ORF 526
Miklos Racz

*Emily Walters**

Friend Center 008

September 12^{th}, 2018 – December 31^{st}, 2018

Contents

1 9/13/2018 3
1.1 LLN and CLT 3
1.2 Exercise 3

2 9/18/2018 3
2.1 Problem from last week 3
2.2 Probability and Measure Theory 4
2.3 Measure Theory 4
2.4 Intersections and Unions of σ-Algebras 4
2.5 Topological Spaces, Borel σ-Algebras, Borel Sets 4
2.6 Measurable Spaces 5

3 9/20/2018 5
3.1 Probability Spaces (review) 5
3.2 Measures 5
3.3 Properties of Measures 6
3.4 Specification of Measures 6
3.5 Completeness and Negligible Sets 7
3.6 Functions 7

4 9/25/2018 7
4.1 Measurable Functions 7
4.2 Approximation of Measurable Functions 9
4.3 Integration 9

5 9/27/2018 10
5.1 The Compactification of $\mathbb{R}$ 10
5.2 Last Time 10
5.3 Continuing 10
5.4 Lebesgue Integration 11

*ewalters@princeton.edu

5.5 Integration Over a Set 11
5.6 Monotone Convergence Theorem 11
5.7 The Insensitivity of the Integral W.R.T Negligible Sets 11
5.8 Fatou's Lemma 11
5.9 Dominated Convergence Theorem 11
5.10 A note on these theorems... 12
5.11 Characterization of the Integral 12
5.12 Product Spaces 12
5.13 Absolute Continuity of Measures 12

6 10/2/2018 13
6.1 Products of Measure Spaces 13
6.2 Probability 14
6.3 Expected Value 14
6.4 Weak Law of Large Numbers 14
6.5 Markov's Inequality 14
6.6 Chebyshev's Inequality 14
6.7 General Markov 15
6.8 Chernoff Bound 15
6.9 Almost Sure Convergence 15
6.10 Strong Law of Large Numbers 15

7 10/4/18 15
7.1 Strong Law of Large Number 15
7.2 Borel-Cantelli Lemmas (Sufficient conditions for Almost Everywhere) 16
7.3 Proof of SLLN 17

8 10/16/2018 17
8.1 Last Time 17
8.2 Characteristic Functions 17
8.3 Characteristic Functions and Moments 18
8.4 Examples of Characteristic Functions 19
8.5 Inversions 19

9 10/18/2018 20
9.1 Last Time 20
9.2 Levy's Continuity Theorem 20
9.3 Central Limit Theorem Proof 20
9.4 Some Comments on Weak Convergence 21
9.5 General Strategy for Proving Weak Convergence (Prohorov) . . 21
9.6 Stochastic Processes 21
9.7 Markov Processes and Markov Chains 22
9.8 Properties of Transition Probabilities Matrix 23

10 10/25/2018 23
10.1 Markov Chains 23
10.2 Graphs 25

11 11/6/2018 26
11.1 Coupling 27

12 11/8/2018 28

13 11/15/2018 29
13.1 Martingales 29
13.2 Conditional Expectation in Simple Discrete Setting 30

14 11/27/2018 31
14.1 Recap 31
14.2 Martingales 32
14.3 Stopping Time 32
14.4 Stopped Martingales 33

1 9/13/2018

1.1 *LLN and CLT*

For a set of coin flip

1.1 THEOREM. *Weak Law of Large Numbers* $\forall\epsilon \lim_{n\to\infty} \mathbb{P}(|\frac{S_n}{n} - \frac{1}{2}| \geq \epsilon) = 0$

1.2 THEOREM. *Strong LLN* $\mathbb{P}(\lim_{n\to\infty} \frac{S_n}{n} = \frac{1}{2}) = 1$

1.3 THEOREM. *General Strong LLN If* $\{x_i\}_{i=1}^{\infty}$ *are I.I.D. random variables such that* $\mathbb{E}[|x_i|] < \infty$ *then if* $S_n = x_1 + \cdots + x_n$, $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[x_i] = \mu$.

1.2 *Exercise*

Come up with some $\{x_i\}_{i=1}^{\infty}$ such that (A) the law of large numbers does not hold and (B) the law of large numbers does hold but the central limit theorem does not.

Problem: Throw a (fair) die until you get a 6. What is the expected number of throws (including the one that gives us a 6) conditioned on the event that all throws give even numbers?

2 9/18/2018

2.1 *Problem from last week*

Problem: Throw a (fair) die until you get a 6. What is the expected number of throws (including the one that gives us a 6) conditioned on the event that all throws give even numbers?

P. Cuff: Let T be the first time when the throw is not a 2 or a 4.

$T \sim Geo(\frac{2}{3})$, $\mathbb{E}[T] = \frac{3}{2}$. However, T is independent of what the die roll actually is, so our solution is $\mathbb{E}[T \mid X_T = 6] = \mathbb{E}[T] = \frac{3}{2}$. This example demonstrates the importance of being careful when dealing with conditions.

2.2 *Probability and Measure Theory*

In 1933, Kolimogorov laid down the mathematical foundations of probability theory using measure theory. Measure theory allows us to unify the theory dealing with discrete and continuous probability cases, deals with mixtures of both discrete and continuous cases, and deals with cases that are neither discrete nor continuous. Measure theory allows us to work with much more exotic probility problems. Ex: Stochastic processes in path space.

2.3 *Measure Theory*

2.1 DEFINITION. A collection C of subsets of E (the universe) is called an algebra if:
1: $\emptyset \in C$
2: $A \in C \implies A^c \in C$
3: $A, B \in C \implies A \cup B \in C$.

C is a σ-algebra if, in addition to (1) and (2),
3′: $A_1, A_2, \cdots \in C \implies \bigcup_{i=1}^{\infty} A_i \in C$.

Examples:
1: $\mathcal{E} = \{\emptyset, E\}$ trivial σ-algebra
2: $\mathcal{E} = 2^E$ discrete σ-algebra

2.4 *Intersections and Unions of σ-Algebras*

- Any (countable or uncountable) intersection of σ-algebras is a σ-algebra.
- The union of two σ-algebras is not necessarily a σ-algebra.

2.2 DEFINITION. Generated σ-algebra Let $\mathcal{C}$ be a collection of subsets of E. Take all σ-algebras that contain $\mathcal{C}$. Take their intersection. This σ-algebra is called the σ-algebra generated by $\mathcal{C}$, and is denoted by $\sigma(\mathcal{C})$.

2.5 *Topological Spaces, Borel σ-Algebras, Borel Sets*

2.3 DEFINITION. If E is a topological space, and $\mathcal{C}$ is the collection of all open sets of E, then $\sigma(\mathcal{C})$ is called the Borel σ-algebra. Its elements are called Borel sets. The Borel σ-algebra is denoted by $\mathcal{B}_E$ or $\mathcal{B}(E)$.

2.6 *Measurable Spaces*

2.4 DEFINITION. A pair $(E, \mathcal{E})$ is a measurable space if $\mathcal{E}$ is a σ-algebra on E. The sets in $\mathcal{E}$ are called measurable sets.

2.5 DEFINITION. Let $(E, \mathcal{E})$, $(F, \mathcal{F})$ are two measurable spaces. If $A \subset E$ and $B \subset F$ are measurable sets, then $A \times B$ is called a measurable rectangle.

2.6 DEFINITION. The product $(E \times F, \mathcal{E} \otimes \mathcal{F})$ where $\mathcal{E} \otimes \mathcal{F} = \sigma(\{A \times B \mid A \in \mathcal{E}, B \in \mathcal{F}\})$, is a measurable space.

2.7 DEFINITION. $\mu : \mathcal{E} \to \mathbb{R}^+$ is a measure on $(E, \mathcal{E})$ if

1. $\mu(\emptyset) = 0$
2. If $A_1, A_2, \cdots \in \mathcal{E}$ are pairwise disjoint, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Property (2) is called countable additivity or σ-additivity.

2.8 DEFINITION. A probability measure is a measure μ such that $\mu(E) = 1$.

2.9 DEFINITION. A probability space is a triple $(E, \mathcal{E}, \mu)$ such that $(E, \mathcal{E})$ is a measurable space, and μ is a probability measure.

Probability spaces are often denoted $(\Omega, \mathcal{F}, \mathbb{P})$.

3 9/20/2018

3.1 *Probability Spaces (review)*

A probability space, written $(\Omega, \mathcal{F}, \mathbb{P})$, is composed of a sample space (Ω), the events ($\mathcal{F}$, a σ-algebra on Ω), and a probability measure ($\mathbb{P}$).

3.2 *Measures*

Suppose $(E, \mathcal{E})$ is a measurable space.

3.1 DEFINITION. We say that $\mu : \mathbf{E} \to \overline{\mathbb{R}_+}$ ($\overline{\mathbb{R}_+}$ is the reals plus the point at positive infinity) such that

- $\mu(\emptyset) = 0$
- If $A_1, A_2, \cdots \in \mathcal{E}$ are pairwise disjoint then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

Examples:

1. The Dirac Measure. $x \in E$, $\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$

2. The Counting Measure. $D \in E$, $\mu(A)$ =# of points in $A \cap D$. If D is countable, then $\mu(A) = \sum_{x \in D} \delta_x(A)$.
3. Discrete Measure. $D \subset E$ countable, $m(x)$ is some real value for every $x \in D$. $\mu(A) = \sum_{x \in D} m(x)\delta_x(A)$.
4. The Uniform Measure on $\{1, 2, \ldots, n\}$. The discrete measure with $m(x) = \frac{1}{n}$.
5. The Lebesgue Measure. $Leb(A) =$ length of A where A is an interval.

3.3 *Properties of Measures*

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

1. Finite Additivity. $A \cap B = \emptyset \implies \mu(A \cup B) = \mu(A) + \mu(B)$.
2. Monotonicity. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$. Note that this is clear because $B = A \cup (B \setminus A)$.
3. Sequential Continuity. If $A_n \subset A$ and A_n converges to A as $n \to \infty$ then $\mu(A_n)$ converges to $\mu(A)$ from below.
4. Boole's Inequality / Union Bound. $A_1, A_2, \cdots \in \mathcal{E}$, $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$. Note: We can prove this by creating a sequence of disjoint subsets of $\bigcup_{i=1}^{\infty} A_i$, then using sequential continuity.
5. If $c > 0$, then $c\mu$ is also a measure, $(c\mu)(A) = c \cdot \mu(A)$.
6. If μ_1, μ_2 are measures then $\mu_1 + \mu_2$ is a measure.

