Math 456: Mathematical Modeling

Tuesday, March 6th, 2018

Markov Chains:
Exit distributions and the Strong Markov Property

Tuesday, March 6th, 2018

Last time

1. Weighted graphs.
2. Existence of stationary distributions (irreducible chains).
3. Informal definition of transient and recurrent states
4. Definition of stopping times.
5. Statement of the Strong Markov Property.

Today

Today we will see:

1. Stopping times and Exit distributions
2. The proof of the Strong Markov Property.
3. Probability of return: properties and computations.

Notation reminder

Probability with respect to initial state
Hitting time

Recall the notation from last time: Given a Markov Chain $X_0, X_1, X_2, \ldots$ we will write

$$\mathbb{P}_x[A] := \mathbb{P}[A \mid X_0 = x]$$
$$\mathbb{E}_x[Y] := \mathbb{E}[Y \mid X_0 = x]$$

Also recall: given a state y, we have the **first hitting time**

$$T_y := \min\{n \geq 1 \mid X_n = y\}$$

which is a positive, integer valued random variable.

Warm up

Consider the Gambler's ruin with a fair coin

$$\text{What is } \mathbb{P}_x[(\text{Winning})]?$$

Using exit times, and writing $A = \{0, M\}$

$$\text{What is } \mathbb{P}_x[X_{T_A} = M]?$$

Warm up

From the total probability formula, it follows that (for $x \neq 0, M$)

$$\mathbb{P}_x[X_{T_A} = M] = \tfrac{1}{2}\mathbb{P}_{x+1}[X_{T_A} = M] + \tfrac{1}{2}\mathbb{P}_{x-1}[X_{T_A} = M]$$

From here, with a little effort, one can show that

$$\mathbb{P}_x[X_{T_A} = M] = \frac{x}{M}, \quad x = 0, 1, \ldots$$

Warm up

From the total probability formula, it follows that (for $x \neq 0, M$)

$$\mathbb{P}_x[X_{T_A} = M] = \tfrac{1}{2}\mathbb{P}_{x+1}[X_{T_A} = M] + \tfrac{1}{2}\mathbb{P}_{x-1}[X_{T_A} = M]$$

From here, with a little effort, one can show that

$$\mathbb{P}_x[X_{T_A} = M] = \frac{x}{M}, \;\; x = 0, 1, \ldots$$

What if the coin is not a **fair**?

Exit probabilities

We are given:
An irreducible Markov chain with state space $\mathcal{S}$
A set $A \subset \mathcal{S}$.

The state of the chain at the time it first reaches A is given by

$$X_{T_A}$$

We may think of A as representing "exits" to the state space, so X_{T_A} represents the location of the chain where we exit.

Exit probabilities

To study the distribution of X_{T_A}, fix $y \in A$, and define

$$G_y(x) := \mathbb{P}[X_{T_A} = y \mid X_0 = x]$$

this defines a function $G_y : \mathcal{S} \to \mathbb{R}$.

This is known as the **Green function** of the chain (with Dirichlet conditions on A).

Exit probabilities

Observe: if $x \in A$, then $T_A = 0$, so

$$\mathbb{P}[X_{T_A} = y \mid X_0 = x] = \mathbb{P}[X_0 = y \mid X_0 = x]$$

and this is 1 or 0 depending on whether $x = y$ or not, so

$$G_y(x) = \begin{cases} 1 \text{ if } x \in A, x = y \\ 0 \text{ if } x \in A, x \neq y \end{cases}$$

What can we say for $x \notin A$?

Exit probabilities

For $x \notin A$, we divide and conquer.

In this case, the chain will move to a new state X_1, and we can condition on this new state.

