Course notes for Advanced Probability

Instructor: Zhen-Qing Chen
Scribe: Avi Levy

April 9, 2015

These are my notes for Math 521, taught by Zhen-Qing Chen at the UW in Autumn 2014. They are updated online at:

http://www.math.washington.edu/~avius

Contents

Chapter 1

Autumn 2014

Lecture 1: September 24

What we will cover

Definition 1.1. *A* **random variable** *is a measurable function defined on a measure space.*

Let $X_1, \cdots$ be random variable with the same distribution. For example, a digit shown on a die. If X_n is the digit shown on throw n, then $X_1, \cdots$ is a sequence of observations.

One can ask: What is $\frac{1}{n}\sum_{i=1}^n X_i$? We will see that as $n \to \infty$, it converges to $\mathbb{E}X_1$. This is the law of large numbers. More specifically, the weak law says that they converge in measure, and the strong law says that they converge almost everywhere.

Example 1.2. *Let $X_k = 1$ if flip k yields heads, and 0 otherwise, and $p = \mathbb{E}X_k = \mathbb{P}(X_k = 1)$ is independent of k.*

Note that $\frac{1}{n}\sum_{j=1}^n X_j$ is a random variable that converges to a deterministic number. We want to study fluctuations of the random average from the mean, that is

$$a_n \left(\frac{1}{n} \sum_{j=1}^n X_j - \mu \right) \to \eta.$$

In most cases $a_n = \sqrt{n}$, here η is a normal distribution.

Random Walk

Simple random walk on $\mathbb{Z}$: A particle moves left or right with equal probability. For higher d there are more options for where a point can move (2^d). We're interested in the probability of reaching a certain point. If we can reach a point with probability 1, the walk is **recurrent** (this occurs when $d \le 2$). Otherwise the walk is **transient** ($d > 2$).

Markov Chains

In the winter

Martingales

These are fair games. You have a sequence of random variables indexed by time: $X_0, X_1, \cdots$ where X_n is the net wealth at time n. Let $\mathcal{F}_n$ denote the information up to time n. In other words, it is the smallest σ-field generated by $X_1, \cdots, X_n$. A "fair game" is when $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$. This is the stochastic version of a constant function. If the = is replaced with $\ge, \le$ it is called a sub martingale and sup martingale respectively. Simple random walk is a martingale.

Laws of large numbers, CLT, asymptotic behaviour of random variables

Example 1.3 (Coupon collector's problem). *Let $X_1, \cdots, X_n$ be iid random variables uniform on $\{1, \cdots, n\}$. Set τ_n to be the time needed to get all of $\{1, \cdots, n\}$. The asymptotic behavior is*

$$\frac{\tau_n}{n \log n} \to 1$$

in probability.

Example 1.4 (Head runs). *Flip a coin, and assign 1 for heads and −1 for tails. If you flip 100 times, what is the probability of 10 consecutive heads?*

Here X_b are iid with $\mathbb{P}(X_n = \pm 1) = 1/2$. Let $l_n = \max\{m : X_{n-m+1} = \cdots = X_n = 1\}$. For instance when $n = 10$ we could have

$$+1 - 1 - 1 + 1 + 1 - 1 - 1 + 1 + 1 + 1$$

has $l_n = 3$. We will show

$$\frac{l_n}{\log_2 n} \to 1.$$

Theorem 1.5 (Weak LLN). *Let $X_1, \cdots$ be* **uncorrelated** *random variables with $\mathbb{E}$*

$$\lim_{n\to\infty} \mathbb{P}(|\frac{1}{n}\sum X_i - \mu| > \varepsilon) = 0.$$

This is called convergence in measure.

Definition 1.6. *Random variables X_i, X_j are uncorrelated if $\mathbb{E}(X_i X_j) = \mathbb{E}X_i \mathbb{E}X_j$ where $i \neq j$. This is weaker than independence, and follows clearly from independence.*

Proof. Use Chebyshev's inequality:

$$P(|\xi| > \lambda) \le \frac{\mathbb{E}\xi^2}{\lambda^2} \qquad \xi \text{ is a random variable.}$$

□

Therefore we have

$$\begin{aligned}
\mathbb{P}(|\frac{1}{n}\sum_j X_j - \mu| > \varepsilon) &= \mathbb{P}(|\frac{S_n - \mathbb{E}S_n}{n}| > \varepsilon) \\
&\le \frac{1}{n^2\varepsilon^2}\mathbb{E}|S_n - \mathbb{E}S_n|^2 \\
\text{(uncorrelated)} \qquad &= \frac{1}{n^2\varepsilon^2}\sum_j \mathrm{Var}(X_j) \\
&\le \frac{C}{n\varepsilon^2} \to 0.
\end{aligned}$$

On Friday we'll see a nice application to real analysis: $f \in C[a, b]$ can be approximated by a polynomial (Weierstrass).

Lecture 2: September 26

Lecture 3: September 29

Lecture 4: October 1

Lecture 5: October 3

Lecture 6: October 6

Lecture 7: October 8

Recall $S_n = X_1 + \cdots + X_n$.

Theorem 7.1. *Let* $\mu_n = \mathbb{E}S_n, \sigma_n^2 = \mathrm{Var}(S_n)$. *If* $\sigma_n = o(b_n)$ *(i.e.* $\sigma_n/b_n \to 0$*), then* $\frac{S_n-\mu_n}{b_n} \to 0$ *in probability.*

Proof. Same as the proof of the Weak LLN; just use Chebyshev. □

Example 7.2 (Coupon Collector's Problem). $X_1, X_2, \cdots$ *i.i.d.* $\sim U\{1,\cdots,n\}$. *Here* $\tau_1 = 1$ *and*

$$\tau_k = \inf\{j > \tau_{k-1} : \#\{X_1,\cdots,X_j\} = k\}, \qquad k \geq 2.$$

Proof. Set $Y_1 = 1$ and $Y_k = \tau_k - \tau_{k-1}$ for $k \geq 2$. Observe that Y_k has geometric distribution with $p_k = \frac{n-k+1}{n}$, where $2 \leq k \leq n$. Then

$$\mathbb{E}Y_k = \frac{1}{p_k} = \frac{n}{n-k+1}, \qquad \mathrm{Var}(Y_k) \leq \frac{1}{p_k^2} = \frac{n^2}{(n-k+1)^2}.$$

Here $\tau_n = Y_1 + \cdots + Y_n$ is a sum of independent variables. We compute

$$\mu_n = \mathbb{E}\tau_n = \sum_1^n \mathbb{E}Y_k = n\sum_{j=1}^n \frac{1}{j} \sim n\log n.$$

(Recall that $f(n) \sim g(n)$ means $\frac{f(n)}{g(n)} \to 1$ as $n \to \infty$.)

By independence of the Y_k, we compute

$$\sigma_n^2 = \mathrm{Var}(\tau_n) = \sum_{k=1}^n \mathrm{Var}(Y_k) \leq \sum_{k=1}^n \frac{n^2}{j^2} \leq cn^2.$$

In the theorem, we want to take $b_n = \mu_n$. Since $n = o(n\log n)$, this is a valid choice. Hence it follows that

$$\frac{S_n - \mu_n}{\mu_n} \to 0 \qquad \text{in probability.}$$

By additivity of limits in probability, it follows that $S_n \sim n\log n$. □

A precise statement of the additivity statement:

Claim 7.3. *If* $\xi_n \to \xi$ *in probability and* $\eta_n \to \eta$ *in probability, then* $\xi_n + \eta_n \to \xi + \eta$ *in probability.*

Proof. Use the triangle inequality.

$$\begin{aligned}\mathbb{P}(|\xi_n + \eta_n - \xi - \eta| > \varepsilon) &\leq \mathbb{P}(|\xi_n - \xi| + |\eta_n - \eta| > \varepsilon) \\ &\leq \mathbb{P}(|\xi_n - \xi| > \frac{\varepsilon}{2}) + \mathbb{P}(|\eta_n - \eta| > \frac{\varepsilon}{2}).\end{aligned}$$

A sequence of random variables converges in probability iff for any subsequence, there is a sub-subsequence on which they converge almost surely. This gives an alternate proof of the preceeding result.

Triangular Arrays

Let X_{nk}, $1 \le k \le n$ be random variables. Let $S_n = \sum_{k=1}^n X_{nk}$.

$$\begin{array}{cccc} n=1 & X_{11} & & \\ n=2 & X_{21} & X_{22} & \\ n=3 & X_{31} & X_{32} & X_{33} \\ & \vdots & \vdots & \vdots \end{array}$$

Here S_n is the sum of each row. This is a very general framework that specializes to the coin collectors problem.

Theorem 7.4 (Weak Law for triangular arrays). *For each n, assume that each row is pairwise independent. Let $b_n > 0$ be a sequence such that $b_n \to \infty$ as $n \to \infty$. Use b_n to truncate the random variables as follows:*

$$\overline{X}_{nk} = X_{nk} 1_{\{|X_{nk}| \le b_n\}}$$

Suppose that:

(a) $\sum_{k=1}^n \mathbb{P}(|X_{nk}| > b_n) \to 0$ *as* $n \to \infty$

(b) $\frac{1}{b_n^2} \sum_{k=1}^n \mathbb{E}[\overline{X}_{nk}^2] \to 0$ *as* $n \to \infty$

Then $\frac{S_n - a_n}{b_n} \to 0$ *in probability, where* $a_n = \sum_{k=1}^n \mathbb{E}\overline{X}_{nk}$.

We use the notation $\mathbb{P}(A, B) = \mathbb{P}(A \cap B)$.

Proof. Let $\overline{S}_n = \sum_{k=1}^n \overline{X}_{nk}$. Fix any $\varepsilon > 0$. Then

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon, S_n = \overline{S}_n\right) + \mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon, S_n \ne \overline{S}_n\right) \\
&\le \mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon, S_n = \overline{S}_n\right) + \mathbb{P}(S_n \ne \overline{S}_n) \\
\text{(Chebyshev)} \qquad &\le \frac{\mathbb{E}(\overline{S}_n - a_n)^2}{b_n^2 \varepsilon^2} + \sum_{k=1}^n \mathbb{P}(X_{nk} \ne \overline{X}_{nk}) \\
\text{(Independence)} \qquad &= \frac{\sum_{k=1}^n \operatorname{Var}(\overline{X}_{nk})}{b_n^2 \varepsilon^2} + \sum_{k=1}^n \mathbb{P}(|X_{nk}| > b_n) \\
&\le \frac{\sum_{k=1}^n \mathbb{E}(\overline{X}_{nk}^2)}{b_n^2 \varepsilon^2} + \sum_{k=1}^n \mathbb{P}(|X_{nk}| > b_n) \to 0.
\end{aligned}$$

Theorem 7.5 (WLLN). *Let $X_1, X_2, \cdots$ be pairwise independent with the same distribution, and suppose*

$$\mathbb{P}(|x_1| > n) = o(1/n), \qquad n \to \infty.$$

Let $S_n = \sum_1^n X_k$ *and* $\mu_n = \mathbb{E}[X_1] 1_{\{|X_1| \le n\}}$. *Then* $\frac{S_n}{n} - \mu_n \to 0$ *in probability.*

Note that this implies the classical weak law. If X_1 has finite first moment, then the first condition is satisfied automatically by Chebshev. Also by LDCT, you can pass the limit in for μ_n. Basically this let's us apply the result even when there's no finite second moment, just a finite first moment.

Proof. Let $X_{nk} = X_k$ and $b_n = n$ in the triangular arrays statement. Observe that (a) holds, since it just says $n\mathbb{P}(|X_1| > n) \to 0$. For (b), observe that

$$\frac{1}{b_n^2}\sum_{k=1}^n \mathbb{E}(\overline{X}_{nk}^2) = \frac{1}{n}\mathbb{E}(X_1^2 1_{\{|X_1|\le n\}}).$$

Recall that for $Y \ge 0$, we have

$$\mathbb{E}(Y^p) = \int_0^\infty py^{p-1}\mathbb{P}(Y > y)\ dy.$$

Continuing the previous calculation,

$$\begin{aligned}\frac{1}{b_n^2}\sum_{k=1}^n \mathbb{E}(\overline{X}_{nk}^2) &= \int_0^n 2y\mathbb{P}(|X_1|1_{\{|X_1|\le n\}} > y)\ dy \\ &\le 2y\mathbb{P}(|X_1| > y)\ dy.\end{aligned}$$

Call the integrand $g(y)$. Observe that $g(y) \to 0$ as $y \to \infty$, and $g(y) \le 2y$. Since g is supported in $[0,\infty)$ and it decays, we have a finite bound $C = \sup g(y) < \infty$. Now choose k such that $\int_k^\infty g(y)\ dy \le \varepsilon$. Then

$$\begin{aligned}\frac{1}{n}\mathbb{E}(X_i^2 1_{\{|X_1|\le n\}}) &= \frac{1}{n}\left[\int_0^k g(y)\ dy + \int_k^n g(y)\ dy\right] \\ &\le Ck/n + \varepsilon(n-k)/n.\end{aligned}$$

Fix ε and hence k, then take $n \to \infty$, then take $\varepsilon \to 0$. We get the bound. $\square$

Lecture 8: October 10

Last time we proved the theorem:

Theorem 8.1 (WLL). *Let $X_1, \cdots$ be pairwise independent with the same distribution. Suppose*

$$n\mathbb{P}(|X| > n) \to 0, \qquad S_n = \sum_{i=1}^{n} X_i, \qquad \mu_n = \mathbb{E}(X_1 1_{\{|X_1| \le n\}})$$

Recall the following fact.

Fact 8.2. *For any $\xi \ge 0$,*

$$\mathbb{E}(\xi^p) = \int_0^\infty p y^{p-1} \mathbb{P}(\xi > y)\ dy, \qquad p > 0.$$

Proof.

$$y^p = p \int_0^y x^{p-1}\ dx = p \int_0^\infty x^{p-1} 1_{\{x<y\}}\ dx$$
$$\mathbb{E}\xi^p = p\mathbb{E} \int_0^\infty x^{p-1} 1_{\{\xi > x\}}\ dx$$
$$\text{Fubini} \qquad = p \int_0^\infty x^{p-1} \mathbb{P}(\xi > x)\ dx.$$

Note that we could replace strict inequalities with weak inequalities, and we have used Fubini since everything was positive. □

Now we prove the class weak law.

Corollary 8.3. *Let $X_1, \cdots$ be pairwise indpendent with the same distribution such that $\mathbb{E}|X_1| < \infty$. Then*

$$\frac{S_n}{n} \to \mu = \mathbb{E}X_1 \qquad \textit{in probability.}$$

Notice that we don't need any second moment assumption.

Proof. It suffices to verify

$$x\mathbb{P}(|X_1| > x) \to 0.$$

Indeed, observe that

$$x 1_{\{|X_1| > x\}} \le |X_1| 1_{\{|X_1| > x\}} \le |X_1|.$$

Since $x1_{\{|X_1|>x\}} \to 0$ pointwise, it follows by LDCT that

$$\mathbb{E}[x 1_{\{|X_1| > x\}}] \to 0.$$

Similarly by LDCT, it follows that

$$\mu_n = \mathbb{E}[X_1 1_{\{|X_1| \le n\}}] \to \mathbb{E}X_1 = \mu.$$

Hence $\frac{S_n}{n} - \mu_n \to 0$ and $\mu_n \to \mu$, so $\frac{S_n}{n} \to \mu$ in probability. □

This concludes the weak law of large number for triangular arrays, which is the most general form. Everything follows from the Chebyshev inequality. There is some care in applying the theorem, for b_n must be chosen appropriately. We do some examples. Early probability was motivated by gambling.

Example 8.4 (St. Petersburg Paradox). *Model a sequence of games by iid random variables $X_1, \cdots$ such that*

$$\mathbb{P}(X_1 = 2^j) = 2^{-j} \qquad j \geq 1.$$

Notice that $\mathbb{E}X_1 = \sum_j 2^j \mathbb{P}(X_1 = 2^j) = \infty$.

How much should the owner charge you to play? You can restrict the number of games played, for instance.

Let $S_n = X_1 + \cdots + X_n$ be the income from n games. Find c_n such that $s_n/c_n \to 1$ in probability. If you play n games, you will get roughly c_n, so the fair price is $\frac{c_n}{n}$. We want to find $b_n \sim c_n$. Recall the WLL for triangular arrays (where each row is the singleton X_k:

(a) $\sum_{i=1}^n \mathbb{P}(|X_k| > b_n) \to 0$

(b) $\frac{1}{b_n^2} \sum_{k=1}^n \mathbb{E}[|X_k|^2 1_{\{|X_k| \leq b_n\}}] \to 0$

Recall $a_n = \sum_{k=1}^n \mathbb{E}[X_k 1_{\{|X_k| \leq b_n\}}]$ and $\frac{S_n - a_n}{b_n} \to 0$ in probability.

Let $I = \sum_{i=1}^n \mathbb{P}(|X_k| > b_n) = n\mathbb{P}(|X_1| > b)$. Take $b_n = 2^{m_n}$. Then $I = n \sum_{j>m} 2^{-j} = n2^{-m_n} \to 0$. Choose $k_n = m_n - \log_2 n$, so that $2^{-k_n} = n2^{-m_n}$ and the condition becomes $k_n \to \infty$. Form the truncation

$$\overline{X}_{nk} = X_k 1_{\{|X_k| \leq b_n\}}.$$

Then the second condition becomes

$$\begin{aligned}
\frac{1}{b_n^2} \sum_{k=1}^n \mathbb{E}(\overline{X}_{nk}^2) &= \frac{n}{b_n^2} \mathbb{E}(X_1^2 1_{\{|X_1| \leq b_n\}}) \\
&= \frac{n}{b_n^2} \sum_{j=1}^{m_n} (2^j)^2 2^{-j} \\
&= \frac{n}{b_n^2} \sum_{j=1}^{m_n} 2^j \\
&\leq n 2^{m_n+1} b_n^2 = \frac{2n}{b_n} \to 0.
\end{aligned}$$

Hence the problem reduces to finding sequences for which:

$$b_n = 2^{m_n}, \qquad m_n = \log_2 n + k_n, \qquad k_n \to \infty, \qquad b_n = o(n).$$

Then we will have

$$a_n = n\mathbb{E}[X_1 1_{\{|X_1| \leq b_n\}}] = nm_n.$$

By the weak law,

$$\frac{S_n - a_n}{b_n} = \frac{S_n - nm_n}{n2^{k_n}} \to 0.$$

So taking $c_n = n2^{k_n}$, the condition $\frac{S_n}{c_n} \to 1$ becomes

$$\begin{aligned} \frac{m_n}{2^{k_n}} &\to 1 \\ \sim \frac{\log_2 n}{2^{k_n}} &\to 1 \\ k_n &= \sup\{k \colon k \le \log_2 \log_2 n, \log_2 n + k \in \mathbb{N}\} \end{aligned}$$

So $k_n \sim \log_2 \log_2 n \to \infty$ and

$$\frac{2n}{b_n} \sim 2^{-k_n} \sim \frac{1}{\log_2 n} \to 0.$$

Hence you should charge $n \log_2 n$.

Lecture 9: October 13

Example 9.1 (St. Petersburg Paradox). *Let* $X_1, \cdots$ *be iid with* $\mathbb{P}(X_1 = 2^j) = 2^{-j}$*. Then* $\mathbb{E}X_1 = \infty$ *and* $S_n = \sum_{i=1} X_i$*. Then*

$$\frac{S_n}{n \log n} \to 1$$

in probability.

This is typical of WLL: it grows to ∞ faster than n. We derived it using the "truncation" (triangular arrays) method.

Take $b_n = 2^{m_n}$, we want $n2^{-m_n} = 2^{-k_n}$ with $k_n \to \infty$ i.e. $m_n = \log_2 n + k_n$. He took m_n to be an integer. In fact, we can simplify by not insisting that m_n is an integer. Then

$$\sum_{k=1}^{n} \mathbb{P}(|X_k| > b_n) = n\mathbb{P}(X_1 > 2^{\lfloor m_n \rfloor}) = n2^{-\lfloor m_n \rfloor} \le \frac{2n}{b_n} \to 0.$$

Now take $k_n = \log_2 \log_2 n$ straightaway.

Strong LLN

Theorem 9.2. *Let* $X_1, \cdots$ *be pairwise independent and identically distributed and* $\mathbb{E}X_1 < \infty$*. Then*

$$\frac{S_n}{n} \to \mu = \mathbb{E}X_1 \qquad \textit{a.s. (which means a.e.)}$$

The idea is to start with

$$\frac{S_n}{n} \to \mu$$

in probability. Then extract a subsequence n_k for which the convergence is a.s. Use this to bootstrap the original convergence to a.s.

Back to the St. Petersburg paradox. We should charge them ∞ to play, but no one would pay. So we only allow players to play N games, and charge them $S_N = N \log_2 N$ to play. Then we charge them $\log_2 N$.

Brief foray into Martingales: $\mathbb{E}M_k = \mathbb{E}M_{k-1}$. The winning strategy is to wait for first winning game them stop. So let T be the stopping time; we compute $\mathbb{E}M_T$, and ask whether $T < \infty$.

Chapter 3: Central Limit Theorem

Let $X_1, \cdots$ be iid with $\mathbb{E}|X_1|^2 < \infty$. Let $\mu = \mathbb{E}X_1$ and $\sigma^2 = \mathrm{Var}(X_1)$. We know that $\frac{S_n}{n} \to \mu$ but we want more precise knowledge about the convergence. We can consider quantities such as $a_n\left(\frac{S_n}{n} - \mu\right)$ and find a sequence for which convergence occurs. The result is:

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \to \sigma\chi$$

where $\chi \sim N(0,1)$ is the standard normal distribution. The way this is stated in textbooks is frequently:

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) \to \mathbb{P}(\chi \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2} 2 \, dy.$$

The textbook develops the theory of convergence of characteristic function to show convergence in probability. It states that

$$F_n(x) \to F(x) \iff \mu_n \to \mu \iff \mathbb{E}e^{i\eta\xi_n} \to \mathbb{E}e^{i\eta\xi}$$

where x is point of continuity of $F(x)$ and ξ_n, ξ are $F_n(x), F(x)$.

We're going to do a different proof developed by Stein that is more probabiistic. It allows one to work in the case with dependence present, instead of the Fourier Transform method (read textbook chapter 3). Note that the characteristic method breaks down in the presence of dependence, because it is difficult to compute the characteristic function in this case.

Lemma 9.3. *Suppose W is a random variable with $\mathbb{E}|W| < \infty$. Then $W \sim N(0,1)$ iff for any $f \in C_b^1$,*

$$\mathbb{E}(f(W)) = \mathbb{E}(Wf(W)).$$

Here $C_b^1 = \{f \in C_b \colon f' \in C_b$ where C_b denotes bounded continuous functions.

In particular elements of C_b^1 are Lipschitz.

Proof. $\Rightarrow$ If $W \sim N(0,1)$, then

$$\begin{aligned}
\mathbb{E}(f') &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(x) e^{-\chi^2/2}\, dx \\
\text{(take limits to justify)} \qquad &= \frac{1}{\sqrt{2\pi}} \left[e^{-x^2/2} f' \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f \, d(e^{-x^2/2}) \right] \\
&= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} x f e^{-x^2/2} \right] \\
&= \mathbb{E}(Wf(W)).
\end{aligned}$$

$\Leftarrow$ Consider the ODE

$$f' - xf = 1_{\{(-\infty, z]\}} - \Phi(z) \qquad \text{where } \Phi(z) = \mathbb{P}(\chi \le z) \qquad (*)$$

Suppose for every z, we can find f for which $(*)$ holds. Then

$$\begin{aligned}
f'(W) - Wf(W) &= 1_{\{W \le z\}} - \Phi(z) \\
\mathbb{E}[f(W) - Wf(W)] &= \mathbb{E}[1_{\{W \le z\}} - \Phi(z)] \\
0 &= \mathbb{P}(W \le z) - \Phi(z).
\end{aligned}$$

Hence $W = \chi$ in distribution. The trouble is that $1_{\{W \le z\}}$ is not continuous, so we replace it with a continuous function.

For $h \in C_b$ and $z \in \mathbb{R}$, consider the ODE

$$f' - xf = h - \mathbb{E}h(Z).$$

Then we solve the ODE as follows:

$$\begin{aligned}
\left(fe^{-x^2/2}\right)' e^{x^2/2} &= h(x) - \mathbb{E}h(Z) \\
\left(fe^{-x^2/2}\right)' &= e^{-x^2/2}(h(x) - \mathbb{E}h(Z)) \\
fe^{-x^2/2} &= \int_{\infty}^{x} e^{-y^2/2}(h(y) - \mathbb{E}h(Z))\, dy \\
f_h(x) = f(x) &= e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2}(h(y) - \mathbb{E}h(Z))\, dy.
\end{aligned}$$

Verify that $f_h \in C_b^1$. Then we have

$$f_h'(W) - W f_h(W) = h(W) - \mathbb{E}h(Z).$$

Taking expectation yields $\mathbb{E}h(W) = \mathbb{E}h(Z)$. Let $h_n \in C_b$ such that $h_n \searrow 1_{\{(-\infty,z]\}}$. By BCT it follows that

$$\mathbb{P}(W \le z) = \mathbb{E}1_{\{(-\infty,z]\}}(W) = \mathbb{E}h(Z).$$ □

How to show that W is "close" to χ? We will compute

$$\mathbb{P}(W \le z) - \mathbb{P}(\chi \le z) = \mathbb{E}[f_z(W) - W f_z(W)].$$

Lecture 10: October 15

CLT via Stein's method

Note: This is not Elias Stein from the books but a different Stein.

Lemma 10.1 (His 2.1). *For a random variables W with $\mathbb{E}|W| < \infty$, then*

$$W \sim N(0,1) \iff \mathbb{E}[f'(W)] = \mathbb{E}[Wf(W)] \qquad \forall f \in C_b^1.$$

Sketch/reminder. For $h \in C_b$ and $z \in \mathbb{R}$, we construct f_h such that

$$f_h'(x) - xf_h(x) = h_z(x) - \mathbb{E}h(Z).$$

We solve it out and find

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2}(h_z(y) - \mathbb{E}h(Z))\, dy.$$

The idea is to approximate $h = 1_{\{(-\infty,z]\}}$ with fuctions in C_b. □

The case when $W \sim N(0,1)$ is the "perfect" case, but many times we want to work with something "close" to normal.

Example 10.2. *Here $W = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. We want to measure $\chi \sim N(0,1)$. Here*

$$1_{\{(-\infty,z]\}}(W) - \mathbb{E}h(Z) = f_z'(W) - Wf_z(W)$$
$$\mathbb{P}(W \le z) - \mathbb{E}h(Z) = \mathbb{E}[f_z'(W) - Wf_z(W)].$$

When the function is not smooth, the estimate isn't so good. But this is not a serious problem, we can smooth it like last time. We are using the left side to approximate the right side.

Lemma 10.3 (His 2.2). *$xf_z(x)$ increases in x, and*

$$|xf_z| \le 1, \qquad |xf_z(x) - yf_z(y) \le 1$$
$$|f_z'| \le 1, \qquad |f_z'(x) - f_z'(y)| \le 1$$
$$0 < f_z \le \min\{\frac{\sqrt{2\pi}}{4}, \frac{1}{|z|}\}$$
$$|(x+y)f_z(x+y) + (x+v)f_z(x+v)| \le (|x| + \frac{\sqrt{2\pi}}{4})(|y| + |v|).$$

Let $\chi = Z \sim N(0,1)$.

Lemma 10.4 (His 2.3). *If h is AC (absolutely continuous), then*

$$\|f_h\|_\infty \le \min\{\sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(\chi)\|_\infty, 2\|h'\|_\infty\}$$
$$\|f_h'\|_\infty \le \min\{2\|h - \mathbb{E}h(\chi)\|_\infty, 4\|h'\|_\infty\}$$
$$\|f_h''\|_\infty \le 2\|h'\|_\infty.$$

Lemma 10.5. *Assume there is a* $\delta > 0$ *such that*

$$|\mathbb{E}[h(W) - \mathbb{E}h(Z)]| \le \delta \|h'\|_\infty$$

for all Libschitz h.

Recall that Lipschitz implies AC implies BV implies derivative exists a.e., so we can take essential supremum and everything is well-defined.

Definition 10.6. *Consider* $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$. *Then the induced distribution* μ *is called the* **law of** ξ *and is denoted* $\mathcal{L}(\xi)$.

Note that the laws of X, Y agree whenever X, Y agree in distribution.

Definition 10.7. *Define the Wasserstein and Kolmogorov metrics as follows:*

$$\begin{aligned} d_W(\mathcal{L}(W), \mathcal{L}(Z)) &= \sup_{h \in \text{Lip}(1)} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \le \delta \\ d_K(\mathcal{L}(W), \mathcal{L}(Z)) &= \sup_z |\mathbb{P}(W \le z) - \mathbb{P}(Z \le z)| \le 2\sqrt{\delta} \end{aligned}$$

Proof. It suffices to prove it for d_K. Take $\delta \in (0, \frac{1}{4})$ for otherwise the result is trivial. Take $a = \sqrt{\delta}(2\pi)^{1/4}$. For fixed $z \in \mathbb{R}$, let

$$h_a(x) = \begin{cases} 1 & x \le z \\ 0 & x \ge z + a \\ \text{linear} & x \in [z, z+a] \end{cases}$$

Note that $a < 1$, and $\text{Lip}(h_a) = 1/a$. Observe that

$$\begin{aligned} \mathbb{P}(W \le z) - \mathbb{P}(Z \le z) &\le \mathbb{E}h_z(W) - \mathbb{E}h_a(Z) + (\mathbb{E}h_a(Z) - \mathbb{E}h(Z)) \\ &\le \delta \cdot \frac{1}{a} + \mathbb{P}(z \le Z \le z + a) \\ &\le \frac{\delta}{a} + \frac{a}{\sqrt{2\pi}} = \frac{2\delta}{a} \\ &= \sqrt{\delta} \frac{2}{(2\pi)^{1/4}} \\ &\le 2\sqrt{\delta}. \end{aligned}$$

Now for the lower bound, we use the same kind of argument.

$$\mathbb{P}(W \le z) - \mathbb{P}(Z \le z) \ge -2\sqrt{\delta}.$$

□

Hence when you only need convergence, you only need to show the simple estimate in Lemma 10.5. If you need to know the rate of convergence, it's better to appeal to Lemma 10.3 directly. Now we'll discuss how to verify the bound in the hypothesis of Lemma 10.5.

Goal 10.8. *$X_1, \cdots$ iid. Let $W_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$. Then $W_n \to Z \sim N(0,1)$ in probability.*

Note 10.9. *We define*

$$W_n = \sum_{i=1}^n \xi_i, \qquad \xi_i = \frac{X_i - \mu}{\sqrt{n}\sigma}.$$

Then $\xi_1, \cdots$ are iid, $\mathbb{E}\xi_1 = 0$, and $\sum_{i=1}^n \mathbb{E}[\xi_i^2] = 1$.

Now let ξ_i be any independent random variables such that $\mathbb{E}\xi_k = 0$ and $\sum_i \mathbb{E}\xi_i^2 = 1$. Set $W = \sum \xi_i$ and $W^{(i)} = W - \xi_i$.

