A \in \Sigma$ and observe,

$$\int_A a_1X_1 + a_2X_2d\mathbb{P} = a_1\int_A X_1d\mathbb{P} + a_2\int X_2d\mathbb{P}$$

$$= a_1\int \mathbb{E}[X_1|\Sigma]d\mathbb{P} + a_2\int \mathbb{E}[X_2|\Sigma]d\mathbb{P} = \int a_1\mathbb{E}[X_1|\Sigma] + a_2\mathbb{E}[X_2|\Sigma]d\mathbb{P}$$

To prove 5, we use simple functions to approximate Z from below via the following formula. Let M be such that $|Z| \leq M$,

$$\varphi_n = \sum_{k=-nM}^{nM-1} \frac{k}{n}1_{Z^{-1}([\frac{k}{n}, \frac{k+1}{n}))}$$

Therefore we have that $\varphi_n(\omega) \leq Z(\omega) < \varphi(\omega) + \frac{1}{n}$ and furthermore, note that because Z is Σ-measurable, $Z^{-1}([\frac{k}{n}, \frac{k+1}{n})) \in \Sigma$. We know that $|\varphi_nX| \leq M|X|$, so for any $A \in \Sigma$, we can apply the dominated convergence theorem to see that,

$$\int_A Y\mathbb{E}[X|\Sigma]d\mathbb{P} = \lim_{n\to\infty}\int \varphi_n\mathbb{E}[X|\Sigma]d\mathbb{P} = \lim_{n\to\infty}\sum_{k=-nM}^{nM-1}\frac{k}{n}\int_{A\cap Z^{-1}[\frac{k}{n}, \frac{k_1}{n})]}\mathbb{E}[X|\Sigma]d\mathbb{P}$$

$$= \lim_{n\to\infty}\sum_{k=-nM}^{nM-1}\frac{k}{n}\int_{A\cap Z^{-1}[\frac{k}{n}, \frac{k_1}{n})]}Xd\mathbb{P} = \lim_{n\to\infty}\int \varphi_nXd\mathbb{P} = \int YXd\mathbb{P}$$

So $Y\mathbb{E}[X|\Sigma]$ is Σ measurable and satisfies $\int_A Y\mathbb{E}[X|\Sigma]d\mathbb{P} = \int_A YXd\mathbb{P}$ for all $A \in \Sigma$. 6 then follows from 5 and 2.

For 7, we observe that for $A \in \Sigma' \subset \Sigma$,

$$\int_A \mathbb{E}[\mathbb{E}[X|\Sigma]|\Sigma']d\mathbb{P} = \int_A \mathbb{E}[X|\Sigma]d\mathbb{P} = \int_A Xd\mathbb{P}$$

Lastly, to prove the conditional Jensen's inequality, notice that for φ convex, we can write,

$$\varphi(x) = \max_{\substack{h\leq\varphi \\ h \text{ linear}}} h(x)$$

By property 4 for any $h \leq \varphi$ we have

$$\mathbb{E}[\varphi(X)|\Sigma] \geq \mathbb{E}[h(X)|\Sigma] = h(\mathbb{E}[X|\Sigma])$$

which implies,

$$\mathbb{E}[\varphi(X)|\Sigma] \geq \max_{\substack{h\leq\varphi \\ h \text{ linear}}} h(\mathbb{E}[X|\Sigma]) = \varphi(\mathbb{E}[X|\Sigma])$$

□

The following examples should help to connect this definition of conditional expectation with our usual understanding. Intuitively, we want to think of Σ as information we know.

Example 12 (No Information)**.** Suppose that X is such that $\sigma(X)$ is independent from Σ. Then in this case, $\mathbb{E}[X|\Sigma] = \mathbb{E}[X]$. To see this, note that for any $A \in \Sigma$, X and 1_A are independent, therefore,

$$\int_A Xd\mathbb{P} = \mathbb{E}[X1_A] = \mathbb{E}[X]\mathbb{E}[1_A] = \int_A \mathbb{E}[X]d\mathbb{P}$$

Example 13 (Averaging)**.** Let $\Omega_1, \Omega_2, \ldots$ be a (finite or infinite) partition of Ω into disjoint sets of positive measure, and let $\Sigma = \sigma(\Omega_1, \Omega_2, \ldots)$. Then for $\omega \in \Omega_i$, we have,

$$\mathbb{E}[X|\Sigma] = \frac{\mathbb{E}[X1_{\Omega_i}]}{\mathbb{P}(\Omega_i)} = \frac{\int_{\Omega_i} Xd\mathbb{P}}{\mathbb{P}(\Omega_i)}$$

So in words, in this case $\mathbb{E}[X|\Sigma]$ is the random variable such that it's value on $\omega \in \Omega_i$ is the average of X over Ω_i.

Example 14 (Conditioning on a Random Variable)**.** For random variables X and Y we can obtain the usual definiton of $\mathbb{E}[X|Y]$ from our definition by taking $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

Example 15. Suppose that X and Y are independent, and φ is such that $\mathbb{E}[|\varphi(X,Y)|] < \infty$. Let $g(x) = \mathbb{E}[\varphi(x,Y)]$. Then,

$$\mathbb{E}[\varphi(X,Y)|X] = g(X)$$

Clearly $g(X)$ is $\sigma(X)$-measurable. To demonstrate this, we see that for any $A \in \sigma(X)$, $A = \{X \in C\}$ for some $C \in \mathcal{B}$. Then,

$$\int_A \varphi(X,Y)d\mathbb{P} = \int\int \varphi(x,y)1_C(x)\nu(dy)\mu(dx) = \int 1_C(x)g(x)\mu(dx) = \int_A g(X)d\mathbb{P}$$

1.7.2 Conditional Probability

Proposition 16. *Suppose that X and Y are random variables with joint density $f(x,y)$, and suppose for simplicity that $\mathbb{P}(Y=y) = \int f(x,y)dx > 0$ for all y. If $\mathbb{E}[|g(X)|] < \infty$, then, $\mathbb{E}[g(X)|Y] = h(Y)$ for,*

$$h(y) = \frac{\int g(x)f(x,y)dx}{\int f(x,y)dx}$$

To drop the condition that $\int f(x,y)dx > 0$ for all y, we simply define h by,

$$h(y)\int f(x,y)dx = \int g(x)f(x,y)dx$$

namely, $h(y)$ can take on any values when $\int f(x,y)dx = 0$.

Proof. $h(Y)$ is clearly $\sigma(Y)$-measurable. Therefore we just need to check that for $A \in \sigma(Y)$, $\int h(Y)d\mathbb{P} = \int g(X)d\mathbb{P}$. It is sufficient to check this for $A = Y^{-1}([a,b])$. Then we have,

$$\int_A h(Y)d\mathbb{P} = \int_a^b \int h(y)f(x,y)dxdy = \int_a^b h(y)\int f(x,y)dxdy$$

$$= \int_a^b [\frac{\int g(x)f(x,y)dx}{\int f(x,y)dx}]\int f(x,y)dxdy = \int_a^b \int g(x)f(x,y)dxdy = \int_A g(X)d\mathbb{P}$$

□

Example 16 (Conditional Probability)**.** In particular in the above proposition, if we take g to be an indicator function of X, where $\int f(x,y)dx > 0$ we can write,

$$\mathbb{E}[1_A(X)|Y](y) = \frac{\int 1_{X(A)}(x)f(x,y)dx}{\int f(x,y)dx} = \frac{\mathbb{P}(X\in A, Y=y)}{\mathbb{P}(Y=y)}$$

which is the familiar formula for conditional probability.

This example leads us to the following definition of conditional probability.

Definition 27 (Regular Conditional Distribution/Probability)**.** Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, $X : (\Omega,\mathcal{F}) \to (S,\mathcal{S})$ a measurable map, and $\mathcal{G} \subset \mathcal{F}$ a sub σ-algebra. $\mu : \Omega \times \mathcal{S} \to [0,1]$ is said to be a regular conditional distribution for X given $\mathcal{G}$ if,

1. For each $A \in \mathcal{S}$, $\mu(-,A)$ is a version of $\mathbb{P}(X\in A|\mathcal{G}) = \mathbb{E}[1_{X\in A}|\mathcal{G}]$.
2. For almost every $\omega \in \Omega$, $\mu(\omega,-)$ is a probability measure on $(S,\mathcal{S})$.

when $(S,\mathcal{S}) = (\Omega,\mathcal{F})$ and X is the identity map, this is called a regular conditional probability.

