Probability Theory

Mathematics C218A/Statistics C205A
Fall 2016

Professor David Aldous
Scribe: Sinho Chewi

Contents

Lecture 1

August 25

1.1 Measure Theory (MT): Conceptual Overview

MT is useful because the definitions from measure theory can be adapted for probability theory. The freshman definition of a random variable (RV) is an object with a range of possible values, the actual value of which is determined by chance. In MT, a RV is a measurable function.

We have already seen:

- Precalculus: $\sum_n f(n)$
- Calculus 1: $\int_a^b f(x)\,\mathrm{d}x$
- Calculus 2: $\iint f(x,y)\,\mathrm{d}x\,\mathrm{d}y$
- Probability: EX

MT provides the abstract integral, $f \mapsto I(f)$ (a definite integral), which unifies the above concepts. MT also answers questions such as: if $f_n \to f_\infty$ (in some sense), does $I(f_n) \to I(f_\infty)$?

"Pick a point x uniformly at random in the unit square." In basic probability theory, the answer is

$$P(x \in A) = \frac{\text{area}(A)}{1}$$

However, we need MT in order to formalize "area(A)".

1.2 Abstract Measure Theory

Denote the universal set by S.

A, B, and C denote subsets of S.

$\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots, \mathcal{F}, \ldots \mathcal{S}$ denote *collections* of subsets. For example, we can have $\mathcal{F} = \{\varnothing, A, B, S\}$.

An element of S is denoted by $s \in S$.

Definition 1.1. $\mathcal{S}$ is a **field** (or **algebra**) if $\mathcal{S}$ is closed under Boolean operations. That is,

1. If $A, B \in \mathcal{S}$, then $A \cap B, A \cup B, A \backslash B, \ldots$ must be in $\mathcal{S}$.

2. $\mathcal{S}$ is non-empty.

$\mathcal{F} = \{\varnothing, S\}$ *is* a field. $\mathcal{F} = \{\varnothing, A, A^c, S\}$ *is* a field.

Exercise. To determine whether a collection is a field, it is enough to check:

- $A \in \mathcal{S} \implies A^c \in \mathcal{S}$
- $A, B \in \mathcal{S} \implies A \cup B \in \mathcal{S}$

Let S be fixed.

Lemma 1.2. *If $\mathcal{S}_1$ and $\mathcal{S}_2$ are fields, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is a field.*

More generally, if $\{\mathcal{S}_\theta : \theta \in \Theta\}$ is any collection of fields in S, then $\bigcap_{\theta\in\Theta} \mathcal{S}_\theta$ is a field.

The above statement is *not true* for $\mathcal{S}_1 \cup \mathcal{S}_2$.

Definition 1.3. Let $\mathcal{A}$ be any collection of subsets of S. Then

$$\mathcal{F}(\mathcal{A}) \overset{\text{def}}{=} \bigcap_{\substack{\mathcal{F} \text{ a field} \\ \mathcal{F} \supseteq \mathcal{A}}} \mathcal{F}$$

is a field by (1.2).

$\mathcal{F}(\mathcal{A})$ is called **the field generated by** $\mathcal{A}$.

Exercise. "$\mathcal{F}(\mathcal{A})$ is the collection of subsets that can be obtained from sets in $\mathcal{A}$ via a finite number of Boolean operations."

Example 1.4. Let $S = \mathbb{R}^1$ and $\mathcal{A}$ be the collection of intervals $(-\infty, x]$, $x \in \mathbb{R}$.

Then $\mathcal{F}(\mathcal{A})$ is the collection of finite disjoint intervals in $\mathbb{R}^1$.

Example 1.5. Let $S = [0,1]^2$ and $\mathcal{A}$ be the collection of rectangles $(x_1, x_2] \times (y_1, y_2]$.

Then $\mathcal{F}(\mathcal{A})$ includes finite unions of connected areas which are made up of finite numbers of horizontal and vertical lines.

Definition 1.6. $\mathcal{S}$ is a σ**-field** (σ**-algebra**) if

1. $\mathcal{S}$ is a field.
2. $\mathcal{S}$ is closed under countable unions and under countable intersections. (If $A_i \in \mathcal{S}$, $1 \le i < \infty$, then $\bigcup_i A_i$ and $\bigcap_i A_i$ are in $\mathcal{S}$.)

Exercise. For 2, it is enough to prove closure under increasing unions: If $A_i \in \mathcal{S}$, $A_1 \subset A_2 \subset A_3 \subset \cdots$, then $\bigcup_i A_i \in \mathcal{S}$.

Lemma 1.7. *If $\mathcal{S}_1$ and $\mathcal{S}_2$ are σ-fields, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is a σ-field.*

More generally, $\{\mathcal{S}_\theta : \theta \in \Theta\}$ is any collection of σ-fields in S, then $\bigcap_{\theta\in\Theta} \mathcal{S}_\theta$ is a σ-field.

Definition 1.8. Let $\mathcal{A}$ be any collection of subsets of S.

Then

$$\sigma(\mathcal{A}) \stackrel{\text{def}}{=} \bigcap_{\substack{\mathcal{G} \text{ a } \sigma\text{-field} \\ \mathcal{G} \supseteq \mathcal{A}}} \mathcal{G}$$

is a σ-field, called the **σ-field generated by $\mathcal{A}$**.

However, there is no useful explicit description of a σ-field.

Definition 1.9. A **measurable space** is a pair $(S, \mathcal{S})$ where S is a set and $\mathcal{S}$ is a σ-field on S.

If S is a topological space and $\mathcal{G}$ is the collection of open sets, then $\sigma(\mathcal{G})$ is called the **Borel σ-field** on S.

Exercise. On $\mathbb{R}^1$ or $\mathbb{R}^d$, the Borel σ-field is the same σ-field generated by the d-dimensional cubes

$$(x_1, y_1] \times (x_2, y_2] \times \cdots \times (x_d, y_d].$$

Lecture 2

August 30

2.1 Measurable Functions

Last time, we talked about a measurable space $(S, \mathcal{S})$.

If S is a topological space, we use $\mathcal{B} = \sigma(\{\text{open sets}\})$ as $\mathcal{S}$, in particular for $S = \mathbb{R}$.

Take sets S_1, S_2 and a function $f : S_1 \to S_2$. For $A \subseteq S_1$, we can define $f(A) = \{f(s_1) : s_1 \in A\} \subseteq S_2$. For $B \subseteq S_2$, we can define $f^{-1}(B) = \{s_1 : f(s_1) \in B\} \subseteq S_1$.

- f^{-1} commutes with Boolean operations and monotone limits:

$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2) \tag{2.1}$$

$$f^{-1}\left(\bigcup_n B_n\right) = \bigcup_n f^{-1}(B_n) \tag{2.2}$$

Note: Given $f : S_1 \to S_2$, given $\mathcal{S}_2$, then $\{f^{-1}(B) : B \in \mathcal{S}_2\}$ is a σ-field on S_1.

Take two measurable spaces $(S_1, \mathcal{S}_2)$ and $(S_2, \mathcal{S}_2)$.

Definition 2.1. A function $f : S_1 \to S_2$ is **measurable** if

$$f^{-1}(B) \in \mathcal{S}_1, \quad \forall B \in \mathcal{S}_2 \tag{2.3}$$

Lemma 2.2. *To check if f is measurable, it is sufficient to check (2.3) for all $B \in \mathcal{B}$, where $\mathcal{B}$ is some collection such that $\sigma(\mathcal{B}) = \mathcal{S}_2$.*

Proof. Consider $\{B \subseteq S_2 : f^{-1}(B) \in \mathcal{S}_1\}$. This is a σ-field because of 2.1 and 2.2 and is a subset of $\mathcal{B}$. If a σ-field $\mathcal{S}$ is a subset of a collection $\mathcal{B}$, then $\mathcal{S} \supseteq \sigma(\mathcal{B})$. Hence, this is a subset of $\sigma(\mathcal{B})$. □

Lemma 2.3. *If S_1, S_2 are toplogical spaces and $f : S_1 \to S_2$ is continuous, then f is measurable.*

Proof. A function f is continuous if and only if $f^{-1}(G_2) \in \{\text{open sets in } S_1\}$, where G_2 is open in S_2. Then 2.2 implies that f is measurable with respect to $\sigma(\{\text{open sets in } S_1\}) = \mathcal{S}_1$. □

Lemma 2.4. *If $S_2 = \mathbb{R}$, it is sufficient to check $f^{-1}((-\infty, x]) \in \mathcal{S}_1 \; \forall x \in \mathbb{R}$.*

Lemma 2.5. *Suppose $h : S_1 \to S_2$ and $g : S_2 \to S_3$, with $f(s_1) = g(h(s_1))$. If g and h are measurable, then $f = g \circ h$ is measurable.*

Lemma 2.6. *Suppose $f_i : (S, \mathcal{S}_1) \to \mathbb{R}$ is a measurable function, $1 \le i \le d$. Suppose $g : \mathbb{R}^d \to \mathbb{R}$ is measurable. Then $g(f_1(s_1), f_2(s_1), \dots, f_d(s_1))$ is a measurable function $S \to \mathbb{R}$.*

Proof. Apply 2.5 to $(S_1, \mathbb{R}^d, \mathbb{R})$ and $h(s_1) = (f_1(s_1), \dots, f_d(s_1))$. All we need to prove is that $h : S \to \mathbb{R}^d$ is measurable. Use the fact that

$$\mathcal{B}^d = \text{Borel } \sigma\text{-field on } \mathbb{R}^d = \sigma\text{-field generated by } \{(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_d]\} = B$$

$$h^{-1}(B) = \bigcap_{i=1}^{d} \{s_1 : f_i(s_1) \subseteq x_i\} \in \mathcal{S}_1$$

We are done by 2.2. □

Corollary 2.7. *If $f_i : S \to \mathbb{R}$ are measurable, then $f_1 + f_2$, $f_1 f_2$, and $\max(f_1, f_2)$ are measurable.*

Proof. The functions $g(x_1, x_2) = x_1 + x_2$, $g(x_1, x_2) = x_1 x_2$, and $g(x_1, x_2) = \max(x_2, x_2)$ for $x_i \in \mathbb{R}$ are continuous, which implies that the functions are measurable. □

Reminder: Let $\bar{\mathbb{R}} = [-\infty, \infty]$. For arbitrary $x_n \in \bar{\mathbb{R}}$, $1 \le n < \infty$, $\limsup_n x_n$ exists in $\bar{\mathbb{R}}$.

$$\lim_{N \to \infty} \sup_{n \ge N} x_n = \limsup x_n$$

$$\lim_{N \to \infty} \inf_{n \ge N} x_n = \liminf x_n$$

$\lim_n x_n$ exists iff $\limsup x_n = \liminf x_n$.

Lemma 2.8. *Given measurable $f_i : S \to \bar{\mathbb{R}}$, $1 \le i < \infty$, define $f^*(s) = \limsup_{n \to \infty} f_n(s)$ and $f_*(s) = \liminf_{n \to \infty} f_n(s)$. Then f^* and f_* are measurable functions $S \to \bar{\mathbb{R}}$.*

Proof. Consider

$$\left\{ s : \limsup_n f_n(s) \le x \right\} = \{s : f_n(s) \le x + 1/i \text{ ultimately (for all sufficiently large } n\text{), for each } i\}$$

$$= \bigcap_{i=1}^{\infty} \{s : f_n(s) \le x + 1/i \text{ ultimately}\}$$

$$= \bigcap_{i=1}^{\infty} \bigcup_{N=1}^{\infty} \{s : f_n(s) \le x + 1/i \; \forall n \ge N\}$$

$$= \bigcap_{i=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \underbrace{\{s : f_n(s) \le x + 1/i\}}_{\in \mathcal{S}}$$

□

2.2 On $\mathbb{R}$-Valued Measurable Functions $(S, \mathcal{S}) \to \mathbb{R}$

For $A \in \mathcal{S}$, the indicator function

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

is a measurable function.

Given real numbers c_i, $1 \le i \le n$ and given a partition $(A_i, 1 \le i \le n)$ of S into measurable sets, define $f(s) = \sum_i c_i 1_{A_i} = c_i$ for $s \in A_i$ (a "simple function").

Lemma 2.9. *Let $h : S \to [0, L]$ be measurable. For $i \ge 1$, define*

$$0 \le h_i(s) = \max_{j \ge 0} \left\{ \frac{j}{2^i} : \frac{j}{2^i} \le h(s) \right\} = 2^{-i} \lfloor 2^i h(s) \rfloor \le h(s)$$

Then $h_i(s) \uparrow h(s)$ as $i \to \infty$, and each h_i is a simple function.

Proof. "Obvious". □

2.3 Measures

Take a measurable space $(S, \mathcal{S})$.

Definition 2.10. A measure μ is a function $\mu : \mathcal{S} \to [0, \infty]$ such that

1. $\mu(\varnothing) = 0$
2. For countable disjoint $A_i \in \mathcal{S}$, $\mu(\bigcup_i A_i) = \sum_i \mu(A_i) \le \infty$.

Condition 2 is *countable additivity*.

- If $\mu(S) = 1$, we call μ a **probability measure**.
- If $\mu(S) < \infty$, call μ a **finite measure**.
- If $\exists S_n \uparrow S$ such that $\mu(S_n) < \infty \; \forall n$, then μ is a **σ-finite measure**.

2.3.1 Elementary Properties

- If $A \subseteq B$, then $\mu(A) \le \mu(B)$.
- $\mu(A \cup B) \le \mu(A) + \mu(B)$
- For a probability measure, $\mu(A^c) = 1 - \mu(A)$.

2.3.2 Monotonocity

If $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A) \le \infty$. If $A_n \downarrow A$ and $\mu(A_n) < \infty$, then $\mu(A_n) \downarrow \mu(A)$.

"Continuity": If $A_n \downarrow \varnothing$, if some $\mu(A_n) < \infty$, then $\mu(A_n) \downarrow 0$.

Lecture 3

September 1

3.1 Probability Measure μ on $(S, \mathcal{S})$

- $\mu(\varnothing) = 0$
- For disjoint $(A_i, 1 \le i < \infty)$, $A_i \in \mathcal{S}$,

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i), \quad 0 \le \mu(A) \le 1$$

Take the case of $S = \{0, 1, 2, \dots\}$ and $\mathcal{S} =$ all subsets of S.

- Given $p_0, p_1, p_2, \dots \ge 0$, with

$$\sum_i p_i = 1 \tag{3.1}$$

  Define, for $A \subset S$, $\mu(A) = \sum_{i \in A} p_i$. This μ is a probability measure (PM).

- Given a PM μ on this S, define $p_i = \mu(\{i\})$ and (3.1) holds.

Consider a set S and let $\mathcal{A}$ and $\mathcal{C}$ denote classes of subsets of S.

Call $\mathcal{A}$ a π-class if $A_1, A_2 \in \mathcal{A} \implies A_1 \cap A_2 \in \mathcal{A}$.

Call $\mathcal{C}$ a λ-class if

1. $S \in \mathcal{C}$
2. If $A, B \in \mathcal{C}$, if $A \subset B$, then $B \setminus A \in \mathcal{C}$.
3. If $A_n \in \mathcal{C}$, if $A_n \uparrow A$, then $A \in \mathcal{C}$.

Lemma 3.1 (Dynkin's π-λ Class Lemma)**.** *If $\mathcal{C}$ is a λ-class, if $\mathcal{A}$ is a π-class, and if $\mathcal{C} \supseteq \mathcal{A}$, then $\mathcal{C} \supseteq \sigma(\mathcal{A})$.*

Proof. See text for proof. □

Lemma 3.2 (Identification Lemma for PMs)**.** *If μ_1 and μ_2 are PMs on $(S, \mathcal{S})$, if $\mu_1(A) = \mu_2(A) \forall A \in \mathcal{A}$, if $\mathcal{A}$ is a π-class, and if $\mathcal{S} = \sigma(\mathcal{A})$, then $\mu_1 = \mu_2$ ($\mu_1(B) = \mu_2(B)\ \forall B \in \mathcal{S}$).*

Proof. Consider the collection $\mathcal{C} \stackrel{\text{def}}{=} \{A : \mu_1(A) = \mu_2(A)\}$, so $\mathcal{C} \supseteq \mathcal{A}$ by hypothesis. To apply 3.1, we only need to check $\mathcal{C}$ is a λ-class (clear from the definition of a PM). □

Theorem 3.3.
- *There exists a σ-finite measure λ on $(\mathbb{R}^1, \mathcal{B}^1)$ such that $\lambda([a,b]) = b - a$ for all $-\infty < a < b < \infty$. This is the **Lebesgue measure on** $\mathbb{R}$ ("length").*
- *There exists a PM λ_1 on $[0,1]$ such that $\lambda_1([a,b)) = b - a$ for all $0 \le a \le b \le 1$. This is the **Lebesgue measure on** $[0,1]$ or the **uniform distribution on** $[0,1]$.*

Proof. See text for proof. □

Consider $f : (S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$, a measurable function. We know that for $B \in \mathcal{S}_2$, $f^{-1}(B) \in \mathcal{S}_1$. Given a PM μ on $(S_1, \mathcal{S}_2)$, we can define a PM $\hat{\mu}$ on $(S_2, \mathcal{S}_2)$ by

$$\hat{\mu}(B) = \mu(f^{-1}(B))$$

This $\hat{\mu}$ is a PM because f^{-1} commutes with Boolean operations.

3.2 Probability Measures on $\mathbb{R}^1$

Given a PM μ on $\mathbb{R}$, define $F(x) = \mu((-\infty, x])$. This F has the properties

- increasing: $x_1 \le x_2 \implies F(x_1) \le F(x_2)$
- right-continuous: if $x_n \downarrow x$, then $F(x_n) \downarrow F(x)$
- $\lim_{x\to\infty} = 1$ and $\lim_{x\to-\infty} F(x) = 0$

A function F with these properties is called a **distribution function**.

Theorem 3.4. *Given a distribution function F, there exists a unique PM μ such that*

$$F(x) = \mu((-\infty, x]) \quad \forall x$$

Undergraduate Version. Take U a RV Uniform$[0,1]$. Then $F^{-1}(U)$ is a RV with distribution function F.

Define G (a version of F^{-1}):

$$\begin{aligned} G(y) &= \sup\{x : F(x) < y\}, && 0 < y < 1 \\ &= \inf\{x : F(x) \ge y\}, && 0 < y < 1 \end{aligned}$$

G is increasing, so G is measurable. For each x:

$$G^{-1}((-\infty, x]) = \{y : G(y) \le x\} = \{y : y \le F(x)\} = [0, F(x)]$$

The "push-forward" lemma says that there exists a PM $\hat{\mu}$ on $\mathbb{R}$ such that

$$\hat{\mu}((-\infty, x]) = \lambda_1(G^{-1}((-\infty, x])) = \lambda_1([0, F(x)]) = F(x)$$

3.3 Coin-Tossing Space

Take a 2-element set $B = \{H, T\}$ or $\{0, 1\}$.

The infinite product space $B^\infty = B^{\mathbb{N}}$ is the set of all $\mathbf{b} = (b_1, b_2, b_3, \dots)$, $b_i \in B$. Given a finite string $\pi = (\pi_1, \dots, \pi_n)$, $\pi_i \in B$, the length is $n = |\pi|$.

Set $A_\pi \subseteq B^\infty$, where $A_\pi = \{\mathbf{b} : (b_1, \dots, b_{|\pi|}) = (\pi_1, \dots, \pi_{|\pi|})\}$.

Define a σ-field $\mathcal{B}^\infty$ on B^∞ as $\sigma(\text{all } A_\pi;\ \pi \text{ a finite string})$.

Theorem 3.5. *There exists a PM μ on $(B^\infty, \mathcal{B}^\infty)$ such that*

$$\mu(A_\pi) = \frac{1}{2^{|\pi|}}, \quad \forall \pi$$

Conceptual Point. This theorem is equivalent to the theorem that λ_1 exists.

The binary expansion of real $x \in (0,1)$ (for example, $x = 0.110110010001\ldots$) is given by

$$x = 0.b_1(x)b_2(x)b_3(x)\ldots, \quad b_i(x) = \begin{cases} 1, & \text{if } 2^i x \text{ is odd} \\ 0, & \text{if } 2^i x \text{ is even} \end{cases}$$

The function $x \mapsto b_i(x)$ is measurable.

Define $g : [0,1] \to B^\infty$ by $g(x) = (b_1(x), b_2(x), \ldots)$. It is easily checked that g is measurable. Use the push-forward lemma to set a PM μ on B^∞ with

$$\mu(A_\pi) = \lambda_1(\{x : g(x) \in A_\pi\}) = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \frac{1}{2^n}$$

for some k, if $|\pi| = n$.

Given μ on B^∞, define $h : B^\infty \to [0,1]$ by

$$h(k) = \sum_i 2^{-i} b_i$$

The push-forward is λ_1.

Lecture 4

September 6

4.1 Abstract Integration (MT Version)

Setting. Let μ be a measure (finite or σ-finite) on $(S, \mathcal{S})$.

Let $\mathcal{H}_+$ be the set of measurable $h : S \to [0, \infty]$.

Theorem 4.1 (Basic Theorem)**.** *There exists a unique map $I : \mathcal{H}_+ \to [0, \infty]$ such that*

1. $I(1_A) = \mu(A)$, $\forall A \in \mathcal{S}$
2. $I(h_1 + h_2) = I(h_1) + I(h_2)$, $\forall h_i \in \mathcal{H}_+$
3. $I(ch) = cI(h)$, $\forall h \in \mathcal{H}_+$, $\forall c \geq 0$
4. *If* $0 \leq h_n \uparrow h \in \mathcal{H}_+$, *then* $I(h_n) \uparrow I(h) \leq \infty$

Background. $h \mapsto \int_{-\infty}^{\infty} h(x)\,\mathrm{d}x$ will be the case $S = \mathbb{R}^1$, μ is the Lebesgue measure.

In practice, we write

$$I(h) = \int_S h\,\mathrm{d}\mu = \int_S h(s)\,\mu(\mathrm{d}s)$$

For $A \in \mathcal{S}$,

$$\int_A h\,\mathrm{d}\mu \stackrel{\text{def}}{=} \int_S (h 1_A)\,\mathrm{d}\mu$$

These are *definite integrals.* We associate integrals with the area under curves. The area of a rectangle of height c and length $\mu(A)$ is $c\mu(A) = c\int_S 1_A\,\mathrm{d}\mu$.

Steps:

1. Define $I(1_A) = \mu(A)$.
2. For simple $h = \sum_i c_i 1_{A_i}$, define $I(h) = \sum_i c_i \mu(A_i)$.
3. For $0 \leq h \leq m$, for m a constant, we can write $h = \lim_n h_n$, with h_n simple (2.9) and define $I(h) = \lim_n I(h_n)$.
4. For general $h \in \mathcal{H}_+$, set $h_m = \min(h, m)$, so $h_m \uparrow h$. Define $I(h) = \lim_{m\uparrow\infty} I(h_m)$.

Note: Consider

$$h(s) = \begin{cases} \infty, & s \in A \\ 0, & s \notin A \end{cases}$$

where $\mu(A) = 0$. Here, $h_m(s) = \min(h(s), m) = m1_A$, so $I(h_m) = m \cdot \mu(A) = 0$. Then

$$I(h) = \lim_{m\uparrow\infty} I(h_m) = 0$$

Notation. (Almost Everywhere)

$$h_1 = h_2 \quad \text{a.e.}$$

means $\{s : h_1(s) \neq h_2(s)\}$ has μ-measure 0.

Notation. For $x \in \mathbb{R}$, $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. Thus, $x = x^+ - x^-$, $|x| = x^+ + x^-$, and $|x - y| \leq |x| + |y|$.

Definition 4.2. A measurable $h : S \to \bar{\mathbb{R}}$ is **integrable** (w.r.t. μ) if $\int_S |h| \,\mathrm{d}\mu < \infty$. For integrable h, define $I(h) = I(h^+) - I(h^-)$ (but finite).

Lemma 4.3. *Suppose h_1, h_2 are integrable.*

1. *(Linearity) For $c_1, c_2 \in \mathbb{R}$, $h \overset{\text{def}}{=} c_1h_1 + c_2h_2$, then h is integrable and $\int h \,\mathrm{d}\mu = c_1 \int h_1 \,\mathrm{d}\mu + c_2 \int h_2 \,\mathrm{d}\mu$.*
2. *If $h_1 = 0$ a.e., then $\int h_1 \,\mathrm{d}\mu = 0$.*
3. *If $h_1 \geq 0$ a.e., then $\int h_1 \,\mathrm{d}\mu \geq 0$.*
4. *If $h_1 \leq h_2$ a.e., then $\int h_1 \,\mathrm{d}\mu \leq \int h_2 \,\mathrm{d}\mu$.*
5. *$\left|\int h \,\mathrm{d}\mu\right| \leq \int |h| \,\mathrm{d}\mu$.*

Proof. 5.

$$\begin{aligned}\left|\int h \,\mathrm{d}\mu\right| &= \left|\int h^+ \,\mathrm{d}\mu - \int h^- \,\mathrm{d}\mu\right| \\ &\leq \left|\int h^+ \,\mathrm{d}\mu\right| + \left|\int h^- \,\mathrm{d}\mu\right| \\ &= \int (h^+ + h^-) \,\mathrm{d}\mu = \int |h| \,\mathrm{d}\mu\end{aligned}$$

□

4.2 Probability Theory (MT Version)

Freshman Version. A RV X is a quantity with a range of possible values, the actual value of which is determined somehow by chance.

$P(X \leq 4)$ is "the chance it turns out that $X \leq 4$".

A **probability space** is

$$(\underbrace{\Omega}_{\text{states of universe}}, \underbrace{\mathcal{F}}_{\text{events, } \sigma\text{-field on } \Omega}, \underbrace{P}_{\text{PM}})$$

Events $A \in \mathcal{F}$ have probabilities $P(A)$.

A **random variable** (RV) is a measurable function $X : \Omega \to (S, \mathcal{S})$ or often $\mathbb{R}$.

For a measurable set $B \in \mathcal{S}$, $\{\omega : X(\omega) \in B\}$ is an event in $\mathcal{F}$ and so has a probability

$$P(\{\omega : X(\omega) \in B\}) = P(X \in B)$$

A given RV $X : \Omega \to (S, \mathcal{S})$ has a **distribution** (or **law**) μ, defined by $\mu(B) = P(X \in B)$. Given a PM P and the RV X, we obtain the push-forward PM μ.

Notation. By example: If X, Y, Z are $\mathbb{R}$-valued RVs, we define **almost surely** (a.s.):

$$X^2 + Y^2 \le Z + 4 \quad \text{a.s.} \qquad \text{means} \qquad \begin{aligned} &P(X^2 + Y^2 \le Z + 4) = 1 \\ &P(\{\omega : X^2(\omega) + Y^2(\omega) \le Z(\omega) + 4\}) = 1 \end{aligned}$$

Given $\mathbb{R}$-valued RVs X_n, X,

$$X_n \to X \quad \text{a.s.} \qquad \text{means} \qquad P(\{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = 1$$

Note: Given arbitrary $\mathbb{R}$-valued X_n, $1 \le n < \infty$, we can deifne $X^* = \limsup_n X_n$ ($X^*(\omega) = \limsup_{n\to\infty} X_n(\omega)$) and X^* is a RV.

Take a RV $Y : (\Omega, \mathcal{F}, P) \to \mathbb{R}$. Then

$$E[Y] \overset{\text{def}}{=} \int_\Omega Y \, \mathrm{d}P$$

provided $E|Y| \equiv \int_\Omega |Y| \, \mathrm{d}P < \infty$. "Y is Ω-integrable."

4.2.1 "Change of Variable" Lemmas

Consider $X : (\Omega, P) \to (S, \mathcal{S})$ and $h : (S, \mathcal{S}) \to \mathbb{R}$.

Lemma 4.4. *If $h(X)$ is integrable, then $Eh(X) = \int_S h \, \mathrm{d}\mu$ for $\mu =$ distribution of X.*

Lemma 4.5. *If ν is a PM on $\mathbb{R}$ with density f, then $\int_\mathbb{R} h \, \mathrm{d}\nu = \int_{-\infty}^{\infty} h(x) f(x) \, \mathrm{d}x$, provided h is ν-integrable.*

Proof. Consider the collection of h for which the stated equality is true.

1. Consider $h = 1_B$, $B \in \mathcal{S}$.

$$\text{LHS} = Eh(X) = E1_{X\in B} = P(X \in B) = \mu(B) = \int 1_B \, \mathrm{d}\mu = \text{RHS}$$

2. Consider $h = 1_B$, $B \subseteq \mathbb{R}$.

$$\text{LHS} = \int 1_B \, \mathrm{d}\nu = \nu(B) = \int_B f(x) \, \mathrm{d}x = \text{RHS} \qquad \text{(definition of density } f(x) \text{ of } \nu)$$

Go through the steps of the sketch proof of 4.1. Both sides of the equalities are integrals. Then:

true for 1_B $\implies$ true for simple h $\implies$ true for bounded measurable h $\implies$ true for integrable h

See the textbook: "monotone class theorem". □

We can combine 4.4 and 4.5.

Lemma 4.6. *Suppose X is $\mathbb{R}$-valued, and its distribution has density f. Then $Eh(X) = \int h(x) f(x) \, \mathrm{d}x$, provided that $h(X)$ is integrable.*

$$\begin{aligned} EX &= \int x f(x)\,\mathrm{d}x \\ EX^2 &= \int x^2 f(x)\,\mathrm{d}x \\ &\text{etc.} \end{aligned}$$

Lecture 5

September 8

5.1 Expectation (Undergraduate Version)

1. EX is the limit of $(X_1 + X_2 + \cdots + X_n)/n$ for IID RVs. *We will prove this later as the SLLN.*
2. EX is the fair stake for a random payoff X. *This is the conceptual basis of martingale theory.*
3. $EX = \sum_i iP(X = i)$ or $\int xf(x)\,\mathrm{d}x$.
4. $Eh(X) = \sum_i h(i)P(X = i)$ or $\int h(x)f(x)\,\mathrm{d}x$. *We checked these in MT (last class).*
5. Abstract rules: $E(X + Y) = EX + EY$, even if X and Y are dependent.

5.2 Expectation & Inequalities (MT Version)

If $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, then

$$EX \overset{\text{def}}{=} \int_\Omega X(\omega)P(\mathrm{d}\omega) \tag{5.1}$$

EX is well-defined if

1. $E|X| < \infty$ $(-\infty < EX < \infty)$,
2. or $0 \leq X \leq \infty$, where $0 \leq EX \leq \infty$.

From the definition (5.1), we can use the properties of the abstract integral.

- $E1_A = P(A)$
- $E(c_1X_1 + c_2X_2) = c_1EX_1 + c_2EX_2$
- **Monotone convergence**: If $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$, so $X_n \uparrow X_\infty$ a.s. (holds for all ω outside some A, $P(A) = 0$), then $EX_n \uparrow EX_\infty \leq \infty$. Consider $0 \leq X_1 1_{A^c} \leq X_2 1_{A^c} \leq \cdots$. Then $X_n 1_{A^c} \uparrow X_\infty 1_{A^c}\ \forall\omega$ and $EX_n = EX_n 1_{A^c}$.
- If $X \geq 0$, if $EX < \infty$, then $P(X < \infty) = 1$. If $P(X < \infty) = 1$, it may not be true that $EX < \infty$. For example, consider $P(X = i) \sim ci^{-3/2}$.

