# Stats 150, Fall 2019
Lecture 6, Tuesday, 9/17/2019

**CLASS ANNOUNCEMENTS:** Pitman notes that Homework 2 was intentionally hard to push us; however we may have relief at times like Homework 3.

Pitman opens to questions regarding irreducible, aperiodic, recurrent (including positive or null recurrent), or transient. For a nice transition probability matrix, there exists a stationary probability $\pi$ so that:

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix}$$

Pitman asks us to recall the single most important formula regarding recurrent chain and its expected time $\mathbb{E}_x T_x$ for returning to $x$, starting from $x$.
For a nice (irreducible, positive recurrent) chai,

$$\mathbb{E}_x T_x = \frac{1}{\pi(x)},$$

where $\pi(x)$ is the long run average (fraction of) time spent in state $x$. Recall that we defined that we hit $x$ exactly once per $x$-cycle on average, which is equal to once per $\mathbb{E}$ cycle.
This makes sense intuitively, where expecting to take a long time before returning to state $x$ corresponds to not being in state $x$ as often.

# 1 Transient Aspects of Chains

We want to discuss various transient aspects of chains, especially the distributions of hitting places and times. Pitman notes that both are interesting and we approach them with similar techniques. Hitting places is a bit easier to treat, so we start with those.

## 1.1 Hitting Places

Recall that we used the notation

$$T_A := \min\{n \geq 1 : X_n \in A\}$$
$$T_X := \min\{n \geq 1 : X_n = x\}$$

This is not trivial for $X$ with $X_0 = x$. For analysis of hitting places (and time), it's often easier to have our discrete-time sequence start at 0.

Hence we define:

$$V_A := \min\{n \geq 0 : X_n \in A\}.$$

Pitman notes that this is not a universal notation, and we may see $T, V, \tau$ used for this definition, but for this text and course, we will use $V_A$ for this purpose.

**Theorem 1.1.** (Durett 1.28, p. 55) Consider a Markov chain with state space $S$. Take two (necessarily disjoint) $A, B \subseteq S$.
We place $A$, the set of target states, at the top of a 2 (or higher dimensional) lattice, and place $B$ as all three remaining boundaries of the lattice (left, right, bottom edges).

## 1.2 Assumptions

Suppose we have $h : S \to \mathbb{R}$ so that:

$$h(a) = 1, \forall_{a \in A}$$
$$h(b) = 0, \forall_{b \in B}.$$

Then by definition, $C := S - (A \cup B)$, intuitively we think of $C$ as an interior set and $A, B$ as boundary sets.
Suppose that $h(x) = \sum_y \mathbb{P}(x, y)h(y)$. Very commonly, we'll write this in matrix notation, where $h$ is a column vector with $h = Ph$ (unlike we have done with row vectors like $\pi$).
Additionally, for all $x \in S$, suppose:

$$\mathbb{P}_x(V_A < \infty \text{ or } V_B < \infty) = 1 \iff \mathbb{P}_x(V_{A \cup B} < \infty) = 1.$$

Further, assume that the set of interior states are finite.
Because we put 0 in the definition of $V_A$, then this is trivial for $x \in A \cup B$.
The question is if we start inside the lattice, what conditions do we have to hit the boundaries?

We hold off on the Theorem's claim until after some discussion.

## 1.3 Side Example:

Find the distribution of $X_{V_{A \cup B}}$, where $V_{A \cup B}$ is the first hit of $A \cup B$. Notice that if we start at $x \in A \cup B$, we are there already, hence there is nothing to find. If we start at $x \notin A \cup B$, we may ask: What is $\mathbb{P}_x(X_{V_{A \cup B}} \in A)$? That is, starting at an interior state, what is the probability that we end up at the

## 1.4 Method of Solution (p. 54):

Pitman notes that the method is more important than the solution here. From the text, "Let $h(x)$ be the probability of hitting $A$ before $B$, starting from $X$". We call this technique **first step analysis**. "By considering what happens at the first step..."
That is, we assert that we start at $X_0 := x$, and we condition on time 1, $X_1$. Generalizing, let $Y$ be any nonnegative (for simplicity) random variable, and let $X_0, X_1, X_2, \dots$ be a Markov chain with transition matrix $P$. Consider $\mathbb{E}_x$ as a function of $Y$, and notice we can write, by summing out all $y \in Y$:

$$\begin{aligned}
\mathbb{E}_x Y &= \mathbb{E}_x \underbrace{\sum_y \mathbb{1}(X_1 = z)}\, Y \\
&= \sum_z \mathbb{E}_x \left[\mathbb{1}(X_1 = z) Y\right] \\
&= \sum_z \mathbb{P}_x(X_1 = z)\mathbb{E}(Y \mid X_1 = z).
\end{aligned}$$