NEED MORE HERE

1.7.3 Radon-Nikodym Theorem

Definition 28 (Absolute Continuity)**.** Let μ and ν be measures on $(\Omega, \mathcal{F})$. We say ν is absolutely continuous with respect to μ (denoted $\nu << \mu$ Find the right TeX) if $\mu(E) = 0$ implies $\nu(E) = 0$ for all $E \in \mathcal{F}$.

Theorem 32 (Radon-Nikodym Theorem)**.** *Let μ, ν be 2 σ-finite measures on $(\Omega, \mathcal{F})$ such that $\nu << \mu$. Then there exists a function $f \geq 0$ such that,*

$$\int_E f d\mu = \nu(E)$$

This function f is called that Radon-Nikodym derivative denoted $\frac{d\nu}{d\mu}$ and is unique up to a μ-measure 0 set.

Before beginning the proof, we will show a couple of simple examples.

Example 17. Let $\mu = \frac{1}{2}(\delta_x + \delta_y)$ and let $\nu = \delta_x$. Clearly $\nu << \mu$. We can easily check that $f = 21_x$ satisfies the conditions of the Radon-Nikodym derivative,

$$\int_E f d\mu = 21_x(E)\frac{1}{2}(\delta_x(E) + \delta_y(E)) = \begin{cases} 1 & x \in E \\ 0 & \text{else} \end{cases}$$

Example 18. Let μ and ν be discrete distributions defined on a countable space (for simplicity, we'll say that they are defined on the space $\mathbb{N}$). Then we can write,

$$\mu = \sum_{k=0}^{\infty} p_k \delta_k \qquad \nu = \sum_{k=0}^{\infty} q_k \delta_k$$

Then $\nu << \mu$ if and only if $q_k = 0$ anywhere $p_k = 0$. Define $f = \sum_{k=0}^{\infty} \frac{q_k}{p_k} 1_k$ (where we just set terms for which $p_k = q_k = 0$ to 0). Then we have,

$$\int_E f d\mu = \sum_{k \in E} f(k) p_k = \sum_{k \in E} \frac{q_k}{p_k} p_k = \sum_{k \in E} q_k = \nu(E)$$

Example 19. Let μ be a measure that is absolutely continuous with respect to the Lesbegue measure (in this case we simply say that μ is absolutely continuous). Then there is an f such that,

$$\mu(E) = \int_E f(x) dx$$

In the case where μ is a probability distribution, then f is the probability densitiy function of μ. Examples of this include $\frac{1}{\sqrt{2\pi}} e^{x^2/2}$ for the normal distribution and others.

If $\nu << \mu$ are 2 absolutely continuous measures with denisty functions g and h respectively, then their Radon-Nikodym derivative is given by $\frac{d\nu}{d\mu} = \frac{g}{h}$ (where $h(x) \neq 0$, we may take it to be 0 otherwise). To check this, we see,

$$\nu(E) = \int_E g d\nu = \int_E g(x) dx = \int_E \frac{g(x)}{h(x)} h(x) dx = \int_E \frac{g}{h} d\mu$$

Now we need to introduce a few definitions before moving onto the proof of the thoerem.

Definition 29 (Singular Measures)**.** Measures μ_1 and μ_2 are said to be mutually singular (or μ_1 is singular with respect to μ_2) if there is a set A such that $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$. This is notated $\mu_1 \perp \mu_2$.

Definition 30 (Signed Measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\alpha : \mathcal{F} \to (-\infty, \infty]$ is a signed measure if,

1. $\alpha(\phi) = 0$
2. If $E = \bigcup_i E_i$ where the E_i are disjoint, then $\alpha(E) = \sum_i \alpha(E_i)$ in the following sense: If $\alpha(E) < \infty$, then the sum converges absolutely, and if $\alpha(E) = \infty$, then $\sum_i \alpha^-(E_i) < \infty$ and $\sum_i \alpha(E_i) = \infty$.

Definition 31 (Positive/Negative Set)**.** Let $(\Omega, \mathcal{F})$ be a measurable space and let α be a signed measure. A set $A \in \mathcal{F}$ is positive if every measurable $B \subset A$ has $\mu(B) \geq 0$. Likewise a set A is negative if every measurable $B \subset A$ has $\alpha(B) \leq 0$.

Next we will prove 3 decomposition theorems for signed measures.

Theorem 33 (Hahn Decomposition)**.** *Let α be a signed measure. Then there is a positive set A and a negative set B such that $\Omega = A \cup B$ and $\emptyset = A \cap B$. Furthermore, this decomposition is unique up to a measure 0 set.*

Proof. We first introduce the following 2 lemmas,

Lemma 10. *If the sets A_n are positive, then $A = \bigcup_n A_n$ are positive, and likewise for negative sets.*

Proof. Let $B \subset A$. Then we can write we can write $B = \cup_n B_n$ where $B_n = A_n \cap (\bigcup_{m=1}^{n-1} A_m^c) \subset A_n$. These sets are all disjoint and, as they are subsets of a positive set, we have that,

$$\alpha(B) = \sum_n \alpha(B_n) \geq 0$$

□

Lemma 11. *Let E be a measurable set with $\alpha(E) < 0$ Then there is a negative set $F \subset E$ with $\alpha(F) < 0$ (and likewise for positive sets).*

Proof. If E is negative, this is true trivially, so assume E is not negative. Let n_1 be the smallest possible integer such that there exists a $E_1 \subset E$ such that $\alpha(E_1) > 1/n_1$. We now inductively construct the set F by setting $F_1 = E - E_1$ if $E - E_1$ is not negative, we repeat the procedure with n_2 and $E_2 \subset F_1$. If the construction does not stop in finite steps, we define $F = \bigcap_k F_k = E - (\bigcup_k E_k)$.

First note that because $0 > \alpha(E) > -\infty$ and $\alpha(E_k) \geq 0$ for all k, we have that $\alpha(F) < 0$. Furthermore, if there were some $F' \subset F$ such that $alpha(F') > 0$, then we could write $\alpha(F') \geq 1/N$ for some N, which is a contradiction by construction of F. The proof for positive sets is equivalent. □

Now, define $c = \inf\{\alpha(B) : B \text{ is negative}\}$. By defintion of a signed measure, $0 \geq c > -\infty$. Choose B_i such that $\alpha(B_1) \searrow c$ and define $B = \bigcup_i B_i$. B is negative by lemma 10, therefore $\alpha(B) \geq c$. additionally, we have that,

$$\alpha(B) = \alpha(B_i) + \alpha(B - B_i) \leq \alpha(B_i)$$

This inequality holds for all B_i and since $\alpha(B_i) \searrow c$, we have that $\alpha(B) \leq c$, thus $\alpha(B) = c$. Set $A = B^c$. To check that A is positive, assume there exists some $A' \subset A$ such that $\alpha(A') < 0$. Then $\alpha(B \cup A') = \alpha(B) + \alpha(A') < c$, which contradicts the defintion of c. Therefore A is a positive set.

To prove uniqueness, let $A \cup B$ and $C \cup D$ be 2 different Hahn decompositions, we want to prove that A and C differ only by a set of measure 0 (which then clearly implies the same for B and D). Let $A' = A - (A \cap C)$ and $C' = C - (A \cap C)$. Clearly both A' and C' are positive and we can write $B = (B \cap D) \cup C'$. This implies that $C' \subset B$. Since B is negative and C' is positive, this tells us that $\alpha(C') = 0$, and by symmetry, $\alpha(A') = 0$ as well. □

Theorem 34 (Jordan Decomposition Theorem)**.** *If α is a signed measure, then $\alpha = \mu_1 - \mu_2$ for μ_1, μ_2 measures with $\mu_2(\Omega) < \infty$. Moreover, this pair is unique.*

Proof. Let $\Omega = A \cup B$ be a Hahn decomposition, and defined,

$$\alpha_+(E) = \alpha(E \cap A) \qquad \alpha_-(E) = -\alpha(E \cap B)$$

Since A is positive and B is negative, α_+ and α_- are both measures. $\alpha_+(A^c) = 0$ and $\alpha_-(A) = 0$ so they are mutually singular.