Definition 10.10.

$$K_i(t) = \mathbb{E}\xi_i \left[1_{\{0 \le t \le \xi_i\}} - 1_{\{\xi_i \le t \le 0\}}\right]$$

Note that $K_i \ge 0$. Moreover

$$\int_{-\infty}^{\infty} K_i(t)\ dt = \mathbb{E}[\xi_i^2], \qquad \int_{-\infty}^{\infty} |t| K_i(t)\ dt = \frac{1}{2}\mathbb{E}[|\xi_i|^3].$$

Next we have

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}[f_h'(W) - W f_h(W)].$$

We compute that

$$\begin{aligned}
\mathbb{E}[W f_h(W)] &= \sum_{i=1}^n \mathbb{E}[\xi_i f_h(W)] \\
&= \sum_{i=1}^n \mathbb{E}[\xi_i (f_h(W) - f_h(W^{(i)}))] \\
&= \sum_{i=1}^n \mathbb{E}[\xi_i \int_0^{\xi_i} f'(W^{(i)} + t)\ dt] \\
&= \sum_i \mathbb{E}[\xi_i \int_{-\infty}^{\infty} f'(W^{(i)} + t)(1_{\{0 \le t \le \xi_i\}} - 1_{\{\xi_i \le t \le 0\}})] \\
\text{Use Fubini and indep.} \qquad &= \sum_i \int_{-\infty}^{\infty} \mathbb{E}[f'(W^{(i)} + t) K_i(t)\ dt]
\end{aligned}$$

Lecture 11: October 17

Given h, the function

$$f_h(x) = e^{x^2/2}\int_{-\infty}^{x} e^{-y^2/x}(h(y) - \mathbb{E}h(z))\ dy$$

is a solution to the differential equation

$$f_h'(x) - xf_h(x) = h(x) - \mathbb{E}h(Z),$$

where $Z \sim N(0,1)$. We are approximating $1_{\{-\infty,\xi]\}}$ with a smooth function, so to be consistent we approximate $\phi(\xi)$.

Lemma 11.1 (Lemma 2.4). *$\|f_h'\| \le \min\{2\|h - \mathbb{E}h(Z)\|_\infty, 2\|h'\|_\infty\}$ and $\|f_n'\|_\infty \le 2\|h'\|_\infty$.*

Recall that $K_i(t) = \mathbb{E}(\xi_i(1_{\{0\le t\le \xi_i\}} - 1_{\{\xi_i \le t \le 0\}})$ and

$$\int_{-\infty}^{\infty} K_i(t)\ dt = \mathbb{E}(\xi_i^2),\ \int_{-\infty}^{\infty} |t|K_i(t)\ dt = \frac{1}{2}\mathbb{E}(|\xi_i|^3).$$

Also for $f = f_n$, we have

$$\mathbb{E}(Wf(W)) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\mathbb{E}(f'(W^{(i)} + t))K_i(t)\ dt.$$

Recall that we want to show that W is close to Z, so we are considering

$$\mathbb{E}(h(W)) - \mathbb{E}(h(Z)) = \mathbb{E}(f_h'(W) - Wf_h(W)).$$

Now we compute $\mathbb{E}f'(W)$. Recall that $\sum \mathbb{E}(\xi_i^2) = 1$ and

$$\mathbb{E}(f'(W)) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\mathbb{E}(f'(W))K_i(t)\ dt.$$

Thus $\mathbb{E}(f'(W) - Wf(W)) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\mathbb{E}(f'(W) - f'(W^{(i)} + t))K_i(t)\ dt$. Note that we have yet to use any property specific to f_h so far.

Theorem 11.2 (2.5). *Let ξ_i be independent with $\mathbb{E}(\xi_i) = 0$ and $\sum_{i=1}^{n}\mathbb{E}(\xi_i^2) = 1$. Assume that $\mathbb{E}(|\xi_i|^3) < \infty$. Then Lemma 2.4 holds with $\delta = 3\sum_{i=1}^{n}\mathbb{E}(|\xi_i|^2)$. That is*

$$\sup_{h\in \mathrm{Lip}(1)} |\mathbb{E}(h(W)) - \mathbb{E}h(Z)| \le \delta.$$

Here $W = \sum_{i=1}^{n}\xi_i$ and $W^{(i)} = W - \xi_i$.

Proof. For $h \in \mathrm{Lip}(1)$, i.e. $\|h'\|_\infty \le 1$, we have

$$\|\mathbb{E}(h(W)) - \mathbb{E}(h(Z))| = \|\mathbb{E}(f_h'(W) - Wf_h(W))\|$$

And so forth. □

Lecture 12: October 20

Theorem 12.1 (2.4). *Assume there exists $\delta > 0$ such that for any uniformly Lipschitz function h, we have*

$$\|\mathbb{E}h(W) - \mathbb{E}h(Z)| \le \delta \|h'\|_\infty \qquad (1)$$

where $Z \sim N(0,1)$. Then the Kolmogorov distance is

$$\sup_z \|\mathbb{P}(W \le z) - \mathbb{P}(Z \le z)\| \le 2\sqrt{\delta}.$$

This follows from Stein's method.

Theorem 12.2 (2.6). *Let $\xi_1, \cdots, \xi_n$ be iid such that $\mathbb{E}\xi_1 = 0$ and $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$. Then (1) holds with*

$$\delta = 4(4\beta_2 + 3\beta_3), \qquad \beta_2 = \sum_{i=1}^n \mathbb{E}(\xi_i^2 1_{\{|\xi_i| \ge 1\}}), \qquad \beta_3 = \sum_{i=1}^n \mathbb{E}(|\xi_i|^3 1_{\{|\xi_i| \le 1\}})$$

and $W = \sum_{i=1}^n \xi_i$.

Theorem 12.3 (Lindeberg-Feller, 2.7). *For each n, let X_{nk} for $1 \le k \le n$ be independent with $\mathbb{E}X_{nk} = 0$. Suppose the following conditions hold:*

(a) $\sum_{i=1}^n \mathbb{E}(X_{ni}^2) \to \sigma^2 > 0$

(b) $\sum_{i=1}^n \mathbb{E}(|X_{ni}|^2 1_{\{|X_{ni}| > \varepsilon\}}) \to 0$

Then $S_n = \sum_{i=1}^n X_{ni}$ converges to σZ, where $Z \sim N(0,1)$.

Condition (b) is called "Lindeberg's Condition". It says that individual terms don't contribute very much. If you have one term dominate, you can get Poisson limits.

Proof. Let $\xi_k = \frac{X_{nk}}{\sigma_n}$ where $\sigma_n^2 = \sum_{i=1}^n \mathbb{E}(X_{ni}^2)$, and set $W_n = \sum_{i=1}^n \xi_k$. Next, we have

$$\beta_2 + \beta_3 = \frac{1}{\sigma_n^2} \sum_{k=1}^n \mathbb{E}(X_{nk}^2 1_{\{|X_{nk}| > \sigma_n\}}) + \frac{1}{\sigma_n^3} \sum_{k=1}^n \mathbb{E}(|X_{nk}|^3 1_{\{|X_{nk}| \le \sigma_n\}}).$$

The σ_n's are a little annoying since we want things in terms of ε (from Lindeberg's Condition). We bound in terms of ε. β_2 is fine, but β_3 is all wrong. We have the inequality going the wrong direction, and we have cubes. Now take ε sufficiently small. Trading ε with $|X_{nk}|$ over and over, we obtain

$$\begin{aligned}
\beta_2 + \beta_3 &\le \frac{1}{\sigma_n^2} \sum_{k=1}^n \mathbb{E}(X_{nk}^2 1_{\{|X_{nk}| > \sigma_n\}}) + \frac{1}{\sigma_n^3} \sum_{k=1}^n \mathbb{E}(|X_{nk}|^3 1_{\{|X_{nk}| < \varepsilon\}}) + \frac{1}{\sigma_n^3} \sum_{k=1}^n \mathbb{E}(|X_{nk}|^3 1_{\{\varepsilon < |X_{nk}| \le \sigma_n\}}) \\
&\le \frac{1}{\sigma_n^2} \sum_{k=1}^n \mathbb{E}(X_{nk}^2 1_{\{|X_{nk}| > \varepsilon\}}) + \frac{\varepsilon}{\sigma_n^3} \sum_{k=1}^n \mathbb{E}(|X_{nk}|^2 1_{\{|X_{nk}| < \varepsilon\}}) \\
&\le \frac{1}{\sigma_n^2} \sum_{k=1}^n \mathbb{E}(X_{nk}^2 1_{\{|X_{nk}| > \varepsilon\}}) + \frac{\varepsilon}{\sigma_n^3} \sum_{k=1}^n \mathbb{E}|X_{nk}|^2 \\
&= \frac{1}{\sigma_n^2} \sum_{k=1}^n \mathbb{E}(X_{nk}^2 1_{\{|X_{nk}| > \varepsilon\}}) + \frac{\varepsilon}{\sigma_n} \\
&\to \frac{\varepsilon}{\sigma}.
\end{aligned}$$

By Lindeberg's Condition, this tends to $\frac{\varepsilon}{\sigma}$. Since $\varepsilon > 0$ was arbitrary, we get the result. □

Theorem 12.4 (Classical CLT). *Suppose $X_1, \cdots$ are iid such that $\mathbb{E}X_i = \mu$ and $\operatorname{Var}(X_i) = \sigma^2$. Then*

$$\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \to Z, \qquad Z \sim N(0,1).$$

Proof. Let $X_{nk} = \frac{X_k - \mu}{\sigma\sqrt{n}}$. Then

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}(X_{ni}^2 1_{\{|X_{ni}|>\varepsilon\}}) &= n\left(\mathbb{E}\left(\frac{X_1-\mu}{\sqrt{n}\sigma}\right)^2 1_{\{|X_1-\mu|>\sqrt{n}\sigma\varepsilon\}}\right) \\ &= \frac{1}{\sigma^2}\mathbb{E}(X_1-\mu)^2 1_{\{|X_1-\mu|>\sqrt{n}\}}\sigma\varepsilon \\ &\to 0.\end{aligned}$$

Now for a slightly non-standard example. Let π_n be a uniformly random permutation of $[n]$. Let K_n be the number of cycles in π_n.

Claim 12.5. $\frac{K_n - \log n}{\sqrt{\log n}} \to Z, \qquad Z \sim N(0,1)$.

Proof. The nice things about uniformly random permutations is that there is a very nice way to build them. We introduce the Chinese restaurant process. There are lots of tables. The first person sits at table 1, choosing any spot. Person 2 either sits at table 1 or chooses table 2. Each time someone enters, they either sit at an existing table or form a new one. Form a permutation by letting the tables be cycles and reading off clockwise. By induction, it is not hard to see that the resuting permutation is uniformly random.

Let $X_k = 1_{\{k \text{ forms a new table}\}}$. By construction the X_k are independent. Now X_k is Bernoulli with $\mathbb{P}(X_k = 1) = \frac{1}{k}$, and $K_n = \sum_{i=1}^n X_i$. Observe that

$$\mathbb{E}K_n = \sum_{k=1}^n \frac{1}{k} \sim \log k, \qquad \operatorname{Var}(K_n) = \sum_{k=1}^n \operatorname{Var}(X_k) = \sum_{k=1}^n \frac{1}{k} - \frac{1}{k^2} \sim \log n.$$

Let $X_{nk} = \frac{X_k - \frac{1}{k}}{\sqrt{\log n}}$. Then $\mathbb{E}X_{nk} = 0$. Moreover,

$$\sum_{i=1}^n \mathbb{E}(X_{ni}^2) \to 1.$$

Now $|X_{ni}| \le \frac{1}{\sqrt{\log n}}$. For large n, observe that

$$\sum_{i=1}^n \mathbb{E}(X_{ni}^2 1_{\{|X_{ni}|>\varepsilon\}}) = 0.$$

By Lindeberg-Feller,

$$\frac{K_n - \sum_{k=1}^n \frac{1}{k}}{\sqrt{\log n}} \to Z.$$

□

Lecture 13: October 22

Before we move away from CLT to talk about Borel-Cantelli, there's something we should know. There is a complete characterization of distributions F such that if $X_1, \cdots$ are iid $\sim F$, then there exist constants a_n, b_n such that

$$a_n\left(\sum_{i=1}^n X_i - b_n\right)$$

converges in distribution to a non-trivial limit.

The classical CLT says that if F has finite variance, you can get a standard normal. In fact you can get other distributions: it is a completely solved problem, ask Chen for more info.

Borel-Cantelli

There are lots of convergences in probability theory, and sometimes one is easier to work with than another. We would like to leverage convergence in probability and get almost sure convergence out of it.

If $\{A_n\}_{n\geq 1}$ are a sequence of subsets of Ω, define

$$\limsup_n A_n = \lim_{m\to\infty} \bigcup_{n=m}^\infty A_n = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty A_n = \{w : w \in A_n \text{ i.o.}\} = \{A_n \text{ i.o.}\}.$$

Here i.o. stands for infinitely often.

Similary we can form the set

$$\liminf A_n = \bigcup_{m=1}^\infty \bigcap_{n=m}^\infty A_n = \{w : w \in A_n \text{ for all but finitely many } n\}.$$

Lemma 13.1 (Borel-Cantelli). *If $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$.*

The proof is really simple but also important.

Proof. Let $N = \sum_{n=1}^\infty 1_{\{A_n\}}$. Note that $\{N = \infty\} = \{A_n \text{ i.o.}\}$. Fubini's Theorem implies that

$$\begin{aligned} \mathbb{E}N &= \mathbb{E}\sum_{n=1}^\infty 1_{\{A_n\}} \\ &= \sum_{n=1}^\infty \mathbb{E}1_{\{A_n\}} \\ &= \sum_{n=1}^\infty \mathbb{P}(A_n). \end{aligned}$$

By hypothesis this is finite, so $\mathbb{E}N < \infty$. THus $\mathbb{P}(N = \infty) = 0$. □

Showing finite expectation yields finite a.e., this is a general technique.

Theorem 13.2. *$X_n \to X$ in probability iff every subsequence X_{n_k} has a subsequence $X_{n(m_k)} \to X$ a.s.*

This gives an indication that almost sure convergence is really weird. It doesn't arise from a metric space structure, in particular. Most of the other convergences people use comes from a topology.

Proof. Let $\varepsilon_k \searrow 0$. Choose $n(m_k) > n(m_{k=1})$ such that

$$\mathbb{P}(|X_{n(m_k)} - X| > \varepsilon_k) \leq 2^{-k}.$$

Let $A_k = \{|X_{n(m_k)} - X| > \varepsilon_k\}$. Now $\sum \mathbb{P}(A_k) < \infty$, so be Borel-Cantelli we get the result. The other direction is trivial, since convergence in probability is metrizable. □

Corollary 13.3. *If f is continuous and $X_n \to X$ in probability, then $f(X_n) \to f(X)$ in probability. If f is also bounded, then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.*

Not that these are hard to prove without the theorem, but it's the lazy person's way of proving them.

Theorem 13.4. *Let $X_1, \cdots$ be iid with $\mathbb{E}X_1 = \mu$ and $\mathbb{E}(X_1^4) < \infty$. Then*

$$\frac{\sum_{i=1}^n X_i}{n} \to \mu$$

almost surely.

Now the fourth moment assumption is rather mild in practice. The most general case allows all the moments to be finite or infinite. If they are undefined, things are a little more weird. Basically this is an extremely intuitive and basic fact that should hold under very mild assumptions.

Proof. Assume $\mu = 0$. We only need the X_i to be sufficiently uncorrelated. This is just like in the L^2 weak law proof. We really need cancellation to happen, since otherwise we could have ± 1 and they would hit n^4 and cause problems. By Chebyshev,

$$\begin{aligned}
\mathbb{P}(|S_n| > n\varepsilon) &\leq \frac{\mathbb{E}(S_n^4)}{n^4\varepsilon^4} \\
&= \mathbb{E}(S_n^4) = \sum_{1\leq i,j,k,l\leq n} \mathbb{E}(X_i X_j X_k X_l) \\
&= \sum \mathbb{E}(X_i^4) + \sum \mathbb{E}(X_i^2 X_j^2) \\
&= n\mathbb{E}(X_1^4) + \binom{4}{2}\binom{n}{2}\mathbb{E}(X_1^2 X_2^2).
\end{aligned}$$

Hence the quantity grows like n^2 and we get

$$\mathbb{P}(|S_n/n| > \varepsilon \text{ i.o.}) = 0.$$

Hence for all ε, we have $\mathbb{P}(\limsup_n |S_n/n| \leq \varepsilon) = 1$, whereupon $\limsup_n |S_n/n| = 0$ almost surely. □

In the large scheme of things, this was a fairly pleasant proof. On Friday we'll do the more general version, which requires a little more work but is also not terrible.

There are other Borel-Cantellis.

Lemma 13.5 (Second Borel-Cantelli). *If $A_1, \cdots$ are independent and $\sum \mathbb{P} A_i = \infty$, then $\mathbb{P}(A_n \text{ } i.o.) = 1$.*

There is a stronger statement which is not referred to as a Borel-Cantelli lemma, but it is similar in spirit.

Lemma 13.6. *If $A_1, \cdots$ are pairwise independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ then*

$$\frac{\sum_{i=1}^n 1_{\{A_i\}}}{\sum_{i=1}^n \mathbb{P}(A_i)} \to 1.$$

Note that the hypotheses are weaker and results stronger than the Second Borel-Cantelli lemma.

Proof of Second Borel-Cantelli.

Fix $M < N < \infty$. Since $1 - x \le e^{-x}$, we have

$$\begin{aligned} \mathbb{P}(\bigcap_{i=M}^{N} A_i^c) &= \prod_{i=M}^{N} (1 - \mathbb{P}(A_i)) \\ &\le \prod_{i=M}^{N} e^{-\mathbb{P}(A_i)} \\ &= \exp\left(- \sum_{i=M}^{N} \mathbb{P}(A_i) \right) \to 0. \end{aligned}$$

Hence $\mathbb{P}\left(\bigcup_{i=M}^N A_i\right) \to 1$ as $N \to \infty$.

Lecture 14: October 24

We weaken the hypotheses of the Strong Law of Large Numbers. This is not the original proof, it's a proof from the 80s. Non-obvious even in retrospect. This is a very careful proof of the Weak LLN for triangular arrays where everything is kept track of at every stage.

Theorem 14.1. *Let $X_1, \cdots$ be pairwise independent and identically distributed. Suppose $\mathbb{E}|X_1| < \infty$ and let $\mu = \mathbb{E}X_1$ and $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{S_n}{n} \to \mu \ a.s.$$

Proof. The main issue we have with our X_i is that they are unbounded, so we start with a truncation. So let $Y_k = X_k 1_{\{|X_k| \le k\}}$. Each Y_k is bounded, but non-uniformly.

Lemma 14.2. *If $T_n = \sum_{i=1}^n Y_i$, then*

$$\frac{|S_n - T_n|}{n} \to 0 \ a.s.$$

Proof. Well we only know one tool to prove a.s. convergence: Borel-Cantelli.

$$\begin{aligned}\sum_{i=1}^{\infty} \mathbb{P}(|X_k| > k) &\le \int_0^\infty \mathbb{P}(|X_1| > t)\ dt \\ &= \mathbb{E}|X_1| < \infty.\end{aligned}$$

This is the typical way of comparing infinite decreasing sequences to integrals, and recall that the left side is independent of k (due to the identical distribution.)

By Borel-Cantelli, $\mathbb{P}(X_k \ne Y_k \text{ i.o.}) = 0$. Now for any $\omega \in \{X_k = Y_k \text{ i.o.}\}^c$, we have $S_n(\omega) = T_n(\omega)$ only differ at finitely many summands (indeed, X_k and Y_k have to agree beyond some finite point). Now the result follows. □

Now we prove a lemma whose motivation is somewhat mysterious right now (but we'll eventually use it in a Chebyshev and Borel-Cantelli argument).

Lemma 14.3.

$$\sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} \le 4\mathbb{E}|X_1| < \infty.$$

Proof.

$$\begin{aligned}\text{Var}(Y_k) \le \mathbb{E}(Y_k^2) &= \int_0^\infty 2y\mathbb{P}(|Y_k| > y)\ dy \\ &\le \int_0^k 2y\ \mathbb{P}(|X_1| > y)\ dy.\end{aligned}$$

Consequently by your favorite integration theorem (Fubini, MCT, DCT, etc.)

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{\mathbb{E}(Y_k^2)}{k^2} &\le \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^\infty 1_{\{y<k\}} 2y\ \mathbb{P}(|X_1| > y)\ dy \\ &= \int_0^\infty \left(\sum_{k=1}^{\infty} \frac{1}{k^2} 1_{\{y<k\}} \right) 2y\mathbb{P}(|X_1| > y)\ dy.\end{aligned}$$

Since $\mathbb{E}|X_1| = \int_0^\infty \mathbb{P}(|X_1| > y)\ dy$, we need to show that

$$2y \sum_{k=1}^{\infty} \frac{1}{k^2} 1_{\{y<k\}} \le 4.$$

This is just a bit of analysis, we've finished the probability portion of this computation.

If $m \ge 2$, we have

$$\sum_{k\ge m} \frac{1}{k^2} \le \int_{m-1}^{\infty} \frac{1}{x^2}\ dx = \frac{1}{m-1}.$$

When $y \ge 1$, the sum starts with $\lfloor y \rfloor + 1 \ge 2$, and we have

$$2y \sum_{k=\lfloor y \rfloor+1}^{\infty} \frac{1}{k^2} \le \frac{2y}{\lfloor y \rfloor} \le 4,$$

since $y \le 2\lfloor y \rfloor$ for $y \ge 1$.

For $0 \le y < 1$, we have

$$2y \sum_{k=1}^{\infty} \frac{1}{k^2} \le 2\left(1 + \sum_{k=2}^{\infty} \frac{1}{k^2}\right) \le 2(1+1) = 4.$$

□

Note that $X_1 = X_1^+ - X_1^-$ and the sequences X_n^+, X_n^- satisfy the same hypotheses. To show pairwise independence, note that taking positive parts is the same as applying the measurable function $x \mapsto x1_{\{x>0\}}$. Thus it suffices to assume $X_1 \ge 0$ a.s.

Fix $\alpha > 1$ and $\varepsilon > 0$. Define $k(n) = \lfloor \alpha^n \rfloor$. This k is going to be our spacing.

Using pairwise independence and interchange theorems,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(|\tau_{k(n)} - \mathbb{E}\tau_{k(n)}| > \varepsilon k(n)) &\le \frac{1}{\varepsilon}^2 \sum_{n=1}^{\infty} \frac{\mathrm{Var}(\tau_{k(n)})}{k(n)^2} \\
&= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{j=1}^{k(n)} \mathrm{Var}(Y_j) \\
&= \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \left(\mathrm{Var}(Y_j) \sum_{n:k(n)\ge j} \frac{1}{k(n)^2} \right) \qquad (\star)
\end{aligned}$$

Introducing the spacing is similar to what we did when extracting a subsequence to get a.s. convergence in Borel-Cantelli.

Now we've reduced to the analysis realm again. At this point it's good to actually plug in to our definition of $k(n)$, since we actually have to do something with it. Observe that

$$\begin{aligned}
\sum_{n:k(n)\ge j} \frac{1}{k(n)^2} &= \sum_{n:\lfloor \alpha^n \rfloor \ge j} \frac{1}{\lfloor \alpha^n \rfloor^2} \\
&\le \sum_{n:\alpha^n \ge j} \frac{4}{\alpha^{2n}}.
\end{aligned}$$

We've used $\lfloor \alpha^n \rfloor \ge \frac{\alpha^n}{2}$. Now the first n for which $\alpha^n \ge j$ will satisfy $\frac{1}{\alpha^{2n}} \le \frac{1}{j^2}$. Hence the geometric series is bounded by $\frac{4}{j^2}\frac{1}{1-\alpha^{-2}}$.

Substituting back into $(\star)$ yields

$$\sum_{n=1}^{\infty} \mathbb{P}(|\tau_{k(n)} - \mathbb{E}\tau_{k(n)}| > \varepsilon k(n)) \le \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \mathrm{Var}(Y_j) \frac{4}{j^2(1-\alpha^{-2})}$$
$$< \infty \text{ by (11)}$$

By Borel-Cantelli, we have

$$\mathbb{P}(|\tau_{k(n)} - \mathbb{E}\tau_{k(n)}| > \varepsilon k(n) \text{ i.o.}) = 0.$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$\frac{\tau_{k(n)} - \mathbb{E}(\tau_{k(n)})}{k(n)} \to 0 \text{ a.s.}$$

By DCT, $\mathbb{E}Y_k \to \mathbb{E}X_1$ as $k \to \infty$. Since $\mathbb{E}\tau_{k(n)} = \sum_{i=1}^{k(n)} \mathbb{E}(Y_k)$, it follows by the trivial case of Cesaro summation that $\frac{\mathbb{E}\tau_{k(n)}}{k(n)} \to \mathbb{E}X_1$. Consequently, $\frac{\tau_{k(n)}}{k(n)} \to \mathbb{E}X_1$ a.s.

We need to deal with what happens for $k(n) \le m < k(n+1)$. Since $Y_k \ge 0$ a.s. (the first time we're using this hypothesis), we have

$$\frac{\tau_{k(n)}}{k(n+1)} \le \frac{\tau_m}{m} \le \frac{\tau_{k(n+1)}}{k(n)}$$
$$\frac{k(n)\tau_{k(n)}}{k(n)k(n+1)} \le \frac{\tau_m}{m} \le \frac{k(n+1)\tau_{k(n+1)}}{k(n)k(n+1)}.$$

Note that

$$\frac{k(n+1)}{k(n)} = \frac{\lfloor \alpha^{n+1} \rfloor}{\lfloor \alpha^n \rfloor} \to \alpha$$

as $n \to \infty$. Thus

$$\frac{1}{\alpha}\mathbb{E}X_1 \le \liminf_{m\to\infty} \frac{\tau_m}{m} \le \limsup_{m\to\infty} \le \alpha \mathbb{E}X_1 \text{ a.s.}$$

Now since $\alpha > 1$ was arbitrary, by letting $\alpha \searrow 1$ we obtain

$$\lim_{m\to\infty} \frac{\tau_m}{m} = \mathbb{E}X_1 \text{ a.s.}$$

By (10), it follows that $\frac{S_n}{n} \to \mathbb{E}X_1$ a.s. □

This proof is convoluted and unmotivated, even in retrospect. The theorem dates back to the 30s, but this proof was constructed in 1981. The original proof came from the study of sums of random series. We start with an iid sequence. Suppose $X_1, \cdots$ are independent. Under what conditions does

$$\sum_{n=1}^{\infty} X_n$$

converge a.s.? It turns out that there is a fairly complete characterization of when such sequences converge.

Lecture 15: October 29

Exam chapter 1 to chapter 3 (measure theory), weak LLN, central limit theorem. In class exam, 50 minutes, basic problems, basic techniques and concepts. Except everyone will do well. The purpose of this course is for students to learn the basic idea sand techniques in advanced probability. Grad eam not taken as seriously as in undergrda study. Also challenging for him, writing a level that can be done in 50 minutes. Some proofs and calculations, not as long as homework problem, for example showing measurability of ϕ_0 map from homework 1.

Ask us to prove the Borel Cantelli lemma (5 minutes). Basic things, important techniques. Stein's method is going to be in it, we lectured on it and it's a simple technique. Weak convergence, defined in terms of convergence of the distribution function.

Lindeberg Feller you need to know, for how else can you apply it. He will supply the paper.

Random Walks

Kolmogorov's 0-1 Law. The first one, pretty useful. Tail sigma-field condition.

For $X_1, \cdots$ iid with $\mathbb{X}_1 = 0$ and $\sigma^2 = \text{Var}(X_1) = 1$, set $S_n = X_1 + \cdots + X_n$. This is a more general random walk.

Theorem 15.1. $\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty$ *a.s.*, $\qquad \liminf_n \frac{S_n}{\sqrt{n}} = -\infty$ *a.s.*

This is an easy application of Kolmogorov's 0-1 Law. Compute even it is larger than k and apply CLT. Thus SRW crosses a parabolar infinitely many times, and in particular is transient. We can ask for a faster growing function which is asymptotically the growth rate? In other words, find $f(n) \to \infty$ such that

$$\limsup \frac{S_n}{f(n)} = 1, \qquad \liminf \frac{S_n}{f(n)} = -1 \ a.s.$$

Due to symmetry, we only need to show one of these holds. Clearly $\lim \frac{f(n)}{\sqrt{n}} \to \infty$. Now we look at $\mathbb{P}(S_n \geq k\sqrt{n})$. It turns out that $f(n) = \sqrt{2n \log\log n}$, and this theorem is called the **Law of the Iterated Logarithm**.

This is a universality result, because we don't need to know the specific distribution of the random variables.

Stopping Times

Let $X_1, \cdots$ be a sequence of r.v. For intuition, we might think of them as being independent although this is not necessary. Think of X_i as stock price for day i, so independence is not important. This is an outcome indexed by time. Let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. This is the information up to time n.

Definition 15.2. *A* **filtration** *is an increasing sequence of σ-algebras.*

Definition 15.3. *A random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a* **stopping time** *with respect to the filtration $\{\mathcal{F}_n\}$ if for any $n < \infty$, the event $\{T = n\} \in \mathbb{F}_n$.*

Note that this is equivalent to $\{T \le n\} \in \mathcal{F}_n$. Notice that every deterministic function is a stopping time (it is a constant random variable). Why is this called a stopping time? Imagine T is the time to sell a stock. In the absence of insider trading, one does not know the future. Hence the measurability criterion.

Example 15.4 (Hitting time). *Given a Borel set $A \subset \mathbb{R}$, the quantity*

$$\sigma_A = \inf\{n: X_n \in A\}$$

is a stopping time.

Proof. For any $n \in \mathbb{N}$, we have

$$\begin{aligned}\{\sigma_A = n\} &= \{X_n \in A, \text{ but } X_j \notin A \text{ for } j < n\} \\ &= \{X_n \in A\} \cap \bigcap_{j=1} n-1\{X_j \notin A\} \qquad \subset \mathcal{F}_n \cap \mathcal{F}_j \subset \mathcal{F}_n.\end{aligned}$$

Example 15.5 (Exit time). *Let $\tau_A = \inf\{n: X_n \notin A\} = \sigma_{A^c}$.*

These are pretty easy proofs when our processes have discrete time. Imagine replacing the sequence X_i be a family $\{X_t, t \in [0,\infty)\}$. We still have the concept of stopping time and hitting time. But the general proof uses capacity theory for the proof. While stoppping times depend on the σ-algebra, when there is no confusion we just omit it.

