STATS 310B Notes

Kenneth Tay

Conditional expectation (Lec 1-2)

- (Lec 1) $L^2(\Omega, \mathcal{F}, P)$ is complete.
- (Lec 1) For $X \in L^2(\Omega, \mathcal{F}, P)$, the **conditional expectation** of X given $\mathcal{G}$ is the $\mathcal{G}$-measurable random variable in L^2 such that $\mathbb{E}XZ = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]Z]$ for all $Z \in L^2(\Omega, \mathcal{G}, P)$. Conditional expectation is unique, and is the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, P)$.
- (Lec 2) For $X \in L^1(\Omega, \mathcal{F}, P)$, the **conditional expectation** of X given $\mathcal{G}$ is the unique $\mathcal{G}$-measurable random variable in L^1 such that for all $A \in \mathcal{G}$, $\mathbb{E}[X; A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]; A]$, where $\mathbb{E}[X; A] = \mathbb{E}[X 1_A]$.
- (Lec 2) If $\mathcal{H}$ independent of $\sigma(X, \mathcal{G})$, then $\mathbb{E}[X \mid \sigma(\mathcal{H}, \mathcal{G})] = \mathbb{E}[X \mid \mathcal{G}]$.
- (Lec 2) **Conditional Jensen's inequality:** If $\phi : \mathbb{R} \mapsto \mathbb{R}$ convex, and $\phi(X)$ and X both integrable, then $\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \phi(\mathbb{E}[X \mid \mathcal{G}])$ a.s.
- For $p \in [1, \infty]$ and any sub-σ-algebra $\mathcal{G}$, $\|\mathbb{E}[X \mid \mathcal{G}]\|_p \leq \|X\|_p$.
- (Dembo Prop 4.2.33) For any $X \in L^1(\Omega, \mathcal{F}, P)$, the collection $\{\mathbb{E}[X \mid \mathcal{H}], \mathcal{H} \subseteq \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$ is U.I.

Martingales (Lec 3-11)

- **Examples of martingales:**
  - (Lec 3) Let $Y_1, Y_2, \ldots$ be independent (integrable) random variables with $\mathbb{E}Y_i = \mu_i$, and let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Then $S_n = \displaystyle\sum_{i=1}^n (Y_i - \mu_i)$ is a martingale.
  - (Lec 3) Same setting as above, assume further that $\text{Var}(Y_i) = \sigma_i^2 < \infty$. Then $Z_n = S_n^2 - \displaystyle\sum_{i=1}^n \sigma_i^2$ is a martingale.
  - (Lec 3) Let $Y_1, Y_2, \ldots$ be independent non-negative random variables with $\mathbb{E}Y_i = 1$ for all i. Then $Z_n = \displaystyle\prod_{i=1}^n Y_i$ is a martingale.
  - (Lec 3) If $X_1, X_2, \ldots$ i.i.d. and $M(\theta) = \mathbb{E}[e^{\theta X_i}] < \infty$ for some θ, then $Z_n = \dfrac{e^{\theta \sum_{i=1}^n X_i}}{M(\theta)^n}$ is a martingale. (θ is usually chosen such that $M(\theta) = 1$ and $\theta \neq 0$.)
- **Uncorrelated differences:** For any martingale $\{X_n\}$, $X_i - X_{i-1}$ and $X_j - X_{j-1}$ are uncorrelated for $i \neq j$, i.e. $\mathbb{E}[(X_i - X_{i-1})(X_j - X_{j-1})] = 0$.

- (HW3) If $\{X_n\}$ is a martingale with $\sup_n \mathbb{E}|X_n| < \infty$, we can write it as the difference of two non-negative martingales.
- (Lec 3) **Stopping time:** A random variable $T : \Omega \mapsto \{1, 2, \dots\} \cup \{\infty\}$ is a **stopping time adapted to** $\{\mathcal{F}_n\}$ if $\{T = n\} \in \mathcal{F}_n$ for all n.
- (Lec 3) If T is a stopping time, then $\{T \leq n\} \in \mathcal{F}_n$ and $\{T \geq n\} \in \mathcal{F}_{n-1}$. For any integer n, $T \wedge n$ is also a stopping time.
- (Lec 3) **Stopped σ-algebra**: $\mathcal{F}_T$ is the set of all $A \in \mathcal{F}$ such that for all $n \in \mathbb{N}$, $A \cap \{T = n\} \in \mathcal{F}_n$. (You can think of $\mathcal{F}_T$ as "all events that are dependent only on information up to the stopping time". It is **not** the same as the σ-algebra generated by T.)
- (Lec 3) **Stopped random variable:** Z_T is defined as $Z_T(\omega) := Z_{T(\omega)}(\omega)$, that is, $Z_T = \sum_{n=1}^{\infty} Z_n 1_{\{T=n\}}$.

