Probability

Zixi Zhou [1]

May 2, 2026

[1]This note was done by 2026 spring.

Contents

Chapter 1

Motivation Example

Example 1.0.1 (Branching process)**.**

1. Z_n: number of people of nth generation
2. X_i^{n+1} number of offspring the ith person in nth generation has
3. Assume X_i^{n+1} are iid (independent identically distributed).

$$Z_{n+1} = X_1^{n+1} + X_2^{n+1} + \cdots + X_{Z_n}^{n+1}.$$

Assume $Z_0 = 1$. What is the extinction probability π for given X?

$$\pi = \mathbb{P}(\exists\ n \text{ s.t. } Z_n = 0)$$

Generating function of X

$$f(\theta) = \mathbb{E}[\theta^X] = \sum_{k\in\mathbb{Z}} \theta^X \Pr(X = k).$$

We have

$$f_n(\theta) = \mathbb{E}[\theta^{Z_n}].$$

What we want to know is that $\pi_n = \Pr(Z_n = 0)$.

Claim 1.0.2. $\pi = \lim_{n\to\infty} \pi_n$

We can prove this by Dominated Convergence Theorem.

Claim 1.0.3. $f_{n+1}(\theta) = f_n(f(\theta))$.

Proof. By law of total expectation, we have

$$\begin{aligned} f_{n+1}(\theta) &= \mathbb{E}[\mathbb{E}[\theta^{Z_{n+1}} \mid Z_n]] \\ &= \mathbb{E}[\mathbb{E}[\theta^{X_1^{n+1}+X_2^{n+1}+\cdots+X_{Z_n}^{n+1}} \mid Z_n]] \\ &= \mathbb{E}[f(\theta)^{Z_n}] \\ &= f_n(f(\theta)). \end{aligned}$$

□

Now we have $\pi_n = f_n(0) = f_{n-1}(f(0)) = f(f_{n-1}(0)) = f(\pi_{n-1})$. Since f is continuous, we have

$$\pi = f(\pi).$$

Theorem 1.0.4. If $\mathbb{E}X > 1$, then π is the unique solution to $x = f(x)$ that lies $(0, 1)$. If $\mathbb{E}X \leq 1$, then $\pi = 1$.

Can we describe more precisely what happens to Z_n as $n \to \infty$?

Observation 1.0.5. $Z_1, Z_2, \ldots$ is a Markov chain.

$$\Pr(Z_{n+1} = j \mid Z_0 = i_0, \ldots, Z_n = i_n) = \Pr(Z_{n+1} = j \mid Z_n = i_n),$$

and we have

$$\mathbb{E}(Z_{n+1} = j \mid Z_0, \ldots, Z_n) = \mathbb{E}(Z_{n+1} \mid Z_n)$$

Observation 1.0.6. Define $M_n = \frac{z_n}{\mu_n}$ and $\mu = \mathbb{E}X$ i.e. $\mathbb{E}[M_n \mid M_{n-1}] = M_{n-1}$. M_n is a martingale.

Proof.

$$\begin{aligned}
\mathbb{E}[M_n \mid M_{n-1}] &= \mathbb{E}[\frac{z_n}{\mu_n} \mid Z_{n-1}] \\
&= \frac{1}{\mu_n}\mathbb{E}[X_1^{n+1} + X_2^{n+1} + \cdots + X_{Z_n}^{n+1} \mid Z_{n-1}] \\
&= \frac{Z_{n-1} \cdot \mu}{\mu_n} \\
&= M_{n-1}.
\end{aligned}$$

By Martingale Convergence, $M_\infty(\omega) = \lim_{n\to\infty} M_n(\omega)$. Then $M_n \to M_\infty$ a.s. $\square$

Example 1.0.7. if $X \sim Geom(p)$, $P(X = k) = (1-p)^k p$ for $0 < p < 1$.

How do we know which solution is it?

If $\frac{p}{1-p} \geq 1 \iff p \geq \frac{1}{2}$ so $\pi = 1$. If $\frac{p}{1-p} < 1$ so $\pi =?$.

Chapter 2

Preliminaries

2.1 Measure

Definition 2.1.1 (Probability measure)**.** Let Ω be a sample space. A probability measure on Ω is a function $\mathbb{P} : 2^\Omega \to [0, 1]$ s.t.

1. $\mathbb{P}(\Omega) = 1$
2. If $\{E_j\}_{j=1}^\infty$ in 2^Ω are disjoint, then

$$\mathbb{P}\left(\cup_{j=1}^\infty E_j\right) = \sum_{j=1}^\infty \mathbb{P}(E_j)$$

Definition 2.1.2 (Field)**.** A collection $\mathcal{F} \subseteq 2^\Omega$ is a field if

1. $\Omega \in \mathcal{F}$
2. If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
3. If $E_1, \ldots, E_n \in \mathcal{F}$, then $\cup_{j=1}^n \in \mathcal{F}$.

Definition 2.1.3 (σ Field)**.** A collection $\mathcal{F} \subseteq 2^\Omega$ is a field if

1. $\Omega \in \mathcal{F}$
2. If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
3. If $\{E_j\}_{j=1}^\infty$ is a countable set of events in $\mathcal{F}$, then $\cup_{j=1}^\infty E_j \in \mathcal{F}$.

Lemma 2.1.4. If I is any index set and $\{\mathcal{F}_i : i \in I\}$ are σ-fields over Ω, then $\cap_{i=I} \mathcal{F}_i$ is a σ-fields.

Proof. Just check the definition of σ-fields. □

Proposition 2.1.5. Let $\mathcal{E} \subseteq 2^\Omega$ be any collection of subsets of Ω. There exists a unique smallest σ-field $\sigma(\mathcal{E})$ that contains $\mathcal{E}$. This σ-field is called the σ-field generated by $\mathcal{E}$.

Proof.

$$\sigma(\mathcal{E}) = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma - \text{field over } \Omega \text{ s.t. } \mathcal{E} \subseteq \mathcal{F}\}.$$

□

Definition 2.1.6. Let X be a topological space (e.g. Euclidean space $\mathbb{R}^d$). The Borel σ-field on X, denoted by $\mathcal{B}(X)$, is the σ-field generated by the collection of open subsets of X, that is,

$$\mathcal{B}(X) = \sigma\big(\{\, U \subseteq X : U \text{ is open} \,\}\big).$$

The elements of $\mathcal{B}(X)$ are called Borel sets.

Fact 2.1.7. Let $d \in \mathbb{N}$ and let $X = \mathbb{R}^d$ equipped with the Euclidean topology. Then every open set $U \subseteq \mathbb{R}^d$ can be written as a countable union of open balls, that is,

$$U = \bigcup_{i=1}^{\infty} B(x_i, r_i)$$

for some sequences $\{x_i\}_{i=1}^{\infty} \subseteq \mathbb{R}^d$ and $\{r_i\}_{i=1}^{\infty} \subseteq (0, \infty)$. Consequently, the Borel σ-field on $\mathbb{R}^d$ satisfies

$$\mathcal{B}(\mathbb{R}^d) = \sigma\big(\{\, B(x, r) : x \in \mathbb{R}^d,\ r > 0 \,\}\big).$$

In particular, for $d = 1$,

$$\mathcal{B}(\mathbb{R}) = \sigma\big(\{\, (a, b) : a < b \,\}\big) = \sigma\big(\{\, [a, b) : a < b \,\}\big) = \sigma\big(\{\, (a, b] : a < b \,\}\big).$$

Definition 2.1.8. Measurable space: $(\Omega.\mathcal{F})$ where Ω denotes any set, and $\mathcal{F}$ is σ-field.

Definition 2.1.9. Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \to [0, \infty]$ is called a measure if it is countably additive, that is, for any sequence $\{E_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ of disjoint sets,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n).$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space. If $\mu(\Omega) < \infty$, then μ is called a finite measure. If $\mu(\Omega) = 1$, then μ is called a probability measure, and $(\Omega, \mathcal{F}, \mu)$ is called a probability space.

Theorem 2.1.10. Let $(\Omega, \mathcal{F})$ be a measurable space. If μ is a measure on $(\Omega, \mathcal{F})$ and $\alpha \geq 0$, then the function $\alpha\mu : \mathcal{F} \to [0, \infty]$ defined by

$$(\alpha\mu)(E) = \alpha\, \mu(E), \qquad E \in \mathcal{F},$$

is a measure.

If $\{\mu_j\}_{j=1}^{\infty}$ is a countable family of measures on $(\Omega, \mathcal{F})$, then the function $\mu : \mathcal{F} \to [0, \infty]$ defined by

$$\mu(E) = \sum_{j=1}^{\infty} \mu_j(E), \qquad E \in \mathcal{F},$$

is a measure.

Theorem 2.1.11 (Basic Properties)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

1. Monotone: If $A, B \in \mathcal{F}$ and $A \subseteq B$, then

$$\mu(A) \leq \mu(B).$$

2. Rule of addition: For any $A, B \in \mathcal{F}$,

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B),$$

and in particular,

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

3. Subadditive: If $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$, then

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} \mu(B_n).$$

We should construct genuine measures.

Definition 2.1.12. The pair $(\Omega, \mathcal{A})$ is called a premeasurable space if $\mathcal{A}$ is a field over Ω.

A countably additive function $\mu : \mathcal{A} \to [0, \infty]$ is called a premeasure, that is, for any sequence $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ such that $\bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$ and the sets E_i are pairwise disjoint,

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

If a function $\chi : \mathcal{A} \to [0, \infty]$ satisfies finite additivity, that is,

$$\chi(A \cup B) = \chi(A) + \chi(B) \quad \text{for all disjoint } A, B \in \mathcal{A},$$

then χ is called a finitely additive measure.

Proposition 2.1.13. Let $(\Omega, \mathcal{A}, \chi)$ be a finitely additive measure space. Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ be a sequence of pairwise disjoint sets such that $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Then

$$\chi(A) \geq \sum_{i=1}^{\infty} \chi(A_i).$$

Proof. Fix $n \in \mathbb{N}$, and $\cup_{i=1}^{n} A_i \in \mathcal{A}$ or $\subseteq A$. Then we have

$$\chi\left(\cup_{i=1}^{n} A_i\right) \leq \chi(A).$$

where A_i are disjoint with each other, we have

$$\sum_{i=1}^{n} \chi(A_i) \leq \chi(A) \Rightarrow \lim_{n\to\infty} \sum_{i=1}^{n} \chi(A_i) \leq \chi(A).$$

□

Definition 2.1.14. A $\mathcal{S} \subseteq 2^{\Omega}$ is semi-algebra or elementary class if the following conditions hold:

1. $\emptyset \in \mathcal{S}$
2. $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$
3. If $A \in \mathcal{S}$ hten A^c is a finite disjoint union of elements from $\mathcal{S}$.

Proposition 2.1.15. Let Ω be a set and let $\mathcal{S} \subseteq 2^{\Omega}$ be a semi-algebra over Ω. Then the field $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$ is equal to the collection of all finite disjoint unions of sets from $\mathcal{S}$.

Proof. 1 □

Proposition 2.1.16. Let Ω be a set and let $\mathcal{S} \subseteq 2^\Omega$ be a semi-algebra over Ω. Let $\chi : \mathcal{S} \to [0, \infty]$ be a finitely additive set function, that is,

$$\chi(E \cup F) = \chi(E) + \chi(F) \quad \text{for all disjoint } E, F \in \mathcal{S}.$$

Then χ extends uniquely to a finitely additive measure on the field $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$. The extension is given by

$$\chi(A) := \sum_{i=1}^{n} \chi(E_i), \quad \text{whenever } A = \bigsqcup_{i=1}^{n} E_i$$

for pairwise disjoint sets $E_1, \ldots, E_n \in \mathcal{S}$.

Denote semi-algebra as $\mathcal{S}$, field as $\mathcal{A}$. Understanding:

1. Before constructing the countably additive measure, we need to construct the premeasure that is defined the function on the field
2. Firstly, we choose the interval $(a, b]$ or $\mathcal{B}_{(]}(\mathbb{R})$, which is also called semi-algebra.
3. Constructing the field $DU(\mathcal{S})$.
4. Semi-algebra + function $\chi_F((a, b]) = F(b) - F(a)$, and the finitely additive function extends to a unique finitely additive measure on the field $\mathcal{A}$. But we need continuity to make sure it is a premeasure.

2.1.1 3.1 Stieltjes-Premeasure

Theorem 2.1.17. Let $F : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function and let χ_F be the finitely additive set function on the field $\mathcal{B}_{(]}(\mathbb{R})$ defined by $\chi_F((a, b]) := F(b) - F(a), -\infty \leq a \leq b \leq \infty$, and extended to $\mathcal{B}_{(]}(\mathbb{R})$ by finite additivity.

Then χ_F is a premeasure on $\mathcal{B}_{(]}(\mathbb{R})$ if and only if F is right-continuous on $\mathbb{R}$, that is,

$$\lim_{\delta \downarrow 0} F(a + \delta) = F(a) \quad \text{for all } a \in \mathbb{R}.$$

Proposition 2.1.18. Let $\mathcal{S} \subseteq 2^\Omega$ be a semi-algebra. A finitely additive measure $\chi : \mathcal{A}(\mathcal{S}) \to [0, \infty]$ is a premeasure if and only if it is countably subadditive on $\mathcal{S}$, that is, whenever

$$E = \bigcup_{j=1}^{\infty} E_j \in \mathcal{S} \quad \text{with } E_j \in \mathcal{S},$$

one has

$$\chi(E) \leq \sum_{j=1}^{\infty} \chi(E_j).$$

2.2 Random Variables

Definition 2.2.1.

1. $(\Omega_1, \mathcal{A})$ and $(\Omega_2, \mathcal{B})$ are measurable spaces. $X : \Omega_1 \to \Omega_2$ is a measurable mapping if

$$X^{-1}(\mathcal{B}) = \{X \in B\} \in \mathcal{A}, \quad \forall B \in \mathcal{B}.$$

2. X is a **measurable function** if $(\Omega_2 . w) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ in (a).
3. X is a **Borel measurable function** if $\Omega_1 . \mathcal{A}) = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ and $(\Omega_2 . w) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Theorem 2.2.2. $X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ is measurable mapping if $\mathcal{B} = \sigma(\mathcal{C})$ and $X^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.

Theorem 2.2.3. If $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ and $f : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ are measurable mappings, then $f(X) = f \cdot X : (\Omega_1, \mathcal{A}_1) \to (\Omega_3, \mathcal{A}_3)$ is also measurable.

Definition 2.2.4. A random variable X is a measurable function from $(\Omega, \mathcal{A}) \to (\Omega, \mathcal{B}) \iff \{X \in B\} = X^{-1}(B) \in \mathcal{A}$ for all Borel sets $B \in \mathcal{B}$.

There is an other definition of random variable

Definition 2.2.5. A random variable X is a measurable mapping from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ such that $\mathbb{P}(|X| = \infty) = \mathbb{P}(\{\omega : |X(\omega)| = \infty\}) = 0$.

Theorem 2.2.6. X is a random variable from $(\Omega, \mathcal{A})$ to $(\Omega, \mathcal{B})$, (i.e. $x \in \mathcal{A}$) if and only if $\{X \leq x\} = X^{-1}([-\infty, x]) \in A$ for all $x \in \mathbb{R}$.

Proof. $\Rightarrow$: Suppose X is a random variable from $(\Omega, \mathcal{A})$ to $(\Omega, \mathcal{B})$. By definition 2.2.1, we know that $X^{-1}(B) \in \mathcal{A}$. Suppose $B = (-\infty, x]$ (because the Borel set is constructed by open intervals in $\mathbb{R}$). Then we have

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}([-\infty, x]) \in \mathcal{A}.$$

$\Leftarrow$: Suppose we have $\{X \leq x\} = X^{-1}([-\infty, x]) \in A$ for all $x \in \mathbb{R}$. Suppose $\mathcal{C} = \{[-\infty, x] : x \in \mathbb{R}\}$. We know that Borel algebra is generated by open intervals, then $\mathcal{B} = \sigma(\mathcal{C})$. We have

$$X^{-1}(C) = X^{-1}([-\infty, x]) \in \mathcal{A}.$$

Hence,

$$X^{-1}(\mathcal{B}) = X^{-1}(\sigma(\mathcal{C})) \subset \mathcal{A}.$$

By theorem 2.2.2, we know that X is a random variable from $(\Omega, \mathcal{A})$ to $(\Omega, \mathcal{B})$. □

2.3 Expecation and Integration

Theorem 2.3.1. Given a nonnegative r.v. X, there exists simple r.v.'s $0 \leq X_1 \leq X_2 \leq \dots$ such that $X_n(\omega) \uparrow X(\omega)$.

Proof. □

Theorem 2.3.2. $X \geq 0$, then $\mathbb{E}X = 0 \iff X = 0$ a.s..

2.4 $\mathcal{L}^1$ and $\mathcal{L}^p$

Definition 2.4.1. $\mathcal{L}^1$ convergence means that

$$\lim_{n \to \infty} \mathbb{E}[|X_n - X_{-\infty}|] = 0.$$

Chapter 3

Conditional Expectation

3.1 Jan 14

by Tatiana (Tanya) Brailovskaya

Definition 3.1.1 (Conditional expectation)**.** Suppose $X : \Omega \to \mathbb{R}$ is a random variable on the probability space $(\Omega, \mathcal{F}_0, \mathbb{P})$ and X is integrable ($\mathbb{E}[|X|] < \infty$). $\mathcal{F} \subseteq \mathcal{F}_0$ is a sub-σ-field. We call r.v. $Y \to \Omega \to \mathbb{R}$ the conditional expectation of X given $\mathcal{F}$ if is satisfies two conditions:

1. Y is $\mathcal{F}$-measurable (or $Y \in \mathcal{F}$)
2. $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ for any $A \in \mathcal{F}$ or $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$.

Intuition 3.1.2. $\mathbb{E}[X \mid \mathcal{F}]$ is a projection of X onto the space of $\mathcal{F}$-measurable r.v.'s.

Claim 3.1.3. $\mathbb{E}[X \mid \mathcal{F}]$ exists and is unique (i.e. if Y and Y' satisfy definition 3.1.1, then $Y = Y'$ a.s.; we call Y, Y' "versions" of $\mathbb{E}[X \mid \mathcal{F}]$).

Lemma 3.1.4. If Y is a version of $\mathbb{E}[X \mid \mathcal{F}]$, then $\mathbb{E}[|Y|] < \infty$ which means Y is integrable. Hint: show that $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|]$.

Proof. Let $A = \{Y > 0\} \in \mathcal{F}$. By definition 3.1.1, we use the second condition:

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \leq \int_A |X| d\mathbb{P}$$

$$\int_{A^c} -Y d\mathbb{P} = \int_{A^c} -X d\mathbb{P} \leq \int_{A^c} |X| d\mathbb{P}.$$

Hence, we have $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|]$. $\square$

Uniqueness of conditional expectation. Suppose Y' also satisfy the 3.1.1, then we have

$$\int_A Y d\mathbb{P} = \int_A Y' d\mathbb{P} \qquad \text{for all } A \in \mathcal{F}.$$

We consider the set $A = Y - Y' \geq \varepsilon > 0$. We have

$$0 = \int_A X - X d\mathbb{P} = \int_A Y - Y' d\mathbb{P} = \mathbb{E}[(Y - Y')\mathbb{1}_A] \geq \mathbb{E}[\varepsilon \mathbb{1}_A] = \varepsilon \mathbb{P}(A).$$

Hence $\mathbb{P}(A) = 0$, and $Y = Y'$ a.s. $\square$

Definition 3.1.5 (Absolutely continuous)**.** ν is absolutely continuous with respect to μ ($\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$.

Definition 3.1.6 (σ-finite measure)**.** μ is a σ-finite measure on $(\Omega, \mathcal{F})$ if there exists a partition of Ω into $A_1, A_2, \dots$ so that $\mu(A_i) < \infty$ for every i.

Theorem 3.1.7 (Radon-Nikodym)**.** Let μ and ν be σ-finite measure on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a function $f \in \mathcal{F}$ so that for all $A \in \mathcal{F}$

$$\int_A f d\mu = \nu(A), \qquad f = \frac{d\nu}{d\mu}.$$

Existence of conditional expectation. Suppose $X > 0$. Because we are looking for Y so that for any $A \in \mathcal{F}$

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}.$$

Define

$$\nu(A) = \int_A X d\mathbb{P}.$$

Now we should verify $\nu \ll \mathbb{P}$ and ν is σ-finite measure. $\nu(\emptyset) = \int_\emptyset X d\mathbb{P} = 0$, and it also satisfies countable addtive condition. Hence, ν is σ-finite measure.

ν is a σ-finite measure on $(\Omega, \mathcal{F})$ and $\nu \ll \mathbb{P}$. By theorem 3.1.7, we can find a $\mathcal{F}$-measurable function f s.t. for any $A \in \mathcal{F}$

$$\int_A f d\mathbb{P} = \nu(A) = \int_A X d\mathbb{P}.$$

□

Example 3.1.8. Suppose X, Z r.v.'s are joint probability density function (pdf) $f_{X,Z}(x,z)$ and $\mathbb{E}[|X|] < \infty$. Then, we define conditional pdf:

$$f_{X|Z}(x \mid z) = \begin{cases} \dfrac{f_{X,Z}(x,z)}{f_Z(z)}, & f_Z(z) \neq 0, \\ 0, & f_Z(z) = 0, \end{cases},$$

we define

$$Y(z) = \int_{\mathbb{R}} x f_{X|Z}(x \mid z) dx.$$

We show that $Y(z)$ is a version of $\mathbb{E}[X \mid \sigma(Z)]$.

For any $A \in \sigma(Z)$

$$\int_A Y(z) f_Z(z) dz = \int_A \left(\int_{\mathbb{R}} x f_{X|Z}(x \mid z) dx \right) f_Z(z) dz = \int_A \int_{\mathbb{R}} f_{X,Z}(x,z) dx dz.$$

Proof. First, Y is Borel measurable as a pointwise (a.e.) ratio of measurable functions integrated over x, and $Y(Z)$ is $\sigma(Z)$-measurable. Moreover, by Tonelli's theorem[a] and $\mathbb{E}[|X|] < \infty$,

$$\int_{\mathbb{R}} |Y(z)| f_Z(z)\, dz \le \int_{\mathbb{R}} \int_{\mathbb{R}} |x|\, f_{X|Z}(x \mid z)\, dx\, f_Z(z)\, dz = \int_{\mathbb{R}} \int_{\mathbb{R}} |x|\, f_{X,Z}(x,z)\, dx\, dz = \mathbb{E}[|X|] < \infty,$$

so $Y(Z)$ is integrable.

Let $A \in \sigma(Z)$. Then there exists a Borel set $B \subset \mathbb{R}$ such that $A = \{Z \in B\}$. Using Fubini's theorem[b] and the definition of $f_{X|Z}$, we compute

$$\mathbb{E}\big[Y(Z)\mathbf{1}_{\{Z \in B\}}\big] = \int_B Y(z) f_Z(z)\, dz = \int_B \left(\int_{\mathbb{R}} x\, f_{X|Z}(x \mid z)\, dx \right) f_Z(z)\, dz = \int_B \int_{\mathbb{R}} x\, f_{X,Z}(x,z)\, dx\, dz.$$

On the other hand,

$$\mathbb{E}\big[X \mathbf{1}_{\{Z \in B\}}\big] = \int_{\mathbb{R}} \int_{\mathbb{R}} x\, \mathbf{1}_{\{z \in B\}}\, f_{X,Z}(x,z)\, dx\, dz = \int_B \int_{\mathbb{R}} x\, f_{X,Z}(x,z)\, dx\, dz.$$

Therefore, for every $A \in \sigma(Z)$,

$$\mathbb{E}\big[Y(Z)\mathbb{1}_A\big] = \mathbb{E}\big[X\mathbb{1}_A\big].$$

By the defining property of conditional expectation, this implies that $Y(Z)$ is a version of $\mathbb{E}[X \mid Z]$. □

[a] I don't what is this theorem
[b] I should check this theorem

Proposition 3.1.9. X is a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Assume $\mathbb{E}|X| < \infty$, and $\mathcal{G}, \mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$.

1. If Y is any version of $\mathbb{E}[X \mid \mathcal{G}] \Rightarrow \mathbb{E}[Y] = \mathbb{E}[X]$.
2. If X is $\mathcal{G}$-measurable $\Rightarrow \mathbb{E}[X \mid \mathcal{G}] = X$ a.s.
3. Linearity: $\mathbb{E}[a_1X_1 + a_2X_2 \mid \mathcal{G}] = a_1\mathbb{E}[X_1 \mid \mathcal{G}] + a_2\mathbb{E}[X_2 \mid \mathcal{G}])$. Then $a_1X_1 + a_2X_2$ is a version of $\mathbb{E}[a_1X_1 + a_2X_2 \mid \mathcal{G}]$.
4. Positive: If $X \geq 0$, then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s.
5. Conditional MCT: If $0 \leq X_n \uparrow X$, then $\mathbb{E}[X_n \mid \mathcal{G}] \uparrow \mathbb{E}[X \mid \mathcal{G}]$ a.s.
6. Conditional Fatou's Lemma: If $X_n \geq 0$, then $\mathbb{E}[\liminf X_n \mid \mathcal{G}] \leq \liminf \mathbb{E}[X_n \mid \mathcal{G}]$ a.s.
7. Conditional DCT: If $|X_n(\omega)| \leq V(\Omega)$ for any n, $\mathbb{E}[V] < \infty$ and $X_n \to X$ a.s. then $\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}]$ a.s.
8. Conditional Jensen's: If $c : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[c(X)] < \infty$ then $\mathbb{E}[c(X) \mid \mathcal{G}] \geq c(\mathbb{E}[X \mid \mathcal{G}])$ a.s.
9. Tower property: $\mathcal{F}_1 \subseteq \mathcal{F}_2$

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2 \mid \mathcal{F}_1] = \mathbb{E}[X \mid \mathcal{F}_1], \qquad \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1 \mid \mathcal{F}_2] = \mathbb{E}[X \mid \mathcal{F}_1]$$

Question: Is it true for any $\mathcal{F}_1, \mathcal{F}_2$? Of course not (give an example), we can think $\Omega = \{a, b, c\}$ with

Chapter 4

Martingales

4.1 Prepare

We introduce some definiton of martingale. There is a definition of superharmonic. When we talk about superharmonic, we can start with a 1 dimension. The function is concave, then, the value of function at the specific centor point x will be larger than the average value of the function calculated over a ball surrounding that centor point.