Notice that we haven't used the Markov property, so we use that now:

$$\mathbb{E}_x Y = \sum_z P(x,z)\mathbb{E}(Y \mid X_1 = z),$$

which is simply computing the expectation by conditioning for general $Y$. Commonly, take the case where $Y$ is an indicator, for example $Y = \mathbb{1}(V_A < V_B)$. Then

$$\mathbb{E}_x Y = \mathbb{P}_x(V_A < V_B).$$

Pitman notes that something else is true as well via first step analysis. Take $x \notin A \cup B$. Look at the probability that $V_A$ happens before $V_B$, provided that we know $X_1 = z$. Now if $z$ is one of the boundary cases, this is trivial. So we treat in cases, using the Markov property:

$$\mathbb{P}_x(V_A < V_B \mid X_1 = z) = \begin{cases} 1, & z \in A \\ 0, & z \in B \\ \mathbb{P}_z(V_A < V_B) \end{cases},$$

where Pitman notes this boils down to simply being familiar with the formal notation for this to be clear.

**Remark:** Pitman poses a question: Does this probability of hitting $A$ before $B$ have anything to do with $P(c,\cdot)$ for $c \in A \cup B$? We agree on the edge cases, for starting in $A$ or $B$.
Now we make this key observation (which is not mentioned in the text). Because of our definitions (the possibility of being there at time zero), the answer is no! With this in mind, we modify the problem at hand to make the entire set of states, $A \cup B$ absorbing. That is, $P(c,c) := 1, \forall_{c \in A \cup B}$. That is to say we arrive, we stick there, and we solve the problem under these circumstances.

Not, notice that we agreed $h(x) := \mathbb{P}_x(V_A < V_B)$, for $x \notin A \cup B$, solves the equation:

$$\boxed{h(x) = \sum_y P(x,y)h(y)} \quad \text{(harmonic equation)}$$

Notice that if we make $A \cup B$ absorbing, then this harmonic equation above is true for ALL $x \in A \cup B$.
Now we finally arrive at the conclusion of the theorem.

# 2 Pitman's version of Durett's Theorem

Assume $P$ has $A \cup B$ as absorbing states. Assume further that $\mathbb{P}_x(\text{hit } A \cup B \text{ eventually}) = 1, \forall_{x \in S}$.
Then $h(x) := \mathbb{P}_x(\text{hit } A \text{ before } B)$ is the **UNIQUE** solution of $h = Ph$, where $h = \mathbb{1}_A$ is the indicator function on $A \cup B$.

This is fundamentally the same as Durett's theorem, but with some tinkering, we have a more elegant statement as here.

Break time.

Notice that $h = Ph$ is a very special equation, as a solution to certain problems. In order to understand this equation, it is important to understand what is $Pf$ for a function (column vector) $f$ (assume nonnegative and bounded so that we can make sense of the summations). Then the action of a matrix on a column vector simply gives us:

$$(Pf)(x) = \sum_{x \in S} P(x, y) f(y),$$

summing over all $x$ in the state space. $P$ gives the probability distribution, and $f$ simply gives the evaluations. Hence directly by our notation, we have:

$$(Pf)(x) = \mathbb{E}_x f(X_1).$$

Hence

$$(Ph)(x) = \mathbb{E}_x h(X_1)$$

as the meaning of $(Ph)(x)$. Another way to say this is by looking at the conditional expectation (knowing $X_0$):

$$\mathbb{E}[h(X_1) \mid X_0] = (Ph)(X_0)$$

Pitman makes the following claim:
If $h = Ph$ (that is, $h$ solves the harmonic equation), then the expectation (starting at $x$) of $h$ of any variable $(X_n)$ is:

$$\mathbb{E}_x[h(X_n)] = h(x),$$

which is true by $n = 1$ by $(Ph)(x) = \mathbb{E}_x h(X_1)$ from above (that is, $h = Ph$). Now, this is true for $n = 1, 2, 3, \dots$ by induction and the Markov property. If we trust this for now (we may revisit this later), we may want to assume that

$$h = \begin{cases} 1, & \text{on } A \\ 0, & \text{on } B, \end{cases}$$

then we can write:

$$h(x) = \mathbb{E}_x h(X_n) = \sum_{y \in S} P^n(x, y) h(y),$$

as our familiar notation for a Markov chain. Then we can equivalently write this as a summation over the three state cases:

$$h(x) = \sum_{y \in A} P^n(x, y) h(y) + \sum_{y \in B} P^n(x, y) h(y) + \sum_{y \in S - A - B} P^n(x, y) h(y).$$