Theorem 15.6. *Let $X_1, \cdots$ be iid, $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ and T is an $\{\mathcal{F}_n\}$ stopping time. Conditioned on $\{T < \infty\}$, the event $\{X_{T+n}, n \ge 1\}$ is independent of $\mathcal{F}_T$ and has the same distribution as the original process $\{X_n, n \ge 1\}$.*

Definition 15.7. *For T a stopping time, set*

$$\begin{aligned}\mathcal{F}_T &= \{A: A \cap \{T \le n\} \in \mathcal{F}_n \qquad \forall n \in \mathbb{N}\} \\ &= \{A: A \cap \{T = n\} \cdots\}\end{aligned}$$

Note that this is a σ-algebra. There are two statements in the theorem: independence, and identical distribution. Recall that we showed in the homework that sequences (r.v. in $\mathbb{R}^{\mathbb{N}}$) have the same distribution iff their finite truncations have the same distribution. Thus we need to show the following:

(a) $\mathbb{P}(A \cap B \mid T < \infty) = \mathbb{P}(A \mid T < \infty)\mathbb{P}(B \mid T < \infty)$ where $A \in \mathcal{F}_T$ and $B \in \sigma(X_{T+n}, n \ge 1)$.

(b) For all $A_1, \cdots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(\bigcap_{j=1}^n \{X_{T+j} \in A_j\} \mid T < \infty) = \mathbb{P}(\bigcap_{j=1}^n \{X_j \in A_j\}).$$

Some example events in $\mathcal{F}_T$ include $\{T \le k\}$, also if $A \in \mathcal{F}_n$ the event $B = A \cap \{T \ge n\}$ belongs to $\mathcal{F}_T$.

Example 15.8. *Let X_i be iid rv with $\mathbb{P}(X_1 = \pm 1) = \frac{1}{2}$. Let $S_n = \sum_{j=1}^n X_j$ and set $T = \inf\{n: S_n = 1\}$. Show that $\mathbb{P}(T < \infty) = 1$ but $\mathbb{E}T = \infty$.*

You can write down the density of T. Look at the probability that it takes $2k+1$ steps to reach 1.

The information in the first n steps is the same as the information of $S_1, \cdots, S_n$. The recentered random walker after time T, after recentering around S_T, has the same distribution as before we stopped it. This is called the **strong Markov property**. We will prove it on Monday.

Lecture 16: November 5

Theorem 16.1 (3.2). *Conditioned on $\{T < \infty\}$, the sequence $\{X_{T+n}, n \geq 1\}$ is independent of $\mathcal{F}_T$ and has the same distribution as the original process.*

Proof. Last time we saw that $\forall B_i \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}(X_{T+j} \in B_j, j = 1, \cdots, k \mid T < \infty) = \mathbb{P}(X_j \in B_j, 1 \leq j \leq k).$$

Thus all the finite sequences agree, so the infinite sequences agree by a homework problem. It remains to show independence from $\mathcal{F}_T$.

For all $A \in \mathcal{F}_T$ and $B_i \in \mathcal{B}(\mathbb{R}^n)$, we compute

COMPUTATION WAS COMMENTED OUT

Let $S_n = \sum_{k=1}^n X_k$, a random walk at time n. Then $\{S_1, \cdots, S_n\} \iff \{X_1, \cdots, X_n\}$ so

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

Corollary 16.2 (Strong Markov property for S_n). *For any $B_j \in \mathcal{B}(\mathbb{R}^d)$, $k \geq 1$, and $T < \infty$,*

$$\begin{aligned}\mathbb{P}(\{S_{T+j} \in B_j, 1 \leq j \leq k\} \mid \mathcal{F}_T) &= \mathbb{P}(\{S_{T+j} \in B_j, 1 \leq j \leq k\} \mid \sigma(S_T)) \\ &= \mathbb{P}(\{S_j + x \in B_j, 1 \leq j \leq k\})\big|_{x=S_T}\end{aligned}$$

Proof. Note that $S_{T+j} = S_T + \sum_{i=1}^j X_{T+i}$. Now the first term is x and the remaining part agrees in distribution with $\sum_{i=1}^j X_i$. For all $A \in \mathcal{F}_T$, we have

$$\begin{aligned}\int_{A \cap \{T<\infty\} \in \mathcal{F}_T} LHS \; dP &= \mathbb{P}(A \cap \{T < \infty\} \cap \{S_{T+j} \in B_j, 1 \leq j \leq k\}) \\ &= \mathbb{P}(A \cap \{T < \infty\} \cap \{S_T + \sum_{i=1}^j X_{T+i} \in B_j, 1 \leq j \leq k\}) \\ &= \int_{A \cap \{T<\infty\}} \mathbb{P}(\sum_{i=1}^j X_{T+i} \in B_j - x, 1 \leq j \leq k)\Big|_{x=S_T} dP \\ &= \int_{A \cap \{T<\infty\}} \mathbb{P}(\sum_{i=1}^j X_i \in B_j - x, 1 \leq j \leq k)\Big|_{x=S_T} dP \\ &= \int_{A \cap \{T<\infty\}} RHS \; dP\end{aligned}$$

Lecture 17: November 7

Stopping time T. Conditioned on $\{T < \infty\}$, we have

$$\{S_{T+n} - S_T, n \ge 1\} = \{\sum_{j=1}^n X_{T+j}, n \ge 1\} \stackrel{d}{=} \{\sum_{j=1}^n X_j, n \ge 1\} = \{S_n, n \ge 1\}$$

Blumenthal & Getoor discusses the differences between the Markov and Strong Markov properties in chapter 1.

We construct the **canonical sample space** for a sequence of iid random variables. Start with a random variables X_1 on $\mathbb{R}^d$ and let $\Omega = (\mathbb{R}^d)^{\mathbb{N}}$ and $X_j(\omega) = \omega_j$. Introduce the shift operator

$$\theta : \Omega \to \Omega \qquad \theta(\omega_1, \omega_2, \cdots) = (\omega_2, \cdots).$$

Let $S_n = \sum_{i=1}^n X_i$ and $S_0 = 0$. Define $T_1 = \inf\{n \ge 1 : S_n = 0\}$, the first return time to 0. For $n \ge 2$ let

$$T_n = \inf\{n > T_{n-1} : S_n = 0\}, \qquad n^{th} \text{ return time to } 0.$$

Using the Strong Markov property,

$$\begin{aligned}
\mathbb{P}(T_2 < \infty) &= \mathbb{P}(T_1 < \infty, T_2 < \infty) \\
&= \mathbb{P}(T_1 < \infty, T_1 + T_1 \circ \theta^{T_1} < \infty) \\
&= \mathbb{E}[\mathbb{P}(T_1 + T_1 \circ \theta^{T_1} < \infty \mid T_1 < \infty), T_1 < \infty] \\
&= \mathbb{E}[\mathbb{P}(T_1 \circ \theta^{T_1} < \infty \mid T_1 < \infty), T_1 < \infty] \\
&= \mathbb{E}[\mathbb{P}(T_1 < \infty), T_1 < \infty] \\
&= \mathbb{P}(T_1 < \infty)^2.
\end{aligned}$$

This is a rigorous derivation, but it is obvious intuitively: you have to return twice, and the events are independent by the Strong Markov property. However you should be a little careful, because $T_2 < \infty$ really says that you return at least twice, not exactly twice.

Similarly by induction, $\mathbb{P}(T_n < \infty) = \mathbb{P}(T_1 < \infty)^n$.

Recurrent and Transience (3.3)

Definition 17.1. *For a simple random walk (SRW) on $\mathbb{Z}^d$, we say it is* **recurrent** *if it returns to 0 infinitely many times. Otherwise it is called* **transient**.

Note that we mean "almost surely", that is with probabilit y1. In symbols,

$$\text{recurent} = \{T_n < \infty \text{ for all } n \ge 1\}$$

We compute the probability as follows:

$$\begin{aligned}
\mathbb{P}(T_n < \infty, \forall n \ge 1) &= \mathbb{P}(\bigcap_{n=1}^{\infty} \{T_n < \infty\}) \\
&= \lim_{n\to\infty} \mathbb{P}(T_n < \infty) \\
&= \lim_{n\to\infty} \mathbb{P}(T_1 < \infty)^n \\
&= \begin{cases} 1 & \mathbb{P}(T_1 < \infty) = 1 \\ 0 & \mathbb{P}(T_1 < \infty) < 1 \end{cases}
\end{aligned}$$

Thus S_n is recurrent if and only if $\mathbb{P}(T_1 < \infty) = 1$.

Theorem 17.2 (3.3). *For any random walk, TFAE:*

(a) $\mathbb{P}(T_1 < \infty) = 1$

(b) $\mathbb{P}(S_n = 0 \ io) = 1$

(c) $\sum_{n=1}^{\infty} \mathbb{P}(S_n = 0) = \infty$

Proof. It is clear that $(a) \iff (b)$. Let

$$\begin{aligned} V = \sum_{n=1}^{\infty} 1_{\{S_n=0\}} = \sum_{n=1}^{\infty} \mathbb{P}(T_n < \infty) \\ = \sum_{n=1}^{\infty} \mathbb{P}(T_1 < \infty)^n \\ = \frac{1}{1 - \mathbb{P}(T_1 < \infty)}. \end{aligned}$$

Thus $(a) \implies \mathbb{E}V = \infty$, which leads to $(a) \iff (c)$. □

Note that (c) is the most amenable to computation.

Theorem 17.3 (3.4). *SRW on $\mathbb{Z}^d$ is recurrent if and only if $d \le 2$.*

Proof. For $d = 1$, we have

$$\begin{aligned} \mathbb{P}(S_{2k} = 0) &= \binom{2k}{k} 2^{-2k} \\ \sum_{k=1}^{\infty} \mathbb{P}(S_{2k} = 0) &\approx \sum_{k=1}^{\infty} \frac{(k/e)^{2k}\sqrt{4\pi k}}{(k/e)^{2k}\sqrt{2\pi k}^2} \\ &\approx \sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty. \end{aligned}$$

Here we have used Stirling's Approximation and the following fact:

Fact 17.4. *If $f(k) \sim g(k)$, then $\sum f(k) < \infty \iff \sum g(k) < \infty$.*

Now consider $d = 2$. Then

$$\begin{aligned} \mathbb{P}(S_{2k} = 0) &= \sum_{i=0}^{n} \binom{2n}{i}\binom{2n-i}{i}\binom{2n-2i}{j} 4^{-2n} \\ &= \sum_{i=0}^{n} \binom{2n}{i, i, j, j} 4^{-2n} \end{aligned}$$

Lecture 18: November 10

Theorem 18.1 (3.4). *SRW on $\mathbb{Z}^d$ is recurrent when $d \le 2$ and transient when $d \ge 3$.*

Proof. We already did the $d = 1$ case, using the following criterion:

$$\sum_{n=1}^{\infty} \mathbb{P}(S_n = 0) = \infty \iff S_n \text{ is recurrent}$$

For $d = 2$, we take i steps left/right and j steps up/down. Thus

$$\begin{aligned}\mathbb{P}(S_{2n} = 0) &= \binom{2n}{i}\binom{2n-i}{i}\binom{2j}{j}4^{-2n} \\ &= \frac{(2n)!}{i!^2 j!^2}4^{-2n}.\end{aligned}$$

We can see this immediately as follows. Say we have $2n$ flags, in 4 colors (one for each direction). There are $(2n)!$ flag combinations, but $i!^2 j!^2$ you can't distinguish. In a second we apply Vandermonde's Identity, which says that choosing n people from $2n$ is the same as dividing up the $2n$ into two groups of n and choosing $i, n-i$ from each. Recalling that $j = n - i$,

$$\begin{aligned}\mathbb{P}(S_{2n} = 0) &= \binom{2n}{n}\sum_{i=0}^{n}\binom{n}{i}\binom{n}{j}4^{-2n} \\ &= \left(\binom{2n}{n}2^{-2n}\right)^2 \\ \text{(by } d = 1 \text{ case)} \qquad &\approx \left(\frac{1}{\sqrt{\pi n}}\right)^2 \\ &\approx \frac{1}{\pi n}.\end{aligned}$$

This sequence is convergent, so the walk with $d = 2$ is recurrent.

Now consider $d = 3$. We compute as follows:

$$\begin{aligned}\mathbb{P}(S_{2n} = 0) &= \sum_{\substack{i,j=1 \\ i+j\le n}} \binom{2n}{i, i, j, j, n-i-j, n-i-j}6^{-2n} \\ &= \binom{2n}{n}\sum_{\substack{i,j=1 \\ i+j\le n}}\binom{n}{i, j, n-i-j}^2 6^{-2n} \\ &\le \binom{2n}{n}\max_{i+j\le n}\binom{n}{i, j, n-i-j}\sum_{i+j\le n}\binom{n}{i, j, n-i-j}6^{-2n} \\ &= \binom{2n}{n}\max_{i+j\le n}\binom{n}{i, j, n-i-j}3^n 6^{-2n}\end{aligned}$$

Observe that the maximal multinomial cofficient occurs when all parts are equal:

$$\binom{n}{a_1, \cdots, a_k} \le \binom{n}{n/k, \cdots, n/k}$$

where each n/k is rounded up or down so as to make the sum n. To prove this fact, observe that we can decrease the largest and increase the smallest to increase the multinomial coefficient.

Consequently

$$\begin{aligned}
\binom{2n}{n}\max_{i+j\le n}\binom{n}{i,j,n-i-j}3^n6^{-2n} &\le \binom{2n}{n}\frac{n!}{[n/3]!^3}12^{-n} \\
\text{(Sterling)} \qquad &\approx \frac{4^n}{\sqrt{\pi n}}\frac{(n/e)^n\sqrt{2\pi n}}{\left((n/3e)^{n/3}\sqrt{2\pi n/3}\right)^3}12^{-n} \\
&= \frac{\sqrt{2}}{\sqrt{2\pi n/3}^3}
\end{aligned}$$

which is summable. Hence the walk in $d = 3$ is transient.

For $d \ge 4$, we embed the transient $d = 3$ walk inside the first three coordinates. Let $S_n = (S_n^1, S_n^2, S_n^3, \cdots, S_n^d)$ denote the SRW and set $Y_n = (S_n^1, S_n^2, S_n^3)$. Define $T_0 = 0$ and

$$T_k = \inf\{j > T_{k-1} : Y_j \ne Y_{T_{k-1}}\}$$

Then $\{Y_{T_k} : k \ge 1\}$ is an SRW for $d = 3$. Hence $\mathbb{P}(S_n = 0) \le \mathbb{P}(Y_{T_k} = 0)$ for $T_k \le n < T_{k+1}$ and the result follows.

Recurrent walks visit any point infinitely many times. □

Lecture 19: November 12

$S_n = X_1 + \cdots + X_n$ with $X_i \overset{d}{=} X$ and independent describes a random walk. SRW occurs when $\mathbb{P}(X = \pm 1) = \frac{1}{2}$. We can also consider $\mathbb{P}(X = \pm k) = Ck^{-p}$ for $k \in \mathbb{N}$, where $p > 1$ and C is a normalizing constant. This is a **long range random walk**. Different behavior occurs depending on if the second moment is finite.

$$\mathbb{E}X^2 = \sum 2ck^{2-p}, \qquad \text{summable} \iff p > 3.$$

When $\mathbb{E}X^2 < \infty$, apply CLT to obtain

$$\frac{S_{[nt]}}{\sqrt{n}} \Rightarrow X_t \sim N(0, t\,\mathrm{Var}(X)), \qquad t \in \mathbb{N}.$$

Let $\sigma^2 = \mathrm{Var}(X)$. For $s < t$, $S_{[nt]} - S_{[ns]}$ is independent of $S_{[ns]}$. Hence by the CLT,

$$\frac{S_{[nt]} - S_{[ns]}}{\sqrt{n}} \Rightarrow X_t - X_s \sim N(0, \sigma^2(t-s)).$$

The random variables X_t satisfy the following properties:

(a) $X_0 = 0$

(b) $X_t \sim N(0, \sigma^2 t)$

(c) For all $0 < s < t$, $X_t - X_s$ is independent of X_s and $X_t - X_s \sim N(0, \sigma^2(t-s))$

(d) The map $t \colon X_t$ is continuous

Extending t from $\mathbb{N}$ to $\mathbb{R}^+$ yields the definition of **Brownian Motion**.

Claim 19.1 (Donsker's Invariance Principle, Functional CLT).

$$\{\frac{S_{[nt]}}{\sqrt{n}}, t \geq 0\} \Rightarrow \{X_t, t \in \mathbb{R}^+\}.$$

Let $\alpha = p - 1$ in the long range random walk. Our previous arguments treated the case $\alpha > 2$. Now suppose $0 < \alpha < 2$. Then $\mathbb{E}X_1^2 = \infty$.

Claim 19.2 (Extended CLT). *Suppose $C(n) \to \infty$ such that $S_n / C(n) \Rightarrow Y$. Then Y is of stable distribution and X_1 is the domain of attraction.*

Fact 19.3.

$$S_n / n^{1/\alpha} \Rightarrow Y \sim \textit{ symmetric } \alpha\textit{-stable distribution}$$

$\mathbb{E}e^{-i\gamma Y} = e^{-\lambda|\gamma|^\alpha}$. If brownian motion $\sim \Delta$ then symmetric α-stable distribution $\sim \Delta^{\alpha/2}$.

Lecture 20: November 14

Permutable Events

$(S, \mathcal{S})$ is a measurable space, μ a probability measure, $\Omega = S^{\mathbb{N}}$, $\mathcal{F} = \mathcal{S}^{\mathbb{N}}$, $\mathbb{P} = \mu^{\mathbb{N}}$, $X_i(\omega) = \omega_i$. Then $X_1, \cdots$ are iid with distribution μ.

Definition 20.1. *A* **finite permutation** *of $\mathbb{N}$ is a bijection $\pi : \mathbb{N} \to \mathbb{N}$ such that $\pi(i) \neq i$ for only finitely many i.*

For $\omega \in \Omega$, let $\pi\omega = (\omega_{\pi(1)}, \cdots)$.

Definition 20.2. *An event A is* **permutable** *if $\pi^{-1}(A) = A$ for any finite permutation π.*

Definition 20.3. *The collection of permutable events is a σ-field $\mathcal{E}$, called the* **exchangeable σ-field***.*

Example 20.4 (Examples of permutable events).

(a) $\{\omega : S_n \in A io\} \in \mathcal{E}$. If S is a vector space (or just an abelian group), then $S_n = X_1 + \cdots + X_n$. Not in **tail** σ-field.

(b) $\{\omega : \limsup \frac{S_n}{a_n} \geq b\} \in \mathcal{E}$, where $\{a_n\}$ is any sequence. Recall that if $a_n \to \infty$, this is in the tail σ-field.

(c) $\mathcal{T} \subsetneq \mathcal{E}$, where $\mathcal{T}$ is the tail σ-field.

Theorem 20.5 (Hewitt-Savage 0-1 Law, 3.8). *If $X_1, \cdots$ are iid and $A \in \mathcal{E}$, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.*

Remark 20.6. *$(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ and $X = (X_1, \cdots)$. Then $X : \widetilde{\Omega} \to ((\mathbb{R}^d)^{\mathbb{N}}, (\mathcal{B}(\mathbb{R}^d))^{\mathbb{N}}, \mu^{\mathbb{N}})$, so the pullback σ-field of $\mathcal{E}$ defines exchangeable events on $\widetilde{\Omega}$.*

Idea 20.7. *Suppose π interchanges $\{1, \cdots, n\}$ and $\{n+1, \cdots, 2n\}$. If $A \in \mathcal{E}$ and $A \in \sigma(X_1, \cdots, X_n)$, $\pi A \in \sigma(X_{n+1}, \cdots, X_{2n})$.*

Proof. By measure theory (Thm *A*2.1), for all $A \in \sigma(X_1, \cdots)$ there are $A_n \in \sigma(X_1, \cdots, X_n)$ such that $\mathbb{P}(A \Delta A_n) \to 0$ as $n \to \infty$ (symmetric difference).

Fix n and define the map

$$\pi_n(i) = \begin{cases} n+i & 1 \leq i \leq n \\ i-n & n+1 \leq i \leq 2n \\ i & else \end{cases}$$

Note that π_n is a finite permutation such that $\pi_n^2 = \text{id}$. Since A is permutaable, $\pi_n A = A$. Now $\pi_n A_n \in \sigma(X_{n+1}, \cdots, X_{2n})$ which is independent of A_n since $A_n \in \sigma(X_1, \cdots, X_n)$.

Idea 20.8. *$A_n \to A$ and $\pi_n A_n \to A$.*

Since A_n independent of $\pi_n A_n$, we obtain

$$\mathbb{P}(A_n \Delta A) = \mathbb{P}(\pi(A_n \Delta A)) = \mathbb{P}(\pi_n A_n \Delta A) \qquad (\star)$$

since the X_i are iid.

To make things precise, observe that $|\mathbb{P}(B) - \mathbb{P}(C)| \le \mathbb{P}(B \Delta C)$. Thus by $(\star)$,

$$\mathbb{P}(\pi_n A_n) \to \mathbb{P}(A).$$

Note that $\mathbb{P}(A \Delta C) \le \mathbb{P}(A \Delta B) + \mathbb{P}(B \Delta C)$. This implies

$$\mathbb{P}(A_n \Delta \pi_n A_n) \le \mathbb{P}(A_n \Delta A) + \mathbb{P}(A \Delta \pi_n A_n) \to 0$$

Now by independence, $\mathbb{P}(A_n \cap \pi_n A_n) = \mathbb{P}(A_n)\mathbb{P}(\pi_n A_n)$. We will show that the LHS tends to $\mathbb{P}(A)$ and the RHS tends to $\mathbb{P}(A)^2$.

$$\begin{aligned} 0 &\le \mathbb{P}(A_n) - \mathbb{J}P(A_n \cap \pi_n A_n) \\ &\le \mathbb{P}(A_n \Delta \pi_n A_n) \to 0 \\ \mathbb{P}(A) &= \lim_{n\to\infty} \mathbb{P}(A_n) \\ &= \lim_{n\to\infty} \mathbb{P}(A_n \cap \pi_n A_n) \\ &= \lim_{n\to\infty} \mathbb{P}(A_n)\mathbb{P}(\pi_n A_n) \\ &= \mathbb{P}(A)^2. \end{aligned}$$

Consequenty $\mathbb{P}(A) \in \{0, 1\}$ as desired. □

Theorem 20.9 (3.6). *For any RW on $\mathbb{R}$, only one of the following happens a.s.:*

(a) $S_n = 0$ *for any* $n \ge 1$

(b) $\lim S_n = \infty$

(c) $\lim S_n = -\infty$

(d) $-\infty = \liminf S_n < \limsup S_n = \infty$

Note that we are now working in the extended reals.

Proof. By Hewitt-Savage, $\limsup S_n \in \mathcal{E}$ so $\limsup S_n = c \in [-\infty, \infty]$. Let $\widetilde{S}_n = S_{n+1} - X_1 \overset{d}{=} S_n$. Then $\limsup \widetilde{S}_n = c$ as well, so $c = c - X_1$. If c is finite, then $X_1 = 0$ (case (a)). If $c = -\infty$, this is case (c).

So suppose $c = \infty$. By the same reasoning to $b = \liminf S_n$, if b is finite we're in case (a) and if $b = \infty$ we're in case (c). The only remaining case is $b = -\infty, c = \infty$ which is (d). □

Lecture 21: November 17

In dimension 1, there are four possibilites: it stands still, it runs to infinity (2 directions), or it fluctuates. What about higher dimension?

Definition 21.1. *For $x \in \mathbb{R}^d$ and RW S_n, x is called:*

(a) a **recurrent value** *if* $\forall \varepsilon > 0$, $\mathbb{P}(|S_n - x| < \varepsilon \ io) = 1$

(b) a **possible value** *if* $\forall \varepsilon > 0$, $\exists n \geq 1$ *such that* $\mathbb{P}(|S_n - x| < \varepsilon) > 0$.

Let $\mathcal{V}$ denote the set of recurrent values and $\mathcal{U}$ denote the set of possible values.

Theorem 21.2. *$\mathcal{V}$ is either $\varnothing$ or is a closed subgroup of $\mathbb{R}^d$. In the latter case, $\mathcal{V} = \mathcal{U}$.*

Note that $\{|S_n - x| < \varepsilon \ io\}$ is permutable, so the 0-1 law applies to it. We say that RW is recurrent if $\mathcal{V} \neq \varnothing$.

Theorem 21.3. *RW is recurrent iff $\forall \varepsilon > 0$, $\sum \mathbb{J}P(|S_n| < \varepsilon) = \infty$. This occurs iff it holds for some ε.*

Suppose the step size of SRW is a Gaussian distribution. One way is to choose $\varepsilon = 1$ and compute the sum: not so easy (no combinatorics). For continuous random variable it's unclear (but in the Gaussian case, it's easy since the sum is also Gaussian).

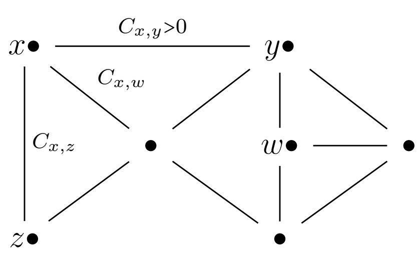
Let $\phi(t) = \mathbb{E}e^{itX}$ be the characteristic function for X, the common distribution of all X_i.

Theorem 21.4. *Let $\delta > 0$. Then S_n is recurrent iff*

$$\int_{[-\delta,\delta]^d} \Re \frac{1}{1 - \phi(t)} \, dt = \infty.$$

Note that the integrand is nonnegative, and for nondegenerate ϕ it is positive (just bound absolute value). When $d = 1$, this was proved by Chung-Fuchs (1951). Extended to higher d by Ornstein and Stone (1969). Random walk on trees or other graphs are more related to markov chains (next quarter).

On $\mathbb{Z}^d$, there is a stationarity property (at least for SRW).



RW on (abelian) groups as well.

Let $T = \inf\{n \geq 1 : S_n = 1\}$. Consider SRW on $\mathbb{R}$. Somewhat perplexingly, $\mathbb{P}(T < \infty) = 1$ yet $\mathbb{E}T = \infty$.

Theorem 21.5 (Wald's Identity). *Let $X_1, \cdots$ iid with $\mathbb{E}|X_1| < \infty$. If T is any stopping time with $\mathbb{E}T < \infty$, then $\mathbb{E}S_T = \mathbb{E}X_1\mathbb{E}T$.*

Notice that if $T = n$ is constant, then $\mathbb{E}S_n = n\mathbb{E}X_1$. So with a leap of faith, we get the identity. Note that things can go wrong, for instance for SRW the left side is 1 and the right is $0 \cdot \infty$.

Proof. Note that part of the proof is that the LHS is well-defined (i.e. S_T is integrable).

$$\begin{aligned}
\mathbb{E}S_T &= \sum_{n=0}^{\infty} \mathbb{E}[S_t; T = n] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[S_n; T = n] \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \mathbb{E}[X_k; T = n] \\
\text{(Fubini)} \qquad &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{E}[X_k; T = n] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[X_k; T \ge k] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[X_k; (T \le k-1)^c] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[X_k]\mathbb{P}(T \ge k) \\
&= \mathbb{E}X_1 \sum_{k=1}^{\infty} \mathbb{P}(T \ge k) \\
&= \mathbb{E}X_1 \mathbb{E}T.
\end{aligned}$$

Now we justify the interchange. Observe that all $|X_i|$ are iid. Set $\widetilde{S}_n = \sum |X_i|$ and apply Tonelli. This dominates the sum so Fubini applies. □

For SRW, how to show $\mathbb{P}(T < \infty) = 1$? Let $a < 0 < b$ be integers. Let $V = \inf\{n \ge 0 \colon S_n \in \{a, b\}\}$. Then $V < \infty$, since SRW oscillates between both infinities. By Wald's identity, $\mathbb{E}S_{V \wedge n} = 0$. Then we pass limits by BCT, so $\mathbb{E}S_V = 0$. But we have

$$\mathbb{E}S_V = a\mathbb{P}(S_V = a) + b\mathbb{P}(S_V = b).$$

Let $T_b = \inf\{n \ge 1 \colon S_n = b\}$. Then we have $a\mathbb{P}(T_a < T_b) + b\mathbb{P}(T_b < T_a) = 0$. These events are complementary, so solving yields

$$\mathbb{P}(T_b < T_a) = \frac{a}{a-b}.$$

Letting $a \to -\infty$, we obtain $\mathbb{P}(T_b < \infty) = 1$.

Lecture 22: November 19

Recall the characteristic function $\phi(y) = \mathbb{E}[e^{i\vec{y}\cdot x}]$

Theorem 22.1. *Then RW S_n is recurrent iff $\exists \delta > 0$*

$$\int_{[-\delta,\delta]^d} \Re \frac{1}{1-\phi(y)}\ dy = \infty \qquad (1).$$

Theorem 22.2 (Chung-Fuchs 1951). *Then S_n is recurrent iff $\exists \delta > 0$ such that*

$$\sup_{r<1} \int_{[-\delta,\delta]^d} \Re \frac{1}{1-r\phi(y)}\ dy = \infty \qquad (2).$$

Note that if $\int_{-[\delta,\delta]^d} \Re \frac{1}{1-\phi(y)}\ dy = \infty$, by Fatou (2) holds so S_n is recurrent. The direction $\to$ was independently proven by D. ornstein and C. Stone in 1969.

Theorem 22.3 (Chung-Fuchs). *For $d = 1$, if $S_n/n \to 0$ in probability, then S_n is recurrent.*

Martingales (Chapter 4)

Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Take X a random variable with $\mathbb{E}|X| < \infty$. Suppose $\mathcal{G} \subset \mathcal{F}$ is a sub σ-field (the information you have). Informallyy, the conditional expectation of X given $\mathcal{G}$, called $\mathbb{E}(X \mid \mathcal{G})$, is the "best possible prediction" of X given the information you have in $\mathcal{G}$.

Certainly when there is a finite second moment, we see that $\mathbb{E}X$ is the least squares estimator. Note that we must assume X is integrable, that is $\mathbb{E}|X| < \infty$.

Definition 22.4. *$\mathbb{E}(X \mid \mathcal{G})$ is any rv $Y \in \mathcal{G}$ such that:*

(a) $\mathbb{E}|Y| < \infty$

(b) $\forall A \in \mathcal{G}$,

$$\int_A Y\ dP = \int_A X\ dP.$$

We must show that Y is unique (a.s.) and that it exists.

Uniqueness Suppose $\int_A (Y_1 - Y_2) = 0$ for all measurable A. Taking $A^+ = \{\omega : Y_1 \geq Y_2\}$. Then the integral condition means $Y_1 = Y_2$ on A^+ a.s. Similarly for A^-. Since $A \cup B$ is the whole space, it follows that $Y_1 = Y_2$ a.e.

Existence Define $\mu(A) = \int_A X\ dP = \mathbb{E}[X 1_A]$. This is a finite signed measure (since $\mathbb{E}|X| < \infty$).

If you haven't seen signed measures before, do the following:

$$\mu^{\pm}(A) = \int_A X^{pm}\ dP, \qquad |\mu|(A) = \int_A |X|\ dP.$$

Then write $\mu = \mu^+ - \mu^-$: this is the definition of a signed measure. Moreover, $|\mu| \ll \mathbb{P}$ (absolute continuity). This means that whenever $\mathbb{P}(A) = 0$, then $|\mu|(A) = 0$.

There are ways to extend, for instance to μ-extended integrable or locally integrable, but we won't get into these technicalities.