To prove uniqueness, suppose that $\alpha = \nu_1 - \nu_2$ and D is a set with $\nu_1(D) = 0$ and $\nu_2(D^c) = 0$. Then if we set $C = D^c$, $C \cup D$ is a Hahn decomposition, since for any $E \subset D$, $\alpha(E) = \nu_1(E) - \nu_2(E) = \nu_1(E) \geq 0$ (and similarly we can show that D^c is negative) and it follows that $\nu_1(E) = \alpha(E \cap C)$ and $\nu_2(E) = -\alpha(E \cap D)$. By the uniqueness of the Hahn decomposition, it follows that $\mu_1 = \nu_1$ and $\mu_2 = \nu_2$. □

Theorem 35 (Lebesgue Decomposition Theorem)**.** *Let μ, ν be σ-finite measures. Then ν can be written as $\nu_r + \nu_s$ where ν_s is singular with repsect to μ and,*

$$\nu_r(E) = \int_E g d\mu$$

Proof. First, by decomposing Ω into $\bigsqcup_i \Omega_i$ such that $\mu(\Omega_i), \nu(\Omega_i) < \infty$ for all i, we may assume that μ and ν are finite.

Define $\mathcal{G} = \{g \geq 0 : \int_E g d\mu \leq \nu(E) \text{ for all } E \subset \Omega\}$. First we want to show that if $g, h \in \mathcal{G}$, then $\max(g, h) \in \mathcal{G}$. Let $A = \{g > h\}$ and $B = \{g \leq h\}$. Then,

$$\int_E \max(g, h) d\mu = \int_{E \cap A} g d\mu + \int_{E \cap B} h d\mu \leq \nu(E \cap A) + \nu(E \cap B) = \nu(E)$$

Let $\kappa = \sup\{\int g d\mu : g \in G\}$. For each n, choose g_n such that $\int g_n d\mu > \kappa - 1/n$ and set $h_n = \max(g_1, ..., g_n)$. The sequence h_n is monotone. Let h be such that $h_n \nearrow h$. Then by the monotone convergence theorem,

$$\kappa \geq \int h d\mu \geq \lim_{n \to \infty} \int g_n d\mu = \kappa$$

Set $\nu_r(E) = \int_E h d\mu$ and $\nu_s(E) = \nu(E) - \nu_r(E)$. All that remains to be shown is that ν_s is singular with respect to μ.

For all n, let $A_n \cup B_n$ be a Hahn decompositon for $\nu_s - \frac{1}{n}\mu$. Since A_n is positive on $\nu_s - \frac{1}{n}\mu$, we know that for all E, $\nu_s(E \cap A_n) \geq \frac{1}{n}\mu(E \cap A_n)$, therefore,

$$\int_E h + \frac{1}{n} 1_{A_n} d\mu = \nu_r(E) + \frac{1}{n}\mu(E \cap A_n) \leq \nu(E)$$

which implies that $h + \frac{1}{n} 1_{A_n} \in \mathcal{G}$. Since h is such that $\int h d\mu = \kappa$, we have that $\int 1_{A_n} d\mu = \mu(A_n) = 0$, otherwise $\int h + \frac{1}{n} 1_{A_n} d\mu > \kappa$, thus contradicting the definition of κ. Set $A = \bigcup_n A_n$. All that remains to show is that $\nu_s(A^c) = 0$. If $\nu_s(A^c) > 0$, then this implies that $(\nu_s - \frac{1}{n}\mu)(A_n) > 0$ for large n, however this is a contradiction since $A^c \subset A_n$ which is a negative set with respect to $\nu_s - \frac{1}{n}\mu$. □

Proof of the Radon-Nikodym Theorem. Let $\nu = \nu_r + \nu_s$ be a Lebesgue decomposition, and let A be chosen such that $\nu_s(A^c) = 0$ and $\mu(A) = 0$. Since $\nu << \mu$, $0 = \nu(A) \geq \nu_s(A)$, which then implies that $\nu_s = 0$, so $\nu(E) = \nu_r(E) = \int_E f d\mu$ for some $f \geq 0$.

Next we prove uniqueness up to a μ-measure 0 set. Suppose that $\int_E g d\mu = \int_E h d\mu$. Then if we let $E_n = \{g > h, g \leq n\}$, we conclude that E_n must have measure 0. Since the choice of n is arbitrary, $\mu(\{g > h\}) = 0$ and similarly we can show that $\mu(\{g < h\}) = 0$. □

Now we are able to give a proof of the existence of conditional expectations.

Proposition 17. *Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}' \subset \mathcal{F}$ be a sub σ-algebra. Then there exists a random variable W such that W is $\mathcal{F}'$-measurable, and for all $A \in \mathcal{F}'$,*

$$\int_A W d\mathbb{P} = \int_A X d\mathbb{P}$$

Proof. First suppose that $X \geq 0$ and for $A \in \mathcal{F}'$ define,

$$\mathbb{P}'(A) = \int_A X d\mathbb{P}$$

$\mathbb{P}'$ is a measure on $(\Omega, \mathcal{F}')$ (one can check the countably additive condition using the dominated convergence theorem, other conditions are trivial), and clearly we have $\mathbb{P}' << \mathbb{P}$ (here we are considering the restriction of $\mathbb{P}$ to $\mathcal{F}$). Let W be the Radon-Nikodym derivative $\frac{d\mathbb{P}'}{d\mathbb{P}}$. Then we have that W is $\mathcal{F}'$ measurable and,

$$\int_A X d\mathbb{P} = \nu(A) = \int W d\mathbb{P}$$

For the general case, we write $X = X^+ - X^-$ and take $\mathbb{E}[X|\mathcal{F}'] = \mathbb{E}[X^+|\mathcal{F}'] - \mathbb{E}[X^-|\mathcal{F}']$. One can easily check that this satisfies the properties of conditional expectation for X. □

1.8 Markov Chains

1.8.1 Definition and Examples

Definition 32 (Transition Probability)**.** Let $(S, \mathcal{S})$ be a measurable space. A function $\pi : S \times \mathcal{S} \to \mathbb{R}$ is said to be a transition probability if,

1. For all $s \in S$, $A \mapsto \pi(s, A)$ is a probability measure.
2. For all $A \in \mathcal{S}$, $s \mapsto \pi(s, A)$ is a measurable function

Definition 33 (Markov Chain)**.** A sequence of random varaibles $X_0, X_1, ...$ on $(S, \mathcal{S})$ with corresponding filtration $\mathcal{F}_k = \sigma(X_0, ..., X_k)$ is a Markov chain with transition probabilities $\pi_{k,k+1}$ if,

$$\mathbb{P}(X_{k+1} \in B|\mathcal{F}_k) = \mathbb{P}(X_{k+1} \in B|X_0, ..., X_k) = \pi_{k,k+1}(X_k, B)$$

(i.e. the probability distribution of X_{k+1} depends only on X_k). If $\pi_{k,k+1}$ does not depend on k, then the Markov process is time-homogenous and we denote the transition probability π.

Remark 19. Note that we could generalize this definition for a general filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset ... \subset \mathcal{F}$.

Given a sequence of transition probabilities and an initial distribution μ on our state space $(S, \mathcal{S})$ we can define a consistent set of finite dimensional distributions on $X_0, ..., X_n$ for all n by,

$$\mathbb{P}_\mu(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0} ... \int_{B_n} \pi_{n-1,n}(x_{n-1}, dx_n)...\pi_{0,1}(x_0, dx_1)\mu(dx_0)$$

If our space $(S, \mathcal{S})$ is sufficiently nice, then we can apply Kolmogorov's extension theorem to construct a probability measure $\mathbb{P}_\mu$ on the sequence space $(\Omega, \mathcal{F}) = (S^{\{0,1,...\}}, \mathcal{S}^{\{0,1,...\}})$ so that for any sequence element $\omega \in \Omega$, the coordinate projection $X_n(\omega) = omega_n \in S$ has the desired distribution. This is useful as it allows us to use shift operators, $\theta_n(\omega_0, \omega_1, ...) = (\omega_n, \omega_{n+1}, ...)$.

