18.615: Introduction to Stochastic Processes

Rachel Wu

Spring 2017

These are my lecture notes from 18.615, Introduction to Stochastic Processes, at the Massachusetts Institute of Technology, taught this semester (Spring 2017) by Professor Alexey Bufetov[1].

I wrote these lecture notes in LaTeX in real time during lectures, so there may be errors and typos. I have lovingly pillaged Tony Zhang's[2] formatting commands and style. Should you encounter an error in the notes, wish to suggest improvements, or alert me to a failure on my part to keep the web notes updated, please contact me at rmwu@mit.edu.

This document was last modified 2018-05-01.

[1] abufetov@mit.edu
[2] txz@mit.edu

Contents

1 February 28, 2017

1.1 Markov chains

Markov chains have state space $S = \{1, 2, \ldots, N\}$ and transition matrix P. We start at

$$\varphi_0 = (P(X_0 = 1), P(X_0 = 2), \ldots, P(X_0 = N)). \tag{1.1}$$

In addition,

$$(P(X_n = 1), \ldots, P(X_n = N)) = \varphi_0 P^n \tag{1.2}$$

and

$$P_n(i, j) = P(X_n = j | X_0 = i). \tag{1.3}$$

1.2 Classification of states

Definition 1.1. For $i, j \in S$, we write

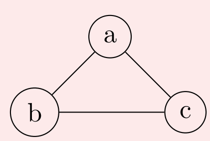
1. $i \to j$ if $\exists m \in \mathbb{Z}_+$ such that $P_m(i, j) > 0$,
2. $i \leftrightarrow j$ if $\exists m_1, m_2 \in \mathbb{Z}_+$ such that $P_{m_1}(i, j) > 0$ and $P_{m_2}(j, i) > 0$.

The former implies that i is reachable from j, and the latter implies that j is *also* reachable from i.

Lemma 1.2 (Transitivity)
If $i \leftrightarrow j$, and $j \leftrightarrow k$, then $i \leftrightarrow k$.

The state space S can be partitioned into $S = \tilde{S}_1 \cup \tilde{S}_2 \cup \cdots \cup \tilde{S}_l$, where $\tilde{S}_i$ are disjoint, nonempty, **communication classes**. Each of these classes has the property that all vertices i, j have $i \leftrightarrow j$?

Example 1.3 (Communication classes)
How many communication classes?



One communication class, since all states are reachable from all others. How about for gambler's ruin? There are three classes: $0, N$, and the rest.

Since Markov chains are directed graphs, communication classes are just maximal connected subgraphs.

Definition 1.4. A Markov chain is called **irreducible** if it has only one communication class.

Definition 1.5. Let $\tilde{S}$ be a communication class, and let $X_0 \in \tilde{S}$. If $\Pr\{X_n\} \notin \tilde{S} \to 1$, then $\tilde{S}$ is called **transient**, and all states $s \in \tilde{S}$ are **transient states**. Otherwise, $\tilde{S}$ is **recurrent**.

If there is any out degree of $\tilde{S}$, then $\tilde{S}$ is transient.

Definition 1.6. For each $s \in S$, we define $T = \{n \in \mathbb{Z}\}$[3] such that $P_n(s, s) > 0$. Then, $d = \gcd(T_1 \cup T_2 \cup \cdots \cup T_n)$ is the **period** of the Markov chain. If $d = 1$, then this Markov chain is **aperiodic**.

Example 1.7

If we are given sets $\{2, 4, 6, 8, \ldots\}$, then the period is 2. For gambler's ruin, the chain is aperiodic.

In fact, if there is any loop (node to self) in the graph, then the Markov chain is aperiodic.

Proposition 1.8

If there exists $a \in \mathbb{Z}_+$, such that all entries of P^a are strictly positive, then the Markov chain is irreducible and aperiodic.

Proof. If $P_a(i, j) > 0, \forall i, j$, then there is only one communication class, and it is irreducible by definition. In addition, $a \in T_1$, and $(a + 1) \in T_1$ as well, and $\gcd(a, a + 1) = 1$, so this is aperiodic. □

Theorem 1.9 (Convergence theorem)

If all entries of P^a are strictly positive, for some a, then there exists a unique invariant distribution π, and for any initial distribution φ_0, $\varphi_0 P^a \to \pi$.

Example 1.10 (Phone call)

Suppose we have a phone that is either free or busy.

$$\begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix} \tag{1.4}$$

The invariant distribution for this is $\pi = (2/3, 1/3)$.

Example 1.11 (Symmetric random walk with reflective boundaries)

Take gambler's ruin, except at the ends, we bounce back with $p = 1$. This has a stationary distribution $\pi = (1/2N, 1/N, \ldots, 1/N, 1/2N)$.

Example 1.12 (Lazy random walks)

The point is lazy, so it doesn't like moving.

$$p(i, i + 1) = 1/4 \quad p(i, i) = 1/2 \quad p(i, i - 1) = 1/4 \tag{1.5}$$

[3] The professor used J.

2 March 2, 2017

Pset 2 is due next lecture. Office hours March 6, 2-3p.

2.1 Markov chain convergence

We prove theorem 1.9 from the previous lecture.

If P^n converges to some matrix, as $n \to \infty$, then the theorem holds.

$$P^n \to \begin{pmatrix} \pi(1) & \pi(2) & \dots & \pi(n) \\ \pi(1) & \pi(2) & \dots & \pi(n) \\ \vdots & & & \vdots \\ \pi(1) & \dots & \dots & \pi(n) \end{pmatrix} \tag{2.1}$$

Then for any φ_0,

$$\varphi_0 P^n = \varphi_0 \begin{pmatrix} \pi(1) & \dots & \pi(n) \\ \vdots & \dots & \vdots \\ \pi(1) & \dots & \pi(n) \end{pmatrix} = (\pi(0), \pi(1), \dots, \pi(n)). \tag{2.2}$$

Now assume that $\pi_2 P = \pi_2$. Then $\pi_2 = \pi_2 P = \pi_2 P^2 = \dots \pi_2 P^n$, which only works if P is invariant (?).

Let $d = \min_{i,j} P_0(i,j) > 0$, $g_{n,j} = \max_i P_n(i,j)$, and $s_{n,j} = \min_i P_n(i,j)$. We need to prove that $g_{n,j} - s_{n,j} \to 0$. Essentially, this means that the elements of P^n converge.

Lemma 2.1

$g_{n,j} - s_{n,j} \to 0$ as $n \to \infty$.

Proof of lemma. We write out the matrix entry form of $P^{a(n+1)}$, and see that

$$g_{a(n+1),j} = \sum_k P_a xxx \cdot P_{n+1} \leq (1-d) g_{xxx,j} + d s_{xx,j}. \tag{2.3}$$

TODO: fill in all the indices... The professor uses a lot of indices I can't see, so basically, all the largest numbers are multiplied by smallest numbers sometime during the matrix multiplication, so the larger numbers shrink and smaller numbers grow.