Finishing the proof of existence: by Radon-Nikodym, $\exists Y \in \mathcal{G}$ such taht $\mu(A) = \int_A Y \, dP$. Then Y satisfies both conditions in the definition.

Fact 22.5. *(a)* $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}X$

(b) $|\mathbb{E}(X \mid \mathcal{G})| \le \mathbb{E}(|X| \mid \mathcal{G})$

(c) $\mathbb{E}(aX + b \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b$

Example 22.6. *(a) If* $\mathcal{G} = \{\varnothing, \Omega\}$ *then* $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}X$*.*

(b) If $\mathcal{G} = \{\varnothing, A, A^c, \Omega\}$ *where* $0 < \mathbb{P}(A) < 1$*, then*

$$Y = \mathbb{E}(X \mid \mathcal{G}) = \frac{\int_A X \, dP}{\mathbb{P}(A)} 1_{\{A\}} + \frac{\int_{A^c} X \, dP}{\mathbb{P}(A^c)} 1_{\{A^c\}}.$$

(c) If $A_1, \cdots, A_n$ *are disjoints sets in* $\mathcal{G}$ *with* $\bigcup A_k = \Omega$*,* $\mathbb{P}(A_k) > 0$*, and* $\mathcal{G} = \sigma(A_1, \cdots, A_n)$ *then*

$$\mathbb{E}(X \mid \mathcal{G}) = \sum_{k=1}^{n} \frac{\mathbb{E}(X 1_{\{A_k\}})}{\mathbb{P}(A_k)} 1_{\{A_k\}}.$$

(d) If (ξ, η) *has a joint density function* $f(x, y)$ *and* $f_Y(y) = \int f(x, y) \, dx$ *then what is*

$$\mathbb{E}[\phi(\xi, \eta) \mid \sigma(\eta)] = g(\eta)?$$

Lecture 23: November 21

Conditional Expectation 4.1

Consider a r.v. X on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}|X| < \infty$. For $\mathcal{G} \subset \mathcal{F}$, we define $\mathbb{E}(X \mid \mathcal{G}) \in \mathcal{G}$ as follows:

$$\int_A \mathbb{E}(X \mid \mathcal{G})\ dP = \int_A X\ dP \qquad (\star)$$

Let $Y = \mathbb{E}(X \mid \mathcal{G})$. Taking $A = \{Y \geq 0\}$ and $A = \{Y < 0\}$ in turn yields $\mathbb{E}|Y| \leq \mathbb{E}|X|$ so Y is integrable. Moreover note that $(\star)$ is equivalent to requiring $\mathbb{E}[X\xi] = \mathbb{E}[Y\xi]$ for all bounded $\xi \in \mathcal{G}$.

Proof. Let $\xi_n = \sum_{k\in\mathbb{Z}} \frac{k}{2^n} 1_{\{\frac{k}{2^n} \leq \xi < \frac{k+1}{2^n}\}}$. Then the sum for ξ_n is finite (by boundedness) and $|\xi_n - \xi| \leq 2^{-n}$. □

Example 23.1. *If X, Y have joint density function $f(x,y)$, let $f_Y(y) = \int f(x,y)\ dx$. Then $\mathbb{E}[g(X) \mid \sigma(Y)] = h(y)$, but what is h?*

To find $h(Y)$, observe that for all $\Phi(Y)$ bounded, we have

$$\begin{aligned}
\mathbb{E}[h(Y)\phi(Y)] &= \mathbb{E}[g(X)\phi(Y)] \\
&= \int g(x)\phi(y)f(x,y)\ dxdy \\
&= \int \phi(y) \int g(x)f(x,y)\ dxdy \\
\implies \int \phi(y)h(y)f_Y(y)\ dy &= \int \phi(y) \int g(x)f(x,y)\ dxdy.
\end{aligned}$$

Thus after some approximation arguments we obtain

$$h(Y) = \frac{\int g(x)f(x,y)\ dx}{f_Y(y)}, \qquad f_Y(y) \neq 0.$$

Note that h is not unique, but $h(Y)$ is unique. This is because h is only determined on $\operatorname{Im} Y = \{y : f_Y(y) \neq 0\}$.

Formally speaking, we have $\mathbb{E}[g(X) \mid Y = y] = h(y)$. But we can write

$$\frac{\mathbb{E}[g(X); Y = y]}{\mathbb{P}(Y = y)} = \frac{\int g(x)f(x,y)\ dx\Delta y}{f_Y(y)\Delta y} = \frac{\int g(x)f(x,y)\ dx}{f_Y(y)},$$

giving intuition for the formula.

The following notation will be used:

$$\mathbb{E}[\cdots \mid X] \text{ means } \mathbb{E}[\cdots \mid \sigma(X)].$$

Example 23.2. *Suppose X, Y are independent such that $\mathbb{E}|\phi(X,Y)| < \infty$. Then $\mathbb{E}[\phi(X,Y) \mid X] = h(X)$. What is h?*

Proof. For all bounded $\xi = g(X) \in \sigma(X)$, we have

$$\begin{aligned}\mathbb{E}[h(X)g(X)] &= \mathbb{E}[\phi(X,Y)g(X)] \\ &= \int \phi(x,y)g(x)\mu_X(dx)\mu_Y(dy).\end{aligned}$$

Here we define

$$\int f(x)d\mu = \int f(x)\mu(dx).$$

□

Instead of $\mu(A)$, we put in an "infinitesimal set": $\mu(A) \to \mu(dx)$. This is heuristic. Continuiing,

$$\begin{aligned}\int g(x)h(x)\mu_X(dx) &= \int g(x)\int \phi(x,y)\mu_Y(dy)\mu_X(dx) \\ \implies h(x) &= \int \Phi(x,y)\mu_Y(dy) \\ &= \mathbb{E}[\Phi(x,Y)].\end{aligned}$$

So our formula says that to compute $\mathbb{E}[\phi(X,Y) \mid X]$, you simply fix x then do your calculation.

$$\mathbb{E}[\phi(X,Y) \mid X] = (\mathbb{E}[\phi(x,Y)])_{X=x}.$$

Theorem 23.3 (4.1). *Suppose* $\mathbb{E}[X^2] < \infty$. *Then* $\mathbb{E}[X \mid \mathcal{G}]$ *is the rv in* $\mathcal{G}$ *that minimizes the "mean square error"* $\mathbb{E}[(X-\xi)^2]$ *for* $\xi \in \mathcal{G}$.

That is to say,

$$\mathbb{E}[(X - \mathbb{E}(X \mid \mathcal{G}))^2] = \min_{\xi \in \mathcal{L}^2(\mathcal{G})} \mathbb{E}[(X-\xi)^2].$$

This says that $\mathbb{E}(X \mid \mathcal{G})$ is the orthogonal projection of X into $\mathcal{L}^2(\mathcal{G})$.

Proof. Take $\mathcal{H} = \mathcal{L}^2(\mathcal{F})$ and $\mathcal{K} = \mathcal{L}^2(\mathcal{G})$, a closed subspace of $\mathcal{H}$. By Hilbert space theory, it suffices to show that $\eta = X - \mathbb{E}(X \mid \mathcal{G}) \perp \mathcal{L}^2(\mathcal{G})$. For notational convenience let $Y = \mathbb{E}(X \mid \mathcal{G})$ and $\eta = X - Y$. Then

$$\begin{aligned}\mathbb{E}(X-\xi)^2 &= \mathbb{E}[\eta + (Y-\xi)]^2 \\ &= \mathbb{E}\eta^2 + 2\mathbb{E}\eta(Y-\xi)] + \mathbb{E}(Y-\xi)^2\end{aligned}$$

Now observe that $Y - \xi \in \mathcal{G}$. From before, it follows that $\mathbb{E}[(X-Y)\xi] = 0$ when ξ is bounded in $\mathcal{G}$. We needed boundedness in order to pass the limit from simple functions. Here we know that X has finite second moment, so another tool is available: Cauchy-Schwarz. Thus it suffices to show that $Y \in \mathcal{L}^2(\mathcal{G})$. This is an exercise he leaves to us.

Consequently $\mathbb{E}(X-\xi)^2 = \mathbb{E}\eta^2 + \mathbb{E}(Y-\xi)^2$ and the result follows. □

Another way to proceed is via the tower identity:

$$\mathbb{E}[(X-Y)(Y-\xi)] = \mathbb{E}[\mathbb{E}[(X-Y)(Y-\xi) \mid \mathcal{G}]] = \mathbb{E}[(Y-\xi)\mathbb{E}[(X-Y) \mid \mathcal{G}]]$$

Lecture 24: November 24

Theorem 24.1 (4.1). *If $\mathbb{E}(X^2) < \infty$ and $\mathcal{G} \subset \mathcal{F}$, then $Y = \mathbb{E}(X \mid \mathcal{G})$ is the r.v. in $\mathcal{G}$ which minimizes $\mathbb{E}(X - \xi)^2$.*

Remark 24.2. *(a) If X is bounded, then $Y = \mathbb{E}(X \mid \mathcal{G})$ is bounded and the result follows. Indeed, if $X \geq 0$ then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$.*

Proof. Let $Y = \mathbb{E}(X \mid \mathcal{G})$. Then $\int_A Y \, dP = \int_A X \, dP \geq 0$. Choose $A = \{Y < 0\}$ and observe $\mathbb{P}(A) = 0$. Hence $Y \geq 0$ a.e.

Now if $m < X < M$ where m, M are constant, it follows that $m < \mathbb{E}(X \mid \mathcal{G}) < M$ (since the expectation of a constant is itself). □

(b) Assuming L^2 knowledge, we can proceed. Let Y be the orthogonal projection of X onto $\mathcal{L}^2(\mathcal{G})$. Check that $\mathbb{E}(X\mathcal{G}) = Y$. Indeed, for all $\in \mathcal{G}$ we have $1_A \in \mathcal{L}^2(\mathcal{G})$ (because $\mathbb{P}(A) < \infty$)). Hence $X - Y \perp 1_A$, so $\int_A X \, dP = \int_A Y \, dP$ so $Y = \mathbb{E}(X \mid \mathcal{G})$.

Remark 24.3. *We can extend the theory beyond the L^1 case by taking limits of cutoffs: $X_n = X \wedge n$.*

We can prove the result without relying on L^2 theory.

Lemma 24.4 (4.2). *Let X, Y be r.v. with $\mathbb{E}|X| + \mathbb{E}|Y| < \infty$. Then*

(a) $\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(X \mid \mathcal{G})$, for any rv $a, b \in \mathcal{G}$

(b) If $X \geq 0$ then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$. Consequently if $X \geq Y$ then $\mathbb{E}(X \mid \mathcal{G}) \geq \mathbb{E}(Y \mid \mathcal{G})$.

(c) If $X_n \geq 0$ and $X_n \nearrow X$, then $\mathbb{E}(X_n \mid \mathcal{G}) \nearrow \mathbb{E}(X \mid \mathcal{G})$.

Jensen If ϕ is convex, $\mathbb{E}|X| < \infty, \mathbb{E}|\phi(X)| < \infty$ then $\phi(\mathbb{E}(X \mid \mathcal{G})) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}]$.

We already proved (a) and (b). For (c), let $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$. By (b) Y_n is monotone, so we may define $Y = \sup_n Y_n = \lim_n Y_n$. Then since $\int_A X_n \, dP = \int_A Y_n \, dP$, by monotone convergence we have $\int_A X = \int_A Y$ and therefore $Y = \mathbb{E}(X \mid \mathcal{G})$ and the result follows. (Everything is a.e.)

Proof of Jensen's Inequality. It is known that convex functions are the envelope of the lines beneath them:

$$\phi(x) = \sup\{ax + b : (a, b) \in S\}, \qquad S = \{(a, b) \in \mathbb{Q}^2 : ax + b \leq \phi(x) \forall x\}.$$

Then for $(a, b) \in S$, we have $\phi(X) \geq aX + b$. Consequently $\mathbb{E}(\phi(X) \mid \mathcal{G}) \geq a\mathbb{E}(X \mid \mathcal{G}) + b$ (by the first two properties). Consequently

$$\mathbb{E}(\phi(X) \mid \mathcal{G}) \leq \sup_{(a,b) \in S} (a\mathbb{E}(X \mid \mathcal{G}) + b) = \phi(\mathbb{E}(X \mid \mathcal{G})).$$

□

Note that we needed S to be countable in the supremum because the inequalities hold a.s., and thus when we take supremum we must ensure that only countable many operations are performed to preserve measure 0.

In particular when $\phi \geq 0$, this implies

$$\mathbb{E}[\phi(E(X \mid \mathcal{G}))] \leq \mathbb{E}[\phi(X)].$$

Example 24.5. *Take $\phi(x) = |x|^p$. For $p \geq 1$ this is a convex function. Thus we have*

$$\mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})|^p) \leq \mathbb{E}|X|^p.$$

When $p = 2$ it yields our earlier claim that $\mathbb{E}(X \mid \mathcal{G})$ has finite second moment.

Theorem 24.6 (Tower property, 4.3). *If $\mathcal{F}_1 \subset \mathcal{F}_2$ and $\mathbb{E}|X| < \infty$, then*

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_1) \mid \mathcal{F}_2) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = \mathbb{E}(X \mid \mathcal{F}_1).$$

Mnemonically, the "smaller one wins". It is clear that the leftmost and rightmost agree, since $\mathcal{F}_2$ has more information than $\mathcal{F}_1$. If the rv is square integrable, we can project twice in a Hilbert space to get the result immediately.

General case. For all $A \in \mathcal{F}_1$, we have

$$\begin{aligned}\int_A \mathbb{E}(X \mid \mathcal{F}_2)\, dP &= \int_A X\, dP \\ &= \int_A \mathbb{E}(X \mid \mathcal{F}_1)\, dP.\end{aligned}$$

Hence $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = \mathbb{E}(X \mid \mathcal{F}_1)$. □

Imagine you have a rv $X \in \mathcal{L}^2(\mathcal{P})$ and you have an increasing sequence of σ-fields $\{\mathcal{F}_n\}$. Let $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$. Then by Theorem 4.3, we have

$$X_n = \mathbb{E}(X_k \mid \mathcal{F}_n), \qquad \forall n < k.$$

Martingales

Definition 24.7. *Let $\{\mathcal{F}_n\}$ be a filtration. A sequence of rv $\{X_n\}$ is* **adapted** *to $\{\mathcal{F}_n\}$ if $X_n \in \mathcal{F}_n$ for any $n \geq 1$.*

Definition 24.8. *If $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$ with*

(a) $\mathbb{E}|X_n| < \infty$ *for* $n \geq 1$

(b) $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ *for any* $n \geq 1$

we say X_n is a martingale.

Replacing the = with $\leq, \geq$ makes is a super/sub martingale, and then $\mathbb{E}X_n$ is decreasing/increasing. So one can think of martingales as stochastic versions of constant functions. SRW is a martingale.

Lecture 25: November 26

Definition 25.1 (Martingale). *$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}X_n$ for any $n \geq 0$.*

Think of X_n as the wealth you have at time n. A supermartingale occurs when $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n$ annd the other inequality is a submartingale.

The martingale property is equivalent to $\mathbb{E}(X_m \mid \mathcal{F}_n) = X_n$ for any $m > n$. For this equivalence, it suffices to observe that

$$\mathbb{E}(X_{n+2} \mid \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_{n+2} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} \mid \mathcal{F}_n).$$

For super- and submartingales the same result holds, but now by monotonicity of expectation.

Clearly:

(a) $\{X_n\}$ is a supermartingale iff $\{-X_n\}$ is a submartingale

(b) $\{X_n\}$ is a martingale iff it is both a super- and submartingale

When does a monotone sequence have a limit? Whenever it is bounded. How about for submartingales? Whenever $\mathbb{E}|X_n|$ is bounded.

If $\{x_n\}$ is a deterministic sequence then $\lim x_n$ exists iff $\liminf x_n = \limsup x_n$. Then we can find a, b such that

$$\liminf_n x_n \leq a < b \leq \limsup_n x_n.$$

Then x_n crosses the interval $[a, b]$ infinitely many times. Let $U(a, b)$ denote the upcrossings of $[a, b]$. Then $\lim_n x_n$ exists iff $U(a, b) < \infty$ for any $a < b$. At the end we will consider $\mathbb{E}U(a, b)$ and show it is finite to prove the result.

Theorem 25.2. *If $\{X_n\}$ is a martingale, ϕ is convex, and $\mathbb{E}|\phi(X_n)| < \infty$ for each n, then $\phi(X_n)$ is a submartingale.*

Proof. We know that $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$. Consequently

$$\begin{aligned} \phi(X_n) &= \phi(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \\ \text{(Jensen)} \qquad &\leq \mathbb{E}[\phi(X_{n+1}) \mid \mathcal{F}_n]. \end{aligned}$$

Theorem 25.3 (4.4). *If X_n is a submartingale, ϕ is convex and increasing, and $\mathbb{E}|\phi(X_n)| < \infty$ for any n, then $\phi(X_n)$ is a submartingale.*

Proof. Same as before, except $\phi(X_n) \leq \phi(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$. □

Corollary 25.4 (4.5). *(a) If X_n is a submartingale then so is $(X_n - a)^+$ and $X_n \vee a$.*

(b) If X_n is a supermartingale, then so is $X_n \wedge a$.

Proof. Just look at functions like $(x - a)^+$ and $x \vee a$. □

Definition 25.5. *A sequence of rv $\{H_n, n \geq\}$ is said to be predictable w.r.t. $\{\mathcal{F}_n\}$ if $H_n \in \mathcal{F}_{n-1}$.*

Theorem 25.6 (4.6). *Let $\{X_n\}$ be a supermartingale. If $H_n \geq 0$ are predictable (wrt the filtration used for $\{X_n\}$) and each H_n is bounded, then the stochastic integral $H.X$ is a supermartingale.*

Definition 25.7. *The* **stochastic integral** *$H.X$ is defined as follows:*

$$(H.X)_n = \begin{cases} 0 & n = 0 \\ \sum_{k=1}^n H_k(X_k - X_{k-1}) & n \geq 1 \end{cases}$$

Proof. Clearly $(H.X)_n \in \mathcal{F}_n$, and by the triangle inequality $(H.X)_n$ is integrable (by boundedness of H_k). Now we compute

$$\begin{aligned} \mathbb{E}((H.X)_{n+1} \mid \mathcal{F}_n) - (H.X)_n &= \mathbb{E}((H.X)_{n+1} - (H.X)_n \mid \mathcal{F}_n) \\ &= \mathbb{E}(H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n) \\ &= H_{n+1}\mathbb{E}(X_{n+1} - X_n \mid \mathcal{F}_n)) \\ &\leq 0. \end{aligned}$$

Remark 25.8. *If X_n is a martingale, H_n is predictable and bounded. Then $H.X$ is a martingale.*

Corollary 25.9 (4.7). *If T is a stopping time for $\{\mathcal{F}_n\}$ and X_n is a supermartingale (martingale) thne $\{X_{T\wedge n}\}$ is a supermartingale (martingale).*

Proof. We compute

$$\begin{aligned} X_{T\wedge n} &= \sum_{k=1}^n (X_{T\wedge k} - X_{T\wedge(k-1)}) + X_0 \\ &= \sum_{k=1}^n 1_{\{T\geq k\}}(X_k - X_{k-1}) + X_0 \\ (H_k = 1_{\{T\geq k\}} \in \mathcal{F}_{k-1}) \qquad &= (H.X)_n + X_0. \end{aligned}$$

Lecture 26: December 1

Recall 26.1. *If $\{X_n\}$ is a supmartingale and T is a stopping time, then so is $\{X_{n\wedge T}\}$.*

Suppose $\{X_n, n \geq 0\}$ is a submartingale. For $a < b$, define

$$T_1 = \inf\{k \leq 0 \mid X_k \leq a\}, \qquad T_2 = \inf\{k \geq T_1 \mid X_k \geq b\}$$

and recursively let

$$T_{2k+1} = \inf\{j \geq T_{2k} \mid X_j \leq a\} \qquad T_{2k+2} = \inf\{j \geq T_{2k+1} \mid X_j \geq b\}$$

It is clear that T_1 is a stopping time w.r.t. the filtration $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. Then by induction, it follows that all T_k are stopping times.

Now observe $\{T_{2k-1<m\leq T_{2k}\}=\{T_{2k-1}\leq m-1}\} \cap \{T_{2k} \leq m-1\}^c \in \mathcal{F}_{m-1}$. Consequently $1_{\{T_{2k-1}<m\leq T_{2k}\}}$ is predictable. Since the predictable processes are closed under addition, it follows that

$$H_m = \sum_{k=1}^{\infty} 1_{\{T_{2k-1}<m\leq T_{2k}\}} \text{ is predictable.}$$

Let $U_n = \sup\{k \mid T_{2k} \leq n\}$. This is the number of upcrossings of $[a,b]$ at time n.

Theorem 26.2 (Upcrossing inequality). *If $\{X_k \mid k \geq 0\}$ is a submartingale, then*

$$\mathbb{E}U_n \leq \frac{\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+}{b-a}$$

Note that this is the first place we use martingale properties in the calculation.

Proof. Let $Y_n = a + (X_n - a)^+$. You just flatten the process when it dips below a. Then $\{Y_n \mid n \geq 0\}$ has the same number of upcrossings as $\{X_n \mid n \geq 0\}$. Moreover all the T_k values are the same for X_n and Y_n, since nothing has been really changed.

Claim 26.3. $(b-a)U_n \leq (H.Y)_n$

Proof. By definition, $(H.Y)_n = \sum_{m=1}^n H_m(Y_m - Y_{m-1})$. Now observe H_m vanishes unless you're in a region between crossings. Each time we make a positive crossing, we gain at least $b-a$. Therefore

$$\begin{aligned}(H.Y)_n &= \sum_{m=1}^{T_{2U_n}} H_m(Y_m - Y_{m-1}) + \sum_{m=T_{2U_n}+1}^{n} H_m(Y_m - Y_{m-1}) \\ &= \sum_{k=1}^{U_n}(Y_{T_{2k}} - Y_{T_{2k-1}}) + 1_{\{T_{2U_n}+1<n<T_{2U_n+2}\}}.(Y_n - Y_{2U_n+1}) \qquad \geq (b-a)U_n.\end{aligned}$$

Here the second term is non-negative because $Y_n \geq a$ and $Y_{T_{2U_n+1}} = a$. □

At last, we use the submartingale property. Note that as $H_m \geq 0$, from last time it follows that $(H.Y)_n$ is a submartingale (Thm 4.6). Actually, we really use that since $H_m \leq 1$, it follows that $1 - H_m \geq 0$ and therefore $1 - H_m$ is a submartingale (since 1 is constant it is predictable, and a difference of predictable variables is still predictable). Taking expected values yields

$$\begin{aligned}(b-a)\mathbb{E}U_n &\leq \mathbb{E}(H.Y)_n \\ &= \mathbb{E}[(1.Y)_n - ((1-H).Y)_n] \\ &= \mathbb{E}(Y_n - Y_0) - \mathbb{E}[((1-H).Y)_n] \\ &\leq \mathbb{E}Y_n - \mathbb{E}Y_0 \\ &= \mathbb{E}[(X_n - a)^+ - (X_0 - a)^+].\end{aligned}$$

Theorem 26.4 (Martingale convergence theorem, 4.9). *If $\{X_n \mid n \geq 0\}$ is a submartingale with $\sup_n \mathbb{E}X_n^+ < \infty$, then $X_n \to X$ a.s. for some rv X with $\mathbb{E}|X| < \infty$.*

We need the stochastic version of bounded from above with positive part, just consider SRW to see what goes wrong otherwise.

Remark 26.5. *For a submartingale $\{X_n\}$ the condition $\sup_n \mathbb{E}X_n^+ < \infty$ is equivalent to $\sup_n \mathbb{E}|X_n| < \infty$. Indeed, we have $|X_n| = 2X_n^+ - X_n$, so*

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_n^+ - \mathbb{E}X_0.$$

Recall that in any martingale, the variables are integrable by definition.

Proof of Thm 4.9. For any $a < b$, let $U_n(a,b)$ denote the upcrossings of $[a,b]$ by time n, and let $U(a,b)$ denote the total upcrossings. Clearly $U_n(a,b) \nearrow U(a,b)$. Therefore

$$\begin{aligned}
\mathbb{E}U(a,b) = \lim_{n\to\infty} \mathbb{E}U_n(a,b) \\
&\leq \liminf_{n\to\infty} \frac{\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+}{b-a} \\
&\leq \liminf_{n\to\infty} \frac{\mathbb{E}(X_n - a)^+}{b-a} \\
&\leq \liminf_{n} \frac{\mathbb{E}(X_n^+) + a^+)}{b-a} \\
&\leq \frac{\sup \mathbb{E}X_n^+ + a^+)}{b-a} \\
&< \infty.
\end{aligned}$$

Hence $U(a,b) < \infty$ a.s., which means $\mathbb{P}(U(a,b) < \infty \forall a,b \in \mathbb{Q}) = 1$. Thus we have shown that for all $\omega \in \Omega_0$, we have $\lim_{n\to\infty} X_n(\omega)$ exists.

It might go to $\pm\infty$, but we rule that out as follows: first $|X_n| to |X|$ a.s., and by Fatou we have

$$\mathbb{E}[|X|] \leq \liminf_n \mathbb{E}|X_n| \leq \sup_n \mathbb{E}|X_n| < \infty.$$

□

Lecture 27: December 3

Recall 27.1. *If $\{X_n\}$ is a submartingale with $\sup \mathbb{E}X_n^+ < \infty$ (equivalent to $\sup \mathbb{E}|X_n| < \infty$) then $X_n \to X$ a.s. with $\mathbb{E}|X| < \infty$.*

Multiplying by -1 yields a corresponding theorem for supmartingales.

Corollary 27.2 (4.10). *If $X_n \geq 0$ is a supmartingale, then $X_n \to X$ a.s. with $\mathbb{E}|X| < \infty$.*

This follows since $\mathbb{E}X_n$ is decreasing for supmartingales, and $|X_n| = X_n$ in this case. So the sup is bounded by $\mathbb{E}|X_1| < \infty$.

Remark 27.3. *We cannot replace a.s. convergence by L^1 convergence.*

Example 27.4. *Stopped SRW on $\mathbb{Z}$*

Let S_n be SRW on $\mathbb{Z}$ starting from 1 (that is, $S_0 = 1$). Let $T = \inf\{n \mid S_n = 0\}$. This is a stopping time and $T < \infty$ (since SRW in $d = 1$ is recurrent). We showed that truncation of a martingale by a stopping time is still a martingale. Thus set $X_n = S_{n \wedge T}$. Note that $X_n \geq 0$ since we always stop at 0. Thus by the corollary it converges a.s., and we identify the limit as $X_T = 0$. We could easily show this without the theorem, since for n sufficiently large the path is frozen at 0.

On the other hand, the convergence cannot occur in L^1 (since $\mathbb{E}X_n = \mathbb{E}X_0 = 1$). So we need more conditions to ensure stronger convergence.

Definition 27.5. *A family of rv $\{X_\alpha \mid \alpha \in \Lambda\}$ is said to be* **uniformly integrable** *if*

$$\lim_{M \to \infty} \sup_{\alpha \in \Lambda} \mathbb{E}[|X_\alpha| 1_{\{|X_\alpha| > M\}}] = 0$$

Remark 27.6. *If $\{X_\alpha \mid \alpha \in \Lambda\}$ is uniformly integrable, then $\sup_{\alpha \in \Lambda} \mathbb{E}|X_\alpha| < \infty$*

Theorem 27.7 (4.11). *$X_n \to X$ in L^1 iff $X_n \to X$ in probability and $\{X_n\}$ is uniformly integrable.*

As an application, if $\{X_n\}$ is a uniformly integrable martingale then $X_n \to X$ a.s. and in L^1. Moreover $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$. We say that the martingale is **closed** when this occurs (combination of Thm 4.9 and 4.10). Partial explanation: for any $n, k \geq 1$ we have $X_n = \mathbb{E}(X_{n+m} \mid \mathcal{F})$ (iterating the definition of a martingale). Letting $k \to \infty$ yields the result. This is a statement of continuity of martingales in L^1.

A martingale $\{X_n\}$ is uniformly integrable iff there exists $X \in L^1$ such that $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$. By the tower property this is a martingale. To check uniform integrability, just check the definition.

Theorem 27.8 (4.12). *Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $\xi \in L^1(\mathbb{P})$, then $\{\mathbb{E}(\xi \mid \mathcal{G}) \mid \mathcal{G} \subset \mathcal{F}\}$ is uniformly integrable.*

Backward Martingale (4.3)

Definition 27.9. *A* **backward martingale** *is a martingale indexed by negative integers.*

We have $\{X_n \mid n \le 0\}$ adapted to a filtration $\mathcal{F}_n$ with $\mathbb{E}|X_n| < \infty$ and $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ for any $n \le -1$.

Theorem 27.10. *Let $\{X_n \mid n \le 0\}$ be a backward martingale. Then $X_n \to X$ a.s. and in L^1 as $n \to \infty$.*

Proof. We have $X_{-n} = \mathbb{E}(X_0 \mid \mathcal{F}_{-n})$ for all $n \ge 0$. Let $U_n(a,b)$ be the number of upcrossings of (a,b) by $X_{-n}, \cdots, X_0$. Let $U(a,b)$ be the total number of upcrossings. Notice that $U_n(a,b) \nearrow U(a,b)$. Moreover

$$\begin{aligned}\mathbb{E}U_n(a,b) &\le \frac{1}{b-a}\mathbb{E}((X_0-a)^+)\\ &\le \frac{\mathbb{E}|X_0|+|a|}{b-a}\\ \mathbb{E}U(a,b) &\le \frac{\mathbb{E}|X_0|+|a|}{b-a} < \infty.\end{aligned}$$

Hence $\lim_{n\to\infty} X_{-n}$ exists a.s. and in L^1. □

Application to the strong law of large numbers:

Let $\xi_1, \cdots$ be iid rv with $\mathbb{E}|\xi_1| < \infty$. Let $S_n = \sum_{i=1}^n \xi_k$. Then $\frac{S_n}{n} \to \mathbb{E}\xi_1$ a.s. and in L^1.