This construction gives us a large family of probability distributions for this sequence, namely a different distribution for each initial distribution μ, however we can note that to specify this family we only need a distribution $\mathbb{P}_x$ for each state $x \in S$ (or more specifically, for the distribution μ_x that assigns probability 1 to x and 0 else) since we have,

$$\mathbb{P}_\mu(A) = \int \mathbb{P}_x(A)\mu(dx)$$

To provide some intuition for what this is saying, in the discrete case we have that,

$$\mathbb{P}_\mu(\omega) = \sum_{x \in S} \mathbb{P}_x(\omega)\mathbb{P}_\mu(X_0 = x)$$

Theorem 36. *In the above construction, the coordinate functions $X_0, X_1, ...$ form a markov chain with respect to the filtration $\mathcal{F}_n = \sigma(X_0, ..., X_n)$ for all μ with transition probabilities $\pi_{k,k+1}$.*

Proof. We need to show,

$$\mathbb{P}_\mu(X_{n+1} \in B|\mathcal{F}_n) = \pi_{n,n+1}(X_n, B)$$

Recall from our definition of conditional probability that,

$$\mathbb{P}_\mu(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{E}_\mu[1_{X_{n+1} \in B}|\mathcal{F}_n]$$

Therefore, we need to check that for all $A \in \sigma(X_0, ..., X_n)$,

$$\int_A 1_{X_{n+1} \in B} d\mathbb{P}_\mu = \int_A \pi_{n,n+1}(X_n, B) d\mathbb{P}_\mu$$

Let $A = \{X_0 \in B_0, ..., X_n \in B_n\}$ (since this is a π system, proving the equality for A of this type will imply the equality for all $A \in \mathcal{F}_n$). Then,

$$\int_A 1_{X_{n+1} \in B} d\mathbb{P}_\mu = \mathbb{P}_\mu(X_j \in B_j, 0 \leq j \leq n, X_{n+1} \in B)$$

$$= \int_{B_0} ... \int_{B_n} \int_B \pi_{n,n+1}(x_n, dx_{n+1})...\pi_{0,1}(x_0, dx_1)\mu(dx_0)$$

$$= \int_{B_0} ... \int_{B_n} \pi_{n,n+1}(x_n, B)\pi_{n-1,n}(x_{n-1}, dx_n)...\pi_{0,1}(x_0, dx_1)\mu(dx_0) = \int_A \pi_{n,n+1}(X_n, B)d\mathbb{P}_\mu$$

□

For the rest of this section we will consider primarily time-homogenous Markov processes.

Example 20 (Random Walk on $\mathbb{Z}^n$)**.** Let $\xi_1, \xi_2, ...$ be i.i.d. random vectors in $\mathbb{Z}^n$. Then,

$$S_n = X_0 + \sum_{i=1}^n \xi_i$$

is a time homogenous Markov Chain with transition probability,

$$\mathbb{P}(S_{n+1} = k | S_n = j) = \mathbb{P}(X_{n+1} = k - j)$$

One commonly studied example is the simple symmetric random walk where ξ_i is ± 1 with probability $\frac{1}{2}$ each.

Example 21 (Random Walk on a Graph)**.** Let G be a directed graph with node set $N(G)$ and edge set $E(G)$. Let w be a weighting of the edges $E(G)$ such that $w \geq 0$ and for all $n \in (G)$, $\sum_{(n,n') \in E(G)} w((n, n')) = 1$ (for simplicity of notation, we state $w(n, n') = 0$ if $(n, n') \notin E(G)$). Then we can define a simple random walk on G with transition probabilities,

$$\mathbb{P}(X_{k+1} = n' | X_k = n) = w((n, n'))$$

Example 22 (Ehrenfest Chain)**.** The Ehrenfest chain models the motion of air particles in 2 chambers separated by a small opening. Let r be the number of particles in total and X_n the number in one chamber. At each time step we uniformly choose a particle from either chamber and move it to the other. The transition probabilities for the Markov chain are then given by,

$$\pi(k, k+1) = \frac{(r-k)}{r}$$
$$\pi(k, k-1) = \frac{k}{r}$$
$$\pi(i, j) = 0 \text{ else}$$

Example 23 (Branching Process)**.** Let $\{\xi_{i,j}\}_{i,j \geq 0}$ be i.i.d. nonnegative integer valued random variables. We inductively define a markov process by setting $X_0 = k > 0$ and,

$$X_{n+1} = \sum_{i=1}^n \xi_{n,i}$$

The transition probabilities are,

$$\pi(i,j) = \mathbb{P}(\sum_{k=1}^{i} \xi_k) = j$$

This process can be thought of as a simple model for asexual population growth, where $\xi_{n,i}$ is the number of offspring the i-th individual in generation n produces.

Example 24 (Wright-Fisher Model)**.** This Markov process is defined inductively by setting X_{n+1} to be a $Bin(N, \frac{X_n}{N})$ random variable. That is, the transition probabilities are given by,

$$\pi(i,j) = \binom{N}{j} (\frac{i}{N})^j (1 - \frac{i}{N})^{N-j}$$

This model can be useful for studying the evolution of population genetics.

Definition 34 (Absorbing State)**.** An absorbing state of a markov chain $X_0, X_1, ...$ is a state x such that $\pi(x,x) = 1$.

Example 21 and example 22 both contain absorbing states. In example 20 the absorbing state is 0 and in example 21 both 0 and N are absorbing states.

Example 25 (Queuing model)**.** Let $\xi_1, \xi_2, ...$ be i.i.d. nonnegative integer valued random variables (think of these as the number of customers arriving at time i). Then we define,

$$X_{k+1} = \begin{cases} X_k + \xi_{k+1} - 1 & X_k \neq 0 \\ \xi_{k+1} & \text{else} \end{cases}$$

which models the number of people in line at time k (assuming a single person is served in each time step if there is a person to be served).

Definition 35 (Transition Matrix)**.** For a markov model on a finite statespace $S = \{1, ..., N\}$ we can define the transition matrix,

$$(\pi)_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$$

Therefore if μ is a distribution vector for X_n, then we obtain a distribution vector for X_{n+1} by taking $\mu^T \pi$.

1.8.2 Stopping Times

Definition 36. A stopping time τ is a nonnegative integer valued random variable on $(\Omega, \mathcal{F}, \mathbb{P}) = (S^{\{0,1,...\}}, \mathcal{S}^{\{0,1,...\}}, \mathbb{P})$ (which may take on the value ∞) such that for all n, $\tau^{-1}(n) \in \sigma(X_0, ..., X_n)$.

Stopping times should be thought of as the time at which an event happens in your Markov chain. The measurability condition requires that if this event happens at time step n, we know it by time step n without any information about the future. The following examples and non-examples illustrate this.

Example 26. The following are examples of stopping times on a Markov chain:

1. $\tau =$ the first n such that $X_n = s$ for $s \in S$
2. $\tau = 2$ steps after the first n such that $X_n = s$

The following are *not* examples of stopping times:

1. $\tau =$ The last time we reach a state s
2. $\tau =$ The time step right before the first time we reach a state s

Proposition 18. *If τ_1, τ_2 are stopping times, then the following are stopping times as well,*

1. $\tau_1 + \tau_2$
2. $\min(\tau_1, \tau_2)$
3. $\max(\tau_1, \tau_2)$

However $\tau_1\tau_2$ and $\tau_1 - \tau_2$ are not necessarily stopping times.

Example 27. For every state $x \in S$ of our Markov chain, the random variable $\tau_x =$ the first visit (after time 0) to state x is a stopping time.

1.8.3 Markov Properties

Theorem 37 (Monotone Class Theorem)**.** *Let $\mathcal{A}$ be a π-system that contains Ω and let $\mathcal{H}$ be a collection of real-valued functions that satisfies,*

1. *If $A \in \mathcal{A}$, then $1_A \in \mathcal{H}$*
2. *If $f, g \in \mathcal{H}$ then $f + g \in \mathcal{H}$ and $cf \in \mathcal{H}$.*
3. *If $f_n \in \mathcal{H}$ are nonnegative and increase to a bounded function f, then $f \in \mathcal{H}$*

Then $\mathcal{H}$ contains all bounded functions that are measurable with respect to $\sigma(\mathcal{A})$.