Let X, Y be $\mathbb{R}$-valued RVs.

Markov's Inequality: If $X \geq 0$, $EX < \infty$, then

$$P(X \geq x) \leq \frac{EX}{x}, \quad 0 < x < \infty$$

Chebyshev's Inequality: If $EX^2 < \infty$, then $\text{var}(X) \overset{\text{def}}{=} EX^2 - (EX)^2 = E(X - EX)^2$ and $0 \le \text{var}(X) < \infty$. If $\text{var}(X) < \infty$, then

$$P(|X - EX| \ge x) \le \frac{\text{var}(X)}{x^2}, \quad 0 < x < \infty$$

Theorem 5.1 (General Form of Markov's Inequality)**.** *Let $\phi : \mathbb{R} \to [0, \infty)$ be increasing. Then*

$$P(X \ge x) \le \frac{E\phi(X)}{\phi(x)}, \quad -\infty < x < \infty$$

provided that the quantity is not $0/0$*.*

Proof. Define

$$h(y) = \begin{cases} 0, & y < x \\ \phi(x), & y \ge x \end{cases}$$

so $h(y) = \phi(x) 1_{(y \ge x)}$. Then $h(y) \le \phi(y)\ \forall y$. Therefore,

$$E\phi(X) \ge Eh(X) = \phi(x) E1_{(X \ge x)} = \phi(x) P(X \ge x)$$ □

The "special" Markov's inequality is the case of $\phi(x) = x^+ = \max(0, x)$.

To prove Chebyshev: set $Y = |X - EX|$ and $\phi(x) = (x^+)^2$.

$$P(Y \ge x) \le \frac{EY^2}{x^2} = \frac{\text{var}(X)}{x^2}$$

Another case is to take $\phi(x) = e^{\theta x}$ for a parameter $\theta > 0$.

$$P(X \ge x) \le \inf_{\theta > 0} \frac{Ee^{\theta X}}{e^{\theta x}} \le \infty, \quad 0 < x < \infty$$

This is called the **Basic Large Deviation Inequality**. The inequality is only useful if $P(X > x) \to 0$ exponentially fast.

Suppose $X \sim \text{Poisson}(\lambda)$. Then $EX = \lambda$ and $\text{var}\, X = \lambda$. Taking $x > \lambda$, Markov gives $P(X > x) \le \lambda/x$ and Chebyshev gives $P(X > x) \le \lambda/(x - \lambda)^2$. We have

$$Ee^{\theta X} = \sum_i \frac{e^{\theta i} e^{-\lambda} \lambda^i}{i!} = e^{-\lambda} \exp(\lambda e^\theta)$$

Minimizing this, we obtain $0 = -x + \lambda e^\theta$. Take θ with $\lambda e^\theta = x$.

$$\begin{aligned} P(X \ge x) &\le \inf_\theta \exp(-\theta x - \lambda + \lambda e^\theta) \\ &= \exp\left(-x \log \frac{x}{\lambda} - \lambda + x\right) \end{aligned}$$

Theorem 5.2 (Cauchy-Schwarz Inequality)**.**

$$|E(XY)| \le \sqrt{(EX^2)(EY^2)}$$

Proof. (Trick!)

$$a > 0, ax^2 + 2bx + c \geq 0 \,\forall x \Leftrightarrow b^2 \leq ac$$

$$E(X + xY)^2 = \underbrace{EY^2}_{a>0} \cdot x^2 + 2\underbrace{E(XY)}_{b} \cdot x + \underbrace{EX^2}_{c}$$

Since $b^2 \leq ac$, we are done. □

Note: Given $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in \mathbb{R}$, take $P(X = x_i, Y = y_i) = 1/n, 1 \leq i \leq n$. C-S says

$$\left|\frac{1}{n}\sum_i x_i y_i\right| \leq \sqrt{\left(\frac{1}{n}\sum_i x_i^2\right)\left(\frac{1}{n}\sum_i y_i^2\right)}$$

Similarly for the next inequalities.

Definition 5.3. ϕ is **convex** if $\forall x < y$, $\forall 0 \leq \lambda \leq 1$, $\phi(x + \lambda(y - x)) \leq \phi(x) + \lambda(\phi(y) - \phi(x))$.

In practice: $\phi''(x) \geq 0 \implies \phi$ is convex.

Theorem 5.4 (Jensen's Inequality)**.** *Consider an interval $I \subseteq \mathbb{R}$. Let $\phi : I \to \mathbb{R}$ be convex. Suppose $P(X \in I) = 1$. Then $\phi(EX) \leq E\phi(X)$ provided both expectations are well-defined.*

Proof. Given x and convex ϕ, there exists a "tangent line" $l(y) \leq \phi(y) \,\forall y$ such that $l(x) = \phi(x)$.

Set $x = EX$, take the tangent $l(\cdot)$ at x.

$$E\phi(X) \geq El(X) \underbrace{=}_{\text{linear}} l(EX) = l(x) = \phi(x) = \phi(EX)$$ □

Consider the distribution of $(X, \phi(X))$. Then

$$\begin{aligned} x &= \text{center of mass} \\ &= (EX, E\phi(X)) \end{aligned}$$

Example 5.5. Take $\phi(x) = |x|^p$, $1 \leq p$. Jensen's inequality says $|EY|^p \leq E|Y|^p$. Apply the inequality with $0 \leq a < b < \infty$, $Y = |X|^a$, $p = b/a$. Then $(E|X|^a)^{b/a} \leq E|X|^b$, so

$$(E|X|^a)^{1/a} \leq (E|X|^b)^{1/b} \tag{5.2}$$

Notation. The "L^p norm" is $\|x\|_p \stackrel{\text{def}}{=} (E|X|^p)^{1/p}, 1 \leq p < \infty$ and (5.2) says that $p \mapsto \|x\|_p$ is increasing on $1 \leq p < \infty$.

Example 5.6. Let

$$\phi(x) = 1/x \tag{5.3}$$

or

$$\phi(x) = -\log x \tag{5.4}$$

with $0 < x < \infty$. If $X > 0$, then $E\phi(X) \geq \phi(EX)$.

1. $$E\frac{1}{X} \geq \frac{1}{EX} \Leftrightarrow EX \geq \frac{1}{E(1/X)}$$

2.
$$-E\log X \geq -\log EX \Leftrightarrow EX \geq \exp(E\log X)$$

Consider $x_1, x_2, \ldots, x_n > 0$, $P(X = x_i) = 1/n$, $1 \leq i \leq n$.

$$\underbrace{\frac{1}{n}\sum_i x_i}_{\textbf{arithmetic mean}} \geq \underbrace{\frac{1}{(1/n)\sum_i 1/(x_i)}}_{\textbf{harmonic mean}}$$

and

$$\frac{1}{n}\sum_i x_i \geq \exp\left(\frac{1}{n}\sum_i \log x_i\right) = \left(\prod_i x_i\right)^{1/n} \quad (\textbf{geometric mean})$$

Lecture 6

September 13

6.1 Independence (Undergraduate)

Events A, B are independent if and only if $P(A \cap B) = P(A)P(B)$.

RVs X and Y are independent if and only if $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$.

Idea: Knowing the value of X doesn't change the probabilities for Y.

6.2 MT Setup $(\Omega, \mathcal{F}, P)$

Consider $\mathcal{B}_1$, $\mathcal{B}_2$, sub-σ-fields of $\mathcal{F}$. Call $\mathcal{B}_1$ and $\mathcal{B}_2$ **independent** if $P(B_1 \cap B_2) = P(B_1)P(B_2) \, \forall B_i \in \mathcal{B}_i$.

View X as a map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$. Since X is measurable, $X^{-1}(D) \in \mathcal{F} \, \forall D \in \mathcal{S}$. The collection $\{X^{-1}(D) : D \in \mathcal{S}\}$ is a sub-σ-field of $\mathcal{F}$. Call this $\sigma(X)$, the "σ-field generated by X".

Call the RVs X_1, X_2 independent if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

Theorem 6.1. *For RVs X_1, X_2, where X_i takes on values in $(S_i, \mathcal{S}_i)$, the following are equivalent:*

(i) X_1 *and* X_2 *are independent.*

(ii) $P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2) \, \forall B_i \in \mathcal{S}_i$

(iii) $P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2) \, \forall B_i \in \mathcal{A}_i$, *where $\mathcal{A}_i$ is a π-class, $\sigma(\mathcal{A}_i) = \mathcal{S}_i$.*

(iv) $E[h_1(X_1)h_2(X_2)] = (Eh_1(X_1))(Eh_2(X_2))$ *for all bounded measurable $h_i : S_i \to \mathbb{R}$.*

Comments.

1. (iv) extends to integrable $h_i(X_i)$.
2. If the X_i are $\mathbb{R}$-valued, independence is equivalent to

$$P(X_1 \leq x_2, X_2 \leq x_2) = P(X_1 \leq x_1)P(X_2 \leq x_2) \quad \forall x_i \in \mathbb{R}$$

3. The fact

   If X_1, X_2 are independent, then $g_1(X_1)$, $g_2(X_2)$ are independent (for arbitrary measurable g_i).

   is true because $\sigma(g(X)) \subseteq \sigma(X)$.

Outline Proof. *(i)* $\Leftrightarrow$ *(ii)* by definition.

(iv) $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii)*: Each is a special case of the previous one.

(ii) $\Rightarrow$ *(iv)* by the "monotone class argument". (iv) holds for $h_i = 1_{B_i}$, and therefore holds for h_i simple, and therefore holds for h_i which are bounded and measurable.

What remains is to prove *(iii)* $\Rightarrow$ *(ii)*.

We want to use Dynkin's π-λ Lemma.

- *Step 1.* Fix $B_2 \in \mathcal{A}_2$. Consider the collection

$$\mathcal{L} = \{A \in \mathcal{S}_1 : P(X_1 \in A, X_2 \in B_2) = P(X_1 \in A)P(X_2 \in B_2)\}$$

  Check $\mathcal{L}$ is a λ-class. By hypothesis, $\mathcal{L} \supseteq \mathcal{A}_1$. The Dynkin Lemma implies $\mathcal{L} = \mathcal{S}_1$.

- *Step 2.* Consider $\mathcal{L}' = \{B_2 \in \mathcal{S}_2 : P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)\ \forall B_1 \in \mathcal{S}_1\}$. Check that $\mathcal{L}'$ is a λ-class. Step 1 implies $\mathcal{L}' \supseteq \mathcal{A}_2$. The Dynkin Lemma implies that

$$\mathcal{L}' \supseteq \sigma(\mathcal{A}_2) = \mathcal{S}_2,$$

  which implies (ii).

□

Definition 6.2. $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ are **independent** means

$$P\left(\bigcap_{i=1}^n B_i\right) = \prod_{i=1}^n P(B_i) \quad \forall B_i \in \mathcal{B}_i$$

This is *stronger* than pairwise independence.

Example 6.3. Let X, Y be fair die throws. The events $\{X = 3\}$, $\{Y = 6\}$, and $\{X = Y\}$ are pairwise independent, but not independent.

Example 6.4. Let X_1, X_2 be independent and uniform on $\{0, 1, \ldots, n-1\}$. Define $X_3 = X_1 + X_2$ modulo n. Then (X_1, X_2, X_3) are pairwise independent, but not independent.

Fact. If X_1, X_2, X_3, X_4, X_5 are independent, then $f(X_1, X_2, X_3)$ and $g(X_4, X_5)$ are independent. The important part is that X_1, X_2, X_3 and X_4, X_5 are distinct.

Exercise. Formalize and verify the "hereditary property of independence".

Exercise. To show that events $A_1, A_2, \ldots, A_n$ are independent, it is enough to show

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i) \quad \forall I \subseteq \{1, 2, \ldots, n\}$$

6.3 Real-Valued RVs X_i, Y_i

We know that $X_n \to X_\infty$ a.s. means $P(\{\omega : X_n(\omega) \to X_\infty(\omega) \text{ as } n \to \infty\}) = 1$.

Definition 6.5. **"Convergence in probability"**, $X_n \underset{P}{\longrightarrow} X_\infty$, means that

$$\lim_{n\to\infty} P(|X_n - X_\infty| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

For $1 \le p < \infty$, we say "$X_n \to X_\infty$ in L^p" or "$X_n \xrightarrow{L^p} X_\infty$ to mean $E|X_n - X_\infty|^p \to 0$ as $n \to \infty$ (and $E|X_n|^p < \infty\,\forall n$), that is $\|X_n - X_\infty\|_p \to 0$ (the L^p norm).

Lemma 6.6. *If $X_n \to X_\infty$ in L^p, then $X_n \underset{P}{\longrightarrow} X_\infty$.*

Proof. Use the general form of Markov's inequality, with $\phi(x) = |x|^p$. Apply this to $X_n - X_\infty$.

$$P(|X_n - X_\infty| > \varepsilon) \le \frac{E(|X_n - X_\infty|^p)}{\varepsilon^p} \to 0$$

as $n \to \infty$. □

6.3.1 Variance

If $E(X^2) < \infty$, define $\mathrm{var}(X) = E(X^2) - E(X)^2 = E(X - EX)^2$.

Definition 6.7. If $EX_i^2 < \infty$, if $E(X_1X_2) = (EX_1)(EX_2)$, we say X_1 and X_2 are **uncorrelated**.

Independence implies uncorrelated.

Fact. If $X_1, X_2, \ldots, X_n$ are pairwise uncorrelated, then $\mathrm{var}(\sum_i X_i) = \sum_i \mathrm{var}(X_i)$. (Exercise)

6.3.2 Weak Law of Large Numbers

Theorem 6.8 (L^2 Weak Law of Large Numbers)**.** *Given X_i, $i \ge 1$, suppose that $\sup_i EX_i^2 \le c$, and suppose they are uncorrelated. Write $\mu_i = EX_i$. Write*

$$S_n = \sum_{i=1}^n X_i$$
$$\bar{\mu}_n = \frac{1}{n}\sum_{u=1}^n \mu_i$$

Then $S_n/n - \bar{\mu}_n \to 0$ in L^2 as $n \to \infty$.

Proof.

$$\frac{1}{n}ES_n = \bar{\mu}_n$$
$$\mathrm{var}(S_n) = \sum_{i=1}^n \mathrm{var}(X_i) \le cn$$
$$\mathrm{var}\left(\frac{1}{n}S_n\right) \le \frac{c}{n}$$
$$E\left(\frac{S_n}{n} - \bar{\mu}_n\right)^2 = \mathrm{var}\left(\frac{S_n}{n}\right) \le \frac{c}{n} \to 0$$

as $n \to \infty$. This is convergence in L^2. □

If $\mu_i \to \mu$ as $i \to \infty$, then $\bar{\mu}_n \to \mu$ as $n \to \infty$ and

$$E\left(\frac{S_n}{n} - \mu\right)^2 \to 0$$

Lecture 7

September 15

7.1 Polynomial Approximation Theorem

"X is Bernoulli(p)" means

$$P(X = 1) = p$$
$$P(X = 0) = 1 - p$$

IID means independent and identically distributed.

Theorem 7.1 (Bernstein's Theorem). *Given a continuous function $f : [0,1] \to \mathbb{R}$, define*

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

$f_n(x)$ is a polynomial of degree n. Then $\sup_x |f_n(x) - f(x)| \to 0$ as $n \to \infty$.

Proof. Fix x. Take IID Bernoulli(x) RVs $(X_i, 1 \le i < \infty)$. Write $S_n = \sum_{i=1}^{n} X_i$ and note that

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

We want to bound

$$\begin{aligned}
|f_n(x) - f(x)| &= \left|Ef\left(\frac{S_n}{n}\right) - f(x)\right| \\
&\le E\left|f\left(\frac{S_n}{n}\right) - f(x)\right| \\
&= E\left|f\left(\frac{S_n}{n}\right) - f(x)\right| 1_{(|S_n/n - x| \le \delta)} + E\left|f\left(\frac{S_n}{n}\right) - f(x)\right| 1_{(|S_n/n - x| > \delta)} \\
&\le \varepsilon + 2MP\left(\left|\frac{S_n}{n} - x\right| > \delta\right) \\
&\le \varepsilon + \frac{2M}{\delta^2} \operatorname{var}\left(\frac{S_n}{n}\right) \\
&\le \varepsilon + \frac{2M}{\delta^2} \frac{x(1-x)}{n}
\end{aligned}$$

Explanation: We used $|EY| \le E|Y|$ and $S_n/n \to x$ in probability by the WWLN. From analysis: set

$M \stackrel{\text{def}}{=} \sup |f| < \infty$. "Uniform continuity" of f says that given $\varepsilon > 0$, $\exists \delta > 0$ such that

$$|y_1 - y_2| \le \delta \Rightarrow |f(y_1) - f(y_2)| \le \varepsilon$$

Choose $\varepsilon > 0$ and take δ as in the definition of uniform continuity. Also, $\operatorname{var}(S_n) = n \operatorname{var}(X) = nx(1-x)$ and $x(1-x) \le 1/4$. Then, we know:

$$\sup_n |f_n(x) - f(x)| \le \varepsilon + \frac{M}{2\delta^2}\frac{1}{n}$$
$$\lim_{n\to\infty} \sup_n |f_n(x) - f(x)| \le \varepsilon, \qquad \text{true } \forall \varepsilon > 0$$
$$\lim_{n\to\infty} \sup_n |f_n(x) - f(x)| = 0$$

□

7.2 Background to Proving a.s. Limits

7.2.1 Axioms

If we have events $B_n \uparrow B$, then $P(B_n) \uparrow P(B)$. If $B_n \downarrow B$, then $P(B_n) \downarrow P(B)$.

For arbitrary events A_n, the event that "A_n happens infinitely often" means $\bigcap_{m=1}^\infty \bigcup_{n=m}^\infty A_n$. "$A_n$ ult." means $\bigcup_{m=1}^\infty \bigcap_{n=m}^\infty A_n$. These events are opposites: $(A_n \text{ inf. often})^c = (A_n^c \text{ ult.})$.

If $P(B_m) = 1$, $1 \le m < \infty$, then $P(\bigcap_{m=1}^\infty B_m) = 1$.

Lemma 7.2 (Weak)**.** *(i)* $P(A_n$ *inf. often*$) \ge \limsup_n P(A_n)$

(ii) $P(A_n$ *ult.*$) \le \liminf_n P(A_n)$

Proof.

$$P\left(\bigcup_{n=m}^{Q} A_n\right) \ge \max_{m \le n \le Q} P(A_n)$$

Take $Q \to \infty$.

$$P\left(\bigcup_{n=m}^{\infty}\right) \ge \sup_{n \ge m} P(A_n)$$

Take $m \to \infty$.

$$P(A_n \text{ inf. often}) \ge \limsup_n P(A_n)$$

□

7.2.2 Borel-Cantelli Lemmas

Lemma 7.3 (First Borel-Cantelli-Lemma)**.** *For arbitrary events* $(A_n, 1 \le n < \infty)$, *if* $\sum_n P(A_n) < \infty$, *then* $P(A_n$ *inf. often*$) = 0$.

Proof. Let $X_n = \sum_{i=1}^n 1_{A_i}$ be the number of events that occur. Let $X_\infty = \sum_{i=1}^\infty 1_{A_i} \le \infty$. Then $EX_\infty = \sum_{i=1}^\infty P(A_i) < \infty$ (by hypothesis), which implies that $P(X_\infty = \infty) = 0$. □

Lemma 7.4 (Second Borel-Cantelli Lemma)**.** *For independent events* $(A_i, 1 \le i < \infty)$, $\sum_i P(A_i) = \infty$, *then* $P(A_n$ *inf. often*$) = 1$.

(There are many variants under alternate assumptions.)

Proof. Fix m. We will prove $P(\bigcup_{n=m}^{\infty} A_n) = 1$, or prove $P(\bigcap_{n=m}^{\infty} A_n^c) = 0$.

Fact: If $0 \leq x \leq 1$, then $1 - x \leq e^{-x}$.

Independence implies that

$$\begin{aligned} P\left(\bigcap_{n=m}^{Q} A_n^c\right) &= \prod_{n=m}^{Q} P(A_n^c) \\ &= \prod_{n=m}^{Q} (1 - P(A_n)) \\ &\leq \exp\left(-\sum_{n=m}^{Q} P(A_n)\right) \end{aligned}$$

Let $Q \uparrow \infty$.

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) \leq \exp\left(-\sum_{n=m}^{\infty} P(A_n)\right) = 0$$

□

Lemma 7.5. *Consider arbitrary* $\mathbb{R}$*-valued RVs* (Y_n) *and arbitrary* $-\infty < y < \infty$*. If*

$$\sum_n P(Y_n \geq y + \varepsilon) < \infty$$

for each $\varepsilon > 0$*, then* $\limsup_n Y_n \leq y$ *a.s.*

Corollary 7.6. *If* $\sum_n P(|Y_n| \geq \varepsilon) < \infty$ *for each* $\varepsilon > 0$*, then* $Y_n \to 0$ *a.s.*

Deterministic Fact. For reals (y_n) and y, "$\limsup_n y_n \leq y$" is equivalent to "$y_n \leq y + \varepsilon$ ultimately, for each $\varepsilon > 0$", which is equivalent to "$y_n \leq y + 1/j$ ultimately, for each $j \geq 1$".

Proof. The hypothesis and 7.3 imply that $P(Y_n \leq y + 1/j \text{ ult.}) = 1$ for each j. Since

$$P(B_j) = 1 \quad \forall j \qquad \Longrightarrow \qquad P(B_j \text{ for all } j) = 1$$

then $P(Y_n \leq y + 1/j \text{ ult., for each } j \geq 1) = 1$. By the deterministic fact, $P(\limsup_n Y_n \leq y) = 1$. □

7.3 4th Moment SLLN

SLLN means the **strong law of large numbers**.

Theorem 7.7 (4th Moment SLLN)**.** *Let* $(X_i, 1 \leq i < \infty)$ *be IID,* $EX = 0$*, and* $EX^4 < \infty$*. Write* $S_n = \sum_{i=1}^{n} X_i$*. Then*

(i) $ES_n^4 \leq 3n^2 EX^4$

(ii) $S_n/n \to 0$ *as* $n \to \infty$*.*

If $EX = \mu$, applying the theorem to $X - \mu$ shows that $S_n/n \to \mu$ a.s.

Proof. (i)

$$ES_n^4 = \sum_i \sum_j \sum_k \sum_l E[X_i X_j X_k X_l]$$

Note that $E[X_i X_j X_k X_l] = 0$ if some index "j" appears only once. For example,

$$E(X_4 X_6 X_6 X_6) = E(X_4)E(\cdot) = 0$$

Therefore,

$$\begin{aligned} ES_n^4 &= nEX^4 + \binom{4}{2}\binom{n}{2}E[X_1^2 X_2^2] \\ &= nEX^4 + 3n(n-1)\underbrace{(EX^2)^2}_{\le EX^4} \end{aligned}$$

since $(EY)^2 \le E(Y^2)$.

(ii) Fix $\varepsilon > 0$.

$$\begin{aligned} P\left(\left|\frac{S_n}{n}\right| \ge \varepsilon\right) &\le E\left|\frac{S_n}{n}\right|^4 \cdot \frac{1}{\varepsilon^4} \\ &\le \varepsilon^{-4} n^{-4} \cdot 3n^2 EX^4 \\ &\le 3\varepsilon^{-4} EX^4 n^{-2} \end{aligned}$$

This implies that

$$\sum_n P\left(\left|\frac{S_n}{n} \ge \varepsilon\right|\right) \le \sum_n 3\varepsilon^{-4} EX^4 n^{-2} < \infty$$

By 7.6, $S_n/n \to 0$ a.s. We used the fact that $s^4 = |s|^4$ and $s^2 = |s|^2$, but this does not work for the third moment: $s^3 \ne |s|^3$.

□

Corollary 7.8. *If $(A_i, 1 \le i < \infty)$ are independent* Bernoulli(p), *$S_n = \sum_{i=1}^n 1_{A_i}$, then $S_n/n \to p$ a.s. as $n \to \infty$.*

We say "data" for n real numbers $x_1, \ldots, x_n$. The empirical distribution is the uniform distribution on $(x_1, \ldots, x_n)$. The empirical distribution function is

$$G(x) = \frac{1}{n}\sum_{i=1}^n 1_{(x_i \le x)}$$

Theorem 7.9 (Glivenko-Cantelli Theorem)**.** *If $X_i, 1 \le i < \infty$ are IID with an arbitrary distribution function F, let $G_n(\omega, x)$ be the empirical distribution of $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$, or*

$$G_n(\omega, x) = \frac{1}{n}\sum_{i=1}^n 1_{(X_i(\omega) \le x)}$$

For fixed x, the events $\{X_i \le x\}$ are IID Bernoulli$(G(x))$. *Using the SLLN for events, $G_n(\omega, x) \to G(x)$ as $n \to \infty$.*

Lecture 8

September 20

8.1 Glivenko-Cantelli Theorem

Lemma 8.1. *Let F_n and F be distribution functions. If*

(i) $F_n(x) \to F(x)$ *for each rational* x

(ii) $F_n(x) \to F(x)$ *and* $F_n(x-) \to F(x-)$ *for each* ***atom*** *of* F ($F(x) - F(x-) = P(X = x) > 0$)

then $\sup_x |F_n(x) - F(x)| \to 0$.

Theorem 8.2 (Glivenko-Cantelli Theorem)**.** *Let* $(X_i, 1 \le i < \infty)$ *be IID with distribution function* F*. Let* $G_n(\omega, x)$ *be the empirical distribution function of* $(X_1, \ldots, X_n)$*.*

$$G_n(\omega, x) = \frac{1}{n} \sum_{i=1}^{n} 1_{(X_i(\omega) \le x)}$$

Then $\sup_x |G_n(\omega, x) - F(x)| \to 0$ *a.s. as* $n \to \infty$*.*

Proof. Fix x. The events $\{X_1 \le x\}$, $\{X_2 \le x\}$, etc. are IID events, with probability $F(x)$. The SLLN implies that $G_n(\omega, x) \to F(x)$ a.s. as $n \to \infty$.

If $S = \{\text{rationals}\} \cup \{\text{atoms of } F\}$ (which is countable), then $P(\{G_n(\omega, x) \to F(x)\, \forall x \in S\}) = 1$. Then 8.1 implies that

$$P\left(\sup_x |G_n(\omega, x) - F(x)| \to 0\right) = 1$$

□

8.2 Gambling on a Favorable Game

Example 8.3 (Betting on a Favorable Game)**.** Take a stake s, where you gain s with probability $1/2 + \alpha$ and lose s with probability $1/2 - \alpha$. (Imagine $\alpha = 1\%$.)

Strategy: Bet some proportion q of your total, each time.

Let X_n be your fortune after n bets. Then

$$X_{n+1} = (1-q)X_n + \begin{cases} 2qX_n & \text{if you win} \\ 0 & \text{if you lose the } (n+1)\text{th bet} \end{cases}$$

$$= (1-q)X_n + 2qX_n 1_{A_{n+1}}$$
$$= (1 - q + 2q1_{A_{n+1}})X_n$$

where A_{n+1} is the event that you win the $(n+1)$th bet. Then

$$X_n = X_0 \prod_{i=1}^{n} (1 - q + 2q1_{A_i})$$
$$\frac{\log X_n}{n} = \frac{\log X_0}{n} + \frac{1}{n}\sum_{i=1}^{n} Y_i$$

where $Y_i = \log(1 - q + 2q1_{A_i})$. As $n \to \infty$,

$$\frac{\log X_n}{n} \to EY \quad \text{a.s.}$$

If $(1/n)\log X_n \to c$, then $X_n \to e^{cn}$, where c is the asymptotic growth rate. The optimal choice of q is to maximize EY.

$$EY = \left(\frac{1}{2} + \alpha\right)\log(1+q) + \left(\frac{1}{2} - \alpha\right)\log(1-q)$$
$$\approx 2\alpha q - \frac{1}{2}q^2$$

for α, q small. Choose $q = 2\alpha$.

$EX_n = X_0(1 + 2q\alpha)^n \to \infty$, but $X_n \to 0$ a.s. if $q \geq q_{\text{crit}} \approx 4\alpha$.

8.3 a.s. Limits for Maxima

Lemma 8.4 (Deterministic Lemma)**.** *If $x_n \geq 0$ and $0 < b_n \uparrow \infty$, then*

$$\limsup_n \frac{\max(x_1, \ldots, x_n)}{b_n} = \limsup_n \frac{x_n}{b_n}$$

Proof. "$\geq$" is obvious. Fix j.

$$\text{LHS} = \limsup \frac{\max(x_j, x_{j+1}, \ldots, x_n)}{b_n} \leq \limsup_{n\to\infty} \max_{j \leq i \leq n} \frac{x_i}{b_i}$$
$$= \sup_{i \geq j} \frac{x_i}{b_i} \quad \forall j$$

Let $j \to \infty$. Then

$$\text{LHS} \leq \limsup_i \frac{x_i}{b_i}$$

□

Example 8.5. Let $(X_i, i \geq 1)$ be IID Exponential(1), so $P(X > x) = e^{-x}$. Write $M_n = \max_{1 \leq i \leq n} X_i$. Then

$$\limsup_n \frac{X_n}{\log n} = 1 \quad \text{a.s.} \tag{8.1}$$

and
$$\frac{M_n}{\log n} \to 1 \quad \text{a.s.}$$

Proof. Fix $\varepsilon > 0$. Then
$$P\left(\frac{X_n}{\log n} > 1+\varepsilon\right) = \exp(-(1+\varepsilon)(\log n)) = n^{-(1+\varepsilon)}$$
and $\sum_n n^{-(1+\varepsilon)} < \infty$. The First Borel-Cantelli Lemma implies that
$$\limsup_n \frac{X_n}{\log n} \le 1+\varepsilon \quad \text{a.s.} \implies \limsup_n \frac{X_n}{\log n} \le 1 \quad \text{a.s.}$$
Now fix $\varepsilon > 0$.
$$P\left(\frac{X_n}{\log n} \ge 1-\varepsilon\right) = n^{-(1-\varepsilon)}$$
where $\sum_n n^{-(1-\varepsilon)} = \infty$. The Second Borel-Cantelli Lemma implies that
$$\limsup_n \frac{X_n}{\log n} \ge 1-\varepsilon \quad \text{a.s.} \implies \limsup_n \frac{X_n}{\log n} \ge 1 \quad \text{a.s.}$$
The result (8.1) and 8.4 imply that
$$\limsup_n \frac{M_n}{\log n} = 1 \quad \text{a.s.}$$
Fix $\varepsilon > 0$.
$$\begin{aligned} P(M_n \le (1-\varepsilon)\log n) &= [P(X \le (1-\varepsilon)\log n)]^n \\ &= (1-n^{-(1-\varepsilon)})^n \\ &\le \exp\left(-n\cdot n^{-(1-\varepsilon)}\right) = \exp(-n^{\varepsilon}) \end{aligned}$$
where we have used $1-x \le e^{-x}$. The First Borel-Cantelli Lemma implies that $M_n \ge (1-\varepsilon)\log n$, ultimately, a.s., which implies that
$$\liminf_n \frac{M_n}{\log n} \ge 1-\varepsilon \quad \text{a.s.} \implies \liminf_n \frac{M_n}{\log n} \ge 1 \quad \text{a.s.}$$ □

Here, $X_n/\log n \to 0$ in probability, but not a.s.
$$P\left(\frac{X_n}{\log n} \ge \varepsilon\right) = n^{-\varepsilon} \to 0$$

8.4 2nd Moment SLLN

Lemma 8.6 (Deterministic Lemma)**.** *Let S_n be real. To prove $S_n/n \to 0$, it is enough to prove $\exists n(j) \uparrow \infty$ such that*

(i) $S_{n(j)}/n(j) \to 0$ *as* $j \to \infty$,

(ii) $d_j/n(j) \to 0$ *as* $j \to \infty$,

for $d_j = \max_{n(j)\le n<n(j+1)} |S_n - S_{n(j)}|$.