4.2 Jan 21

by Tatiana (Tanya) Brailovskaya

Warm-up (Durrett, Exercise 4.1.5): For r.v. X and σ-algebras $\mathcal{F}_1, \mathcal{F}_2$ s.t. $\mathcal{F}_1 \subseteq \mathcal{F}_2$ then by Tower Property

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1] = \mathbb{E}[X \mid \mathcal{F}_1], \quad \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_2] = \mathbb{E}[X \mid \mathcal{F}_1].$$

Give an example on $\Omega = \{a, b, c\}$ on which the above equation is not true. Suppose $X(a) = 1, X(b) = 2, X(c) = 3$, $\mathcal{F}_1 = \{\{\emptyset\}, \{\Omega\}, \{a\}, \{b, c\}\}$, and $\mathcal{F}_2 = \{\{\emptyset\}, \{\Omega\}, \{b\}, \{a, c\}\}$. We need to compute the $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_2]$, and $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1]$. Assume $\mathbb{P}(a) = \mathbb{P}(b) = \mathbb{P}(c) = 1/3$. First we need to compute $\mathbb{E}[X \mid \mathcal{F}_1]$. By $\int_{\{a\}} X d\mathbb{P} = X(a)\mathbb{P}(a)$, and $\int_{\{a\}} Y d\mathbb{P} = Y(a)\mathbb{P}(a)$

$$\mathbb{E}[X \mid \mathcal{F}_1](a) = \frac{X(a)1/3}{1/3} = 1, \quad \mathbb{E}[X \mid \mathcal{F}_1](b) = \mathbb{E}[X \mid \mathcal{F}_1](c) = \frac{X(b)1/3 + X(c)1/3}{2/3} = 2.5.$$

Suupose $Y(a) = 1, Y(b) = Y(c) = 2.5$. Then we have

$$\mathbb{E}[Y \mid \mathcal{F}_2](b) = 2.5, \quad \mathbb{E}[Y \mid \mathcal{F}_2](a) = \mathbb{E}[Y \mid \mathcal{F}_2](c) = \frac{Y(a)1/3 + Y(c)1/3}{2/3} = 1.75.$$

Hence, we have $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_2] = (1.75, 2.5, 1.75)$. In the same way, we can compute $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1]$.

Martingales (also sub/super martingales) are from gambling definitions. Examples

Definition 4.2.1. We call a collection of σ-algebras: $\mathcal{F} = \{\mathcal{F}_n\}_{n=1}^{\infty}$ a filtration if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. A sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is adapted to $\mathcal{F}$ if for every n, X_n is $\mathcal{F}_n$ measurable. If X_n further satisfies

1. $\mathbb{E}|X_n| < \infty$
2. $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$

then X_n is a martingale and if instead of (2)

1. $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] \le X_{n-1} \Rightarrow \{X_n\}$ is supermartingale.

2. $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] \ge X_{n-1} \Rightarrow \{X_n\}$ is subermartingale.

Example 4.2.2 (Random Walk)**.** Let $Y_1, \ldots, Y_n$ be iid r.v's. $S_n = S_0 + \sum_{i=1}^n Y_i$ and S_0 is a constant. $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. $\{\mathcal{F}_n\}_{n=0}^\infty$ is a filtration, and S_n is adapted to it.

Assume $\mathbb{E}|Y_1| < \infty \Rightarrow \mathbb{E}|S_n| < \infty$, we have

$$\mathbb{E}[S_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[S_0 + \sum_{i=1}^n Y_i \mid \mathcal{F}_{n-1}] = \underbrace{S_0 + \sum_{i=1}^{n-1} Y_i}_{S_{n-1}} + \underbrace{\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}]}_{\mathbb{E}[Y_n]}.$$

If $\mathbb{E}[Y_n] = 0$, then S_n is a martingale. If $\mathbb{E}[Y_n] \le 0$, then S_n is a supermartingale. If $\mathbb{E}[Y_n] \ge 0$, then S_n is a subermartingale.

Example 4.2.3. Let $Y_1, \ldots, Y_n$ be iid r.v's. $M_n = \prod_{i=1}^n Y_i$ and S_0 is a constant. $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. $\{\mathcal{F}_n\}_{n=0}^\infty$ is a filtration, and M_n is adapted to it.

Assume $\mathbb{E}|Y_1|^2 < \infty \Rightarrow \mathbb{E}|M_n| < \infty$, we have

$$\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[\prod_{i=1}^n Y_i \mid \mathcal{F}_{n-1}] = \prod_{i=1}^{n-1} Y_i \mathbb{E}[Y_n].$$

If $\mathbb{E}[Y_n] = 1$, then M_n is a martingale. If $\mathbb{E}[Y_n] \le 1$, then M_n is a supermartingale. If $\mathbb{E}[Y_n] \ge 1$, then M_n is a subermartingale.

Claim 4.2.4. If X_n is a supermartingale, then for all $n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] \le X_m$.

Proof. Because X_n is a supermartingale, we have $\mathbb{E}[X_{m+1} \mid \mathcal{F}_m] \le X_m$. We just need to let $n = m + k, k \ge 2$.

$$\mathbb{E}[X_{m+k} \mid \mathcal{F}_m] = \mathbb{E}[\mathbb{E}[X_{m+k} \mid \mathcal{F}_{m+k-1}] \mid \mathcal{F}_m] \le \mathbb{E}[X_{m+k-1} \mid \mathcal{F}_m].$$

$\square$

Corollary 4.2.5. 1. if X_n is submartingale, then for every $n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] \ge X_m$

2. if X_n is martingale, then for every $n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m$

Definition 4.2.6 (predictable sequence)**.** For a fltration $\{\mathcal{F}_n\}_{n=0}^\infty$, we call a sequence of r.v. $\{H_n\}_{n=1}^\infty$ a predictable sequence if H_n is measurable w.r.t. $\mathcal{F}_{n-1}$.

Is a predictable sequence adapted? H_n is $\mathcal{F}_{n-1}$ measurable $\Rightarrow H_n$ is $\mathcal{F}_n$ measurable. For any random variable that is measurable w.r.t. $\mathcal{F}_{n-1}$ is automatically measurable w.r.t. to the larger σ-algebra $\mathcal{F}_n$.

Example 4.2.7 (Example)**.** At each round, you bet a stake X \$ then w.p. p you win bet X \$ and w.p. $1 - p$ you lose stake X \$.

1. X_n: net winnings after n rounds if you bet 1 \$ each time

2. H_n: amount of money your stake in round $n + 1$

3. $(H \cdot X)_n = \sum_{m=1}^n H_n(X_m - X_{m-1})$

4. if $(H \cdot X)_n$ is a submartingale $\Rightarrow \mathbb{E}[H \cdot X]_n \ge 0$.

Theorem 4.2.8. Let X_n, $n \ge 0$ be a supermartingale. If $H_n \ge 0$ is predictable and each H_n is bounded. Then $(H \cdot X)_n$ is a supermartingale.

Proof.

$$\begin{aligned}\mathbb{E}[(H \cdot X)_{n+1} \mid \mathcal{F}_n] &= (H \cdot X)_n + \mathbb{E}[H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1}\mathbb{E}[(X_{n+1} - X_n) \mid \mathcal{F}_n] \\ &\leq (H \cdot X)_n\end{aligned}$$

□

Definition 4.2.9. We call a r.v. N with values in $\{0, 1, 2, \dots\}$ stopping time if $\{N = n\} \in \mathcal{F}_n$.

Theorem 4.2.10. If N is a stopping time and X_n is a supermartingale, then $X_{N \wedge n}$ is a supermartingale.

Proof.

$$X_{N \wedge n} = \sum_{m=1}^{n} H_m(X_m - X_{m-1}) + X_0,$$

and $H_m = \mathbb{1}_{N \geq m}$.

We want to show that $\mathbb{E}[X_{N\wedge n} \mid \mathcal{F}_{n-1}] \leq X_{N\wedge n-1}$. We have

$$X_{N\wedge n} = X_{N\wedge n-1} + H_n(X_n - X_{n-1}).$$

Then we

$$\mathbb{E}[X_{N\wedge n} \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_{N\wedge n-1} + H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] = X_{N\wedge n-1} + H_n\mathbb{E}[(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}].$$

Since N is stopping time and X_n is a supermartingale, we have $H_n\mathbb{E}[(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] \leq 0$. Thus, we have

$$\mathbb{E}[X_{N\wedge n} \mid \mathcal{F}_{n-1}] \leq X_{N\wedge n-1}.$$

□

4.3 Jan 26

by Tatiana (Tanya) Brailovskaya

Warm-up (Durrett, Exercise 4.2.2): Give an example of a sub-MG X_n s.t. X_n^2 is a super-MG. Hint: X_n doesn't have to be random.

Maybe we can let $X_n = -\frac{1}{n}$, then we have $X_n^2 = \frac{1}{n^2}$. (Assume $\mathcal{F}_n = \{\emptyset, \Omega\}$).

Theorem 4.3.1 (MG Convergence Theorem)**.** If X_n is a sub-MG s.t. $\sup_n \mathbb{E}X_n^+ < \infty$, then $X_n \to X$ a.s. for some X with $\mathbb{E}|X| < \infty$.

Before proving 4.3.1, we introduce upcrossing inequality.

Recall: Suppose X_n, $n \geq 0$ is a sub-MG. Let $a < b$ and $N_0 = -1$. For $k \geq 1$, let

$$N_{2k-1} = \inf\{m > N_{2k-2} : X_m \leq a\}, \quad N_{2k} = \inf\{m > N_{2k-1} : X_m \geq b\}$$

Theorem 4.3.2 (Upcrossing Inequality)**.** If $X_m, m \geq 0$ is a sub-MG, then

$$(b - a)\mathbb{E}u_n \leq \mathbb{E}[X_n - a]^+ - \mathbb{E}[X_0 - a]^+$$

Proof. Let $Y_m = a + (X_m - a)^+$ is a sub-MG, and Y_m makes the same number of upcrossings of $[a, b]$ on X_m. Then we have

$$H_m = \begin{cases} 1, & \text{if } N_{2k-1} < m \leq N_{2k} \text{ for some } k, \\ 0, & \text{otherwise.} \end{cases},$$

and H_m is a predictable sequence. Then we have (total profit is more expensive than upcroosing)

$$(b - a)u_n \leq (H \cdot Y)_n.$$

Let $K_m = 1 - H_m$, we have

$$Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n = \sum_{i=1}^{n} (Y_i - Y_{i-1}).$$

Because $(K \cdot Y)_n$ is a sub-MG, then $\mathbb{E}[K \cdot Y]_n \geq \mathbb{E}[K \cdot Y]_0 = 0$, and $\mathbb{E}[H \cdot Y]_n \leq \mathbb{E}[Y_n - Y_0]$. Combining the equations, we have

$$(b-a)\mathbb{E}u_n \leq \mathbb{E}Y_n - \mathbb{E}Y_0.$$

□

Now we can prove 4.3.1.

Proof. Let $(X-a)^+ \leq X^+ + |a|$. By upcrossing inequality 4.3.2, we have

$$\mathbb{E}u_n \leq \frac{|a| + \mathbb{E}X_n^+}{b-a}.$$

Then $\sup_n \mathbb{E}u_n < \infty$. Because u_n is a monotonically increasing sequence, then $u_n \to u$ a.s. And $\mathbb{E}u < \infty$, $u < \infty$ a.s. Now we consider the probability

$$\mathbb{P}\left(\cup_{a,b\in\mathbb{Q}}\{\liminf X_n < a < b < \limsup X_n\}\right) = 0.^{a}$$

Then we have $\liminf X_n = \limsup X_n$ i.e. $X = \lim X_n$ a.s.

Now we shoule show $\mathbb{E}|X| < \infty$. We write

$$X_n = X_n^+ - X_n^-, X = X^+ - X^-.$$

Suffices to check that $\mathbb{E}X^+ < \infty$ and $\mathbb{E}X^- < \infty$. By Fatou's Lemma, we have

$$\mathbb{E}[\liminf X_n^+] \leq \liminf \mathbb{E}[X_N^+] < \infty.$$

In the same way, we have

$$\mathbb{E}[\liminf X_n^-] \leq \liminf \mathbb{E}[X_N^-] < \infty.$$

And

$$\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \leq \mathbb{E}X_n^+ - \mathbb{E}X_0,$$

since X_n is a sub-MG, then the inequality holds. Then as $n \to \infty$, $X_n \to X$ and $\mathbb{E}X \leq \mathbb{E}X_0$. Thus, we have

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2M - \mathbb{E}X_0 < \infty,$$

where $M = \sup\{\mathbb{E}X_N^+ : n \geq 0\}$. □

[a] We should think about $\mathbb{P}(u_\infty = \infty) = 0$ which means the probability of infinity number of upcrossing is zero. $\liminf X_n < a$ means that the sequence will always occasionally fall below a.

4.3.1 Application of Martingale Convergence Theorem 4.3.1

An urn contains r red and g green balls. At each step draw a ball and replace it with c balls of the same color. Let X_n be fraction of green balls after nth draw.

Lemma 4.3.3. $\{X_n\}_{n=1}^{\infty}$ is MG.

Proof. Suppose i red balls, and j green balls at time n. Then we have

$$X_{n+1} = \begin{cases} \frac{j+c}{i+j+c}, & \text{w.p. } \frac{j}{i+j} \\ \frac{j}{i+j+c}, & \text{w.p. } \frac{i}{i+j} \end{cases}.$$

Define

$$\mathbb{1}_G = \begin{cases} 1, & \text{if } n+1\text{th ball is green} \\ 0, & \text{otherwise} \end{cases}.$$

$$\begin{aligned}\mathbb{E}[X_{n+1} \mid X_n] &= \mathbb{E}[\mathbb{E}[X_{n+1} \mid \sigma(\mathbb{1}_G, X_n)] \mid X_n] \\ &= \mathbb{E}[\frac{j+c}{i+j+c}\mathbb{1}_G + \frac{j}{i+j+c}(1-\mathbb{1}_G) \mid X_n] \\ &= \frac{j+c}{i+j+c}\frac{j}{i+j} + \frac{j}{i+j+c}\frac{i}{i+j} \\ &= \frac{j}{i+j} \\ &= X_n.\end{aligned}$$

By Corollary, $X_n \to X$ a.s. □

We can also derive the distribution of X_n:

$\{X_n\}_{n=0}^{\infty}$ is a MG s.t. $X_n \to X$ a.s. Does it follow that $X_n \to X$ in L^p? **Not necessarily!**

Example 4.3.4. Let $S_n = 1 + \sum_{n=1}^{\infty} Y_n, Y_n = \pm 1$ w.p. $1/2$. $N = \inf\{n : S_n = 0\}$. S_n is MG, N is a stopping since $\Rightarrow X_n = S_{N\wedge n}$ is a MG. By Corollary, $X_n \to X a.s.$

Lemma 4.3.5. $X = 0$ a.s.

If $X = K > 0$

4.4 Jan 28

by Tatiana (Tanya) Brailovskaya

Warm up: last time we proved X_n-sub-MG, N is stopping time, s.t. $\mathbb{P}(N \le k) = 1$, then $\mathbb{E}X_0 \le \mathbb{E}X_N \le \mathbb{E}X_k$. M - another stopping time s.t. $M \le N$ then

$$\mathbb{E}X_0 \le \mathbb{E}X_M \le \mathbb{E}X_N \le \mathbb{E}X_k.$$

Hint: modify the pf of $\mathbb{E}X_0 \le \mathbb{E}X_N \le \mathbb{E}X_k$

Proof. Consider $K_n = 1_{\{M<n\}}$

$$(K \cdot X)_n = X_n - X_{M\wedge n} \implies \mathbb{E}(K \cdot X)_N = \mathbb{E}X_N - \mathbb{E}X_M$$

$$0 = \mathbb{E}(K \cdot X)_0 \le \mathbb{E}(K \cdot X)_N \text{ because } (K \cdot X)_n \text{ is a sub-MG \& } \mathbb{P}(N \le k) = 1$$

□

Theorem 4.4.1 (Doob's Inequality)**.** X_m-sub-MG, $\widehat{X}_n = \max_{m\le n} X_m^+$, $\lambda > 0$, $A = \{\widehat{X}_n \ge \lambda\}$.

$$\lambda\mathbb{P}(A) \le \mathbb{E}[X_n \mathbb{1}_A] \le \mathbb{E}X_n^+.$$

Proof. Let $N = \inf\{m : X_m \ge \lambda \text{ or } m = n\}$

$$\lambda\mathbb{P}(A) \le \mathbb{E}X_N\mathbb{1}_A \le \mathbb{E}X_n\mathbb{1}_A.$$

Explanatory Notes from Image [a]: $\mathbb{E}X_N \le \mathbb{E}X_n$, $\mathbb{E}[X_N\mathbb{1}_A + X_N\mathbb{1}_{A^c}] = \mathbb{E}[X_N\mathbb{1}_A + X_n\mathbb{1}_{A^c}]$ □

[a] If the event A happened, we have $N \le n$. If A^c, $N = n$. Thus, $N \le n$. By previous theorem, for sub-MG X_n and N is a stopping time, we have $\mathbb{E}X_N \le \mathbb{E}X_n$.

$$\mathbb{P}(\max_{m\le n} X_m \ge \lambda) \le \frac{\mathbb{E}X_n^+}{\lambda}$$

Compare Thm 1 with Markov Inequality ($X_n \ge 0$),

$$\mathbb{P}(\widehat{X}_n \ge \lambda) \le \frac{\mathbb{E}[\widehat{X}_n]}{\lambda} \quad (MI), \qquad \text{vs} \qquad \mathbb{P}(\widehat{X}_n \ge \lambda) \le \frac{\mathbb{E}X_n^+}{\lambda} \quad (DI).$$

Example 4.4.2. $S_n = 1 + \sum_{i=1}^n Y_i$ where $Y_i = \pm 1$ w.p. 0.5. $N = \inf\{n : S_n = 0\}, X_n = S_{N\wedge n}$ is a MG too $\implies$ $\mathbb{E}X_n = \mathbb{E}X_0 = 1$. We can show: $\mathbb{E}\left[\max_{0\le m\le n} X_n\right] \sim \log n \implies$ MI is trivial, while DI $\mathbb{P}(\widehat{X}_n \ge \lambda) \le 1/\lambda$. Takeaway: $\mathbb{E}X_n^+$ & $\mathbb{E}\widehat{X}_n$ are not always comparable

Theorem 4.4.3 (L^p maximum inequality)**.** 1. X_n-sub-MG then for $1 < p < \infty$

$$\mathbb{E}|\widehat{X}_n|^p \le \left(\frac{p}{p-1}\right)^p \cdot \mathbb{E}(X_n^+)^p.$$

2. If Y_n is a MG and $Y^* = \max_{0\le m\le n}|Y_m|$ then

$$\mathbb{E}(Y_n^*)^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|Y_n|^p.$$

Proof. We want the arg to work regardless of $\mathbb{E}\widehat{X}_n^p < \infty$ or ∞

$$\mathbb{E}|\widehat{X}_n \wedge M|^p = \int_0^\infty p\lambda^{p-1}\mathbb{P}(\widehat{X}_n \wedge M \ge \lambda)d\lambda$$

Read the book page 205, Durrett.[a] □

[a] We should know some Lemma 2.2.13, Fubini's theorem, Theorem 1.6.3.
Lemma 2.2.13: If $y \ge 0, p > 1$, then we have $\mathbb{E}[Y^p] = \int py^{p-1}\mathbb{P}(Y > y)d\lambda$.

Theorem 4.4.4. If X_n is a MG with $\sup_n \mathbb{E}|X_n|^p < \infty$ for $p > 0$ then $X_n \to X$ a.s. & in L^p.

Proof. Why $X_n \to X$ a.s.? (Recall: we need $\sup_n \mathbb{E}X_n^+ < \infty$)

$$\sup_n \mathbb{E}X_n^+ \le \sup_n \mathbb{E}|X_n| \le (\sup_n \mathbb{E}|X_n|^p)^{1/p} < \infty.$$

Why does L^p convergence follow? Use DCT:

$$|X_n - X|^p \le (|X_n| + |X|)^p \le (2\max(|X_n|, |X|))^p \le 2^p(|X_n|^p + |X|^p)$$

$$\sup_n |X_n - X|^p \le 2^p(\sup_n |X_n|^p + |X|^p)$$

$$\mathbb{E}\max_{m\le n}|X_m|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^+)^p$$

$$\downarrow \lim n \to \infty$$

$$\mathbb{E}\sup_n |X_n|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|X|^p < \infty.$$

□

Example 4.4.5. Z_n: # of individuals in gen n, with $Z_0 = 1$. Y_i^n: # of offspring ith individual in nth gen produces, Y_i^n are iid across i & n.

$$Z_{n+1} = Y_1^n + Y_2^n + \cdots + Y_{Z_n}^n, \quad \mu = \mathbb{E}Y_1^n, \quad \sigma^2 = \mathrm{Var}(Y_1^n)$$

Assume $\mu > 1$. $X_n = Z_n/\mu^n$. Check: X_n is MG

$$\mathbb{E}X_n^+ = \mathbb{E}X_n = \mathbb{E}X_0 = 1 \implies X_n \to X \text{ a.s.}$$

Claim: $X_n \to X$ in L^2.

Proof. $\mathbb{E}X_n^2 = \mathbb{E}[\mathbb{E}[X_n^2|X_{n-1}]]$.

$$\mathbb{E}[X_n^2|X_{n-1}] = \mathbb{E}[(X_n - X_{n-1})^2 + 2X_nX_{n-1} - X_{n-1}^2|X_{n-1}] = X_{n-1}^2 + \mathbb{E}[(X_n - X_{n-1})^2|X_{n-1}].$$

$$\mathbb{E}[(X_n - X_{n-1})^2|X_{n-1}] = \mathbb{E}\left[\left(\frac{Z_n}{\mu^n} - \frac{Z_{n-1}}{\mu^{n-1}}\right)^2 \middle| X_{n-1}\right] = \mu^{-2n}\mathbb{E}[(Z_n - \mu Z_{n-1})^2|X_{n-1}]$$

on $\{Z_{n-1} = k\}$:

$$\mathbb{E}[(Z_n - \mu Z_{n-1})^2|X_{n-1}]\mathbb{1}_{\{Z_{n-1}=k\}} = \mathbb{E}\left(\sum_{i=1}^k Y_i^n - \mu k\right)^2 = k\sigma^2$$

$$\implies \mathbb{E}[(Z_n - \mu Z_{n-1})^2|X_{n-1}] = Z_{n-1}\sigma^2$$

$$\implies \mathbb{E}X_n^2 = \mathbb{E}X_{n-1}^2 + \mathbb{E}(Z_{n-1}\sigma^2/\mu^{2n}) \xrightarrow{\text{by induction}} \sup_n \mathbb{E}X_n^2 = \left[1 + \sigma^2\sum_{k=2}^{n+1}\mu^{-k}\right] < \infty.$$

□

Moreover, $\mathbb{E}X^2 = \lim_{n\to\infty}\mathbb{E}X_n^2 = 1 + \frac{\sigma^2}{\mu^2}\overbrace{\left(\frac{1}{1-1/\mu}\right)}^{\sup_n} > 0$. And, $X_n \to X$ in $L^1 \implies \mathbb{E}X = \lim_{n\to\infty}\mathbb{E}X_n = 1$. We can check: $\{X > 0\} = \{Z_n > 0 \text{ for } \forall n\}$

$$\implies \text{when } \mu > 1, \mathbb{P}(\text{extinction}) \le 1 - \mathbb{P}(\{X > 0\}) < 1 \quad (\text{in fact, } \mathbb{P}(\text{extinction}) < 1).$$

4.5 Feb 2

by Tatiana (Tanya) Brailovskaya

Warm up:

Lemma 4.5.1. Let $S_n = Y_1 + \cdots + Y_n$, $Y_1, Y_2, \ldots$ is iid, $\mathbb{E}Y_1 = 0, \mathrm{Var}(Y_1) = \sigma^2 < \infty$. Show that

$$\mathbb{P}(\max_{1\le m\le n}|S_m| \ge x) \le \frac{\mathrm{Var}(S_n)}{x^2}$$

Proof. Because $f(x) = x^2$ is convex, we can apply Jensen's inequaliy $\mathbb{E}[S_n^2 \mid \mathcal{F}_n] \ge (\mathbb{E}[S_{n+1} \mid \mathcal{F}_n])^2$. Since S_n is a martingale, then $\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n$. Then we know that S_n^2 is a sub-MG. By Doob's inequality 4.4.1, we have

$$\begin{aligned}\mathbb{P}(\max_{1\le m\le n}|S_m| \ge x) &= \mathbb{P}(\max_{1\le m\le n}|S_m|^2 \ge x^2)\\ &= \frac{\mathbb{E}S_n^2}{x^2}\\ &= \frac{\mathrm{Var}(S_n)}{x^2}.\end{aligned}$$

□

Remark 4.5.2. $\mathrm{Var}(S_n) = n\sigma^2 \Rightarrow$ for $x = C\sigma^2\sqrt{n}$

$$\mathbb{P}(\max_{1\le m\le n}|S_m| \ge C\sigma\sqrt{n}) \le \frac{1}{C^2}.$$

i.e.

$$\max_{1\le m\le n}|S_m| \ge C\sigma\sqrt{n} \quad \text{w.r.} \quad 1 - \frac{1}{C^2}.$$

Can also show that: $\max_{1\le m\le n}|S_m| \ge C\sigma\sqrt{n}$.

Now we should talk about the condition of L^p, $1 < p < \infty$.

