Probability Theory

Mathematics C218B/Statistics C205B
Spring 2017

Professor David Aldous
Scribe: Sinho Chewi

Contents

Lecture 1

January 17

1.1 Convergence in Distribution

We have two definitions:

- Probability measure (PM) μ on $\mathbb{R}$,
- Distribution function F on $\mathbb{R}$.

Given μ, $F(x) \stackrel{\text{def}}{=} \mu(-\infty, x]$ is a distribution function.

Given F, there exists a μ such that $F(x) = \mu(-\infty, x]$.

x is a **continuity point** of F if $F(x) = F(x-)$, which means $\mu\{x\} = 0$.

Theorem 1.1. *For PMs $(\mu_n, 1 \leq n < \infty)$ and μ on $\mathbb{R}$, the following are equivalent.*

1. *$F_{\mu_n}(x) \to F_\mu(x)$ as $n \to \infty$ for all continuity points x of F.*
2. *$\int_{-\infty}^{\infty} g(x)\mu_n(\mathrm{d}x) \xrightarrow{n\to\infty} \int_{-\infty}^{\infty} g(x)\mu(\mathrm{d}x)$ for all bounded continuous $g : \mathbb{R} \to \mathbb{R}$.*
3. *There exist, on some probability space, RVs $(\hat{X}_n, 1 \leq n < \infty)$ and $(\hat{X})$ such that for all $1 \leq n < \infty$, $\mathrm{dist}(\hat{X}_n) = \mathrm{dist}(X_n)$, $\mathrm{dist}(\hat{X}) = \mathrm{dist}(X)$, and $\hat{X}_n \to \hat{X}$ a.s. as $n \to \infty$.*

Note: 2 and 3 make sense for PMs on a metric space $\mathcal{S}$ and define "weak convergence" on $\mathcal{S}$. In fact, 2 $\Leftrightarrow$ 3 on general $\mathcal{S}$ ("Skorohod representation theorem"). The theorem shows that 1 is not just arbitrary.

$\mathrm{dist}(X)$ is often written as $\mathcal{L}(X)$ (for "law"). Write $X_n \xrightarrow{\mathrm{d}} X$ "in distribution" to mean $\mathrm{dist}(X_n) \to \mathrm{dist}(X)$. Call this "weak convergence" $\mu_n \to \mu$.

Proof. 3 $\implies$ 2: $\hat{X}_n \to \hat{X}$ a.s. implies that $g(\hat{X}_n) \to g(\hat{X})$ a.s. (g is continuous), which implies that $Eg(\hat{X}_n) \to Eg(\hat{X})$ (g is bounded), which implies that $Eg(X_n) \to Eg(X)$. 2 is equivalent to saying $Eg(X_n) \to Eg(X)$ for all bounded, continuous g.

2 $\implies$ 1: Fix x_0 and define $f_j(x)$ by 1 when $x \leq x_0$, 0 when $x \geq x_0 + 1/j$, and linear in between.

$$\begin{aligned} F_{\mu_n}(x_0) &= \int_{-\infty}^{\infty} 1_{(x \leq x_0)} \mu_n(\mathrm{d}x) \\ &\leq \int_{-\infty}^{\infty} f_j(x)\mu_n(\mathrm{d}x) \end{aligned}$$

$$\limsup_n F_{\mu_n}(x_0) \le \lim_n \int_{-\infty}^{\infty} f_j(x)\mu_n(\mathrm{d}x) = \int_{-\infty}^{\infty} f_j(x)\mu(\mathrm{d}x) \le F_\mu(x_0 + 1/j)$$

by 2. Let $j \to \infty$ to obtain

$$\limsup_n F_{\mu_n}(x_0) \le F_\mu(x_0).$$

Define $g_j(x)$ by 1 when $x \le x_0 - 1/j$, 0 when $x \ge x_0$, and linear in between.

$$\begin{aligned}\liminf_n F_{\mu_n}(x_0) &\ge \lim_n \int_{-\infty}^{\infty} g_j(x)\mu_n(\mathrm{d}x) \\ &= \int_{-\infty}^{\infty} g_j(x)\mu(\mathrm{d}x) \\ &\ge F_\mu(x_0 - 1/j)\end{aligned}$$

Let $j \to \infty$.

$$\liminf_n F_{\mu_n}(x_0) \ge F_\mu(x_0-)$$

If x_0 is a continuity point, we have shown $F_{\mu_n}(x_0) \to F_\mu(x_0)$.

1 $\implies$ 3: Recall the inverse function of F_μ.

$$F_\mu^{-1}(y) \stackrel{\text{def}}{=} \sup\{x : F_\mu(x) < y\} = \inf\{x : F_\mu(x) \ge y\}$$

If U is uniform on $[0,1]$, then F_μ^{-1} is a RV whose distribution is μ.

Exercise. 1 implies $F_{\mu_n}^{-1}(y) \to F_\mu^{-1}(y)$ for all y such that $\{x : F_\mu(x) = y\}$ is either empty or a single point x.

The other case is when $\{x : F_\mu(x) = y\}$ is a non-trivial interval. This can only happen for countably many y. $F_{\mu_n}^{-1}(U) \to F_\mu^{-1}(U)$ a.s. (all U outside a countable set). This is 3. □

1.2 Elementary Examples

Here are elementary examples where we show 1 by calculation.

Example 1.2. If X_n has the uniform distribution on $\{1, 2, \ldots, n\}$, then $X_n/n \xrightarrow{\mathrm{d}} U$, which is uniform on $[0,1]$.

Example 1.3. X_θ has the Geometric(θ) distribution. $P(X > i) = (1-\theta)^i$, $i = 0, 1, 2, \ldots$. Then, $\theta X_\theta \xrightarrow{\mathrm{d}} Y$ with the Exponential(1) distribution, $P(Y > y) = e^{-y}$, $0 \le y < \infty$.

Example 1.4. B_n is the "birthday RV", $\min\{j : \xi_j = \xi_i \text{ for some } 1 \le i < j\}$ for IID ξ_i uniform on $\{1, 2, \ldots, n\}$. Then $n^{-1/2}B_n \xrightarrow{\mathrm{d}} R$ with Rayleigh distribution $P(R > x) = \exp(-x^2/2)$.

1.2.1 Artificial Examples

Example 1.5. For any X: $X + 1/n \xrightarrow{\mathrm{d}} X$ as $n \to \infty$.

Note: $F_{X+1/n}(x) = F_X(x - 1/n) \to F_X(x)$ iff $F_X(x) = F_X(x-)$.

Example 1.6. If X_n is uniform on the interval $[x_0 - 1/n, x_0 + 1/n]$, then $X_n \xrightarrow{\text{d}} x_0$. Above, we had examples of discrete distributions converging to continuous distributions. This example shows that continuous distributions can converge to discrete distributions.

Example 1.7. X_n has density $f_n(x) = (1/2)(1 + \sin(2\pi n x))$ on $0 \le x \le 1$. $X_n \xrightarrow{\text{d}} U$, uniform on $[0, 1]$, with $f_U(x) \equiv 1$. Here, it is *not* true that $f_{X_n}(x) \to f_U(x)$.

1.3 Consequences of Weak Convergence

For a function $g : \mathbb{R} \to \mathbb{R}$, write $D_g = \{x : g \text{ is not continuous at } x\}$ and assume D_g is measurable.

Corollary 1.8. *If $X_n \xrightarrow{\text{d}} X$, if $P(X \in D_g) = 0$, then $g(X_n) \xrightarrow{\text{d}} g(X)$. Then, if g is bounded, we have $Eg(X_n) \to Eg(X)$.*

Proof. Use 3. There exist $\hat{X}_n \to \hat{X}$ a.s. (outside some Ω_0, $P(\Omega_0) = 0$), so $g(\hat{X}_n) \to g(\hat{X})$ a.s. (outside $\Omega_0 \cup \{X \in D_g\}$), which by 3 implies $g(X_n) \xrightarrow{\text{d}} g(X)$. By bounded convergence, $Eg(X_n) \to Eg(X)$. □

If $X_n \xrightarrow{\text{d}} X$, then $1/X_n \xrightarrow{\text{d}} 1/X$, provided $P(X = 0) = 0$.

Corollary 1.9. *If $X_n \ge 0$, if $X_n \xrightarrow{\text{d}} X$, then $EX \le \liminf_n EX_n$.*

Proof. This is Fatou's Lemma for $\hat{X}_n \to \hat{X}$ a.s., $\hat{X}_n \ge 0$. Apply 3. □

Theorem 1.10 (Scheffe's Theorem)**.** *Let θ be a σ-finite measure on $(S, \mathcal{S})$. Suppose that measurable $h_n, h : S \to [0, \infty]$ are such that $\int_S h_n \,\text{d}\theta = 1$ for all n, $\int_S h \,\text{d}\theta = 1$, and $h_n(s) \to h(s)$ a.e. (θ). Then $\int_S |h_n(s) - h(s)| \theta(\text{d}s) \to 0$.*

Proof.

$$\int_S |h_n(s) - h(s)| \theta(\text{d}s) = 2 \int_S (h - h_n)^+ \theta(\text{d}s),$$

but $0 \le (h - h_0)^+ \le h$ and $(h - h_n)^+ \to 0$ a.e. The Dominated Convergence Theorem implies the result. □

Lecture 2

January 19

2.1 Conditions for Weak Convergence

Theorem 2.1 (Scheffe's Theorem). *Let $\theta(\cdot)$ be a σ-finite measure on S. If $h_n, h : S \to [0, \infty)$ satisfy $\int_S h_n \,\mathrm{d}\theta = 1$, $\int_S h \,\mathrm{d}\theta = 1$, and $h_n(s) \to h(s)$ θ-a.e., then $\int_S |h_n(s) - h(s)| \theta(\mathrm{d}s) \to 0$.*

Proposition 2.2. *Suppose $(X_n, 1 \le n < \infty)$ and X are integer-valued. The following are equivalent:*

(a) $X_n \xrightarrow{\mathrm{d}} X$.

(b) $P(X_n = i) \xrightarrow{n\to\infty} P(X = i)$, *for all* i.

(c) $\sum_i |P(X_n = i) - P(X = i)| \to 0$.

Proof. (a) $\implies$ (b): $P(X_n \le i + 1/2) \to P(X \le i + 1/2)$. Then,

$$P(X_n = i) = P(X_n \le i + 1/2) - P(X_n \le i - 1/2) \to P(X \le i + 1/2) - P(X \le i - 1/2) = P(X = i).$$

(b) $\implies$ (c): Scheffe's Theorem 2.1 for $\theta(i) \equiv 1$ for all i, $h_n(i) = P(X_n = i)$.

(c) $\implies$ (a):

$$\begin{aligned} |P(X_n \le x) - P(X \le x)| &= \left| \sum_{i \le x} (P(X_n = i) - P(X = i)) \right| \\ &\le \sum_i |P(X_n = i) - P(X = i)| \end{aligned}$$

□

Proposition 2.3. *If X_n and X have probability densities $f_n(x)$ and $f(x)$, if $f_n(x) \to f(x)$ for almost all x, then $X_n \xrightarrow{\mathrm{d}} X$.*

Proof. Scheffe's Theorem 2.1:

$$|P(X_n \le x) - P(X \le x)| \le \int |f_n(x) - f(x)| \,\mathrm{d}x \to 0$$

□

2.2 Tight Distributions

Consider $\mathbb{R}$-valued $(X_n, 1 \le n < \infty)$.

Definition 2.4. Say (X_n) is **tight** if $\lim_{B\uparrow\infty} \sup_n P(|X_n| \ge B) = 0$.

Definition 2.5. Say (X_n) is **uniformly integrable** if $\lim_{B\uparrow\infty} \sup_n E[|X_n| 1_{(|X_n| \ge B)}] = 0$.

Actually, the above definitions are properties of $\mu_n = \text{dist}(X_n)$.

Lemma 2.6 (Easy)**.** *(a) If $\sup_n E|X_n| < \infty$, or more generally if $\sup_n E\phi(|X_n|) < \infty$ for some $0 \le \phi(x) \uparrow \infty$ as $x \uparrow \infty$, then (X_n) is tight.*

(b) If $\sup_n EX_n^2 < \infty$, or more generally if $\sup_n E\phi(|X_n|) < \infty$ for some $0 \le \phi(x) \uparrow \infty$ such that $\phi(x)/x \to \infty$ as $x \to \infty$, then (X_n) is UI.

Proof. (a) Markov's inequality:

$$P(|X_n| \ge B) \le \frac{E\phi(|X_n|)}{\phi(B)}$$

□

Lemma 2.7 (205A)**.** *If $X_n \to X$ a.s., if $(X_n, 1 \le n < \infty)$ is UI, then $E|X| < \infty$ and $EX_n \to EX$.*

Corollary 2.8. *If $X_n \xrightarrow{d} X$, if $(X_n, 1 \le n < \infty)$ is UI, then $E|X| < \infty$ and $EX_n \to EX$.*

(Apply the lemma to $\hat{X}_n$.)

Distribution functions F, or equivalently, PMs μ on $(-\infty, \infty)$, satisfy:

- $0 \le F(x) \le 1$, $\forall x \in (-\infty, \infty)$.
- $x \mapsto F(x)$ is increasing.
- $F(x+) = F(x)$ (right-continuity).
- $\lim_{x\uparrow\infty} F(x) = 1$, $\lim_{x\downarrow-\infty} F(x) = 0$.

An **extended distribution function** (EDF) F has the first three properties above.

$$\lim_{x\uparrow\infty} F(x) = \text{“}F(\infty)\text{”} \le 1$$
$$\lim_{x\downarrow-\infty} F(x) = \text{“}F(-\infty)\text{”} \ge 0$$

There is a one-to-one correspondence between PMs μ on $[-\infty, \infty]$ and EDFs. Think of an RV X with values in $[-\infty, \infty]$.

Theorem 2.9 (Helly's Selection Theorem)**.** *Let $F_1, F_2, \ldots$ be distribution functions on $(-\infty, \infty)$.*

- *There exists $n_j \to \infty$ and an EDF G such that $F_{n_j}(x) \to G(x)$ for all continuity points x of G.*
- *If $(F_n, 1 \le n < \infty)$ is tight, then G is a distribution function on $(-\infty, \infty)$.*

Suppose Z is standard Normal, with distribution function $\Phi(z)$. J is uniform on $\{1, 2, 3\}$, and

$$X_n = \begin{cases} -n, & \text{if } J = 1, \\ Z, & \text{if } J = 2, \\ n, & \text{if } J = 3. \end{cases}$$

Then, the distribution function of X_n does not converge to a distribution function.

Proof. (a) Let $q_1, q_2, q_3, \ldots,$ be the rationals. The sequence $F_1(q_1), F_2(q_2), F_3(q_3), \ldots$ is in $[0, 1]$ so (compactness) there exists a subsequence $m(1, 1), m(1, 2), m(1, 3), \ldots$ such that

$$F_{m(1,i)}(q_1) \xrightarrow[i\to\infty]{} \text{some limit } G_0(q_1).$$

Then, we use a *diagonal argument.* $F_{m(1,i)}(q_2)$, $i = 1, 2, \ldots$ is a sequence in $[0, 1]$; there exists a subsequence $m(2, 1), m(2, 2), m(2, 3), \ldots$ such that $F_{m(2,i)}(q_2) \to$ some $G_0(q_2)$.

Repeat for each $k \geq 1$: find a subsequence $(m(k, i), i \geq 1)$ of $(m(k-1, i), i \geq 1)$ such that

$$F_{m(k,i)}(q_k) \xrightarrow[i\to\infty]{} \text{some } G_0(q_k).$$

Consider $m(i, i)$ (the "diagonal"): this has the property $F_{m(i,i)}(q_k) \xrightarrow[i\to\infty]{} G_0(q_k)$ for *all* k.

Now, define an EDF G by

$$G(x) = \inf_{\substack{q \text{ rational} \\ q > x}} G_0(q).$$

Check that G *is* an EDF.

Fix x. For any $q > x$,

$$\begin{aligned} \limsup_i F_{m(i,i)}(x) &\leq \limsup_i F_{m(i,i)}(q) = G_0(q) \\ &\leq G(x) \end{aligned}$$

by letting $q \downarrow x$. By the same argument, $\liminf_i F_{m(i,i)}(x) \geq G(x-)$. So, if $G(x) = G(x-)$, then $F_{m(i,i)}(x) \to G(x)$.

(b) Tight implies that there exists $K(B)$ such that $\limsup_n P(X_n \leq B) \geq 1 - K(B)$, $K(B) \downarrow 0$ as $B \uparrow \infty$. Consider $F_{m(i,i)}(q) \to G(q)$ $\forall q$, which implies that $G(B) \geq 1 - K(B)$, so G puts 0 mass on $+\infty$.

□

Corollary 2.10. *Given $(X_n, 1 \leq n < \infty)$ and X ($\mathbb{R}$-valued RVs), suppose (X_n) is tight. Suppose that, whenever $X_{n_j} \xrightarrow{d}$ some Y as $j \to \infty$ for some (n_j), we have $Y \overset{d}{=} X$. Then, $X_n \xrightarrow{d} X$ as $n \to \infty$.*

Proof. By contradiction. If $X_n \not\to X$ in distribution, then there exists x_0, a continuity point of X, such that $P(X_n \leq x_0) \not\to P(X \leq x_0)$. $\exists \varepsilon > 0$ and $m_j \to \infty$ such that $\left|P(X_{n_j} \leq x) - P(X \leq x)\right| \geq \varepsilon$ for all j. Apply Helly 2.9 to (X_{n_j}): there exists a subsequence $X_{n_j} \xrightarrow{d}$ some Y. But $Y \overset{d}{=} X$ by hypothesis, so $\left|P(X_{n_j} \leq x) - P(X \leq x)\right| \to 0$, which is a contradiction. □

Lemma 2.11. *Suppose $EX = 0$, $EX^2 = 1$, and $EX^4 \leq K$. Then, there exists $c(K) > 0$, depending on K, such that $P(X > 0) \geq c(K)$.*

Proof. By contradiction. There exists K such that the statement is false. So, there exists X_n such that $EX_n = 0$, $EX_n^2 = 1$, $EX_n^4 \leq K$, but $P(X_n > 0) \leq 1/n$. Helly 2.9 implies that there exists a subsequence $X_{n_j} \xrightarrow{\text{d}}$ some X. So, $EX = 0$, $EX^2 = 1$, and $P(X > 0) = 0$, which is impossible. □

Lecture 3

January 24

3.1 Transforms

There are three variants of the same idea.

1. Let X take values in $\{0, 1, 2, \dots\}$. The **probability generating function** is

$$h_X(z) = \sum_{n=0}^{\infty} P(X = n) z^n = E z^X$$

for $0 \le z \le 1$.

2. If X takes values in $[0, \infty)$, the **Laplace transform** is $L_X(\theta) = Ee^{-\theta X} = \int_0^\infty e^{-\theta x} f_X(x)\,\mathrm{d}x$ if X has density $f_X(x)$. If X has distribution μ, then $L_X(\theta) = \int_0^\infty e^{-\theta x} \mu_X(\mathrm{d}x)$. The Laplace transform is finite for $0 \le \theta < \infty$.

3. For X, an arbitrary $\mathbb{R}$-valued random variable, the **characteristic function** (Fourier transform) is $\phi_X(t) = Ee^{itX} = E\cos(tX) + iE\sin(tX)$. If X has a density, then $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,\mathrm{d}x$.

Point. If $S = X_1 + X_2$ for independent X_1, X_2, then

$$\begin{aligned} h_S(z) &= h_{X_1}(z) h_{X_2}(z), \\ L_S(\theta) &= L_{X_1}(\theta) L_{X_2}(\theta), \\ \phi_S(t) &= \phi_{X_1}(t) \phi_{X_2}(t), \end{aligned}$$

since

$$\begin{aligned} Ee^{itS} &= E[e^{itX_1} e^{itX_2}] \\ &= (Ee^{itX_1})(Ee^{itX_2}) \\ \phi_S(t) &= \phi_{X_1}(t) \phi_{X_2}(t) \end{aligned}$$

by the product rule.

Notation. $t, x, y \in \mathbb{R}$. $z \in \mathbb{C}$, $z = x + iy$. $|z| = \sqrt{x^2 + y^2}$, $|z_1 z_2| = |z_1||z_2|$. $\left|e^{itx}\right| = 1$. For a $\mathbb{C}$-valued RV $Z = X + iY$, $EZ = EX + iEY$. $|EZ| \le E|Z|$. $\phi_X(t) = Ee^{itX}$, where $\phi_X : \mathbb{R} \to \mathbb{C}$. The modulus is

$$|\phi_X(t)| = \left|Ee^{itX}\right| \le E\left|e^{itX}\right| = 1.$$

$\phi_X(t+h) - \phi_X(t) = E[e^{i(t+h)X} - e^{itX}] = E[e^{itX}(e^{ihX} - 1)]$, so

$$|\phi_X(t+h) - \phi_X(t)| \le E\left[\left|e^{itX}\right| \cdot \left|e^{ihX} - 1\right|\right] = E\left|e^{ihX} - 1\right| = \psi(h),$$

say. As $h \downarrow 0$, then $e^{ihX} - 1 \to 0$. Use bounded convergence to see that $t \mapsto \phi_X(t)$ is uniformly continuous.

3.2 Inversion

Theorem 3.1 (Inversion Formulas)**.** *Let $\phi(t)$ be the CF of a PM μ.*

(a)

$$\mu(a,b) + \frac{1}{2}(\mu\{a\} + \mu\{b\}) = \lim_{T\to\infty} \frac{1}{2\pi} \underbrace{\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t)\,\mathrm{d}t}_{I(T)}, \qquad -\infty < a < b < \infty.$$

(b) If $\int_{-\infty}^{\infty} |\phi(t)|\,\mathrm{d}t < \infty$, then μ has a bounded continuous density

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\phi(t)\,\mathrm{d}t.$$

$$S(T) \overset{\text{def}}{=} \int_0^T \frac{\sin x}{x}\,\mathrm{d}x \to \frac{\pi}{2} \qquad \text{as } T \to \infty.$$

$$\frac{e^{-ita} - e^{-itb}}{it} = \int_a^b e^{-ity}\,\mathrm{d}y,$$

so the modulus is at most $b - a$.

Proof. By Fubini,

$$I(T) = \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot e^{itx}\,\mathrm{d}t \right) \mu(\mathrm{d}x).$$

The inner integral contains a term

$$\int_{-T}^{T} \frac{e^{it(x-a)}}{it}\,\mathrm{d}t = \int_{-T}^{T} \frac{\sin(t(x-a))}{t}\,\mathrm{d}t + \underbrace{\frac{1}{i}\int_{-T}^{T} \frac{\cos(t(x-a))}{t}\,\mathrm{d}t}_{=0 \text{ by symmetry}},$$

since $e^{it} = \cos t + i \sin t$. The first term is

$$\begin{aligned} \int_{-T}^{T} \frac{\sin(\theta t)}{t}\,\mathrm{d}t = 2\int_0^T \frac{\sin(\theta t)}{t}\,\mathrm{d}t &= 2S(\theta T), && \theta > 0, \\ &= 2\,\mathrm{sgn}(\theta) \cdot S(T|\theta|) = R(T,\theta), && -\infty < \theta < \infty, \\ &\to \pi\,\mathrm{sgn}(\theta) && \text{as } T \to \infty. \end{aligned}$$

Here,

$$\mathrm{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

Therefore,

$$I(T) = \int_{-\infty}^{\infty} (R(x-a,T) - R(x-b,T))\mu(\mathrm{d}x).$$

The integrand is bounded by $2\sup_{\theta,T} R(\theta T) \equiv K < \infty$. Let $T \to \infty$.

$$\lim_{T\to\infty} I(T) = \int_{-\infty}^{\infty} \chi_{a,b}(x)\mu(\mathrm{d}x),$$

where

$$\chi_{a,b}(x) = \begin{cases} 0, & x < a \text{ or } x > b, \\ 2\pi, & a < x < b, \\ \pi, & x = a \text{ or } x = b. \end{cases}$$

(Check.) This is (a).

In case (b), the integral

$$\int_{-\infty}^{\infty} \underbrace{\frac{e^{-ita} - e^{-itb}}{it}}_{|\cdot| \le b-a} \phi(t)\,\mathrm{d}t$$

is absolutely convergent. Use (a):

$$\mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\} \le \frac{b-a}{2\pi}\int_{-\infty}^{\infty} |\phi(t)|\,\mathrm{d}t.$$

Note that if $(a', b') \downarrow \{x\}$, then $\mu\{x\} = 0\ \forall x$. By (a),

$$\begin{aligned} \mu(a,b) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_a^b e^{-ity}\,\mathrm{d}y\right)\phi(t)\,\mathrm{d}t \\ &= \int_a^b \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\phi(t)\,\mathrm{d}t\right)\mathrm{d}y \end{aligned}$$

by Fubini. The integrand is the density function $f(y)$ for μ, and

$$f(y) \le \frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi(t)|\,\mathrm{d}t. \qquad \square$$

Comments.

1. If $\phi_\mu(t) \equiv \phi_\nu(t)\ \forall t$, then $\nu = \mu$. (*Uniqueness*)
2. In principle, we can calculate the distribution of $S_n = X_1 + X_2 + \cdots + X_n$ for independent X_i using $\phi_{S_n} = \prod_{i=1}^n \phi_{X_i}(t)$.

Example 3.2. If X has Normal$(0,\sigma^2)$ distribution, then $\phi_X(t) = \exp(-\sigma^2 t^2/2)$.

So, if X_1, X_2 are independent Normal$(0,\sigma_i^2)$, then $S = X_1 + X_2$ has

$$\begin{aligned} \phi_S(t) &= \phi_{X_1}(t)\phi_{X_2}(t) = \exp(-\sigma_1^2 t^2/2 - \sigma_2^2 t^2/2) = \exp(-(\sigma_1^2+\sigma_2^2)/2) \\ &= \text{CF of Normal}(0, \sigma_1^2 + \sigma_2^2). \end{aligned}$$

Example 3.3. X has the Exponential(1) distribution, $f(x) = e^{-x}$, $x > 0$.

$$\phi_X(t) = \int_0^\infty e^{itx} e^{-x}\,\mathrm{d}x = \int_0^\infty e^{-(1-it)x}\,\mathrm{d}x = \frac{1}{1-it}$$

For $c > 0$, $\phi_{cX}(t) = \phi_X(ct) = Ee^{ictX}$.

Example 3.4. Y has density

$$f_Y(y) = \frac{1}{2}e^{-|y|}, \qquad -\infty < y < \infty.$$

Since

$$\mu_Y = \frac{1}{2}\mu_X + \frac{1}{2}\mu_{-X},$$

which implies that

$$\begin{aligned} \phi_Y(t) &= \frac{1}{2}\phi_X(t) + \frac{1}{2}\phi_{-X}(t) = \frac{1}{2}(\phi_X(t) + \phi_X(-t)) \\ &= \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{(1-it)(1+it)} = \frac{1}{1+t^2}. \end{aligned}$$

3.3 Parseval Identity

Theorem 3.5 (Parseval Identity)**.** *Let μ and ν be PMs with CFs ϕ_μ and ϕ_ν. Then*

$$\int_{-\infty}^\infty \phi_\nu(t)\mu(\mathrm{d}t) = \int_{-\infty}^\infty \phi_\mu(t)\nu(\mathrm{d}t).$$

Proof. Take X, Y independent, $\mathrm{dist}(X) = \mu$, $\mathrm{dist}(Y) = \nu$.

$$E[e^{iXY} \mid Y = y] = Ee^{iyX} = \phi_\mu(y),$$

so

$$E[e^{iXY}] = E\phi_\mu(Y) = \int_{-\infty}^\infty \phi_\mu(y)\nu(\mathrm{d}y) = \text{right side}.$$

Also,

$$E[e^{iXY}] = E[E[e^{iYX} \mid X]] = \text{left side}.$$ □

By choice of "simple" ν, we get general identities between μ and $\phi(\mu)$.

Example 3.6. ν is uniform on $[-c, c]$.

$$\phi_\nu(t) = \frac{\sin(ct)}{ct}.$$

For any μ,

$$\frac{1}{2c}\int_{-c}^{c} \phi_\mu(t)\,\mathrm{d}t = \int_{-\infty}^\infty \frac{\sin(ct)}{ct}\mu(\mathrm{d}t).$$

Example 3.7. Take ν to be Normal$(0, \sigma^2)$. For any μ,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/(2\sigma^2)} \phi_\mu(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} e^{-\sigma^2 t^2/2} \mu(\mathrm{d}t).$$

Lecture 4

January 26

4.1 Applications of Inversion Formula

Inversion Formula: If a PM μ has CF ϕ such that $\int_{-\infty}^{\infty} |\phi(t)| \, \mathrm{d}t < \infty$, then μ has a bounded continuous density

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) \, \mathrm{d}t.$$

In general, $\phi_{aW}(t) = \phi_W(at)$.

Corollary 4.1. *Given a PM μ with CF ϕ and density f, suppose ϕ is $\mathbb{R}$-valued, $\phi \geq 0$, and*

$$\int_{-\infty}^{\infty} \phi(t) \, \mathrm{d}t < \infty.$$

Then,

$$g(x) \overset{\text{def}}{=} \frac{\phi(x)}{2\pi f(0)}$$

*is a density function, and its CF is $f(t)/f(0)$. Here, f and g are called **dual pairs**.*

Proof. By the inversion formula,

$$\frac{f(y)}{f(0)} = \int_{-\infty}^{\infty} e^{-ity} \underbrace{\frac{\phi(t)}{2\pi f(0)}}_{g(t)} \, \mathrm{d}t = \text{CF of } g(x).$$

For $y = 0$,

$$1 = \int_{-\infty}^{\infty} \underbrace{\frac{\phi(t)}{2\pi f(0)}}_{=g(t)} \, \mathrm{d}t.$$

□

Example 4.2 (Last Class)**.** If

$$f(x) = \frac{1}{2} e^{-|x|},$$

then

$$\phi(t) = \frac{1}{1+t^2}.$$

The dual is

$$g(x) = \frac{\phi(x)}{\pi} = \frac{1}{\pi(1+x^2)},$$

the standard Cauchy distribution, and this has CF $f(t)/f(0) = e^{-|t|}$, for $-\infty < t < \infty$. Write W for a RV with the standard Cauchy distribution. Take $W_1, W_2, \ldots$, IID copies of W.

$$\phi_{W_1+W_2+\cdots+W_n}(t) = (e^{-|t|})^n = e^{-n|t|} = \phi_{nW}(t).$$

By uniqueness, $\sum_{i=1}^n W_i \stackrel{\text{d}}{=} nW$, so

$$\frac{1}{n}\sum_{i=1}^n W_i \stackrel{\text{d}}{=} W.$$

The LLN does not hold. $E|W| = \infty$.

