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 $A^c \in \mathcal{F}$ as well
3. For every countable collection $\{A_i\} \subset \mathcal{F}$, $\bigcup A_i \subset \mathcal{F}$ as well

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?.

1.1.2 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

$\square$

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 6. 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

1.1.3 Measurable Functions and Random Variables

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

Definition 7 (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 8 (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 9. 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 10 (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. $F(x) - F(x-) = \mathbb{P}(X = x)$

1.1.4 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 11 (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 12. 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,*

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

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.

Proposition 8 (Markov Inequality)**.** *Need statement and proofs for all of the above ALSO CHEBY-SHEV!*

Definition 13 (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.

Theorem 4 (Fubini's Theorem)**.** *If* $f \geq 0$ *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. Proof Needed (in Durrett, not in my notes) □

1.2 Convergence

1.2.1 Convergence Theorems for Integrals

Definition 14 (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 5 (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. Proof needed □

Theorem 6 (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. Proof needed □

Theorem 7 (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. Proof needed □

Theorem 8 (Dominated Convergence Theorem)**.** *If $f_n \xrightarrow{a.e.} f$, and $|f_n| \leq g$ for g integrable, then $\lim_{n\to\infty} f_n d\mu \to \int f d\mu$*

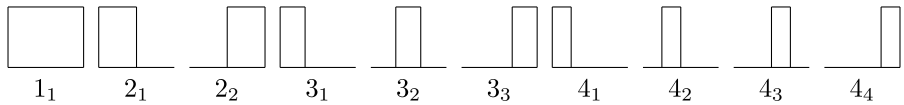
Proof. Proof needed □

1.2.2 Convergence of Random Variables

Definition 15 (Almost Sure Convergence)**.** A sequence of random variables $\{X_n\}_n$ converges almost surely to X (denoted $X_n \xrightarrow{a.s.} X$) if $\mathbb{P}(\{\omega \in \Omega | X_n(\omega)$ does not converge to $X(\omega)\}) = 0$

Definition 16 (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 9. 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 $1_1, 2_1, 2_2, 3_1, 3_2, 3_3, 4_1, \ldots$. 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 2. *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 17 (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 10. 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,

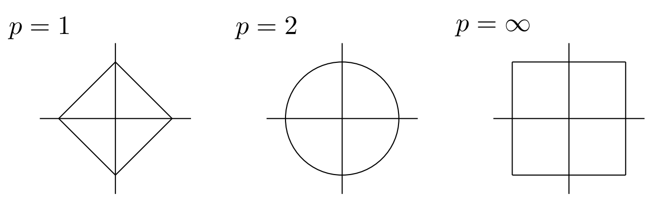
$$\liminf_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \liminf_{n\to\infty} \mathbb{E}[g_{x-\varepsilon,\varepsilon}(X_n)] = \mathbb{E}[g_{x-\varepsilon.\varepsilon}(X)] \geq \mathbb{P}(X \leq x - \varepsilon)$$

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. □

1.3 L^p Space

Definition 18 (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 11 (Minkowski's Theorem)**.** $||f + g||_p \leq ||f||_p + ||g||_p$.

Proof. Proof needed □

Theorem 12 (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 19 (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 20 (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 21 (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 11. Random variables $X_1, ..., X_n$ independent is equivalent to the σ-algebras $\sigma(X_1)...\sigma(X_n)$ independent.

Remark 12. 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 13. *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 3 (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 22. 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 13. This is weaker than independence.

Theorem 14 (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 14. 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 15 (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 $X_i^c=X_i1_{|X_i|<c}$, 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,

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

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,

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

$Y_i^c \xrightarrow{a.s.} 0$, 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 16 (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,

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

Then we have that

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

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

Theorem 17 (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 $Y_k = X_k 1_{|X_k|\leq k}$. We first show that it is sufficient to prove the convergence of the truncated variables.

Lemma 4. *Let $T_n = Y_1 + ... + Y_n$. It is sufficient to show that $\frac{T_n}{n} \xrightarrow{a.s.} \mu$.*

Proof. $\sum_{k=1}^{\infty} \mathbb{P}(|X_k| > k) \leq \int_0^{\infty} \mathbb{P}(|X_1| > t)dt = \mathbb{E}[|X_i|] < \infty$ so by Borel-Cantelli we have that $\mathbb{P}(X_k \neq Y_k$ 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 5. $\sum_{k=0}^{\infty} \frac{Var(Y_k)}{k^2} \leq 4\mathbb{E}[X_1] < \infty$

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

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

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

So to complete the proof we show that if $y \geq 0$, then $2y \sum_{k>y} \frac{1}{k^2} \leq 4$. (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 18 (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 15. 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 19 (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 20 (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,

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

So we have,

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

□

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

Lemma 6 (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 $Y_k = X_k 1_{|X_k|\le k}$, 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

1.6 Central Limit Theorem

1.6.1 Characteristic Functions

1.7 Markov Chains

1.8 Martingales

1.8.1 Radon-Nikodym Theorem

Definition 23 (Absolute Continuity)**.** Definition

Definition 24 (Radon-Nikodym Derivative)**.**

Theorem 21 (Radon-Nikodym Theorem)**.**

1.9 Large Deviation Principles

1.10 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 25. 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 3. 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 13. *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 26. 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 14. *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 16. 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.

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

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 27. 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 17. 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