  (Note: Z_T is always $\mathcal{F}_T$-measurable.)
- (Lec 4) The convex transformation of a martingale is a submartingale. The convex **and** non-decreasing transformation of a submartingale is a submartingale.
- (Lec 5) **Uniform integrability:** A sequence of random variables $\{X_n\}$ is **uniformly integrable** if (i) each X_n is integrable, and (ii) $\lim_{a\to\infty} \sup_n \mathbb{E}(|X_n|; |X_n| \geq a) = 0$. A single integrable X is U.I.
- (Lec 6) $\{X_n\}$ is U.I. iff for all $\varepsilon > 0$, there is $\delta > 0$ such that

$$A \in \mathcal{F} \text{ with } P(A) < \delta \quad \Rightarrow \quad \mathbb{E}(|X_n|; A) < \varepsilon \text{ for all } n.$$

- (Dembo Prop 4.2.33) For any $X \in L^1(\Omega, \mathcal{F}, P)$, the collection $\{\mathbb{E}[X \mid \mathcal{H}], \mathcal{H} \subseteq \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$ is U.I.
- (Dembo Prop 5.4.4) Suppose $\{Y_n\}$ integrable and τ a stopping time for filtration $\{\mathcal{F}_n\}$. Then $\{Y_{n\wedge\tau}\}$ is U.I. if any one of the following hold:
  1. $\mathbb{E}\tau < \infty$ and a.s. $\mathbb{E}[|Y_n - Y_{n-1}| | \mathcal{F}_{n-1}] \leq c$ for some finite, non-random c.
  2. $\{Y_n I_{\tau > n}\}$ is U.I. and $Y_\tau I_{\tau<\infty}$ is integrable.
  3. $(Y_n, \mathcal{F}_n)$ is a U.I. submartingale (or supermartingale).
- If a collection $\{X_\alpha\}$ is dominated by an integrable random variable, then it is U.I.
- (Lec 5) If $\{X_n\}$ U.I., then $\sup_n \mathbb{E}|X_n| < \infty$.
- (Lec 6) If $\{X_n\}$ is U.I. and $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{L^1} X$.
- (Dembo Dfn 5.3.13) For integrable X, the sequence $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$ is called the **Doob martingale** of X w.r.t. $\{\mathcal{F}_n\}$.
- (Dembo Cor 5.3.14) A martingale $\{X_n\}$ is U.I. iff $X_n = \mathbb{E}[X_\infty \mid \mathcal{F}_n]$ is a Doob's martingale, or equivalently iff $X_n \to X_\infty$ in L^1.
- (Durrett Thm 5.7.1) If X_n is a U.I. submartingale, then for any stopping time N, $X_{n\wedge N}$ is a U.I. submartingale as well.
- (Lec 6) **Definition of absolutely continuous:** Let P and Q be two probability measures on measurable space $(\Omega, \mathcal{F})$. We say that Q **is absolutely continuous w.r.t.** P, and write $Q << P$, if $P(A) = 0 \quad \Rightarrow \quad Q(A) = 0$.

- (Lec 6) Assume that $Q << P$. Then for all $\varepsilon > 0$, there exists $\delta > 0$ such that $P(A) < \delta \Rightarrow Q(A) < \varepsilon$.

- (Lec 6) **Radon-Nikodym Theorem:** Let $(\Omega, \mathcal{F})$ be a measurable space and suppose that $\mathcal{F}$ is countably generated. Let P and Q be two probability measures on this space such that $Q << P$.

  Then, there exists a non-negative random variable L on this space such that for all $A \in \mathcal{F}$, $Q(A) = \int_A dQ = \int_A L dP$. We write $L := \frac{dQ}{dP}$.

- (Lec 7) **Dynamic Programming:** Let $\{X_n\}_{1 \leq n \leq N}$ be an $\mathcal{F}_n$-measurable sequence of integrable random variables. We want to maximize $\mathbb{E}X_T$ over all stopping times T.

  Define adapted $\{V_n\}$ by $V_N = X_N$, and $V_n = \max\{X_n, \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]\}$, and let $\tau = \inf\{X_k = V_k\}$. Then τ is the solution. (Note that $\{V_n\}$ is a supermartingale.)

- (Lec 9) **Lévy's form of the Borel-Cantelli Lemma:** Let $\{\mathcal{F}_n\}_{n \geq 1}$ be a filtration and events $A_n \in \mathcal{F}_n$ for all n. Then
$$\sum_{n=1}^{\infty} 1_{A_n} = \infty \text{ if and only if (a.s.) } \sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1}) = \infty.$$

- (Lec 9) **Almost supermartingales:** Let Z_n be a sequence of $\mathcal{F}_n$-measurable, non-negative, integrable random variables. Z_n is an **almost supermartingale** if there are 2 other non-negative, integrable, adapted sequences ξ_n and ζ_n such that $\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] \leq Z_n + \xi_n - \zeta_n$ a.s. for all n.

- (Lec 9) If Z_n is an almost supermartingale, then $Y_n = Z_n - \sum_{k=1}^{n-1}(\xi_k - \zeta_k)$ is a supermartingale.