Proof. By $\Omega \in \mathcal{A}$, and conditions 2 and 3, $\mathcal{G} = \{1_A | 1_A \in \mathcal{H}\}$ is a $\pi - \lambda$ system. 1 and 2 implies that $\mathcal{H}$ contains all simple functions and 3 implies that $\mathcal{H}$ contains all bounded measurable functions. □

Theorem 38 (Markov Property)**.** *Let $X_0, X_1, ...$ be a time homogenous Markov chain on state space $(S, \mathcal{S})$ and let $(\Omega, \mathcal{F}) = (S^{\{0,1,...\}}, \mathcal{S}^{\{0,1,...\}})$. Let $Y : \Omega \to \mathbb{R}$ be bounded and measurable. Then,*

$$\mathbb{E}_\mu[Y \circ \theta_m | \mathcal{F}_m] = \mathbb{E}_{X_m}[Y]$$

where $\mathbb{E}_\mu$ is expectation with respect to $\mathbb{P}_\mu$ and $\mathbb{E}_{X_m}[Y]$ should be thought of as the function $\varphi(x) = \mathbb{E}_x[Y]$ (i.e. expectation taken with respect to $\mathbb{P}_x$) evaluated at X_m. In particular, if we take Y to be a product of indicator random variables of the form $\prod_{i=1}^n 1_{X_i \in A_i}$ then,

$$\mathbb{P}_\mu(X_{m+1} \in A_1, ..., X_{m+n} \in A_n | \mathcal{F}_m) = \mathbb{P}_{X_m}(X_1 \in A_1, ..., X_n \in A_n)$$

In words, this means that the probability of things that happen after X_m depends only on X_m (i.e. the process is *memoryless*).

Proof. We will prove this theorem first for the class of functions $Y(\omega) = \prod_{0 \le k \le n} g_k(\omega_k)$ where g_k is bounded an measurable. We start by checking the conditional expectation property against $A = \{\omega : \omega_0 \in A_0, ..., \omega_m \in A_m\} \in \mathcal{F}_m$. We have,

$$\int_A \prod_{k=0}^{n} g_k(X_{m+k}) d\mathbb{P}_\mu$$

$$= \int_{A_0} \mu(dx_0) \int_{A_1} \pi(x_0, dx_1)... \int_{A_m} \pi(x_{m-1}, dx_m) g_0(x_m) \int \pi(x_m, dx_{m+1}) g_1(x_{m+1})...$$

$$\int \pi(x_{m+n-1}, dx_{m+n}) g_n(x_{m+n})$$

$$= \int_A \mathbb{E}_{X_m}[\prod_{k=0}^{n} g_k(X_k)] d\mathbb{P}_\mu$$

Now since A of this type form a π-system and sets satisfying this property form a λ system, the property is true for all $A \in \mathcal{F}_m$ and therefore we have that $\mathbb{E}[\prod_{k=0}^{n} g_k \circ \theta_m | \mathcal{F}_m] = \mathbb{E}_{X_m}[\prod g_k]$.

Now, let $\mathcal{H}$ be the class of functions Y such that $\mathbb{E}[Y \circ \theta_m | \mathcal{F}_m] = \mathbb{E}_{X_m}[Y]$. From the proof above we see that $\mathcal{H}$ contains functions of the form $\prod_{k=0}^{n} g_k$. Let $\mathcal{A}$ be the collection of sets of the form $\{\omega : \omega_0 \in A_0, ..., \omega_k \in A_k\}$. Taking $g_k = 1_{A_k}$ we see that $\mathcal{H}$ contains all 1_A for $A \in \mathcal{A}$. Furthermore, by the linearity property of conditional expectation and monotone convergence, we see that $\mathcal{H}$ satisfies all conditions of the monotone class theorem, and therefore contains all bounded measurable functions with respect to $\sigma(\mathcal{A}) = \mathcal{F}$. □

Theorem 39 (Strong Markov Property)**.** *Let τ be a stopping time and define $\mathcal{F}_\tau = \{A \in \mathcal{F} | A \cap \{\tau = n\} \in \mathcal{F}_n$ for all $n\}$. On $\tau < \infty$ we can define the shift operator $\theta_\tau(\omega_0, \omega_1, ...) = (\omega_\tau, \omega_{\tau+1}, ...)$. Suppose that for each n, $Y_n : \omega \to \mathbb{R}$ is a bounded measurable function. Then on $\tau < \infty$,*

$$\mathbb{E}_\mu[Y_\tau \circ \theta_\tau | \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[Y_\tau]$$

Similar to the normal Markov property, this in particular tells us that,

$$\mathbb{P}_\mu(X_{\tau+1} \in A_1, ..., X_{\tau_n \in A_n} | \mathcal{F}_\tau) = \mathbb{P}_{X_\tau}(X_1 \in A_1, ..., X_n \in A_n)$$

In words, this states that the probability of things that happen after X_τ for τ a stopping time depends only on X_τ.

Proof. Let $A \in \mathcal{F}_\tau$. Then we have,

$$\int_A Y_\tau \circ \theta_\tau d\mathbb{P}_\mu = \sum_{m=0}^{\infty} \int_{A \cap \{\tau = n\}} Y_\tau \circ \theta_\tau d\mathbb{P}_\mu = \sum_{m=0}^{\infty} \int_{A \cap \{\tau = m\}} Y_m \circ \theta_m d\mathbb{P}_\mu$$

By the Markov property and the fact that $A \cap \{\tau = m\} \in \mathcal{F}_m$,

$$\sum_{m=0}^{\infty} \int_{A \cap \{\tau = m\}} Y_m \circ \theta_n d\mathbb{P}_\mu = \sum_{m=0}^{\infty} \int_{A \cap \{\tau = m\}} \mathbb{E}_{X_m}[Y_m] d\mathbb{P}_\mu = \int_A \mathbb{E}_{X_\tau}[Y_\tau] d\mathbb{P}_\mu$$

Therefore $\mathbb{E}_\mu[Y_\tau \circ \theta_\tau | \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[Y_\tau]$. □

1.8.4 Categorization of States

Definition 37 (Transient/Recurrent State)**.** A state $x \in S$ of a Markov chain $X_0, X_1, ...$ with transition probability π is transcient if $\mathbb{P}_x(\tau_x < \infty) < 1$. A state x is recurrent if $\mathbb{P}_x(\tau_x < \infty) = 1$.

Example 28. Consider a random walk on the following graph,

$$\begin{tikzcd} A \arrow[r, "1"] & B \arrow[r, bend left, "1"] & C \arrow[l, bend left, ".5"] \arrow[loop right, ".5"] \end{tikzcd}$$

The state A is transcient since $\mathbb{P}_A(\tau_A < \infty) = 0$. The state C is recurrent because, starting at C we can either loop back to C or move to B and then back to C with probability 1, therefore, $\mathbb{P}_C(\tau_C \le 2) = 1$. B is slightly trickier since it is theoretically possible that we could continue to loop at C, however,

$$\mathbb{P}_B(\tau_B = B) = \mathbb{P}_C(X_k = C, k \ge 0) = \lim_{k \to \infty} (\frac{1}{2})^k = 0$$

Definition 38 (Positive/Null Recurrent State)**.** A state x is positive recurrent if it is recurrent and $\mathbb{E}_x[\tau_x] < \infty$. x is null recurrent if it is recurrent and $\mathbb{E}[\tau_x] = \infty$.

Definition 39 (Irreducible Markov Chain)**.** A Markov chain is irreducible if for all $x, y \in S$, There exists some $1 \le n < \infty$ such that $\pi^{(n)}(x, y) > 0$ (i.e. the transition probability of going from x to y in n steps). Equivalently, if we let τ_y be the stopping time of the first visit after 0 to y, $\mathbb{P}_x(\tau_y < \infty) > 0$.

