Math 218a Lecture Notes, Fall 2020

Probability Theory

Professor: Shirshendu Ganguly
Vishal Raman

Contents

§1 August 27th, 2020

§1.1 Introduction

Consider a **random experiment** - this involves a state space Ω and some "probability" on it. The outcome of an experiment would be $\omega \in \Omega$.

Example 1.1 (Fair Coin Toss)
$\Omega = \{0, 1\}, P(0) = 1/2, P(1) = 1/2$ models a fair coin toss. The outcomes are $\omega \in \Omega, \omega = 0$ or $\omega = 1$.

Example 1.2 (Continuous State Space)
$\Omega = [0, 1]$, X is the outcome of a random experiment. Suppose X is uniformly distributed random variable. $P(X \in [0, \frac{1}{2}]) = 1/2$. Take $A = \mathbb{Q} \cap [0, 1]$. $P(x \in A) = 0$, since A has no "volume". Similarly, taking $A_1 = \mathbb{R} \setminus \mathbb{Q} \cap [0, 1]$, then $P(x \in A_1) = 1 - P(x \in A) = 1$. Finally, take $E \subset [0, 1]$. $P(x \in E) =$"volume" of E.

The issue: we need to define some notion of volume. Some properties we would like are the following:

- Translation Invariance
- Countable Additivity: $A_1, A_2, \ldots$ disjoint with $A = \bigcup A_i$, then $P(A) = \sum_{i=1}^{\infty} P(A_i)$.

§1.2 Nonmeasurable Sets

Take $I = [-1, 2]$, and define $x \sim y$ iff $x - y \in \mathbb{Q}$. [Exercise: check that $\sim$ is an equivalence relation.] This decomposes I into equivalence classes $I/\sim$. Note that the equivalence classes are countable, since any class is $x + A, A \subset Q$.

For each equivalence class B, pick $x_B \in B \cap [0, 1]$. Define $E = \{x_B\}$ over all the equivalence classes. Note that x_B is a representative of B in E, so $B = \{x_b + q : x_b + q \in I, q \in \mathbb{Q}\}$.

Now, consider the set $[0, 1] \subset \bigcup_{q \in [-1,1]} E + q \subset [-1, 2]$. Equality doesn't hold, because there can be B s. t. x_b is close to 0. Then $E + (\mathbb{Q} \cap [-1, 1])$ will only recover elements of B near 1 and will not go up to 2.

Proposition 1.3
We claim that $E + q$ are disjoint for different values of q.

Proof. Suppose $E + q_1 \cap E + q_2 \neq \emptyset$ for some q_1, q_2. Then, there exists $x, y \in E$ such that $x + q_1 = y + q_2$. This implies that $x - y = q_2 - q_1 \in \mathbb{Q}$, so $x \sim y$, but by definition, there is exactly one member of each equivalence class in E. $\square$

The big question: What is $P(E)$? Suppose $P(E) > 0$. Then $\bigcup_{q \in [-1,1]} E + Q \subset [-1, 2]$ and $P(E + q_1) = P(E + q_2) = P(E)$ for all q_1, q_2. Furthermore, by countable additivity,

$$1 \geq P(\bigcup_{q \in [-1,1] E+q}) = \sum_{q \in [-1,1]} P(E + q) = \infty \cdot P(E).$$

This would imply that $P(E) = 0$. However,

$$[0,1] \subseteq \bigcup_{q \in [-1,1]} E + q \Rightarrow P([0,1]) = 1/3 \leq \sum_{q \in [-1,1]} P(E+q) = 0.$$

Hence, $P(E)$ cannot be defined.

The issue is the step where we pick x_B, since we need to pick x_B from uncountably many points, which assumes the axiom of choice. It was proved by Robert M. Solovay that all models of set theory excluding the axiom of choice have the property that all sets are Lebesgue measurable.

Our goal is thus to come up with a general framework where things can be consistently defined for a large class of sets.

§1.3 Measure Theory Beginnings

For the definitions, we take Ω to be the state space.

Definition 1.4 (Sigma-Algebra)**.** Suppose Σ follows the following properties:

1. $\emptyset \in \Sigma$
2. $A \in \Sigma \Rightarrow A^c \in \Sigma$
3. $A_1, A_2, \cdots \in \Sigma$, then $\bigcup A_i \in \Sigma$

Note that 2 and 3 imply 1 since $(A \cup A^c)^c = \emptyset$. Then Σ is a sigma-algebra.

Note that we also have countable intersections(this is an easy exercise).

§2 September 1st, 2020

Last time:

- We discussed the notation of a Σ-algebra, a reasonable class of sets on which we will define measures.
- Properties: $\emptyset \in \mathcal{A}, A \in \mathcal{A}, A^c \in \mathcal{A}, \bigcup A_i \in \mathcal{A}$.

§2.1 Measures

We are working in a space (Ω, Σ).

Definition 2.1 (Measure)**.** A measure is a function $\mu : \Sigma \to [0, \infty]$ with the following properties:

- $\mu(\emptyset) = 0$
- "Countable Additivity": $\mu(\bigcup A)i) = \sum \mu(A_i)$ for disjoint $A_i \in \Sigma$.

Example 2.2

If Ω is finite, $1, 2, \ldots, n$, $\Sigma = 2^\Omega$, then all possible measures on (Ω, Σ) are given by fixing $a_1, a_2, \ldots, a_n \in [0, \infty]$ and $\mu(A) = \sum_{i \in A} a_i$.

Properties of measures:

- Monotonicity: $A \subset B$, then $\mu(A) \leq \mu(B)$.

  Proof. $B = A \cup (B \setminus A)$ and $B \setminus A \in \Sigma$, so
  $$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A).$$
  □

- Countable Subadditivity: $A \subseteq \bigcup_{i=1}^\infty B_i$, then $\mu(A) \leq \sum \mu(B_i)$.

  Proof. We disjointify the B_i: Define $C_1 = B_1, C_i = B_i \setminus B_{i-1}$. Then
  $$\mu(A) \leq \mu(\bigcup C_i) = \sum \mu(C_i) \leq \sum \mu(B_i).$$
  □

- : Continuity from below: If $A_i \uparrow A$, then $\mu(A_i) \uparrow \mu(A)$.

  Proof. $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \ldots$, so by countable additivity
  $$\mu(A) = \sum_{i=1}^\infty \mu(C_i) = \lim_{n \to \infty} \sum_{i=1}^n \mu(C_i) = \lim_{n \to \infty} \mu(A_n).$$
  □

- Continuity from above, if $A_i \downarrow A$, and $\mu(A_1) < \infty$, then $\mu(A_i) \to \mu(A)$

  Proof. We need the condition $\mu(A_i) < \infty$. Take $A_i = [i, \infty)$ as a counterexample if we don't have that condition.

  Define $A_1 \setminus A_i = B_i$, so $B_i \uparrow A_1 \setminus A$. Then, use the continuity from below. □

§2.2 Sigma algebras

Fact 2.3. For any $A \subset 2^\Omega$, define

$$\Sigma(A) = \bigcap_{A \in \Sigma} \Sigma.$$

Then, $\Sigma(A)$ is a sigma-algebra.

Note that $\Sigma(A)$ is the smallest sigma-algebra containing A. For this reason, we call it the sigma-algebra **generated** by A.

Example 2.4

Take $X, Y \subset 2^\Omega$. We want to prove $\Sigma(X) = \Sigma(Y)$. It suffices to show $X \subseteq \Sigma(Y)$ and $Y \subseteq \Sigma(X)$.

Definition 2.5 (Borel Sigma-Algebra)**.** $(\Omega, \mathcal{U})$, a topological space with a family of open sets. The **Borel Sigma-Algebra** is $\mathcal{B} = \Sigma(\mathcal{U})$.

Example 2.6

For $\Omega = \mathbb{R}$, $\mathcal{B}$ is the sigma algebra generated by open sets in $\mathbb{R}$. We also have $\mathcal{B}$ is the sigma-algebra generated by open intervals in $\mathbb{R}$, which follows from the fact that any open set can be written as a countable union of open intervals. Furthermore,

$$\Sigma((a, b) : a, b \in \mathbb{Q}, \mathbb{R}) = \Sigma([a, b] : a, b \in \mathbb{Q}, \mathbb{R}),$$

since $[a, b] = \bigcap (a - 1/n, b + 1/n)$ and $(a, b) = \bigcup [a + 1/n, b - 1/n]$.

§2.3 Uniform Measure on the Borel Sets

We will attempt to define the uniform measure on Borel sets of $\mathbb{R}$. Broadly, we do it as follows:

1. Define it on a semi-algebra containing the intervals.
2. Extend the definition to an algebra.
3. Extend it to a sigma-algebra.

Definition 2.7 (Semi-algebra)**.** $\Sigma \subset 2^\Omega$ is a semi-algebra if

- $A_1, A_2 \in \Sigma$ implies $A_1 \cap A_2 \in \Sigma$
- $A_1 \in \Sigma$ implies that $A_1^c = \bigcup_{i=1}^n B_i$ for $B_i \in \Sigma$.

Note: The set of intervals $\{(a, b) : a, b \in \mathbb{R}\}$ is not a semi-algebra. If $(a, b)^c = [b, \infty)$ which is not finitely coverable by disjoint open sets. Similarly, $\{[a, b] : a, b \in \mathbb{R}\}$ is not a semi-algebra.

Claim: $\Sigma = \{(a, b] : a, b \in \mathbb{R}\}$ is a semi-algebra. [This is left as an exercise].

Now, $\mu((a, b]) = b - a$. The proof that μ is countable additive on Σ. If $A = \bigcup_{i=1}^\infty B_i$, B_i disjoint, $A, B_i \in \Sigma$, then $\mu(A) = \sum_{i=1}^\infty \mu(B_i)$.

Proof. We first show that $\mu(A) \geq \sum_{i=1}^{\infty} \mu(B_i)$. This is an easy exercise, show $\mu(A) \geq \sum_{i=1}^{n} \mu(B_i)$, and we pass to the limit.

It suffices to show $\mu(A) \leq \sum_{i=1}^{\infty} \mu(B_i)$. We do this by exploiting compactness.

Let $A = (a, b] \supset [a + 1/n, b] = A;$, take $B_i = (c_i, d] \subset c_i, d + \frac{\epsilon}{2^i} = B_i'$. Note that

$$A' \subset \bigcup_{i=1}^{\infty} B_i',$$

so there exists a finite subcover $A' \subset \bigcup_{j=1}^{k} B_{i_j}'$. It is easy to show that $b - (a + 1/n) \leq \sum_{j=1}^{k} (d_{i_j}' - c_{i_j}')$. But note that

$$\sum_{j=1}^{k} (d_{i_j}' - c_{i_j}') \leq \sum_{j=1}^{k} d_{i_j} - c_{i_j} + \epsilon,$$

which implies that

$$\mu(A) - 1/n \leq \sum_{i=1}^{\infty} \mu(B_i) + \epsilon \Rightarrow \mu(A) \leq \sum_{i=1}^{\infty} \mu(B_i).$$

□

Definition 2.8. $\mathcal{A}$ is an algebra if

- $\emptyset \in \mathcal{A}$
- $A_1 \in \mathcal{A}$ implies $A^c \in \mathcal{A}$
- $A_1, \ldots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^{n} A_i \in A$.

The algebra generated by a semi-algebra is given by taking all possible disjoint finite unions.

Claim: $\Sigma_a = \{\bigcup_{i=1}^{n} A_i\}$ for disjoint A_i semialgebras is an algebra.

Proof. We show $A, B \in \Sigma_a \Rightarrow A \cup B \in \Sigma_a$ and $A^c \in \Sigma_a$. Note that $A = \bigcup_{i=1}^{n} C_i, B = \bigcup_{j=1}^{k} D_j$, so

$$A \cap B = \bigcup_{i=1}^{n} \bigcup_{j=1}^{k} C_i \cap D_j,$$

and $C_i \cap D_j$ are disjoint. Then $C_i, D_j \in \Sigma$ implies $C_i \cap D_j \in \Sigma$.

Then, if $A = \bigcup_{i=1}^{k} C_i$, then $A^c = \bigcap_{i=1}^{k} C_i^c$, and $C_i^c = \bigcup_{j=1}^{\ell} E_j \in \Sigma_a$. □

We extend μ to an algebra by $\mu(A) = \sum_{i=1}^{k} \mu(C_i)$, where $A = \bigcup C_i$ in the semi-algebra.

§3 September 3rd, 2020

Recall that we are aiming to define the uniform measure on $(\mathbb{R}, \mathcal{B})$.

Last time:

1. We defined a **premeasure** on a semi-algebra, which $\Sigma_{semi} = \{(a, b] : -\infty \le a \le b \le \infty\}$, $\mu((a, b]) = b - a$ was countably additive.

2. Extend μ to an algebra $\Sigma_a =$ disjoint union of elements of Σ_{semi}.

3. For $A = \bigcup_{i=1}^k C_i \in \Sigma_a$,
$$\mu(A) = \sum_{i=1}^k \mu(C_i).$$

§3.1 Uniform Measure on the Borel Sets

We first need show show μ is well defined. Suppose $A = \bigcup_{i=1}^l C_i, \bigcup_{j=1}^\ell B_j$ for $C_i, B_i \in \Sigma_{semi}$. We want
$$\mu(A) = \sum_{i=1}^k \mu(C_i) = \sum_{j=1}^\ell \mu(B_j).$$

Note that $C_i = \bigcup_{j=1}^\ell (C_i \cap B_j)$, which are all disjoint. Similarly, $B_j = \bigcup_{i=1}^k (B_j \cap C_i)$, disjoint. Thus, from the finite additivity of Σ_{semi}, we have
$$\sum_{i=1}^k \mu(C_i) = \sum_{i=1}^k \sum_{j=1}^\ell \mu(C_i \cap B_k) = \sum_{j=1}^\ell \mu(B_j),$$
as desired.

We will next show that μ is finitely additive additive. First, if we have $A_1, A_2, \ldots, A_n \in \Sigma_a$ disjoint, we show $\mu\left(\bigcup A_i\right) = \sum \mu(A_i)$.

Note that each $A_i = \bigcup_{j=1}^{m_i} C_j^i$, which are disjoint, so
$$\mu\left(\bigcup A_i\right) = \mu\left(\bigcup_i \bigcup_j C_j^i\right) = \sum_{i=1}^n \sum_{j=1}^{m_i} \mu(C_j^i) = \sum_{i=1}^n \mu(A_i).$$

Next, we show μ is monotonic. For $A, B \in \Sigma_a$, $A \subseteq B$, $B = A \cup (B \setminus A)$, so
$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$

We show countably additivity: Let $A = \bigcup_{i=1}^\infty \mu(A_i)$. We need to show $\mu(A) = \sum_{i=1}^\infty \mu(A_i)$.

We first show $\mu(A) \ge \sum_{i=1}^\infty \mu(A_i)$. It suffices to show $\mu(A) \ge \sum_{i=1}^n \mu(A_i)$. Since $\bigcup_{i=1}^n A_i \subseteq A$, monotonicity gives $\mu\left(\bigcup_{i=1}^n A_i\right) \le \mu(A)$.