- (Lec 9) **Stochastic gradient descent:** Assume that we have a measurable function $y = M(x)$ with the following conditions:

  - $M(x)$ is positive if $x > \theta$, $M(x)$ negative if $x < \theta$,
  - $|M(x)| \leq m$ for all x,
  - For all $\varepsilon > 0$, $\inf_{\varepsilon < x < 1/\varepsilon} M(x + \theta) > 0$ and $\sup_{-1/\varepsilon < x < -\varepsilon} M(x + \theta) < 0$.
  - Given x, we can't generate $M(x)$ exactly, but we can generate a random variable Y with mean $M(x)$ and variance $\leq \sigma^2$.

  Consider the following procedure:

  1. Choose a sequence of predetermined non-negative real numbers $\{a_n\}$ and some arbitrary X_1.
  2. (Loop) Given X_n, generate Y_n with mean $M(X_n)$ and variance $\leq \sigma^2$. Set $X_{n+1} = X_n - a_n Y_n$.

  If $\sum a_n = \infty$ and $\sum a_n^2 < \infty$, then $\lim_{n \to \infty} X_n = \theta$ a.s.

- (Lec 10) A martingale Z_n is **square-integrable** if $\mathbb{E}Z_n^2 < \infty$ for all n.

- (Lec 10) A sequence $\{Y_n\}$ is a **predictable process** if Y_n is $\mathcal{F}_{n-1}$-measurable for all n.

- (Lec 10) Let Z_n be a square-integrable martingale. Define $\sigma_n^2 = \mathbb{E}[(Z_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}]$. It can be checked that $Z_n^2 - \sum_{i=1}^{n} \sigma_i^2$ is a martingale.

  $\left\{\sum_{i=1}^{n}(Z_i - Z_{i-1})^2\right\}$ is called the **quadratic variation** of $\{Z_n\}$.

$\{A_n\} := \left\{ \sum_{i=1}^{n} \sigma_i^2 \right\}$ is called the **predictable quadratic variation** of $\{Z_n\}$. (Also called **square variation**, sometimes denoted $\langle Z \rangle_n$. Note that $\mathbb{E}[\langle Z \rangle_n] = \text{Var}\,(Z_n - Z_0)$.)

- (Lec 11) **Doob decomposition:** For a submartingale $\{Y_n\}$, let $A_n = \sum_{j=1}^{n} \mathbb{E}[Y_j - Y_{j-1} \mid \mathcal{F}_{j-1}]$. A_n is a non-negative, increasing predictable process and $Y_n - A_n$ is a martingale. $Y_n = (Y_n - A_n) + A_n$ is the **Doob decomposition** of Y_n. (Every submartingale can be decomposed into the sum of a martingale and a non-negative increasing predictable process.)

- (Dembo Thm 5.2.6) **Doob's Inequality:** For any submartingale $\{X_n\}$ and $x > 0$, let $\tau_x = \min\{k \geq 0 : X_k \geq x\}$. Then for any finite $n \geq 0$,

$$\mathbb{P}\left(\max_{0 \leq k \leq n} X_k \geq x \right) \leq \frac{1}{x} \mathbb{E}\left[X_n 1_{\{\tau_x \leq n\}} \right] \leq \frac{1}{x} \mathbb{E}[(X_n)_+].$$

- (Dembo Cor 5.2.13) **Doob's L^p inequality:** If X_n is a submartingale then for any n and $p > 1$, $\mathbb{E}\left[\left(\max_{k \leq n} X_k \right)_+^p \right] \leq q^p \mathbb{E}[(X_n)_+^p]$, where $q = p/(p-1)$.

  If X_n is in fact a martingale, then we also have $\mathbb{E}\left[\left(\max_{k \leq n} |X_k| \right)^p \right] \leq q^p \mathbb{E}[|X_n|^p]$.

- (Dembo Dfn 5.1.27) **Martingale transform:** Let $\{V_n\}$ be a predictable sequence and let $\{X_n\}$ be a sub or supermartingale. The martingale transform of $\{V_n\}$ w.r.t. $\{X_n\}$ is $Y_0 = 0$, $Y_n = \sum_{k=1}^{n} V_k(X_k - X_{k-1})$.

- (Dembo Thm 5.1.28) In the setting above,
  1. If Y_n integrable and X_n a martingale, then Y_n is also a martingale.
  2. If Y_n integrable, $V_n \geq 0$ and X_n a submartingale (or supermartingale), then Y_n is also a submartingale (or supermartingale).
  3. To have Y_n integrable, it suffices to have $|V_n| \leq c_n$ for non-random finite constants c_n, or to have $V_n \in L^q$ and $X_n \in L^p$ for all n and some $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Optional sampling theorems

- (Lec 3) **Wald's Lemma for bounded stopping times:** Let $\{\mathcal{F}_n\}$ be a filtration, $\{Z_n\}$ a martingale adapted to this filtration, T a stopping time w.r.t. this filtration. Suppose that there exists some $N \in \mathbb{N}$ such that $T \leq N$ a.s. Then $\mathbb{E}Z_T = \mathbb{E}Z_1$.

- (Durrett Thm 4.1.5) **Wald's equation:** Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X_i| < \infty$. If N is an integrable stopping time, then $\mathbb{E}S_N = \mathbb{E}X_1 \cdot \mathbb{E}N$.