Now we show that the smallest element increases.

$$s_{a(n+1),j} = \sum_k P_a(\hat{i}, k) \cdot P_{n+1}(k,j) \geq (1-d) s_{an,j} + d g_{an,j}. \tag{2.4}$$

So

$$g_{a(n+1),j} - s_{a(n+1),j} \leq (1-2d)(g_{n,j} - s_{n,j}) \tag{2.5}$$

Lemma 2.2

$g_{n+1,j} \leq g_{n,j}$ and $s_{n+1,j} \geq s_{n,j}$.

Proof of lemma. P^{n+n} multiples rows and columns, which sum to 1, with elements each greater than 0. So $g_{n+1,j} \leq 1 \cdot g_{n,j} + 0 = g_{n,j}$, and also $s_{n+1,j} \geq s_{n,j}$ for the same reason. So here, $g_{n+1,j} - s_{n+1,j} \leq g_{n,j} - s_{n,j}$. (This only doesn't work in generality since they may be equal). □

□

Question 2.3. Couldn't we just have used eigenvalues, one $\lambda = 1$? Well yes, but this is more general, for arbitrary dimensions. In general, they may also be more invariant distributions.

2.2 Implications of convergence

We make the following observations.

- If THMC is irreducible and aperiodic, then $\exists a$, such that P^a has all entries strictly positive. If $J \subset \mathbb{Z}$, and $\gcd(J) = 1$, and J is closed under addition, then J forms a group of the periods?

Corollary 2.4

Suppose we have function $f : S \to \mathbb{R}$. Then

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right] \to \pi(1)f(1) + \pi(2)f(2) + \dots \tag{2.6}$$

since when $n \to \infty$, the Markov chain does not change from π.

Corollary 2.5 (Ergodic theorem)

Let $f = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$. Then as a generalization of the convergence theorem,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to f. \tag{2.7}$$

Now consider a few more things.

- Let $\mathbb{1}$ be an indicator function for state r. Then $\mathbb{E}\left[\frac{1}{n}\sum_i \mathbb{1}_{X_i=r}\right] \to \pi(r)$, or the **number of visits** to state r is proportional to $n\pi(r)$.
- Now let τ be the **return time**, which is the time it takes to hit X_i after the first time. Suppose $X_0 = r$, and $\tau(r) = \min\{i \geq 1, X_i = r\}$ Then $\mathbb{E}[\tau_r] = \frac{1}{\pi(r)}$.

We can just use these facts on psets. However, we must show that the Markov chain is irreducible and aperiodic, and find the invariant distribution (and of course, show that it is a Markov chain).

3 March 7, 2017

3.1 Markov chains (continued)

Today we continue Markov chains.

Example 3.1 (Ehrenfest Urn model)
There are two urns of balls. At each step, an urn is selected and one ball is transferred to the other urn. The states are the number of balls in urn 1, $k = \{0, \dots, n\}$. The transition probabilities are

$$\Pr\{k, k+1\} = \frac{n-k}{n} \tag{3.1}$$

$$\Pr\{k, k-1\} = \frac{k}{n} \tag{3.2}$$

We observe that the urns are the same, so $\pi(k) = \pi(n-k)$. By definition of a stationary distribution,

$$\pi(k-1)\Pr\{k-1, k\} + \pi(k+1)\Pr\{k+1, k\} = \pi(k), \tag{3.3}$$

for $k = 1, \dots, n-1$. We will show that $\pi(k) = \frac{1}{2^n}\binom{n}{k}$ is a solution.

$$\begin{aligned} &\pi(k-1)\Pr\{k-1, k\} + \pi(k+1)\Pr\{k+1, k\} \\ &= \frac{n!}{(k-1)!(n-k+1)!} \cdot \frac{n-k+1}{n} + \frac{n!}{(k+1)!(n-k-1)!} \cdot \frac{k+1}{n} \\ &= \frac{n!}{k!(n-k)!} = \pi(k) \end{aligned} \tag{3.4}$$

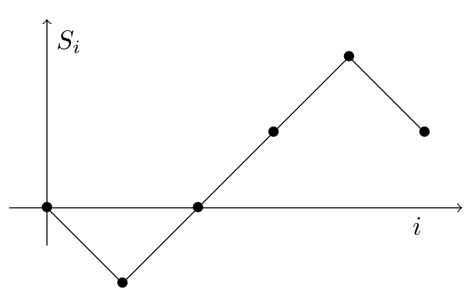
This concept is similar to the second law of thermodynamics.

3.2 Simple random walks in $\mathbb{Z}$

Let $\{\epsilon_0, \dots, \epsilon_n\}$ be i.i.d. random variables,

$$\epsilon_i = \begin{cases} 1 & p = 1/2 \\ -1 & p = 1/2, \end{cases} \tag{3.5}$$

and let S_n be the sum of the first n random variables.



How many paths are there from (0,0) to (n, x)? Well if we take p steps up and q steps down,

$$\begin{aligned} n &= p + q \\ x &= p - q. \end{aligned}$$

Then the number of paths is equal to

$$N_{(0,0)\to(n,x)} = \begin{cases} \binom{p+q}{p} = \binom{n}{\frac{n+x}{2}} & \frac{n+x}{2} \in \mathbb{Z}, |x| \le n \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

To quantify this, we introduce our best friend.

Theorem 3.2 (Stirling's formula)
Most useful approximation ever,

$$n! \sim \sqrt{2\pi n}(\frac{n}{e})^n \tag{3.7}$$

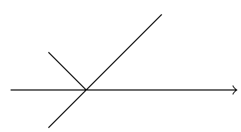
Using Stirling's $\Pr\{S_n = 0\}$ approaches $1/\sqrt{\pi n}$. Then we find that

$$\mathbb{E}[\# \text{ of returns to } 0] = \sum_{n-1}^{\infty} \Pr\{S_n = 0\} = \infty. \tag{3.8}$$

3.3 Counting paths examples

Example 3.3 (Burning houses)
Imagine that there is a river, a firetruck, and a burning house. The firetruck must visit the river before rescuing the house. How do we determine the fastest route?

Solution. Reflect the firetruck across the river, connect the house and reflection, and determine where that intersects the river.



Example 3.4
We have two points (a, α) and (b, β), such that $\alpha, \beta > 0$, $b > a$, and $a, b, \alpha, \beta \in \mathbb{Z}$. How many paths between a and b touch or cross the x axis?