Proof. Define $X_{-n} = \frac{S_n}{n}$. Set $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \cdots) = \sigma(S_n, \xi_{n+1}, \cdots)$. We claim that $(X_{-n}, \mathcal{F}_{-n})$ is a backward martingale. Indeed, we compute

$$\begin{aligned}\mathbb{E}(X_{-n} \mid \mathcal{F}_{-n-1}) &= \mathbb{E}[\frac{S_n}{n} \mid \mathcal{F}_{-n-1}]\\ &= \mathbb{E}[\frac{S_{n+1}-\xi_{n+1}}{n} \mid \mathcal{F}_{-n-1}]\\ &= \frac{S_{n+1} - \mathbb{E}(\xi_{n+1} \mid \mathcal{F}_{-n-1})}{n}.\end{aligned}$$

Now observe that $\mathbb{E}(\xi_i \mid \mathcal{F}_{-n-1}) = \mathbb{E}(\xi_j \mid \mathcal{F}_{-n-1})$ for $i, j \le n+1$, since if you know the summation all of its members were equally likely. Continuing,

$$\begin{aligned}\mathbb{E}(X_{-n} \mid \mathcal{F}_{-n-1}) &= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)}\\ &= \frac{S_{n+1}}{n+1} = X_{-n-1}.\end{aligned}$$

Therefore X_{-n} converges to $X \in \bigcap_n \mathcal{F}_{-n}$ a.s. and in L^1. Now $\bigcap_n \mathcal{F}_{-n} \subset \mathcal{E}$, the exchangeable σ-field. Now by Hewitt-Savage 0-1 law, it follows that X is the constant $\mathbb{E}X_1$. □

Chapter 2

Winter 2015

Lecture 1: January 7

Recall the last problem from the final. We are showing the a.s. convergence. Since the event is exchangeable, by Hewitt-Savage it suffices to show that $\mathbb{P}(V = \infty) > 0$. Let X_i be iid, $\mathbb{E}X_i = 0$ and $\sigma^2 = \mathbb{E}X_i^2 \in \mathbb{R}^+$. Set $S_n = \sum_{k=1}^n X_n$ and $V = \sum_{n=1}^\infty 1_{\{S_{n^2}=0\}}$.

Claim 1.1. *$V = \infty$ a.s.*

Proof. Set $V_n = \sum_{k\ge n} 1_{\{S_{k^2}=0\}}$ and $V_n^M = \sum_{k=n}^M 1_{\{S_{k^2}=0\}}$. We have

$$\{V = \infty\} = \bigcap_{n=1}^\infty \{V_n \ge 1\} = \bigcap_{n=1}^\infty \{V_n > 0\}.$$

Therefore

$$\mathbb{P}(V_n^M \ge 1) = \mathbb{P}(V_n^M > 0) \ge \frac{(\mathbb{E}V_n^M)^2}{\mathbb{E}(V_n^M)^2}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}V_n^M &= \sum_{k=n}^M \mathbb{P}(S_{k^2} = 0) \\
\mathbb{P}(S_{k^2} = 0) &= \mathbb{P}\left(\frac{|S_{k^2}|}{k} < \frac{1}{k}\right) \sim \frac{c}{k} \\
\implies \mathbb{E}V_n^M &\approx \sum_{k=n}^M \frac{c}{k} \\
&\approx c(\log M - \log n) \\
\implies \mathbb{E}(V_n^M)^2 &= \mathbb{E}V_n^M + 2\sum_{n\le k<j\le M} \mathbb{P}(S_{k^2} = 0, S_{j^2} = 0) \\
&= \mathbb{E}(V_n^M) + 2\sum \mathbb{P}(S_{k^2} = 0)\mathbb{P}(S_{k^2} = 0)\mathbb{P}(S_{j^2-k^2} = 0) \\
&\approx c(\log M - \log n) + 2\sum_{k=n}^M \sum_{j=k+1}^M \frac{c}{k}\frac{c}{\sqrt{j^2-k^2}} \\
&\le c(\log M - \log n) + c' \sum_{k=n}^M \frac{1}{k}\sum_{k+1}^M \frac{1}{j-k} \\
&\le (\log M - \log n)(c + c' \log M) \\
\implies \lim_{M\to\infty} \mathbb{P}(V_n^M \ge 1) &\ge \lim_{M\to\infty} \frac{c^2(\log M - \log n)^2}{(\log M - \log n)(c + c'\log M)} \\
&\ge \frac{c^2}{c_1} > 0.
\end{aligned}$$

Thus $\mathbb{P}(V = \infty) > 0$, whereupon $\mathbb{P}(V = \infty) = 1$ as desired. □

Definition 1.2. *A family of rv $\{X_k\}_{k\in I}$ is* **uniformly integrable** *if*

$$\lim_{M\to\infty} \sup_{k\in I} \int_{|X_k|>M} |X_k| \, dP = 0 \qquad (1)$$

Note that I need not be countable.

Remark 1.3 (u.i. $\Longrightarrow \mathcal{L}^1$ bounded). *If $\{X_k\}$ is u.i., then* $\sup_{k\in I}\mathbb{E}|X_k|<\infty$.

Indeed, there is some M for which $\sup_{k\in I}\int_{|X_k|>M}|X_k|\,dP<1$. Truncating yields

$$\begin{aligned}\mathbb{E}|X_k| &= \mathbb{E}\left[|X_k|1_{\{|X_k|\le M\}}+|X_k|1_{\{|X_k|>M\}}\right]\\ &\le M+1.\end{aligned}$$

Note that the property of being **tight** is a more general (weaker) notion than u.i. (take $X_k=1$).

Remark 1.4. (1) *always holds for a finite family of integrable rv (homework).*

Theorem 1.5 (4.11). *$X_n\to X$ in $\mathcal{L}^1$ iff $\{X_n\}$ is u.i. and $X_n\to X$ in probability.*

Proof. $\Rightarrow$ By Chebyshev, $\mathbb{P}(|X_n-X|>\varepsilon)\le\frac{\mathbb{E}|X_n-X|}{\varepsilon}\to 0$, so $X_n\to X$ in probability. Applying Fatou, it follows that for all $M>0$

$$\varliminf_{n\to\infty}\mathbb{E}|X_n|1_{\{|X_1|\le M\}}\ge\mathbb{E}|X|1_{\{|X|\le M\}}.$$

Since $\mathbb{E}|X_n|\to\mathbb{E}|X|$,

$$\varlimsup_n\mathbb{E}|X_n|1_{\{|X_n|>M\}}\le\mathbb{E}|X|1_{\{|X|>M\}}$$

Hence $\{X_n\}$ is u.i.

$\Leftarrow$ For all $M>0$, let $\phi(x)=(-M)\vee x\wedge M$. Note $|\phi(x)-x|=(|x|-M)1_{\{|X|>M\}}$. Then we have

$$\begin{aligned}X_n-X &\le |X_n-\phi(X_n)|+|\phi(X_n)-\phi(X)|+|\phi(X)-X|\\ \mathbb{E}|X_n-X| &\le \mathbb{E}(|X_n|-M;|X_n|>M)+\mathbb{E}|\phi(X_n)-\phi(X)|+\mathbb{E}(|X|-M;|X|>M)\end{aligned}$$

By u.i., $\forall\varepsilon$ there exists M such that $\sup_n\mathbb{E}|X_n|1_{\{|X_n|\ge M\}}\le\varepsilon$. Consequently

$$\mathbb{E}|X_n-X|\le 2\varepsilon+\mathbb{E}|\phi(X_n)-\phi(X)|\le 2\varepsilon+\varlimsup\mathbb{E}|\phi(X_n)-\phi(X)|$$

A sequence convergences in probability iff every subsequence converges a.s., so the $\varlimsup$ vanishes. □

Lecture 2: January 9

Recall:

Theorem 2.1 (4.11). *$X_n \to X_1$ in $\mathcal{L}^1$ (i.e. $\lim_n \mathbb{E}|X_n - X| = 0$) iff $\{X_n\}$ u.i. and $X_n \to X$ in probability.*

Theorem 2.2 (4.12). *Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $\xi \in \mathcal{L}^1(\mathbb{P})$, then*

$$\{\mathbb{E}[\xi \mid \mathcal{G}] : \mathcal{G} \subset \mathcal{F}\}$$

is a u.i. family.

Proof. Fix $M \geq 1$. Now we use triangle inequality (i.e. Jensen's for $|\cdot|$) to obtain

$$\begin{aligned}\int_{|\mathbb{E}(\xi \mid \mathcal{G}| > M)} |\mathbb{E}(\xi \mid \mathcal{G})| \, dP &\leq \int_{|\mathbb{E}(\xi \mid \mathcal{G}| > M)} \mathbb{E}(|\xi| \mid \mathcal{G}) \, dP \\ &= \int_{|\mathbb{E}(\xi \mid \mathcal{G}| > M)} |\xi|| \, dP.\end{aligned}$$

Moreover by Chebyshev,

$$\begin{aligned}\mathbb{P}(|\mathbb{E}(\xi \mid \mathcal{G})| > M) &\leq \frac{\mathbb{E}(|\mathbb{E}(\xi \mid \mathcal{G})|)}{M} \\ &\leq \frac{\mathbb{E}|\xi|}{M}.\end{aligned}$$

Lemma 2.3 (4.13). *If $\xi \in \mathcal{L}^1(\mathbb{P})$, then $\forall \varepsilon \exists \delta$ such that*

$$\mathbb{P}(A) < \delta \implies \mathbb{E}[|\xi|; A] < \varepsilon.$$

Proof. Suppose not. Then we have $\varepsilon > 0$ and A_n such that $\mathbb{P}(A_n) < 1/n$ but $\mathbb{E}[|\xi|; A_n] \geq \varepsilon$. Consequently $|\xi| 1_{\{A_n\}} \to 0$ in probability. Passing to a subsequence yields $|\xi| 1_{\{A_{n_k}\}} \to 0$ p.w.a.e. Since $|\xi| 1_{\{A_n\}} \leq |\xi|$, we may apply DCT to obtain $\lim_k \mathbb{E}[|\xi| 1_{\{A_{n_k}\}}] = 0$, contradiction. □

Theorem 2.4 (4.14). *For a martingale $\{X_n, \mathcal{F}_n\}$ TFAE:*

(a) $\{X_n\}$ is u.i.

(b) X_n converges a.e. and in $\mathcal{L}^1$

(c) X_n converges in $\mathcal{L}^1$

(d) $\exists X \in \mathcal{L}^1(\mathbb{P})$ such that $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$

Property (d) is historically called a **closed** martingale.

Proof. By Theorems 4.9 (26.4) and 4.11 (1.5), $(a) \implies (b)$. Then $(b) \implies (c)$ is clear, and $(c) \implies (d)$ by Thm 4.9 (26.4) (needs explanation). Lastly $(d) \implies (a)$ by Theorem 4.12 (2.2).

Explanation of $(c) \implies (d)$: Suppose $X_n \to \widetilde{\xi}$ in $\mathcal{L}^1$. Then $\sup_n \mathbb{E}|X_n| < \infty$. By Thm 4.9 (26.4), $X_n \to \widetilde{\xi}$ a.s. For any $m > n$, we have

$$X_n = \mathbb{E}(X_m \mid \mathcal{F}_n) \to_{\mathcal{L}^1} \mathbb{E}(\widetilde{\xi} \mid \mathcal{F}_n).$$

Indeed, observe that

$$\begin{aligned}\mathbb{E}|\mathbb{E}(X_m \mid \mathcal{F}_n) - \mathbb{E}(\xi \mid \mathcal{F}_n)| &= \mathbb{E}|\mathbb{E}(X_m - \xi \mid \mathcal{F}_n)| \\ &\le \mathbb{E}\mathbb{E}(|X_m - \xi| \mid \mathcal{F}_n) \\ &= \mathbb{E}|X_m - \xi| \to 0.\end{aligned}$$

Why do we call the martingale **closed**? Write X_∞ for ξ and let $\mathcal{F}_\infty = \mathcal{F}$ (or it could be $\bigcup_n \mathcal{F}_n$, it doesn't affect martingale convergence. Under the conditions of Thm 4.14 (2.4), the family $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}\cup\{\infty\}}$ is a martingale.

Corollary 2.5. *Suppose $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ (that is, $\mathcal{F}_n \subset \mathcal{F}_{n+1} \forall n$ and $\mathcal{F}_\infty = \sigma(\mathcal{F}_1, \mathcal{F}_2, \cdots)$). Then for any $\xi \in \mathcal{L}^1(\mathbb{P})$, $\mathbb{E}(\xi \mid \mathcal{F}_n) \to \mathbb{E}(\xi \mid \mathcal{F}_\infty)$ a.s. and in $\mathcal{L}^1$.*

Proof. Since $\{\mathbb{E}(\xi \mid \mathcal{F}_n)\}$ is a u.i. martingale, it follows that $X_n := \mathbb{E}(\xi \mid \mathcal{F}_N) \to X_\infty$ a.s. and in $\mathcal{L}^1$. Now $\forall A \in \bigcup_n \mathcal{F}_n$, $\exists m$ such that $A \in \mathcal{F}_m$. Since $X_m = \mathbb{E}(X_\infty$, it follows that

$$\begin{aligned}\int_A X_m \, dP &= \int_A \mathbb{E}(X_\infty \mid \mathcal{F}_m) \, dP \\ &= \int_A X_\infty \, dP \\ \text{Meanwhile,} \qquad \int_A X_m \, dP &= \int_A \mathbb{E}(\xi \mid \mathcal{F}_m) \, dP \\ &= \int_A \xi \, dP \\ \implies \int_A X_\infty \, dP &= \int_A \xi \, dP\end{aligned}$$

This holds $\forall A \in \sigma(\cup \mathcal{F}_n) = \mathcal{F}_\infty$. Thus by tools from measure theory (π-system, Monotone Class Lemma, Caratheodory Extension) it follows that $X_\infty = \mathbb{E}(\xi \mid \mathcal{F}_\infty)$. ☐

Lecture 3: January 12

Theorem 3.1 (4.15). *Suppose $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ and $\mathbb{E}(|X|) < \infty$. Then $\mathbb{E}(X \mid \mathcal{F}_n) \to \mathbb{E}(X \mid \mathcal{F}_\infty)$ a.s. and in $\mathcal{L}^1$.*

Corollary 3.2 (4.16). *Suppose $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$. Then $\mathbb{E}(1_A \mid \mathcal{F}_n) \to 1_A$ a.s.*

This is called a zero-one law, since 1_A is 0 or 1. Note that this implies convergence in $\mathcal{L}^1$, by bounded convergence. More generally if $X \in \mathcal{F}_\infty$ with $\mathbb{E}|X| < \infty$, then $\mathbb{E}(X \mid \mathcal{F}_n) \to X$ a.s. and in $\mathcal{L}^1$.

Observe that $A \mapsto \mathbb{E}(1_A \mid \mathcal{G})$ looks like a random measure; for every $\omega \in \Omega$, we set $\nu_\omega(A) = \mathbb{E}(1_A \mid \mathcal{G})(\omega)$. The issue preventing ν_ω from being a measure is that the expectation is only defined almost everywhere. Hence we have to discard null sets. Yet such a choice depends on A, the set we're measuring, preventing a universal definition.

Definition 3.3. *If there are a family of probability measure $\mathbb{P}(\omega, \cdot)$ for $\omega \in \Omega$ such that:*

(a) For every ω, $\mathbb{P}(\omega, \cdot)$ is a measure

(b) $\forall A \in \mathcal{F}$, $\mathbb{P}(\omega, A) = \mathbb{E}(1_A \mid \mathcal{G})$ a.s.

Then $\mathbb{P}(\cdot, \cdot)$ is called a **regular conditional probability**.

Theorem 3.4 (DCT for Conditional Expectation, 4.17). *Suppose $Y_n \to Y$ a.s. and $|Y_n| \le Z$ for all n, a.s. with $\mathbb{E}|Z| < \infty$. If $\mathcal{F}_n \nearrow \mathcal{F}$ then $\mathbb{E}(Y_n \mid \mathcal{F}_n) \to \mathbb{E}(Y \mid \mathcal{F}_\infty)$ a.s.*

There is a subtlety: is the exceptional set uniform for all n? But a countable union of null sets is null, so we can take their union.

Proof.

$$\begin{aligned} |\mathbb{E}(Y_n \mid \mathcal{F}_n) - \mathbb{E}(Y \mid \mathcal{F}_\infty)| &= |\mathbb{E}(Y_n - Y \mid \mathcal{F}_n)| + |\mathbb{E}(Y \mid \mathcal{F}_n) - \mathbb{E}(Y \mid \mathcal{F}_\infty)| \\ &\le \overline{\lim_{n\to\infty}}\, \mathbb{E}(|Y_n - Y| \mid \mathcal{F}_n) + \mathbb{E}(Y \mid \mathcal{F}_n) - \mathbb{E}(Y \mid \mathcal{F}_\infty)| \end{aligned}$$

Let $W_N = \sup_{j,m \ge N} |Y_j - Y_m|$. Then the first term is bounded by $\overline{\lim}_{N\to\infty} \overline{\lim}_{n\to\infty} \mathbb{E}(W_N \mid \mathcal{F}_n)$. Note that $|W_N| \le 2Z$ and $W_N \searrow 0$ a.s., so by DCT $W_N \to 0$ in $\mathcal{L}^1$. It follows by Theorem 4.15 that

$$\begin{aligned} \overline{\lim_{n\to\infty}}\, \mathbb{E}(|Y_n - Y| \mid \mathcal{F}_n) &\le \overline{\lim_{N\to\infty}}\, \overline{\lim_{n\to\infty}}\, \mathbb{E}(W_N \mid \mathcal{F}_n) \\ &= \overline{\lim_{N\to\infty}}\, \mathbb{E}(W_N \mid \mathcal{F}_\infty) \end{aligned}$$

Since $W_N \searrow 0$ is monotone, the limit exists so we only need to identify it. This gives us a.s. convergence. Homework: Figure out what conditions you need. This is the difference between convergence in probability and convergence a.s. □

Optional Stopping (4.4)

Given a martingale X_n (i.e., $\mathbb{E}|X_n| < \infty, \mathbb{E}(X_m \mid \mathcal{F}_n) = X_n$ for $m \ge n$), we may ask two questions:

(a) Given two stoping times $T \ge S$, is it true that $\mathbb{E}(X_T \mid \mathcal{F}_S) = X_S$?

(b) Simpler question: is $\mathbb{E}X_T = \mathbb{E}X_0$.

Optimal Stopping

We would like to extend the martingale property from deterministic times to random times.

Theorem 3.5 (4.19). *If $\{X_n, \mathcal{F}_n\}$ is a u.i. submartingale, then for any stopping time T, so is $X_{T\wedge n}$ i*

Proof. By Corollary 4.7, $X_{T\wedge n}$ is a submartingale, so we just have to check u.i. As $x \mapsto x^+$ is a convex increasing function, it follows (by Jensen and monotonicity) that X_n^+ is also a submartingale. Now we compute

$$\begin{aligned}
\mathbb{E}X_{T\wedge n}^+ &= \sum_{k=0}^{n} \mathbb{E}(X_k^+; T = k) + \mathbb{E}(X_n^+; T > n) \\
&\le \sum_{k=0}^{n} \mathbb{E}(X_n^+; T = k) + \mathbb{E}(X_n^+; T > n) \\
&= \mathbb{E}(X_n^+).
\end{aligned}$$

Consequently $\sup_n mE(X_{T\wedge n}^+) \le \sup_n \mathbb{E}(X_n^+) < \infty$. We claim that $X_T = \lim_{n\to\infty} X_{n\wedge T}$. To make sense of this statement, we have to define what we mean when $T = \infty$. But since $\{X_n\}$ is u.i., it converges to some X, which we set as X_∞.

As we saw last quarter, the condition $\sup_n \mathbb{E}(X_{T\wedge n}^+) < \infty$ is equivalent to requiring $\sup_n \mathbb{E}(|X_{T\wedge n}|) < \infty$. Consequently by Fatou,

$$\mathbb{E}(|X_T|) \le \sup_n \mathbb{E}|X_n| < \infty.$$

Hence for any $M > 0$,

$$\begin{aligned}
\mathbb{E}(|X_{T\wedge n}|; |X_{T\wedge n}| \ge M) &= \mathbb{E}(|X_T|; |X_T| \ge M, T \le n) + \mathbb{E}(|X_n|; |X_n| \ge M, T > n) \\
&\le \mathbb{E}(|X_T|; |X_T| \ge M) + \mathbb{E}(|X_n|; |X_n| \ge M).
\end{aligned}$$

Taking suprema over n, it follows that both parts vanish as $M \to \infty$. The first follows from $X_T \in \mathcal{L}^1(\mathbb{P})$, and the second from u.i. □

Remark 3.6. *From the proof, we see that all we need is $\mathbb{E}|X_T| < \infty$ and $\{X_n 1_{\{T>n\}}\}$ is u.i. Just rewrite the last inequality as*

$$\le \mathbb{E}(|X_T|; |X_T| \ge M) + \mathbb{E}(|X_n| 1_{\{T>n\}}; |X_n| \ge M, T > n)$$

But the question arises: how do we check these two conditions? Well, the easiest (and common) case is when we actually have a submartingale anyway. It's a convenience theorem, not a tight theorem.

Theorem 3.7 (4.20). *If $\{X_n\}$ is a ui. submartingale, then for any stopping time T, $\mathbb{E}X_0 \le \mathbb{E}X_T \le \mathbb{E}X_\infty$, where X_∞ is its limit (by martingale convergence).*

Proof. Since $X_{T\wedge n}$ is a submartingale, we have $\mathbb{E}X_0 \le \mathbb{E}X_{T\wedge n} \le \mathbb{E}X_n$ by the same calculation as in the last proof. By martingale coonvergence (4.14 and 4.19), it follows that $X_n \to X$ in $\mathcal{L}^1$ and $X_{T\wedge n} \to X_T$ in $\mathcal{L}^1$. □

Theorem 3.8 (4.21). *If $S \ge T$ are stopping times and $\{X_{T\wedge n}; n \ge\}$ is a u.i. submartingale, then $X_S \le \mathbb{E}(X_T \mid \mathcal{F}_S)$ and $\mathbb{E}X_S \le \mathbb{E}X_T$.*

Proof. Observe that $X_S \le X_T$ follows immediately, by considering the martingale $Y_n = X_{T\wedge n}$. For the first part, recall that $X_T(\omega) = X_{T(\omega)}(\omega)$. Thus it is clear that X_T is $\mathcal{F}_T$ measurable. □

Lecture 4: January 16

Martingales

Corollary 4.1 (4.22). *(a) If $\{X_n\}$ is a u.i. submartingale, then $\{X_T\}$ over all stopping times is a u.i. family.*

(b) If $\{X_n\}$ is a u.i. martingale then for any stopping time, $X_T = \mathbb{E}(X_\infty \mid \mathcal{F}_T)$.

Proof. (a) Let $X_\infty = \lim_n X_n$. By Theorem 3.8, $0 \le X_T \le \mathbb{E}(X_\infty \mid \mathcal{F}_T)$. Hence $\{X_T\}$ is u.i.

(b) Follows from 3.8 by applying it to X_n and $-X_n$ with $S \leftarrow T$ and $T \leftarrow \infty$. □

Theorem 4.2 (4.23). *Suppose $\{X_n\}$ is a submartingale and $\mathbb{E}(|X_{n+1} - X_n| \mid \mathcal{F}_n) \le B < \infty$ a.s. If T is a stopping time with $\mathbb{E}T < \infty$, then $\{X_{T\wedge n}\}$ is u.i. and $\mathbb{E}X_T \ge \mathbb{X}_0$.*

Theorem 4.3 (Recalling Theorem 3.8). *$S \le T$, $\{X_{T\wedge n}\}$ u.i. submartingale. Then $X_S \le \mathbb{E}(X_T \mid \mathcal{F}_s)$.*

Example 4.4. *Let $X_n = \sum_{i=1}^n \xi_i$ with ξ_i iid. It's a martingale if $\mathbb{E}\xi_1 = 0$, and a submartingale if $\mathbb{E}\xi_i \ge 0$. Then the* **martingale difference** *$X_{n+1} - X_n = \xi_{n+1}$, motivating the expression which appears in Theorem 4.2.*

Proof of Theorem 4.2. We have the following:

$$\begin{aligned}
X_T &= X_0 + \sum_{k=0}^{\infty}(X_{k+1} - X_k)1_{\{T>k\}} \\
\implies X_{T\wedge n} &= X_0 + \sum_{k=0}^{n-1}(X_{k+1} - X_k)1_{\{T>k\}} \\
\implies |X_{T\wedge n}| &\le |X_0| + \sum_{k=0}^{\infty}|X_{k+1} - X_k|1_{\{T>k\}} \\
\implies \mathbb{E}|X_{T\wedge n}| &\le \mathbb{E}|X_0| + \sum_{k=0}^{\infty}\mathbb{E}\mathbb{E}(|X_{k+1} - X_k|1_{\{T>k\}} \mid \mathcal{F}_k) \\
&\le \mathbb{E}|X_0| + \sum_{k=0}^{\infty}\mathbb{E}(B1_{\{T>k\}}) \\
&= \mathbb{E}|X_0| + B\mathbb{E}T < \infty.
\end{aligned}$$

Corollary 4.5. *If ξ_i are iid with $\mathbb{E}|\xi_i| < \infty$ and $S_n = \sum_{i=1}^n \xi_i$ and T is a stopping time with $\mathbb{E}T < \infty$ then $\mathbb{E}S_T = \mathbb{E}T\mathbb{E}\xi_1$ (Wald's Identity).*

Lecture 5: January 21

Theorem 5.1 (4.23). *Suppose X_n is a submartingale and $\mathbb{E}(|X_{k+1} - X_k \mid \mathcal{F}_k) \le B$ a.s., where B is a constant.*

If T is a stopping time with $\mathbb{E}T < \infty$, then $\{X_{T\wedge n}\}$ is u.i. and hence $\mathbb{E}X_T \ge \mathbb{E}X_0$.

Recall that the truncation is a submartingale, as we showed a while ago.

Example 5.2. *If ξ_i are iid with $\mathbb{E}|\xi_i| < \infty$ and let $S_n = \sum_{k=1}^n \xi_k$. For any stopping time T with $\mathbb{E}T < \infty$, we have $\mathbb{E}S_T = \mathbb{E}X_1 \mathbb{E}T$.*

Proof. Let $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$. Then $\{X_n, \mathcal{F}_n\}$ is a martingale, where $X_n = \sum(\xi_k - \mathbb{E}\xi_k)$. Hence

$$\mathbb{E}(|X_{k+1} - X_k| \mid \mathcal{F}_k) = \mathbb{E}|\xi_1 - \mathbb{E}\xi_1| \le 2\mathbb{E}|\xi_1| < \infty.$$

By Theorem 4.2, $\mathbb{E}X_T = \mathbb{E}X_1 = 0$. Since $X_T = S_T - \mathbb{E}\xi_1 T$ on $T = n$ (and $T < \infty$ a.s.), it follows that $\mathbb{E}(S_T - \mathbb{E}X_1 T) = \mathbb{E}X_T = 0$. This implies the result. □

Example 5.3 (Biased random walk). *Let ξ_i be iid with $\mathbb{P}(\xi_1 = 1) = p$. Then $\mathbb{P}(\xi_1 = -1) = 1-p = q$. Say $p > \frac{1}{2}$. Then $\xi_1 = 2p - 1 > 0$, so the walk drifts to the right.*

Let $S_n = \sum_{k=1}^n \xi_k$. We would like to calculate $\mathbb{P}(T_a < T_b)$. If we just center by translating, we get a martingale that doesn't tell us quite enough (we don't know the expected exit time out of an interval). However, there is another way to rescale multiplicatively that works out.

Define $X_k = \phi(S_n) = \frac{q}{p}^{S_n}$. Then we compute

$$\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[\left(\frac{q}{p}\right)^{S_n + \xi_{n+1}} \mid \mathcal{F}_n] \\
&= X_n \mathbb{E}[\left(\frac{q}{p}\right)^{S_n + \xi_{n+1}} \mid \mathcal{F}_n] \\
&= X_n \left(\frac{q}{p} p + \frac{p}{q} q\right) \\
&= X_n.
\end{aligned}$$

Note that $\phi(k) \le 1$ for $k \ge 0$. Given $a < 0 < b$, let $T = T_a \wedge T_b$: the exit time of (a, b) by S_n. Then $\phi(S_{n\wedge T})$ is a bounded martingale, since it is either $\phi(a)$ or $\phi(b)$. So $\mathbb{E}\phi(S_T) = \mathbb{E}\phi(S_0) = 1$. On the other hand, we have $S_T = a 1_{\{T_a < T_b\}} + b 1_{\{T_b < T_a\}}$. Hence

$$\mathbb{P}(T_a < T_b) = \frac{\phi(0) - \phi(b)}{\phi(a) - \phi(b)} = \frac{\left(\frac{q}{p}\right)^b - 1}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$$

Consequently $\mathbb{P}(T_a < \infty) = \left(\frac{q}{p}\right)^{|a|}$, whereas $\mathbb{P}(T_b < \infty) = 1$.

If $a < 0$, we see that $\mathbb{E}(\inf S_n)] = \frac{1}{2p-1} < \infty$ - thus not only is it pointwise finite, it also has finite expectation.

Lecture 6: January 23

Example 6.1 (Biased Random Walk). *Consider $\xi_1, \cdots$ iid having values ± 1, and wlog set $p = \mathbb{P}(\xi_1 = 1) > \frac{1}{2}$. Set $S_n = \sum_{k=1}^n \xi_k$ and $X_n = \sum_{k=1}^n (\xi_k - \mathbb{E}\xi_k)$. Then X_n is a martingale.*

(a) Let $\phi(k) = \left(\frac{q}{p}\right)^k$, then $\phi(S_n)$ is a martingale

(b) For $a < 0 < b$ we computed

$$\mathbb{P}(T_a < T_b) = \frac{\phi(b) - \phi(0)}{\phi(b) - \phi(a)}$$

(c) Letting $b \to \infty$, we have

$$\begin{aligned} \mathbb{P}(\inf S_n \le a) &= \mathbb{P}(T_a < \infty) \\ &= \frac{1}{\phi(a)} = \left(\frac{q}{p}\right)^{|a|} \\ \implies \mathbb{E}[(\inf S_n)^-] &= \sum_{k=0}^{\infty} \mathbb{P}(T_{-k} < \infty) \\ &= \frac{1}{1 - \frac{q}{p}} = \frac{p}{2p-1}. \end{aligned}$$

(d) Let $a \to -\infty$. Then $\mathbb{P}(T_b < \infty) = 1$. Is $\mathbb{E}T_b < \infty$?

Let's first compute $\mathbb{E}T$ where $T = T_a \wedge T_b$. Since $S_{T\wedge n}$ is a bounded and $X_{T\wedge n}$ is a martingale, we have $\mathbb{E}X_{T\wedge n} = \mathbb{E}X_0 = 0$. We claim that any bounded family is u.i. To be precise, suppose we have $\{\xi_t, t \in \Lambda\}$. If $|\xi_t| \le \xi \in \mathcal{L}^1$ for any t, then $\{\xi_t, t \in \Lambda\}$ is u.i. We want to show that

$$\sup_{t\in\Lambda} \int_{|\xi_t|>M} |\xi_t| \, dP \to 0, \qquad \text{as } M \to \infty.$$

But this follows, since $\{|\xi_t| > M\} \subset \{\xi > M\}$, which means that

$$\int_{|\xi_t|>M} |\xi_t| \, dP \le \int_{\xi > M} \xi \, dP$$

Returning to the application, we obtain that

$$\begin{aligned} \mu\mathbb{E}[T \wedge n] &= \mathbb{E}S_{T\wedge n} \\ \text{(Bounded Convergence)} \qquad &\to \mathbb{E}S_T \\ \text{(Monotone Convergence)} \qquad &\to \mu ET. \end{aligned}$$

Now we can proceed in our computations, recalling that $\mathbb{E}S_T = a\mathbb{P}(T_a < T_b) + b\mathbb{P}(T_b < T_a)$. As $T_a \wedge T_b \nearrow T_b$ as $a \searrow -\infty$, we have that

$$\begin{aligned} \text{(Monotone Convergence)} \qquad \mathbb{E}T_b &= \lim_n \frac{\mathbb{E}[T_{-n} \wedge T_b]}{\mu} \\ &= \frac{b}{2p-1}. \end{aligned}$$

Thus using martingales, we can compute hitting very explicitly. Note that the $p \to \frac{1}{2}$ limit heuristically recovers our earlier results.