4.2 Another Proof of Inversion

Exercise: If $Y_n \xrightarrow{\text{d}} c$, then $Y_n \to c$ in probability. If $Y_n \xrightarrow{\text{d}} c$, then $X + Y_n \xrightarrow{\text{d}} X + c$ (for any X).

2nd Proof of Inversion Formula. Take X with $\text{dist}(X) = \mu$. Take $Z_\sigma \stackrel{\text{d}}{=} \text{Normal}(0, \sigma^2)$, independent of X. $X + Z_\sigma \xrightarrow{\text{d}} X$ as $\sigma \downarrow 0$. Note: $X + Z_\sigma$ has density

$$\begin{aligned} f_{X+Z_\sigma}(0) &= \int_{-\infty}^{\infty} f_{Z_\sigma}(t)\mu(\text{d}t) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2/(2\sigma^2)}\mu(\text{d}t). \end{aligned}$$

Use Parseval's Identity for the normal distribution, $\theta = 1/\sigma$.

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-t^2\sigma^2/2}\phi(t)\,\text{d}t$$

Then, $\phi_{X-x}(t) = e^{-ixt}\phi_X(t)$. Applying the above to $X - x$ instead of X, we have

$$f_{X+Z_\sigma}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-t^2\sigma^2/2}e^{-itx}\phi(t)\,\text{d}t.$$

Let $\sigma \downarrow 0$. Appeal to bounded convergence.

$$\lim_{\sigma\downarrow 0} f_{X+Z_\sigma}(x) = \underbrace{\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,\text{d}t}_{f(x),\text{ say}}.$$

Final detail:

$$P(a \le X \le b) = \lim_{\sigma\downarrow 0} P(a \le X + Z_\sigma \le b)$$

at continuity points a, b of X. The limit is $\int_a^b f(x)\,\mathrm{d}x$. This is enough to prove that f is the density of X. □

4.3 Continuity Theorem

Theorem 4.3 (Continuity Theorem)**.** *Suppose X_n has CF ϕ_n.*

(a) If $X_n \xrightarrow{\mathrm{d}} X_\infty$, then $\phi_n(t) \to \phi_\infty(t)$, for each t.

(b) Suppose $\lim_{n\to\infty} \phi_n(t)$ exists ($= \phi(t)$, say), for each t. If either

(1) ϕ is a CF, or

(2) $\phi(t) \to 1$ as $t \to 0$, or

(3) $(X_n, n \geq 1)$ are tight,

then $X_n \xrightarrow{\mathrm{d}} X_\infty$, and X_∞ has CF ϕ.

Proof. (a) $X_n \xrightarrow{\mathrm{d}} X_\infty$ implies that $Eg(X_n) \to Eg(X_\infty)$ for bounded, continuous g. Take $g(x) = e^{itx}$, which shows that $\phi_n(t) \to \phi_\infty(t)$ as $n \to \infty$, for t fixed.

(b) Suppose (3). Helly's Theorem implies that there exists a subsequence $X_{n_j} \xrightarrow{\mathrm{d}}$ some $\hat{X}$. By (a) and the hypothesis, $\hat{X}$ has CF ϕ. By a previous lemma (every convergent subsequence has the same limit distribution) implies that the whole sequence $X_n \xrightarrow{\mathrm{d}} \hat{X}$ with CF ϕ, which is a proof of (b).

Claim: (1) $\implies$ (2). A CF ϕ is continuous, with $\phi(0) = 1$.

We need to prove that (2) and the hypothesis imply (3).

Fix K, put $c = 2/K$. (Trick)

$$\begin{aligned}
P(|X_n| \geq K) &\leq E2\left(1 - \frac{1}{c|X_n|}\right) 1_{(|X_n| \geq K)} \\
&\leq 2E\left(1 - \frac{\sin(c|X_n|)}{c|X_n|}\right) 1_{(|X_n| \geq K)} \\
&\leq 2E\left(1 - \frac{\sin(c|X_n|)}{c|X_n|}\right),
\end{aligned}$$

because $\sin y \leq 1$ and

$$\frac{\sin y}{y} \leq 1.$$

Use the Parseval Identity for the Uniform$[-c, c]$ distribution.

$$P(|X_n| \geq K) \leq 2\left(1 - \frac{1}{2c}\int_{-c}^{c} \phi_n(t)\,\mathrm{d}t\right) = \frac{1}{c}\int_{-c}^{c} (1 - \phi_n(t))\,\mathrm{d}t$$

Use bounded convergence as $n \to \infty$.

$$\limsup_n P(|X_n| \geq K) \leq \frac{1}{c}\int_{-c}^{c} (1 - \phi(t))\,\mathrm{d}t$$

On the LHS, we can take the limit as $K \uparrow \infty$. On the RHS, we can take the limit as $c \downarrow 0$. Then, the RHS is 0 by (2), which gives tightness. □

4.4 CFs & Moments

$$e^{itx} = \sum_{m=0}^{\infty} \frac{(itx)^m}{m!}$$

This suggests that the CF ϕ of X is

$$\phi_X(t) = \sum_{m=0}^{\infty} \frac{E(itX)^m}{m!} = 1 + itEX - \frac{t^2}{2}EX^2 + \cdots$$

However, EX^m may be infinite.

Lemma 4.4 (Technical Lemma, Durrett 3.3.7)**.**

$$\left| e^{iy} - \sum_{m=0}^{n} \frac{(iy)^m}{m!} \right| \le \min\left(\frac{|y|^{n+1}}{(n+1)!}, \frac{2|y|^n}{n!} \right)$$

Apply this to $y = tX$.

$$\begin{aligned} \left| \phi_X(t) - \sum_{m=0}^{n} \frac{E(itX)^m}{m!} \right| &\le E \min\left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right) \\ &= \frac{|t|^n}{n!} \underbrace{E \min\left(\frac{|t||X|^{n+1}}{n+1}, 2|X|^n \right)}_{Z_t} \end{aligned}$$

Corollary 4.5. *Suppose* $E|X|^n < \infty$. *Then,*

$$\phi_X(t) = \sum_{m=0}^{n} \frac{E(itX)^m}{m!} + o(|t|^n) \qquad \text{as } t \to 0.$$

Proof. $Z_t \to 0$ a.s. as $t \to 0$, and is dominated by $2|X|^n$ which is integrable. Hence, $EZ_t \to 0$ as $t \to 0$. □

Lecture 5

January 31

5.1 Characteristic Function Proofs

Convergence Theorem: X_n has CF ϕ_n. If $\phi_n(t) \to \phi_\infty(t)$ as $n \to \infty$, for each t, if $\phi_\infty(t)$ is the CF of some X_∞, then $X_n \xrightarrow{\mathrm{d}} X_\infty$.

Suppose $E|X|^n < \infty$. Then

$$\left| \phi_X(t) - \sum_{m=1}^{n} \frac{E(itX)^m}{m!} \right| = o(|t|^n) \qquad \text{as } t \to 0.$$

Theorem 5.1 (Weak Law of Large Numbers). *Let $X_1, X_2, \dots$ be IID with $EX = \theta$, $S_n = \sum_{i=1}^n X_i$, then $S_n/n \to \theta$ in distribution, and hence convergence in probability.*

Proof. The PM σ_θ has CF $e^{i\theta t}$. It is enough to prove $\phi_{S_n/n}(t) \to e^{i\theta t}$ as $n \to \infty$, for a fixed t. Since $\phi_{S_n}(t) = (\phi_X(t))^n$,

$$\phi_{S_n/n}(t) = \left(\phi_X\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{n(\phi_X(t/n) - 1)}{n}\right)^n.$$

If $z_n \to z \in \mathbb{C}$, then $(1 + z_n/n)^n \to e^z$. It is enough to prove

$$\underbrace{n\left(\phi_X\left(\frac{t}{n}\right) - 1\right)}_{\text{Left}} \to i\theta t.$$

The bound for $n = 1$ gives $|\phi_X(s) - (1 + is\theta)| = o(|s|)$. Apply the bound with $s = t/n$. Then, we know

$$\text{Left} = n\left(i\frac{t}{n}\theta + o\left(\frac{|t|}{n}\right)\right) = it\theta + n \cdot o\left(\frac{|t|}{n}\right) \to it\theta. \qquad \square$$

Remarks. The proof shows that

$$\phi'_X(0) = \theta \tag{5.1}$$

is sufficient for the WLLN 5.1.

Fact. In fact, (5.1) is also necessary. The property $EX = \theta$ implies $\phi'_X(0) = \theta$, but *not* conversely.

5.2 Central Limit Theorems

Theorem 5.2 (IID Central Limit Theorem)**.** *Let* $(X_i, i \geq 1)$ *be IID,* $EX = \mu$*,* $\text{var}(X) = \sigma^2 < \infty$*. Then,*

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{\text{d}} \text{Normal}(0, \sigma^2).$$

Proof. WLOG take $\mu = 0$. It is enough to show

$$\underbrace{\phi_{S_n/\sqrt{n}}(t)}_{\text{Left}} \to \exp\left(-\frac{\sigma^2 t^2}{2}\right).$$

Also,

$$\phi_{S_n/\sqrt{n}}(t) = \left(\phi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{n(\phi_X(t/\sqrt{n}) - 1)}{n}\right)^n.$$

It is enough to show $n(\phi_X(t/\sqrt{n}) - 1) \to \sigma^2 t^2/2$. The bound for $n = 2$ and $EX = 0$ is

$$\left|\phi_X(s) - \left(1 - \frac{s^2\sigma^2}{2}\right)\right| = o(s^2).$$

Then, with $s = t/\sqrt{n}$,

$$\text{Left} = n\left(\frac{t^2}{n}\frac{\sigma^2}{2} + o\left(\frac{t^2}{n}\right)\right) = \frac{t^2\sigma^2}{2} + n \cdot o\left(\frac{t^2}{n}\right) \to \frac{t^2\sigma^2}{2}.$$

□

Theorem 5.3 (Lindeberg's Theorem)**.** *For each* n*, let* $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ *be independent,* $EX_{n,m} = 0$*,* $\text{var}\, X_{n,m} = \sigma^2_{n,m} < \infty$*. Write* $S_n = \sum_{m=1}^n X_{n,m}$*,* $\sigma_n^2 = \sum_{m=1}^n \sigma^2_{n,m} = \text{var}(S_n)$*,* $ES_n = 0$*. Suppose*

(i) $\sigma_n^2 \to \sigma^2 < \infty$ *as* $n \to \infty$*,*

(ii) $\lim_{n\to\infty} \sum_{m=1}^n E[X^2_{n,m} 1_{(|X_{n,m}| > \varepsilon)}] = 0$*, for each* $\varepsilon > 0$*. This is known as the* ***Lindeberg condition****: UAN = uniformly asymptotically negligible.*

Then, $S_n \xrightarrow{\text{d}} \text{Normal}(0, \sigma^2)$*.*

Proof. $\phi_{n,m}(t)$ is the CF of $X_{n,m}$. The more precise bound is

$$\left|\phi_{n,m}(t) - \left(1 - \frac{t^2\sigma^2_{n,m}}{2}\right)\right| \leq E\min\left(\frac{|tX_{n,m}|^3}{6}, |tX_{n,m}|^2\right)$$

Cheap: If $|x| \leq \varepsilon$, then $|x|^3 \leq \varepsilon x^2$.

$$\leq \frac{\varepsilon|t|^3}{6} E[X^2_{n,m}] + |t|^2 E[X^2_{n,m} 1_{(|X_{n,m}| \geq \varepsilon)}].$$

Then,

$$\limsup_n \underbrace{\sum_{m=1}^{n} \left| \phi_{n,m}(t) - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \right|}_{=B_n(t)} \leq \frac{\varepsilon|t|^3}{6} \cdot \sigma^2 + 0, \tag{5.2}$$

by hypothesis (ii). Let $\varepsilon \downarrow 0$. Then, the summation goes to 0 as $n \to \infty$.

Claim.

(a) $\max_m \sigma_{n,m}^2 \to 0$ as $n \to \infty$.

(b) $\sum_m \sigma_{n,m}^4 \to 0$ as $n \to \infty$.

Proof.

(a)

$$\begin{aligned}\sigma_{n,m}^2 &= EX_{n,m}^2 1_{(|X_{n,m}| \geq \varepsilon)} + EX_{n,m}^2 1_{(|X_{n,m}| \leq \varepsilon)} \\ &\leq \sum_m EX_{n,m}^2 1_{(|X_{n,m}| \geq \varepsilon)} + \varepsilon^2,\end{aligned}$$

so

$$\limsup_n \max_m \sigma_{n,m}^2 \leq 0 + \varepsilon^2,$$

by (ii). Let $\varepsilon \downarrow 0$.

(b)

$$\sum_m \sigma_{n,m}^4 \leq \left(\max_m \sigma_{n,m}^2\right) \sum_m \sigma_{n,m}^2 \to 0,$$

by (i).

$\phi_{S_n}(t) = \prod_{m=1}^n \phi_{n,m}(t)$. By 5.4,

$$\left| \phi_{S_n}(t) - \prod_{m=1}^{n} \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \right| \leq B_n(t) \to 0$$

by (5.2), using Claim (a).

So, it is enough to prove

$$\prod_{i=1}^{n} \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \to \exp\left(-\frac{t^2\sigma^2}{2}\right).$$

This will follow from 5.5 applied to $a_{n,m} = \sigma_{n,m}^2$. Assumption (i) is (i), while (ii) is Claim (b). ☐

Lemma 5.4. *If $w_i, z_i \in \mathbb{C}$, $|w_i| \leq 1$, $|z_i| \leq 1$, then $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \leq \sum_i |w_i - z_i|$.*

Proof.

$$\begin{aligned}|z_1 z_2 \cdots z_i w_{i+1} \cdots w_n - z_1 \cdots z_{i+1} w_{i+2} \cdots w_n| &= |(z_{i+1} - w_{i+1}) \cdot A| \\ &\le |z_{i+1} - w_{i+1}|,\end{aligned}$$

where $|A| \le 1$. □

Lemma 5.5. *Let $a_{n,m} \in \mathbb{R}$. If*

(i) $\sum_m a_{n,m} \to a$ *as* $n \to \infty$,

(ii) $\sum_m a_{n,m}^2 \to 0$.

Then, $\prod_{m=1}^n (1 - a_{n,m}) \to e^{-a}$.

Proof. We know that $\max_m |a_{n,m}| \to 0$ by (ii). Since $|\log(1-x) + x| \le Cx^2$ for $|x| \le 1/2$,

$$\begin{aligned}\left|\sum_{m=1}^n \log(1 - a_{n,m}) + \sum_{m=1}^n a_{n,m}\right| &\le C \sum_m a_{n,m}^2 \qquad \text{for large } n, \\ &\to 0 \qquad \text{as n} \to \infty.\end{aligned}$$

Hence, $\log \prod_{m=1}^n (1 - a_{n,m}) \to -a$. □

Theorem 5.6 (Equivalent Form of Lindeberg CLT)**.** *For each n, let $X_{n,m}$, $1 \le m \le n$, be independent, $EX_{n,m} = 0$. Let $S_n = \sum_{m=1}^n X_{n,m}$ and $s_n^2 = \mathrm{var}(S_n) = \sum_{i=1}^n \mathrm{var}(X_{n,m})$. Suppose*

$$\sum_{m=1}^n E\left[\frac{X_{n,m}^2}{s_n^2} 1_{(|X_{n,m}| \ge \varepsilon s_n)}\right] \to 0 \qquad \text{as } n \to \infty.$$

Then, $S_n / s_n \xrightarrow{\text{d}} \mathrm{Normal}(0,1)$.

This is the previous theorem 5.3 applied with $\hat{X}_{n,m} = X_{n,m}/s_n$. Now, it looks more like the IID version.

Lecture 6

February 2

6.1 Lindeberg Theorem

Restatement of the Lindeberg Theorem without prior rescaling:

Lindeberg Theorem: For each n, assume that the $(X_{n,m}, 1 \le m \le n)$ are independent, $EX_{n,m} = 0$, $s_n^2 = \sum_{m=1}^n \text{var}(X_{n,m}) < \infty$, $S_n = \sum_{m=1}^n X_{n,m}$ (so $ES_n = 0$). If

$$\sum_{m=1}^n E\left[\frac{X_{n,m}^2}{s_n^2} 1_{(|X_{n,m}| \ge \varepsilon s_n)}\right] \to 0$$

as $n \to \infty$, for each $\varepsilon > 0$ (UAN), then $S_n / s_n \xrightarrow{\text{d}} \text{Normal}(0,1)$.

This is the previous version applied to $X_{n,m}/s_n$.

Corollary 6.1. *Suppose* $(Y_1, Y_2, \ldots)$ *are independent,* $EY_i = 0$. *Suppose* $s_n^2 = \sum_{i=1}^n \text{var}(Y_i) < \infty$. *If* $|Y_i| \le M$ *a.s. and if* $s_n \to \infty$ *as* $n \to \infty$, *then*

$$\frac{1}{s_n} \sum_{i=1}^n Y_i \xrightarrow{\text{d}} \text{Normal}(0,1).$$

Proof. Apply the Lindeberg Theorem 5.3 to $X_{n,m} = Y_m$. The event $|X_{n,m}| \ge \varepsilon s_n$ can only happen if $M \ge \varepsilon s_n$, that is, $s_n \le M/\varepsilon$. The event has probability 0 for large n, which implies UAN. □

Corollary 6.2. *In 5.3, we may replace UAN by* ***Lyapunov's condition****:* $\exists \delta > 0$ *such that*

$$L_n \stackrel{\text{def}}{=} \frac{\sum_{m=1}^n E|X_{n,m}|^{2+\delta}}{s_n^{2+\delta}} \to 0 \qquad \text{as } n \to \infty.$$

Proof.

$$x^2 1_{(|x| \ge \varepsilon s_n)} \le \frac{|x|^{2+\delta}}{|\varepsilon s_n|^\delta} = x^2 \left(\frac{|x|}{\varepsilon s_n}\right)^\delta \qquad \forall x$$

So,

$$\sum_{m=1}^{n} E\left[\frac{X_{n,m}^2}{s_n^2}1_{(|X_{n,m}|\geq \varepsilon s_n)}\right] \leq \frac{\sum_{m=1}^{n} E|X_{n,m}|^{2+\delta}}{s_n^2(\varepsilon s_n)^\delta} = \frac{L_n}{\varepsilon^\delta} \to 0 \qquad \text{as } n\to\infty,$$

which checks UAN. □

Corollary 6.3. *Let $(Y_i, i\geq 1)$ be independent, $EY_i = 0$. Suppose $\text{var}(Y_i)\to\sigma^2<\infty$ as $i\to\infty$. Suppose $\exists\delta>0$ such that $M := \sup_i E|Y_i|^{2+\delta}<\infty$. Then*

$$\frac{\sum_{i=1}^n Y_i}{\sigma\sqrt{n}} \xrightarrow{\text{d}} \text{Normal}(0,1).$$

Proof. Set $X_{n,m} = Y_m$ and check Lyapunov's condition.

$$s_n^2 = \text{var}\left(\sum_{i=1}^n Y_i\right) \sim n\sigma^2,$$

$$L_n \leq \frac{Mn}{s_n^{2+\delta}} \sim \frac{Mn}{\sigma^{2+\delta}n^{1+\delta/2}} = \frac{M}{\sigma^{2+\delta}}n^{-\delta/2} \to 0 \qquad \text{as } n\to\infty.$$

We conclude

$$\frac{\sum_{i=1}^n Y_i}{s_n} \xrightarrow{\text{d}} \text{Normal}(0,1)$$

and $s_n \sim \sigma\sqrt{n}$. □

Corollary 6.4. *If $(X_i, i\geq 1)$ are independent, $|X_i|\leq A$, $\mu_i = EX_i$, $\sigma_i^2 = \text{var}(X_i)<\infty$, and if $S_n = \sum_{i=1}^n X_i \xrightarrow[a.s.]{}$ some S_∞, which is finite, then $\sum_{i=1}^n \mu_i$ and $\sum_{i=1}^n \sigma_i^2$ converge to a finite limit.*

Proof. By contradiction. Suppose $s_n = \sum_{i=1}^n \sigma_i^2 \to\infty$ as $n\to\infty$. We can apply 6.1 to $X_i - \mu_i$.

$$\frac{1}{s_n}\sum_{i=1}^n (X_i-\mu_i) \xrightarrow{\text{d}} \text{Normal}(0,1)$$

$$\frac{S_n}{s_n} - \frac{1}{s_n}\sum_{i=1}^n \mu_i \xrightarrow{\text{d}} \text{Normal}(0,1)$$

The first term converges in distribution to 0. The second term is a constant. The LHS can only converge in distribution to a constant. This contradiction implies that $s_n\to s_\infty<\infty$.

By 205A, this implies $\sum_{i=1}^n (Y_i-\mu_i)$ converges a.s., so $S_n - \sum_{i=1}^n \mu_i$ converges a.s. Since $S_n\to S_\infty$ a.s., this implies $\sum_{i=1}^n \mu_i$ converges a.s. □

6.2 3 Series Theorem

Theorem 6.5 (Classical "3 Series Theorem")**.** *Suppose (X_i) are independent. Then $\sum_{i=1}^n X_i$ converges a.s. to a finite limit if and only if, for some A,*

1. *$\sum_i P(|X_i|\geq A)<\infty$,*
2. *For $Y_i = X_i 1_{(|X_i|\leq A)}$, we have $\sum_{i=1}^n EY_i$ converges,*

3. $\sum_i \operatorname{var}(Y_i) < \infty$.

Proof. "If": We implicitly proved this part in 205A.

For "only if", assume $\sum_{i=1}^n X_i$ converges. The events $\{|X_n| > A\}$ occur only finitely often. By Borel-Cantelli 2, $\sum_i P(|X_n| > A) < \infty$. Also, $\sum_i Y_i$ converges a.s. Apply 6.4 to (Y_i): $\sum_i EY_i$ and $\sum_i \operatorname{var}(Y_i)$ converge. □

6.3 Classical Theory: "Infinitely Divisible Distributions"

What are all possible limits

$$\frac{\sum_{i=1}^n X_i - a_n}{b_n} \xrightarrow{\mathrm{d}} Y?$$

See Durrett 3.7 and 3.8.

6.4 Poisson Limits

For PMs μ_1, μ_2 on measurable $(S, \mathcal{S})$,

$$\|\mu_2 - \mu_1\| \stackrel{\text{def}}{=} \sup_{A \in \mathcal{S}} |\mu_1(A) - \mu_2(A)|.$$

This is the **variational distance**.

(Easy) If S is countable, then

$$\|\mu_1 - \mu_2\| = \frac{1}{2} \sum_{s \in \mathcal{S}} |\mu_1\{s\} - \mu_2\{s\}|.$$

If $S = \mathbb{R}$ and μ_i has density f_i, then

$$\|\mu_1 - \mu_2\| = \frac{1}{2} \int_{-\infty}^{\infty} |f_1(x) - f_2(x)| \, \mathrm{d}x.$$

Know. Class 2 says for *countable* S,

$$\mu_n \to \mu_\infty \text{ weakly} \iff \|\mu_n - \mu_\infty\| \to 0.$$

Let $f_\infty = 1$ and f_n be a sinusoid on $[0, 1]$ with period $1/n$. Here, $\mu_n \to \mu_\infty$ weakly but $\|\mu_n - \mu_\infty\| \not\to 0$.

Lemma 6.6 (Easy?). *(a) If* $\operatorname{dist}(X_i) = \mu_i$, $i = 1, 2$, *then* $P(X_1 \neq X_2) \geq \|\mu_1 - \mu_2\|$.

(b) Given μ_1, μ_2, *there exist* (X_1, X_2) *with* $\operatorname{dist}(X_i) = \mu_i$ *and* $P(X_1 \neq X_2) = \|\mu_1 - \mu_2\|$.

This uses a **coupling** argument.

"X is Bernoulli(p)" means $P(X = 1) = p$, $P(X = 0) = 1 - p$.

Theorem 6.7 (Le Cam's Theorem)**.** *Suppose* $(X_r, 1 \leq r \leq n)$ *are independent* Bernoulli(p_r). *Write* $S = \sum_{r=1}^n X_r$, $\lambda = \sum_{r=1}^n p_r$. *Then* $\|\operatorname{dist}(S) - \operatorname{Poisson}(\lambda)\| \leq \sum_{r=1}^n p_r^2$.

Proof. Given p (small), we want (X, Y), $X \stackrel{d}{=} \text{Bernoulli}(p)$, $Y \stackrel{d}{=} \text{Poisson}(p)$, and $P(X \neq Y)$ is small. Define

$$\begin{aligned} P(X = 0, Y = 0) &= 1 - p, \\ P(X = 1, Y = y) &= \frac{e^{-p} p^y}{y!}, \qquad y \geq 1, \\ P(X = 1, Y = 0) &= e^{-p} - (1 - p). \end{aligned}$$

(Check that this works.)

$$\begin{aligned} P(Y \neq X) &= e^{-p} - (1 - p) + P(Y \geq 2) \\ &= e^{-p} - (1 - p) + (1 - e^{-p} - pe^{-p}) \\ &= p(1 - e^{-p}) \leq p^2 \end{aligned}$$

For each r, construct the coupled pair $(\hat{X}_r, \hat{Y}_r)$ for $p = p_r$. Let the pairs be independent as r varies.

$$\left\| \text{dist}\left(\sum_{r=1}^n \hat{X}_r\right) - \text{dist}\left(\sum_{r=1}^n \hat{Y}_r\right) \right\| \leq P\left(\sum_{i=1}^n \hat{X}_i \neq \sum_{i=1}^n \hat{Y}_i\right) \leq \sum_{i=1}^n P(\hat{X}_i \neq \hat{Y}_i) \leq \sum_{i=1}^n p_r^2$$

Then, $\text{dist}(\sum_{r=1}^n \hat{X}_r)$ has the same distribution as S and $\text{dist}(\sum_{r=1}^n \hat{Y}_r) = \text{Poisson}(\sum p_r = \lambda)$. □

Lecture 7

February 7

7.1 Method of Moments

Say that $\text{dist}(X)$ is **determined by its moments** if $E|X|^k < \infty\ \forall k$ and for all Y, if $EY^k = EX^k\ \forall k$, then $Y \stackrel{\text{d}}{=} X$.

Lemma 7.1 (Method of Moments)**.** *To prove $X_n \xrightarrow{\text{d}} X$, it is sufficient to prove*

(i) X is determined by its moments,

(ii) $EX_n^k \to EX^k$ as $n \to \infty$, for each $k \geq 1$.

Proof. EX_n^2 is bounded, so $(X_n, n \geq 1)$ is tight. If $X_{j_n} \xrightarrow{\text{d}}$ some Y, then $EY^k = EX^k$ implies that $Y \stackrel{\text{d}}{=} X$. By the old "subsequence trick" lemma, $X_n \xrightarrow{\text{d}} X$. □

Not all distributions are determined by moments.

Theorem 7.2 (Durrett Theorem 3.3.11)**.** *If*

$$\limsup_{\substack{k\to\infty,\\ k \text{ even}}} \frac{(EX^k)^{1/k}}{k} < \infty, \tag{7.1}$$

then $\text{dist}(X)$ *is determined by its moments.*

Consider $X \stackrel{\text{d}}{=} \text{Normal}(0,1)$.

$$EX^{2m} = \frac{(2m)!}{2^m m!}$$

Also $(n!)^{1/n} \sim n/e$ as $n \to \infty$. Set $k = 2m$.

$$\limsup \frac{2m/e}{2^{1/2}(m/e)^{1/2} 2m} \sim m^{-1/2} \to 0.$$

So, (7.1) holds for Normal$(0,1)$.

7.2 Application to Poisson Limits

It is easy to check (7.1).

Notation. $x(x-1)(x-2)\cdots(x-k+1) = [x]_k$. $[x]_1 = x$, $[x]_2 = x(x-1)$, etc.

For $X \geq 0$, integer-valued,

$$E[X]_k = E\left[\frac{X!}{(X-k)!}1_{(X\geq k)}\right].$$

For $X \stackrel{\mathrm{d}}{=} \text{Poisson}(\lambda)$,

$$\begin{aligned} E[X]_k &= \sum_{m=k}^{\infty} \frac{m!}{(m-k)!}\frac{e^{-\lambda}\lambda^m}{m!} \underbrace{=}_{m=k+i} e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\lambda^k \\ &= \lambda^k. \end{aligned}$$

x^k can be written as a linear combination of $[x]_1, [x]_2, \ldots, [x]_k$.

Corollary 7.3 (Method of Moments Adapted to Poisson)**.** *For positive integer-valued X_n, to prove $X_n \xrightarrow{\mathrm{d}} \text{Poisson}(\lambda)$, it is enough to prove $E[X_n]_k \to \lambda^k$ as $n \to \infty$, for all k.*

Consider a counting RV $X = \sum_i 1_{(A_i)}$ for events A_i. $[X]_k = \sum_{(i_1,\ldots,i_k)} 1_{A_{i_1}}1_{A_{i_2}}\cdots 1_{A_{i_k}}$ over ordered distinct $(i_1,\ldots,i_k)$. Then, $E[X]_k = \sum_{(i_1,\ldots,i_k)} P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})$.

Example 7.4. Put M balls at random (uniformly, independently) into N boxes. Let $X = X_{M,N}$ be the number of empty boxes, $\sum_{i=1}^N A_i$, where A_i is the event "box i is empty".

$$E[X]_k = [N]_k P(A_1\cap A_2\cap\cdots\cap A_k) = [N]_k\left(1-\frac{k}{M}\right)^M.$$

Consider $N, M \to \infty$ in some way, and we want to prove $X_{N,M} \xrightarrow{\mathrm{d}} \text{Poisson}(e^{-c})$. We must prove $E[X]_k \to e^{-ck}$. Asymptotically, we want

$$N^k \exp\left(-k\frac{M}{N}\right) \to e^{-ck}.$$

This is true, provided $M = o(N^2)$. Hence, we want to show $N\exp(-M/N) \to e^{-c}$, so we want to show $\log N - M/N \to -c$. Rearranging,

$$\frac{M - N\log N}{N} \to c. \tag{7.2}$$

Define $M = M_N$ by (7.2) and check that the argument works.

7.2.1 Coupon Collector Problem

Put balls uniformly independently into N boxes. Let L_N be the number of balls until there are no empty boxes. $P(L_N \leq M) = P(X_{N,M} = 0)$ because they are the same events. Under relation (7.2), the probability goes to $\exp(-e^{-c})$ because $X_{N,M} \xrightarrow{\mathrm{d}} \text{Poisson}(e^{-c})$. Then, $P(L_N \leq N\log N + cN) \to \exp(-e^{-c})$, so

$$P\left(\frac{L_N - N\log N}{N} \leq c\right) \to \exp(-e^{-c}),$$

that is,

$$\frac{L_N - N\log N}{N} \xrightarrow{\mathrm{d}} \xi,$$

where ξ has distribution function $P(\xi \leq c) = \exp(-e^{-c})$ for $-\infty < c < \infty$. This is known as the **Gumbel distribution**.