Recall that we've set $A \cup B$ to be absorbing, so the first two terms are simply:

$$\sum_{y \in A} P^n(x, y) h(y) = \mathbb{P}(V_A \leq n)$$
$$\sum_{y \in B} P^n(x, y) h(y) = 0$$

Hence

$$h(x) = \mathbb{P}(V_A \le n) + 0 + \sum_{y \in S-A-B} P^n(x,y)h(y).$$

Now if we take $n \to \infty$, then

$$\lim_{n\to\infty} h(x) = \lim_{n\to\infty} \mathbb{E}_x h(X_n) = P_x(V_A \le \infty) + \underbrace{\lim_{n\to\infty} \sum_{y\in S-A-B} P^n(x,y)h(y)}_{=0}$$
$$= \mathbb{P}_x(V_A < \infty),$$

because $\mathbb{P}_x(\text{hit } A \cup B \text{ eventually}) = 1$ via our assumption.

# 3 Canonical Example: Gambler's Ruin for a fair coin

The state space is $S := \{0, 1, 2, \ldots, N\}$, and the goal state is $A := \{N\}$, and the bad state is $B = \{0\}$. The transition matrix then is:

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & \\ 0 & \frac{1}{2} & \frac{1}{2} & & 0\cdots \\ 0 & 0 & \ddots & & \end{bmatrix}$$

Now let $(X_n)$ be the simple random walk with absorbing states $\{0, N\}$. Then

$$\mathbb{P}_x(\text{hit } N \text{ before } 0) = h(x)$$

is desired. That is, $h = Ph$. Hence:

$$h(x) = \frac{1}{2}h(x+1) + \frac{1}{2}h(x-1), 0 < x < N$$
$$h(0) = h(0), h(N) = h(N)$$

We set the conditions:

$$h(N) := 1, \qquad h(0) := 0.$$

Now the harmonic equation $h(x) = \frac{1}{2}h(x+1) + \frac{1}{2}h(x-1)$ says that the graph of $h(x)$ is a straight line, passing through 0 and 1. Hence $h(x) = \frac{x}{N}$ as the unique solution to this system of equations. The theory is a bit more clever in that if we are certain to hit a boundary (as we have shown in lecture 5), then:

$$\mathbb{P}_x(\text{hit } N \text{ before } 0) = \frac{x}{N},$$

which Pitman notes is a famous result as due to Abraham deMoivre around 1730.

## 3.1 Now what about a biased coin?

Pitman asks what are the harmonic equations? We reason that this results in the same equations, with slight modifications:

$$h(x) = ph(x+1) + qh(x-1), 0 < x < N,$$

which we may solve via algebra as done in Durett (p. 58). Pitman shows us a more clever way, related to the idea of a Martingale. There is a discussion of this problem in the context of Martingale at the end of the text, as aspects of a hitting-time problem (we will revisit this at the end of the course). The idea is to say that $h(x) = x$ is no longer harmonic when $q \neq p$ (biased coin). Now we believe that

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

We check this:

$$\begin{aligned} Ph(x) &= P\left(\frac{q}{p}\right)^{x+1} + q\left(\frac{q}{p}\right)^{x-1} \\ &= \left(\frac{q}{p}\right)^x \left[\cancel{p}\frac{q}{\cancel{p}} + \cancel{q}\frac{p}{\cancel{q}}\right] \\ &= \left(\frac{q}{p}\right)^x. \end{aligned}$$

Now Pitman notes this is a bit clever, but it is not a bad idea to try that the harmonic equation is a power. As soon as we have found this, we can play this game again.
Pitman concludes that by $h = P^n h$, we have:

$$\begin{aligned} h(x) &= \mathbb{E}_x \left(\frac{q}{p}\right)^{X_n}, \forall_n \\ &= \left(\frac{q}{p}\right)^N \mathbb{P}_x(\text{hit } N \text{ before } n) + \left(\frac{q}{p}\right)^0 \mathbb{P}_x(\text{hit } 0 \text{ before } n) + \sum_{y \notin 0, N} \cdots \end{aligned}$$

Now taking $n \to \infty$, this final term goes to zero. Hence in the limit,

$$h(x) = \left(\frac{q}{p}\right)^X = \left(\frac{q}{p}\right)^N \mathbb{P}_x(hitN) + \left(\frac{q}{p}\right)^0 \mathbb{P}_x(\text{hit } 0),$$

and additionally $\mathbb{P}_x(\text{hit } N) + \mathbb{P}_x(\text{hit } 0) = 1$. Now we have two equations and two unknowns, and we can solve as normal.

Lecture ends here.