Proposition 19. *Let $X_0, X_1, ...$ be an irreducible Markov chain. Then all states are of the same type (i.e. transient or positive/null recurrent).*

Proof. First we begin by showing that if $x \in S$ is recurrent, then all other states $y \in S$ are recurrent. By the irreducible property we have that there exists a k such that $\pi^{(k)}(x, y) > 0$. This implies that $\mathbb{P}_x(\tau_y < \infty) > 0$. Let m be such that $\pi^{(m)}(y, x) > 0$, then,

$$\mathbb{P}_y(\tau_y < \infty) \ge \pi^{(m)}(y, x)\mathbb{P}_x(\tau_y < \infty) > 0$$

Now we prove that if $x \in S$ is positive recurrent, then all other states $y \in S$ are positive recurrent as well. First note that since there exists a k such that $\pi^{(k)}(x, y) > 0$, then $\mathbb{P}_x(\tau_y < \tau_x) > 0$. We have,

$$\infty > \mathbb{E}_x[\tau_x] = \mathbb{E}_x[\tau_x(1_{\tau_y < \tau_x} + 1_{\tau_x < \tau_y})] \ge \mathbb{E}_x[\tau_x 1_{\tau_y < \tau_x}]$$

Now we let $\tau_{y \to x}$ be the number of steps between the first visit to y and the first visit to x after visiting y. On the event space $\tau_y < \tau_x$ we can write $\tau_x = \tau_y + \tau_{y \to x}$. We use this substitution to get,

$$\mathbb{E}_x[1_{\tau_y < \tau_x}\tau_x] = \mathbb{E}_x[1_{\tau_y < \tau_x}(\tau_x + \tau_{y \to x})] \ge \mathbb{E}_x[1_{\tau_y < \tau_x}\tau_{y \to x}]$$

Now we note that by the memoryless property, $1_{\tau_x < \tau_y}$ and $\tau_{y \to x}$ are independent. This is because the distribution of $\tau_{y \to x}$ is the same whether or not x is visited before y. Therefore we can apply the independence property and the Markov property to get,

$$\mathbb{E}_x[1_{\tau_y < \tau_x}\tau_{y \to x}] = \mathbb{E}_x[1_{\tau_y < \tau_x}]\mathbb{E}_x[\tau_{y \to x}] = \mathbb{P}_x(\tau_y < \tau_x)\mathbb{E}_y[\tau_x]$$

Letting $p = \mathbb{P}_x(\tau_y < \tau_x)$, we have that,

$$\mathbb{E}_y[\tau_x] \leq \frac{1}{p}\mathbb{E}_x[\tau_x] < \infty$$

Next we want to prove a similar inequality for $\mathbb{E}_x[\tau_y]$.

$$\mathbb{E}_x[\tau_y] = \mathbb{E}_x[\tau_y(1_{\tau_y<\tau_x} + 1_{\tau_x<\tau_y})] \leq \mathbb{E}_x[\tau_y 1_{\tau_y<\tau_x}] + \mathbb{E}_x[(\tau_x + \mathbb{E}_x[\tau_y])1_{\tau_x<\tau_y}]$$
$$= \mathbb{E}_x[\tau_x] + \mathbb{E}_x[\tau_y]\mathbb{P}(\tau_x < \tau_y) = \mathbb{E}_x[\tau_x] + (1-p)\mathbb{E}_x[\tau_y]$$

Therefore we have that, $\mathbb{E}_x[\tau_y] \leq \frac{1}{p}\mathbb{E}_x[\tau_x]$. Combining these bounds we get,

$$\mathbb{E}_y[\tau_y] \leq \mathbb{E}_y[\tau_x] + \mathbb{E}_x[\tau_y] \leq \frac{2}{p}\mathbb{E}_x[\tau_x] < \infty$$

□

Theorem 40. *Let $X_0, X_1, \ldots$ be a irreducible Markov chain on a countable state space S. Then $\{X_n\}_n$ is transcient if and only if for all $x, y \in S$,*

$$G(x,y) = \sum_{k=0}^{\infty} \pi^k(x,y) < \infty$$

Moreover, $G(x,x) - \frac{1}{1-f(x,x)}$ where $f(x,y) = \mathbb{P}_x(\tau_y < \infty)$ and $G(x,y) = f(x,y)G(y,y)$ for $x \neq y$.

Proof.

$$\begin{aligned} G(x,x) &= \sum_{k=1}^{\infty} \pi^k(x,x) \\ &= \sum_{k=0}^{\infty} \mathbb{P}_x(X_k = x) \\ &= \sum_{k=0}^{\infty} \mathbb{E}_x[1_{X_k=x}] \\ &= \mathbb{E}_x[\sum_{k=0}^{\infty} 1_{X_k=x}] && \text{By monotone convergence theorem} \end{aligned}$$

Let $N(x) = \sum_{k=0}^{\infty} 1_{X_k=x} = \#$ of visits to x. If $\{X_n\}_n$ is recurrent, then by the strong Markov property, $\mathbb{E}[N(x)] = \infty$ (i.e. letting τ_x^k be the time between the $k-1$st and kth visit to x, by the strong Markov property, all τ_x^k are finite with probability 1. Clearly this implies that $N(x) = \infty$ with probability 1).

Next we prove that if our Markov chain is transcient, then $\mathbb{E}[N(x)] = \frac{1}{1-f(x,x)} < \infty$. We write,

$$\mathbb{E}_x[\sum_{k=0}^{\infty} 1_{X_k=x}] = \sum_{j=0}^{\infty} j\mathbb{P}_x(j \text{ visits to } x) = \sum_{j=0}^{\infty} jf(x,x)^{j-1}(1-f(x,x)) = \frac{1}{1-f(x,x)} \textit{Why?}$$

Finally, we note that because $f(x,y) = \mathbb{E}_x[N(y)] = \mathbb{P}_x(\tau_y < \infty)\mathbb{E}_y[N(y)] = f(x,y)G(y,y)$, we have that the off diagonal terms are smaller than the diagonal terms, thus concluding the proof. □

Example 29 (SSRW on $\mathbb{Z}^d$)**.** We generalize the simple symmetric random walk on $\mathbb{Z}$ to $\mathbb{Z}^d$ by setting $X_0 = \vec{0}$ and $\xi_i = \pm e_k$ with probability $\frac{1}{2^d}$.

First we check the state type for $d = 1$.

FINISH

1.8.5 Invariant Measures

Definition 40 (Invariant/Stationary Measure)**.** Given a Markov chain with state space S and transition probability π, then a measure μ is invariant or stationary if,

$$\int \pi(x, dx)\mu(dx) = \mu(dy)$$

in words, this means that the distribution of X_{n+1} is the same as that of X_n under $\mathbb{P}_\mu$.

Remark 20. Note that when $S = \{1, ..., N\}$, this is equivalent to $\mu^T \pi = \mu^T$ or $\pi^T \mu = \mu$. Thus μ is an invariant measure on a finite state space Markov chain if and only if it is a normal eigenvector of the transition matrix with eigenvalue 1.

Example 30. In the simplest example, we can take the random walk on the graph G with 2 nodes A, B and edge set (A, B) and (B, A). That is, at each time step we move to the node we are not currently at with probability 1. This has invariant measure $\mu = \frac{1}{2}(\delta_A + \delta_B)$.

For the remainder of this section we will consider only Markov chains on countable state spaces.

FINISH

1.9 Martingales

1.9.1 Martingales Definitions and Examples

Definition 41 (Martingale)**.** Let $X_1, X_2, ...$ be a (finite or infinite) sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[|X_i|] < \infty$ for all i. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset ...$ be a filtration of $\mathcal{F}$. The sequence $(X_i)_i$ is a martingale with respect to $(\mathcal{F}_i)_i$ if for each i, $\mathbb{E}[X_{i+1}|\mathcal{F}_i] = X_i$.

Remark 21. Typically we will take our filtration to be $\mathcal{F}_i = \sigma(X_1, ..., X_i)$ (the "natural filtration" of $(X_i)_i$). When this is the case, the filtration itself will not be specified.

Definition 42 (Super/Sub-Martingale)**.** A sequence $(X_i)_i$ is a super-martingale if $\mathbb{E}[X_{i+1}|\mathcal{F}_i] \leq X_i$ and a sub-martingale if $\mathbb{E}[X_{i+1}|\mathcal{F}_i] \geq X_i$

Example 31 (Linear Martingale)**.** Let $Y_1, Y_2, ...$ be i.i.d. centered random variables. Then $X_i = c + \sum_{k=1}^{i} Y_k$ is a martingale with respect to $\mathcal{F}_m = \sigma(Y_1, ..., Y_m)$. In particular, the simple symmetric random walk on $\mathbb{Z}$ is a Martingale. If $\mathbb{E}[Y_i] < 0$, it is a super-martingale and if $\mathbb{E}[Y_i] > 0$ it is a sub-martingale.

