Stationary Distributions of Markov Chains

Will Perkins

April 4, 2013

Back to Markov Chains

Recall a discrete time, discrete space Markov Chain, is a process X_n so that

$$\Pr[X_n = x_n | X_{n-1} = x_{n-1}, \ldots X_1 = x_1] = \Pr[X_n = x_n | X_{n-1} = x_{n-1}]$$

A time-homogeneous Markov Chain, the transition rates do not depend on time. I.e.

$$\Pr[X_n = j | X_{n-1} = i] = \Pr[X_k = i | X_{k-1} = i] =: p_{ij}$$

Transition Matrix

The transition matrix P of a MC has entries
$P_{ij} = \Pr[X_n = j | X_{n-1} = i]$.

- The entires are non-negative.
- The rows sum to 1.
- Called a stochastic matrix.

If X_0 has distribution μ_0 then X_1 has distribution $\mu_0 P$ (matrix-vector multiplication), and X_n has distribution $\mu_0 P^n$.

Some Terminology

- A state i is recurrent if the probability X_n returns to i given that it starts at i is 1.
- A state that is not recurrent is transient.
- A recurrent state is positive recurrent if the expected return time is finite, otherwise it is null recurrent.
- We proved that i is transient if and only if $\sum_n p_{ii}(n) < \infty$.
- A state i is periodic with period r if the greatest common divisors of the times n so that $p_{ii}(n) > 0$ is r. If $r > 1$ we say i is periodic.

Classifying and Decomposing Markov Chains

We say that a state i *communicates* with a state j (written $i \to j$) if there is a positive probability that the chain visits j after it starts at i.

i and j *intercommunicate* if $i \to j$ and $j \to i$.

Theorem

If i and j intercommunicate, then

- *i and j are either both (transient, null recurrent, positive recurent) or neither is.*
- *i and j have the same period.*

Classifying and Decomposing Markov Chains

We call a subset C of the state space $\mathcal{X}$ *closed*, if $p_{ij} = 0$ for all $i \in C$, $j \notin C$. I.e. the chain cannot escape from C.
Eg. A branching process has a closed subset consisting of one state.

A subset C of $\mathcal{X}$ is *irreducible* if i and j intercommunicate for all $i, j \in C$.

Classifying and Decomposing Markov Chains

Theorem (Decomposition Theorem)

The state space $\mathcal{X}$ of a Markov Chain can be decomposed uniquely as

$$\mathcal{X} = T \cup C_1 \cup C_2 \cup \cdots$$

where T is the set of all transient states, and each C_i is closed and irreducible.

Decompose a branching process, a simple random walk, and a random walk on a finite, disconnected graph.

Random Walks on Graphs

One very natural class of Markov Chains are random walks on graphs.

- A simple random walk on a graph G moves uniformly to a random neighbor at each step.
- A lazy random walk on a graph remains where it is with probability $1/2$ and with probability $1/2$ moves to a uniformly chosen random neighbor.
- If we allow directed, weighted edges and loops, then random walks on graphs can represent all discrete time, discrete space Markov Chains.

We will often use these as examples, and refer to the graph instead of the chain.

Stationary Distribution

Definition

A probability measure μ on the state space $\mathcal{X}$ of a Markov chain is a stationary measure if

$$\sum_{i \in \mathcal{X}} \mu(i) p_{ij} = \mu(j)$$

If we think of μ as a vector, then the condition is:

$$\mu P = \mu$$

Notice that we can always find a vector that satisfies this equation, but not necessarily a probability vector (non-negative, sums to 1).

Does a branching process have a stationary distribution? SRW?

The Ehrenfest Chain

Another good example is the Ehrenfest chain, a simple model of gas moving between two containers.
We have two urns, and R balls. A state is described by the number of balls in urn 1. At each step, we pick a ball at random and move it to the other urn.

Does the Ehrenfest Chain have a stationary distribution?

Existence of Stationary Distributions

Theorem

An irreducible Markov Chain has a stationary distribution if and only if it is positive recurrent.

Proof: Fix a positive recurrent state k. Assume that $X_0 = k$. Let T_k be the first return time to state k. Let N_i be the number of visits to state i before time T_k. And let $\rho_i(k) = \mathbb{E}N_i$. (note that $\rho_k(k) = 1$).

We will show that $\rho(k)P = \rho(k)$. Notice that $\sum_i \rho_i(k) < \infty$ since k is positive recurrent.