7.3 Weak Convergence in Metric Spaces

Recall the definition of a complete, separable metric space (S, d). As an example, take $\mathbb{R}^k$, with

$$d(x, y) = |x - y| = \sqrt{\sum_i (x_i - y_i)^2},$$

for $x = (x_1, \ldots, x_k)$.

On $\mathbb{R}^k$, we have a partial order $x \le y \iff x_i \le y_i$, $1 \le i \le k$. We can define a distribution function for $\mathbb{R}^k$-valued $X = (X_1, \ldots, X_k)$.

$$F(x) = P(X \le x) = P(X_i \le x_i, \text{all } 1 \le i \le k)$$

However, this is less useful than in one dimension.

Theorem 7.5 (Portmanteau Theorem)**.** *On (S, d), let μ_n, $1 \le n \le \infty$ be PMs on (S, d). The following are equivalent, and define weak convergence $\mu_n \to \mu_\infty$.*

(a) $\int_S f \, \mathrm{d}\mu_n \xrightarrow{n\to\infty} \int_S f \, \mathrm{d}\mu_\infty$ *for all bounded continuous* $f : S \to \mathbb{R}$.

(b) $\limsup_n \mu_n(C) \le \mu_\infty(C)$ *for all closed* C.

(c) $\liminf_n \mu_n(G) \ge \mu_\infty(G)$ *for all open* G.

(d) $\mu_n(A) \to \mu_\infty(A)$ *for all* A *such that* $\mu_\infty(\bar{A} \setminus A^0) = 0$. *(This is the analog of continuity points.)*

(e) There exist S-valued RVs $\hat{X}_n$ such that $\operatorname{dist}(\hat{X}_n) = \mu_n$, $1 \le n \le \infty$, *and* $\hat{X}_n \to \hat{X}_\infty$ *a.s.*

The hard part is $\implies$ *(e)*, which is the Skorokhod Representation Theorem.

We will state analogs of $\mathbb{R}^1$ results.

Lemma 7.6 (Continuous Mapping Theorem)**.** *If $X_n \xrightarrow{\mathrm{d}} X_\infty$, then $f(X_n) \xrightarrow{\mathrm{d}} f(X_\infty)$ for any $f : S \to S'$ such that $P(X_\infty \in \mathcal{D}_f) = 0$, where $\mathcal{D}_f = \{x \in S : f \text{ is not continuous at } x\}$.*

Theorem 7.7. *For $\mathbb{R}^k$-valued (X_n), $X_n \xrightarrow{\mathrm{d}} X_\infty$ if and only if $F_n(x) \to F_\infty(x)$ for all continuity points x of F_∞.*

Definition 7.8. $(X_n, 1 \le n < \infty)$ is **tight** if for all $\varepsilon > 0$, there exists a compact $K_\varepsilon \subseteq S$ such that $\sup_n P(X_n \notin K_\varepsilon) \le \varepsilon$.

In $\mathbb{R}^k$, $(X_n, 1 \le n < \infty)$ is tight if and only if $\forall \varepsilon > 0\ \exists B_\varepsilon < \infty$ such that $\sup_n P(|X_n| \ge B_\varepsilon) \le \varepsilon$.

Theorem 7.9 (Prohorov's Theorem)**.** *(a) If $X_n \xrightarrow{\mathrm{d}}$ some X_∞, then $(X_n, 1 \le n < \infty)$ is tight.*

(b) If $(X_n, 1 \le n < \infty)$ is tight, then there exists a subsequence $X_{n_j} \xrightarrow{\mathrm{d}}$ some X_∞.

See the section in Billingsley on Convergence of PMs.

Lecture 8

February 9

8.1 Characteristic Functions in $\mathbb{R}^k$

For $t \in \mathbb{R}^k$, $x \in \mathbb{R}^k$, $t \cdot x = \sum_{i=1}^k t_i x_i$.

$X = (X_1, \ldots, X_k)$ is a $\mathbb{R}^k$-valued RV. $t \cdot X = \sum_{i=1}^k t_i X_i$ is a $\mathbb{R}^1$-valued RV.

The CF of X is a function $\phi(t) = E \exp(it \cdot X)$ as a function from $\mathbb{R}^k$ to $\mathbb{C}$.

The Uniqueness and Continuity Theorems are the same as in $\mathbb{R}^1$ (see the Billingsley textbook).

Theorem 8.1. *Let $X^{(n)}$, $n \geq 1$, be $\mathbb{R}^k$-valued RVs. Suppose $\phi_{X^{(n)}}(t) \to$ some limit $\phi(t)$ $\forall t \in \mathbb{R}^k$. If either*

(i) $(X^{(n)}, n \geq 1)$ is tight, or

(ii) ϕ is a CF,

then $X^{(n)} \xrightarrow{\text{d}} X$, where X has CF ϕ.

Theorem 8.2 (Cramér-Wold Device)**.** *Let $(X^{(n)})$ be $\mathbb{R}^k$-valued RVs. Suppose $t \cdot X^{(n)} \xrightarrow{\text{d}}$ some W_t (convergence in $\mathbb{R}^1$) as $n \to \infty$, for all $t \in \mathbb{R}^k$. If either*

(i) $(X^{(n)}, n \geq 1)$ is tight, or

(ii) $\exists X$ such that $t \cdot X \stackrel{\text{d}}{=} W_t$ $\forall t \in \mathbb{R}^k$.

Then, $X^{(n)} \xrightarrow{\text{d}} X$, where $t \cdot X \stackrel{\text{d}}{=} W_t$ $\forall t$.

Proof.

$$\phi_{X^{(n)}}(t) = E \exp(it \cdot X^{(n)}) \to E \exp(iW_t) \stackrel{\text{def}}{=} \phi(t)$$

Under (i), 8.1 implies that $X^{(n)} \xrightarrow{\text{d}}$ some X. We know that $t \cdot X^{(n)} \xrightarrow{\text{d}} W_t$. By the Continuous Mapping Theorem, $t \cdot X \stackrel{\text{d}}{=} W_t$.

Under (ii), $\phi(t) = E \exp(it \cdot X)$, and so is a CF. Apply (ii) of 8.1. □

Corollary 8.3. *To show $X^{(n)} \xrightarrow{d} X$ in $\mathbb{R}^k$, it is enough to show $E\prod_{j=1}^k f_j(X_j^{(n)}) \to E\prod_{j=1}^k f_j(X_j)$ for all bounded, continuous $f_j : \mathbb{R} \to \mathbb{R}$.*

Proof. This extends to $f_j : \mathbb{R} \to \mathbb{C}$. However, $x \mapsto e^{it\cdot x} \equiv \prod_{j=1}^n e^{it_j x_j}$ is of this multiplicative form. So, we have $E\exp(it\cdot X^{(n)}) \to E\exp(it\cdot X)$ $\forall t \in \mathbb{R}^k$. ☐

8.2 Central Limit Theorem in $\mathbb{R}^k$

Theorem 8.4 (IID CLT in $\mathbb{R}^k$)**.** *Consider X, $\mathbb{R}^k$-valued, $EX = 0$. Let $E[X_j X_\ell] = \Gamma_{j,\ell} < \infty$ (Γ is the covariance matrix). Let $X^{(n)}$ be IID copies of X, $S^{(n)} = \sum_{i=1}^n X^{(i)}$, $\mathbb{R}^k$-valued, $ES^{(n)} = 0$. Then $n^{-1/2}S^{(n)} \xrightarrow{d} Y$, where Y has CF*

$$\phi_Y(t) = \exp\left(-\frac{1}{2}\sum_j\sum_\ell t_i t_\ell \Gamma_{j,\ell}\right) = \exp\left(-\frac{1}{2}t^\top\Gamma t\right). \tag{8.1}$$

Proof.

$$E\left|S^{(n)}\right|^2 = \sum_{j=1}^k E\left|S_j^{(n)}\right|^2 = n\sum_{j=1}^n E|X_j|^2 = nE|X|^2$$

$E\left|n^{-1/2}S^{(n)}\right|^2 = E|X|^2$, so $(n^{-1/2}S^{(n)}, n \geq 1)$ is tight in $\mathbb{R}^k$. To apply Cramér-Wold 8.2, we need to show $t\cdot(n^{-1/2}S^{(n)}) \xrightarrow{d}$ some W_t.

$$\begin{aligned} n^{-1/2}\sum_{i=1}^n t\cdot X^{(i)} &\xrightarrow{d} \text{Normal}(0, E(t\cdot X)^2) \\ &= \text{Normal}(0, t^\top\Gamma t) \\ &= W_t, \end{aligned}$$

by the 1-dimensional CLT, since

$$E(t\cdot X)^2 = E\left[\left(\sum_{j=1}^k t_j X_j\right)\left(\sum_{\ell=1}^k t_\ell X_\ell\right)\right] = \sum_j\sum_\ell t_j t_\ell \Gamma_{j,\ell} = t^\top\Gamma t,$$

and

$$E\exp(iW_t) = \exp\left(-\frac{1}{2}t^\top\Gamma t\right).$$ ☐

Definition 8.5. A $\mathbb{R}^k$-valued Y has Normal$(0,\Gamma)$ distribution if its CF is (8.1).

Let A be an arbitrary non-random $k\times k$ matrix. Let $Z = (Z_1, Z_2, \ldots, Z_k)$ have IID Normal$(0,1)$ components. Consider $Y = AZ$, $Y_i = \sum_j A_{i,j}Z_j$.

$$t\cdot Y = \sum_i t_i Y_i = \sum_i\sum_j t_i A_{i,j} Z_j$$

$$E(t\cdot Y)^2 = E\left(\sum_i\sum_j t_i A_{i,j} Z_j\right)\left(\sum_\ell\sum_m t_\ell A_{\ell,m} Z_m\right) = \sum_j\sum_i\sum_\ell t_i A_{i,j} A_{\ell,j} t_\ell$$

$$= t^\top A A^\top t$$

This says Y has $\text{Normal}(0, AA^\top)$ distribution.

Check: $t \cdot Y$ is Normal.

Proposition 8.6. *For a $k \times k$ matrix Γ, the following are equivalent:*

1. $\Gamma = AA^\top$ *for some A.*
2. *The* $\text{Normal}(0, \Gamma)$ *distribution exists, and can be constructed as AZ for $Z = (Z_1, \ldots, Z_k)$ IID* $\text{Normal}(0, 1)$ *and for A as in 1.*
3. Γ *is the covariance matrix of some X with $EX = 0$.*
4. Γ *is symmetric and non-negative definite: $t^\top \Gamma t \geq 0\ \forall t$.*

Proof. 1 $\implies$ 2: We already proved this.

2 $\implies$ 3: Specialization.

3 $\implies$ 4: $t^\top \Gamma t$ is $\text{var}(t \cdot X)$.

4 $\implies$ 1 is matrix theory.

$$\begin{aligned}
\Gamma &= U^\top D U \\
&= U^\top D^{1/2} D^{1/2} U \\
&= AA^\top
\end{aligned}$$

for U orthonormal, D diagonal, $D \geq 0$. □

The CLT 8.4 gives 3 $\implies$ 2.

8.3 Weak Convergence in $\mathbb{R}^k$

Example 8.7 (Artificial Example)**.** Consider a probability measure on the unit square which is uniform on parallel diagonal lines. U is uniform on $[0, 1]$, $X_n = U$,

$$\begin{aligned}
Y_n &= nU - \lfloor nU \rfloor = \text{decimal part of } nU \\
&= nU \bmod 1.
\end{aligned}$$

As $n \to \infty$, $(X_n, Y_n) \xrightarrow{\text{d}} (U, \hat{U})$, with $\hat{U}$ uniform of $[0, 1]$, independent of U, which means that

$$(X_n, Y_n) \xrightarrow{\text{d}} \text{uniform on square } [0, 1]^2.$$

Simple Facts. For $\mathbb{R}$-valued Xs and Ys, a statement like

$$(X_n, Y_n) \xrightarrow{\text{d}} (X, Y) \tag{8.2}$$

is a statement about weak convergence on $\mathbb{R}^2$. Consider

$$X_n \xrightarrow{\text{d}} X \text{ and } Y_n \xrightarrow{\text{d}} Y. \tag{8.3}$$

(8.2) $\implies$ (8.3):

Continuous Mapping Theorem. If $(X_n, Y_n) \xrightarrow{d} (X, Y)$, then $g(X_n, Y_n) \xrightarrow{d} g(X, Y)$ for continuous g.

$(x, y) \mapsto x$ is continuous.

Not conversely!

So, $(X_n, Y_n) \xrightarrow{d} (X, Y)$ implies $X_n + Y_n \xrightarrow{d} X + Y$, $X_n/Y_n \xrightarrow{d} X/Y$ provided $P(Y = 0) = 0$.

Lemma 8.8. *Suppose* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{d} Y$. *If either*

(i) $P(Y = y_0) = 1$ *for some* y_0, *or*

(ii) X_n *and* Y_n *are independent (each* n*),*

then $(X_n, Y_n) \xrightarrow{d} (X, Y)$, *where* X *and* Y *are independent.*

Example 8.9 (Artificial Example)**.** ID and pairwise independence are *not* enough for the CLT.

$$19 = 10011 \text{ in binary}$$

$$n = \sum_{i=1}^{\infty} b_i(n) 2^{i-1}$$

Take $\xi_0, \xi_1, \xi_2, \ldots$, IID, $P(\xi = 1) = 1/2 = P(\xi = -1)$. Define $X_n = \xi_0 \prod_{i: b_i(n)=1} \xi_i$, $n \geq 0$. X_n takes values $\{\pm 1\}$. Check that the (X_n) are pairwise independent.

$$S = \sum_{n=0}^{2^j - 1} X_n = \xi_0(1 + \xi_1)(1 + \xi_2) \cdots (1 + \xi_j)$$

$ES = 0$, $\text{var}(S) = 2^j$. Then, $P(S = 2^j) = P(S = -2^j) = (1/2)2^{-j}$, and $S = 0$ otherwise. $2^{-j/2} S$ does not converge to $\text{Normal}(0, 1)$.

Lecture 9

February 14

9.1 Markov Chains: Big Picture

state space S	**discrete time**	**continuous time**
finite	very similar	very similar
countable	our focus	similar
general: measure theory	(∗)	doesn't exist
general: topology	(∗)	SDE (starting from BM) semigroup setting

We will do a little of the sections marked (∗).

9.2 Measure Theory Background

"X and Y are independent" means "$\sigma(X)$ and $\sigma(Y)$ are independent".

Definition 9.1. X and Y are **conditionally independent (CI) given** $\mathcal{G}$ means

$$E[h(X)g(Y) \mid \mathcal{G}] = E[h(X) \mid \mathcal{G}]E[g(Y) \mid \mathcal{G}] \qquad \forall \text{ bounded, measurable } g, h.$$

This is equivalent to

$$E[g(Y) \mid \mathcal{G}, X] = E[g(Y) \mid \mathcal{G}] \qquad \forall \text{ bounded, measurable } g. \tag{9.1}$$

Idea: Given $\mathcal{G}$, knowing also X gives no extra information about Y.

Easy Fact. If X and Y are CI given $\mathcal{G}$, if V is $\mathcal{G}$-measurable, then X and (Y, V) are CI given $\mathcal{G}$.

Recall. μ is a PM on $S_1 \times S_2$. μ_1 is the marginal PM on S_1. Q is a kernel $Q(s_1, B)$ from $S_1 \to S_2$. There is a one-to-one correspondence $\mu \leftrightarrow (\mu_1, Q)$.

Lemma 9.2 (The Splice Lemma). *Given spaces S_1, S_2, S_3 (Borel spaces), given a PM $\mu_{1,2}$ on $S_1 \times S_2$ and a PM $\mu_{2,3}$ on $S_2 \times S_3$ such that their marginals on S_2 are identical, then there exists a unique PM μ on $S_1 \times S_2 \times S_3$ such that, for $\mu = \mathrm{dist}(X_1, X_2, X_3)$,*

- $\mathrm{dist}(X_1, X_2) = \mu_{1,2}$ *and* $\mathrm{dist}(X_2, X_3) = \mu_{2,3}$*, and*
- X_1 *and* X_2 *are CI given* X_2*.*

Proof. Consider $(S_1 \times S_2) \times S_3$. Specify μ by

- the marginal on $S_1 \times S_2$ is $\mu_{1,2}$,
- the kernel Q from $S_1 \times S_2$ to S_3 is $Q((s_1, s_2), B) = Q_{2,3}(s_2, B)$, where $Q_{2,3}$ is the kernel $S_2 \to S_3$ associated with $\mu_{2,3}$.

This specifies μ. Then,

$$\begin{aligned} E[h(X_3) \mid (X_1, X_2)] = \int h(x) Q((X_1, X_2), \mathrm{d}x) &= \int h(x) Q_{2,3}(X_2, \mathrm{d}x) \\ &= E[h(X_3) \mid X_2]. \end{aligned}$$

We have checked (9.1), which implies CI. The calculation also says that $\operatorname{dist}(X_2, X_3) = \mu_{2,3}$.

Exercise: Prove uniqueness. □

9.3 Existence of General Markov Chains (Borel Spaces)

Theorem 9.3 (Existence of General Markov Chains (Borel Spaces))**.** *Given Borel $S_0, S_1, S_2, \ldots$, given a PM μ_0 on S_0, given kernels Q_n from S_n to S_{n+1} (each $n \ge 0$), there exists $(X_0, X_1, X_2, \ldots)$, unique in distribution, such that*

(a) $\operatorname{dist}(X_0) = \mu_0$,

(b) Q_n is the conditional probability kernel for X_{n+1} given X_n,

(c) X_{n+1} and $(X_0, X_1, \ldots, X_{n-1})$ are CI given X_n (all $n \ge 1$),

(d) $(X_n, X_{n+1}, \ldots)$ and $\mathcal{F}_n$ are CI given X_n.

Proof. Suppose (induction) we have constructed $(X_0, X_1, \ldots, X_n)$. Apply the Splice Lemma 9.2 to $(X_0, X_1, \ldots, X_{n-1})$ and X_n and X_{n+1}. We have a joint distribution for $(X_0, X_1, \ldots, X_{n-1})$ and X_n. The joint distribution of X_n and X_{n+1} is specified by $\operatorname{dist}(X_n)$ and the kernel Q_n. The Splice Lemma implies the existence of $\operatorname{dist}(X_0, X_1, \ldots, X_n, X_{n+1})$ with the CI property. Apply the Kolmogorov Extension Theorem to get $\operatorname{dist}(X_0, X_1, X_2, \ldots)$.

(c) gives the "one-step ahead"property, but we want the analog for the entire future.

Write $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. (c) and the "Easy Fact" imply

(c') $\mathcal{F}_n$ and (X_n, X_{n+1}) are CI given X_n.

Claim (d): By MT, it is enough to prove, for each m,

(d') $(X_n, X_{n+1}, \ldots, X_{n+m})$ and $\mathcal{F}_n$ are CI given X_n, for all n.

Use induction on m. The statement is true for $m = 1$ by (c'). We will prove the statement for $m = 2$. The same argument (exercise) gives the inductive step $m \to m + 1$.

Apply (c') to $n + 1$.

$$\begin{aligned} E[g(X_{n+1}, X_{n+2}) \mid F_{n+1}, X_{n+1}] &= E[g(X_{n+1}, X_{n+2}) \mid X_{n+1}] \\ &= h(X_{n+1}), \qquad \text{say.} \end{aligned}$$

Condition on $\mathcal{F}_n$.

$$E[g(X_{n+1}, X_{n+2}) \mid \mathcal{F}_n] = E[h(X_{n+1}) \mid \mathcal{F}_n] = E[h(X_{n+1}) \mid X_n]$$

using the $m = 1$ case of CI. Condition on X_n.

$$E[g(X_{n+1}, X_{n+2}) \mid X_n] = E[h(X_{n+1}) \mid X_n],$$

so (X_{n+1}, X_{n+2}) and $\mathcal{F}_n$ are CI given X_n. This is (d') for $m = 2$. □

In practice, we usually consider time-homogeneous chains: $S_n = S$, $Q_n = Q$.

(Idea). Given $X_{n_0} = x_0$, the future process $(X_{n_0+n}, n \geq 0)$ has the same distribution as the process $(x_0 = X_0, X_1, X_2, \dots)$.

Formula. For bounded, measurable $h : S^\infty \to \mathbb{R}$,

$$E[h(X_{n_0}, X_{n_0+1}, \dots) \mid \mathcal{F}_{n_0}] = g(X_{n_0}),$$

where $g(x) \stackrel{\text{def}}{=} Eh(\hat{X}_0, \hat{X}_1, \dots)$, where $(\hat{X}_n, n \geq 0)$ is the chain with $X_0 = x$.

9.4 Elementary Examples

Recall the following elementary examples for S countable.

The kernel is specified by transition probabilities $p_{i,j} \equiv p(i,j) = \mathbb{P}(X_1 = j \mid X_0 = i)$ which form a transition matrix $\mathbf{P} = (p_{i,j} : i, j \in S)$.

Example 9.4 (Random Walk on $\mathbb{Z}^d$)**.** Given IID ξ_i, $i \geq 1$, $\mathbb{Z}^d$-valued, $X_n = \sum_{t=1}^n \xi_t$. Then, (X_n) is Markov, $p(i,j) = \mathbb{P}(\xi = j - i)$. Here, $S = \mathbb{Z}^d$.

Example 9.5 (Renewal Chain)**.** $S = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$. Take $(\xi_i, i \geq 1)$ to be IID, $\mathbb{P}(\xi \geq 1) = 1$, $S_n = \sum_{t=1}^n \xi_t$. Define $X_n = \min\{n - S_m : S_m \leq n\}$. This is Markov on $\mathbb{Z}^+$. Then,

$$\begin{aligned} p(i, i+1) &= \mathbb{P}(\xi > i + 1 \mid \xi > i), \\ p(i, 0) &= \mathbb{P}(\xi = i + 1 \mid \xi > i). \end{aligned}$$

Example 9.6 (Galton-Watson Branching Process)**.** Given a PM μ on $\{0, 1, 2, \dots\}$, $X_0 = 1$ (1 individual in generation 0). In each generation, each individual has a random (dist $= \mu$) number of offspring in the next generation. X_n is the population in generation n. This is Markov. $p(i,j) = \mathbb{P}(\xi_1 + \xi_2 + \cdots + \xi_i = j)$ for IID(μ) RVs (ξ_i).

S is infinite in 9.4 to 9.6.

Example 9.7 (Ehrenfest Urn Model)**.** There are B balls and 2 boxes. Pick a random ball and move it to the other box. X_n is the number of balls in the left box.

$$\begin{aligned} p(i, i-1) &= \frac{i}{B}, \\ p(i, i+1) &= \frac{B - i}{B}. \end{aligned}$$

Lecture 10

February 16

10.1 Markov Chains: Some Classical Methods

We have a countable $S = \{i, j, k, \dots\}$ and a transition matrix $\mathbf{P} = (p_{i,j})_{i,j \in S}$ satisfying $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1$. The Markov chain $(X_0, X_1, X_2, \dots)$ has

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{i,j}.$$

We write

$$\begin{aligned} \mathbb{P}_i(\cdot) &= \mathbb{P}(\cdot \mid X_0 = i), \\ \mathbb{E}_i[\cdot] &= \mathbb{E}[\cdot \mid X_0 = i]. \end{aligned}$$

Write $\mu_n = \operatorname{dist}(X_n)$. μ_n can be viewed as a vector $\boldsymbol{\mu}_n = (\mu_n(i), i \in S)$, where $\mu_n(i) = \mathbb{P}(X_n = i)$. Then,

$$\mu_{n+1}(j) = \sum_i \mu_n(i) p_{i,j}.$$

In matrix form, we have the **forwards equation** $\boldsymbol{\mu}_{n+1} = \boldsymbol{\mu}_n \mathbf{P}$, a vector-matrix product. Then,

$$\begin{aligned} \boldsymbol{\mu}_1 &= \boldsymbol{\mu}_0 \mathbf{P}, \\ \boldsymbol{\mu}_2 &= \boldsymbol{\mu}_1 \mathbf{P} = \boldsymbol{\mu}_0 \mathbf{P}^2, \\ &\vdots \end{aligned}$$

so $\boldsymbol{\mu}_n = \boldsymbol{\mu}_0 \mathbf{P}^n$, where $\mathbf{P}^n = \mathbf{P}\mathbf{P}\cdots\mathbf{P}$ is matrix multiplication. We obtained these equations by conditioning on X_n.

Fix a function $f : S \to \mathbb{R}$. Consider $\nu_n(i) = \mathbb{E}_i f(X_n)$. Condition on X_1.

$$\begin{aligned} \nu_{n+1}(i) = \mathbb{E}_i f(X_{n+1}) &= \sum_j p_{i,j} \underbrace{\mathbb{E}_j[f(X_{n+1}) \mid \cancel{X_0 = i}, X_1 = j]}_{= \mathbb{E}_j f(X_n)} \\ &= \sum_j p_{i,j} \nu_j(j), \end{aligned}$$

by the Markov property. Write $\nu_n(i) = \mathbb{E}_i f(X_n)$. The **backwards equation** is $\boldsymbol{\nu}_{n+1} = \mathbf{P}\boldsymbol{\nu}_n$, so

$$\boldsymbol{\nu}_n = \mathbf{P}^n \boldsymbol{\nu}_0.$$

We see that $(\mathbf{P}^n)_{i,j} = \mathbb{P}(X_n = j \mid X_0 = i)$.

The analog of $(p_{i,j})$ on general S is the kerenel $Q = Q(x, A)$, which defines two maps.

For $\mu \in \mathscr{P}(S)$, we have a map $\mu \mapsto \hat{\mu}$, where $\hat{\mu}(\cdot) = \int \mu(\mathrm{d}x)Q(x, \cdot)$. Here, $\mu \mapsto \mu Q$.

For a function $f : S \to \mathbb{R}$, we have a map $f \mapsto \hat{f}$, where $\hat{f}(x) = \int Q(x, \mathrm{d}y)f(y)$. Here, $f \mapsto Qf$.

Many questions about finite-state MCs can be answered in terms of the matrix $\mathbf{P}$.

10.1.1 Hitting Times

For $A \subseteq S$, write

$$\tau_A = \min\{n \geq 0 : X_n \in A\},$$
$$T_A = \min\{n \geq 1 : X_n \in A\}.$$

In either case, the hitting time could equal ∞ if $X_n \notin A \ \forall n$. Consider $h_A(i) = \mathbb{P}_i(T_A < \infty)$.

First way to study h_A: Define the matrix $\mathbf{Q}$, the "$\mathbf{P}$-chain killed after entering A".

$$q_{i,j} = \begin{cases} p_{i,j}, & i \notin A, \\ 0, & i \in A. \end{cases}$$

Easy: $\mathbb{P}_i(\tau_A = n, X_{\tau_A} = j) = (\mathbf{Q}^n)_{i,j}$ for $j \in A$.

$$[(\mathbf{I} - \mathbf{Q})^{-1}]_{i,j} = \left[\sum_{n=0}^{\infty} \mathbf{Q}^n\right]_{i,j} = \mathbb{P}_i(\tau_A < \infty, X_{\tau_A} = j), \qquad j \in A$$

This is the matrix form of the identity

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

From this, we obtain

$$\mathbf{h}_A = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{1}_A.$$

Second way to consider $\mathbf{h}_A$:

Proposition 10.1. *(a)* $\mathbf{h} = \mathbf{h}_A$ *satisfies*

(i) $h(i) = \sum_j p_{i,j}h(j)$ *for* $i \in A$*,*

(ii) $h(i) = 1$ *for* $i \in A$*,*

(iii) $\mathbf{h} \geq 0$*.*

(b) If $\mathbf{h}$ *satisifes (i) to (iii), then* $\mathbf{h}_A \leq \mathbf{h}$*, so* $\mathbf{h}_A$ *is the minimal solution of (i) to (iii).*

Proof. (a) Condition on the first step for (i).

(b) Define $\mathbf{P}^A$ and $(X_n^A, n \geq 0)$, the "$\mathbf{P}$-chain stopped on A", by

$$p_{i,j}^A = p_{i,j}, \ i \notin A, \qquad p_{i,j} = \delta_{i,j}, \ i \in A.$$

Given $\mathbf{h}$ which satisfies (i) to (iii), (i) and (ii) imply $\mathbf{h} = \mathbf{P}^A\mathbf{h}$ and $\mathbf{h} \geq \mathbf{1}_A$. Therefore,

$$\mathbf{h} = \mathbf{P}^A\mathbf{h} \geq \mathbf{P}^A\mathbf{1}_A.$$

Repeat n times to obtain $\mathbf{h} \geq (\mathbf{P}^A)^n \mathbf{1}_A$. Then,

$$h(i) = ((\mathbf{P}^A)^n \mathbf{1}_A)_i = \mathbb{P}_i(X_n^A \in A) = \mathbb{P}_i(\tau_A \leq n),$$

since

$$X_n^A = \begin{cases} X_n, & \text{if } \tau_A > n, \\ X_{\tau_A}, & \text{if } \tau_A \leq n. \end{cases}$$

Let $n \to \infty$. $h(i) \geq \mathbb{P}_i(\tau_A < \infty) \equiv h_A(i)$.

□

10.1.2 Generating Functions

Let $T_y = \min\{n \geq 0 : X_n = y\}$, $p_{x,y}^n = \mathbb{P}_x(X_n = y) = (\mathbf{P}^n)_{i,j}$. The "Strong Markov Property" says

$$\mathbb{P}_x(X_n = y) = \sum_{m=0}^{n} \mathbb{P}_x(T_y = m)\mathbb{P}_y(X_{n-m} = y),$$

$$p_{x,y}^n = \sum_{m=0}^{\infty} \mathbb{P}_x(T_y = m) p_{y,y}^{n-m},$$

$$\phi_{x,y}(z) = \sum_{n=0}^{\infty} p_{x,y}^n z^n = \sum_{0 \leq m \leq n < \infty} \sum \mathbb{P}_x(T_y = m) z^m p_{y,y}^{n-m} z^{n-m}, \qquad n = m + i,$$

$$= \underbrace{\sum_{m=0}^{\infty} \mathbb{P}_x(T_y = m) z^m}_{\stackrel{\text{def}}{=} \psi_{x,y}(z)} \underbrace{\sum_{i=0}^{\infty} p_{y,y}^i z_i}_{= \phi_{y,y}(z)}.$$

Now, we have a formula for the GF of (T_y).

$$\psi_{x,y}(z) = \frac{\phi_{x,y}(z)}{\phi_{y,y}(z)}.$$

Consider the matrix $\mathbf{\Phi}(z)$ with entries $\phi_{x,y}(z)$.

$$\mathbf{\Phi}(z) = \sum_{n=0}^{\infty} \mathbf{P}^n z^n = (\mathbf{I} - \mathbf{P}z)^{-1}$$

This, in principle, is a formula for the distribution of T_y in terms of $\mathbf{P}$.