Proof. Given n, for some j where $n(j) \le n < n(j+1)$,

$$\left|\frac{S_n}{n}\right| \le \left|\frac{S_n}{n(j)}\right| \le \frac{\left|S_{n(j)}\right| + d_j}{n(j)} \to 0$$

as $j \to \infty$. □

Theorem 8.7 (2nd Moment SLLN)**.** *Given $(X_i, 1 \le i < \infty)$, with $EX_i \equiv 0$, let $\sup_i EX_i^2 = B < \infty$ and the X_i be* ***orthogonal****, $E(X_iX_j) = 0$, $j \ne i$. (We are not assuming independence!) Write*

$$S_n = \sum_{i=1}^{n} X_i$$

Then $S_n/n \to 0$ a.s.

Proof. Since $\mathrm{var}(S_n) \le nB$, Chebyshev's inequality implies

$$P\left(\frac{|S_n|}{n} \ge \varepsilon\right) \le \frac{nB}{n^2\varepsilon^2} = \frac{B}{n\varepsilon^2}$$

Take $n(j) = j^2$.

$$P\left(\left|\frac{S_{n(j)}}{n(j)}\right| \ge \varepsilon\right) \le \frac{B}{\varepsilon^2}\frac{1}{j^2}$$

Use Borel-Cantelli.

$$\frac{S_{n(j)}}{n(j)} \to 0 \quad \text{a.s.} \quad \text{as } j \to \infty$$

By 8.6, it is enough to prove $D_j/j^2 \to 0$ a.s., for

$$D_j = \max_{j^2 \le n < (j+1)^2} \left|S_n - S_{j^2}\right|$$

Then

$$D_j^2 = \max_{j^2 \le n < (j+1)^2} (S_n - S_{j^2})^2$$

$$ED_j^2 \underbrace{\le}_{\text{crude}} \sum_{n=j^2}^{(j+1)^2-1} E(S_n - S_{j^2})^2$$

Since

$$E(S_n - S_{j^2})^2 = \mathrm{var}\left(\sum_{j^2+1}^{n} X_i\right) \le B(n - j^2)$$

Letting $n = j^2 + i$, we have

$$ED_j^2 \le B\sum_{i=1}^{2j+1} i = \frac{1}{2}(2j+1)(2j+2)B$$

We have

$$P\left(\frac{D_j}{j^2} \geq \varepsilon\right) \leq \frac{ED_j^2}{\varepsilon^2 j^4} \in O(j^{-2})$$

The First Borel-Cantelli Lemma implies that $D_j/j^2 \to 0$ as $j \to \infty$. □

Theorem 8.8 (Dominated Convergence Theorem)**.** *If $X_n \to X$ a.s., if $\exists Y \geq 0$ such that $|X_n| \leq Y$ a.s. for all n, and if $EY < \infty$, then $EX_n \to EX$, $E(X_n - X) \to 0$, and $E(X) < \infty$.*

Proof. Fix $\varepsilon > 0$. Define $A_N = \{|X_n - X| \leq \varepsilon, \text{all } n \geq N\}$. Then $A_N \uparrow A_\infty$, say, and $P(A_\infty) = 1$. Also, $A_N^c \downarrow A_\infty^c$, and $P(A_\infty^c) = 0$.

$$\begin{aligned} E|X_N - X| &= E|X_N - X|1_{A_N} + E|X_N - X|1_{A_N^c} \\ &\leq \varepsilon + 2E\underbrace{Y1_{A_N^c}}_{\downarrow 0 \text{ a.s.}} \end{aligned}$$

$$\limsup_N E|X_N - X| \leq \varepsilon + 0, \quad \text{by monotone convergence}$$

This is true for all ε, so $E|X_N - X| \to 0$. □

Lecture 9

September 22

9.1 SLLN

Theorem 9.1 (Kolmogorov's Maximal Inequality)**.** *Let* $(X_i, 1 \le i \le n)$ *be independent,* $EX_i = 0$*, and* $EX_i^2 < \infty$*. Let* $S_m = \sum_{i=1}^m X_i$ *and* $S_n^* = \max_{1 \le m \le n} |S_m|$*. Then*

$$P(S_n^* \ge x) \le \frac{ES_n^2}{x^2}, \quad x > 0$$

Comments:

1. Markov's inequality gives
$$P(|S_n| \ge x) \le \frac{ES_n^2}{x^2}$$
The theorem gives a stronger result.
2. Idea: There is a "first time" that something happens.
3. Martingale theory develops better notation.

Proof. Fix x. Consider the event $\{S_n^* \ge x\} = \bigcup_{k=1}^m A_k$, where $A_k = \{|S_k| \ge x, |S_i| < x, \text{ all } 1 \le i < k\}$. The events A_k are disjoint. Note that (S_k, A_k) is independent of $S_n - S_k$. $S_n - S_k$ depends on $X_{k+1}, X_{k+2}, \dots, X_n$, while (S_k, A_k) depends on $(X_1, \dots, X_n)$. Then, since $S_n = S_k + (S_n - S_k)$,

$$\begin{aligned}
ES_n^2 &\ge \sum_{k=1}^n E[S_n^2 1_{A_k}] \\
&= \sum_{k=1}^n [E(S_k^2 1_{A_k}) + 2E(\underbrace{S_k 1_{A_k}(S_n - S_k)}_{=0}) + \underbrace{E((S_n - S_k)^2 1_{A_k})}_{\ge 0}] \\
ES_n^2 &\ge \sum_{i=1}^n E(S_k^2 1_{A_k}) \\
&\ge \sum_{k=1}^n E(x^2 1_{A_k}) \\
&= x^2 P\left(\bigcup_{k=1}^n A_k\right) \\
&= x^2 P(|S_n^*| \ge x)
\end{aligned}$$

because $S_k 1_{A_k}$ and $S_n - S_k$ are independent, $E(S_n - S_k) = 0$, and $|S_k| \geq x$ on A_k. ☐

"$\sum_{i=1}^\infty x_i$ converges" means that $\lim_{N\to\infty} \sum_{i=1}^N x_i$ exists and is finite. The Cauchy criterion says that this is equivalent to $\sup_{n\geq K} \left|\sum_{i=k+1}^n x_i\right| \to 0$ as $k \to \infty$. "$\sum_{i=1}^\infty X_i$ converges a.s." means

$$P\left(\omega : \lim_{N\to\infty} \sum_{i=1}^N X_i(\omega) \text{ exists and is finite}\right) = 1$$

Theorem 9.2. *Let (X_i) be independent, with $EX_i = 0$ and $\sigma_i^2 = \mathrm{var}(X_i) < \infty$. If $\sum_{i=1}^\infty \sigma_i^2 < \infty$, then $\sum_{i=1}^\infty X_i$ converges a.s.*

Comment. Consider the following argument: $\mathrm{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \sigma_i^2$. Taking $n \to \infty$, then

$$\mathrm{var}\left(\sum_{i=1}^\infty X_i\right) = \sum_{i=1}^\infty \sigma_i^2 < \infty \tag{9.1}$$

which shows that $\sum_{i=1}^\infty X_i$ is finite a.s. This argument is incorrect because *a priori*, we do not know that we have a convergent random variable.

Exercise. Knowing 9.2, show (9.1).

Proof. Define $M_k = \sup_{n>k} \left|\sum_{i=k+1}^n X_i\right|$. It is enough to show that $M_k \to 0$ a.s. as $k \to \infty$. Define also $W_k = \sup_{n_2>n_1>k} \left|\sum_{i=n_1+1}^{n_2} X_i\right|$ and note that $M_k \leq W_k \leq 2M_k$ and W_k decreases as k increases.

$$P\left(\sup_{k<n\leq N}\left|\sum_{i=k+1}^N X_i\right| \geq \varepsilon\right) \underbrace{\leq}_{9.1} \varepsilon^{-2}\,\mathrm{var}\left(\sum_{i=k+1}^N X_i\right) = \varepsilon^{-2}\sum_{i=k+1}^N \sigma_i^2$$

Taking $N \to \infty$, $P(M_k > \varepsilon) \leq \varepsilon^{-2} \sum_{i=1}^\infty \sigma_i^2$.

$$P(W_k > \varepsilon) \leq P\left(M_k > \frac{\varepsilon}{2}\right) \leq 4\varepsilon^{-2} \sum_{i=k+1}^\infty \sigma_i^2 \to 0 \text{ as } k \to \infty$$

Taking $k \to \infty$, then $W_k \downarrow W_\infty$ for some W_∞ a.s. Then $P(W_\infty > \varepsilon) = 0$, which implies that $W_\infty = 0$ a.s., which implies that $W_k \downarrow 0$ a.s. and $M_k \to 0$ a.s. ☐

Lemma 9.3 (Deterministic Lemma (Kronecker))**.** *Let (x_n) be a sequence of reals, $S_n = \sum_{i=1}^n x_i$, $0 < a_n \uparrow \infty$ as $n \uparrow \infty$. If $\sum_i x_i/a_i$ converges, then $S_n/a_n \to 0$.*

Proof. Exercise/Textbook. ☐

Corollary 9.4. *Let (X_i) be independent, $EX_i = 0$, $EX_i^2 < \infty$, and $S_n = \sum_{i=1}^n X_i$. If $0 < a_n \uparrow \infty$ as $n \uparrow \infty$ and if $\sum_n EX_n^2/a_n^2 < \infty$, then $S_n/a_n \to 0$ a.s.*

Proof. 9.2 implies that $\sum_n X_n/a_n$ converges a.s. Then 9.3 implies that $S_n/a_n \to 0$ a.s. ☐

Specialization. Suppose also that $EX_n^2 \sim cn^{2\alpha}$, $\alpha > 0$. Take $a_n^2 = n^{1+2\alpha+2\varepsilon}$ ($\varepsilon > 0$ is small). Then 9.4 implies that $S_n/n^{1/2+\alpha+\varepsilon} \to 0$ a.s.

Specialization. Suppose that $\sup_n EX_n^2 < \infty$. Take $a_n^2 = n(\log n)^{1+\varepsilon}$. Then 9.4 implies that $S_n/\sqrt{n\log^{1+\varepsilon} n} \to 0$ a.s. We know implicitly from the CLT that if (X_i) are IID, then $S_n/\sqrt{n} \to 0$ a.s. is **not** true. The law of iterated logarithm gives the proper borderline.

Theorem 9.5 (SLLN)**.** *Let (X_i) be IID with $E|X| < \infty$. Then $S_n/n \to EX$ a.s. as $n \to \infty$.*

Proof. The idea is to truncate, center, and then apply 9.4.

If $Z \geq 0$, then

$$EZ^k = \int_0^\infty kz^{k-1} P(Z \geq z)\,\mathrm{d}z \qquad \text{because } \approx \int_0^\infty x^k f(x)\,\mathrm{d}x$$

Define $Y_k = X_k 1_{(|X_k| \leq k)}$. Then

$$\sum_k P(Y_k \neq X_k) = \sum_{k=1}^\infty P(|X| > k) \leq \int_0^\infty P(|X| > x)\,\mathrm{d}x = E|X| < \infty$$

Then the First Borel-Cantelli Lemma implies that $P(Y_k = X_k, \text{ ultimately}) = 1$. It is enough to prove that $(1/n)\sum_{k=1}^n Y_k \to EX$ a.s.

Center: define $X'_k = Y_k - EY_k$. *Claim*: $\sum_k \mathrm{var}(X'_k)/k^2 < \infty$.

$$EY_k^2 = \int_0^\infty 2yP(|Y_k| > y)\,\mathrm{d}y = \underbrace{\int_0^\infty 2yP(k \geq |X_k| \geq y) 1_{(y \leq k)}\,\mathrm{d}y}_{\text{Check this!}}$$

$$\leq \int_0^\infty 2yP(|X_k| \geq y) 1_{(y \leq k)}\,\mathrm{d}y$$

$$\sum_k \frac{\mathrm{var}(X_k)}{k^2} \leq \sum_k \frac{EY_k^2}{k^2} \leq \sum_k \frac{1}{k^2} \int_0^\infty 2yP(|X| \geq y) 1_{(y \leq k)}\,\mathrm{d}y$$

$$= \int_0^\infty \underbrace{\left(\sum_k \frac{1}{k^2} 1_{(y \leq k)} 2y \right)}_{G(y)} P(|X| \geq y)\,\mathrm{d}y$$

Claim: $G(y) \leq 4$, for all $0 < y < \infty$. Since $G(y) \leq \sum_k 1/k^2 \leq 2$ for $y \leq 1$, this is true for $y \leq 1$. Take $y > 1$.

$$\frac{1}{k^2} \leq \int_{k-1}^k \frac{1}{x^2}\,\mathrm{d}x$$

so

$$\sum_k \frac{1}{k^2} 1_{(y \leq k)} = \sum_{k \geq \lceil y \rceil} \frac{1}{k^2} \leq \int_{\lceil y \rceil - 1}^\infty \frac{1}{x^2}\,\mathrm{d}x = \frac{1}{\lceil y \rceil - 1}$$

Since $y > 1$,

$$G(y) \leq \frac{2y}{\lceil y \rceil - 1} \leq 4$$

(by a picture). Then

$$\sum_k \frac{\mathrm{var}(X'_k)}{k^2} \leq 4 \int_0^\infty P(|X| \geq y)\,\mathrm{d}y = 4E|X| < \infty$$

Apply 9.4 to (X'_n): $(1/n)\sum_{i=1}^n X'_i \to 0$ a.s., so $(1/n)\sum_{i=1}^n (Y_i - EY_i) \to 0$ a.s. Note that

$$EY_i = EX 1_{(|X| \leq i)} \to EX$$

as $i \to \infty$. By dominated convergence, $(1/n)\sum_{i=1}^n (EY_i - EX) \to 0$ a.s. Add the two equations to get $(1/n)\sum_{i=1}^n (Y_i - EX) \to 0$ a.s., which implies that $(1/n)\sum_{i=1}^n Y_i \to EX$ a.s. □

Lecture 10

September 27

10.1 Truncation

Corollary 10.1 (SLLN). *Take IID* (X_i)*, where* $EX^+ = \infty$, $EX^- < \infty$ $(X = X^+ - X^-)$*. Let* $S_n = \sum_{i=1}^n X_i$*. Then* $S_n/n \to \infty$ *a.s.*

Proof. Fix a large $B < \infty$. Define $Y_i = X_i 1_{(X_i \le B)}$. Then the (Y_i) are IID, with $E|Y_i| < \infty$, so we can apply the SLLN to (Y_i).

$$\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow[a.s.]{} EY = EX1_{(X \le B)}$$

Then

$$\liminf_n \frac{1}{n} S_n \ge \liminf_n \frac{1}{n}\sum_{i=1}^n Y_i \underbrace{=}_{a.s.} EX1_{(X\le B)}$$

for each B. As $B \uparrow \infty$, then $E[X1_{(X\le B)}] \uparrow -EX^- + EX^+ = +\infty$. Therefore, letting $B \uparrow \infty$,

$$\liminf_n \frac{1}{n} S_n \ge \infty$$

□

10.2 Renewal SLLN

If we travel halfway at 60 mph and halfway at 20 mph, the average speed is 30 mph. To see this, traveling 120 miles takes 1 hour + 3 hours = 4 hours.

Lemma 10.2 (Deterministic Lemma). *Consider real numbers* $s_0 = 0$, $s_n/n \to a \in (0, \infty)$ *as* $n \to \infty$*. Let* $h(t) = \min\{n : s_n \ge t\}$ *and* $m(t) = \max\{n : s_n \le t\}$*. Note that* $m(t) \ge h(t) - 1$*. Then* $h(t)/t \to 1/a$ *and* $m(t)/t \to 1/a$ *as* $t \to \infty$*.*

Proof. Fix $\varepsilon > 0$. Then $s_n \le (a+\varepsilon)n$ ultimately, which implies that $h(t) \ge t/(a+\varepsilon)$ ultimately. Then

$$\liminf_t \frac{h(t)}{t} \ge \frac{1}{a+\varepsilon} \underbrace{\Rightarrow}_{\varepsilon\downarrow 0} \liminf_t \frac{h(t)}{t} \ge \frac{1}{a}$$

Similarly, $m(t) \leq t/(a+\varepsilon)$ ultimately, which implies that $\limsup_t m(t)/t \leq 1/a$. We have

$$\frac{1}{a} \leq \liminf_t \frac{h(t)}{t} \leq \limsup_t \frac{m(t)}{t} \leq \frac{1}{a} \qquad \square$$

Corollary 10.3 (Renewal SLLN)**.** *Let (X_i) be IID, with $EX = \mu \in (0,\infty)$. Let $S_n = \sum_{i=1}^n X_i$. Define $N_t = \max\{n : S_n \leq t\}$ and $H_t = \min\{n : S_n \geq t\}$. Then $N_t/t \to 1/\mu$ and $H_t/t \to 1/\mu$ a.s. as $t \to \infty$.*

Proof. Use the SLLN and 10.2 $\square$

Story. Light bulbs have IID lifetimes $X_1, X_2, \ldots > 0$. We have a new bulb at time 0, and let N_t be the number of bulbs replaced by time t.

10.3 Stopping Times

A random variable is a measurable function $X_i : (\Omega, \mathcal{F}, P) \to \mathbb{R}$. Given $X_0, X_1, \ldots, X_n$, we define the σ-field $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, the collection of events of the form $\{\omega : (X_0(\omega), \ldots, X_n(\omega)) \in B\}$ for some measurable $B \subseteq \mathbb{R}^{n+1}$ (so $\mathcal{F}_n \subseteq \mathcal{F}$). $\mathcal{F}_n$ is the "information at time n".

A **stopping time** is a RV $T : (\Omega, \mathcal{F}, P) \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ such that

$$\{T = n\} \in \mathcal{F}_n, \qquad 0 \leq n < \infty \tag{10.1}$$

This is equivalent to the definition

$$\{T \leq n\} \in \mathcal{F}_n, \qquad 0 \leq n < \infty \tag{10.2}$$

Given (10.1), $\{T \leq n\} = \{T = 0\} \cup \{T = 1\} \cup \cdots \cup \{T = n\} \in \mathcal{F}_n$, since each event is in $\mathcal{F}_n$. Given (10.1), $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n$, since both events are in $\mathcal{F}_n$.

Example 10.4. $T = \min\{n : X_n \in B\}$ for some measurable $B \subseteq \mathbb{R}^1$ is a stopping time because

$$\{T \leq n\} = \bigcup_{i=0}^{n} \{X_i \in B\} \in \mathcal{F}_n$$

Note. Given arbitrary $X_1, X_2, \ldots, X_n$, define $S_0 = 0$, $S_m = \sum_{i=1}^m X_i$. Then

$$\sigma(X_1, \ldots, X_n) = \sigma(S_0, S_1, \ldots, S_n) = \mathcal{F}_n$$

and so $T = \min\{n : S_n \geq b\}$ is a stopping time.

$T = \infty$ if the event never happens.

Given $X_1, \ldots, X_N$ (for a given N), $T = \max\{n : n \leq N, X_n \geq a\}$ is *not* a stopping time.

Theorem 10.5 (Wald's Equation/Identity/Formula)**.** *Let (X_i) be IID with $EX = \mu$ and $S_n = \sum_{i=1}^n X_i$. Let T be a stopping time with $ET < \infty$. Then $ES_T = \mu \cdot ET$.*

Note. This is an undergraduate result under the assumption that T is independent of (X_i).

Fact. $E\sum_{i=1}^\infty Y_i = \sum_{i=1}^\infty EY_i$, provided $\sum_i E|Y_i| < \infty$.

Proof: $\sum_{i=1}^n Y_i \to \sum_{i=1}^\infty Y_i$ a.s., and the summation is dominated by $\sum_{i=1}^n |Y_i|$. Use dominated convergence.

Proof.

$$S_n = \sum_{i=1}^{\infty} X_i 1_{(i \le n)} \Rightarrow S_T = \sum_{i=1}^{\infty} X_i 1_{(i \le T)}$$

Since $\{i \le T\}^c = \{T \le i-1\} \in \mathcal{F}_{i-1}$, $\{i \le T\}$ is independent of the X_i. Then

$$E[X_i 1_{(i \le T)}] = \mu P(T \ge i)$$
$$\sum_{i=1}^{\infty} E[X_i 1_{(i \le T)}] = \mu ET \tag{10.3}$$

We need to show that $ES_T = \mu \cdot ET$. By the Fact, it is enough to show that $\sum_{i=1}^{\infty} E|X_i|1_{(i \le T)} < \infty$. We can apply (10.3) to $|X_i|$. Then

$$\sum_{i=1}^{\infty} E[|X_i|1_{(i \le T)}] = (E|X|)ET < \infty$$

□

10.4 Fatou's Lemma

Lemma 10.6 (Fatou's Lemma)**.** *Take arbitrary* $X_n \ge 0$*. Then* $E[\liminf_n X_n] \le \liminf_n EX_n \le \infty$*.*

Proof. Define $Y_N = \inf_{n \ge N} X_n$. Then $0 \le Y_N \uparrow \liminf X_n$, so $0 \le EY_N \uparrow E[\liminf X_n]$. Since $Y_N \le X_N$,

$$\begin{aligned} E[\liminf X_n] &= \liminf_N EY_N \\ &\le \liminf_N EX_N \end{aligned}$$

□

Corollary 10.7. *Take arbitrary* $X_n \ge 0$*. If* $X_n \to X_\infty$ *a.s., then* $EX_\infty \le \liminf_n EX_n \le \infty$*.*

Recall the aggressive "gambling on a favorable game" example. There, $X_n \ge 0$, $X_n \to 0$ a.s., but $EX_n \to \infty$.

10.5 Back to Renewal Theory

Under the assumptions of 10.3, with the additional assumption that $X \ge 0$ a.s., then $E[N(t)/t] \to 1/\mu$ as $t \to \infty$.

Proof. By 10.6,

$$\begin{aligned} \frac{1}{\mu} &\le \liminf_{\substack{t \to \infty \\ t \text{ integer}}} E\left[\frac{N(t)}{t}\right] \\ &= \liminf_{t \to \infty} E\left[\frac{N(t)}{t}\right] \end{aligned}$$

It is enough to show the upper bound

$$\limsup_t E\left[\frac{N(t)}{t}\right] \le \frac{1}{\mu}$$

Since $X \ge 0$, $N(t)+1 = \min\{n : S_n > t\}$ is a stopping time. $\min(N(t)+1, m)$ is also a stopping time. Apply 10.5 to obtain

$$ES_{\min(N(t)+1,m)} = \mu E \min(N(t)+1, m)$$

Let $m \uparrow \infty$.

$$ES_{N(t)+1} = \mu E[N(t) + 1] \leq \infty \tag{10.4}$$

Fix k. Let $\hat{X}_i = \min(X_i, k)$. Define $\hat{S}_n$ and $\hat{N}(t)$ similarly. Then

$$\hat{S}_n \leq S_n \Rightarrow \hat{N}(t) \geq N(t)$$

We can apply (10.4) to $(\hat{X}_i)$.

$$E[\hat{N}(t) + 1] \cdot E \min(X, k) = E\hat{S}_{\hat{N}(t)+1} \leq t + k < \infty$$

Therefore,

$$\frac{E[N(t) + 1]}{t} \leq \frac{t + k}{t} \frac{1}{E \min(X, k)}$$

This implies that

$$\limsup_t \frac{EN(t)}{t} \leq \frac{1}{E \min(X, k)}$$

which is true for all k. Let $k \uparrow \infty$ to obtain

$$\limsup_t \frac{EN(t)}{t} \leq \frac{1}{EX} = \frac{1}{\mu}$$ □

Lecture 11

September 29

11.1 Miscellaneous Measure Theory Related Topics

11.1.1 Kolgomorov's 0-1 Law

Theorem 11.1 (Kolmogorov's 0-1 Law)**.** *Consider $X_1, X_2, \dots$ mapping onto any range space. Define $\tau_n = \sigma(X_n, X_{n+1}, X_{n+2}, \dots)$ and $\bigcap_{n\geq 1} \tau_n = \tau$ (the "tail σ-field"). If $(X_1, X_2, \dots)$ are independent, then $A \in \tau$ implies that $P(A)$ is 0 or 1, that is, τ is a trivial σ-field.*

Note. $\limsup_n X_n$ is τ_n-measurable for all n, so it is τ-measurable.

Proof. Define $\mathcal{F}_{n-1} = \sigma(X_1, \dots, X_{n-1})$. $\mathcal{F}_{n-1}$ is independent of τ_n, which implies that $\mathcal{F}_{n-1}$ is independent of τ, which implies that the field $\bigcup_n \mathcal{F}_n$ is independent of τ. By the π-λ Lemma, $\sigma(\bigcup \mathcal{F}_n) = \sigma(X_1, X_2, \dots)$ is independent of τ, which implies that τ is independent of τ. Then, $A \in \tau$ implies that $P(A \cap A) = P(A)P(A) = P(A)$. $x^2 = x$ implies that $x = 0$ or 1. □

Lemma 11.2. *If $\mathcal{A}$ is a trivial σ-field, and if X, a RV that takes on values in $[-\infty, \infty]$, is $\mathcal{A}$-measurable, then there exists x_0 such that $P(X = x_0) = 1$.*

Proof. Define $x_0 = \inf\{x : P(X \leq x) = 1\}$. For the case where $x_0 \in (-\infty, \infty)$, then $P(X \leq x_0 + \varepsilon) = 1$ and $P(X \leq x_0 - \varepsilon) = 0$ for all ε. □

11.1.2 "Modes of Convergence" for $\mathbb{R}$-Valued RVs

$X_n \xrightarrow{a.s.} X$ means $P(\omega : X_n(\omega) \to X(\omega)) = 1$.

$X_n \xrightarrow{P} X$ means $P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$, for all $\varepsilon > 0$.

$X_n \xrightarrow{L^p} X$ means that $E|X_n - X|^p \to 0$ and $\sup_n E|X_n|^p < \infty$ ($\infty > p \geq 1$).

Facts:

1. We showed before that $\xrightarrow{L^p}$ implies $\xrightarrow{P}$, but not conversely.
2. $\xrightarrow{a.s.}$ implies $\xrightarrow{P}$, but not conversely.

Example 11.3. Let U be uniform on $[0, 1]$. Let $X_n = n1_{(U\leq 1/n)}$. Then $X_n \xrightarrow{P} 0$, but $EX_n = 1$, so $X_n \to 0$ in L^1 is false.

If $X_n \xrightarrow{a.s.} X$, since $P(A_n \text{ inf. often}) \geq \limsup_n P(A_n)$,

$$\begin{aligned} 0 &= P(|X_n - X| \geq \varepsilon \text{ inf. often}) \\ &\geq \limsup_n P(|X_n - X| \geq \varepsilon) = 0 \end{aligned}$$

which implies that $X_n \to X$ in probability.

Example 11.4. Take independent events (A_n) with $P(A_n) \to 0$, which implies that $1_{A_n} \to 0$ in probability. $\sum_n P(A_n) = \infty$ implies, by the Second Borel-Cantelli Lemma, that $P(A_n \text{ inf. often}) = 1$, which implies that $1_{A_n} \to 0$ a.s. is false.

Recall the Dominated Convergence Theorem (DCT): If $X_n \to X$ a.s., if $\exists Y \geq 0$ with $EY < \infty$, and $|X_n| \leq Y$ for all n, then $E|X_n - X| \to 0$ and $EX_n \to EX$.

Lemma 11.5. *If $X_n \xrightarrow{P} X$, then there exists a subsequence, $n_1 < n_2 < n_3 < \cdots$ such that $X_{n_j} \xrightarrow{a.s.} X$ as $j \to \infty$.*

Proof. Choose n_j inductively.

$$n_j = \min\{n > n_{j-1}, P(|X_n - X| \geq 2^{-j}) \leq 2^{-j}\}$$

Then $\sum_j P(|X_n - X| \geq 2^{-j}) < \infty$. The First Borel-Cantelli Lemma implies that $\left|X_{n_j} - X\right| \leq 2^{-j}$, ultimately in j, a.e., which implies that $X_{n_j} \to X$ a.s. □

Aside. The result is related to the fact that "a.s. convergence" is not convergence in a metric.

Corollary 11.6. *The DCT remains true under the assumption that $X_n \to X$ in probability.*

Proof. Suppose that the statement is false: $\exists \varepsilon > 0$ and a subsequence $m_1 < m_2 < m_3 < \cdots$ such that $E\left|X_{m_j} - X\right| \geq \varepsilon \, \forall j$. Now $X_{m_j} \to X$ in probability, so 11.5 implies that there exists a subsequence (n_j) of (m_j) such that $X_{n_j} \to X$ a.s. and $E\left|X_{n_j} - X\right| \geq \varepsilon \, \forall j$. This contradicts the DCT. □

This proof uses the "subsequence trick".

Exercise. Obvious: If f is continuous, $X_n \to X$ a.s. implies that $f(X_n) \to f(X)$ a.s. Less obvious: If f is continuous, $X_n \to X$ in probability implies that $f(X_n) \to f(X)$ in probability. (This can be proven with the subsequence trick.)

11.1.3 Radon-Nikodym Derivative

There are two views of integration in calculus.

1. Given f, a, b, then $\int_a^b f(x)\,\mathrm{d}x$ is a number.
2. $$F(x) = \int_0^x f(y)\,\mathrm{d}y \quad \Leftrightarrow \quad f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

   Integration is an operation $f \mapsto F$, which is the opposite of $F \mapsto F'$.

In MT, given a PM μ, integration is a map $h \mapsto I(h) = \int h\,\mathrm{d}\mu$. The analog in MT involves *measures*, not functions.

Take a measurable space $(S, \mathcal{S})$. Fix a σ-finite measure μ on $(S, \mathcal{S})$. Consider a measurable $h : S \to [0, \infty)$. For $A \in \mathcal{S}$, define $\nu(A) = \int_A h \,\mathrm{d}\mu \leq \infty$.

Claim. ν is a σ-finite measure on $(S, \mathcal{S})$.

The fact that μ is σ-finite implies that there exists $A_n \uparrow S$, with $\mu(A_n) < \infty$. Define $B_n = A_n \cap \{s : h(s) \leq n\}$. Then $B_n \uparrow S$ and $\nu(B_n) \leq n\mu(A_n) < \infty$.

The two measures ν and μ have a relationship. For all A, if $\mu(A) = 0$, then $\nu(A) = 0$. This property has a name: ν is **absolutely continuous** with respect to μ, denoted $\nu \ll \mu$.

Theorem 11.7 (Radon-Nikodym Theorem)**.** *If μ and ν are σ-finite measures on $(S, \mathcal{S})$, if $\nu \ll \mu$, then there exists a measurable $h : S \to [0, \infty)$ such that $\nu(A) = \int_A h \,\mathrm{d}\mu \;\forall A \in \mathcal{S}$.*

Notation. Write

$$h = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$$

and

$$h(s) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(s)$$

and call $h = \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ the Radon-Nikodym **density** of ν with respect to μ.

