Theory of Probability Notes

These are notes made in preparation for oral exams involving the following topics in probability: Random walks, Martingales, and Markov Chains. Textbook used: "Probability: Theory and Examples," Durrett.

Durrett Chapter 4 (Random Walks).

Random Walk: Let $X_1, X_2, \ldots$ be iid taking values in $\mathbb{R}^d$ and let $S_n = X_1 + \ldots + X_n$. S_n is a random walk.

This section is concerned with the properties of the sequence $S_1(\omega), S_2(\omega), \ldots$

Stopping Times

Stopping Time: Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. A stopping time $T : \Omega \to \mathbb{Z}_+ \cup \{+\infty\}$ is a random variable such that $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$, or equivalently, $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

- The constant times (e.g., $T \equiv 10$) are always stopping times.
- Let $(X_n)_{n\geq 0}$ be an adapted process. Fix $A \in \mathcal{B}_{\mathbb{R}}$. Then the first entry time into A,

  notated as $T_A := \inf\{n \geq 0 : X_n \in A\}$, with the convention that $\inf \emptyset = +\infty$ is a stopping time.

Associated with each stopping time N is a σ-field $\mathcal{F}_N$ = information known at time N. $\mathcal{F}_N$ is the collection of sets A that have $A \cap \{N = n\} \in \mathcal{F}_n$, $\forall n < \infty$. When $N = n$, A must be measurable with respect to information known at time n.

Stopping Times Lemma: Let $S, T, (T_n)_{n\geq 0}$ be stopping times on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$. Then the following are stopping times: *i)* $S \wedge T := \min(S, T)$, *ii)* $S \vee T := \max(S, T)$,

iii) $S + T$, *iv)* $\liminf_n T_n$ and $\inf_n T_n$, *v)* $\limsup_n T_n$ and $\sup_n T_n$.

Proof of i) $\{S \wedge T \leq n\} = (\{S \leq n\}^c \cap \{T \leq n\}^c)^c \in \mathcal{F}_n$. ■

Notational Conventions for Measurable Space $(S, \mathcal{S})$: Assuming the random sequence $S_1(\omega), S_2(\omega), \ldots$ defined above, Let $\Omega := \{(\omega_1, \omega_2, \ldots) : \omega_i \in S\}$, $\mathcal{F} := \mathcal{S} \times \mathcal{S} \times \ldots$, $\mathbb{P} := \mu \times \mu \times \ldots$ where μ is the distribution of X_i, and $X_n(\omega) =: \omega_n$.

Finite Permutation of $\mathbb{N}$: A map π from $\mathbb{N}$ onto $\mathbb{N}$ so that $\pi(i) \neq i$ for only finitely many i.

Permutable Event A: An event $A \in \mathcal{F}$ is permutable if $\pi^{-1}A \equiv \{\omega : \pi\omega \in A\} = A$ for any finite permutation π.

Symmetric Function: A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be symmetric if $f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$ for each $(x_1, \ldots, x_n) \in \mathbb{R}$ and for each permutation π of $\{1, 2, \ldots, n\}$.

Exchangeable σ-field ε: Let $X_1, X_2, \ldots$ be a sequence of r.v.s on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $F_n := \left\{f : R^n \to R \text{ symmetric m'ble}\right\}$, and $\varepsilon_n := \sigma(F_n, X_{n+1}, X_{n+2}, \ldots)$. The exchangeable σ-field ε is defined as $\varepsilon := \cap_{n=1}^{\infty} \varepsilon_n$.

The next results shows that for an iid sequence, there is no difference between ε and $\mathcal{T}$, they are both trivial.

Hewitt Savage 0-1 Law (Gen. of Kolmogorov 0-1) **D4.1.1**: Let ε be the exchangeable σ-field of iid $X_1, X_2, \ldots$ Then $\mathbb{P}(A) \in \{0, 1\}$ for any $A \in \varepsilon$.

Random Walk Possibilities (Consequence of the HSL) **D4.1.2**: For random walks on $\mathbb{R}$, there are 4 possibilities, one of which has probability 1. i) $S_n = 0$ for all n, ii/iii) $S_n \to \pm\infty$, iv) $-\infty = \liminf S_n < \limsup S_n = \infty$.

Proof: Theorem D4.1.1 implies $\limsup S_n$ is a constant $c \in [-\infty, \infty]$. Let $S'_n = S_{n+1} - X_1$. Since S'_n has the same distribution as S_n, it follows that $c = c - X_1$. If c is finite, subtracting c from both sides we conclude $X_1 \equiv 0$ and i) occurs. Turning the last statement around, we see that if X_1 is not equivalently zero, then $c = -\infty$ or ∞. The same analysis applies to the $\liminf$. Discarding the impossible combination $\limsup S_n = -\infty$ and $\liminf S_n = +\infty$, we have proved the results. ■

Exercises D4.1.1-7

Theorem D4.1.3: Let $X_1, X_2, \ldots$ be iid, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and T be a stopping time with $\mathbb{P}(T < \infty) > 0$. Conditional on $\{T < \infty\}$, $\left\{X_{T+n}, n \geq 1\right\}$ is independent of $\mathcal{F}_T$ and has the same distribution as the original sequence.

Let T a stopping time, $T_0 := 0,\ T_n(\omega) := T_{n-1}(\omega) + T(\theta^{T_{n-1}}(\omega))$ for $n \geq 1$, and $t_n(\omega) := T(\theta^{T_{n-1}}(\omega))$, we now extend D4.1.3:

Theorem D4.1.4: Suppose $\mathbb{P}(T < \infty) = 1$. Then the "random vectors" $V_n = (t_n, X_{T_{n-1}+1}, \ldots, X_{T_n})$ are independent and identically distributed.

Proof: It is clear from theorem D4.1.3 that V_n and V_1 have the same distribution. The independence also follows from theorem D4.1.3 and induction since $V_1, \ldots, V_{n-1} \in \mathcal{F}(T_{n-1})$. ■

Exercises D4.1.8-11

Wald's Identity: Let $\xi_1, \xi_2, \ldots$ be iid with finite mean $\mu := \mathbb{E}[\xi_n]$. Set ξ_0 and let $S_n = \xi_1 + \ldots + \xi_n$. Let T be a stopping time with $\mathbb{E}[T] < \infty$. Then, $\mathbb{E}[S_T] = \mu\mathbb{E}[T]$.

Exercises D4.1.12-14

Theorem D4.1.6 (Wald's 2nd Equation): Let $\xi_1, \xi_2, \ldots$ be iid with $\mu := \mathbb{E}[\xi_n] = 0$, $\mathbb{E}[\xi_n^2] =: \sigma^2 < \infty$, and $S_n = \xi_1 + \xi_2 + \ldots$. If T is a stopping time with $\mathbb{E}[T] < \infty$, then $\mathbb{E}[S_T^2] = \sigma^2\mathbb{E}[T]$.

Theorem D4.1.7 (used in the proof of D4.1.6): Let $X_1, X_2, \ldots$ be iid with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = 1$, and let $T_c = \inf\left\{n \geq 1 : |S_n| > cn^{\frac{1}{2}}\right\}$, then $\mathbb{E}[T_c]\begin{cases} < \infty \text{ for } c < 1, \\ = \infty \text{ for } c \geq 1. \end{cases}$

Lemma D4.1.8: If T is a stopping time with $\mathbb{E}[T] = \infty$, then $\frac{\mathbb{E}[X_{T\wedge n}^2]}{\mathbb{E}[T\wedge n]} \to 0$.

Recurrence

Recurrent Value: $x \in S$ is considered recurrent if, for every $\varepsilon > 0$, we have $\mathbb{P}\left(|S_n - x| < \varepsilon \text{ i.o.}\right) = 1$.

Possible Value (of random walk): $x \in S$ is a possible value if, for any $\varepsilon > 0$, $\exists n$ such that $\mathbb{P}(|S_n - x| < \varepsilon) > 0$.

Let $\mathcal{V}$ be the set of all recurrent values, and $\mathcal{U}$ be the set of all possible values, then:

Theorem D4.2.1: $\mathcal{V}$ is either $\emptyset$ or a closed a subgroup of $\mathbb{R}^d$. If $\mathcal{V}$ is a closed subset, then $\mathcal{V} = \mathcal{U}$.

Transient/Recurrent: If $\mathcal{V} = \emptyset$, the random walk is said to be transient, otherwise it is called recurrent.

Let $\tau_0 = 0$ and $\tau_n = \inf\{m > \tau_{n-1} : S_m = 0\}$ be the time of the nth return to 0.

Theorem D4.2.2: For any random walk, the following are equivalent:

i) $\mathbb{P}(\tau_1 < \infty) = 1$, ii) $\mathbb{P}\left(S_m = 0 \text{ i.o.}\right) = 1$, and iii) $\sum_{m=0}^{\infty} \mathbb{P}(S_m = 0) = \infty$.

Proof: If $\mathbb{P}(\tau_1 < \infty) = 1$, then $\mathbb{P}(\tau_n < \infty) = 1$ for all n and $\mathbb{P}\left(S_m = 0 \text{ i.o.}\right) = 1$. Let $V := \sum_{m=0}^{\infty} 1_{\{S_m=0\}} = \sum_{n=0}^{\infty} 1_{\{\tau_n<\infty\}}$ be the number of visits to 0, counting the visit at time 0. Taking expected value and using Fubini's theorem to put the expected value inside the sum: $\mathbb{E}[V] = \sum_{m=0}^{\infty} \mathbb{P}(S_m = 0) = \sum_{n=0}^{\infty} \mathbb{P}(\tau_n < \infty) = \sum_{n=0}^{\infty} \mathbb{P}(\tau_1 < \infty)^n = \frac{1}{1-\mathbb{P}(\tau_1<\infty)}$. The 2nd equality shows that ii ⇒ iii and, in combinations with the last two, shows that if i) is false, then iii) is false (i.e.,iii ⇒ i). ■

Theorem D4.2.3 (6.33): The Simple (lattice) Symmetric Random Walk (SSRW) S_n on $\mathbb{Z}^d$ is recurrent for $d = 1,2$ and is transient for $d \geq 3$.

Examples 6.27-6.32

Exercise 6.34: Consider SSRW on $\mathbb{Z}$. Show that $\mathbb{E}_0[T_0] = +\infty$. Prove the same for SSRW on $\mathbb{Z}^2$.

Theorem D4.2.6: The convergence (divergence) of $\Sigma_n\mathbb{P}(|S_n| < \varepsilon)$ for any single value of $\varepsilon > 0$ is sufficient to determine the transience (recurrence) of S_n.

The proof of this theorem uses the following lemmas:

Lemma D4.2.4 (Gen. of D4.2.2): If $\sum_{n=1}^{\infty} \mathbb{P}(|S_n| < \varepsilon) < \infty$, then $\mathbb{P}\left(|S_n| < \varepsilon \text{ i.o.}\right) = 0$. If $\sum_{n=1}^{\infty} \mathbb{P}(|S_n| < \varepsilon) = \infty$, then $\mathbb{P}\left(|S_n| < 2\varepsilon \text{ i.o.}\right) = 1$.

Lemma D4.2.5: Let m be an integer ≥ 2. Then, $\sum_{n=0}^{\infty} \mathbb{P}(|S_n| < m\varepsilon) \leq (2m)^d \sum_{n=0}^{\infty} \mathbb{P}(|S_n| < \varepsilon)$.

Theorem D4.2.7 (Chung-Fuchs Theorem): Suppose $d = 1$. If the weak law of large numbers holds in the form $\frac{S_n}{n} \xrightarrow{p} 0$, then S_n is recurrent.

Theorem D4.2.8: If S_n is a random walk in $\mathbb{R}^2$ and $\frac{S_n}{n^{\frac{1}{2}}} \Rightarrow$ a non-degenerate normal distribution (non-degenerate means $\dim\left(\text{support}(\lim_n S_n/\sqrt{n}\,)\right) = 2$), then S_n is recurrent.