10.2 More Examples of MCs

Example 10.2 (Random Walk on an Undirected Finite Graph $G = (V, E)$)**.** The state space is V. $v \in V$ has some degree $d(v)$, the number of edges at v. Suppose $d(v) \geq 1$. Then,

$$p_{i,j} = \frac{1}{d(i)}, \qquad \text{if } (i,j) \in E.$$

Example 10.3 (Card-Shuffling "Random Transposition" Model)**.** Consider a n card deck. S is the set of $n!$ orderings. Pick two random cards and interchange them; this is one step of the chain. For configurations $\mathbf{x}$ and $\mathbf{y}$,

$$p_{\mathbf{x},\mathbf{y}} = \frac{2}{n^2}, \quad \text{if it is possible to reach } x \to y \text{ by a transposition},$$

$$p_{\mathbf{x},\mathbf{x}} = \frac{1}{n}.$$

Lecture 11

February 21

11.1 Strong Markov Property

Let $(X_n, n = 0, 1, 2, \dots)$ be a MC on a countable $S = \{x, y, z, \dots\}$. Let $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$.

Markov property: for bounded, measurable $f : S^\infty \to \mathbb{R}$, write $g(x) = \mathbb{E}_x f(X_0, X_1, X_2, \dots)$. Then, $\mathbb{E}_\mu[f(X_n, X_{n+1}, X_{n+2}, \dots) \mid \mathcal{F}_n] = g(X_n)$.

Recall: A stopping time $T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ is such that $\{T \le n\} \in \mathcal{F}_n$, for all $0 \le n < \infty$. This is equivalent to $\{T = n\} \in \mathcal{F}_n$, for all $0 \le n < \infty$.

Theorem 11.1 (Strong Markov Property)**.** *Write $g(x) = \mathbb{E}_x f(X_0, X_1, X_2, \dots)$, where $f : S^\infty \to \mathbb{R}$ is bounded and measurable. Then, $\mathbb{E}_\mu[f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T] = g(X_T)$ a.s. on $\{T < \infty\}$.*

Proof. $X_T 1_{(T<\infty)}$ is $\mathcal{F}_T$-measurable. We need to check: for $B \in \mathcal{F}_T$,

$$\mathbb{E}_\mu[f(X_T, X_{T+1}, \dots) 1_B 1_{(T<\infty)}] = \mathbb{E}_\mu[g(X_T) 1_B 1_{(T<\infty)}].$$

Break over $n = 0, 1, 2, \dots$. $1_B 1_{(T<\infty)} = \sum_{n=0}^\infty 1_{B \cap \{T=n\}} = \sum_{n=0}^\infty 1_{A_n}$, where $A_n = B \cap \{T = n\} \in \mathcal{F}_n$ by the definition of $\mathcal{F}_T$. So, it is enough to show

$$\mathbb{E}_\mu[f(X_T, X_{T+1}, \dots) 1_{A_n}] \mathbb{E}_\mu[g(X_T) 1_{A_n}],$$

which is

$$\mathbb{E}_\mu[f(X_n, X_{n+1}, \dots) 1_{A_n}] = \mathbb{E}_\mu[g(X_n) 1_{A_n}] \qquad \text{on } A_n \ (T = n).$$

This is the Markov property. □

Special Case. Suppose T is such that $X_T = y$ (non-random y) on $\{T < \infty\}$. Then,

$$\mathbb{E}_\mu[f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T] = g(y) \qquad \text{on } \{T < \infty\}, \ \forall f.$$

This implies $f(X_T, X_{T+1}, \dots)$ and $\mathcal{F}_T$ are independent on $\{T < \infty\}$ and

$$\text{dist}((X_T, X_{T+1}, \dots) \mid \mathcal{F}_T) = \text{dist}_y(X_0, X_1, \dots).$$

11.2 Recurrence Times

Consider $T_y^+ \stackrel{\text{def}}{=} \min\{n \ge 1 : X_n = y\}$ and $\rho_{x,y} = \mathbb{P}_x(T_y^+ < \infty)$.

Lemma 11.2. *For distinct x, y, z, $\rho_{x,z} \geq \rho_{x,y}\rho_{y,z}$.*

Proof.

$$\begin{aligned}\rho_{x,z} &\geq \mathbb{P}_x(\text{visit } z \text{ sometime after } T_y^+) \\ &= \rho_{x,y}\mathbb{P}_x(\text{visit } z \text{ sometime after } T_y^+ \mid T_y^+ < \infty).\end{aligned}$$

We want to say the second factor is $\rho_{y,z}$ by the SMP. Take $f(x_0, x_1, x_2, \dots) = 1_{(x_i = z \text{ for some } i)}$.

$$\mathbb{E}[f(X_{T_y^+}, X_{T_y^+ + 1}, \dots) \mid \mathcal{F}_{T_y^+}] = g(y) = \mathbb{E}_y f(X_0, X_1, \dots) = \rho_{y,z}.$$

Take the expectation over $1_{(T<\infty)}$.

$$\mathbb{P}(\text{visit sometime after } T_y^+, \text{and } T_y^+ < \infty) = g(y)\mathbb{P}(T_y^+ < \infty) = \rho_{y,z}\mathbb{P}(T_y^+ < \infty).$$ □

Define T_y^k to be the time of the kth visit to y, $T_y^0 = 0$, and $T_y^{k+1} = \min\{n : n > T_y^k, X_n = y\}$. Then, $\rho_{x,y} = \mathbb{P}_x(T_y^1 < \infty)$.

Theorem 11.3 (Theorem 6.4.1)**.**

$$\mathbb{P}_x(T_y^k < \infty) = \rho_{x,y}\rho_{y,y}^{k-1}, \qquad k \geq 1.$$

Proof. It is true for $k = 1$. By induction, suppose it is true for k.

$$\mathbb{P}_x(T_y^{k+1} < \infty) = \mathbb{P}_x(T_y^1 < \infty, T_y^{k+1} < \infty) = \mathbb{E}_x[1_{(T_y^+ < \infty)} \underbrace{\mathbb{P}(T_y^{k+1} < \infty \mid \mathcal{F}_{T_y^1})}_{*}]$$

However,

$$\begin{aligned} * &= \mathbb{E}_x[f(X_{T_y^1}, X_{T_y^1+1}, \dots) \mid \mathcal{F}_{T_y^1}] \quad \text{for } f(x_0, x_1, \dots) = 1_{(x_i = y \text{ for at least } k \text{ values of } i)} \\ &\underbrace{=}_{\text{SMP}} g(y) = \mathbb{E}_y f(X_0, X_1, \dots) = \mathbb{P}_y(T_y^k < \infty).\end{aligned}$$

By induction, this is $\rho_{y,y}^k$. Hence,

$$\mathbb{P}_x(T_y^{k+1} < \infty) = \rho_{y,y}^k \rho_{x,y},$$

so the statement is true for $k + 1$. □

Definition 11.4. A state y is **recurrent** if $\rho_{y,y} = 1$ and **transient** if $\rho_{y,y} < 1$.

Consider the number of visits to y, $\sum_{n=1}^{\infty} 1_{(X_n = y)} = N(y)$.

Lemma 11.5. *If y is recurrent, then $\mathbb{P}_y(N(y) = \infty) = 1$, so $\mathbb{E}_y N(y) = \infty$. If y is transient, then $\mathbb{P}_y(N(y) \geq k) = \mathbb{P}_y(T_y^k < \infty) = \rho_{y,y}^k$, for $k = 0, 1, 2, \dots$, and so*

$$\mathbb{E}_y N(y) = \frac{1}{1 - \rho_{x,y}} - 1 = \frac{\rho_{y,y}}{1 - \rho_{y,y}} < \infty.$$

Proof. In either case, $\mathbb{P}_y(N(y) \geq k) = \mathbb{P}_y(T_y^k < \infty) = \rho_{y,y}^k$. Then,

$$\mathbb{P}_y(N(y) = k) = \rho_{y,y}^k - \rho_{y,y}^{k+1} = 0 \quad \text{if } \rho_{y,y} = 1,$$

and

$$\mathbb{P}_y(N(y) < \infty) = 0.$$

□

Note.

$$\mathbb{E}_x N(y) = \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=1}^{\infty} p_{x,y}^{(n)}.$$

Corollary 11.6.

$$y \text{ is recurrent} \iff \mathbb{E}_y N(y) = \infty \iff \sum_n p_{y,y}^{(n)} = \infty.$$

Theorem 11.7 (Theorem 6.4.3)**.** *Suppose x is recurrent and $\rho_{x,y} > 0$. Then, y is recurrent and $\rho_{y,x} = 1$. [So, $\rho_{x,y} = 1$ by switching x and y.]*

Proof.

$$\underbrace{\mathbb{P}_x(N(x) < \infty)}_{0,\ x \text{ recurrent}} \geq \mathbb{P}_x(T_y < \infty, \text{never visit } x \text{ after } T_y)$$

$$\underbrace{=}_{\text{SMP}} \underbrace{\rho_{x,y}}_{>0,\text{ hypothesis}} \underbrace{(1 - \rho_{y,x})}_{\text{must be } 0},$$

so $\rho_{y,x} = 1$.

$\rho_{x,y} > 0 \implies \exists K$ such that $p_{x,y}^{(K)} > 0$.

$\rho_{y,x} > 0 \implies \exists L$ such that $p_{y,x}^{(L)} > 0$.

Then,

$$p_{y,y}^{(K+L+m)} \geq \mathbb{P}_y(X_L = x, X_{L+m} = x, X_{L+m+K} = y)$$

$$\underbrace{=}_{\text{Markov}} p_{y,x}^{(L)} p_{x,x}^{(m)} p_{x,y}^{(K)}.$$

Now, sum over m.

$$\underbrace{\sum_m p_{y,y}^{(m)}}_{=\infty} \geq \underbrace{p_{y,x}^{(L)}}_{>0} \underbrace{p_{x,y}^{(K)}}_{>0} \sum_m p_{x,x}^{(m)} = \infty,$$

since x is recurrent, so y is recurrent. □

11.3 Elementary Graph Theory

Consider a directed graph on countable S, the set of vertices. Given $\mathbf{P} = (p_{i,j})$, put the edge $i \to j$ if $p_{i,j} > 0$.

We can define an equivalence relation R by

$$i \, R \, j \iff i = j \text{ or } \exists \text{ directed path from } i \text{ to } j \text{ and from } j \text{ to } i.$$

This partitions S into "strongly connected components" (SCC).

A SCC "C" is **open** if $\exists i \in C, j \notin C$ with $i \to j$ ($p_{i,j} > 0$), **closed** if not.

Corollary 11.8. *In a SCC C, either all $x \in C$ are recurrent or all $x \in C$ are transient.*

Proof. Suppose some $x \in C$ is recurrent. Take any $y \in C$. Then, $\rho_{x,y} > 0$, so by 11.7, y is recurrent. □

Example 11.9. Suppose $S = \{0, 1, 2, \dots\}$ and suppose $p_{0,0} = 1$, $p_{i,i+1} > 0$, $p_{i,i-1} > 0$. There are two SCCs, one open (and therefore transient) and one closed. So,

$$\mathbb{P}(X_n = 0 \text{ ultimately}) + \mathbb{P}(X_n \to \infty \text{ as } n \to \infty) = 1,$$

since $N(y) < \infty$ for each $y \geq 1$.

Lecture 12

February 23

12.1 Classification of States

Let S be the state space, $T_x^+ = \min\{n \geq 1 : X_n = x\}$, $\rho_{x,y} = \mathbb{P}_x(T_y^+ < \infty)$, and $N(x) = \sum_{n=1}^{\infty} 1_{(X_n = x)}$.

$$\begin{aligned} x \text{ is recurrent} \overset{\text{def}}{=} \rho_{x,x} = 1 &\implies \mathbb{P}_x(N(x) = \infty) = 1 \\ &\implies \mathbb{E}_x N(x) = \infty. \end{aligned}$$

$$x \text{ is transient} \overset{\text{def}}{=} \rho_{x,x} < 1 \implies \mathbb{E}_x N(x) = \frac{\rho_{x,x}}{1 - \rho_{x,x}} < \infty.$$

It is aways the case that

$$\mathbb{E}_x N(y) = \frac{\rho_{x,y}}{1 - \rho_{y,y}}.$$

Define the relation $x \sim y$ by $x = y$ or $(\rho_{x,y} > 0$ and $\rho_{y,x} > 0)$. The equivalence class C is a "SCC". Define C is open if $\exists x \in C, y \notin C,\ \rho_{x,y} > 0$, and C is closed if not.

Fact. Given a SCC "C", either x is transient for all $x \in C$ or x is recurrent for all $x \in C$. Call C transient or recurrent respectively.

Theorem. If x is recurrent and $\rho_{x,y} > 0$, then y is recurrent and $\rho_{y,x} = 1$.

Proposition 12.1. *Let C be a SCC.*

(a) If C is open, then C is transient (if C is recurrent, then C is closed).

(b) If C is closed and finite, then C is recurrent.

(c) If S is finite, then $R = \{$recurrent states$\}$ is non-empty and $\mathbb{P}_x(T_R < \infty) = 1\ \forall x$.

Proof. (a) follows from 11.7. If C is open, then $\exists x \in C, y \notin C\ \rho_{x,y} > 0$. If x is recurrent, by the Theorem, $\rho_{y,x} > 0$ implies $x \sim y$, which implies $y \in C$, which is a contradiction.

(b): Fix $x \in C$. For a chain started at x, since C is closed,

$$\sum_{y \in C} 1_{(X_n = y)} = 1 \quad \forall n,$$

$$\mathbb{E}_x \sum_{n=1}^{\infty} \sum_{y \in C} 1_{(X_n = y)} = \infty,$$

$$\mathbb{E}_x \sum_{y \in C} N(y) = \sum_{y \in C} \mathbb{E}_x N(y) = \infty.$$

If C is finite, then $\mathbb{E}_x N(y) = \infty$ for some $y \in C$, so y is recurrent, so C is recurrent.

(c): Fix x. Consider a transient y. Then, $\mathbb{E}_x N(y) < \infty$, so

$$\mathbb{P}_x(N(y) < \infty) = 1,$$

so $\mathbb{P}_x(\sum_{y \text{ transient}} N(y) < \infty) = 1$. However, $T_R \le \sum_{y \text{ transient}} N(y) + 1$, so $\mathbb{P}_x(T_R < \infty) = 1$. □

Note: At T_R, we are at state X_{T_R}, which is some closed C, which implies that $X_n \in C \; \forall n \ge T_R$.

Definition 12.2. A chain is **irreducible** if $\rho_{x,y} > 0 \; \forall x, y$.

12.1 implies: if S is finite and irreducible, then the chain is recurrent. If S is infinite and irreducible, then the chain may be recurrent or transient.

12.2 Birth-and-Death Chains

Let $S = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$, $p(i, i+1) = p_i > 0$, $p(i, i-1) = q_i > 0$ (for $i \ge 1$), $p(i, i) = r_i = 1 - p_i - q_1 \ge 0$. Set $q_0 = 0$.

Write $\tau_j = \min\{n \ge 0 : X_n = j\}$.

Analysis. Fix $m \ge 1$. Study $f(i) = \mathbb{P}_i(\tau_m < \tau_0)$, $0 \le i \le m$, $f(0) = 0$, $f(m) = 1$. Condition on the first step: for $1 \le i \le m-1$, $f(i) = p_i f(i+1) + q_i f(i-1) + r_i f(i)$. Solve: $p_i(f(i+1) - f(i)) = q_i(f(i) - f(i-1))$, or

$$f(i+1) - f(i) = \frac{q_i}{p_i}(f(i) - f(i-1)),$$

$$f(i+1) - f(i) = \left(\prod_{j=1}^{i} \frac{q_j}{p_j} \right) f(1),$$

$$f(x) = f(1) \underbrace{\sum_{i=0}^{x-1} \prod_{j=1}^{i} \frac{q_j}{p_j}}_{\phi(x)}.$$

We know $1 = f(m) = f(1)\phi(m)$, so $f(1) = 1/\phi(m)$. Hence,

$$\mathbb{P}_i(\tau_m < \tau_0) = \frac{\phi(i)}{\phi(m)}, \qquad 0 \le i \le m.$$

Can we say

$$\mathbb{P}_i(\tau_m > \tau_0) = 1 - \frac{\phi(i)}{\phi(m)}?$$

Make the chain absorbing at 0 and m. The states $\{1, \ldots, m-1\}$ are transient, so $\mathbb{P}_i(\tau_0 \text{ or } \tau_m < \infty) = 1$.

Is the chain recurrent or transient? recurrent $\iff \rho_{0,0} = 1 \iff \rho_{1,0} = 1$.

$$\{\tau_0 < \infty\} = \bigcup_{m=1}^{\infty} \{\tau_0 < \tau_m\},$$
$$\mathbb{P}_1(\tau_0 < \infty) = \lim_{m \to \infty} \mathbb{P}_1(\tau_0 < \tau_m) = \lim_{m \uparrow \infty} \left(1 - \frac{\phi(1)}{\phi(m)}\right).$$

Thus,

$$\text{recurrent} \iff \phi(\infty) \equiv \lim_{m \uparrow \infty} \phi(m) = \infty \iff \sum_i \prod_{j=1}^{i} \frac{q_j}{p_j} = \infty.$$

For a simple RW, $p_i = p > 0$, $q_i = q = 1 - p$. Then, the chain is recurrent if $p \geq 1/2$, transient if $p > 1/2$.

More Delicate Case. Fix C, take

$$p_i = \frac{1}{2} + \frac{C}{i} \quad \text{for large } i,$$
$$q_i = \frac{1}{2} - \frac{C}{i} \quad \text{for large } i.$$

Then,

$$\frac{q_j}{p_j} = \frac{1 - 2C/j}{1 + 2C/j} \approx \exp\left(-\frac{4C}{j}\right),$$
$$\prod_{j=1}^{i} \frac{q_j}{p_j} \approx \exp(-4C \cdot \log i) \approx i^{-4C}.$$

Then, if $C > 1/4$, the chain is transient, and if $C < 1/4$, the chain is recurrent.

12.3 Invariant Measures

Setting. We have an irreducible $\mathbf{P}$ on a countable S.

Definition 12.3. A measure $\mu \geq 0$ on S is **invariant** if $\boldsymbol{\mu}\mathbf{P} = \boldsymbol{\mu}$, that is, $\sum_i \mu(i) p_{i,j} = \mu(j)\ \forall j$.

Note: We may have $\mu(S) = \infty$. Ignore the trivial case $\mu \equiv 0$.

If μ is invariant, then $c\mu$ is invariant, $0 < c < \infty$.

If invariant μ has $\mu(S) = 1$, call it **stationary**.

If invariant μ has $0 < \mu(S) < \infty$, then

$$\hat{\mu}(i) \equiv \frac{\mu(i)}{\mu(S)}$$

is stationary.

Definition 12.4. A general process $(X_n, n = 0, 1, 2, \dots)$ is **stationary** if $\forall n \geq 1$,

$$(X_n, X_{n+1}, \dots) \stackrel{\text{d}}{=} (X_0, X_1, \dots).$$

If $(X_n, n \geq 0)$ is a MC and $\text{dist}(X_0)$ is a stationary distribution, then the process $(X_n, n \geq 0)$ is stationary.

Aside. If μ is invariant, $\mu(S) = \infty$, take (at time 0) independent Poisson$(\mu(i))$ particles at i and run each particle as an independent MC. This particle process is stationary.

$\mu_n = \text{dist}(X_n)$ always evolves as $\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n-1}\mathbf{P}$.

Two Special Settings. $\mu(S) \leq \infty$.

1. $\mu \equiv 1$ is invariant $\iff \sum_i p_{i,j} = 1 \; \forall j \iff$ **doubly stochastic matrix**.
2. If $\mu(x)p(x,y) = \mu(y)p(y,x) \; \forall x, y$, then μ is invariant (**reversible** case).

Proof.

$$(\boldsymbol{\mu}\mathbf{P})_y = \sum_x \mu(x)p(x,y) = \sum_x \mu(y)p(y,x) = \mu(y).$$ □

Example 12.5 (Simple RW on $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$)**.**

$$p(x, x+1) = p,$$
$$p(x, x-1) = q = 1 - p.$$

What is an invariant μ?

$\mathbf{P}$ *is* doubly stochastic: $\mu(i) \equiv 1$ is invariant.

$\mu(x) = (p/q)^x$ is a reversible invariant measure.

For $p \neq 1/2$, the chain is transient and has 2 different σ-finite invariant measures.

Example 12.6 (Birth-Death Chain on $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$)**.**

$$p(i, i+1) = p_i > 0,$$
$$p(i, i-1) = q_i = 1 - p_i > 0, \quad i \geq 1.$$

This has the reversible invariant measure

$$\mu(i) = \prod_{j=1}^{i} \frac{p_{j-1}}{q_j}.$$

Check:

$$\mu(i)p_i = \mu(i+1)q_{i+1} \iff \frac{\mu(i+1)}{\mu(i)} = \frac{p_i}{q_{i+1}}.$$

This is the *unique* invariant measure (up to scaling). Looking at $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{P}$ at 0:

$$\mu(0) = \mu(0)p(0,0) + \mu(1)p(1,0)$$
$$\mu(1) = \mu(0)p(0,1) + \mu(1)p(1,1) + \mu(2)p(2,1).$$

The first equation determines $\mu(1)$ in terms of $\mu(0)$, and the second equation determines $\mu(2)$ in terms of $\mu(0)$, and so forth.

Lecture 13

February 28

13.1 Periodicity

Consider the directed graph associated with $\mathbf{P}$ on countable S.

For state x, $d(x) \stackrel{\text{def}}{=}$ greatest common divisor of $\{n : p^n_{x,x} > 0\}$.

Theorem 13.1 (Text, Exercise)**.** *Suppose that the Markov chain is irreducible.*

(a) $d(x) = d \geq 1$ *for each* $x \in S$.

The case $d = 1$ *is* ***aperiodic****, and the case* $d \geq 2$ *is* ***periodic*** *with period* d.

(b) $\exists n(x) < \infty$ *such that* $p^n_{x,x} > 0$ *for all* $n \geq n(x)$ *with* $d \mid n$.

(c) S *can be partitioned into* d *"cyclic classes"* $C_0, C_1, \ldots, C_{d-1}$ *such that if* $x \in C_u$, $\mathbb{P}_x(X_n \in C_v)$ *is* 1 *if* $n = v - u$ *modulo* d, 0 *if not.*

(d) If the Markov chain is aperiodic, $\forall (x,y)\ \exists n(x,y)$ *such that* $p^n_{x,y} > 0\ \forall n \geq n(x,y)$.

(e) If the period is $d \geq 2$, *then* $\mathbf{P}^d$ *defines a MC on each* C_u, *which is irreducible on* C_u.

(f) If $\exists x$ *with* $p_{x,x} > 0$, *then by (a),* $d = 1$ *and the chain is aperiodic.*

13.2 Existence of Invariant Measures

If μ and ν are PMs on measurable S, the variation distance is $\|\mu - \nu\| \stackrel{\text{def}}{=} \sup_A |\mu(A) - \nu(A)|$. If S is countable, then

$$\|\mu - \nu\| = \frac{1}{2} \sum_i |\mu(i) - \nu(i)|$$

and $\|\mu_n - \mu_\infty\| \to 0 \iff \mu_n(i) \to \mu_\infty(i)\ \forall i \in S$.

$\boldsymbol{\mu}\mathbf{P}$ is $\text{dist}(X_1)$ when $\mu = \text{dist}(X_0)$.

Lemma 13.2. *For a MC with transition matrix* $\mathbf{P}$,

$$\|\mu P - \nu P\| \leq \|\mu - \nu\|.$$

Proof.

$$\begin{aligned}\text{Left} = \frac{1}{2}\sum_i \left|\sum_j (\mu(j)-\nu(j))p_{j,i}\right| &\le \frac{1}{2}\sum_i\sum_j |\mu(j)-\nu(j)|p_{j,i} \\ &= \frac{1}{2}\sum_j |\mu(j)-\nu(j)| = \|\mu-\nu\|,\end{aligned}$$

since $\sum_i p_{j,i} \equiv 1$. ☐

Lemma 13.3. *Let $(X_n, 0 \le n < \infty)$ be the (μ_0, P) chain. Write $\boldsymbol{\mu}_n = \operatorname{dist}(X_n) = \boldsymbol{\mu}_0\mathbf{P}^n$. If $\|\mu_n - \mu_\infty\| \to 0$ for some PM μ_∞, then μ_∞ is a stationary distribution for $\mathbf{P}$, $\boldsymbol{\mu}_\infty = \boldsymbol{\mu}_\infty\mathbf{P}$.*

Proof.

$$\begin{aligned}\|\mu_\infty P - \mu_n P\| &\overbrace{\le}^{13.2} \|\mu_\infty - \mu_n\| \to 0 \quad \text{as } n\to\infty \\ \|\mu_\infty P - \mu_{n+1}\| &\to 0 \quad \text{as } n \to \infty \\ \|\mu_\infty - \mu_{n+1}\| &\to 0 \quad \text{as } n\to\infty\end{aligned}$$

By the Triangle Inequality,

$$\begin{aligned}\|\mu_\infty - \mu_\infty P\| &\to 0 \quad \text{as } n \to \infty \\ &= 0.\end{aligned}$$

☐

So, the possible $n \to \infty$ limit distributions are exactly the stationary distributions.

Let $T_x = T_x^+ = \min\{n \ge 1 : X_n = x\}$. Fix state b. Define

$$\mu(b,x) = \mathbb{E}_b[\text{number of visits to } x \text{ before } T_b] = \mathbb{E}_b\sum_{n=0}^{\infty} 1_{(X_n = x, T_b > n)},$$

which implies that $\mu(b,b) = 1$. $\mu(b,\cdot)$ is a measure on S and $\mathbb{E}_b T_b = \mu(b,S) \le \infty$.

Proposition 13.4 (No Assumptions)**.** *Consider these equations for an unknown measure μ:*

$$\mu(y) = \sum_x \mu(x)p(x,y)\ \forall y \ne b, \qquad \mu(b) = 1. \tag{13.1}$$

Then, $\mu(b,\cdot)$ is the minimal solution of (13.1) and $\mathbb{P}_b(T_b < \infty) = \sum_x \mu(b,x)p(x,b)\ \forall x$.

Proof. Let the matrix $\mathbf{K}$ be the "chain killed at T_b". $K_{x,y} = P_{x,y}$ for $y \ne b$ and $K_{x,y} = 0$ for $y = b$. Write $\alpha_n(y) = \mathbb{P}_b(X_n = y, T_b > n)$. Check that $\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n\mathbf{K}$. $\alpha_0(y) = \delta_b(y) = 1_{(y=b)}$. Therefore, $\boldsymbol{\alpha}_n = \boldsymbol{\delta}_b\mathbf{K}^n$. By definition, $\mu(b,y) = \sum_{n=0}^\infty \alpha_n(y)$, so $\boldsymbol{\mu}(b,\cdot) = \sum_{n=0}^\infty \boldsymbol{\delta}_b\mathbf{K}^n$. Rewrite (13.1) as $\boldsymbol{\mu} = \boldsymbol{\delta}_b + \boldsymbol{\mu}\mathbf{K}$. Hence, $\boldsymbol{\mu}(b,\cdot)$ satisfies (13.1).

Let $\boldsymbol{\mu}$ be some solution of (13.1). Then, $\boldsymbol{\mu} = \boldsymbol{\delta}_b + (\boldsymbol{\delta}_b + \boldsymbol{\mu}\mathbf{K})\mathbf{K} = \boldsymbol{\delta}_b + \boldsymbol{\delta}_b\mathbf{K} + \boldsymbol{\mu}\mathbf{K}^2$. Inductively, $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{K}^{m+1} + \sum_{n=0}^m \boldsymbol{\delta}_b\mathbf{K}^n \ge \sum_{n=0}^m \boldsymbol{\delta}_b\mathbf{K}^n \uparrow \sum_{n=0}^\infty \boldsymbol{\delta}_b\mathbf{K}^n = \boldsymbol{\mu}(b,\cdot)$, which implies $\boldsymbol{\mu} \ge \boldsymbol{\mu}(b,\cdot)$.

$$\mathbb{P}_b(T_b < \infty) = \sum_{n=0}^{\infty} \mathbb{P}_b(T_b = n+1) = \sum_{n=0}^{\infty} \sum_y \mathbb{P}_b(X_n = y, T_b = n+1)$$

However, $\mathbb{P}_b(X_n = y, T_b = n+1) = \mathbb{P}_b(T_b > n, X_n = y, T_b = n+1) = \alpha_n(y)p(y,b)$ by conditioning on $\mathcal{F}_n$, so:

$$\begin{aligned} &= \sum_y \left(\sum_{n=0}^{\infty} \alpha_n(y) \right) p(y,b) \\ &= \sum_y \mu(b,y)p(y,b). \end{aligned}$$

□

Lemma 13.5. *Suppose the Markov chain is irreducible. Suppose $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{P}$, where $0 \le \mu(x) \le \infty$.*

(a) If $\mu(b) = 0$ for some b, then $\mu \equiv 0$.

(b) If $\mu(b) = \infty$ for some b, then $\mu \equiv \infty$.

Proof. Fix x. There exist n, m such that $p^n_{x,b} > 0$ and $p^m_{b,x} > 0$. $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{P}^n$ implies that $\mu(b) \ge \mu(x)p^n_{x,b}$, so if $\mu(b) = 0$, then $\mu(x) = 0$. $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{P}^m$ implies that $\mu(x) \ge \mu(b)p^m_{b,x}$, which implies that if $\mu(b) = \infty$, then $\mu(x) = \infty$. □

Theorem 13.6. *Suppose the Markov chain is irreducible and recurrent. Then, there exists an invariant μ which satisfies $0 < \mu(x) < \infty\ \forall x \in S$. This μ is unique up to scaling. Either*

*(i) $\mu(S) = \infty$ and $\mathbb{E}_x T_x = \infty\ \forall x$ (**null-recurrent**), or*

*(ii) $\mu(S) < \infty$ and $\mathbb{E}_x T_x < \infty\ \forall x$ (**positive-recurrent**).*

Proof. Fix b. Define $\mu(\cdot) = \mu(b,\cdot)$, which satisfies (13.1). Then, $\mu(x) = (\mu P)(x)$ for $x \ne b$ and $(\mu P)(b) = \mathbb{P}_b(T_b < \infty) = 1 = \mu(b)$, since the chain is recurrent. Therefore, $\boldsymbol{\mu} = \boldsymbol{\mu}\mathbf{P}$ is invariant. Since $\mu(b) = 1$, 13.5 and the assumption that the chain is irreducible implies $0 < \mu(x) < \infty\ \forall x$.