Solution. There is a bijection between the paths from a to b and the paths from a' to b, where a' is a reflected across the x axis. Simply reflect across the x axis at the first point the $a' \to b$ line crosses the x axis. There are

$$\binom{b-a}{\frac{(b-a)+(\beta-\alpha)}{2}} \tag{3.9}$$

paths that satisfy the constraint.

Corollary 3.5

The number of paths from $a \to b$ which do not touch the x axis is equal to the number of paths from $a \to b$ minus the number of paths from $a' \to b$.

Corollary 3.6

The number of paths from $(1,1) \to (n,x)$ which are strictly above the x axis is $N_{n-1,x-1} - N_{n-1,x+1}$.

Example 3.7

How many paths from $(1,1) \to (5,1)$ do not touch the x axis?

Solution. Trivially we could draw them and see that there are 2. Using the formula,

$$N_{4,0} - N_{4,2} = \binom{4}{2} - \binom{4}{3} = 6 - 4 = 2. \tag{3.10}$$

Example 3.8 (Dyck paths)

How many paths from $(1,1) \to (2n+1,1)$ do not touch the x axis?

Solution. We again use the formula and see that

$$N_{2n,0} - N_{2n,2} = \binom{2n}{n} - \binom{2n}{n+1} = \frac{1}{n+1}\binom{2n}{n}. \tag{3.11}$$

In particular, we have discovered Catalan numbers! This is also equivalent to the number of Dyck paths from $(0,0) \to (2n,0)$.

4 March 9, 2017

4.1 More simple random walks in $\mathbb{Z}$

Recall from last lecture that the number of paths from $(0,0) \to (n,x) = N_{n,x}$

$$N_{n,x} = \begin{cases} \binom{p+q}{p} = \binom{n}{\frac{n+x}{2}} & \frac{n+x}{2} \in \mathbb{Z}, |x| \leq n \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The number of paths from $(a,\alpha) \to (b,\beta)$ which do not touch the x axis is

$$N_{b-a,\beta-\alpha} - N_{b-a,\beta+\alpha}. \tag{4.2}$$

At the end of the course, we will model Brownian motion with such a simple random walk.

> "We will suffer all this combinatorics..."—abufetov

Finally, we also found Catalan numbers in Dyck paths.

Theorem 4.1 (Ballot's theorem)
The number of paths from $(0,0) \to (n,x)$ such that $s_i > 0, \forall i$ is

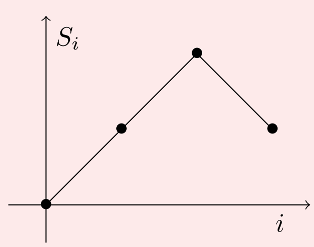
$$N_{n-1,x-1} - N_{n-1,x+1} = \frac{x}{n} N_{n,x}$$

Proof. The left side is equal to the number of paths from $(1,1) \to (n,x)$ since from (0,0), we can only go up to (1,1). For the right side, let us take p steps up and q steps down, such that $x = p - q, n = p + q$. Then

$$\binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{(p+q-1)!}{(p-1)!?}\left(\frac{1}{q} - \frac{1}{p}\right) \tag{4.3}$$

□

Example 4.2 (Application of Ballot's theorem)
Of the paths from $(0,0) \to (3,1)$, only one satisfies the non-negativity.



Proposition 4.3
The number of paths of length $2n$, such that $s_i > 0, \forall i \neq 0$, is

$$N_{2n-1,1} = \frac{1}{2} N_{2n,0}.$$

Proof. The number of paths satisfying this condition is

$$\begin{aligned}
&\sum_{r=1}^{\infty} \#\text{ paths } (s_0 = 0, s_1 > 0, \ldots, s_{2n-1} > 0, s_{2n} = 2r) \\
&= \sum_{r=1}^{\infty} N_{2n-1,2n-1} - N_{2n-1,2n+1} \\
&= (N_{2n-1,1} - N_{2n-1,3}) + (N_{2n-1,3} - N_{2n-1,5}) + \ldots \\
&= N_{2n-1,1}
\end{aligned}$$

since we have a telescoping series. □

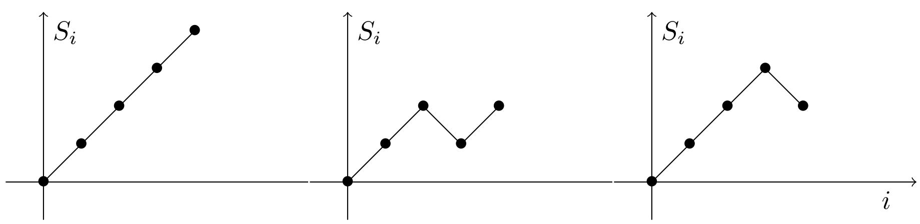
Corollary 4.4

The number of paths that do not return to 0 in the first $2n$ steps, $(s_0 = 0, s_1 \neq 0, \ldots, s_{2n-1} \neq 0, s_{2n} = 2r)$ is

$$2 \cdot \frac{1}{2} N_{2n,0} = N_{2n,0},$$

since points are all positive or all negative.

For $n = 2$, we can have the positive or negative versions of the following.



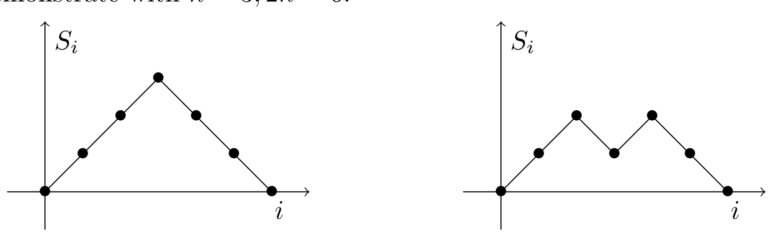
Proposition 4.5

The number of paths that start from 0 and return to 0 at $2n$ is

$$4N_{2n-2,0} - N_{2n,0} = \frac{1}{2n-1} N_{2n,0}$$

.

We demonstrate with $n = 3, 2n = 6$.



And we also have the negative versions of these.

4.2 Applications to probability

Recall that ϵ are distributed i.i.d. Bernoulli. Let

$$\mu_{2n} = \Pr\{S_{2n} = 0\} = \frac{N_{2n,0}}{2^{2n}} \sim \frac{1}{\pi n} \to 0$$

and

$$f_{2n} = \Pr\{\text{return to } 0\} = \frac{1}{2n-1}\mu_{2n}.$$

Corollary 4.6

We can derive that $f_{2n} = \mu_{2n-2} - \mu_{2n}$ from proposition 4.5.

Corollary 4.7

The probability that a simple random walk returns to 0 is

$$\sum_{i=0}^{\infty} f_i \to 1$$

since we have another telescoping series.