Remark: The conclusion is also true if the limit is degenerate, but in that case, the random walk is essentially one-(or zero)-dimensional, and the result follows from the Chung-Fuchs Theorem.

Let $\varphi := \mathbb{E}[e^{itX_j}]$ be the characteristic function of one of the steps of the random walk, then we have:

Theorem D4.2.9: Let $\delta > 0$. S_n is recurrent if and only if $\int_{(-\delta,\delta)^d} \text{Re}\, \frac{1}{1-\varphi(y)} dy = \infty$.

Theorem D4.2.10: Let $\delta > 0$. S_n is recurrent if and only if $\sup_{r<1} \int_{(-\delta,\delta)^d} \text{Re}\, \frac{1}{1-r\varphi(y)} dy = \infty$.

The next two lemmas are used in the proof of D4.2.10:

Lemma D4.2.11 (**Parseval Relation**): Let μ and ν be probability measures on $\mathbb{R}^d$ with characteristic functions φ and ψ. Then: $\int \psi(t)\mu(dt) = \int \varphi(x)\nu(dx)$. **Proof**: Since $e^{it\cdot x}$ is bounded, by Fubini:
$\int \psi(t)\mu(dt) = \int\int e^{itx}\nu(dx)\mu(dt) = \int\int e^{itx}\mu(dt)\nu(dx) = \int \varphi(x)\nu(dx)$. ■

Lemma D4.2.12: If $|x| \leq \frac{\pi}{3}$, then $1 - \cos x \geq \frac{x^2}{4}$. **Proof**: It suffices to prove the results for $x > 0$. If $z \leq \frac{\pi}{3}$, then $\cos z \geq \frac{1}{2}$, so $\sin y = \int_0^x \cos z dz \geq \frac{y}{2}$, and $1 - \cos x = \int_0^x \sin y dy \geq \int_0^x \frac{y}{2} dy = \frac{x^2}{4}$. ■

Truly Three Dimensional Random Walk: A random walk S_n in $\mathbb{R}^3$ is considered truly three-dimensional if the distribution of X_1 is $\mathbb{P}(X_1 \cdot \theta \neq 0) > 0$, for all nonzero vectors θ.

Theorem D4.2.13: No truly three-dimensional random walk is recurrent.

In conclusion:

- S_n is recurrent in $d = 1$ if $S_n/n \xrightarrow{p} 0$.
- S_n is recurrent in $d = 2$ if $S_n/\sqrt{n} \Rightarrow$ a non-degenerate normal 2D distribution.
- S_n is transient in $d \geq 3$ if it is "truly N-dimensional."

Durrett Chapter 5 (Martingales).

Conditional Expectation: Consider a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X \in L^1(\Omega, \mathcal{F}, P)$. Let $G \subseteq \mathcal{F}$ be a sub-σ-field. We define $\mathbb{E}[X|G]$, the conditional expectation of X given G, as a random variable Y such that:

i) Y is G-measurable and $\mathbb{E}|Y| < \infty$.

ii) $\mathbb{E}[\mathbb{E}[X|G]1_A] = \mathbb{E}[Y1_A] = \mathbb{E}[X1_A]$ for any $A \in G$.

Y defined in this way is also unique.

Conditional Expectation Characterizations:

- $\mathbb{E}[X|A]$, where A is an event, is a scalar and the expected value of X given that A occurs.
- $\mathbb{E}[X|Y]$, where Y is a r.v., is a random variable whose value at $\omega \in \Omega$ is $\mathbb{E}[X|A]$, where A is the event $\{Y = Y(\omega)\}$.
- $\mathbb{E}[X|1_A]$ is the case $Y = 1_A$, and $1_A(\omega)$ is 1 if $\omega \in A$ and 0 otherwise.

This is the random variable that returns $\mathbb{E}[X|A]$ if $\omega \in A$ and $\mathbb{E}[X|A^c]$ if $\omega \notin A$.

Exercise D5.1.1: Generalize the last argument (not shown on this document, see Durrett) to show that if $X_1 = X_2$ on $B \in \mathcal{F}$, then $\mathbb{E}[X_1|\mathcal{F}] = \mathbb{E}[X_2|\mathcal{F}]$ a.s. on B.

Absolute Continuity: Let ν and μ be two σ-finite measures on $(\Omega, \mathcal{F})$. We say ν is absolutely continuous with respect to μ, and write $\nu << \mu$, if $\mu(A) = 0 \Rightarrow \nu(A) = 0$, for each $A \in \mathcal{F}$.

Radon-Nikodym Lemma: Let ν and μ be two σ-finite measures on $(\Omega, \mathcal{F})$. Then $\nu << \mu$ if and only if there exists a $\mathcal{F}$-measurable function $f : \Omega \to [0, \infty)$ such that $\nu(B) = \int_B f d\mu$, for all $B \in \mathcal{F}$.

Existence of Conditional Expectation: Let $X \in L^1(\Omega, \mathcal{F}, P)$ and G be a sub-σ-field. Then $\mathbb{E}[X|G]$ exists.

Exercise D5.1.2: Baye's formula - Let $G \in \mathcal{G}$ and show that $\mathbb{P}(G|A) = \frac{\int_G \mathbb{P}(A|\mathcal{G}) d\mathbb{P}}{\int_\Omega \mathbb{P}(A|\mathcal{G}) d\mathbb{P}}$.

Properties of Conditional Expectation

Trivial Conditional Expectations:

a) If $X \in G$, then $\mathbb{E}[X|G] = X$ a.s.

b) If $G = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|G] = \mathbb{E}[X]$.

c) If X is independent of G, then $\mathbb{E}[X|G] = \mathbb{E}[X]$ a.s.. To prove this, observe that $\mathbb{E}[X]$ is G-measurable (in that the preimage of a constant is $\in \{\emptyset, \Omega\} \in G$) and for any $A \in G$ we have: $\mathbb{E}[X1_A] = \mathbb{E}[X]\mathbb{E}[1_A] = \mathbb{E}[\mathbb{E}[X]1_A]$.

Lemma 6.2.1: Suppose X and Z are independent random variables. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that $\mathbb{E}|\varphi(X, Y)| < \infty$ and let $g(z) = \mathbb{E}[\varphi(X, z)]$. Then: $\mathbb{E}[\varphi(X, Z)|Z] = g(Z)$ a.s.

Linearity of Conditional Expectation: For L^1 random variables.

Monotonicity: If $X \leq Y$ and $X, Y \in L^1$, then $\mathbb{E}[X|G] \leq \mathbb{E}[Y|G]$ a.s.

Pre-Tower Property: If $\mathcal{F} \subset \mathcal{G}$ and $\mathbb{E}[X|\mathcal{G}] \in \mathcal{F}$, then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X|\mathcal{G}]$.

Tower Property: Let $H \subseteq G$ be sub-σ-fields of $\mathcal{F}$. Then: $\mathbb{E}[\mathbb{E}[X|G]|H] = \mathbb{E}[X|H]$ a.s.

Take out what is known: If X is G-measurable, then for any random variable Y such that $\mathbb{E}|Y| < \infty$ and $\mathbb{E}|XY| < \infty$, we have: $\mathbb{E}[XY|G] = X\mathbb{E}[Y|G]$ a.s.

Conditional MCT: Let $X, X_n \geq 0$ be a sequence of integrable random variables and $X_n \uparrow X$. Then $\mathbb{E}[X_n|G] \uparrow \mathbb{E}[X|G]$ a.s.

Conditional Jensen's Inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function and $\mathbb{E}|X| < \infty$ and $\mathbb{E}|\varphi(X)| < \infty$, then $\mathbb{E}[\varphi(X)|G] \geq \varphi(\mathbb{E}[X|G])$ a.s.

L^p Contraction: For $p \geq 1$, $\mathbb{E}[|\mathbb{E}[X|G]|^p] \leq \mathbb{E}[|X|^p]$.

Conditional Fatou's Lemma: Let $X_n \geq 0$ be integrable random variables and let $\liminf_n X_n$ be integrable. Then $\mathbb{E}[\liminf_n X_n|G] \leq \liminf_n \mathbb{E}[X_n|G]$ a.s.

Conditional DCT: If $X_n \to X$ a.s. and $|X_n| \leq Y$ for some integrable random variable Y. Then $\mathbb{E}[X_n|G] \to \mathbb{E}[X|G]$ a.s.

Chebyshev's Inequality: If $a > 0$, then $\mathbb{P}(|X| \geq a|\mathcal{F}) \leq a^{-2}\mathbb{E}[X^2|\mathcal{F}]$.

Exercises D5.1.3-7

Mean Square Error: Suppose that $X \in L^2(\Omega, \mathcal{F}, P)$. Then for any $Y \in L^2(\Omega, G, P)$, we have: $\mathbb{E}\left[(X-Y)^2\right] \geq \mathbb{E}\left[(X - \mathbb{E}[X|G])^2\right]$. The equality holds if and only if $Y = \mathbb{E}[X|G]$ a.s.

Exercises D5.1.8-11

Symmetrization by Conditional Expectation: Suppose that $X_1, X_2, \ldots$ is an iid sequence. Then for any $n \geq k$, we have: $\mathbb{E}[f(X_1, X_2, \ldots, X_k)|\varepsilon_n] = \frac{1}{(n)_k}\mathbb{E}[\Sigma_i f(X_{i_1}, X_{i_2}, \ldots, X_{i_k})|\varepsilon_n]$, where the sum is over all k-tuples of distinct integers $1 \leq i_1, \ldots, i_k \leq n$ and $(n)_k = n(n-1)\ldots(n-k+1)$ is the number of such k-tuples.

5.2 Martingales, Almost Sure Convergence

Discrete Time Martingale

Filtration: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration $(\mathcal{F}_n)_{n\geq 0}$ is an increasing sequence of sub sigma fields $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$.

Adapted: A sequence of random variables $(X_n)_{n\geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be adapted to the filtration $(\mathcal{F}_n)_{n\geq 0}$ if X_n is $\mathcal{F}_n$-measurable for each $n \geq 0$.

Martingale: A sequence of random variables $(X_n)_{n\geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$ if : a) $(X_n)_{n\geq 0}$ is adapted to $(\mathcal{F}_n)_{n\geq 0}$. b) $\mathbb{E}|X_n| < \infty$ for each n. c) $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ a.s. for each n.

In c) if "=" is replaced by "$\leq$" or $\geq$", then $(X)_{n\geq 0}$ is said to be a super-martingale or sub-martingale respectively.

Exercise 5.2: Show that if $(X_n)_{n\geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, then it is also a martingale with respect to its canonical filtration $(\sigma(X_0, X_1, \ldots, X_n))_{n\geq 0}$.

Theorem D5.2.1: If X_n is a super-martingale, then for $n > m$, $\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m$.

Theorem D5.2.2: i) If X_n is a sub-martingale, then for $n > m$, $\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m$. ii) If X_n is a martingale, then for $n > m$, $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$.

Theorem D5.2.3: If X_n is a martingale with respect to $\mathcal{F}_n$ and φ is a convex function with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n then $\varphi(X_n)$ is a sub-martingale with respect to $\mathcal{F}_n$. Consequently, if $p \geq 1$ and $\mathbb{E}|X_n|^p < \infty$ for all n, then $|X_n|^p$ is a sub-martingale with respect to $\mathcal{F}_n$.

Theorem D5.2.4: If X_n is a sub-martingale with respect to $\mathcal{F}_n$ and φ is a convex function with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n then $\varphi(X_n)$ is a sub-martingale with respect to $\mathcal{F}_n$. Consequently, i) If X_n is a sub-martingale then $(X_n - a)^+$ is a sub-martingale. ii) If X_n is a super-martingale, then $X_n \wedge a$ is a super-martingale.

Exercise D5.2.3: Give an example of a sub-martingale X_n so that X_n^2 is a super-martingale. Hint: X_n does not have to be random.