- (Durrett Thm 4.1.6) **Wald's second equation:** Let $X_1, X_2, \ldots$ be i.i.d. with mean 0 and $\mathbb{E}X_n^2 = \sigma^2 < \infty$. If T is an integrable stopping time, then $\mathbb{E}S_T^2 = \sigma^2 \mathbb{E}T$.

- **Optional Sampling Theorem:** Let $\{X_n\}$ be a supermartingale, and let $\sigma \leq \tau$ be two stopping times.
  1. Assume there is a constant T s.t. $\tau \leq T$. Then $X_\sigma \geq \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$, and in particular, $\mathbb{E}[X_\sigma] \geq \mathbb{E}[X_\tau]$. If X is a martingale, then equality holds.

2. If X is non-negative and $\tau < \infty$ a.s., then $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_0] < \infty$, $\mathbb{E}[X_\sigma] \leq \mathbb{E}[X_0] < \infty$, and $X_\sigma \geq \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$.
3. If X is only adapted and integrable, then X is a martingale iff $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ for any bounded stopping time τ.

- **Martingale subsequences:** If τ_n is a monotonically increasing sequence of bounded stopping times and X is a martingale, then $\{X_{\tau_n}\}$ is a martingale w.r.t. $\{\mathcal{F}_{\tau_n}\}$.
- (Lec 8) **Stopped processes:** If Z_n is a martingale and τ is any stopping time, then $Z_{\tau\wedge n}$ is a martingale as well. (Same applies for submartingales and supermartingales.)
- Let $\{X_n\}$ be a U.I. martingale (sub-martingale resp.) and $\sigma \leq \tau$ be finite stopping times. Then $\mathbb{E}[|X_\tau|] < \infty$ and $X_\sigma = \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$ ($X_\sigma \geq \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$ resp.).
- **Optional sampling in infinite time:** Let τ be a stopping time and X a martingale. Then X_τ is integrable and $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ if one of the following conditions hold:
  1. τ is bounded.
  2. $\{X_{n\wedge\tau}\}$ is dominated by an integrable random variable Z, i.e. $|X_{n\wedge\tau}| \leq Z$ a.s.
  3. $\mathbb{E}[\tau] < \infty$ and there exists $K \geq 0$ such that $\sup_n |X_n - X_{n-1}| \leq K$.

Convergence of martingales

- (Lec 5) **(Sub)martingale Convergence Theorem**: Let $\{Z_n, \mathcal{F}_n\}_{n=1}^\infty$ be a submartingale. Suppose $\sup_n \mathbb{E}Z_n^+ < \infty$. Then there is a random variable Z taking values in $[-\infty, \infty)$ such that $Z_n \to Z$ a.s.
- (Lec 6) If $\{Z_n, \mathcal{F}_n\}$ is a U.I. martingale (or sub/super-martingale), then there is a random variable Z which is finite a.s. and $Z_n \to Z$ a.s. and in L^1.
- (Lec 6) **Lévy's Upward Convergence Theorem:** Let $\{\mathcal{F}_n\}$ be a filtration and $\mathcal{F}_\infty = \sigma\left(\bigcup_{n=1}^\infty \mathcal{F}_n\right)$. Let Y be an integrable random variable. Then $\mathbb{E}(Y \mid \mathcal{F}_n) \to \mathbb{E}(Y \mid \mathcal{F}_\infty)$ a.s. and in L^1.
- (Lec 8) If Z_n is a supermartingale with $\mathbb{E}|Z_1| < \infty$ that is uniformly bounded below by a constant, then $\lim Z_n$ exists and is finite a.s. (Same for submartingale bounded above by a constant.)
- (Lec 9) **Convergence for martingales with bounded increments:** Let Z_n be a martingale with bounded increments, i.e. there is a constant c such that for all n, $|Z_n - Z_{n-1}| \leq c$ a.s. With probability 1, either $\lim Z_n$ exists and is finite, or $\limsup Z_n = \infty$ and $\liminf Z_n = -\infty$.
- (HW4) If M_n is a martingale with bounded increments, then $\dfrac{M_n}{n} \overset{a.s.}{\to} 0$. (See Dembo Ex 5.3.41 for generalization.)
- (HW4) If $\{Z_n\}$ is a supermartingale with bounded increments, then $\limsup_{n\to\infty} \dfrac{Z_n}{n} \leq 0$ a.s.
- (Lec 9) **Convergence for almost supermartingales:** On the set $\left\{\sum_{n=1}^\infty \xi_n < \infty\right\}$, $\lim Z_n$ exists and is finite and $\sum \zeta_n < \infty$ a.s.
- (Lec 10) **Martingale SLLN:** Let

- $\{S_n, \mathcal{F}_n\}_{n\geq 1}$ be a martingale with mean 0,
- $X_n = S_n - S_{n-1}$ (where $S_0 = 0$), with the assumption that $\mathbb{E}[X_n^2] < \infty$ for all n,
- $\sigma_n^2 = \mathbb{E}[X_n^2 \mid \mathcal{F}_{n-1}]$,
- c_n an $\mathcal{F}_{n-1}$-measurable random variable, with the assumption that $c_n > 0$ and increasing.