Next, we show $\mu(A) \le \sum_{i=1}^\infty \mu(A_i)$. First $A = \bigcup_{j=1}^k C_j$ for $C_j \in \Sigma_{semi}$, $A_i = \bigcup_\ell^{m_i} C_\ell^i$ for $C_\ell^j \in \Sigma_{semi}$.

Thus,
$$\mu(A) = \sum_{j=1}^k \mu(C_j).$$

Hence, it suffices to show $\mu(C_j) \le \sum_{i=1}^{\infty} \mu(C_j \cap A_i)$, since

$$\mu(A) = \sum \mu(C_j) \le \sum_{j=1}^{k} \sum_{i=1}^{\infty} \mu(C_j \cap A_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{k} \mu(C_j \cap A_i).$$

Note that $C_j = \bigcup_{i=1}^{\infty} \bigcup_{\ell}^{m_i} C_j \cap C_\ell^i$, and we finish by using countable additivity for Σ_{semi}.

It suffices extend to $\Sigma(\Sigma_a)$ which is the sigma-algebra generated by Σ_a.

Theorem 1 (Caratheodory's Extension Theorem)

We have the following:

- Given a countably additive measure μ on an algebra Σ_a, it can be extended to a measure on $\Sigma(\Sigma_a)$.
- If μ is σ-finite on Σ_a, the extension is unique.

A measure μ is σ-finite on Σ_a if there exists $A_1 \subseteq A_2 \subseteq \cdots \in \Sigma_a$ so that $\bigcup A_i = \Omega$, $\mu(A_i) \le \infty$ for all i.

Proof. For example, consider $\Sigma_{semi} = \{(a, b] \cap \mathbb{Q}\}$. Then $\Sigma = 2^{\mathbb{Q}}$. The cardinality of every element in Σ_{semi} is either ∞ or 0. Define $\mu(A) = \infty$ if $|A| = \infty$, else 0. One can check μ is a measure on Σ_{semi}. We can also take the counting measure ν. This agrees on Σ_{semi}, but not on Σ. We can check that ν is not sigma-finite.

We now show uniqueness, given σ-finiteness. For simplicity, assume $\mu(\Omega) < \infty$. If we have two measures μ_1, μ_2 on Σ with $\mu_1(A) = \mu_2(A)$ for all $A \in \Sigma_a$, then we show $\mu_1(B) = \mu_2(B)$ for all $B \in \Sigma$.

A general strategy to show some property is true for a sigma-algebra is to show that the sets satisfying those properties must be closed under some natural operations and that any such family of sets must contain a sigma-algebra.

Theorem 2 (π - λ)

A class of sets P is said to be a π-system if $A, B \in P$ implies $A \cap B \in P$. A class of sets G is said to be a λ-system if $\Omega \in G$, $A \subset B, A, B \in G$, then $B \setminus A \in G$, and $A_i \in G$, $A_i \uparrow A \to A \in G$. If P is a π system contained in G, a λ-system, then $\Sigma(P) \subset G$.

Note that a semi-algebra is a π system. It suffices to consider the set $\mathscr{C} = \{A : \mu_1(A) = \mu_2(A)\}$. We know that $\Sigma_{semi} \subset \mathscr{C}$. To show $\Sigma \subset \mathscr{C}$, it suffices to show that $\mathscr{C}$ is a λ-system.

Note that a sigma-algebra is a λ-system, so given any π-system P, $\Sigma(P)$ is the smallest λ-system containing P.

We have already verified $\Sigma_a \subset \mathscr{C}$. Furthermore, $\Omega \in \mathscr{C}$ because Ω is an algebra. Finally, suppose we have $A \subset B$, $A, B \in \mathscr{C}$. We need $B \setminus A \in \mathscr{C}$. $\mu_1(A) = \mu_2(A), \mu_1(B) = \mu_2(B)$ and $\mu(\Omega) < \infty$. Since $\mu_1(\Omega) = \mu(\Omega) = \mu_2(\Omega) < \infty$,

$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A).$$

For $A_i \uparrow A$, by continuity from below, $\mu_1(A_i) \to \mu_1(A), \mu_2(A_i) \to \mu_2(A)$, so $A \in \mathscr{C}$. $\square$

An easy exercise is to modify the above prove to include the sigma-finite case. The proof of the $\pi - \lambda$ theorem involves some set theory.

We'll sketch the existence proof. Suppose we have μ on Σ_a. For example, take $B \subset \mathbb{R}$. How do we define $\mu(B)$? We could try to approximate B by the union of intervals.

Define the outer measure, $\mu_*(B) = \inf \sum_{i=1}^{\infty} \mu(A_i)$ defined over covers of B. Some properties are $\mu_*(B_1) \leq \mu_*(B_2)$ if $B_1 \subseteq B_2$, $\mu_*(\emptyset) = 0$, and $\mu_*(\bigcup C_i) \leq \sum_{i=1}^{\infty} \mu_*(C_i)$.

Define $\mathcal{A} = \{A : \mu_*(E) = \mu_*(E \cap A) + \mu_*(E \cap A^C) \forall E\}$. $\mathcal{A}$ is a sigma algebra containing Σ_a and μ_* is a measure when restricted to $\mathcal{A}$.

§4 September 8th, 2020

Last time, we completed the construction of the uniform measure on the Borel sets.

§4.1 The Outer Measure

On an algebra Σ_a, let Σ_a^σ be the elements formed by taking countable unions of elements of Σ_a. Let $\Sigma_a^{\sigma\delta}$ contain countable intersections of elements of Σ_a^σ. Notice that from the definition of an outer measure, for any set B, there exists a set $B' \in \Sigma_a^{\sigma\delta}$ such that

$$B \subset B', \mu_*(B) = \mu_*(B').$$

This implies that $\mu_*(B' \setminus B) = 0$ and for every N such that $\mu_*(N) = 0$, we can check that N belongs to $\mathcal{A}$. Remark: The construction defines the measure μ_* on sets of the form $A \cup B$, where A is a Borel set and $\mu_*(B) = 0$. It is easy to check that $\mu_*(B) = 0$ implies that there exists a Borel set C such that $\mu_*(C) = 0$ and $B \subseteq C$. Thus, we call it the completion of Borel sets. This is a strictly larger sigma-algebra than the Borel sets, which follows from comparing cardinalities. Namely, the cardinality of the Borel sets is $2^{\aleph_0}$. Observe the Lebesgue sigma algebra contains $2^{\text{Cantor Set}} = 2^C$.

Meausres on the real line are characterized by distribution functions, which are non-decreasing right continuous functions F. One can adopt the same strategy to define a measure on the real line by defining $\mu((a, b]) = F(b) - F(a)$. Similarly, given μ on $(R, B(R))$, we can check that $F(b) = \mu((-\infty, b])$ is a distribution function.

We can also consider $(\mathbb{R}^d, B(\mathbb{R}^d))$, the Borel sets on $\mathbb{R}^d$. We claim that this is

$$\Sigma((a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d)) = \Sigma(B_1, B_2, \ldots, B_d : B_i \in B(R)).$$

For distribution functions on $\mathbb{R}^d$, consider the lexigraphical partial order. We would like them to satisfy,

- $F(x)$ is non-decreasing
- $F(x)$ is right continuous: If $x_i \downarrow x$, then $F(x_i) \to F(x)$.
- $F(x) \to 0$ as $x \downarrow -\infty$, $F(x) \to 1$ as $x \to \infty$.

The properties above are not actually enough to define a measure. (Consider the semibox in $\mathbb{R}^2$).

However, for any F such that $F(A) \geq 0$ for any $A \in \Sigma_{semi}$, the strategy to build a measure on $B(\mathbb{R}^d)$ from $\Sigma_{semi} \to \Sigma_a \to B(R^d)$ works.

§4.2 Functions Between Measure Spaces

Suppose we have two measure spaces $(\Omega_1, \Sigma_1), (\Omega_2, \Sigma_2)$ and a function $f : \Omega_1 \to \Omega_2$.

Definition 4.1 (Measurable Function)**.** f is said to be **measurable** if $f^{-1}(A_2) \in \Sigma_1$ for all $A \in \Sigma_2$. If $(\Omega_2, \Sigma_2) = (\mathbb{R}, B(\mathbb{R}))$, then f will be called a **random variable**.

Proposition 4.2

If $\Sigma_2 = \Sigma(\mathcal{A})$, then to check f is measurable, it suffices to check $f^{-1}(B) \in \Sigma_1$ for all $B \in \mathcal{A}$.

Proof.

$$\Sigma' = \{B : f^{-1}(B) \in \Sigma_1\}$$

is a sigma-algebra: If $B \in \Sigma'$, then $B^c \in \Sigma'$, since $f^{-1}(B^c) = (f^{-1}(B))^c$. $\Omega \in \Sigma'$ since $f^{-1}(\Omega_2) = \Omega_1$. It is easy to check countable unions. Σ' is a sigma algebra containing $\mathcal{A}$, so Σ' contains $\Sigma(\mathcal{A})$. □

Fact 4.3. If we have $f : (\Omega_1, \Sigma_1, \mu_1) \to (\Omega_2, \Sigma_2)$, f induces a measure μ_2 on (Ω_2, Σ_2) where $\mu_2(B) = \mu(f^{-1}(B))$ for all $B \in \Sigma_2$.

Some properties:

- If we have $f : (\Omega_1, \Sigma_1) \to (\Omega_2, \Sigma_2), g : (\Omega_2, \Sigma_2) \to (\Omega_3, \Sigma_3)$, then $h = g \circ f$ is measurable.
- If X_1, X_2 are two random variables, then (X_1, X_2) is a measurable function into $(\mathbb{R}^2, B(\mathbb{R}^2))$.

  We know that $B(\mathbb{R}^2) = \Sigma(I_1 \times I_2)$ for intervals. Finally,

  $$(X_1, X_2)^{-1}(I_1 \times I_2) = X_1^{-1}(I_1) \cap X_2^{-1}(I_2)$$

  , so$(X_1, X_2)^{-1}(I_1 \times I_2)$ is measurable.
- Suppose F is a continuous function from $(\Omega_1, B(\Omega_1)) \to (\Omega_2, B(\Omega_2))$. Then F is measurable, since the preimage of open sets is open.
- if $X_1, X_2, \ldots, X_d$ is a random variable, then $X_1 + X_2 + \cdots + X_d$ is a random variable.
- If f_n are random variables and $f_n \to f$ pointwise. Then, f is measurable.

  Proof. Consider the set $\{f > x\} = \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \{f_m > x\}$ is measurable. Then $(x, \infty)_{x \in \mathbb{R}}$ is a generating set. □
- $X : (\Omega_1, \Sigma_1, \mu_1) \to (R, B(R))$ induces a measure μ on $(R, B(R))$, where $\mu(B) = \mu_1(X^{-1}(B))$ for all $B \in B(R)$. It also induces a distribution function F, which is nondecreasing, right continuous, and $F(x) \uparrow 1$ and $x \to \infty$, $F(x) \downarrow 0$ as $x \to -\infty$.

Given a distribution function, is there a random variable? Given a distribution function, we can construct a measure on $\mathbb{R}$ by the Caratheodory Extension Theorem. Let $I : (R, B(R), \mu) \to (R, B(R))$. We could also take $X : ([0,1], B([0,1], \mu) \to (R, B(R))$, where μ is uniform. Suppose F is continuous and strictly monotone. We want X to induce the distribution F, so it suffices to show $X^{-1}(\infty, y) = [0, F(y)]$. If we define $X(F(y)) = y$,, we get the above, but that's not always well defined. For general distributions, one can come up with various definitions of an "inverse" which induces the desired properties. One particular choice is

$$w \in (0,1), X(\omega) = \sup\{y : F(y) < \omega\}.$$

It is an exercise to check that $\{\omega : X(\omega) < x\} = \{\omega : \omega \leq F(x)\}$, which implies that X is measureable.

§5 September 10th, 2020

§5.1 Integration

We work in (Ω, Σ, μ) a sigma-finite measure space. Often, we take it to be a probability measure. The goal is to define a notion of integration for measurable functions and the behavior of integration with limits. Consider a function $f : \Omega \to \mathbb{R}$ measurable with the Borel Sigma-algebra.

Our strategy is as follows:

1. Consider simple functions.

2. Extend to bounded functions.

3. Extend to general functions.

§5.2 Simple Functions

Definition 5.1 (Simple Function)**.** Consider $f = 1_E$, where $\mu(E) < \infty$, an indicator function. A **simple function** is a linear combination of indicator functions,

$$f = \sum_{i=1}^{k} c_i 1_{A_i}, \mu(A_i) < \infty,$$

where A_i are disjoint.

We'll define the integral of a simple function as

$$\int f d\mu = \sum_{i=1}^{k} c_i \mu(A_i).$$

First, note that any $f = \sum_{i=1}^{k} d_i 1_{B_i}$ can be represented as $\sum_{i=1}^{k} c_i 1_{A_i}$, where they are disjoint. We verify that our definition is well defined. Suppose

$$f = \sum_{i=1}^{k} c_i 1_{A_i} = \sum_{j=1}^{m} e_j 1_{F_j}.$$

Then, observe that for i, j such that $A_i \cap F_j \neq \emptyset$, then $c_i = e_j$. So, we have $f = \sum_{i,j} d_{i,j} 1_{A_i \cap F_j}$, and we can check that $d_{ij} = c_i = e_j$. Thus,

$$\sum_{i=1}^{k} c_i \mu(A_i) = \sum_{i,j} d_{ij} \mu(A_i \cap F_j) = \sum_{j=1}^{m} e_j \mu(F_j).$$

Some properties:

- If $f \geq 0$ then $\int f \geq 0$.
- $\int af = a \int f$.
- $\int (f + g) = \int f + \int g$.
- If $g \leq f$ then $\int g \leq \int f$.
- if $g = f$, $\int g = \int f$.
- $|\int f| \leq \int |f|$.

§5.3 Bounded Functions

Suppose $|f| \leq M$ and f vanishes outside E and $\mu(E) < \infty$.

We can approximate from above or below:

$$\sup \int_{g \leq f} g \leq \inf \int_{h \geq f} h$$

To prove equality, it suffices to show that there exists g, h such that $\int h - \int g \leq \epsilon$. It suffices to construct h such that $h - f < \epsilon$ and $f - g < \epsilon$. Then $\int h - \int g \leq 2\epsilon\mu(E)$.

Note that the range of f is $[-M, M]$, so we can discretize the interval into smaller intervals $I_1, \ldots, I_k$ of size ϵ.

Then, define $A_1 = f^{-1}(I_1) \cap E$, and $f^{-1}(I_j) \cap E = A_j$. Then.

$$h = \sum ((j+1)\epsilon - M)1_{A_j}, g = \sum j\epsilon - M)1_{A_j}.$$

Thus, $\int f = \sup \int g = \inf \int h$.

Observe that the new definition agrees with the old definition when f is simple.

As an exercise, we'll verify $\int f + g = \int f + \int g$. Take Suppose $\int h_1 \geq \int f \geq \int g_1$ with $\int h_1 - g_1 < \epsilon$, and $\int h_2 \geq \int g \geq \int g_2$ with $\int h_2 - g_2 < \epsilon$.