So the expected return time is

$$\mathbb{E}[\tau_0] = \sum_{n=1}^{\infty} 2n f_{2n} = \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} \to \infty = O(\sqrt{n}). \tag{4.4}$$

5 March 21, 2017

We had an exam and a snow day, so we continue discussing simple random walks now.

5.1 Last return

Let $\tau = \max_i \{S_{2i} = 0\}$, which is the last return to 0 for $i \in 0, 1, \dots, 2n$. So

$$\Pr\{\tau = 2k\} = \frac{N_{2k,0}N_{2n-2k,0}}{2^{2n}} = \mu_{2k}\mu_{2n-2k}.$$

Bashing with Stirling's,

$$\Pr\{\tau = 2k\} \sim \frac{1}{\sqrt{\pi k}}\frac{1}{\sqrt{\pi(n-k)}}. \tag{5.1}$$

Proposition 5.1

$N_{2k,0} \cdot N_{2n-2k,0}$ is the number of paths that last return to 0 at $2k$.

However, probabilities at individual points tend to 0, so we need to determine the cumulative distribution. For $0 < \alpha < 1$,

$$\begin{aligned}\Pr\{\tau_0 \le 2\alpha n\} &= \sum_{k=0}^{\alpha n} \Pr\{\tau_0 = 2k\} \\ &\approx \sum_{k=1}^{\alpha n} \frac{1}{\pi\sqrt{k(n-k)}} \\ &= \sum_{k=1}^{\alpha n} \frac{1}{n}\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}. \end{aligned} \tag{5.2}$$

As n tends to ∞, this summation becomes an integral.

$$\int_0^\alpha \frac{1}{\pi\sqrt{x(1-x)}}\,dx = \frac{2}{\pi}\arcsin\sqrt{\alpha}$$

Theorem 5.2

As n tends to ∞, $\Pr\{\tau_\alpha \le 2\alpha n\} \to \frac{2}{\pi}\arcsin\sqrt{\alpha}$.

For example, take $\alpha = 0.1$. Then approximately 20.4% of length n paths last return to 0 at 10% of n. It is very likely for last returns to be close to 0 or n, and the least likely for last returns to occur in the middle.

I'm tired so I'll fill in the plots over spring break....... Fill in a u shaped plot

5.2 Maximum value of simple random walks

We now consider paths with a given maximum value.

Proposition 5.3

The number of paths from $(0,0) \to (n,x)$ of length n, which touch the $q = r$ line is $N_{n,2r-x}$.

This is trivially true by bijection and reflection.

Proposition 5.4

The number of paths from $(0,0) \to (n,x)$ of length n such that $\max_k S_k = r$ is $N_{n,2r-x} - N_{n,2r+2-x}$

This is equivalent to the number of paths that touch $y = r$ minus the number of paths that touch $y = r+1$. All paths going beyond must also touch $y = r+1$.

Proposition 5.5

The number of paths of length n such that $\max_k S_k = r$ is

$$\sum_{x=r}^{1} \#\text{ paths } (0,0) \to (n,x) = \max(N_{n,r}; N_{n,r+1}),$$

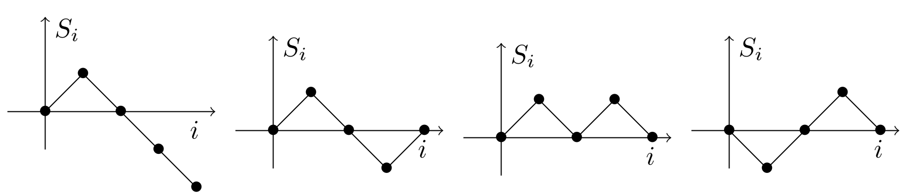
depending on parity.

This comes from the typical telescoping series.

Example 5.6

Let $n = 4, r = 1$. How many paths are there?

Solution. There are $\max(N_{4,1}, N_{4,2}) = 4$.



Theorem 5.7

As n tends to ∞,

$$\Pr\left\{S_k \leq \alpha\sqrt{k}\right\} \to \frac{1}{2\pi}\int_{-\infty}^{\alpha} e^{-t^2/2}\,dt$$

by the central limit theorem.

6 Mach 23, 2017

6.1 Maximum value, ctd.

Recall that we are dealing with a simple random walk of length n.

Theorem 6.1

By the central limit theorem, for $\alpha \geq 0, n \to \infty$,

$$\Pr\left\{\max_k S_k \geq \alpha\sqrt{k}\right\} \to \frac{1}{\sqrt{2\pi}} \int_\alpha^{+\infty} e^{-t^2/2}\, dt\,.$$

This is equivalent to the sum

$$\frac{N_{n,n} + N_{n,n-1} + N_{n,n-2} + \cdots + N_{n,r\alpha\sqrt{n}}}{2^n}.$$

In particular, note that $N_{n,n-1}$ and all other "odd" terms are 0, since parity. In words, that is

$$\frac{\#[\text{paths max } n] + \#[\text{paths max } n-1] + \cdots + \#[\text{paths max } \lceil \alpha\sqrt{n}\rceil]}{2^n}$$

Proof. From central limit theorem,

$$\int_\alpha^\infty e^{-t^2/2}\, dt$$

Combinatorially, we can treat this quantity as

$$N_{n,n} + N_{n,n-1} + \cdots + N_{n,grosschunkintegral}$$

However, we want to find

$$\begin{aligned} &\frac{\#\text{paths max} = n + \#\text{paths max} = n-1 \ldots}{2^n} \\ &= \frac{\max(N_{n,n}; N_{n,n+1}) + \max(N_{n,n}; N_{n,n+1}) \ldots}{2^n}, \end{aligned}$$

but by parity, we notice that the $N_{n,n+1}$ etc. terms are 0, so this value is twice the central limit theorem result. □

6.2 Time above line

We wonder how much time a path spends above the x axis.

Lemma 6.2

Let F_{2k} be the number of paths which return to 0 for the first time at $2k$. Then

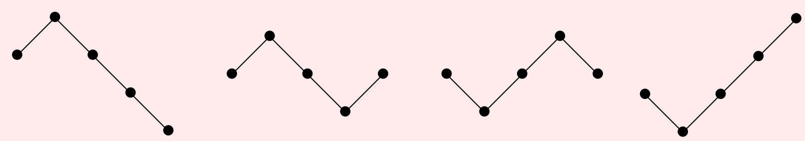
$$N_{2k,0} = F_{2k} + F_{2k-2}N_{2,0} + F_{2k-4}N_{4,0} + \cdots + F_2 N_{2k-2,0}$$

Each step of a path with length $2n$ is either on the positive or negative half plane. We study paths of length $2n$, $(S_1, S_2, \ldots, S_{2n})$. Let N_+ be the number of "positive steps," and N_- be "negative." Observe that both N_+ and N_- must be even, as we must return to 0 to change sign. Let

$$B_{2k,2n} = \#[\text{paths length } 2n \text{ such that } N_+ = 2k.$$

Example 6.3

Find $B_{2,4}$. We draw out the possibilities. There are 4.