In particular, if μ is a probability measure on $\mathbb{R}^1$ and if $\mu \ll \text{Leb}$, then $h = \dfrac{\mathrm{d}\mu}{\mathrm{dLeb}}$ exists (the density function, e.g. Normal, Exponential, etc.).

Proof of Radon-Nikodym. See the MT text. We will prove this via martingales later. □

11.1.4 Probability Measures on $\mathbb{R}$

We know there is a 1-1 correspondence between probability measures μ and distribution functions F.

$$F(x) = \mu(-\infty, x]$$

"x is an **atom** of μ" means that $\mu(\{x\}) > 0$. μ can have only countably many atoms.

There are three basic types of PMs μ:

1. $\mu \ll \text{Leb}$, so it can be described by its density f.

$$F(x) = \int_{-\infty}^{x} f(y) \,\mathrm{d}y$$

   Here, f can be any measurable function with $f \geq 0$ and $\int_{-\infty}^{\infty} f(x) \,\mathrm{d}x = 1$.

2. μ is **purely atomic** if there exists a countable set of atoms $x_1, x_2, \ldots$ and $\sum_i \mu(\{x_i\}) = 1$, which implies that $\mu(\mathbb{R} \setminus \cup_i \{x_i\}) = 0$ (discrete).

3. **Singular measures**: there exists A such that $\text{Leb}(A) = 0$, $\mu(A) = 1$, but there are no atoms.

Take $x \in [0, 1]$ with a binary expansion, e.g. $0.10110100011\ldots$. Say that $b_i(x)$ is the ith digit of the binary expansion of x ($\lfloor 2^i n \rfloor \bmod 2$), which defines a map from $[0, 1]$ to B^∞. Next, map to $\{0, 1, 2\}^\infty$ by converting 1s to 2s, and then map back to $[0, 1]$ by interpreting the result base 3, to obtain $\sum_{i=1}^{\infty} 3^{-i}(2b_i(x))$. Putting these together yields a measurable map $H : [0, 1] \to [0, 1]$. Take U to be Uniform$[0, 1]$. What is the distribution of $H(U)$?

$F(x) = P(H(U) \leq x)$ is the **Cantor function**, which is continuous. The set of possible values of H is "the base-3 expansion has no "1"" is the **Cantor set**, C, and $\text{Leb}(C) = 0$ while $P(H(U) \in C) = 1$.

The distribution of $H(U)$ is called the "uniform distribution on the Cantor set".

Fact. Any PM μ on $\mathbb{R}^1$ has a unique decomposition

$$\mu = a_1 \underbrace{\mu_1}_{\text{type 1}} + a_2 \underbrace{\mu_2}_{\text{type 2}} + a_3 \underbrace{\mu_3}_{\text{type 3}}$$

where $a_i \geq 0$, $a_1 + a_2 + a_3 = 1$.

Lecture 12

October 4

12.1 Large Deviations Theorem (Durrett)

If $a_n \sim ce^{\beta n}$ as $n \to \infty$, then $(1/n)\log a_n \to \beta$, where β is the asymptotic growth (decrease) rate. Today, $\beta < 0$.

Assumptions. Let (X_i) be IID, with $S_n = \sum_{i=1}^n X_i$, $EX = \mu$. Fix $a > \mu$, $P(X \geq a) > 0$. Define $\phi(\theta) = E\exp(\theta X)$, and assume $\theta^* = \sup\{\theta : \phi(\theta) < \infty\} > 0$.

Consider $P(S_n/n \geq a)$. We know that $P(S_n/n \geq a) \to 0$ as $n \to \infty$ by the WLLN. How fast?

Our general LD inequality gives

$$P(Y \geq y) \leq \inf_{\theta \geq 0} \frac{Ee^{\theta Y}}{e^{\theta y}}$$

Therefore,

$$P\left(\frac{S_n}{n} \geq a\right) = P(S_n \geq an) \leq \inf_{\theta} \frac{E\exp(\theta S_n)}{\exp(\theta an)}$$

On the other hand,

$$\exp(\theta S_n) = \exp\left(\theta \sum_{i=1}^n X_i\right) = \prod_{i=1}^n \exp(\theta X_i)$$

$$E\exp(\theta S_n) = \prod_{i=1}^n E\exp(\theta X_i) = (\phi(\theta))^n$$

This implies:

$$P\left(\frac{S_n}{n} \geq a\right) \leq \inf_{\theta > 0} \frac{E\exp(\theta S_n)}{\exp(\theta an)} \leq \left(\inf_{\theta} \frac{\phi(\theta)}{e^{\theta a}}\right)^n$$

$$\frac{1}{n}\log P\left(\frac{S_n}{n} \geq a\right) \leq \inf_{\theta}\left[\log\phi(\theta) - a\theta\right]$$

Theorem 12.1. *As $n \to \infty$,*

$$\lim_{n\to\infty} \frac{1}{n}\log P\left(\frac{S_n}{n} \geq a\right) = \inf_{\theta}\left(\log\phi(\theta) - a\theta\right) = \inf_{\theta} G(\theta)$$

There are three steps in the proof.

- analysis of $\phi(\theta)$

- tilting lemma
- put it together

Lemma 12.2.

$$\phi'(0+) = \mu$$

We believe this because

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\phi(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}Ee^{\theta X} \underbrace{=}_{\text{how to justify in detail?}} E\frac{\mathrm{d}}{\mathrm{d}\theta}e^{\theta X} = E[Xe^{\theta X}] \qquad \forall\theta$$

Taking $\theta = 0$, $\phi'(0+) = EX$.

Proof. We know that $(e^{\theta X} - 1)/\theta \to X$ a.s. as $\theta \downarrow 0$. We want

$$E\left[\frac{e^{\theta X} - 1}{\theta}\right] \to EX \tag{12.1}$$

We seek to use the Dominated Convergence Theorem. For $x > 0$,

$$e^{\theta x} - 1 = \int_0^{\theta x} e^y \,\mathrm{d}y \le \theta x e^{\theta x}$$

For $x < 0$,

$$\left|e^{\theta x} - 1\right| = \int_{\theta x}^0 e^y \,\mathrm{d}y \le |\theta x|$$

These imply

$$\left|e^{\theta x} - 1\right| \le \theta|x|\max(1, e^{\theta x})$$

For $0 < \theta \le \theta_0$,

$$(12.1) \le |x|\max(1, e^{\theta_0 x}) \tag{12.2}$$

By hypothesis, there exists θ_1 such that $Ee^{\theta_1 X} < \infty$. Choose $\theta_0 < \theta_1$, so that $E[|X|\max(1, e^{\theta_0 X})] < \infty$. Now, the RVs are bounded by (12.2). Apply the DCT. □

The same argument applies to

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}Ee^{\theta X} = E\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}e^{\theta X} = EX^2e^{\theta X}$$

Lemma 12.3. *$\phi'(0+) = \mu$, and for $0 < \theta < \theta^*$,*

$$\phi'(\theta) = E[Xe^{\theta X}]$$
$$\phi''(\theta) = E[X^2e^{\theta X}]$$

Suppose X is discrete. Fix θ. Define a distribution for $\hat{X}$ by

$$P(\hat{X} = x) = \frac{e^{\theta x}P(X = x)}{\phi(\theta)}$$

Fix θ. Then

$$\phi(\theta) = \sum_x e^{\theta x}P(X = x)$$

Also,

$$\begin{aligned} E\hat{X} &= \sum_x xP(\hat{X} = x) = \frac{\sum_x xe^{\theta x}P(X = x)}{\phi(\theta)} \\ &= \frac{EXe^{\theta X}}{\phi(\theta)} = \frac{\phi'(\theta)}{\phi(\theta)} = \frac{\mathrm{d}}{\mathrm{d}\theta} \log \phi(\theta) \end{aligned}$$

and

$$\begin{aligned} E[\hat{X}^2] &= \frac{E[X^2e^{\theta X}]}{\phi(\theta)} = \frac{\phi''(\theta)}{\phi(\theta)} \\ \mathrm{var}(\hat{X}) &= E[\hat{X}^2] - (E\hat{X})^2 \\ &= \frac{\phi''(\theta)}{\phi(\theta)} - \left(\frac{\phi'(\theta)}{\phi(\theta)}\right)^2 \\ &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\phi'(\theta)}{\phi(\theta)}\right) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log \phi(\theta) \end{aligned}$$

For general X, define the distribution of $\hat{X}$ by the Radon-Nikodym density

$$\frac{\mathrm{d}P(\hat{X} \in \cdot)}{\mathrm{d}P(X \in \cdot)}(x) = \frac{e^{\theta x}}{\phi(\theta)}$$

Lemma 12.4 (Tilting Lemma)**.**

$$E\hat{X} = \frac{\mathrm{d}}{\mathrm{d}\theta} \log \phi(\theta)$$

and

$$\mathrm{var}(\hat{X}) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log \phi(\theta)$$

Now, we study $G(\theta) = \log \phi(\theta) - a\theta$.

$$\begin{aligned} G'(0+) &= \frac{\phi'(0+)}{\phi(0)} - a = \mu - a < 0 \\ G''(\theta) &= \mathrm{var}\, \hat{X}_\theta > 0 \qquad \text{on } 0 < \theta < \theta^* \\ G(0) &= 0 \end{aligned}$$

It is easy to see that $G(\theta) \to \infty$ as $\theta \to \infty$. G is strictly convex.

Find $\inf_\theta G(\theta)$ by solving $G'(\theta) = 0$, or

$$\frac{\phi'(\theta)}{\phi(\theta)} = a$$

Case 1. There exists a solution $\theta_a \in (0, \theta^*)$ of the equation $\phi'(\theta)/\phi(\theta) = a$.

Bad Case. Take the density $f(x) \sim x^{-2}e^{-\lambda x}$ as $x \to \infty$. Then $\phi(\lambda) < \infty$, but $\phi(\lambda+) = \infty$.

Assume case 1. Choose $\theta \in (\theta_a, \theta^*)$. Consider the tilted distribution $\hat{X} = \hat{X}_\theta$.

$$E\hat{X} = \frac{\mathrm{d}}{\mathrm{d}\theta}(\log \phi(\theta)) > \frac{\mathrm{d}}{\mathrm{d}\theta}(\log \phi(\theta))\Big|_{\theta=\theta_a}$$

because $E\hat{X}_\theta > a$ and $E\hat{X}_\theta \downarrow a$ as $\theta \downarrow \theta_a$. (Check!)

Fix $b > E\hat{X}_\theta$. The trick is to apply the WLLN to the tilted $(\hat{X}_i)$. Since

$$\frac{P(\hat{X} = x)}{P(X = x)} = \frac{e^{\theta x}}{\phi(\theta)}$$

we have

$$\frac{P(\hat{X}_1 = x_1, \dots, \hat{X}_n = x_n)}{P(X_1 = x_1, \dots, X_n = x_n)} = \frac{e^{\theta \sum_{i=1}^n X_i}}{\phi^n(\theta)}$$

which gives

$$\frac{P(\hat{S}_n = s)}{P(S_n = s)} = \frac{e^{\theta s}}{\phi^n(\theta)}$$

Therefore,

$$\frac{P(y_1 \le \hat{S}_n \le y_2)}{P(y_1 \le S_n \le y_2)} \le \frac{e^{\theta y_2}}{\phi^n(\theta)}$$

with $y_1 = an$, $y_2 = bn$, so

$$P\left(a \le \frac{S_n}{n} \le b\right) \ge e^{-\theta b n}\phi^n(\theta) \underbrace{P\left(a \le \frac{\hat{S}_n}{n} \le b\right)}_{\to 1 \text{ as } n \to \infty}$$

Hence,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \ge a\right) &\ge -b\theta + \log\phi(\theta) \\ &\ge -b\theta_a + \log\phi(\theta_a) \\ &\ge -a\theta_a + \log\phi(\theta_a) \\ &= G(\theta_a) \end{aligned}$$

(Let $\theta \downarrow \theta_a$. Since this is true for all $b > a$, let $b \downarrow a$.)

12.2 Conditional Distributions

Undergraduate Version. Consider (X, Y):

	discrete	continuous
	$p(x,y) = P(X = x, Y = y)$	$f(x,y)$ joint density
marginal distribution	$p_X(x) = P(X = x)$	$f_X(x) =$ density of X
	conditional distribution of Y given $X = x$	conditional density of Y given $X = x$
	$p_{Y \mid X}(y \mid x) = P(Y = y \mid X = x)$	$y \mapsto f_{Y \mid X}(y \mid x)$
	$p(x,y) = p_X(x) p_{Y \mid X}(y \mid x)$	$f(x,y) = f_X(x) f_{Y \mid X}(y \mid x)$

Lecture 13

October 6

13.1 Conditional Distributions

Consider two measurable spaces $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$. Then

$$(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2) = \sigma(A \times B : A \in \mathcal{S}_1, B \in \mathcal{S}_2)$$

Consider two RVs, $X : (\Omega, \mathcal{F}, P) \to (S_1, \mathcal{S}_1)$ and $Y : (\Omega, \mathcal{F}, P) \to (S_2, \mathcal{S}_2)$. (X, Y) is one RV with values in $S_1 \times S_2$. (X, Y) has a distribution μ, a PM on $S_1 \times S_2$. X has a distribution μ_1, a PM on S_1. What is the conditional distribution of Y given X?

Suppose that $S_1 = S_2 = S$ is countable. Then $P(Y = y \mid X = x) = f(y \mid x)$ has the following properties:

- $f(y \mid x) \geq 0$
- $\sum_y f(y \mid x) = 1 \; \forall x$

These properties define a stochastic matrix. The joint distribution is

$$P(X = x, Y = y) = P(X = x)P(Y = y \mid X = x)$$

Definition 13.1. A **kernel** Q from S_1 to S_2 is a map $Q : (S_1 \times \mathcal{S}_2) \to [0, 1]$ such that

(a) for fixed s_1, $B \mapsto Q(s, B)$ is a PM on S_2,

(b) for fixed $B \in \mathcal{S}_2$, $s_1 \mapsto Q(s_1, B)$ is a measurable function $S_1 \to \mathbb{R}$.

For $S_1 = S_2 = S$ countable, we have a 1-1 correspondence between Q and $f(y \mid x)$ given by

$$Q(s_1, B) = \sum_{y \in B} f(y \mid s_1)$$

Warning. If $h : S_1 \times S_2 \to \mathbb{R}$, consider:

1. h is measurable.
2. $\forall s_1$, $s_2 \mapsto h(s_1, s_2)$ is measurable $S_2 \to \mathbb{R}$ and
   $\forall s_2$, $s_1 \mapsto h(s_1, s_2)$ is measurable $S_1 \to \mathbb{R}$.

Fact. 1 implies 2, but 2 does not imply 1.

Example 13.2. Let $S_1 = S_2 = [0,1]$, with some non-measurable $A \subset [0,1]$, and consider

$$h(x,x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

Comment. We interpret $P(Y \in B \mid X = s_1) = Q(s_1, B)$.

Proposition 13.3. *Given a PM μ on $S_1 \times S_2$, a PM μ_1 on S_1, and a kernel Q from S_1 to S_2, the following are equivalent:*

$$\mu(A \times B) = \int_A Q(s_1, B)\mu_1(\mathrm{d}s_1) \qquad \forall A \in \mathcal{S}_1,\ \forall B \in \mathcal{S}_2 \tag{BR1}$$

$$\mu(D) = \int_{S_1} Q(s_1, D_{s_1})\mu_1(\mathrm{d}s_1) \qquad \forall D \in \mathcal{S}_1 \otimes \mathcal{S}_2 \tag{BR2}$$

Here, $D_{s_1} = \{s_2 : (s_1, s_2) \in D\}$.

$$\int_{S_1 \times S_2} h(s_1, s_2)\mu(\mathrm{d}\mathbf{s}) = \int_{S_1} \left(\int_{S_2} h(s_1, s_2) Q(s_1, \mathrm{d}s_2) \right) \mu_1(\mathrm{d}s_1) \tag{BR3}$$

where $\mathbf{s} = (s_1, s_2)$, *provided that h is measurable with $h \geq 0$ or h is μ-integrable.*

First, a technical lemma.

Lemma 13.4. *For each $D \in \mathcal{S}_1 \otimes \mathcal{S}_2$,*

(i) $D_{s_1} \in \mathcal{S}_2\ \forall s_1 \in \mathcal{S}_2$

(ii) The map $s_1 \mapsto Q(s_1, D_{s_1})$ is measurable.

Proof. Let $\mathcal{D}$ be the collection of all D satisfying (i) and (ii). The rectangles $A \times B$ are in $\mathcal{D}$. Apply the π-λ Theorem. If $D^n \uparrow D$, then $D^n_{s_1} \uparrow D_{s_1}$, which implies that $Q(s_1, D^n_{s_1}) \uparrow Q(s_1, D_{s_1})$. We check the λ-class property for $\mathcal{D}$. □

Outline Proof. (BR1) $\Rightarrow$ (BR2): Consider $\mathcal{D}'$, the collection of D where (BR2) holds. Use the π-λ Theorem.

(BR2) $\Rightarrow$ (BR3): Use a monotone class argument. □

Theorem 13.5 (Easy Theorem)**.** *Given a PM μ_1 on S_1, given a kernel Q from S_1 to S_2, the definition*

$$\mu(D) = \int_{S_1} Q(s_1, D_{s_1})\mu_1(\mathrm{d}s_1), \qquad D \in \mathcal{S}_1 \otimes \mathcal{S}_2$$

defines a PM μ on $S_1 \times S_2$.

Proof. The proof follows from the definitions and the properties of integrals. □

Theorem 13.6 (Hard Theorem)**.** *Given a PM μ on $S_1 \times S_2$, define the marginal PM μ_1 on S_1 by $\mu_1(A) = \mu(A \times S_2)$. If S_2 is a Borel space, then there exists a kernel Q from S_1 to S_2 such that* (BR1) *to (BR3) hold.*

Proof. Fix $B \in \mathcal{S}_2$. Consider $\nu(A) \stackrel{\text{def}}{=} \mu(A \times B)$, $A \in \mathcal{S}_1$. ν is a (sub-probability) measure on S_1. Also,

$$\nu(A) \leq \mu(A \times S_2) = \mu_1(A)$$

This implies that $\nu \ll \mu_1$. Consider the Radon-Nikodym density

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu_1}(s_1) = Q(s_1, B) \qquad \text{(definition of } Q(s_1, B))$$

which has the properties: $s_1 \mapsto Q(s_1, B)$ is measurable (requirement for a kernel), and

$$\nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu_1}(s_1)\mu_1(\mathrm{d}s_1) \qquad \Leftrightarrow \qquad \mu(A \times B) = \int_{S_1} Q(s_1, B)\,\mathrm{d}s_1 \quad \forall A \in \mathcal{S}_1$$

which is (BR1). Repeat for every $B \in \mathcal{S}_2$ to set $Q(s_1, B)$ defined. We need the second property of a "kernel", which is: $\forall s_1$, the map $B \mapsto Q(s_1, B)$ is a PM on S_2.

Issue. If $h_1 = h_2$ a.e. (with respect to μ_1), then $\int_A h_1\,\mathrm{d}\mu_1 = \int_A h_2\,\mathrm{d}\mu_1$.

Take the case where $S_2 = \mathbb{R}$. For each rational $r \in \mathbb{R}$, do the construction for $B = (-\infty, r]$. Write $F(s_1, r) = Q(s_1, (-\infty, r])$. This has the properties: $s_1 \mapsto F(s_1, r)$ is measurable, and

$$\mu(A \times (-\infty, r_1]) = \int_A F(s_1, r)\mu_1(\mathrm{d}s_1) \qquad \forall A$$

Given $r_1 < r_2$,

$$\begin{aligned} \mu(A \times (r_1, r_2]) &= \int_A (F(s_1, r_2) - F(s_1, r_1))\mu(\mathrm{d}s_1) && \forall A \\ &\geq 0, && \forall A \end{aligned}$$

which implies that $F(s_1, r_2) \geq F(s_1, r_1)$ a.e. in S_1.

Redefine $F(s_1, r) = \Phi(r)\ \forall r$ for s_1 in the null set. Repeat for all pairs (r_1, r_2). We now have a version of $(F(s_1, r))$ such that $r \mapsto F(s_1, r)$ is monotone on rational r, for all s_1 (Property A).

Easy. Modify F again to make

$$\lim_{r \uparrow \infty} F(s_1, r) = 1 \qquad \forall s_1$$
$$\lim_{r \downarrow -\infty} F(s_1, r) = 0 \qquad \forall s_1$$

(Property B). Consider $r_n \downarrow r$ (for all rationals). Then $\mu(A \times (r, r_n]) \to 0\ \forall A$, so $F(s_1, r_n) \downarrow F(s_1, r)$ a.e. Modify F again so that (Property C) $r_n \downarrow r$ (for all rationals) implies that $F(s_1, r_n) \downarrow F(s_1, r)\ \forall s_1$.

Deterministic Fact. If $r \mapsto F(r)$, where r is rational, has the properties A, B, and C, then

$$\hat{F}(x) = \lim_{\substack{r \downarrow x \\ r > x \\ r \text{ rational}}} F(r)$$

is a distribution function, with $\hat{F}(r) = F(r)$.

Use the fact to define $\hat{F}(s_1, x) = \lim_{r \downarrow x} F(s_1, r)$ $\forall x \in \mathbb{R}$. Here, $S_1 \mapsto \hat{F}(s_1, x)$ is measurable, and $x \mapsto \hat{F}(s_1, x)$ is a distribution function. Define Q by $Q(s_1, \cdot)$ is the PM with distribution function $F(s_1, x)$. $\square$

Lecture 14

October 11

14.1 Recap

Given a PM μ on $S_1 \times S_2$, there exists a marginal PM μ_1 on S_1, and (if S_2 is Borel) there exists a kernel Q from S_1 to S_2 such that (BR1) to (BR3) hold.

Interpretation: If μ is the distribution of (X, Y), then μ_1 is the distribution of X, and

$$Q(x, B) = P(Y \in B \mid X = x)$$

14.2 Product Measure

Given PMs μ_1 on $(S_1, \mathcal{S}_2)$, μ_2 on $(S_2, \mathcal{S}_2)$, there exists a "**product measure**" $\mu = \mu_1 \otimes \mu_2$ on $S_1 \times S_2$.

1. $\mu(A \times B) = \mu_1(A) \times \mu_2(B)$ for $A \in \mathcal{S}_1$ and $B \in \mathcal{S}_2$.
2. If $D \in \mathcal{S}_1 \otimes \mathcal{S}_2$, then $\mu(D) = \int \mu_2(D_{s_1})\mu_1(\mathrm{d}s_1)$.
3. For measurable $h : S_1 \times S_2 \to \mathbb{R}$,

$$\int_{S_1 \times S_2} h(s_1, s_2)\mu(\mathrm{d}\mathbf{s}) = \int_{S_1} \left[\int_{S_2} h(s_1, s_2)\mu_2(\mathrm{d}s_2) \right] \mu(\mathrm{d}s_1)$$

   provided $h \geq 0$ or $|h|$ is μ-integrable. This is **Fubini's Theorem**.

Define $Q(s_1, B) = \mu_2(B)\ \forall s_1\ \forall B$. Use (BR1) through (BR3).

Saying $\mathrm{dist}(X, Y) = \mu_1 \otimes \mu_2$ is equivalent to X and Y are independent, with $\mathrm{dist}(X) = \mu_1$ and $\mathrm{dist}(Y) = \mu_2$.

Comment. 3 works for σ-finite measures, such as λ, the Lebesgue measure on $\mathbb{R}^1$.

3, in terms of expectations, says that $Eh(X_1, X_2) = Eh_1(X_1)$, where $h_1(x_1) = Eh(x_1, X_2)$. The general identity is (usually) best viewed as calculating the same quantity in two different ways.

Example 14.1. If $X \geq 0$, then $EX = \int_0^\infty P(X \geq t)\,\mathrm{d}t$.

To prove this, let $D = \{(x, t) : x \geq t\}$ and μ be the distribution of X. $\lambda(D_x) = x$ and $D_t = (t, \infty)$, so

$$(\mu \times \lambda)(D) = \int \underbrace{\lambda(D_x)}_{x} \mu(\mathrm{d}x) = EX$$

$$(\mu \times \lambda)(D) = \int \underbrace{\mu(t, \infty)}_{P(X \geq t)} \lambda(\mathrm{d}t)$$

Example 14.2. Let X_1, X_2 be independent. For $j = 1, 2$, $\mu_j = \mathrm{dist}(X_j)$ and $\phi_j(t) = \exp(itX_j)$ for $t \in \mathbb{R}$. (Here, $i = \sqrt{-1}$.) We can prove **Parseval's identity**:

$$\int \phi_2(t)\mu_1(\mathrm{d}t) = \int \phi_1(t)\mu_2(\mathrm{d}t)$$

We know that

$$E\exp(iX_1X_2) = Eh_1(X)$$

where

$$h_1(x_1) = E\exp(ix_1X_2) = \phi_2(x_1)$$
$$E\exp(iX_1X_2) = E\phi_2(X_1) = \int \phi_2(t)\mu_1(\mathrm{d}t)$$

Do this for the other way too, and we get $E\exp(iX_1X_2) = E\phi_1(X_2)$.

Example 14.3 (Convolution Formula (Undergraduate))**.** Suppose X and Y have independent densities f_X and f_Y, with distribution functions F_X and F_Y. Then $S = X + Y$ has density

$$f(s) = \int_{-\infty}^{\infty} f_Y(s - x)f_X(x)\,\mathrm{d}x$$

Now, suppose that we have no regularity assumptions. Let $D = \{(x, y) : x + y \leq s\}$, μ_X be the distribution of X, and μ_Y be the distribution of Y.

$$P(S \leq s) = \mu_X \otimes \mu_Y(D) = \int \underbrace{\mu_Y(D_x)}_{F_Y(s-x)} \mu_X(\mathrm{d}s)$$

This implies

$$P(S \leq s) = \int F_Y(s - x)\mu_X(\mathrm{d}x)$$

Informally, differentiate with respect to s, provided that μ_Y has a density f_Y.

$$f_S(s) = \int f_Y(s - x)\mu_X(\mathrm{d}x)\,\mathrm{d}x \tag{14.1}$$

How do we justify (14.1)? Justify identities involving differentiation by checking the integrated form. We need to show

$$\begin{aligned} P(S \leq s_0) &= \int_{-\infty}^{s_0} \left(\int_{-\infty}^{\infty} f_Y(s - x)\mu_X(\mathrm{d}x) \right) \mathrm{d}s \\ &= \int \left(\int_{-\infty}^{s_0} f_Y(s - x)\,\mathrm{d}s \right) \mu_X(\mathrm{d}x) = \int F_Y(s_0 - x)\mu(\mathrm{d}x) = P(S \leq s_0) \end{aligned}$$

With a "change of variables",

$$\int \mu_X(\mathrm{d}x) = \int f_X(x)\,\mathrm{d}x$$

so if μ_X has a density f_X, then the change of variables gives

$$f_S(s) = \int f_Y(s-x) f_X(s)\,\mathrm{d}x$$

Example 14.4. Suppose (X, Y) has joint density $f(x, y)$ and a marginal density $f_1(x)$. We can define $f(y \mid x) = f(x, y)/f_1(x)$. Define the kernel Q by $Q(x, \cdot)$ is the PM with density $y \mapsto f(y \mid x)$. Then this Q is the kernel in the general theorem about $\mu = \operatorname{dist}(X, Y)$.

We need to verify (BR1).

$$P(X \in A, Y \in B) = \int_A Q(x, B)\mu_X(\mathrm{d}x)$$

$$\text{Left} = \iint 1_{(X\in A)} 1_{(Y\in B)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$\text{Right} = \int 1_{(X\in A)} \left(\int 1_{(Y\in B)} f(y \mid x)\,\mathrm{d}y \right) f_1(x)\,\mathrm{d}x \underbrace{=}_{\text{Fubini}} \iint 1_{(X\in A)} 1_{(Y\in B)} f_1(x) f(y \mid x)\,\mathrm{d}x\,\mathrm{d}y$$

14.3 RVs & PMs

Know. $X = (\Omega, \mathcal{F}, P) \to (S, \mathcal{S})$ has a distribution $\mu = \operatorname{dist}(X)$, a PM on $(S, \mathcal{S})$.

"Given μ, is there an X with $\operatorname{dist}(X) = \mu$?" has a trivial "yes" answer. We can take $(S, \mathcal{S}, \mu)$.

Know. There exists a RV U with a uniform distribution on $[0, 1]$.

Know. For any PM μ on $\mathbb{R}$, the RV $X = F_\mu^{-1}(U)$ has $\operatorname{dist}(X) = \mu$.

Know. The binary expansion $U = 0.b_1(U)b_2(U)b_3(U)\dots$ gives an infinite sequence of RVs $(b_i(U))$ which are independent,

$$P(b_i(U) = 1) = \frac{1}{2}$$
$$P(b_i(U) = 0) = \frac{1}{2}$$

Definition 14.5. $(S, \mathcal{S})$ is a **Borel space** if there exists a Borel-measurable $A \subseteq R$ and a bijection $\phi : A \to S$ such that both ϕ and ϕ^{-1} are measurable.

ϕ, the identity map from $(S_0, \mathcal{S}_1)$ to $(S_0, \mathcal{S}_2)$ is measurable iff $\mathcal{S}_2 \subseteq \mathcal{S}_1$. ϕ^{-1} is measurable iff $\mathcal{S}_1 \subseteq \mathcal{S}_2$. ϕ and ϕ^{-1} are measurable is equivalent to $\mathcal{S}_1 = \mathcal{S}_2$.

Outsource to analysis:

Theorem 14.6. *Every complete separable metric space is a Borel space.*

Consider a PM ν on a Borel space $(S, \mathcal{S})$. Let μ be the PM on A, the push-forward of ν under ϕ^{-1}. $X = F_\mu^{-1}(U)$ is a RV with distribution μ. ν is the push-forward of μ under ϕ. Then $\phi(F_\mu^{-1}(U))$ has distribution ν.

We have proved:

Lemma 14.7. *Given a PM ν on a Borel space $(S, \mathcal{S})$, there exists a measurable $h : [0,1] \to S$ such that $h(U)$ has distribution ν.*

Observation. Let π_k be the kth prime number, and $I^{(k)} = \{\pi_k, \pi_k^2, \pi_k^3, \dots\}$ is an infinite set. Then $I^{(2)}, I^{(3)}, I^{(4)}, \dots$ are disjoint. Given a sequence μ_k of PMs on $\mathbb{R}$, define $U_k = \sum_{i=1}^{\infty} 2^{-i} b_{\pi_k^i}(U)$. Then U_k is Uniform$[0,1]$, independent as k varies. Define $X_k = F_{\mu_k}^{-1}(U_k)$. We get an infinite sequence of independent RVs with the given distribution μ_k, which are all functions of *some* U. If $\mathbf{X} = (X_1, X_2, \dots)$, then $\text{dist}(\mathbf{X})$ is a PM on $\mathbb{R}^\infty$ with distribution $\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \cdots$.