Then

$$h_1 + h_2 \geq f + g \geq g_1 + g_2,$$

and

$$\int h_1 + \int h_2 < \int g_1 + g_2 + 2\epsilon = \int g_1 + \int g_2 + 2\epsilon.$$

§5.4 General Functions

Assume $f \geq 0$. Note that we can no longer approximate from above, so we approximate from below:

$$\int f = \sup\{\int h, h \leq f, \text{ bounded }\}.$$

Clearly, the definition agrees with the old one for bounded functions with finite support.

As an exercise, we'll prove that $\int f + g = \int f + \int g$. If we have a bounded $h_1 \leq f, h_2 \leq g$, then $h_1 + h_2 \leq f + g$, which implies that $\int h_1 + h_2 = \int h_1 + \int h_2 \leq \int f + g$. and $\sup \int h_1 + \sup \int h_2 = \int f + \int g$, so $\int f + \int g \leq \int f + g$.

> **Lemma 5.2**
>
> Suppose $E_n \uparrow \Omega$, $\mu(E_n) < \infty$. Now, consider $(f \wedge n)1_{E_n}$, where $f \wedge n$ is the minimum of f, n. Then, the function is bounded and has finite support. Note that $(f \wedge n)1_{E_n}$. We claim that
>
> $$\int (f \wedge n)1_{E_n} \uparrow \int f.$$

Proof. It is clear that

$$\lim \int (f \wedge n)1_{E_n} \leq \int f,$$

since $h_n = (f \wedge n)1_{E_n}$ is contained in the set of bounded functions for which the supremum is $\int f$.

If suffices to show that $\lim \int (f \wedge n)1_{E_n} \geq \int f$. Take h bounded such that $\int f < \int g + \epsilon$. There is a set E such that $g \leq M$ on E and g is 0 on E^c.

$g \le f$, so for any $n \ge M$, $h_n \ge g$ on $E_n \cap E$. We claim that

$$\int h_n \ge \int g - M\mu(E \setminus E_n).$$

Then $E_n \uparrow \Omega$, so $\mu(E \setminus E_n) \to 0$. □

Now, we conclude the original proof. Note that $\int f + g = \lim \int ((f+g) \wedge n) 1_{E_n}$, so

$$\int ((f+g) \wedge n) 1_{E_n} \le \int (f \wedge n) 1_{E_n} + \int (g \wedge n) 1_{E_n}.$$

Taking limits gives the desired result.

§5.5 Arbitrary Measurable Functions

Define $\int f$ only when $\int |f| < \infty$. Define

$$f = f^+ - f^-,$$

where $f^+ = f \vee 0, f^- = f \wedge 0$.

Then

$$\int f = \int f^+ - \int f^-.$$

Lemma 5.3

If f_1, f_2 nonnegative and $f = f_1 - f_2$, then

$$\int f = \int f_1 - \int f_2.$$

Proof. $f = f^+ - f^- = f_1 - f_2$, so

$$f^+ + f_2 = f_1 + f^-,$$

then

$$\int f^+ + \int f_2 = \int f_1 + \int f^-,$$

so

$$\int f = \int f^+ - \int f^- = \int f_1 - \int f_2.$$

□

§6 September 15th, 2020

§6.1 Properties of Integrals with Limits

We assume for simplicity that our measure space (Ω, Σ, μ) is finite. Last time, we defined integrals for measurable functions starting with indicators, to simple functions, to non-negative functions, and finally to general functions.

Observe that if f is 0 almost surely, then $\int f = 0$. Suppose $\{f_n\}$ is a set of measurable functions and $f_n \to f$ pointwise almost everywhere, then $\lim f_n$ is measurable. In other words, there exists a set E such that $\mu(E) = 0$ and f_n converges on E^c. Define f to be 0 on E and $\lim f_n$ on E^c. Note that f is measurable. Suppose f_n "converge" to f. When can one expect $\int f_n$ to converge to $\int f$?

Definition 6.1 (Convergence in Measure)**.** We say $f_n \to f$ in measure if given $\epsilon > 0$,

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

We denote this by $f_n \xrightarrow{\mu} f$.

Exercise 6.2. If $\mu(\Omega) < \infty$, then $f_n \to f$ almost everywhere implies that $f_n \xrightarrow{\mu} f$.

Example 6.3

Suppose $f_n = 1_{[-n,n]}$ over $\mathbb{R}$. Then $f_n \to f = 1_{\mathbb{R}}$, but $\mu(|f_n - f| > \epsilon) = \infty$.

Recall continuity of measure from below: If $A_n \uparrow A$ then $\mu(A_n) \uparrow \mu(A)$, but if $\mu(A) = \infty$, then $\mu(A) - \mu(A_n) = \infty$ for all n. This doesn't happen for $\mu(\Omega) < \infty$.

Example 6.4

If $f_n \xrightarrow{\mu} f$, then does $f_n \to f$ almost everywhere? No: Take $\Omega = [0,1]$, $f = 0$ and $f_1 = 1_{[0,1/n]}, f_2 = 1_{[1/n,2/n]}, \ldots, f_n = 1_{[n-1/n,1]}, f_{n+1} = 1_{[0,1/(n+1)]}, \cdots$

There is always some interval where $f_n = 1$, so it does not converge pointwise to 0.

Example 6.5

If $f_n \xrightarrow{\mu} f$, then does $\int f_n \to \int f$? No. Take $f_n = \frac{1}{n}1_{[0,n]}$.

Theorem 3 (Bounded Convergence Theorem)

Suppose μ is finite and f_n are supposed on E such that $\mu(E) < \infty$.

If $|f_n| < M$ and $f_n \xrightarrow{\mu} f$, then $\int f_n \to \int f$.

Proof. f must be 0 almost everywhere outside E. Define $F = \{|f_n - f| < \epsilon\}$. Note that

$$\begin{aligned} \left|\int_E f_n - \int_E f\right| &\le \int_{F\cap E} |f_n - f| + \int_{E\cap F^c} |f_n - f| \\ &\le \epsilon\mu(E) + 2M\mu(F^c \cap E) \xrightarrow{\epsilon\to 0} 0. \end{aligned}$$

□

Theorem 4 (Fatou's Lemma)
If $f_n \geq 0$ then
$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \left(\liminf_{n\to\infty} f_n\right) d\mu.$$

Proof. Let $g_n(x) = \inf_{m\geq n} f_m(x)$. Then $g_n(x) \uparrow g = \liminf f_n$.

We know that $f_n \geq g_n$. This implies that $\int f_n \geq \int g_n$. We have that
$$\liminf \int f_n \geq \liminf \int g_n = \lim \int g_n.$$

Hence, It suffices to show that
$$\lim \int g_n \geq \int g.$$

Recall that $g_n \uparrow g$ so $\lim g_n = g$. Consider $g_n \wedge m$, a bounded function. Note that $g_n \wedge m \uparrow g \wedge m$, so by the Bounded Convergence Theorem,
$$\int (g_n \wedge m) \uparrow \int (g \wedge m).$$
Furthermore, we have
$$\int g_n \geq \int (g_n \wedge m),$$
so
$$\lim \int g_n \geq \lim \int (g_n \wedge m) \uparrow \int (g \wedge m) \uparrow \int g,$$
where the last inequality comes from approximation by bounded functions of finite support. □

Theorem 5 (Monotone Convergence Theorem)
If $f_n \geq 0$ and $f_n \uparrow f$, then $\int f_n \uparrow \int f$.

Proof. Note that $\int f \geq \lim_{n\to\infty} \int f_n$ since $\int f \geq \int f_n$. Then $\int f \leq \lim_{n\to} \int f_n$ by Fatou's lemma. □

Theorem 6 (Dominated Convergence Theorem)
If $|f_n| \leq g$ where $\int g \leq \infty$ and $f_n \to f$ pointwise, then
$$\int f_n \to \int f.$$

Proof. Note that $f_n + g \geq 0$, and $f_n + g \to f + g$ so by Fatou's lemma,
$$\liminf_{n\to\infty} \int f_n + g \geq \int f + g d\mu,$$
which implies that $\liminf_{n\to\infty} \int f_n \geq \int f$.

Then, applying the result to $g - f_n$, we have
$$\liminf_{n\to\infty} -f_n \geq \int -f \Rightarrow \limsup_{n\to\infty} f_n \leq \int f,$$
which implies that $\lim \int f_n = \int f$, as desired. □

§6.2 Expected Value

We have been discussing measurable functions, but these can easily be translated into the language of random variables. Namely, if X is a random variable and $\int |X| < \infty$, then $\int X = E(X)$, the expectation of X.

- $X_n \geq 0$, then $X_n \uparrow X \to E(X_n) \to E(X)$.
- $|X_n| < Y$, $X_n \to X$, $E(Y) < \infty$, then $E(X_n) \to E(X)$.

§6.3 Change of measure for Integrals

We have a random measurable map

$$X : (\Omega_1, \Sigma_1, \mu_1) \to (\Omega_2, \Sigma_2) \xrightarrow{f} (\mathbb{R}, B(\mathbb{R})).$$

Then $f \circ X : (\Omega_1, \Sigma_1) \to (\mathbb{R}, B(\mathbb{R}))$, and if $\int |f \circ X| < \infty$, then X induces a measure μ_2 on (Ω_2, Σ_2) with $\mu_2(A) = \mu_1(X^{-1}(A))$. Hence, we can discuss

$$\int_{\Omega_2} f d\mu_2.$$

Theorem 7 (Change of Measure)

$$\int_{\Omega_1} f \circ X d\mu_1 = \int_{\Omega_2} f d\mu_2.$$

Proof. Let $f = 1_E$ for $E \in \Sigma_2$.

$$\int_{\Omega_2} f d\mu_2 = \mu_2(E) = \mu_1(X^{-1}(E)).$$

Then $f \circ X = 1(X^{-1}(E))$, so

$$\int f \circ X = \mu_1(X^{-1}(E)).$$

For simple functions, we can use linearity of integrals for the result. For nonnegative functions, we construct a monotone sequence of functions which increase to f. One possible choice is

$$f_n = \frac{\lfloor 2^n f \rfloor}{2^n} \wedge n.$$

We know that $f_n \uparrow f$ and $\int_{\Omega_1} f_n \circ X = \int_{\Omega_2} f_n$, and $f_n \circ X \uparrow f \circ X$, so by the monotone convergence theorem,

$$\int_{\Omega_1} f_n \circ X \to \int_{\Omega_1} f \circ X,$$

and

$$\int_{\Omega_2} f_n \to \int_{\Omega_2} f,$$

so it follows that $\int_{\Omega_1} f \circ X = \int_{\Omega_2} f$, as desired. $\square$

§6.4 Product Measures

We will relate high dimensional integrals with low dimensional ones with the notion of Product Measures.

Let $(\Omega_1, \Sigma_1, \mu_1), (\Omega_2, \Sigma_2, \mu_2)$ be measure spaces. Consider $(\Omega_1 \times \Omega_2, \Sigma(\Sigma_1 \times \Sigma_2))$. Note that $\Sigma_1 \times \Sigma_2$ is a semialgebra. From here, we construct the product measure.

Theorem 8

There exists a unique measure on $\Sigma_p = \Sigma_1 \times \Sigma_2$, μ such that

$$\mu(A \times B) = \mu_1(A)\mu_2(B)$$

for all $A \times B \in \Sigma_p$. We will call this the product measure.

Proof. Given a countable additive sigma-finite measure on a semi-algebra, it admits a unique extension to the generated sigma-algebra by Caratheodory's extension theorem. To prove the existence and uniqueness, it suffices to check countable additivity on $\Sigma_1 \times \Sigma_2$.

If we have $A \times B = \bigcup A_i \times B_i$, we want

$$\mu(A \times B) = \mu_1(A)\mu_2(B) = \sum \mu_1(A_i)\mu_2(B_i).$$

Our strategy is to product to one dimension less. Fix $x \in A$. Note that $B = \{y : (x, y) \in A \times B\}$. But $A \times B = \bigcup A_i \times B_i$. Consider all A_i's that contain x. Then, the corresponding B_i's form a disjoint partition of B. We know that $\mu_2(B) = \sum_{x \in A_i} \mu_2(B_i)$ for all $x \in A$, so in particular

$$1_A \mu_2(B) = \sum 1_A \mu_2(B_i).$$

We claim that the two functions are pointwise same on Ω_1. Then, we integrate (uses MCT) with respect to μ_1 to get

$$\mu_1(A)\mu_2(B) = \lim_{n \to \infty} \sum_{i=1}^{n} \mu_1(A_i)\mu_2(B_i).$$

$\square$

§7 September 17th, 2020

§7.1 Product Measures, Continued

We have a semialgebra $\Sigma_1 \times \Sigma_2$ and we denote $\Sigma_{1\times 2} = \Sigma(\Sigma_1 \times \Sigma_2)$. Let $E \in \Sigma_{1\times 2}$.

Lemma 7.1

For any $X \in \Omega_1$, the set

$$E_x = \{y \in \Sigma_2 : (x, y) \in E\}$$

is measurable.

Proof. If $E = E_1 \times E_2$, then either $E_x = \emptyset$ or $E_x = E_2$. We show that the set of E with this property forms a sigma-algebra. If $E \in \mathcal{A}$, then $(E^c)_x = (E_x)^c$ so $E^c \in \mathcal{A}$.

If $E_1, E_2, \dots \in \mathcal{A}$, then

$$\left(\bigcup_{i=1}^{\infty} E_i\right)_x = \bigcup_{i=1}^{\infty} (E_i)_x.$$

Hence, $\mathcal{A}$ is a sigma-algebra containing $\Sigma_1 \times \Sigma_2$, so $\mathcal{A} = \Sigma_{1\times 2}$. □

Theorem 9

For any $E \in \Sigma_{1\times 2}$,

$$\mu_1 \times \mu_2(E) = \mu(E) = \int_{\Omega_1} \mu_2(E_x) d\mu_1,$$

and $\mu_2(E_x)$ is a measurable function from $\Omega_1 \to \mathbb{R}$.

Proof. The result is clear for rectangles. If $E_1, E_2 \in \mathcal{A}$, then $\mu_2(E_x) = \mu_2(E_{1x}) + \mu_2(E_{2x}) - \mu((E_1 \cap E_2)_x)$ We use the $\pi - \lambda$ theorem. It suffices to show that $\mathcal{A}$ is a λ-system and we have that $A \supset \Sigma_1 \times \Sigma_2$, which is a π-system.

It is clear that $\Omega_1 \times \Omega_2 \in \mathcal{A}$. We claim that if $E_n \in \mathcal{A}$, then $E_n \uparrow E$ implies $E \in \mathcal{A}$. Note that $(E_n)_x \uparrow E_x$ for all x then $\mu_2(E_{nx}) \uparrow \mu_2(E_x)$ and if we define $\mu_2(E_{nx}) = f_n(x)$, then $f_n(x) \uparrow E_x$ is measurable, as desired. Finally, we show that if $E_1 \supset E_2$ and $E_1, E_2 \in \mathcal{A}$, then $E_1 \setminus E_2 \in \mathcal{A}$. This is clear since $(E_1 \setminus E_2)_x = E_{1x} \setminus E_{2x}$ so $\mu((E_1 \setminus E_2)_x) = \mu_2(E_{1x}) - \mu_2(E_{2x})$ is measurable as the difference of finite measurable functions.

The same argument shows that $\mathcal{A}$ is a λ-system if we define $\mathcal{A}$ to be all E so that both conclusions of the theorem hold.