3.2 DEFINITION. If $\mu(E) < \infty$ then μ is called a finite measure.

3.3 DEFINITION. We say that a measure μ is σ-finite if there exists a measurable countable partition $\{E_n\}$ of E such that $\mu(E_n) < \infty$ for all n. Ex: Leb is σ-finite.

3.4 *Specification of Measures*

3.4 THEOREM. *Let $(E, \mathcal{E})$ be a measurable space. Let μ and ν be two measures on $(E, \mathcal{E})$ with $\mu(E) = \nu(E) < \infty$. If μ and ν agree on a collection of subsets that is closed under intersections, that generate $\mathcal{E}$, then $\mu = \nu$.*

3.5 COROLLARY. *Cor: If μ and ν are two probability measures on $\mathbb{R}$ with the same cumulative distribution functions, then $\mu = \nu$.*

3.6 DEFINITION. The cumulative distribution at a point x is $\mu([-\infty, x])$.

Assume that $\{x\} \in \mathcal{E}$ if $x \in E$. This is true of all standard measurable spaces.

3.7 DEFINITION. x is an atom of μ if $\mu(\{x\}) > 0$.

3.8 DEFINITION. μ is purely atomic if $\exists D \subset E$ such that $\forall x \in D$, $\mu(\{x\}) > 0$ and $\mu(E \setminus D) = 0$.

3.9 DEFINITION. μ is diffuse if it has no atoms. Ex: *Leb*.

3.10 LEMMA. *If μ is a σ-finite measure on $(E, \mathcal{E})$ then we can write $\mu = \lambda + \nu$ where λ is diffuse and ν is purely atomic.*

3.5 *Completeness and Negligible Sets*

3.11 DEFINITION. A measurable set A is negligible if $\mu(A) = 0$. An arbitrary subset of E is negligible if it is contained in a measurable set that is negligible.

3.12 DEFINITION. A measure space is complete if every negligible set is measurable.

3.13 LEMMA. *To make a measure space complete, take $\overline{\mathcal{E}} = \sigma(\mathcal{E} \cup \mathcal{N})$, where $\mathcal{N}$ is the collection of negligible sets. $\forall A \subset \overline{\mathcal{E}}$, $A = B \cup N$ with $B \in \mathcal{E}, N \in \mathcal{N}$. Define $\overline{\mu}(A) = \mu(B)$. This is called the completion of the measure space, $(E, \overline{\mathcal{E}}, \overline{\mu})$. In the case of $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, Leb)$, the elements of $\overline{\mathcal{B}}_{\mathbb{R}}$ are called Lebesgue-measurable.*

3.6 *Functions*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measure space. $\Omega = \{1, 2, 3, 4, 5, 6\}$, $X =$ outcome mod 5. $X(\omega) = \omega \bmod 5$, $X : \{1, 2, 3, 4, 5, 6\} \to \{0, 1, 2, 3, 4\}$. $\mathbb{P}(X = 1) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = 1\})$.

... Basic set theory stuff.

3.14 DEFINITION. $(E, \mathcal{E})$ and $(F, \mathcal{F})$ are two measure spaces. $f : E \to F$ is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if $f^{-1}(A) \in \mathcal{E}$ for every $A \in \mathcal{F}$.

3.15 THEOREM. *$(E, \mathcal{E})$, $(F, \mathcal{F})$ measurable spaces. $f : E \to F$ is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if and only if there exists a collection $\mathcal{F}_0$ of subsets of F such that $f^{-1}(B) \in \mathcal{E}$ $\forall B \in \mathcal{F}_I$, and $\mathcal{F}_0$ generates $\mathcal{F}$.*

Proof. Left as an exercise □

3.16 THEOREM. *Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, $(G, \mathcal{G})$ be measure spaces. $f : E \to F$, $g : F \to G$. If f and g are measurable, then $g \circ f$ is measurable.*

4 9/25/2018

4.1 *Measurable Functions*

Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F})$ be measure spaces. Let $f : E \to F$.

4.1 DEFINITION. f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if $f^{-1}(B) \in \mathcal{E}\ \forall B \in \mathcal{F}$.

Generally, we will focus on measurable functions $f : E \to \mathbb{R}$ (Real Valued function), $f : E \to \overline{\mathbb{R}} = [-\infty, \infty]$ (Numerical function), or similar.

4.2 DEFINITION. $f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable if it is measurable relative to $\mathcal{E}$ and $\mathcal{B}_{\mathbb{R}}$.

4.3 DEFINITION. If E is a topological space and $\mathcal{E}$ is the Borel σ-algebra, then we simply say that f is a Borel function.

4.4 LEMMA. *$f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable, if and only if $f^{-1}((-\infty, r]) \in \mathcal{E}$ for all $r \in \mathbb{R}$.*

Proof. From HW1: $\sigma(\{(-\infty, r] \mid r \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R})$. Then it follows from claim stated last time wrt the inverse of a generating set.

$\square$

4.5 DEFINITION. $f^+ := \max\{f, 0\}$, $f^- := -\min\{f, 0\}$. Note that $f = f^+ - f^-$.

4.6 LEMMA. *f is $\mathcal{E}$-measurable if and only if f^+ and f^- are $\mathcal{E}$-measurable.*

Proof. Left as an exercise $\square$

4.7 DEFINITION. An indicator function is of the form

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Check: $\mathbb{1}_A$ is $\mathcal{E}$-measurable if and only if $A \in \mathcal{E}$.

4.8 DEFINITION. A function is simple if $f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$, $a_i \in \mathbb{R}$. Where $A_1, A_2, \ldots A_n$ are $\mathcal{E}$-measurable.

4.9 DEFINITION. The canonical form of a simple function is $f = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$ where $\{B_j\}$ is a partition of $\mathcal{E}$.

4.10 FACT. Conversely, if a function is $\mathcal{E}$-measurable and takes only finitely many real values, then it is a simple function.

4.11 FACT. If f and g are simple, then so are $f + g$, $f - g$, fg, f/g, $\max\{f, g\}$, $\min\{f, g\}$.

4.12 THEOREM. *The class of measurable functions is closed under limits.*

Let $\{f_n\}$ *be a sequence of* $\mathcal{E}$*-measurable functions then* $\inf f_n$, $\sup f_n$, $\liminf f_n$, *and* $\limsup f_n$, *defined pointwise, are* $\mathcal{E}$*-measurable.*

Proof. For $\sup f_n = f$, we want to show that $f^{-1}(-\infty, r] \in \mathcal{E}$. Since intersections can be rewritten as unions, and $f(x) \leq r \iff f_n(x) \leq r \; \forall n$, we have

$$f^{-1}(-\infty, r] = \bigcap_{n=1}^{\infty} f_n^{-1}(-\infty, r]$$

But we know that $f_n^{-1}(-\infty, r] \in \mathcal{E}$ and since this is a countable intersection, $f^{-1}(-\infty, r] \in \mathcal{E}$. □

4.2 *Approximation of Measurable Functions*

let $f : \overline{\mathbb{R}}_+ \to \overline{\mathbb{R}}_+$. We can approximate f by a sequence of simple functions by subdividing $\overline{\mathbb{R}}_+$ into a partition with partitions of length $\frac{1}{2^n}$. If $x \in A$, $f_n(x)$ is the lower bound of $f(A)$. (???)

Alternatively, let $d_n(x) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{[\frac{k-1}{2^n}, \frac{k}{2^n}]} + n\mathbb{1}_{[n,\infty]}$. Then $f_n = d_n \circ f$.

4.13 THEOREM. *A function* f *is* $\mathcal{E}$*-measurable if and only if it is the increasing limit of simple functions.*

Note: Lookup monotone classes of functions.

4.3 *Integration*

Suppose $(E, \mathcal{E}, \mu)$ is a measure space. Define $f : E \to \mathbb{R}$. We want to find $\int f d_\mu$. That is, the integral of f relative to the measure μ. We denote this $\mu f = \mu(f) = \int f d_\mu = \int \mu(dx) f(x) = \int_E \mu(dx) f(x)$.

How we will do this is we will first define integrals over measure spaces for simple functions, then extend this definition by taking limits.

4.14 DEFINITION. If f is a simple function, $f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$, where $\{A_i\}$ is a partition of E we define the integral as

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

Now, suppose that f is a measurable positive function, and let $f_n = d_n \circ f$. Then $\int f d\mu = \lim_{n\to\infty} f_n d\mu$.

Finally, if $f = f^+ - f^-$, then $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$, provided that at least one of the two integrals on the right are finite. Otherwise, $\int f d\mu$ is undefined.

5 9/27/2018

5.1 *The Compactification of* $\mathbb{R}$

$\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{B}(R) \cup \{+\infty\} \cup \{-\infty\})$.

$\overline{\mathbb{R}}$ has the topology where $A \subset \overline{\mathbb{R}}$ is open if $A \setminus \{-\infty, +\infty\}$ is open in $\mathbb{R}$.

5.2 *Last Time*

For $(E, \mathcal{E}, \mu)$, $f : E \to \overline{\mathbb{R}}$, we defined $\mu(f) = \int f d\mu$.

5.3 *Continuing*

5.1 DEFINITION. f is integrable if $\int f d\mu$ exists and is finite.

Note: f is integrable $\iff \int |f| d\mu < \infty$. Notice $|f| = f^+ + f^-$.

5.2 EXERCISE. Every integrable function is real-valued almost everywhere (a.e.).

5.3 DEFINITION. A statement holds almost everywhere (for almost every $x \in E$) if it holds for all x except for x in a negligible set. Denoted μ-a.e. or (a.e.). For probability measures, we say "almost surely".

Properties of Integrals:
$a, b \in \mathbb{R}^+$, $f, g \in \mathcal{E}_+$ ($\mathcal{E}$-measurable positive functions).

1. Positivity: $\mu(f) \geq 0$. $\mu(f) = 0 \implies f = 0$ a.e.
2. Linearity: $\mu(af + bg) = a\mu(f) + b\mu(g)$.
3. Monotonicity: If $f \leq g$ a.e., then $\mu(f) \leq \mu(g)$.