Why is μ unique? Suppose $\hat{\mu}$ is invariant: rescale to make $\hat{\mu}(b) = 1$. By minimality in 13.4, $\hat{\boldsymbol{\mu}} \ge \boldsymbol{\mu}(b,\cdot)$. Since they are both invariant, $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(b,\cdot) \ge \mathbf{0}$ is invariant and equals 0 at b. Then, 13.5 implies that $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(b,\cdot) \equiv \mathbf{0}$, so $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(b,\cdot)$.

Consider some invariant μ. $\mu(x,\cdot)$ is a scaled version of μ, so $\mu(x,\cdot) = c_x\mu(\cdot)$, $0 < c_x < \infty$, by uniqueness.

$$\mathbb{E}_x T_x = \sum_y \mu(x,y) = c_x \mu(S),$$

which implies that either (i) or (ii) occur. □

Corollary 13.7. *A finite-state irreducible chain is positive-recurrent.*

Proof. Last class, we showed that the chain is recurrent, so an invariant μ exists, so

$$\mu(S) = \sum_{x \in S} \mu(x) < \infty.$$

Therefore, we are in case (ii). □

Lecture 14

March 2

14.1 Stationary Measures

Consider $\mathbf{P}$ on countable S.

Proposition: Consider the equations

$$\mu(y) = \sum_x \mu(x)p(x,y)\ \forall y \neq b, \qquad \mu(b) = 1. \tag{14.1}$$

Then, $\mu(b,\cdot) = \mathbb{E}_b[\text{number of visits to } \cdot \text{ before } T_b]$ is the minimal solution to (14.1) and $\mathbb{P}_b(T_b < \infty) = \sum_x \mu(b,x)p(x,b)$.

$\mu(b,\cdot)$ is the "**b-block occupation measure**".

Theorem: Suppose the Markov chain is irreducible and recurrent. Then, there exists an invariant μ, unique up to scaling. Either

(i) $\mu(S) = \infty$ and $\mathbb{E}_x T_x = \infty\ \forall x$ (null-recurrent) or

(ii) $\mu(S) < \infty$ and $\mathbb{E}_x T_x < \infty\ \forall x$ (positive-recurrent).

In case (ii),

$$\pi(x) = \frac{\mu\{x\}}{\mu(S)}$$

is a stationary distribution.

Theorem 14.1. *Suppose the Markov chain is irreducible. Then, it is positive-recurrent if and only if a stationary distribution π exists. If so, then*

$$\pi(x) = \frac{1}{\mathbb{E}_x T_x}.$$

Mystery. Starting by defining

$$\pi(x) = \frac{1}{\mathbb{E}_x T_x}$$

and showing π is stationary is not so easy.

Proof. Suppose that a stationary distribution π exists. Fix b. Define

$$\mu(j) = \frac{\pi(j)}{\pi(b)}.$$

μ is invariant, $\mu(b) = 1$. "Minimality" in 13.4 implies that $\mu(b, y) \le \mu(y)\ \forall y$. Therefore,

$$\mathbb{E}_b T_b = \sum_y \mu(b, y) \le \sum_y \mu(y) = \frac{1}{\pi(b)} < \infty,$$

so the chain is positive-recurrent.

Suppose that the chain is positive-recurrent. 13.6 implies that π exists.

Fix b. We know that $\mu(b, \cdot)$ is invariant, so

$$\pi(x) = \frac{\mu(b, x)}{\mu(b, S)}$$

is the unique stationary distribution. This is true for $x = b$, so

$$\pi(b) = \frac{\mu(b, b)}{\mu(b, S)} = \frac{1}{\mathbb{E}_b T_b}.$$ □

Warning. Suppose S is infinite, the chain is irreducible, and an invariant μ exists with $\mu(S) = \infty$. This does *not* imply that the chain is recurrent. Also, this does *not* imply that the invariant measure is unique up to scaling.

Example 14.2 (SRW on $\mathbb{Z}$)**.**

$$p(i, i+1) = p, \qquad p(i, i-1) = q = 1 - p.$$

For $p \ne 1/2$, there are two invariant measures:

$$\mu(i) \equiv 1, \qquad \mu(i) = \left(\frac{p}{q}\right)^i.$$

Also, the chain is transient.

For $p = 1/2$, the chain is recurrent and there is a unique (up to scaling) invariant $\mu(i) \equiv 1$.

Example 14.3 (Reflecting RW on $\mathbb{Z}^+$)**.**

$$\begin{aligned} p(i, i+1) &= p, \quad i \ge 1, \\ p(i, i-1) &= 1 - p, \quad i \ge 1, \\ p(0, 1) &= 1. \end{aligned}$$

If $p > 1/2$, the chain is transient.

If $p = 1/2$, the chain is null-recurrent.

If $p < 1/2$, the chain is positive-recurrent.

14.2 Convergence to the Stationary Distribution

Know. If $\exists \mu_0$ such that $\mathbb{P}_{\mu_0}(X_n = j) \xrightarrow{n\to\infty} \pi(j)\ \forall j$ for some probability distribution π, then π is stationary.

Theorem 14.4 (The MC Convergence Theorem)**.** *Suppose the chain is irreducible and positive-recurrent, so the stationary π exists. If the chain is also aperiodic, then $\mathbb{P}_{\mu_0}(X_n = j) \xrightarrow{n\to\infty} \pi(j)\ \forall j\ \forall \mu_0$.*

Proof. Fix μ_0. We shall construct a Markov chain on $S \times S$, call it $((X_n, Y_n), n = 0, 1, \dots)$, such that

(i) $(X_n, n \geq 0)$ is the $(\mu_0, \mathbf{P})$-chain,

(ii) $(Y_n, n \geq 0)$ is the stationary $(\pi, \mathbf{P})$-chain,

(iii) $X_n = Y_n\ \forall n \geq T$, where $T < \infty$ a.s.

This will prove the theorem because

$$|\mathbb{P}_{\mu_0}(X_n = j) - \pi(j)| = |\mathbb{P}_{\mu_0}(X_n = j) - \mathbb{P}(Y_n = j)| \leq \mathbb{P}(X_n \neq Y_n) \leq \mathbb{P}(T > n) \to 0 \qquad \text{as } n \to \infty.$$

This is the **MC coupling method**.

The transition matrix on $S \times S$ is

$$\begin{aligned} (x_1, y_1) &\to (x_2, y_2) \text{ with probability } p(x_1, x_2)p(y_1, y_2), \quad x_1 \neq y_1, \\ (x, x) &\to (y, y) \text{ with probability } p(x, y). \end{aligned}$$

The initial distribution is $\mu_0 \otimes \pi$. Two particles initially move as independent MCs, but after meeting, they stick together and move as a single MC.

Fussy argument: why is $(X_n, n \geq 0)$ Markov?

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, Y_n = y_n, \text{past of both proceses}) &= \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, Y_n = y_n) \\ &\underbrace{=}_{\text{form of TM}} \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n). \end{aligned}$$

Condition on the past of X_n.

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \text{past of } X) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n),$$

which is the Markov property for (X_n).

Define $T_{\text{meet}} = \min\{n : X_n = Y_n\}$. Then, $X_n = Y_n\ \forall n \geq T_{\text{meet}}$. It is enough to prove $T_{\text{meet}} < \infty$ a.s. Consider $((\hat{X}_n, \hat{Y}_n), n \geq 0)$ with $(\hat{X}_n)$ the $(\mu_0, \mathbf{P})$-chain, $(\hat{Y}_n)$ the $(\pi, \mathbf{P})$-chain, independent. This is a **product chain**. The distribution of T_{meet} is the same.

Let $\hat{\mathbf{Q}}$ be the transition matrix for the product chain. $\mathbf{P}$ is aperiodic, so

$$\hat{\mathbf{Q}}^{(n)}((x_0, y_0), (x_n, y_n)) = p^{(n)}_{x_0, x_n} p^{(n)}_{y_0, y_n} > 0$$

for large n by aperiodicity of $\mathbf{P}$. Hence, $\hat{\mathbf{Q}}$ is irreducible.

It is easy to see that $\pi \otimes \pi$ is invariant and stationary for $\hat{\mathbf{Q}}$. 14.1 implies that the product chain $\hat{\mathbf{Q}}$ is positive-recurrent. Take state (b, b). $T_{(b,b)} < \infty$ a.s. in the product chain, so $T_{\text{meet}} \leq T_{(b,b)} < \infty$ a.s. in the product chain, so also in the coupled chain. □

Note for later: In order to show

$$|\mathbb{P}_\mu(X_n = x) - \mathbb{P}_\nu(X_n = x)| \to 0 \ \forall x \qquad \text{as } n \to \infty,$$

it is enough to show that the product chain is irreducible and recurrent.

Proposition 14.5. *Suppose that the chain is irreducible and not positive-recurrent. Then,*

$$\mathbb{P}_{\mu_0}(X_n = j) \xrightarrow{n\to\infty} 0 \quad \forall j \quad \forall \mu_0.$$

Proof. Reduce to the aperiodic case. First, suppose that the chain is transient.

$$\sum_n \mathbb{P}_\mu(X_n = j) = \mathbb{E}_\mu N(j) \le 1 + \mathbb{E}_j N(j) = 1 + \frac{1}{1 - \rho_{j,j}} < \infty,$$

by transience. Therefore, $\mathbb{P}_\mu(X_n = j) \to 0$.

So, suppose that the chain is null-recurrent. Consider the product chain. Suppose that the product chain is transient. As above, $\mathbb{P}_{\mu\otimes\mu}(X_n = j, Y_n = j) \to 0$, so $(\mathbb{P}_\mu(X_n = j))^2 \to 0$. Suppose that $\hat{\mathbf{Q}}$ is recurrent. If the result is false, then $\exists \mu_0$ $\exists b$ $\exists$subsequence (j_n) such that

$$\mathbb{P}_\mu(X_{j_n} = b) \to \alpha_b > 0.$$

By compactness, there exists a subsequence k_n such that

$$\mathbb{P}_\mu(X_{k_n} = y) \to \text{some } \alpha_y \ge 0 \quad \forall y. \tag{14.2}$$

By the coupling argument, (14.2) holds for all μ. [See notes.] This implies that (α_y) is a stationary distribution, so the chain is positive-recurrent. □

Lecture 15

March 7

15.1 Coupling & Mixing Times

For PMs μ, ν on countable S,

$$\|\mu - \nu\| = \frac{1}{2}\sum_{s}|\mu(s) - \nu(s)|$$

$$\overbrace{=}^{\text{Lemma}} \inf\{\mathbb{P}(X \neq Y) : \text{over joint distributions } (X, Y) \text{ with } \operatorname{dist}(X) = \mu, \operatorname{dist}(Y) = \nu\}.$$

Consider an irreducible, positive-recurrent $\mathbf{P}$ with stationary distribution π. Suppose that we construct $((X_n, Y_n), n \geq 0)$ such that:

- (X_n) is the $(\mu_0, \mathbf{P})$-chain. Write $\mu_n = \operatorname{dist}(X_n)$.
- (Y_n) is the $(\pi, \mathbf{P})$-chain.
- $X_n = Y_n$ for all $n \geq T$ (for some T).

Then, $\|\mu_n - \pi\| \leq \mathbb{P}(X_n \neq Y_n) \leq \mathbb{P}(T > n)$. This is the **MC coupling inequality**.

Note: We did not assume that (X_n, Y_n) is Markov or T is a stopping time, but in almost every example, these hold.

15.1.1 Card Shuffling by Random Transposition

Example 15.1 ("Card Shuffling by Random Transposition")**.** Consider a deck of C cards. Rule: pick two cards uniformly, independently (they may be the same). Interchange them.

This is a MC on the state space of all $C!$ decks. The $\mathbf{P}$ is symmetric, so π is uniform. $\mathbf{P}$ is aperiodic because $p_{\mathbf{x},\mathbf{x}} \geq 0$. $\mathbf{P}$ is irreducible by group theory. So (for arbitrary μ_0), the convergence theorem implies $\|\mu_n - \pi\| \to 0$ as $n \to \infty$ (C is fixed). How large must n be (in terms of C) for $\|\mu_n - \pi\|$ to be small? This is the **mixing time**.

We will show a coupling with $\mathbb{E}T \leq C^2$. Then,

$$\|\mu_n - \pi\| \leq \mathbb{P}(T > n) \leq \frac{C^2}{n},$$

so order C^2 shuffles are enough. The correct mixing time is order $C \log C$.

X deck	Y deck
e	a
b	f
c	c
a	e
d	d
f	b

The following rule on **P** is the same as the previous rule:

- Pick the card label uniformly at random.
- Pick a position uniformly at random.
- Switch the card with the position.

The rule for coupling is: make the same choices in both decks.

Suppose we pick card a and position 3.

X deck	Y deck
e	c
b	f
a	a
c	e
d	d
f	b

Instead, if we pick card b and position 4:

X deck	Y deck
e	a
a	f
c	c
b	b
d	d
f	e

We will study Z_n, the number of unmatched cards. In our first choice, we went from $Z_n = 4$ to $Z_{n+1} = 4$. In our second choice, we went to $Z_{n+1} = 3$.

Easy:

$$\begin{aligned} &Z_{n+1} \le Z_n \text{ always,} \\ &Z_{n+1} \le Z_n - 1 \text{ if the position and card were both unmatched.} \end{aligned} \tag{15.1}$$

Study $T = \min\{n : Z_n = 0\}$. Write $S_m = \min\{n : Z_n \le m\}$. Then, $T = S_0 = S_1$. (15.1) implies that

$$\mathbb{P}(Z_{n+1} \le Z_n - 1 \mid Z_n = m, \text{past}) \ge \left(\frac{m}{C}\right)^2,$$

so

$$\mathbb{E}[S_{m-1} - S_m] \le \frac{1}{(m/C)^2} = \frac{C^2}{m^2},$$

where $S_C = 0$. Hence,

$$\mathbb{E}T = \mathbb{E}S_1 = \sum_{m=2}^{C} \mathbb{E}[S_{m-1} - S_m] \le C^2 \sum_{m=2}^{C} \frac{1}{m^2} \le C^2 \left(\frac{\pi^2}{6} - 1\right) \le C^2.$$

Comment: Use the structure of $\mathbf{P}$ to try to construct a coupling so that some notion like Z_n ("distance" between states) tends to decrease.

15.2 Ergodic Theorem for Markov Chains

Theorem 15.2 (Ergodic Theorem for Markov Chains)**.** *Consider an irreducible, positive-recurrent MC. Let π be the stationary distribution. Take $f : S \to \mathbb{R}$ such that $\sum_x \pi(x)|f(x)| < \infty$. Then,*

$$\frac{1}{t} \sum_{n=1}^{t} f(X_n) \xrightarrow[a.s.]{} \bar{f} := \sum_x \pi(x) f(x) \qquad \text{as } t \to \infty.$$

Proof. We can reduce to the IID SLLN. We can assume that $f \ge 0$ (write $f = f^+ - f^-$). Fix state b. Let T^j be the time of the jth visit to b.

If we consider a typical sequence for the chain:

$$xz \underbrace{\overbrace{b}^{T^1}\ wae}_{\Lambda_1} \underbrace{\overbrace{b}^{T^2}\ c}_{\Lambda_2} \underbrace{\overbrace{b}^{T^3}\ qrsaw}_{\Lambda_3} b \dots$$

Define $\Lambda_j = (X(T^j), X(T^j + 1), \dots, X(T^{j+1} - 1))$. The Strong Markov Property implies that the $(\Lambda_j, j \ge 1)$ are IID. Λ_j takes values in $\bigcup_{d=1}^{\infty} S^d = S^{(\infty)}$. Define $R_j = \sum_{i=T^{j-1}}^{T^j - 1} f(X_i)$, the sum of the f-values over Λ_{j-1}. The SMP implies $(R_1, R_2, R_3, \dots)$ are IID and $(T^2 - T^1, T^3 - T^2, \dots)$ are IID. Apply the IID SLLN.

$$\frac{1}{n} \sum_{i=2}^{n} R_i \to \mathbb{E}R_2 \text{ a.s.} \qquad \text{and} \qquad \frac{1}{n} \sum_{i=1}^{n} (T^i - T^{i-1}) \xrightarrow{\text{a.s.}} \mathbb{E}[T^2 - T^1],$$

$$\frac{1}{n} T^n \to \mathbb{E}[T^2 - T^1] \equiv \mathbb{E}_b T_b^+ = \frac{1}{\pi(b)},$$

where T_b^+ is the return time to b. We can calculate $\mathbb{E}R_2 = \sum_x \mu(b, x) f(x)$. We know that $\mu(b, \cdot)$ is a multiple of $\pi(\cdot)$, so

$$\mu(b, x) = \frac{\pi(x)}{\pi(b)} \implies \mathbb{E}R_2 = \frac{\bar{f}}{\pi(b)} \implies \frac{1}{n} \sum_{i=1}^{n} R_i \to \frac{\bar{f}}{\pi(b)} \text{ a.s.}$$

Now, apply 15.3 with $r(t) = \sum_{i=1}^{t} f(X_i)$, $t_n = T^n$, $r_n = R_n$ (each ω). Conclude that

$$\frac{1}{t} \sum_{i=1}^{t} f(X_i) \xrightarrow[\text{a.s.}]{} \frac{\mathbb{E}R_2}{\mathbb{E}[T^2 - T^1]} = \bar{f}.$$

□

Lemma 15.3 (Deterministic Lemma, 205A)**.** *Let $0 < t_n \uparrow \infty$, $t_n/n \to \bar{t} > 0$. Let $r_i \ge 0$, such that*

$n^{-1}\sum_{i=1}^{n} r_i \to \bar{r} > 0$ and $\sum_{i=1}^{n(t)} r_i \leq r(t) \leq \sum_{i=1}^{n(t)+1} r_i$, where $n(t) = \max\{n : t_n \leq t\}$. Then,

$$\frac{r(t)}{t} \to \frac{\bar{r}}{\bar{t}} \qquad \text{as } t \to \infty.$$

Special Case: Fix y. Set $f(x) = 1_{(x=y)}$. Then,

$$\frac{1}{t} N_t(y) \xrightarrow{\text{a.s.}} \pi(y),$$

where $N_t(y)$ is the number of visits to y before t.

Lecture 16

March 9

16.1 Renewal Reward Theorem

Proposition 16.1. *Let $(X_n, n \geq 0)$ be irreducible and positive-recurrent, where π is the stationary distribution. Fix x. Let $0 < S < \infty$ be a stopping time such that $X_S = x$ a.s. Then,*

$$\underbrace{\mathbb{E}_x \sum_{t=0}^{S-1} 1_{(X_t=y)}}_{\text{number of visits to } y \text{ before } S} = \pi(y)\mathbb{E}_x S.$$

Proof. $S = f(X_0, X_1, X_2, \dots)$ for some f. Define $S_0 = 0$, $S_1 = S$, and

$$S_{j+1} - S_j \stackrel{\text{def}}{=} f(X_{S_j}, X_{S_j+1}, X_{S_j+2}, \dots).$$

$R_j = \sum_{t=S_{j-1}}^{S_j-1}$ is the number of visits to y during $[S_{j-1}, S_j)$. The Strong Markov Property implies that the blocks $\Lambda_1, \Lambda_2, \dots$ are IID. Therefore, the $(R_1, R_2, \dots)$ are IID and the $(S_j - S_{j-1}, j \geq 1)$ are each IID. By the SLLN,

$$\frac{1}{n}\sum_{i=1}^{n} R_i \to \mathbb{E}R_1 \text{ a.s.}, \qquad \frac{1}{n}S_n \to \mathbb{E}S \text{ a.s.}$$

If $N_t(y) = \sum_{i=0}^{t-1} 1_{(X_i=y)}$, then 15.3 implies that

$$\frac{1}{t}N_t(y) \to \frac{\mathbb{E}_x R_1}{\mathbb{E}_x S},$$

but we know from the MC Ergodic Theorem, 15.2, that the LHS converges to $\pi(y)$ a.s., so the RHS and $\pi(y)$ are equal. □

We can replace the "x" by a PM θ. (Use the "general" ergodic theorem.)

16.2 Finite Markov Chains: Matrix Theory

Consider an irreducible, positive-recurrent, aperiodic chain.

$$p^t(x, y) = \mathbb{P}_x(X_t = y) \to \pi(y) \qquad \text{as } t \to \infty.$$

If S is finite, then (easy) the convergence is geometrically fast.

$$\sum_{t=0}^{\infty}(p^t(x,y)-\pi(y)) = z(x,y), \qquad \text{say.}$$

(The sum converges.) Assume $z_{x,y} = \sum_{t=0}^{\infty}(p^t(x,y)-\pi(y))$ exists. The matrix $\mathbf{Z}$ is determined by $\mathbf{P}$. How?

Let $\mathbf{I}$ be the identity matrix and $\mathbf{\Pi}$ be the matrix where $\Pi_{x,y} = \pi_y$. Saying $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ means that $\mathbf{\Pi}\mathbf{P} = \mathbf{\Pi}$.

$$\mathbf{Z} \stackrel{\text{def}}{=} \sum_{t=0}^{\infty}(\mathbf{P}^t - \boldsymbol{P}) \implies \mathbf{ZP} = \mathbf{Z} - (\mathbf{I}-\mathbf{\Pi}) \implies \mathbf{Z}(\mathbf{I}-\mathbf{P}) = \mathbf{I}-\mathbf{\Pi},$$

so $\mathbf{Z} = (\mathbf{I}-\mathbf{\Pi})(\mathbf{I}-\mathbf{P})^{-1}\ldots$ but $\boldsymbol{\pi}(\mathbf{I}-\mathbf{P}) = \mathbf{0}$ implies that $(\mathbf{I}-\mathbf{P})$ is *not* invertible. $\mathbf{Z}$ can be interpreted as a "generalized inverse". Kemeny-Snell, *Finite Markov Chains* treats this topic.

Let $T_x = \min\{n \geq 0 : X_n = x\}$.

Step 1. Let $y \neq x$. Consider $S = \min\{t > T_y : X_t = x\}$. 16.1 implies that $\pi(y)\mathbb{E}_x S = \mathbb{E}_y N_{T_x}(y)$. Note that $\mathbb{E}_x S = \mathbb{E}_x T_y + \mathbb{E}_y T_x$.

Lemma 16.2.

$$\mathbb{E}_x[\textit{number of visits to } x \textit{ before } T_y] = \pi(x)(\mathbb{E}_x T_y + \mathbb{E}_y T_x).$$

Step 2: Fix a constant k, and consider $S = \min\{t \geq k : X_t = x\}$. 16.1 implies

$$\pi(y)(k + \mathbb{E}_{\rho^{(k)}} T_x) = \mathbb{E}_x[\text{number of visits to } y \text{ before } k] + \mathbb{E}_{\rho^{(k)}}[\text{number of visits to } y \text{ before } T_x].$$

Then,

$$\pi(y)\mathbb{E}_{\rho^{(k)}} T_x = \sum_{t=0}^{k-1}(p_{x,y}^{(t)} - \pi(y)) + \mathbb{E}_{\rho^{(k)}}[\text{number of visits to } y \text{ before } T_x].$$

Let $k \to \infty$. $\rho^{(k)} \to \pi$, so $\pi(y)\mathbb{E}_\pi T_x = z_{x,y} + \mathbb{E}_\pi[\text{number of visits to } y \text{ before } T_x]$.

Lemma 16.3.

$$\pi(x)\mathbb{E}_\pi T_x = z_{x,x}.$$

Lemma 16.4.

$$\mathbb{E}_\pi[\textit{number of visits to } y \textit{ before } T_x] = \pi(y)\mathbb{E}_\pi T_x - z_{x,y}.$$

Step 3. Consider $S = \min\{n \geq T_y + k : X_n = x\}$. 16.1 implies that

$$\pi(y)(\mathbb{E}_x T_y + k + \mathbb{E}_{\theta^{(k)}} T_x) = \mathbb{E}_y[\text{number of visits to } y \text{ before } k] + \mathbb{E}_{\theta^{(k)}}[\text{number of visits to } y \text{ before } T_x],$$

$$\pi(y)(\mathbb{E}_x T_y + \mathbb{E}_{\theta^{(k)}} T_x) = \sum_{t=0}^{k-1}(p^t(y,y) - \pi(y)) + \mathbb{E}_{\theta^{(k)}}[\text{number of visits to } y \text{ before } T_x].$$

Let $k \to \infty$.

$$\pi(y)(\mathbb{E}_x T_y + \mathbb{E}_\pi T_x) = z_{y,y} + \underbrace{\mathbb{E}_\pi[\text{number of visits to } y \text{ before } T_x]}_{\pi(y)\mathbb{E}_\pi T_x - z_{x,y}}.$$

Also, 16.3 says $\pi(x)\mathbb{E}_\pi T_x = z_{x,x}$.

The point of this is:

Lemma 16.5.

$$\pi(y)\mathbb{E}_x T_y = z_{y,y} - z_{x,y}.$$

Example 16.6 (Patterns in Coin-Tossing)**.** Fix a sequence, say, $HHTHH$. Toss a fair coin until we see this pattern. What is the expected number of tosses?

In 205A, we had a martingale proof.

We can use a 32-state MC, $(X_n, n \geq 0)$, of overlapping 5-tuples. π is uniform, $\pi(x) = 1/32$. Study $\mathbb{E}_\pi T_x$ for $x = HHTHH$.

$$\begin{aligned} p^{(0)}(x,x) &= 1, \\ p^{(1)}(x,x) &= 0, \\ p^{(2)}(x,x) &= 0, \\ p^{(3)}(x,x) &= \frac{1}{8}, \\ p^{(4)}(x,x) &= \frac{1}{16}, \\ p^{(t)}(x,x) &= \frac{1}{32}, \qquad t \geq 5. \end{aligned}$$

Then,

$$z_{x,x} = \sum_{t=0}^{\infty} \left(p^{(t)}(x,x) - \frac{1}{32} \right) = 1 + \frac{1}{8} + \frac{1}{16} - \frac{5}{32}.$$

Then, by the formula,

$$\mathbb{E}_\pi T_x = \frac{z_{x,x}}{\pi(x)} = 32 z_{x,x} = 32 + 4 + 2 - 5.$$

16.3 The MC CLT & Variance of Sums

Consider a chain on finite S, irreducible and aperiodic, with stationary distribution π. Consider a function $f : S \to \mathbb{R}$ with $\bar{f} = \sum_i \pi_i f(i) = 0$. Write $S_t = \sum_{n=1}^{t} f(X_n)$. We can prove (using IID blocks) that

$$\frac{S_t}{\sqrt{t}} \xrightarrow{\mathrm{d}} \operatorname{Normal}(0, \sigma^2(t)).$$

Instead, we will directly study $\operatorname{var} S_t$. Consider the stationary chain.

$$\begin{aligned} \sigma^2(t) = \lim_{t\to\infty} \frac{\operatorname{var}(S_t)}{t} &= \lim_{t\to\infty} \sum_{u=1}^{t} \sum_{v=1}^{t} \mathbb{E}_\pi[f(X_u)f(X_v)], \qquad u - v = s, \\ &= \sum_{s=-\infty}^{\infty} \mathbb{E}_\pi[f(X_0)f(X_s)], \\ \mathbb{E}_\pi[f(X_0)f(X_s)] &= \sum_i \sum_j f(i)f(j)\pi(i)[p^{(s)}(i,j) - \pi_j] \qquad \text{because } \sum_j \pi_j f(j) = \bar{f} = 0 \end{aligned}$$

$$\sigma^2(t) = \sum_{s=-\infty}^{\infty} \mathbb{E}_\pi[f(X_0)f(X_s)] = \sum_i \sum_j f(i)f(j)\pi(i) z_{i,j}.$$

We are using $\mathbb{E}_\pi[f(X_u)f(X_v)] = \mathbb{E}_\pi[f(X_0)f(X_s)]$. Given a stationary process $(X_0, X_1, X_2, \dots)$, Kolmogorov extension says that there exists a process $(X_n, -\infty < n < \infty)$.

The sum over $s \geq 0$ of $\pi_x(p^{(s)}_{x,y} - \pi_y) = \pi_x z_{x,y}$. The sum over $s \leq 0$ is $\pi_y z_{y,x}$ because

$$\pi(i)p^{(-s)}(i,j) \overbrace{=}^{\text{stationarity}} \pi(j)p^{(s)}(j,i).$$

The sum over $s = 0$ is $\pi_x(\delta_{x,y} - \pi_y)$.

Conclusion: $\sigma^2(t) = f^\top \Gamma f$ for $\Gamma_{i,j} = \pi_i z_{i,j} + \pi_j z_{j,i} - \pi_i(\delta_{i,j} - \pi_j)$ (symmetric).

Lecture 17

March 14

17.1 Martingale Methods for Markov Chains

17.1.1 Harmonic Functions

Setting. (X_n) is an irreducible MC on countable S. $\mathbf{P} = (p(x,y))$. We have $h : S \to [0,\infty)$ and let $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. Suppose $\mathbb{E}h(X_0) < \infty$. Then, $(h(X_n), 0 \le n < \infty)$ is a MG if and only if $h(x) = \sum_y p(x,y)h(y)\ \forall x \in S$.

$$\begin{aligned}\mathbb{E}[h(X_{n+1}) \mid \mathcal{F}_n] &\underbrace{=}_{\text{Markov}} \mathbb{E}[h(X_{n+1}) \mid X_n] \\ &= \mathbb{E}[h(X_{n+1}) \mid X_n = x] = \sum_y p(x,y)h(y) = h(x) \qquad \text{on } \{X_n = x\} \\ &= h(X_n) \qquad \text{a.s.}\end{aligned}$$

which is the MG property.

h is **harmonic** w.r.t. $\mathbf{P}$.

$(h(X_n), 0 \le n < \infty)$ is a super-MG if and only if $h(x) \ge \sum_y p(x,y)h(y)$. h is **superharmonic** w.r.t. $\mathbf{P}$.

Lemma 17.1. *If (X_n) is recurrent and $h \ge 0$ is superharmonic, then h is constant.*

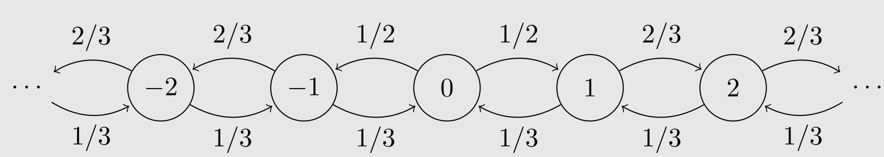
Proof. $h(X_n) \ge 0$ is a super-MG, so (MG convergence) $h(X_n) \to$ some $H_\infty \ge 0$ a.s.