□

Theorem 10 (Fubini)

Let $f \geq 0$ or $\|f\|_1 < \infty$. Then

1. For all x, $f(x, \cdot)$ is measurable on Σ_2.
2. $\int f(x, \cdot)$ is measurable on Σ_1.
3. $\int \int f(x, \cdot) = \int f$.

Proof. We have verified this for $f = 1_E$. Suppose $f \geq 0$. By linearity of integrals and the fact that the sum of measurable functions is measurable, the claim holds for simple functions. For general $f \geq 0$, take a sequence of simple functions $f_n \uparrow f$.

Then,

$$\int \left(\int f_n(x, \cdot) d\mu_2 \right) d\mu_1 = \int f_n d(\mu_1 \times \mu_2),$$

so the result follows from the monotone convergence theorem.

For general $f \in L^1(\mathbb{R})$, $f = f^+ - f^-$, so we use the above to conclude. □

Example 7.2 (Not-integrable Function)

Let $\Omega_1 = \Omega_2 = \mathbb{N}$. Suppose μ_1, μ_2 are counting measures. Let $f(m, m) = 1$, $f(m + 1, m) = -1$ and $f(m, n) = 0$. Then

$$\sum_m \sum_n f(m, n) = 1, \sum_n \sum_m f(m, n) = 0.$$

The failure is that $f \notin \ell^1$.

Example 7.3 (Not σ-finite)

Let $\Omega_1 = \Omega_2 = (0, 1)$. Let μ_1 be the uniform measure and μ_2 be the counting measure. Let $E = \{(x, x) : x \in (0, 1)\}$. Then $\int \int \mu_2(E_x) d\mu_1 = 1$, but $\int \int \mu_1(E_y) d\mu_2 = 0$.

§7.2 Independence

Definition 7.4 (Naive Independence)**.** If X_1, X_2 are random variables, then X_1 and X_2 are independent if

$$P(X_1 \in E, X_2 \in F) = P(X_1 \in E)P(X_2 \in F).$$

We will generalize this notion.

Definition 7.5 (Independence)**.** For a (Ω, Σ, μ), if $\Sigma_1, \Sigma_2, \ldots, \Sigma_k \subset \Sigma$ are said to be **mutually independent** if for any subset $\{i_1, i_2, \ldots, i_\ell\} \subset \{1, \ldots, k\}$ and sets $A_{i_1}, A_{i_2}, \ldots, A_{i_\ell}$, $A_{i_j} \in \Sigma_{i_j}$

$$\mu(A_{i_1} \cap \cdots \cap A_{i_\ell}) = \prod \mu(A_{i_j}).$$

This is the same as the condition

$$\mu(A_1 \cap A_2 \cap \ldots A_k) = \prod_{i=1}^k \mu(A_i),$$

since we take some of the $A_i = \Omega$.

Definition 7.6 (Independent Random Variables)**.** $X_1, \ldots, X_k$ are mutually independent of $\{\Sigma(X_i)\}$ are mutually independent.

Theorem 11

Suppose $A_1, A_2, \cdots \subset \Sigma$ are mutually independent π-systems. Then $\Sigma(A_i)$ are also mutually independent.

Proof. Wlog, we can assume $\Omega \in A_i$ for all i. Fix $B_2 \in A_2, \ldots, B_\ell \in A_\ell$. For $B_1 \in \Sigma(A_1)$, define the two measures μ', μ'' as

$$\mu'(B_1) = \mu(B_1 \cap B_2 \cap \cdots \cap B_\ell),$$

$$\mu''(B_1) = \mu(B_1) \prod_{i=2}^{\ell} \mu(B_i).$$

We claim that $\mu' = \mu''$. Observe that μ' and μ'' agree on A_1 by hypothesis, so the claim holds by the uniqueness part of the Caratheodory Extension theorem on $\Sigma(A_1)$.

$\Sigma(A_1), A_2, \ldots, A_\ell$ are mutually independent π systems. We iterate to get that $\Sigma(A_i)$ are mutually independent. □

Example 7.7 (Pairwise Independent $\neq$ Mutually Independent)

Take $X_1, X_2, X_3 \in \{0, 1\}$, Pick (X_1, X_2, X_3) uniformly from all triples (x_1, x_2, x_3) such that $x_1 + x_2 + x_3 = 0 \pmod 2$. Note that $P(X_i = 1) = P(X_i = 0) = 1/2$. It is clear that (X_i, X_j) are independent, but (X_1, X_2, X_3) are not independent since $P((X_1, X_2, X_3) = (1, 1, 1)) = 0 \neq (1/2)^3$.

Theorem 12 (Kolmogorov's 0-1 Law)

Suppose $X_1, X_2, \ldots$ are independent random variables. Consider

$$T_n = \sigma(X_n, X_{n+1}, \ldots),$$

and let

$$T = \cap_{n=1}^{\infty} T_n$$

(this is known as a tail-sigma algebra). Then T is a μ-trivial sigma algebra: for all $E \in T$, $\mu(E) = 0$ or 1.

Proof. The idea is E is independent of $X_1, \ldots, X_{n-1}$, so E is independent of $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_{n-1})$. Hence E is independent of $\bigcap_{i=1}^{n-1} \sigma(X_i)$, so E is independent of $\bigcap_{i=1}^{\infty} \sigma(X_i)$, so E is independent of $\Sigma(X_1, X_2, \ldots)$. But $E \in T \subset \Sigma(X_1, \ldots)$, so $P(E \cap E) = P(E)P(E) = P(E)$, so $P(E) = 0$ or 1.

Claim: If A_{ij} for $j = 1, \ldots, m_i$ such that A_{ij} are all π-systems containing Ω are mutually independent, then $\Sigma(A_{i1}, A_{i2}, \ldots, A_{im_i})$ are also mutually independent. To prove this, let $A_i = \{B_1 \cap B_2 \cap \cdots \cap B_{m_i} : B_j \in A_{ij}\}$.

We know that $\Sigma(X_1), \ldots$ are independent π systems,so $\Sigma(X_1, \ldots, X_n)$ and $\Sigma(X_{n+1}, \ldots)$ are independent. Hence E is independent of $\bigcap_{i=1}^{\infty} \Sigma(X_i)$, so E is independent of T, which gives the result. □

Theorem 13 (Kolmogorov Extension)
Take $(\mathbb{R}^n, B(\mathbb{R}^n), \mu_n)$ a consistent family of measures on $\mathbb{R}^n$: for $A \in B(\mathbb{R}^n)$

$$\mu_{n+1}(A \times \mathbb{R}) = \mu_n(A).$$

Then there exists a measure μ on $(\mathbb{R}^{\mathbb{N}}, B(\mathbb{R}^{\mathbb{N}}))$ such that μ agrees with μ_n on $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \ldots$.

§8 September 22nd, 2020

Last time, we discussed product measures, independent random variables/sigma algebras, and how to construct infinitely many independent random variables. We also proved the $0-1$ law for tail-sigma algebras.

If we have $(\Omega, \Sigma, \mathbb{P})$ and random variables $X_1, X_2, \ldots$, $T_n = \Sigma(X_n, X_{n+1}, \ldots)$ and $T_\infty = \bigcap T_n$ is a sigma algebra that is P-trivial.

Any event that does not depend on any finite set of X_i's is in the tail-sigma algebra. For example, let $Y = \limsup X_i$ and $E = \{Y < t\}$. Note that Y does not depend on finitely many X_i's. Another example is $S_n = \sum_{i=1}^n x_i$ and we define $Y = \limsup \frac{S_n}{n}$.

When does $\frac{S_n}{n}$ have a limit?

§8.1 Law of Large Numbers

We have $X_1, X_2, \ldots$ independent random variables. What is the asymptotic behavior of $\frac{S_n}{n}$?

Suppose $X_1, X_2, \ldots$ have $E(X_i^2) < C$, $E(X_iX_j) = 0$ and $E(X_i) = 0$. Then,

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0 \Leftrightarrow \mathbb{P}\left(|\frac{S_n}{n}| > \epsilon\right) \to 0.$$

Proof. We first note Markov's Inequality: Suppose X is a nonnegative random variable. For any positive c,

$$P(X > c) \leq \frac{E(X)}{c}.$$

Furthermore, note that

$$\left\{\left|\frac{S_n}{n}\right| > \epsilon\right\} = \left\{\left(\frac{S_n}{n}\right)^2 > \epsilon^2\right\}.$$

By Markov's Inequality,

$$\mathbb{P}((\frac{S_n}{n})^2 > \epsilon^2) \leq \frac{1}{n^2\epsilon^2}E(S_n^2),$$

and finally,

$$E(S_n)^2 = E((X_1 + \ldots X_n)^2) = \sum EX_i^2 + \sum E(X_iX_j) \leq nC$$

so

$$\frac{1}{n^2\epsilon^2}E(S_n^2) \leq \frac{nC}{n^2\epsilon^2} = \frac{C}{n\epsilon^2} \to 0.$$

□

Corollary 8.1

If $X_1, X_2, \ldots$ are independent with the same distribution and $E(X_i) = \mu$ $E(X_i^2) = \sigma^2$, then

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

Proof. Note that $E(\overline{X_i X_j}) = E(\overline{X_i})E(\overline{X_j}) = 0$ by Fubini's theorem. Hence we apply the previous theorem to $\bar{X}_i = X_i - \mu$. □

Fact 8.2. Chebyshev's Inequality: For any RV X,

$$P(|X| > t) \leq \frac{E(X^2)}{t^2}.$$

Example 8.3 (Polynomial Approximation)

Task: Given $f : [0,1] \to \mathbb{R}$ continuous, and $\epsilon > 0$, find a polynomial $f_n(x)$ such that

$$|f_n(x) - f(x)| < \epsilon$$

for all $x \in [0,1]$.

Let

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

We expect $f_n(x) \approx f(x)$ by the Binomial Theorem. Precisely,

$$f_n(x) = E\left(f\left(\frac{S_n}{n}\right)\right)$$

where $S_n \sim Bin(n, x)$ with $S_n = \sum_{i=1}^n X_i$ for $X_i \sim Ber(x)$. It suffices to show that $\frac{S_n}{n} \approx x$.

By the Law of Large Numbers,

$$P\left(\left|\frac{S_n}{n} - x\right| > \epsilon\right) \to 0.$$

Since f is continuous on $[0,1]$, it is uniformly continuous, so that given δ, there exists ϵ such that for all x, y with $|x-y| < \epsilon$, $|f(x) - f(y)| < \delta$. If we let the event above be A^c, then,

$$\begin{aligned} E(f(S_n/n)) &= E(f(S_n/n)1_A) + E(f(S_n/n)1_{A^c}) \\ &= f(x)P(A) + E(f(S_n/n) - f(x))1_A) + E(f(S_n/n))1_{A^c} \\ &\leq f(x)P(A) + \delta P(A) + \sup_{x \in [0,1]} f(x)P(A^c) \qquad \to f(x). \end{aligned}$$

Note that

$$P(A^c) \leq \frac{\operatorname{Var}(S_n)}{n^2\epsilon^2} \leq \frac{1}{n\epsilon^2}$$

since $\operatorname{Var}(X_i) \leq 1$ for $X_i \in [0,1]$.

Hence, for any $x \in [0,1]$, $f_n(x) \to f(x)$ uniformly as $n \to \infty$.

Now, our goal is to prove the law of Large Numbers without the second moment assumption. Namely, for $X_1, X_2, \ldots$ iid with $E|X_i| < \infty$, $EX_i = 0$,

$$\frac{S_n}{n} \xrightarrow{P} 0.$$

Our strategy is truncation.

Definition 8.4. For any random variable X, we will consider the random variable from $X_M = X1_{|X|<M}$. Note that we have $E(X_M^2) < \infty$ for all M even if $E(X^2) = \infty$.

Theorem 14

Suppose that for each n there exists a constant b_n such that

$$\sum_{i=1}^{n} P(|X_{n_i}| > b_n) \to 0$$

and

$$\sum_{i=1}^{n} \frac{E(\overline{X_{n_i}})^2}{b_n^2} \to 0.$$

Then

$$\sum_{i=1}^{n} \frac{X_{n_i} - E(\overline{X_{n_i}})}{b_n} \to 0.$$

Proof. We first prove that

$$Y = \sum_{i=1}^{n} \frac{\overline{X_{n_i}} - E(\overline{X_{n_i}})}{b_n} \to 0.$$

This follows from Chebyshev, since $E(Y) = 0$ and

$$\mathrm{Var}(Y) \le \sum_{i=1}^{n} \frac{E(\overline{X_{n_i}}^2)}{b_n^2}.$$

Then $\sum_{i=1}^{n} P(|X_{n_i} > bn|) \to 0$ so if $X_{n_i} < b_n$, $X_{n_i} = \overline{X_{n_i}}$. Let $B = \{X_{n_i} \neq \overline{X_{n_i}} : i \in \{1, \dots, n\}\}$. Then

$$P(B) \le \sum_{i=1}^{n} P(|X_{n_i}| > b_n) \to 0,$$

so it follows that

$$\sum_{i=1}^{n} \frac{X_{n_i} - E(\overline{X_{n_i}})}{b_n} \to 0.$$

□

Lemma 8.5

Suppose $X_1, X_2, \dots$ are iid. Suppose that

$$KP(|X_1| > K) \to 0.$$

Then

$$\frac{\sum_{i=1}^{n} X_i - nE(X_1 1\{|X_1| < n\})}{n} \to 0$$

in measure.

Proof. Note that this does not imply $E(X_1) < \infty$. Form a triangular sequence from the X_i's and let $b_n = n$. We show that $\sum_{i=1}^{n} P(|X_i| > n) \to 0$ and $\sum_{i=1}^{n} E(\overline{X_i}^2) \to 0$.

For 2, it suffices to show

$$E(\overline{X_i}^2)/n \to 0.$$

Note that $|\overline{X_i}| = |X_i| 1\{|X_i| < n\}$. Suppose X is a non-negative random variable. Note that

$$E(X) \approx \sum_{n=1}^{\infty} P(X > n).$$

Similarly,

$$E(X^2) \approx \sum_{n=1}^{\infty} nP(X > n).$$

Then

$$E(\overline{X_i}^2) \approx \sum_{K=1}^{n} KP(X_1 > K).$$

It suffices to show that

$$\frac{\sum_{k=1}^{n} kP(|X_1| > k)}{n} \to 0,$$

which follows from the fact that $kP(|X_1| > k) \to 0$. $\square$

Theorem 15 (Law of Large Numbers)

If $X_1, X_2, \ldots$ iid and $E(|X_1|) < \infty$ and $E(X_1) = 0$, then $S_n/n \to 0$ in measure.

Proof. Note that $kP(|X_1| > k) \leq E(|X_1| 1\{|X_1| > k\}) \to 0$, by the dominated convergence theorem. By the lemma,

$$\frac{S_n}{n} - E(\overline{X_1}) \to 0,$$

and note that $E(\overline{X_1}) \to E(X_1) = 0$ by the dominated convergence theorem. $\square$

§9 September 24th, 2020

§9.1 Law of Large Numbers, continued

Last time, we began discussing the Law of large numbers. Recall:

- Markov's Inequality:
$$P(|X| > c) \leq \frac{E(|X|)}{c}.$$

- With $X_1, X_2, \ldots$ iid, $E(X_i) = 0$. When $E(X_1^2) < c$,
$$S_n/n \xrightarrow{P} 0.$$

- Under 1st Moment condition, we used truncation to make thinks bounded and have second moments. We discussed triangular arrays and saw a theorem which proves a LLN type of statement for truncated variables.