Monotone Convergence Theorem: If $f_n \to f$ from below, then $\mu(f_n) \to \mu(f)$ from below.

1. Dirac measure: using the Dirac delta $\delta_{x_0}(f) = f(x_0)$.

2. $\mu = \sum_{x \in D} m(x)\delta_x$, $D \subset E$, then $\mu(f) = \sum_{x \in D} m(x) f(x)$. Note that if E is countable, then every measure is of this form ($m(x) = \mu(\{x\})$).

Note that if E is a vector space, we can think of $\mu(f)$ as the inner product $\langle \mu, f \rangle$.

5.4 *Lebesgue Integration*

$E \subset \mathbb{R}^d$ a Borel set; $\mathcal{E} = \mathcal{B}(E)$. $Leb_E :=$ restriction of Leb to $(E, \mathcal{E})$.

$Leb_E(f) = \int_E Leb_E(dx) f(x) = \int_E dx f(x) = \int_E f(x) dx$

If the Reiman integral of f exists, then Leb integral of f does as well and things are equal. However, the converse is false. Notice that if $E = [0,1]$, $f = \mathbb{1}_{\mathbb{Q}}$, then $Leb(f) = 0$.

5.5 *Integration Over a Set*

$A \subset E$, $A \in E$, $f \in \mathcal{E}$. Then $f\mathbb{1}_A \in \mathcal{E}$ and so $\mu(f\mathbb{1}_A) = \int f\mathbb{1}_A d\mu = \int_A f d\mu$.

5.6 *Monotone Convergence Theorem*

5.4 THEOREM. *Let $\{f_n\}$ be a monotone increasing sequence of measurable positive functions. Then $\mu(\lim_{n\to\infty} f_n) = \lim_{n\to\infty} \mu(f_n)$.*

Proof. $f := \lim f_n$ is well defined, so $\mu(f)$ is well defined. For all n, $f_n \leq f$, so $\mu(f_n) \leq \mu(f)$ and $\lim_n \mu(f_n) \leq \mu(f)$.

For the other direction, we want to show that $\lim_{n\to\infty} \mu(f_n) \geq \mu(d_k \circ f)$ $\forall k$. □

5.7 *The Insensitivity of the Integral W.R.T Negligible Sets*

5.5 LEMMA. *If $A \in \mathcal{E}$ is negligible, then $\int_A f d\mu = 0$ for all measurable f.*

If $f = g$ a.e., then $\mu(f) = \mu(g)$.

If $f \in \mathcal{E}_+$, $\mu(f) = 0$ then $f = 0$ a.e.

5.8 *Fatou's Lemma*

5.6 THEOREM. *Fatou's Lemma Let $(f_n)_{n\geq 1}$ be a sequence of functions in $\mathcal{E}_+$. Then $\mu(\liminf f_n) \leq \liminf \mu(f_n)$. This follows from MCT (HW).*

5.9 *Dominated Convergence Theorem*

5.7 THEOREM. *Dominated Convergence Theorem If f_n is a sequence of functions and there exists a function g such that (a) $|f_n| \leq g$ $\forall n \geq$, and (b) g is integrable, then $f := \lim f_n$ (if it exists) is integrable and $\mu(f) = \lim \mu(f_n)$. This follows from Faton's (HW).*

Terminology: g dominates f_n for every n.

5.8 COROLLARY. *Bounded Convergence Theorem Suppose that μ is a finite measure, and $|f_n| \leq c < \infty$ (c a constant), and $f := \lim f_n$ exists. Then $\mu(f) = \lim \mu(f_n)$.*

5.10 *A note on these theorems...*

For the majority of these theorems, there exists an always everywhere version.

5.11 *Characterization of the Integral*

$f \mapsto \mu(f)$, maps from $\mathcal{E}_+$ into $\overline{\mathbb{R}}_+$.

5.9 THEOREM. *Let $(E, \mathcal{E})$ be a measurable space. $L : \mathcal{E}_+ \to \overline{\mathbb{R}}_+$. Then there exists a unique measure μ on $(E, \mathcal{E})$ such that $L(f) = \mu(f)$ if and only if*

- $f = 0 \implies L(f) = 0$.
- $L(af + bg) = aL(f) + bL(g)$.
- *If $f_n \to f$ from below, then $L(f_n) \to L(f)$ from below.*

5.12 *Product Spaces*

- $(E, \mathcal{E})$, $(F, \mathcal{F})$: $(E \times F, \mathcal{E} \otimes \mathcal{F})$.
- $(E, \mathcal{E}, \mu)$, $(F, \mathcal{F}, \nu)$: $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \times \nu)$.
- $(\mu \times nu)(A \times B) = \mu(A) \times \nu(B)$.

5.10 THEOREM. *Fubini: Suppose that $f : E \times F \to \overline{\mathbb{R}}$ such that $\int\int_{E \times F} |f| d(\mu \times \nu) < \infty$. Then $\int\int_{E \times F} f d(\mu \times \nu) = \int_F (\int_E f(x,y)\mu(dx))\nu(dy) = \int_E (\int_F f(x,y)\nu(dy))\mu(DX)$.*

5.11 THEOREM. *Tonelli: If $f \geq 0$ then the same conclusions hold.*

5.13 *Absolute Continuity of Measures*

5.12 DEFINITION. $(E, \mathcal{E})$ with measures μ and ν. We say that μ is absolutely continuous with regard to ν if $\forall A \in \mathcal{E}$, $\nu(A) = 0 \implies \mu(A) = 0$. Denote this by $\mu << \nu$.

Example: If a measure on $\mathbb{R}$ has a density (e.g., $\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, the standard Gaussian measure), then it is absolutely continuous with regard to the Lebasgue measure.

Example: Discrete distributions with the same support.

5.13 THEOREM. *Suppose that μ is σ-finite, and that $\nu << \mu$. Then there exists a positive $\mathcal{E}$-measurable function p such that $\int_E \nu(dx)f(x) = \int_E \mu(dx)p(x)f(x)$, $\forall f \in \mathcal{E}_+$.*

Moreover, p is unique up to equivalence (if this holds for p', then $p = p'$ a.e.).

5.14 DEFINITION. This function p is called the Radon-Nikodym derivative of ν with regard to μ. We write this as $p(x) = \frac{\nu(dx)}{\mu(dx)}(x)$, or $p = \frac{d\nu}{d\mu}$.

If we care about ν, but it is difficult to use. If $\nu << \mu$, then we can perform calculations using the nicer μ.

5.15 DEFINITION. μ is singular with regard to μ if there exists some set $D \in \mathcal{E}$ such that $\mu(D) = 0$ and $\nu(E \setminus D) = 0$.

6 10/2/2018

6.1 *Products of Measure Spaces*

A product of measure spaces

$$\bigotimes_{i=1}^{n}(E_i, \mathcal{E}_i, \mu_i)$$

can be seen as n mutually independent random variables (i.e. n coin tosses).

How do we define a countably infinite product of measure spaces ($\bigotimes_{i=1}^{\infty}(E_i, \mathcal{E}_i, \mu_i)$)?

Let $\mathcal{R}$ be the collection of all finite dimension measurable rectangles. That is, all sets of the form $\{x \mid x_1 \in B_1, \ldots, x_n \in B_n, x_{n+1} \in \mathbb{R}, x_{n+2} \in \mathbb{R}, \ldots\}$ where $n \in \mathbb{N}$ and $B_i \in \mathcal{B}(\mathbb{R})$. Then $\mathcal{B}_C = \sigma(\mathcal{R})$. We define the measure as

$$\mu(\{x \mid x_1 \in B_1, \ldots, x_n \in B_n, \ldots\}) = \mu_1(B_1)\mu_2(B_2)\ldots\mu_n(B_n)$$

6.1 THEOREM. *Kolmogarov's Extension Theorem: Suppose $\{mu_n\}_{n\geq 1}$ is a sequence of probability measures, where μ_n is a probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ that is consistent. That is*

$$\mu_{n+1}(\{x_1 \in B_1, x_2 \in B_2, \ldots, x_n \in B_n, x_{n+1} \in \mathbb{R}^n\}) = \mu_n(\{x_1 \in B_1, x_2 \in B_2, \ldots, x_n \in B_n\})$$

For all $n \in \mathbb{N}$ and all $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. Then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_C)$ such that $\mathbb{P}(\{w \mid w_1 \in B_1, \ldots, w_n \in B_n\}) = \mu_n(B_1 \times B_2 \times \cdots \times B_n)$.

6.2 *Probability*

$(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. $X : \Omega \to \mathbb{R}$ a random variable. A distribution of X is the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Note that $\mu(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}) = \mathbb{P}(X^{-1}A)$.

Reminder: Refresh yourself on the definitions of the common probability distributions

- Binomial
- Geometric
- Poisson
- Exponential
- Gaussian

6.3 *Expected Value*

6.2 DEFINITION. $\mathbb{E}[X] = \int_{\mathbb{R}} x\mu(dx) = \int_{\omega} X(\omega)\mathbb{P}(d\omega)$

6.4 *Weak Law of Large Numbers*

6.3 THEOREM. *Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables, and that $\mathbb{E}[|X_n|] < \infty$. Let $m = \mathbb{X}_{\mathbb{1}}$. Then $\lim(n \to \infty)\mathbb{P}\left(|\frac{X_1+\cdots+X_n}{n} - m| \geq \epsilon\right) = 0$.*

6.5 *Markov's Inequality*

6.4 THEOREM. *Let X be a non-negative random variable ($X : \Omega \to \mathbb{R}_+$), and let $\lambda > 0$. Then $\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}$.*

Proof. $\mathbb{P}(X \geq \lambda) = E[\mathbb{1}_{\{x \geq \lambda\}}] \leq \mathbb{E}[\frac{X}{\lambda}]$. □

6.6 *Chebyshev's Inequality*

6.5 THEOREM. *Let X be a random variable, such that $\mathbb{E}[X^2] < \infty$. Then*

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

where $Var(X) = \mathbb{E}[(X - \mathbb{E}[X]^2)]$.