Existence of Stationary Distributions

$$\rho_i(k) = \sum_{n=1}^{\infty} \Pr[X_n = i \wedge T_k \geq n | X_0 = k]$$

$$= \sum_{n=1}^{\infty} \sum_{j \neq k} \Pr[X_n = i, X_{n-1} = j, T_k \geq n | X_0 = k]$$

$$= \sum_{n=1}^{\infty} \sum_{j \neq k} \Pr[X_{n-1} = j, T_k \geq n | X_0 = k] p_{ji}$$

$$= p_{ki} + \sum_{n=2}^{\infty} \sum_{j \neq k} \Pr[X_{n-1} = j, T_k \geq n | X_0 = k] p_{ji}$$

$$= p_{ki} + \sum_{j \neq k} \sum_{n=2}^{\infty} \Pr[X_{n-1} = j, T_k \geq n | X_0 = k] p_{ji}$$

Existence of Stationary Distributions

$$= p_{ki} + \sum_{j \neq k} \sum_{n=1}^{\infty} \Pr[X_n = j, T_k \geq n-1 | X_0 = k] p_{ji}$$

$$= p_{ki} + \sum_{j \neq k} \rho_j(k) p_{ji}$$

$$= \rho_k(k) p_{ki} + \sum_{j \neq k} \rho_j(k) p_{ji}$$

$$\rho_i(k) = \sum_{j \in \mathcal{X}} \rho_j(k) p_{ji}$$

which says,

$$\rho(k) = \rho(k) P$$

Existence of Stationary Distributions

Now define $\mu(i) = \frac{\rho_i(k)}{\sum_j \rho_j(k)}$ to get a stationary distribution.

Uniqueness of the Stationary Distribution

Assume that an irreducible, positive recurrent MC has a stationary distribution μ. Let X_0 have distribution μ, and let $\tau_j = \mathbb{E}T_j$, the mean recurrence time of state j.

$$\mu_j \tau_j = \sum_{n=1}^{\infty} \Pr[T_j \geq n, X_0 = j]$$

$$= \Pr[X_0 = j] + \sum_{n=2}^{\infty} \Pr[X_m \neq j, 1 \leq m \leq n-1] - \Pr[X_m \neq j, 0 \leq m \leq n-1]$$

$$= \Pr[X_0 = j] + \sum_{n=2}^{\infty} \Pr[X_m \neq j, 0 \leq m \leq n-2] - \Pr[X_m \neq j, 0 \leq m \leq n-1]$$

a telescoping sum!

$$= \Pr[X_0 = j] + \Pr[X_0 \neq j] - \lim_{n \to \infty} \Pr[X_m \neq j, 0 \leq m \leq n-1]$$

$$= 1$$

Uniqueness of the Stationary Distribution

So we've shown that for any stationary distribution of an irreducible, positive recurrent MC, $\mu(j) = 1/\tau_j$. So it is unique.

Convergence to the Stationary Distribution

If μ is a stationary distribution of a MC X_n, then if X_n has distribution μ, X_{n+1} also has distribution μ.
What we would like to know is whether, for any starting distribution, X_n converges in distribution to μ.

Negative example: a simple periodic markov chain.

The Limit Theorem

Theorem

For an irreducible, aperiodic Markov chain,

$$\lim_{n \to \infty} p_{ij}(n) = \frac{1}{\tau_j}$$

for any $i, j \in \mathcal{X}$.

Note that for a irreducible, aperiodic, positive recurrent chain this implies

$$\Pr[X_n = j] \to \frac{1}{\tau_j}$$

Proof of the Limit Theorem

We prove the theorem in the positive recurrent case with a 'coupling' of two Markov chains. A coupling of two processes is a way to define them on the same probability space so that their marginal distributions are correct.

In our case X_n will be our markov chain with $X_0 = i$ and Y_n the same Markov chain with $Y_0 = k$. We will do a simple coupling: X_n and Y_n will be independent.

Proof of the Limit Theorem

Now pick a state $x \in \mathcal{X}$. Let T_x be the smallest n so that $X_n = Y_n = x$. Then

$$p_i j(n) \leq p_{kj}(n) + \Pr[T_x > n]$$

since conditioned on $T_x \leq n$, X_n and Y_n have the same distribution.

Now we claim that $\Pr[T_x > n] \to 0$. Why? Use aperiodic, irreducible, postive recurrent. Aperiodicity is need to show that $Z_n = (X_n, Y_n)$ is irreducible.