- We showed that the truncation has no limiting effect. Then, we considered
$$\sum X_i 1_{|X_i|<n}/n \to 0,$$
which implied the law of large numbers.

Example 9.1

Let $X_1, X_2, \ldots$ be iid with $X_i \geq 0$. Suppose $E(X_1) = \infty$. Then
$$\frac{\sum_{i=1}^n X_i}{n} = ?$$

Let $Y_i \sim X_i 1_{|X_i|<M}$. Then $S_n'/n \sum_{i=1}^n Y_i/n \to E(Y_i)$ by the weak law of large numbers. But by nonnegativity, $S_n/n > S_n'/n \to E(Y_i)$, but $E(Y_i)$ can be made arbitrarily large by choosing M very large.

For any c,
$$P\left(\frac{S_n}{n} > c\right) \to 1,$$
so $S_n/n \to \infty$.

Example 9.2

Let $X = 2^i$ with probability $1/2^i$ for $i \geq 1$. Note $E(X) = \infty$.

Let $X_1, X_2, \ldots$ be iid X. What is the growth rate of S_n? One expects to see some X_i's take value comparable to n since $P(X_1 = n) = \frac{1}{n}$.

We will control the growth with truncation. Let $\alpha_n = \log n + k(n)$, $b_n = 2^{\alpha_n}$. We need to show that

$$\sum_{i=1}^n P(X_i > b_n) \to 0,$$

and

$$\frac{\sum_{i=1}^n E(X_i^2 1_{X_i < b_n})}{b_n^2} \to 0.$$

Note that

$$P(X_i > b_n) \approx \frac{1}{b_n} = \frac{1}{n 2^{k(n)}},$$

so

$$\sum P(X_i > b_n) = \frac{1}{2^{k(n)}} \to 0.$$

Then,

$$E(X_i^2 1_{X_i < b_n}) \approx \sum_{i=1}^{\alpha(n)} 2^{2i}/2^i = \sum_{i=1}^{\alpha(n)} 2^i \approx 2^{\alpha(n)} = b_n.$$

Then,

$$\frac{\sum_{i=1}^n E(X_i^2 1_{X_i < b_n})}{b_n^2} \approx \frac{n b_n}{b_n^2} = \frac{1}{2^{k(n))}} \to 0.$$

Therefore,

$$\frac{S_n - nE(\overline{X_i})}{b_n} \to 0.$$

Note that $E(\overline{X_i}) = \alpha(n)$, so

$$\frac{S_n - n(\log n + k(n))}{n 2^{k(n)}}$$

If we choose $\log\log n$, then

$$\frac{S_n - n(\log n + \log\log n)}{n \log n} \to 0,$$

so

$$\frac{S_n}{n \log n} \to 1 \implies S_n = \Theta(n \log n).$$

§9.2 Almost Sure Convergence

Let $X_1, X_2, \ldots$ iid, $E(X_i) = 0$, $E(X_i^2) < C$.

We know that

$$\frac{S_n}{n} \xrightarrow{P} 0,$$

but do we have

$$\frac{S_n}{n} \to 0,$$

almost surely?

Lemma 9.3 (Borel-Cantelli)
If events E_i satisfy $\sum_{i=1}^{\infty} P(E_i) < \infty$, then $P(E_i \text{ infinitely often}) = 0$.

Example 9.4
Let $\epsilon > 0$. We want
$$P\left(\left|\frac{S_n}{n}\right| > \epsilon, i.o.\right) = 0.$$
In order to apply BC, we have to show
$$\sum P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \infty,$$
but
$$\sum P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) \approx \frac{1}{\epsilon^2 n} \to \infty.$$
We try to get around this by assuming a higher moment. Suppose $E(X^4) < \infty$. Then,
$$\frac{E(S_n^4)}{n^4} = \frac{E((\sum_{i=1}^n X_i)^4)}{n^4} = \frac{nE(X_1^4) + n^2 E(X_1^2 X_2^2)}{n^4} \approx \frac{1}{n^2}.$$
So
$$P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) \le \frac{1}{\epsilon^4} E((S_n/n)^4) \approx \frac{1}{\epsilon^4 n^2},$$
which gives that
$$\sum P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \infty.$$

Can one use naive Markov to show a subsequence converges?
If we let $K(n) = n^2$,
$$P(|\frac{S_{k(n)}}{k(n)}| > \epsilon) \approx \frac{1}{n^2}$$
so we can take the infinite sum and it approaches 0.
Define
$$D(n) = \sup_{k(n) \le i \le k(n+1)} |S_i - S_{k(n)}|.$$
It suffices to show that
$$\frac{D(n)}{k(n)} \to 0.$$
We know that
$$P(|D_n/k(n)| > \epsilon) \le \sum_{i=k(n)}^{k(n+1)} P\left(\frac{|S_i - S_{k(n)}|}{k(n)} > \epsilon\right),$$
by subadditivity. By Chebyshev,
$$\sum_{i=k(n)}^{k(n+1)} P\left(\frac{|S_i - S_{k(n)}|}{k(n)} > \epsilon\right) \le \sum \frac{i - k(n)}{k(n)^2 \epsilon^2} \le \frac{(k(n+1) - k(n))^2}{2k(n)^2} \approx \frac{1}{n^2},$$
so $\frac{D_n}{k(n)} \to 0$ almost surely by BC.

§10 September 29th, 2020

Recall from last time:

- Weak law of Large Numbers, using triangular arrays and truncation,
- The Borel-Cantelli Lemma,
- The Strong Law of Large Numbers, assuming 4th moments with Markov, and assuming 2nd moments, we proved convergence along a subsequence and controlled oscillations.

Fact 10.1. $X_n \to X$ in probability if and only if for any sequence of X_n, there exists a subsequence which converges almost surely.

Today, we will prove the most general version of SLLN, under the first moment assumption.

§10.1 General Law of Large Numbers

Let $E_1, E_2, \ldots$ be pairwise independent events, where $p_i = P(E_i)$. Assume that $\sum P_i \to \infty$. We have $S_n = \sum_{i=1}^n 1_{E_i}$, and we would like to consider $\frac{S_n}{E(S_n)}$. We claim that

$$S_n/E(S_n) \to 1,$$

almost surely.

Proof. Let $a_n = \{|S_n - E(S_n)| > \epsilon E(S_n)\}$. We want to bound $P(a_n)$. It suffices to prove that for any $\epsilon > 0$, $\sum a_n < \infty$, by the Borel-Cantelli lemma.

Note that

$$P(|S_n - E(S_n)| > \epsilon E(S_n)) \le \frac{\text{Var}(S_n)}{\epsilon^2 (E(S_n))^2},$$

and

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(1_{E_i}) = \sum_{i=1}^n p_i(1-p_i) \le 1.$$

Hence, $\text{Var}(S_n) \le E(S_n)$ and

$$a_n \le \frac{E(S_n)}{\epsilon^2(E(S_n))^2} = \frac{1}{\epsilon^2}\frac{1}{E(S_n)}.$$

Denote $E(S_n) = g_n$. Let $k(n)$ be the least element such that $g_{k(n)} \ge n^2$.

We have that

$$S_{k(n)}/g_{k(n)} \to 1,$$

almost surely, by applying Borel-Cantelli. It suffices to control the error between the subsequence.

Let $k(n) \le m \le k(n+1)$. We would like to show that

$$S_m/g_m \to 1.$$

But notice that $S_{k(m)} \le S_m \le S_{k(m+1)}$ since indicator functions are nonnegative. So,

$$S_{k(n)}/g_{k(n+1)} \le S_m/g_m \le S_{k(n+1)}/g_{k(n)} = \frac{S_{k(n+1)}}{g_{k(n+1)}}\frac{g_{k(n+1)}}{g_{k(n)}} \approx \frac{S_{k(n+1)}}{g_{k(n+1)}}\frac{(n+1)^2}{n^2} \to 1.$$

We can have a similarly bound for the bottom term, and the result follows from the squeeze theorem. □

Theorem 16 (Strong Law of Large Numbers)
Let $X_1, X_2, \ldots$ be iid and $E|X_i| < \infty$, $E(X_i) = 0$. Then

$$S_n/n \to 0$$

almost surely.

Proof. We will prove convergence of S_n^+/n and S_n^-/n, where $S_n^+ = \sum_{i=1}^n X_i^+$ and similarly for the negative. Hence, we can assume without loss of generality that $X_i \geq 0$. We start by applying truncation: $\bar{X}_i = X_i 1_{|X_i| \leq i}$.

It suffices to prove $\bar{S}_n/n \to 0$ almost surely. This follows from the fact that $\bar{X}_i = X_i$ for large enough i almost surely, since $\sum P(\bar{X}_i \neq X_i) = \sum_{i=1}^\infty P(X_i > i) = E(X_i) < \infty$.

By Markov,

$$P\left(\left|\frac{\bar{S}_n - E(\bar{S}_n)}{n}\right| > \epsilon\right) \approx \frac{Var(\bar{S}_n - E(\bar{S}_n))}{\epsilon^2 n^2}.$$

This will not necessarily be summable over all n, so we choose a subsequence. We will choose $k(n) = \alpha^n$ for a fixed $\alpha > 1$.

We need to show that $\sum Var(\overline{S_{k(n)}})/k(n)^2 < \infty$. Note that

$$Var(\overline{S_{k(n)}})/k(n)^2 \leq \sum_{i=1}^{k(n)} E\bar{X}_i{}^2/k(n)^2$$

and

$$E\bar{X}_i{}^2 \approx \sum_{j=1}^{i} jP(X > j).$$

We can hence rewrite our expression as

$$\sum_{n=1}^\infty \frac{\sum_{i=1}^{k(n)} \sum_{i=1}^{i} jP(X > j)}{k(n)^2} \leq \sum_{j=1}^\infty \sum_{n=1}^\infty \frac{jP(X > j)}{k(n)} 1(k(n) \geq j) \approx \sum_{j=1}^\infty jP(X > j) \cdot \frac{1}{j} \approx E(X),$$

and we have that $E(X) < \infty$.

Hence

$$\frac{\bar{S}_{k(n)} - E(\bar{S}_{k(n)})}{k(n)} \to 0,$$

almost surely.

Then, $E(\bar{S}_{k(n)})/k(n) \to E(x)$ since $E(\bar{X}_i) \to E(X)$ by DCT, so $\bar{S}_{k(n)}/k(n) \to E(X)$ almost surely. Since $\bar{X}_i > 0$, we have

$$\bar{S}_{k(n)} \leq \bar{S}_m \leq \bar{S}_{k(n+1)},$$

and we can apply the squeezing argument from before but within bounds of $1/\alpha, \alpha$. It suffices to choose any $\alpha > 1$, so choosing α arbitrarily close to 1 gives that $\limsup$ and $\liminf$ are both equal to $E(X)$.

Hence, $\bar{S}_m/m \to E(X)$ almost surely, and by BC we have that

$$\frac{S_m}{m} \to E(X),$$

when $X \geq 0$ almost surely. The full theorem comes from splitting $S_m = S_m^+ - S_m^-$. $\square$

Corollary 10.2

If X_i iid with $E(X_i) = \infty$ then $S_n/n \to \infty$.

Proof. Take $X_i^m = X_i 1(X_i < m)$. $E(X_i^m) < \infty$ and $\frac{S_n^m}{n} \to E(X_i^m)$ and $S_n/n \geq S_n^m/n$ for all n, m so choosing large enough m gives the result. $\square$

§10.2 Second Proof of SLLN

We will use the following:

Theorem 17 (Kolmogorov's Maximal Inequality)

Let $X_1, X_2, \ldots$ be independent with $E(X_i^2) < \infty$, $E(X_i) = 0$.

$$P(\max_{k \leq n} |S_k| > \epsilon) \leq \frac{Var(S_n)}{\epsilon^2}.$$

Proof. Let $A_n = \{\max_{k \leq n} |S_k| > \epsilon\}$/ Suppose T_k is the event that k is the smallest index such that $|S_k| > \epsilon$. Then,

$$A_n = \bigcup_{k=1}^n T_k.$$

$$\begin{aligned} E(S_n^2) \geq E(S_n^2 1_{A_n}) &= E(S_n^2 \sum_{k=1}^n 1_{T_k}) \\ &= E((S_k + (S_n - S_k))^2 1_{T_k}) \\ &= E(S_k^2 1_{T_k}) + E((S_n - S_k)^2 1_{T_k}) + E(S_k(S_n - S_k) 1_{T_k}). \end{aligned}$$

The last term is 0 since S_k and 1_{T_k} are measurable functions with respect to $\{X_1, \ldots, X_k\}$, but $S_n - S_k = \sum_{j=k+1}^n X_j$ is measurable with respect to $\{X_{k+1} \ldots, X_n\}$.

Hence,

$$E(S_n^2 1_{T_k}) \geq E(S_k^2 1_{T_k}) \geq \epsilon^2 P(T_k),$$

since $|S_k| > \epsilon$ on T_k, which gives

$$P(A_n) \leq \frac{Var(S_n)}{\epsilon^2}$$

by the union bound. $\square$

Theorem 18

Suppose $X_1, \ldots$ are independent mean 0 random variables. If $\sum_{i=1}^\infty Var(X_i) < \infty$, then $\sum_{i=1}^n X_i$ converges almost surely.

§11 October 1st, 2020

Last time:

- We covered the SLLN using an exponentially growing convergent subsequence.
- Kolmogorov's Maximal Inequality: Let $X_1, X_2, \dots$ be independent with $E(X_i^2) < \infty$, $E(X_i) = 0$.
$$P(\max_{k \le n} |S_k| > \epsilon) \le \frac{Var(S_n)}{\epsilon^2}.$$

§11.1 Another Proof of SLLN, continued

Theorem 19

Suppose $X_1, \dots$ are independent mean 0 random variables. If $\sum_{i=1}^{\infty} Var(X_i) < \infty$, then $\sum_{i=1}^{n} X_i$ converges almost surely.

Proof. We would like to show that $S_1, S_2, \dots$ converges. It suffices to show that (S_i) is almost surely Cauchy. Given any $\epsilon > 0$, there exists n_0 such that for all $n_1, n_2 > n_0$ $|S_{n_1} - S_{n_2}| < \epsilon$.

Then,
$$P\left(\sup_{n \le k \le m} |S_k - S_n| > \epsilon\right) = \epsilon^{-2} \sum_{k=n}^{m} Var(X_k).$$
Hence,
$$P\left(\sup_{n \le k} |S_k - S_n| > \epsilon\right) = \epsilon^{-2} \sum_{k=n}^{\infty} Var(X_k),$$
where we take the limit $m \to \infty$ and use continuity from below. Denote $A_n = \{\sup_{n \le k} |S_k - S_n| > \epsilon\}$. If we let $B_n = \{\sup_{k_1, k_2 \ge n} |S_{k_1} - S_{k_2}| < 2\epsilon\}$, then $P(B_n) \ge 1 - P(A_n)$. Note that B_n increases and $P(\bigcup B_n) \ge 1 - \lim P(A_n) = 1$, and since ϵ is arbitrary, taking the intersection over $\epsilon = 1/m$ implies that S_k is Cauchy almost surely. □

Theorem 20

Suppose X_i are iid with $E|X_i| < \infty$, $E(X_i) = 0$. Then $\sum_{k=1}^{n} \frac{X_k}{k}$ converges almost surely.