The total probability formula + Markov property then says that

$$\begin{aligned} &\mathbb{P}[X_{T_A} = y \mid X_0 = x] \\ &= \sum_{x' \in \mathcal{S}} \mathbb{P}[X_{T_A} = y \mid X_1 = x'] \mathbb{P}[X_1 = x' \mid X_0 = x] \end{aligned}$$

Exit probabilities

Now, the identity

$$\begin{aligned}&\mathbb{P}[X_{T_A} = y \mid X_0 = x] \\ &= \sum_{x' \in \mathcal{S}} \mathbb{P}[X_{T_A} = y \mid X_1 = x'] \mathbb{P}[X_1 = x' \mid X_0 = x]\end{aligned}$$

can be rewritten as

$$G_y(x) = \sum_{x' \in \mathcal{S}} G_y(x') \mathrm{p}(x, x')$$

or, using that $\sum_{x'} \mathrm{p}(x, x') = 1$

$$\sum_{x' \in \mathcal{S}} (G_y(x') - G_y(x)) \mathrm{p}(x, x') = 0 \;\; \forall\, x \in \mathcal{S} \setminus A.$$

Exit probabilities

Example

Let us go back to the Gambler's ruin.
Assume we have a **biased coin** ($p \neq q$)

Take $y = M$. Then for $x \in \{1, \dots, M-1\}$

$$q(G_M(x+1) - G_M(x)) + p(G_M(x-1) - G_M(x)) = 0$$

With some cleverness, we can write this as a recurrence relation:

$$G_M(x+1) - G_M(x) = \frac{p}{q}(G_M(x) - G_M(x-1))$$

Exit probabilities

Example

Thus, we have for $x \in \{1, \ldots, M-1\}$

$$\begin{aligned} G_M(x+1) - G_M(x) &= \left(\frac{p}{q}\right)^x (G_M(1) - G_M(0)) \\ &= \left(\frac{p}{q}\right)^x G_M(1) \end{aligned}$$

this being since $G_M(0) = 0$.
Now, we can add up all these identities, from $x = 0$ up to some arbitrary element of $\{1, \ldots, M\}$, obtaining

$$G_M(x) - G_M(0) = \sum_{i=0}^{x-1} \left(\frac{p}{q}\right)^i G_M(1)$$

Exit probabilities

Example

Therefore (again, since $G_M(0) = 0$)

$$G_M(x) = \sum_{i=0}^{x-1} \left(\frac{p}{q}\right)^i G_M(1)$$

What about $G_M(1)$?. We use the elementary identity

$$\sum_{i=0}^{x-1} \left(\frac{p}{q}\right)^i = \frac{(p/q)^x - 1}{(p/q) - 1}$$

Substituting…

Exit probabilities

Example

. . . we obtain for every $x \in \{0, \ldots, M\}$ the formula

$$G_M(x) = \frac{(p/q)^x - 1}{(p/q) - 1} G_M(1)$$

Since $G_M(M) = 1$, it follows that

$$1 = \frac{(p/q)^M - 1}{(p/q) - 1} G_M(1)$$

Therefore,

$$G_M(1) = \frac{(p/q) - 1}{(p/q)^M - 1}$$

Exit probabilities

Example

Thus, we have for $x \in \{1, \dots, M-1\}$

$$\begin{aligned} G_M(x+1) - G_M(x) &= \left(\frac{p}{q}\right)^x (G_M(1) - G_M(0)) \\ &= \left(\frac{p}{q}\right)^x G_M(1) \end{aligned}$$

this being since $G_M(0) = 0$. At the same time, we have $G_M(M) = 1$, so

$$1 - G_M(M-1) = \left(\frac{p}{q}\right)^{M-1} G_M(1)$$

Exit probabilities

Example

In conclusion, for every x

$$G_M(x) = \frac{(p/q)^x - 1}{(p/q)^{M+1} - 1}$$

The Laplacian Operator

To every function $f : \mathcal{S} \to \mathbb{R}$ we associate a new one,

$$\Delta f(x) := \sum_{x' \in \mathcal{S}} (f(x')) - f(x))\mathrm{p}(x, x')$$

and known as the Laplacian of f.

The Laplacian Operator

Then, for a given set $A \in \mathcal{S}$ and $y \in A$, we have

$$\begin{cases} G_y(x) & = \delta_{xy} \text{ if } x \in A, \\ \Delta G_y(x) & = 0 \text{ if } x \in \mathcal{S} \setminus A. \end{cases}$$

Solving the above amounts to a linear algebra problem involving the transition matrix.

The Laplacian Operator

This means we can compute an exact formula for the exit distribution for a set A, without even running a single simulation.

The flip side is, if one wants to approximate the solution to the above problem, one could run lots of simulations, and tally the “exit points”.