Then, on the set $\left\{ \sum_{n=1}^{\infty} \frac{\sigma_n^2}{c_n^2} < \infty \text{ and } \lim_{n\to\infty} c_n = \infty \right\}$, $\frac{S_n}{c_n} \to 0$ a.s.

- (Lec 10) Let $X_n \in \{0,1\}$ for all n, X_n $\mathcal{F}_n$-measurable. Let $S_n = \sum_{i=1}^{n} X_i$, and $p_n = P(X_n = 1 \mid \mathcal{F}_{n-1})$.

  Then on the set $\left\{ \sum_{i=1}^{\infty} p_i = \infty \right\}$, $\lim_{n\to\infty} \frac{S_n}{\sum_{i=1}^n p_i} = 1$.

- (Lec 10) **Convergence of square-integrable martingales:** If $\{Z_n\}$ is a square-integrable martingale and $\sup_n \mathbb{E}Z_n^2 < \infty$, then there exists Z such that $Z_n \to Z$ a.s. and in L^2.

- (Lec 10) For a square-integrable martingale $\{Z_n\}$, $\lim Z_n$ exists and is finite a.s. on the set $\left\{ \sum_{n=1}^{\infty} \sigma_n^2 < \infty \right\}$.

Ergodic theory (Lec 11-13)

- (Lec 11) Let $(\Omega, \mathcal{F}, P)$ be a probability space. A measurable map $\varphi : \Omega \mapsto \Omega$ is called a **measure-preserving transform** if $P(\varphi^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$.

- (Lec 12) If φ is measure-preserving, $\int Y dP = \int Y \circ \varphi dP$ for all Y.

- (Lec 11) Examples of measure-preserving transforms:
  - $(\Omega, \mathcal{F}, P) = ([0,1), \mathrm{Borel}, \mathrm{Leb})$, $\theta \in \Omega$, and let $\varphi(x) = x + \theta \pmod 1$. This is ergodic iff θ is irrational.
  - Probability space as above, $\varphi(x) = 2x \pmod 1$.
  - **Bernoulli shift:** $\Omega = \{0,1\}^{\mathbb{N}}$ (where $\mathbb{N} = \{0, 1, \dots\}$), $\mathcal{F}$ the product σ-algebra (i.e. the smallest σ-algebra under which all coordinate maps are measurable), P be the law of an infinite i.i.d. Bernoulli(1/2) sequence, $\varphi((\omega_0, \omega_1, \dots)) := (\omega_1, \omega_2, \dots)$. This is ergodic.

- (Lec 11) Let φ be measure-preserving on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{I} = \{A \in \mathcal{F} : \varphi^{-1}(A) = A\}$. Then $\mathcal{I}$ is a σ-algebra, and is called the **invariant σ-algebra of** φ.

- (HW5) For $A \in \mathcal{I}$, $1_A \circ \varphi = 1_A$, and for $\mathcal{I}$-measurable Y, $Y \circ \varphi = Y$ a.s.

- (Lec 11) A measure-preserving transform φ is called **ergodic** if for all $A \in \mathcal{I}$, $P(A) = 0$ or 1.

- (Lec 12) **Maximal Ergodic Theorem:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, φ a measure-preserving transform with invariant σ-algebra $\mathcal{I}$, and X a random variable in L^1.

  Let $X_j(\omega) := X(\varphi^j(\omega))$, $S_k = X_0 + X_1 + \dots + X_k$, and $M_k = \max\{0, S_0, S_1, \dots, S_k\}$.

  Then $\mathbb{E}[X; M_k > 0] \geq 0$.

- (Lec 12) **Birkhoff's Ergodic Theorem:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, φ a measure-preserving transform with invariant σ-algebra $\mathcal{I}$, and X a random variable in L^1. Then $\dfrac{1}{n}\displaystyle\sum_{m=0}^{n-1} X(\varphi^m(\omega)) \to \mathbb{E}[X \mid \mathcal{I}]$ a.s. and in L^1.

  In particular, if φ is ergodic, then $\dfrac{1}{n}\displaystyle\sum_{m=0}^{n-1} X(\varphi^m(\omega)) \to \mathbb{E}X$ a.s. and in L^1 ("space average is equal to time average").

- (Lec 13) Let $\Omega = [0, 1)$, $\varphi(x) = x + \theta \pmod 1$, where θ is irrational. Take any subinterval $[a, b)$ of $[0, 1)$. Then, for any $x \in [0, 1)$, $\dfrac{1}{n}\displaystyle\sum_{m=0}^{n-1} 1\{\varphi^m(x) \in [a, b)\} \to b - a$.

- (Lec 13) **von Neumann's Ergodic Theorem:** Same as Birkhoff's, except $X \in L^2$ and the convergence is in L^2.

Markov chains (Lec 14-18)

Set-up: $(\Omega, \mathcal{F}, P)$ a probability space, $\{\mathcal{F}_n\}$ a filtration, $\{X_n\}$ an adapted sequence of random variables taking values in $(S, \mathcal{S})$.