Lecture 15

October 13

15.1 More "RVs & Distributions"

Corollary 15.1. *Given a PM μ on $S \times \mathbb{R}$, given a RV $X : \Omega \to S$ where $\operatorname{dist}(X) = \mu_1$ is the marginal of μ, given a RV $U : \Omega \to [0,1]$, where $\operatorname{dist}(U)$ is $\operatorname{Uniform}(0,1)$ and U is independent of X, then $\exists f : S \times [0,1] \to \mathbb{R}$ such that, writing $Y = f(X,U)$, $\operatorname{dist}(X,Y) = \mu$.*

Proof. Let Q be the kernel $S \to \mathbb{R}$ associated with μ. Let $f(s,u)$ be the inverse distribution function of the PM $Q(s,\cdot)$. $f(s,U)$ has the distribution $Q(s,\cdot)$.

Check this f works. The above statement is equivalent to $Q(s,B) = \lambda\{u : f(s,u) \in B\}$.

$$P(X \in A, Y \in B) = P(X \in A, f(X,U) \in B) = \iint 1_{(X \in A)} 1_{(f(x,u) \in B)} \, \mu(\mathrm{d}x) \otimes \lambda(\mathrm{d}u)$$

$$\underbrace{=}_{\text{Fubini}} \int 1_{(X \in A)} Q(x,B) \, \mu(\mathrm{d}x) \underbrace{=}_{\text{def. of Q}} \int \mu(A \times B)$$

□

Consider the map

$$\tilde{\pi}_{n,m} : \underbrace{(x_1, x_2, \ldots, x_n)}_{\mathbb{R}^n} \to \underbrace{(x_1, \ldots, x_m)}_{\mathbb{R}^m}$$

for $1 \le m < n < \infty$. $\pi_{m,n}$ is the associated map $\mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(\mathbb{R}^m)$ given by

$$\operatorname{dist}(X_1, \ldots, X_n) \mapsto \operatorname{dist}(X_1, \ldots, X_m)$$

Theorem 15.2 (Kolmogorov Extension (Consistency) Theorem)**.** *Given PMs μ_n on $\mathbb{R}^n$, $1 \le n < \infty$, which are consistent in the sense that $\pi_{n,m}\mu_n = \mu_m$, $1 \le m < n < \infty$, then there exists a PM μ_∞ on $\mathbb{R}^\infty$ such that $\pi_{\infty,m}\mu_\infty = \mu_m$, $1 \le m < \infty$.*

To define $(x_i, 1 \le i < \infty)$, it is enough to define x_i for each i.

To define $(X_i, 1 \le i < \infty)$, it is enough to define each X_i.

Proof. Take $U_1, U_2, \ldots,$ independent U$[0,1]$. Define $X_1 = F_{\mu_1}^{-1}(U_1)$. Inductively, suppose we have defined $\mathbf{X}_n = (X_1, \ldots, X_n)$ as functions of $(U_1, \ldots, U_n)$, such that $\operatorname{dist}(\mathbf{X}_n) = \mu_n$. We will show that there exists f_{n+1} such that, defining $X_{n+1} = f_{n+1}(\mathbf{X}_n, U_{n+1})$, we have

$$\operatorname{dist}(\mathbf{X}_{n+1} = (\mathbf{X}_n, X_{n+1})) = \mu_{n+1}$$

This constructs an infinite sequence $(X_n, 1 \le n < \infty)$. Define $\mu_\infty = \text{dist}(X_n, 1 \le n < \infty)$. Use 15.1 with $S = \mathbb{R}^n$, $X = \mathbf{X}_n$, $U = U_{n+1}$, and $\mu = \mu_{n+1}$ on $\mathbb{R}^n \times \mathbb{R}$. □

Example 15.3. Given a measurable $h : \mathbb{R} \to \mathbb{R}$, and a PM μ that is invariant under h ($\text{dist}(X) = \mu$ implies that $\text{dist}(h(X)) = \mu$), for each n, take $\text{dist}(X_n) = \mu$. Define $X_i = h(X_{i+1})$, $1 \le i \le n-1$. Let $\mu_n = \text{dist}(X_1, \ldots, X_n)$. (This is a separate construction for different n.) Then 15.2 implies that $\exists \mu_\infty = \text{dist}(Y_1, Y_2, \ldots)$ such that $\text{dist}(Y_1, \ldots, Y_n) = \text{dist}(X_1, \ldots, X_n)$ $\forall n$, where $Y_i = h(Y_{i+1})$ for all $1 \le i < \infty$.

15.2 Intermission: Example Relevant to Data

Hypothesis: Probabilities from gambling odds are indistinguishable from "true probabilities" as formalized in math.

Does this hypothesis make predictions that can be checked against data?

Consider $P(\text{home team wins})$, which starts off at 50%. The probability fluctuates over time, eventually reaching 0% or 100%. Suppose there is a half-time break. The perceived probability at half-time will change from game to game.

Model. Let Z_1 be the point difference at half-time (home team − away team) in the first half. Let Z_2 be the point difference in the second half. The home team wins if and only if $Z_1 + Z_2 > 0$. Assume $Z_1 \overset{\text{d}}{=} -Z_1$ (symmetric), with Z_1 and Z_2 independent. Suppose that Z_1 has a continuous distribution.

$$\begin{aligned}
P(\text{home team wins} \mid Z_1 = z) &= P(Z_2 \ge -z \mid Z_1 = z) \\
&= P(Z_2 \ge -z) && \text{by independence} \\
&= P(Z_2 \le z) && \text{by symmetry} \\
&= F_2(z) \\
P(\text{home team wins} \mid Z_1) &= F_2(Z_1) \\
&\overset{\text{d}}{=} \text{Uniform}[0,1]
\end{aligned}$$

15.3 Conditional Expectation in a Measure Theory Setting

Undergraduate Version. Let X, Y be $\mathbb{R}$-valued and A be an event. EX is a number. $E[X \mid A]$ is a number. $E[X \mid Y = y]$ is a number depending on y (is a function of y), which equals $h(y)$, say. Write $E[X \mid Y] = h(Y)$, which we view as a RV. This is useful because $EE[X \mid Y] = EX$.

MT Setup. X is a map from $(\Omega, \mathcal{F}, P)$ to $\mathbb{R}$, with $E|X| < \infty$. Consider a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$. We will define $E[X \mid \mathcal{G}]$ to be a certain $\mathcal{G}$-measurable RV.

$\mathcal{G}$ is "information".

EX is the *fair stake* now to get the payoff X tomorrow. The gain is $X - a$, and in order for the stake to be fair, $E[\text{gain}] = 0$ means that $a = EX$.

Suppose that we know the information in $\mathcal{G}$. The fair stake now is Y, say.

Strategy: Choose $G \in \mathcal{G}$. Bet if G happens, not if G^c happens. We gain $(X - Y)1_G$. The stake is fair if $E[\text{gain}] = 0$ for all stakes, which is equivalent to $E(X - Y)1_G = 0$ $\forall G$.

Define $E[X \mid \mathcal{G}]$ to be the RV Y satisfying:

$$Y \text{ is } \mathcal{G}\text{-measurable} \tag{15.1}$$

$$EY1_G = EX1_G \quad \forall G \in \mathcal{G} \tag{15.2}$$

15.3.1 Existence

For $G \in \mathcal{G}$, define $\nu(G) = EX1_G$. If $P(G) = 0$, then $\nu(G) = 0$, which says that $\nu \ll P$ as measures on $(\Omega, \mathcal{G})$. The Radon-Nikodym Theorem says that there is a density

$$\frac{\mathrm{d}\nu}{\mathrm{d}P}(\omega) = Y(\omega)$$

which is $\mathcal{G}$-measurable. The defining property of the Radon-Nikodym density is (15.2). (This works when ν is a signed measure.)

15.3.2 Uniqueness

Lemma 15.4. *If Y is $\mathcal{G}$-measurable, if $E|Y| < \infty$, if $E[Y1_G] \geq 0\ \forall G \in \mathcal{G}$, then $Y \geq 0$ a.s.*

Proof. If not, $G \overset{\text{def}}{=} \{Y < 0\}$ has $P(G) > 0$ and $EY1_G < 0$. Contradiction. □

Corollary 15.5. *If Y_1 and Y_2 each satisfy (15.1) and (15.2), then $Y_1 = Y_2$ a.s.*

Proof. $E(Y_1 - Y_2)1_G = 0\ \forall G$, which by 15.4 implies that $Y_1 \geq Y_2$ a.s. and $Y_1 \leq Y_2$ a.s. □

Lemma 15.6 (Technical Lemma). *(a) If $Z = E[X \mid \mathcal{G}]$, then $E[VZ] = E[VX]$ for all bounded $\mathcal{G}$-measurable V. Use the definition for $V = 1_G$ and the Monotone Class Theorem.*

(b) If Z is $\mathcal{G}$-measurable, then to prove that $Z = E[X \mid \mathcal{G}]$, it is enough to prove

$$EZ1_A = EX1_A \quad \forall A \in \mathcal{A}$$

where $\mathcal{A}$ is a π-class, $\mathcal{G} = \sigma(\mathcal{A})$. (Dynkin π-λ Lemma)

Lecture 16

October 18

16.1 Conditional Expectation

Let $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, $E|X| < \infty$, and $\mathcal{G} \subseteq \mathcal{F}$. $E[X \mid \mathcal{G}]$ is the RV Z such that

(i) Z is $\mathcal{G}$-measurable.

(ii) $E[Z1_G] = E[X1_G]\ \forall G \in \mathcal{G}$

Conditional expectation is only unique up to a null set. For example, if we write $Z = Z_1 + Z_2$ (where these are RVs as in the definition of conditional expectation), then the statement is implicitly qualified as $Z = Z_1 + Z_2$ a.s.

Lemma 16.1. *For $Z = E[X \mid \mathcal{G}]$, we have $E[VZ] = E[VX]$ for all bounded $\mathcal{G}$-measurable RVs V.*

16.1.1 General Properties of Conditional Expectation

Setting: Take a fixed $\mathcal{G}$.

Idea: The general properties of CE mimic the general properties of ordinary expectation, but with $\mathcal{G}$-measurable RVs playing the role of constants.

Properties of expectation:

- $E[X_1 + X_2] = E[X_1] + E[X_2]$
- $E[cX] = cE[X]$
- $|EX| \leq E|X|$
- $E[c] = c$

Properties of conditional expectation:

(a) $E[X_1 + X_2 \mid \mathcal{G}] = E[X_1 \mid \mathcal{G}] + E[X_2 \mid \mathcal{G}]$

(b) $E[VX \mid \mathcal{G}] = VE[X \mid \mathcal{G}]$ for all bounded $\mathcal{G}$-measurable V

(c) If $0 \leq X_n \uparrow X$ a.s., then $E[X_n \mid \mathcal{G}] \uparrow E[X \mid \mathcal{G}]$ a.s.

(d) If $X \geq 0$ a.s., then $E[X \mid \mathcal{G}] \geq 0$ a.s.

(e) $|E[X \mid \mathcal{G}]| \leq E[|X| \mid \mathcal{G}]$ a.s.

(f) $E[E[X \mid \mathcal{G}]] = EX$ (use $G = \Omega$ in the definition)

(g) If X is $\mathcal{G}$-measurable, then $E[X \mid \mathcal{G}] = X$ by definition. If $\mathcal{G}$ is trivial, then $E[X \mid \mathcal{G}] = EX$ ($\mathcal{G}$ trivial implies that $E[X \mid \mathcal{G}]$ is constant, which equals EX).

(h) If $\mathcal{G} \subseteq \mathcal{H}$, then $E[X \mid \mathcal{G}] = E[E[X \mid \mathcal{H}] \mid \mathcal{G}]$. This is called the **tower property**.

In fact, the properties above are true provided that $E|VZ| < \infty$.

Proofs. (a) Write $Z_i = E[X_i \mid \mathcal{G}]$. We need to show that $Z \stackrel{\text{def}}{=} Z_1 + Z_2 = E[X_1 + X_2 \mid \mathcal{G}]$. Is Z $\mathcal{G}$-measurable? Yes, since Z_i is $\mathcal{G}$-measurable. For the second part of the definition,

$$E[Z 1_G] = E[Z_1 1_G] + E[Z_2 1_G] = E[X_1 1_G] + E[X_2 1_G] = E[(X_1 + X_2) 1_G] \quad \forall G \in \mathcal{G}$$

(b) Define $Z = VE[X \mid \mathcal{G}]$. We need to show $Z = E[VX \mid \mathcal{G}]$. Is Z $\mathcal{G}$-measurable? Yes, since V and $E[X \mid \mathcal{G}]$ are $\mathcal{G}$-measurable.

$$E[E[X \mid \mathcal{G}] V 1_G] = E[XV 1_G] \quad \forall G \in \mathcal{G}$$

The equality is true by 16.1 applied to $V1_G$, since $V1_G$ is $\mathcal{G}$-measurable.

(c) Easy exercise.

(d) Easy exercise.

(e) Easy exercise.

(h) Write $Z = E[X \mid \mathcal{G}]$. We need to check:

$$E[Z 1_G] = E[X 1_G] = E[E[X \mid \mathcal{H}] 1_G]$$

by the definition of Z. The second equality is because of the definition of $E[X \mid \mathcal{H}]$ and $\mathcal{G} \subseteq \mathcal{H}$, so $G \in \mathcal{G}$ implies that $G \in \mathcal{H}$.

□

16.1.2 Orthogonality

$X \mapsto E[X \mid \mathcal{G}]$ is an orthogonal projection in Hilbert space. Recall from 16.1 that

$$E[(X - E[X \mid \mathcal{G}])V] = 0$$

for V $\mathcal{G}$-measurable and $EV^2 < \infty$. (By the Cauchy-Schwarz Inequality, $E|VX| \le \sqrt{(EX^2)(EV^2)} < \infty$.)

(i) $X - E[X \mid \mathcal{G}]$ and V are orthogonal for all $\mathcal{G}$-measurable V.

16.1.3 Conditional Variance

Recall that $\text{var}(X) = E\left[X - E[X]\right]^2$.

Definition 16.2. Define **conditional variance** by

$$\text{var}(X \mid \mathcal{G}) = E\left[(X - E[X \mid \mathcal{G}])^2 \mid \mathcal{G}\right]$$

(j) If Y is $\mathcal{G}$-measurable, $EY^2 < \infty$, then $E[(X-Y)^2 \mid \mathcal{G}] = \text{var}(X \mid \mathcal{G}) + (E[X \mid \mathcal{G}] - Y)^2$.

Proof.

$$\text{Left} = E[(\underbrace{(X - E[X \mid \mathcal{G}])}_{a} + \underbrace{(E[X \mid \mathcal{G}] - Y)}_{b})^2 \mid \mathcal{G}]$$

Expand the square. We have $E[ab|\mathcal{G}] = bE[a|\mathcal{G}] = 0$, so the cross-terms vanish. Since b is $\mathcal{G}$-measurable, $E[a^2 + b^2 \mid \mathcal{G}] = \mathrm{var}(X \mid \mathcal{G}) + b^2$. ☐

The constant c that minimizes $E(X-c)^2$ is $c = EX$.

(k) The $\mathcal{G}$-measurable RV that minimizes $E(X-Y)^2$ is $Y = E[X \mid \mathcal{G}]$.

Take the expectation of (j). Then

$$E(X-Y)^2 = E\,\mathrm{var}(X \mid \mathcal{G}) + E(E[X \mid \mathcal{G}] - Y)^2$$

(l) $\mathrm{var}(X) = E\,\mathrm{var}(X \mid \mathcal{G}) + \mathrm{var}\,E[X \mid \mathcal{G}]$

Proof. Replacing X by $X - c$ changes no terms, so we can assume $EX = 0$.

$$\begin{aligned}
\mathrm{var}\,X &= E[X^2] = E\left[E[X^2 \mid \mathcal{G}]\right] \\
E[X^2 \mid \mathcal{G}] &= E[(\underbrace{(X - E[X \mid \mathcal{G}])}_{a} + \underbrace{E[X \mid \mathcal{G}]}_{b})^2 \mid \mathcal{G}] \\
&= E[a^2 \mid \mathcal{G}] + b^2 \\
&= \mathrm{var}(X \mid \mathcal{G}) + (E[X \mid \mathcal{G}])^2 \\
\mathrm{var}(X) &= E\left[\mathrm{var}(X \mid \mathcal{G}) + (E[X \mid \mathcal{G}])^2\right] \\
&= E\,\mathrm{var}(X \mid \mathcal{G}) + \underbrace{E(E[X \mid \mathcal{G}] - 0)^2}_{=\mathrm{var}\,E[X \mid \mathcal{G}]}
\end{aligned}$$

since $E[ab \mid \mathcal{G}] = 0$ and $E[E[X \mid \mathcal{G}]] = EX = 0$. ☐

16.1.4 Independence

What is the connection with independence?

(m) X is independent of $\mathcal{G}$ iff

$$E[h(X) \mid \mathcal{G}] = Eh(X) \qquad \text{for all bounded measurable } h : S \to \mathbb{R} \tag{16.1}$$

Here, X can be S-valued.

Proof. Suppose X is independent of $\mathcal{G}$. We need to show:

$$E[(Eh(X))1_G] = (Eh(X))(E1_G) = E[h(X)1_G]$$

This holds by independence.

Suppose that (16.1) holds. Take $h = 1_B$ for $B \subseteq S$. (16.1) implies (by the same argument as above)

$$P(X \in B, G) = E[h(X)1_G] = E[h(X)]E[1_G] = P(X \in B)P(G)$$

for all B and G, which implies that X and $\mathcal{G}$ are independent. ☐

Recall that X and Y are independent if and only if $E[h_1(X)h_2(Y)] = (Eh_1(X))(Eh_2(Y))\ \forall h_1, h_2$.

16.2 Background to Conditional Independence

There are three general contexts in which this idea arises.

1. Bayes
   (a) Take a random Θ, which takes values in $\{\text{PMs on } \mathbb{R}^1\} = \mathcal{P}(\mathbb{R})$.
   (b) Conditional on $\Theta = \theta \in \mathcal{P}(\mathbb{R})$, take $X_1, X_2, X_3, \ldots$ which are IID θ.

   The (X_i) are conditionally independent given Θ.

2. The simple Markov property for $(X_n, n \geq 0)$

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

   (X_{n+1}) and $(X_{n-1}, X_{n-2}, \ldots, X_0)$ are conditionally independent given X_n.

3. Given $(W_{\mathbf{x}}, \mathbf{x} = (x_1, x_2) \in \mathbb{Z}^2)$, let $N(\mathbf{x})$ be the neighbors of $\mathbf{x}$. The idea is that $W_{\mathbf{x}}$ depends only on $\{W_{\mathbf{y}}, \mathbf{y} \in N(\mathbf{x})\}$ and not on the other Ws. We formalize the idea as $W_{\mathbf{x}}$ and $(W_{\mathbf{z}}, \mathbf{z} \notin N(\mathbf{x}) \cup \{\mathbf{x}\})$ are conditionally independent given $\{W_{\mathbf{y}}, \mathbf{y} \in N(\mathbf{x})\}$.

Lecture 17

October 20

17.1 Two Final "Conditioning" Topics

Recall Jensen's inequality: $E\phi(X) \geq \phi(EX)$ if ϕ is convex, if $E|X| < \infty$ and $E|\phi(X)| < \infty$.

(n) Conditional Jensen's inequality: $E[\phi(X) \mid \mathcal{G}] \geq \phi(E[X \mid \mathcal{G}])$ a.s.

17.1.1 Conditional Independence

Recall that in MT, independence is a property of $\mathcal{G}_1$ and $\mathcal{G}_2$. The RVs X_1 and X_2 are independent if $\sigma(X_1)$ and $\sigma(X_2)$ are independent. Recall that for $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{S})$, $\sigma(X) \subseteq \mathcal{F}$. Independence is also equivalent to $E[h_1(X_1)h_2(X_2)] = (Eh_1(X_1))(Eh_2(X_2))$ for all $h_i : S_i \to \mathbb{R}$ which are bounded and measurable. This is also equivalent to $E[h_1(X_1) \mid X_2] = Eh_1(X_1)$ a.s. for all h_1.

Undergraduate Setting. Given a discrete RV V, define $P(X_1 = x_1 \mid V = v)$ and define $P(X_2 = x_2 \mid V = v)$. Then, we can construct (X_1, X_2, V) such that

$$P(X_1 = x_1, X_2 = x_2 \mid V = v) = P(X_1 = x_1 \mid V = v) \times P(X_2 = x_2 \mid V = v)$$

Definition 17.1. X_1 and X_2 are **conditionally independent** (CI) given $\mathcal{G}$ means

$$E[h_1(X_1)h_2(X_2) \mid \mathcal{G}] = E[h_1(X_1) \mid \mathcal{G}] \times E[h_2(X_2) \mid \mathcal{G}] \quad \forall h_i$$

We can replace X_1 with a σ-field $\mathcal{H}_1$, and $h_1(X_1)$ with a bounded $\mathcal{H}_1$-measurable RV.

Homework (Later). This is equivalent to $E[h_1(X_1) \mid \mathcal{G}, X_2] = E[h_1(X_1) \mid \mathcal{G}]$ a.s. for all h_1. Once you know $\mathcal{G}$, knowing also X_2 gives no *extra* info about X_1.

17.1.2 Conditional Probability & Conditional Expectation

Undergraduate. We define a conditional P by

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

and a conditional E by

$$E[h(Y) \mid X = x] = \sum_y h(y)P(Y = y \mid X = x)$$

and the two concepts are related.

From $(X,Y):(\Omega,\mathcal{F},P)\to S_1\times S_2$, we get a kernel Q from S_1 to S_2, where $Q(x,B)$ means $P(Y\in B\,|\,X=x)$. Given $W:(\Omega,\mathcal{F},P)\to\mathbb{R}$, $E|W|<\infty$, $\mathcal{G}\subseteq\mathcal{F}$, we defined $E[W\,|\,\mathcal{G}]=Z$, specified by $E[Z1_G]=E[W1_G]$ for all $G\in\mathcal{G}$. What is the relationship between these two concepts?

Write $W=h(Y)$, where $h:S_2\to\mathbb{R}$. Write $\mathcal{G}=\sigma(X)$. Write $I:(\Omega,\mathcal{F})\to(\Omega,\mathcal{G})$, the identity function. We have $(I,Y):\Omega\to(\Omega,\mathcal{G})\times(S_2,\mathcal{S}_2)$. Write $\alpha(\omega,B)$ for the kernel associated with (I,Y). Then $\alpha(\omega,B)$ means $P(Y\in B\,|\,\mathcal{G})(\omega)$.

We can start from conditional expectation: let $P(A)=E[1_A]$. Define $P(A\,|\,\mathcal{G})(\omega)=E[1_A\,|\,\mathcal{G}](\omega)$. Then $\alpha(\cdot,B)=P(Y\in B\,|\,\mathcal{G})$. This is the **regular conditional distribution for Y given $\mathcal{G}$**. It is "regular" in the sense that $B\mapsto\alpha(\omega,B)$ is a PM.

What is this in MT?

$$E[h(Y)\,|\,\mathcal{G}](\omega)=\int h(y)\alpha(\omega,\mathrm{d}y)$$

(Homework)

17.2 Martingales

A σ-field $\mathcal{G}$ is a collection of events: $A\in\mathcal{G}$, where A is an event. For a RV X, "X is $\mathcal{G}$-measurable" means $\sigma(X)\subseteq\mathcal{G}$. We use the shorthand $X\in\mathcal{G}$. (This can, in principle, cause confusion: consider $J\in\mathcal{F}$?)

17.2.1 General Setup $(\Omega,\mathcal{F},P)$

Sub-σ-fields $\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\mathcal{F}_2\subseteq\cdots\subseteq\mathcal{F}$ form a **filtration**. We interpret $\mathcal{F}_n$ as the "information known at time n".

A sequence $(X_n,n\geq 0)$ is **adapted** to $(\mathcal{F}_n)$ means $X_n\in\mathcal{F}_n\ \forall n$.

Definition 17.2. A $\mathbb{R}$-valued process $(X_n,0\leq n<\infty)$ is a **martingale** (MG) if

(i) $E|X_n|<\infty\ \forall n$

(ii) (X_n) is adapted to $(\mathcal{F}_n)$

(iii) $E[X_{n+1}\,|\,\mathcal{F}_n]=X_n,\ 0\leq n<\infty$

In condition (iii), if we have $E[X_{n+1}\,|\,\mathcal{F}_n]\geq X_n$, we have a **submartingale**. If $E[X_{n+1}\,|\,\mathcal{F}_n]\leq X_n$, we have a **supermartingale**.

Note that (iii) can be rewritten as $E[X_{n+1}-X_n\,|\,\mathcal{F}_n]=0\ \forall n$.

Typical Use of Theory: We have a complicated system (Y_n) and we look for h such that $h(Y_n)$ is a MG. Take $\mathcal{F}_n=\sigma(Y_0,Y_1,\ldots,Y_n)$. If we take $X_n=h(Y_n)$, then (X_n) *is* adapated to $(\mathcal{F}_n)$.

If we define X_n and we say "X_n is a MG", then we are taking $\mathcal{F}_n=\sigma(X_0,X_1,\ldots,X_n)$, the **natural filtration** for (X_n).

17.2.2 Examples Based on Independent RVs $\xi_1,\xi_2,\xi_3,\ldots$, $\mathcal{F}_n=\sigma(\xi_1,\ldots,\xi_n)$

Example 17.3. If $E|\xi_i| < \infty$ and $E\xi_i = 0\ \forall i$, then $S_n = \sum_{i=1}^n \xi_i$ is a MG.

$$E[S_{n+1} \mid \mathcal{F}_n] = E[S_n + \xi_{n+1} \mid \mathcal{F}_n] = S_n + E[\xi_{n+1} \mid \mathcal{F}_n] = S_n + \underbrace{E\xi_{n+1}}_{0} = S_n$$

because $S_n \in \mathcal{F}_n$ and S_{n+1} is independent of $\mathcal{F}_n$.

Example 17.4. As in 17.3, suppose also $\sigma_i^2 = E\xi_i^2 < \infty$. Then $Q_n = S_n^2 - \sum_{i=1}^n \sigma_i^2$ is a MG.

$$\begin{aligned}
Q_{n+1} - Q_n &= S_{n+1}^2 - S_n^2 - \sigma_{n+1}^2 = 2S_n\xi_{n+1} + \xi_{n+1}^2 - \sigma_{n+1}^2 \\
E[Q_{n+1} - Q_n \mid \mathcal{F}_n] &= \underbrace{E[2S_n\xi_{n+1} \mid \mathcal{F}_n]}_{S_n \in \mathcal{F}_n} + \underbrace{E[\xi_{n+1}^2 \mid \mathcal{F}_n]}_{\text{indep.}} - \sigma_{n+1}^2 \\
&= 2S_n \underbrace{E[S_{n+1} \mid \mathcal{F}_n]}_{=0} + E[\xi_{n+1}^2] - \sigma_{n+1}^2 \\
&= 0
\end{aligned}$$

Example 17.5. Suppose (ξ_i) are independent, $E\xi_i = 1$. Then $M_n = \prod_{i=1}^n \xi_i$ is a MG.

$$\begin{aligned}
M_{n+1} &= M_n\xi_{n+1}, \qquad M_n \in \mathcal{F}_n \\
E[M_{n+1} \mid \mathcal{F}_n] &= E[M_n\xi_{n+1} \mid \mathcal{F}_n] = M_n E[\xi_{n+1} \mid \mathcal{F}_n] \\
&= M_n E[\xi_{n+1}] \\
&= M_n \cdot 1
\end{aligned}$$

Example 17.6. Suppose that (ξ_i) are independent. Fix t, $S_n = \sum_{i=1}^n \xi_i$, and suppose that we have $\phi_i(t) \overset{\text{def}}{=} E\exp(t\xi_i) < \infty$. Then

$$X_n = \frac{\exp(tS_n)}{\prod_{i=1}^n \phi_i(t)}$$

is a MG.

$$\begin{aligned}
X_i &= \prod_{i=1}^n Y_i \\
Y_i &= \frac{\exp(t\xi_i)}{\phi_i(t)}
\end{aligned}$$

By independence, the expectation is 1.

Example 17.7. Take (ξ_i) IID. Take density functions f and $g > 0$. Define the likelihood ratio

$$L_n = \prod_{i=1}^n \frac{g(\xi_i)}{f(\xi_i)}$$

(a) If the (ξ_i) have density f, then $(\forall g)$ (L_n) is a MG.

$$\begin{aligned}
L_n &= \prod_{i=1}^n Y_i, \\
Y_i &= \frac{g(\xi_i)}{f(\xi_i)}
\end{aligned}$$

$$EY_i = \int \frac{g(y)}{f(y)} \cdot f(y)\,\mathrm{d}y$$
$$= \int g(y)\,\mathrm{d}y = 1$$

(b) If the (ξ_i) have density g, then, provided that $EL_n < \infty$, (L_n) is a sub-MG. (a) implies that $(1/L_n, n \geq 0)$ is a MG.

$$\frac{1}{L_n} = E\left[\frac{1}{L_{n+1}} \,\middle|\, \mathcal{F}_n\right] \geq \frac{1}{E[L_{n+1} \mid \mathcal{F}_n]}$$

by Conditional Jensen's Inequality. Therefore, $E[L_{n+1} \mid \mathcal{F}_n] \geq L_n$, so this is a sub-MG.

Lecture 18

October 25

18.1 General Constructions of MGs

Consider a filtration $(\mathcal{F}_n, 0 \le n < \infty)$ on $(\Omega, \mathcal{F}, P)$. Recall that $(X_n, 0 \le n < \infty)$ is **adapted** to $(\mathcal{F}_n)$ means $X_n \in \mathcal{F}_n$, $0 \le n < \infty$.

We can define $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n) \subseteq \mathcal{F}$. We are usually *not* given a RV X_∞. When we consider X_T for a stopping time T, we need to care about $\{T = \infty\}$.

Example 18.1. Consider any X with $E|X| < \infty$, then $X_n = E[X \mid \mathcal{F}_n]$, $0 \le n < \infty$ is a MG.

$$\begin{aligned} E[X_n \mid \mathcal{F}_{n-1}] &= E\left[E[X \mid \mathcal{F}_n] \mid \mathcal{F}_{n-1}\right] \\ &= E[X \mid \mathcal{F}_{n-1}] = X_{n-1} \end{aligned}$$

by the Tower Property, since $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$.

Similarly, for any event A, $Y_n = P(A \mid \mathcal{F}_n)$ is a MG.

Notation. For any $X = (X_n)$, define $\Delta_n^X = X_n - X_{n-1}$, $n \ge 1$. Then $(X_n, n \ge 0)$ is a MG if and only if $\Delta_n^X \in \mathcal{F}_n$ for $n \ge 1$, $E\left|\Delta_n^X\right| < \infty$ for $n \ge 1$, $E[\Delta_n^X \mid \mathcal{F}_n] = 0$ a.s. for $n \ge 1$, and $X_0 \in \mathcal{F}_0$, $E|X_0| < \infty$. Call $(\Delta_n^X, n \ge 1)$ a **martingale difference sequence**. To get the sub-MG property, $E[\Delta_n^X \mid \mathcal{F}_{n-1}] \ge 0$ a.s. for $n \ge 1$.

Example 18.2. Consider any $(X_n, n \ge 0)$, adapted to $(\mathcal{F}_n)$ and $E|X_n| < \infty$ $\forall n$. Define (Y_n) by $Y_0 = X_0$, $\Delta_n^Y = \Delta_n^X - E[\Delta_n^X \mid \mathcal{F}_{n-1}]$. Define (Z_n) by $Z_0 = 0$, $\Delta_n^Z = E[\Delta_n^X \mid \mathcal{F}_{n-1}]$. Then

(i) $X_n = Y_n + Z_n$

(ii) (Y_n) is a MG.