Math 456: Mathematical Modeling

Thursday, March 8th, 2018

Markov Chains: The Strong Markov Property (…and a bit more about the Laplacian)

Thursday, March 8th, 2018

Last time

1. A bit on calculus on weighted graphs
2. Definitions: gradient, divergence, and Laplacian.
3. The Laplacian as a measure of a function's smoothness.
4. Exit distributions. Example: Gambler's ruin.
5. Exit distributions and the associated graph Laplacian.

Today

1. The Laplacian and stationary distributions.
2. The Strong Markov property.
3. Probability of return: properties and computations.

Δ vs. Stationary distributions

Last time we defined the Laplacian on a weighted graph.
For the graph associated to a Markov chain ($G = \mathcal{S}$, weights given by transition probabilities), for a function f on the state space, its Laplacian is given by

$$\Delta f(x) = \sum_{y \in G} (f(y) - f(x))\mathrm{p}(x, y)$$

Δ vs. Stationary distributions

Let π be a stationary distribution for the chain, and let f be a function solving $\Delta f = 0$.

How do the two equations compare?

We have

$$\pi(x) = \sum_{y \in G} \pi(y) \mathrm{p}(y, x)$$

$$\sum_{y \in G} (f(y) - f(x)) \mathrm{p}(x, y) = 0$$

Δ vs. Stationary distributions

Let π be a stationary distribution for the chain, and let f be a function solving $\Delta f = 0$.

How do the two equations compare?

...rewriting the second one,

$$\pi(x) = \sum_{y \in G} \pi(y)\mathrm{p}(y, x)$$

$$f(x) = \sum_{y \in G} f(y)\mathrm{p}(x, y)$$

where we have used that $\sum_y \mathrm{p}(x, y) = 1$ for every x.
What is the difference? We have $\mathrm{p}(x, y)$ in one, and $\mathrm{p}(y, x)$ in the other (and we are summing in y in both cases).

Δ vs. Stationary distributions

What does this look like in vector notation?

Solving $\Delta f = 0$ in all of G means

$$\mathrm{p} f = f$$

and as we know, π being stationary means

$$\mathrm{p}^t \pi = \pi$$

The two conditions are very similar, but they are not **exactly** the same in general!

If $\pi^t = \pi$, they of course coincide! In general, if p is **doubly stochastic** then they coincide.

Recall the **Strong Markov Property** from last time.

Theorem (Strong Markov Property)

Let X_n denote a Markov Chain. If T is a stopping time and $Y_n := X_{T+n}$, then Y_n is also a Markov chain. Moreover, the transition probability for Y_n is the same as the one for X_n.

Theorem (Strong Markov Property)

Let X_n denote a Markov Chain. If T is a stopping time and $Y_n := X_{T+n}$, then Y_n is also a Markov chain. Moreover, the transition probability for Y_n is the same as the one for X_n.

What does this mean?

It means that given states $\alpha_1, \ldots, \alpha_n$, we have

$$\mathbb{P}(Y_{n+1} = \alpha_{n+1} \mid Y_n = \alpha_n, \ldots Y_1 = \alpha_1) = \mathrm{p}(\alpha_n, \alpha_{n+1})$$

where $\mathrm{p}(x, y)$ is the transition matrix for the original chain.

Classification of states

The Strong Markov Property

Theorem (Strong Markov Property)

Let X_n denote a Markov Chain. If T is a stopping time and $Y_n := X_{T+n}$, then Y_n is also a Markov chain. Moreover, the transition probability for Y_n is the same as the one for X_n.

What does this mean?
Equivalently, it means that

$$\mathbb{P}(Y_{n+1} = \alpha_{n+1}, Y_n = \alpha_n, \dots Y_1 = \alpha_1)$$

is equal to $\mathrm{p}(\alpha_n, \alpha_{n+1})\mathrm{p}(\alpha_{n-1}, \alpha_n) \dots \mathrm{p}(\alpha_1, \alpha_2)\mathbb{P}(Y_1 = \alpha_1)$.

(we shall skip the proof of this for now and revisit it in a couple of lectures)

Probability of return

Now, armed with the Strong Markov Property, let us put it to use to analyze various questions about the long time behavior of a chain.