Proof. Let $\overline{X_k} = X_k 1_{|X_k| \le k}$. We first show that $\sum \overline{X_k}/k$ converges almost surely, and X_k and $\overline{X_k}$ at finitely many points(by BC), which gives the desired result.

It suffices to show that
$$\sum_{k=1}^{\infty} Var(\overline{X_k})/k^2 = \sum_{k=1}^{\infty} E(\overline{X_k}^2)/k^2 < \infty.$$

Then
$$E(\overline{X_k}^2) = \sum_{i=1}^{\infty} iP(|\overline{X_k}| > i) \le \sum_{i=1}^{k} iP(|X_1| > i).$$

Hence,

$$\sum_{k=1}^{\infty} E(\overline{X_k}^2)/k^2 \leq \sum_{i=1}^{\infty} iP(|X_1 > i|) \sum_{k=i}^{\infty} \frac{1}{k^2} \approx \sum_{i=1}^{\infty} iP(|X_1 > i|)\frac{1}{i} \approx E|X_1| < \infty.$$

□

Lemma 11.1 (Kronecker's Lemma)

Suppose $a_n \uparrow \infty$ and y_k are real numbers such that $\sum_{k=1} y_k/a_k$ converges. Then $\sum_{k=1}^{n} y_k/a_n \to 0$.

Proof. Note that $\sum_{k=1}^{n} y_k/a_k = b_n$ with $b_n \to b$. Then, $y_k = (b_k - b_{k-1})a_k$. Then

$$\begin{aligned} a_n^{-1} \sum_{i=1}^{n} y_k &= a_n^{-1} \sum_{k=1}^{n} (b_k - b_{k-1})a_k \\ &= a_n^{-1}\left(a_n b_n + \sum_{k=1}^{n-1} b_k(a_k - a_{k-1})\right) \\ &= b_n - \sum_{k-1}^{n-1} b_k \left(\frac{a_k - a_{k-1}}{a_n}\right) \end{aligned}$$

Hence, $\sum_{k-1}^{n-1} b_k \left(\frac{a_k - a_{k-1}}{a_n}\right) \to b$ and $b_n \to b$, so the difference converges to 0. □

This implies SLLN, since $\sum X_k/k$ converges almost surely. We claimed that this was quantitative.

Example 11.2 (Tighter Bound)

Suppose $X_1, X_2, \ldots$ are iid and $E(X_i = 0)$, $E(X_1^2) < c$. We have already proved that $\sum_{i=1}^{n} X_i/n \to 0$ almost surely. We can also show that

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n} \log n^{1/2+\epsilon}} \to 0$$

almost surely, for any $\epsilon > 0$.

Proof. Let $a_k = \sqrt{k} \log k^{1/2+\epsilon}$. It suffices to show the convergence of

$$\sum_{k=1}^{n} \frac{X_k}{a_k}.$$

Note that

$$\sum_{k=1}^{\infty} Var(X_k/a_k) \approx \sum_{k=1}^{\infty} \frac{C}{k(\log k)^{1+2\epsilon}} < \infty.$$

□

Example 11.3

Let $X_1, X_2, \dots$ be iid with $EX_i = 0$. Assume that $E|X_1|^p < \infty$ for some $p \in (1,2)$. Then $S_n/n^{1/p} \to 0$.

Proof. Let $Y_k = X_k 1_{|X_k|<k^{1/p}}$. Note that

$$\sum_{k=1}^{\infty} P(Y_k \neq X_k) = \sum_{k=1}^{\infty} P(|X_k|^p > k) \le E|X_1|^p < \infty.$$

We then show the convergence of

$$\sum \frac{Y_k - EY_k}{k^{1/p}}.$$

It suffices to show that $\sum_{k=1}^{\infty} Var(Y_k)/k^{2/p} < \infty$. Then

$$\sum_{k=1}^{\infty} Var(Y_k)/k^{2/p} \le \sum_{k=1}^{\infty} EY_k^2/k^{2/p},$$

where $m = k_0^{1/p}$. Then, $EY_k^2 = \sum_{m=1}^{k^{1/p}} m^2 P(x \in [m, m+1])$, so this is approximately

$$\sum_{m=1}^{\infty} m^2 P(x \in [m, m+1]) \sum_{k=k_0}^{\infty} \frac{1}{k^{2/p}} \approx \sum_{m=1}^{\infty} m^2 P(x \in [m, m+1]) \frac{1}{k_0^{2/p-1}},$$

and $k_0^{2/p-1} = m^{2-p}$, so this simplifies to

$$\sum_{m=1}^{\infty} m^p P(x \in [m, m+1]) \approx E|X|^p < \infty.$$

Finally $E(Y_k) = -E(X 1_{|X_k|>k^{1/p}})$ and

$$E(X 1_{|X_k|>k^{1/p}}) = k^{1/p} E((X/k^{1/p}) 1_{|X|>k^{1/p}}) \le k^{1/p} E(|X/k^{1/p}|^p 1_{|X|>k^{1/p}}).$$

This can be written as

$$k^{1/p-1} E(|X|^p 1_{|X|>k^{1/p}}) \to 0,$$

so

$$|\sum EY_k| \le \sum_{k=1}^{n} k^{1/p-1} b_k \approx_{k=1}^{n} n^{1/p} b_k \to 0.$$

□

Lemma 11.4 (Glivenko-Cantelli)

Consider a $X_1, X_2, \ldots$ iid for some distribution F. Define

$$F_n(y) = \frac{1}{n}\sum_{i=1}^n 1_{X_i \le y}.$$

Then, almost surely,

$$\sup_y |F_n(y) - F(y)| \xrightarrow{n\to\infty} 0.$$

Proof. The uniformity follows because F and F_n are monotone. □

§12 October 6th, 2020

§12.1 Convergence of Distributions

Lemma 12.1 (Glivenko-Cantelli)

Consider a $X_1, X_2, \ldots$ iid for some distribution F. Define

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \le y}.$$

Then, almost surely,

$$\sup_y |F_n(y) - F(y)| \xrightarrow{n\to\infty} 0.$$

Proof. Note that F_n is monotonically increasing from 0 to 1. We divide the real line into intervals depending on F as follows: Fix $\epsilon > 0$. Say $Z_0 = -\infty$, $Z_1 = \inf\{y : F(y) \ge \epsilon\epsilon\}$, and $Z_n = \inf\{y : F(y) \ge n\epsilon\}$. We eventually have a sequence of random variables $\{Z_n\}$. We also define $Z_{1/\epsilon} = \infty$.

For any i, $\sup |F_n(x) - F(x)|$ is small for all $x \in [Z_i, Z_{i+1})$. Note that $F(Z_{i+1}) - F(Z_i) \le \epsilon$. Then, we apply SLLN: We let $w_j = 1(X_j \le Z_i)$ and $w_j' = 1(X_j < Z_{i+1})$. Then $E(w_j) = F(Z_i)$ and $E(w_j') = F(Z_{i+1})$. Then we use SLLN to show that the end points are uniformly close and we sandwich the limit inside $[Z_i, Z_{i+1})$. $\square$

Definition 12.2. Let F_n, F be distribution functions. We say $F_n \to F$ in the sense of distributions in $F_n(x) \to F(x)$ such that F is continuous at x.

Example 12.3

Let F_n be the distribution induced by the point mass at $1/n$. We see that $F_n(0) = 0$, which does not converge to $F(0)$, which is 1.

Example 12.4

Suppose X_i are iid $Ber(\pm 1)$. Let $S_n = \sum_{i=1}^{n} X_i$ and consider $\frac{S_n}{\sqrt{n}} \to$ Standard Gaussian.

Take any random variable with distribution F. Let $X_n = X + \frac{1}{n}$, then $X_n \to X$ in distribution. For any continuity point z of F, $F_n(z) \to F(z)$ and $F_n(z) = F(z - \frac{1}{n})$, so fro all $X_n \le Z, X \le z - \frac{1}{n}$ which implies that

$$P(X_n \le z) = P(X \le z - 1/n) = F(z - 1/n).$$

We investigate the relationship between notions of convergence. If $X_n \to X$ in probability, does $X_n \to X$ in distribution? Yes.

If $X_n \to X$ in probability, $P(|X_n - X| > \epsilon) \to 0$. Note that

$$P(X_n \le z) \le P(X \le z + \epsilon) + P(|X_n - X| > \epsilon),$$

so $\limsup P(X_n \le z) \le P(X \le z + \epsilon) \le P(X \le z)$.

Then

$$P(X_n \le z) \ge P(x \le z - \epsilon) - P(|X_n - X| > \epsilon),$$

so $\liminf P(X_n \le Z) \ge P(X \le z - \epsilon) \ge P(X \le z-)$, so $P(X_n \le z) = P(X \le z)$.

The converse does not make sense. One example that fails where $X_n \to X$ in distribution but $X_n \not\to X$ where $X = Ber(\pm 1)$ and $X_1, X_2, \dots$ are all $-X$. Then $X_i \sim X$, but $X_i - X = -2X$.

Theorem 21 (Skorokhod Representation)
If $F_n \to F$ in distribution, then there exists a probability space (Ω, Σ, P) and random variables X_n, X defined on Ω such that $X_n \sim F_n$ and $X_n \to X$ almost surely.

Proof. We did some similar when given a distribution F, we constructed a random variable with distribution F. We work in the uniform space $([0,1], B, P)$. Then $X_n = F_n^{-1}(\omega) = \sup\{y : F_n(y) < \omega\}$ and $X(\omega) = F^{-1}(\omega) = \sup\{y : F(y) < \omega\}$. It suffices to show that $F_n^{-1}(\omega) \to F^{-1}(\omega)$.

Define $F_n'^{-1}(\omega) = \inf\{z : F_n(z) > \omega\}$ and similarly $F_n'^{-1}$. Let $A = \{\omega : F^{-1}(\omega) < F'^{-1}(\omega)\}$. We showed in homework that A is countable. We now prove that $X_n \to X$ in A^c. Suppose that $F_n^{-1}(\omega) \to z$. Let $u < z < v$ where u, v are continuity points and $F(u) < \omega - \epsilon$, $F(v) > \omega + \epsilon$. These can be chosen because the set of discontinuities of a function is countable. For all large n, $F_n(u) < \omega - \epsilon/2$ and $F_n(v) > \omega + \epsilon/2$ because $F_n \to F$ at all continuity points. Then $F_n^{-1}(\omega) \geq u$ and $F_n^{-1}(\omega) \leq v$, hence

$$\limsup_n |F_n^{-1}(\omega) - z| \leq u - v.$$

But u, v were arbitrary continuity points, so

$$\limsup_n |F_n^{-1}(\omega) - z| = 0.$$

$\square$

Fact 12.5. $X_n \to X$ in distribution if and only if for any bounded continuous function g, $E(g(X_n)) \to E(g(X))$.

Proof. If $X_n \to X$ in distribution, then we have $Y_n \to Y$ with $Y_n \sim X_n$, $Y \sim X$ so $g(Y_n)$ has the same distribution as $g(X_n)$. Thus $E(g(Y_n)) = E(g(X_n))$. Then, g is continuous so $g(Y_n) \to g(Y)$ almost surely since $Y_n \to Y$ almost surely, and g is bounded so $E(g(X_n)) = E(g(Y_n)) \to E(g(Y)) = E(g(X))$.

For the other direction, if $E(g(X_n)) \to E(g(x))$ for all bounded continuous functions g, we choose a mollified indicator function. Then $E(g(X_n)) \geq P(X_n \leq z)$ and $E(g(X_n)) \to E(g(X)) \leq P(x \leq z + \epsilon)$. Hence,

$$\limsup P(X_n \leq z) \leq P(x \leq z + \epsilon).$$

Similarly,

$$\liminf P(X_n \leq z) \geq P(X \leq z - \epsilon),$$

which gives that $P(X_n \leq z) \to P(X \leq z)$ by using the fact that z is a continuity point. $\square$

This gives an alternate definition of convergence in distribution.

For any topological space and a sequence of measures μ_n and μ on Ω, we can say $\mu_n \to \mu$ in distribution if for all bounded continuous $g : \Omega \to \mathbb{R}$, $E_{\mu_n} g \to E_\mu g$.

Theorem 22 (Continuous Mapping Theorem)

Suppose $X_n \to X$ and $g : \mathbb{R} \to \mathbb{R}$ is a measurable function such that C_g is a set of continuity points of g such that $P(X \in C_g) = 1$, then $g(X_n) \to g(X)$ in distribution.

Proof. Note that the set of continuity points for a measurable function g, C_g is measurable. $\square$

§13 October 8th, 2020

Last time,

- We defined distributional convergence.
- We proved the Skorokhod Representation Theorem.
- We showed an equivalent notion of distributional convergence, namely for bounded continuous functions g, $E(g(x_n)) \to E(g(x))$ implies $x_n \to x$ in distribution.

§13.1 Weak Convergence

Theorem 23 (Continuous Mapping Theorem)
Suppose $X_n \to X$ and $g : \mathbb{R} \to \mathbb{R}$ is a measurable function such that C_g is a set of continuity points of g such that $P(X \in C_g) = 1$, then $g(X_n) \to g(X)$ in distribution.

Proof. We claimed that C_g, the set of continuity points is measurable. We know that $X_n \to X$ in distribution, so there exists $Y_n \to Y$ almost surely with $X_n \sim Y_n$, $X \sim Y$. Then $g(X_n) \sim g(Y_n)$ and $g(X) \sim g(Y)$. Then $Y \in C_g$ almost surely because $X \sim Y$. Therefore, $g(Y_n) \to g(Y)$ almost surely, which implies distributional convergence. $\square$

§13.2 Portmanteau's Lemma

We show many equivalent conditions for distributional convergence.

1. $F_n \xrightarrow{d} F$
2. For any open U, $\liminf P_n(U) \geq P(u)$.
3. For any closed V, $\limsup P_n(V) \leq P_n(V)$.
4. For any A such that $P(\partial A) = 0 (\partial A = \overline{A} \setminus A^o)$, we have $P_n(A) \to P(A)$.

Proof. We first show 1 implies 2. Let $Y_n \sim F_n$, $Y \sim F$ with $Y_n \to Y$ almost surely. Let $F_n = 1(Y_n \in U)$, $f = 1(Y \in U)$. Then $P(Y_n \in U) = P_n(U)$. Finally,

$$\liminf_{n\to\infty} f_n \geq f.$$

This is because $f(\omega) = 1$ if $Y(\omega) \in U$ and 0 otherwise. Pick ω such that $f(\omega) = 1$. Then $Y(\omega) \in U$ and $Y_n(\omega) \to Y(\omega)$ and U is open, we know that for large n, $Y_n(\omega) \in U$. Hence, $f_n(\omega) = 1$ for large n, so $\liminf f_n = 1$.

Finally, by Fatou's Lemma,

$$\liminf P_n(U) \geq \int \liminf f_n \geq \int f = P(U).$$

2 implies 3 is easy. We take V^c, then $P_n(V) = 1 - P_n(V^c)$ and $\limsup P_n(V) = 1 - \liminf P_n(V^c)$.