(iii) $Z_n \in \mathcal{F}_{n-1}$, for $n \ge 1$ and $Z_n = 0$. (Z_n) is **predictable** and $E|Z_n| < \infty$.

This is the *unique* decomposition with these properties.

Why is this unique?

$$\begin{aligned} E[\Delta_n^X \mid \mathcal{F}_{n-1}] &= E[\Delta_n^Y \mid \mathcal{F}_{n-1}] + E[\Delta_n^Z \mid \mathcal{F}_{n-1}] \\ &= 0 + \Delta_n^Z \end{aligned}$$

since (Y_n) is a MG and Z is predictable.

This is called the **Doob decomposition**.

If (X_n) is a MG, then $(X_n - X_0, n \geq 0)$ is a MG. We often say "WLOG assume $X_0 = 0$".

For a MG, $E[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$ implies that $EX_n = EX_{n-1}$, which implies that $EX_n = EX_0$ $\forall n$. For a sub-MG, $E[X_n \mid \mathcal{F}_{n-1}] \geq X_{n-1}$, which implies that $EX_n \geq EX_{n-1}$, which implies that $EX_n \geq EX_0$ $\forall n$.

Theorem 18.3 (Convexity Theorem)**.** *Let (X_n) be adapted to $(\mathcal{F}_n)$, ϕ be convex, and $E|\phi(X_n)| < \infty$.*

(a) If (X_n) is a MG, then $\phi(X_n)$ is a sub-MG.

(b) If (X_n) is a sub-MG and if ϕ is increasing, then $\phi(X_n)$ is a sub-MG.

Proof. (b)

$$\begin{aligned} E[\phi(X_{n+1}) \mid \mathcal{F}_n] &\geq \phi(\underbrace{E[X_{n+1} \mid \mathcal{F}_n]}_{\geq X_n}) \\ &\geq \phi(X_n) \end{aligned}$$

where we used Conditional Jensen, (X_n) is a sub-MG, and ϕ is increasing. Hence, $\phi(X_n)$ *is* a sub-MG. We have equality if (X_n) is a MG.

□

Example 18.4. If (X_n) is a MG, then (provided integrable)

(i) $|X_n|^p$ $(p \geq 1)$ is a sub-MG, because $x \mapsto |x|^p$ is convex

(ii) X_n^2 is a sub-MG

(iii) $\exp(\theta X_n)$, $(-\infty < \theta < \infty)$ is a sub-MG, because $x \mapsto e^{\theta x}$ is convex

(iv) $\max(X_n, c)$ is a sub-MG, because $x \mapsto \max(x, c)$ is convex

(v) $\min(X_n, c)$ is a super-MG

18.2 Stopping Times

Definition 18.5. A RV $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is a **stopping time** if

$$\{T = n\} \in \mathcal{F}_n, \quad 0 \leq n < \infty \tag{18.1}$$

This implies that $\{T = \infty\} \in \mathcal{F}_\infty$. Equivalently,

$$\{T \leq n\} \in \mathcal{F}_n, \quad 0 \leq n < \infty \tag{18.2}$$

Definition 18.6. For a stopping time T, define $\mathcal{F}_T$ as the collection of sets $A \in \mathcal{F}$ such that

$$A \cap \{T = n\} \in \mathcal{F}_n, \quad 0 \leq n < \infty \tag{18.3}$$

or equivalently,

$$A \cap \{T \leq n\} \in \mathcal{F}_n, \quad 0 \leq n < \infty \tag{18.4}$$

This is the **pre-T σ-field**.

There are many "obvious" properties.

1. If (X_n) is adapted, if T is a stopping time, $T < \infty$, then X_T is $\mathcal{F}_T$-measurable.

Proof. *Want*: $\{X_T \in B\} \in \mathcal{F}_T \ \forall B$.

Want: $\{X_T \in B\} \cap \{T = n\} \in \mathcal{F}_n$. This is the same as $\{X_n \in B\} \cap \{T = n\}$. $\{X_n \in B\} \in \mathcal{F}_n$ since X_n is adapted. $\{T = n\} \in \mathcal{F}_n$ by the definition of a stopping time. □

2. If $T_1 \le T_2$ are stopping times, then $\mathcal{F}_{T_1} \subseteq \mathcal{F}_{T_2}$.
3. If S and T are stopping times, then $\{S = T\} \in \mathcal{F}_S \cap \mathcal{F}_T$, and for $A \subseteq \{S = T\}$,

$$A \in \mathcal{F}_S \Leftrightarrow A \in \mathcal{F}_T$$

Given an adapted (X_n) and a stopping time T, the process $\hat{X}_n = X_{\min(n,T)}$ is adapted. Call $\hat{X}$ the "**stopped process**".

Story. $\mathcal{F}_n$ is the information at the end of day n. You can buy a stock at the end of any day n. X_n is the price of 1 share at the end of day n. H_n is the number of shares I hold during day n (they must be bought at day $n-1$ or earlier). Therefore, $H_n \in \mathcal{F}_{n-1}$. Y_n is my accumulated profit at the end of day n. What is the relation? $\Delta_n^Y = H_n \Delta_n^X$. Also, $Y_0 = 0$. Write $Y = H \cdot X$, a "**martingale transform**" or a "**discrete-time stochastic integral**".

Theorem 18.7 (Durrett 2.7)**.** *Suppose (X_n) is adapted and (H_n) is predictable. Consider $Y = H \cdot X$ (for simplicity, assume H_n is bounded).*

(i) If (X_n) is a MG, then (Y_n) is a MG.

(ii) If (X_n) is a sub-MG and $H_n \ge 0$, then (Y_n) is a sub-MG.

Proof. (ii)

$$\begin{aligned} E[\Delta_n^Y] &= \underbrace{E[H_n \Delta_n^X \mid \mathcal{F}_{n-1}]}_{H_n \in \mathcal{F}_{n-1}} \\ &= \underbrace{H_n}_{\ge 0} \underbrace{E[\Delta_n^X \mid \mathcal{F}_{n-1}]}_{\ge 0} \\ &\ge 0 \end{aligned}$$

since (X_n) is a sub-MG. Therefore, (Y_n) is a sub-MG.

□

Corollary 18.8. *If (X_n) is a (sub-)MG, if T is a stopping time, then $\hat{X}_n = X_{\min(n,T)}$ is a (sub-)MG.*

Proof. Buy 1 share at the end of day 0 and sell at the end of day T.

$$H_n = 1_{(0 \le n \le T)}$$

(H_n) is *predictable* because $\{n \le T\} = \{T \le n-1\}^c \in \mathcal{F}_{n-1}$. The process $Y = H \cdot X$ is explicitly

$Y_n = X_{\min(n,T)} - X_0$. Apply 18.7. □

Lecture 19

October 27

19.1 Optional Sampling Theorem

Last class: let (X_n) be a sub-MG w.r.t. $(\mathcal{F}_n)$. (H_n) is a predictable process which is bounded. Define $Y = H \cdot X$ by $Y_0 = 0$, $\Delta_n^Y = H_n \Delta_n^X$. Then (Y_n) is a sub-MG, provided $H_n \geq 0$. H_n is the number of shares held on day n.

The sub-MG property is

$$E[X_n \mid \mathcal{F}_{n-1}] \geq X_{n-1}$$

which implies that EX_n is increasing.

Corollary 19.1. *Let (X_n) be a sub-MG. Let $0 \leq T_1 \leq T_2 \leq t_0$ be stopping times. Then*

$$E[X_{T_2} \mid \mathcal{F}_{T_1}] \geq X_{T_1}$$

Proof. Fix an event $A \in \mathcal{F}_{T_1}$. The strategy is: "If A happens, buy 1 share at T_1 and sell at T_2. If A does not happen, do nothing." $H_n = 1_A 1_{(T_1 < n \leq T_2)}$. We want to check that H_n is predictable. In other words, we want to check $A \cap \{T_1 < n \leq T_2\} \in \mathcal{F}_{n-1}$. We can write $\{T_1 < n \leq T_2\}$ as $\{T_1 \leq n-1\} \setminus \{T_2 \leq n-1\}$, because $T_2 \geq T_1$, so we have $(A \cap \{T_1 \leq n-1\}) \setminus (A \cap \{T_2 \leq n-1\})$. By the definition of $A \in \mathcal{F}_{T_1}$, the two events are in $\mathcal{F}_{n-1}$.

So, (Y_n) is a sub-MG. $Y_n = (X_{T_2 \wedge n} - X_{T_1 \wedge n}) 1_A$, where $a \wedge b = \min(a, b)$. The sub-MG property implies that $EY_{t_0} \geq EY_0 = 0$. We have shown

$$E[(X_{T_2} - X_{T_1}) 1_A] \geq 0 \quad \forall A \in \mathcal{F}_{T_1}$$

Fact. If $E[Z 1_A] \geq 0 \; \forall A \in \mathcal{G}$, then $E[Z \mid \mathcal{G}] \geq 0$ a.s. Therefore,

$$E[X_{T_2} - X_{T_1} \mid \mathcal{F}_{T_1}] \geq 0 \quad \text{a.s.}$$ □

"OST" is the **Optional Sampling Theorem**.

Theorem 19.2 (Basic Version of OST)**.** *If (X_n) is a (sub-)MG, $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$ are stopping times, if $T_i \leq t_i$ (a constant), then $(X_{T_i}, i = 0, 1, 2, \ldots)$ is a (sub-)MG w.r.t. $(\mathcal{F}_{T_i}, i = 0, 1, 2, \ldots)$.*

In particular, $EX_{T_i} \geq EX_0$ for a sub-MG and $EX_{T_i} = EX_0$ for a MG, and $EX_{T_2} \geq EX_{T_1}$ if $T_2 \geq T_1$. There are many other versions without the restriction that $T \leq t_0$.

Write $X_N^* = \max(X_0, X_1, \dots, X_N)$. We know that $P(X_N^* \geq x) \leq \sum_{n=0}^N P(X_n \geq x)$ is always true. If the (X_i) are independent, then $P(X_N^* \geq x) = 1 - \prod_{n=0}^N P(X_n < x)$. With MGs, we can get better bounds than the former.

19.2 Maximal Inequalities

Lemma 19.3. *Let (X_n) be a super-MG, $X_n \geq 0$ a.s. Write $X^* = \sup_n X_n$, so $X_N^* \uparrow X^*$ as $N \to \infty$. Then $P(X^* \geq \lambda) \leq EX_0/\lambda$, for all $\lambda > 0$.*

Proof. Define $T = \min\{n : X_n \geq \lambda\}$. Apply the OST 19.2 to 0 and $T \wedge N$. Then

$$\begin{aligned} EX_0 \geq EX_{T\wedge N} &= EX_T 1_{(T\leq N)} + EX_N 1_{(T>N)} \\ &\geq \lambda P(T \leq N) + 0 \end{aligned}$$

This implies

$$P(T \leq N) \leq \lambda^{-1} EX_0$$

The event $\{T \leq N\}$ is the same as the event $\{X_N^* \geq \lambda\}$. Therefore

$$P(X_N^* \geq \lambda) \leq \lambda^{-1} EX_0$$

Let $N \to \infty$. Then we only have

$$P(X^* > \lambda) \leq \lambda^{-1} EX_0$$

Apply this to $\lambda_j \uparrow \lambda$ to obtain

$$P(X^* \geq \lambda) \leq \lambda^{-1} EX_0$$

(Check this.) □

Lemma 19.4 (Doob's (L^1) Maximal Inequality)**.** *Let (X_n) be a sub-MG. For $\lambda > 0$,*

$$\lambda P(X_N^* \geq \lambda) \leq E[X_N 1_{(X_N^* \geq \lambda)}] \leq EX_N^+ = E\max(X, 0)$$

Note that

$$\max_A E[Y 1_A] = E[Y 1_{(Y\geq 0)}] = E\max(Y, 0) = EY^+$$

which implies that $E[Y1_A] \leq EY^+$.

Proof. Let $T = \min\{n : X_n \geq \lambda\}$. Apply the OST 19.2 to $T \wedge N$ and N: $EX_{T\wedge N} \leq EX_N$. Therefore,

$$EX_T 1_{(T\leq N)} + EX_N 1_{(T>N)} \leq EX_N 1_{(T\leq N)} + EX_N 1_{(T>N)}$$

$X_T \geq \lambda$, so

$$\lambda P(T \leq N) \leq EX_N 1_{(T\leq N)} = EX_N 1_{(X_N^* \geq \lambda)}$$ □

Corollary 19.5. *If (X_n) is a MG, then (because $Y_n = |X_n|$ is also a sub-MG)*

$$P\left(\max_{0\leq n\leq N} |X_n| \geq \lambda\right) \leq \frac{E|X_N|}{\lambda}$$

Also, $Z_n = X_n^2$ is a sub-MG (provided $EX_n^2 < \infty$). Apply 19.4 to (Z_n):

$$\lambda^2 P\left(\max_{0\le n\le N} X_n^2 \ge \lambda^2\right) \le EX_N^2$$

or

$$P\left(\max_{0\le n\le N} |X_n| \ge \lambda\right) \le \frac{EX_N^2}{\lambda^2}$$

These are two different bounds for the same quantity.

Use the notation

$$a \vee b = \max(a,b)$$
$$a \wedge b = \min(a,b)$$

Theorem 19.6 (Doob's L^2 Maximal Inequality)**.** *Let (X_n) be a sub-MG. Then*

$$E[(0\vee X_N^*)^2] \le 4E[(X_N^+)^2]$$

Proof.

$$\begin{aligned}
E[(0\vee Z)^2] &= 2\int_0^\infty \lambda P(Z\ge\lambda)\,\mathrm{d}\lambda \\
\underbrace{E[(0\vee X_N^*)^2]}_{a} &= 2\int_0^\infty \lambda P(X_N^*\ge\lambda)\,\mathrm{d}\lambda \le 2\int_0^\infty E[X_N 1_{(X_N^*\ge\lambda)}]\,\mathrm{d}\lambda \\
&\le 2\int_0^\infty E[X_N^+ 1_{(X_N^*\ge\lambda)}]\,\mathrm{d}\lambda \\
&\le 2E\left[X_N^+\int_0^\infty 1_{(X_N^*\ge\lambda)}\,\mathrm{d}\lambda\right] \\
&= 2E[X_N^+(0\vee X_N^*)] \\
&\le 2(\underbrace{E[(X_N^+)^2]}_{b}\times\underbrace{E[(0\vee X_N^*)^2]}_{a})^{1/2}
\end{aligned}$$

by the Cauchy-Schwarz Inequality. The inequality is saying $a \le 2\sqrt{ba}$, so $a\le 4b$. There is also a special case when $a=\infty$. □

If we use the Hölder Inequality instead of the Cauchy-Schwarz Inequality, then we obtain

$$E[(0\vee X_N^*)^p] \le \left(\frac{p}{p-1}\right)^p E[(X_N^+)^p]$$

for $1<p<\infty$. This is *not* true for $p=1$.

Example 19.7. Let $X_0 = 1$ and consider a simple symmetric RW on $\mathbb{Z}$, stopping at

$$T = \min\{n\ge 1 : X_n = 0\}$$

(X_n) is a MG. $EX_n = 1\ \forall n$. Also, $X_N^* \uparrow X^* = \sup_n X_n$. Elementary fact: $P(X^*\ge m) = 1/m$. Therefore, $EX^* = \infty$, so $EX_N^* \uparrow \infty$, but $EX_N = 1\ \forall N$.

Lecture 20

November 1

20.1 Upcrossing Inequality

Take any $\mathbb{R}$-valued $(X_n, n \geq 0)$ and any $a < b$. Define $S_1 = \min\{n : X_n \leq a\}$, $T_1 = \min\{n : X_n \geq b\}$, $S_2 = \min\{n > T_1 : X_n \leq a\}$, $T_2 = \min\{n > S_2 : X_n \geq b\}$, etc.

Define $U_n = U_n[a,b] = \max\{k : T_k \leq n\}$, the number of upcrossings over $[a,b]$ completed by time n.

Theorem 20.1 (The Upcrossing Inequality). *Suppose (X_n) is a sub-MG. Then*

$$\begin{aligned}(b-a)EU_n &\leq E(X_n - a)^+ - E(X_0 - a)^+ \\ &\leq EX_n^+ + |a|\end{aligned}$$

Proof. Note that $(x-a)^+ \leq x^+ + |a|$, so $E(X-a)^+ \leq EX^+ + |a|$.

(*Trick*) In the case that $X_n \geq a \; \forall n$, we will prove $(b-a)EU_n \leq EX_n^+ - EX_0^+$. For general (X_n), apply the result to $\max(X_n, a) - a$, which is a sub-MG.

Use the "buy low, sell high" strategy: buy 1 share at S_i, and sell 1 share at T_i. Consider $Y = H \cdot X$, where $H_n = 1_{(S_1 < n \leq T_1)} + 1_{(S_2 < n \leq T_2)} + \cdots$. This is a predictable process, so (Y_n) is a sub-MG.

$$\begin{aligned}Y_n &= \sum_{i=1}^{U_n} (X_{T_i} - X_{S_i}) + (X_n - X_{S_{U_n+1}}) 1_{(n > S_{U_n+1})} \\ &\geq (b-a)U_n + 0\end{aligned}$$

Take expectations.

$$(b-a)EU_n \leq EY_n$$

Consider the opposite strategy K: $K_n = 1 - H_n$. $(X_n - Y_n) = (K \cdot X)_n + X_0$ is a sub-MG.

$$\begin{aligned}E[X_0 - Y_0] &\leq E[X_n - Y_n] \\ EX_0 &\leq EX_n - EY_n \\ (b-a)EU_n &\leq EX_n - EX_0\end{aligned}$$

□

20.2 Martingale Convergence

Theorem 20.2 (Martingale Convergence Theorem)**.** *If (X_n) is a sub-MG, if $\sup_n EX_n^+ < \infty$, then $X_n \to X_\infty$ a.s., for some X_∞ with $E|X_\infty| < \infty$.*

Proof. $U_n[a,b] \uparrow U_\infty[a,b]$, so

$$EU_\infty[a,b] = \lim_n EU_n[a,b] \le \frac{\sup_n EX_n^+ + |a|}{b-a}$$

which implies that $U_\infty[a,b] < \infty$ a.s. This implies

$$P(U_n[a,b] < \infty, \text{all rational pairs } a < b) = 1$$

For reals (x_n), if $\limsup_n x_n > \liminf_n x_n$, then $U_\infty[a,b] = \infty$, for some $a < b$. Since $U_\infty[a,b] < \infty$ for all rational $a < b$, then $\limsup x_n = \liminf x_n \in [-\infty, \infty]$. Therefore, $X_n \to X_\infty$ a.s., but $X_\infty \in [-\infty, \infty]$.

Recall Fatou's Lemma: If $Y_n \ge 0$,

$$E \liminf_n Y_n \le \liminf_n EY_n$$

$X_n^+ \to X_\infty^+$ a.s. implies (by Fatou's Lemma) that $EX_\infty^+ \le \liminf_n EX_n^+ < \infty$. Also,

$$EX_n^- = EX_n^+ - EX_n \le EX_n^+ - EX_0$$

since $EX_0 \le EX_n$. Since $X_n^- \to X_\infty^-$ a.s., by Fatou's Lemma,

$$EX_\infty^- \le \liminf_n EX_n^- \le \sup_n EX_n^+ - EX_0 < \infty$$

Since $EX_\infty^+ < \infty$ and $EX_\infty^- < \infty$, then $E|X_\infty| < \infty$. □

Corollary 20.3. *If (X_n) is a super-MG, if $X_n \ge 0$ a.s., then $X_n \to X_\infty$ a.s. and $0 \le EX_\infty \le EX_0$.*

Proof. Apply 20.2 to $(-X_n)$, so $X_n \to X_\infty$ a.s. Use Fatou's Lemma: $EX_\infty \le \liminf_n EX_n \le EX_0$. □

Recall the simple RW $X_0 = 1$, stopped at $T = \min\{n : X_n = 0\}$. Let $Y_n = X_{\min(T,n)}$. Then $Y_n \to 0 = Y_\infty$ a.s., but $EY_n = 1\ \forall n$ but $EY_\infty = 0$.

20.3 Facts About Uniform (Equi-)Integrability

Consider $\mathbb{R}$-valued RVs.

Definition 20.4. A family (Y_α) is **UI** if

$$\lim_{b\to\infty} \sup_\alpha E\left[|Y_\alpha| 1_{(|Y_\alpha| \ge b)}\right] = 0$$

If $E|Y| < \infty$, then $\lim_{b\to\infty} E[|Y|1_{(|Y|>b)}] = 0$.

We will quote some facts (see Durrett or Billingsley).

1. If $\sup_\alpha E|Y_\alpha|^q < \infty$ for some $q > 1$, then (Y_α) is UI, which implies that $\sup_\alpha E|Y_\alpha| < \infty$.
2. if $Y_n \to Y_\infty$ a.s., if (Y_n) is UI, then $E|Y_\infty| < \infty$ and $E|Y_n - Y_\infty| \to 0$, i.e. $Y_n \to Y_\infty$ in L^1.
3. If $Y_n \to Y_\infty$ in L^1, then (Y_n) is UI.

4. If $E|Y| < \infty$, the family of $\{E[Y \mid \mathcal{G}], \text{all } \mathcal{G}\}$ is UI.

Theorem 20.5. *For a MG (X_n), the following are equivalent.*

(i) (X_n) is UI.

(ii) X_n converges in L^1.

(iii) There exists a RV X_∞ with $E|X_\infty| < \infty$ such that $X_k = E[X_\infty \mid \mathcal{F}_k]\ \forall k$.

If these conditions hold, then $\exists X_\infty$ such that $X_n \to X_\infty$ both a.s. and in L^1.

Proof. *(iii)* $\Rightarrow$ *(i)*, by 4.

(i) implies, by 1, $\sup_n E|X_n| < \infty$, which by 20.2 implies X_n converges to some X_∞ a.s., which implies by 2 that $X_n \to X_\infty$ in L^1, which implies (ii).

Given (ii), $X_n \to X_\infty$ in L^1, which implies that $E|X_n - X_\infty| \to 0$ with $E|X_\infty| < \infty$. We need to prove that $EX_\infty 1_A = EX_k 1_A\ \forall A \in \mathcal{F}_k$. Fix A and k. By the MG property, for $n > k$, $E[X_n \mid \mathcal{F}_k] = X_k$, so $EX_n 1_A = EX_k 1_A$. Hence, $|EX_\infty 1_A - EX_n 1_A| \le E|X_\infty - X_n| \to 0$ as $n \to \infty$, so $|EX_\infty 1_A - EX_k 1_A| = 0$. □

Theorem 20.6 (Levy's 0-1 Law)**.** *Take any process $(Y_n, n \ge 0)$. Take any RV Z with $E|Z| < \infty$ and $Z \in \sigma(Y_n, n \ge 0)$. Then $X_n = E[Z \mid Y_1, \dots, Y_n]$ is a UI martingale, so by 20.5, $X_n \to X_\infty$ a.s. and in L^1. In fact, $X_\infty = Z$ because*

$$\begin{aligned} X_n &= E[X_\infty \mid Y_1, \dots, Y_n] \\ &= E[Z \mid Y_1, \dots, Y_n] \end{aligned}$$

so $E[X_\infty - Z \mid \mathcal{F}_n] = 0$, so $E[X_\infty - Z \mid \mathcal{F}_\infty] = 0 = X_\infty - Z$ (since $X_\infty - Z$ is $\mathcal{F}_\infty$-measurable).

Remark: In particular, take $Z = 1_A$. Then

$$P(A \mid Y_1, \dots, Y_n)(\omega) \to 1_A(\omega) \quad \text{a.s.}$$

for all $A \in \sigma(Y_n, n \ge 0)$.

For independent (Y_n), suppose A is in the tail σ-field.

$$P(A \mid Y_1, \dots, Y_n)(\omega) = P(A) \to 1_A \quad \text{a.s.} \quad \text{as } n \to \infty$$

which implies that 1_A is a constant a.s., which implies that $P(A) = 0$ or 1.

Lecture 21

November 3

21.1 "Converge or Oscillate Infinitely"

Lemma 21.1. *Let (X_n) be a MG such that $|X_n - X_{n-1}| \leq K \ \forall n$. Then $P(C \cup D) = 1$ for the events*

$$C = \left\{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists and is finite}\right\}$$
$$D = \left\{\omega : \limsup_n X_n(\omega) = +\infty \text{ and } \liminf_n X_n(\omega) = -\infty\right\}$$

Proof. WLOG $X_0 = 0$. Fix $L > 0$. Define $T = \min\{n : X_n < -L\}$. The stopped process $(X_{T\wedge n}, n \geq 0)$ is a MG which is always at least $-L - K$. By the (positive super-MG) convergence theorem, $X_{T\wedge n}$ converges to some finite limit a.s. as $n \to \infty$. This implies $\{\inf_n X_n > -L\} = \{T = \infty\} \subseteq C$. This is true for all L, so let $L \to \infty$. Therefore,

$$A_1 = \left\{\inf_n X_n > -\infty\right\} \subseteq C$$

The same argument applied to $(-X_n)$ gives

$$A_2 = \left\{\sup_n X_n < \infty\right\} \subseteq C$$

so we are done because $(A_1 \cap A_2)^c = D$. □

21.2 Conditional Borel-Cantelli

Lemma 21.2 (Conditional Borel-Cantelli Lemma)**.** *Consider events (A_n) adapted to $(\mathcal{F}_n)$. Define $B_n = \bigcup_{m\geq n} A_m$ and $B = \bigcap_n B_n = \{A_n \text{ inf. often}\}$. Then*

(a) $\{A_n \text{ inf. often}\} = \{\sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1}) = \infty\}$ *a.s.*

(b) $P(B_{n+1} \mid \mathcal{F}_n) \to 1_B$ *a.s. as $n \to \infty$.*

$B_1 = B_2$ a.s. means $P(B_1 \triangle B_2) = 0$.

Proof. (b) Consider $K < n$. Then $B \subseteq B_n \subseteq B_K$ and

$$P(B \mid \mathcal{F}_n) \leq P(B_{n+1} \mid \mathcal{F}_n) = P(B_{n+1} \mid \mathcal{F}_n) \leq P(B_K \mid \mathcal{F}_n)$$

Take the limit as $n \to \infty$.

$$1_B \leq \liminf_n P(B_{n+1} \mid \mathcal{F}_n) \leq \limsup_n P(B_{n+1} \mid \mathcal{F}_n) \leq 1_{B_K}$$

Let $K \uparrow \infty$. Then $1_{B_K} \downarrow 1_B$.

(a) Consider $X_n = \sum_{m=1}^n (1_{A_m} - P(A_m \mid \mathcal{F}_{m-1}))$, which is a MG, and $|X_{n+1} - X_n| \leq 1$. Then 21.1 implies that $P(C \cup D) = 1$. We want to prove

$$\left\{\sum_m 1_{A_m} = \infty\right\} = \left\{\sum_m P(A_m \mid \mathcal{F}_{m-1}) = \infty\right\} \quad \text{a.s.}$$

Observe that $X_n = \sum_{m=1}^n 1_{A_m} - \sum_{m=1}^n P(A_m \mid \mathcal{F}_{m-1})$. On event D, both sums are ∞. On event C, either both sums are finite or both sums equal ∞.

□

21.3 "Product" MGs

21.3.1 Convergence for "Multiplicative" MGs

Theorem 21.3 (Kakutani's Theorem)**.** *Take $(X_i, i \geq 1)$ to be independent, $X_i > 0$, $EX_i = 1$. We know that $M_n = \prod_{i=1}^n X_i$ is a MG and so $M_n \xrightarrow{a.s.} M_\infty$, with $EM_\infty \leq 1$. Then properties (i) to (v) below are equivalent:*

(i) $EM_\infty = 1$.

(ii) $M_n \to M_\infty$ *in* L^1.

(iii) $(M_n, n \geq i)$ *is UI.*

(iv) Set $a_i = EX_i^{1/2}$ *and note that* $0 \leq a_i \leq 1$. $\prod_{i=1}^\infty a_i > 0$.

(v) $\sum_i (1 - a_i) < \infty$.

Proof. Conditions (i), (ii), (iii) are equivalent by the L^1 MG convergence theorem.

Conditions (iv), (v) are equivalent by calculus. Use $1 - x + x^2 \geq e^{-x} \geq 1 - x$ for small $x > 0$.

Suppose (iv) holds. Consider

$$N_n = \frac{X_1^{1/2}}{a_1} \cdot \frac{X_2^{1/2}}{a_2} \cdots \frac{X_n^{1/2}}{a_n}$$

which is a MG.

$$E[N_n^2] = \frac{EM_n}{\prod_{i=1}^n a_i^2} \leq \frac{1}{\prod_{i=1}^\infty a_i^2} = K < \infty$$

Apply the Doob L^2 maximal inequality.

$$E\left[\sup_n N_n\right] \le 4K$$

Note that $M_n \le N_n^2$ since $M_n = N_n^2 \prod_{i=1}^n a_i^2$. Therefore, $E[\sup_n M_n] \le (4K)^2 < \infty$. This implies that $(M_n, n \ge 1)$ is UI. If $Z \ge 0$, $EZ < \infty$, the family $\{X : 0 \le X \le Z\}$ is UI. This gives (iii).

Suppose that (iv) is false, so $\prod_{i=1}^\infty a_i = 0$. For the MG (N_n), we have $N_n \to N_\infty$ a.s. We must have

$$N_\infty = \frac{\prod_{i=1}^\infty X_i^{1/2}}{\prod_{i=1}^\infty a_i}$$

Since the denominator is 0, then $\prod_{i=1}^\infty X_i^{1/2} = M_\infty^{1/2} = 0$ a.s., so (i) fails. □

21.3.2 Likelihood Ratios (Absolute Continuity of Infinite Product Measures)

Given densities $f_i, 1 \le i < \infty$ and $g_i, 1 \le i < \infty$, assume $f_i > 0$ and $g_i > 0$. Take $\Omega = \mathbb{R}^\infty$ with $X(\boldsymbol{\omega}) = \omega_i$. Work with P, the product measure where the (X_i) are independent with densities f_i. Consider Q, where the (X_i) have densities g_i. The "likelihood ratio"

$$L_n = \prod_{i=1}^n \frac{g_i(X_i)}{f_i(X_i)}$$

is the Radon-Nikodym density

$$\frac{\mathrm{d}Q_n}{\mathrm{d}P_n}$$

(Q_n is the probability measure with corresponding density $f_1 \otimes f_2 \otimes \cdots \otimes f_n$.)

Know. $(L_n, n \ge 1)$ is a MG w.r.t. P.

Suppose that $(L_n, n \ge 1)$ is UI. Then $L_n \to L_\infty$ in L^1 and $L_n = E[L_\infty \mid \mathcal{F}_n]$. What this means, from the definition of the R-N density, is

$$\begin{aligned} Q(A) &= EL_n 1_A \quad \forall A \in \mathcal{F}_n \\ &= EL_\infty 1_A \quad \forall A \in \bigcup_n \mathcal{F}_n \\ &= EL_\infty 1_A \quad \forall A \in \mathcal{F}_\infty \end{aligned}$$

so L_∞ is the R-N density

$$\frac{\mathrm{d}Q}{\mathrm{d}P}$$

on $\mathbb{R}^\infty$. Therefore, $Q \ll P$.