For states y_1, y_2, X_n visits y infinitely often, so $H_\infty = h(y_1) = h(y_2)$ a.s., so h is constant. ∎

Fact. A transient chain may or may not have the property

$$\text{there exists a non-constant harmonic } h \text{ with } 0 \le h \le 1. \tag{17.1}$$

Example 17.2. Consider the following chain.



$\mathbb{P}(X_n \to \infty \text{ or } X_n \to -\infty) = 1$.

$h(x) \stackrel{\text{def}}{=} \mathbb{P}_x(X_n \to +\infty)$. Note that

$$h(X_{n+1}) = \mathbb{P}_x(X_n \to \infty \mid X_m, X_{m-1}, X_{m-2}, \dots) \equiv \mathbb{P}(A \mid \mathcal{F}_m) \text{ is always a MG.}$$

Hence, h is harmonic. It is easy to see that $h(x) \to 1$ as $x \to \infty$ and $h(x) \to 0$ as $x \to -\infty$.

Example 17.3. $(\xi_i, i \geq 1)$ are IID $\mathbb{Z}^d$-valued. $X_n = \sum_{i=1}^n \xi_i$ is a MC on $\mathbb{Z}^d$.

Suppose h is harmonic, $0 \leq h \leq 1$. $h(X_n)$ is a MG, so $h(X_n) \xrightarrow{\text{a.s.}} H_\infty$, say. H_∞ is in the exchangeable σ-field of $(\xi_i, 1 \leq i < \infty)$, which is trivial by the Hewitt-Savage 0-1 law. So, H_∞ is constant. Since $h(X_n)$ is a MG, then $h(X_n) = \mathbb{E}[H_\infty \mid \mathcal{F}_n]$ is constant.

Remark. "Martin boundary theory" discusses extreme harmonic functions and the number of ways that a countable-state chain can go to infinity.

17.1.2 Mean Hitting Times

Lemma 17.4. *Fix $A \subseteq S$. $T_A = \min\{n \geq 0 : X_n \in A\}$.*

(a) Suppose $h(x) \stackrel{\text{def}}{=} \mathbb{E}_x T_A < \infty \ \forall x \in S$. Define $Y_n = h(X_n) + n$. Then, $(Y_{n \wedge T_A}, 0 \leq n < \infty)$ is a MG.

(b) If $0 \leq h < \infty$ satisfies $h(x) \geq \sum_y p(x,y)h(y) + 1 \ \forall x \notin A$, then $\mathbb{E}_x T_A \leq h(x) \ \forall x$.

Proof. (a) For $x \notin A$, then condition on the first step $h(x) = 1 + \mathbb{E}_x h(X_1) = 1 + \sum_y p(x,y)h(y)$. Then, $Y_0 = \mathbb{E}[Y_1 \mid X_0]$ on $\{T_A > 0\}$. By the same argument, $Y_n = \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]$ on $\{T_A > n\}$, which implies that $(Y_{n \wedge T_A}, n \geq 0)$ is a MG.

(b) Given such an h, write $Y_n = h(X_n) + n$. The above argument implies that $(Y_{n \wedge T_A}, n \geq 0)$ is a super-MG. By MG convergence, $Y_{n \wedge T_A} \to$ some Z a.s. as $n \to \infty$ and $\mathbb{E}Z \leq \mathbb{E}Y_0$. However, $Y_n \to \infty$ as $n \to \infty$, so $T_A < \infty$ a.s. So, $Z = Y_{T_A} \geq T_A$.

$$\mathbb{E}_x T_A \leq \mathbb{E}_x Z \leq \mathbb{E}_x Y_0 = h(x).$$ □

17.1.3 Criteria for Recurrence on Infinite S

We can use these ideas to prove recurrence/transience.

Idea. $h(x)$ is the distance from x to a reference state. If h tends to decrease, then we have recurrence. If h tends to increase, then we have transience.

Proposition 17.5. *If there exists $h : S \to [0, \infty)$ and a finite $B \subseteq S$ such that*

(i) $h(x) \geq \sum_y p(x,y)h(y) \ \forall x \notin B$,

(ii) $|\{x : h(x) \leq M\}| < \infty \ \forall M < \infty$,

then the chain is recurrent.

Proof. (i) implies that $h(X_{n \wedge T_B})$ is a super-MG, so $h(X_{n \wedge T_B}) \xrightarrow{\text{a.s.}}$ some Z as $n \to \infty$. By contradiction: X_n visits each state only finitely often. So, (ii) implies $h(X_n) \to \infty$ a.s. Therefore, $T_B < \infty$ a.s. Since this is true for every initial state, $\mathbb{P}(X_n \text{ visits } B \text{ infinitely often}) = 1$. However, B is finite, so X_n visits

B only finitely often, which is a contradiction. □

Proposition 17.6. *As above, but strengthen (i) to* $\exists \delta > 0$ *such that*

(iii) $h(x) \geq \sum_y p(x,y)h(y) + \delta \ \forall x \notin B$

and also assume

(iv) $|\{y : p(x,y) > 0\}| < \infty$ *for* $x \in B$.

Then, the chain is positive-recurrent.

Proof. We can assume $\delta = 1$ ($h \leftarrow h/\delta$). By 17.4, $\mathbb{E}_x T_B \leq h(x)$. Let $T_B^+ = \min\{n \geq 1 : X_n \in B\}$.

$$\begin{aligned} (x \notin B) \quad & \mathbb{E}_x T_B^+ = \mathbb{E}_x T_B \leq h(x), \\ (x \in B) \quad & \mathbb{E}_x T_B^+ \leq 1 + \max\{h(y) : p(x,y) > 0\} < \infty \qquad \text{by (iv)}. \end{aligned}$$

Consider Z_m = "the chain watched only on B" $= X_{S_m}$, where S_m is the time of the mth visit to B. (Z_m) is an irreducible, finite-state chain, so it has a stationary distribution $\hat{\pi}$.

$$\begin{aligned} \mu(x,y) &\stackrel{\text{def}}{=} \mathbb{E}_x \sum_{n=0}^{\infty} 1_{(X_n = y, n < T_B^+)} < \infty, \\ \pi(y) &\stackrel{\text{def}}{=} \sum_{x \in B} \hat{\pi}(x)\mu(x,y). \end{aligned}$$

From the homework, π is an invariant measure for $\mathbf{P}$ and

$$\begin{aligned} \sum_{y \in S} \pi(y) &= \sum_{y \in S} \sum_{x \in B} \hat{\pi}(x)\mu(x,y) \\ &= \sum_{x \in B} \hat{\pi}(x) \mathbb{E}_x T_B^+ < \infty, \end{aligned}$$

so the chain is positive-recurrent. □

Later Homework. Show the corresponding sufficient condition for transience.

Lecture 18

March 16

18.1 Rejection Sampling

Undergraduate. $F^{-1}(U)$ has the distribution function F.

"Rejection sampling".

Want: To simulate from a given density $g(x)$.

Know: How to simulate from some density $f(x)$.

Know:

$$\sup_x \frac{g(x)}{f(x)} \le C \text{ is known.}$$

- x is a sample from f.
- With probability $g(x)/(Cf(x))$, output x.
- Else, repeat.

On each step,

$$\mathbb{P}(\text{output} \in [x, x + \mathrm{d}x]) = f(x)\,\mathrm{d}x \cdot \frac{g(x)}{Cf(x)} = \frac{1}{c} g(x)\,\mathrm{d}x.$$

$\mathbb{P}(\text{some output}) = 1/C$, so the density given that we have an output is $g(x)$.

18.2 Markov Chains on Measurable State Spaces

Consider a MC $(X_n, n \ge 0)$ on measurable S, specified by the kernel $Q(s, A) = \mathbb{P}(X_1 \in A \mid X_0 = s)$.

$$\mu_n(\cdot) = \text{dist}(X_n) = \int Q(s, \cdot)\mu_{n-1}(\mathrm{d}s).$$

Lemma 18.1. *Let β be a PM on S with the following assumption:*

(H1) Suppose that $\forall x \in S$, there exists a stopping time $T_x < \infty$ a.s. for the (δ_x, Q)-chain such that $\mathbb{P}_x(X_{T_x} \in \cdot) = \beta(\cdot)$.

Then, for the (β, Q)-chain, $\exists T < \infty$ such that $\mathbb{P}_\beta(X_T \in \cdot) = \beta(\cdot)$ and define

$$\mu(A) \stackrel{\text{def}}{=} \mathbb{E}_\beta[\textit{number of visits to } A \textit{ before } T].$$

Suppose $\exists A_n \uparrow S$ such that $\mu(A_n) < \infty$. This defines a (maybe σ-finite) invariant measure μ.

Proof. Condition on the first step. □

Consider the following assumption:

(H2) There exists a PM β and $\exists \delta > 0$ such that $Q(x, \cdot) \geq \delta\beta(\cdot)\ \forall x \in S$.

Lemma 18.2. *(H2) $\implies$ (H1).*

Proof. This is rejection sampling.

Write $Q(x, \cdot) = \delta\beta(\cdot) + (1-\delta)R(x, \cdot)$, which is the definition of the kernel $R(x, \cdot)$. Let $(\xi_i, i \geq 1)$ be independent, $\mathbb{P}(\xi_i = 1) = \delta$, $\mathbb{P}(\xi_i = 0) = 1 - \delta$. Construct a Q-chain: given $X_{n-1} = x$, if $\xi_n = 1$, then X_n has distribution β; if $\xi_n = 0$, then X_n has distribution $R(x, \cdot)$. Define $T = \min\{n : \xi_n = 1\}$. T has the Geometric(δ) distribution, and X_T has the distribution β. □

Useful Version. Consider the assumptions:

(H3) There exists a subset $A \subseteq S$ and a PM β and $\delta > 0$ such that

- (i) $\mathbb{P}_x(T_A < \infty) = 1\ \forall x \in S$,
- (ii) $Q(x, \cdot) \geq \delta\beta(\cdot)\ \forall x \in A$.

This is a **Harris chain**.

Lemma 18.3. *(H3) $\implies$ (H1).*

Proof. Define V_j to be the time of the jth visit to A, $V_{j+1} = \min\{n > V_j : X_n \in A\}$. Define

$$Y_j = X_{(1+V_j)}.$$

Then, (Y_j) is a MC with some kernel $\hat{Q}$, and by (ii), $\hat{Q}$ satisfies (H2). Therefore, (Y_j) satisfies (H1), so (X_n) satisfies (H1). □

We can derive limit theorems from (H1) analogously to the countable state case. In particular, if we have $\mu(S) < \infty \iff$ positive-recurrent, then

$$\pi(\cdot) = \frac{\mu(\cdot)}{\mu(S)}$$

is a stationary distribution and

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_{(X_i \in A)} \xrightarrow{\text{a.s.}} \pi(A) \qquad \text{as } n \to \infty$$

(for any initial distribution) and

$$\|\text{dist}(X_n) - \pi\|_{\text{VD}} \to 0 \qquad \text{as } n \to \infty \text{ if aperiodic.}$$

See Durrett, section 6.8.

Example 18.4. $S = \mathbb{R}^d$. $Q(x, \cdot)$ has the density $q(x, y) > 0$ everywhere which is a continuous function of (x, y).

Take $A = \text{ball}(\mathbf{0}, B)$. Then, $\inf_{x,y \in A} q(x, y) \equiv \varepsilon > 0$ by uniform continuity, so (ii) holds for the choice $\beta = \text{Uniform}(A)$ and

$$\delta = \frac{\varepsilon}{\text{Leb}(A)}.$$

We need to show $T_A < \infty$ a.s. It is enough to show $\exists B$ $\mathbb{E}_x|X_1| \le |x|$ for all x with $|x| > B$. By super-MG convergence, $T_A < \infty$.

This method cannot work if there are only a countable number of possible transitions from a state.

18.3 Markov Chains as Iterated Random Functions

This follows the posted Diaconis-Freedman paper. It is also known as **coupling from the past**.

Background. Given $f : S \to S$, we can iterate: if we have $f(s)$, $f^{(2)}(s) = f(f(s))$, and

$$f^{(n+1)}(s) = f(f^{(n)}(s)) = f^{(n)}(f(s)).$$

Let S be measurable and μ be a PM invariant under f. This is the structure of **ergodic theory**.

If S is a topological space, and f is continuous, consider s_0, $s_1 = f(s_0)$, $s_{n+1} = f(s_n) = f^{(n)}(s_0)$. Consider μ_n, the empirical distribution on $(S_0, S_1, \ldots, S_n)$:

$$\frac{1}{n}\sum_{i=0}^{n-1} \delta_{s_i}.$$

Suppose $\mu_n \to$ some μ weakly. Then, μ is invariant. This is the study of dynamical systems or "chaos".

Lemma 18.5 (Old Lemma)**.** *Given a PM μ on $S \times S$, the first marginal μ_1, given independent X and U such that* $\text{dist}(X) = \mu_1$ *and* $U = \text{Uniform}(0, 1)$*, then* $\exists f : S \times [0, 1] \to S$ *such that* $\text{dist}(X, f(X, U)) = \mu$.

Given a MC, take some explicit representation as $X_{n+1} = f(X_n, \xi_{n+1}) = f_{\xi_{n+1}}(X_n)$ for IID $(\xi_i, i \ge 1)$, $\hat{S}$-valued, where f is continuous $S \times \hat{S} \to S$. We want to show $\text{dist}(X_n) \to$ some π weakly.

$$X_0 = x_0, \qquad X_n(x_0) = f_{\xi_n}(f_{\xi_{n-1}}(\cdots f_{\xi_2}(f_{\xi_1}(x_0))\cdots)).$$

Instead, consider

$$Y_n(x_0) = f_{\xi_1}(f_{\xi_2}(\cdots f_{\xi_{n-1}}(f_{\xi_n}(x_0))\cdots)).$$

Here, $Y_n(x_0) \overset{\text{d}}{=} X_n(x_0)$.

If we can prove $Y_n(x_0) \xrightarrow[\text{a.s.}]{} $ some $Y_\infty(x_0)$ as $n \to \infty$, then $\text{dist}(X_n(x_0)) \to \pi$ weakly.

Example 18.6. Let (A_i, B_i) be IID $\mathbb{R}^2$-valued. Define a $\mathbb{R}^1$-valued MC X_n by

$$X_{n+1} = A_{n+1}X_n + B_{n+1}.$$

For $X_0 = x_0$,

$$X_n = \sum_{j=0}^{n} B_j \prod_{k=j+1}^{n} A_k, \qquad B_0 = x_0,$$

$$Y_n = \sum_{j=0}^{n} B_j \prod_{k=1}^{j-1} A_k, \qquad B_0 = x_n.$$

By the IID SLLN,

$$\frac{1}{j} \log \left| \prod_{i=1}^{j-1} A_i \right| \to \mathbb{E} \log |A_1| \qquad \text{a.s.}$$

Easy. If $\mathbb{E} \log |A_1| < 0$, then $\prod_{i=1}^{j} A_i \to 0$ geometrically fast. If also $\mathbb{E} \log |B_1| < \infty$, then

$$Y_n \xrightarrow{\text{a.s.}} Y_\infty = \sum_{j=0}^{\infty} B_j \prod_{k=1}^{j-1} A_k$$

so $\operatorname{dist}(X_n) \to \operatorname{dist}(Y_\infty)$ weakly.

The analog for $\mathbb{R}^d$-valued

$$X_{n+1} = \underbrace{A_{n+1}}_{d \times d \text{ matrix}} X_n + \underbrace{B_n}_{d\text{-vector}}$$

works. We get a stationary distribution π on $\mathbb{R}^d$.

Lecture 19

March 21

19.1 Another MC Example

Setting. $X_n = f(X_{n-1}, \xi_n)$ for prescribed f and IID (ξ_i).

Suppose we have a metric space (S, d). For $f : S \to S$,

$$\|f\|_{\text{Lip}} \stackrel{\text{def}}{=} \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x, y)}.$$

For a random function $f(x, \xi)$, consider $\mathbb{E} \log \|f(\cdot, \xi)\|_{\text{Lip}} \equiv \kappa$, say.

Theorem 19.1 (Diaconis-Freedman Paper)**.** *For a MC of form $X_n = f(X_{n-1}, \xi_n)$, if $\kappa < 0$ (and side conditions), then the "coupling from the past" method shows there exists a unique stationary distribution π and $\text{dist}(X_n) \to \pi$ weakly.*

Example 19.2. $S = (0, 1)$. Given $X_0 = x$, flip a fair coin $\{L, R\}$. If L, take X_1 to be Uniform$[0, x]$, and if R, take X_1 to be Uniform$[x, 1]$.

Define

$$f(x, u, L) = ux,$$
$$f(x, u, R) = x + u(1 - x).$$

Take $\xi = (U, I)$, U is Uniform$[0, 1]$, I is Uniform$\{L, R\}$, independent. This represents the chain as $X_n = f(X_{n-1}, \xi_n)$.

$$\|f(\cdot, u, L)\|_{\text{Lip}} = u = \|f(\cdot, u, R)\|_{\text{Lip}} \implies \kappa = \mathbb{E} \log U < 0.$$

19.1 implies that a stationary π exists.

(*Exercise*). Find π explicitly.

19.2 Ergodic Theory

19.2.1 "Probability" Set-Up

$(X_0, X_1, X_2, \dots)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{R}$-valued, are **stationary** if

$$(X_0, X_1, \dots, X_{n-1}) \stackrel{\text{d}}{=} (X_1, X_2, \dots, X_n) \quad \forall n. \tag{19.1}$$

This is equivalent to $(X_0, X_1, X_2, \dots) \overset{d}{=} (X_1, X_2, X_3, \dots)$ and equivalent to

$$(X_0, X_1, X_2, \dots) \overset{d}{=} (X_n, X_{n+1}, X_{n+2}, \dots) \quad \forall n.$$

Given stationary $(X_n, 0 \le n < \infty)$, there exists (Kolmogorov Extension Theorem) a two-sided stationary sequence $(\hat{X}_n, -\infty < n < \infty)$, such that $(\hat{X}_n, n \ge 0) \overset{d}{=} (X_n, n \ge 0)$.

Example 19.3. IID random variables are stationary.

Example 19.4. Exchangeable random variables are stationary.

Example 19.5. A stationary Markov chain is stationary.

Example 19.6 ("Moving Average")**.** Let (ξ_i) be IID. Fix $L \ge 2$. Let

$$A_i = \frac{\xi_i + \xi_{i+1} + \cdots + \xi_{i+L-1}}{L}.$$

Then, $(A_i, i \ge 0)$ is stationary.

Theorem 19.7 (Easy)**.** *If $(X_n, 0 \le n < \infty)$ is stationary, if $g : \mathbb{R}^\infty \to \mathbb{R}$ is measurable, then for $Y_n = g(X_n, X_{n+1}, X_{n+2}, \dots)$, $(Y_n, 0 \le n < \infty)$ is stationary.*

This starts with very random ingredients.

19.2.2 Ergodic Theory Set-Up

A probability space $(S, \mathcal{S}, \mu)$ is "concrete". For a measurable $\phi : S \to S$, the push-forward measure is $\hat{\mu}(A) = \mu(\phi^{-1}(A))$. Suppose μ is invariant under ϕ: $\mu(A) = \mu(\phi^{-1}(A))\ \forall A$. [Given μ, say ϕ is a **measure-preserving transformation**.]

Now, for any measurable $f : S \to \mathbb{R}$, we can define

$$X_0(s) = f(s), \tag{19.2}$$
$$X_1(s) = f(\phi(s)), \tag{19.3}$$
$$X_2(s) = f(\phi^{(2)}(s)), \tag{19.4}$$
$$X_n(s) = f(\phi^{(n)}(s)), \tag{19.5}$$
$$\phi^{(m)}(s) = \phi(\phi^{(m-1)}(s)). \tag{19.6}$$

We can define RVs $(X_n, 0 \le n < \infty)$ on a probability space $(S, \mathcal{S}, \mu)$.

Lemma 19.8. *Given μ, ϕ as above, for any f, the sequence $(X_n, n \ge 0)$ is stationary.*

Proof. To check (19.1), we need to check

$$\mu\{s : X_0(s) \in A_0, X_1(s) \in A_1, \dots, X_{n-1}(s) \in A_{n-1}\} = \mu\{s : X_1(s) \in A_0, \dots, X_n(s) \in A_{n-1}\}.$$

Let $B = \{s : X_0(s) \in A_0, X_1(s) \in A_1, \dots, X_{n-1}(s) \in A_{n-1}\}$.

$$\text{Left} = \mu\{s : s \in B\}, \tag{19.7}$$

$$\begin{aligned}\text{Right} &= \mu\{s : X_0(\phi(s)) \in A_0, X_1(\phi(s)) \in A_1, \dots, X_{n-1}(\phi(s)) \in A_{n-1}\} \\ &= \mu\{s : \phi(s) \in B\} = (19.7) \qquad \text{by measure-preserving.}\end{aligned} \tag{19.8}$$

□

Here, we start with deterministic objects.

Example 19.9 ("Rotation on a Circle")**.** $S = [0,1]$. Fix $\theta \in (0,1)$. Take

$$\phi(s) = s + \theta \pmod 1, \qquad \mu = \text{Lebesgue measure on } S.$$

Example 19.10 (Baker's Transformation)**.** $S = [0,1]^2$ and $\mu = \text{Leb}^2$.

$$\phi(x,y) = \begin{cases} \left(2x, \dfrac{y}{2}\right), & \text{if } x < \dfrac{1}{2}, \\ \left(2x-1, \dfrac{1}{2} + \dfrac{y}{2}\right), & \text{if } x \geq \dfrac{1}{2}. \end{cases}$$

Given stationary $(\hat{X}_n, n \geq 0)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, there is a "canonical" way to set it up in the ergodic theory set-up.

Define $S = \mathbb{R}^\infty$, $\mu = \text{dist}(\hat{X}_n, n \geq 0)$ on S. Define $\phi : S \to S$ by $\phi(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$. The function $f : S \to \mathbb{R}$ is $f(x_0, x_1, \dots) = x_0$. Then, define X_n as in (19.5) gives $X_n(x_0, x_1, x_2, \dots) = x_n$ and $(X_n, n \geq 0) \stackrel{\text{d}}{=} (\hat{X}_n, n \geq 0)$. The former are RVs on $(\mathbb{R}^\infty, \mu)$ and the latter are RVs on $(\Omega, \mathcal{F}, \mathbb{P})$.

19.2.3 Invariant Events

Definition 19.11. In the ergodic theory set-up, an event A is **invariant** if $\phi^{-1}(A) = A$ a.s.

Easy Fact: If $A = \phi^{-1}(A)$ a.s., then $A^* = \bigcup_{n=1}^\infty \bigcap_{i>n} \phi^{-i}(A)$ satisfies $A^* = A$ a.s. and $\phi^{-1}(A^*) = A^*$ always.

The collection of all invariant events forms the **invariant σ-field** $\mathcal{I}$.

Definition 19.12. A measure-preserving transformation ϕ on $(S, \mathcal{S}, \mu)$ is **ergodic** if $\mathcal{I}$ is trivial. That is, $\mu(A) = 0$ or 1 for each invariant A.

Given a stationary $(\hat{X}_n, n \geq 0)$, go to the canonical set-up to use these definitions. The notion of invariant $A \subseteq \mathbb{R}^\infty$ says that

$$\{\omega : (X_0(\omega), X_1(\omega), \dots) \in A\} \stackrel{\text{a.s.}}{=} \{\omega : (X_1(\omega), X_2(\omega), \dots) \in A\}. \tag{19.9}$$

The process $(\hat{X}_n)$ is ergodic $\iff$ $\mathbb{P}((X_0, X_1, \dots) \in A) = 0$ or 1 for each invariant A.

Lemma 19.13. *For stationary $(X_n, n \geq 0)$ in the canonical set-up, $\mathcal{I} \stackrel{a.s.}{\subseteq} \tau =$ tail σ-field of (X_n).*

Proof.

$$A \subseteq \mathbb{R}^\infty \text{ invariant} \implies A = \phi^{-1}(A) \text{ a.s.}$$

(where ϕ is the shift map $(x_0, x_1, \dots) \mapsto (x_1, x_2, \dots)$)

$$\implies A = \phi^{-n}(A) \text{ a.s.}$$

Therefore,

$$\phi^{-n}(A) = \{\omega : (X_n(\omega), X_{n+1}(\omega), \dots) \in A\} \in \sigma(X_n, X_{n+1}, \dots) \equiv \tau_n,$$

so $A \in \bigcap_n \tau_n = \tau$ a.s. $\square$

For example, consider alternating coin flips, $HTHTHTH\dots$ or $THTHTHT\dots$. Then, $X_0 \in \tau$, but we have $X_0 \notin \mathcal{I}$.

Recall: **Theorem**. If $(X_n, n \geq 0)$ is stationary, if $g : \mathbb{R}^\infty \to \mathbb{R}$ is measurable, then $(Y_n, n \geq 0)$ is stationary for $Y_n = g(X_n, X_{n+1}, \dots)$ *and* if (X_n) is ergodic, then (Y_n) is ergodic.

If B is invariant for (Y_n),

$$\{\omega : (Y_0(\omega), Y_1(\omega), \dots) \in B\} \overset{\text{a.s.}}{=} \{\omega : (Y_1(\omega), Y_2(\omega), \dots) \in B\},$$

and this reduces to (19.9) for a certain A depending on B.

Lecture 20

March 23

20.1 Ergodic Theory & Markov Chains

Proposition 20.1. *Any stationary irreducible Markov countable state (S) Markov chain is ergodic.*

Proof. Let π be the stationary distribution. We know that the chain is positive-recurrent, which implies that it visits every state infinitely often. Consider an invariant set $A \subseteq S^\infty$. Define the function $h(x) = \mathbb{E}_x 1_{((X_0,X_1,\dots)\in A)}$.

$$\begin{aligned} \mathbb{E}_\pi[1_{((X_0,X_1,\dots)\in A)} \mid \mathcal{F}_n] &\overset{\text{a.s}}{=} \mathbb{E}_\pi[1_{((X_n,X_{n+1},\dots)\in A)} \mid \mathcal{F}_n] \qquad \text{(definition of "invariant")} \\ &\underbrace{=}_{\text{Markov}} h(X_n). \end{aligned}$$

The LHS is a MG, so it converges a.s. to $1_{((X_0,X_1,\dots)\in A)}$. Since $h(X_n)$ is also converging a.s., $h(x)$ is constant for all x, so $h(x) = 0$ for all x or $h(x) = 1$ for all x. Therefore, $\mathbb{E}_\pi 1_{((X_0,X_1,\dots)\in A)}$ is 0 or 1, so the chain is ergodic. □

Fact. Here, the *tail* σ-field is trivial $\iff$ the chain is aperiodic.

20.2 Ergodic Theorem

Theorem 20.2 (The Ergodic Theorem)**.** *For stationary $(X_i, 0 \le i < \infty)$ with $\mathbb{E}|X_0| < \infty$ and*

$$S_n = \sum_{i=0}^{n-1} X_i,$$

we have $n^{-1}S_n \to \mathbb{E}[X_0 \mid \mathcal{I}]$ a.s. and in L^1 as $n \to \infty$.

Ergodic implies that the limit is $\mathbb{E}X_0$.

Lemma 20.3 (Maximal Lemma)**.** *Write $M_k = \max(0, S_1, S_2, \dots, S_k)$. Then, $\mathbb{E}[X_0 1_{(M_k>0)}] \ge 0$.*

Proof. See the text. □

Easy. $\mathbb{E}[X_k \mid \mathcal{I}] \overset{\text{a.s}}{=} \mathbb{E}[X_0 \mid \mathcal{I}]$ and $(X_k - \mathbb{E}[X_k \mid \mathcal{I}], k \ge 0)$ is stationary.

Classic Proof of 20.2. Reduce to the case $\mathbb{E}[X_0 \mid \mathcal{I}] = 0$. Write

$$\bar{X} = \limsup \frac{S_n}{n} \in \mathcal{I}.$$

It is enough to prove $\bar{X} \le 0$ a.s. (then apply this to $-\bar{X}$).

Fix $\varepsilon > 0$. Consider $X_i^* = (X_i - \varepsilon)1_{(\bar{X} > \varepsilon)}$. Check that $(X_i^*, i \ge 0)$ is stationary, and define S_k^*, M_k^* as in 20.3. Define $F_n = \{M_n^* > 0\}$. Let

$$F = \bigcup_n F_n = \left\{\sup_{n\ge1} \frac{S_n^*}{n} > 0\right\} = \left\{\sup_{n\ge1} \frac{S_n^*}{n} > \varepsilon, \bar{X} > \varepsilon\right\} = \{\bar{X} > \varepsilon\}.$$

Apply the Maximal Lemma 20.3 to (X_i^*).

$$\mathbb{E}[X_0^* 1_{F_n}] \ge 0.$$

Note that $F_n \uparrow F$ and $\mathbb{E}|X_0^*| \le \mathbb{E}|X_0| + \varepsilon < \infty$. Therefore,

$$\mathbb{E}[X_0^*] = \mathbb{E}[X_0^* 1_F] = \lim_n \mathbb{E}[X_0^* 1_{F_n}] \ge 0. \tag{20.1}$$

However, $F \in \mathcal{I}$ and $\mathbb{E}[X_0 \mid \mathcal{I}] = 0$, which implies that $\mathbb{E}[X_0 1_F] = 0$. $X_0^* = (X_0 - \varepsilon)1_F$ implies that $\mathbb{E}X_0^* = \mathbb{E}[X_0 1_F] - \varepsilon\mathbb{P}(F)$, so $\mathbb{P}(F) = 0$. Hence, $\mathbb{P}(\bar{X} > \varepsilon) = 0$, so $\bar{X} \le 0$ a.s. □

20.3 Applications to Range/Recurrence of "Stationary Increment" Random Walks

Setting. Let $(X_1, X_2, X_3, \dots)$ be stationary, $\mathbb{Z}^d$-valued. $S_0 = 0$ and $S_k = \sum_{i=1}^k X_i$. The event A is the event that we "never return to 0": $\{S_k \ne 0, \forall k \ge 1\} = \{(X_1, X_2, \dots) \in \hat{A}\}$, where

$$\hat{A} = \left\{(x_1, x_2, \dots) : \sum_{i=1}^j x_i \ne 0 \ \forall j \ge 1\right\},$$

$$\hat{A}_k = \left\{(x_1, \dots, x_k) : \sum_{i=1}^j x_i \ne 0 \ \forall 1 \le j \le k\right\}.$$

So, $\hat{A} \subseteq (\mathbb{Z}^d)^\infty$.

Theorem 20.4. *In the setting above, R_n is the number of distinct sites in $\mathbb{Z}^d$ that $(S_1, \dots, S_n)$ visits. Then, $n^{-1}R_n \xrightarrow[L^1]{a.s.} \mathbb{E}[1_A \mid \mathcal{I}]$.*

Idea. R_n counts the number of events. We will sandwich R_n between two stationary processes of events.