Proof. $\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda) = \mathbb{P}(|X - \mathbb{E}[X]^2| \geq \lambda^2)$. Now apply Markov's Inequality. □

6.7 *General Markov*

6.6 theorem. *Let X be a random variable and $f : \mathbb{R} \to \mathbb{R}_+$ an increasing function.*

Then $\mathbb{P}(X \geq \lambda) = \mathbb{P}(f(X) \geq f(\lambda)) \leq \frac{\mathbb{E}[f(X)]}{f(\lambda)}$.

6.8 *Chernoff Bound*

6.7 theorem. *Suppose $X_1, X_2, \ldots, X_n$ independent Bernoulli random variables, with $\mathbb{E}[X_i] = p_i$. Let $S_n = \sum_{i=1}^{n} X_n$, and $\mu = \sum_{n=1}^{n} p_i$. Then*

$$\mathbb{P}(S_n \geq \mu + \lambda) \leq e^{-\frac{2\lambda^2}{n}}$$

and

$$\mathbb{P}(S_n \leq \mu - \lambda) \leq e^{-\frac{2\lambda^2}{n}}$$

In general, Chernoff provides a much tighter bound than Chebyshev.

Proof. Homework. □

6.9 *Almost Sure Convergence*

$Y_1, Y_2, \ldots$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

6.8 definition. $\lim_{n\to\infty} Y_n = 0$ almost everywhere $\iff \mathbb{P}(\{\omega \mid \lim_{n\to\infty} Y_n(\omega) = 0\}) = 1$.

6.10 *Strong Law of Large Numbers*

6.9 theorem. *Under the same conditions as before, ($X_1, X_2, \ldots$ I.I.D., $\mathbb{E}[|X_1|] < \infty$, $m = \mathbb{E}[X_1]$) we have that $\frac{S_n}{n} \to m$ almost everywhere as $n \to \infty$.*

7 10/4/18

7.1 *Strong Law of Large Number*

7.1 theorem. *Let $X_1, X_2, \ldots$ be I.I.D. random variables with $\mathcal{E}[X_1] < \infty$ and $m = \mathcal{E}[X_1]$. Then $\frac{1}{n}\sum_{i=1}^{n} X_i \to m$ almost everywhere.*

In other words

$$\mathbb{P}(\{\omega \mid \lim_{n\to\infty} \sum_{i=1}^{n} X_i(\omega) = m\}) = 1$$

Equivalently $\frac{1}{n}\sum_{i=1}^{n}(X_i - m) \to 0$ almost everywhere. Without loss of generality, assume $m = 0$.

7.2 *Borel-Cantelli Lemmas (Sufficient conditions for Almost Everywhere)*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_i \in \mathcal{F}$ for $i \geq 1$. Define the following event

$$\limsup A_n = \bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} A_i = \{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n\}$$

$$\liminf A_n = \bigcup_{j=1}^{\infty} \bigcap^{\infty} i = jA_i = \{\omega \in \Omega \mid \omega \in A_n \text{ for all but finitely many } n\}$$

We can think of unions and intersections of sets representing event as follows:

$$\bigcap \iff \forall \mid \bigcup \iff \exists$$

This terminology comes from $\limsup_{n\to\infty} \mathbb{1}_{A_k}(\omega) = \mathbb{1}_{\limsup A_n}(\omega)$.

Fattou: $\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n)$.

7.2 LEMMA. *Borel-Cantelli I: If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup A_n) = 0$. That is, almost surely only finitely many of the events happen.*

Proof. Let $\epsilon > 0$ be arbitrary. Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, there exists some N_ϵ such that $\sum_{n=N_\epsilon}^{\infty} \mathbb{P}(A_n) \leq \epsilon$. Suppose $\sum_{n=N}^{\infty} \mathbb{P}(A_n) \leq \epsilon$. Then we have

$$0 \leq \mathbb{P}(\limsup A_n) = \mathbb{P}(\bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} A_i) \leq \mathbb{P}(\bigcup_{i=N}^{\infty} A_i) \leq \sum_{i=N}^{\infty} \mathbb{P}(A_i) \leq \epsilon$$

□

7.3 LEMMA. *Borel-Cantelli II: If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$ and A_n are mutually independent then $\mathbb{P}(\limsup A_n) = 1$.*

That is, almost surely infinitely many of A_n will occur.

Proof. Want to show $\mathbb{P}(\limsup A_n) = 1$. That is equivalent to saying that $\mathbb{P}((\limsup A_n)^C) = 0$. Now,

$$(\limsup A_n)^C = (\bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} A_i)^C = \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} A_i^c = \liminf A_n^c$$

Need to show that $\mathbb{P}(\bigcap_{i=j}^{\infty} A_i^C = 0)$. Fix a large number M, then $\mathbb{P}(\bigcap_{i=j}^{\infty} A_i^C) \leq \mathbb{P}(\bigcap_{i=j}^{M} A_i^C) = \prod_{i=j}^{M} \mathbb{P}(A_i^C) = \prod_{i=j}^{M}(1 - \mathbb{P}(A_i)) \leq \prod_{i=j}^{M} e^{-\mathbb{P}(A_i)} = e^{-\sum_{i=j}^{M} \mathbb{P}(A_i)}$.

Now let $M \to \infty$ and it goes to 0.

$\square$

7.3 *Proof of SLLN*

7.4 theorem. *Let $X_1, X_2, \ldots$ be I.I.D. random variables with $\mathcal{E}[|X_1|] < \infty$ and $\mathcal{E}[X_1] = 0$. Then $\frac{1}{n}\sum_{i=1} X_i \to 0$ almost everywhere.*

Step 1: Kolmogorov's inequality
Step 2: Apply inequality to show that the summable variances imply a.s. convergence.
Step 3: Kronecker's lemma and SLLN with summable variances condition
Step 4: Full SLLN proof using truncation argument

7.5 theorem. *Kolmogorov's Inequality: Let $X_1, X_2, \ldots, X_n$ be mutually independent. Assume $\mathbb{E}[X_i] = 0$ and $\sigma_i^2 = \mathbb{E}[X_i^2] < \infty$. Then for any $\lambda > 0$ we have $\mathbb{P}(\max_{1\le i\le n} |X_1 + X_2 + \cdots + X_n| \ge \lambda) \le \frac{\sum_{i=1}^n \sigma_i^2}{\lambda^2}$. Note: this is a strengthening of Chebyshev.*

Proof. Let $S_k = X_1 + X_2 + \cdots + X_k$. Define $A = \{\omega \mid \max_{1\le i\le n} |S_i| \ge \lambda\}$. Define $A_k = \omega \mid \max_{1\le i\le k-1} |S_i| < \lambda, |S_k| \ge \lambda$.

Notice that $A = \bigcup_{k=1}^n A_k$, $\mathbb{1}_A = \sum_{k=1}^n \mathbb{1}_{A_k}$, $A_k \cap A_l = \emptyset$ for $k \ne l$.

$$\mathbb{P}(A) = \mathbb{E}\mathbb{1}_A = \sum_{k=1}^n \mathbb{E}\mathbb{1}_{A_k} \le \sum_{k=1}^n \mathbb{E}[\frac{S_k^2}{\lambda^2}\mathbb{1}_{A_k}] \le \frac{1}{\lambda^2}\sum_{k=1}^n \mathbb{E}[S_k^2\mathbb{1}_{A_k}] + \mathbb{E}[(S_n - S_k)^2\mathbb{1}A_k]$$

$\square$

8 10/16/2018

8.1 *Last Time*

Last time we proved the central limit theorem.

8.2 *Characteristic Functions*

8.1 definition.

$$\Phi_X(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$$

Is the characteristic function of X, with X a random variable.

Properties:

1. $\Phi_X(0) = 1$

2. $\Phi_{-X}(t) = \Phi_X(-t) = \overline{\Phi_X(t)}$
3. $|\Phi_X(t)| \leq 1$
4. $t \mapsto \Phi_X(t)$ is uniformly continuous on $\mathbb{R}$.
5. Φ_X is positive definite. That is $\forall n$ and $\forall t_1, t_2, \ldots, t_n \in \mathbb{R}$, the matrix $\{M_{i,j} = \Phi_X(t_i - t_j)\}_{i,j=1}^n$ is positive definite. That is, $\forall z_1, z_2, \ldots, z_n \in \mathbb{C}$, $\sum_{i,j=1}^n z_i \Phi_X(t_i - t_j)\overline{z_j} \geq 0$.

Note: The distribution of X is symmetric around 0 if and only if Φ_X is real.

8.2 THEOREM. *Bochner's Theorem: Let $\Phi : \mathbb{R} \to \mathbb{C}$. Suppose that*

- $\Phi(0) = 1$.
- *$t \mapsto \Phi(t)$ is continuous at $t = 0$.*
- *Φ is positive definite.*

Then Φ is the characteristic function of some random variable X. In other words, there exists a CDF F such that $\Phi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$.

Further Properties:

- $\Phi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{ibt}\Phi_X(at)$.
- If X_1 and X_2 are independent, then $\Phi_{X_1+X_2}(t) = \Phi_{X_1}(t) + \Phi_{X_2}(t)$.
- If $F' = f$ (F a CDF), then $\Phi = \hat{f}$, the Fourier transform. That is, in this case, $\Phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx$.

8.3 *Characteristic Functions and Moments*

8.3 THEOREM. *Suppose that $\mathbb{E}[|X|^k] < \infty$. Then $\Phi_X \in \mathbb{C}^k$ and $\Phi_X^{(k)}(t) = \mathbb{E}[(iX)^k e^{itx}]$.*

Note: $\exists$ a random variable X such that $\mathbb{E}|X| = \infty$, but Φ_X is differentiable at $t = 0$.

8.4 CLAIM. *Taylor Approximation Around $t = 0$: Assume that $\mathbb{E}[|X|^m] < \infty$. Then $\Phi_X(t) = \sum_{k=0}^m \mathbb{E}[X^k] \cdot \frac{(it)^k}{k!} + o(t^m)$.*

8.5 CLAIM. *Analyticity: Suppose that $\mathbb{E}[|X|^k] < \infty\ \forall k$ and $R^{-1} := \limsup_{m\to\infty} (\frac{|\mathbb{E}[X^m]|}{m!})^{\frac{1}{m}} < \infty$, then Φ extends analytically to the strip $\{t + is \mid |s| < R\}$.*

Example: $\mathcal{N}(0,1)$, $R = \infty$.