Similarly, if $Q \ll P$, then we can prove $(L_n, n \ge 1)$ is UI. So $Q \ll P \Leftrightarrow (L_n, n \ge 1)$ is UI $\Leftrightarrow \sum_i (1 - a_i) < \infty$.

$$\begin{aligned} a_i &= E\left(\frac{g_i}{f_i}(X_i)\right)^{1/2} = \int g_i^{1/2}(x) f_i^{1/2}(x)\,\mathrm{d}x \\ 1 - a_i &= \frac{1}{2}\int \left(g_i^{1/2}(x) - f_i^{1/2}(x)\right)^2 \mathrm{d}x \end{aligned}$$

(by algebra). Our condition is

$$\sum_{i=1}^\infty \int \left(g_i^{1/2}(x) - f_i^{1/2}(x)\right)^2 \mathrm{d}x < \infty$$

"f_i and g_i become close for large i."

We know that for $f \not\equiv g$, then Q and P are singular.

Lecture 22

November 8

22.1 Setup for OST

Let $(X_n, n \geq 0)$ be a sub-MG and $T < \infty$ a.s. be a stopping time. We want to conclude that $EX_0 \leq EX_T$. What *extra* assumptions do we need?

Know. It is sufficient that $T \leq t_0 < \infty$ a.s., so it is sufficient that $E|X_T - X_{T\wedge n}| \to 0$ as $n \to \infty$.

Theorem 22.1. *(See Durrett.) It is sufficient that*

(a) $E|X_n|1_{(T>n)} \to 0$ *as* $n \to \infty$, *and*

(b) $E|X_T| < \infty$.

Theorem 22.2 (Useful Version of OST)**.** *Suppose:* (X_n) *is a sub-MG,* T *is a stopping time, and* $ET < \infty$. *Write* $\Delta_n = X_n - X_{n-1}$. *If there exists a constant* b *such that*

$$E[|\Delta_n| \mid \mathcal{F}_{n-1}] \leq b \quad \text{on } \{n \leq T\} \tag{22.1}$$

then $EX_0 \leq EX_T$.

Proof.

$$X_T = X_0 + \sum_{n=1}^{T} \Delta_n$$

Consider $Y = |X_0| + \sum_{n=1}^{T} |\Delta_n|$. Note that $|X_T| \leq Y$ and $|X_{T\wedge n}| \leq Y$. Then

$$EY = E|X_0| + \sum_{n=1}^{\infty} E|\Delta_n| 1_{(T \geq n)}$$

We have

$$\begin{aligned} E[|\Delta_n| 1_{(T\geq n)} \mid \mathcal{F}_{n-1}] &= 1_{(T\geq n)} E[|\Delta_n| \mid \mathcal{F}_{n-1}] \\ &\leq b 1_{(T\geq n)} \quad \text{by (22.1)} \end{aligned}$$

Take expectations of both sides.

$$E[|\Delta_n| 1_{(T\geq n)}] \leq bP(T \geq n)$$

Therefore,

$$\begin{aligned} EY &\le E|X_0| + \sum_{n=1}^{\infty} bP(T \ge n) \\ &= E|X_0| + bET < \infty \end{aligned}$$

so $E|X_T| \le EY < \infty$, which checks (b). For condition (a),

$$\begin{aligned} E|X_n|1_{(T>n)} &= E|X_{T\wedge n}|1_{(T>n)} \\ &\le EY1_{(T>n)} \to 0 \end{aligned}$$

as $n \to \infty$, since $EY < \infty$. (We are using the fact that $E|W| < \infty$ and $P(A_n) \to 0$ imply $E[W1_{A_n}] \to 0$.) □

22.2 Martingale Proofs

Principle. Given a MG proof of an exact formula, one can often get inequality conclusions out of inequality assumptions.

Corollary 22.3 (Inequality Version of Wald's Identity)**.** *Suppose (ξ_i) are independent, $\mu_i \le E\xi_i \le \mu_2$, and $\sup_i E|\xi_i| < \infty$. Let $S_n = \sum_{i=1}^{n} \xi_i$. Then, for any stopping time T with $ET < \infty$,*

$$\mu_1 ET \le ES_T \le \mu_2 ET$$

Wald: If the (ξ_i) are IID, then $ES_T = (E\xi) \cdot (ET)$.

Proof. Apply 22.2 to $X_n = S_n - n\mu_1$, so that $\Delta_n = \xi_n - \mu_1$. $E[\Delta_n \mid \mathcal{F}_{n-1}] = E\xi_n - \mu_1 \ge 0$, so (X_n) is a sub-MG. We have

$$E[|\Delta_n| \mid \mathcal{F}_{n-1}] = E|\Delta_n| \le E|\xi_n| + |\mu_1| \le b$$

by hypothesis. Therefore, $EX_0 \le EX_T$, so $0 \le ES_T - \mu_1 ET$, so $ES_T \ge \mu_1 ET$. □

Lemma 22.4. *Take (ξ_i) IID, $S_n = \sum_{i=1}^{n} \xi_i$. Fix $a > 0$ and $b > E\xi$. Suppose $\exists\theta > 0$ such that $E\exp(\theta\xi) = e^{\theta b}$. Then $P(S_n \ge a + bn$ for some $n \ge 0) \le e^{-\theta a}$.*

Proof. Set $\hat{\xi}_i = \xi_i - b$. Then $\hat{S}_n = S_n - nb$ and $E\exp(\theta\hat{\xi}) = 1$ by definition. Then $(\exp(\theta\hat{S}_n), n \ge 0)$ is a MG. Apply the L^1 maximal inequality, so

$$P\left(\sup_n \exp(\theta\hat{S}_n) \ge \lambda\right) \le \frac{1}{\lambda}$$

Set $\lambda = e^{\theta a}$. Then

$$P\left(\sup_n \hat{S}_n \ge a\right) \le e^{-\theta a}$$

which implies the result. □

Lemma 22.5. *Suppose (ξ_i) are IID and let $S_n = \sum_{i=1}^{n} \xi_i$. Suppose $\exists\theta > 0$ such that*

$$\phi(\theta) = E\exp(\theta\xi) = 1$$

Suppose T is a stopping time with $ET < \infty$ and $S_n \le B$ on $\{n < T\}$ for all n. Then $E\exp(\theta S_T) = 1$.

Proof. $X_n \stackrel{\text{def}}{=} \exp(\theta S_n)$ is a MG. We need to check (22.1) from 22.2.

$$\Delta_n = X_n - X_{n-1} = X_{n-1}(\exp(\theta\xi_n) - 1)$$
$$|\Delta_n| \le X_{n-1}|\exp(\theta\xi_n) - 1|$$
$$E[|\Delta_n| \mid \mathcal{F}_{n-1}] \le X_{n-1}E|\exp(\theta\xi) - 1| \le 2X_{n-1}$$

On $\{n \le T\} = \{n-1 < T\}$, we have $S_{n-1} \le B$, so $X_{n-1} \le e^{\theta B}$. Therefore, $2X_{n-1} \le 2e^{\theta B}$ on $\{n \le T\}$. This verifies (22.1). □

22.3 Boundary Crossing Inequalities

Setting. Let (ξ_i) be IID with $S_n = \sum_{i=1}^n \xi_i$. Suppose that $|\xi_i| \le L$ and assume $E\xi < 0$, with $P(\xi > 0) > 0$. Fix $a < 0 < b$, and consider $T = \min\{n : S_n \ge b \text{ or } S_n \le a\}$.

Exercise. $ET < \infty$.

So, $P(S_T \ge b \text{ and } S_T \le b + L) = x$, say, and $P(S_T \le a \text{ and } S_T \ge a - L) = 1 - x$.

Consider $\phi(\theta) = E\exp(\theta\xi) < \infty$. We know that $\phi(0) = 1$, $\phi'(0) = E\xi < 0$, and $\phi(\theta) \to \infty$ as $\theta \to \infty$. Therefore, $\exists\theta > 0$ such that $\phi(\theta) = 1$.

Apply 22.5 to conclude that $E\exp(\theta S_T) = 1$.

$$xe^{\theta b} + (1-x)e^{\theta(a-L)} \le 1 \le xe^{\theta(b+L)} + (1-x)e^{\theta a} \tag{22.2}$$

With some algebra,

$$\frac{1 - e^{\theta a}}{e^{b+L} - e^{\theta a}} \le x \le \frac{1 - e^{\theta(a-L)}}{e^{\theta b} - e^{\theta(a-L)}}$$

Special Case. If $P(\xi = 1) = p < 1/2$ and $P(\xi = -1) = q = 1 - p$ and $a < 0 < b$ are integers, then (22.2) is an equality, so

$$x = \frac{1 - e^{\theta a}}{e^{\theta b} - e^{\theta a}} = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}$$

Write $\phi(\theta) = pe^{\theta} + qe^{-\theta} = 1$ and solve, so $e^{\theta} = q/p$. This yields the result that we see in an undergraduate course.

Lecture 23

November 10

23.1 Patterns in Coin-Tossing

We will say this in words. (*Exercise*: Rewrite the argument with mathematical notation for a general pattern.)

Fix the pattern $HTTHT$. Toss a fair coin until we see this pattern: this requires W tosses. W is random and $5 \leq W < \infty$ a.s. What is EW?

Strategy 7: Bet \$1 that Toss 7 is H. If we win the bet, bet \$2 that Toss 8 is T. If we win again, bet \$4 that Toss 9 is T. If we win again, bet \$8 that Toss 10 is H. If we win again, bet \$16 that Toss 11 is T.

Overall Strategy: Do strategy i for each $1 \leq i \leq W$ and then stop after toss W.

The OST tells us that $E[\text{profit}] = 0$. The cost is W and our return is $32 + 4 = 36$ (because HT is the start of the pattern). Therefore, $E[36 - W] = 0$ so $EW = 36$.

For a pattern of $HHHHH$, we would have $EW = 32 + 16 + 8 + 4 + 2 = 62$.

We can also show the existence of "non-transitive dice": 3 patterns such that no matter what pattern you choose, I can choose a pattern such that the odds will be favorable that my pattern comes up before yours.

23.2 MG Proof of Radon-Nikodym

Theorem 23.1. *Consider $(S, \mathcal{S}, \mu)$, a probability space, where $\mathcal{S} = \sigma(A_1, A_2, A_3, \ldots)$ (generated by countable events). If $\nu \ll \mu$, $\nu(S) < \infty$, then there exists a measurable $h : S \to [0, \infty)$ such that $\nu(A) = \int_A h \, \mathrm{d}\mu$, for all $A \in \mathcal{S}$.*

Proof. Heuristics:

$$h(s) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(s) = \lim_{A \downarrow \{s\}} \frac{\nu(A)}{\mu(A)}$$

Define $\mathcal{F}_n = \sigma(A_1, A_2, \ldots, A_n)$, a finite field with 2^n atoms. Define

$$X_n(s) = \begin{cases} \dfrac{\nu(F)}{\mu(F)}, & \text{for the atom } s \in F \\ 0, & \text{if } \mu(F) = 0 \end{cases}$$

(Recall that $\nu \ll \mu$ means "$\mu(A) = 0 \Rightarrow \nu(A) = 0$".)

$$E_\mu X_n 1_F = \frac{\nu(F)}{\mu(F)} \times \mu(F) \quad \text{for atom } F$$

so

$$E_\mu X_n 1_F = \nu(F) \quad \text{for each } F \in \mathcal{F}_n \tag{23.1}$$

Claim: $(X_n, \mathcal{F}_n)$ is a MG.

Why: Take $G \in \mathcal{F}_{n-1}$. Then

$$G = \underbrace{(G \cap A_n)}_{G_1} \cup \underbrace{(G \cap A_n^c)}_{G_2}$$

$$EX_n 1_G = EX_n 1_{G_1} + EX_n 1_{G_2} = \nu(G_1) + \nu(G_2) = \nu(G) = EX_{n-1} 1_G$$

by 23.1 for all $G \in \mathcal{F}_{n-1}$, so $X_{n-1} = E[X_n \mid \mathcal{F}_{n-1}]$. By the MG convergence theorem, $X_n \to X_\infty \geq 0$ a.s. If we prove $(X_n, n \geq 1)$ is UI, then by the theorem we have proven, $X_n = E[X_\infty \mid \mathcal{F}_n]$.

For $F \in \mathcal{F}_n$,

$$EX_\infty 1_F = EX_n 1_F = \nu(F)$$

which implies

$$EX_\infty 1_F = \nu(F) \quad \forall F \in \bigcup_n F_n$$

which implies that this holds $\forall F \in \sigma(\bigcup_n \mathcal{F}_n) = \mathcal{S}$. Then

$$\nu(F) = E_\mu X_\infty 1_F = \int_F X_\infty \, \mathrm{d}\mu$$

which shows that X_∞ is the R-N density $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Proof that (X_n) *is UI.* We know that

$$EX_n 1_{(X_n \geq b)} = \nu(X_n \geq b)$$

by (23.1). Given $\varepsilon > 0$, take b such that $\nu(S)/b \leq \delta(\varepsilon)$. Then

$$\mu(X_n \geq b) \leq \frac{EX_n}{b} = \frac{\nu(S)}{b} \leq \delta(\varepsilon)$$

by Markov's inequality. Then 23.2 implies that $\nu(X_n \geq b) \leq \varepsilon$, so $\sup_n EX_n 1_{(X_n \geq b)} \leq \varepsilon$, which implies UI. □

Lemma 23.2. *Suppose $\nu \ll \mu$. $\forall \varepsilon > 0$ $\exists \delta(\varepsilon) > 0$ such that $\mu(A) \leq \delta(\varepsilon)$ implies $\nu(A) \leq \varepsilon$.*

Proof. If the statement is false for ε, $\exists A_n$ $\mu(A_n) \leq 2^{-n}, \nu(A_n) > \varepsilon$. Consider $\Lambda = \{A_n \text{ inf. often}\}$. Then $\mu(\Lambda) = 0$, $\nu(\Lambda) \geq \varepsilon$, which contradicts the definition of $\nu \ll \mu$. □

23.3 Azuma's Inequality

Theorem 23.3 (Azuma's Inequality)**.** *Let* $S_n = \sum_{i=1}^n X_i$ *be a MG with* $|X_i| \le 1$ *a.s. Then we have* $P(S_n \ge \lambda\sqrt{n}) \le e^{-\lambda^2/2}$ *for all* $\lambda > 0$*, so* $P(|S_n| \ge \lambda\sqrt{n}) \le 2e^{-\lambda^2/2}$ *for* $\lambda > 0$*.*

Lemma 23.4. *If* $EY = 0$ *and* $|Y| \le 1$*, then* $Ee^{\alpha Y} \le e^{\alpha^2/2}$ $\forall\alpha$*.*

Proof. $e^{\alpha x}$ is convex, so draw the straight line L between $e^{-\alpha}$ and e^{α}. By convexity,

$$\begin{aligned} Ee^{\alpha Y} \le EL(Y) = L(EY) = L(0) &= \frac{e^{\alpha} + e^{-\alpha}}{2} \\ &\le e^{\alpha^2/2} \end{aligned}$$

by calculus. Look at the coefficient of α^{2n} in the series expansion.

$$\frac{1}{(2n)!} \le \frac{1}{2^n n!}$$ □

Proof of Azuma's Inequality. Apply 23.4 to the conditional distribution of X_i given $\mathcal{F}_{i-1}$. Then we obtain $E[e^{\alpha X_i} \mid \mathcal{F}_{i-1}] \le e^{\alpha^2/2}$.

$$E[e^{\alpha S_n} \mid \mathcal{F}_{n-1}] = e^{\alpha S_{n-1}} E[e^{\alpha X_n} \mid \mathcal{F}_{n-1}] \le e^{\alpha^2/2} e^{\alpha S_{n-1}}$$

so

$$\begin{aligned} Ee^{\alpha S_n} &\le e^{\alpha^2/2} Ee^{\alpha S_{n-1}} \\ Ee^{\alpha S_n} &\le (e^{\alpha^2/2})^n = \exp\left(\frac{\alpha^2 n}{2}\right) \end{aligned}$$

Then, by the large deviation inequality,

$$P(S_n \ge \lambda\sqrt{n}) \le \frac{Ee^{\alpha S_n}}{e^{\alpha\lambda\sqrt{n}}} \le \exp\left(\frac{\alpha^2 n}{2} - \alpha\lambda\sqrt{n}\right) = \exp\left(-\frac{\lambda^2}{2}\right)$$

We minimize over α, so take $\alpha = \lambda/\sqrt{n}$. □

23.4 Method of Bounded Differences

Corollary 23.5. *Take* $(\xi_1, \xi_2, \ldots, \xi_n)$ *to be independent in arbitrary state spaces. Take a* $\mathbb{R}$*-valued* $Z = f(\xi_1, \xi_2, \ldots, \xi_n)$ *such that* f *has the property: if* $\mathbf{x} = (x_1, \ldots, x_n)$ *and* $\mathbf{y} = (y_1, \ldots, y_n)$ *are such that* $|\{i : y_i \ne x_i\}| = 1$*, then* $|f(\mathbf{x}) - f(\mathbf{y})| \le 1$*. Then* $P(|Z - EZ| \ge \lambda\sqrt{n}) \le 2e^{-\lambda^2/2}$*,* $\lambda > 0$*.*

Proof. WLOG, take $EZ = 0$. Write $S_m = E[Z \mid \mathcal{F}_m]$, with $\mathcal{F}_m = \sigma(\xi_1, \xi_2, \ldots, \xi_m)$, so $(S_m, 1 \le m \le n)$ is a MG. We need to prove $|S_m - S_{m-1}| \le 1$ and then apply Azuma's inequality 23.3.

Fix m. If we know all $(\xi_i, i \ne m)$, then apply 23.6 conditionally.

$$|Z - \underbrace{E[Z \mid \xi_i, i \ne m]}_{Z^*}| \le 1 \tag{23.2}$$

and

$$E[Z^* \mid \mathcal{F}_m] = E[Z^* \mid \mathcal{F}_{m-1}, \xi_m] = E[Z^* \mid \mathcal{F}_{m-1}] = E[Z \mid \mathcal{F}_{m-1}]$$

since Z^* and $\mathcal{F}_{m-1}$ are in $\sigma(\xi_i, i \neq m)$, ξ_m is independent of the two, and 23.7. Then, we applied the tower property. This implies

$$\begin{aligned} |S_m - S_{m-1}| &= |E[Z \mid \mathcal{F}_m] - E[Z^* \mid \mathcal{F}_m]| \\ &\leq E[|Z - Z^*| \mid \mathcal{F}_m] \\ &\leq 1 \end{aligned}$$

by (23.2). □

Lemma 23.6 (Obvious Lemma)**.** *If Y is such that any 2 possible values are within 1 of each other, then $|Y - EY| \leq 1$.*

Lemma 23.7 (Obvious Lemma)**.** *If W is independent of $(Y, \mathcal{G})$, then $E[Y \mid \mathcal{G}, W] = E[Y \mid \mathcal{F}]$.*

Lecture 24

November 15

24.1 Examples Using "Method of Bounded Differences"

Last class: **Theorem**. Suppose $\xi_1, \xi_2, \ldots, \xi_n$ are independent, $Z = f(\xi_1, \ldots, \xi_n)$, where f has the property

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq 1 \tag{24.1}$$

whenever $|\{i : x_i \neq y_i\}| = 1$. Then $P(|Z - EZ| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/2}$ for $\lambda > 0$.

Example 24.1. Put n balls "at random" into m boxes. Consider $Z(n,m)$, the number of empty boxes. $EZ(n,m) = m(1-1/m)^n$. There is a complicated formula for the distribution. However, we can apply the theorem to ξ_i, the box containing ball i, for $1 \leq i \leq n$. Then (24.1) holds.

Example 24.2. Take two independent Bernoulli(1/2) sequences of length n (e.g. 10100110 and 01101000). Let Z_n be the length of the longest common subsequence.

Fact. $Z_n/n \xrightarrow{\text{a.s.}} c$ as $n \to \infty$, but there is no formula for c.

Take ξ_i to be the pair of digits in the two strings at position i. Any change $\mathbf{x} \mapsto \mathbf{y}$ has $f(\mathbf{y}) - f(\mathbf{x}) \geq -2$, which also implies that $f(\mathbf{y}') - f(\mathbf{x}') \leq 2$ for any $\mathbf{x}', \mathbf{y}'$. Therefore, $Z_n/2$ satisfies (24.1).

Recall: A c**-coloring** of G means assigning one of c colors to each vertex such that $\text{color}(v) \neq \text{color}(v')$ whenever (v, v') is an edge. The **chromatic number** is $\chi(G) = \min\{c : \exists c\text{-coloring}\}$.

Recall: An Erdős–Renyi random graph model $\mathcal{G}(n,p)$ has n vertices and each of the $\binom{n}{2}$ possible edges is present with probability p.

Let $Z = \chi(\mathcal{G}(n,p))$. Order the vertices as $1, 2, 3, \ldots, n$. For $i \geq 2$, let

$$\xi_i = (1_{(i,1)\text{ is an edge}}, \ldots, 1_{(i,i-1)\text{ is an edge}})$$

Then (24.1) holds for $Z = f(\xi_2, \xi_3, \ldots, \xi_n)$.

(To check (24.1), we are using the trick $\sup_{i \neq j} |x_i - x_j| = \sup_{i \neq j} (x_i - x_j)$.)

Example 24.3. Put n points IID uniform in the unit square. Fix $0 < c < 1$. Let $Z(n,c)$ be the maximum number of disjoint $c \times c$ squares containing 0 points. Let ξ_i be the position of the ith point. (24.1) holds.

24.2 Reversed MGs

Consider sub-σ-fields $\mathcal{G}_0 \supseteq \mathcal{G}_1 \supseteq \mathcal{G}_2 \supseteq \cdots$, where $\mathcal{G}_\infty = \bigcap_n \mathcal{G}_n$. We say that (X_n) is a **reversed MG** if: $E|X_n| < \infty$, $E[X_m \mid \mathcal{G}_n] = X_n$, for $m \leq n$, and (X_n) is adapted to $(\mathcal{G}_n)$. (In Durrett, $\mathcal{G}_n = \mathcal{F}_{-n}$.) The definition implies that $X_n = E[X_0 \mid \mathcal{G}_n]$.

Theorem 24.4. *For a reversed MG, $X_n \to E[X_0 \mid \mathcal{G}_\infty]$ a.s. and in L^1.*

Proof. $(X_N, X_{N-1}, \ldots, X_0)$ is a MG. If U_N is the number of upcrossings of the martingale over $[a, b]$, the upcrossing inequality says

$$EU_N \leq \frac{E|X_0| + |a|}{b - a}$$

(As in the proof for MGs:) $U_N \uparrow U_\infty$, where

$$EU_\infty \leq \frac{E|X_0| + a}{b - a}$$

which implies that $U_\infty < \infty$ a.s., which implies that $X_n \to X_\infty \in [-\infty, \infty]$ a.s. However, we have $X_n = E[X_0 \mid \mathcal{G}_n]$, so (X_n) is UI, so $X_n \to X_\infty$ in L^1 (also), with $E|X_\infty| < \infty$.

We need to show $X_\infty = E[X_0 \mid \mathcal{G}_\infty]$. $X_n \in \mathcal{G}_n \subseteq \mathcal{G}_K$ for $n > K$. Take $n \to \infty$, so $X_\infty \in \mathcal{G}_K$. Take $K \to \infty$, so $X_\infty \in \mathcal{G}_\infty$. We need to show $EX_\infty 1_G = EX_0 1_G$ for $G \in \mathcal{G}_\infty$. $X_n = E[X_0 \mid \mathcal{G}_n]$ implies that $EX_n 1_G = EX_0 1_G$ for $G \in \mathcal{G}_\infty$. $X_n \to X_\infty$ in L^1 implies that $EX_n 1_G \to EX_\infty 1_G$, so $EX_0 1_G = EX_\infty 1_G$. □

24.3 Exchangeable Sequences

A sequence of RVs $(X_1, X_2, X_3, \ldots)$ is called **exchangeable** if

$$(X_1, X_2, \ldots, X_n) \stackrel{\mathrm{d}}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$$

for all n and all permutations π of $(1, 2, \ldots, n)$.

Clearly, IID implies exchangeable.

Theorem 24.5. *Suppose $(X_i, 1 \leq i < \infty)$ are exchangeable and $\mathbb{R}$-valued and $E|X_1| < \infty$. Write $S_n = \sum_{i=1}^n X_i$. Then $S_n/n \to E[X_1 \mid \tau]$ a.s. and in L^1, where $\tau = \mathrm{tail}(X_i, i \geq 1)$.*

Corollary 24.6. *If (X_i) are IID, $E|X_1| < \infty$, then τ is trivial, which implies that $E[X_1 \mid \tau] = EX_1$ and 24.5 implies that $S_n/n \to EX_1$.*

Fact. If $(Z_1, W) \stackrel{\mathrm{d}}{=} (Z_2, W)$ and $E|Z_1| < \infty$, then $E[Z_1 \mid W] = E[Z_2 \mid W]$ a.s.

Proof. Let Q be the kernel associated with the distribution (Z_1, W). $E[Z_1 \mid W] = \phi(W)$, where the function $\phi(w) = \int z Q(\omega, \mathrm{d}z)$, and $E[Z_2 \mid W] = \phi(W)$. □

Exercise. Let $E|X| < \infty$. If $X \stackrel{\mathrm{d}}{=} E[X \mid \mathcal{G}]$, then $X = E[X \mid \mathcal{G}]$ a.s.

Comment. The proof is easy if $EX^2 < \infty$.

Proof of 24.5. Define

$$\begin{aligned}\mathcal{G}_n &= \sigma(S_n, X_{n+1}, X_{n+2}, \dots) \\ &= \sigma(S_n, S_{n+1}, S_{n+2}, \dots)\end{aligned}$$

$\mathcal{G}_n \supseteq \mathcal{G}_{n-1} \supseteq \cdots$ are decreasing. Then

$$S_n = E[S_n \mid \mathcal{G}_n] = \sum_{i=1}^{n} E[X_i \mid \mathcal{G}_n] = nE[X_1 \mid \mathcal{G}_n]$$

by 24.7. Therefore, $S_n/n = E[X_1 \mid \mathcal{G}_n] \to E[X_1 \mid \mathcal{G}_\infty]$ a.s. and in L^1. Note that $\mathcal{G}_\infty \supseteq \tau$. However, $\lim S_n/n$ is τ-measurable. Therefore, $E[X_1 \mid \mathcal{G}_\infty]$ is τ-measurable, so

$$E[X_1 \mid \tau] = E[E[X_1 \mid \mathcal{G}_\infty] \mid \tau] = E[X_1 \mid \mathcal{G}_\infty]$$ □

Lemma 24.7. *$E[X_i \mid \mathcal{G}_n] = E[X_1 \mid \mathcal{G}_n]$ a.s., $1 \le i \le n$.*

Proof. Take a permutation π of $(1, \dots, n)$.

$$(X_{\pi(1)}, \dots, X_{\pi(n)}, X_{n+1}, X_{n+2}, \dots) \stackrel{\mathrm{d}}{=} (X_1, \dots, X_n, X_{n+1}, X_{n+2}, \dots)$$

Set $W = (S_n, X_{n+1}, X_{n+2}, \dots)$. Then $(X_{\pi(i)}, \dots, X_{\pi(n)}, W) \stackrel{\mathrm{d}}{=} (X_1, \dots, X_n, W)$, which implies that $(X_{\pi(i)}, W) \stackrel{\mathrm{d}}{=} (X_1, W)$, which implies that $(X_i, W) \stackrel{\mathrm{d}}{=} (X_1, W)$ for $1 \le i \le n$. By the Fact proven above, $E[X_i \mid W] = E[X_1 \mid W]$, and $\mathcal{G}_n = \sigma(W)$. □

Lecture 25

November 17

25.1 "Play Red"

Consider a finite set S and let $X_1, X_2, \ldots, X_N$ be a uniform random ordering. This is clearly a (finite) exchangeable sequence.

Proposition 25.1. *If $(X_1, \ldots, X_N)$ is an exchangeable sequence, if $0 \le T \le N-1$ is a stopping time, then $X_{T+1} \stackrel{\text{d}}{=} X_1$.*

Proof. Recall from last lecture:

Lemma: If $(Z_1, W) \stackrel{\text{d}}{=} (Z_2, W)$, then $E[\phi(Z_1) \mid W] = E[\phi(Z_2) \mid W]$ a.s.

$(X_{n+1}, X_1, \ldots, X_n) \stackrel{\text{d}}{=} (X_N, X_1, \ldots, X_n)$. By the Lemma, $P(X_{n+1} \in A \mid \mathcal{F}_n) = P(X_N \in A \mid \mathcal{F}_n)$ a.s., which implies that $P(X_{n+1} \in A \mid \mathcal{F}_T) = P(X_N \in A \mid \mathcal{F}_T)$ a.s. on $\{T = n\}$, for all n, so they equal each other everywhere. Now, take expectations:

$$P(X_{T+1} \in A) = P(X_N \in A)$$

$$X_{T+1} \stackrel{\text{d}}{=} X_N \stackrel{\text{d}}{=} X_1$$

□

25.2 de Finetti's Theorem

Given random A and $B > 0$, form the following construction: given $A = a$ and $B = b$, let $(X_i, 1 \le i < \infty)$ be IID Normal(a, b). This is a **parametric Bayes** formulation.

Let $\mathcal{P}(R)$ be the space of all PMs on $\mathbb{R}$. M is a random variable with values in $\mathcal{P}(R)$. Construction: given $M = \mu$, let $(X_i, i \ge 1)$ be IID(μ). This gives an infinite exchangeable sequence.

Theorem 25.2 (de Finetti's Theorem)**.** *Let $(X_i, 1 \le i < \infty)$ be exchangeable and $\mathbb{R}$-valued. Let τ be the tail σ-field. Then, conditionally on τ, the (X_i) are IID. That is,*

(a) $X_1, X_2, \ldots$ are CI given τ.

(b) There exists a kernel $Q(\omega, \cdot)$ (a random PM) such that $Q(\omega, \cdot)$ is the regular conditional distribution of X_i given τ, for each i.

$$P(X_i \in A \mid \tau)(\omega) = Q(\omega, A) \qquad \forall i$$

Proof (Sophisticated). Let $\phi : \mathbb{R} \to \mathbb{R}$ be bounded and measurable. Exchangeable implies that

$$(X_1, \dots, X_n) \stackrel{\text{d}}{=} (X_1, X_k, X_{k+1}, \dots, X_{n+k-1})$$

Let $n \to \infty$. Then, $(X_1, X_2, \dots) \stackrel{\text{d}}{=} (X_1, X_k, X_{k+1}, \dots)$. Therefore,

$$E[\phi(X_1) \mid X_2, X_3, \dots] \stackrel{\text{d}}{=} E[\phi(X_1) \mid X_k, X_{k+1}, \dots]$$

$\sigma(X_k, X_{k+1}, \dots) \downarrow \tau$ as $k \to \infty$. Apply reversed MG convergence, so the RHS converges to $E[\phi(X_1) \mid \tau]$ a.s. We conclude that $E[\phi(X_1) \mid X_2, X_3, \dots] \stackrel{\text{d}}{=} E[\phi(X_1) \mid \tau]$.