Proof. R_n is at least the number of m's ($1 \le m \le n$) such that $(S_{m+1}, S_{m+2}, \dots)$ are all different from S_m. The latter is $\sum_{m=1}^n 1_{((X_{m+1}, X_{m+2}, \dots) \in \hat{A})}$ and the $m = 0$ case is 1_A. The Ergodic Theorem 20.2 implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n 1_{((X_{m+1}, X_{m+2}, \dots) \in \hat{A})} = \mathbb{E}[1_A \mid \mathcal{I}] \le \liminf_n n^{-1} R_n.$$

(This is one side of the theorem.)

Fix k. Observe

$$\begin{aligned} R_n &\le k + \text{number of } m\text{'s } (1 \le m \le n-k) \text{ such that } S_{m+1}, S_{m+2}, \dots, S_{m+k} \text{ are all different from } S_m \\ &= k + \sum_{m=1}^{n-k} 1_{((X_{m+1},\dots,X_{m+k})\in \hat{A}_k)}, \qquad \text{where } \hat{A}_k \text{ is the analog of } \hat{A}. \end{aligned}$$

Apply the Ergodic Theorem 20.2 to the stationary process of indicators.

$$\limsup_n n^{-1} R_n \le \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n-k} 1_{((X_{m+1},\dots,X_{m+k})\in \hat{A}_k)} = \mathbb{E}[1_{A_k} \mid \mathcal{I}] \quad \text{a.s. and in } L^1$$

Let $k \uparrow \infty$, $A_k \downarrow A$.

$$\le \mathbb{E}[1_A \mid \mathcal{I}] \quad \text{a.s. and in } L^1$$

□

Theorem 20.5. *In the setting above, assume the random variables are $\mathbb{Z}^1$-valued and $\mathbb{E}|X_1| < \infty$.*

(i) If $\mathbb{E}[X_1 \mid \mathcal{I}] = 0$, then $\mathbb{P}(A) = 0$ ("recurrence").

(ii) If $\mathbb{P}(A) = 0$, then $\mathbb{P}(S_n = 0$ infinitely often$) = 1$.

Proof. (i) By 20.4, it is enough to prove $R_n/n \to 0$ a.s. (then $\mathbb{E}[1_A \mid \mathcal{I}] = 0 \implies \mathbb{P}(A) = 0$). However, $R_n \le 1 + \max_{m\le n} S_m - \min_{m \le n} S_m$. So, it is enough to show

$$\frac{1}{n} \max_{m\le n} S_m \to 0 \qquad \text{a.s.} \tag{20.2}$$

The Ergodic Theorem 20.2 says

$$\frac{S_n}{n} \to 0 \qquad \text{a.s.} \tag{20.3}$$

and it is a deterministic fact that (20.3) $\implies$ (20.2).

(ii) We will show $\mathbb{P}(X_n = 0$ for at least 2 values of $n) = 1$. A similar argument will work for any B. Write T_n for the time of the nth return to 0. $\{T_1 = j, T_2 = j+k\} = \{T_1 = j\} \cap G_{j,k}$, where

$$G_{j,k} = \{S_{j+i} - S_j \ne 0, 1 \le i \le k-1, S_{j+k} = S_j\}.$$

Stationarity implies $\mathbb{P}(G_{j,k}) = \mathbb{P}(G_{0,k}) = \mathbb{P}(T_1 = k)$. The hypothesis implies that

$$\sum_{k=1}^{\infty} \mathbb{P}(T_i = k) = 1 \implies \sum_{k=1}^{\infty} \mathbb{P}(G_{j,k}) = 1 \implies \bigcup_{k\ge 1} G_{j,k} = \Omega \qquad \text{a.s.},$$

so $\bigcup_{k=1}^{\infty} (G_{j,k} \cap \{T_1 = j\}) = \{T_1 = j\}$ a.s. So, $\{T_1 = j, T_2 < \infty\} = \{T_1 = j\}$ a.s. Take the union over j, and we have $\{T_1 < \infty, T_2 < \infty\} = \{T_1 < \infty\}$ a.s. The latter has probability 1, so the former has probability 1.

□

Lecture 21

April 4

21.1 Entropy

Definition 21.1. If π is a PM on finite S,

$$H(\pi) = -\sum_{s \in S} \pi(s) \log \pi(s)$$

is the **entropy** of π.

Easy: $0 \le H(\pi) \le \log|S|$. $H(\text{uniform distribution on } S) = \log|S|$.

$(X_0, X_1, X_2, \dots)$ is a S-valued process. $p(x_0, x_1, \dots, x_{n-1}) = \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1})$. Then, $L_n \stackrel{\text{def}}{=} p(X_0, X_1, \dots, X_{n-1})$ is the empirical likelihood.

For IID (X_i), $\text{dist}(X_i) = \pi$, then

$$p(x_0, x_1, \dots, x_{n-1}) = \prod_{s \in S} (\pi(s))^{m(n,s)}$$

where $m(n,s) = \sum_{i=0}^{n-1} 1_{(x_i = s)}$,

$$\log p(x_0, x_1, \dots, x_{n-1}) = \sum_s m(n,s) \log \pi(s)$$

$$\frac{1}{n} \log p(X_0, \dots, X_{n-1}) = \sum_s F(n,s) \log \pi(s), \qquad F(n,s) = \frac{1}{n} \sum_{i=0}^{n-1} 1_{(X_i = s)} \xrightarrow{\text{a.s.}} \pi(s) \quad \text{as} \quad n \to \infty$$

$$\xrightarrow[n \to \infty]{\text{a.s.}} \sum_s \pi(s) \log \pi(s) = -H(\pi).$$

Informally, for a typical realization $x_0, x_1, \dots, x_{n-1}$, $p(x_0, x_1, \dots, x_{n-1}) \approx \exp(-nH(\pi))$.

Theorem 21.2 (Shannon-McMillan-Breiman Theorem)**.** *If $(X_i, i \ge 0)$ is stationary and ergodic, then*

$$-\frac{1}{n} \log L_n \xrightarrow{a.s.} H$$

for a constant $0 \le H < \infty$.

The proof uses:

- MG convergence

- Ergodic Theorem
- K-step Markov process

Proof. Embed $(X_i, i \geq 0)$ into a doubly-infinite process $(X_i, -\infty < i < +\infty)$, which is stationary and ergodic. Write $p(x_n \mid x_{n-1}, \dots, x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0)$. Consider $p(x_0 \mid X_{-1}, X_{-2}, \dots, X_{-n}) \xrightarrow[\text{a.s.}]{n\to\infty} p(x_0 \mid \mathcal{F}_\infty)$ since $p(x_0 \mid X_{-1}, X_{-2}, \dots, X_{-n})$ is a MG. Here, $\mathcal{F}_\infty = \sigma(X_{-1}, X_{-2}, \dots)$. Define $H_k = \mathbb{E}[-\log p(X_0 \mid X_{-1}, \dots, X_{-k})]$. Then,

$$\begin{aligned} H_k = \mathbb{E}[-\log p(X_0 \mid X_{-1}, \dots, X_{-k})] &= \mathbb{E}[\mathbb{E}[-\log p(X_0 \mid X_{-1}, \dots, X_{-k}) \mid X_{-1}, \dots, X_{-k}]] \\ &= \mathbb{E}\left[\sum_x -\log p(x \mid X_{-1}, \dots, X_{-k}) \cdot p(x \mid X_{-1}, \dots, X_{-k})\right] \\ &\to \mathbb{E}\sum_x (-\log p(x \mid \mathcal{F}_\infty) \cdot p(x \mid \mathcal{F}_\infty)) \end{aligned}$$

(as $k \to \infty$, by MG convergence)

$$\begin{aligned} &= \mathbb{E}[\mathbb{E}[-\log p(X_0 \mid \mathcal{F}_\infty) \mid \mathcal{F}_\infty]] \\ &= \mathbb{E}[-\log p(X_0 \mid \mathcal{F}_\infty)] \\ &\overset{\text{define}}{=} H. \end{aligned}$$

The Ergodic Theorem says

$$\frac{1}{n}\sum_{m=0}^{n-1} F(\underbrace{X_m, X_{m-1}, X_{m-2}, \dots}_{Y_m}) \to \mathbb{E}F(X_0, X_{-1}, X_{-2}, \dots)$$

for bounded measurable F. Apply the Ergodic Theorem to $F(X_0, X_1, \dots) = -\log p(X_0 | X_{-1}, X_{-2}, \dots)$.

$$\frac{1}{n}\sum_{m=0}^{n-1} -\log p(X_m \mid X_{m-1}, X_{m-2}, \dots) \xrightarrow{\text{a.s.}} H \tag{21.1}$$

Elementary: $p(x_0, x_1, \dots, x_{n-1} \mid x_{-1}, \dots, x_{-k}) = \prod_{m=0}^{n-1} p(x_m \mid x_{m-1}, x_{m-2}, \dots, x_0, \dots, x_{-k})$. Substitute in $(X_{-1}, \dots, X_{-k})$, let $k \to \infty$, and use MG convergence.

$$p(x_0, \dots, x_{n-1} \mid \mathcal{F}_\infty) = \prod_{m=0}^{n-1} p(x_m \mid x_{m-1}, \dots, x_0, \mathcal{F}_\infty)$$

Substitute $X_1, \dots, X_{n-1}$, take $(1/n)\log(\cdot)$, and apply (21.1).

$$-\frac{1}{n}\log p(X_0, \dots, X_{n-1} \mid \mathcal{F}_\infty) \xrightarrow{\text{a.s.}} H \quad \text{by} \quad (21.1). \tag{21.2}$$

Given a distribution (Y_0, Y_1) with $\text{dist}(Y_0) = \text{dist}(Y_1)$ on S^*, we can construct a stationary Markov $(\hat{X}_0, \hat{X}_1, \hat{X}_2, \dots)$ with $(\hat{X}_n, \hat{X}_{n+1}) \overset{\text{d}}{=} (Y_0, Y_1)$. Take $\hat{X}_0 = Y_0$ and for the transitions, use the kernel $Q(x_0, x_1) = \mathbb{P}(Y_1 = x_1 \mid Y_0 = x_0)$.

Given stationary S-valued $(X_0, X_1, X_2, \dots)$, set $S^* = S^k$. Set $Y_0 = (X_0, \dots, X_{k-1})$, $Y_1 = (X_1, \dots, X_k)$. We can construct $(\hat{Y}_i, i \geq 0)$ as above which is stationary. The process $(\hat{Y}_0, \hat{Y}_1, \hat{Y}_2, \dots)$ has the Markov property $\mathbb{P}(\hat{Y}_m = \cdot \mid Y_{m-1}, Y_{m-2}, \dots)$ depends only on Y_{m-1}. Extract the coordinates: $(\hat{X}_0, \hat{X}_1, \dots)$ is a stationary sequence with the "k-step Markov" property. $\mathbb{P}(\hat{X}_m = x_m \mid \hat{X}_{m-1} = x_{m-1}, \dots)$ depends only on $x_{m-1},\dots,x_{m-k}$ and $(\hat{X}_m,\dots,\hat{X}_{m+k-1}) \overset{\mathrm{d}}{=} (X_0,X_1,\dots,X_{k-1})$.

Fix k. Apply the Ergodic Theorem to $F(x_0,x_{-1},\dots,x_{-k}) = -\log p(x_0 \mid x_{-1},\dots,x_{-k})$.

$$-\frac{1}{n}\sum_{m=0}^{n-1} \log p(X_m \mid X_{m-1},\dots,X_{m-k}) \xrightarrow{\text{a.s.}} H_k$$
$$= -\frac{1}{n}\log \prod_{m=0}^{n-1} p(X_m \mid X_{m-1},\dots,X_{m-k})$$

Write $p^{(k)}(x_0,x_1,\dots,x_{n-1}) = p(x_0,\dots,x_{k-1})\prod_{m=k}^{n-1} p(x_m \mid x_{m-1},\dots,x_{m-k})$. This is the distribution of the k-step Markov process.

$$-\frac{1}{n}\log p^{(k)}(X_0,\dots,X_{n-1}) \xrightarrow[n\to\infty]{\text{a.s.}} H_k \qquad \text{by the above argument.} \tag{21.3}$$

Because the $H_k \to H$, to prove the theorem, it is enough to prove

$$H \le \liminf_n -\frac{1}{n}\log p(X_0,\dots,X_{n-1}) \tag{21.4}$$

and

$$\limsup_n -\frac{1}{n}\log p(X_0,\dots,X_{n-1}) \le H_k. \tag{21.5}$$

Check: Given 21.3,

$$(21.3) + (21.6) \implies (21.5),$$
$$(21.2) + (21.7) \implies (21.4).$$

□

Lemma 21.3. *(a) If $W_n \ge 0$, $\mathbb{E}W_n \le 1$, then $\limsup_n n^{-1}\log W_n \le 0$ a.s.*

(b)

$$\mathbb{E}\frac{p^{(k)}(X_0,\dots,X_{n-1})}{p(X_0,\dots,X_{n-1})} = 1. \tag{21.6}$$

(c)

$$\mathbb{E}\frac{p(X_0,\dots,X_{n-1})}{p(X_0,\dots,X_{n-1} \mid \mathcal{F}_\infty)} \le 1. \tag{21.7}$$

Proof. (a)

$$\mathbb{P}(n^{-1}\log W_n \ge \varepsilon) = \mathbb{P}(W_n \ge e^{\varepsilon n})$$
$$\le e^{-\varepsilon n}\mathbb{E}W_n \le e^{-\varepsilon n}.$$

Use Borel-Cantelli.

(c) To prove (21.7), it is enough to prove

$$\mathbb{E}\frac{p(X_0,\dots,X_{n-1})}{p(X_0,\dots,X_{n-1} \mid X_{-1},\dots,X_{-k})} = 1,$$

and then let $k \to \infty$ and use Fatou's Lemma.

$$\begin{aligned}\mathbb{E}\frac{p(X_0,\ldots,X_{n-1})}{p(X_0,\ldots,X_{n-1}\mid X_{-1},\ldots,X_{-k})} &= \mathbb{E}\frac{p(X_0,\ldots,X_{n-1})p(X_{-1},\ldots,X_{-k})}{p(X_{-k},\ldots,X_0,\ldots,X_{n-1})}\\ &= 1 \qquad \text{by (21.8) because the numerator } \textit{is} \text{ a PM.}\end{aligned}$$

Recall: If Y has distribution π, then

$$\mathbb{E}\frac{\hat{\pi}(Y)}{\pi(Y)} = 1 \tag{21.8}$$

for any distribution $\hat{\pi}$.

(b) Again, by (21.8) because $p^{(k)}$ is a PM.

□

Lecture 22

April 6

22.1 Entropy Rate

Setting: $(X_i, i \geq 0)$ is stationary, ergodic, S-valued.

Theorem: $L_n = p(X_0, X_1, \ldots, X_{n-1})$, where $p(x_0, x_1, \ldots, x_{n-1}) = \mathbb{P}(X_i = x_i, 0 \leq i \leq n-1)$. Then,

$$-\frac{1}{n} \log L_n \xrightarrow{\text{a.s.}} H \text{ (constant)} \qquad \text{as} \qquad n \to \infty.$$

Call H the **entropy rate** of the process (X_i).

Recall that for a PM π on S, $H(\pi) \stackrel{\text{def}}{=} \sum_s \pi(s) \log \pi(s) = -\mathbb{E}[\log \pi(X)]$ if $X \stackrel{d}{\sim} \pi$ is the **entropy** of π.

The proof of the Shannon-McMillan-Breiman Theorem 21.2 gave a formula for the entropy rate

$$H = -\mathbb{E}[\log p(X_0 \mid X_{-1}, X_{-2}, \ldots)]$$

in terms of the function $p(x_0 \mid x_{-1}, x_{-2}, \ldots, x_{-n})$.

Example 22.1. If (X_i) is IID(π), then $p(x_0 \mid x_{-1}) = \pi(x_0)$, so $H = -\mathbb{E}[\log \pi(X_0)] = H(\pi)$.

Example 22.2. Let (X_i) be stationary Markov, $\mathbb{P}(X_i = x, X_{i+1} = y) = \pi(x)q(x,y)$, where $\mathbf{Q}$ is the transition matrix.

$$p(x_0 \mid x_{-1}, x_{-2}, \ldots) \stackrel{\text{Markov}}{=} p(x_0 \mid x_{-1}),$$

so

$$H = -\mathbb{E} \log p(\underbrace{X_0}_{y} \mid \underbrace{X_{-1}}_{x}) = \sum_x \sum_y \pi(x) q(x,y) \log q(x,y).$$

Corollary 22.3. *Let* $\hat{H}_k = H(\text{dist}(X_0, X_1, \ldots, X_{k-1}))$. *Then,*

$$\frac{1}{k} \hat{H}_k \to H \qquad \textit{as} \qquad k \to \infty.$$

22.2 Asymptotic Equipartition Property

Different Viewpoint. What do we know about (X_i) if we are told H but don't know $p(x_0, \ldots, x_n)$?

Consider $B_k \subseteq S^k$.

$$|B_k| \min_{\mathbf{x}\in B_k} p(\mathbf{x}) \le \mathbb{P}((X_0, X_1, \ldots, X_{k-1}) \in B_k) \le |B_k| \max_{\mathbf{x}\in B_k} p(\mathbf{x}).$$

Theorem 22.4 (Asymptotic Equipartition Property)**.** *Fix* $\delta > 0$.

(a) If $|B_k| = o(\exp(k(H-\delta)))$*, then* $\mathbb{P}((X_0, \ldots, X_{k-1}) \in B_k) \to 0$ *as* $k \to \infty$.

(b) $\exists B_k$ *with* $|B_k| = O(\exp(k(H+\delta)))$ *such that* $\mathbb{P}((X_0, \ldots, X_{k-1}) \in B_k) \to 1$ *as* $k \to \infty$.

Proof. (a)

$$\begin{aligned}
\mathbb{P}((X_0, \ldots, X_{k-1}) \in B_k) &\le \mathbb{P}\left((X_0, \ldots, X_{k-1}) \in B_k \text{ and } -\frac{1}{k}\log L_k \ge H - \delta\right) \\
&\quad + \underbrace{\mathbb{P}\left(-\frac{1}{k}L_k \le H - \delta\right)}_{o(1) \text{ as } k\to\infty} \\
&= \mathbb{P}((X_0, \ldots, X_{k-1}) \in B_k \cap B_k') \\
&\le |B_k| \exp(-k(H-\delta)) \to 0 \qquad \text{as} \qquad k \to \infty,
\end{aligned}$$

where

$$\begin{aligned}
B_k &= \left\{\mathbf{x} : -\frac{1}{k}\log p(\mathbf{x}) \ge H - \delta\right\} \\
&= \{\mathbf{x} : p(\mathbf{x}) \le \exp(-k(H-\delta))\}.
\end{aligned}$$

(b) Choose

$$B_k = \left\{\mathbf{x} : -\frac{1}{k}\log p(\mathbf{x}) \le H + \delta\right\}.$$

Then, 21.2 implies $\mathbb{P}((X_0, \ldots, X_{k-1}) \in B_k) \to 1$ as $k \to \infty$.

$$1 \ge \mathbb{P}((X_0, \ldots, X_k) \in B_k) \ge |B_k| \exp(-k(H+\delta)).$$

□

In fact, (b) holds for

$$B_k = \left\{\mathbf{x} : -\frac{1}{k}\log p(\mathbf{x}) \in [H-\delta, H+\delta]\right\}.$$

22.3 Subadditive Ergodic Theorem

Background:

1. Consider $\mathbb{R}$-valued RVs (ξ_i). Define $X_{m,n} = \sum_{i=m+1}^{n} \xi_i$.

$$X_{0,n} = X_{0,m} + X_{m,n}, \qquad 0 \le m \le n.$$

2. If the (ξ_i) are stationary, then for any fixed $k \ge 1$,

$$\text{dist}(X_{m,n} : 0 \le m \le n < \infty) = \text{dist}(X_{m+k,n+k} : 0 \le m \le n < \infty). \tag{22.1}$$

Theorem 22.5 (Kingman's Subadditive Ergodic Theorem)**.** *Suppose we have $\mathbb{R}$-valued random variables $(X_{m,n} : 0 \le m < n < \infty)$ satisfy (22.1) and*

$$X_{0,n} \le X_{0,m} + X_{m,n} \qquad 0 \le m < n < \infty \tag{22.2}$$

and

$$\mathbb{E}X_{0,1}^{+} < \infty \qquad \text{and} \qquad \inf_n \frac{\mathbb{E}X_{0,n}}{n} > -\infty. \tag{22.3}$$

Then

$$\frac{1}{n} X_{0,n} \xrightarrow[L^1]{a.s.} X,$$

say, and $\mathbb{E}X = \lim_n n^{-1}\mathbb{E}X_n = \inf_n n^{-1}\mathbb{E}X_n > -\infty$.

The Durrett text gives an alternate "Liggett" version. See the text for the proof.

Often, it is useful to show limits exist without explicit calculation.

Example 22.6 (Products of Random Matrices)**.** Let $A_1, A_2, \dots$ be a stationary sequence of random $s \times s$ matrices, with entries $A_m(i,j) > 0$. Consider the random matrix $\alpha_{m,n} = A_{m+1}A_{m+2}\cdots A_n$.

Proposition 22.7. *If $\mathbb{E}|\log A_1(i,j)| < \infty\ \forall i,j$, then*

$$\frac{1}{n}\log \alpha_{0,n}(i,j) \to -X \qquad a.s.$$

(some X).

Proof. Define $X_{m,n} = -\log \alpha_{m,n}(1,1)$. $\alpha_{0,n} = \alpha_{0,m}\alpha_{m,n}$, so $\alpha_{0,n}(1,1) \ge \alpha_{0,m}(1,1)\alpha_{m,n}(1,1)$. Therefore, $X_{m,n}$ has property (22.2) and property (22.1) follows from the fact that (A_i) is stationary.

$$\mathbb{E}X_{0,1}^{+} \le \mathbb{E}|\log A_1(1,1)| < \infty \qquad \text{by assumption.}$$

Note that $\alpha_{0,n}(1,1)$ is the sum of s^{n-1} terms of the form $A_1(1,i_1)A_2(i_1,i_2)\cdots A_n(i_{n-1},1)$, so

$$\alpha_{0,n}(1,1) \le s^{n-1}\prod_{m=1}^{n}\max_{i,j} A_m(i,j),$$

so

$$\mathbb{E}\frac{1}{n}\log\alpha_{0,n}(1,1) \le \log s + \mathbb{E}\log\max_{i,j} A_1(i,j) \equiv \beta < \infty,$$

which is (22.3). 22.5 implies

$$\frac{1}{n}\log\alpha_{0,n}(1,1) \to -X \qquad \text{a.s.}$$

For the general (i,j) entry,

$$\alpha_{0,n}(i,j) \ge A_1(i,1)\alpha_{1,n-1}(1,1)A_n(1,j),$$

so

$$-\frac{1}{n}\log \alpha_{0,n}(i,j) \to -X \qquad \text{a.s.}$$ □

Example 22.8 (First Passage Percolation on Square Lattice)**.** Let $(\tau_e, e \in E)$ be IID, $0 < \tau_e < \infty$, $\mathbb{E}\tau_e < \infty$, where E is the edges of the $\mathbb{Z}^2$ lattice. Define $X_{m,n}$ to be the time to travel from $(m,0)$ to $(n,0)$: $X_{m,n} = \min\{\sum_{e\in\pi} \tau_e : \pi \text{ a path from } (m,0) \text{ to } (n,0)\}$.

Check Hypotheses. (22.1) holds because the (τ_e) are invariant under translation by k.

$$\begin{aligned} X_{m,n} &\leq \text{minimum time route from } (0,0) \text{ to } (0,n) \text{ via } (m,0) \\ &= X_{0,m} + X_{m,n}, \end{aligned}$$

which checks (22.2). $X_{0,1} \leq \tau_e$, so $\mathbb{E}X_{0,1}^+ < \infty$, which checks (22.3). 22.5 implies

$$\frac{1}{n}X_{0,n} \to \text{some } X.$$

Note that changing a finite number of the τ_e does not change X. Therefore, $X \in \text{tail}(\tau_e, e \in E)$, which is trivial by the 0-1 Law, so X is constant.

Lecture 23

April 11

23.1 Law of Iterated Logarithm

Let $B(t), 0 \le t < \infty$ be standard Brownian motion.

Curious Fact: $\hat{B}(t) = tB(1/t)$ is also standard BM (calculate the covariance $\mathbb{E}[\hat{B}(s)\hat{B}(t)]$). So, limits as $t \to \infty$ are "equivalent" to limits as $t \to 0$.

Theorem 23.1 (Law of Iterated Logarithm). *(a)*

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{2t \log\log t}} = 1 \qquad a.s.$$

(b)

$$\limsup_{t\downarrow 0} \frac{B(t)}{\sqrt{2t \log\log(1/t)}} = 1 \qquad a.s. \tag{23.1}$$

Harder Result: If (X_i) are IID, $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, and $S_n = \sum_{i=1}^n X_i$, then

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \qquad \text{a.s.}$$

We will prove (23.1). Recall:

Lemma 23.2. *If $c > 0$, $d > 0$,*

$$\mathbb{P}\left(\sup_{0\le t<\infty} (B_t - td) \ge c\right) = \exp(-2cd).$$

Proof of 23.1. Write $h(t) = \sqrt{2t \log\log(1/t)}$. Fix $0 < \delta < \theta < 1$. Apply 23.2 with

$$d = \frac{1}{2}\theta^{-n}(1+\delta)h(\theta^n), \qquad c = \frac{1}{2}h(\theta).$$

So,

$$2cd = (1+\delta)\log\log\frac{1}{\theta^n} = (1+\delta)\log n + K_{\delta,\theta}.$$

23.2 implies

$$\mathbb{P}\left(\sup_t\left(B_t-\frac{1}{2}(1+\delta)\theta^{-n}h(\theta^n)t\right)\geq\frac{1}{2}h(\theta^n)\right)\leq\hat{K}_{\delta,\theta}n^{-(1+\delta)}.$$

Borel-Cantelli 1 implies

$$\sup_t\left(B_t-\frac{1}{2}(1+\delta)\theta^{-n}h(\theta^n)t\right)\leq\frac{1}{2}h(\theta^n)\qquad\text{for all } n\geq n_0(\omega).$$

Consider small t, say $\theta^{n+1}<t<\theta^n$, $n>n_0(\omega)$. Then,

$$B_t\leq\frac{1}{2}h(\theta^n)+\frac{1}{2}(1+\delta)\theta^{-n}h(\theta^n)t\leq\frac{1}{2}(2+\delta)h(\theta^n)\leq\frac{1}{2}(2+\delta)\theta^{-1/2}h(t),$$

since $h(t)\geq h(\theta^{n+1})\geq\theta^{1/2}h(\theta^n)$ for n large (check). Hence,

$$\limsup_{t\downarrow 0}\frac{B_t}{h(t)}\leq\frac{1}{2}(2+\delta)\theta^{-1/2}\quad\text{a.s.}$$

Let $\delta\downarrow 0$ and $\theta\uparrow 1$.

$$\leq 1\quad\text{a.s.}\qquad\text{(upper bound).}$$

Lower Bound. Fix $\theta>0$. Suppose we prove

$$\mathbb{P}(B(\theta^n)-B(\theta^{n+1})>(1-\theta)^{1/2}h(\theta^n)\text{ infinitely often})=1.\tag{23.2}$$

Then, by the upper bound (applied to $-B(t)$), $-B(\theta^{n+1})\leq 2h(\theta^{n+1})$ ultimately. Combining these two facts, $B(\theta^n)\geq(1-\theta)^{1/2}h(\theta^n)-2h(\theta^{n+1})$ infinitely often. But,

$$\frac{h(\theta^{n+1})}{h(\theta^n)}\to\theta^{1/2}\implies h(\theta^{n+1})\leq 2\theta^{1/2}h(\theta^n)\qquad\text{ultimately,}$$

since $h(t)=\sqrt{2t\log\log t}$, so $B(\theta^n)\geq((1-\theta)^{1/2}-4\theta^{1/2})h(\theta^n)$ infinitely often. Hence,

$$\limsup_{t\downarrow 0}\frac{B(t)}{h(t)}\geq\limsup_{n\to\infty}\frac{B(\theta^n)}{h(\theta^n)}\geq(1-\theta)^{1/2}-4\theta^{1/2}\qquad\text{a.s.}$$

Let $\theta\downarrow 0$.

Proof of (23.2): For Z, Normal$(0,1)$,

$$\begin{aligned}\mathbb{P}(Z>x)&\sim\frac{\phi(x)}{x}\\&\sim(2\pi)^{-1/2}x^{-1}\exp\left(-\frac{x^2}{2}\right)\qquad\text{as}\qquad n\to\infty.\end{aligned}$$

So,

$$\begin{aligned}\mathbb{P}(B(\theta^n)-B(\theta^{n+1})>(1-\theta)^{1/2}h(\theta^n))&=\mathbb{P}((\theta^n-\theta^{n+1})^{1/2}Z>(1-\theta)^{1/2}h(\theta^n))=\mathbb{P}(Z>\theta^{-n/2}h(\theta^n))\\&=\mathbb{P}(Z>\sqrt{2\log\log(1/\theta^n)})\qquad\text{(definition of } h(t))\\&\sim\text{constant}\cdot(\log n)^{-1/2}\cdot\frac{1}{n\log(1/\theta)}.\end{aligned}$$

Since the summation $\sum_n(\cdot)=\infty$, Borel-Cantelli 2 implies (23.2). □

23.2 Embedding Distributions into BM

Consider $B(t)$. Take $U \le 0 \le V$ (dependent), but independent of $B(t)$ with $\mathbb{E}U + \mathbb{E}V = 0$. Let

$$T = \inf\{t : B(t) = U \text{ or } V\}.$$

(205A) Conditional on $(U = u, V = v)$, $\mathbb{E}B_T^2 = \mathbb{E}T = -uv$, $\mathbb{E}B_T = 0$.

$$\mathbb{P}(B_T = u) = \frac{v}{v-u}, \qquad \mathbb{P}(B_T = v) = \frac{-u}{v-u}.$$

($B^2(t) - t$ is a MG.) Since $\mathbb{E}[B_T^2 \mid UV] = \mathbb{E}[T \mid UV]$, then $\mathbb{E}B_T^2 = \mathbb{E}T$, $\mathbb{E}B_T = 0$.

$$\mathbb{P}(B_T \in (u, u + \mathrm{d}u)) = \mathbb{E}\left[\frac{V}{V-u} 1_{(U \in (u, u+\mathrm{d}u))}\right], \qquad (u < 0), \tag{23.3}$$

$$\mathbb{P}(B_T \in (v, v + \mathrm{d}v)) = \mathbb{E}\left[\frac{-U}{v-U} 1_{(V \in (v, v+\mathrm{d}v))}\right], \qquad (v > 0). \tag{23.4}$$

Proposition 23.3. *Given* $\operatorname{dist}(X)$ *with* $\mathbb{E}X = 0$, *there exists a joint distribution* (U, V) *such that* $B_T \overset{\mathrm{d}}{=} X$.