1. $\sup_n \mathbb{E}|X_n|^p < \infty$ is sufficient for MG convergence in L^p, $1 < p < \infty$.
2. $\sup_n \mathbb{E}|X_n| < \infty$ is not going to be sufficient for L^1 convergence.
3. You can always find an example satisfying $\mathbb{E}|X_n| < \infty$, but the limit does not equal the same number.

Definition 4.5.3. A collection of r.v. X_i, $i \in I$ is uniformly integrable (U.I.) if

$$\lim_{M\to\infty}\left(\sup_{i\in I}\mathbb{E}[|X_i|; |X_i| > M]\right) = 0.$$

We should clarify that $\mathbb{E}[A; B]$ means $\mathbb{E}[A \cdot \mathbb{1}_B]$

Theorem 4.5.4. Suppose that $\mathbb{E}|X_n| < \infty$ for all n. If $X_n \to X$ in probability, TFAE (then the following are equivalent): For a sub-MG TFAE:

1. It is U.I.
2. It converges in a.s. and L^1: $X_n \to X$ in L^1
3. It converges in L^1: $\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$

Proof. $1 \Rightarrow 2$: Let

$$\varphi_M(x) = \begin{cases} M, & x \geq M \\ x, & |x| < M \\ -M, & x \leq -M \end{cases}.$$

Note that $\varphi_M(x_n) \to \varphi_M(x)$ in probability, and

$$|X_n - X| \leq |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X)| + |\varphi_M(X) - X|.$$

We can also note that

$$|Y - \varphi_M(Y)| = (|Y| - M)^+ \leq |Y|\mathbb{1}(|Y| > M).$$

Then we have $|\varphi_M(X_n) - \varphi_M(X)| \to 0$, $|X_n - \varphi_M(X_n)| \to 0$ by U.I, and $|\varphi_M(X) - X| \to 0$ since $\mathbb{E}|X| < \infty$ (Fatou's Lemma - Durrett Exercise 2.3.4)[a].

$2 \Rightarrow 3$:

$$|\mathbb{E}|X_n| - \mathbb{E}|X|| \leq \mathbb{E}|X_n - X|.$$

$3 \Rightarrow 1$: Let

$$\psi_M(x) = \begin{cases} x, & x \in [0, M-1], \\ 0, & x \in [M, \infty), \\ \text{linear}, & x \in [M-1, M]. \end{cases}$$

Then we have $x\mathbb{1}(x > M) \leq x - \psi_M(x)$. By DCT as $M \to \infty$

$$\mathbb{E}|X| - \mathbb{E}\psi_M(|X|) \to 0$$

and $\mathbb{E}\psi_M(|X_n|) \to \mathbb{E}\psi_M(|X|)$

For n large enough, we have

$$\begin{aligned}\mathbb{E}\big[|X_n|; |X_n| > M\big] &\leq \mathbb{E}|X_n| - \mathbb{E}\psi_M(|X_n|) \\ &\leq \mathbb{E}|x| - \mathbb{E}\psi_M(|x|) + \frac{\varepsilon}{2}\end{aligned}$$

□

[a] When the proof says $|\varphi_M(X_n) - \varphi_M(X)| \to 0$ and so on, it is actually referring to the expectations of these terms. Because we want to prove $X_n \to X$ in L^1 (which means $\mathbb{E}|X_n - X| \to 0$), we take the expected value of both sides of your triangle inequality:

$$\mathbb{E}|X_n - X| \leq \mathbb{E}|X_n - \varphi_M(X_n)| + \mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| + \mathbb{E}|\varphi_M(X) - X|$$

The goal is to show that we can make the total sum arbitrarily small by first choosing a large cutoff M, and then letting $n \to \infty$. Here are the exact details for how each of those three expected values goes to zero: 1. The middle term: $\mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| \to 0$ as $n \to \infty$. We are given that $X_n \to X$ in probability. Because the truncation function $\varphi_M(x)$ is continuous, the Continuous Mapping Theorem dictates that $\varphi_M(X_n) \to \varphi_M(X)$ in probability. Notice that the output of φ_M is bounded by M and $-M$. This means the difference $|\varphi_M(X_n) - \varphi_M(X)|$ can never be larger than $2M$. Because the random variables are strictly bounded by a constant, we can apply the **Bounded Convergence Theorem** (a special case of the Dominated Convergence Theorem) to conclude that convergence in probability implies convergence in L^1. Therefore, for any fixed M, $\lim_{n\to\infty} \mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| = 0$

2. The first term: $\mathbb{E}|X_n - \varphi_M(X_n)| \to 0$ by U.I. as $M \to \infty$. Why it works: Using the inequality you noted, we have $\mathbb{E}|X_n - \varphi_M(X_n)| \leq \mathbb{E}[|X_n|\mathbb{1}(|X_n| > M)]$. This expression is exactly the "tail expectation" used in the definition of Uniform Integrability (U.I.). By the definition of U.I., $\lim_{M\to\infty} (\sup_n \mathbb{E}[|X_n|; |X_n| > M]) = 0$. This means we can pick an M large enough to make this term arbitrarily small simultaneously for all n.

3. The last term: $\mathbb{E}|\varphi_M(X) - X| \to 0$ by Fatou's Lemma as $M \to \infty$. Similar to the first term, $\mathbb{E}|\varphi_M(X) - X| \leq \mathbb{E}[|X|\mathbb{1}(|X| > M)]$. For any single random variable X, if its mean is finite ($\mathbb{E}|X| < \infty$), the expected value of its extreme tails must shrink to 0 as the cutoff M goes to infinity.

$\mathbb{E}|X| < \infty$: This is where Fatou's Lemma comes in. Uniform integrability guarantees that the expected values of X_n are uniformly bounded by some constant K (i.e., $\sup_n \mathbb{E}|X_n| < K < \infty$). Since $X_n \to X$ in probability, Fatou's Lemma: $\mathbb{E}|X| \leq \liminf_{n\to\infty} \mathbb{E}|X_n| \leq K < \infty$. Because X is integrable, we can make this third term arbitrarily small by choosing M large enough.

Lemma 4.5.5. Let $\{X_i\}_{i\in I}$ be a family of random variables. If $\{X_i\}_{i\in I}$ is uniformly integrable (U.I.), then the family is bounded in L^1, i.e. $\sup_{i\in I} \mathbb{E}|X_i| < \infty$

Proof. By the definition of U.I. , for M sufficiently large, we have:

$$\sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}\{|X_i| > M\}] \leq 1$$

We can decompose the expectation as:

$$\sup_{i\in I} \mathbb{E}[|X_i|] = \sup_{i\in I} (\mathbb{E}[|X_i|\mathbb{1}\{|X_i| \leq M\}] + \mathbb{E}[|X_i|\mathbb{1}\{|X_i| > M\}])$$

The first term is bounded by M (since $|X_i| \leq M$ on the set $\{|X_i| \leq M\}$), and the second term is bounded by 1 based on the uniform integrability condition. Thus, we obtain:

$$\sup_{i\in I} \mathbb{E}[|X_i|] \leq M + 1 < \infty$$

□

Question 4.5.6. Does $\sup_{i\in I} \mathbb{E}|X_i| < \infty \Rightarrow X_i, i \in I$ U.i.?

Proof. NO!. Counterexample: There exists a sequence of random variables $(X_n)_{n\geq 1}$ such that

$$\sup_{n\geq 1} \mathbb{E}\left[|X_i|\right] < \infty,$$

but $(X_n)_{n\geq 1}$ is not uniformly integrable. In particular, let X_n be defined by

$$X_n = \begin{cases} n, & \text{with probability } \frac{1}{n}, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{E}[|X_i|] = n \cdot \frac{1}{n} = 1 \quad \text{for all } n \geq 1, \qquad \text{hence } \sup_{n\geq 1} \mathbb{E}[|X_i|] = 1.$$

However, for any $M > 0$,

$$\mathbb{E}\left[X_n \mathbb{1}_{\{X_n > M\}}\right] = 0 \quad \text{if } n < M, \qquad \mathbb{E}\left[X_n \mathbb{1}_{\{X_n > M\}}\right] = 1 \quad \text{if } n \geq M,$$

and therefore

$$\sup_{n\geq 1} \mathbb{E}\left[X_n \mathbb{1}_{\{X_n > M\}}\right] = 1 \quad \text{for all } M > 0,$$

so $(X_n)_{n\geq 1}$ is not uniformly integrable. □

Theorem 4.5.7. Given a probabiliy space $(\Omega, \mathcal{F}_0, \mathbb{P})$ and r.v. $X \in L^1$[1]. We have $\{\mathbb{E}[X \mid \mathcal{F}] : \mathcal{F} \subseteq \mathcal{F}_0 \text{ is a } \sigma\text{-algebra}\}$ is U.I.

Proof. Proof see the Durret 4.6.1. □

Theorem 4.5.8. For a sub-MG TFAE (then the following are equivalent):

1. It is U.I.
2. It converges in a.s. and L^1.
3. It converges in L^1: $\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$

Proof. (i) $\Rightarrow$ (ii): U.I. $\Rightarrow \sup_n \mathbb{E}|X_n| < \infty \Rightarrow$ a.s. convergence $\Rightarrow$ in probability convergence $\Rightarrow$ Thm 3 implies the desired conclusion (ii) $\Rightarrow$ (iii): we have already proves before. (iii) $\Rightarrow$ (i) convergence in $L^1 \Rightarrow$ convergence in probability $\Rightarrow$ Thm 3 implies desired conclusion. □

Preparation for MG version of theorem 4.5.8.

Lemma 4.5.9. If $\mathbb{E}|X_n| < \infty$, $\mathbb{E}|X| < \infty$ & $X_n \to X$ in L^1, then $\mathbb{E}[X_n; A] \to \mathbb{E}[X; A]$.

Proof. $|\mathbb{E}[X_n \mathbb{1}_A - X \mathbb{1}_A]| \leq \mathbb{E}|X_n - X| \mathbb{1}_A \leq \mathbb{E}|X_n - X| \to 0$ □

Lemma 4.5.10. If a MG $X_n \to X$ in L^1 then $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$.

Proof. By MG prop. for $m \geq n$ $\mathbb{E}[X_m \mid \mathcal{F}_n] = X_n \Rightarrow$ if $A \in \mathcal{F}_n$ $\mathbb{E}[X_n; A] = \mathbb{E}[X_m; A]$.
By Lemma 1, $\mathbb{E}[X_m; A] \to \mathbb{E}[X; A]$ for $\forall A \in \mathcal{F}_n$. □

Theorem 4.5.11. For a MG TFAE

1. It is U.I.
2. It converges in a.s. & and L^1.
3. It converges in L^1.
4. $\exists$ a r.v. X w $\mathbb{E}|X| < \infty$ s.t. $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$.

Proof. (i)$\to$(ii)$\to$(iii) by Thm 4.5.8
(iii)$\to$(iv) by Lemma 2
(iv)$\to$(i) Example 2 of u.i. families (Durrett theorem 4.6.1). □

[1]The notation $X \in L^1$ simply means that X is an integrable random variable, which mathematically means that its expected absolute value is finite: $\mathbb{E}|X| < \infty$

4.6 Feb 4

by Tatiana (Tanya) Brailovskaya

Warm-up: Consider the following random graph: on the gird $\mathbb{Z}^d$, for every edge, independently delete that edge w.p. $1-p$ or keep it otherwise (this is called bond perolation). Let

$$A = \{\text{there exists an infty component.}\}.$$

Show that $\mathbb{P}(A) = 0$ or 1.

Proof. Kolmogorov's 0-1 Law. □

Remark 4.6.1. Moreover there exists p' s.t. for all $p < p'$, $\mathbb{P}(A) = 0$, $p > p'$ $\mathbb{P}(A) = 1$. p' is percolation threshold.

Theorem 4.6.2. Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, i.e. $\{\mathcal{F}_n\}_{n=0}^\infty$ form a filtration with $\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right)$. Then for $X \in \mathcal{L}^1$:

$$\mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{F}_\infty] \quad \text{a.s. and in } \mathcal{L}^1.$$

Proof. For $m > n$,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_m] \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n].$$

Let $Y_n = \mathbb{E}[X \mid \mathcal{F}_n]$. And it's easily to check Y_n is a MG: $\mathbb{E}[|Y_n|] < \infty$, $\{Y_n\}$ is adapted and $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = Y_n$. From, we know that Y_n is UI. By Jensen inequality and the function $|\cdot|$ is convex, we have $|Y_n| = |\mathbb{E}[X \mid \mathcal{F}_n]| \leq \mathbb{E}[|X| \mid \mathcal{F}_n]$. Then we know that

$$Y_n \text{ is U.I.} \implies Y_n \text{ converges a.s. \& in } \mathcal{L}^1 \text{ to some } Y,$$

by 4.5.11. For any $A \in \mathcal{F}_n$, by the definition of conditional expectation:

$$(*) \int_A X \, d\mathbb{P} = \int_A Y_n \, d\mathbb{P}.$$

Since $Y_n \to Y$ in $\mathcal{L}^1$, we have $\int_A Y_n \, d\mathbb{P} \to \int_A Y \, d\mathbb{P}$. Thus:

$$\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P} \quad \text{for all } A \in \bigcup_n \mathcal{F}_n.$$

Applying π-λ Theorem: Since $(*)$ holds for the algebra $\bigcup_n \mathcal{F}_n$, by the π-λ theorem, it holds for all $A \in \mathcal{F}_\infty$. $\implies$ $Y = \mathbb{E}[X|\mathcal{F}_\infty]$ a.s.. □

Theorem 4.6.3 (Levy's 0-1 Law)**.** If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $\mathbb{E}[\mathbb{1}_A|\mathcal{F}_n] \to \mathbb{E}[\mathbb{1}_A|\mathcal{F}_\infty] = \mathbb{1}_A$.

Proof. □

Definition 4.6.4 (Tail σ-algebra)**.** For a collection of σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \ldots$, let $\mathcal{G}_n = \sigma\left(\bigcup_{m=n}^\infty \mathcal{F}_m\right)$. The tail σ-algebra is defined as $\mathcal{T} = \bigcap_{n=1}^\infty \mathcal{G}_n$.

Corollary 4.6.5 (Kolmogorov's 0-1 Law)**.** Let $X_1, X_2, \ldots$ be independent. If $A \in \mathcal{T}$, then $\mathbb{P}(A) = 0$ or 1.

Proof. $S_n = \sigma(X_1, \ldots, X_n)$. If $A \in \mathcal{T} \implies A \in S_\infty \implies \mathbb{E}[\mathbb{1}_A|S_n] \to \mathbb{1}_A$ a.s. (Lévy's 0-1 Law 4.6.3).

Claim: A is independent of S_n for $\forall n$. Pf: $\mathcal{G}_{n+1} = \sigma(X_{n+1}, X_{n+2}, \ldots)$. Since $\mathcal{G}_{n+1}$ is independent of S_n and $A \in \mathcal{G}_{n+1} \implies A$ is independent of S_n. $\implies \mathbb{P}(A) = \mathbb{1}_A \implies \mathbb{P}(A) = 0$ or 1 □

Remark 4.6.6. We can introduce the exchangeable σ-algebra and Hewitt-Savage 0-1 law.

Theorem 4.6.7 (Dominated Convergence Theorem and Expecation)**.** If $Y_n \to Y$ a.s. and $|Y_n| \le Z$ for $\forall n$ with $\mathbb{E}[Z] < \infty$. Given $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, we have that:

$$\mathbb{E}[Y_n|\mathcal{F}_n] \to \mathbb{E}[Y|\mathcal{F}_\infty] \quad \text{a.s.}$$

Lemma 4.6.8. Let $X_n \ge 0$ and let $D = \{X_n = 0 \text{ for some } n\}$. Assume:

$$\mathbb{P}(D|X_1, \dots, X_n) \ge \delta(x) > 0 \quad \text{a.s. on } \{X_n \le x\}.$$

Then:

$$\mathbb{P}\left(D \cup \{\lim_n X_n = \infty\}\right) = 1.$$

Example 4.6.9 (branching process)**.** Z_n - branching process, s.t. # of offspring each individual has is k w.p. p_k for $k = 0, 1, 2, \dots$. If $p_0 > 0$, $\mathbb{P}\left(\lim_{n\to\infty} Z_n = 0 \text{ or } \infty\right) = 1$.

Use Lemma: Let $D = \{Z_n = 0 \text{ for some } n \ge 1\}$ (the extinction event).

$$\mathbb{P}(D|Z_1, \dots, Z_n) \ge \mathbb{P}(Z_{n+1} = 0|Z_1, \dots, Z_n) = p_0^{Z_n} \ge p_0^x$$

This holds a.s. on the event $\{Z_n \le x\}$.

Is $p_0 = 0$, is the same result true? Generally not, if $p_1 = 1$

Recall: if X_n is a sub-MG, and N is a bounded stopping time $\Rightarrow \mathbb{E}X_0 \le \mathbb{E}X_N$. This can fail if N is unbounded.

Theorem 4.6.10. If X_n is U.I. and sub-MG, then for any stopping time N, $X_{N\wedge n}$ is UI.

Proof.

$$\mathbb{E}\big[|X_N|\mathbb{1}\{|X_N| < K\}\big] = \underbrace{\mathbb{E}\big[|X_N|\mathbb{1}\{|X_N| < K\}, N \le n\big]}_{1} + \underbrace{\mathbb{E}\big[|X_N|\mathbb{1}\{|X_N| < K\}, N > n\big]}_{2}$$

and

$$\lim_{K\to\infty} \sup(2) = 0 \quad \text{because } X_n \text{ is UI}$$

Equation is sufficient to check that $\mathbb{E}|X_N| < \infty$. Sub-MG convergence theorem $\Rightarrow X_{N\wedge n} \to X_N$ a.s. $\Rightarrow \mathbb{E}|X_N| < \infty$.

Need to check: $\sup_n \mathbb{E}X^+_{N\wedge n} < \infty$.

$$\mathbb{E}X^+_{N\wedge n} \le \mathbb{E}X^+_n \overset{\text{UI}}{\Longrightarrow} \sup_n \mathbb{E}X^+_{N\wedge n} \le \sup_n \mathbb{E}X^+_n < \infty.$$

□

Theorem 4.6.11. If X_n is a sub-MG and for stopping itme N, $X_{N\wedge n}$ is UI, then $\mathbb{E}X_0 \le \mathbb{E}X_N$.

Proof. $X_{N\wedge n} \to X_N$ a.s. and in $\mathcal{L}^1$. $\mathbb{E}X_0 \le \mathbb{E}X_{N\wedge n} \to \mathbb{E}X_N$ □

Theorem 4.6.12. If X_n is a UI and sub-MG then for any stopping time N we have $\mathbb{E}X_0 \le \mathbb{E}X_N \le \mathbb{E}X_\infty$ where $X_\infty = \lim_{n\to\infty} X_n$.

Proof. By last theorem, we have $\mathbb{E}X_{N\wedge n} \le \mathbb{E}X_n$. By $\mathcal{L}^1$ convergence and UI, we have $\mathbb{E}X_N \le \mathbb{E}X_\infty$. □

Theorem 4.6.13. If $X_n \ge 0$ is super-MG and N is any stopping time, then $\mathbb{E}X_0 \ge \mathbb{E}X_N$.

Proof. We know that $\mathbb{E}X_0 \geq \mathbb{E}X_{N\wedge n}$. By Fatou's Lemma, we have $\liminf \mathbb{E}X_{N\wedge n} \geq \mathbb{E}[\liminf X_{N\wedge n}] = \mathbb{E}X_N$. The last equality is because $X_{N\wedge n} \geq 0$ is a super-MG. □

4.7 Feb 9

by Tatiana (Tanya) Brailovskaya

Warm-up: Suppose that X_n is a sub-MG and $\mathbb{E}(|X_{n+1} - X_n| \mid \mathcal{F}_n) \leq B$ a.s. Show that if N is a stopping time with $\mathbb{E}N < \infty$ then $X_{N\wedge n}$ is UI and hence $\mathbb{E}X_N \geq \mathbb{E}X_0$

Hint: Write

$$X_{N\wedge n} = X_0 + \sum_{m=0}^{n-1}(X_{m+1} - X_m)\mathbb{1}_{(N>m)}$$

then,

$$|X_{N\wedge n}| \leq |X_0| + \sum_{m=0}^{\infty}|X_{m+1} - X_m|\mathbb{1}_{(N>m)}.$$

Since RHS in the above inequality does not depend on n, to conclude that $X_{N\wedge n}$ is UI (?), it suffices to show that RHS is integrable.

Proof.

$$\mathbb{E}[|X_{m+1} - X_m|\mathbb{1}_{(N>m)}] = \mathbb{E}[\mathbb{E}[|X_{m+1} - X_m| \mid \mathcal{F}_m]\mathbb{1}_{(N>m)}] \leq B\mathbb{P}(N > m) \quad \text{a.s.}$$

$$\implies \sum_{m=0}^{\infty}\mathbb{E}[|X_{m+1} - X_m|\mathbb{1}_{(N>m)}] \leq B\sum_{m=0}^{\infty}\mathbb{P}(N > m) = B\mathbb{E}N < \infty$$

By definition of sub-MG, $\mathbb{E}|X_0| < \infty \implies X_{N\wedge n}$ is UI. Why $X_{N\wedge n}$ UI $\implies \mathbb{E}X_N \geq \mathbb{E}X_0$? Because $\mathbb{E}X_{N\wedge n} \geq \mathbb{E}X_0$ since $N \wedge n \leq n$. $\mathbb{E}[X_{N\wedge n}] \geq \mathbb{E}[X_{N\wedge 0}] = \mathbb{E}[X_0]$ Since the definition of sub-MG $\implies$ first inequality. Then $\mathbb{E}X_{N\wedge n} \to \mathbb{E}X_N$ because UI $\implies \mathcal{L}^1$ convergence. □

This lecture, we will talk about the application of optimal stopping theorems.

Applications to the Random Walk (RW). Let $Y_1, \ldots, Y_n$ be iid, $S_n = S_0 + Y_1 + \cdots + Y_n$ where S_0 is a constant. $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. We will use 3 different types of martingales one can produce from S_n: (i) Linear MG: let $\mu = \mathbb{E}Y_i$, then $X_n = S_n - n\mu$ is a MG.
(ii) Quadratic MG: Suppose $\mathbb{E}Y_i = 0$ & $\mathbb{E}Y_i^2 = \sigma^2 < \infty$. Then $X_n = S_n^2 - n\sigma^2$ is a MG.
(iii) Exponential MG: Suppose $\phi(\theta) = \mathbb{E}\exp(\theta Y_i) < \infty$. Then $X_n = \frac{\exp(\theta S_n)}{\phi(\theta)^n}$ is a MG.

First application of optional stopping to RW: Wald's equation.

Question 4.7.1. Consider the RW S_n with $\mathbb{E}Y_i = \mu$ & stopping time N s.t. $\mathbb{E}N < \infty$. Show that $\mathbb{E}S_N = \mu\mathbb{E}N$. *Hint: use the optional stopping thm we proved in the warm-up.*

Proof. Use the MG (i). Then

$$\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] = \mathbb{E}[|Y_{n+1} - \mu|] \leq \underbrace{\mathbb{E}|Y_1| + \mu}_{B}.$$

Now by optional stopping thm from warm-up: $\mathbb{E}X_N = \mathbb{E}X_0 = 0$. Recall: $X_N = S_N - \mu N \implies \mathbb{E}S_N = \mu\mathbb{E}N$. □

Theorem 4.7.2. Consider the symmetric simple RW, i.e. the case in which $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = 1/2$. Suppose $S_0 = x$ and let $N = \{n : S_n \notin (a, b)\}$. Then,

1. $$\mathbb{P}(S_N = a) = \frac{b - x}{b - a}, \qquad \mathbb{P}(S_N = b) = \frac{x - a}{b - a}$$

2. $\mathbb{E}_0 N = -ab$ and more generally, $\mathbb{E}_x N = (b - x)(x - a)$ here $\mathbb{E}_x N$ denotes $\mathbb{E}N$ for given starting point x.

Proof. (a) We want to use Optional Stopping thm (e.g. thm from warm-up) to obtain the following eq

$$x = \mathbb{E}S_0 = \mathbb{E}S_N = a\mathbb{P}(S_N = a) + b(1 - \mathbb{P}(S_N = a)) \quad (*).$$

Solving $(*)$ for $\mathbb{P}(S_N = a)$, the desired result follows. To verify $(*)$ holds, we will need to check: $\mathbb{E}[|S_{n+1} - S_n| \mid \mathcal{F}_n] \leq B$ a.s.$\mathbb{E}N < \infty$ (i) we already checked when deriving the Wald Eq.
(ii) we only consider the case $x \in (a, b)$. Notice that if the RW takes $(b - a)$ consecutive $+1$ steps, it will necessarily exit the interval. This happens with probability $2^{-(b-a)}$.

$$\implies \mathbb{P}(N > m(b - a)) \leq (1 - 2^{-(b-a)})^m$$

$$\mathbb{E}N \leq \sum_{m=0}^{\infty} (b - a)\mathbb{P}(N > m(b - a)) = (b - a)\frac{1}{1 - (1 - 2^{-(b-a)})} = (b - a)2^{(b-a)} < \infty.$$

(b) To prove this we may want to use Wald Eq. $\mathbb{E}_0 S_N = \mu \mathbb{E}N$. Why wouldn't that work here? (Note: for the symmetric case, $\mu = 0$, so the equation becomes $0 = 0$, which provides no information about $\mathbb{E}N$).Instead we work with the quadratic MG $S_n^2 - n$:

$$\implies \underbrace{\mathbb{E}_0(S_{0\wedge N}^2 - 0 \wedge N)}_{=0} = \mathbb{E}_0(S_{n\wedge N}^2 - n \wedge N)$$

By Monotone convergence:

$$\mathbb{E}_0(n \wedge N) \to \mathbb{E}_0 N$$

By bounded convergence:

$$\mathbb{E}_0 S_{n\wedge N}^2 \to \mathbb{E}_0 S_N^2$$

By part (a):

$$\mathbb{E}_0 S_N^2 = a^2 \frac{b}{b - a} + b^2 \left(\frac{-a}{b - a}\right) = -ab = \mathbb{E}_0 N$$

How can we now get the answer for arbitrary x given $\mathbb{E}_0 N$? □

Interprelation: Thm 1 is only meaningful if $x \in (a, b)$. (think about what happens when $x \notin (a, b)$). Unsurprisingly, if $|x - a| < |b - x|$ the RW is more likely to visit a before b (and vice versa). Also, the farther x is from the endpoints of the interval (a, b), the longer it takes (on average) for the RW to exit (a, b).