Theorem 6.2 (Doob's Maximum Inequality, 4.24). *Let $\{X_n\}$ be a submartingale. Define $\overline{X}_n = \max_{k \le n} X_k^+$. Then for any $\lambda > 0$, we have*

$$\lambda \mathbb{P}(\overline{X}_n \ge \lambda) \le \mathbb{E}(X_n; \overline{X}_n \ge \lambda) \le \mathbb{E}(X_n^+).$$

Proof. We could have stated the theorem for non-negative submartingales, since if $\{X_n\}$ is a submartingale then so is $\{X_n^+\}$. Let $T = \inf\{k \mid X_k \ge \lambda\}$ and set $A = \{\overline{X}_n \ge \lambda\}$. Then $T \le n$ on A and $T > n$ on A^c. So

$$\begin{aligned} \lambda \mathbb{P}(\overline{X}_n \ge \lambda) &\le \mathbb{E}(X_{T \wedge n}; \overline{X}_n \ge \lambda) \\ &\le \mathbb{E}(X_n; \overline{X} \ge \lambda). \end{aligned}$$

Indeed, this follows since

$$\begin{aligned} & \mathbb{E} X_{T \wedge n} \le \mathbb{E} X_n \\ \implies & \mathbb{E}(X_{T \wedge n; A}) + \mathbb{E}(X_{T \wedge n}; A^c) \le \mathbb{E}(X_n; A) + \mathbb{E}(X_n; A^c) \\ \text{(cancelling)} \qquad \implies & \mathbb{E}(X_{T \wedge n; A}) \le \mathbb{E}(X_n; A). \end{aligned}$$

Chaining with earlier inequalities, we obtain

$$\begin{aligned} \lambda \mathbb{P}(\overline{X}_n &\le \mathbb{E}(X_n; \overline{X} \ge \lambda) \\ &\le \mathbb{E}(X_n^+; \overline{X}_n \ge \lambda) \\ &\le \mathbb{E} X_n^+. \end{aligned}$$

In applications, we mostly ignore the middle term. This is called a **weak 1-1 type** inequality.

Theorem 6.3 (L^p maximum inequality, 4.25). *If $\{X_n\}$ is a submartingale and $p \in (1, \infty)$ then $\|\overline{X}_n\|_p \le \frac{p}{p-1} \|X_n^+\|_p$.*

Note that we set $\|\xi\|_p = (\mathbb{E}|\xi|^p)^{1/p}$.

Proof. If we knew a priori that $\overline{X}_n$ was in $\mathcal{L}^p$, we could apply Hölder and be done. Instead, we have to truncate:

$$\mathbb{E}[(\overline{X}_n \wedge a)^p] = \int_0^a p \lambda^{p-1} \mathbb{P}(\overline{X}_n \ge \lambda) \, d\lambda.$$

Lecture 7: January 26

Tail inequalities for martingales, 4.5

Example 7.1 (Tail inequality for random walk). *Let $S_n = \sum_{k=1}^n \xi_k$ with ξ_k iid. Take $\mathbb{E}\xi_1 = 0$ and $\sigma^2 = \mathbb{E}\xi_1^2 = \mathrm{Var}(\xi_1)$. Then $\frac{S_n}{\sqrt{n}} \Rightarrow \sigma\chi$ where $\chi \sim N(0,1)$. Can we bound the probability that $\left|\frac{S_n}{\sqrt{n}}\right| \geq \lambda$?*

Roughly speaking, we have

$$\mathbb{P}\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq \lambda\right) \approx \mathbb{P}(\sigma|\chi| \geq \lambda) = \int_{|x|\geq\lambda} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}\, dx.$$

We bound it using the same trick as if we were computing it:

$$\left(\int_{|x|\geq\lambda} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}\, dx\right)^2 \leq \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_{\sqrt{2}\lambda}^\infty e^{-r^2/2\sigma^2} r\, dr\, d\theta = e^{-\lambda^2/\sigma^2}.$$

Thus $\mathbb{P}(|\sigma X| \geq \lambda) \leq e^{-\lambda^2/2\sigma^2}$.

Theorem 7.2 (Azuma-Hoeffding Inequality, 4.26). *Suppose X_n is a martingale with $|X_k - X_{k-1}| \leq c_k$. Then for all $n \geq 0$ and $\lambda > 0$ we have*

$$\mathbb{P}(|X_n - X_0| \geq \lambda) \leq 2e^{-\lambda^2/2\sum c_k^2}$$

To appreciate the inequality, let X_n be SRW (as in Example 7.1). Then $|X_k - X_{k-1}| = 1 = c_k$. Take $\lambda = a\sqrt{n}$. Then the tail inequality yields

$$\mathbb{P}\left(\frac{|X_n - X_0|}{\sqrt{n}} \geq a\right) \leq 2e^{-a^2/2}.$$

The key inequality in this proof is the following:

Claim 7.3 (Azuma's Inequality). *If X is a r.v. with $\mathbb{E}X = 0$ and $|X| \leq c$, then $\mathbb{E}e^{aX} \leq e^{a^2c^2/2}$.*

Proof of Theorem 7.2. Let $\xi_k = X_k - X_{k-1}$ (the marginal difference). Then $\xi_k \in [-c_k, c_k]$, so we may write $\xi_k = \lambda(-c_k) + (1-\lambda)c_k$ for some $\lambda \in [0,1]$. In fact, we have $\lambda = \frac{\xi_k + c_k}{2c_k}$. By convexity of e^{ax},

$$\begin{aligned} e^{a\xi_k} &\leq e^{-ac_k}\frac{c_k - \xi_k}{2c_k} + e^{ac_k}\frac{\xi_k + c_k}{2c_k} \implies \\ \mathbb{E}(e^{a\xi_k} \mid \mathcal{F}_{k-1}) &\leq \mathbb{E}\left(e^{-ac_k}\frac{c_k - \xi_k}{2c_k} + e^{ac_k}\frac{\xi_k + c_k}{2c_k} \,\middle|\, \mathcal{F}_{k-1}\right) \\ &= \frac{e^{-ac_k} + e^{ac_k}}{2} \\ \text{(Taylor)} \qquad &\leq e^{a^2c_k^2/2}. \end{aligned}$$

The same reasoning extends to handle the telescoping sum $X_n - X_0 = \sum_{k=1}^n \xi_k$.

$$\begin{aligned}
\mathbb{E}e^{a(X_n - X_0)} &= \mathbb{E}\exp\left[a\sum_{k=1}^n \xi_k\right] \\
\text{(towering)} \qquad &= \mathbb{E}\left\{\mathbb{E}\left(\exp\left[a\sum_{k=1}^n \xi_k\right] \,\middle|\, \mathcal{F}_{n-1}\right)\right\} \\
&\le \mathbb{E}\left\{\exp\left[\frac{a^2}{2}\sum_{k=1}^n c_k^2\right]\right\} \\
&= e^{a^2 \sum c_k^2/2}.
\end{aligned}$$

Now introduce the parameter $\lambda > 0$. Then

$$\begin{aligned}
\mathbb{P}(X_n - X_0 \ge \lambda) &= \mathbb{P}(e^{a(X_n - X_0)} \ge e^{a\lambda}) \\
&= \mathbb{P}(e^{a(X_n - X_0 - \lambda)} \ge 1) \\
\text{(Markov)} \qquad &\le \mathbb{E}e^{a(X_n - X_0 - \lambda)} \\
&\le e^{-a\lambda + a^2 \sum c_k^2/2}.
\end{aligned}$$

Choosing $a = \frac{\lambda}{\sum c_k^2}$ minimizes the quadratic. Plugging in yields

$$\mathbb{P}(X_n - X_0) \le e^{-\lambda^2/2\sum c_k^2}$$

Applying the same analysis to the martingale $-X_n$ yields the bound

$$\mathbb{P}(|X_n - X_0| \ge \lambda) \le 2e^{-\lambda^2/2\sum c_k^2}. \qquad \square$$

The bound is rather sharp.

Corollary 7.4 (4.27). *Let X_n be a martingale with $|X_k - X_{k-1}| \le c$. Then*

$$\mathbb{P}\left(\frac{|X_n - X_0|}{\sqrt{n}} \ge \lambda\right) \le 2e^{-\lambda^2/2c^2}$$

Theorem 7.5 (Azuma-Hoeffding II, 4.28). *Let X_k be a martingale with $\xi_k \le X_k - X_{k-1} \le \xi_k + d_k$ for some constant $d_k > 0$ and random variable $\xi_k \in \mathcal{F}_{k-1}$. Then for $n > 0$ and $\lambda > 0$,*

$$\mathbb{P}(|X_n - X_0| \ge \lambda) \le 2\exp\left(-2\lambda^2 / \sum_k d_k^2\right).$$

Lecture 8: January 28

Theorem 8.1 (Extended version of Azuma-Hoeffding Inequality II, 4.28). *If $\{X_n\}$ is a martingale such that $\xi_k \le X_k - X_{k-1} \le \xi_k + d_k$ for some constant d_k and ξ_k is predictable, then*

$$\mathbb{P}(|X_n - X_0| \le \lambda) \le 2e^{-2\lambda^2/\sum_{i\le n} d_i^2}$$

Recall that for ξ_k to be **predictable**, we have $\xi_k \in \mathcal{F}_{k-1}$. Note that Theorem 7.2 corresponds to the case $\xi_k = -c_k$ and $d_k = 2c_k$.

Lemma 8.2 (Hoeffding's Lemma). *If X is an r.v. with $\mathbb{E}X = 0$ and $c_- \le X \le c_+$, then for any $a > 0$*

$$\mathbb{E}e^{aX} \le e^{a^2(c_+-c_-)^2/8}.$$

Proof. Homework 1, Problem 2. □

Applications

These types of inequalities are called **tail inequalities**.

Example 8.3. *Suppose $\{Z_i\}_{i\le n}$ are independent r.v. and f is a Lipschitz function with condition C.*

Definition 8.4. *A function f is* **Lipschitz with condition** C *if $|f(x)-f(\widehat{x}_i)| \le C$ for any $i, x, \widehat{x}_i$.*

Let $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and $\mathcal{F}_k = \sigma(Z_1, \cdots, Z_k)$. Set $X_k = \mathbb{E}[f(Z_1, \cdots, Z_n \mid \mathcal{F}_k)$. Then $X_0 = \mathbb{E}f(Z_1, \cdots, Z_n)$ and $X_n = f(Z_1, \cdots, Z_n)$.

Claim 8.5. *$\{X_n\}$ satisfies the conditions of Theorem 7.5 with $d_k = C$.*

Proof. Let μ_k be the distribution of Z_k induced on $\mathbb{R}$. Then

$$\begin{aligned} X_k &= \mathbb{E}f(Z_1, \cdots, Z_n) \mid \mathcal{F}_k) \\ &= \int_{\mathbb{R}^{n-k}} f(Z_1, \cdots, Z_k, \xi_{k+1}, \cdots, \xi_n)\mu_{k+1}(d\xi_{k+1})\cdots\mu_n(d\xi_n). \end{aligned}$$

Let $\xi_k = \inf y \int_{\mathbb{R}^{n-k}} f(Z_1, \cdots, Z_{k-1}, y, \xi_{k+1}, \cdots, \xi_n)\mu_{k+1}(d\xi_{k+1})\cdots\mu_n(d\xi_n) - X_{k-1}$. Then $\xi_k \le X_k - X_{k-1} = \xi_k + (X_k - X_{k-1} - \xi_k)$. But by the hypotheses on f, we obtain $X_k - X_{k-1} - \xi_k \le C$. Hence by Theorem 7.5, it follows that

$$\mathbb{P}(|f(Z_1, \cdots, Z_n) - \mathbb{E}f(Z_1, \cdots, Z_n)| \ge \lambda) \le 2e^{-2\lambda^2/nC^2}.$$

This gives us a nice Gaussian-type tail bound:

$$\mathbb{P}\left(\frac{|f(Z_1, \cdots, Z_n) - \mathbb{E}f(Z_1, \cdots, Z_n)|}{C\sqrt{n}} \ge \lambda\right) \le 2e^{-2\lambda^2}.$$

□

Example 8.6 (Pattern Matching). *Let $\{X_i\}_{i\le n}$ be iid uniform from an alphabet Σ of size N. Take $A = \{a_1, \cdots, a_k\}$ to be a particular string of k characters from Σ, and let Y denote the number of occurrences of A in the random string $\{X_1, \cdots, X_n\}$. What is $\mathbb{P}(|Y - \mathbb{E}Y| \ge \lambda)$ bounded by?*

We observe that $\mathbb{E}Y = \frac{n-k+1}{N^k}$. Let $\mathcal{F}_k = \sigma(X_1, \cdots, X_k)$ with $Z_k = \mathbb{E}[Y \mid \mathcal{F}_k]$ where $Y = f(X_1, \cdots, X_n)$. We claim that f is Lipstchitz of constant k. Indeed, it only effects k possible words. Thus $|Z_k - Z_{k-1}| \le k$ (knowing an additional character only affects your prediction by at most k). Hence we may apply Theorem 7.2 to obtain

$$\mathbb{P}(|Y - \mathbb{E}Y| \ge \lambda) \le 2e^{-\lambda^2/2k^2 n}.$$

But in view of the last example and Theorem 7.5, we have the sharper bound

$$\mathbb{P}(|Y - \mathbb{E}Y| \ge \lambda) \le 2e^{-2\lambda^2/k^2 n}.$$

Lecture 9: January 30

Balls and Bins

We have m balls and n bins, place them iid uniform (we refer to the set of bins as V). Let $Y_i = 1_{\{\text{bin } i \text{ empty}\}}$ and $Y = \sum_{i=1}^n Y_i$. Note that the Y_i are not independent. Observe that

$$\mathbb{E}Y_i = \mathbb{P}(\text{bin } i \text{ empty}) = \left(1 - \frac{1}{n}\right)^m \implies \mathbb{E}Y = n\left(1 - \frac{1}{n}\right)^m.$$

Now let's determine the tail distribution $\mathbb{P}(|Y - \mathbb{E}Y| \le \lambda)$. Let X_i denote the bin that ball i is placed in. Then $Y = f(X_1, \cdots, X_m)$ and the $\{X_i\}$ are iid. More concretely, start by defining $f: V^m \to \mathbb{N}$ as follows:

$$f(x_1, \cdots, x_m) = \#\text{ non-zero entries of } \sum_{i=1}^m \delta_{x_i}.$$

Notice that f is Lipschitz with $C = 1$. By Theorem 7.5, $\mathbb{P}(|Y - \mathbb{E}Y| \le \lambda) \le 2e^{-2\lambda^2/m}$.

Random Graphs

We consider the Erdős-Rényi random graph on n vertices: add edges with probability p (so $G = ([n], E)$ where E is a random subset of $2^{[n]}$). The random graph is called $G_{n,p}$. Define the chromatic number $\chi(G)$ to be the minimal number of colors needed to color G (or in fancy terms, $\chi(G) = \min\{q \mid \exists f: G \to K_q\}$).

Let G_i denote the subgraph of G induced by $[i] \subset [n]$. Let $Y_k = \mathbb{E}(Y \mid \sigma(G_1, \cdots, G_k))$. This is called the **vertex exposure martingale**. Observe that for some $\xi_k \in \mathcal{F}_{k-1}$, we have $\xi_k \le Y_k - Y_{k-1} \le \xi_k + 1$. Hence by Theorem 7.5,

$$\mathbb{P}(|Y - \mathbb{E}Y| \ge \lambda) = \mathbb{P}(|Y_n - Y_0| \ge \lambda) \le 2e^{-2\lambda^2/n}.$$

So the chromatic number of the Erdős-Rényi graph has a Gaussian tail bound, which we can determine even before saying anything about its mean.

Chapter 5: Markov Chains

Martingales are a certain type of generalization of random walk (having mean 0). Markov Chains are another generalization.

5.1: Definitions and Examples

Let $(S, \mathbb{S})$ be a measure space. A sequence of r.v. $\{X_n\}$ taking values in S and a filtration $\{\mathcal{F}_n\} \subset \mathcal{S}$ is said to be a **Markov chain** if $X_n \in \mathcal{F}_n$ and for any $A \in \mathcal{S}$,

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \mid X_n).$$

In other words, the next step depends only on the current position.

Example 9.1. *Consider independent* $X_0, \xi_1, \cdots, \xi_n \in \mathbb{R}^n$ *with* $X_n - X_0 = \sum_{i \le n} \xi_i$.

Set $\mathcal{F}_n = \sigma(X_0, \xi_1, \cdots, \xi_n) = \sigma(\{X_k\}_{k=0}^n)$. Then for any $A \in \mathcal{B}(\mathbb{R}^n)$, we have

$$\begin{aligned}\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) &= \mathbb{P}(X_{n+1-X_n} \in A - X_n \mid \mathcal{F}_n) \\ &= \mathbb{P}(\xi_{n+1} \in A - X_n) \\ &= \mu_{n+1}(A - X_n) \\ &= \mathbb{P}(X_{n+1} \in A \mid X_n).\end{aligned}$$

In words, we have a random walk starting at X_0, where we allow the distribution at distinct steps to vary. You can extend this to manifolds, by taking a geodesic ball at each time, then choose the next point according to some distribution.

Lecture 10: February 4

If X_n is a Markov Chain with transition probability $p_n(x, dy)$ with initial distribution μ, then

$$\mathbb{P}(X_0 \in A_1, \cdots, X_n \in A_n) = \int_{A_0 \times \cdots \times A_n} \mathbb{P}_0(x_0, dx_1) \cdots \mathbb{P}_{n-1}(x_{n-1}, dx_n) \mu(dx_0).$$

Consider $X = \{X_n\} \in S^{\mathbb{N}}$. The induced measure on $S^{\mathbb{N}}$ is denoted by $\mathbb{P}_\mu$, called the **law of** X. When $\mu(dx) = \delta_x$, we write $\mathbb{P}_\mu$ as $\mathbb{P}_x$. By Kolmogorov Extension, there is an equivalence of data between markov chains and their transition kernels.

Consider a time-homogeneous markov chain with transition kernel p. If S is countable, we can write $S = \mathbb{N}$. Then $\mathbb{P}(x, dy)$ is uniquely determined by the probability mass function

$$\{p(x,y); y \in S\}_{x \in S} \cong (p_{x,y})_{x,y \in S} \qquad (\mathbb{P}).$$

For each $x \in S$, we have $p_{x,y} \geq 0$ and $\sum_y \mathbb{P}_{x,y} = 1$.

When S is countable, markov chains are equivalent to specifying $\mathbb{P} = (p_{ij})$ with $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$ for any $i \in S$. Really think of $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$.

Examples of Markov Chains

There's a fancy word for markov chains with countable state space: random walks on graphs. We will fix a (deterministic) graph, which moreover has a finite number of vertices. The simplest way to do this is via simple random walk, i.e. each neighboring vertex is chosen uniformly.

Alternatively, we can move to different neighbors with different probabilities. Then a convenient way to encode this is via electrical networks, treating the transition probabilities like resistances (or more precisely, conductivities - the inverse of resistance). We say that $x \sim y$ if $c_{x,y} > 0$. Such a theory is symmetrical (i.e., undirected) for physical reasons.

More generally, we can think of direct graphs. Here we define $x \to y$ if $p_{x,y} > 0$. There are plenty of other examples. For instance, the most general random walk is just a sum of independent r.v., called a **time inhomogeneous markov chain**.

(a) Branching Process: After one unit of time, an individual in the population dies and gives birth to n children with probability p_n, independent of any other individual. (This is a Galton-Watson process.) Let $\{\xi_i^{(k)}\}_{i,k \in \mathbb{N}}$ be iid, with $\mathbb{P}(\xi_1^{(1)} = n) = p_n$ (so ξ_i are the number of children). Take Z_n to be the population at time n. Then we have the recursive formula

$$Z_n = \sum_{j=1}^{Z_{n-1}} \xi_j^n.$$

Claim 10.1. *Z_n is a markov chain.*

Proof. $\mathbb{P}(Z_{n+1} \mid Z_n = i) = \mathbb{P}(\sum_{k \leq i} \xi_k = j)$, where we have dropped a subscript to emphasize that they're all iid. ☐

Question: Does $\lim_{n \to \infty} Z_n$ exist? Is it positive? Is it 1? This is called the **extinction problem**, or sometimes the **family name problem**.

(a) Renewal Chain: Let $p_n \geq 0$ with $\sum_n p_n = 1$. Take $S = \{0, \cdots\}$ and define

$$\begin{cases} \mathbb{P}(0,n) = p_n \\ \mathbb{P}(i,i-1) = 1 \quad i \geq 1 \\ \mathbb{P}(i,j) = 0 \end{cases}$$

Lecture 11: February 9

Theorem 11.1 (Strong Markov Property, 5.3). *For any stopping time T and bounded r.v. Y, then on $\{T < \infty\}$ we have $\mathbb{E}_\mu[Y \circ \theta_T \mid \mathcal{F}_T] = \mathbb{E}_{X_T} Y$. As always, this is $\mathbb{P}_\mu$ a.s.*

We can relax some conditions, for instance $Y \geq 0$ (just work with truncations, then use monotone convergence).

Proof. It suffices to show that for any $A \in \mathcal{F}_T$, we have

$$\mathbb{E}[Y \circ \theta_T; A \cap \{T < \infty\}] = \mathbb{E}[\mathbb{E}_{X_T} Y; A \cap \{T < \infty\}] \qquad (\star)$$

Observe that another way of writing the statement is

$$\mathbb{E}[1_{\{T<\infty\}} Y \circ \theta_T \mid \mathcal{F}_T] = 1_{\{T<\infty\}} \mathbb{E}_{X_T} Y.$$ □

Why is $x \mapsto \mathbb{E}_x Y$ a $(S, \mathcal{S})$-measurable function? If Y is cylindrical, the result is clear. Then by Monotone Class Theorem, the result extends to all monotone Y.

Now to prove $(\star)$, we enumerate the possibilities for T. Therefore

$$\begin{aligned} LHS &= \sum_{k=0}^{\infty} \mathbb{E}[Y \circ \theta_k; A \cap \{T = k\}] \\ \text{(simple markov)} \qquad &= \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{E}_{X_k} Y; A \cap \{T = k\}] \\ &= \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{E}_{X_T} Y; A \cap \{T = k\}] \\ &= \mathbb{E}[\mathbb{E}_{X_T} Y; A \cap \{T < \infty\}]. \end{aligned}$$

Corollary 11.2 (Chapman-Kolmogorov, 5.4). *When S is discrete, $\mathbb{P}_x(X_{n+k} = \xi) = \sum_{y \in S} \mathbb{P}_x(X_n = y)\mathbb{P}_y(X_k = \xi)$.*

Proof.

$$\begin{aligned} \mathbb{P}_x(X_{n+k} = \xi) &= \mathbb{E}_x[\mathbb{E}_x(1_{\{\xi\}}(X_{n+k}) \mid \mathcal{F}_n)] \\ &= \mathbb{E}_X[\mathbb{E}_{X_n} 1_{\{\xi\}}(X_k)] \\ &= \mathbb{E}_x \mathbb{P}_{X_n}(X_k = \xi) \\ &= \sum_{y \in S} \mathbb{P}_x(X_n = y)\mathbb{P}_y(X_k = \xi). \end{aligned}$$

When S is discrete, we can represent it by $\{1, \cdots\}$ and represent the transition kernel $p(x,y) = \mathbb{P}_x(X_1 = y)$ by the matrix $P = (p_{ij})_{i,j \in S}$. Let $\mathbb{P}_{xy}^{(n)} = \mathbb{P}_x(X_n = y)$. By Chapman-Kolmogorov, we see that $P^{(2)} = P^2$ (matrix multiplication). So one may say that the study of discrete markov chains is a subfield of matrix theory, but this is misguided.

For any bounded function f on S, we define $T_n f(x) = \mathbb{E}_x[f(X_n)]$ for $x \in S$. Here T_n is a linear operator on the space of bounded functions, and by Chapman-Kolmogorov $T_n \circ T_k = T_{+k}$.

Recurrence and Transience

From now on, consider markov chains on a countable state space S. For $y \in S$, let $T_y^0 = 0$ and $T_y^1 = \inf\{n \geq 1 \mid X_n = y\}$, and let $T_y^k = \inf\{n > T_y^{k-1} \mid X_n = y\}$ for $k \geq 2$ be the kth return time to y. By convention, $\inf \varnothing = \infty$.

Lecture 12: February 13

Recurrent and Transience, 5.3

Consider a Markov chain $\{X_n\}$ on a countable state space S. For $y \in S$, set $T_y^0 = 0$ and $T_y^k = \inf\{k > T_y^{k-1}, X_{T_y^{k-1}+k} = y\}$. This is the kth visit time to y. Also set $N(y) = \sum_{n=1}^\infty 1_{\{X_n=y\}} = \sum_{k=1}^\infty 1_{\{T_y^k<\infty\}}$. Write T_y for T_y^1, and for $x, y \in S$ define $f_{xy} = \mathbb{P}_x(T_y < \infty)$. This is called the **hitting time** of y when you start your Markov chain at x (we don't count 0), whereas the **entering time** means that you don't count time 0.

Theorem 12.1 (5.5). *Starting a Markov chain at x, what is the probability that the Markov chain will visit y at least k times?*

$$\mathbb{P}_x(N(y) \ge k) = \mathbb{P}_x(T_y^k < \infty) = f_{xy} f_{yy}^{k-1}.$$

Proof. We use the strong Markov property as follows. For $k \ge 2$,

$$\begin{aligned}
\mathbb{P}_x(T_y^k < \infty) &= \mathbb{P}_x(T_y < \infty, T_y^k < \infty) \\
&= \mathbb{P}_x(T_y < \infty, T_y^{k-1} \circ \theta_{T_y} < \infty) \\
&= \mathbb{E}_x[\mathbb{P}(T_y < \infty, T_y^{k-1} \circ \theta_{T_y} \mid \mathcal{F}_{T_y})] \\
&= \mathbb{E}_x[1_{\{T_y<\infty\}} \mathbb{P}_{X_{T_y}}(T_y^{k-1} < \infty)] \\
&= f_{xy} \mathbb{P}_y(T_y^{k-1} < \infty).
\end{aligned}$$

Setting $x = y$ yields $\mathbb{P}_y(T_y^{k-1} < \infty) = f_{yy}\mathbb{P}_y(T_y^{k-2} < \infty)$, so by induction $\mathbb{P}_y(T_y^{k-1} < \infty) = f_{yy}^{k-1}$. Using the calculation again yields

$$\mathbb{P}_x(T_y^k < \infty) = f_{xy} f_{yy}^{k-1}.$$ □

Therefore

$$\mathbb{P}_x(N(y) = \infty) = \begin{cases} f_{xy} & f_{yy} = 1 \\ 0 & f_{yy} < 1 \end{cases}$$

Definition 12.2. *A state y is said to be* **recurrent** *if $f_{yy} = 1$ and* **transient** *if $f_{yy} < 1$*

Equivalently, this says $N(y) = \infty$ (resp. $<$) a.s. on $\mathbb{P}_y$.

For random walk, we saw that recurrence could be computed in terms of summability of a certain sequence. We extend this to Markov chains.

Claim 12.3. $N(y) = \infty \iff \mathbb{E}N(y) = \infty$

Proof. Note by Fubini that $\mathbb{E}_x N(y) = \sum_{n=1}^\infty \mathbb{E}_X[1_{\{X_n=y\}}] = \sum_{n=1}^\infty p_n(x, y)$. This is known as the Green's function $G(x, y)$. (It is used to solve the Dirichlet problem.)

On the other hand, we compute

$$\begin{aligned}\mathbb{E}_x N(y) &= \sum_{k=1}^{\infty} \mathbb{E}_x[1_{\{T_y^k<\infty\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) \\ &= \sum_{k=1} f_{xy} f_{yy}^{k-1} \\ &= \begin{cases} 0, & f_{xy} = 0 \\ \frac{f_{xy}}{1-f_{yy}}, & f_{yy} < 1 \\ \infty, & else \end{cases}\end{aligned}$$

Theorem 12.4 (5.6). *(a) TFAE:*

(a) y is recurrent

(b) $\mathbb{E}_y N(y) = \infty$

(c) $G(x,y) = \infty$

(b) $\sum_{n=1}^{\infty} p_n(x,y) < \infty$ *if* $\mathbb{P}_x(N(y) = \infty) = 0$

(c) $\sum_{n=1}^{\infty} p_n(x,y) = \infty$ *if* $\mathbb{P}_x(N(y) = \infty) > 0$

Lecture 13: February 18

Discrete (countable state space S) and a Markov chain on S is irreducible if $f_{xy} > 0$ for all $x, y \in S$.

Theorem 13.1 (5.7 (6.4.3)). *x is recurrent and $f_{xy} > 0$ then y is recurrent and $f_{yx} = 1$ (so $f_{xy} = 1$)*

Proof. Let $k = \inf\{j : p_j(x,y) > 0\}$. Then take a shortest path $x \to y_1 \to \cdots \to y_{k-1} \to y$ with $p(x,y_1)p(y_1,y_2)\cdots p(y_{k-1},y) > 0$ where $y_i \neq x, y$. If $f_{yx} < 1$ then $\mathbb{P}_x(\tau_x = \infty) \geq p(x,y_1)\cdots p(y_{k-1},y)(1 - f_{yx}) > 0$. Because of Theorem 5.5, $\mathbb{P}_x(\tau_x < \infty) = f_{xx} = 1$. Hence $f_{yx} = 1$. □

For some L, $p_L(y,x) > 0$. Write $p_{L+n+k}(y,y) = p_L(y,x)p_n(x,x)p_k(x,y)$. Then

$$\begin{aligned}\sum_{m=1}^{\infty} p_m(y,y) &\geq \sum_{n=1}^{\infty} p_{L+n+k}(y,y) \\ &\geq \sum_{n=1}^{\infty} p_L(y,x)p_n(x,x)p_k(x,y) \\ &= p_L(y,x)\left[\sum_{n=1}^{\infty} p_n(x,x)\right] p_k(x,y) \\ &= \infty.\end{aligned}$$

Thus y is recurrent. Now interchange x, y.

Theorem 13.2 (5.8). *A finite Markov chain contains a recurrent state.*

Proof. Suppose no recurrent states exist. Then $\mathbb{E}_x N(y) = \frac{f_{xy}}{1-f_{xx}} < \infty$, so

$$\begin{aligned}\infty &> \sum_{y\in S} \mathbb{E}_x N(y) \\ &= \sum_{y\in S}\sum_{n=1}^{\infty} p_n(x,y) \\ &= \sum_{n=1}^{\infty}\sum_{y\in S} p_n(x,y) \\ &= \infty.\end{aligned}$$

Durrect section 6.5

We make the following definitions:

(a) μ is **stationary** for $\mathbb{P}$ if $\sum_x \mu(x)p(x,y) = \mu(y)$ (i.e. $\mu = \mu P$)

(b) If $\mu(S) = 1$, we call it a **stationary distribution**

(c) μ is **reversible** if $\mu(x)p(x,y) = \mu(y)p(y,x)$.