8.4 *Examples of Characteristic Functions*

- $X \sim Bernoulli(p) : \Phi_X(t) = pe^{it} + (1-p)$.
- $X \sim Bin(n,p) : \Phi_X(t) = (pe^{it} + (1-p))^n$.
- $X \sim Radenmacher : \Phi_X(t) = \cos(t)$.
- $X \sim Uni(a,b) : \Phi_X(t) = \frac{e^{itb}-e^{ita}}{it(b-a)}$.
- $X \sim \mathcal{N}(m,\sigma^2) : \Phi_X(t) = e^{itm-\frac{\sigma^2 t^2}{2}}$.
- $X \sim Exp(\lambda) : \Phi_X(t) = \frac{\lambda}{\lambda - it}$.
- $X \sim Cauchy : \Phi_X(t) = e^{-|t|}$.

Want to compute $\Phi_X(t)$ with $X \sim \mathcal{N}(0,1)$. Let f_X be the density function of the standard normal distribution X. We know $\Phi'_X(t) = \mathbb{E}[iXe^{itX}] = \int_{\mathbb{R}} -x\sin(tx)f_X(x)dx = \int_{\mathbb{R}} \sin(tx)f'_X(x)dx = -t\int_{\mathbb{R}} \cos(tx)f_X(x)dx = -t\Phi_X(t)$. So,

$$\Phi'_X(t) = -t\Phi_X(t); \Phi_X(0) = 1$$

This ODE problem has a unique solution, $\Phi_X(t) = e^{-\frac{t^2}{2}}$.

8.5 *Inversions*

8.6 theorem. *Levy's Inversion Theorem: Suppose that X has CDF F_X and characteristic function Φ_X. For every real numbers $a < b$ and t. Let*

$$\Psi_{a,b}(t) := \frac{1}{2\pi}\int_a^b e^{-itu}du = \frac{e^{-tb} - e^{-ita}}{-i2\pi t}$$

Then $\lim_{T\to\infty} \int_{-T}^{T} \Psi_{a,b}(t)\Phi_X(t)dt = \frac{1}{2}[F_X(b) + F_X(b-)] - \frac{1}{2}[F_X(a) + F_X(a-)]$.

In particular, if a and b are continuity points of F_X, then the limit is $F_X(b) - F_X(a)$.

Furthermore, if $\int_{\mathbb{R}} |\Phi_X(t)|dt < \infty$ then X has the following bounded and continuous probability density function:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi_X(t)dt$$

This is a special case of the standard Fourier inversion. The formula here is an integrated version which holds also in the absence of a density.

8.7 corollary. *If $\Phi_X(t) = \Phi_Y(t)\ \forall t \in \mathbb{R}$, then X and Y have the same distribution.*

Proof. Proof left as an exercise. □

9 10/18/2018

9.1 *Last Time*

The "Law" of X, written $\mathcal{L}(X)$, is in one to one correspondence with Φ_X.

9.2 *Levy's Continuity Theorem*

9.1 THEOREM. *Let $\{F_n\}_{n\geq 1}$ be a sequence of CDFs on $\mathbb{R}$ and let $\{\Phi_n\}_{n\geq 1}$ be the corresponding characteristic functions.*

- *If $F_n \to F$ then $\Phi_n(t) \to \Phi(t)$ for all $t \in \mathbb{R}$ (pointwise).*
- *Suppose that for every $t \in \mathbb{R}$, the limit $\lim_{n\to\infty} \Phi_n(t)$ exists and denote it by $\Phi(t)$. Suppose that Φ is continuous at $t = 0$. Then $\exists$ CDF F such that $\Phi(t) = \int e^{itx} dF(x)$ and $F_n \to F$.*

9.3 *Central Limit Theorem Proof*

9.2 THEOREM. *Suppose $X_1, X_2, \ldots$ are I.I.D. such that $\mathbb{E}[X_1^2] < \infty$. Assume $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1^2] = 1$. Let $S_n = X_1 + \cdots + X_n$. Then $\frac{S_n}{\sqrt{n}} \to \mathcal{N}(0,1)$.*

Proof. $\Phi_{X_1}(t) = \mathbb{E}[e^{itX_1}]$. As $t \to 0$,

$$\Phi_{X_1}(t) = 1 + \mathbb{E}[X_1]\frac{it}{1!} + \mathbb{E}[X_1^2]\frac{(it)^2}{2!} + o(t^2) = 1 - \frac{t^2}{2 + o(t^2)}$$

where $f(t) = o(t^2)$ if $\lim_{t\to 0} \frac{f(t)}{t^2} = 0$. Now,

$$\begin{aligned}
&\Phi_{\frac{S_n}{\sqrt{n}}}\\
=&\mathbb{E}[e^{it(\frac{S_n}{\sqrt{n}})}]\\
=&\mathbb{E}[e^{(\frac{it}{\sqrt{n}})(X_1+\cdots+X_n)}]\\
=&\mathbb{E}[e^{(\frac{it}{\sqrt{n}})X_1}]\cdot\cdots\cdot\mathbb{E}[e^{(\frac{it}{\sqrt{n}})X_n}]\\
=&(\Phi_{X_1}(\frac{t}{\sqrt{n}}))^n\\
=&(1 - \frac{(\frac{t}{\sqrt{n}})^2}{2} + o((\frac{t}{\sqrt{n}})^2))^n\\
=&(1 - \frac{\frac{t^2}{1}}{2} + o(\frac{t^2}{2}))^n\\
\to& e^{-\frac{t^2}{2}}
\end{aligned}$$

□

9.4 *Some Comments on Weak Convergence*

9.3 definition. A sequence of probability measures $\{\mu_n\}_{n\geq 1}$ on $\mathbb{R}$ is tight if $\forall\epsilon > 0$, $\exists K < \infty$ such that $\forall n \geq 1$, $\mu_n([-K, K]) > 1 - \epsilon$. Intuitively, mass doesn't "disappear to infinity" if $\mu_n = \mu$.

Counterexamples:

- $\mu_n = \delta_n$
- $\mu_n : \mathcal{N}(0, \sigma_n^2 = n)$
- $Uni(-n, n)$.

9.4 definition. Let $\{\mu_n\}_{n\geq 1}$ be a probability measure on a complete metric space S. This sequence is called "tight" if $\forall\epsilon > 0$, $\exists K \subset S$ such that $\forall n \geq 1$, $\mu_n(K) > 1 - \epsilon$.

9.5 claim. *If $F_n \to F$, then $\{F_n\}_{n\geq 1}$ is tight.*

Proof. Proof left as an exercise.

□

9.6 theorem. *Helly: If $\{F_n\}_{n\geq 1}$ are tight, then $\exists\{n_k\}_{k\geq 1}$ (subsequence) and a CDF F such that $F_{n_k} \to F$.*

9.7 theorem. *Prohovor: Let S be a complete separable metric space. If $\{\mu_n\}_{n\geq 1}$ are tight, then $\exists$ a weakly convergent subsequence $\{\mu_{n_k}\}_{k\geq 1}$.*

9.5 *General Strategy for Proving Weak Convergence (Prohorov)*

1. Show that $\{\mu_n\}_{n\geq 1}$ is tight.
2. Identify the limit μ.
3. Show uniqueness of the limit.

9.6 *Stochastic Processes*

- Index set $\mathbb{T}$ representing time (EX: $\mathbb{R}$, $\mathbb{R}_+$, $[a, b]$, $\mathbb{Z}$, $\mathbb{Z}_+$, …)
- State space S (locally compact complete metric space)

$$t \mapsto X_t; X : \mathbb{T} \times \Omega \to S$$

Where $\forall t \in \mathbb{T}; X(t, \cdot) : \Omega \to S$ is measurable.

1. $t \in \mathbb{T}$, $X_t : \Omega \to S$ (marginal)

2. $\omega \in \Omega$ fixed, $X(\cdot, \omega) : \mathbb{T} \to S$ a (random) function.

3. $X : \Omega \to S^{\mathbb{T}} = \{\text{functions } \mathbb{T} \to S\}$. A random variable taking values in function space.

Examples:

- $\{X_n\}_{n \geq 1}$ I.I.D.
- $S_n = \sum_{i=1}^{n} Y_i$ where $\{Y_i\}_{n \geq 1}$ are I.I.D.
- Random walks on a graph.

9.7 *Markov Processes and Markov Chains*

9.8 DEFINITION. Markov Process: A stochastic process where the distribution of the future given the past and present only depends on the present.

Discrete Time and Discrete Space: In this case, Markov Processes are called Markov Chains.

9.9 DEFINITION. $X_0, X_1, \dots$ is a Markov Chain on a discrete state space S if $\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = X_{n-1})$ $\forall n \geq 1, \forall x_0, x_1, \dots, x_n \in S$. That is, the distribution of X_n depends only on the result of X_{n-1}.

9.10 REMARK. Notation: $X_0^n = (X_0, X_1, \dots, X_n)$.

Notice that, in general,

$$\mathbb{P}(X_0^n = x_0^n) = \mathbb{P}(X_n = x_n \mid x_0^{n-1} = x_0^{n-1}) \cdot \mathbb{P}(X_0^{n-1} = x_0^{n-1}) = \cdots = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i \mid X_0^{i-1} = x_0^{i-1})$$

So for Markov chains,

$$\mathbb{P}(X_0^n = x_0^n) = \mathbb{P}(X_0 = x_0) \cdot \prod_{i=1}^{n} (X_i = x_i \mid X_{i-1} = X_{i-1})$$

9.11 REMARK. Notation:

$$\mathbb{P}(X_n = y \mid X_{n-1} = x) = P_{x,y}(n)$$

This is called the transition probabilities index. That is, the probability of going from x to y at time n. This gives the transition probability matrix at time n:

$$P(n) = \{P_{x,y}(b)\}_{x,y \in S}$$

9.12 DEFINITION. Often the transition probabilities are not a function of n. In that case, we say that the Markov chain is time-homogeneous, and write

$$P = \{P_{x,y}\}_{x,y \in S}$$

9.8 *Properties of Transition Probabilities Matrix*

Notice that for a Markov chain, $\mathbb{P}(X_0^n = x_0^n) = \mathbb{P}(X_0 = x_0) \cdot \prod_{i=1}^n P_{x_{i-1},x_i} P(i)$. So, two things determine a Markov chain

- The initial distribution X_0.
- Transition probabilities.