Remark 4.7.3. let $T_x = \min\{n : S_n = x\}$. Taking $a = 0, x = 1, b = M$ we have: $\mathbb{P}(T_M < T_0) = 1/M$. Let S_n denote the RW with $S_0 = 1$. Then

$$\mathbb{P}(T_M < T_0) = \mathbb{P}\left(\sup_n S_{T_0 \wedge n} \geq M\right) \leq \frac{1}{M},$$

inequality holds by Doob's inequality + Monotone Convergence $\implies$ Doob's inequality is sharp for this example (we discussed this before today we see the pf.!) $\mathbb{P}(T_0 < T_M) = \frac{M-1}{M}$ Letting $M \to \infty$, $\mathbb{P}(T_0 < \infty) = 1$, i.e. with probability 1 the RW started 1 will reach 0 in finite time.

Question 4.7.4. Is it also the case that $\mathbb{E}_1 T_0 < \infty$? Why or why not?

Proof. $\mathbb{E}_1 T_0 = \infty$. By Thm 1 (b) We see that if $a = 0, x = 1, b = M$ $\mathbb{E}_1(T_0 \wedge T_M) = M - 1$. By Monotone convergence, we can take $M \to \infty$ to get the desired conclusion. □

4.8 Feb 11

by Tatiana (Tanya) Brailovskaya

Question 4.8.1. A monkey is typing a letter (A-Z) independently uniformly every second. How long in expectation until the monkey types out ABRACADABRA?

How about AB?

Warm-up: Come up with a MG X_n such that

$$T = \{\text{time by which monkey types AB is a stopping time for } X_n\}.$$

Hint: think about betting strategies.

Proof. A player bets 1\$ on letter A & if they're correct, they win 26\$. Next, they again bet 1\$ on letter A & an additional bet of 26\$ that next letter is B. If B shows up the player wins 26^2. If any of the bets are incorrect, the player loses the stakes, and a new player joins the game. X_n is the amount of money the Casino makes after n rounds.

First, we should verify X_n is a MG. If player at time n is betting only on A:

$$X_n = \begin{cases} X_{n-1} + 1 & \text{w.p. } 25/26 \\ X_{n-1} - 25 & \text{w.p. } 1/26 \end{cases}.$$

If player at time n is betting on both A & B:

$$X_n = \begin{cases} X_{n-1} + 27 & \text{w.p. } 24/26 \\ X_{n-1} + 1 & \text{w.p. } 1/26 \\ X_{n-1} + 27 - 26^2 & \text{w.p. } 1/26 \end{cases}.$$

By optional stopping: $\mathbb{E}[X_T] = X_0 \leftarrow$ initial amount of money the casino has. How do we represent X_T in terms of T?

$$X_T = X_0 + \underbrace{T_1}_{\substack{\#\text{ of players} \\ \text{that lost after 1 round}}} + 2 \cdot \underbrace{T_2}_{\substack{\#\text{ of players} \\ \text{who lost after 2 rounds}}} - (26^2 - 2) = X_0 + T - 2 - (26^2 - 2).$$

Then we have $\mathbb{E}T = 26^2$.

What about the typing AA? The $\mathbb{E}[T] =$. I still don't udnerstand the whole system. $\square$

Idea: in a filtration, the σ-algebra is like $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$. what if instead of $\mathcal{F}_0 \supseteq \mathcal{F}_1 \supseteq \dots$?

Definition 4.8.2. $X_0, X_{-1}, X_{-2}, \dots$ is a backward martingale if it is adapted $\mathcal{F}_0 \supseteq \mathcal{F}_{-1} \supseteq \mathcal{F}_{-2} \dots$ and $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$.

Theorem 4.8.3. $X_{-\infty} = \lim_{n \to -\infty} X_n$ exists in a.s. & $\mathcal{L}^1$.

Proof. We used upcrossing inequality 4.3.2 to show that # of upcrossings $< \infty_{\text{a.s.}}$. We can think of $X_{-n} \dots X_0$ as a MG $Y_0 \dots Y_n$. Then we have

$$(b-a)\mathbb{E}[\underbrace{U_n}_{\#\text{ of upcrossings of (a,b) 'between' (-n, 0)}} \leq \mathbb{E}(X_0 - a)^+ - \mathbb{E}(X_{-n} - a)^+ \leq \mathbb{E}(X_0 - a)^+.$$

By MCT, we have $\mathbb{E}U_\infty \leq \frac{\mathbb{E}(X_0 - a)^+}{(b-a)} < \infty$. Then $U_\infty < \infty$ a.s. $\lim_{n \to -\infty} X_n$ exists. Next, we should state the $\mathcal{L}^1$ convergence. For $m > n, \mathbb{E}[X_m | \mathcal{F}_n] = X_n$. In particular, $\mathbb{E}[X_0 | \mathcal{F}_n] = X_n \Rightarrow X_n$ U.I.$\Rightarrow X_n$ converges in $\mathcal{L}^1$. $\square$

Theorem 4.8.4. If $X_{-\infty} = \lim_{n \to -\infty} X_n$ and $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ then: $X_{-\infty} = \mathbb{E}[X_0 | \mathcal{F}_{-\infty}]$.

Proof. Note: $X_{-\infty}$ is measurable w.r.t. $\mathcal{F}_{-\infty}$. This is because $X_{-\infty}$ is measurable w.r.t. $\mathcal{F}_{-n}$ (consider $\lim_{k>n} X_{-k} = X_{-\infty}$). Recall: $X_n = \mathbb{E}[X_0 | \mathcal{F}_n] \implies$ for $\forall A \in \mathcal{F}_{-\infty}$. By the property of conditional expectation

$$\int_A X_n d\mathbb{P} = \int_A X_0 d\mathbb{P}.$$

By $\mathcal{L}^1$ convergence (as shown in the previous proof)

$$\int_A X_n d\mathbb{P} \xrightarrow{\mathcal{L}^1 \text{ conv.}} \int_A X_{-\infty} d\mathbb{P}.$$

Combining these, we have

$$\int_A X_{-\infty} d\mathbb{P} = \int_A X_0 d\mathbb{P} \implies X_{-\infty} = \mathbb{E}[X_0|\mathcal{F}_{-\infty}].$$

□

Application SLLN Let $Y_1, Y_2, \dots$ be i.i.d. random variables with $\mathbb{E}|Y_1| < \infty$. Let $S_n = Y_1 + \cdots + Y_n$ and $X_{-n} = \frac{S_n}{n}$. Let $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \dots) = \sigma(S_n, Y_{n+1}, Y_{n+2}, \dots)$. Claim: X_{-n} is a backwards martingale.

Proof.

$$\mathbb{E}[X_{-n}|\mathcal{F}_{-n-1}] = \mathbb{E}\left[\frac{S_{n+1} - Y_{n+1}}{n}\middle|\mathcal{F}_{-n-1}\right] = \frac{S_{n+1}}{n} - \frac{1}{n}\mathbb{E}[Y_{n+1}|\mathcal{F}_{-n-1}]$$

Note: For $j, k \leq n+1$

$$\implies \mathbb{E}[Y_j|\mathcal{F}_{-n-1}] = \mathbb{E}[Y_k|\mathcal{F}_{-n-1}].$$

$$\implies \mathbb{E}[Y_{n+1}|\mathcal{F}_{-n-1}] = \frac{1}{n+1}\sum_{k=1}^{n+1}\mathbb{E}[Y_k|\mathcal{F}_{-n-1}] = \frac{S_{n+1}}{n+1}.$$

Substituting back into the first equation:

$$\mathbb{E}[X_{-n}|\mathcal{F}_{-n-1}] = \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = X_{-n-1}$$

□

Chapter 5

Markov Chain

5.1 Feb 16

by Tatiana (Tanya) Brailovskaya

Warm-up: show that if X_n is a backward MG with $X_0 \in L^p$ for $p > 1$ then X_n converges in L^p. Hint: Recall pf of L^p convergence for regular MG.

Proof. $\mathcal{L}^p$ maximum inequaliy: if Y_n is a MG then

$$\mathbb{E} \max_{0 \le m \le n} |Y_m|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|Y_n|^p$$

$$\implies \mathbb{E} \max_{-n \le m \le 0} |X_m|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_0|^p$$

$$\xrightarrow{\text{MCT}} \mathbb{E} \max_{m} |X_m|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_0|^p.$$

Since $X_n \to X_{-\infty}$ a.s. suffice to apply DCT

$$|X_n - X_{-\infty}|^p \le \underbrace{(2 \sup_n |X_n|)^p}_{\text{integrable r.v.}},$$

then we have $X_n \to X_{-\infty}$ in $\mathcal{L}^p$. □

So far:

1. Conditions for a.s. + L^p convergence $1 \le p < \infty$.
2. Proving tail bounds via Doob's inequality & Doob MG
3. More general 0-1 laws than Kolmogorov
4. Optional stopping thms for computing $\mathbb{E}T$ for stopping time T

Examples: random walks, branching processes, betting strategies.

Definition 5.1.1. We say that X_n with a countable state space is a Markov Chain (MC), if it satisfies the Markov property for $\forall i_0, i_1, \ldots, i_n, i \& j$ in the state space we have that

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \underbrace{\mathbb{P}(X_{n+1} = j \mid X_n = i)}_{p(i,j) \text{ - transition probability}}$$

Example 5.1.2. 1. Random Walk. You are standing at some position i on an integer grid ($\mathbb{Z}^d$). At each step, you take a jump of size Y_m (which the Durrett text denotes as ξ_m), drawn from a fixed probability distribution μ. Your position at time $n+1$ (X_{n+1}) only depends on your current position X_n, plus the completely independent new jump Y_{n+1}. The path you took to get to X_n is irrelevant. To get from state i to state j, your next random jump must be exactly the distance between them, which is $j-i$. Since the jumps are drawn from the distribution μ, the probability of making exactly this jump is simply $\mu(\{j-i\})$. A standard random walk doesn't tend to an "equilibrium" because it usually wanders off to infinity or keeps oscillating forever without settling down into a stable probability distribution

2. Braching Process. Let $Y_1, Y_2, \dots$ iid r.v.'s (offspring #'s) then

$$p(i,j) = \mathbb{P}\left(\sum_{m=1}^{i} Y_m = j\right).$$

What happens as $n \to \infty$? (in (1) + (2)). (1) + (2) X_n doesn't tend to an "equilibrium".

3. Wright-Fisher model. Population of $N/2$ diploid individuals, i.e. they have 2 alleles of same gene: A & a. we think of gene pool at time $n+1$ as a sample with replacement from the gene pool at time n. $X_n = \#$ of alleles A at time n.

$$p(i,j) = \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j} \binom{N}{j} \qquad X_n \sim \mathrm{Bin}(N, \frac{i}{N}).$$

As $n \to \infty$, $X_n = 0$ or $X_n = N$ "absorbing states".

Definition 5.1.3. Let $(S, \mathcal{F})$ $\mathcal{F}$ is a σ-alg on S be a measurable space. A function $p : S \times \mathcal{F} \to \mathbb{R}$ is a transition probabiltiy if

1. For $\forall x \in S : A \to p(x, A)$ is a probability measure on $(S, \mathcal{F})$.
2. For each $A \in \mathcal{F}, x \to p(x, A)$ is a measurable function.

We say that X_n is a MC (Markov Chain) with transition probability p if

$$\mathbb{P}(X_{n+1} \in B \mid \sigma(X_0, \dots, X_n)) = p(X_n, B).$$

Q: How do we know X_n satisfying definition exists?

1. We need to construct a probability space on $S^{\mathbb{N}}$
2. let $\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \mathbb{R}\}$. We equip $\mathbb{R}^{\mathbb{N}}$ with the product σ-algebra $\mathbb{R}^{\mathbb{N}}$ generated by

$$\{\omega : \omega_i \in (a_i, b_i] \text{ for } i = 1, \dots, n\}.$$

Theorem 5.1.4. Suppose we are given probability measure μ_n on $(\mathbb{R}^n, \mathcal{R}^n)$ that are consistent[1], meaning

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

$$\implies \exists \text{ a unique probability measurable } P \text{ on } (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}) \text{ w/}$$
$$P(\omega : \omega_i \in (a_i, b_i], 1 \le i \le n) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

[1]Consider $\mu_{n+1}(A \times \mathbb{R}) = \mu_n(A)$ The probability of the first n steps landing in region A, and the $(n+1)$-th step landing anywhere in $\mathbb{R}$, must equal the probability of just the first n steps landing in A

Remark 5.1.5. Thm 5.1.4 holds $(S, \mathcal{F})$ if $\exists$ 1-1 map $\psi : S \to \mathbb{R}$ s.t. ψ & ψ^{-1} are measurable. (see more: Thm 2.1.22 in Durrett).

Definition 5.1.6. Let $(S, \mathcal{F})$ be a measurable space ($\mathcal{F}$ is a σ-alg on S). A function $p : S \times \mathcal{F} \to \mathbb{R}$ is a transition probability if:

1. For $\forall x \in S : A \to p(x, A)$ is a probability measure on $(S, \mathcal{F})$.
2. For each $A \in \mathcal{F}, x \to p(x, A)$ is a measurable function.

We say that X_n is a MC with transition probability p if

$$\mathbb{P}(X_{n+1} \in B \mid \sigma(X_0, \dots, X_n)) = p(X_n, B).$$

We need a family of consistent meas. to apply Thm 1: X_0 has init μ, then

$$\mu_n(B_0 \times B_1 \times \cdots \times B_n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n) \implies P_\mu \text{ exists.}$$

Theorem 5.1.7. let $X_n(\omega) = \omega_n$ is MC with transition probability p, i.e.

$$\mathbb{P}_\mu(X_{n+1} \in B \mid \underbrace{\sigma(X_0, \dots, X_n)}_{\mathcal{F}_n}) = p(X_n, B).$$

Proof. Goal: show for $\forall A \in \mathcal{F}_n : P_\mu(X_{n+1} \in B; A) = \int_A p(x_n, B) dP_\mu$.

Suffices to check these A's only by π-λ thm. Let $A = \{X_0 \in B_0, X_1 \in B_1, \dots, X_n \in B_n\} = B_0 \times B_1 \times \cdots \times B_n$

$$\begin{aligned}\mathbb{P}_\mu(X_{n+1} \in B; A) &= \mathbb{P}_\mu(X_0 \in B_0, X_1 \in B_1, \dots, X_n \in B_n, X_{n+1} \in B) \\ &= \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n) p(x_n, B)\end{aligned}$$

Recall: $x_n \to p(x_n, B)$ is a bdd measurable function. For any $C \in \mathcal{F}$

$$\int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n) \mathbb{1}_C(x_n) = \int_A \mathbb{1}_C(x_n) dP_\mu.$$

By linearity, extend[a] to simple function, then bdd measurable function. ☐

[a] I am not sure about the extension. Need to check

5.2 Feb 18

by Tatiana (Tanya) Brailovskaya

Warm-up: Wright-Fisher $X_n = 0$ & N are abosrbing states, i.e. $p(0,0) = 1$ & $p(N,N) = 1$. Show that $\mathbb{P}_X(X_n \in \{0, N\}) \to 1$ as $n \to \infty$ for all X.

We will use Lévy's 0-1 Law to show that for any MC X_n if

$$(*) \begin{cases} \mathbb{P}\left(\bigcup_{m=n+1}^{\infty} \{X_m \in B_m\} \mid X_n\right) \geq \delta > 0 & \text{on } \{X_n \in A_n\} \text{ then} \\ \mathbb{P}(\{X_n \in A_n \text{ i.o.}\} \setminus \{X_n \in B_n \text{ i.o.}\}) = 0 \end{cases}$$

1. Show $(*)$ implies $\mathbb{P}_x(X_n \in \{0, N\}) \to 1$
2. Apply 0-1 Law to the event $E = \bigcup_{m=n+1}^{\infty} \{X_m \in B_m\}$ to prove $(*)$

Recall: Levy's 0-1 Law. If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$ then $\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_n] \to \mathbb{1}_A$ a.s.

Proof. 1. Applying the Lemma ($*$) to the Wright-Fisher Model.

To show that $\mathbb{P}_x(X_n \in \{0, N\}) \to 1$, we need to map the Wright-Fisher model to the events in the lemma.

Let our target absorbing states be $B_n = \{0, N\}$. Let the remaining "transient" states be $A_n = \{1, 2, \ldots, N-1\}$. For the Wright-Fisher model, the transition probabilities from any state i follow a binomial distribution: $p(i,j) = \binom{N}{j}(i/N)^j(1-i/N)^{N-j}$. If the chain is currently in some state $X_n = i \in A_n$, what is the probability it jumps **directly** into one of the absorbing states on the very next step? Probability of jumping to 0: $p(i,0) = (1-i/N)^N$. Probability of jumping to N: $p(i,N) = (i/N)^N$.

Because $i \in \{1, \ldots, N-1\}$, we know the fractions satisfy $1/N \le i/N \le (N-1)/N$. This strict boundary guarantees a minimum probability of jumping to an absorbing state:

$$p(i,0) \ge \left(\frac{1}{N}\right)^N \quad \text{and} \quad p(i,N) \ge \left(\frac{1}{N}\right)^N$$

Therefore, the probability of getting absorbed on the very next step is strictly bounded away from zero:

$$\mathbb{P}(X_{n+1} \in \{0, N\} \mid X_n = i) = p(i,0) + p(i,N) \ge 2\left(\frac{1}{N}\right)^N$$

Let $\delta = 2(1/N)^N > 0$. Because hitting the absorbing states in the next step is a subset of hitting them at some point in the future ($\{X_{n+1} \in B_{n+1}\} \subset \bigcup_{m=n+1}^{\infty}\{X_m \in B_m\}$), we have:

$$\mathbb{P}\left(\bigcup_{m=n+1}^{\infty}\{X_m \in B_m\} \mid X_n\right) \ge \mathbb{P}(X_{n+1} \in B_{n+1} \mid X_n) \ge \delta > 0 \quad \text{on } \{X_n \in A_n\}$$

This perfectly satisfies the condition of lemma ($*$). The lemma immediately grants us:

$$\mathbb{P}(\{X_n \in \{1, \ldots, N-1\} \text{ i.o.}\} \setminus \{X_n \in \{0, N\} \text{ i.o.}\}) = 0.$$

Since the state space is finite, the Markov chain must visit some states infinitely often (i.o.). If the chain never hits $\{0, N\}$, it is forced to stay in $\{1, \ldots, N-1\}$ forever, meaning it visits A_n infinitely often. But the lemma just proved the probability of doing this without hitting B_n is 0. Therefore, with probability 1, the chain must eventually hit $\{0, N\}$. Because 0 and N are absorbing states ($p(0,0) = 1$ and $p(N,N) = 1$), once the chain hits them, it stays there forever. This proves $\mathbb{P}_x(X_n \in \{0, N\}) \to 1$.

2. Proving the Lemma ($*$) using Lévy's 0-1 Law

Let $E_n = \bigcup_{m=n+1}^{\infty}\{X_m \in B_m\}$. This is a decreasing sequence of events ($E_1 \supset E_2 \supset \ldots$) that shrinks down exactly to the tail event $E = \{X_m \in B_m \text{ i.o.}\}$. Because $E \in \mathcal{F}_\infty$, we can apply Lévy's 0-1 Law, which states that our best guess of the event E given the information up to time n will converge to the actual outcome:

$$\mathbb{P}(E \mid \mathcal{F}_n) \to \mathbb{1}_E \quad \text{a.s.}$$

We can split the conditional probability of E_n into two pieces:

$$\mathbb{P}(E_n \mid \mathcal{F}_n) = \mathbb{P}(E \mid \mathcal{F}_n) + \mathbb{P}(E_n \setminus E \mid \mathcal{F}_n).$$

The first term $\mathbb{P}(E \mid \mathcal{F}_n)$ converges to $\mathbb{1}_E$ almost surely. The second term concerns the event $E_n \setminus E$, which is a sequence of events decreasing to $\emptyset$. Because the events vanish in the limit, it can be shown that $\mathbb{P}(E_n \setminus E \mid \mathcal{F}_n) \to 0$ almost surely.

Adding these limits together, we get:

$$\mathbb{P}(E_n \mid \mathcal{F}_n) \to \mathbb{1}_E \quad \text{a.s.}$$

The Contradiction Argument: Suppose for the sake of contradiction that there is an outcome ω that belongs to the event $\{X_n \in A_n \text{ i.o.}\} \setminus E$. 1. Because $\omega \notin E$, the indicator function $\mathbb{1}_E(\omega) = 0$. By our limit above, this requires that $\mathbb{P}(E_n \mid \mathcal{F}_n)(\omega) \to 0$ as $n \to \infty$. 2. Because $\omega \in \{X_n \in A_n \text{ i.o.}\}$, the event $X_n \in A_n$ happens for infinitely many time steps n. By the core assumption of the lemma, at all of these infinitely many steps, we are guaranteed that $\mathbb{P}(E_n \mid \mathcal{F}_n)(\omega) \ge \delta > 0$.

It is mathematically impossible for a sequence to eventually converge to 0 while simultaneously spiking up to be $\ge \delta > 0$ infinitely often. This contradiction proves that the event $\{X_n \in A_n \text{ i.o.}\} \setminus \{X_n \in B_n \text{ i.o.}\}$ cannot occur, so its probability is strictly 0.

Would you like to move on to the second half of the Wright-Fisher midterm question and compute the exact absorption probabilities $\mathbb{P}_x(V_N < V_0)$ using the optional stopping theorem? □

Remark 5.2.1. If X_n is a MC with absorbing states A then we have $\mathbb{P}_X(X_n \in A) \to 1$ if $\exists m < \infty$ such that $\mathbb{P}_X(X_m \in A) \geq \delta > 0$.

Proof. Step 1: Define the target event. Let E be the event that the Markov chain eventually hits the absorbing states A at some point in the future. Because A is an absorbing state, hitting A means staying in A forever. Thus, the event E can be written as $E = \{X_n \in A \text{ eventually}\}$. Let $F_n = \sigma(X_0, \dots, X_n)$ be the history of the chain up to time n.

Step 2: Apply Lévy's 0-1 Law. Because E is an event in the infinite future σ-algebra (F_∞), we have

$$\mathbb{P}(E \mid F_n) \to \mathbb{1}_E \quad \text{almost surely.}$$

Step 3: Apply the assumption. We are given that if the chain has not yet been absorbed, there is always a probability of at least $\delta > 0$ that it will hit the absorbing states A within the next m steps. Since hitting A in the next m steps implies that E happens, we have:

$$\mathbb{P}(E \mid F_n) \geq \mathbb{P}(X_{n+m} \in A \mid X_n) \geq \delta > 0$$

as long as the chain is not yet in A.

Step 4: Proof by Contradiction. Suppose for the sake of contradiction that the chain never hits A. Because the chain never hits A, the event E did not happen, which means the indicator function $\mathbb{1}_E = 0$. By Lévy's 0-1 Law from Step 2, this requires that the sequence of conditional probabilities $\mathbb{P}(E \mid F_n)$ must converge to 0 as $n \to \infty$. However, because the chain never hits A, it is always in the unabsorbed transient states. Thus, by Step 3, the condition $\mathbb{P}(E \mid F_n) \geq \delta > 0$ holds for all n.

It is mathematically impossible for a sequence to strictly remain $\geq \delta > 0$ and simultaneously converge to 0. This contradiction proves that the probability of the chain "never hitting A" is exactly 0. Therefore, the chain must hit A almost surely. Since A is an absorbing state (once you enter, you cannot leave), the probability that the chain is in A goes to 1 as $n \to \infty$: $\lim_{n\to\infty} \mathbb{P}_x(X_n \in A) = 1$. □

We are goting to talk about Markov Properties

$$\mathbb{P}(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B), \qquad \mathbb{E}[I(X_{n+1} \in B) \mid \mathcal{F}_n].$$

what if we instead plug-in $h(X_{n+1}, X_{n+2}, \dots)$ or if we let n be a stopping time.

Theorem 5.2.2 (Monotone Class Theorem)**.** Let $\mathcal{A}$ be a π-system that contains Ω & $\mathcal{H}$ be a collection of $\mathbb{R}$-valued func s.t.

1. if $A \in \mathcal{A} \implies \mathbb{1}_A \in \mathcal{H}$
2. $f, g \in \mathcal{H} \implies af + bg \in \mathcal{H}$ for $\forall a, b \in \mathbb{R}$
3. if $\exists f_n \geq 0$ & $f_n \uparrow f$ with f bounded $\implies f \in \mathcal{H}$ then $\mathcal{H}$ contains all function measurable w.r.t. $\sigma(\mathcal{A})$

Lemma 5.2.3. For any bounded f, $\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \int_\Omega f(y)p(X_n, dy)$

Proof. $\mathbb{E}[f(X_{n+1}); A] \overset{(\Delta)}{=} \int_A \int_\Omega f(y)p(X_n, dy)d\mathbb{P}$. We want to show for $\forall A \in \mathcal{F}_n$. Let $\mathcal{A} = \{X_0 \in A_0, \dots, X_n \in A_n\}$ for $A_i \in \mathcal{F}$ which is σ-algebra for state space. A is π-system, and $\sigma(\mathcal{A}) = \mathcal{F}_n$.