Example 32 (Quadratic Martingale)**.** If X_n is a linear martingale as described above and let $\sigma^2 = Var(Y_i) < \infty$. Then $X_n^2 - n\sigma^2$ is a martingale with respect to $\mathcal{F}_m = \sigma(Y_1, ..., Y_m)$. To verify this, we expand out the conditional expectation,

$$\begin{aligned}\mathbb{E}[X_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n] &= \mathbb{E}[(X_n + Y_{n+1})^2 - n\sigma^2 - \sigma^2|\mathcal{F}_n] \\ &= X_n^2 - n\sigma^2 + X_n\mathbb{E}[Y_{n+1}] + \mathbb{E}[Y_{n+1}^2] - \sigma^2 = X_n^2 - n\sigma^2\end{aligned}$$

Example 33 (Exponential Martingale)**.** Let $Y_1, Y_2, ...$ be nonnegative i.i.d. random variables with $\mathbb{E}[Y_i] = 1$. If $\mathcal{F}_m = \sigma(Y_1, ..., Y_m)$, then $M_n = \prod_{k=1}^n Y_k$ is a martingale. To verify this,

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = M_n\mathbb{E}[Y_{n+1}] = M_n$$

Now if we let $Y_i = e^{\theta\xi_i}$ and $\varphi(\theta) = \mathbb{E}[e^{\theta\xi_i}] < \infty$, then $Y_1 = e^{\theta\xi_1}/\varphi(\theta)$ has mean 1, therefore,

$$M_n = \frac{e^{\theta\sum_{k=1}^n \xi_k}}{\varphi(\theta)^n}$$

is a martingale.

Example 34 (Polya's Urn)**.** Need example

1.9.2 Properties of Martingales

Proposition 20. *If X_n is a sub(super)-martingale and $n > m$, then$\mathbb{E}[X_n|\mathcal{F}_m] \leq (\geq) X_m$. It follows that if X_n is a strict martingale, $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$.*

Proof. The proof is a simple induction. □

Proposition 21. *Let $(X_i, \mathcal{F}_i)_i$ be a martingale, and let φ be a real valued convex (concave) function and $\varphi(X_i) \in L^1$ for all i. Then $(\varphi(X_i), \mathcal{F}_i)_i$ is a sub(super)-martingale.*

Proof. The proof is a direct application of the conditional Jensen's inequality. We prove this for convex φ as the concave case follows by negation. By conditional Jensen's inequality we can write,

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = \varphi(X_n)$$

□

Remark 22. As the previous 2 propositions illustrate, results for sub-martingales will have a dual result for super-martingales that can be inferred by the property that $(X_n)_n$ a super-martingale implies that $(-X_n)_n$ is a sub-martingale. From this point on we will only state the sub-martingale statements of theorems, however the corresponding super-martingale result should be implied.

Theorem 41 (Doeb's Inequality)**.** *Suppose $X_1, .., X_n$ is a finite martingale. Then*

$$\mathbb{P}(\sup_{1\leq j\leq n} |X_j| \geq l) \leq \frac{1}{l}\mathbb{E}[|X_n| 1_{\sup |X_j| \geq l}] \leq \frac{1}{l}\mathbb{E}[|X_n|]$$

Remark 23. Note that this is a generalization of Kolmogorov's maximal inequality, since, as we have shown above, $S_n = \sum_{i=1}^n Y_i$ is a martingale.

Proof. NEED PROOF □

Theorem 42. *If N is a stopping time and X_n is a super-martingale, then $X_{n\wedge N}$ is a super-martingale (where $n \wedge N = \min(n, N)$).*

1.9.3 Martingale Convergence Theorems

Example 35. Given a random variable $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \subset ...\mathcal{F}$ we can construct a Martingale by defining $X_n = \mathbb{E}[X|\mathcal{F}_n]$.

Given this example, we may now ask if all Martingales can be defined this way. In the finite case the answer is clearly yes, since $X_1, ..., X_n$ a finite martingale implies $X_i = \mathbb{E}[X_n|\mathcal{F}_i]$. In the infinite case, the answer is more subtle. This section of the notes will be devoted to answering the following 2 questions:

1. If $(X_n)_n$ is realized as $\mathbb{E}[X|\mathcal{F}_n]$ for some X, does it necessarily follow that $X_n \to X$ in some sense?
2. What do we need to know about $(X_n)_n$ to guarantee that there exists an X such that $\mathbb{E}[X|\mathcal{F}_n] = X_n$.

Theorem 43. *If there exists an $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$, then $\lim_{n\to\infty} ||X - X_n||_p = 0$ and $X_n \to X$ almost surely.*

Proof. Varadhan 5.5 (which I don't have and kinda sucks anyway) □

Theorem 44. *If for some $p > 1$ we have $\sup_n ||X_n||_p < \infty$, then there exists an $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$.*

Proof. Varadhan 5.6 □

Next we address the L^1 case.

Definition 43 (Uniform Integrability)**.** A sequence $X_1, X_2, ...$ of random variables is uniformly integrable if,

$$\lim_{k\to\infty} \sup_n \mathbb{E}[|X_n| 1_{|X_n| \geq k}] \to 0$$

Intuitively speaking, uniform integrability is a condition which requires that a small amount of mass does not escape to infinity, causing issues with convergence in expectation, as illustrated by the following non-example,

Example 36. The sequence $X_n = n$ with probability $1/n$ and 0 else is not uniformly integrable. Note that even though this sequence converges almost surely to 0, it does not converge to 0 in L^1.

Theorem 45. *If (X_n) is uniformly integrable, then there exists an $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$.*

Proof. Need to prove, also need to check the statement for the additional condition of converging in L1 □

1.9.4 Optional Stopping Theorem

FINISH SECTION

1.10 Point Processes

1.10.1 Poisson Point Processes

NEED SECTION

1.10.2 Determinental Point Processes

Let E be a one-particle space what does this mean. Generally speaking E can be a seperable Hausdorff space, however we will only consider,

$$E = \prod_{j=1}^{m} E_j \text{ where } E_j n \cong \mathbb{R}^d \text{ or } \mathbb{Z}^d$$

why write it in this specific decomposed format? For the remainder of this section we will assume, unless otherwise stated, that $E = \mathbb{R}^d$, as generalizations to the above cases are easy. Let X be the space of all finite or countable configurations of points in E. That is, for $\xi \in X$, we can write $\xi = (x_i)_{i \in \mathbb{Z}_+}$, or when $d = 1, \mathbb{Z}$ to allow for a natural ordering $x_i \leq x_{i+1}$. When $d > 1$, we assume ξ is ordered according to $||x_i||_2 < ||x_{i+1}||_2$ why strict inequality? Does this ever cause any issues?. Furthermore, we assume that all $\xi \in X$ are locally finite. That is, for any $K \subset E$ compact, $|\xi \cap K| < \infty$.

We now want to construct a σ-algebra on our set X. Let $B \subset E$ be any bounded Borel set and $n \geq 0$. We define the cylindrical set $C_n^B = \{\xi \in X : |\xi \cap B| = n\}$. We then consider the σ-algebra generated by these sets, $\mathcal{C} = \sigma(\{C_n^B\}_{n \in \mathbb{Z}_+, B \in \mathcal{B}})$.

Definition 44 (Random Point Field)**.** A random point field is a triplet $(X, \mathcal{C}, \mathbb{P})$ where X and $\mathcal{C}$ are as above and $\mathbb{P}$ is a probability measure on $(X, \mathcal{C})$.

Include discussion of constructing these probability measures given that this is a very uncountable space?

Definition 45 (k-Point Correlation Function)**.** A locally integrable function

1.11 Brownian Motion

1.12 Large Deviation Principles

1.13 Stein's Method

2 Random Graph Theory

2.1 Simple Random Graphs

2.2 Geometric Random Graphs

3 Persistent Homology

3.1 Definitions

3.1.1 Complex Constructions

Cech Construction

Vietoris-Rips Complex

Alpha Complex

Witness Complex

Mapper

3.1.2 Persistent Homology

3.2 Main Results

3.2.1 Bounds for Reconstructing a Riemannian Manifold

3.2.2 Structure Theorem for Persistent Vector Spaces

Fix k a field, $\mathbb{A}$ a pure submonoid of $\mathbb{R}_+$. The goal of these structure theorems is to classify the isomorphism classes of $\mathbb{A}$-parametrized persistent k-vector spaces using the barcode construction. In general this is complicated, but doable for objects of $\mathfrak{B}Vect(k)$ that satisfy a finiteness condition, that is always satisfied for the Vietoris-Rips complexes associated with finite metric spaces (check this later).