Doob's Martingale Transform: Call the sequence of random variables $(H_n)_{n\geq 1}$ **predictable** with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$ if H_n is $\mathcal{F}_{n-1}$ measurable for each $n \geq 1$. Let $(X_n)_{n\geq 0}$ be a $(\mathcal{F}_n)_{n\geq 0}$–martingale. Define $(H \cdot X)_0 = 0$, $(H \cdot X)_n = \sum_{k=1}^n H_k(X_k - X_{k-1})$.

Doob's Martingale Transform Lemma: Assume that X_n is a martingale and $(H \cdot X)_n \in L^1$ for each n. Then, $(H \cdot X)_n$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale.

Theorem D5.2.5: Let X_n be a super-martingale. If H_n is predictable and each H_n is bounded, then $(H \cdot X)_n$ is a super-martingale (similarly for sub-martingales and for martingales).

Doob's Decomposition: Any sub-martingale X_n with respect to $\mathcal{F}_n$ can be uniquely written as the sum of a martingale M_n with respect to $\mathcal{F}_n$ and an increasing predictable process A_n with $A_0 = 0$.

Let $D_0 = X_0$ and $D_j = X_j - \mathbb{E}[X_j | \mathcal{F}_{j-1}] \in \mathcal{F}_j$ for $j \geq 1$. Set $M_n = D_0 + D_1 + \ldots + D_n \in \mathcal{F}_n$, and $A_0 = 0$, $A_n = X_n - M_n = \mathbb{E}[X_n | \mathcal{F}_{n-1}] - (D_0 + \ldots + D_{n-1}) \in \mathcal{F}_{n-1}$ for $n \geq 1$.

Stopping Time Martingale Proposition: If T is a stopping time and $(X_n)_{n\geq 0}$ is a super-martingale, then $(X_{T\wedge n})_{n\geq 0}$ is a super-martingale.

Stopped Martingale Corollary: If T is a stopping time and $(X_n)_{n\geq 0}$ is a martingale, then $(X_{T\wedge n})_{n\geq 0}$ is a martingale.

Let T be a stopping time with $\mathbb{E}[T] < \infty$, then $\mathbb{E}[T] = \sum_{i=1}^{\infty} \mathbb{P}(T \geq i)$.

Martingale Convergence

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space and let $(X_n)_{n\geq 0}$ be any adapted process. Let $a < b \in \mathbb{R}$. Denote by $U_n[a,b]$ the number of up crossings from a to b by time n, i.e., the largest $k \geq 0$ such that there are (random) times $0 \leq s_1 < t_1 < \ldots < s_k < t_k \leq n$ and $X_{s_i} \leq a, X_{t_i} \geq b$ for each i.

Doob's Upcrossing Inequality: If $(X_n)_{n\geq 0}$ is a sub-martingale, then $\mathbb{E}[U_n[a,b]] \leq \frac{\mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+]}{b-a}$.

We use the Upcrossing Inequality in the following theorem to show that the nonnegative sub-martingale has a finite number of of crossings, and therefore converges.

Martingale Convergence: Suppose that $(X_n)_{n\geq 0}$ is a sub-martingale with $\sup_n \mathbb{E}[X_n^+] < \infty$. Then for some X, we have $X_n \to X$ a.s., where $\mathbb{E}|X| < \infty$.

This gives us the following 2 corollaries:

L^1-**Bounded Martingale Convergence**: If $(X_n)_{n\geq 0}$ is a martingale with $\sup_n \mathbb{E}|X_n| < \infty$, then $X_n \to X$ a.s. and $\mathbb{E}|X| < \infty$.

Non-negative Super-Martingale Convergence: If $(X_n)_{n\geq 0}$ is a super-martingale with $X_n \geq 0$, then $X_n \to X$ a.s. and $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.

Exercises D5.2.4-D5.2.14
Exercise D5.2.5: Let $X_n = \Sigma_{m\leq n} 1_{Bm}$ and suppose $B_n \in \mathcal{F}_n$. What is the Doob decomposition for X_n?
Exercise D5.2.13: The switching principle. Suppose X_n^1 and X_n^2 are super-martingales with respect to F_n, and N is a stopping time so that $X_N^1 \geq X_N^2$. Then show that: $Y_n = X_n^1 1_{\{N>n\}} + X_n^2 1_{\{N\leq n\}}$ is a super-martingale.

Examples

Theorem D5.3.1 (**Bounded Increments**): Let $X_1, X_2, \ldots$ be a martingale with $|X_{n+1} - X_n| \leq M < \infty$. Let $C := \left\{\lim X_n \text{ exists and is finite}\right\}$, and $D := \left\{\limsup X_n = +\infty \text{ and } \liminf X_n = -\infty\right\}$. Then, $\mathbb{P}(C \cup D) = 1$.

Exercises D5.3.1-D5.3.4

Theorem D5.3.2 (**2nd Borel-Cantelli Lemma**): Let $\mathcal{F}_n$, $n \geq 0$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and A_n, $n \geq 1$ a sequence of events with $A_n \in \mathcal{F}_n$. Then, $\left\{A_n \text{ i.o.}\right\} = \left\{\sum_{n=1}^{\infty} \mathbb{P}(A_n | \mathcal{F}_{n-1}) = \infty\right\}$.

Exercises D5.3.5-6

Polya's Urn Scheme

Radon-Nikodym Derivatives

Let μ be a finite measure and ν a probability measure on $(\Omega, \mathcal{F})$. Let $\mathcal{F}_n \uparrow \mathcal{F}$ be σ-fields (i.e., $\sigma(\cup \mathcal{F}_n) = \mathcal{F}$). Let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$.

Theorem D5.3.3: Suppose $\mu_n << \nu_n$ for all n. Let $X_n = d\mu_n / d\nu_n$ and let $X = \limsup X_n$. Then, $\mu(A) = \int_A X d\nu + \mu(A \cap \{X = \infty\})$.

Lemma D5.3.4: $X_n = d\mu_n/d\nu_n$ (Define: $(\Omega, \mathcal{F}, \nu)$) is a martingale with respect to $\mathcal{F}_n$.

Branching Processes: Galton-Watson

Let $\xi_i^n, i \geq 1, n \geq 0$ be iid nonnegative integer valued random variables with a common $\mu := \mathbb{E}[\xi_i^n] \in (0,\infty)$. Define
$Z_0 = 1$ and $Z_{n+1} = \begin{cases} \xi_1^n + \ldots + \xi_{Z_n}^n, & \text{if } Z_n > 0, \\ 0, & \text{if } Z_n = 0. \end{cases}$

Then, $\left(\frac{Z_n}{\mu^n}\right)_{n\geq 0}$ is a martingale with respect to $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 0 \leq m < n)$.

Theorem D5.3.7: If $\mu < 1$, then $Z_n = 0$ for all n sufficiently large, so $Z_n/\mu^n \to 0$.

Theorem D5.3.8: If $\mu = 1$ and $\mathbb{P}(\xi_i^m = 1) < 1$, then $Z_n = 0$ for all n sufficiently large.

Generating Function for the Offspring Distribution $p_k : \varphi(s) := \sum_{k\geq 0} p_k s^k$ on $s \in [0,1]$, where $p_k = \mathbb{P}(\xi_i^m = k)$.

This generating function is used in the proof of the following theorem:

Theorem D5.3.9 (probability of extinction): $\mathbb{P}\left(Z_n = 0 \text{ for some } n\right) = \mathbb{P}(\cup_{n=0}^{\infty} \{Z_n = 0\}) = \rho$ the unique fixed point of φ in $[0,1)$.

So, If $\mu > 1$, then $\rho < 1$, that is, $\mathbb{P}\left(Z_n > 0 \text{ for all } n\right) > 0$.

Theorem D5.3.10: $W = \lim Z_n/\mu^n$ is not $\equiv 0$ if and only if $\Sigma p_k k \log k < \infty$.

Exercises D5.3.10-12. **Exercise D5.3.12**: Show that if $\mathbb{P}(\lim Z_n/\mu^n = 0) < 1$, then it is $= \rho$ and hence $\{\lim Z_n/\mu^n > 0\} = \left\{Z_n > 0 \text{ for all } n\right\}$ a.s.

Martingale Inequalities

Stopping Time Submartingale Inequality (Proposition 5.24): If $(X_m)_{m\geq 0}$ is a sub-martingale and T is a stopping time with $\mathbb{P}(T \leq k) = 1$, for some $k \in \mathbb{Z}_+$, then $\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_k]$.

Corollary: If $(X_m)_{m\geq 0}$ is a martingale and T is a stopping time with $\mathbb{P}(T \leq k) = 1$, for some $k \in \mathbb{Z}_+$, then $\mathbb{E}[X_0] = \mathbb{E}[X_T] = \mathbb{E}[X_k]$.

Exercise D5.4.1-3

Doob's Maximal Inequality: Let $(X_m)_{m\geq 0}$ be a nonnegative sub-martingale, $X_n^* := \max_{0\leq m\leq n} X_m$, $\lambda > 0$, and $A := \{X_n^* \geq \lambda\}$. Then, $\mathbb{P}(A) \leq \frac{1}{\lambda}\mathbb{E}[X_n 1_A] \leq \frac{1}{\lambda}\mathbb{E}[X_n]$.

Relatedly: If (X_n) is a sub-martingale, then (X_n^+) is a nonnegative sub-martingale, and if (X_n) is a martingale, then $(|X_n|)$ is a nonnegative sub-martingale.

Observe that $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1} \iff \mathbb{E}[X_n 1_A] = \mathbb{E}[X_{n-1} 1_A]$ for all $A \in \mathcal{F}_{n-1}$.

Exercise D5.4.4. **Exercise D5.4.5**: Let X_n be a martingale with $X_0 = 0$ and $\mathbb{E}[X_n^2] < \infty$. Show that: $\mathbb{P}\left(\max_{1\leq m\leq n} X_m \geq \lambda\right) \leq \frac{\mathbb{E}[X_n^2]}{\mathbb{E}[X_n^2]+\lambda^2}$. Hint: Use the fact that $(X_n + c)^2$ is a submartingale and optimize over c.

L^p-Maximal Inequality: Let $(X_n)_{n\geq 0}$ be a nonnegative sub-martingale, $X_n^* = \max_{0\leq m\leq n} X_m$. Fix $p > 1$. Then, $\mathbb{E}[(X_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_n^p]$.

Theorem D5.4.4: Let X_n be a sub-martingale and $\log^+ x := \max(\log x, 0)$, then $\mathbb{E}[X_n^*] \leq (1 - \frac{1}{e})^{-1}\{1 + \mathbb{E}[X_n^+ \log^+(X_n^+)]\}$.

Exercise D5.4.6: Prove Theorem D5.4.4 by carrying out the following steps: i) Imitate the proof of D5.4.2 but use the trivial bound $\mathbb{P}(A) \leq 1$ for $\lambda \leq 1$ to show that $\mathbb{E}[X_n^* \wedge M] \leq 1 + \int X_n^+ \log(X_n^* \wedge M) d\mathbb{P}$.

L^p-Convergence Theorem for Martingales (See L^1 Bdd Cnvg Thm): Suppose that $(X_n)_{n\geq 0}$ is a martingale with $\sup \mathbb{E}[|X_n|^p] < \infty$ for some $p > 1$. Then, $X_n \to X$ a.s. and in L^p.

Theorem D5.4.6 - Orthogonality of Martingale Increments: Let X_n be a martingale with $\mathbb{E}[X_n^2] < \infty$ for all n. If $m \leq n$ and $Y \in \mathcal{F}_m$ has $\mathbb{E}[Y^2] < \infty$ then $\mathbb{E}[(X_n - X_m)Y] = 0$.

Theorem D5.4.7 - Conditional Variance Formula: If X_n is a martingale with $\mathbb{E}[X_n^2] < \infty$ for all n, then $\mathbb{E}\left[(X_n - X_m)^2 | \mathcal{F}_m \right] = \mathbb{E}[X_n^2 | \mathcal{F}_m] - X_m^2$.