So

- $0 \leq P_{x,y} \leq 1$.
- $\sum_{y \in S} P_{x,y} = 1 \; \forall x \in S$.

10 10/25/2018

10.1 *Markov Chains*

For the time being we will consider finite state spaces (time-homogeneous Markov chains).

Say we want to understand the probability that we end up at a particular state at timestep n. We can do so by taking powers of the transition probability matrix. That is, $\mathbb{P}(X_n = y \mid X_0 = x) = (P^n)_{x,y}$.

Proof. This is true for $n = 1$ by definition. For $n > 1$, we can use conditioning

$$\begin{aligned}
\mathbb{P}(X_n = y \mid X_0 = x) &= \sum_{z \in S} \mathbb{P}(X_n = y, X_{n-1} = z \mid X_0 = x) \\
&= \sum_{z \in S} \mathbb{P}(X_n = y \mid X_{n-1} = z, X_0 = x) \\
&= \sum_{z \in S} \mathbb{P}(X_n = y \mid X_{n-1} = z)\mathbb{P}(X_{n-1} = z \mid X_0 = x) \\
&= \sum_{z \in S} P_{zy}(P^{n-1})_{xy} = (P^n)_{xy}
\end{aligned}$$

□

Now, think of what P defines. P is a matrix,

- It acts on column vectors to its right. Think of these column vectors as functions.
- It acts on row vectors to the left. Think of these as probability measures.

Suppose $f : S \to \mathbb{R}$. Then $(P^n f)(x) = \mathbb{E}[f(X_n) \mid X_0 = x]$.

Proof. $(Pf)(x) = \sum_{y \in S} P_{xy} f(y) = \sum_{y \in S} \mathbb{P}(X_1 = y \mid X_0 = x) f(y) = \mathbb{E}[f(X_1) \mid X_0 = x]$.

Now suppose $\mu : S \to \mathbb{R}$ is a measure. Then $(\mu P^n)(x) = \mathbb{P}_\mu(X_n = x)$ here $\mathbb{P}_\mu$ indicates that the initial distribution is μ.

$$(\mu P^n)(x) = \sum_{y \in S} \mu(y)(P^n)_{yx} = \sum_{y \in S} \mu(y) \mathbb{P}(X_n = x \mid X_0 = y)$$

□

Note: P acts to the right on functions.

$$l^\infty(S) = \{f : S \to \mathbb{R} \mid ||f||_\infty < \infty$$

Properties:

- Keeps positivity: If $f \geq 0$, then $Pf \geq 0$.
- P is a contraction: $||P||_{\infty,\infty} \leq 1$.
- $||Pf||_\infty = \max_{x \in S} |(Pf)(x)| = \max_{x \in S} |\sum_y P_{xy} f(y)| \leq \max_{x \in S} \sum_y P_{xy} |f(y)| \leq \max x \in S \sum_y P_{xy} ||f||_\infty = ||f||_\infty$.
- $P1 = 1$ (That is, vector of all 1s).

Note: P acts to the left on measures.

$l^1(S) = \{\mu : S \to \mathbb{R} \mid ||\mu||_1 < \infty\} = \sum_{x \in S} |\mu(x)|$.

1. Keeps positivity.
2. Is a contraction. $||\mu P||_1 \leq ||\mu||_1$.
3. $\exists \mu$ such that $\mu P = \mu$.

10.1 DEFINITION. If $\mu P = \mu$, and μ is a probability measure, then μ is a stationary distribution.

10.2 CLAIM. *If $\mu P = \mu$, then also $|\mu| P = |\mu|$ where $|\mu|(x) = |\mu(x)|$.*

10.3 COROLLARY. *If $\mu P = \mu$, then $\mu^+ P = \mu^+$, $\mu^{-1} P = \mu^{-1}$.*

10.4 COROLLARY. *If S is finite, then a stationary probability measure with respect to P always exists.*

Proof.

$$(|\mu|P)(x) = \sum_{y \in S} |\mu(y)| P_{yx} \geq |\sum_{y \in S} \mu(y) P_{yx}| = |\mu(x)|$$

Suppose that $\exists x \in S$ such that $(|\mu|P)(x) > |\mu(x)|$. Then $|||\mu|P|_1 = \sum_{y \in S}(|\mu|P)(y) > \sum_{y \in S} |\mu|(y) = |||\mu|||_1$. This contradicts the contraction property.

□

10.2 *Graphs*

Let $G + (V, E)$ be a directed graph with $(i, j) \in E$ (indicating a change of state from i to j) if and only if $P_{ij} > 0$.

10.5 DEFINITION. $A \subset S$ is closed if $\mathbb{P}(A \to A^C) = 0$. That is, once the state is in A, it cannot enter a state outside of A.

10.6 REMARK. Note: $\emptyset, S$ are closed. If A, B are closed, then $A \cap B$ and $A \cup B$ are also closed.

10.7 DEFINITION. $A \subset S$, $\overline{A} = \bigcap_{B \supset A; B \text{ closed}} B$. This is the closure of A.

10.8 DEFINITION. Suppose that $A = \overline{A}$ and $A \neq \emptyset$. Then A is said to be irreducible if $\forall B \subset A$, $B \neq \emptyset$ we have $\overline{B} = A$. Otherwise, A is called reducible.

10.9 DEFINITION. $x \in S$ is an absorbing state if $\{x\}$ is irreducible.

10.10 DEFINITION. $x \in S$ is an inessential state if it is not part of any irreducible component. Otherwise, the state is essential.

Often we will focus on irreducible Markov chains, because over a long time scale, only the irreducible components of any Markov chain are relevant.

Periodicity:

10.11 DEFINITION. The period of $x \in S$ is the greatest common divisor of all walks that start in x and come back to x.

10.12 CLAIM. *Suppose $x, y \in S$ such that there exists some path $x \to y$ and a path*

$y \to x$. Then $per(x) = per(y)$.

Proof. Proof left as an exercise. $\square$

10.13 COROLLARY. *The period of every state in an irreducible component is the same.*

10.14 DEFINITION. The period of an irreducible component is equal to the period of its states.

10.15 DEFINITION. An irreducible Markov chain (a chain composed of a single irreducible component) has a period equal to the period of its states.

10.16 FACT. The number of stationary measures of a Markov chain is equal to the number of irreducible components of that chain.

10.17 DEFINITION. P is ergodic if it is irreducible and is aperiodic.

10.18 DEFINITION. If a state, component, or chain has a period of 1, we call it aperiodic.

10.19 THEOREM. *Suppose that $X_0, X_1, \ldots$ is an ergodic Markov chain on a finite state space S with transition probability matrix P.For all $x \in S$, if $X_0 = x$, $X_n \to \Pi$ as $n \to \infty$ where Π is the unique stationary distribution.*

11 11/6/2018

11.1 THEOREM (Convergence of Markov Chains). *Let S be a finite state space. let $X_0, X_1, X_2, \ldots$ be an S-valued ergodic (irreducible and aperiodic) Markov chain with $X_0 = x \in S$. Then $X_n \to \Pi$ as $n \to \infty$ whhere Π is the unique stationary distribution of the Markov chain.*

11.2 THEOREM. *$dim\{f \mid Pf = f\}$ = number of irreducible components.*

Proof. Suppose that P is irreducible. $f = c \cdot \mathbb{1}$ is an eigenvalue.

Proof by contradiction for $\leq$: Assume $\exists f \mid Pf = f$ such that f is nonconstant. Let $A = \{x \in S \mid f(x) = \max f(y); y \in S\} \subsetneq S$, $A \neq \emptyset$. $x_0 \in A$ and $y_0 \in A^C$ both exist. There exists a path from $x_0 \to y_0$ because P is irreducible. There exists x, y along this path such that $x \in A$, $y \in A^C$, $P_{xy} > 0$. Now,

$$(Pf)(x) = \sum_{z \in S} P_{xz} f(z) \leq P_{xy} f(y) + (1 - P_{xy}) \max_{z \in S} f(z) < \max z \in S f(z) = f(x)$$

f is not an eigenvector with eigenvalue 1, a contradiction.

Proof for $\geq$: If $dim\{f \mid Pf = f\} >$ number of irreducible components, then $\exists f \mid Pf = f$ and f is non-constant on some irreducible component (by the pigeonhole principle). Reach a contradiction by applying the maximum principle on this component □

11.1 *Coupling*

11.3 DEFINITION (Coupling). If X and Y are random variables then a coupling (X', Y') is a pair of random variables over a single probability space such that X' has the same distribution as X and Y' has the same distribution as Y.

11.4 REMARK. A trivial coupling always exists: take X' and Y' to be independent with appropriate marginals. This is typically not useful.

Often we want a coupling such that some "nice" property is satisfied.

Examples:

1. $X' \leq Y'$ A.S. (Stochastic Domination)
2. Want a coupling that minimizes $\mathbb{P}(X' \neq Y')$.

11.5 DEFINITION (Stochastic Domination). We say that X (first-order) stochastically dominates Y, denoted $X \succeq Y$, if $\forall x \in \mathbb{R}$, $\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x)$.

11.6 THEOREM. *$X \succeq Y$ if and only if there exists a coupling (X', Y') such that $X' \geq Y'$ a.s.*

Proof. Proof left as an exercise. □

11.7 THEOREM. *$X \sim Bin(m, p)$, $Y \sim Bin(n, p)$, $m < n$. Then $X \preceq Y$.*

Proof. Consider the following coupling: Let $Z_i \sim Bernoulli(p)$, $\{Z_i\}$ iid. Let $X' = \sum i = 1^m Z_i$, $Y' = \sum_{i=1}^n Z_i$. We have that $X' \sim X$, $Y' \sim Y$ and $X' \leq Y'$ by construction. □

The idea of stochastic domination is very important to economics. For example, if we wanted to compare two lotteries X and Y, we would choose the one that dominates the other.

Convergence of Markov Chains. $X_0 = x, X_1, X_2, \ldots$. Let $Y_0, Y_1, \ldots$ be a markov chain on S with probability transition matrix P (same as for $\{X_i\}$), and $Y_0 \sim \Pi$ with $\{X_n\}, \{Y_n\}$ independent.