Definition 46. An $\mathbb{A}$-graded k-vector space is a k-vector space V equipped with a decomposition,

$$V \cong \bigoplus_{\alpha \in \mathbb{A}} V_\alpha$$

Given 2 $\mathbb{A}$ graded vector spaces V_*, W_*, we place a grading on their tensor product $V \otimes W$ by,

$$(V \otimes W)_\alpha = \bigoplus_{\alpha_1 + \alpha_2 = \alpha} V_{\alpha_1} \otimes W_{\alpha_2}$$

An $\mathbb{A}$ graded k-algebra is a $\mathbb{A}$-vector space R_* equipped with a homomorphism $R_* \otimes R_* \to R_*$ satisfying associativity and distributivity.

Example 37. The monoid k-algebra $k[\mathbb{A}]_*$, for which the grading is given by,

$$k[\mathbb{A}]_\alpha = k \cdot \alpha$$

We will denote the element $\alpha \in k[\mathbb{A}]$ as t^α. This is meant to mirror the polynomial construction. Namely, we have that $t^\alpha t^{\alpha'} = t^{\alpha + \alpha'}$ (i.e. the grading does shift additively when you multiply), so in the case where $\mathbb{A} = \mathbb{N}$, we have that $k[\mathbb{N}]$ is just the graded module $k[t]$ with the usual grading t^n is grade n.

How exactly do we define graded modules over a graded ring?

First we prove a proposition that shows that persistent k-vector spaces can be identified with a category for which we will later be able to prove a structure theorem.

Proposition 22. *Let $\underline{G}(\mathbb{A}, k)$ denote the category of $\mathbb{A}$-graded $k[\mathbb{A}]_*$-modules. Then there is an equivalence of categories,*

$$\mathfrak{B}\mathbb{A}Vect(k) \cong \underline{G}(\mathbb{A}, k)$$

Proof. Proof needed pg. 8 of Carlsson ∎

Definition 47. For $\alpha \in \mathbb{A}$, define $F(\alpha)$ to be the free $\mathbb{A}$-graded $k[\mathbb{A}]_*$-module on a single generator in grading α (i.e. all elements have their grading shifted by an additive factor of α). For any pair $\alpha, \alpha' \in \mathbb{A}$, define $F(\alpha, \alpha')$ to be the quotient,

$$F(\alpha)/(t^{\alpha'-\alpha}F(\alpha))$$

Proposition 23. *Any finitely presented object of $\underline{G}(\mathbb{A}, k)$ is isomorphic to a module of the form,*

$$\bigoplus_{s=1}^{m} F(\alpha_s) \oplus \bigoplus_{t=1}^{n} F(\alpha_t, \alpha_t')$$

and furthermore, this decomposition is unique up to reordering of the summands.

Remark 24. An R-module M is finitely presented if there exists a surjection $R^{\oplus n} \to M$. The claim is that this always holds true in the case of the Vietoris-Rips complexes for finite metric spaces.

Proof. The proof is given as a sketch by analogy with the special case of $k[t]$. Should probably have an understanding of the details of the generalization for the exam. For the case of a nongraded PID there is a proof using matrix equivalence outlined below.

Two $m \times n$ matrices M, N over a commutative ring A are said to be *equivalent* if there are invertible matrices R and S such that,

$$M = RNS$$

Any matrix P determines a module $Q(P) = A^{\oplus m}/P(A^{\oplus n}$ such that

$$A^{\oplus n} \xrightarrow{P} A^{\oplus m} \to Q(P) \to 0$$

is a presentation on $Q(P)$. Moreover we can see that when P and P' are equivalent, the corresponding $Q(P)$ and $Q(p')$ are isomorphic by examining the following commutative diagram of exact sequences.

$$\begin{tikzcd}
A^{\oplus n} \arrow[r, "P"] \arrow[d, "R", "\sim"'] & A^{\oplus m} \arrow[r] & Q(P) \arrow[r] & 0 \\
A^{\oplus n} \arrow[r, "P'"] & A^{\oplus m} \arrow[u, "S", "\sim"'] \arrow[r] & Q(P') \arrow[r] & 0
\end{tikzcd}$$

When the ring is a PID we can show that every $m \times n$ matrix is equivalent to a matrix the form,

$$\left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right]$$

with D diagonal, which clearly gives the result for modules over a PID.

Are all of the $k[\mathbb{A}]$ modules PIDs? I feel like this should be the case as they are analogous to polynomial rings

To generalize to the graded case we need to make the following adaptations to the proof:

We first consider "$\mathbb{A}$-labeled matrices" as opposed to regular matrices. That is, since a homogeneous basis (i.e. a basis such that each element lies entirely in a single component of the grading decomposition) are equipped with a map to $\mathbb{A}$ (selecting their grade) we can also equip any matrix that describes a graded homomorphism with a labeling by elements in $\mathbb{A}$ (via the labeling of the basis element they correspond to?). We write $r_i, c_j \in \mathbb{A}$ for the labeling of the ith row and jth collumn respectively.

The corresponding entries in a matrix describing a graded homomorphism written in homogeneous bases are homogeneous (since a homogeneous element would be sent to a homogeneous element, represented as such in a homogeneous basis). The grading of the element in the ij-spot in the matrix is $c_j - r_i$ (i.e. what the r_i-graded basis element in the domain needs to be shifted by to be the graded c_j in the codomain) and is therefore of the form $xt^{c_j - r_i}$ for $x \in k$. Since the element $t^{c_j - r_i}$ is determined by the labels, we can uniquely represent this graded homomorphism by an $\mathbb{A}$-labeled matrix with entries in k, such that the ijth entry is 0 if $c_j - r_i < 0$. Such a matrix is referred to as $\mathbb{A}$-adapted. For square matrices representing automorphisms, we have that $r_i = c_i$.

A square matrix P_{ij} is said to be elementary if $P_{ii} = 1$ for all i and there is only 1 nonzero off-diagonal element (these are the matrices such that multiplying on the left (right) corresponds to adding a multiple of a row (column) to another row (column). In the $\mathbb{A}$-adapted case it corresponds to adding a row (column) with smaller (larger) labeling (since elements are 0 unless $c_j > r_i$).

To prove the result in the graded setting, we just now need to show that given any $\mathbb{A}$-labeled $m \times n$ matrix, it is possible to apply a sequence of adapted row/column operations such that we get a diagonal matrix, and permuting the rows/columns gives us an upper diagonal matrix which gives the result (keeping track of the corresponding labels that refer to the gradings). Uniqueness is proved in reference [22] of Carlsson's notes. □

These propositions now set us up to define barcodes on the outputs of persistent homology.

Definition 48. An $\mathbb{A}$-valued barcode is a finite set of elements,

$$(\alpha, \alpha') \in \mathbb{A} \times (\mathbb{A} \cup \{+\infty\}$$

satisfying the condition that $\alpha < \alpha'$. An $\mathbb{A}$ valued barcode is said to be finite if all right hand endpoints are $< \infty$. If $\mathbb{A} = \mathbb{R}_+$, we refer to it as simply a barcode, without specifying the monoid.

Remark 25. Via the equivalences defined in the above propositions, we have shown that isomorphism classes of elements in $\mathfrak{B}Vect(k)$ correspond to isomorphism classes of $k[\mathbb{A}]_*$-graded modules, which correspond to a decomposition uniquely determined by a barcode labeling where we identify the module,

$$\bigoplus_{s=1}^{m} F(\alpha_s) \oplus \bigoplus_{t=1}^{n} F(\alpha_t, \alpha'_t)$$

with the barcode, $\{(\alpha_s, +\infty) | 1 \leq s \leq m\} \cup \{(\alpha_t, \alpha'_t) | 1 \leq t \leq n\}$

3.2.3 Stability Theorems