Argue (Δ) for $\forall A \in \mathcal{A}$ & $f = \mathbb{1}_C$ for $C \in \mathcal{A}$. Then argue that $\mathcal{H}$ for which (Δ) holds on $\mathcal{A}$ satisfies conditions of theorem 5.2.2. Conclude that (Δ) holds for all $A \in \mathcal{F}_n$ by π-λ theorem. □

Corollary 5.2.4.

$$\mathbb{E}_\mu\left[\prod_{k=0}^{n} f_k(X_k)\right] = \int_\Omega \mu(dx_0)f_0(x_0)\int_\Omega p(x_0, dx_1)f_1(x_1)\cdots\int_\Omega p(x_{n-1}, dx_n)f_n(x_n)$$

Proof. First, we have
$$\mathbb{E}[f_0(X_0)] = \int_\Omega \mu(dx_0) f_0(x_0).$$
We suppose it holds for $n-1$, we have
$$\begin{aligned}
\mathbb{E}_\mu\Big[\prod_{k=0}^n f_k(X_k)\Big] &= \mathbb{E}_\mu\Big[\mathbb{E}\Big[\prod_{k=0}^n f_k(X_k) \mid \mathcal{F}_{n-1}\Big]\Big] \\
&= \mathbb{E}_\mu\Big[\prod_{k=0}^{n-1} f_k(X_k)\mathbb{E}\Big[f_n(X_n) \mid \mathcal{F}_{n-1}\Big]\Big] \\
&= \mathbb{E}_\mu\Big[\prod_{k=0}^{n-1} f_k(X_k)\int_\Omega f_n(y)p(X_{n-1}, dy)\Big] \\
&= \mathbb{E}_\mu\Big[\prod_{k=0}^{n-2} f_k(X_k)\cdot f_{n-1}(X_{n-1})\int_\Omega f_n(y)p(X_{n-1}, dy)\Big] \\
&= \int_\Omega \mu(dx_0)f_0(x_0)\int_\Omega p(x_0, dx_1)f_1(x_1)\cdots\int_\Omega p(x_{n-2}, dx_{n-1})\cdot f_{n-1}(x_{n-1})\int_\Omega f_n(y)p(x_{n-1}, dy) \\
&= \int_\Omega \mu(dx_0)f_0(x_0)\int_\Omega p(x_0, dx_1)f_1(x_1)\cdots\int_\Omega p(x_{n-2}, dx_{n-1})\cdot f_{n-1}(x_{n-1})\int_\Omega f_n(x_n)p(x_{n-1}, dx_n).
\end{aligned}$$
□

Theorem 5.2.5. Let $h : \Omega^{\mathbb{N}} \to \mathbb{R}$ be bounded & measurable. Then, $\mathbb{E}_\mu[h(X_m, X_{m+1}\dots) \mid \mathcal{F}_m] = \mathbb{E}_{X_m}[h(X_0, X_1\dots)]$

Proof. First, we start at simple function h
$$h(y_0, y_1, \dots) = f_0(y_0)\cdots f_k(y_k),$$
where f is bounded and measurable function. We look at LHS[a],
$$\begin{aligned}
\text{LHS} &= \mathbb{E}[f_0(X_m)f_1(X_{m+1})\cdots f_k(X_{m+k}) \mid \mathcal{F}_m] \\
&= f_0(X_m)\mathbb{E}[f_1(X_{m+1})\cdots f_k(X_{m+k}) \mid \mathcal{F}_m] \\
&= f_0(X_m)\int_\Omega p(X_m, dx_1)f_1(x_1)\cdots \\
&= \text{RHS}.
\end{aligned}$$
Let $\mathcal{H}$ contains all the function satisfying the above equation. We should test 5.2.2's conditions.

1. Of course holds, we just let $f = \mathbb{1}$.
2. Linearity of expectation.
3. By conditional MCT, it holds.

□

[a] By the definition of Markov Chain, we have
$$\mathbb{E}[f_k(X_{m+k}) \mid \mathcal{F}_{m+k-1}] = \int_S p(X_{m+k-1}, dy)f_k(y).$$

Let $\mathcal{F}_N = \{A : A \cap \{N = n\} \in \mathcal{F}_n\}$

Theorem 5.2.6 (Strong Markov Property)**.** Suppose for each n, $h_n : \Omega^{\mathbb{N}} \to \mathbb{R}$ is measurable & $|h_n| \le M$ for $\forall n$; N - stopping time for $\{\sigma(X_0, \dots, X_n)\}_{n=0}^\infty$ Then, $\mathbb{E}_\mu[h_N(X_N, X_{N+1}\dots) \mid \mathcal{F}_N] = \mathbb{E}_{X_N}[h_N(X_0, X_1\dots)]$ on $\{N < \infty\}$.

Proof. Let $A \in \mathcal{F}_N$. $\mathbb{E}_\mu[h_N(X_N, X_{N+1}\dots); A \cap \{N < \infty\}] = \sum_{n=0}^{\infty} \mathbb{E}_\mu[h_n(X_n, X_{n+1}\dots); A \cap \{N = n\}]$. Define $E_n = A \cap \{N = n\}$. We know that E_n is $\mathcal{F}_n$-measurable. By theorem 5.2.5, we have

$$\mathbb{E}_\mu[h_n(X_n, X_{n+1}, \dots); E_n] = \mathbb{E}[\mathbb{E}_{X_n}[h_n(X_0, X_1, \dots)]; E_n].$$

Then we have

$$\sum_{n=0}^{\infty} \mathbb{E}_\mu[h_n(X_n, X_{n+1}\dots); A \cap \{N = n\}] = \sum_{n=0}^{\infty} \mathbb{E}_\mu[\mathbb{E}_{X_n}[h_n(X_0, X_1, \dots)]; A \cap \{N = n\}].$$

Thuw we have

$$\mathbb{E}_\mu[h_N(X_N, X_{N+1}\dots); A \cap \{N < \infty\}] = \mathbb{E}_\mu[\mathbb{E}_{X_n}[h_n(X_0, X_1, \dots)]; A \cap \{N < \infty\}].$$

Because it holds for any $A \in \mathcal{F}_N$ and by definition of conditional expectation 3.1.1, we have

$$\mathbb{E}_\mu[h_N(X_N, X_{N+1}\dots) \mid \mathcal{F}_N] = \mathbb{E}_{X_N}[h_N(X_0, X_1\dots)].$$

□

5.3 Feb 23

by Tatiana (Tanya) Brailovskaya

Warm-up: $T_y^k = \inf\{n > T_y^{k-1} : X_n = y\}, k \geq 1$[2] which means that time of kth visit to y and $T_y^0 = 0$. Let $\rho_{xy} = \mathbb{P}_X(T_y^1 < \infty)$. Show $\mathbb{P}_X(T_y^k < \infty) = \rho_{XY}\rho_{YY}^{k-1}$. Hint: apply strong markov property, and how to choose h?

Proof. $h(\omega) = 1$ where $\omega = (X_0, X_1, \dots)$ if there exists n such that $X_n = y$. Let $N = T_y^{k-1}$

$$\begin{aligned}
\mathbb{P}_x(T_y^k < \infty) &= \mathbb{E}_x\left[h(X_N, X_{N+1}\dots); N < \infty\right] \\
&= \mathbb{E}_x\Big[\underbrace{\mathbb{E}_{X_N = y} h(X_0, X_1 \dots)}_{\rho_{yy}}; N < \infty\Big] \\
&= \rho_{yy}\mathbb{E}_x[N < \infty] \\
&= \rho_{yy}\mathbb{P}_x(T_y^{k-1} < \infty)
\end{aligned}$$

When $k = 1$, $\mathbb{P}_x(T_y^1 < \infty) = \rho_{xy}$.
Suppose the equation holds at $k-1$, we have $\mathbb{P}_x(T_y^{k-1} < \infty) = \rho_{xy}\rho_{yy}^{k-2}$.
At k, we have $\mathbb{P}_x(T_y^k < \infty) = \rho_{yy}\mathbb{P}_x(T_y^{k-1} < \infty) = \rho_{yy}\rho_{xy}\rho_{yy}^{k-2} = \rho_{xy}\rho_{yy}^{k-1}$. □

Definition 5.3.1. We say that state y is recurrent if $\rho_{yy} = 1$ and we call y transient if $\rho_{yy} < 1$.

Keep thinking about the **Warm-up** problem. We need to show $\mathbb{P}_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$. If y is recurrent, $\mathbb{P}_y(T_y^k < \infty) = \rho_{yy}^k = 1$ for all k. Then we have

$$\mathbb{P}_y(X_n = y \text{ i.o.}) = \mathbb{P}_y\left(\bigcap_{k=1}^{\infty}\{T_y^k < \infty\}\right).$$

In particular $\mathbb{E}_y N(y) = \infty$, where $N(y)$ means the number of visits of y.

If y is recurrent, $\mathbb{P}_y(T_y^k < \infty) = \rho_{yy}^k = 1$ for all k. Then we have $\mathbb{P}(X_n = y \text{ i.o. }) = \mathbb{P}_y(\cap_{k=1}^{\infty}\{T_y^k < \infty\}) = 1$. In particular $\mathbb{E}_y N(y) = \infty$ which $N(y)$ is number of visits to y.

Question 5.3.2. If y is transient what is $\mathbb{E}_y N(y)$? Hint: $\mathbb{E}_y N(y) = \sum_{k=1}^{\infty} \mathbb{P}_y\left(N(y) \geq k\right)$, where $N(y) \geq k$ means that $\{T_y^k < \infty\}$.

[2] I am a litte confused here. why should be like $k-1$th?

Proof. We directly have

$$\mathbb{E}_y N(y) = \sum_{k=1}^{\infty} \mathbb{P}_y\left(N(y) \geq k\right) = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < \infty) = \sum_{k=1}^{\infty} \rho_{yy}^k = \frac{\rho_{yy}}{1-\rho_{yy}} < \infty.$$

Thus, if y is transient, $\mathbb{E}_y N(y) < \infty$. □

Theorem 5.3.3. Assume the state space is countable. y is recurrent if and only if $\mathbb{E}_y N(y) = \infty$.

Proof. Step 1: The Probability of the k-th Visit. From the Warm-up problem, we know that starting from state x, the probability of visiting state y for the k-th time is $\mathbb{P}_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$. If we start the chain already at state y (so $x = y$), this formula perfectly simplifies to

$$\mathbb{P}_y(T_y^k < \infty) = \rho_{yy}\rho_{yy}^{k-1} = \rho_{yy}^k$$

Since the definition of a recurrent state is $\rho_{yy} = 1$, we substitute it into the equation

$$\mathbb{P}_y(T_y^k < \infty) = 1^k = 1$$

This means that for a recurrent state, the probability of returning to it for the 1st time, the 2nd time, the 100th time, or the k-th time is always exactly 1 (100 % certainty).

Step 2: Translating "Infinitely Often" (i.o.) into Set Theory. The phrase $X_n = y$ i.o. means the Markov chain visits state y infinitely many times. If a chain visits state y infinitely many times, it means for every positive integer k, the k-th visit must happen. Therefore, the event of visiting y infinitely often is exactly the intersection of all the events of the k-th visits happening:

$$\{X_n = y \text{ i.o.}\} = \bigcap_{k=1}^{\infty} \{T_y^k < \infty\}$$

Step 3: Calculating the Probability of the Intersection. Notice that the events $\{T_y^k < \infty\}$ are a nested, decreasing sequence. In plain English: if you visit state y for the $(k+1)$-th time, you must have already visited it for the k-th time. Therefore:

$$\{T_y^1 < \infty\} \supset \{T_y^2 < \infty\} \supset \{T_y^3 < \infty\} \supset \dots$$

Because these events are decreasing, we can use the Continuity of Probability to pull the limit outside the probability measure

$$\mathbb{P}_y\left(\bigcap_{k=1}^{\infty} \{T_y^k < \infty\}\right) = \lim_{k\to\infty} \mathbb{P}_y(T_y^k < \infty)$$

From Step 1, we know $\mathbb{P}_y(T_y^k < \infty) = 1$ for all k. Thus:

$$\mathbb{P}_y(X_n = y \text{ i.o.}) = \lim_{k\to\infty} 1 = 1$$

(This formally proves Lemma: A state is recurrent if and only if the probability of visiting it infinitely often is 1).

Step 4: Formalizing $N(y)$ (The Number of Visits). Let $N(y)$ be the total number of visits to state y at times $n \geq 1$. Mathematically, this is the sum of indicator functions for every time step n:

$$N(y) = \sum_{n=1}^{\infty} 1_{\{X_n = y\}}$$

However, there is a second, brilliant way to count the total number of visits. Instead of checking every time step n, we can just check how many k-th visits were successfully completed:

$$N(y) = \sum_{k=1}^{\infty} 1_{\{T_y^k < \infty\}}$$

Both sums will yield the exact same integer.

Step 5: Computing the Expectation $\mathbb{E}_y[N(y)]$. Now we take the expected value of the second definition of $N(y)$

$$\mathbb{E}_y[N(y)] = \mathbb{E}_y\left[\sum_{k=1}^{\infty} 1_{\{T_y^k < \infty\}}\right].$$

By the Monotone Convergence Theorem, we can safely swap the expectation and the infinite sum for non-negative random variables

$$\mathbb{E}_y[N(y)] = \sum_{k=1}^{\infty} \mathbb{E}_y \left[1_{\{T_y^k < \infty\}} \right].$$

The expected value of an indicator function is simply the probability of the event, so

$$\mathbb{E}_y[N(y)] = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < \infty) = \sum_{k=1}^{\infty} \rho_{yy}^k.$$

Since state y is recurrent, we substitute $\rho_{yy} = 1$

$$\mathbb{E}_y[N(y)] = \sum_{k=1}^{\infty} 1^k = 1 + 1 + 1 + \cdots = \infty$$

This concludes the proof. For a recurrent state, because the chance of returning is perfectly 1, the sum of these probabilities is an infinite series of 1s, making the expected number of visits exactly ∞. □

Theorem 5.3.4 (Recurrenece is contagion)**.** If X is recurrent & $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$.

Proof. Goal: Show $\rho_{yx} = 1$. Why $\rho_{yx} = 1 \implies y$ is recurrent? Use theorem 5.3.3[a].

$$\mathbb{E}_y N(y) = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < \infty) \geq \underset{\overset{\shortparallel}{1}}{\rho_{yx}} \underbrace{\left(\sum_{k=1}^{\infty} p_{xx}^k \right)}_{\infty} \underbrace{\rho_{xy}}_{>0} = \infty,$$

in fact, sufficient to show $\rho_{yx} > 0$. Next: proof of goal (Contradiction): if $\rho_{xy} > 0$ and $\rho_{yx} < 1 \implies \rho_{xx} < 1$.

Let $K = \inf \left\{ k : p^k(x,y) > 0 \right\}$, and $p^k(x,y)$ means that k-step transition probability. Then, $\exists y_1 \dots y_{K-1}$ s.t.

$$p^K(x,y) \geq p(x,y_1)p(y_1,y_2)\dots p(y_{K-1},y) > 0$$

$$\mathbb{P}_x(T_x^1 = \infty) \geq p(x,y_1)\dots p(y_{K-1},y)(1-\rho_{yx}) > 0.$$

□

[a] This is "Chapman-Kolmogorov equation (Theorem 5.2.4)". We choose the specific way to achieve. Then the probabilty of this specific way should be smaller than the total probability. Since $\rho_{xy} = 1$, there must be some number of steps L such that $p^L(y,x) > 0$. We also know there is a path from x to y of length K, so $p^K(x,y) > 0$. If the chain starts at y and we want it to be at y exactly $L + n + K$ steps later, one specific way to achieve this is: Go from y to x in L steps; Go from x to x in n steps; Go from x to y in K steps. Thus we have

$$p^{L+n+K}(y,y) \leq p^L(y,x)p^n(x,x)p^K(x,y) \implies \sum_{n=1}^{\infty} p^{L+n+K}(y,y) \leq p^L(y,x) \sum_{n=1}^{\infty} p^n(x,x)p^K(x,y) = \infty.$$

Definition 5.3.5. C is closed if for any $y \in C$, $\mathbb{P}_y(X_n \in C) = 1$ for all n. D is irreducible if $x, y \in D \implies \rho_{xy} > 0$.

Theorem 5.3.6. Let C be a finite closed set. Then C contains a recurrent state. If C is also irreducible then $\forall x \in C$ are recurrent

Proof. If C is irreducible & contains a recurrent state, by theorem before all $y \in C$ are recurrent.

$N(C) = \#$ of visits to C and $\mathbb{E}_x N(C) = \sum_{y \in C} \mathbb{E}_x N(c) y = \infty$, for $x \in C$ (by definition of being closed). Since C is finite, $\exists y$ s.t. $\mathbb{E}_x N(y) = \infty$. Thus $\mathbb{E}_x N(y) = \rho_{xy} \sum_{k=0}^{\infty} \rho_{yy}^k = \infty \implies \rho_{xy} > 0$ and y is recurrent. □

Example 5.3.7.

Theorem 5.3.8. Let R denote the set of recurrent states. Then, $R = \bigcup_i R_i$ where R_i is closed & irreducible for $\forall i$.

Proof. For $\forall x \in R$, set $C_x = \{y : \rho_{xy} > 0\}$. By Thm 2, $\rho_{yx} = 1 \implies C_x$ is closed & irreducible.Can show: for $\forall x, y \in C$

$$C_x = C_y \quad \text{or} \quad C_x \cap C_y = \emptyset.$$

□

Observation 5.3.9. If MC is ∞ & irreducible $\implies \forall$ states are recurrent or $\forall$ states are transient.

Example 5.3.10. Customers arrive according to a Poisson process with rate λ.

Summary of Equivalences.
For a Recurrent State y, The Following Are Equivalent (TFAE):

1. $\rho_{yy} = 1$ (Guaranteed to return).
2. $P_y(X_n = y \text{ i.o.}) = 1$ (Visits y infinitely many times).
3. $E_y[N(y)] = \infty$ (Expected number of visits is infinite).
4. $\sum_{n=1}^{\infty} p_n(y, y) = \infty$ (The sum of n-step transition probabilities diverges).

For a Transient State y, The Following Are Equivalent (TFAE):

1. $\rho_{yy} < 1$ (Chance of never returning is > 0).
2. $P_y(X_n = y \text{ i.o.}) = 0$ (Visits y only finitely many times).
3. $E_y[N(y)] < \infty$ (Expected number of visits is finite).
4. $\sum_{n=1}^{\infty} p_n(y, y) < \infty$ (The sum of n-step transition probabilities converges).

5.4 Feb 25

by Tatiana (Tanya) Brailovskaya
MC that is ∞ & irreducible either has all recurrent or all transient states.
Recall: M/G/1 queuing

- Customers arrive according to Poisson(λ) process, service for time $\sim F$
- X_n is the number of customers waiting in line when nth customer arrives

$a_k = \mathbb{P}(k \text{ customers arrive in a single service time})$

Claim: let $\mu = \sum_k a_k k$.

Warm-up: If $\mu > 1 \implies$ MC is transient[3]. If $\mu \leq 1 \implies$ MC is recurrent. What's the intuition behind Claim?

Proof. Let $Y_1, Y_2 \ldots$ iid with $\mathbb{P}(Y_i = j) = a_{j+1}$. $S_n = Y_1 + \cdots + Y_n$. We can observe that $X_n = X_0 + S_n$ until $N = \inf\{n > 0 : X_0 + S_n = 0\}$ when $X_0 > 0$.
Goal (i): $\mu > 1 \implies \mathbb{P}_x(N < \infty) < 1$.

$$S_n \longrightarrow \infty \text{ a.s.} \implies \inf S_n > -\infty \text{ a.s., in particular } \implies \exists x > 0.$$

Because $\frac{S_n}{n} \to \mu - 1 > 0$ a.s., we know $S_n \longrightarrow \infty$. Then we have

$$\mathbb{P}(\inf S_n > -x) \geq \delta > 0 \implies \mathbb{P}_x(N < \infty) = \mathbb{P}(\inf S_n \leq -x) \leq 1 - \delta < 1.$$

[3] it means that the line will never goes back to state 0.

Therefore, we have

$$\underbrace{\mathbb{P}_0(N<\infty)}_{\rho_{00}} = \sum_x p_{0x}\mathbb{P}_x(N<\infty) < 1.$$

Goal (2): $\mu \le 1 \implies \mathbb{P}_0(N<\infty) = 1$.

Observation: $X_{n\wedge N}$ is a super-MG & $X_{n\wedge N} \ge 0$. Let $T_M = \inf\{n : X_{n\wedge N} \ge M\}$. Optional Stopping: $\forall M > 0 : \mathbb{E}X_0 \ge \mathbb{E}X_{T_M\wedge N} \ge M\mathbb{P}_{X_0}(T_M < N) \implies \forall x \ne 0 : \mathbb{P}_x(T_M < N) \le x/M \longrightarrow 0$ as $M \to \infty$. Can check: $N \wedge T_M < \infty\ \forall M \implies \mathbb{P}_x(N<\infty) = 1$.

Conclude: $\mathbb{P}_0(N<\infty) = \sum_x p_{0x}\mathbb{P}_x(N<\infty) = 1$. □

Theorem 5.4.1 (Foster's Theorem)**.** Suppose S is irreducible & $\varphi \ge 0$ with $\mathbb{E}_x\varphi(X_1) \le \varphi(x)$ for $\forall x \notin F$ a finite set, and $\varphi(x) \to \infty$ as $x \to \infty$. Then the chain is recurrent.

Question 5.4.2. Q: For M/G/1 queue, what is φ & F?

$\varphi(x) = x$ $(x \to \infty)$, $F = \{0\}$, $\mathbb{E}_xX_1 = \mathbb{E}[(x+Y_1)^+] = \mathbb{E}(x+Y_1) = x + (\mu - 1) \le x$, $x > 0$ & $Y_1 \ge -1$.

Let $N_n(y) = \sum_{m=1}^n \mathbb{1}(X_m = y)$, where $N_n(y)$ is the number of visits to y satisfying at $X_0 = x$ by time n.

Theorem 5.4.3. Let $N_n(y) = \sum_{m=1}^n \mathbb{1}(X_m = y)$, where $N_n(y)$ is the number of visits to y satisfying at $X_0 = x$ by time n. For y recurrent,

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_yT_y}\mathbb{1}_{\{T_y<\infty\}} \quad \text{a.s. } \mathbb{P}_x,$$

T_y is the 1st visit to y after $n = 0$ (with $1/\infty = 0$).

Proof. First, $x = y$. Let $R(k) = \min\{n \ge 1 : N_n(y) = k\}$ where $R(k)$ is the kth return to y.

$t_k = R(k) - R(k-1)$, let $R(0) = 0 \implies t_1, t_2, \dots$ iid r.v.

$$\frac{R(k)}{k} = \frac{\sum_{l=1}^k t_l}{k} \implies \mathbb{E}[t_1] = \mathbb{E}_yT_y \text{ a.s. (by SLLN)}$$

and

$$\frac{R(N_n(y))}{N_n(y)} \le \frac{n}{N_n(y)} \le \frac{R(N_n(y)+1)}{N_n(y)+1} \times \frac{N_n(y)+1}{N_n(y)} \implies \frac{n}{N_n(y)} \to \mathbb{E}_yT_y \quad \mathbb{P}_y\text{-a.s.}$$

where the factor $\frac{N_n(y)+1}{N_n(y)} \to 1$ because y is recurrent $\implies N_n(y) \to \infty$ a.s.

On $\{T_y < \infty\}$[a]: $t_2, t_3, \dots$ iid

$$\frac{R(k)}{k} = \frac{t_1}{k} + \frac{(t_2 + \cdots + t_k)}{k} \to 0 + \mathbb{E}_yT_y \quad \mathbb{P}_x\text{-a.s.}$$

Repeat the previous argument to show

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_yT_y} \quad \mathbb{P}_x\text{-a.s. on } \{T_y < \infty\}.$$

Q: What happens on $\{T_y = \infty\}$? $N_n(y) = 0\ \forall n$. □

[a] it means that we eventually hit y.

Definition 5.4.4. For a recurrently state y, we say y is null recurrent if $\mathbb{E}_yT_y = \infty \implies N_n(y)/n \to 0$; we say y is positive recurrent if $\mathbb{E}_yT_y < \infty \implies N_n(y)/n \to p > 0$.

Definition 5.4.5. μ is a stationary measure where $\mathbb{P}_\mu(X_1 = y) = \mu(y)$, i.e.

$$\sum_x \mu(x)p(x,y) = \mu(y).$$

If μ is a probability measure we call it stationary distribution.

Question 5.4.6. Can you give an example of a stationary measure that isn't a distribution?

Example 5.4.7. Random walk on $\mathbb{Z}^d$, $p(x,y) = f(x-y)$ and $f(z) \geq 0$ where $\sum_z f(z) = 1 \implies \mu(x) \equiv 1$ is a stationary measure

$$\sum_x p(x,y)\mu(x) = \sum_z f(z) = 1 = \mu(y).$$

Example 5.4.8. Random walk on $\mathbb{Z}$ & $p(x,x+1) = p$, $p(x,x-1) = 1-p = q \implies \mu(x) \equiv 1$ is a stationary measure.

Also: $\mu(x) = (p/q)^x = (q/p)^{-x}$ is also a stationary measure $\sum_x \mu(x) = \sum_x (p/q)^x = \infty$.

Example 5.4.9. Modify (1) or (2) s.t. you have a stationary distribution. Walk on a cycle of len N, $p(x,x+1) = p(x,x-1) = 1/2$ where the movement follows $x+1 \pmod N$ and $x-1 \pmod N)\mu(x) \equiv 1/N$ is a stationary dist.

$$\sum_x \mu(x)p(x,y) = \frac{1}{N}(p(y-1,y) + p(y+1,y)) = \frac{1}{N} = \mu(y)$$

5.5 Mar 2

by Tatiana (Tanya) Brailovskaya

Recall: $\mu(x)$ is a stationary measure if

$$\sum_x \mu(x)p(x,y) = \mu(y).$$

$\mu(x)$ isn't necessarily a probability measure.

Warm-up: Consider the birth and death $S = \{0,1,2,3,\ldots\}$. $p(x,x+1) = p_x, p(x,x) = r_x, p(x,x-1) = q_x$ with $q_0 = 0$ and $p(i,j) = 0$ otherwise. Let $\mu_1(x) = 1 \;\; \forall x.$ and $\mu_2(x) = \prod_{k=1}^{x} \frac{p_{k-1}}{q_k}$ for all x. For what choice of p_x, r_x, q_x is μ_1, μ_2 stationary measure?