- (Lec 14) **Transition probabilities:** $p_n(x, y) = P(X_{n+1} = y \mid X_n = x)$.

- (Lec 14) **Chapman-Kolmogorov equations:** Let $P_n = \Big(p_k(x, y)\Big)_{x,y\in S}$, and $P^{(n)} = P_0 P_1 \dots P_{n-1}$. Then $P(X_n = y \mid X_0 = x) = P^{(n)}(x, y)$.

- (Lec 14) **Time-homogeneous:** The transition probabilities are the same at each time step. In this case, the transition probability matrix is P with entries $P_{ij} = P(X_{n+1} = j \mid X_n = i)$.
  - If p_0 is the initial distribution of X_0 (row vector), then the distribution of X_n is $p_0 P^n$.
  - $P^n f$ is the conditional expectation of $f(X_{i+n})$ given X_i. Thus, given an initial distribution p_0, we have $\mathbb{E}[f(X_n)] = p_0 P^n f$.

- (Lec 14) **Markov property:** Given a function $f : S \times S \times \ldots \to \mathbb{R}$ which is measurable w.r.t. the product σ-algebra, let $g(x) = \mathbb{E}[f(X_0, X_1, X_2, \dots) \mid X_0 = x] =: \mathbb{E}_x[f(X_0, X_1, X_2, \dots)]$. Then, for any n, $\mathbb{E}[f(X_n, X_{n+1}, X_{n+2}, \dots) \mid X_n = x] = g(x)$.

- (Lec 14) **Strong Markov property:** Let T is a stopping time w.r.t. $\{\mathcal{F}_n\}_{n\geq 0}$. Then, on the set $\{T < \infty\}$, $\mathbb{E}[f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T] = g(X_T)$.

- (Lec 14) Every discrete-time Markov chain has the Markov property and strong Markov property.

From here on, assume that the Markov chain is time-homogeneous and takes values on a countable state space S.

- (Lec 15) **Hitting time:** For $x \in S$, the **first hitting time** of x is $T_x := \inf\{n \geq 1 : X_n = x\}$. (Note: Time 0 doesn't count.)

- (Lec 15) Let $\rho_{xy} := P(T_y < \infty \mid X_0 = x)$. In particular, ρ_{xx} is the probability of ever returning to x given that the chain starts at x. A state x is **recurrent** if $\rho_{xx} = 1$, and is **transient** otherwise.

- (Lec 15) Notation: Let $N(x) = \sum_{n=1}^{\infty} 1_{\{X_n = x\}}$, i.e. the number of visits to x (except time 0).

  Let $p_{xy}^{(n)} := P(X_n = y \mid X_0 = x)$.

- (Lec 15) **Theorem for recurrence:** The following are equivalent:

  (a) x is recurrent (i.e. $\rho_{xx} = 1$).

  (b) $\mathbb{E}_x N(x) = \infty$, where $\mathbb{E}_x$ means $\mathbb{E}[\cdot \mid X_0 = x]$.

  (c) $P_x(N(x) = \infty) = 1$.

  (d) $\sum_{n=1}^{\infty} p_{xx}^{(n)} = \infty$.

- (Lec 16) A state y is **accessible** from a state x if $p_{xy}^{(n)} > 0$ for some $n \geq 0$. If so, we write $x \to y$. Two states x and y are **communicating** if $x \to y$ and $y \to x$. We write $x \leftrightarrow y$.

- (Lec 16) The Markov chain is said to be **irreducible** if the number of equivalence classes is 1.

- (Lec 16) Recurrence (and therefore, transience) is a class property.

- (Lec 16) If the state space S is finite, then there exists at least one recurrent state. Hence, if the Markov chain is irreducible and state space is finite, then all states are recurrent.

- (Lec 16) The **period** of a state x is the greatest common divisor of all $n \geq 1$ such that $p_{xx}^{(n)} > 0$. It is denoted by $d(x)$. A state is said to be **aperiodic** if its period is 1. States that communicate have the same period.

- (Lec 16) For any recurrent state x, let μ_{xx} := expected time of first return to x starting from x. A recurrent state x is called **positive recurrent** if $\mu_{xx} < \infty$. If $\mu_{xx} = \infty$, it is called **null recurrent**.

- (Lec 17) If x is a recurrent state, then $\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p_{xx}^{(m)} = \frac{1}{\mu_{xx}}$.

- (Lec 17) If x and y communicate and are recurrent, then $\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p_{yx}^{(m)} = \frac{1}{\mu_{xx}}$.

- (Lec 18) Positive recurrence and null recurrence are also class properties.

- (Lec 18) A Markov chain with finite state space cannot have a null recurrent state.

- (Lec 18) If S' is a positive recurrent equivalence class and $|S'| < \infty$, then $\sum_{y \in S'} \frac{1}{\mu_{yy}} = 1$.

  In particular, if S is finite and the chain is irreducible (i.e. $|S| = |S'|$), then $\sum_{x \in S} \frac{1}{\mu_{xx}} = 1$.