Proposition 6.4

We know that $B_{2n,2n} = N_{2n,0} = \#$ of non-negative paths. Then $\frac{1}{2}N_{2n,0} =$ $\#$ of positive paths, as we can just shift up-right and fill in the last step.

Proposition 6.5

$B_{2k,2n} = N_{2k,0} \cdot N_{2n-2k,0}$

This can be proved by induction on n.

7 April 4, 2017

7.1 Infinite-state Markov chains

We are given an unbounded but countable state space $S = \{0, 1, 2, \ldots\} \in \mathbb{Z}$, with a probability distribution $\{\alpha(x)\}, \sum_x \alpha(x) = 1$. A Markov chain $X_0, X_1, \ldots$ has transition probability

$$\Pr\{X_{n+1} = y | X_n = x\} = p(x, y), \text{ for } x, y \in S.$$[4]

So by total probability,

$$\Pr\{X_{n+1} = y\} = \sum_{x \in S} p(X_n = x) p(x, y).$$

An invariant distribution π has the property that

$$\pi(y) = \sum_{x \in S} \pi(x) p(x, y).$$

Example 7.1

We have an infinite Markov chain where

$$p(i, i+1) = 1, i \geq 0.$$

As a graph,

$$\boxed{0} \to \boxed{1} \to \boxed{2} \to \boxed{3} \to \boxed{4} \to \text{etc.}$$

This has *no* stationary distribution.

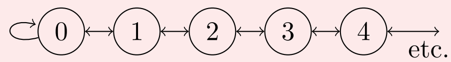
[4] In addition, $\Pr\{X_{n+k} = y | X_n = x\} = p_k(x, y)$.

Example 7.2

Consider a Markov chain, similar to simple random walks.

$$p(i, i+1) = p(i, i-1) = 1/2$$
$$p(0, 0) = 1/2$$

As a graph,



So

$$\pi(n) = \pi(n-1)/2 + \pi(n+1)/2.$$

This is a bit harder, but we can say that $\pi(0) = \pi(1)$, and thus

$$\pi(1) = \pi(2) = \cdots = \pi(n-1) = \pi(n).$$

This isn't very useful since we have an infinite number of states, so we can say that there's no invariant distribution.

These examples show two types of new behavior: there could be no convergence like 7.1, or just no distribution like 7.2.

Definition 7.3. Let $X_0, X_1, \ldots$ be a sequence of random variables. Let random variable τ be the **stopping time**, such that event $\{\tau = n\}$ depends on $X_0, X_1, \ldots, X_n$ only.

Definition 7.4. Given some state $x \in S$, then the **hitting time** the first time the Markov chain visits x, $\tau_n = \min\{n, X_n = x\}$.

Proposition 7.5 (Strong Markov property)

Let $X_0, X_1, \ldots$ be a Markov chain, and let τ be a stopping time. Then

$$\Pr\{X_{\tau+k} = y | X_\tau = x\} = p_k(x, y)$$

for any $x, y \in S$.

Proof. By total probability,

$$\begin{aligned}\Pr\{X_{\tau+k} = y | X_\tau = x\} &= \sum_{n=1}^{\infty} \Pr\{\tau = n\} \Pr\{X_{n+k} = y | X_n = x\} \\ &= \sum_{n=1}^{\infty} \Pr\{\tau = n\} p_k(x, y)\end{aligned}$$

We can pull $p_k(x, y)$ out of the summation, and realize that $\sum_{n=1}^{\infty} \Pr\{\tau = n\} = 1$, so all that is left is $p_k(x, y)$. $\square$

Communications classes are the same as before. An irreducible Markov chain is one with one communication class.

Definition 7.6. A ITHMC[5] is **recurrent** if for some state x, $\Pr\{X_n = x\}$ for infinitely many times n has probability 1. Otherwise, the Markov chain is **transient**.

Proposition 7.7

If an irreducible Markov chain is recurrent, then $\Pr\{X_n = y\}$ for infinitely many n is equal to 1, $\forall y$.

Proof. For some $x \in S$, event $\{X_n = x\}$ happens infinitely often. By the definition of a communication class, $\exists k, p_k(x, y) = \epsilon > 0$. So if $X_n = x$, then $\Pr\{X_{n+k} = y\} = \epsilon$, and y also occurs infinitely often. $\square$

[5] infinite time-homogenous Markov chain

8 April 6, 2017

8.1 Infinite-state Markov chains, ctd.

We continue discussion of infinite-state Markov chains.

Proposition 8.1

If for some state $x \in S$, then

$$\sum_n p_n(x,x) < \infty$$

implies that the Markov chain is transient, while

$$\sum_n p_n(x,x) = \infty$$

implies that the Markov chain is recurrent.

Definition 8.2. If for any two states $x, y \in S$, $p_n(x,y) \to 0$ as $n \to \infty$, then this Markov chain is known as **null recurrent**, and it does *not* have an invariant distribution. Otherwise, the Markov chain is **positive recurrent**, and it has a invariant distribution.

Proposition 8.3

If a recurrent ITHMC has an invariant distribution, then it is positive recurrent.

Proof. Take state $y \in S$, with $\pi(y) > 0$. By the definition of an invariant distribution,

$$\pi(y) = \sum_{x \in S} \pi(x) p(x,y).$$

That implies that after n steps, we are *still* at the same distribution.

$$\pi(y) = \sum_{x \in S} \pi(x) p_n(x,y).$$

There exists a finite set $F \subset S$, $|F| < \infty$, such that

$$\sum_{x \in F} \pi(x) < \epsilon \pi(y).$$

We take an n large enough so that

$$\pi(y) = \sum_{x \in S} \pi(x) p_n(x,y) = \sum_{x \in F} \pi(x) p_n(x,y) + \sum$$

type up lol □

Claim 8.4. A positive recurrent ITHMC has a unique invariant distribution. For any $y \in S$, $p(x,y) \to \pi(y)$ (convergence theorem).

8.2 Applications of Infinite-state MCs

For the rest of class, we will go through examples.

Example 8.5 (Simple random walk)

Here, $S = \mathbb{Z}$ and transition probabilities are $1/2$ to each direction. How do we characterize this Markov chain?