We show 3 implies 4. Note that $\overline{A} = A \cup \partial A$ and $P(\overline{A}) = P(A)$. Then $A^o = A \setminus \partial A$ and $P(A^o) = P(A)$. By 2, $\liminf P_n(A^o) \geq P(A)$ and $\limsup P_n(\overline{A}) \leq P(A)$ and $P_n(A^o) \leq P_n(A) \leq P_n(\overline{A})$. Therefore, $P_n(A) \to P(A)$.

Finall 4 implies 1. If $P_n(A) \to P(A)$ for $P(\partial A) = 0$, then choose $A = (-\infty, x]$ for a continuity point x. Then $P(\{x\}) = 0$, so we get $F_n(x) = P_n(A) \to P(A) = F(x)$. $\square$

§13.3 Helly's Selection Theorem

Theorem 24

Given a sequence of distributions $F_1, F_2, \ldots$, there exists a subsequence $n_1, n_2, \ldots$ and a right continuous non-decreasing function F so that $F_{n_k}(x) \to F(x)$ for all continuity points.

Note that F may not be a distribution since $F(\infty) - F(-\infty) \neq 1$. For example, take $F_n(x) = 1_{x \geq n}$. Then $F_n(x) = 0$ for $n \geq x$ so $F_n \to 0$, which is not a distribution.

Proof. We want $F_{n_k}(x) \to F(x)$. If we fix x, $0 \leq F_n(x) \leq 1$ so we can pick a subsequence so that $F_{n_k}(x) \to a$ from the compactness of $[0,1]$. We first ensure convergence for the rationals. Let $\{q_n\}$ be an enumeration of the rationals. Iteratively choose subsequences n_k^i such that $\{n_k^i\}$ is a subsequence of $\{n_k^{i-1}\}$ and $F_{n_k}^(q_i)$ converges.

Hence, $F_{n_k^k}(q)$ converges for all $q \in \mathbb{Q}$. We will call this $\{n^k\}$ for convenience. Let the limit be $\lim_{k\to\infty} F_{n^k} = G(q)$.

Then, define $F(x) = \inf_{q > x} G(q)$. We show $F_{n_k}(x) \to F(x)$ for all x with F continuous at x. Pick $q_1, q_2, q_3 \in Q$ so that $q_1 > x > q_2 > q_3$. Then $F_{n_k}(x) \leq F_{n_k}(q_1)$ so $\limsup f_{n_k}(x) \leq G(q_1)$ so $\limsup F_{n_k}(x) \leq F(x)$.

Since x is a continuity point, there exists q_3 close to x so that $F(q_3) \geq F(x) - \epsilon$, but $F(q_3) = \inf_{q > q_3} G(q)$, so there exists q_2 with $q_3 < q_2 < x$ so that $G(q_2) \geq F(q_3) + \epsilon$ so $G(q_2) \geq F(x) - 2\epsilon$. But $F_{n_k}(x) \geq F_{n_k}(q_2)$ so $\liminf F_{n_k}(x) \geq G(q_2) \geq F(x) - 2\epsilon$ so $\liminf F_{n_k}(x) \geq F(x)$. $\square$

When does the limit preserve mass?

Definition 13.1. Given $\epsilon > 0$, $\{F_n\}$ is tight if there exists M_ϵ so that F_n in the sequence

$$F_n(M_\epsilon) - F_n(-M_\epsilon) \geq 1 - \epsilon.$$

Tightness is a necessary and sufficient condition for preserving mass.

Proof. We have $F_{n_k} \to F$ for continuity points. We show F has mass 1. We can assume $M_\epsilon, -M_\epsilon$ are continuity points. Then

$$F_{n_k}(M_\epsilon) - F_{n_k}(-M_\epsilon) \to F(M_\epsilon) - F(-M_\epsilon) \geq 1 - \epsilon.$$

$\square$

If a sequence is not tight, then there exists subsequential limits with mass less then 1.

§13.4 Fourier Transforms

Let X be a random variable with distribution F. For any t, define $\phi(t) = E(e^{itx})$.

We have some properties:

- $\phi(0) = 1$.
- $\overline{\phi(t)} = \phi(-t)$.
- $|\phi(t)| \leq 1$.

Proof. We use Jensen's Inequality: For a convex function $\phi(x_1, x_2, \ldots, x_n)$ with $E|\phi| < \infty$, $E|X_i| < \infty$ for all i, then $E(\phi) \geq \phi(EX_1, EX_2, \ldots)$.

Take $(x, y) \mapsto (x^2 + y^2)^{1/2}$. Then

$$|\phi(t)| \leq E|e^{itx}| = 1.$$

□

- ϕ is uniformly continuous.

$$|\phi(t+h) - \phi(t)| \leq |E(e^{i(t+h)} - e^{itx})| \leq E|e^{itx}(e^{ihx} - 1)| = E|e^{ihx} - 1| \to 0,$$

by the Bounded Convergence Theorem.

§14 October 13, 2020

§14.1 Fourier Transforms, Continued

Recall $\phi(t) = E(e^{itx})$. The broad goal is to study distributional convergence using the characteristic functions.

§14.2 Differentiability of Characters

When is $\phi(t)$ differentiable? If ϕ is differentiable at t, then $\phi'(t) = E(ixe^{itx})$, so $\phi'(0) = E(ix)$. For this to make sense, we might need $E|X| < \infty$.

Theorem 25

If $E|X| < \infty$, then ϕ is continuously differentiable.

Proof.

$$\lim_{h \to 0} \frac{\phi(t+h) - \phi(t)}{g} = E\left(\frac{e^{i(t+h)X} - e^{itX}}{h}\right)$$

It suffices to show that

$$E\left(\frac{e^{i(t+h)X} - e^{itX}}{h} - ixe^{itX}\right) \to 0.$$

$$\begin{aligned}
E\left(\frac{e^{i(t+h)X} - e^{itX}}{h} - iXe^{itX}\right) &= E(iXe^{itX}(\frac{^{itX} - 1}{ihX} - 1)) \\
&= E\left(iXe^{itX}\left(\frac{\int_0^h e^{inX} - 1}{h} - 1\right)\right) \\
&\leq E(|iXe^{itX}| \sup_{u \leq h} |e^{iuX} - 1|) \to 0
\end{aligned}$$

by DCT, since the argument is at most $2|X|$.

Then, ϕ' is continuous since

$$\phi'(t+h) - \phi'(t) = E(ixe^{i(t+h)x} 0 e^{itx}) = E(ixe^{itx}[e^{ihx} - 1]) \to 0$$

by DCT. □

Similarly if $E|X|^k < \infty$, then $\phi(t) \in C^k$. The proof follows straightforwardly by induction on the moment.

Lemma 14.1

If ϕ is twice-differentiable, then $E(X^2) < \infty$.

Proof. Note that
$$\frac{2\phi(0)-\phi(h)-\phi(-h)}{h^2}\xrightarrow{h\to 0}-\phi''(0).$$
But this quantity is exactly
$$\frac{2-2E(\cos(hx))}{h^2}\to -\phi''(0)$$
since $\phi(h)=E(\cos(hx)+i\sin(hx))$ and $\phi(-h)=E(\cos(hx)-i\sin(hx))$.

Finally, $\cos(x)=1-\frac{x^2}{2}$, so we $f_h(X)\to X^2$, $f_h(X)\geq 0$ and $E(f_h(x))\to\phi''(0)$(where $f_h(X)$ is the term inside the expectation). It follows that $E(X^2)\leq \lim E(f_h(X))=-\phi''(0)$. Furthermore, $\phi''(0)=E(-X^2)$. □

§14.3 Fourier Inversion Formula

Let $X\sim\mu$ and $\phi(t)=E(e^{itx})$. Let F be the distribution induced by μ. We would like to approximate $F(b)-F(a)$.

We consider
$$\int_{-T}^{T}\int_a^b e^{-iut}\phi(t)dudt=\int_{-T}^{T}\frac{e^{-iat}-e^{-bt}}{it}\phi(t).$$
Then
$$\int_{-T}^{T}\frac{e^{-iat}-e^{-bt}}{it}\int_{\mathbb{R}}e^{itx}d\mu=\int_{\mathbb{R}}\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}=2\int_{\mathbb{R}}\int_0^T\frac{\sin(t(x-a))-\sin(t(x-b))}{t}dt.$$
Hence, we have
$$2\int_{\mathbb{R}}d\mu\int_{0^T}\frac{\sin(t(x-a))-\sin(t(x-b))}{t}dt.$$
Then, for any x,
$$\int_0^T\frac{\sin(\theta x)}{x}$$
is bounded as $T\to\infty$ and converges to $\frac{\pi}{2}\operatorname{sgn}(\theta)$. It follows that the integral converges to $\pi\mu(\{a\})+\pi\mu(\{b\})+2\pi\mu(a,b)$. Dividing by 2π throughout, gives
$$\mu(a,b)+\frac{\mu(\{a\}\cup\{b\})}{2}.$$

Lemma 14.2

If F_1,F_2 are two distributions with $\phi_1(t)=\phi_2(t)$, where ϕ_1 and ϕ_2 are the characteristics functions, then $F_1=F_2$.

Proof. Consider the set $A=\{x:\mu_1(x),\mu_2(x)>0\}$, which is countable. For all $a,b,\in A^c$, $\mu_1([a,b])=\mu_2([a,b])$. Then, send $a\to-\infty$ to show $F_1(b)=F_2(b)$ for all $b\in A^c$, but A^c is dense so $F_1=F_2$ everywhere by right-continuity. □

Theorem 26

If $\phi(t)$ is integrable, then F has a density: $F[a,b]=\int_a^b f dx$ for all a,b.

Namely,
$$f(x)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{-ixt}\phi(t)dx.$$

Proof. We first show there are no atoms, $\mu(x) = 0$ for all x. Then for $a < x < b$

$$\mu(x) \leq \int |\phi(t)||b-a| \to 0.$$

Finally,

$$\int_a^b \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iut}\phi(t) = \mu(a,b) = \mu[a,b].$$

It is easy to show that f is real valued.

$$\overline{f}(u) = \int e^{iut}\overline{\phi}(t) = \int_{\mathbb{R}} e^{iut}\phi(-t) = \int_R e^{-iut}\phi(t)dt = f(u).$$

□

§15 October 15th, 2020

Recall

- Moments of a Random Variable implies smoothness of the characteristic functions In particular, the existence of the k-th moment implies that $\phi^{(k)}$ exists and is continuous.
- The Inversion Formula:
$$\frac{1}{2\pi}\int_{-T}^{T}\int_{a}^{b} e^{-itu}\phi(t)dt \xrightarrow{T\to\infty} \frac{1}{2}\mu(\{a,b\}) + \mu(a,b).$$
- If $\phi(t)$ is integrable, then a density exists - namely
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t).$$
This implies that $\int_a^b f = \mu[a,b]$. We can also show that f is continuous since $|f(t+h) - f(h)| \to 0$ uniformly by the dominated convergence theorem.

§15.1 Characteristic Functions

Proposition 15.1

If X, Y are independent random variables, $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

Proof.
$$E(e^{it(X+Y)}) = E(e^{itX}e^{itY}) = E(e^{itX})E(e^{itY}).$$
□

Proposition 15.2

$$\phi(t) = E\left(\sum_{m=0}^{n}\frac{(itX)^m}{m!}\right) + o(t^n).$$

Proof. Note that
$$\left|\phi(t) - E\left(\sum_{m=0}^{n}\frac{(itx)^m}{m!}\right)\right| \le E\min(2|x|^n/n!, |x|^{n+1}/(n+1)!).$$
This follows from the fact that
$$\int_0^x e^{is}(x-s)^n ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int e^{is}(x-s)^{n+1}ds.$$
Putting $n = 0$ gives
$$e^{ix} = 1 + ix + i^2\int e^{is}(x-s).$$

It follows by induction that

$$e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!} \int_0^x e^{is}(x-s)^n dx.$$

Finally,

$$\left| \frac{i^{n+1}}{n!} \int_0^x e^{is}(x-s)^n dx \right| \leq \int_0^x (x-s)^n/n! = \frac{x^{n+1}}{(n+1)!}.$$

Then

$$\begin{aligned} \frac{1}{n!} \int_0^x e^{is}(x-s)^n &= \frac{1}{n!}(e^{is}/i)(x-s)^n|_0^x + \int_0^x ne^{is}/i(x-s)^{n-1}) \\ &= \frac{1}{n!}(ix^n - in \int_0^x e^{is}(x-s)^{n-1}) \\ &= \frac{1}{(n-1!)} \int (x-s)^{n-1}(1-e^{is}) \\ &\leq 2x^n/n!. \end{aligned}$$

It follows that $|\phi(t) - \sum E(itX)^m/m!| \leq E \min(2(tX)^n/n!, (tx)^{n+1}/(n+1)!)$ □

§15.2 Weak Convergence

Theorem 27

For F_n distributions with $\phi_n(t)$, F_n converge in distribution if and only if $\phi_n \to \phi$(where ϕ is continuous at 0).

Remark: Consider $N(0,1)$. Then $\phi(t) = e^{-t^2}/2$. It's easy to prove that $\phi'(t) = E(ixe^{itx}) = -t\phi(t)$, using integration by parts to simplify the right side with $\phi(0) = 1$. If $X \sim N(0,1)$ $\sigma X \sim N(0, \sigma^2)$ so $\phi(t) = e^{-\sigma^2 X^2/2}$. Then $\phi_n \to \delta_0$. But $F_n(x) \to \frac{1}{2}$ for all x. We will see that continuity at 0 gives the tightness condition.

Proof. We prove the forward direction. $F_n \to F$ in distribution implies that $E(g(X_n)) \to E(g(X))$ for bounded continuous functions. The result follows since e^{itx} is a bounded continuous function.

We prove the converse. We first show that $\{F_n\}$ is a tight sequence. Note that

$$\int_{-u}^{u} (1-e^{itx}) = 2u - \int_{-u}^{u} (\cos tx + i \sin tx)dt = 2u - \frac{2\sin ux}{x}.$$

Then

$$\frac{1}{2u} \int_{-u}^{u} (1-e^{itx}) = 1 - \frac{\sin ux}{ux}.$$

Then $|\sin x| \leq |x|$, so

$$\begin{aligned} \frac{1}{2u} \int_{-u}^{u} (1-\phi_n(t)) &= \int 1 - \frac{\sin ux}{ux} d\mu_n \\ &\geq \frac{1}{2}\mu_n(|x| \geq 2/u). \end{aligned}$$

Then $\phi_n(0) \to \phi(0) = 1$, so

$$\mu_n(|x| > 2/u) \leq u^{-1} \int_{-u}^{u} (1-\phi_n(t))dt \to u^{-1} \int_{-u}^{u} (1-\phi(t))dt \leq 2\epsilon,$$

for some choice of u from the continuity of ϕ at 0.