- (Lec 18) **Stationary distribution:** A probability measure π on S is called a **stationary distribution/invariant measure** for the chain if, for all $y \in S$, $\sum_{x} \pi_x p_{xy} = \pi_y$.

- (Lec 18) If S is finite and the chain is irreducible, then $\pi_x = \frac{1}{\mu_{xx}}$ is the unique stationary distribution.

- (Lec 18) If S is finite and the chain is irreducible and aperiodic, then $\pi_x = \dfrac{1}{\mu_{xx}}$ is the unique stationary distribution. Moreover, $\lim_{n\to\infty} p_{xy}^{(n)} = \dfrac{1}{\mu_{yy}}$.
- (305C) If S is finite, then there exists at least one stationary distribution. (This is a corollary of Perron-Frobenius Theorem v2.)
- (305C) If S finite, irreducible, aperiodic and has stationary distribution π, then for all starting points ω_0, $\lim_{n\to\infty} \mathbb{P}_{\omega_0}(X_n = \omega) = \pi(\omega)$ for all ω. Also,

$$\mathbb{P}_{\omega_0}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \sum_{\omega} \pi(\omega) f(\omega)\right) = 1.$$

Renewal theory (Lec 17)

Set-up: $X_1, X_2, \ldots$ i.i.d. non-negative random variables with $P(X_1 = 0) < 1$. $\mu := \mathbb{E}X_1$ (possibly infinite), $S_n = \sum_{i=1}^{n} X_i$ with $S_0 = 0$.

- (Lec 17) For real number $t \geq 0$, let $N(t) := \sup\{n : S_n \leq t\}$. Then $\{N(t) : t \geq 0\}$ is called a **renewal process**. (We can think of this process as replacing lightbulbs, X_i is the lifetime of the i^{th} lightbulb, and when it dies, we replace it with a new lightbulb. S_n can be thought of as the time till the n^{th} lightbulb goes off.)
- (Lec 17) **Renewal function:** $m(t) := \mathbb{E}[N(t)]$. For all t, $m(t) < \infty$.
- (Lec 17) **Elementary renewal theorem:** $\lim_{t\to\infty} \dfrac{m(t)}{t} = \dfrac{1}{\mu}$.

First-Passage Percolation (Lec 19)

Set-up: We have the lattice $\mathbb{Z}^d$.

- On each edge e, we have a non-negative random variable X_e, called the weight of the edge. Assume that the X_e are i.i.d.
- The weight of a path is equal to the sum of edge weights along the path.
- The **first-passage time** from x to y, denoted T_{xy}, is the minimum of the weights of all paths from x to y.
- Let $T_n = T_{0,ne_1}$, where $e_1 = (1, 0, \ldots, 0)$.
- Assume that there exist $0 < a < b$ such that $\mathbb{P}(a \leq X_e \leq b) = 1$. Then Var $T_n \leq Cn$, where C depends only on a, b and d. (We can take $C = b^3/a$.)
- Assume that $\mathbb{E}X_e < \infty$. Then $\mu = \lim_{n\to\infty} \dfrac{\mathbb{E}T_n}{n}$ exists. Moreover, if $P(X_e = 0) = 0$, then $\mu > 0$.

Concentration inequalities (Lec 19-20)

- (Lec 19) **Efron-Stein Inequality:** Let $X_1, \ldots, X_n$ be independent random variables. Let $X'_1, \ldots, X'_n$ be another set of independent random variables, independent of $X_1, \ldots, X_n$, such that X'_i has the same distribution as X_i for all i. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a measurable function such that $\mathbb{E}[W^2] < \infty$, where $W = f(X_1, \ldots, X_n)$. Then,

$$\text{Var } W \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E}\left[(\Delta_i f)^2\right] = \sum_{i=1}^n \mathbb{E}\left[(f(X_1, \ldots, X_n) - f(X_1, \ldots, X'_i, \ldots, X_n))^2\right].$$

- (Lec 20) **Azuma-Hoeffding Inequality:** Let $\{M_k\}_{0 \leq k \leq n}$ be a martingale adapted to some filtration. Let $X_k = M_k - M_{k-1}$. Suppose that $|X_k| \leq c_k$ a.s. for each k, where $c_1, \ldots, c_n$ are constants. Then for all $t > 0$,

$$P(M_n - M_0 \geq t) \leq \exp\left(-\frac{t^2}{2\sum c_k^2}\right), \qquad P(M_n - M_0 \leq -t) \leq \exp\left(-\frac{t^2}{2\sum c_k^2}\right).$$

- (Lec 20) **Bounded Difference Inequality:** Let $X_1, \ldots, X_n$ be independent random variables. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a measurable function such that there exist $c_1, \ldots, c_n$ with the property that for all $x_1, \ldots, x_n$, $x'_1, \ldots, x'_n$, $|f(x_1, \ldots, x_n) - f(x_1, \ldots, x'_i, \ldots, x_n)| \leq c_i$.

  Let $W = f(X_1, \ldots, X_n)$. Then for all $t > 0$,

$$P(W - \mathbb{E}W \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right), \qquad P(W - \mathbb{E}W \leq -t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Other Stuff

General random walks

Let $X_1, X_2, \ldots$ be iid, and $S_n = X_1 + \cdots + X_n$. $\{S_n\}$ is a random walk.