Solution. The probability of returning to 0 is

$$p_{2n}(0,0) = p^n q^n \binom{2n}{n}.$$

By Stirling's, $\binom{2n}{n} \sim 2^{2n}/\sqrt{\pi n}$. Then

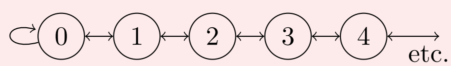
$$\sum_{n=1}^{\infty} p_{2n}(0,0) \approx \sum_{n=1}^{\infty} p^n q^n 2^{2n} \frac{1}{\sqrt{\pi n}} = \sum_{n=1}^{\infty} (4p(1-p))^n \frac{1}{\sqrt{\pi n}}.$$

This sum converges if $p \neq 1/2$. Otherwise, it diverges since $4 \cdot \frac{1}{4} = 1$. Therefore, it is transient if $p \neq 1/2$ and recurrent otherwise.

[TODO fill in from picture]

Example 8.6 (Simple random walk in $\mathbb{Z}_{>0}$)

Let S be the set of non-negative integers.



We have transition probabilities

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

where 0 bounces back. What is the invariant distribution?

Solution. The invariant distribution can be found as

$$\pi(x) = \pi(x+1)q + \pi(x-1)p.$$

This is a difference equation, with general solution

$$\pi(x) = c_1 + c_2 \left(\frac{p}{q}\right)^x.$$

The base condition is $\pi(0) = q\pi(0) + q\pi(1)$. Plugging in and bashing, we find that there are 3 possible outcomes.

Case 1 If $p > q$, then $\pi(x) > 1$, which is impossible. This is transient.

Case 2 If $p = q$, then $\pi(x)$ is also impossible. This is null recurrent.

Case 3 If $p < q$, then we have an invariant distribution

$$\pi(x) = \frac{q-p}{q} \left(\frac{p}{q}\right)^x.$$

Example 8.7 (d-dimensional lattice walk)
Let $S = \mathbb{Z}^d$, which is a d-dimensional lattice. Transition probabilities are $p(v_i, v_j) = \frac{1}{2d}$. What is $p_{2n}(0,0)$?

Solution. We have $2n/d$ steps of simple random walk in each dimension. The probability of returning in each dimension is $1/\sqrt{\pi n/d}$. If the simple random walks are independent, then the probability of returning is $(1/\sqrt{\pi n/d})^d$. For large n, this can be made rigorous, though they are not *actually* independent.

$$\sum_{n=1}^{\infty} p_{2n}(0,0) \sim c \sum \frac{1}{n^{d/2}}.$$

If $d > 2$, then this Markov chain is transient. If $d = 1, 2$, then this is a harmonic series that diverges, and this Markov chain is recurrent.

9 April 11, 2017

9.1 Probability generating functions

Let X be a random variable that takes on values $\{-k, -k+1, \ldots, 0, 1, \ldots\}$. Then the probability generating function is

$$f_X(s) = \sum_{n=-k}^{\infty} \Pr\{X = n\} s^n.$$

There are two useful properties of probability generating functions.

Proposition 9.1

If X_1 and X_2 are independent, then $f_{X_1} f_{X_2}(s) = f_{X_1+X_2}(s)$

Proof. If X_1 and X_2 are independent, then

$$\Pr\{X_1 = x_1\} \Pr\{X_2 = x_2\} = \Pr\{X_1 = x_1, X_2 = x_2\}.$$

We use this in the probability generating functions.

$$\begin{aligned} f_{X_1}(s) f_{X_2}(s) &= \sum_{n_1=-k_1}^{\infty} \Pr\{X_1 = n_1\} s^{n_1} \sum_{n_2=-k_2}^{\infty} \Pr\{X_2 = n_2\} s^{n_2} \\ &= \sum_{n=-k_1-k_2}^{\infty} s^n \left(\sum_{n_1=-k_1}^{n+k} \Pr\{X_1 = n_1, X_2 = n - n_1\} \right) \\ &= \sum_{n=-k_1-k_2}^{\infty} s^n \Pr\{X_1 + X_2 = n\} = f_{X_1+X_2}(s). \end{aligned}$$

We let $n_2 = n - n_1$. □

Proposition 9.2

The expected value can be found as $\mathbb{E}[X] = (f_X(s))'|_{s=1}$.

Proof. This follows by definition.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=-k}^{\infty} n \Pr\{X = n\} \\ &= (\sum \Pr\{X = n\} s^n)'|_{s=1} \end{aligned}$$

□

We look at some examples.

Example 9.3 (Bernoulli)
Let X be a Bernoulli random variable $\in \{-1, 1\}$. Then

$$f_X(s) = \frac{1}{2}s^{-1} + \frac{1}{2}s^1.$$

The expectation is 0. If $X_1 \dots X_n$ are i.i.d., then

$$f_{X_1+\dots+X_n} = \frac{1}{2^n}(s^{-1} + s)^n.$$

Example 9.4 (Poisson)
Let $X \sim \text{Poisson}(\alpha)$. That is, $\Pr\{X = k\} = e^{-\alpha}\frac{\alpha^k}{k!}$. Then

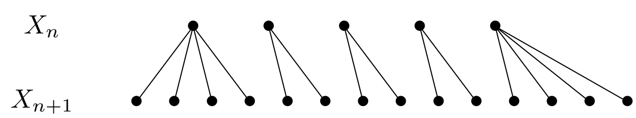
$$f_X(s) = \sum_{k=0}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!} = e^{-\alpha}e^{\alpha s}$$

The expectation is α. If $X_1 \dots X_n$ are i.i.d., then

$$f_{X_1+\dots+X_n} = e^{(s-1)(\alpha_1+\alpha_2+\dots+\alpha_n)}.$$

9.2 Branching processes

At time n, the process produces X_n particles, each of which produces random offspring and then dies.



In this class, we assume that each particle reproduces independently and with the same distribution ξ, such that $\Pr\{\xi = i)\} = p_i, p_i \geq 0 \sum p_i = 1$. A branching process is a THMC with transition probabilities

$$P(k, j) = \Pr\{y_1 + \dots + y_k = j\},$$

where $y_1, y_2, \dots, y_k$ are i.i.d. $\sim \xi$.

$$\begin{aligned}
\Pr\{X_{n+1} = j\} &= \sum_{k=0}^{\infty} \Pr\{X_n = k, X_{n+1} = j\} \\
&= \sum_{k=0}^{\infty} \Pr\{X_n = k\}\Pr\{X_{n+1} = j | X_n = k\} \\
&= \sum_{k=0}^{\infty} \Pr\{X_n = k\}\Pr\{y_1 + \dots + y_k = j\} \qquad (9.1)
\end{aligned}$$

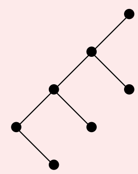
Theorem 9.5

$f_{X_{n+1}}(s) = f_{X_n}(f_\xi(s))$

Proof. Use equation 9.1 and propositions. □

Example 9.6

Let $p_0 = 1/2, p_2 = 1/2, X_0 = 1$.