Remember: we were trying to study the even that starting from x, we eventually reach some other state y at some (random) time T_y, as well as the probability that starting from x, we eventually return to x itself at some later (random) time T_x.

Probability of return

Recall
The first hitting time,

$$T_y = \min\{n \geq 0 \mid X_n = y\}$$

with the convention that

$$T_y = \infty$$

in the event that $\{X_n \neq y \; \forall n\}$.

Probability of return

also recall (with the same convention in case the respective event is empty)

$$T_x^k := \min\{n > T_x^{k-1} \mid X_n = x\}$$
$$T_A^k := \min\{n > T_A^{k-1} \mid X_n \in A\}$$

This is known as the **time of the k-th return to x** or the **k-th hitting time for A**.

Probability of return

also recall (with the same convention in case the respective event is empty)

$$T_x^k := \min\{n > T_x^{k-1} \mid X_n = x\}$$
$$T_A^k := \min\{n > T_A^{k-1} \mid X_n \in A\}$$

This is known as the **time of the k-th return to** x or the k**-th hitting time for** A.

Evidently, each T_x^k is a stopping time.

Probability of return

The **probability of returning to** x

$$\rho_{xx} := \mathbb{P}_x[T_x < \infty]$$

The probability of **visiting** y **starting from** x.

$$\rho_{xy} := \mathbb{P}_x[T_y < \infty]$$

Note: $\rho_{xy} > 0$ if and only if $x \mapsto y$.

Probability of return

Using the Strong Markov Property, we can show the following:

Lemma (See equation (1.4) in Durrett)

$$\mathbb{P}_x[T_x^k < \infty] = \rho_{xx}^k$$

Probability of return

Using the Strong Markov Property, we can show the following:

Lemma (See equation (1.4) in Durrett)

$$\mathbb{P}_x[T_x^k < \infty] = \rho_{xx}^k$$

Note: In particular, if $\rho_{xx} < 1$, the probability of returning to the state k times goes to zero exponentially fast as $k \to 0$.

Probability of return

Proof.

Observe that,

$$\begin{aligned}\mathbb{P}_x[T_x^k < \infty] &= \mathbb{P}_x[T_x^k < \infty, T_x^{k-1} < \infty] \\ &= \mathbb{P}_x[X_{T_x^{k-1}+n} = x \ \text{ for some } \ n, T_x^{k-1} < \infty]\end{aligned}$$

Therefore

$$\begin{aligned}&\mathbb{P}_x[T_x^k < \infty] \\ &= \mathbb{P}_x[X_{T_x^{k-1}+n} = x \ \text{ for some } \ n \mid T_x^{k-1} < \infty]\mathbb{P}[T_x^{k-1} < \infty]\end{aligned}$$

Probability of return

Proof.

Let $Y_n := X_{T_x^{k-1}+n}$. By the Strong Markov Property

$$\begin{aligned}
&\mathbb{P}_x[X_{T_x^{k-1}+n} = x \text{ for some } n \mid T_x^{k-1} < \infty] \\
&= \mathbb{P}[Y_n = x \text{ for some n} \mid T_x^{k-1} < \infty] \\
&= \mathbb{P}[X_n = x \text{ for some n} \mid X_0 = x] \\
&= \mathbb{P}[T_x < \infty \mid X_0 = x] = \mathbb{P}_x[T_x < \infty]
\end{aligned}$$

Repeating this argument, we obtain the desired formula.

□

Probability of return

Now, reviewing the definition of transient and recurrent from last time, we agree to say that

1. a state x will be called **recurrent** if $\rho_{xx} = 1$.
2. a state x will be called **transient** if $\rho_{xx} < 1$.

Probability of return

The previous lemma and these definitions lead us to the following realizations (one trivial, two not so trivial)

1. If x is transient, then the probability of returning to x k times decreases geometrically with k.
2. If x is recurrent and $k \in \mathbb{N}$, then with probability equal to 1, the chain will be on the state x at least k times.
3. Actually, if x is recurrent, then with probability 1, starting from x, the chain will revisit x infinitely manytimes.

This last point is particularly delicate, and the proof uses an important tool in probability known as the Borel-Cantelli lemma.