Exercises D5.4.7-5.4.9

Uniform Integrability and L^1-Convergence of Martingales

Uniform Integrability: A family of random variables $(X_\alpha)_{\alpha \in \Lambda}$ is said to be uniformly integrable (*UI*) if $\sup_{\alpha \in \Lambda} \mathbb{E}[|X_\alpha| 1_{\{|X_\alpha| > M\}}] \to 0$ as $M \to \infty$.

Remark: Since $\mathbb{E}|X_\alpha| \leq M + \mathbb{E}[|X_\alpha| 1_{\{|X_\alpha| > M\}}]$, we have that $(X_\alpha)_{\alpha \in \Lambda}$ is $UI \Rightarrow (X_\alpha)_{\alpha \in \Lambda}$ is L^1-bounded, i.e., $\sup_{\alpha \in \Lambda} \mathbb{E}|X_\alpha| < \infty$.

Sub σ-field UI Lemma: Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then, $\left\{ \mathbb{E}[X|\mathcal{G}] : \mathcal{G} \text{ a } \sigma\text{-field } \subset \mathcal{F} \right\}$ is uniformly integrable.

Exercise D5.5.1: Let $\varphi \geq 0$ be any function with $\frac{1}{x}\varphi(x) \to \infty$ as $x \to \infty$, for example, $\varphi(x) = x^p$ with $p > 1$ or $\varphi(x) = x \log^+ x$. If $\mathbb{E}[\varphi(|X_i|)] \leq C$ for all $i \in I$, then $\{X_i : i \in I\}$ is uniformly integrable.

Convergence in Probability Equivalency Theorem: If $X_n \to X$ in probability, then *TFAE* :

- ♦ $\{X_n : n \geq 0\}$ is uniformly integrable.
- ♦ $X_n \to X$ in L^1 $\Rightarrow$ $\mathbb{E}[|X_n - X|] \to 0$.
- ♦ $\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$.

Convergence in Probability Corollary: As a consequence of the previous theorem, $X_n \overset{p}{\to} X$ and $(X_n)_{n \geq 0}$ is *UI* iff $X_n \overset{L^1}{\to} X$. In particular, if $X_n \overset{p}{\to} X$ and $|X_n| \leq Y$ for some $Y \in L^1$, then $X_n \overset{L^1}{\to} X$.

Sub-martingale Equivalencies Theorem: for a sub-martingale $(X_n)_{n \geq 0}$, TFAE:

- ♦ $(X_n)_{n \geq 0}$ is *UI*,
- ♦ X_n converges a.s. and in L^1,
- ♦ X_n converges in L^1.

If $(X_n)_{n \geq 0}$ is a martingale, then these are also equivalent to:

- ♦ there exists an integrable random variable X so that $X_n = \mathbb{E}[X|\mathcal{F}_n]$.

Lemma D5.5.4: If integrable random variables $X_n \to X$ in L^1, then $\mathbb{E}[X_n 1_A] \to \mathbb{E}[X 1_A]$.

Exercise D5.5.2-4.

Theorem D5.5.8 - Levy's 0-1 Law: Suppose that $\mathcal{F}_n \uparrow \mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$ and $A \in \mathcal{F}_\infty$, then $\mathbb{E}[1_A | \mathcal{F}_n] \to 1_A$ a.s..

Levy's Forward Law: Suppose that $\mathcal{F}_n \uparrow \mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$. If $X \in L^1$, then $\mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{F}_\infty]$ a.s. and in L^1.

Kolmogorov's 0-1 Law: Let $\xi_1, \xi_2, \ldots$ be independent random variables and let $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \ldots, \xi_n)$ for each n. Let $\mathcal{T} = \cap_{k=1}^{\infty} \sigma(\xi_k, \xi_{k+1}, \ldots)$ be the tail σ-field. Then for any $A \in \mathcal{T}$, $\mathbb{P}(A) \in \{0, 1\}$.

Exercises D5.5.5-7

Theorem D5.5.9 (Dominated Convergence for Converging σ-Algebra): Suppose $Y_n \to Y$ a.s. and $|Y_n| \leq Z$ for all n where $\mathbb{E}[Z] < \infty$. If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ then $\mathbb{E}[Y_n | \mathcal{F}_n] \to \mathbb{E}[Y|\mathcal{F}_\infty]$ a.s.

Exercise D5.5.8: Show that if $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $Y_n \to Y$ in L^1, then $\mathbb{E}[Y_n|\mathcal{F}_n] \to \mathbb{E}[Y|\mathcal{F}_\infty]$ in L^1.

Backward Martingale

Let $(\mathcal{F}_{-n})_{n \geq 0}$ be a sequence of sub-σ-fields, with the property $\ldots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0$. A sequence of random variables $(X_{-n})_{n \geq 0}$ is said to be a backward (or reverse) martingale if:

- ♦ $X_{-n} \in \mathcal{F}_{-n}$ for each $n \in \mathbb{Z}_+$.
- ♦ $X_{-n} \in L^1$ for each $n \in \mathbb{Z}_+$.
- ♦ $\mathbb{E}[X_{-n} | \mathcal{F}_{-(n+1)}] = X_{-(n+1)}$ for each $n \in \mathbb{Z}_+$.

Clearly, $\mathbb{E}[X_0 | \mathcal{F}_{-n}] = X_{-n}$ for each $n \in \mathbb{Z}_+$. Hence, if $(X_{-n})_{n \in \mathbb{Z}_+}$ is a reverse martingale, then it is UI.

Convergence of Reverse Martingale Theorem: Let $(X_{-n})_{n \geq 0}$ be a reverse martingale. Then $X_{-n} \overset{n \to \infty}{\to} X_{-\infty}$ a.s. and in L^1. Moreover, $\mathbb{E}[X_0 | \mathcal{F}_{-\infty}] = X_{-\infty}$ where $\mathcal{F}_{-\infty} = \cap_{n \in \mathbb{Z}_+} \mathcal{F}_{-n}$.

Exercise D5.6.1: Show that if $X_0 \in L^p$, the convergence $X_{-n} \overset{n \to \infty}{\to} X_{-\infty}$ occurs in L^p.

Levy's Backward Law: Let $Y \in L^1$. Suppose that there is a decreasing sequence of σ-fields $\mathcal{G}_0 \supseteq \mathcal{G}_1 \supseteq \mathcal{G}_2 \supseteq \ldots$ and $\mathcal{G}_\infty = \bigcap_{n=0}^{\infty} \mathcal{G}_n$. Then, $\mathbb{E}[Y|\mathcal{G}_n] \to \mathbb{E}[Y|\mathcal{G}_\infty]$ a.s. and in L^1.

> **Exercise D5.6.2**: Prove the backwards analog of theorem D5.5.9. Suppose $Y_n \to Y_{-\infty}$ a.s. as $n \to -\infty$ and $|Y_n| \leq Z$ a.s. where $\mathbb{E}[Z] \leq \infty$. If $\mathcal{F} \downarrow \mathcal{F}_{-\infty}$, then $\mathbb{E}[Y_n|\mathcal{F}_n] \to \mathbb{E}[Y_{-\infty}|\mathcal{F}_{-\infty}]$ a.s.

Exchangeable Sequence: We say a sequence of random variables $X_1, X_2, \ldots$ is exchangeable if for each $n \geq 1$, we have $(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$, for all permutations π of $\{1, 2, \ldots, n\}$, i.e., their joint law is invariant under any finite permutation of coordinates.

- Exchangeable Sequence $\Rightarrow$ Identical Distribution (but not the reverse).
- IID $\Rightarrow$ Exchangeable Sequence (but not the reverse).

de Finetti's Theorem: If $X_1, X_2, \ldots$ are exchangeable, then conditional on ε, $X_1, X_2, \ldots$ are iid.

Lemma D5.6.4 - Consistency of *U*-Statistics (used in proof of HW Law): Suppose $X_1, X_2, \ldots$ are i.i.d. and $f : \mathbb{R}^k \to \mathbb{R}$ is bounded, then $A_n(\varphi) := \frac{1}{(n)_k} \Sigma_i f(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \stackrel{\text{a.s.}+L^1}{\to} \mathbb{E}[f(X_1, X_2, \ldots, X_k)]$ a.s.

For example, if $X_1, X_2, \ldots$ is a iid sequence with finite mean, then $\frac{1}{n(n-1)} \sum_{i \neq j \leq n} |X_i - X_j| \stackrel{\text{a.s.}+L^1}{\to} \mathbb{E}|X_1 - X_2|$.

Theorem D5.6.6: If $X_1, X_2, \ldots$ are exchangeable and take values in $\{0, 1\}$, then there is a probability distribution $F(\theta)$ on $[0, 1]$ so that $\mathbb{P}(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta)$.

> **Exercises D5.6.3-5**

Optional Stopping Theorem

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$ be a filtered probability space and let T be a stopping time. We denote by $\mathcal{F}_T$, the σ-field of "events which have occurred prior to time T." In symbols: $\mathcal{F}_T := \left\{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n, \ \forall n \geq 0\right\}$. In the above definition of $\mathcal{F}_T$, the event $\{T \leq n\}$ can be replaced by $\{T = n\}$.

Optional Stopping Proposition:

i) If T is a stopping time, then $\mathcal{F}_T$ is a σ-field and T is $\mathcal{F}_T$-measurable.

ii) If $S \leq T$ are stopping times, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

iii) Let T be a stopping time with $\mathbb{P}(T < \infty) = 1$ and $(X_n)_{n \geq 0}$ be an adapted sequence. Then $X_T \in \mathcal{F}_T$.

UI Sub-martingale Stopping Time Closure: If $(X_n)_{n \geq 0}$ is a *UI* sub-martingale, then for any stopping time T, $(X_{T \wedge n})_{n \geq 0}$ is *UI*.

Theorem D5.7.2: If $\mathbb{E}[X_T] < \infty$ and $X_n 1_{\{T > n\}}$ is *UI*, then $X_{T \wedge n}$ is *UI*.

Theorem D5.7.3 (see prop. 5.24): If X_n is a uniformly integrable sub-martingale, then for any stopping time $T \leq \infty$, we have: $\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_\infty]$, where $X_\infty = \lim X_n$.

Optional Stopping Theorem for Submartingales: If S and T are stopping times with $\mathbb{P}(S \leq T < \infty) = 1$ and $(X_{T \wedge n})_{n \geq 0}$ is a uniformly integrable sub-martingale, then $\mathbb{E}[X_T|\mathcal{F}_S] \geq X_S$ a.s. Consequently, $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.

Optional Stopping Theorem for Martingales: If S and T are stopping times with $\mathbb{P}(S \leq T < \infty) = 1$ and $(X_{T \wedge n})_{n \geq 0}$ is a uniformly integrable martingale, then $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$ a.s. Consequently, $\mathbb{E}[X_S] = \mathbb{E}[X_T]$.

Theorem D5.7.5 (a generalization of Wald): Suppose X_n is a sub-martingale and $\mathbb{E}[|X_{n+1} - X_n| : \mathcal{F}_n] \leq B$ a.s. If T is a stopping time w/$\mathbb{E}[T] < \infty$, then $X_{T \wedge n}$ is uniformly integrable and hence $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$.

Nonnegative Supermartingale Stopping Time Theorem D5.7.6: If X_n is a nonnegative super-martingale and $T \leq \infty$ is a stopping time, then $\mathbb{E}[X_0] \geq \mathbb{E}[X_T]$ where $X_\infty = \lim X_n$, which exists by Theorem D5.2.9.

Theorem D5.7.7 (**Asymmetric Simple Random Walk**): Let $\xi_1, \xi_2, \ldots$ be iid random variables and $X_n := \xi_1 + \ldots + \xi_n$. Let $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = -1) = q \equiv 1 - p$ with $p \neq q$. Without loss of generality, we assume $\frac{1}{2} < p < 1$.

a) If $\varphi(x) = \left\{\frac{1-p}{p}\right\}^x$ then $\varphi(S_n)$ is a martingale.

b) If we let $T_x = \inf\{n : S_n = x\}$ then for $a < 0 < b$, we have: $\mathbb{P}(T_a < T_b) = \frac{\varphi(b)-\varphi(0)}{\varphi(b)-\varphi(a)}$.

c) If $a < 0$, then $\mathbb{P}(\min_n S_n \leq a) = \mathbb{P}(T_a < \infty) = \left\{\frac{1-p}{p}\right\}^{-a}$.

d) If $b > 0$, then $\mathbb{P}(T_b < \infty) = 1$ and $\mathbb{E}[T_b] = \frac{b}{2p-1}$.