Proof. We prove the case where X has some density $f(x)$. Recall $x = x^+ - x^-$. Then,

$$\mathbb{E}X = 0 \iff \mathbb{E}X^+ = \mathbb{E}X^- = c, \text{ say.}$$

Take the joint density for (U, V)

$$f_{U,V}(u, v) = \frac{f(u)f(v)(v-u)}{c}, \qquad u < 0 < v.$$

Check that the total mass is 1.

$$\int_0^\infty \int_{-\infty}^0 f_{U,V}(u, v)\,\mathrm{d}u\,\mathrm{d}v \overset{?}{=} 1.$$

The inner integral is

$$\int_{-\infty}^0 \frac{f(v)f(u)(v-u)}{c}\,\mathrm{d}u = \frac{vf(v)}{c}\mathbb{P}(X < 0) + \frac{f(v)}{c}\mathbb{E}X^-.$$

So

$$\begin{aligned}\int_0^\infty \left[\frac{vf(v)}{c}\mathbb{P}(X<0) + \frac{f(v)}{c}\mathbb{E}X^-\right]\mathrm{d}v &= \frac{\mathbb{E}X^+\mathbb{P}(X<0)}{c} + \frac{\mathbb{E}X^-\mathbb{P}(X>0)}{c} \\ &= \mathbb{P}(X<0) + \mathbb{P}(X>0) = 1.\end{aligned}$$

Also,

$$\begin{aligned}\frac{\mathbb{P}(B_T \in (u, u+\mathrm{d}u))}{\mathrm{d}u} &\overset{(23.3)}{=} \int_0^\infty \frac{v}{v-u} \cdot f_{U,V}(u, v)\,\mathrm{d}v = \int_0^\infty \frac{vf(u)f(v)}{c}\,\mathrm{d}v \\ &= f(u)\int_0^\infty \frac{vf(v)}{c}\,\mathrm{d}v \\ &= f(u)\frac{\mathbb{E}X^+}{c} = f(u).\end{aligned}$$

□

The Morters-Peres book section 5.3 gives other embeddings $B(T) \overset{\mathrm{d}}{=}$ given X.

23.3 Donsker's Invariance Principle

Donsker's Invariance Principle says that BM is the scaling limit of random walks.

Set-Up: We have IID (X_i), $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, $S_n = \sum_{i=1}^n X_i$. Interpolate to continuous $S(t)$.

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lceil t \rceil} - S_{\lfloor t \rfloor}).$$

Rescale time and space.

$$S_n^*(t) = \frac{S(nt)}{\sqrt{n}}, \qquad 0 \le t \le 1.$$

We can regard S_n^* as a random function, a RV taking values in the space $C[0,1]$ of continuous functions $f : [0,1] \to \mathbb{R}$. We can consider $(B(t), 0 \le t \le 1)$ as a RV B taking values in $C[0,1]$.

The theory of weak convergence on metric spaces formalizes the idea "$S_n^* \xrightarrow{\text{d}} B$".

The assertion $S_n^*(1) \xrightarrow{\text{d}} B(1)$ is the assertion

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{d}} \text{Normal}(0,1),$$

which is the CLT.

Lecture 24

April 13

24.1 Donsker's Invariance Principle

Setting. $(X_i, 1 \leq i < \infty)$ are IID, $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$, $S_n = \sum_{i=1}^n X_i$. $S(t)$ is the linear interpolation.

$$S_n^*(t) = \frac{1}{\sqrt{n}} S(nt), \qquad 0 \leq t \leq 1.$$

"As $n \to \infty$, the process S_n^* converges in distribution to BM."

(Last Class) Given $\text{dist}(X_1)$ and standard BM $(B(t), 0 \leq t < \infty)$, there exists a stopping time T_1 with $B(T_1) \stackrel{\text{d}}{=} X_1$ and $\mathbb{E}T_1 = 1$.

Use the Strong Markov Property. If $\tilde{B}(u) \stackrel{\text{def}}{=} B(T_1 + u) - B(T_1)$, then the process $(\tilde{B}(u), 0 \leq u < \infty)$ is distributed as BM independent of $\mathcal{F}(T_1)$. There exists a stopping time T_2 for $\tilde{B}$ such that $\tilde{B}(T_2) \stackrel{\text{d}}{=} X_2$ and is independent of $B(T_1)$. Now, $(B(T_1), B(T_1 + T_2)) \stackrel{\text{d}}{=} (X_1, X_1 + X_2)$.

Conclusion: There exist IID $(\tilde{T}_i, 1 \leq i < \infty)$ such that

$$\begin{aligned}(B(\tilde{T}_1), B(\tilde{T}_1 + \tilde{T}_2), B(\tilde{T}_1 + \tilde{T}_2 + \tilde{T}_3), \dots) &\stackrel{\text{d}}{=} (S_1, S_2, S_3, \dots) \\ &\stackrel{\text{d}}{=} (B(T_1), B(T_2), \dots),\end{aligned}$$

where $T_k = \sum_{i=1}^k \tilde{T}_i$.

Trick: Work with this construction of $S(t)$ and $S_n^*(t)$.

Idea: $S_k \approx B(k)$ to first-order.

Proposition 24.1.

$$\forall \varepsilon > 0 \qquad \lim_{n\to\infty} \mathbb{P}\left(\sup_{0 \leq t \leq 1} \left| \frac{B(nt)}{\sqrt{n}} - S_n^*(t) \right| > \varepsilon \right) = 0.$$

Why?

$$S_n^*(t) = \frac{S_{nt}}{\sqrt{n}} \approx \frac{B(nt)}{\sqrt{n}}$$

by the SLLN for the (T_i).

Proof. Set

$$W_n(t) = \frac{B(nt)}{\sqrt{n}}. \tag{24.1}$$

$W_n(t)$ is distributed as BM. Study every $A_n \stackrel{\text{def}}{=} \{\exists 0 \le t \le 1 \ |S_n^*(t) - W_n(t)| > \varepsilon\}$. Set $k = k(t)$ such that

$$\frac{k-1}{n} \le t \le \frac{k}{n}.$$

Note that

$$A_n \subseteq \left\{\exists 0 \le t \le 1 : \left|\frac{S_k}{\sqrt{n}} - W_n(t)\right| > \varepsilon\right\} \cup \left\{\exists 0 \le t \le 1 : \left|\frac{S_{k-1}}{\sqrt{n}} - W_n(t)\right| > \varepsilon\right\}.$$

If an average is $> \varepsilon$, then one of the items $> \varepsilon$. Rewrite (24.1):

$$S_k = B(T_k) = \sqrt{n} W_n\left(\frac{T_k}{n}\right).$$

Then,

$$\begin{aligned} A_n \subseteq &\left\{\exists 0 \le t \le 1 : \left|W_n\left(\frac{T_k}{n}\right) - W_n(t)\right| > \varepsilon\right\} \cup \left\{\exists 0 \le t \le 1 : \left|W_n\left(\frac{T_{k-1}}{n}\right) - W_n(t)\right| > \varepsilon\right\} \\ &\equiv A_n^*, \text{ say.} \end{aligned}$$

Repeat the "continuity of BM" argument.

Claim: Take $\delta > 0$. If A_n^*, then

$$D_n(\delta) \stackrel{\text{def}}{=} \left\{\exists 0 \le t \le 1 : \max\left(\left|\frac{T_{k-1}}{n} - t\right|, \left|\frac{T_k}{n} - t\right|\right) \ge \delta\right\}$$

or

$$D_n^*(\delta) \stackrel{\text{def}}{=} \{\exists 0 \le s, t \le 2 : |s - t| \le \delta, |W_n(s) - W_n(t)| > \varepsilon\}.$$

$\mathbb{P}(D_n^*(\delta)) \to 0$ as $\delta \to 0$ (uniformly in n) because BM paths are continuous.

Need to show: $\mathbb{P}(D_n(\delta)) \to 0$ as $n \to \infty$, for fixed $\delta > 0$. By the SLLN,

$$\frac{T_n}{n} \to 1 \qquad \text{a.s.}$$

By 24.2,

$$\mathbb{P}\left(\sup_{0 \le k \le n} \frac{|T_k - k|}{n} \ge \delta\right) \to 0 \qquad \text{as} \qquad n \to \infty. \tag{24.2}$$

In $D_n(\delta)$, we have

$$\frac{k-1}{n} \le t \le \frac{k}{n}.$$

Take

$$n > \frac{2}{\delta}.$$

In $D_n(\delta)$, the maximum must be attained with

$$t = \frac{k}{n} \qquad \text{or} \qquad t = \frac{k-1}{n}.$$

Therefore,

$$\mathbb{P}(D_n(\delta)) \le \mathbb{P}\left(\sup_{1\le k\le n} \max\left(\frac{T_k-(k-1)}{n}, \frac{k-T_{k-1}}{n}\right) > \delta\right).$$

Because

$$\frac{1}{n} < \frac{\delta}{2},$$

one has

$$\mathbb{P}(D_n(\delta)) \le \mathbb{P}\left(\sup_{1\le k\le n} \frac{T_k-k}{n} > \frac{\delta}{2}\right) + \mathbb{P}\left(\sup_{1\le k\le n} \frac{(k-1)-T_{k-1}}{n} > \frac{\delta}{2}\right)$$
$$\to 0 \qquad \text{as} \qquad n\to\infty \quad \text{by (24.2)}.$$

□

Lemma 24.2 (Deterministic Lemma)**.** *If*

$$\frac{a(n)}{n} \to 1,$$

then

$$\sup_{1\le k\le n} \frac{|a(k)-k|}{n} \to 0 \qquad \textit{as} \qquad n\to\infty.$$

Consider the metric space $(C[0,1], d)$ on the space of continuous functions $f : [0,1]\to\mathbb{R}$, with

$$d(f_1,f_2) = \sup_{0\le t\le 1} |f_1(t)-f_2(t)|.$$

We have seen a little about "weak convergence on metric spaces".

Easy general fact, applied to our setting: If S_n^*, W_n^*, and W (W is the BM process) satisfy

(i) $d(S_n^*, W_n^*) \to 0$ in probability as $n\to\infty$,

(ii) $W_n^* \overset{\mathrm{d}}{=} W \ \forall n$,

then $S_n^* \xrightarrow{\mathrm{d}} W$.

Here, we have

$$W_n^* = \frac{B(nt)}{\sqrt{n}}$$

and $\mathbb{P}(d(W_n^*, S_n^*) > \varepsilon) \to 0\ \forall\varepsilon$: 24.1 says $d(W_n^*, S_n^*) \to 0$ in probability. This is Donsker's Invariance Principle. $S_n^* \to W$ in distribution on $C[0,1]$.

As a general "weak convergence" fact, applied to Donsker's Theorem:

Corollary 24.3. *If $\psi : C[0,1] \to \mathbb{R}$ is continuous, or more generally, if $\mathbb{P}(W \in \mathcal{D}_\psi) = 0$ for*

$$\mathcal{D}_\psi \stackrel{\text{def}}{=} \{f : \psi \text{ is not continuous at } f\},$$

then $\psi(S_n^*) \xrightarrow{\text{d}} \psi(W)$ on $\mathbb{R}$.

Example 24.4. $\psi(f) \stackrel{\text{def}}{=} \sup_{0\le t\le 1} f(t)$. This *is* everywhere continuous because

$$|\psi(f) - \psi(g)| \le \sup_t |f(t) - g(t)| \equiv d(f,g).$$

Example 24.5. $\psi(f) = \text{Leb}\{t \in [0,1] : f(t) > 0\}$. If we take

$$f_n(t) \equiv \frac{1}{n}, \qquad \psi(f_n) = 1,$$
$$f(t) \equiv 0, \qquad \psi(f) = 0,$$

but $f_n \to f$, so ψ is not continuous. If f satisfies

$$\text{Leb}\{t : f(t) = 0\} = 0, \tag{24.3}$$

then ψ *is* continuous at f. If $f_n \to f$, then $1_{(f_n>0)} \to 1_{(f>0)}$ outside $\{f = 0\}$. If $f_n \to f$ and f satisfies (24.3), then $1_{(f_n(t)>0)} \to 1_{(f(t)>0)}$ a.e. $\implies \int_0^1 1_{(f_n(t)>0)}\,\text{d}t \to \int_0^1 1_{(f(t)>0)}\,\text{d}t$, so $\psi(f_n) \to \psi(f)$.

$$\mathcal{D}_\psi = \{f : \text{Leb}\{t : f(t) = 0\} > 0\}.$$

To use 24.3, we need to show $\mathbb{P}(\text{Leb}\{t : W(t) = 0\} > 0) = 0$. It is enough to show

$$\mathbb{E}[\text{Leb}\{t : W(t) = 0\}] = 0,$$

but we have $\int_0^1 \mathbb{P}(W_t = 0)\,\text{d}t = 0$ because $\mathbb{P}(W_t = 0) = 0$ for $t > 0$.

Example 24.6.

$$\psi(f) = \inf\left\{s : f(s) = \sup_{0\le t\le 1} f(t)\right\}.$$

Exercise: If f has the property

$$\left\{s : f(s) = \sup_t f(t)\right\} \text{ is a single point} \tag{24.4}$$

then ψ is continuous at f. To apply 24.3, we need to show $\mathbb{P}(B \text{ has property (24.4)}) = 1$.

Lecture 25

April 25

25.1 Martingale Central Limit Theorem

Take standard Brownian motion $(B(t), 0 \le t < \infty)$. Given dist(X) with $\mathbb{E}X = 0$, there exists a stopping time T such that $B(T) \stackrel{d}{=} X$, which implies $\mathbb{E}T = \mathbb{E}X^2 = \mathrm{var}(X)$. We can show $\mathbb{E}T^2 \le c\mathbb{E}X^4$ for constant c.

Theorem 25.1 (Martingale Embedding into BM)**.** *Take a MG $0 = S_0, S_1, S_2, \ldots$. Then, there exists stopping times $0 = T_0 \le T_1 \le T_2$ such that $(S_0, S_1, S_2, \ldots) \stackrel{d}{=} (B(T_0), B(T_1), B(T_2), \ldots)$.*

Proof. By induction on k. Condition on $(S_0 = 0, S_1 = s_1, \ldots, S_k = s_k)$ (or condition on $\mathcal{F}_k$). The conditional distribution of $(S_{k+1} - S_k)$ given $\mathcal{F}_k$ is a mean-0 distribution. Apply the embedding to the conditional distribution and $(B(T_k + t) - B(T_k), t \ge 0)$ to get $T_{k+1} - T_k = \hat{T}_k$. □

Note: $\mathbb{E}[T_{k+1} - T_k \mid \mathcal{F}_k] = \mathbb{E}[(S_{k+1} - S_k)^2 \mid \mathcal{F}_k]$ and

$$\mathbb{E}[(T_{k+1} - T_k)^2 \mid \mathcal{F}_k] \le c\mathbb{E}[(S_{k+1} - S_k)^4 \mid \mathcal{F}_k]. \tag{25.1}$$

Theorem 25.2 (Lindeberg-Feller CLT for Martingales)**.** *For each n, let $(X_{n,m}, \mathcal{F}_{n,m}, m = 0, 1, \ldots, n)$ be a martingale difference sequence, that is, $(S_{n,m}, \mathcal{F}_{n,m}, m = 0, 1, \ldots, n)$ is a MG, $S_{n,m} = \sum_{i=1}^m X_{n,i}$, that is, $X_{n,m}$ is $\mathcal{F}_{n,m}$-measurable, $\mathbb{E}[X_{n,m+1} | \mathcal{F}_{n,m}] = 0$. Write $V_{n,k} = \sum_{m=1}^k \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}]$. Suppose*

(i) $V_{n,nt} \underset{\mathbb{P}}{\to} t$ as $n \to \infty$, $0 \le t \le 1$ fixed ($V_{n,nt}$ defined by linear interpolation),

(ii) $\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 1_{(|X_{n,m}| > \varepsilon)} \mid \mathcal{F}_{n,m-1}] \underset{\mathbb{P}}{\to} 0$ as $n \to \infty$.

Then, $(S_{n,nt}, 0 \le t \le 1) \xrightarrow{d} (B(t), 0 \le t \le 1)$ as $C[0,1]$-valued random functions. In particular, $S_{n,n} \xrightarrow{d}$ Normal$(0,1)$.

Outline Proof. (See Durrett 3rd Edition).

We prove this under the stronger assumption $|X_{n,m}| \le \varepsilon_n$, $\varepsilon_n \downarrow 0$. For a single sequence $(\xi_i, i \ge 1)$, then

$$X_{n,i} = \frac{\xi_i}{\sqrt{n}},$$

so the stronger assumption is saying $|\xi_n| \le \varepsilon_n \sqrt{n}$.

If we stop the process if $V_{n,\cdot}$ reaches $3/2$, take $\varepsilon_n < 1/2$, then we can assume $V_{n,n} \le 2$ by (i).

Regard the embedding $(B(T_{n,m}), m = 0,1,\dots,n)$ as the definition of $(S_{n,m}, m = 0,1,\dots,n)$. So, $(S_{n,nt}, 0 \le t \le 1) \stackrel{\mathrm{d}}{=} (B(T_{n,nt}), 0 \le t \le 1)$. It is enough to show $T_{n,nt} \underset{\mathbb{P}}{\to} t$ as $n \to \infty$ (for fixed t), and then use continuity of BM paths as in Donsker's Theorem. Write $t_{n,m} = T_{n,m} - T_{n,m-1}$.

$$\mathbb{E}[t_{n,m} \mid \mathcal{F}_{m,m-1}] = \mathbb{E}[X_{n,m}^2 \mid \mathcal{F}_{n,m-1}] \stackrel{(i)}{\Longrightarrow} \sum_{m=1}^{nt} \mathbb{E}[T_{n,m} \mid \mathcal{F}_{n,m-1}] \underset{\mathbb{P}}{\to} t \quad \text{as} \quad n \to \infty.$$

By orthogonality of the increments of the MDS $t_{n,m} - \mathbb{E}[T_{n,m} \mid \mathcal{F}_{n,m-1}]$,

$$\begin{aligned}\mathbb{E}(T_{n,nt} - V_{n,nt})^2 = \mathbb{E}\left(\sum_{m=1}^{nt} t_{n,nt} - \mathbb{E}[t_{n,m} \mid \mathcal{F}_{n,m-1}]\right)^2 &\underset{\text{orthogonality}}{=} \mathbb{E}\sum_{m=1}^{nt}(t_{n,m} - \mathbb{E}[t_{n,m} \mid \mathcal{F}_{n,m-1}])^2 \\ &\underset{(25.1)}{\le} c\mathbb{E}\left[\sum_{m=1}^{nt} \mathbb{E}[X_{n,m}^4 \mid \mathcal{F}_{n,m-1}]\right] \\ &\le c\mathbb{E}\left[\varepsilon_n^2 \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 \mid \mathcal{F}_{n,m-1}]\right] \le c\varepsilon_n^2 \mathbb{E}V_{n,n} \\ &\to 0 \quad \text{as} \quad n \to \infty.\end{aligned}$$

So, $T_{n,nt} - V_{n,nt} \underset{\mathbb{P}}{\to} 0$. □

25.2 The 3 Arcsine Laws

The 3 arcsine RVs associated with $(B(t), 0 \le t \le 1)$:

1. Consider $L = \sup\{t \le 1 : B(t) = 0\}$, $0 \le L \le 1$.

   $$\mathbb{P}(L \le t \mid B(t) = a) = \mathbb{P}(T_{|a|} > 1 - t),$$

   so

   $$\mathbb{P}(L \le t) = \int_0^\infty \mathbb{P}(T_{|a|} > 1 - t) f_{B(t)}(a)\,\mathrm{d}a.$$

   We know how to calculate these quantities since

   $$\mathbb{P}(T_b \le s) = \mathbb{P}\left(\max_{0 \le u \le s} B_s \ge b\right) = \mathbb{P}(|B_s| \ge b).$$

   From calculus, the density is

   $$f_L(t) = \frac{1}{\pi t^{1/2}(1-t)^{1/2}}, \qquad 0 < t < 1. \tag{25.2}$$

2. Consider $M(t) \stackrel{\text{def}}{=} \sup_{0 \le s \le t} B(s)$.

   Fact: The process $(M(t) - B(t), 0 \le t < \infty)$ has the same distribution as $(|B(t)|, 0 \le t < \infty)$. This is different from the fact $M(t) \stackrel{\mathrm{d}}{=} |B(t)|$, which holds for fixed t.

   The RV L applied to $(M(t) - B(t))$ is some RV $\hat{L}$ applied to $(B(t))$.

   $$\hat{L} = \sup\{t \le 1 : B(t) = M(t)\} = \inf\{t : B(t) = M(1)\}.$$

So, $\hat{L}$ also has the same arcsine density $f_L(t)$ at (25.2).

Rewrite: $\psi_2 : C[0,1] \to \mathbb{R}$, where

$$\psi_2(f) = \inf \left\{ t : f(t) = \sup_{0 \le s \le t} f(s) \right\}.$$

$\psi_2(B)$ has the arcsine density.

3. We considered (last class) $\psi_3(t) = \text{Leb}\{0 \le t \le 1 : f(t) > 0\}$.

   Fact: $\psi_3(B)$ also has the arcsine density.

   History: The original proof is based on a combinatorial identity for a simple symmetric RW

$$S_m = \sum_{i=1}^{m} \xi_i.$$

   The combinatorial identity is

$$\#\{1 \le k \le n : S_k > 0\} \stackrel{\text{d}}{=} \min \left\{ k \le n : S_k = \max_{0 \le j \le n} S_j \right\}.$$

   Multiply by $1/n$.

$$\frac{1}{n}\#\{1 \le k \le n : S_k > 0\} \stackrel{\text{d}}{=} \frac{1}{n} \min \left\{ k \le n : S_k = \max_{0 \le j \le n} S_j \right\}. \tag{25.3}$$

   Rescale to

$$S_n^*(t) = \frac{S_{nt}}{\sqrt{n}}.$$

   The LHS of (25.3) is close to $\psi_3(S_n^*)$ and the RHS of (25.3) is close to $\psi_2(S_n^*)$. As $n \to \infty$, the differences converge in probability to 0. Donsker's Theorem implies that

$$\psi_2(S_n^*) \xrightarrow{\text{d}} \psi_2(B),$$
$$\psi_3(S_n^*) \xrightarrow{\text{d}} \psi_3(B),$$

   which implies $\psi_2(B) \stackrel{\text{d}}{=} \psi_3(B)$.

Lecture 26

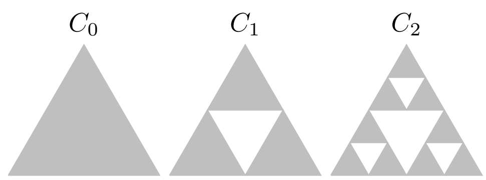
April 27

26.1 Local Time for Brownian Motion

[Morters-Peres book, Chapter 6.]

26.1.1 Existence

The classic example of a fractal set is $C_0, C_1, C_2, \ldots$. The C_n are closed and $C_n \downarrow C_\infty$, so C_∞ is closed and non-empty. $\text{area}(C_\infty) = 0$.



Instead, consider PMs where μ_n is a *uniform* (relative to area) PM on C_n. Then, $\mu_n \to \mu_\infty$ weakly, with $\text{supp}(\mu_\infty) = C_\infty$. Intuitively, μ_∞ is a "uniform" PM on C_∞.

For $(B(t), 0 \le t < \infty)$, the zero-set $Z(\omega) = \{t : B(t,\omega) = 0\}$ is a random closed subset of $[0,\infty)$. We know $\text{Leb}(Z(\omega)) = 0$ a.s. since $\mathbb{P}(B(t) = 0) = 0$, $t > 0$. [MP] proves that the Hausdorff dimension of $Z(\omega)$ is $1/2$ a.s. If we have any measure on $Z(\omega)$, we can describe it via

$$L(t,\omega) = \text{measure of } Z(\omega) \restriction [0,t],$$

which must have the property

$$t \mapsto L(t,\omega) \text{ increases only on } \{t : t \in Z(\omega)\} = \{t : B(t) = 0\}. \tag{26.1}$$

We will give a construction of a process called "local time at 0" which has the property (26.1).

Study $D(a,b,t)$, the number of downcrossings completed by time t.

Theorem 26.1. *There exists a process $(L(t), 0 \le t < \infty)$ such that for all $a_n \uparrow 0$, $b_n \downarrow 0$,*

$$\lim_{n\to\infty} Z(b_n - a_n) D(a_n, b_n, t) = L(t) \qquad \textit{a.s.}$$

Clearly, such $L(t)$ has property (26.1).

Key Idea: Take $a < m < b$. Look at one downcrossing over $[a, b]$ followed by an upcrossing. X^* is the number of downcrossings of $[a, m]$ and Y^* is the number of downcrossings of $[m, b]$. We know

$$\mathbb{P}_m(T_a < T_b) = \frac{b-m}{b-a}.$$

Then,

$$Y^* = 1 + \begin{cases} 0 & \text{with probability } p = \dfrac{b-m}{b-a}, \\ 1 + Y^{**} & \text{with probability } 1-p, \end{cases}$$

where $Y^{**} \stackrel{\text{d}}{=} Y^*$. So,

$$X^* \stackrel{\text{d}}{=} \text{Geometric}\left(\frac{m-a}{b-a}\right)$$

$$Y^* \stackrel{\text{d}}{=} \text{Geometric}\left(\frac{b-m}{b-a}\right),$$

independent.

Lemma 26.2. *Take $a < m < b$ and a stopping time T with $B(T) \geq b$. Write $D = D(a, b, T)$ and $D(a, m, T)$ and $D(m, b, T)$. These are related by*

$$D(a, m, T) = X_0 + \sum_{i=1}^{D} X_j,$$

$$D(m, b, T) = Y_0 + \sum_{j=1}^{D} Y_j,$$

where the X's, Y's, and D are independent, $X_j \stackrel{\text{d}}{=} X^*$, $j \geq 1$, $Y_j \stackrel{\text{d}}{=} Y^*$, $j \geq 1$, $X_0 \geq 0$, and $Y_0 \geq 0$.

Lemma 26.3. *Take $a_n \uparrow 0$, $b_n \downarrow 0$, and $b > b_1 > b_2 > \cdots$. The discrete-"time" process*

$$(2(b_n - a_n) D(a_n, b_n, T_b), n = 1, 2, \dots)$$

is a submartingale and converges a.s. to $L(T_b)$, say, as $n \to \infty$.

Proof. We can assume $a_{n+1} = a_n$, $b_{n+1} < b_n$.

$$\mathbb{E}[D(a_n = a_{n+1}, b_{n+1}, T_b) \mid \mathcal{F}_n] = (\geq 0) + (\mathbb{E}X^*) \cdot D(a_n, b_n, T_b).$$

Note that

$$\mathbb{E}X^* = \frac{b_n - a_n}{b_{n+1} - a_{n+1}}.$$

This is the sub-MG property.

If $G \stackrel{\text{d}}{=} \text{Geometric}(p)$, then $\mathbb{E}G^2 \leq 2/p^2$.

[MP] says

$$D(a_n, b_n, T_b) \stackrel{\text{d}}{=} \text{Geometric}\left(\frac{b_n - a_n}{b - a_n}\right).$$

Actually, the LHS is smaller. So,

$$\mathbb{E}(2(b_n - a_n)D(a_n, b_n, T_n))^2 \leq 8(b - a_n)^2 \to 8b^2.$$

Apply the Sub-MG Convergence Theorem. □

After the stopping time $\hat{T}_t$, then $\hat{B}(u) \stackrel{\text{def}}{=} B(\hat{T}_t + u)$, $u \geq 0$ is BM. Apply the construction to $\hat{B}(n)$ to get $\hat{L}(T_b)$.

Trick: Define

$$L(t) = \lim_{b \to \infty} L(T_b) - \hat{L}(T_b).$$

We can show that the paths $t \mapsto L(t, \omega)$ are continuous.

26.1.2 Connection with the Maximum Process

Why is $L(t)$ interesting?

Recall $|B(t)|$ is "reflecting BM". Given $B(t)$, consider $M(t) = \sup_{0 \leq s \leq t} B(s)$.

Fact: Given BM $B_1(t)$ and $M_1(t)$, the process $B_2(t) \stackrel{\text{def}}{=} M_1(t) - B_1(t)$ is distributed as reflecting BM. We have the "same" $L(t)$ for $B(t)$ and $|B(t)|$.

Given this fact, consider $L(t)$, local time at zero for $B_2(t)$. $t \mapsto L(t)$ has the property (26.1): it is increasing only at t such that $B_2(t) = 0$, that is, when $B_1(t) = M_1(t)$. But, $t \mapsto M_1(t)$ has the same property (26.1). This suggests:

Fact: The process $(L(t), 0 \leq t < \infty) \stackrel{\text{d}}{=} (M(t), 0 \leq t < \infty)$.

26.1.3 Occupation Density

Consider $f : [0, t] \to \mathbb{R}$. There is always an "occupation measure" on $\mathbb{R}$

$$\mu_t(\cdot) = \text{Leb}\{t : f(t) \in \cdot\}.$$

This may or may not have a density

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}\,\text{Leb}}(y) = \ell_t(y), \qquad y \in \mathbb{R}.$$

If f is smooth,

$$\ell_t(y) = \sum_{\substack{x \leq t \\ f(x) = y}} \frac{1}{f'(x)}.$$

It is not obvious if "occupation density" exists for BM paths.

Theorem 26.4 (Local Time = Occupation Density)**.** *There exists* $(L(t, y, \omega), 0 \leq t < \infty, y \in \mathbb{R})$ *such that* $y \mapsto L(t, y, \omega)$ *is the occupation density of the function* $(s \mapsto B(s, \omega), 0 \leq s \leq t)$ *and also* $(t, y) \mapsto L(t, y, \omega)$ *is jointly continuous.*

Idea: $L(t, 0, \omega)$ is the $L(t)$ process we constructed.

For each y, we repeat the construction with $a_n \uparrow y$, $b_n \downarrow y$ to get $L(t, y, \omega)$.

Fact: For BM, $\mathbb{E}$[time spent within $[a, b]$ during downcrossings over $[a, b]$] $= (b-a)^2$.

In the limit,

$$L(t) = \lim_n 2(b_n - a_n) D(a_n, b_n, t).$$

By the SLLN, the total amount of time spent in $[a_n, b_n] \sim 2(b_n - a_n) D(a_n, b_n, t)$.