Note that reversible implies stationary.

Example 13.3. *Birth/death chain (Example 6.5.4)*

Here $S = \{0, \cdots\}$ and we set $\mu(x) = \prod_{k=1}^{x} \frac{p_{k-1}}{q_k}$ with $q_0 = 0$.

Theorem 13.4 (5.9 (6.5.1)). *Suppose P is irreducible. Then there exists a reversible μ iff:*

(a) $p(x,y) > 0 \implies p(y,x) > 0$

(b) For any loop $x_0, \cdots, x_n = x_0$ with $\prod_{i=1}^{n} p(x_i, x_{i-1}) > 0$ we have

$$\prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})} = 1$$

Note that the quantity appearing in the cycle condition states that the probability of traversing a path is independent of orientation.

Proof. For the direct part, assume that μ does not vanish identically. By the definition of stationarity, $\mu(x) > 0$ for all $x \in S$. Hence if there is a reversible μ, the cyclic condition holds:

$$\prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})} = \prod_{i=1}^{n} \frac{\mu(x_i)}{\mu(x_{i-1})} = 1.$$

For the converse, fix $a \in S$ and set $\mu(a) = 1$. For $x \in S$, take a path $x_0 = a, x_1, \cdots, x_n = x$ with positive probability and set

$$\mu(x) = \prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})}.$$

This is well-defined due to the cycle condition. More specifically, let w and w' be two paths from a to x. Then we are reduced to showing $p(a)p(\overline{x}) = p(\overline{a})p(x)$ where $\bar{\cdot}$ means the reverse path. But this follows from the cycle condition on $a\overline{x}$. Lastly, it follows that $\mu(x)p(x,y) = \mu(y)p(y,x)$. □

Theorem 13.5 (5.10 (6.5.2)). *If x is recurrent and $T = \inf\{n \geq 1 : X_n = x\}$ then*

$$\begin{aligned} \mu_x(y) &= \mathbb{E}_x\left[\sum_{n=0}^{T-1} 1_{\{X_n = y\}}\right] \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y; T > n) \end{aligned}$$

is a stationary measure.

Proof. Observe that $\mu_x(Y)$ is the expected number of visits to y before returning to x. Then

$$\begin{aligned} \sum_z \mu_x(z)p(z,y) &= \sum_z \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = z; T > n)p(z,y) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n+1} = y; T \geq n+1) \\ &= \sum_{k=1}^{\infty} \mathbb{P}_x(X_k = y; T \geq k) \\ &= \mathbb{E}_x\left[\sum_{k=1}^{T} 1_{\{X_k = y\}}\right] \\ &= \mu_x(y). \end{aligned}$$

Note that we need to show that μ is finite.

Lecture 14: February 20

Going over the midterm problems. Problem 2:

ξ_i indep with $\mathbb{P}(0 \le \xi_i \le 1) = 1$ and $S_n = \sum_{k=1}^n \xi_k$. Show that

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > \lambda) \le 2e^{-2\lambda^2/n}.$$

Proof. Observe that $S_n - \mathbb{E}S_n = \sum_{k=1}^n(\xi_k - \mathbb{E}\xi_k)$ and $X_i = \sum_{k=1}^i(\xi_k - \mathbb{E}\xi_k)$ is a martingale. But we have the bounds

$$-\mathbb{E}\xi_{i+1} \le X_{i+1} - X_i = \xi_{i+1} - \mathbb{E}\xi_{i+1} \le 1 - \mathbb{E}\xi_{i+1}.$$

Since $0, 1$ are predictable, we may apply Azuma-Hoeffding II to obtain $\mathbb{P}(|X_n| > \lambda) \le 2e^{-2\lambda^2/n}$. □

Note that one may also apply the general setup with $S_n = f(\xi_1, \cdots, \xi_n$ and $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$ (which is Lipschitz with $C = 1$).

For problem 3:

(a) X_n = largest number shown on first n rolls of a die. Let ξ_k be the number shown in roll k. Then $X_n = \max_{1\le k\le n} \xi_k$ and $X_{n+1} = \max\{X_n, \xi_{n+1}\}$ is a Markov chain. *Note: This is different than the homework problem with simple random walk.*

(b) Y_n = number of 5's you get in the first n rolls. Here $Y_{n+1} = Y_n + 1_{\{\xi_{n+1}=5\}}$, which again has the desired form.

(c) $Z_n = \inf\{k \ge 0 : \xi_{n+k} = 5\}$ Here $Z_{n+1} = Z_n - 1$ if $Z_n > 0$ and otherwise it is the number of rolls to see a 5.

The first two are finite Markov chains, and the third has a countably infinite state space.

Recurrence and transience: Note that the first two chains are not even irreducible.

Corollary 14.1 (5.11; derived from the following Theorem 5.12). *Any finite irreducible Markov chain has a unique* **stationary distribution**.

Proof. Any finite irreducible Markov chain is recurrent by Theorems 5.7 and 5.8. By Theorem 5.10 we have existence, and the following result gives existence. □

Theorem 14.2 (5.12). *If P is irreducible and recurrent, then the stationary measure is unique up to a constant multiple.*

Proof. Suppose ν is any stationary measure of $\mathbb{P}$ and $a \in S$. Let μ_a be the stationary measure constructed in Theorem 5.10, i.e.

$$\mu_a(x) = \mathbb{E}_a[\sum_{n=0}^{T_a-1} 1_{\{X_n=x\}}] = \sum_{n=0}^{\infty} \mathbb{P}(X_n = x; T_a > n).$$

This follows by Fubini (it is related to the Green's function).

For $x \neq a$, we compute

$$\begin{aligned}
\nu(x) &= \sum_{y\in S} \nu(y)\mathbb{P}(y,x) \\
&= \nu(a)\mathbb{P}(a,x) + \sum_{y\in S, y\neq a} \nu(y)\mathbb{P}(y,x) \\
&= \nu(a)\mathbb{P}(a,x) + \sum_{y\in S, y\neq a} \left(\nu(a)\mathbb{P}(a,y) + \sum_{y_2\in S, y_2\neq a} \nu(y_2)\mathbb{P}(y_2,y)\right)\mathbb{P}(y,x) \\
&= \nu(a)\left(\mathbb{P}(a,x) + \sum_{y\in S, y\neq a} \mathbb{P}(a,y)\mathbb{P}(y,x)\right) + \sum_{y\in S, y\neq a}\sum_{y_2\in S, y_2\neq a} \nu(y_2)\mathbb{P}(y_2,y)\mathbb{P}(y,x) \\
&= \cdots \\
&\geq \nu(a)\left(\mathbb{P}(a,x) + \sum_{y\in S, y\neq a} \mathbb{P}(a,y)\mathbb{P}(y,x) + \sum_{y\in S, y\neq a}\sum_{y_2\in S, y_2\neq a} \mathbb{P}(a,y_2)\mathbb{P}(y_2,y)\mathbb{P}(y,x)\right) \\
&+ \sum_{y\in S, y\neq a}\sum_{y_2\in S, y_2\neq a}\sum_{y_3\in S, y_3\neq a} \mathbb{P}(a,y_3)\mathbb{P}(y_3,y_2)\mathbb{P}(y_2,y)\mathbb{P}(y,x) \\
&= \nu(a)\left(\mathbb{P}_a(X_1 = x) + \mathbb{P}_a(X_2 = x, 2 < T_a) + \mathbb{P}_a(X_3 = x, 3 < T_a) + \cdots\right).
\end{aligned}$$

Comparing with out earlier expression for $\mu_a(x)$, it follows that $\nu(x) \geq \nu(a)\mu_a(x)$ (we showed it for $x \neq a$, but for $x = a$ it is clear). Since ν is stationary, we have

$$\begin{aligned}
\nu(a) &= \sum_x \nu(x)\mathbb{P}(x,a) \\
&\geq \sum_x \nu(a)\mu_a(x)\mathbb{P}(x,a) \\
&= \nu(a)\mu_a(a) = \nu(a).
\end{aligned}$$

Thus $\nu(x) = \nu(a)\mu_a(x)$. To be precise, this works if $\mathbb{P}(x,a) > 0$. In general we have only $\mathbb{P}^n(x,a) > 0$ for some n. So we apply the above with $\nu = \nu\mathbb{P}^n$. □

Lecture 15: February 23

At the end of last lecture, we prove that if the Markov chain is irreducible and recurrent, then the stationary distribution exists and is unique up to constants. Recall that **stationary distribution** refers to a probability measure, whereas a **stationary measure** could have infinite mass.

Theorem 15.1 (5.13). *If there is a stationary distribution π then for state y with $\pi(y) > 0$ is recurrent.*

Proof. Recall that y is recurrent iff $\sum_{n=1}^{\infty} p_n(y,y) = \infty$ and that $\pi(y) = \sum_x \pi(x)p_n(x,y)$ and that $N(y) = \sum_{n=1}^{\infty} 1_{\{X_n=y\}}$ is the number of visits to y and that $\mathbb{E}_x N(y) = f_{xy} \sum_{n\geq 0} f_{yy}^n$. Consequently

$$\begin{aligned} \infty &= \sum_{n=1}^{\infty} \pi(y) \\ \text{(Fubini)} \qquad &= \sum_x \pi(x) \sum_n \text{ı}_n(x,y) \\ &= \sum_x \pi(x)\mathbb{E}_x N(y) \\ &= \sum_x \pi(x) f_{xy} \sum_{n\geq 0} f_{yy}^n. \end{aligned}$$

Suppose for contradiction that $f_{yy} < 1$. Then the right side is bounded by $\sum_x \frac{\pi(x)}{1-f_{yy}} = \frac{1}{1-f_{yy}}$ (since π is a probability measure). Thus we must have $f_{yy} = 1$ and the result follows. □

Theorem 15.2 (5.14). *If P is irreducible and recurrent, then $|mP$ has a stationary distribution π iff $\mathbb{E}_x T_x < \infty$ for some (and hence all) $x \in S$. In this case $\pi(x) = \frac{1}{\mathbb{E}_x T_x}$ for any $x \in S$.*

Proof. By Theorems 5.10 and 5.12, $\mathbb{P}$ has a stationary measure unique up to a constant multiple. In fact by THeorem 5.10, fix $x \in S$. Then

$$\mu_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} 1_{\{X_n=y\}} \right]$$

gives a stationary measure. Now its total mass is

$$\begin{aligned} \sum_{y\in S} \mu_x(y) &= \sum_{y\in S} \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} 1_{\{X_n=y\}} \right] \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \sum_{y\in S} 1_{\{X_n=y\}} \right] \\ &= \mathbb{E}_x T_x. \end{aligned}$$

Thus we have a stationary distribution iff $\mathbb{E}_x T_x < \infty$ for all x.

We compute the stationary distribution

$$\begin{aligned} \pi(y) &= \frac{\mu_x(y)}{\sum_{y\in S} \mu_x(y)} = \frac{\mu_x(y)}{\mathbb{E}_x T_x} \\ \pi(x) &= \frac{1}{\mathbb{E}_x T_x}. \end{aligned}$$

By uniqueness of the stationary distribution, the claim follows. □

Definition 15.3. *We say that a state $x \in S$ is* **positive recurrrent** *if $\mathbb{E}_x T_x < \infty$. A recurrent state x is said to be* **null recurrent** *if $\mathbb{E}_x T_, = \infty$.*

Note that any finite Markov chain is positive recurrent, whereas SRW on $\mathbb{Z}$ is null recurrent. Indeed, the stationary distribution is counting measure.

Example 15.4. *Renewal chain*

We start with $p \in (0,1)$ (we considered $p = \frac{1}{6}$). Here $S = \{0, 1, \cdots\}$. For $k > 0$, we set $p_{k,k-1} = 1$ and $p_{0,k}$ is a geometric distribution $\sum q^{k-1}p$. Let's find the stationary measure μ_k.

From the equation $\mu_i = \sum_k \mu_k p_{k,i}$ we see that

$$\begin{aligned} \mu_0 &= \mu_1 \\ \mu_i &= \mu_0 q^{i-1} p + \mu_{i+1} \\ \mu_{i+1} = \mu_i - \mu_i q^{i-1} p. \end{aligned}$$

Suppose $\mu_k = \mu_0 q^{k-1}$. Then we have $\mu_{k+1} = \mu_0 q^{k-1} - \mu_0 q^{k-1} p = \mu_0 q^{k-1}$. Consequently the total mass is

$$\mu(S) = \mu_0 + \mu_0 + \mu_0 q + \cdots = \mu_0 \left(1 + \frac{1}{1-q}\right) = \frac{p+1}{p}\mu_0.$$

Consequently the stationary distribution is

$$\pi(k) = \frac{pq^{k-1}}{p+1}, \qquad k \geq 1.$$

Thus the Markov chain is positive recurrent. Moreover, $\mathbb{E}_k T_k = \frac{p+1}{pq^{k-1}}$ for $k \geq 1$.

The same calculation generalizes to all renewal chains (i.e., for any probability distribution out of state 0).

Time Reversal

Suppose X_n is a Markov chain with stationary measure μ. Run the chain X_n with X_0 distributed according to μ. Fix $N \geq 1$ and consider the **time reversal** Markov chain $Y_k = X_{N-k}$ for $0 \leq k \leq N$. Note that $Y_0 = X_n$ which has distribution μ. Is Y_k a Markov chain, and if so, what is its transition probability?

Lecture 16: February 25

Stationary measure and time reversal Markov chain

Let X_n be a Markov chain with stationary measure μ (not necessarily a probability measure). Run X_n with initial 'distribution' μ. What does it mean to run a Markov chain with a non-probability measure μ? Think of putting mass $\mu(x)$ at state x, then let the mass run independently according to $\mathbb{P}$. (Note: see Durrett comment immediately preceeding Theorem 6.1.1).

Then $\mathbb{P}(X_n = y)$ = the expected mass you find at state y at time n.

Example 16.1. *Run SRW on $\mathbb{Z}$, start with the counting measure μ on $\mathbb{Z}$. Then $\mathbb{P}(X_n = y)$ is the expected number of particles in state y at time n.*

Find time N, run stationary Markov chain X_n backward. Thus we set $Y_k = X_{N-k}$ for $0 \le k \le n$. Note $Y_0 = X_N$ as distribution μ.

Claim 16.2. *$\{Y_k : 0 \le k \le N\}$ is a Markov chain.*

Proof sketch:.

$$\begin{aligned}
\mathbb{P}(Y_{k+1} = y \mid Y_k = x; Y_j = x_j \text{ for } 0 \le j \le k-1) &= \frac{\mathbb{P}(Y_{k+1} = y, Y_k = x, Y_j = x_j, 0 \le j \le k-1)}{\mathbb{P}(Y_k = x, Y_j = x_j, 0 \le j \le k-1)} \\
&= \frac{\mathbb{P}(X_{N-(k+1)} = y, X_{N-k=x}, X_{N-j} = x_j, 0 \le j \le k-1)}{\mathbb{P}(X_{N-k=x}, X_{N-j} = x_j, 0 \le j \le k-1)} \\
&= \frac{\mu(y)p(y,x)p(x,x_{k-1})\cdots p(x_1,x_0)}{\mu(x)p(x,x_{k-1})\cdots p(x_1,x_0)} \\
&= \frac{\mu(y)p(y,x)}{\mu(x)} \\
&= \mathbb{P}(Y_{k+1} = y \mid Y_k = x) \\
&= q(x,y).
\end{aligned}$$

Note that is not rigorous, since $\mathbb{P}(Y_{k+1} = y, Y_k = x, \cdots)$ is not a probability at all so the theory doesn't apply. But now that we have 'guessed' the transition probability, we can *a posteriori* use $q(x,y)$ to define the Markov chain.

Claim 16.3. *$q(x,y)$ is a transition probability:*

(a) $q(x,y) \ge 0$

(b) $\forall x$, $\sum_y q(x,y) = \frac{\sum_y \mu(y)p(y,x)}{\mu(x)} = 1$

Let Y be the Markov chain wth transition probability $q(x,y)$ and observe that Y_n has stationary measure μ.

$$\sum \mu(x)q(x,y) = \sum \mu(y)p(y,x) = \mu(y).$$

Now we make the 'time rerversal' rigorous. We want to show that the joint distribution $(X_0, X_1) \overset{d}{=} (Y_1, Y_0)$. This is equivalent to showing

$$\mathbb{E}_\mu[f(X_0)g(X_1)] = \mathbb{E}_\mu[g(Y_1)f(Y_0)].$$

Just calculate it:

$$\sum_y \sum_x \mu(x) f(x) \mathbb{P}(x,y) g(y) = \sum_x \sum_y \mu(y) g(y) q(y,x) f(x).$$

Similarly,

$$\mathbb{E}_\mu[\prod_{i=0}^n f_i(X_i)] = \mathbb{E}_\mu[\prod_{i=0}^n f_i(Y_{n-i})].$$

Time reversal is always taken w.r.t. a measure, which we think of as the distribution at time N.

Duals

Definition 16.4. *Suppose* $(X_0, \cdots, X_n) \overset{=}{d} (Y_n, \cdots, Y_0)$ *for all* n*, where* X_n *and* Y_n *are Markov chains. Then* X_n *and* Y_n *are* **dual**.

Definition 16.5. *The* **transition operator** *associated to a transition kernel* $p(x,y)$ *is*

$$Pf(x) = \sum_y p(x,y) f(y) = \mathbb{E}_x f(X_1),$$

To be precise, we take the transition operator to act on the space of bounded measurable function $\mathcal{B}_b(S)$. Then

$$\langle Pf, g \rangle_{L^2(S,\mu)} = \langle f, Qg \rangle_{L^2(S,\mu)}$$

which means $P^* = Q$.

Observe that if P, Q are dual in this sense, then $\mu(x) q(x,y) = \mu(y) p(y,x)$. Summing over x, it follows that μ is stationary for Q and vice versa. Thus in the case of **conservative** Markov chains, existence of duals is equivalent to the existence of a stationary measure.

Definition 16.6. *If* $q(x,y) = p(x,y)$ *for all* x, y*, then* $\mu(x) p(x,y) = \mu(y) p(y,x)$*. Thus we say* μ *is a* **reversible** *measure for* X_n*.*

This is equivalent to saying that the operator P is selfadjoint.

Asymptotic Behavior, 5.5 (6.6 in Durrett)

Consider a Markov chain X_n on S. If y is transient, then $\sum_n p_n(x,y) < \infty$ for every x. Hence

$$\lim +n \to \infty p_n(x,y) = 0.$$

Assume that y is recurrent. Do the following limits exist and agree?

$$\lim_{n\to\infty} p_n(x,y) = \lim_{n\to\infty} \mathbb{P}_x(X_n = y)$$

By considering biased random walk (for the parity issues), we can construct examples where one of the limits doesn't exist. But instead we can consider the Cesaro limit:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n p_n(x,y) = \frac{1}{n} \mathbb{E}_x \sum_{k=1}^n 1_{\{X_k = y\}}.$$

The quantity $N_n(y) = \sum_{k=1}^n 1_{\{X_k=y\}}$ is the expected number of visits to y by time n.

Theorem 16.7 (5.16). *If y is recurrent, then for any $x \in S$*

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y} 1_{\{T_y < \infty\}}.$$

Use the Strong Markov property. This is essentially the Strong Law of large numbers.

Lecture 17: February 27

Theorem 17.1 (5.16). *Suppose y is recurrent. Then for any $x \in S$,*

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y} 1_{\{T_y < \infty\}},$$

$\mathbb{P}_x$ almost surely.

Proof. First take $x = y$. Define $T_y^{(0)} = 0, T_y^{(1)} = T_y$ and generally $T_y^{(k)} = T_y \circ \theta_{T_y^{(k-1)}} + T_y^{(k-1)}$ to be the kth returning time to y. Let $T_k = T_y^{(k)} - T_y^{(k-1)}$ be the inter-returning time. By the strong Markov property of Markov chains, $\{T_k\}$ are i.i.d. under $\mathbb{P}_y$.

$$0 \underbrace{\to}_{T_1} T_y^{(1)} \underbrace{\to}_{T_2} T_y^{(2)} \underbrace{\to}_{T_3} T_y^{(3)} \to \cdots$$

By the Strong Law,

$$\frac{T_y^{(k)}}{k} = \frac{\sum_{j=1}^k T_j}{k} \to \mathbb{E}_y T_y.$$

Note that we have used a truncation argument to handle the case with $\mathbb{E}_y T_y = \infty$. The number of visits to y by time n is $T_y^{(N_n(y))} \le n < T_y^{(N_n(y)+1)}$. Therefore

$$\begin{aligned} \frac{T_y^{(N_n(y))}}{N_n(y)} &\le \frac{n}{N_n(y)} \\ &\le \left(\frac{T_y^{(N_n(y)+1)}}{N_n(y)+1}\right)\left(\frac{N_n(y)+1}{N_n(y)}\right). \end{aligned}$$

As y is recurrent, $N_n(y) \to \infty$ as $n \to \infty$. By sandwiching, we have $\lim_{n\to\infty} \frac{n}{N_n(y)} = \mathbb{E}_y T_y$, at least $\mathbb{P}_y$-a.s. (after interchanging a.s. with a countable limit).

Now suppose $x \neq y$. On $T_y = \infty$, it is clear that $N_n(y) = 0$ and the result follows. Conditioned on $\{T_y < \infty\}$, we have $\{T_k\}_{k=2}^\infty$ are iid. They also have the same disitribution as T_y under $\mathbb{P}_y$. More precisely,

$$\mathbb{P}_x(T_k = j \mid T_y < \infty) = \mathbb{P}_y(T_y = j), \qquad k \ge 2.$$

Observe that

$$\frac{T_y^{(k)}}{k} = \frac{T_1}{k} + \frac{\sum_{j=2}^k T_j}{k} \to \mathbb{E}_y T_y$$

on $\{T_y < \infty\}$. Then proceed as above to get the result. □

Corollary 17.2 (5.17).

$$\mathbb{E}_x \frac{N_n(y)}{n} = \frac{1}{n}\sum_{k=1}^n p_k(x,y) \to \frac{f_{xy}}{\mathbb{E}_y T_y}$$

Recall that $f_{xy} = \mathbb{P}_x(T_y < \infty)$. This result is obtained by taking expectations of both sides. Note that we dnoo't need to assume the Markov chain is irreducible, and we don't need to assume that y is positive recurrent.

We would like to determine if $p_n(x,y)$ converges. We already saw that it does not for SRW, but this is not a convincing example since we don't have a stationary distribution. Here is a better counterexample:

Example 17.3. *Take* $S = \{1,2\}$ *and*

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let $x \to y$ denote $\mathbb{P}_x(T_y < \infty) > 0$.

Definition 17.4. *Suppose* $x \to x$. *The* **period** *of* x *is*

$$dx = \gcd(\{k \geq 1 : p_k(x,x) > 0\}).$$

Lemma 17.5 (5.18). *If* $x \leftrightarrow y$ *then* $dx = dy$.

Proof. Let $K, L \geq 1$ be integers for which $p_K(x,y) > 0$ and $P_L(y,x) > 0$. Then by Chapman-Kolmogorv,

$$P_{K+L}(x,x) \geq P_K(x,y)P_L(y,x) > 0$$
$$P_{K+L}(y,y) \geq P_L(y,x)P_K(x,y) > 0.$$

Hence $dx \mid K+L$ and $dy \mid K+L$. For any n such that $p_n(x,x) > 0$, we have $p_{K+L+n}(y,y) > 0$. Indeed, by Chapman-Kolmogorv

$$p_{K+L+n}(y,y) \geq p_L(y,x)p_n(x,x)p_K(x,y) > 0.$$

Consequently $dy \mid K+L+n$ whereupon $dy \mid n$ and thus $dy \mid dx$. By symmetry, $dx = dy$. □

This is called the **class property**. We can partition S into equivalence classes under the relation $x \leftrightarrow y$, and this result states that the period is constant in each equivalence class. These equivalence classes are precisely the strongly connected components of the Markov chain.

Definition 17.6. *A Markov chain is* **aperiodic** *if every stsate has period 1.*

Theorem 17.7 (Convergence Theorem, 5.18). *Suppose* P *is irreducible and aperiodic with stationary distribution* π. *Then*

$$\lim_{n\to\infty} p_n(x,y) = \pi(y) = \frac{1}{\mathbb{E}_y T_y}, \qquad \textit{for all } x, y.$$

To prove this, we use a coupling argument: we run two Markov chains from different points. Then we show they meet at a finite time, and bound the total variation distance by the probability they meet after n.

Lecture 18: March 2

Applications of Ergodicity

We have $\mu_{n+1} = \mu_n \mathbb{P}$. Thus if μ_n converges to some limiting distribution π, then $\pi = \pi\mathbb{P}$. This is like the central limit theorem: you start out with a bunch of randomness, but it dies down and gives you something predictable.

Example 18.1 (Service system). *Let i be the number of customers. Then we have a markov chain: $i \overset{.05}{\to} i+1$ with $i \overset{.1}{\to} i-1$.*

The time average and the ensemble average both tend to the same stationary probability (i.e., either sample proportion of time at which a single restaurant is in a state, or sample a bunch of restaurants at the same time).

We know that a reversible process is ergodic, but not every ergodic process is reversible.

Example 18.2 (Hastings-Metropolis=MCMC). $S = \{1, \cdots, m\}$. *Here $b(j) > 0$ and $B = \sum_{j=1}^m b(j)$ with invariant distribution $\pi(j) = b(j)/B$.*

In general its hard to compute the invariant distribution, so instead we do a simulation and look at time averages. Here MCMC means **Markov Chain Monte Carlo** method, used in Bayesian statistics.

Take an irreducible Markov chain with matrix $\mathbb{P}$. Let $p(i,j) = \mathbb{P}(i \to j)$. Choose $k \in S$ and set $n = 0, X_0 = k$. Generate r.v. X with pdf $\mathbb{P}(X = j) = p(X_n, j)$. Generate $U \sim U[0,1]$. If $U < \frac{b(x)p(x,X_n)}{b(X_n)p(X_n,x)}$ then we set $X_{n+1} = x$, else $X_{n+1} = X_n$.

This produces a new Markov chain with $\widetilde{p}(i,j) = p(i,j)\alpha(i,j)$, where

$$\alpha(i,j) = \min\left(\frac{b(j)p(j,i)}{b(i)p(i,j)}, 1\right).$$

Claim 18.3. *$\widetilde{p}$ has stationary measure π (i.e. b)*

This will show that $\pi(i)\widetilde{p}(i,j) = \pi(j)\widetilde{p}(j,i)$, which means it is reversible. This is ergodic with stationary measure π.

Proof. Suppose $b(j)p(j,i) < b(i)p(i,j)$. Then

$$\begin{aligned} \pi(i)\widetilde{p}(i,j) &= \pi(i)p(i,j)\alpha(i,j) \\ &= \pi(i)\frac{b(j)p(j,i)}{b(i)} \qquad = \pi(j)\widetilde{p}(j,i). \end{aligned}$$

The same holds in the remaining case. □

There is an art to setting this up nicely (i.e. choosing a good Markov chain).

Example 18.4 (Gibbs Sampler). *(a) Take n random points in the unit disc $\mathbb{B}^d$ such that $\|x_i - x_j\| \geq d$.*

(b) Let X_i be iid $\exp(\lambda)$ and $S = \sum_{i=1}^n X_i$. Given that $S \geq C$, find the distribution.

For the algorithm, let $X = (X_1, \cdots, X_n)$ be a random vector and $p(x)$ is the probability mass function of X. We want to find the conditional probability distribution

$$\begin{cases} \frac{p(x)}{p(X \in A)}, & x \in A \\ 0, & else \end{cases}$$

Assume we know $\mathbb{P}_i(X_i = x_i) = \mathbb{P}(X_i = x_i \mid X_j = x_j, j \neq i)$. SUppose our current state is $X_t = (x_1, \cdots, x_n)$. Pick $i \in [n]$ uniformly. If $X_t = x$, then we consider $X_{t+1} = y$ with $x_i \to x$. Then

$$\begin{aligned} p(x, y) = \frac{1}{n}\mathbb{P}(X_i = x \mid X_j = x_j, j \neq i) \\ = \frac{\mathbb{P}(y)}{n\mathbb{P}(X_j = x_j, j \neq i)}. \end{aligned}$$

Then we have

$$\begin{aligned} \alpha(i, j) = \alpha(x, y) &= \min\left(\frac{f(y)p(y, x)}{f(x)p(x, y)}, 1\right) \\ &= \begin{cases} 1, & y \in A \\ 0, & else \end{cases} \end{aligned}$$

(by checking each case).

Starting at $x = (x_1, \cdots, x_n) \in A$. Choose I uniformly, i.e. $I = 1 + [nU]$. If $I = i$, select y as before using the given marginals. If $y \in A$, next state is y otherwise next state is x. In general you start with a $U[0, 1]$.

For example 1, use uniform distribution. We have n points in the unit disc that are spaced about by d. Pick a new one uniformly, and if it is too close toss it. You can quickly generate the distribution by looking at histograms.

For example 2, pick i and consider $S - X_i$. What is $\mathbb{P}(X > t + s \mid X > s)$? Since $\exp(\lambda)$ is memoryless, it is $\mathbb{P}(X > t)$: that is $e^{-\lambda t}$. If $S - X_i$ is bigger than C, then the candidate is $E \sim \exp(\lambda)$, and otherwise it is $E + (S - X_i)$.

To extend, we need survival probabilities. You can only get closed forms in certain cases (e.g. Weibull).

Lecture 19: March 4

The metropolis algorithm is useful for many situations when you want to sample from a distribution that its hard to sample from directly. For instance, in the Ising model there are a large number of configurations which must be summed over to find the normalizing constant. Instead, when we use the metropolis algorithm we only need ratios.

We have a long sequence $\{a_1, \cdots, a_N\}$. Then we compute $Z = \sum_{k=1}^{N} a_k$ to form the probability distribution $\{a_i/Z\}$. It's hard to compute Z, but we don't need to if we use metropolis.

Ergodic Theory of Markov Chains

How to construct an ergodic Markov chain with a given stationary ditribution? We start with an easy to sample Markov chain, then we produce a new one using a rejection criterion. There is also simulated annealing and coupling from the past, also known as the Propp-Wilson algorithm.

Theorem 19.1 (5.18). *Suppose P is irreducible, aperiodic, and has stationary distribution π. Then*

$$\lim_{n\to\infty} \mathbb{P}_x(X_n = y) = \lim_{n\to\infty} p_n(x, y) = \pi(y)$$

Definition 19.2. *A* **coupling** *of two processes X, Y consists of a process Z on the product space such that the projections $\pi_1(Z) \sim X$ and $\pi_2(Z) \sim Y$.*

If the processes meet, we say the coupling is **successful**. We need the following lemma:

Lemma 19.3 (5.19). *If $dx = 1$, there exists an N such that $p_n(x, x) > 0$ for all $n > N$.*

Proof. We use coupling, i.e. run two independent Markov chains and see when they meet. Let $S^2 = S \times S$ be the product space. We define a transition probability $\widetilde{p}$ on S^2 as follows:

$$\widetilde{p}((x_1, y_1), (x_2, y_2)) = p(x_1, x_2)p(y_1, y_2). \qquad \square$$

Proof. We break it up into several claims, each trivial to check.