Exercise D5.7.2: Let S_n be an asymmetric simple random walk with $\frac{1}{2} < p < 1$, and let $\sigma^2 = pq$. Use the fact that $X_n = (S_n - (p-q)n)^2 - 4\sigma^2 n$ is a martingale to show $Var(T_b) = \frac{b\sigma^2}{(p-q)^3}$.

Exercises D5.7.3-9

Random Durrett Exercises

D5.1.3: Prove Chebyshev's Inequality for conditional expectations.

D5.1.4: Suppose $X \geq 0$ and $\mathbb{E}[X] = \infty$. (There is nothing to prove when $\mathbb{E}[X] < \infty$.) Show there is a unique $\mathcal{F}$-measurable Y with $0 \leq Y \leq \infty$ such that $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$ for all $A \in \mathcal{F}$. HINT: Let $X_M = X \wedge M$, $Y_M = \mathbb{E}[X_M | \mathcal{F}]$, and let $M \to \infty$.

D5.1.5: Imitate the proof in the remark after Thereom 1.5.2 to prove the conditional Cauchy Schwartz inequality: $\mathbb{E}[XY|\mathcal{G}]^2 \leq \mathbb{E}[X^2|\mathcal{G}]\mathbb{E}[Y^2|\mathcal{G}]$.

D5.1.6: Give an example on $\Omega = \{a, b, c\}$ in which: $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]|\mathcal{F}_2] \neq \mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1]$.

D5.1.7: Show that when $\mathbb{E}|X|$, $\mathbb{E}|Y|$, and $\mathbb{E}|XY|$ are finite, each statement implies the next one and give examples with $X, Y \in \{-1, 0, 1\}$ a.s. that show the reverse implications are false: i) X and Y are independent, ii) $\mathbb{E}[Y|X] = \mathbb{E}[Y]$, ii) label$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

D5.1.8: Show that if $\mathcal{G} \subset \mathcal{F}$ and $\mathbb{E}[X^2] < \infty$, then $\mathbb{E}\left[\{X - \mathbb{E}[X|\mathcal{F}]\}^2\right] + \mathbb{E}\left[\{\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[X|\mathcal{G}]\}^2\right] = \mathbb{E}\left[\{X - \mathbb{E}[X|\mathcal{G}]\}^2\right]$. [*If* we drop the 2nd term on the left, we get an inequality that says geometrically, the larger the subspace the closer the projection is, or statistically, more information means a smaller mean square error.]

D5.1.9: Let $Var(X|\mathcal{F}) = \mathbb{E}[X^2|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}]^2$. Show that $Var(X) = \mathbb{E}[Var(X|\mathcal{F})] + Var(\mathbb{E}[X|\mathcal{F}])$.

D5.1.10: Let Y_1, Y_2 be iid with mean μ and variance σ^2, N an independent positive integer valued random variable with $\mathbb{E}[N^2] < \infty$ and $X = Y_1 + \ldots + Y_N$. Show that $Var(X) = \sigma^2\mathbb{E}[N] + \mu^2 Var(N)$. To understand and help remember the formula, think about the 2 special cases in which N or Y is constant.

D5.1.11: Show that if X and Y are random variables with $\mathbb{E}[Y|\mathcal{G}] = X$ and $\mathbb{E}[Y^2] = \mathbb{E}[X^2] < \infty$, then $X = Y$ a.s.

D5.2.4: Give an example of a martingale X_n with $X_n \to -\infty$ a.s. Hint: Let $X_n = \xi_1 + \ldots + \xi_n$, where the ξ_i are independent (but not identically distributed) with $\mathbb{E}[\xi_i] = 0$.

D5.3.5: Let $p_m \in [0,1)$. Use the Borel-Cantelli lemmas to show that $\prod_{m=1}^{\infty}(1-p_m) = 0$ if and only if $\sum_{m=1}^{\infty} p_m = \infty$.

D5.3.6: Show $\sum_{n=2}^{\infty} \mathbb{P}(A_n | \cap_{m=1}^{n-1} A_m^c) = \infty$ implies $\mathbb{P}(\cap_{m=1}^{n-1} A_m^c) = 0$.

D5.3.10: Galton and Watson who invented the process that bears their names were interested in the survival of family names. Suppose each family has exactly 3 children but coin flips determine their sex. In the eighteen hundreds, only male children kept the family name so following the male offspring leads to a branching process with $p_0 = \frac{1}{8},\ p_1 = \frac{3}{8},\ p_2 = \frac{3}{8},\ p_3 = \frac{1}{8}$. Compute the probability ρ that the family name will die out when $Z_0 = 1$.

D5.4.1: Show that if $j \leq k$, then $\mathbb{E}[X_j 1_{\{T=j\}}] \leq \mathbb{E}[X_k 1_{\{T=j\}}]$ and sum over j to get a 2nd proof of $\mathbb{E}[X_T] \leq \mathbb{E}[X_k]$.

D5.4.2: Generalized the proof of Theorem 5.4.1 to show that if X_n is a sub-martingale and $M \leq T$ are stopping times with $\mathbb{P}(T \leq k) = 1$, then $\mathbb{E}[X_M] \leq \mathbb{E}[X_T]$.

D5.4.3: Use the stopping times from the Exercise 4.1.7 to strengthen the conclusion of the previous exercise to $\mathbb{E}[X_T | \mathcal{F}_M] \geq X_M$.

Durrett Chapter 6 (Markov Chains)

Markov chain: Given a filtration $(\mathcal{F}_n)_{n\geq 0}$, an $(\mathcal{F}_n)_{n\geq 0}$-adapted stochastic process $(X_n)_{n\geq 0}$ taking values in $(S, \mathcal{S})$ is called a Markov chain if it has the **Markov Property**:

$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{P}(X_{n+1} \in B|X_n)$ a.s. for each $B \in \mathcal{S}, n \geq 0$.

Markov Chain Transition Probability: A set function $p : S \times \mathcal{S} \to [0,1]$ is said to be a transition probability if:

i) For each $x \in S,\ B \to p(x,B)$ is a probability measure on $(S, \mathcal{S})$.

ii) For each $B \in \mathcal{S},\ x \to p(x,B)$ is a $\mathcal{S}$-measurable function.

We say a Markov chain $(X_n)_{n\geq 0}$ has transition probabilities $(p_n)_{n\geq 0}$ if $\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = p_n(X_n, B)$ almost surely for each $n \geq 0$ and $B \in \mathcal{S}$.

Transition Matrix: The probability of moving from i to j in one time step is $\mathbb{P}(j|i) =: p_{ij}$, if we put these into a matrix, we have the transition matrix $p = [p_{ij}]$.

Time Homogeneous Markov Chain: A Markov chain in which the transition probabilities are all the same $p_n = p$ for all time $n \geq 0$.

Exercise 6.4: Let p be a transition probability on $S \times \mathcal{S}$ and μ be a probability measure on $(T, \mathcal{T})$. Then there exists a unique probability measure λ on $\left(T \times S,\ \mathcal{T} \otimes \mathcal{S}\right)$ such that $\lambda(A \times B) = \int_A \mu(dx_0)p(x_0, B)$ for all $A \in \mathcal{T},\ B \in \mathcal{S}$.

Theorem 6.5 (**Existence of Markov Chain**): Let $(p_n)_{n\geq 0}$ be a sequence of transition probabilities and μ be a probability measure on $(S, \mathcal{S})$. Then there exists a unique probability measure $\mathbb{P}_\mu$ on $(S^{\mathbb{Z}_+}, \mathcal{S}^{\mathbb{Z}_+})$ such that if $X \in S^{\mathbb{Z}_+}$, the coordinate maps $X_n(x) = x_n : S^{\mathbb{Z}_+} \to S$ form a Markov chain with respect to its canonical σ-field $(\mathcal{F}_n)_{n\geq 0}$ and such that for each $B \in \mathcal{S}$:

i) $\mathbb{P}_\mu(X_0 \in B) = \mu(B)$, ii) $\mathbb{P}_\mu(X_{n+1} \in B|\mathcal{F}_n) = p_n(X_n, B)$ $\mathbb{P}_\mu$- a.s.

Proposition 6.6: If $(X_n)_{n\geq 0}$ is a Markov chain with transition probabilities $(p_n)_{n\geq 0}$ and initial distribution μ, then the finite dimensional distributions are given by (EQ 37 in the class notes), i.e.,
$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_k \in A_k) = \int_{A_0} \mu(dx_0) \int_{A_1} p_0(x_0, dx_1) \ldots \int_{A_k} p_k(x_{k-1}, dx_k)$.

Useful in proving this proposition is the following lemma:

Lemma 6.7: Under the Markov set up, for each $n \geq 0$ and for any bounded measurable function $f : S \to \mathbb{R}$, we have: $\mathbb{E}[f(X_{n+1})|\mathcal{F}_n] = \int p_n(X_n, dy)f(y)$ a.s.

Theorem D6.1.3 (**Monotone Class Theorem**): Let $\mathcal{A}$ be a π-system that contains Ω and let $\mathcal{H}$ be a collection of real valued functions that satisfies:

i) If $A \in \mathcal{A}$, then $1_A \in \mathcal{H}$.

ii) If $f, g \in \mathcal{H}$, then $f + g$, and $cf \in \mathcal{H}$ for any real number c.

iii) If $f_n \in \mathcal{H}$ are nonnegative and increase to a bounded function f, then $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all bounded functions measurable with respect to $\sigma(\mathcal{A})$.

Proposition 6.8: Let $(X_n)_{n\geq 0}$ be a Markov chain on a countable set S with transition matrix p and initial distribution μ. Then:

a) $\mathbb{P}\left(X_0 = i_0,\ X_1 = i_1, \ldots, X_n = i_n\right) = \mu(i_0)p_0(i_0, i_1) \ldots p_{n-1}(i_{n-1}, i_n)$

b) $\mathbb{P}(X_n = j|X_0 = i) = (p^n)(i,j)$.

c) $\mathbb{P}(X_n = j) = \sum_{i\in S} \mu(i)(p^n)(i,j)$.

Section D6.2 Examples

Examples and Exercises D6.2.2-9

Exercise D6.2.4: Let $\xi_0, \xi_1 \ldots$ be iid $\in \{H, T\}$, taking each value with probability $\frac{1}{2}$. Show that $X_n = (\xi_n, \xi_{n+1})$ is a Markov chain and compute its transition probability p. What is p^2?

Exercise D6.2.8: Let $\xi_0, \xi_1 \ldots$ be iid$\in \{-1, 1\}$, taking each value with probability $\frac{1}{2}$. Let $S_0 = 0,\ S_n = \xi_1 + \ldots + \xi_n$, and $X_n = \max\{S_m : 0 \le m \le n\}$. Show that X_n is not a Markov chain.

D6.3 - Strong Markov Property

Let $(X_n)_{n\ge 0}$ be an S-valued (time homogeneous) Markov chain with initial distribution μ. By theorem 6.5, assume that X_n's are the coordinate maps on the space $(S^{\mathbb{Z}_+}, \mathcal{S}^{\mathbb{Z}_+}, \mathbb{P}_\mu)$. Let $\mathcal{F}_n$ be the σ-field generated by $X_0, \ldots, X_n$. Define the shift operator $\theta : S^{\mathbb{Z}_+} \to S^{\mathbb{Z}_+}$ by $\theta(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. So: $(\theta^k(x))_n = x_{k+n}$ for $n, k \ge 0$.