- (Durrett Thm 4.1.2) For a random walk on $\mathbb{R}$, there are only 4 possibilities, one of which has probability 1: (i) $S_n = 0$ for all n, (ii) $S_n \to \infty$, (iii) $S_n \to -\infty$, and (iv) $\liminf S_n = -\infty$ and $\limsup S_n = \infty$.
- If $\mathbb{E}X_i > 0$, then $S_n \overset{a.s.}{\to} \infty$. If $\mathbb{E}X_i < 0$, then $S_n \overset{a.s.}{\to} -\infty$. If $\mathbb{E}X_i = 0$ and $P(X_i = 0) < 1$, then $\limsup S_n = \infty$ with probability 1 and $\liminf S_n = -\infty$ with probability 1.
- (HW5) Let $M_n = \max\limits_{0 \leq k \leq n} S_k$. If X_i's have finite mean, then $\mathbb{E}M_n = \displaystyle\sum_{k=1}^n \frac{\mathbb{E}[S_k^+]}{k}$.
- (Durrett Thm 6.3.5) **Reflection principle:** Assume that the X_i's have a distribution symmetric about 0. Then if $a > 0$, we have $P\left(\sup\limits_{m \leq n} S_m > a\right) \leq 2P(S_n > a)$. (Still holds if $>$s replaced with $\geq$s.)

Simple symmetric random walks on $\mathbb{Z}^d$

- (310B Lec 15) For simple symmetric walk on $\mathbb{Z}^d$, 0 (or any other state) is recurrent if $d = 1, 2$, and transient if $d \geq 3$.

Simple symmetric random walk on $\mathbb{Z}$

$X_i = 1$ or -1, each with probability $1/2$.

- (310B Lec 3) $\{S_n^2 - n\}$ is a martingale.
- (310B HW2) $\{S_n^4 - 6nS_n^2 + 3n^2 + 2n\}$ is a martingale.
- (310B Lec 3 & 4) Let $T = \inf\{k : S_k = a \text{ or } b\}$, where $a < 0 < b$ are 2 integers. Then T is finite a.s., $\mathbb{E}T = ab$, and $P(S_T = a) = \dfrac{b}{b-a}$.
- (310B Lec 16) Let $T = \inf\{n \geq 1 : S_n = 0\}$. Then $\mathbb{E}T = \infty$, i.e. 0 is null recurrent.
- (Durrett Thm 3.1.2) If $2k/\sqrt{2n} \to x$, then $P(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}$.
- (Durrett Thm 3.1.3) **De Moivre-Laplace Theorem:** If $a < b$, then as $m \to \infty$, $P\left(a \leq \dfrac{S_m}{\sqrt{m}} \leq b\right) \to \displaystyle\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.
- (Durrett Thm 4.1.7) Let X_i be i.i.d. with mean 0 variance 1, and let $T_c = \inf\{n \geq 1 : |S_n| > c\sqrt{n}\}$. Then $\mathbb{E}T_c$ is finite for $c < 1$, and infinite for $c \geq 1$.

Biased random walk on $\mathbb{Z}$

Here, S_n increments by 1 with probability p, and decrements by 1 with probability $q = 1-p$. Let $\varphi(x) = (q/p)^x$.

- (310B Lec 4) $\varphi(S_n) = \left(\dfrac{q}{p}\right)^{S_n}$ is a martingale.
- (Durrett Thm 5.7.7 proof) $\{S_n - (p-q)n\}$ is a martingale.
- (310B Lec 4) Let $T = \inf\{k : S_k = a \text{ or } b\}$, where $a < 0 < b$ are 2 integers. Then T is finite a.s. and $P(S_T = a) = \dfrac{1 - (q/p)^b}{(q/p)^a - (q/p)^b}$.
- (Durrett Thm 5.7.7) Let $T_x = \inf\{n : S_n = x\}$. Assume $p > 1/2$.
  - If $a < 0$, then $P\left(\min_n S_n \leq a\right) = P(T_a < \infty) = (q/p)^{-a}$.
  - If $b > 0$, then $P(T_b < \infty) = 1$ and $\mathbb{E}T_n = \dfrac{b}{2p-1}$.

Other

- (Dembo Lem 5.2.7) **Lenglart's bound:** Let $(Z_n, \mathcal{F}_n)$ be a non-negative submartingale with $Z_0 = 0$. Let $V_n = \max_{0 \leq k \leq n} Z_k$ and let A_n be the $\mathcal{F}_n$-predictable sequence in Doob's decomposition of Z_n. Then, for any $\mathcal{F}_n$-stopping time τ and all $x, y > 0$,

$$\mathbb{P}(V_\tau \geq x, A_\tau \leq y) \leq \frac{\mathbb{E}[A_\tau \wedge y]}{x}.$$

Fruther, in this case $\mathbb{E}[V_\tau^p] \leq \left[1 + \frac{1}{1-p}\right] \cdot \mathbb{E}[A_\tau^p]$ for any $p \in (0,1)$.