Fact: If $E[Z \mid \mathcal{G}] \stackrel{\text{d}}{=} Z$, then $E[Z \mid \mathcal{G}] = Z$ a.s. If $\mathcal{G} \subseteq \mathcal{H}$, if $E[Z \mid \mathcal{G}] \stackrel{\text{d}}{=} E[Z \mid \mathcal{H}]$, then $E[Z \mid \mathcal{G}] = E[Z \mid \mathcal{H}]$ a.s.

By the exercise, $E[\phi(X_1) \mid X_2, X_3, \dots] = E[\phi(X_1) \mid \tau]$ a.s. By the same argument: $\forall k \geq 1$,

$$E[\phi(X_k) \mid X_{k+1}, X_{k+2}, \dots] = E[\phi(X_k) \mid \tau] \qquad \text{a.s.}$$

U and V are CI given τ if and only if $E[\phi(U)|V, \tau] = E[\phi(U)|\tau]$ a.s. Therefore, X_k and $(X_{k+1}, X_{k+2}, \dots)$ are CI given τ. This is enough to show that $(X_1, X_2, X_3, \dots)$ are CI given τ.

Exchangeable implies that $(X_1, X_{i+1}, X_{i+2}, \dots) \stackrel{\text{d}}{=} (X_i, X_{i+1}, X_{i+2}, \dots)$. By the Lemma,

$$E[\phi(X_1) \mid X_{i+1}, X_{i+2}, \dots] = E[\phi(X_i) \mid X_{i+1}, \dots] \qquad \text{a.s.}$$

Condition on τ: $E[\phi(X_1) \mid \tau] = E[\phi(X_i) \mid \tau]$ a.s. Therefore, X_1 and X_i have the same conditional distribution given τ. □

Recall Glivenko-Cantelli: Define $F(x_1, x_2, \dots, x_n, t)$ to be the empirical distribution of $(x_1, \dots, x_n)$:

$$F(x_1, x_2, \dots, x_n, t) = \frac{1}{n} \sum_{i=1}^{n} 1_{(x_i \leq t)}$$

If $(X_i, i \geq 1)$ are IID with distribution function F, then $F(X_1, \dots, X_n, t) \xrightarrow{\text{a.s.}} F(t)$, for each t, as $n \to \infty$.

Given exchangeable $(X_i, 1 \leq i < \infty)$, de Finetti's Theorem 25.2 implies that

$$F(X_1, \cdots, X_n, t) \xrightarrow{\text{a.s.}} G(\omega, t) \tag{25.1}$$

which is the distribution function of $Q(\omega, \cdot)$.

We can identify Q with the limit 25.1.

25.3 MGs in Galton-Watson Branching Processes

ξ takes values in $\{0, 1, 2, \dots\}$. Each individual in generation g has ξ offspring in generation $g+1$. The ξ are independent. Z_n is the number of individuals in generation n, with $Z_0 = 1$ as a default. Write $\mu = E\xi < \infty$.

"Extinction" is the event $\{Z_n = 0 \text{ for some } n\}$ and "survival" is the event $\{Z_n \geq 1 \ \forall n\}$.

Let $\mathcal{F}_n = \sigma(Z_0, Z_1, \dots, Z_n)$.

$$E[Z_{n+1} \mid \mathcal{F}_n] = \mu Z_n \tag{25.2}$$

This implies that $EZ_{n+1} = \mu \cdot EZ_n$, so inductively, $EZ_n = \mu^n$.

If $\mu < 1$, then $P(Z_n \geq 1) \leq EZ_n \leq \mu^n \to 0$, so $P(\text{extinction}) = 1$.

Undergraduate: "$P(\text{extinction}) < 1$" if and only if $\mu > 1$ or $P(\xi = 1) = 1$.

Study the case $\mu > 1$. 25.2 implies that $(Z_n/\mu^n, n \geq 0)$ is a MG, since $E[Z_n/\mu^n] = 1$. By the MG convergence theorem, $Z_n/\mu^n \xrightarrow{\text{a.s.}} W \geq 0$, $EW \leq 1$. Suppose $E\xi^2 < \infty$. We will show $(Z_n/\mu^n, n \geq 1)$ is UI. Then, $Z_n/\mu^n \to W$ in L^1 and $EW = 1$. Clearly, $\{\text{extinction}\} \subseteq \{W = 0\}$. We can prove $\{\text{extinction}\} = \{W = 0\}$ a.s. So, either we have extinction, or Z_n grows exponentially fast.

Calculation:

$$\operatorname{var}(Z_n) = E\underbrace{\operatorname{var}(Z_n \mid \mathcal{F}_{n-1})}_{Z_{n-1}\operatorname{var}(\xi)} + \operatorname{var}\underbrace{E[Z_n \mid \mathcal{F}_{n-1}]}_{\mu \cdot Z_{n-1}}$$

$$\operatorname{var}\left(\frac{Z_n}{\mu^n}\right) = \frac{\operatorname{var}(\xi)}{\mu^{n+1}} + \operatorname{var}\left(\frac{Z_{n-1}}{\mu^{n-1}}\right)$$

By induction,

$$\begin{aligned}\operatorname{var}\left(\frac{Z_n}{\mu^n}\right) &= \operatorname{var}(\xi) \cdot \sum_{i=2}^{n+1} \frac{1}{\mu^i} \\ &\leq K < \infty \qquad \text{for all } n\end{aligned}$$

so $(Z_n/\mu^n, n \geq 1)$ is UI.

25.4 L^2 Theory

Topic: L^2 theory. (See Durrett for more.)

Consider $(M_n, n \geq 0)$, $M_0 = 0$, with $\Delta_n = M_n - M_{n-1}$. Suppose $EM_n^2 < \infty$, for all n.

Orthogonality of Increments. $E[\Delta_i \Delta_j] = 0$, for $i < j$, because $E[\Delta_i\Delta_j \mid \mathcal{F}_{j-1}] = \Delta_i E[\Delta_j \mid \mathcal{F}_{j-1}] = 0$. So $EM_n^2 = \sum_{i=1}^n E[\Delta_i^2]$. Say that the martingale is "L^2 **bounded**" if $\sup_n EM_n^2 < \infty$, which is equivalent to $\sum_{i=1}^\infty E[\Delta_i^2] < \infty$. If (M_n) is L^2 bounded, then (L^1 convergence) $M_n \xrightarrow{\text{a.s.}} M_\infty$ and in L^1. In fact, we also have $M_n \to M_\infty$ in L^2.

For $n_1 < n_2$, $E[(M_{n_2} - M_{n_1})^2] = \sum_{i=n_1+1}^{n_2} E[\Delta_i^2]$. If (M_n) is L^2 bounded,

$$\lim_{n\to\infty} \sup_{n_2 > n_1} \|M_{n_2} - M_{n_1}\|_2 = 0$$

"Cauchy criterion $\implies$ convergence" is the definition of a "complete metric space".

Fact. L^2 is a complete metric space.

This implies that $M_n \to M_\infty$ in L^2.

Lecture 26

November 22

26.1 Brownian Motion

A $\mathbb{R}^1$-valued process $(B(t), 0 \le t < \infty)$ is **(standard) Brownian motion (Wiener process)** if $B(0) = 0$ and

(a) $B(t_0), B(t_1) - B(t_0), \dots, B(t_n) - B(t_{n-1})$ are independent, for any $0 \le t_0 < t_1 < \cdots < t_n$ ("independent increments").

(b) $B(t) - B(s)$ has the $\text{Normal}(0, t-s)$ distribution, where $t - s$ is the variance.

(c) The sample paths $t \mapsto B(t)$ are continuous. We have a measurable function $B(\omega, t)$. In other words, for all ω, $t \mapsto B(\omega, t)$ is continuous $[0, \infty) \to \mathbb{R}$.

Write $\mathbb{Q}_2$ for the **dyadic rationals**, the set of $\{i/2^j, i, j \ge 0\}$. We will work on the time interval $[0, 1]$. Enumerate $\mathbb{Q}_2$ as $q_1, q_2, q_3, \dots$. For each n, properties (a) and (b) specify a joint distribution of

$$(B(q_1), B(q_2), \dots, B(q_n))$$

by relabeling the q_i. These are *consistent*, as n increases. Suppose that we add a time s between t_1 and t_2. We check that $\text{Normal}(0, s - t_1) + \text{Normal}(0, t_2 - s) = \text{Normal}(0, t_2 - t_1)$ for independent normals. Use the Kolmogorov Extension Theorem to show that there exists a process $(B(q), q \in \mathbb{Q}_2 \cap [0, 1])$.

For $f : \mathbb{Q}_2 \cap [0, 1] \to \mathbb{R}$ and $\delta > 0$, define

$$w(f, \delta) = \sup_{\substack{0 \le q_1 < q_2 \le 1 \\ q_i \in \mathbb{Q}_2 \\ q_2 - q_1 \le \delta}} |f(q_2) - f(q_1)|$$

Lemma 26.1. *If*

$$w(f, \delta) \to 0 \qquad \text{as} \qquad \delta \to 0 \tag{26.1}$$

then there exists a continuous $\tilde{f} : [0, 1] \to \mathbb{R}$ *such that* $\tilde{f}(q) = f(q)\ \forall q \in \mathbb{Q}_2 \cap [0, 1]$.

Proof. Define

$$\tilde{f}(t) = \limsup_{\substack{q \downarrow t \\ q \in \mathbb{Q}_2}} f(q)$$

If $|t - s| \le \delta$, then $|\tilde{f}(t) - \tilde{f}(s)| \le w(t, \delta)$. Then, (26.1) implies that $\tilde{f}$ is continuous. ☐

Now, it is enough to show $P(w(B(\cdot),\delta) \geq \varepsilon) \to 0$ as $\delta \downarrow 0$, with $\varepsilon > 0$ fixed. This will imply $w(B(\cdot),\delta) \to 0$ a.s. as $\delta \to 0$. Then, we can apply 26.1 to show that there exists $\tilde{B}$ such that ($t \mapsto \tilde{B}(\omega,t)$ is continuous) a.s. It is easy to check that properties (a) and (b) remain true for all real t. Redefine $B(t,\omega) \equiv 0$ $\forall t$ on a null set.

Define

$$\bar{w}(f,2^{-m}) = \max_{0 \leq j \leq 2^m-1} \sup_{j/2^m \leq q \leq (j+1)/2^m} |f(q) - f(j/2^m)|$$

Consider $0 \leq q_1 < q_2 \leq 1$ with $q_2 - q_1 \leq 1/2^m$, which means they are either in the same or adjacent intervals. Then $|f(q_2) - f(q_1)| \leq 3\bar{w}(f,2^{-m})$. It is enough to prove $P(\bar{w}(B(\cdot),2^{-m}) \geq \varepsilon) \to 0$ as $m \to \infty$. ($Y_n \downarrow 0$ in probability implies $Y_n \downarrow 0$ a.s.)

Define $S_m = \sup_{0 \leq q \leq 1/2^m} |B(q)|$. $\bar{w}(B(\cdot),2^{-m})$ is the maximum of 2^m identically distributed RVs. Then $P(\bar{w}(B(\cdot),2^{-m}) \geq \varepsilon) \leq 2^m P(S_m \geq \varepsilon)$.

Fix m and take $n > m$. Consider $B(i/2^n, 0 \leq i \leq 2^n/2^m)$. This is a MG. Therefore, $B^4(i/2^n, i \geq 0)$ is a sub-MG. Use the L^1 maximal inequality.

$$P\left(\max_{i/2^n \leq 1/2^m} B^4\left(\frac{i}{2^n}\right) \geq \varepsilon^4\right) \leq \varepsilon^{-4} E B^4\left(\frac{1}{2^m}\right) = \varepsilon^{-4} 2^{-2m} E Z^4$$

(If Z is Normal$(0,1)$, then $B(t) \overset{\mathrm{d}}{=} t^{1/2} Z$.) Let $n \to \infty$. Then $P(S_m > \varepsilon) \leq \varepsilon^{-4} 2^{-2m} E Z^4$.

$$P(\bar{w}(B(\cdot),2^{-m}) \geq \varepsilon) \leq 2^m P(S_m \geq \varepsilon) \leq 2^{-m} \varepsilon^{-4} E Z^4 \to 0 \qquad \text{as } m \to \infty$$

Theorem 26.2. *For almost all ω, the sample path $t \mapsto B(\omega,t)$ is* ***nowhere*** *differentiable.*

If Brownian motion were differentiable at the origin, we would expect $B(t) \sim O(t)$ as $t \to 0$, which contradicts the fact that $B(t)$ has SD $t^{1/2}$.

Analysis. Consider $f : [0,1] \to \mathbb{R}$. Fix $C < \infty$. Suppose $\exists s$ such that $f'(s)$ exists and $|f'(s)| \leq C/2$. Then, there exists n_0 such that for $n \geq n_0$,

$$|f(t) - f(s)| \leq C|t-s| \qquad \text{for all } t \text{ such that } |t-s| \leq 3/n \tag{26.2}$$

Rewrite the above statement: define $A_n = \{f : (26.2) \text{ holds for some } s\}$. As $n \to \infty$,

$$A_n \uparrow A \supseteq \{f : |f'(s)| \leq C/2 \text{ for some } s\}$$

For $0 \leq k \leq n-1$, define

$$Y(f,k,n) = \max\left(\left|f\left(\frac{k+3}{n}\right) - f\left(\frac{k+2}{n}\right)\right|, \left|f\left(\frac{k+2}{n}\right) - f\left(\frac{k+1}{n}\right)\right|, \left|f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right|\right)$$

Given $f \in A_n$, (26.2) holds for some s, say $k/n \leq s \leq (k+1)/n$. Near s, the slope is C, so the maximum difference is at most $C \cdot 5/n$, so $Y(f,k,n) \leq 5C/n$. Then

$$A_n \subseteq D_n \overset{\text{def}}{=} \{f : Y(f,k,n) \leq 5C/n \text{ for some } k \leq n-1\}$$

Probability.

$$P\left(\left|B\left(\frac{k+1}{n}\right) - B\left(\frac{k}{n}\right)\right| \leq \frac{5C}{n}\right) = P\left(|Z| \leq \frac{5C}{n^{1/2}}\right) \tag{26.3}$$

$$\leq (2\pi)^{-1/2} \cdot \frac{10C}{n^{1/2}} \tag{26.4}$$

since the increment is $\text{Normal}(0, 1/n) = n^{-1/2}Z$. Regard $B(\cdot)$ as a random f.

$$P\left(Y(B,k,n) \leq \frac{5C}{n}\right) \leq (26.3)^3 \leq \frac{1000C^3}{n^{3/2}} \tag{26.5}$$

Then

$$\begin{aligned} P(B(\cdot) \in D_n) &\leq n \cdot (26.5) \\ &\leq \frac{1000C^3}{n^{1/2}} \\ P(B(\cdot) \in A_n) &\leq \frac{1000C^3}{n^{1/2}} \end{aligned}$$

Let $n \to \infty$. $P(B(\cdot) \in A) = 0$.

Lecture 27

November 29

27.1 Aspects of Brownian Motion

- model for many processes fluctuating continuously: stock market, etc.
- (*Theory*) limit of RWs with small step size
- Gaussian process
- "diffusions": continuous-path Markov processes
- martingale properties

We will concentrate on the last aspect.

Definition 27.1. Brownian motion $(B(t), 0 \le t < \infty)$ has the properties

- for $s < t$, $B(t) - B(s) \overset{\mathrm{d}}{=} \mathrm{Normal}(0, t-s)$
- for $0 \le t_1 < t_2 < \cdots < t_n$, the increments $(B(t_{i+1}) - B(t_i), 1 \le i \le n-1)$ are independent
- the sample paths $t \mapsto B(t)$ are continuous
- $B(0) = 0$

27.2 Continuous-Time Martingales

$(M_t, \mathcal{F}_t)$ with the filtration $(\mathcal{F}_t, 0 \le t < \infty)$ is a **MG** if

- $E|M_t| < \infty \ \forall t$
- M_t is adapted to $\mathcal{F}_t$
- for $s < t$, $E[M_t \mid \mathcal{F}_t] = M_s$ a.s.

All of our MGs will have continuous paths. The general theory requires only right-continuity.

$T : \Omega \to [0, \infty)$ is a **stopping time** if $\{T \le t\} \in \mathcal{F}_t$, for all $0 \le t < \infty$. In discrete time, the stopping time property with $\{T \le n\}$ was equivalent to the definition with $\{T = n\}$, but this is not true in continuous time.

Theorem 27.2 (Optional Sampling Theorem)**.** *If (M_t) is a MG, if T is a stopping time, and if (for t_0 an integer, WLOG) $P(T \le t_0) = 1$, then $EM_T = EM_0$.*

Proof. Fix m and look at times that are multiples of 2^{-m}. Define $T_m = \inf\{i/2^m : i/2^m > T\}$. Note that $\{T < t\} = \bigcup_n \{T \leq t - 1/n\} \in \mathcal{F}_t$. This T_m is a stopping time for $(M_{i/2^m}, \mathcal{F}_{i/2^m}, i \geq 0)$, and $T_m \leq t_0 + 1$. Apply the discrete-time OST to obtain $EM_{T_m} = EM_0$ and $M_{T_m} = E[M_{t_0+1} \,|\, \mathcal{F}_{T_m}]$ (which implies that $(M_{T_m}, m \geq 1)$ is UI). As $m \to \infty$, $T_m \downarrow T$, and right-continuity implies that $M_{T_m} \to M_T$ a.s., so $EM_{T_m} \to EM_T$. □

With BM we associate the natural filtration $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$.

Proposition 27.3. *The following are MGs.*

- B_t
- $B_t^2 - t$
- $\exp(\theta B_t - \theta^2 t/2)$*, for* $\theta \in \mathbb{R}$
- $B_t^3 - 3tB_t$
- $B_t^4 - 6tB_t^2 + 3t^2$

Proof. Fix $s < t$.

$$\begin{aligned} B_t &= B_s + (B_t - B_s) \\ E[B_t \mid \mathcal{F}_s] &= B_s + E[B_t - B_s \mid \mathcal{F}_s] \\ &= B_s + E[B_t - B_s] \\ &= B_s + 0 = B_s \end{aligned}$$

$B_t - B_s$ is independent of $(B_{s_1}, B_{s_2}, \ldots, B_{s_n})$ for all $0 \leq s_1 < s_2 < \cdots < s_n \leq s$. We conclude that $B_t - B_s$ is independent of $\mathcal{F}_s \overset{\text{def}}{=} \sigma(B_i, 0 \leq u \leq s)$ using the MT fact about independence: it suffices to prove independence for any finite subcollection.

Write $Y_t = B_t^2 - t = (B_t + (B_t - B_s))^2 - t$.

$$\begin{aligned} Y_t &= Y_s + 2B_s(B_t - B_s) + (B_t - B_s)^2 - (t-s) \\ E[Y_t \mid \mathcal{F}_s] &= Y_s + 2B_s \underbrace{E[B_t - B_s \mid \mathcal{F}_s]}_{=0} + \underbrace{E[(B_t - B_s)^2 \mid \mathcal{F}_s]}_{=E(B_t - B_s)^2 = t-s} - (t-s) = Y_s \end{aligned}$$

Aside. If $W \overset{\text{d}}{=} \text{Normal}(0, \sigma^2)$, then $E\exp(\theta W) = \exp(\theta^2\sigma^2/2)$.

Write $Z_t^\theta = \exp(\theta B_t - \theta^2 t/2)$.

$$\begin{aligned} Z_t &= Z_s \exp(\theta(B_t - B_s)) \exp\left(-\frac{\theta^2}{2}(t-s)\right) \\ E[Z_t \mid \mathcal{F}_s] &= Z_s \exp\left(\frac{\theta^2}{2}(t-s)\right) \underbrace{E\exp(\theta(B_t - B_s))}_{=\exp(\theta^2(t-s)/2)} = Z_s \end{aligned}$$

Informally, $(Z_t^\theta, 0 \leq t < \infty)$ is a MG, so

$$\left(\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} Z_t^\theta, 0 \leq t < \infty\right)$$

should be a MG. If we differentiate k times, and set $\theta = 0$, we get a sequence of polynomials in B_t. □

A typical stopping time is $T_b = \inf\{t : B(t) = b\} = \inf\{t : B(t) \geq b\}$ (for $b > 0$). Also, for $b > 0$, $t > 0$, $\{T_b \leq t\} = \{\sup_{s \leq t} B(s) \geq b\}$. Note that

$$\sup_{s \leq t} B(s) = \sup_{\substack{u \leq t \\ u \text{ rational}}} B(u)$$

is $\mathcal{F}_t$-measurable.

Lemma 27.4. *Fix $-a < 0 < b$. Consider $T = \min\{T_{-a}, T_b\}$. Then*

$$P(B_T = b) = \frac{a}{a+b} = P(T_b < T_{-a}) \tag{27.1}$$

$$P(B_T = -a) = \frac{b}{a+b} \tag{27.2}$$

$$ET = ab \tag{27.3}$$

Proof. $P(T > t) \leq P(B(t) \in [-a, b]) \to 0$ as $t \to \infty$, so $T < \infty$ a.s. Apply OST, 27.2, to 0 and $T \wedge t$.

$$0 = EB_0 = EB_{T \wedge t}$$

As $t \to \infty$, $B_{T \wedge t} \to B_T$ a.s. and

$$|B_{T \wedge t}| \leq \max(a, b)$$

This implies that $0 = EB_T$, but B_T takes values in $\{-a, b\}$ only, so we must have the distribution (27.1) and (27.2).

Apply the OST 27.2 to $B_t^2 - t$. Then $EB_{T \wedge t}^2 = E[T \wedge t]$. Let $t \to \infty$.

$$EB_T^2 = ET = b^2 \left(\frac{a}{a+b}\right) + (-a)^2 \left(\frac{b}{a+b}\right) = ab \qquad \square$$

Note $P(T_b < \infty) \geq P(T_b < T_{-a}) \to 1$ as $a \to \infty$, so $T_b < \infty$ a.s.

Fix $c > 0$ and $-\infty < d < \infty$. Consider $T = \inf\{t : B_t = c + dt\} \leq \infty$.

Lemma 27.5.

$$E \exp(-\lambda T) = \exp\left(-c\left(d + \sqrt{d^2 + 2\lambda}\right)\right)$$

for $0 \leq \lambda < \infty$. This is the Laplace transform of T.

Proof. Consider $\theta > \max(0, 2d)$. Apply the OST 27.2 to $\exp(\theta B_t - \theta^2 t/2)$ and $T \wedge t$.

$$1 = E \exp\left(\theta B_{T \wedge t} - \frac{\theta^2}{2}(T \wedge t)\right) \tag{27.4}$$

Case $d \leq 0$, $\theta > 0$: Here, $B_{T \wedge t} - (\theta^2/2)(T \wedge t) \leq \theta c$, $T \leq T_c < \infty$.

Case $d > 0$, $\theta > 2d$:

$$\theta B_{T\wedge t} - \frac{\theta^2}{2}(T \wedge t) \le \sup_{0\le s<\infty} \left(\theta(c + ds) - \frac{\theta^2}{2} s\right) \equiv \theta c$$

and $\theta B_{T\wedge t} - (\theta^2/2)(T \wedge t) \to \infty$ as $t \to \infty$ on $\{T = \infty\}$.

Let $t \to \infty$. $1 = E[\exp(\theta B_T - (\theta^2/2)T)]1_{(T<\infty)}$. Put $B_T = c + dT$ on $\{T < \infty\}$.

$$1 = \exp\left(\theta c + \left(\theta d - \frac{\theta^2}{2}\right) T\right) 1_{(T<\infty)}$$

Given $\lambda > 0$, define $\theta = \theta(\lambda)$ as the solution of $\theta d - \theta^2/2 = -\lambda$, so $\theta(\lambda) = d + \sqrt{d^2 + 2\lambda} > \max(0, 2d)$.

$$1 = E \exp(c\theta(\lambda) - \lambda T)$$
$$E \exp(-\lambda T) = \exp(-c\theta(\lambda))$$

□

Lecture 28

December 1

28.1 Explicit Calculations with Brownian Motion

Last class:

$$T_a = \inf\{t : B_t = a\} \tag{28.1}$$
$$T_{c,d} = \inf\{t : B_t = c + dt\}, \qquad c > 0, -\infty < d < \infty \tag{28.2}$$
$$E\exp(-\lambda T_{c,d}) = \exp\left(-c\left(d + \sqrt{d^2 + 2\lambda}\right)\right), \qquad 0 < \lambda < \infty \tag{28.3}$$

28.1.1 Consequences of Formula (28.3)

Special Cases.

1. $d = 0$, $c > 0$.

$$E\exp(-\lambda T_c) = \exp(-c\sqrt{2\lambda}), \qquad 0 < \lambda < \infty$$

We can invert this to get the formula for the density.

2.

$$P(T_{c,d} < \infty) = \lim_{\lambda \downarrow 0} E\exp(-\lambda T_{c,d}) = \begin{cases} \exp(-2cd), & d \geq 0 \\ 1, & d \leq 0 \end{cases}$$

We know $0 < T \leq \infty$. As $\lambda \downarrow 0$, $\exp(-\lambda T) \uparrow 1_{(T<\infty)}$, so the expectation converges to $P(T < \infty)$ by monotone convergence.

Define $M_d \stackrel{\text{def}}{=} \sup_{t \geq 0}(B_t - dt)$, which is BM with drift $-d$. The event $\{M_d \geq c\} = \{T_{c,d} < \infty\}$, so $P(M_d \geq c) = \exp(-2dc)$ (for $d > 0$). Therefore, the distribution of M_d is Exponential($2d$),

$$EM_d = \frac{1}{2d}$$

28.1.2 Reflection Principle Formula & Consequences

Theorem 28.1 (Reflection Principle)**.** *For $a, b > 0$, $t > 0$,*

$$P(T_a \leq t, B_t \geq a + b) = P(T_a \leq t, B_t \leq a - b)$$

Condition on $T_a = s$, say. The future process $\tilde{B}_u = B_{s+u} - a$, $0 \leq u < \infty$, is distributed as BM.

$$P(B_t \geq a + b \mid T_a = s) = P(\tilde{B}_{t-s} \geq b)$$

$$P(B_t \le a-b \mid T_a = s) = P(\tilde{B}_{t-s} \le -b)$$
$$P(B_t \ge a+b \mid T_a = s) = P(B_t \le a-b \mid T_a = s)$$

This implies (by integration)

$$P(B_t \ge a+b \mid T_a \le t) = P(B_t \le a-b \mid T_a \le t)$$

$$P(T_a \le t) = 2P(B_t \ge a) = P(|B_t| \ge a) \tag{28.4}$$

The standard Normal density for Z is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$
$$\bar{\Phi}(x) = \int_x^\infty \phi(u)\,\mathrm{d}u$$

and $B_t \stackrel{\mathrm{d}}{=} t^{1/2} Z$.

$$P(T_a \le t) = 2P(Z \ge at^{-1/2}) = 2\bar{\Phi}(at^{-1/2})$$

T_a has density

$$\begin{aligned} f_{T_a}(t) &= 2 \cdot \left(-\frac{1}{2} a t^{-3/2}\right)(-\phi(at^{-1/2})) \\ &= \frac{a}{\sqrt{2\pi}} t^{-3/2} \exp\left(-\frac{a^2}{2t}\right), \qquad 0 < t < \infty \end{aligned}$$

Check that this is consistent with $E\exp(-\lambda T_a) = \exp(-a\sqrt{2\lambda})$. Because $f_{T_a}(t) \approx t^{-3/2}$ as $t \to \infty$, $ET_a = \infty$.

Consider $M_t = \sup_{0 \le s \le t} B_s$. Then the event $\{M_t \ge a\}$ equals the event $\{T_a \le t\}$, so

$$P(M_t \ge a) = P(T_a \le t) = P(|B_t| \ge a)$$

by (28.4). Therefore, $M_t \stackrel{\mathrm{d}}{=} |B_t|$, for each $0 < t < \infty$, but they are not the same as processes.

We can use the Reflection Principle 28.1 formula to find the joint distribution of (M_t, B_t). We know that $P(T_a \le t, B_t \ge a+b) = P(B_t \ge a+b)$, so $P(B_t \ge a+b) = P(M_t \ge a, B_t \le a-b)$. Replace b by $a-b$.

$$P(B_t \ge 2a-b) = P(M_t \ge a, B_t \le b)$$

(for $a > 0$, $a > b$). Hence,

$$P(M_t \ge a, B_t \le b) = \bar{\Phi}\left(\frac{2a-b}{t^{1/2}}\right)$$

So, (M_t, B_t) has joint density

$$\begin{aligned} f_{M_t,B_t}(a,b) &= -\frac{\mathrm{d}}{\mathrm{d}a}\frac{\mathrm{d}}{\mathrm{d}b}\bar{\Phi}\left(\frac{2a-b}{t^{1/2}}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}a}\left(t^{-1/2}\phi\left(\frac{2a-b}{t^{1/2}}\right)\right) \\ &= -t^{-1/2} \cdot 2t^{-1/2}\phi'\left(\frac{2a-b}{t^{1/2}}\right) \\ &= \frac{2}{t}\frac{2a-b}{t^{1/2}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(2a-b)^2}{2t}\right) \end{aligned}$$

$$f_{M_t,B_t}(a,b) = \frac{2(2a-b)}{\sqrt{2\pi}t^{3/2}} \exp\left(-\frac{(2a-b)^2}{2t}\right), \qquad \text{for } \infty > a \geq b > -\infty \tag{28.5}$$

using $\phi'(x) = -x\phi(x)$.

Special Cases of (28.5) *with* $t = 1$.

$$f(a,0) = = \frac{4a}{\sqrt{2\pi}} \exp(-2a^2)$$

The conditional density of M_1 given $B_1 = 0$ is

$$\begin{aligned} f_{M_1 \mid B_1}(a \mid 0) &= \frac{f_{M_1,B_1}(a,0)}{f_{B_1}(0)} \\ &= 4a \exp(-2a^2) \end{aligned}$$

since

$$f_{B_1}(0) = \phi(0) = \frac{1}{\sqrt{2\pi}}$$

Therefore,

$$P(M_1 \geq a \mid B_1 = 0) = \exp(-2a^2)$$

Brownian bridge $(B_t^0, 0 \leq t \leq 1)$ is defined as $(B_t, 0 \leq t \leq 1)$, conditioned on $(B_1 = 0)$. Hence, $P(M^0 \geq a) = \exp(-2a^2)$ for $M^0 = \sup_{0 \leq t \leq 1} B_t^0$.

Another Special Case: $a = 0$. Consider

$$\begin{aligned} f_{B_1|M_1}(-b \mid 0) = \frac{f_{M_1,B_1}(0,-b)}{f_{M_1}(0)} &= \frac{2b}{\sqrt{2\pi}} \exp\left(-\frac{b^2}{2}\right) \cdot \frac{\sqrt{2\pi}}{2} \\ &= b \exp\left(-\frac{b^2}{2}\right) \end{aligned}$$

Therefore,

$$P(B_1 \geq b \mid B_t \geq 0 \; \forall t \in [0,1]) = \exp\left(-\frac{b^2}{2}\right), \qquad b > 0$$

$M_1 \stackrel{\mathrm{d}}{=} |B_1|$ implies $f_{M_1}(0) = 2\phi(0)$.

Brownian meander $(B_t^{(m)}, 0 \leq t \leq 1)$ is defined as BM conditioned on $(B_t \geq 0, 0 \leq t \leq 1)$. The calculation gives the distribution of $B_1^{(m)}$.