Theorem 6.14 (**Strengthened Markov Property**): Let $(X_n)_{n\ge 0}$ be as above. For any bounded measurable function $f : S^{\mathbb{Z}_+} \to \mathbb{R}$, and any $k \ge 0$, $\mathbb{E}_\mu[f \circ \theta^k | \mathcal{F}_k] = \mathbb{E}_{X_k}[f]$ $\mathbb{P}_\mu$- a.s.

Above, $\mathbb{E}_\mu$ is the expectation operator associated with probability measure $\mathbb{P}_\mu$, where $\mu = \delta_{x_0}$.

Exercise 6.15: From the above theorem, deduce that: $\mathbb{P}((X_{k+1}, X_{k+2}, \ldots) \in B, (X_0, X_1, \ldots, X_{k-1}) \in A | X_k) = \mathbb{P}((X_{k+1}, X_{k+2}, \ldots) \in B | X_k)\mathbb{P}((X_0, X_1, \ldots X_{k-1}) \in A | X_k)$.

Exercise D6.3.1: Use the Markov property to show that if $A \in \sigma(X_0, \ldots, X_n)$ and $B \in \sigma(X_n, X_{n+1}, \ldots)$, then for any initial distribution μ, we have: $\mathbb{P}_\mu(A \cap B | X_n) = \mathbb{P}_\mu(A|X_n)\mathbb{P}_\mu(B|X_n)$.
In words, the past and future are conditionally independent given the present. Hint: Write the left-hand side as $\mathbb{E}_\mu(\mathbb{E}_\mu(1_A 1_B | F_n)|X_n)$.

Theorem D6.3.2 (**Chapman-Kolmogorom Equation**): $\mathbb{P}_x(X_{m+n} = z) = \Sigma_y \mathbb{P}_x(X_m = y)\mathbb{P}_y(X_n = z)$. (short proof).

Theorem D6.3.3: Let X_n be a Markov chain and suppose $\mathbb{P}(\cup_{m=n+1}^{\infty} \{X_m \in B_m\} | X_n) \ge \delta > 0$ on $\{X_n \in A_n\}$, then $\mathbb{P}\Big(\big\{X_n \in A_n \text{ i.o.}\big\} - \big\{X_n \in B_n \text{ i.o.}\big\}\Big) = 0$.

Intuitive meaning of this theorem: "If the chance of a pedestrians getting run over is greater than $\delta > 0$ each time he crosses a certain street, then he will not be crossing it indefinitely (since he will be killed first)!"

Exercise D6.3.2

Absorbing: A state a is called absorbing if $\mathbb{P}_a(X_1 = a) = 1$.

Theorem 6.16 (**Strong Markov Property**): Let $(X_n)_{n\ge 0}$ be as above. For any bounded measurable function $f : S^{\mathbb{Z}_+} \to \mathbb{R}$ and for any stopping time T, $\mathbb{E}_\mu[f \circ \theta^T | \mathcal{F}_T] = \mathbb{E}_{X_T}[f]$ on $\{T < \infty\}$ $\mathbb{P}_\mu$- a.s.

Theorem D6.3.5 (**Reflection Principle**): Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed with a distribution that is symmetric about 0. Let $S_n = \xi_1 + \ldots + \xi_n$. If $a > 0$, then $\mathbb{P}(\sup_{m\le n} S_m > a) \le 2\mathbb{P}(S_n > a)$.

Exercises D6.3.3-12
Exercise D6.3.7: Let X_n be a Markov chain with $S = \{0, 1, \ldots, N\}$ and suppose that X_n is a martingale and $\mathbb{P}_x(\tau_0 \wedge \tau_N < \infty) > 0$ for all x. i) Show that 0 and N are absorbing states, that is, $p(0,0) = p(N,N) = 1$.
Exercise D6.3.10: Let $\tau_A = \inf\{n \ge 0 : X_n \in A\}$ and $g(x) = \mathbb{E}_x[\tau_A]$. Suppose that $S - A$ is finite and for each $x \in S - A$, $\mathbb{P}_x(\tau_A < \infty) > 0$. **i) Show that** $g(x) = 1 + \sum_y p(x,y)g(y)$ **for all** $x \notin A$ $(*)$

Recurrence and Transience

Let X_0 be any state. Let $T_y^0 := 0$, and for $k \ge 1$, let $T_y^k := \inf\{n > T_y^{k-1} : X_n = y\}$, the time of the kth visit since possibly X_0 to y. Let $\rho_{yz} := \mathbb{P}_y(T_z < \infty)$ (probability you'll get to z from y in finite time.

Theorem 6.17: For $k \ge 1$, $\mathbb{P}_y(T_z^k < \infty) = \rho_{yz}\rho_{zz}^{k-1}$.

Exercise 6.18 (**iid cycle**): Suppose $y \in S$ such that $\rho_{yy} = 1$. Let $R_k = T_y^k$ be the time of the kth return to y, and for $k \ge 1$, let $r_k = R_k - R_{k-1}$ be the kth inter-arrival time. Use the strong Markov property to conclude that under $\mathbb{P}_y$, the vectors $v_k = (r_k, X_{R_{k-1}}, \ldots, X_{R_k - 1}), k \ge 1$ are i.i.d.

Recurrent: A state $y \in S$ is called **recurrent** if $\rho_{yy} = 1$ and is called **transient** if $\rho_{yy} < 1$.

By Theorem 6.17: If y is recurrent, then $\mathbb{P}_y\big(X_n = y \text{ i.o.}\big) = \lim_{k\to\infty} \mathbb{P}_y(T_y^k < \infty) = \lim_k \rho_{yy}^k = 1.$

If y is transient, then $\mathbb{P}_y\big(X_n = y \text{ i.o.}\big) = \lim_k \rho_{yy}^k = 0.$

Total Visits ($N(y)$): Let the total number of visits to y by the Markov chain X_n be notated as $N(y) := \sum_{n=1}^{\infty} 1_{\{X_n=y\}}$.

Lemma 6.20: For any $x, y \in S,$ we have:

i) $\mathbb{P}_x(N(y) = k) = \rho_{xy}\rho_{yy}^{k-1}(1-\rho_{yy}),$

ii) $\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1-\rho_{yy}} = \sum_{n=1}^{\infty} p^n(x,y)$ (where we interpret $\frac{0}{0} = 0,\ \frac{c}{0} = +\infty$ for $c > 0$).

Recurrent Corollary: A state $x \in S$ is **recurrent** if and only if $\mathbb{E}_x[N(x)] = \sum_{n=1}^{\infty} p^n(x,x) = \infty.$

Exercises D6.4.2-3

Communication

A state x **leads to**, or is **accessible from** another state $y \neq x,$ denoted by $x \to y,$ if $\rho_{xy} > 0$ (or equivalently, for some $n \geq 1,\ p^n(x,y) > 0$). Formally, state y is accessible from state x if there exists an integer $n_{xy} \geq 0$ such that $\mathbb{P}(X_{n_{xy}} = y | X_0 = x) = p_{xy}^{(n_{xy})} > 0.$ This integer is allowed to be different for each pair of states, hence the subscripts in n_{ij}. Allowing n to be zero means that every state is accessible from itself by definition, or $x \to x.$ The accessibility relation is reflexive and transitive, but not necessarily symmetric.

A pair of states x and y are said to **communicate**, denoted by $x \leftrightarrow y,$ if $x \to y$ and $y \to x.$

Communicating Class: "$\leftrightarrow$" is an equivalence relation. Therefore, there is a partition C_1, C_2 of $S,$ with each block C_i being referred to as a communicating class.

Irreducible Subset: A subset $A \subseteq S$ is called irreducible if $x \leftrightarrow y$ for all $x, y \in A.$ By definition, each class is irreducible.

A Markov *chain* is said to be **irreducible** if it is possible to get to any state from any state. More formally, a Markov chain is said to be **irreducible** if its state space is a single communicating class, i.e., $x \leftrightarrow y$ for all $x, y \in s.$

Proposition 6.23: If x is recurrent and $\rho_{xy} > 0,$ then:

i) $\rho_{yx} = 1,$ ii) y is recurrent, iii) $\rho_{xy} = 1.$

Exercise D6.4.4: Use the strong Markup property to show that $\rho_{xz} \geq \rho_{xy}\rho_{yz}$.

Closed Subset of States: We call a subset of states $A \subseteq S$ closed if $\rho_{xy} = 0$ for all $x \in A$ and $y \notin A.$

Lemma 6.24: A recurrent class C is closed.

Lemma 6.25: If C is a finite closed set, then it contains at least one recurrent state. In particular, a finite closed class C is recurrent.

Lemma 6.26: In a finite state Markov chain, a class is recurrent (respectively transient) if and only if it is closed (respectively not closed).

Exercise D6.4.5

Theorem D6.4.5 (**Decomposition Theorem**, **motivated by Example D6.4.1**): Let $R = \{x : \rho_{xx} = 1\}$ be the recurrent states of a Markov chain. R can be written as $\cup_i R_i,$ where each R_i is closed and irreducible. [This results shows that for the study of recurrent states we can, without loss of generality, consider a single irreducible closed set.]

Birth and Death Chains on $\mathbb{N}$

Let $N := \inf\{n : X_n = 0\}$. Let $p(i,i+1) =: p_i,\quad p(i,i-1) =: q_i,\quad p(i,i) =: r_i,$ where $q_0 = 0.$ Define the function φ so that $\varphi(X_{N\wedge n})$ is a martingale. Set $\varphi(0) = 0$ and $\varphi(1) = 1,$ and note that for the martingale property to hold when $X_n = k \geq 1,$ we have: $\varphi(k) = p_k\varphi(k+1) + r_k\varphi(k) + q_k\varphi(k-1).$ And using $r_k = 1-(p_k+q_k)$, we rewrite this as $\varphi(k+1) - \varphi(k) = \frac{q_k}{p_k}(\varphi(k) - \varphi(k-1)).$ Suppose that $p_k, q_k > 0$ for $k \geq 1$ so that the chain is irreducible. We find that $\varphi(m+1) - \varphi(m) = \prod_{j=1}^{m} \frac{q_j}{p_j}$ for $m \geq 1$ and $\varphi(n) = \sum_{m=0}^{n-1}\prod_{j=1}^{m} \frac{q_j}{p_j}$ for $n \geq 1.$

Theorem D6.4.6: Let $T_c = \inf\{n \geq 1 : X_n = c\}$. If $a < x < b$, then $\mathbb{P}_x(T_a < T_b) = \frac{\varphi(b)-\varphi(x)}{\varphi(b)-\varphi(a)}$, and $\mathbb{P}_x(T_b < T_a) = \frac{\varphi(x)-\varphi(a)}{\varphi(b)-\varphi(a)}$.

Theorem D6.4.7: 0 is recurrent if and only if $\varphi(M) \to \infty$ as $M \to \infty$, that is: $\varphi(\infty) \equiv \sum_{m=0}^{\infty} \prod_{j=1}^{m} \frac{q_j}{p_j} = \infty$. If $\varphi(\infty) < \infty$, then $\mathbb{P}_x(T_0 = \infty) = \frac{\varphi(x)}{\varphi(\infty)}$.

Exercise D6.4.6

Theorem D6.4.8: Suppose S is irreducible, and $\varphi \geq 0$ with $\mathbb{E}_x\left[\varphi(X_1) \leq \varphi(x) \text{ for } x \notin F,\right]$ a finite set, and $\varphi(x) \to \infty$ as $x \to \infty$, that is, $\{x : \varphi(x) \leq M\}$ is finite for any $M < \infty$, then the chain X_n is recurrent.

Exercises D6.4.7-10

Stationary Measures

Stationary/Invariant Measure μ:

(invariant) $\mu p = \mu : \mu(y) = \Sigma_{x \in S}\mu(x)p(x,y)$. (i.e., μ is a left eigenvector of p).

The last equation says $\mathbb{P}_\mu(X_1 = y) = \mu(y)$. Using the Markov property and induction, it follows that $\mathbb{P}_\mu(X_n = y) = \mu(y)$ for all $n \geq 1$. If μ is a probability measure, we call μ a stationary distribution, and it represents a possible equilibrium for the chain.

