

Convergence Rate of Markov Chains

Will Perkins

April 16, 2013

Last class we saw that if X_n is an irreducible, aperiodic, positive recurrent Markov chain, then there exists a stationary distribution μ on the state space \mathcal{X} , so that no matter where the chain starts,

$$X_n \rightarrow \mu$$

in distribution as $n \rightarrow \infty$.

A very natural question to ask is “How long does it take?”
Why do we care about the rate of convergence to the stationary distribution?

Sampling Algorithms

How do you sample from a given probability distribution (with a computer)?

It depends on what description you have of the distribution.

- 1 If you know the distribution function $F(t)$, then you can sample a uniform $[0, 1]$ rv U and take $X = F^{-1}(U)$. What is the distribution of X ?
- 2 Rejection sampling: If we can draw a box around the graph of the density function $f(x)$, then we can sample a uniform point under the curve, by taking a uniform point in the box and rejecting if the point is above the curve. [Picture]. Rejection sampling is useful when you have procedure to compute $f(x)$ but the form is difficult to work with analytically. Analogue for discrete distributions?

Sampling Algorithms

Those are the easy cases. Sometimes we only know indirect information about the distribution we would like to sample from. Eg. consider a given graph G on n vertices. Say we want to sample an *independent set* from the graph, uniformly over all independent sets.

Where to start? We don't even know the number of independent sets of a graph (it's a difficult computational problem).

Markov Chain Monte Carlo

One idea comes from stationary distributions. If we could somehow set up a Markov Chain so that our desired distribution is stationary, then we could just run the chain and (eventually) the distribution would converge to the desired distribution. So we could just sample X_N for large enough N .

Two questions:

- 1 How can we set up such a Markov Chain?
- 2 How large does N have to be?

We will start with the second question today.

Total Variation Distance

What does it mean for the distribution of X_n to be close to μ ?

Recall Total Variation Distance:

$$\|P - Q\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$$

Exercise: If $\|P_n - Q\|_{TV} \rightarrow 0$, then $P_n \rightarrow Q$ in distribution.

Alternate formulation:

$$\|P - Q\|_{TV} = \max_{A \subset \mathcal{X}} P(A) - Q(A)$$

Mixing Times

Let X_n have distribution π_n^j when $X_0 = j$, and let μ be the stationary distribution of the MC.

Definition

The mixing time of the Markov Chain X_n is

$$\tau_{1/4} = \inf\{n : \sup_{j \in \mathcal{X}} \|\pi_n^j - \mu\|_{TV} < 1/4\}$$

In other words, no matter in which state the chain starts, after n steps the distribution is distance at most $1/4$ from stationary. We could pick any $1/4 < 1/2$ and τ would be the same up to a constant factor.

Examples

Usually we are interested in the asymptotics of the mixing times of large Markov Chains, indexed by some parameter n . Eg. random walks on graphs of n vertices.

- What is the mixing time of a random walk on a complete graph K_n ?
- What is the mixing time of a random walk on a cycle C_n ?

We are looking for upper and lower bounds that hopefully differ by only a constant factor.

Let π be a probability distribution on \mathcal{X} . The *detailed balance* equations are:

$$\pi(i)p_{ij} = \pi(j)p_{ji}$$

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

Lemma

If π satisfies the detailed balance equations then it is the unique stationary distribution.

A chain with such a distribution is called *reversible*.

Proof:

Reversibility is a good way to find a stationary distribution.

Example: Say the transition matrix of a finite Markov chain is symmetric: $p_{ij} = p_{ji}$. What is the stationary distribution of such a Markov chain?

Exercise (in class): For a given graph G on n vertices, find a Markov Chain on the independent sets of the graph so that the stationary distribution of the chain is the uniform distribution over all independent sets.

Ehrenfest Urn (lazy version): n balls lie in two Urns A and B . At each step, pick a ball uniformly at random and place it in one of the bins uniformly at random.

What is the stationary distribution of the Ehrenfest Urn?

Lazy Random Walk on the hypercube: State space $\{0, 1\}^n$, at each step pick a coordinate and with probability $1/2$ flip it. What is the stationary distribution of this random walk?

How are the two chains related?

The Metropolis Algorithm

Say we want to sample from a different distribution, not necessarily uniform. Can we change the transition rates in such a way that our desired distribution is stationary?

Amazingly, yes. Say we have a distribution π over \mathcal{X} so that

$$\pi(x) = \frac{w(x)}{\sum_{y \in \mathcal{X}} w(y)}$$

I.e. we know the proportions but not the normalizing constant (and \mathcal{X} is much too big to compute it).

The Metropolis Algorithm

Metropolis-Hastings Algorithm

- 1 Create a graph structure on \mathcal{X} so the graph is connected and has maximum degree D .
- 2 Define the following transition probabilities:
 - 1 $p(x, y) = \frac{1}{2D} (\min\{w(y)/w(x), 1\})$ if x and y are neighbors.
 - 2 $p(x, y) = 0$ if x and y are not neighbors
 - 3 $p(x, x) = 1 - \sum_{y \sim x} p(x, y)$
- 3 Check that this Markov chain is irreducible, aperiodic, reversible and has stationary distribution π .

Example

Say we want to sample large independent sets from a graph G . I.e.

$$P(I) = \frac{\lambda^{|I|}}{Z}$$

where $Z = \sum_J \lambda^{|J|}$ where the sum is over all independent sets. Note that this distribution gives more weight to the largest independent sets.

Use the Metropolis Algorithm to find a Markov Chain with this distribution as the stationary distribution.

How to Bound Mixing Times

The first technique we will consider is Coupling. Recall: A coupling of two Markov Chains X_n and Y_n is a way to defined them on the same probability space - i.e. specify a joint distribution so the marginal distributions are correct but somehow the dependence of the chains tells us something.

We can specify a coupling of P and Q both distributions on \mathcal{X} as a probability distribution μ on $\mathcal{X} \times \mathcal{X}$ so that

$$P(x) = \sum_y \mu(x, y)$$

$$Q(y) = \sum_x \mu(x, y)$$

How to Bound Mixing Times

The *maximal coupling* of P and Q is the coupling that maximizes

$$\sum_x \mu(x, x)$$

the probabilities along the diagonal.

Lemma

Let μ be a maximal coupling of P and Q . Then

$$\|P - Q\|_{TV} = 1 - \sum_x \mu(x, x)$$

Proof:

How to Bound Mixing Times

We will look at a particular type of coupling of two Markov chains. Let X_n and Y_n be two copies of the same Markov chain but with different starting positions: $X_0 = x$ for some arbitrary state x and Y_0 will have distribution π , the stationary distribution of the chain.

Proposition

$$\|X_n - Y_n\|_{TV} \leq \Pr[X_n \neq Y_n]$$

Proof:

How to Bound Mixing Times

In particular, Y_n has distribution π for all n , so we have

$$\|X_n - \pi\|_{TV} \leq \Pr[X_n \neq Y_n]$$

One useful feature we can require of our coupling is that once X_n and Y_n collide, they move together. In this case we have:

Proposition

Let τ be the first time $X_n = Y_n$. Then

$$\|X_n - \pi\|_{TV} \leq \Pr[\tau > n]$$

Proof:

The Ehrenfest Urn: How can we couple X_n and Y_n ?

The idea is to look at the refined chain, the random walk on the hypercube. Here there is a good candidate for a coupling: To update, pick one of the n coordinates at random for both chains together, and update to the same value.

Check that this is a valid coupling and that once the chains collide they move together.

What is the collision time?

What about lower bounds?