Proof.

$$\sum_x \mu_1(x)p(x,y) = p_{y-1} + r_y + q_{y+1} = \mu_1(y) = 1$$

where $p_{y-1} + r_y + q_{y+1} = 1$ is the condition for stationarity.

$$\begin{aligned}
\sum_x \mu_2(x)p(x,y) &= \prod_{k=1}^{y-1}\left(\frac{p_{k-1}}{q_k}\right)p_{y-1} + \prod_{k=1}^{y}\left(\frac{p_{k-1}}{q_k}\right)r_y + \prod_{k=1}^{y+1}\left(\frac{p_{k-1}}{q_k}\right)q_{y+1} \\
&= \prod_{k=1}^{y-1}\left(\frac{p_{k-1}}{q_k}\right)\left[p_{y-1} + \left(\frac{p_{y-1}}{q_y}\right)r_y + \left(\frac{p_{y-1}}{q_y}\right)p_y\right] \\
&= \prod_{k=1}^{y}\frac{p_{k-1}}{q_k} \\
&= \mu_2(y)
\end{aligned}$$

implies $\mu_2(y)$ always stationary. □

Definition 5.5.1. We say that μ satisfies detialed balance if $\mu(x)p(x,y) = \mu(y)p(y,x)$ for any y. We call that μ reversible.

Question 5.5.2. why if μ is reversible then μ is a stationary measure?

Proof.

$$\sum_x \mu(x)p(x,y) = \sum_x \mu(x)p(y,x) = \mu(y)\sum_x p(y,x) = \mu(y).$$

□

Claim 5.5.3. μ_2 satisfies detailed balance[4].

Proof. We just need to check $\mu_2(x)p(x,y) = \mu_2(x+1)p(x+1,y)$. Substituting the definition of μ_2, LHS is

$$\prod_{k=1}^{x}\left(\frac{p_{k-1}}{q_k}\right)p_x.$$

We consider the flow from $x+1$ down to x - RHS $= \mu_2(x+1)q_{x+1}$. And we have

$$\mu_2(x+1) = \prod_{k=1}^{x}\left(\frac{p_{k-1}}{q_k}\right)\cdot\frac{p_x}{q_{x+1}}.$$

Thus, we have

$$\mu_2(x)\left(\frac{p_x}{q_{x+1}}\right)q_{x+1} = \text{LHS}.$$

□

Theorem 5.5.4. Suppose μ is stationary distribution with $\mu(X) > 0$ for all x and $X_0 \sim \mu$. Then[5], $Y_m = X_{n-m}, 0 \le m \le n$ is a markov chain with $Y_0 \sim \mu$, and transition probability $q(x,y) = \frac{\mu(y)p(y,x)}{\mu(x)}$. If moreover μ is reversible, $q = p$.

Proof.

$$\mathbb{P}(Y_{m+1} = y \mid Y_m = x) = \mathbb{P}(X_{n-m-1} = y \mid X_{n-m} = x) = \frac{\mathbb{P}(X_{n-m-1} = y \cap X_{n-m} = x)}{\mathbb{P}(X_{n-m} = x)}$$
$$= \frac{\mathbb{P}(X_{n-m} = x \mid X_{n-m-1} = y)\mathbb{P}(X_{n-m-1} = y)}{\mu(x)} = p(y,x)\frac{\mu(y)}{\mu(x)}$$

where it remains to check the Markov Property. □

Example 5.5.5. Caveat: most chains don't have a reversible measure.E.g. $M/G/1$ - queue: X_n is the number of customers arriving during the nth customer's service time.

$$p(x, x+10) > 0$$

where this means 11 customers arrived during a service time. But $p(x+10, x) = 0$ because the queue can shrink by 1 customer at a time.

$$\mu(x)p(x,x+10) = \mu(x+10)p(x+10,x) = 0 \quad \forall x$$

where the term $\mu(x)p(x,x+10)$ is greater than 0 for some x, implying there is no reversible measure.

Theorem 5.5.6 (Kolmogorov Cycle Condition)**.** Suppose p is a transition probability for an irreducible chain. This chain has a reversible measure[6] iff

1. $p(x,y) > 0$ then $p(y,x) > 0$.

[4]A Markov chain is reversible with respect to a measure π if and only if it satisfies the detailed balance equations for all states x and y $\pi(x)p(x,y) = \pi(y)p(y,x)$.
[5]this is the backward MC
[6]This is $\mu(x)p(x,y) = \mu(y)p(y,x)$ for $\forall x, y$

2. for any cycle $x_0, x_1, \dots, x_n = x_0$ with $\prod_{1\le i\le n} p(x_i, x_{i-1}) > 0$ then we have $\prod_{i=1}^n \frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})} = 1$.

Theorem 5.5.7. Let X be a recurrent state and let $T = \inf\{n \ge 1 : X_n = x\}$,

$$\mu_x(y) = \mathbb{E}_x\Big(\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n=y\}}\Big) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T > n)$$

is a stationary measure.

Remark 5.5.8. $\mu_x(y) < \infty$ for all y. $\mu_x P = \mu_x \implies \mu_x P^n = \mu_x$ for $\forall n \ge 1 . \forall n \ge 1$

$$\sum_y \mu_x(y) p^n(y,x) = \mu_x(x) = 1 \implies \exists n \text{ s.t. } p^n(y,x) > 0 \implies \mu_x(y) < \infty$$

where $p^n(y,x) > 0$ is equivalent to the condition $\rho_{yx} > 0$. Suppose $\rho_{yx} = 0$. Then, Case 1: $\rho_{xy} > 0 . x$ is recurrent $\implies \rho_{yx} = 1$, which leads to a contradiction.

Case 2: $\rho_{xy} = 0 \implies \mu_x(y) = 0$.

Proof. Sketch

$$\mu_x(y) = \mathbb{E}[\# \text{ of visits to } y \text{ on } n = 0, 1, \dots, T-1]$$

$$(\mu_x P)(y) \overset{(\Delta)}{=} \mathbb{E}_x[\# \text{ of visits to } y \text{ on } n = 1, 2, \dots, T]$$

where $(\mu_x P)(y) = \sum_z \mu_x(z) p(z,y)$.

$$\mu_x(y) = (\mu_x P)(y) \text{ b/c } X_T = X_0 = x.$$

Need to rigorously check (Δ). Let $\bar{p}_n(x,y) = \mathbb{P}_x(X_n = y, T > n)$, i.e. $\mu_x(y) = \sum_{n=0}^{\infty} \bar{p}_n(x,y)$.

$$\sum_z \mu_x(z) p(z,y) = \sum_{n=0}^{\infty} \sum_y \bar{p}_n(x,y) p(y,z)$$

where we want to show (wts) $\mathbb{P}_x(X_{n+1} = z, T > n+1)$ for $z \ne x$.

Case 1: $z \ne x$,

$$\sum_y \bar{p}_n(x,y) p(y,z) = \sum_y \mathbb{P}_x(X_n = y, T > n, X_{n+1} = z) = \mathbb{P}_x(X_{n+1} = z, T > n+1)$$

where if $X_{n+1} = z$ and $T > n$, then $T > n+1$ because $X_{n+1} \ne x$.

Case 2: $z = x$,

$$\sum_y \bar{p}_n(x,y) p(y,z) = \sum_y \mathbb{P}_x(X_n = y, T > n, X_{n+1} = x) = \mathbb{P}_x(T = n+1)$$

$$\implies \sum_z \mu_x(z) p(z,y) = \sum_{n=0}^{\infty} \mathbb{P}_x(T = n+1) = 1 = \mu_x(x),$$

where we use the property of recurrence to show the sum equals 1. □

5.6 Mar 4

tby Tatiana (Tanya) Brailovskaya

Midterm: Wednesday March 18. Topics: HW 1-4, including Chapter 4, 5.1, 5.2. Can bring cheat sheet, A4, handwritten double-sided.

Recall: for a Markov Chain with recurrent state x, then $T = \inf\{n \ge 1 : X_n = x\}$, then

$$\mu_x(y) = \mathbb{E}_x\Big(\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n=y\}}\Big) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T > n)$$

is a stationary measure.

Warm-up: Consider the renewal chain[7] on $S = \{0, 1, 2, 3, \dots\}$. X_n is the amount of time until the next renewal. What is a recurrent state for this Markov Chain? – 0. Becasue $\mathbb{P}_0(T_0 < \infty) = 0$.

Write down $\mu_0(y)$ for this Markov Chain.

$$\mu_0(y) = \sum_{j=y-1}^{\infty} f_{j+1} = \mathbb{E}_0\left[\sum_{n=0}^{T-1} \mathbb{1}_{X_n=y}\right] = \mathbb{P}_0(X_1 \geq y)$$

$$\mu_0(y) = \sum_{n=0}^{\infty} \mathbb{P}_0(X_{n+1} = y)$$

$$\mathbb{P}_0(X_{n+1} = y) = p(0, y+n) = f_{y+n+1}$$

$$\mu_0(y) = \sum_{n=0}^{\infty} f_{y+n+1} = f_{y+1} + f_{y+2} + f_{y+3} + \dots$$

Is $\mu_0(y)$ a stationary distribution?

$$\sum_y \mu_0(y) = 1 + \sum_{y=1}^{\infty} \mathbb{P}_0(X_1 \geq y)$$

if $\mathbb{E}_0 X_1 < \infty$, $\mu_0'(y) = \mathbb{P}_0(X_1 \geq y)$ which is a stationary distribution. Recall, theorem before: If ρ is irreducible and recurrent, its stationary measure is unique with probabiltiy to a scalar multiple.

Recall the theorem before, if p is irreducible and recurrent, its stationary measure is unique up to a scaler mutipli

For which $\{f_i\}$ is the unique stationary measure up to ? when is this MC irreducible? $f_i > 0$ i.o. iff that is no $k < \infty$ s.t. for $j \geq k$ $f_j = 0$.

Today we are going to talk about the property of stationary distributions - convergence theorem.

Theorem 5.6.1. If p is irreducible TFAE

1. some x is possible reccurent
2. there is a stationary distribution
3. all states are positive reccurent

Proof. (i) → (ii)

$$\mu_x(y) = \frac{\sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n)}{\mathbb{E}_x T_x} \text{ is stationary + prob. dist.}$$

(ii) → (iii)

$$\pi(y) = \frac{1}{\mathbb{E}_y T_y} \text{ (thm 5.6.4)}$$

irreducibility + existence of stat. dist.
$\implies$ all states are recurrent $\implies \pi(y) > 0$ for $\forall y \implies \mathbb{E}_y T_y < \infty$ □

Recall: recurrent states x is positive recurrent if $\mathbb{E}_x T_x < \infty$ and null recurrent if $\mathbb{E}_x T_x = \infty$.

Corollary 5.6.2. if p irreducible, recurrent with stationary measure μ with $\sum_x \mu(x) = \infty$ then all stationary are null recurrent.

[7] $p(0, j) = f_{j+1}$ for $j \geq 0$, $p(i, i-1) = 1$ for $i > 0$, $p(i, j) = 0$

Example 5.6.3. [8] Consider the random walk on $\mathbb{Z}^d$. for $d = 1, 2$ random walk is recurrent. $\mu(x) = 1$ is a stationary measure. Then all $x \in \mathbb{Z}, \mathbb{Z}^2$ are null-recurrent.

By theorem , the random walk on $\mathbb{Z}^d$ doesn't have a stationary distribution.

1. $d = 1, 2$ by theorem 1

2. $d \geq 3$ random walk is transiant by theorem 2.

Theorem 5.6.4. If there is a stationary distribution π an $\pi(y) > 0$ then y is recurrent.

Proof. We will directly $\rho_{yy} = 1$.

Recall

$$\mathbb{E}_x N(y) = \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k) = \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) = \sum_{k}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \frac{\rho_{xy}}{1-\rho_{yy}}.$$

And, we have

$$\mathbb{E}_X N(y) = \mathbb{E}\left[\sum_{n=0]}^{\infty} \mathbb{1}(X_n = y \mid X_0 = x)\right] = \sum_{n=0}^{\infty} \mathbb{P}_X(X_n = y) = \sum_{n=0}^{\infty} \rho^n(x, y).$$

Based on the image provided, here is the LaTeX transcription of the proof on the whiteboard.

Since π is stationary, $(\pi P^n)(y) = \pi(y) \quad \forall n$. Then we have

$$\sum_{n=1}^{\infty}\left(\sum_x \pi(x)P^n(x,y)\right) = \sum_n \pi(y) = \infty, \quad \sum_x \pi(x) \underbrace{\sum_{n=1}^{\infty} P^n(x,y)}_{\mathbb{E}_x N(y)} = \sum_x \overbrace{\pi(x)\rho_{xy}}^{\leq 1}\left(\frac{1}{1-\rho_{yy}}\right) \implies \rho_{yy} = 1.$$

□

Theorem 5.6.5. If p is irreducible and has stationary distribution π, then $\pi(x) = \frac{1}{\mathbb{E}_x T_x}$, and $y_\infty = 0$.

Recall: We expect $\frac{N_n(y)}{n} \approx \pi(y)$ for large n, where $N_n(y)$ is the # of visits to y by n.
We showed that $\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y}$ a.s. when y is recurrent.

Question 5.6.6. Q: if y is transient, $\lim_n \frac{N_n(y)}{n}$?

$$\lim_n \frac{N_n(y)}{n} = 0 \quad \text{a.s.}$$

because transience $\implies$ MC returns to y finitely often.
$\mathbb{E}_y T_y = \infty$ because $\mathbb{P}_y(T = \infty) > 0$.

Proof. If there is a stationary dist., then $\exists x$ w $\pi(x) > 0 \implies x$ is recurrent $\implies$ all states are recurrent

Irreducible + recurrent $\implies$ stationary measure unique up to scaling

$$\mu_x(y) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n), \quad \sum_y \mu_x(y) = \sum_{n=0}^{\infty}\sum_y \mathbb{P}_x(X_n = y, T_x > n) = \sum_{n=0}^{\infty} \mathbb{P}_x(T_x > n) = \mathbb{E}_x T_x < \infty.$$

Thus, we have $\pi(y) = \frac{\mu_X(y)}{\mathbb{E}_x T_x}$, in particular $\pi(x) = \frac{1}{\mathbb{E}_x T_x}$. □

[8] I don't understand here.

5.7 Mar 16

by Tatiana (Tanya) Brailovskaya

Recall the fisrt theorem in the last course.

Thm 1: P is irreducible + recurrent $\implies$ stationary measure is unique up to constant multiples.

Thm 2: P is irreducible TFAE (i) some x is positive recurrent

(ii) there is a stationary dist.

(iii) All states are positive recurrent

Warum-up: birth date

$S = \{0, 1, 2, \dots\}$, $p(x, x+1) = p, p(x, x) = r_x, p(x, x-1) = q_x$. $\mu(x) = \cap_{k=1}^{x} \frac{p_{k-1}}{q_k}$ is stationary measure. if $\sum_x \mu(x) < \infty$, then there exists a stationary distribution $\widehat{\mu}(x) = \frac{1}{\sum_x \mu(x)} \mu(x)$. When is $\widehat{\mu}$ the unique stationary distribution?

- make sure P is irreducible, e.g. $p_x, q_x > 0 \quad \forall x$.
- Thm 2 + existence of $\widehat{\mu} \implies P$ is positive recurrent
- recurrence + irreducibility $\implies \widehat{\mu}$ is unique

Thm 3: P is irreducible w stat. dist $\mu \implies \mu$ is the unique stat. dist.

Remark 5.7.1. irreducible + existence of π, then π is unique.

Example 5.7.2. if a markov chaine ahs a collection A of absorbing states with $|A| > 1$, there are multiple stationary distribution for $t \in A$:

$$\pi_y(x) = \begin{cases} 1, & y = x \\ 0 & y \neq x \end{cases}.$$

- We saw that for many (e.g. Wright-Fisher) MC's converge to a π_y
- Say in WF: $x \neq 0, N$. $p^n(x, y) \to 0$ for $y \neq 0, N$ and $p^n(x, 0) \to 1 - \frac{x}{N}$. $p^n(x, N) \to x/N$. Also π with $\pi(0) = 1 - \frac{x}{N}$, $\pi(N) = x/N$, $\pi(y) = 0$ for $y \neq 0, N$ is not a stationary measure.

Definition 5.7.3. Let x be a recurrent state and $I_x = \{n \geq 1, p^n(x, x) > 0\}$. Let $d_X = \gcd(I_x)$ where d_x is the periodic of x.

Example 5.7.4.

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad P^3 = P \quad P^4 = P^2 \dots$$

i.e. $d_1 = 2$, $d_2 = 2$

$$P^n(x, x) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even} \end{cases}$$

i.e. P^n doesn't converge!

Definition 5.7.5. We say that p is aperiodic if $d_x = 1$ for $\forall\ x \in S$.

Lemma 5.7.6. Assume x & y are recurrent. If $\rho_{xy} > 0$ then $d_y = d_x$.

Proof. First show that dy/dx.

- Note. x is recurrent & $\rho_{xy} > 0 \implies \rho_{yx} = 1 > 0$. Then there exists $K_1,\ L$ s.t. $p^K(x, y) > 0$ and $p^L(y, x) > 0$. Then we have

$$p^{K+L}(y, y) \geq p^L(y, x) p^K(x, y) > 0 \implies dy/(K + L).$$

$$P^{K+n+L}(y, y) \geq P^L(y, x) P^n(x, x) P^K(x, y) > 0 \quad \text{for any } n \text{ s.t. } P^n(x, x) > 0$$

$$\implies d_y \mid K + n + L \implies d_y \mid n \implies d_y \mid d_x.$$

□

Corollary 5.7.7. If p is irreducible and recurrent, all states have the same period d. In particular, if for some x, $dx = 1$ then the chain is aperiodic.

Example 5.7.8.

Lemma 5.7.9. If $dx = 1$, then $p^m(x, x) > 0$ for $m > m_0$.

Theorem 5.7.10. If p is irreducible, aperiodic with a stationary distribution π then for $\forall\ x, y \in S$ $p^n(x, y) \to \pi(y)$ as $n \to \infty$.

Proof. Idea: "Coupling":

- Let X_n & Y_n denote 2 independent copies of MC with $X_0 = x$, $Y_0 \sim \pi$
- Let T be the first time $X_n = Y_n$
- For all $n \geq T$, Set $Y_n = X_n$

Then (X_n, Y_n) forms a MC on $S \times S$ with transition probability q:

$$q((x_1, y_1), (x_2, y_2)) = \begin{cases} p(x_1, x_2)p(y_1, y_2) & \text{if } x_1 \neq y_1 \\ p(x_1, x_2) & \text{if } x_1 = y_1, x_2 = y_2 \\ 0 & \text{o/wise} \end{cases}$$

Obs: for $n \geq T$, $X_n \sim Y_n$, i.e. $X_n \sim \pi$
We will show: (1) $\sum_y |\underbrace{\mathbb{P}(X_n = y)}_{P^n(x,y)} - \underbrace{\mathbb{P}(Y_n = y)}_{\pi(y)}| \leq 2\mathbb{P}(T > n)$
(2) $\lim_{n\to\infty} \mathbb{P}(T > n) = 0$
$\sum_y |P^n(x, y) - \pi(y)| \to 0 \implies P^n(x, y) \to \pi(y)$
(1):

$$\begin{aligned} \mathbb{P}(X_n = y) &= \mathbb{P}(Y_n = y, T \leq n) + \mathbb{P}(X_n = y, T > n) \\ &\leq \mathbb{P}(Y_n = y) + \mathbb{P}(X_n = y, T > n) \end{aligned}$$

Same arg shows:

$$\mathbb{P}(Y_n = y) \leq \mathbb{P}(X_n = y) + \mathbb{P}(Y_n = y, T > n)$$

$$\sum_y |\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)| \leq \sum_y (\mathbb{P}(X_n = y, T > n) + \mathbb{P}(Y_n = y, T > n)) = 2\mathbb{P}(T > n)$$

We will show: (1) $\sum_y |\underbrace{\mathbb{P}(X_n = y)}_{P^n(x,y)} - \underbrace{\mathbb{P}(Y_n = y)}_{\pi(y)}| \leq 2\mathbb{P}(T > n)$
(2) $\lim_{n\to\infty} \mathbb{P}(T > n) = 0$
$\implies \sum_y |P^n(x, y) - \pi(y)| \to 0 \implies P^n(x, y) \to \pi(y)$.
Pf of (2): P is irreducible + recurrent $\xrightarrow{\text{this is where we need aperiodicity}}$ q is irreducible + recurrent
$T_{(y,y)} < \infty$ a.s. b/c q is recurrent, $T = \min_y T_{(y,y)} \implies T < \infty$ a.s. $\iff \mathbb{P}(T > n) \xrightarrow[n\to\infty]{} 0$
you can check that $\widehat{\pi}((i, j)) = \pi(i)\pi(j)$ is a stat. dist. for q □

5.8 Mar 23

by Tatiana (Tanya) Brailovskaya

Recall 5.7.10,

Warm-up Ehrenfest claim. $X_n = \#$ of particle at time n. $S = \{0, 1, \ldots, r\}$. The transition probabiltiy

$$p(k, k+1) = \frac{r-k}{r}, \qquad p(k, k-1) = \frac{k}{r}.$$

(1) Is p irreducible? Yes. Is p aperiodic?

(2) Does p have a stationary distribution? Fact:

$$p^{2n}(x,y) \to \binom{r}{y}\left(\frac{1}{2}\right)^{r-1} \quad \text{if } x,y \text{ have same parity}$$

$$p^{2n+1}(x,y) \to \binom{r}{y}\left(\frac{1}{2}\right)^{r-1} \quad \text{if } x,y \text{ have diff. parity}$$

(3) Does p have a stationary dist? Yes!

$$\mu(x) = \binom{r}{x}\left(\frac{1}{2}\right)^{r}.$$

Intuition: at stationarity, each particle is equally likely to be in (1) or (2), indep. of others

Why? Detailed balance condition. Recursible measure. $\forall x,y: \quad \mu(x)p(x,y) = \mu(y)p(y,x)$

WTS:

$$\mu(x)p(x,x+1) = \mu(x+1)p(x+1,x).$$

Consider the modified Ehrenfest chain:

$$p(k,k) = \frac{1}{2}, \quad p(k,k-1) = \frac{k}{2r}, \quad p(k,k+1) = \frac{r-k}{2r} \quad \forall k$$

(4) Does 5.7.10 apply now? Yes.

We give some details of proof of 5.7.10.

Proof.

$$q((x_1,y_1),(x_2,y_2)) = \begin{cases} p(x_1,x_2)p(y_1,y_2) & x_1 \neq y_1 \\ p(x_1,x_2) & x_1 = y_1, x_2 = y_2 \\ 0 & \text{o/wise} \end{cases}$$

- trans. prob. for (X_n, Y_n) where X_n, Y_n both have trans prob p & are indep until $X_n = Y_n$.
- We control conv. $X_n \to \pi$ by $\mathbb{P}(T > n)$ where $T = \inf\{n \geq 1, X_n = Y_n\}$.

WTS: $T < \infty$ a.s. (this requires aperiodicity!)

Consider $\bar{p}((x_1,y_1),(x_2,y_2)) = p(x_1,x_2)p(y_1,y_2)$.

Claim: $\bar{p}$ is irreducible & has stat. dist.

$$\bar{\pi}((x,y)) = \pi(x)\pi(y)$$

Pf: p is irreducible $\implies$ for given $(x_1,y_1),(x_2,y_2)$ $\exists K, L$ s.t. $P^K(x_1,x_2) > 0$ & $P^L(y_1,y_2) > 0$.

p is aperiodic[a] $\implies$ $\exists m$ s.t. $\forall M \geq m$, $P^M(x_2,x_2) > 0$ & $P^M(y_2,y_2) > 0$. Then we have

$$\begin{aligned} \bar{p}^{K+L+m}((x_1,y_1),(x_2,y_2)) &= P^{K+L+m}(x_1,x_2)P^{K+L+m}(y_1,y_2) \\ &\geq (P^K(x_1,x_2)P^{m+L}(x_2,x_2))(P^L(y_1,y_2)P^{m+K}(y_2,y_2)) \\ &> 0. \end{aligned}$$

Then we have $\bar{p}$ is irreducible.

$$\sum_x \pi(x)p(x,y) = \pi(y)$$

$$\begin{aligned} \sum_{(x_1,y_1)} \bar{\pi}((x_1,y_1))\bar{p}((x_1,y_1),(x_2,y_2)) &= \left(\sum_{x_1} \pi(x_1)p(x_1,x_2)\right)\left(\sum_{y_1} \pi(y_1)p(y_1,y_2)\right) \\ &= \pi(x_2)\pi(y_2) \implies \bar{\pi} \text{ is stationary under } \bar{p} \end{aligned}$$

□

[a] this part should be verified. I think it is directly from the number theory.

Theorem 5.8.1. Suppose p is irreducible with stationary distribution π. Let be a function so that $\sum_y |f(y)|\pi(y) < \infty$. Then, for any initial distribution μ :

$$\frac{1}{n}\sum_{m=1}^{n} f(x_m) \to \sum_y f(y)\pi(y), \qquad \mathbb{P}_\mu - a.s.$$

Corollary 5.8.2. If p is irreducible with stationary distribution π and $X_0 \sim \pi$, then for f s.t. $\mathbb{E}_\pi |f| < \infty$, we have that

$$\lim_n \frac{1}{n}\sum_{m=0}^{\infty} f(x_m) \qquad \exists \mathbb{P}_\pi - a.s.$$

Definition 5.8.3. $\{X_n\}_{n\geq 0}$ is a stationary distribution sequence if for every $k \geq 1$ that shifted sequence $\{x_{n+k}\}_{n\geq 0}$ has thte same distribution as $\{x_n\}_{n\geq 0}$.