Therefore, $Y_n \sim \Pi$ $\forall n \geq 0$. Now I will couple $\{X_n\}, \{Y_n\}$. Let $T = \min\{n \geq 0 \mid X_n = Y_n\}$,

$$Z_n = \begin{cases} X_n & n < T \\ Y_n & n \geq T \end{cases}$$

We claim that $\{Z_n\}$ is a Markov Chain on S with $Z_0 = x$ and probability transition matrix P. Therefore, $(\{Z_n\}, \{Y_n\})$ is a coupling.

$$
\begin{aligned}
&\mathbb{P}\left(X_n = a\right) \\
=&\mathbb{P}\left(Z_n = a\right) \\
=&\mathbb{P}\left(Z_n = a \mid T > n\right) \cdot (1 - \mathbb{P}\left(T \leq n\right)) + \mathbb{P}\left(Z_n = a \mid T \leq n\right) \cdot \mathbb{P}\left(T \leq n\right) \\
=&\Pi(a) \cdot (1 - \mathbb{P}\left(T > n\right)) + \mathbb{P}\left(Z_n = a \mid T > n\right) \cdot \mathbb{P}\left(T > n\right)
\end{aligned}
$$

so

$$
\begin{aligned}
\mathbb{P}\left(()\, X_n = a\right) - \Pi(a) &= \mathbb{P}\left(T > n\right) \left(\mathbb{P}\left(Z_n = a \mid T > n\right) - \Pi(a)\right) \\
\left|\mathbb{P}\left(X_n = a\right) - \Pi(a)\right| &= \mathbb{P}\left(T > n\right) \left|\mathbb{P}\left(Z_n = a \mid T > n\right) - \Pi(a)\right| \leq \mathbb{P}\left(T > n\right)
\end{aligned}
$$

For all $a \in S$. Assume that $\delta = \min_{x',y'} P_{x'y'} > 0$. Then

$$
\begin{aligned}
&\mathbb{P}\left(()\, T > l + 1 \mid T > l\right) \\
\leq &\max_{x',y',x' \neq y'} \mathbb{P}\left(X_{l+y} \neq Y_{l+1} \mid X_l = x', Y_l = y'\right) \\
=&1 - \min_{x',y',x' \neq y'} \mathbb{P}\left(X_{l+1} = Y_{l+1} \mid X_l = x', Y_l = y'\right) \\
\leq&1 - \min_{x',y',x' \neq y'} \mathbb{P}\left(X_{l+1} = Y_{l+1} = x' \mid X_l = x', Y_l = y'\right) \\
\leq&1 - \min_{x',y',x' \neq y'} P_{x'x'} P_{y'x'} \\
\leq&1 - \delta^2
\end{aligned}
$$

Therefore, $\mathbb{P}\left(T > l\right) \leq (1 - \delta^2)^l$. This shows that eventually X_n and Y_n will meet, and after that point Z_n is stationary.

Ergodicity implies that $\exists M < \infty$ such that $(P^M)_{x',y'} > 0$ for all $x', y' \in S$. That is, there is a path between any two states. $\square$

12 11/8/2018

What is the probability that I could end up in a given irreducible component?

We can compute via first step analysis $X_0, X_1, X_2, \ldots$. $T := \min\{n \geq 0 \mid X_n \text{ is an essential state (first time in an irreducible component)}\}$. $f_i(x) = \mathbb{P}\left(()\, X_T \text{ in component } C_i \mid X_0 = x\right)$. Obviously $f_i(x) = 1$ if x is in C_i, and 0 if x is in an irreducible component other than C_i.

To define $f_i(x)$ for other parts of the graph, we can compute $f_i(x)$ by looking at possible single steps starting at x, then define the function inductively.

12.1 THEOREM (Convergence of Markov Chains). *Let S be a finite state space. $X_0, X_1, \ldots$ an S-valued irreducible Markov Chain with period k. Then $\frac{1}{k}\{\mathcal{L}(X_n) + \mathcal{L}(X_{n+1}) + \cdots + \mathcal{L}(X_{n+k-1})\} \to \pi$ as $n \to \infty$ where π is the unique stationary distribution.*

12.2 THEOREM (Ergodic Theorem for Markov Chains). *Let S be a finite state space. Let $X_0, X_1, \ldots$ be an S-valued irreducible Markov chain. Let $f : S \to \mathbb{R}$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \sum_{x\in S} f(x)\pi(x)$$

almost surely.

12.3 REMARK (Ergodic Principle). Time-averages are like space-averages.

Special Case:

$f(x) = \mathbb{1}_{x=x_0}$ where $x_0 \in S$. $N_n(x) := \sum_{k=1}^{n} \mathbb{1}_{X_k=x}$. Ergodic theorem: $\frac{N_n(x_0}{n} \to \pi(x_0)$ as $n \to \infty$.

That is, fraction of time spent at x_0 converges to the stationary probability of x_0.

Observation: It is enough to prove theorem for such functions, since every function is a linear combination of such functions.

12.4 REMARK. The Ergodic theorem is basically the law of large numbers for Markov chains.

Proof. One approach (sketch):
$\mathrm{Var}(\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{X_k=x_0}) = \frac{1}{n^2}\sum_{i,j=1}^{n} Cov(\mathbb{1}_{X_i=x_0}\mathbb{1}_{X_j=x_0}) \leq$ exercise... $\leq \frac{1}{n^2}Cn = \frac{C}{n}$. This process converges in probability. $\square$

13 11/15/2018

13.1 *Martingales*

13.1 DEFINITION (Martingales). A martingale is a stochastic process with the property that in expectation, it does not change. That is

$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] = X_n$$

In this class we will

1. Define conditional expectation.
2. Introduce martingales.
3. Study their properties.

13.2 *Conditional Expectation in Simple Discrete Setting*

Suppose X, Y are discrete random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

13.2 DEFINITION. Conditional Expectation Let

$$f(y) := \mathbb{E}\left[X \mid Y = y\right] = \frac{\mathbb{E}\left[X\mathbb{1}_{\{y=y\}}\right]}{\mathbb{P}\left(Y = y\right)}$$

then the conditional expectation is defined as

$$\mathbb{E}\left[X \mid Y\right] := f(Y)$$

Importantly, this is a random variable.

Conditional expectation has the following properties

1. Linearity: $\mathbb{E}\left[a_1 X_1 + a_2 X_2 \mid Y\right] = a_1 \mathbb{E}\left[X_1 \mid Y\right] + a_2 \mathbb{E}\left[X_2 \mid Y\right]$.
2. Tower rule (a.k.a. law of Total Expectation): $\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X\right]$. Proof: $\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \sum_y \mathbb{E}\left[X \mid Y = y\right] \cdot \mathbb{P}\left(Y = y\right) = \sum_y \sum_x x \cdot \mathbb{P}\left(X = x \mid Y = y\right) \cdot \mathbb{P}\left(Y = y\right) = \sum_x x \cdot \sum_y \mathbb{P}\left(X = x \mid Y = y\right) \cdot \mathbb{P}\left(Y = y\right) = \sum_x x \cdot \mathbb{P}\left(X = x\right) = \mathbb{E}\left[X\right]$.
3. Independence: If X and Y are independent, then $\mathbb{E}\left[X \mid Y\right] = \mathbb{E}\left[X\right]$.
4. Take out what is known: $\mathbb{E}\left[Xg(Y) \mid Y\right] = g(Y) \cdot \mathbb{E}\left[X \mid Y\right]$.

Now suppose that $g := \sigma(Y) = \{Y^{-1}(B) \mid B \in \mathcal{B}\}$. Let $Z := \mathbb{E}\left[X \mid Y\right]$ and $G_y = \{\omega \in \Omega \mid Y(\omega) = y\}$. Then Z is constant on each G_y, g is generated by the G_y, and Z is (Ω, g)-measurable.

Now let $G \in g$, $G = \bigcup_{y \in I} G_y$. $\mathbb{E}\left[Z \cdot 1_G\right] = \sum_{y \in I} \mathbb{E}\left[Z \cdot 1_{G_y}\right] = \sum_{y \in I} \mathbb{E}\left[X \mid Y = y\right] \cdot \mathbb{P}\left(Y = y\right) = \sum_{y \in I} \sum_x x \cdot \mathbb{P}\left(X = x \mid Y = y\right) \cdot \mathbb{P}\left(Y = y\right) = \mathbb{E}\left[X \cdot 1_G\right]$.

13.3 DEFINITION (Conditional Expectation (General Form)). Suppose that X is an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. That is, $\mathbb{E}\left[|X|\right] < \infty$. Let $g \subset \mathcal{F}$ be a σ-algebra. The conditional expectation $\mathbb{E}\left[X \mid g\right]$ is a g-measurable random variable such that for all $G \in g$ we have that

$$\mathbb{E}\left[\mathbb{E}\left[X \mid g\right] \cdot 1_G\right] = \mathbb{E}\left[X \cdot 1_G\right]$$

Conditional expectation has the following properties:

1. Linearity: $\mathbb{E}\left[a_1 X_1 + a_2 X_2 \mid g\right] = a_1 \mathbb{E}\left[X_1 \mid g\right] + a_2 \mathbb{E}\left[X_2 \mid g\right]$.
2. Tower rule (a.k.a. law of Total Expectation): If $g_1 \subset g_2$, then $\mathbb{E}\left[\mathbb{E}\left[X \mid g_2\right] \mid g_1\right] = \mathbb{E}\left[X \mid g_1\right]$. If $g_1 = \{\emptyset, \Omega\}$, then $\mathbb{E}\left[\mathbb{E}\left[X \mid g_2\right]\right] = \mathbb{E}\left[X\right]$.
3. Independence: If g and $\sigma(X)$ are independent, then $\mathbb{E}\left[X \mid g\right] = \mathbb{E}\left[X\right]$.

4. Take out what is known: If Y is g-measurable and X and XY are integrable, then $\mathbb{E}[XY \mid g] = Y \cdot \mathbb{E}[X \mid g]$. That is, if Y is g-measurable, then it can be pulled out from the conditional expectation.

13.4 THEOREM (Existence and Uniqueness of Conditional Expectation). *Given an integrable random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ and a σ-algebra $g \subset \mathcal{F}$, there exists an integrable random variable Z on $(\Omega, g, \mathbb{P})$ such that for every $G \in g$ we have that* $\mathbb{E}[Z \cdot 1_G] = \mathbb{E}[X \cdot 1_G]$.