Since ϵ is arbitrary, it follows that μ_n is tight. Now we show that F_n converges to F in distribution. Note that any subsequence has a convergent subsequence which converges to a distribution. Then the subsequences converge to $\phi_{n_i} \to \phi_F$, and $\phi_n \to \phi$ so $\phi_F = \phi$. Hence, every subsequence has a convergence converges to the same F by the uniqueness of characteristic functions. Then $F_n(x) \to F(x)$ for all continuity points x of F and it follows that for $F_n(x)$, and subsequence admits a further subsequence converging to $F(x)$. Hence $F_n(x) \to F(x)$, since otherwise one can extract a subsequence such that $|F_{n_i}(x) - F(x)| > \delta$ for all i. $\square$

§16 October 20th, 2020

§16.1 Basic Central Limit Theorem

Under natural conditions, we will prove that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

Example 16.1

Let $X_i = \pm 1$ with probability $1/2$. We can analyze binomial coefficients to prove a central limit theorem.

Theorem 28

Let X_i be iid $E(X_i) = 0$, $E(X^2) = 1$. Then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, 1).$$

Proof. Recall that $\phi_z(t)e^{-t^2/2}$. Hence, we show that $\phi_{S_n/\sqrt{n}}(t) \to e^{-t^2/2}$.

Then

$$\phi_{S_n/\sqrt{n}}(t) = \phi_{S_n}(t/\sqrt{n}).$$

Then $S_n = \sum X_i$ iid, so

$$\phi_{S_n}(t/\sqrt{n})) = (\phi_{X_1}(t/\sqrt{n}))^n.$$

Then

$$|\phi_{X_i}(t/\sqrt{n}) - E\sum_{m=0}^{2}(itx/\sqrt{n})^m/m!| \le E(\min\{(itx/\sqrt{n})^2, (itx/\sqrt{n})^3\}).$$

Then, the expectation simplifies to $1 - t^2/2n$ since $E(X_1) = 0, E(X_1^2) = 1$.

Then,

$$E = E(\min\{(itx/\sqrt{n})^2, (itx/\sqrt{n})^3\}) = o(t^2/n),$$

or $\frac{E}{t^2/n} \to 0$ as $t^2/n \to 0$. This is because

$$E = \frac{t^2}{n}E(\min(X^2, t/\sqrt{n}X^3)) \xrightarrow{t^2/n \to 0} 0,$$

by DCT.

Hence,

$$\phi_{X_1}(t/\sqrt{n})^n = (1 - t^2/2n + o(t^2/n))^n.$$

We show that

$$(1 - t^2/2n + o(t^2/n))^n \to e^{-t^2/2}.$$

More generally, we show that if $c_n \to c$, then $(1 + c_n/n)^n \to e^c$.

Lemma 16.2

Suppose we have complex numbers $z_1, z_2, \ldots, z_n, w_1, w_2, \ldots, w_n$ with $|z_i||w_i| \le \theta$. Then

$$\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| \le \theta^{n-1}\sum_{i=1}^n |z_i - w_i|.$$

Proof. We use telescoping.

$$\prod z_i - \prod w_i = \prod_{i=1}^n z_i - \prod_{i=1}^{n-1} z_i w_n + prod_{i=1}^{n-1} z_i w_n - \prod_{i=1}^{n-2} z_i w_{n-1} w_n + \cdots - \prod_{i=1}^n w_i.$$

Then, using the triangle inequality and summing the bound gives that

$$|\prod z_i - \prod w_i| \le \theta^{n-1}\sum |z_i - w_i|.$$

□

We first bound

$$\left|(1 + c_n/n)^n - e^{c_n}\right|.$$

The whole thing is the same as $|(1 + c_n/n)^n - (e^{c_n/n})^n|$. Note that $e^x \ge 1 + x$ for real x and $|e^x - (1 + x)| \le |x|^2$ for complex x.

Note that $1 + |c_n/n| \le e^{|c_n/n|} = \theta$, so the error is bounded by

$$ne^{|c_n/n|(n-1)}|e^{c_n/n} - (1 + c_n/n)| \le e^{|c_n/n|(n-1)}|c_n|^2/n,$$

and it follows that $(1 + c_n/n)^n \to e^c$. □

§16.2 Lindeberg-Feller CLT

Theorem 29 (Lindeberg-Feller)

We prove CLT for a triangular array of random variables. Consider

$$\begin{aligned}&X_{11}\\ &X_{21}, X_{22}\\ &\ldots\\ &X_{n1}, X_{n2}, \ldots, X_{nn}\end{aligned}$$

independent. We suppose that $EX_{ni} = 0$, $\sum_{i=1}^n EX_{ni}^2 \to 1$ and $\sum_{i=1}^n E(X_{ni}^2, 1(|X_{ni}| > \epsilon)) \to 0$ for all $\epsilon > 0$.

Let $S_n = \sum_{i=1}^n X_{ni}$. We show that $S_n \xrightarrow{d} \mathcal{N}(0, 1)$.

Example 16.3

For the previous example, we had $X_1, X_2, \ldots$ and $S_n = \sum_{i=1}^n X_i$. We take $X_{ni} = X_i/\sqrt{n}$. We know that $EX_{ni} = 0$ and $E(X_{ni}^2) = \frac{1}{n}$, so $\sum EX_{ni}^2 = 1$.

Then

$$\sum_{i=1}^n E(X_{ni}^2, 1(|X_{ni}| > \epsilon)) = nE(X_{ni}^2, 1(|X_{ni}| > \epsilon)) = E(X_1^2 1|X_1| > \epsilon\sqrt{n}) \to 0.$$

Proof. Note that $\phi_{S_n}(t) = \prod \phi_{X_{ni}}(t)$. Let $EX_{ni}^2 = \sigma_{ni}^2$. We have that $\sum \sigma_{ni}^2 = 1$. We want to show that $\phi_{S_n}(t) \to e^{-t^2/2}$.

Note that

$$\begin{aligned}|\phi_{X_{ni}}(t) - (1 - \sigma_{ni}^2 t^2/2)| &\leq E(\min(t^2 X_{ni}^2, t^3 X_{ni}^3)) \\ &\leq E(t^2 X_{ni}^2 1|X_{ni} > \epsilon|) + E(t^3 X_{ni}^3 1|X_{ni}| < \epsilon|) \\ &\leq E(t^2 X_{ni}^2 1|X_{ni}| > \epsilon) + \epsilon t^3 E(X_{ni}^2 1|X_{ni}| < \epsilon).\end{aligned}$$

If we sum the left and take limits, we have

$$\sum_{i=1}^{n} E_{ni} \leq 0 + \epsilon t^3 \xrightarrow{n \to \infty, \epsilon \to 0} 0.$$

Now, from the previous lemma, note that $|\phi_{ni}(t)| \leq 1$. We also have $|1 - \sigma_{ni}^2 t^2/2| \leq 1$, since $\sup \sigma_{ni}^2 \to 0$. This is because

$$\sigma_{ni}^2 = EX_{ni}^2 = E(X_{ni}^2 ||X_{ni}| < \epsilon) + E(X_{ni}^2 ||X_{ni}| > \epsilon) \to 0.$$

Hence,

$$\left| \prod \phi_{ni}(t) - \prod (1 - \sigma_{ni}^2 t^2/2) \right| \to 0.$$

Finally,

$$\log \left(\prod (1 - \sigma_{ni}^2 t^2/2) \right) = \sum \log(1 - \sigma_{ni}^2 t^2/2) \approx -\sum \sigma_{ni}^2 t^2/2 \to -t^2/2,$$

so

$$\phi_{S_n}(t) \to e^{-t^2/2}.$$

□

§16.3 Kolmogorov Three-Series Theorem

Theorem 30 (Kolmogorov Three-Series)

Let $X_1, \ldots, X_i, \ldots$ be independent and $\sum_{i=1}^{n}$ converges almost surely. The above and the following are equivalent. For $A > 0$,

1. $\sum_{i=1}^{\infty} P(|X_i| > A) < \infty$.
2. Where $\overline{X_i} = X_i 1(|X_i| < A)$, $\sum_{i=1}^{n} E\overline{X_i}$ converges.
3. $\sum_{i=1}^{\infty} Var(\overline{X_i}) < \infty$.

Proof. We first prove the converse. It suffices to show that $\sum \overline{X_i}$ converges almost surely, since $\sum \overline{X_i}$ converges is equivalent to $\sum \overline{X_i}$ converging by BC.

By the Kolmogorov Maximal Inequality, $\sum Var < \infty$ implies converges of centered independent random variables, so $\sum \overline{X_i} - E(\overline{X_i})$ converges and $\sum E(\overline{X_i})$ converges by assumption, so $\sum \overline{X_i}$ converges.

Now,we prove the forward direction. The first condition is easy since by BC, $|X_i| < A$ eventually, so $\sum P(|X_i| > A) < \infty$. We now prove $\sum Var(\overline{X_i}) < \infty$. Suppose not: let $c_n = \sum_{i=1}^{n} Var(\overline{X_i})$. Then $c_n \to \infty$. Define $\overline{X}_{ni} = \frac{\overline{X_i} - E\overline{X_i}}{\sqrt{C_n}}$. Note that $E\overline{X}_{ni} = 0$, $\sum_{i=1}^{n} E\overline{X_{ni}}^2 = 1$. Finally, for $\epsilon > 0$,

$$\sum E(\overline{X_{ni}}^2, 1(|X_{ni}| > \epsilon)) \to 0,$$

since $\overline{X_{ni}} > \epsilon$ implies that $\overline{X_i} - E\overline{X_i} > \epsilon\sqrt{c_n}$. By Lin-Fell CLT,

$$\sum \frac{\overline{X_i} - E\overline{X_i}}{\sqrt{c_n}} \to \mathcal{N}(0,1).$$

But by hypothesis, $\sum \overline{X_i}$ converges almost surely, so

$$\sum \overline{X_i}/\sqrt{c_n} \to 0,$$

which is a contradiction.

Hence, $\sum Var(\overline{X_i}) < \infty$. By KMI, $\sum \overline{X_i} - E\overline{X_i}$ converges, so $\sum E\overline{X_i}$ converges, giving 2. $\square$

§17 October 22nd, 2020

§17.1 CLT with Unbounded Variance

Let $X_1, X_2, \dots$ by iid symmetric random variables, and $P(|X_i| > x) = \frac{1}{x^2}$ for all $x > 1$. Then

$$E(X^2) = \int 2xP(|X| > n) = \int 2/x = \infty.$$

We will handle this by using truncation, applying BC, and using CLT for truncated variables. We want to truncate at the smallest possible level so that BC can still be applied.

Define $X_i^n = X_i 1(|X_i| \le \sqrt{n}\log\log n)$ for $i \le n$. Then,

$$P(X_1 + \cdots + X_n \ne X_1^n + X_2^n + \cdots + X_n^n) \le \frac{n}{(\sqrt{n}\log\log n)^2} \to 0.$$

By symmetry, $EX_i^n = 0$. Then

$$E((X_i^n)^2) \le \int_1^{\sqrt{n}\log\log n} 2xP(|X| > n) = \int_1^{\sqrt{n}\log\log n} \frac{2}{x} = \log(n(\log\log n)^2) = \log n + 2\log\log\log n.$$

We can also lower bound it by approximately $\log n$, so $E((X_1^n)^2) = \log n + o(\log n)$. Using LF CLT, take

$$\frac{X_1^n}{\sqrt{n\log n}}, \frac{X_2^n}{\sqrt{n\log n}}, \dots, \frac{X_n^n}{\sqrt{n\log n}}.$$

We show the sum of variances to a limit, and contributions from large variances add up asymptotically to 0.

Note that

$$\sum_{i=1}^n Var(\overline{X_{n,i}}) = n\frac{\log n + o(\log n)}{n\log n} \to 1.$$

Then

$$\sum_{i=1}^n E(\overline{X_{n,i}}^2, |\overline{X}_{n,i}| > \epsilon) \to 0,$$

since

$$\frac{X_i 1(|X_i| \le \sqrt{n}\log\log n)}{\sqrt{n\log n}} \le \frac{\log\log n}{\sqrt{\log n}} < \epsilon$$

for large n.

By LF, $\frac{\overline{S_n}}{\sqrt{n\log n}} \xrightarrow{d} N(0,1)$, and by BC it follows that $\frac{S_n}{\sqrt{n\log n}} \xrightarrow{d} N(0,1)$.

§17.2 General Theory of Distributions

We have a separable metric space (S, d) and we define distributions convergence in this context.

Definition 17.1. $X_n \xrightarrow{d} X$ if $Ef(X_n) toEf(X)$ for any bounded continuous function f.

Theorem 31 (Equivalent Conditions for Convergence)
The following are equivalent.

1. $Ef(X_n) \to Ef(X)$ for all bounded continuous f.
2. For closed K, $\limsup P(X_n \in K) \le P(X \in K)$.
3. For open U, $\liminf P(X_n \in U) \ge P(X \in U)$.
4. For A with $P(\partial A) = 0$, $P(X_n \in A) \to P(A)$.
5. If f is a bounded measurable function and D_f denotes the discountinuity points of f, if $P(X \in D_f) = 0$, then $Ef(X_n) \to Ef(X)$.

Remark: We have not stated a Skorokhod Representation Theorem.

Proof. We only show 4 implies 5, the rest are trivial. Suppose $|f(x)| \le K$. Divide $[-K, K]$ into intervals of size ϵ, and call them I_i. Let $A_i = f^{-1}(I_i)$. It suffices to show that $P(X_n \in A_i) \to P(X \in A_i)$. Note that $\partial A_i \subset D_f \cap f^{-1}(i\epsilon) \cup f^{-1}((i+1)\epsilon)$. Then, we choose the partition so that the boundary points have 0 mass(which is possible since the set of boundary points with positive mass is countable). Hence, the boundary probability is 0 and the result follows. □

§17.3 Convergence in $\mathbb{R}^\mathbf{d}$

We can talk about distribution functions $F(X) \uparrow 1$ as $X \uparrow \infty$ and $F(X) \downarrow 0$ with $X \downarrow -\infty$ that are right continuous and monotone. We also require that all rectangles have positive mass.

Then, we define weak convergence as $F_n(x) \to F(x)$ for all continuity points of X. In $d = 1$, there were at most countably many discontinuity points. In $d > 1$, this is false. For example, take $Y = U[0, 1]$, $X = 0$ and $(0, Y)$. The distribution function is given by

$$F(x, y) = \begin{cases} 1, x \ge 0, y \ge 1 \\ y, x \ge 0, 0 \le y \le 1, \\ 0 \text{ else.} \end{cases}$$

which is discontinuous at each $(0, y)$.

Exercise 17.2. For each coordinate, the discontinuity points for each coordinate D_i is countable.

Theorem 32
$F_n(x) \to F(x)$ at continuity points is equivalent to $X_n \to X$ in distribution.

Proof. If $X_n \to X$ in distribution, then $F_n(x) \to F(x)$ and x is a continuity point, since if x is a continuity point, the hyperplane passing through x has mass 0 so $P(X \in \partial A) = 0$ so $P(X_n \in A) \to P(X \in A)$, and by definition, this is $F_n(x) \to F(x)$.

If $F_n(x) \to F(x)$ for all continuity points, then given $A = (a_1 \times b_1] \times \cdots \times (a_d \times b_d]$, then $P(X_n \in A) \to P(X \in A)$. Observing that any open set can be approximated from the inside by a disjoint union of such rectangles, we find that $\liminf P(X_n \in U) \ge \lim P(X_n \in B) = P(X \in B) \approx P(X \in U)$. □