We say that $\{x_n\}_{n\geq 0}$ and $\{y_n\}_{n\geq 0}$ have the same distribution if for $\forall m \in \mathbb{N}$ and $\forall 1_i, \dots, i_m$ we have

$$(X_{i_1}, \dots, X_{i_m}) = (Y_{i_1}, \dots, Y_{i_m})$$

Example 5.8.4. (1) $\{X_n\}_{n\geq 0}$ for a MC with stationary distribution π & $X_0 \sim \pi$
(2) $X_0, X_1 \dots$ are iid
(3) Rotation of the circle:
$\Omega = [0,1)$, $\mathcal{F} =$ Borel subsets, $\mathbb{P} =$ Lebesgue meas.
let $\theta \in (0,1)$ & for $n \geq 0$ then

$$X_n(w) = (w + n\theta) \mod 1, \qquad x \mod 1 = x - \lfloor x \rfloor \quad (\text{e.g. } 2.5 \mod 1 = 0.5).$$

Why is X_0 called a rotation?

$$X_n(\omega) = e^{2\pi i(\omega+n\theta)} = e^{2\pi i(\omega+n\theta) \mod 1}$$

More generally:
(4) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A measurable map $\psi : \Omega \to \Omega$ is measure preserving if $\mathbb{P}(\psi^{-1}(A)) = \mathbb{P}(A)$ for $\forall A \in \mathcal{F}$.
Let $\psi^n = \psi(\psi^{n-1})$ for $n \geq 1$ w $\psi^0(\omega) = \omega$. If X is a r.v. on this prob. space, then

$$X_n(\omega) = X(\psi^n(\omega))$$

is a stationary sequence.

Question 5.8.5. How do we know that a probability space $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$ that realizes $\{X_n(w)\}_{n\geq 0}$ exists?

Kolmogorov Ext Thm

Claim 5.8.6. (3) is of the form $(*)$

Proof. let $\varphi(x) = (x+\theta) \mod 1$. WTS: φ is measure-preserving

$$\varphi^{-1}(y) = \begin{cases} 1+y-\theta & 0 \leq y < \theta \\ y-\theta & \theta \leq y < 1 \end{cases}$$

i.e. if $y \in [0,\theta) \implies \varphi^{-1}(y) \in [1-\theta, 1)$ $y \in [\theta, 1) \implies \varphi^{-1}(y) \in [0, 1-\theta)$
WLOG consider $(a,b) \subseteq [1-\theta, 1)$

$$\mathbb{P}(\varphi^{-1}(y) \in (a,b)) = \mathbb{P}(x \in (1+a-\theta, 1+b-\theta)) = (1+b-\theta) - (1+a-\theta) = b-a$$

$\implies \varphi$ is meas. preserving. $\square$ $\square$

Claim 5.8.7. $X_n(\omega)$ is stationary.

Proof. Let $B \in \sigma(X_0, \ldots, X_n)$ and $A = \{\omega : (X_0(\omega), \ldots, X_n(\omega)) \in B\}$.

$$\begin{aligned}
\mathbb{P}((X_k, \ldots, X_{n+k}) \in B) &= \mathbb{P}((\psi^k(\omega), \psi(\psi^k(\omega)), \ldots, \psi^n(\psi^k(\omega))) \in B) \\
&= \mathbb{P}(\psi^k(\omega) \in A) \xrightarrow{\psi \text{ is measure-pres.}} \mathbb{P}(\omega \in A) \\
&= \mathbb{P}((X_0, \ldots, X_n) \in B)
\end{aligned}$$

$\square$

Chapter 6

Ergodic

6.1 Mar 25

by Tatiana (Tanya) Brailovskaya

$\{X_n\}_{n\geq 0}$ is stationary if $\{X_{n+k}\}_{n\geq 0}$ has the same distribution as $\{X_n\}_{n\geq 0}$ for every $k \geq 0$.

Stationary sequences can be constructed from measure-preserving maps. We say φ is measure preserving on $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A), \qquad A \in \mathcal{F}.$$

If $X_n(\omega) = X(\varphi^n(\omega))$, then $\{X_n\}$ is stationary.

Warm-up. Suppose $\omega = \{\omega_n\}_{n\geq 0}$ is a stationary sequence. Let φ be the shift operator

$$\varphi(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots).$$

Then φ is measure preserving by stationarity. Also, if $X(\omega) = \omega_0$, then

$$X(\varphi^n(\omega)) = \omega_n.$$

Conclusion: every stationary sequence can be realized in this form.

In this section φ denotes a measure-preserving map on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 6.1.1. A set $A \in \mathcal{F}$ is invariant if $\varphi^{-1}A = A$. Let $\mathcal{I}$ denote the collection of invariant sets.

Remark 6.1.2. $\mathcal{I}$ is a σ-algebra.

Definition 6.1.3. φ is ergodic if $\mathcal{I}$ is trivial, i.e. for every $A \in \mathcal{I}$,

$$\mathbb{P}(A) \in \{0, 1\}.$$

Remark 6.1.4. If φ is not ergodic, then there is a nontrivial invariant set A with $0 < \mathbb{P}(A) < 1$. Intuitively, ergodicity is the dynamical analogue of irreducibility.

Example 6.1.5 (Irreducible Markov chain)**.** Let $\{X_n\}$ be an irreducible recurrent Markov chain with stationary distribution π, and let φ be the shift operator on paths. If $A \in \mathcal{I}$, then

$$\mathbb{E}_\pi[\mathbb{1}_A \mid \mathcal{F}_n] = \mathbb{E}_\pi[\mathbb{1}_A \circ \varphi^n \mid \mathcal{F}_n] = \mathbb{E}_{X_n}\mathbb{1}_A.$$

By Levy's theorem,

$$\mathbb{E}_\pi[\mathbb{1}_A \mid \mathcal{F}_n] \to \mathbb{1}_A \qquad \text{a.s.}$$

Since the chain is recurrent and irreducible, every state is visited infinitely often. Hence $\mathbb{E}_y\mathbb{1}_A = \mathbb{1}_A$ a.s. for every state y, and therefore $\mathbb{1}_A$ is a.s. constant. Thus $\mathbb{P}(A) \in \{0, 1\}$.

Example 6.1.6 (IID sequence)**.** For an iid sequence $(X_1, X_2, \dots)$, let φ be the shift operator. If $A \in \mathcal{I}$, then for every n,

$$A = \{\omega : \varphi^n(\omega) \in A\} \in \sigma(X_n, X_{n+1}, \dots).$$

Hence

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots) = \mathcal{T}.$$

By Kolmogorov's 0-1 law, the tail σ-algebra $\mathcal{T}$ is trivial, so φ is ergodic.

Example 6.1.7 (Rotation of the circle)**.** Let $\omega \sim \mathrm{Unif}[0,1)$ and $\theta \in (0,1)$. Define

$$X_n(\omega) = (\omega + n\theta) \pmod 1, \qquad \varphi(x) = (x + \theta) \pmod 1,$$

where $x \pmod 1 = x - \lfloor x \rfloor$.

Claim 6.1.8. φ is ergodic iff $\theta \notin \mathbb{Q}$.

Proof. If $\theta = m/n \in \mathbb{Q}$, choose a Borel set $B \subset [0, 1/n)$ and set

$$A = \bigcup_{k=0}^{n-1} \left(B + \frac{k}{n}\right).$$

Then A is invariant, and by choosing B with nontrivial Lebesgue measure we get a nontrivial invariant set. Thus φ is not ergodic.

Conversely, suppose $\theta \notin \mathbb{Q}$. Let $A \in \mathcal{I}$ and put $f = \mathbb{1}_A$. Write the Fourier expansion

$$f(x) = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k x}, \qquad c_k = \int_0^1 f(x) e^{-2\pi i k x}\, dx.$$

Since A is invariant, $f(\varphi(x)) = f(x)$ a.e. Thus

$$\sum_k c_k e^{2\pi i k (x+\theta)} = \sum_k c_k e^{2\pi i k x}.$$

By uniqueness of Fourier coefficients,

$$c_k = c_k e^{2\pi i k \theta}.$$

Since $\theta \notin \mathbb{Q}$, $e^{2\pi i k \theta} \neq 1$ unless $k = 0$. Hence $c_k = 0$ for $k \neq 0$, so f is constant a.e. Therefore $\mathbb{P}(A) \in \{0, 1\}$. $\square$

Theorem 6.1.9 (Birkhoff's ergodic theorem)**.** For any $X \in L^1(\Omega)$,

$$\frac{1}{n} \sum_{m=0}^{n-1} X(\varphi^m \omega) \to \mathbb{E}[X \mid \mathcal{I}] \qquad \text{a.s. and in } L^1.$$

Remark 6.1.10. If φ is ergodic, then $\mathbb{E}[X \mid \mathcal{I}] = \mathbb{E}X$.

Back to examples:

1. Irreducible Markov chain with stationary distribution π. If $X_0 \sim \pi$ and $f \in L^1(\pi)$, then

$$\frac{1}{n} \sum_{m=0}^{n-1} f(X_m) \to \sum_x f(x)\pi(x) \qquad \text{a.s. and in } L^1.$$

2. IID sequence. If $X(\omega) = \omega_0$, then

$$\frac{1}{n} \sum_{m=0}^{n-1} X_m \to \mathbb{E}X_0 \qquad \text{a.s. and in } L^1.$$

3. Irrational rotation of the circle. If $\theta \notin \mathbb{Q}$ and $A \in \mathcal{B}([0,1))$, then

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m \omega) \to |A| \qquad \text{a.s.}$$

Theorem 6.1.11 (Weyl equidistribution for rotations)**.** If $\theta \notin \mathbb{Q}$ and $A = [a,b) \subset [0,1)$, then

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A((\omega + m\theta) \pmod 1) \to b - a$$

for every $\omega \in [0,1)$.

Proof. Birkhoff gives the result for a set of starting points of probability 1. To upgrade this to pointwise convergence for intervals, let

$$A_k = \left[a + \frac{1}{k}, b - \frac{1}{k}\right)$$

when $b - a > 2/k$. For each k, the convergence for A_k holds on a full measure set Ω_k. Then $G = \cap_k \Omega_k$ is dense in $[0,1)$.

Given any $x \in [0,1)$, choose $\omega_k \in G$ with $|\omega_k - x| \le 1/k$. If $\varphi^m(\omega_k) \in A_k$, then $\varphi^m(x) \in A$. Therefore

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m x) \ge b - a - \frac{2}{k}.$$

A similar outer approximation gives

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m x) \le b - a + \frac{4}{k}.$$

Letting $k \to \infty$ proves the claim. □

6.2 Mar 30

by Tatiana (Tanya) Brailovskaya

Theorem 6.2.1 (Birkhoff's ergodic theorem)**.**

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m \omega) \to \mathbb{E}[X \mid \mathcal{I}] \qquad \text{a.s. and in } L^1.$$

Example 6.2.2 (Benford's law)**.** Let $\theta = \log_{10} 2$ and

$$A_k = [\log_{10} k, \log_{10}(k+1)), \qquad 1 \le k \le 9.$$

Since $\theta \notin \mathbb{Q}$, the sequence $m\theta \pmod 1$ is equidistributed.

Warm-up. Show that the first digit of 2^m is k iff $m\theta \pmod 1 \in A_k$.

Proof. The first digit of 2^m is k iff there exists $\ell \in \mathbb{N}$ such that

$$(k+1)10^\ell > 2^m \ge k10^\ell.$$

Taking $\log_{10}$ gives

$$\log_{10}(k+1) + \ell > m \log_{10} 2 \ge \log_{10} k + \ell.$$

This is exactly $m\theta \pmod 1 \in A_k$. □

Thus, by equidistribution,

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}\{m\theta \pmod 1 \in A_k\} \to |A_k| = \log_{10}\left(\frac{k+1}{k}\right).$$

The distribution

$$p(k) = \log_{10}\left(\frac{k+1}{k}\right), \qquad 1 \le k \le 9,$$

is Benford's law.

The topic today is recurrence.

Let $X_1, X_2, \dots$ be a stationary sequence with values in $\mathbb{R}^d$. Define

$$S_k = X_1 + \cdots + X_k, \qquad A = \{S_k \neq 0 \text{ for all } k \ge 1\},$$

and

$$R_n = |\{S_1, \dots, S_n\}|.$$

Theorem 6.2.3.

$$\frac{R_n}{n} \to \mathbb{E}[\mathbb{1}_A \mid \mathcal{I}] \qquad \text{a.s.}$$

Proof. Realize the stationary sequence on path space and let φ be the shift. Observe that

$$R_n \ge \sum_{m=1}^{n} \mathbb{1}_A(\varphi^m \omega).$$

Indeed, if $\mathbb{1}_A(\varphi^m\omega) = 1$, then after time m the walk never returns to S_m, so m is the last visit to the value S_m. By Birkhoff,

$$\liminf_{n\to\infty} \frac{R_n}{n} \ge \mathbb{E}[\mathbb{1}_A \mid \mathcal{I}] \qquad \text{a.s.}$$

For the reverse inequality, set

$$A_k = \{S_1 \neq 0, \dots, S_k \neq 0\}.$$

If a point among $S_1, \dots, S_n$ is represented by a time $m \le n-k$, then unless $A_k(\varphi^m\omega)$ occurs, it will be revisited within the next k steps. Hence

$$R_n \le k + \sum_{m=1}^{n-k} \mathbb{1}_{A_k}(\varphi^m\omega).$$

Birkhoff gives

$$\limsup_{n\to\infty} \frac{R_n}{n} \le \mathbb{E}[\mathbb{1}_{A_k} \mid \mathcal{I}] \qquad \text{a.s.}$$

Since $A_k \downarrow A$, the martingale convergence theorem gives

$$\mathbb{E}[\mathbb{1}_{A_k} \mid \mathcal{I}] \downarrow \mathbb{E}[\mathbb{1}_A \mid \mathcal{I}],$$

and the proof is complete. □

Example 6.2.4. Recurrence of random walks with stationary increments on $\mathbb{Z}$.

Theorem 6.2.5. Let $X_1, X_2, \dots$ be a stationary sequence with values in $\mathbb{Z}$ and $\mathbb{E}|X_1| < \infty$. For $S_n = X_1 + \cdots + X_n$ and $A = \{S_k \neq 0 \text{ for all } k \ge 1\}$:

1. If $\mathbb{E}[X_1 \mid \mathcal{I}] = 0$, then $\mathbb{P}(A) = 0$.
2. If $\mathbb{P}(A) = 0$, then $\mathbb{P}(S_n = 0 \text{ i.o.}) = 1$.

Proof. By Birkhoff,

$$\frac{S_n}{n} = \frac{1}{n}\sum_{m=1}^{n} X_m \to \mathbb{E}[X_1 \mid \mathcal{I}].$$

Hence under assumption (i), $S_n/n \to 0$ a.s. Moreover,

$$\limsup_{n\to\infty} \frac{\max_{1\le k\le n}|S_k|}{n} = 0.$$

To see this, fix K. Then

$$\max_{1\le k\le n}\frac{|S_k|}{n} \le \max_{1\le k\le K}\frac{|S_k|}{n} + \sup_{k\ge K}\frac{|S_k|}{k},$$

and let first $n\to\infty$, then $K\to\infty$.

Since $S_k \in \mathbb{Z}$,

$$R_n \le 1 + 2\max_{1\le k\le n}|S_k|,$$

so $R_n/n \to 0$. The previous theorem gives

$$\mathbb{E}[\mathbb{1}_A \mid \mathcal{I}] = 0,$$

and therefore $\mathbb{P}(A) = \mathbb{E}\mathbb{E}[\mathbb{1}_A \mid \mathcal{I}] = 0$.

For (ii), let

$$F_k = \{S_1 \ne 0, \dots, S_{k-1} \ne 0, S_k = 0\}.$$

Since $\mathbb{P}(A) = 0$, the first return time to 0 is finite a.s., so

$$\sum_{k=1}^{\infty} \mathbb{P}(F_k) = 1.$$

After the first return, stationarity gives the same conclusion for the next return. Iterating, for every $\ell \ge 1$ the walk returns to 0 at least ℓ times with probability 1. Thus $S_n = 0$ infinitely often a.s. □

Theorem 6.2.6 (Kac recurrence theorem)**.** Let $X_0, X_1, \dots$ be stationary with values in $(S, \mathcal{S})$, and let $A \in \mathcal{S}$. Define

$$T_0 = 0, \qquad T_n = \inf\{m > T_{n-1} : X_m \in A\}.$$

If $\mathbb{P}(X_n \in A \text{ at least once}) = 1$, then under $\mathbb{P}(\cdot \mid X_0 \in A)$ the increments

$$t_n = T_n - T_{n-1}$$

form a stationary sequence and

$$\mathbb{E}[T_1 \mid X_0 \in A] = \frac{1}{\mathbb{P}(X_0 \in A)}.$$

6.3 Apr 1

by Tatiana (Tanya) Brailovskaya

Last time: recurrence of stationary sequences. Let $X_1, X_2, \dots$ be stationary, say $X_i \in \mathbb{Z}$, and

$$S_k = X_1 + \cdots + X_k, \qquad A = \{S_k \ne 0 \text{ for all } k \ge 1\}, \qquad R_n = |\{S_1, \dots, S_n\}|.$$

Warm-up. Suppose $\mathbb{E}[X_1 \mid \mathcal{I}] = 0$ and $|X_i| \le M$ a.s. Show that

$$\sup_{k\le n}\frac{|S_k|}{n} \to 0 \qquad \text{a.s.}$$

This follows from the argument in the proof above. If X_i are iid mean-zero increments, typically $\sup_{k\le n}|S_k|$ is of order $\sqrt{n}$.

Theorem 6.3.1 (Subadditive ergodic theorem)**.** Suppose $X_{m,n}$, $0 \le m < n$, satisfy:

(i) $X_{0,m} + X_{m,n} \geq X_{0,n}$.

(ii) For every $k \geq 1$, $\{X_{nk,(n+1)k}\}_{n\geq 0}$ is stationary.

(iii) The distribution of $\{X_{m,m+k} : k \geq 1\}$ does not depend on m.

(iv) $\mathbb{E}X_{0,n}^{+} < \infty$ for all n, and $\mathbb{E}X_{0,n} \geq \gamma_0 n$ for some $\gamma_0 > -\infty$.

Then

$$\lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_{m\geq 1} \frac{\mathbb{E}X_{0,m}}{m} =: \gamma,$$

and

$$\lim_{n\to\infty} \frac{X_{0,n}}{n}$$

exists a.s. and in L^1. If the stationary sequences in (ii) are ergodic, then the limit is γ a.s.

Example 6.3.2 (Additive sums)**.** Let $Y_1, Y_2, \ldots$ be stationary with $\mathbb{E}|Y_k| < \infty$, and set

$$X_{m,n} = Y_{m+1} + \cdots + Y_n.$$

Then

$$X_{0,m} + X_{m,n} = X_{0,n}.$$

Also

$$\{X_{nk,(n+1)k}\}_{n\geq 0} = \{Y_{nk+1} + \cdots + Y_{(n+1)k}\}_{n\geq 0}$$

is stationary. Hence the theorem applies and gives

$$\frac{X_{0,n}}{n} \to X \qquad \text{a.s. and in } L^1.$$

If the stationary sequence is ergodic, then $X = \mathbb{E}Y_1$ a.s., recovering Birkhoff's theorem.

Example 6.3.3 (Longest common subsequence)**.** Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be ergodic sequences of random characters. Let $L_{m,n}$ be the length of the longest common subsequence between the two strings on the interval $(m, n]$. Then

$$L_{0,m} + L_{m,n} \leq L_{0,n}.$$

Therefore $-L_{m,n}$ is subadditive. The theorem implies that

$$\frac{L_{0,n}}{n} \to \gamma \qquad \text{a.s. and in } L^1$$

for a non-random constant γ.

Example 6.3.4 (Products of random matrices)**.** Let $A_1, A_2, \ldots$ be a stationary sequence of $k \times k$ matrices with positive entries and

$$\mathbb{E}|\log A_m(i,j)| < \infty \qquad \forall i, j.$$

Put

$$\alpha_{m,n}(i,j) = (A_{m+1} \cdots A_n)(i,j).$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_{0,n}(i,j)$$

exists a.s. and in L^1.

Proof. First prove the claim for $(i,j)=(1,1)$. By positivity,

$$\alpha_{0,m}(1,1)\alpha_{m,n}(1,1) \le \alpha_{0,n}(1,1).$$

Thus

$$X_{m,n} := -\log \alpha_{m,n}(1,1)$$

is subadditive. The stationarity assumptions follow from stationarity of the matrices.

For the integrability lower bound, note that

$$\alpha_{0,n}(1,1) \le k^{n-1} \prod_{m=1}^{n} \sup_{i,j} A_m(i,j),$$

so

$$-\log \alpha_{0,n}(1,1) \ge -n\log k - \sum_{m=1}^{n} \log\left(\sup_{i,j} A_m(i,j)\right).$$

Since $\mathbb{E}|\log A_m(i,j)| < \infty$ and there are only finitely many entries,

$$\mathbb{E}\log^+\left(\sup_{i,j} A_m(i,j)\right) < \infty.$$

Hence condition (iv) holds.

The subadditive ergodic theorem gives the a.s. and L^1 limit for $\frac{1}{n}\log \alpha_{0,n}(1,1)$. For general (i,j), positivity lets us sandwich $\alpha_{r,n}(i,j)$ between diagonal products up to fixed multiplicative factors involving only finitely many entries, e.g.

$$A_r(j,i)\,\alpha_{r,n}(i,j) \le \alpha_{r-1,n}(j,j),$$

and similarly from the other side. After taking $-\frac{1}{n}\log$ and letting $n\to\infty$, the fixed factors disappear. Stationarity then transfers the diagonal limit to every pair (i,j). $\square$

Chapter 7

Brownian Motion

7.1 Apr 6

by Tatiana (Tanya) Brailovskaya

Recall a random walk

$$S_n = X_1 + \cdots + X_n,$$

where X_i are iid and $\mathbb{E}|X_i| < \infty$. It is a martingale when $\mathbb{E}X_i = 0$, a Markov chain, and has stationary independent increments: if $n_1 < n_2 < \cdots < n_m$, then

$$(S_{n_2} - S_{n_1}, \dots, S_{n_m} - S_{n_{m-1}}) \stackrel{d}{=} (S_{n_2-n_1}, \dots, S_{n_m-n_{m-1}}),$$

and the increments are independent.

Definition 7.1.1. A Brownian motion B_t, $t \geq 0$, is an $\mathbb{R}$-valued stochastic process such that:

1. If $0 \leq t_0 < t_1 < \cdots < t_n$, then

   $$B(t_0),\ B(t_1) - B(t_0), \dots, B(t_n) - B(t_{n-1})$$

   are independent.

2. If $s, t \geq 0$, then

   $$B(s+t) - B(s) \sim \mathcal{N}(0, t).$$

3. With probability 1, $t \mapsto B_t$ is continuous.

Theorem 7.1.2 (Donsker's theorem, 1951)**.** If X_i satisfy $\mathbb{E}X_i = 0$ and $\mathrm{Var}(X_i) = 1$, and if $S_n = X_1 + \cdots + X_n$, then

$$\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \Rightarrow B_t.$$

The convergence is weak convergence of measures on the Skorokhod space of functions from $[0,1]$ to $\mathbb{R}$.

Remark 7.1.3. A weaker finite-dimensional statement is: for every $0 \leq t_1 < \cdots < t_m$,

$$\left(\frac{S_{\lfloor nt_1 \rfloor}}{\sqrt{n}}, \dots, \frac{S_{\lfloor nt_m \rfloor}}{\sqrt{n}}\right) \Rightarrow (B(t_1), \dots, B(t_m)).$$

How do we know Brownian motion exists?

Definition 7.1.4. For a stochastic process X_t, $t \in \mathcal{T}$, with values in $\mathbb{R}$, the probability measures

$$\mu_{t_1,\dots,t_n}(B) = \mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in B), \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

are called the finite-dimensional distributions.

Definition 7.1.5. A family $\{\mu_{t_1,\dots,t_n}\}$ is consistent if:

1. it is invariant under permutations of the time coordinates, and
2. it is compatible under marginalization, i.e.

$$\mu_{t_1,\dots,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R}) = \mu_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n).$$

Theorem 7.1.6 (Kolmogorov extension theorem)**.** If $\{\mu_{t_1,\dots,t_n}\}$ is a consistent family of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then there exists a unique probability measure ν on

$$\Omega_0 = \{\omega : \mathcal{T} \to \mathbb{R}\}$$

equipped with the cylinder σ-algebra such that

$$\nu\{\omega : \omega(t_i) \in A_i,\ 1 \le i \le n\} = \mu_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n).$$

Construction attempt. Let $x \in \mathbb{R}$. For $0 < t_1 < \cdots < t_n$, define a measure on $\mathbb{R}^n$ by

$$\mu^x_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \int_{A_1} \cdots \int_{A_n} \prod_{m=1}^{n} p_{t_m - t_{m-1}}(x_{m-1}, x_m)\, dx_n \cdots dx_1,$$

where $x_0 = x$, $t_0 = 0$, and

$$p_t(a,b) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(b-a)^2}{2t}\right).$$

The consistency check uses

$$\int_{\mathbb{R}} p_{t_n - t_{n-1}}(x_{n-1}, x_n)\, dx_n = 1$$

and the Chapman-Kolmogorov identity. Kolmogorov's theorem gives a measure ν_x with independent Gaussian increments and $\nu_x\{\omega : \omega(0) = x\} = 1$.

Claim 7.1.7 (Translation invariance)**.** Under ν_x, the process $\{B_t - B_0 : t \ge 0\}$ is independent of B_0 and has the same distribution as Brownian motion started at 0.