TODO fill out the graph to be more fishy

We see that $f_{X_0}(s) = s$ and $f_{X_1}(s) = \frac{1}{2} + \frac{1}{2}s^2$. Continuing,

$$f_{X_n}(s) = f_\xi(f_s(f_\xi(\ldots f_\xi(s)))).$$

Let $A = \lim_{n\to\infty} \Pr\{X_n = 0\}$ be the extinction probability. If $\mathbb{E}[s] < 1$, then $A = 1$.

$$\Pr\{X_n = 0\} = f_{X_n}(0) = f_\xi(f_s(\ldots f_s(0)))$$

Something something $f_\xi(s) : [0,1] \to \mathbb{R}$. $f_\xi(1) = 1, f_\xi(0) = p_0 > 0$. $f_\xi(s)$ is increasing and continuous,

TODO pgfplots this graph

Theorem 9.7

The extinction probability A is the smallest positive root of $f_\xi(s) = s$.

Proof. Note that the limit $A = \lim_{n\to\infty} f_\xi^{(n)}(0)$ exists, since $f_\xi^{(n)}(0)$ is increasing and bounded by ≤ 1. In addition,

$$f_\xi(A) = f_\xi(\lim_{n\to\infty} f_\xi^{(n)}(0)) = \lim_{n\to\infty} f_\xi^{(n)}(0) = A.$$

Finally, if $\hat{A}$ is the smallest positive root of $f_\xi(s) = s$, then $\hat{A} > 0$ and tends to A. □

Example 9.8

Let $p_0 = 1/2, p_2 = 1/2$. Then $f_\xi(s) = 1/2 + 1/2s^2$, so $s = 1$ is the root to $f_\xi(s) = s$. This has extinction probability 1.

Example 9.9

Let $p_0 = 1/4, p_1 = 1/4, p_2 = 1/2$. Then $f_\xi(s) = 1/2 + 1/4s + 1/2s^2$, so $s = 1, 1/2$ are the roots to $f_\xi(s) = s$. This has extinction probability $1/2$, the smaller root.

10 April 13, 2017

10.1 Branching processes, ctd.

Recall that we have a distribution ξ with $\Pr\{\xi = 0\}, \Pr\{\xi = 1\}, \ldots$

Example 10.1

If $\Pr\{\xi = 0\} + \Pr\{\xi = 1\} = 1$, then the next step there are either 1 or 0 offspring. This process dies out with probability 1.

Assume that $X_0 = 1$. Recall that extinction probability

$$A = f_\xi(A) = \Pr\{\xi = 0\} + \Pr\{\xi = 1\} + \cdots + \Pr\{\xi = k\}.$$

This is because each particle dies out independently, so if there are k particles, they need to all die out.

Theorem 10.2

If $\mathbb{E}[\xi] \leq 1$, then $a = 1$, and if $\mathbb{E}[\xi] > 1$, then $f_\xi(s) = s$ has a unique positive root a such that $a < 1$.

[INSERT PICTURE FROM LAST LECTURE]

Graphically, $\mathbb{E}[\xi]$ is the slope at 1, so the left plot is for $\mathbb{E}[\xi] = 1$ and the right is $\mathbb{E}[\xi] < 1$.

We require that $f_\xi(s)$ is a convex function, $f''_\xi(s) > 0$.

$$\begin{aligned} f''_\xi(s) &= \frac{d^2}{ds^2}\left(\sum_{k=0}^{\infty} \Pr\{\xi = k\}\, s^k\right) \\ &= \sum_{k=2}^{\infty} \Pr\{\xi = k\}\, s^{k-2} \end{aligned}$$

We can also prove the theorem analytically. First a useful claim about convexity.

Claim 10.3. If $f(x)$ is convex, then the collection of points $\{(x, y) : y \geq f(x)\}$ is convex. That is, the intersection of this set with any line is a segment.

There are two cases to show.

Case 1 $\mathbb{E}[\xi] = f'_\xi(1) \leq 1$ implies that $f''_\xi(s) < f'_\xi(1) \leq 1$.

$$1 - f_\xi(s) = \int_s^1 f'_\xi(s)\, dt < 1 - s$$

For all $s < 1$, this implies that $f_\xi(s) > s$.

Case 2 $\mathbb{E}[\xi] = f'_\xi(1) > 1$ implies that

$$f_\xi(1 - \epsilon) \approx 1 - f'_\xi(1) \cdot \epsilon \leq 1 - \epsilon$$

for small ϵ. Thus, $f_\xi(s) < s$ when s is close to 1. Then $f_\xi(0) = \Pr\{\xi = 0\} > 0$, $f_\xi(s) > 0$ for $s = 0$, so there should be some intersection, so $f_\xi(s)$ has a root strictly < 1. $f_\xi(s)$ is convex, so $f_\xi(s)$ has at most 2 roots, and 1 is a root, so we have found the only other root.

Note, this proof does not cover the case $\Pr\{\xi = 1\}$ and $a = 0$, but we are handwaving.

Moving on,

$$f_{X_{n+1}}(s) = f_{X_n}(f_\xi(s)) = f_{X_0}(f_\xi(f_\xi(\dots f_\xi(s))))$$

10.2 Conditional probability

Our next subject of study involves conditional probability and expectation, which we introduce here.

Given two random variables X, Y, we know $\Pr\{X = x, Y = y\}$ for each pair (x, y). The conditional probability

$$\Pr\{Y = y | X = x\} = \frac{\Pr\{X = x, Y = y\}}{\Pr\{X = x\}}.$$

if $\Pr\{X = x\} \neq 0$. In the continuous case, this is problematic since $\Pr\{X = x\}$ is always 0. The usual expectation is a real number,

$$\mathbb{E}[f(x, y)] = \sum_{x,y} f(x, y) \Pr\{X = x, Y = y\}.$$

However, that is not true for conditional expectation.

Definition 10.4. Conditional expectation $\mathbb{E}[Y|X]$ is a random variable

$$X \to \sum_y y \Pr\{Y = y | X = x\}.$$

Example 10.5 (Dice)

Let X_1, X_2 be i.i.d. random variables distributed $\Pr\{X_i = i\} = 1/6$ for $i \in \{1, 2, \dots, 6\}$. It's intuitive to see that $\mathbb{E}[X_1 + X_2 | X_1] = \mathbb{E}[X_2] + X_1$.

Definition 10.6. The conditional expectation on several variables is

$$\mathbb{E}[Y | X_1, \dots, X_l] = \sum_y y \Pr\{Y = y | X_1 \dots X_l\}$$

for $\Pr\{X_1 \dots X_l\} > 0$.

There are several useful properties of conditional expectations. Let $a, b \in \mathbb{R}$ and X, Y be random variables.