Moreover, if this property holds for any $G \in g$ and two random variables Z and $\tilde{Z}$ on $(\Omega, g, \mathbb{P})$, then $\mathbb{P}\left(Z = \tilde{Z}\right) = 1$. That is, any two random variables that hold the earlier defined property, then they are equal almost everywhere.

13.5 REMARK. 1. An important special case is when g is generated by a random variable (or multiple random variables). If $g = \sigma(Y)$ then we often just write $\mathbb{E}[X \mid Y]$ for $\mathbb{E}[X \mid g] = \mathbb{E}[X \mid \sigma(Y)]$.

2. If X is g-measurable, then $\mathbb{E}[X \mid g] = X$.

3. If $g = \{\emptyset, \Omega\}$, the trivial σ-algebra, then $\mathbb{E}[X \mid g] = \mathbb{E}[X]$.

13.6 REMARK (Intuition Behind Conditional Expectation). Suppose that $\mathbb{E}\left[X^2\right] < \infty$ and $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Note that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with inner product $\langle X, Y\rangle = \mathbb{E}[XY]$. So, the norm is $||X||_{L_2} = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{E}[X^2]}$.

Now let $g \subset \mathcal{F}$ be a σ-algebra and define $K := L^2(\Omega, g, \mathbb{P})$, a subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then if $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ but $X \notin K$, we can show that the orthogonal projection to the subspace is $Z = \mathbb{E}[X \mid g]$.

Suppose that $Y \in K$. Then $\mathbb{E}[(X-Z)(Y-Z)] = \mathbb{E}[\mathbb{E}[(X-Z)(Y-Z) \mid g]]$ by the tower rule. But $Y - Z$ is g-measurable so we get $\mathbb{E}[(Y-Z)\mathbb{E}[X - Z \mid g]]$ and $\mathbb{E}[X - Z \mid g] = \mathbb{E}[X \mid g] - Z = 0$ so $(X - Z)$ and $(Y - Z)$ are perpendicular, showing that conditional expectation is an orthogonal projection.

As a refresher, recall the following:

13.7 DEFINITION (Absolute Continuity). Let μ, ν be two measures on some measurable space $(E, \mathcal{E})$. We say that ν is absolutely continuous with respect to μ, writing $\nu << \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0\ \forall A \in \mathcal{E}$.

14 11/27/2018

14.1 *Recap*

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

14.1 DEFINITION (Filtration). $\{\mathcal{F}_{n\geq 0}\}$, an increasing sequence of σ-algebras.

14.2 DEFINITION (Martingale). $\{X_n\}_{n\geq 0}$ is a martingale w.r.t $\{\mathcal{F}_n\}$ if

1. $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$.
2. X_n is integrable $\forall n \geq 0$.
3. $\mathbb{E}[X_{n+m} \mid \mathcal{F}_n] = X_n$.

Property 3 of martingales is the main condition. It is enough to show this for $m = 1$. We called $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ the natural filtration.

14.2 *Martingales*

Example: $X_1, X_2, \ldots$ I.I.D. with $\mathbb{P}(X_1 = +1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$.

$S_n = \sum_{i=1}^n X_n$ a simple symmetric random walk.

$Y_n = S_n^2 - n$.

$\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n) = \sigma(S_0, S_1, \ldots, S_n)$, $g_n = \sigma(Y_0, Y_1, \ldots, Y_n)$. Note that $g_n \subset \mathcal{F}_n$. This is due to the fact that Y_n loses information due to the squaring.

14.3 EXERCISE. Is it true that $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = Y_n$?

Yes. By the tower rule:
$\mathbb{E}[Y_{n+1} \mid g_n] = \mathbb{E}[\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \mid g_n] = \mathbb{E}[Y_n \mid g_n] = Y_n$.

14.3 *Stopping Time*

$(\Omega, \mathcal{F}, \mathbb{P})$, $\{F_n\}_{n \geq 0}$ a filtration.

14.4 DEFINITION (Stopping Time). Suppose that $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is $\mathcal{F}$-measurable. T is a stopping time if $\{\omega \mid T(\omega) \leq n\} \in \mathcal{F}_n$. Note that this means that at any step of the stochastic process, T can tell us if we have stopped.

Equivalently: T is a stopping time if $\{\omega \mid T(\omega) = n\} \in \mathcal{F}_n$. The proof of equivalency is left as an exercise.

Examples:

1. (First) Hitting Times: $\{X_n\}$ an S-valued stochastic process. $\{\mathcal{F}_n\}$ the natural filtration. $A \subseteq S$. $T_A := \inf\{m \geq 0 \mid X_m \in A\}$. That is, the first time the stochastic process "hits" A.
2. Special Case, Return Times: $R := \inf\{m \geq 1 \mid X_m = X_0\}$.

Non-Examples:

1. Last Hitting Times: $\sup\{m \geq 0 \mid X_m \in A\}$. This is a random time, but not a stopping time. To determine if we have stopped, we have to look at all future time steps.
2. Sell Before the Market Drops: $\inf\{m \geq 0 \mid X_{m+1} - X_m < 0\}$.
3. Constant Time before a Hitting Time: $\inf\{m \geq 0 \mid X_m \in A\} - 5$.

14.5 CLAIM. *If T_1, T_2 are stopping times, then $\min\{T_1, T_2\}$ and $\max\{T_1, T_2\}$ are also stopping times. Denote these as $T_1 \wedge T_2$ and $T_1 \vee T_2$ respectively.*

14.4 *Stopped Martingales*

$\{X_n\}_{n\geq 0}$ is a martingale, submartingale, or supermartingale. T a stopping time.

14.6 DEFINITION (Stopped Process). The stopped process $\{X_n^T\}_{n\geq 0}$ is defined as $X_n^T := X_{T\wedge n}$. That is, once the stop time is hit, the process stops.

14.7 THEOREM. *If $\{X_n\}_{n\geq 0}$ is a mg/submg/supermg and T is a stopping time, then $\{X_n^T\}_{n\geq 0}$ is always a mg/submg/supermg with respect to the same filtration.*

Proof.

1. Adaptations: $X_n^T = \sum_{k=0}^{n-1} \mathbb{1}_{T=k} X_k + \mathbb{1}_{T\geq n} X_n$. Both terms are $\mathcal{F}_n$ measurable.
2. Integrability: $|X_n^T| \leq \sum_{k=0}^{n} |X_k|$. $|X_k|$ is integrable $\forall k \geq 0$, and the sum of integrable random variables is also integrable, so $|X_n^T|$ is integrable.
3. Mg property: $\mathbb{E}\left[X_n^T \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\sum_{k=0}^{n-1} \mathbb{1}_{T=k} X_k + \mathbb{1}_{T\geq n} X_n \mid \mathcal{F}_{n-1}\right] = \sum k = 0^{n-1} (\mathbb{1}_{T=k} X_k) + \mathbb{1}_{T\geq n} \mathbb{E}\left[X_n \mid \mathcal{F}_{n-1}\right] = \sum_{k=0}^{n-1} (\mathbb{1}_{T=k} X_k) + \mathbb{1}_{T\geq n} X_{n-1} = \sum_{k=0}^{n-2} (\mathbb{1}_{T=k} X_k) + \mathbb{1}_{T\geq n-1} X_{n-1} = X_{n-1}^T$.

□

14.8 COROLLARY. *If $\{X_n\}_{n\geq 0}$ is a mg. and T is a stopping time then $\mathbb{E}\left[X_{T\wedge n}\right] = \mathbb{E}\left[X_{T\wedge 0}\right] = \mathbb{E}\left[X_0\right]$.*

Assume that $\mathbb{P}\left(T < \infty\right) = 1$. Then $\lim_{n\to\infty} T \wedge n = T$ almost surely.

Question: Is it true that $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_0\right]$.

Answer: No. $\{S_n\}_{n\geq 0}$ a symmetric simple random walk starting at 0. $T = \inf\{m \geq 0 \mid S_m = 1\}$.

14.9 THEOREM (Doob's Optimal Stopping Theorem). *$(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, $\{\mathcal{F}_n\}_{n\geq 0}$ a filtration, $\{X_n\}_{n\geq 0}$ a martingale, and T a stopping time with $\mathbb{P}\left(T < 0\right) = 1$.*

If any of the following conditions hold, then $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_0\right]$.

1. *(Bounded Stopping Time) $\exists N < \infty$ such that $T \leq N$ almost surely.*

2. *(Bounded Process)* $\exists K < \infty$ *such that* $\forall n$, $\mathbb{P}\left(|X_n| \leq K\right) = 1$. *This can be relaxed to* $\forall n$, $\mathbb{P}\left(|X_{T\wedge n}| \leq K\right) = 1$.

3. *(Bounded Increments)* $\mathbb{E}\left[T\right] < \infty$ *and* $\exists K < \infty$ *such that* $\forall n \geq 1$, $\mathbb{P}\left(|X_n - X_{n-1}| \leq K\right) = 1$. *This can be relaxed to* $\mathbb{E}\left[T\right] < \infty$ *and* $\exists K < \infty$ *such that* $\forall n \geq 1$, $\mathbb{P}\left(|X_{T\wedge n} - X_{T\wedge n-1}| \leq K\right) = 1$.

Proof.

1. $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_{T\wedge N}\right] = \mathbb{E}\left[X_{T\wedge 0}\right] = \mathbb{E}\left[X_0\right]$ using the fact that a $T \leq N$ almost surely and a stopped martingale is a martingale.

2. Use the dominated converge theorem with the function K as the dominating function. $|X_{T\wedge n}| \leq K$ almost surely, and K is integrable, so $\mathbb{E}\left[X_0\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{T\wedge n}\right] = \mathbb{E}\left[\lim_{n\to\infty} X_{T\wedge n}\right] = \mathbb{E}\left[X_T\right]$.

3. $|X_{T\wedge n} - X_0| = |\sum_{i=1}^{n}(X_{T\wedge i} - X_{T\wedge(i-1)})| \leq \sum_{i=1}^{n}|X_{T\wedge i} - X_{T\wedge(i-1)}| \leq (T\wedge n)K \leq TK$. So we can use the dominated convergence theorem with $T \cdot K$ as the dominating function.

$\square$