Proof. The increments are stationary and independent. The σ-algebra generated by B_0 is independent of the π-system generated by events involving only the increments

$$B(t_1) - B(0), \dots, B(t_n) - B(t_{n-1}),$$

so the monotone class theorem gives the claim. □

There is still a problem: the cylinder σ-algebra on all functions $[0,\infty) \to \mathbb{R}$ is too coarse to contain the set of continuous functions. To fix this, let

$$\mathbb{Q}_2 = \{m2^{-n} : m, n \ge 0\}$$

be the dyadic rationals, and consider

$$\Omega_q = \{\omega : \mathbb{Q}_2 \to \mathbb{R}\}.$$

Applying Kolmogorov extension on Ω_q gives a discrete-time Brownian motion on dyadic times.

Theorem 7.1.8. Let $T < \infty$ and $x \in \mathbb{R}$. The measure ν_x assigns probability 1 to

$$\mathcal{U}_T = \{\omega : t \mapsto \omega(t) \text{ is uniformly continuous on } \mathbb{Q}_2 \cap [0,T]\}.$$

Proof. It is enough to prove the result for $x = 0$ and $T = 1$, by translation and scaling. The key input is Kolmogorov's continuity criterion, proved below. □

Fact 7.1.9. If a function f is uniformly continuous on a dense subset $S \subset [0, T]$, then there exists a unique continuous extension $\tilde{f}$ to $[0, T]$.

Thus for almost every dyadic path there is a unique continuous extension. This completes the construction of Brownian motion on $C[0, \infty)$.

Remark 7.1.10. Durrett defines $\mathcal{F}_x$ to be the smallest σ-algebra such that every coordinate map

$$\pi_t : C[0, \infty) \to \mathbb{R}, \qquad \pi_t(f) = f(t),$$

is measurable. This agrees with the Borel σ-algebra induced by the topology of uniform convergence on compact intervals.

Claim 7.1.11 (Scaling)**.** If $B_0 = 0$, then for every $t > 0$,

$$\{B_{st} : s \geq 0\} \overset{d}{=} \{t^{1/2} B_s : s \geq 0\}.$$

Proof. For $s_0 < s_1 < \cdots < s_n$,

$$B_{s_i t} - B_{s_{i-1} t} \sim \mathcal{N}(0, t(s_i - s_{i-1})) \overset{d}{=} t^{1/2}(B_{s_i} - B_{s_{i-1}}),$$

and the increments are independent. □

Remark 7.1.12 (Equivalent description for $B_0 = 0$)**.** Brownian motion is equivalently a Gaussian process with

$$\mathbb{E} B_s = 0, \qquad \mathbb{E}[B_s B_t] = s \wedge t,$$

and continuous paths. Indeed, for $t > s$,

$$B_t = B_s + (B_t - B_s),$$

and the increment is independent of B_s, giving $\mathbb{E}[B_s B_t] = \mathbb{E} B_s^2 = s$.

7.2 Apr 8

by Tatiana (Tanya) Brailovskaya

Warm-up. Let B_t be Brownian motion with $B_0 = 0$. Compute $\mathbb{E}[B_1^2 B_2 B_3]$.

Proof. Write $B_2 = B_1 + (B_2 - B_1)$ and $B_3 = B_2 + (B_3 - B_2)$. The last increment has mean 0 and is independent of $\mathcal{F}_2$, so

$$\mathbb{E}[B_1^2 B_2 B_3] = \mathbb{E}[B_1^2 B_2^2].$$

Also

$$B_2^2 = (B_1 + (B_2 - B_1))^2,$$

so by independence,

$$\mathbb{E}[B_1^2 B_2^2] = \mathbb{E}[B_1^4] + \mathbb{E}[B_1^2]\mathbb{E}[(B_2 - B_1)^2] = 3 + 1 = 4.$$

□

Last time, we constructed a process with independent Gaussian increments. The remaining issue was to show that the dyadic paths extend to continuous paths.

Theorem 7.2.1 (Kolmogorov continuity criterion)**.** Let X_t, $t \in \mathbb{Q}_2 \cap [0,1]$, be a process such that

$$\mathbb{E}|X_t - X_s|^\beta \le K|t-s|^{1+\alpha}$$

for some $\alpha, \beta > 0$. If $\gamma < \alpha/\beta$, then with probability 1 there exists $C(\omega) < \infty$ such that

$$|X(q) - X(r)| \le C(\omega)|q-r|^\gamma$$

for all $q, r \in \mathbb{Q}_2 \cap [0,1]$.

Proof. Let

$$G_n = \left\{ \left| X\left(\frac{i}{2^n}\right) - X\left(\frac{i-1}{2^n}\right) \right| \le 2^{-\gamma n} \text{ for all } 1 \le i \le 2^n \right\}.$$

By Markov's inequality,

$$\begin{aligned} \mathbb{P}(G_n^c) &\le \sum_{i=1}^{2^n} \mathbb{P}\left(\left| X\left(\frac{i}{2^n}\right) - X\left(\frac{i-1}{2^n}\right) \right| > 2^{-\gamma n} \right) \\ &\le 2^n K 2^{-n(1+\alpha)} 2^{\gamma\beta n} = K 2^{-n(\alpha - \gamma\beta)}. \end{aligned}$$

Since $\gamma < \alpha/\beta$, $\sum_n \mathbb{P}(G_n^c) < \infty$. By Borel-Cantelli, with probability 1 there is N such that G_n holds for all $n \ge N$.

On $H_N = \cap_{n=N}^\infty G_n$, a dyadic chaining argument gives

$$|X(q) - X(r)| \le \frac{2}{1-2^{-\gamma}} |q-r|^\gamma$$

whenever $q, r \in \mathbb{Q}_2 \cap [0,1]$ and $0 < |q-r| < 2^{-N}$. Covering $[q,r]$ by finitely many such small intervals gives a global random constant $C(\omega)$. □

Question 7.2.2. For Brownian motion, which α, β can we use?

Since $B_t - B_s \sim \sqrt{|t-s|}\, g$ for $g \sim \mathcal{N}(0,1)$,

$$\mathbb{E}|B_t - B_s|^{2m} = |t-s|^m \mathbb{E}|g|^{2m}.$$

Thus we may take $\beta = 2m$ and $\alpha = m-1$, so any

$$\gamma < \frac{m-1}{2m}$$

works. Letting $m \to \infty$, Brownian paths are a.s. γ-Holder continuous for every $\gamma < 1/2$.

7.3 Apr 13

by Tatiana (Tanya) Brailovskaya

Last time. With probability 1, Brownian paths are γ-Holder continuous for every $\gamma < 1/2$.

Remark 7.3.1. For $\gamma \ge 1/2$, Brownian paths are a.s. not γ-Holder continuous. The case $\gamma > 1/2$ was in HW7.

Theorem 7.3.2. With probability 1, Brownian paths are not Lipschitz continuous at any point.

Proof. Fix $C < \infty$. Let

$$A_n = \left\{ \omega : \exists s \in [0,1] \text{ such that } |B_t - B_s| \le C|t-s| \text{ whenever } |t-s| \le \frac{3}{n} \right\}.$$

For $1 \le k \le n-2$, define

$$Y_{k,n} = \max_{j=0,1,2} \left| B\left(\frac{k+j}{n}\right) - B\left(\frac{k+j-1}{n}\right) \right|.$$

Let B_n be the event that at least one $Y_{k,n} \le 5C/n$. Then $A_n \subseteq B_n$ and

$$\mathbb{P}(A_n) \le \mathbb{P}(B_n) \le \sum_{k=1}^{n-2} \mathbb{P}\left(Y_{k,n} \le \frac{5C}{n}\right).$$

The three increments in $Y_{k,n}$ are independent, so

$$\mathbb{P}\left(Y_{k,n} \le \frac{5C}{n}\right) = \left[\mathbb{P}\left(|B(1/n)| \le \frac{5C}{n}\right)\right]^3 = \left[\mathbb{P}\left(|B_1| \le \frac{5C}{\sqrt{n}}\right)\right]^3 \le \frac{C'}{n^{3/2}}.$$

Hence $\mathbb{P}(A_n) \le C''/\sqrt{n}$ along the proxy events. Since

$$A_n \subseteq A_{n+1} \subseteq A_{n+2} \subseteq \cdots,$$

for every $m \ge n$ we have $\mathbb{P}(A_n) \le \mathbb{P}(B_m) \to 0$. Thus $\mathbb{P}(A_n) = 0$ for every n, and the Lipschitz condition fails a.s. □

Now we discuss the Markov property of Brownian motion. Let

$$\mathcal{F}_s^0 = \sigma(B_r : r \le s), \qquad \mathcal{F}_s^+ = \bigcap_{t>s} \mathcal{F}_t^0.$$

Observation 7.3.3. The filtration $\{\mathcal{F}_s^+\}$ is right-continuous:

$$\bigcap_{t>s} \mathcal{F}_t^+ = \bigcap_{t>s} \bigcap_{u>t} \mathcal{F}_u^0 = \bigcap_{u>s} \mathcal{F}_u^0 = \mathcal{F}_s^+.$$

For $f(u) > 0$, the random variable

$$\limsup_{t \downarrow s} \frac{B_t - B_s}{f(t-s)}$$

is $\mathcal{F}_s^+$-measurable, but not necessarily $\mathcal{F}_s^0$-measurable.

Let the shift operator Θ_s on path space be

$$(\Theta_s \omega)(t) = \omega(s+t).$$

Theorem 7.3.4 (Markov property)**.** If $s \ge 0$ and Y is bounded and measurable with respect to $(C, \mathcal{C})$, then for every $x \in \mathbb{R}$,

$$\mathbb{E}_x[Y \circ \Theta_s \mid \mathcal{F}_s^+] = \mathbb{E}_{B_s} Y.$$

Proof. First check this for cylinder functions

$$Y = \prod_{m=1}^{n} f_m(B(t_m)), \qquad 0 < t_1 < \cdots < t_n,$$

using independent Gaussian increments. Then apply the monotone class theorem. □

Corollary 7.3.5.

$$\mathbb{E}_x[Y \circ \Theta_s \mid \mathcal{F}_s^+] = \mathbb{E}_x[Y \circ \Theta_s \mid \mathcal{F}_s^0].$$

Proof. The Markov property gives $\mathbb{E}_x[Y \circ \Theta_s \mid \mathcal{F}_s^+] = \mathbb{E}_{B_s} Y$, which is a measurable function of B_s and hence is $\mathcal{F}_s^0$-measurable. Apply the tower property. □

Theorem 7.3.6. If Z is bounded and measurable with respect to $(C, \mathcal{C})$, then for all $s \ge 0$ and $x \in \mathbb{R}$,

$$\mathbb{E}_x[Z \mid \mathcal{F}_s^+] = \mathbb{E}_x[Z \mid \mathcal{F}_s^0].$$

Proof. Prove the statement first for cylinder functions. Split the time coordinates into those $\le s$ and those $> s$, then use the Markov property for the future coordinates. The resulting conditional expectation is $\mathcal{F}_s^0$-measurable. Extend by a monotone class argument. □

Theorem 7.3.7 (Blumenthal's 0-1 law)**.** If $A \in \mathcal{F}_0^+$, then for every $x \in \mathbb{R}$,

$$\mathbb{P}_x(A) \in \{0, 1\}.$$

Proof. Since $A \in \mathcal{F}_0^+$,

$$\mathbb{1}_A = \mathbb{E}_x[\mathbb{1}_A \mid \mathcal{F}_0^+].$$

By the previous theorem,

$$\mathbb{E}_x[\mathbb{1}_A \mid \mathcal{F}_0^+] = \mathbb{E}_x[\mathbb{1}_A \mid \mathcal{F}_0^0].$$

But $\mathcal{F}_0^0 = \sigma(B_0)$ is trivial under $\mathbb{P}_x$, so this conditional expectation is the constant $\mathbb{P}_x(A)$. Thus $\mathbb{1}_A = \mathbb{P}_x(A)$ a.s., and the probability is 0 or 1. □

Theorem 7.3.8. Let

$$\tau = \inf\{t \ge 0 : B_t > 0\}.$$

Under $\mathbb{P}_0$,

$$\mathbb{P}_0(\tau = 0) = 1.$$

Proof. The event $\{\tau = 0\}$ belongs to $\mathcal{F}_0^+$. By Blumenthal's 0-1 law, its probability is 0 or 1. For every $t > 0$,

$$\mathbb{P}_0(\tau \le t) \ge \mathbb{P}_0(B_t > 0) = \frac{1}{2}.$$

Letting $t \downarrow 0$ gives $\mathbb{P}_0(\tau = 0) \ge 1/2$, so it must equal 1. □

7.4 Apr 15

by Tatiana (Tanya) Brailovskaya

Last time. We compared $\mathcal{F}_t^+$ and $\mathcal{F}_t^0$, proved the Markov property, and proved Blumenthal's 0-1 law.

Warm-up. Compute $\mathbb{E}_0[B_1^2 B_2 B_3]$ using the Markov property.

Proof. Conditioning first on $\mathcal{F}_2^+$,

$$\mathbb{E}_0[B_1^2 B_2 B_3] = \mathbb{E}_0\left[B_1^2 B_2\, \mathbb{E}_0[B_3 \mid \mathcal{F}_2^+]\right] = \mathbb{E}_0[B_1^2 B_2^2].$$

Conditioning on $\mathcal{F}_1^+$,

$$\mathbb{E}_0[B_2^2 \mid \mathcal{F}_1^+] = \mathbb{E}_{B_1}[B_1^2] = B_1^2 + 1.$$

Hence

$$\mathbb{E}_0[B_1^2 B_2^2] = \mathbb{E}_0[B_1^2(B_1^2 + 1)] = 3 + 1 = 4.$$

□

To use strong Markov property cleanly, we complete the filtration. Let

$$\mathcal{N}_x = \{A : A \subset D \in \mathcal{F}_\infty^+,\ \mathbb{P}_x(D) = 0\},$$

and set

$$\mathcal{F}_s^x = \sigma(\mathcal{F}_s^+ \cup \mathcal{N}_x), \qquad \mathcal{F}_s = \bigcap_{x \in \mathbb{R}} \mathcal{F}_s^x.$$

This is the usual completed right-continuous Brownian filtration.

Definition 7.4.1. A random variable S with values in $[0,\infty]$ is a stopping time if

$$\{S \le t\} \in \mathcal{F}_t \qquad \forall t \ge 0.$$

Question 7.4.2. Why does $\{S < t\} \in \mathcal{F}_t$ for all t imply $\{S \le t\} \in \mathcal{F}_t$ for all t?

Proof. Since the filtration is right-continuous,

$$\{S \le t\} = \bigcap_{n=1}^{\infty} \{S < t + 1/n\} \in \bigcap_{n=1}^{\infty} \mathcal{F}_{t+1/n} = \mathcal{F}_t.$$

□

Theorem 7.4.3. If G is open or closed and

$$T = \inf\{t \ge 0 : B_t \in G\},$$

then T is a stopping time.

Proof. If G is open, then by continuity,

$$\{T < t\} = \bigcup_{q \in \mathbb{Q} \cap [0,t)} \{B_q \in G\} \in \mathcal{F}_t.$$

Thus T is a stopping time.

If G is closed, let

$$G_n = \{y : |y - x| < 1/n \text{ for some } x \in G\}$$

and

$$T_n = \inf\{t \ge 0 : B_t \in G_n\}.$$

Each G_n is open, so T_n is a stopping time. Also $G_n \downarrow G$, hence $T_n \uparrow T$ by continuity of Brownian paths. Since an increasing limit of stopping times is a stopping time, T is a stopping time. □

For a stopping time S, define the shifted path

$$(\Theta_S \omega)(t) = \begin{cases} \omega(S(\omega) + t), & S(\omega) < \infty, \\ \Delta, & S(\omega) = \infty, \end{cases}$$

where Δ is an extra cemetery state. Define

$$\mathcal{F}_S = \{A : A \cap \{S \le t\} \in \mathcal{F}_t \text{ for all } t \ge 0\}.$$

Theorem 7.4.4. If $S \le T$ are stopping times, then

$$\mathcal{F}_S \subseteq \mathcal{F}_T.$$

Proof. If $A \in \mathcal{F}_S$, then

$$A \cap \{T \le t\} = A \cap \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t.$$

□

Theorem 7.4.5 (Strong Markov property)**.** Let $Y_s(\omega)$ be bounded and measurable with respect to $\sigma(\mathbb{R}_+ \times C)$. If S is a stopping time, then for every $x \in \mathbb{R}$,

$$\mathbb{E}_x[Y_S \circ \Theta_S \mid \mathcal{F}_S] = \mathbb{E}_{B_S} Y_S \qquad \text{on } \{S < \infty\}.$$

Proof. First prove the result when S takes countably many deterministic values t_n. On $\{S = t_n\}$ this is the ordinary Markov property at time t_n, and summing over n proves the claim.

For a general stopping time, approximate from above by dyadic stopping times

$$S_n = (\lfloor 2^n S \rfloor + 1)2^{-n}.$$

Then $S_n \downarrow S$, $\mathcal{F}_S \subseteq \mathcal{F}_{S_n}$, and by path continuity $B_{S_n} \to B_S$. For bounded cylinder functions continuity and dominated convergence pass the identity to the limit. A monotone class argument extends it to all bounded measurable Y. $\square$

7.5 Apr 20

by Tatiana (Tanya) Brailovskaya

Warm-up. If T_n is a sequence of stopping times, then

$$T = \sup_n T_n \qquad \text{and} \qquad S = \limsup_{n\to\infty} T_n$$

are stopping times.

Proof. For $T = \sup_n T_n$,

$$\{T \le t\} = \bigcap_{n=1}^{\infty} \{T_n \le t\} \in \mathcal{F}_t.$$

For $S = \limsup_n T_n$, write

$$S = \lim_{m\to\infty} \sup_{n\ge m} T_n.$$

Each $S_m = \sup_{n\ge m} T_n$ is a stopping time, and $S_m \downarrow S$. By right-continuity,

$$\{S \le t\} = \bigcap_{q>t,\ q\in\mathbb{Q}} \bigcup_{m=1}^{\infty} \{S_m \le q\} \in \mathcal{F}_t.$$

$\square$

Theorem 7.5.1. Let

$$\mathcal{Z}(\omega) = \{t : B_t(\omega) = 0\}.$$

With probability 1, $\mathcal{Z}(\omega)$ is uncountable and has Lebesgue measure 0.

Proof. First show that the zero set has no isolated points. For $t \ge 0$, define

$$R_t = \inf\{u \ge t : B_u = 0\}.$$

By Blumenthal's 0-1 law and the strong Markov property, after any time R_t the Brownian motion returns to 0 immediately. Taking t rational shows that almost surely no zero is isolated. Since Brownian paths are continuous, $\mathcal{Z}(\omega)$ is closed. A closed set with no isolated points is perfect, hence uncountable.

For Lebesgue measure, use Tonelli's theorem:

$$\mathbb{E}_x|\mathcal{Z}(\omega)| = \mathbb{E}_x \int_0^{\infty} \mathbb{1}_{\{B_t=0\}}\, dt = \int_0^{\infty} \mathbb{P}_x(B_t = 0)\, dt = 0.$$

Thus $|\mathcal{Z}(\omega)| = 0$ a.s. $\square$

Let

$$T_a = \inf\{t \ge 0 : B_t = a\}.$$

Theorem 7.5.2. Under $\mathbb{P}_0$, the process $\{T_a : a \ge 0\}$ has stationary independent increments.

Proof. If $0 < a < b$, then by the strong Markov property at T_a,

$$T_b - T_a = T_{b-a} \circ \Theta_{T_a} \stackrel{d}{=} T_{b-a}.$$

This gives stationary increments. For $0 = a_0 < a_1 < \cdots < a_n$ and bounded measurable f_i, repeated conditioning with the strong Markov property gives

$$\mathbb{E}_0 \prod_{i=1}^{n} f_i(T_{a_i} - T_{a_{i-1}}) = \prod_{i=1}^{n} \mathbb{E}_0 f_i(T_{a_i - a_{i-1}}),$$

so the increments are independent. □

Claim 7.5.3. $T_a \stackrel{d}{=} a^2 T_1$ for $a > 0$.

Proof. By Brownian scaling, if $T_a = t$, then $a = B_t \stackrel{d}{=} aB_{t/a^2}$. Hence the first hitting time of a has the same law as a^2 times the first hitting time of 1. □

Let $t_k = T_k - T_{k-1}$. Then t_k are iid copies of T_1 and

$$T_n = t_1 + \cdots + t_n \stackrel{d}{=} n^2 T_1.$$

Consequently,

$$\frac{t_1 + \cdots + t_n}{n^2} \stackrel{d}{=} T_1.$$

If T_1' is an independent copy of T_1, then

$$aT_1 + bT_1' \stackrel{d}{=} (\sqrt{a} + \sqrt{b})^2 T_1.$$

Thus T_1 has a stable law of index $1/2$.

7.6 Apr 22

by Tatiana (Tanya) Brailovskaya

Warm-up. Let X_n be a simple random walk on $\mathbb{Z}$, with $a < x < b$ and

$$X_n = x + Y_1 + \cdots + Y_n, \qquad \mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2}.$$

Let

$$T_a = \inf\{n > 0 : X_n = a\}, \qquad T_b = \inf\{n > 0 : X_n = b\}.$$

Use optional stopping to show

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}.$$

Proof. Let $T = T_a \wedge T_b$. Since X_n is a martingale,

$$\mathbb{E}X_{T\wedge n} = x.$$

The stopping time T is finite a.s. and $\mathbb{E}T < \infty$; for example, in every block of length $b - a$ there is probability at least $2^{-(b-a)}$ of hitting one of the endpoints. Hence $X_{T\wedge n} \to X_T$ and is bounded, so

$$x = \mathbb{E}X_T = a\mathbb{P}_x(T_a < T_b) + b\mathbb{P}_x(T_b < T_a).$$

Solving gives

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}.$$

One can also show

$$\mathbb{E}T = (x-a)(b-x).$$

□

Definition 7.6.1. A process $(X_t)_{t\geq 0}$ is a martingale adapted to a right-continuous filtration $\{\mathcal{F}_t\}$ if:

1. $X_t \in \mathcal{F}_t$,
2. $\mathbb{E}|X_t| < \infty$ for all t, and
3. $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ for all $t > s$.

Theorem 7.6.2. B_t is a martingale adapted to $\mathcal{F}_t$.

Proof. For $t > s$,

$$\mathbb{E}[B_t \mid \mathcal{F}_s] = \mathbb{E}[B_s + (B_t - B_s) \mid \mathcal{F}_s] = B_s.$$

□

Theorem 7.6.3 (Optional stopping)**.** Let X_t be a right-continuous martingale adapted to $\{\mathcal{F}_t\}$. If T is a bounded stopping time, then

$$\mathbb{E}X_T = \mathbb{E}X_0.$$

Proof. Approximate T from above by dyadic stopping times

$$T_m = (\lfloor 2^m T \rfloor + 1)2^{-m}.$$

Then $T_m \downarrow T$ and T_m takes finitely many values on a bounded interval. Applying the discrete-time optional stopping theorem to the martingale sampled on the dyadic grid gives

$$\mathbb{E}X_{T_m} = \mathbb{E}X_0.$$

Since X is right-continuous and $T_m \downarrow T$, $X_{T_m} \to X_T$. The discrete stopped variables are uniformly integrable, so $\mathbb{E}X_T = \mathbb{E}X_0$. □

Theorem 7.6.4. If $a < x < b$, then

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}.$$

Proof. Let $T = T_a \wedge T_b$. Applying optional stopping to the bounded stopping time $T \wedge t$ for the martingale B_t gives

$$x = \mathbb{E}_x B_{T\wedge t}.$$

Letting $t \to \infty$ and using dominated convergence,

$$x = \mathbb{E}_x B_T = a\mathbb{P}_x(T_a < T_b) + b\mathbb{P}_x(T_b < T_a).$$

Solve for the probability. □

Theorem 7.6.5. Let

$$T = \inf\{t : B_t \notin (a,b)\}, \qquad a < 0 < b.$$

Then

$$\mathbb{E}_0 T = -ab.$$

Proof. First,

$$B_t^2 - t$$

is a martingale, since for $t > s$,

$$\mathbb{E}[B_t^2 \mid \mathcal{F}_s] = \mathbb{E}[(B_s + (B_t - B_s))^2 \mid \mathcal{F}_s] = B_s^2 + t - s.$$

Optional stopping at $T \wedge t$ gives

$$\mathbb{E}_0[B_{T\wedge t}^2] = \mathbb{E}_0[T \wedge t].$$

Letting $t \to \infty$,

$$\mathbb{E}_0 T = \mathbb{E}_0[B_T^2].$$

By the previous theorem,

$$\mathbb{P}_0(T_a < T_b) = \frac{b}{b-a}, \qquad \mathbb{P}_0(T_b < T_a) = \frac{-a}{b-a}.$$

Hence

$$\mathbb{E}_0[B_T^2] = a^2 \frac{b}{b-a} + b^2 \frac{-a}{b-a} = -ab.$$

More generally, for $a < x < b$,

$$\mathbb{E}_x T = (x-a)(b-x).$$

□

Theorem 7.6.6. If

$$T_a = \inf\{t : B_t = a\},$$

then for $\lambda \geq 0$,

$$\mathbb{E}_0 \exp(-\lambda T_a) = \exp(-|a|\sqrt{2\lambda}).$$

Proof. For $\theta \in \mathbb{R}$,

$$M_t = \exp\left(\theta B_t - \frac{\theta^2}{2} t\right)$$

is a martingale. Optional stopping at $T_a \wedge t$ gives

$$\mathbb{E}_0 \exp\left(\theta B_{T_a \wedge t} - \frac{\theta^2}{2}(T_a \wedge t)\right) = 1.$$

Let $a > 0$ and choose $\theta = \sqrt{2\lambda}$. Letting $t \to \infty$,

$$1 = \mathbb{E}_0 \exp\left(a\sqrt{2\lambda} - \lambda T_a\right),$$

so

$$\mathbb{E}_0 e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}.$$

The case $a < 0$ follows by symmetry. □

Recall that T_a has a stable law of index $1/2$. Formally setting $\lambda = -it$ in the Laplace transform gives

$$\mathbb{E}_0[e^{itT_a}] = \exp\left(-|a|\sqrt{-2it}\right),$$

which is the characteristic function of this stable law.