Claim 19.4. *$\widetilde{p}$ is irreducible*

Proof. We wish to show that for any $x_1 \leftrightarrow x_2$ and $y_1 \leftrightarrow y_2$, there are K, L for which $p_K(x_1, x_2) > 0$ and $p_L(y_1, y_2) > 0$. Taking N from Lemma 5.19 and applying Chapman-Kolmogorov, it follows that $p_{K+L+N}((x_1, y_1), (x_2, y_2)) > 0$. $\square$

Claim 19.5. *$\widetilde{\pi}(x, y) = \pi(x)\pi(y)$ is the stationary distribution of $\widetilde{p}$*

Proof. It is clear that this is a probability distribution (by Fubini). Now we check

$$\begin{aligned} \pi(x_0, y_0) &= \pi(x_0)\pi(y_0) \\ \text{(stationarity of } \pi) \qquad &= \sum_{x\in S} \pi(x)p(x, x_0) \sum_{y\in S} \pi(y)p(y, y_0) \\ \text{(Fubini)} \qquad &= \sum_{(x,y)\in S^2} \pi(x, y)\widetilde{p}((x, y), (x_0, y_0)) \end{aligned}$$

Hence all states of $\widetilde{p}$ are positive recurrent.

Claim 19.6. *The coupling is successful*

Proof. We let $T = \inf\{n \geq 1 \colon X_n = Y_n\}$. Note that the two Markov chains have arbitrary initial distributions that are irrelevant. Since the Markov chain $\widetilde{p}$ is positive recurrent, it follows that it will hit any point in finite time. In particular, it will hit the diagonal so $T < \infty$ and we are done. □

Claim 19.7. $\mathbb{P}(X_n = y, n \geq T) = \mathbb{P}(Y_n = y, n \geq T)$

Intuitively this should be a consequence of the Strong Markov property, but there is a slight issue because $n - T$ is not a stopping time.

Proof. We get around the measurability issue by summing over all cases. Consider the filtration $\mathcal{F}_k = \sigma(\{Z_j\}_{j \leq k})$

$$\begin{aligned}
\mathbb{P}(X_n = y, n \geq T) &= \sum_{k=1}^{n} \mathbb{P}(X_n = y, T = k) \\
&= \sum_{k=1}^{n} \mathbb{E}\mathbb{P}(X_n = y, T = k \mid \mathcal{F}_k) \\
&= \mathbb{E}[1_{\{T=k\}} \mathbb{P}(X_n = y \mid \mathcal{F}_k)] \\
\text{(for any choice of } Y_k) \qquad &= \sum_{k=1}^{n} \mathbb{E}[1_{\{T=k\}} \mathbb{P}_{(X_k, Y_k)}(X_{n-k} = y)] \\
(\star) \qquad &= \sum_{k=1}^{n} \mathbb{E}[1_{\{T=k\}} \mathbb{P}_{(X_k, Y_k)}(Y_{n-k} = y)] \\
&= \mathbb{P}(Y_n = y, n \geq T).
\end{aligned}$$

$(\star)$: since the distributions of X, Y are the same after they meet.

Now we use the triangle inequality:

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

This gives the total variation bound with $\mathbb{P}(T > n) \to 0$. □

Lecture 20: March 6

The proof from last time used the following lemma which we didn't have time to prove.

Lemma 20.1. *If $dx = 1$ then there is an integer N such that $p_k(x,x) > 0$ for any $k \geq N$.*

Proof. Since $dx = 1$, there are relatively prime loops of size n_1, n_2. Hence by Bezout's Lemma, there are integers a, b for which $an_1 + bn_2 = 1$. Without loss assume $a > 0$ and take $N = -bn_2^2$. For $k \geq N$, write $k = jn_2 + r$. Then we have

$$\begin{aligned} k &= jn_2 + arn_1 + brn_2 \\ &= arn_1 + (j + br)n_2, \end{aligned}$$

where both coefficients are non-negative. Indeed, $j = \left\lfloor \frac{k}{n_2} \right\rfloor \geq \left\lfloor \frac{-bn_2^2}{n_2} \right\rfloor = -bn_2$. □

Convergence Theorem

We have the following assumptions:

(a) **irreducible**

(b) aperiodic (i.e. $dx = 1$)

(c) existence of stationary distribution π

Then $p_n(x,y) \to \pi(y) = \frac{1}{\mathbb{E}_y T_y}$ for all y.

Note that the hypothesis of irreducibility is not actually a restriction. Inside any irreducible component, the period is constant. Say the period is d. Then we decompose the state space S into strongly connected components $C_r(x)$, consisting of states y which travel to x in $d\mathbb{N}_0 + r$ steps. Then the chain p^d is irreducible and aperiodic on each component, and moreover

$$\lim_{n\to\infty} p_{nd}(x,y) = \frac{\pi(y)}{\pi(C_r(x))}.$$

This corresponds to strong mixing. The statement of weak mixing is the following.

Let $N_n(y)$ be the number of visits to y by time n. Then

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \pi(y), \qquad p_x - a.s.$$

This is a statement about Cesaro convergence, which is weaker than the type of convergence in the theorem. Here we speed up the Markov chain by a factor of d, i.e. $p \mapsto p^d$, to obtain $\widetilde{\pi}(y) = d\pi(y)$.

What happens when we drop the assumption of a stationary distribution (and replace it with **recurrence**)? In the null recurrent case, things converge to 0. More generally,

$$p_n(x,y) \to \frac{1}{\mathbb{E}_y T_y}.$$

So far we have only discussed Markov chains on finite or countable state spaces. That is $\{X_n; n = 0, \cdots\}$ on S. Now we an consider a general state space, e.g. $S = \mathbb{R}^n$. For example, one can move uniformly to a point with distance 1 at each step. This is the theory of **general Markov chains** which is relatively recent (e.g., existence of such Markov chains).

Lecture 21: March 9

Consider
$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{P}_x(X_k = y) = \frac{1}{n}\sum_{k=1}^{n}p_k(x,y) \to \frac{f_{xy}}{\mathbb{E}_y T_y}.$$
If π is the **stationary** distribution of a Markov chain, then $\{X_0, X_1, \cdots\} \overset{d}{=} \{X_k, X_{k+1}, \cdots\}$ under $\mathbb{P}_\pi$ for any $k \geq 1$.

Ergodic Theorems, Ch. 6

Definition 21.1. *The process $\{X_n\}$ is* **stationary** *if for any $k \geq 1$, $\{X_n\} \overset{d}{=} \{X_{n+k}\}$.*

A basic fact about stationary sequences, called the ergodic theorem, asserts that if $\mathbb{E}|f(X_0)| < \infty$, then
$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(X_k)$$
exists a.s. and in $\mathcal{L}^1$. What is the limit? (This is related to the Law of Large Numbers.)

Example 21.2. *(a) $\{X_i\}$ iid is clearly stationary.*

(b) Take a Markov chain with $\mathbb{P}_\pi$ (stationary distribution).

(c) Rotation of the circle $\mathbb{S}^1 = [0,1) = \Omega$. Take $\mathbb{P}$ to be Lebesgue measure. Fix any $\theta \in \Omega$ and take $X_n(\omega) = \omega + n\theta \mod 1$.

Fact 21.3. *(a) If $\{X_0, \ldots\}$ is a stationary sequence, then so is $\{g(X_0), g(X_1), \ldots\}$.*

(b) Bernoulli shift

We can see the second as a special case of the first. take $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}[0,1)$ and $\mathbb{P}$ is Lebesgue. Take $X_0(\omega) = \omega$ and $X_n(\omega) = 2X_{n-1}(\omega) \mod 1$. Then $\{X_0, \ldots\}$ is stationary. Then we may write
$$X_0(\omega) = \sum_{n=1}^{\infty}\xi_n 2^{-n}, \qquad \xi_n \in \{0,1\} iid Bernoulli.$$
Next, $X_i(\omega) = \sum_{n=1}^{\infty}\xi_{n+i}2^{-n}$. Hence X_i corresponds to $(\xi_{i+1}, \cdots)$.

Definition 21.4. *A measurable map $\varphi: \Omega \to \Omega$ is said to be* **measure preserving** *if*
$$\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A), \qquad \forall A \in \mathcal{F}.$$

For the Bernoulli shift, $\Omega = \{0,1\}^{\mathbb{N}}$ and $\mathbb{P} = \bigotimes_{n=1}^{\infty}\frac{\delta_0+\delta_1}{2}$. Then the map
$$\varphi: (x_1, \ldots) \mapsto (x_2, \ldots)$$
is measure preserving. Indeed, it suffices to check it on cylinder sets.

Given a measure preserving map φ on Ω, for any r.v. $X \in \mathcal{F}$ define $X_0 = X$ and $X_n = X(\varphi^n)$. Then $\{X_0, \ldots\}$ is stationary. Conversely, given a stationary sequence $\{X_n\}$ taking values in $(S, \mathcal{S})$ there is a measure $\mathbb{P}$ on $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$ such that $\{\omega_n\} \overset{d}{=} \{X_n\}$ (use Kolmogorov Extension).

Fact 21.5. *Any stationary sequence* $\{X_n\}_{n\in\mathbb{N}}$ *can be embedded into a stationary sequence* $\{X_n\}_{n\in\mathbb{Z}}$.

Indeed, take $(X_{-n},\dots,X_n)\overset{=}{d}(X_0,\cdots,X_{2n})$.

From now on, take $(\Omega,\mathcal{F},\mathbb{P})$ with a measure preserving map φ.

Definition 21.6. *(a) We say that* $A\in\mathcal{F}$ *is* **invariant** *if* $\varphi^{-1}(A)=A$ *up to* $\mathbb{P}$*-a.s. equality. More precisely, we mean*

$$\mathbb{P}(A\Delta\varphi^{-1}(A))=0.$$

(b) We say $B\in\mathcal{F}$ *is strictly invariant if* $\varphi^{-1}(B)=B$.

Claim 21.7. *The event* A *is invariant if and only if there is a strictly invariant* B *for which* $\mathbb{P}(A\Delta B)=0$.

Let $\mathcal{I}$ denote the class of invariant sets; it is a σ-field. Moreover an r.v. $\xi\in I$ if and only if $\xi\circ\varphi=\xi$ a.s.

Definition 21.8. *A measure preserving map* φ *is* **ergodic** *if* $\mathcal{I}$ *is trivial.*

Theorem 21.9 (Birkhoff Ergodic Theorem). *For* $X\in\mathcal{L}^1$*, we have the convergence*

$$\frac{1}{n}\sum_{k=1}^{n}X(\varphi^k(\omega))\to\mathbb{E}(X\mid\mathcal{I})$$

a.s. and in $\mathcal{L}^1$.

In the context of stationary Markov chains, we have the following restatement.

Theorem 21.10. *Suppose* $\{X_n\}$ *is a Markov chain with stationary distribution* $\pi>0$*. Then* $\{X_n\}$ *is ergodic iff* X *is irreducible.*

Note that we can always ensure $\pi>0$ by throwing away to states with $\pi(x)=0$.

Observe that $\mathcal{I}\subset\mathcal{T}$, the tail σ-field. Thus if $\mathcal{T}=\varnothing$ we always have ergodicity, but by considering a periodic irreducible Markov chain we obtain an example in which $\mathcal{I}=\varnothing$ but not $\mathcal{T}$.

Lecture 22: March 11

Theorem 22.1. *Assume a Markov chain X_n has stationary distribution π with $\pi(x) > 0$ for every $x \in S$. Then X_n is ergodic if and only if X_n is irreducible under $\mathbb{P}_\pi$.*

Proof. Note all states are positive recurrent. Moreover since $\pi > 0$, the state space is countable. Define an equivalence relation with $x \sim y$ if the Markov chain can transition from x to y, and index the equivalence classes as R_i. If $X_0 \in R_i$, then $X_n \in R_i$ for all $n \geq 1$ since $\pi = \pi P$ and $\pi > 0$. Let $A_i = \{X_0 \in R_i\}$; it follows that $\theta^{-1} A_i = A_i$, so $A_i \in \mathcal{I}$. Thus if X_n is ergodic, we must have $\mathbb{P}_\pi(A_i) \in \{0,1\}$ so there is one equivalence class and X_n is irreducible.

Conversely, suppose X_n is irreducible and consider any invariant event $A \in \mathcal{I}$. Then $A = \theta^{-n} A$ almost surely. Let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Then

$$\begin{aligned}\mathbb{E}_\pi(1_A \mid \mathcal{F}_n) &= \mathbb{E}_\pi(1_A \circ \theta^n \mid \mathcal{F}_n)\\ &= \mathbb{E}_{X_n}(1_A)\\ &= \varphi(X_n).\end{aligned}$$

Here $\varphi(x) = \mathbb{P}_x(A)$. Therefore

$$\mathbb{E}_\pi(1_A \mid \mathcal{F}_\infty) = \lim_{n\to\infty} \mathbb{E}_\pi(1_A \mid \mathcal{F}_n) = \lim_{n\to\infty} \varphi(X_n).$$

For all $y \in S$, the right sides $\varphi(X_n)$ takes value $\varphi(y)$ i.o. a.s.

Here $\varphi(y) = 1_A$ means that φ is deterministic with value 0 or 1, so $\mathbb{P}_\varphi(A)$ is 0 or 1. □

Remark 22.2. *Typically $\mathcal{I} \subsetneq \mathcal{T}$. For example, consider an irreducible Markov chain with period $d \geq 2$. Fix x and consider $C_0(x), C_1(x), \cdots, C_{d-1}(x)$. Then $\mathcal{T} = \sigma(\{X_0 \in C_0(x)\}, \{X_0 \in C_1(x)\}, \ldots, \{X_0 \in C_{d-1}(x)\})$.*

Note that any invariant set is in the tail σ-algebra, because $A = \theta^{-n} A \in \sigma(X_n, X_{n+1}, \ldots)$ for every n.

Lemma 22.3 (Maximal ergodic lemma). *Consider φ measure preserving on $(\Omega, \mathcal{F}, \mathbb{P})$. Given a r.v. X, define $X_n = X \circ \varphi^n$. Set $S_n = \sum_{k=0}^{n-1} X_k$ and $M_n = \max\{0, S_1, \cdots, S_n\}$. Then $\mathbb{E}[X; m_n > 0] \geq 0$.*

Proof. First we claim $X(\omega) \geq M_k(\omega) - M_k(\varphi\omega)$ on $\{M_k > 0\}$. For $j \leq k$, clearly $M_k \circ \varphi \geq S_j \circ \varphi$. Then $X + M_k \circ \varphi \geq X + S_j \circ \varphi = S_{j+1}$ and so $X(\omega) \geq S_{j+1}(\omega) - M_k(\varphi\omega)$ for $j = 1, \ldots, k$. Trivially $X(\omega) \geq S_1(\omega) - M_k(\varphi\omega)$ since $S_1 = X$ and $M_k \geq 0$. Thus

$$\begin{aligned}\mathbb{E}[X; M_k > 0] &\geq \int_{M_k>0} [\max\{S_1, \cdots, S_k\} - M_k(\varphi)]\, d\mathbb{P}\\ &= \int_{M_k>0} (M_k - M_k \circ \varphi)\, d\mathbb{P}\\ (M_k \geq 0) \implies \quad &\geq \int M_k - M_k \circ \varphi\, d\mathbb{P} = 0,\end{aligned}$$

since φ is measure preserving. □

Theorem 22.4 (Birkhoff ergodic theorem). *For $X \in \mathcal{L}^1$, we have the convergence*

$$\frac{1}{n}\sum_{k=0}^{n-1} X \circ \varphi^k \to \mathbb{E}[X \mid \mathcal{I}]$$

a.s. and in $\mathcal{L}^1$.

If $X \in \mathcal{L}^p$ for $p > 1$, the convergence is also in $\mathcal{L}^p$.

Proof. Let $X' = X - \mathbb{E}(X \mid \mathcal{I})$. Then $X' \circ \varphi^n = X \circ \varphi^n - \mathbb{E}(X \mid \mathcal{I})$. So wlog we assume $\mathbb{E}(X \mid \mathcal{I}) = 0$ (i.e., we are shifting to assume the mean is 0).

Let $S_n = \sum_{k=0}^{n-1} X \circ \varphi^k$ and consider $\overline{X} = \overline{\lim}_{n\to\infty} \frac{S_n}{n}$. For all $\varepsilon > 0$, let $D = \{\overline{X} > \varepsilon\}$. We want to show $\mathbb{P}(D) = 0$. Since $\overline{X} \circ \varphi = \overline{X}$, we have $D \in \mathcal{I}$. Define

$$X^* = (X - \varepsilon)1_D, \qquad S_n^* = \sum_{k=0}^{n-1} X^* \circ \varphi^j = (S_n - n\varepsilon)1_D.$$

Define $F_n^* = \{M_n^* > 0\}$ which increases in n to a set F. In fact

$$\begin{aligned} F = \bigcup_{n=1}^{\infty} F_n &= \bigcup_{n=1}^{\infty} \{\frac{S_n^*}{n} > 0\} \\ &= \left\{\sup_n \frac{S_n - n\varepsilon}{n} 1_D\right\}. \end{aligned}$$

Observe that $F = \{\overline{\lim} \frac{S_n}{n} > \varepsilon\}$ and by the maximal ergodic lemma $\mathbb{E}[X^*; F_n] \geq 0$. Taking limits yields $\mathbb{E}[X^*; F] \geq 0$. Consequently

$$\begin{aligned} 0 \leq \mathbb{E}[X^*; D] &= \mathbb{E}[X - \varepsilon; D] \\ &= \mathbb{E}[X; D] - \varepsilon\mathbb{P}(D) \\ &= \mathbb{E}[\mathbb{E}(X \mid \mathcal{I}); D] - \varepsilon\mathbb{P}(D) \\ &= -\varepsilon\mathbb{P}(D). \end{aligned}$$

Hence $\mathbb{P}(D) = 0$ as desired, and by symmetry we get the same bound for $\underline{\lim}$. Thus the limit converges to 0 as desired. □

Chapter 3

Spring 2015

Lecture 1: March 30

Consider $\varphi: (\Omega, \mathbb{P}) \to (\Omega, \mathbb{P})$ measure preserving.

Theorem 1.1 (Birkhoff Ergodic Theorem). *For $X \in \mathcal{L}^1$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} X(\varphi^k \omega) \to \mathbb{E}[X \mid \mathcal{I}]$$

almost surely and in $\mathcal{L}^1$.

Proof. Shifting $X' = X - \mathbb{E}(X \mid \mathcal{I})$, we may assume without loss that $\mathbb{E}(X \mid \mathcal{I}) = 0$. Let $S_n = \sum_{k=0}^{n-1} X(\varphi^k\omega)$. We showed last time that $\lim_{n\to\infty} \frac{S_n}{n} = 0$ a.s. We now show convergence in $\mathcal{L}^1$.

For $M \geq 1$, let $X_M = X1_{\{|X| \leq M\}}$. We know

$$\frac{1}{n}\sum_{k=0}^{n-1} X_m(\varphi^k\omega)) \to \mathbb{E}(X_M \mid \mathcal{I})$$

a.s. The convergence occurs in $\mathcal{L}^1$ by bounded convergence, as $|X_M(\varphi^k)| \leq M$.

Let $X'_M = X - X_M = X1_{\{|X| \geq M\}}$. Then

$$\frac{S_n}{n} = \underbrace{\frac{1}{n}\sum_{k=0}^{n-1} X_M(\varphi^k)}_{I_n^M} + \underbrace{\frac{1}{n}\sum_{k=0}^{n-1} X'_M(\varphi^k)}_{II_n^M}$$

$$\begin{aligned}\implies A_n = \mathbb{E}\left|\frac{S_n}{n} - \mathbb{E}(X \mid \mathcal{I})\right| &\leq \mathbb{E}|I_n^M - \mathbb{E}(X_M \mid \mathcal{I})| + \mathbb{E}|II_n^M - \mathbb{E}(X_M^1 \mid \mathcal{I})| \\ &\leq \mathbb{E}|I_n^M - \mathbb{E}(X_M \mid \mathcal{I})| + \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}|X_M^1(\varphi^k)| + \mathbb{E}|X'_M| \\ &\leq \mathbb{E}|I_n^M - \mathbb{E}(X \mid \mathcal{I})| + 2\mathbb{E}|X'_M|\end{aligned}$$

Now for any $\varepsilon > 0$, choose $M \geq 1$ such that $\mathbb{E}|X'_M| < \varepsilon$. Then $A_n \leq \mathbb{E}|I_n^M - \mathbb{E}(X_M \mid \mathcal{I})| + 2\varepsilon$. Hence $\lim_{n\to\infty} A_n = 0$. ☐

Example 1.2. *Applications of Birkhoff's Ergodic Theorem.*

(a) iid sequence $\frac{1}{n}\sum_{k=0}^{n-1} X_k \to \mathbb{E}X_1$ a.s. and in $\mathcal{L}^1$. We recover the strong law.

(b) irreducible Markov chain with stationary distribution π. In this case, $I = \{\varnothing, |omega\}$. For any $f \in \mathcal{L}^1(S;\pi)$ we have

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \mathbb{E}f(X_0) = \int f\, d\pi$$

a.s. and in $\mathcal{L}^1$.

(c) Rotation of the circle. Given $\theta \in (0,1]$, consider $\varphi_\theta: \omega \to \omega + \theta \mod 1$. Think of it as a circle with $e^{2\pi i\omega} \to e^{2\pi i(\omega+\theta)}$.

Fact 1.3. *(a) φ_ω is measure-preserving on $\Omega = [0,1)$*

(b) φ_θ *is ergodic iff* θ *is irrational*

(c) *For all* $A \in \mathcal{B}([0,1))$,

$$\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{\varphi^k\omega \in A\}} \to |A|$$

a.s. and in $\mathcal{L}^1$

Theorem 1.4 (7.24 in Durrett). *If* $A = [a,b)$ *the above a.s. convergence can be strengthened to* **everywhere** *convergence*

Recall the Bernoulli shifts on $[0,1)$. Express x in dyadic form:

$$x = \sum_{k=1}^{\infty} x_k 2^{-k}, \qquad x_k \in \{0,1\}$$

Then $\varphi(x) = 2x \mod 1$ is just $\varphi\colon (x_1, x_2, \ldots) \to (x_2, x_3, \ldots)$. Let $i_1, \ldots, i_k \in \{0,1\}$ satisfy $r = \sum_{j=1}^k i_j 2^{-j}$. Let $X(\omega) = 1_{\{\omega \in [r, r+2^{-k})\}}$: in other words, we truncate the dyadic decomposition at height k. Then

$$\frac{1}{n}\sum_{k=0}^{n-1} X(\varphi^k\omega) \to 2^{-k}$$

a.s. and in $\mathcal{L}^1$.

Lecture 2: April 6

Definition 2.1. *A one dimensional* **Brownian motion** *(BM) is a real-valued process $B_t, t \geq 0$ such that:*

(a) If $t_0 < t_1 < \cdots < t_n$, then $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent

(b) $B_{t+s} - B_s \sim N(0,t)$

(c) $t \mapsto B_t$ is continuous

Our space is $\Omega = \mathbb{R}^{[0,\infty)}$. However the space of all continuous functions is not measurable, so we have to go via the rationals. Take $\Omega_0 = \mathbb{R}^{[0,\infty)\cap\mathbb{Q}}$. We show that the sample paths are absolutely continuous. In fact, we have

Theorem 2.2 (8.15). *Brownian paths are γ-Hölder continuous for any $\gamma < \frac{1}{2}$.*

Let $\{\varphi_n, n \geq 1\}$ be an ONB of $\mathcal{L}^2([0,1], dx)$. Specifically, we use the Haar base (simple functions with powers of 2). The antiderivatives of the Haar base form the tent functions. Let $\{\xi_n, n \geq 1\}$ be iid $N(0,1)$ and set $B_t = \sum_{n=1}^{\infty} \left\langle 1_{[0,t]}, \varphi_n \right\rangle \xi_n$. This is convergent, because $\mathbb{E}(B_t^2) = t$. More precisely, let

$$X_j(t) = \sum_{n=1}^{j} \left\langle 1_{[0,t]}, \varphi_n \right\rangle \xi_n$$
$$\implies \mathbb{E}(X_j(t) - X_m(t))^2 = \sum_{n=m+1}^{j} \left\langle 1_{[0,t]}, \varphi_n \right\rangle^2 \to 0.$$

Then set $B_t = \lim_{j\to\infty} X_j(t)$ to conclude.

Each $X_j(t)$ is distributed as $N\left(0, \sum_{n=1}^{j} \left\langle 1_{[0,t]}, \varphi_n \right\rangle^2\right)$. Hence $B_t \sim N(0,t)$. For $s < t$, we have

$$(B_s, B_{t-s}) = \lim_{j\to\infty} (X_j(s), X_j(t) - X_j(s)).$$

First, they are jointly Gaussian (since they are a linear combination of jointly Gaussian vectors).

Claim 2.3. *The covariance is 0.*

Proof. We prove this more generally as follows. For any $f \in \mathcal{L}^2[0,1]$, define

$$\begin{aligned} Z_f &= \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle \xi_n \\ \implies Z_f &\sim N(0, \|f\|_2^2) \\ \implies \operatorname{Cov}(Z_f, Z_g) &= \mathbb{E}[Z_f \cdot Z_g] \\ &= \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle \langle g, \varphi_n \rangle \\ &= \langle f, g \rangle_{\mathcal{L}^2[0,1]} \end{aligned}$$

We get a Gaussian field that is isometric to $\mathcal{L}^2[0,1]$. □

Thus B_t has independent stationary increments and $B_t - B_s \sim N(0, t-s)$. It remains to show that $t \mapsto B_t$ is continuous; we leave this as an exercise.

Review of exam from last quarter

(a) Lazy walk on a cube:

Expected number of steps starting from u before returning, i.e. $\mathbb{E}_u T_u = \frac{1}{\pi(u)}$.

Expected number of visits to v before returning to u: this is $X = \sum_{k=0}^{T_u} 1_{\{X_k = v\}}$. But in class we saw $\mathbb{E}X = \mu_u(v) = c\pi(v)$ (since it is irreducible). As $\mu_u(u) = 1$, we get $c = 8$. Hence $\mathbb{E}X = 8\frac{1}{8} = 1$.

The hint suggests finding the pmf for X. Let $p = \mathbb{P}_u(T_v < T_u) = \mathbb{P}_v(T_u < T_v)$. For $j \geq 1$, we have $\mathbb{P}_n(X = j) = p(1-p)^{j-1}p$. Thus $\mathbb{E}X = \sum_{j=1}^{\infty} j\mathbb{P}_n(X = j) = 1$.

Lecture 3: April 8

Going over the Math 522 final exam

#2 Let X_n be the biased RW on $\mathbb{Z}$. There is a lowest point you can go, then shift to get a biased RW and the result follows.

#3 Markov chain X_n with transition probability $p(x,y)$. Then $h \geq 0$ is harmonic if $h(x) = \sum_y p(x,y)h(y) = \mathbb{E}_x[h(X_1)]$ (assume $h \geq 0$ to make r.h.s. well-defined, other conditions also suffice). For example if we use SRW, then $h(x) = \frac{h(x-1)+h(x+1)}{2}$. Also ξ on Ω is invariant if $\xi \circ \theta = \xi$.

Prove the following:

(a) There is a bijection between bounded harmonic functions and bounded invariant r.v

(b) Any bounded harmonic function can be written as $h_1 - h_2$

(c) If X_n is an irreducible and recurrent Markov chain, then any invariant r.v. is trivial.

Proof. (a) If $\xi \geq 0$ is invariant, then

$$\begin{aligned} h(x) \overset{def}{=} \mathbb{E}_x\xi &= \mathbb{E}_x[\xi \circ \theta] \\ &= \mathbb{E}_x[\mathbb{E}_{X_1}\xi] \\ &= \mathbb{E}_x[h(X_1)] \end{aligned}$$

so h is harmonic. Conversely, if $h \geq 0$ is harmonic then $M_n = h(X_n)$ is a martingale. First, observe that M_n is integrable as h is bounded. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then

$$\begin{aligned} \mathbb{E}_x[h(X_{n+1} \mid \mathcal{F}_n)] &= \mathbb{E}_X[h(X_1) \circ \theta_n \mid \mathcal{F}_n] \\ &= \mathbb{E}_{X_n} h(X_1) \\ &= h(X_n). \end{aligned}$$

By Martingale Convergence, $h(X_n) \to M_\infty$ a.s. (and in $\mathcal{L}_1$, but we don't need this). Hence $M_\infty \circ \theta = M_\infty$ a.s. Since h is bounded, M_n is a u.i. martingale and therefore

$$\mathbb{E}_x M_\infty = \mathbb{E}_x M_1 = h(x).$$

(b) By the previous part,

$$\begin{aligned} h(x) &= \mathbb{E}_x\xi \\ &= \mathbb{E}_x\xi^+ - \mathbb{E}_x\xi^-. \end{aligned}$$

(c) For invariant r.v. ξ, let $\xi_n = (-n) \vee \xi \wedge n$. This is bounded and invariant. Let $h_n(x) = \mathbb{E}_x\xi_n$ which is bounded harmonic. Since $k \mapsto h_n(X_k)$ is a bounded martingale, $\lim_{k\to\infty} h_n(X_k)$ exists a.s. Fix x_0 and $y \in X$. Under $\mathbb{P}_{x_0}$ we have

$$X_n(\omega) = y \qquad \text{i.o., a.s.}$$

Now $\lim_{n\to\infty} h(X_n) = h_n(X_0) = h_n(y)$ a.s. Thus h_n is constant (up to a set of measure 0). The result follows, as $\xi_n = \lim_{k\to\infty} h_n(X_k)$ and thus ξ is constant. $\square$

Construction of Brownian Motion

We work only on $[0,1]$. This is Lévy's construction. Let $\{\varphi_n\}$ be an orthonormal basis of $\mathcal{L}^2([0,1],dx)$ and choose $\{\xi_n\}$ iid Gaussian. For any $f \in \mathcal{L}^2[0,1]$, set $X_f = \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle \xi_n$. This is well-defined as the $\mathcal{L}^2$ limit of partial sums.

Then $\mathbb{E}(X_f)^2 = \sum \langle f, \varphi_n \rangle^2$, so by Parseval's Identity $\mathbb{E}(X_f)^2 = \|f\|_2^2$. More generally, $\mathbb{E}[X_f X_g] = \langle f, g \rangle$. We define $B_t = X_{1_{\{[0,t]\}}}$. It is clear that this satisfies all properties of BM besides continuity.

We use the Haar base φ_n (fluctuates between ± 1 on intervals of size 2^{-n}).

To get the Gaussian free field, we use the Hilbert space $W_0^{1,2}(D)$.