Stationary/Invariant Distribution π:

i) (probability distribution) $\Sigma_{x \in S}\pi(x) = 1$

ii) (invariant) $\pi p = \pi : \pi(y) = \Sigma_{x \in S}\pi(x)p(x,y)$. (i.e., π is a left eigenvector of p).

Exercises D6.5.1-2

Reversible Measure: A measure μ that satisfies $\mu(x)p(x,y) = \mu(y)p(y,x)$ for all x, y.

Theorem D6.5.1: Suppose p is irreducible. A necessary and sufficient condition for the existence of a reversible measure is that i) $p(x,y) > 0$ implies $p(y,x) > 0$, and ii) for any loop $x_0, \ldots, x_n = x_0$ with $\prod_{1 \leq i \leq n} p(x_i, x_{i-1}) > 0$, $\prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})} = 1$.

Unique Stationary/Invariant Distribution π: Suppose that S is finite and p is irreducible. Then there exists a unique solution to $\pi p = \pi$ with $\Sigma_{i \in S}\pi(i) = 1$ and $\pi(i) > 0$ for all $i \in S$.

Stationary/Invariant Measure Theorem (D6.5.2): Let x be a recurrent state. Then: $\mu_x(y) := \mathbb{E}_x\left[\sum_{n=0}^{T_x - 1} 1_{\{X_n = y\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n)$, is a stationary measure.

Exercise D6.5.3

Recurrent Time in y: Define $\mu_x(y)$ as the expected time spent in y between visits to x.

Exercise 6.42 (Irreducible Transient Chain with no Stationary Measure). Consider a Markov chain on $\mathbb{Z}_+$ with transition matrix p given by $p(0,1) = 1$, $p(i,i+1) = 1 - \left(\frac{1}{2}\right)^i$ for all $i \geq 1$, and $p(1,0) = \frac{1}{2}$, $p(i,i-1) = \left(\frac{1}{4}\right)^i$, $p(i,0) = \left(\frac{1}{2}\right)^i - \left(\frac{1}{4}\right)^i$ for all $i \geq 2$. Show that this Markov chain is irreducible, transient and does not have any nontrivial stationary measure.

Lemma 6.43: Suppose that a Markov chain is irreducible and recurrent. Let μ be a stationary measure with $\mu(y) > 0$ for all $y \in S$. If ν is another stationary measure, then $\mu = c\nu$ for some $c > 0$.

Exercises D6.5.4-7

Exercise D6.5.5: Show that if p is irreducible and recurrent, then $\mu_x(y)\mu_y(z) = \mu_x(z)$

Positive recurrent: $\mathbb{E}_x[T_x] = \sum_{n=1}^{\infty} n\mathbb{P}(T_x = n) = \sum_{y \in S} \mu_x(y) < \infty \quad \Rightarrow \quad \mathbb{P}_x(T_x < \infty) = 1$.

Null-Recurrent: A state $x \in S$ is said to be null recurrent if $\mathbb{P}_x(T_x < \infty) = 1$, but $\mathbb{E}_x[T_x] = \infty$.

If $\{X_n\}$ is **recurrent** but not **null recurrent** then it is called **positive recurrent**.

A state j is called positive recurrent if the expected amount of time to return to a state j given that the chain started in state j has finite first moment: $\mathbb{E}[\tau_{jj}] < \infty$. A positive recurrent state j is always recurrent: If $\mathbb{E}[\tau_{jj}] < \infty$, then $f_j = \mathbb{P}(\tau_{jj} < \infty) = 1$, but the converse is not true: a recurrent state need not be positive recurrent. A current state j for which $\mathbb{E}[\tau_{jj}] = \infty$ is called null recurrent.

To test whether the state is **postive-recurent** or **null-recurrent**, we compute the mean return time:

$\mathbb{E}_x[T_x] = \sum_{n=1}^{\infty} np^n(x,x) = \infty,$ is null-recurrent.

If a chain is finite and irreducible, then there exists a **unique stationary/invariant distribution** π, and it is **positive recurrent**.

If $\{X_n\}$ is **positive recurrent**, then for every $x, y \in S$, $\lim_{n\to\infty} p_n(x,y) = \pi(y) > 0$ where $\pi : S \to [0,1]$ is the **stationary/invariant distribution**.

Unique Corollary: An irreducible, positive recurrent Markov chain has a unique **stationary/invariant distribution** π.

Stationary/Invariant Corollaries: For an irreducible and recurrent chain, the following are true.

a) The **stationary/invariant measures** are unique up to constant multiples.

b) If μ is a **stationary/invariant measure**, then $\mu(x) > 0$ for all x.

c) The **stationary/invariant distribution** π, if it exists, is unique.

d) If a **stationary/invariant measure** has infinite mass, then the **stationary/invariant distribution** π cannot exist.

Recurrence from Stationary/Invariant Distributions: If π is a **stationary/invariant** distribution of a Markov chain and $\pi(x) > 0$, then x is recurrent.

Calculating Stationary/Invariant Distribution: If p is irreducible and has a stationary distribution π. Then $\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}$.

Exercise D6.5.8: Compute the expected number of moves it takes a night to return to its initial position if it starts in a corner of the chessboard, assuming there are no other pieces on the board, and each time it chooses a move at random from its legal moves. (Note: A chessboard is $\{0,1,\dots,7\}^2$. A Knight's move is L-shaped; 2 steps in one direction, followed by one step in a perpendicular direction.)

Lemma 6.51 (**Thm D6.5.6**): For an irreducible Markov chain, the following are equivalent.

i) There exists $x \in S$ that is positive recurrent.

ii) There exists a stationary distribution π.

iii) Every state is positive recurrent.

Exercise D6.5.9: Suppose p is irreducible and positive recurrent. Then $\mathbb{E}_x[T_y] < \infty$ for all x, y.
Exercises D6.5.10-13

Theorem D6.5.7: If p is irreducible and has a stationary distribution π, then any other stationary measure is a multiple of π.

Doubly Stochastic: A probability transition matrix $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$ is called doubly stochastic if $\Sigma_i p_{ij} = 1$ for all j and $\Sigma_j p_{ij} = 1$ for all i. The uniform distribution is a stationary distribution of p if and only if the probability transition matrix is doubly stochastic.

Stationary Sequence: A sequence of random variables $(X_n)_{n\geq 0}$ is said to be stationary if $(X_n, X_{n+1}, \dots) \overset{d}{=} (X_0, X_1, \dots)$, for each $n \geq 0$, or equivalently, $(X_n, X_{n+1}, \dots, X_{n+m}) \overset{d}{=} (X_0, X_1, \dots, X_m)$, for each $n, m \geq 0$. Exchangeable sequences are stationary.

A **reversible measure** is always stationary since $\Sigma_{x\in S}\mu(x)p(x,y) = \Sigma_{x\in S}\mu(y)p(y,x) = \mu(y)\Sigma\mathbb{P}(y,x) = \mu(y)\cdot 1 = \mu(y)$.

Lemma 6.60: Consider a Markov chain $(X_n)_{n\geq 0}$ started from a stationary distribution π and transition matrix p. Fix $N \geq 1$ and define $Y_n = X_{N-n}$ for $n = 0, 1, \dots, N$. Then $(Y_n)_{0\leq n\leq N}$ is a time-homogeneous Markov chain with initial distribution π and transition matrix q given by $q(x,y) = \frac{\pi(y)p(y,x)}{\pi(x)}$.

Convergence of Markov Chains

Number of Visits to y: Let $N_n(y) := \Sigma_{i=1}^{n} 1_{\{X_i = y\}}$ be the number of times the chain visits y during the time $\{1, \ldots, n\}$.

Theorem 6.62 (**Asymptotic Density of Returns**): Let $y \in S$ be recurrent. Then $\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y[T_y]} 1_{\{T_y < \infty\}}$ $\mathbb{P}_x$- a.s.

Aperiodic. For a state x, let $I_x := \{n \geq 1 : p_n(x,x) > 0\}$. Let d_x be the greatest common divisor of I_x (if $I_x = \emptyset$, then $d_x = +\infty$). We say x has period d_x. If every state of a Markov chain has period 1, then we call the chain aperiodic.

Lemma 6.67: If $x \leftrightarrow y$, then $d_x = d_y$, i.e., periodicity is a class property.

Lemma 6.68: If $d_x = 1$, then there exists $n_0 \geq 1$ such that $p^n(x,x) > 0$ for all $n \geq n_0$.

Lemma 6.69: An irreducible aperiodic Markov chain has the following property: for each $x, y \in S$, there exists $n_0 = n_0(x,y) \geq 1$ such that $p^n(x,y) > 0$ for all $n \geq n_0$.

An irreducible aperiodic chain $\{X_n\}$ is **null recurrent** if it is **recurrent** and $\lim_{n\to\infty} p_n(x,y) = 0$ for all $x, y \in S$.

Theorem 6.70 (**Markov Chain Convergence Theorem**): Consider an irreducible, aperiodic Markov chain with stationary distribution π. Then, $p^n(x,y) \to \pi(y)$ as $n \to \infty$, for all $x, y \in S$.

Exercises D6.6.1-7
Exercise D6.6.1: Show that if S is finite and p is irreducible and aperiodic, then there is an m so that $p^m(x,y) > 0$ for all x, y.

Convergence Theorem

Total Variation Distance: For two probability measures μ, ν on S, their total variation distance is given by: $d_{TV}(\mu,\nu) := 1/2 \sum_{x=S} |\mu(x) - \nu(x)| = sup_{A\subseteq S} |\mu(A) - \nu(A)|$.

Exercise 6.73: Let μ, ν be two probability measures on a countable space S. Then, $d_{TV}(\mu,\nu) = \inf\left\{\mathbb{P}(X \neq Y) : (X,Y) \text{ has a joint distribution with } X \sim \mu,\ Y \sim \nu\right\}$.

Coupled Markov Chain: Let μ, ν be to probability measures on a countable space S, and $(X_n, Y_n)_{n\geq 0}$ on the product space $S \times S$. The chain is considered coupled if:

i) Its marginal processes $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are Markov chains on S with the same transition matrix p and the initial distributions μ and ν respectively, and

ii) $X_n = Y_n$ for $n \geq T$, where $T := \inf\{n \geq 0 : X_n = Y_n\}$.

The next few lemmas are used in the proof of Theorem 6.70:

Lemma 6.74 (**Coupling Inequality**): Let $(X_n, Y_n)_{n\geq 0}$ be a coupled Markov chain as above with $X_0 \sim \mu$ and $Y_0 \sim \nu$. Let μ_n and ν_n be the marginal distribution of X_n and Y_n respectively. Then, $d_{TV}(\mu,\nu) \leq \mathbb{P}(T > n)$.

Lemma 6.75: Let $(X_n, Y_n)_{n\geq 0}$ is a Markov chain on $S \times S$ with transition matrix q and the initial distribution $(X_0, Y_0) \sim \mu \otimes \nu$. Then $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are Markov chains on S with the same transition matrix p and the initial distributions μ and ν respectively.

Lemma 6.76: $\mathbb{P}_{\mu\otimes\nu}(T < \infty) = 1$.

Mixing Time of Markov Chains

Exercise 6.78: Prove that $\mathbb{E}[T] = k(n-k)$. (Hint: Use optional stopping theorem on the martingale $(D_t^2 - t)_{t\geq 0}$).
Exercise 6.80: Show that for any $C > 0$, $\mathbb{P}(S > d\log d + Cd) \leq e^{-C}$ for all $d \geq 1$. Hint: $\mathbb{P}(S > d\log d + Cd) \leq \sum_{i=1}^{d} \mathbb{P}\left(i\text{th coupon is not picked in the first } d\log d + Cd \text{ draws}\right) = d\left(1 - \frac{1}{d}\right)^{d\log d + Cd}$.

Random Durrett Exercises

D6.3.2: Let $D = \left\{X_n = a \text{ for some } n \geq 1\right\}$ and let $h(x) = \mathbb{P}_x(D)$. i) Use theorem D6.3.3 to conclude that $h(X_n) \to 0$ a.s. on D^c. Here, a.s. means $\mathbb{P}_\mu$ a.s. for any initial distribution μ. ii) Obtain the result in Exercise D5.5.5 as a special case.