Claim 10.7. $\mathbb{E}[a|X] = a$

Claim 10.8. $\mathbb{E}[aY_1 + bY_2 | X] = a\mathbb{E}[Y_1|X] + b\mathbb{E}[Y_2|X]$

Claim 10.9. If X and Y are independent, then $\mathbb{E}[Y|X] = \mathbb{E}[Y]$.

Claim 10.10. $\mathbb{E}\left[Yf(x)|X\right] = f(x)\mathbb{E}\left[Y|X\right]$

Proof. We prove this by bashing simple math.

$$\begin{aligned}\mathbb{E}\left[Yf(x)|X\right] &= \sum_y y \Pr\left\{Yf(x) = yf(x)\right\} \\ &= \sum_y yf(x) \Pr\left\{Y = y\right\} \\ &= f(x) \sum_y y \Pr\left\{Y = y\right\}\end{aligned}$$

□

Claim 10.11. $\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\right] = \mathbb{E}\left[Y\right]$

Proof.

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\right] &= \sum_x \Pr\left\{X = x\right\} \sum_y \Pr\left\{Y = y|X = x\right\} \\ &= \sum_x \sum_y y \Pr\left\{Y = y, X = x\right\} \\ &= \sum_y y \Pr\left\{Y = y\right\} = \mathbb{E}\left[Y\right]\end{aligned}$$

□

Claim 10.12. $\mathbb{E}\left[\mathbb{E}\left[Y|X_1, X_2\right]|X_1\right] = \mathbb{E}\left[Y|X_1\right]$

Proof.

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Y|X_1, X_2\right]\right] &= \sum_{x_1} \Pr\left\{X_2 = x_2|X_1 = x_1\right\} \sum_y \Pr\left\{Y = y|X_1 = x_1, X_2 = x_2\right\} \\ &= \sum_{x_1, y} \Pr\left\{X_2 = x_2|X_1 = x_1\right\} y \frac{\Pr\left\{Y = y, X_2 = x_2|X_1 = x_1\right\}}{\Pr\left\{X_2 = x_2|X_1 = x_1\right\}} \\ &= \sum_y y \Pr\left\{Y = y|X_1 = x_1\right\} = \mathbb{E}\left[Y|X_1\right]\end{aligned}$$

□

11 April 20, 2017

11.1 Martingales

In this class, we study stochastic processes, where we are given a sequence of random variables $X_0, X_1, \ldots$. A **martingale** is another type of stochastic process which depends on past events to predict a current mean.

Definition 11.1. $X_0, X_1, \ldots, X_n, X_{n+1}$ is a martingale if for any $k < n$,

$$\mathbb{E}\left[X_{k+1} | X_k, X_{k-1}, \ldots, X_1, X_0\right] = X_k.$$

Proposition 11.2

$\mathbb{E}\left[M_{k+2} | M_0 \ldots M_k\right] = M_k$

Proof.

$$M_k = \mathbb{E}\left[M_{k+1} | M_0 \ldots M_k\right] = \mathbb{E}\left[\mathbb{E}\left[M_{k+1}\right]\right]$$

FILL IT IN FROM PHOTO

□

Proposition 11.3

$\mathbb{E}\left[M_{k+k'} | M_0 \ldots M_k\right] = M_k$

Proposition 11.4

$\mathbb{E}\left[M_k\right] = \mathbb{E}\left[M_0\right]$

Proof. $\mathbb{E}\left[M_k | M_0\right] = M_0$, so $\mathbb{E}\left[M_k\right] = \mathbb{E}\left[\mathbb{E}\left[M_k | M_0\right]\right] = \mathbb{E}\left[M_0\right]$. □

Example 11.5

Let $X_1, X_2, \ldots, X_n$ be independent random variables, such that $\mathbb{E}\left[X_i\right] = 0$ for any i. Then $M_n = \sum_i X_i$ is a martingale.

Example 11.6

We play a game to gain 3 or -1 with equal probability each round, but we must pay 1 to play a round. Let $X_1, X_2, \ldots, X_n$ be independent random variables, such that $\mathbb{E}\left[X_i\right] = \mu$ for any i. Then $M_n = X_1 + \cdots + X_n - n\mu$ is a martingale.

Proof. Change variables with $\tilde{X}_i = X_i - \mu$, which are independent. By the previous example, $M_n = \sum_{i=1}^{n} \tilde{X}_i$ is a martingale. □

Example 11.7

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, where $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\left[X_i^2\right] = V$, where V is a constant. Then $M_n = (X_1 + X_2 + \cdots + X_n)^2 - nV$ is a martingale.

Proof. Let $S_n = \sum_{i=1}^n X_i$.

$$\begin{aligned}\mathbb{E}[M_{n+1}|M_n] &= \mathbb{E}[S_n + X_{n+1} - (n+1)V|S_n] \\ &= -(n+1)V + \mathbb{E}\left[S_n^2|S_n\right] + 2\mathbb{E}[S_n X_{n+1}|S_n]\end{aligned}$$

The last term can be expanded.

$$\begin{aligned}\mathbb{E}\left[X_{n+1}^2|S_n\right] &= -(n+1)V + S_n^2 + 2S_n\mathbb{E}[X_{n+1}|S_n] + \mathbb{E}\left[X_{n+1}^2\right] \\ &= -(n+1)V + S_n^2 + 0 + V \\ &= S_n^2 - nV = M_n\end{aligned}$$

□

Example 11.8

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, and $S_n = \sum_{i=1}^n X_i$. Then $M_1 = \frac{S_n}{n}$, $M_2 = \frac{S_{n-1}}{n-1}$, and $M_k = \frac{S_{n-k}}{n-k}$, down to S_1.

We can find that $\mathbb{E}[S_{k-1}|S_k] = \frac{k-1}{k} S_k$.

Example 11.9

Consider a branching process, $X_0 = 1$, and ξ is the offspring of one particle, where $\mathbb{E}[\xi] = \mu$. X_n is the number of particles at time n.

$\mathbb{E}[X_{n+1}|X_n] = X_n\mu$. Basically every step we expect each particle to produce μ particles, at each step.

"I don't know! I need your help..."—abufetov, in a pouty voice

Then the martingale we are looking for is $M_n = \frac{X_n}{\mu^n}$.

$$\begin{aligned}\mathbb{E}[M_{n+1}|M_n] &= \mathbb{E}\left[\frac{X_{n+1}}{\mu^{n+1}}|X_n\right] \\ &= \frac{\mu X_n}{\mu^{n+1}} = \frac{X_n}{\mu^n} = M_n\end{aligned}$$

12 April 27, 2017

12.1 Rip Liang is a bad influence

13 May 1, 2018

After a year I am returning to finish my notes!