D6.4.5: Show that in the Ehrenfest chain (Example D6.2.5), all states are recurrent.

D6.4.6: A gambler is playing roulette and bedding \$1 on black each time. The probability she wins \$1 is $\frac{18}{38}$, and the probability she loses is \$1 is $\frac{20}{38}$. i) Calculate the probability that starting with \$20 she reaches \$40 before losing her money. ii) Use the fact that $X_n + \frac{2n}{38}$ is a martingale to calculate $\mathbb{E}[T_{40} \wedge T_0]$.

D6.5.3: Use the construction in the proof of Theorem D6.5.2 to show that $\mu(j) = \Sigma_{k \geq j} f_{k+1}$ defines a stationary measure for the renewal chain (Example D6.2.3).

D6.5.10: Suppose p is irreducible and has a stationary measure μ with $\Sigma_x \mu(x) = \infty$. Then p is not positive recurrent.

Ergodic Theory

Measure Preserving Map: Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. A measurable map $T : \Omega \to \Omega$ is called measure preserving if for each $E \in \mathcal{F}$, we have: $\mu(T^{-1}(A)) = \mu(A)$. In other words, $\mu \circ T^{-1} = \mu$. In this case, μ is called the **invariant measure** with respect to T and $(\Omega, \mathcal{F}, \mu, T)$ is called a **measure preserving system**.

Invariant Event: Let $(\Omega, \mathcal{F}, \mu, T)$ be a measure preserving system. Call an event $A \in \mathcal{F}$ invariant if $T^{-1}(A) = A$ μ- a.s., i.e., $\mu(A \Delta T^{-1}(A)) = 0$. Collection of invariant events: $\mathcal{I} := \{A \in \mathcal{F} : A \text{ is invariant}\}$.

Invariant Random Variable: A random variable $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ is said to be invariant if $X \circ T = X$ μ- a.s.

Ergodic System: A measure preserving system $(\Omega, \mathcal{F}, \mu, T)$ is ergodic if $\mu(A) = 0$ or 1 for each $A \in \mathcal{I}$.

Birkhoff's Ergodic Theorem: Let $(\Omega, \mathcal{F}, \mu, T)$ be a measure preserving system. Then for any $X \in L^1(\Omega, \mathcal{F}, \mu)$, we have: $\frac{1}{n}\sum_{k=0}^{n-1} X \circ T^k \overset{\text{a.s.}+L^1}{\to} \mathbb{E}[X|\mathcal{I}]$, as $n \to \infty$.

Moreover, when T is ergodic, then we have $\mathbb{E}[X|\mathcal{I}] = \mathbb{E}[X]$ a.s. since $\mathcal{I}$ is a trivial σ-field. Consequently, $\frac{1}{n}\sum_{k=0}^{n-1} X \circ T^k \overset{\text{a.s.}+L^1}{\to} \mathbb{E}[X]$, as $n \to \infty$.

Brownian Motion

Standard Brownian Motion (BM): A collection of random variables $(B_t)_{t \geq 0}$ defined on a common probability space such that:

1. For any $n \geq 1$ and $0 = t_0 < t_1 < \ldots < t_n$, the **increments** $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq n$ are independent.
2. For any $s < t$, $B_t - B_s \sim N(0, t-s)$. Also, $B_0 = 0$ a.s.
3. For a.s. every $\omega \in \Omega$, the path $t \to B_t(\omega)$ is continuous.

Properties of BM:

♦ If $0 < t_0 < t_1 < \ldots < t_n$, then the joint distribution of $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ follows n-dimensional multivariate Gaussian with mean 0 and the covariance matrix given by $\mathbb{E}[B_{t_i} B_{t_j}] = \min(t_i, t_j)$.

♦ For each $T > 0$ and $\varepsilon > 0$, the Brownian path $t \to B_t(\omega)$ is $\left(\frac{1}{2} - \varepsilon\right)$-Holder continuous on $[0, T]$.

$$|B_{t_0} - B_{t_1}| \leq C|t_0 - t_1|^{\frac{1}{2} - \varepsilon}, \text{ a.s.}$$

♦ **Scale Invariance**: Fix $c > 0$. Define $W_t = \frac{B_{ct}}{\sqrt{c}}$ for $t \in [0, \infty)$. Then $(W_t)_{t \geq 0}$ is again standard BM.

♦ **Negative BM**: $(-B_t)_{t \geq 0}$ is also standard BM.

♦ **Time Shift**: Fix $s \geq 0$ and define $W_t = B_{t+s} - B_s$ for $t \geq 0$. Then, $(W_t)_{t \geq 0}$ is a standard BM, independent of $\mathcal{F}_s := \sigma(B_u : u \leq s)$.

♦ **Time Inversion**: Let $W_t = \begin{cases} tB_{\frac{1}{t}} & \text{if } t > 0, \\ 0 & \text{if } t = 0, \end{cases}$ then $(W_t)_{t \geq 0}$ is a standard BM. Note that $(W_t)_{t \geq 0}$ are jointly Gaussians with mean 0 and for $0 < s < t$, we have covariance matrix given by :
$\mathbb{E}[W_s W_t] = st \min(\frac{1}{s}, \frac{1}{t}) = \min(s,t) = s$.

♦ **Law of Large Numbers**: $\lim_{t \to \infty} \frac{B_t}{t} = 0$ a.s.

♦ **Nowhere Differentiability**: Brownian paths are nowhere differentiable, a.s.

Let $\mathcal{F}_s^0 := \sigma(B_r : 0 \leq r \leq s)$ and $\mathcal{F}_s^+ := \cap_{t > s} \mathcal{F}_t^0$.

$\mathcal{F}_s^+$ are right continuous: $\cap_{t>s} \mathcal{F}_t^+ = \cap_{t>s} (\cap_{u>t} \mathcal{F}_u^0) = \cap_{u>s} \mathcal{F}_u^0 = \mathcal{F}_s^+$.

And therefore allow us an "infinitesimal peak into the future." In other words, $A \in \mathcal{F}_s^+$ if it is in $\mathcal{F}_{s+\varepsilon}^0$, for all ε.

Blumenthal's 0-1 **law**: If $A \in \mathcal{F}_0^+$, then for all $x \in \mathbb{R}^{d,}$ we have: $\mathbb{P}_x(A) \in \{0,1\}$.

Difference between Convergence in Probability and Convergence Almost Surely:

For simplicity, consider the case where $X = 0$ and X_n is the indicator function of an event E_n. "X_n converges almost surely to 0" says that with probability 1, only finitely many of the events E_n occur. "X_n converges in probability to 0" says that the probability of event E_n goes to 0 as $n \to \infty$.

Consider a case where for each m, you partition the sample space into m events, each of probability $1/m$, and take all these events for all m to form your sequence E_n. Then $X_n \to 0$ in probability because the probabilities of the individual events go to 0, but each sample point is in infinitely many E_n (one for each m) so X_n does not go to 0 almost surely.

Markov Examples

- **Dice Game**: A game whose moves are determined entirely by dice is a Markov chain, indeed, an absorbing Markov chain. This is in contrast to card games such as blackjack, where the cards represent a 'memory' of the past moves. Consider probability for a certain event in the game. In the above-mentioned dice games, the only thing that matters is the current state of the board. The next state of the board depends on the current state, and the next roll of the dice. It doesn't depend on how things got to their current state. In a game such as blackjack, a player can gain an advantage by remembering which cards have already been shown, so the next state (or hand) of the game is not independent of the past states.
- **Multivalued Markov Chain** $\{X_n, X_{n+1}\}_{n \geq 1}$: where $X_i \in \{T, H\}$ iid. $\mathbb{P}\{X_i = T\} = \frac{1}{2}$.
- **Random walk Markov chains**: Consider a random walk on the number line where, at each step, the position (call it x) may change by +1 (to the right) or -1 (to the left) with probabilities: $P_{move \text{ left}} = \frac{1}{2} + \frac{1}{2}\left(\frac{x}{c+|x|}\right)$,
$P_{\text{move } right} = 1 - P_{\text{move } left}$, w/$c > 0$.
Example, if constant c equals 1, the probabilities of a move to the left at positions $x = -2, -1, 0, 1, 2$ are $\frac{1}{6}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{5}{6}$ resp. The random walk has a centering effect that weakens as c increases. Since probabilities depend only on the current position (value of x) and not on prior positions, this biased random walk satisfies definition of Markov chain. If $X_n = 1$, then $\mathbb{P}(X_{n+1} = 2 | \mathcal{F}_n) = \mathbb{P}(X_{n+1} = 2 | X_n = 1)$.
- **Gambler's Fortune**: Suppose you start w/\$10, and wager \$1 on an unending fair coin toss indefinitely, or until you lose all your money. If X_n is # of \$s after n tosses, w/$X_0 = 10$, then $\{X_n : n \in \mathbb{N}\}$ is a Markov process. If I know that you have \$12 now, then it is expected that w/even odds, you'll either have \$11 or \$13 after next toss. The fact that the guess is not improved by the knowledge of earlier tosses showcases the Markov property, the memoryless property of a stochastic process. If $X_n = c + 1$, then $\mathbb{P}(X_{n+1} = c | \mathcal{F}_n) = \mathbb{P}(X_{n+1} = c | X_n = c + 1) = \frac{1}{2}$.

Martingale Examples

- **Galton Watson Process**
- **Doobs Martingale Transform** (if $(H \cdot X)_n \in L^1$).
- An **unbiased random walk** (in any number of dimensions): $\mathbf{S}_n := \sum_{i=1}^{n} X_i$ where $X_i \in \{-1, 1\}$ iid and $S_0 = 0$, is an example of a martingale.

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

  $= S_n + \frac{1}{2} - \frac{1}{2} = S_n$ a.s.
- A **gambler's fortune** is a martingale if all betting games which the gambler plays are fair. Suppose S_n is a gambler's fortune after n tosses of a fair coin, where the gambler wins \$1 if the coin comes up heads and loses \$1 if it comes up tails. The gambler's conditional expected fortune after the next trial, given the history, is equal to their present fortune. This sequence is thus a martingale. $\mathbb{E}[S_{n+1} | \mathcal{F}_n] =$ Same as above
- **Gamblers Total Gain/Loss Variance**: Let $Y_n := X_n^2 - n,$ where X_n is gambler's fortune from preceding example. The sequence $\{Y_n : n = 1, 2, 3, \ldots\}$ is a martingale. Used to show that gambler's total gain or loss varies roughly between plus or minus the square root of the number of steps.
- **Pólya's Urn** contains a number of different coloured marbles; at each iteration a marble is randomly selected from the urn and replaced with several more of that same colour. For any given colour, the fraction of marbles in the urn with that colour is a martingale. For example, if currently 95% of the marbles are red then, though the next iteration is more likely to add red marbles than another color, this bias is exactly balanced out by the fact that adding more red marbles alters the fraction much less significantly than adding the same number of non-red marbles would. Let $S_n = \sum_{i=0}^{n} X_i$ the # of marbles that are red, where X_0 is the initial number of red marbles.

  $\mathbb{E}\left[\frac{S_{n+1}}{T_{n+1}} | \mathcal{F}_n \right] = \frac{1}{T_n+1} \mathbb{E}[X_{n+1} + S_n | \mathcal{F}_n]$

  $= \frac{1}{T_n+1}(S_n + \mathbb{E}[X_{n+1} | \mathcal{F}_n]) = \frac{1}{T_n+1}\left(S_n + \mathbb{P}\left(\text{you choose red}\right)\right) = \frac{S_n + \frac{S_n}{T_n}}{T_n+1} = \frac{S_n\left(1 + \frac{1}{T_n}\right)}{T_n+1} = \frac{S_n\left(\frac{T_n+1}{T_n}\right)}{T_n+1} = \frac{S_n}{T_n}$ a.s.