Markov Chains Mixing Times

Ori Gurel-Gurevich Ariel Yadin

December 18, 2007

1 stationary distribution

Definition.

$$\pi = \pi P$$

Example. *SRW on graph. stationary dist = degree.*

Frog example is a special case (even though it looks directed at first glance).

Example. *SRW on directed graph. no formula for stationary dist. special case: if in-degree=out-degree then like former case. note: in these cases, distribution on edges is uniform. In particular, the uniform distribution on the deck of cards is stationary. A special case of RW on groups.*

Note: any Markov Chain can be represented as a RW on a directed weighted graph. Vertices=States, Transitions=Edges.

Does a stationary distribution always exist? Is it unique? Does the distribution converges to a stationary distribution?

1.1 existence

Theorem. *For any Markov chain P on a finite state space, a stationary distribution π exists - $\pi = \pi P$.*

Note: not true for infinite state space, not even if we want a stationary measure. example: shift on $\mathbb{N}$.

There are many proofs: analytic, algebraic, probabilistic.

Proof. Brouwer fixed point theorem. P is a continuous (linear) function from the n-dimensional simplex to itself, and therefore has a fixed point. This is exactly the stationary distribution. □

Proof. pick some distribution μ_0 and let $\mu_t = \mu_0 P^t$. Let $\pi_t = \sum_{s<t} \mu_s/t$ be the sequence of averages. The space of all probability measures on Ω is compact, so this sequence must have a sub-limit, π. It is easy to check that $\pi = \pi P$, since $\pi_t(P - I) = \pi_{t+1} - \pi_t = (\mu_t - \mu_0)/n \rightarrow 0$. □

Actually, the limit of averages (called the *Cesaro limit*) exists, so there's no need to take a sub-limit.

We will see another, more explicit proof later.

1.2 uniqueness

Obviously, sometimes the stationary distribution is not unique.

Trivial example: if the graph is disconnected then each of the stationary distributions of the components of the graph is stationary.

Definition. *If it is possible to get from any state to any state then the chain is called* irreducible. *Equivalently, the directed graph associated with the chain is strongly connected. Equivalently,* $\forall x, v \in \Omega \exists n \in \mathbb{N}\ P^n(x, y) > 0$.

Note that irreducibility is a property of the underlying graph, and is not affected by the actual probability ,as long as they're positive.

Is irreducibility enough for uniqueness? Yes. But first let's give a more constructive proof of the existence of a stationary distribution.

hitting time, τ_x.

return time, τ_x^+

$E_x(\tau_x^+) < \infty$.

The hitting time is a special case of a *stopping time*, a very useful concept later on.

Fix a vertex z. Let $\pi(y) = \mathbf{E}_z(\texttt{\# of visits to y before returning to z}) = \sum_{t\geq 0} \mathbf{P}_z(X_t = y, t < \tau_z^+)$.

Theorem. $\pi P = \pi$.

Recall that our matrix is stochastic, i.e. $P\mathbf{1} = P$

Definition. *a function is* harmonic *if* $h = Ph$*, i.e.* $h(x) = \sum_{y\in\Omega} P(x,y)h(y)$.

A constant function is harmonic.

Theorem. *A harmonic function of an irreducible Markov chain is constant.*

Proof. □

Exercise: prove that one-way-connectivity is enough.

Theorem. *The stationary distribution of an irreducible Markov chain is unique.*

Proof. rank of P_I is $n-1$ because the dimension of $\{h|(P-I)h = 0\}$ is 1. Therefore the dimension of $\{\pi|\pi(P-I) = 0\}$ must also be 1. □

Again, one-way-connectivity is enough.

This is a nice illustration of the usefulness of the algebraic representation of Markov chains.

alternative. If both π and σ are stationary, then so is $(1+x)\pi - x\sigma$. Choosing the right x will produce a stationary distribution with one (or more) the states having 0 probability, which contradicts the next lemma. □

Lemma. *If* π *is a stationary distribution for an irreducible Markov chain, then* $\pi(x) > 0$ *for any* $x \in \Omega$.

Proof. This is very similar to the proof about harmonic functions. If $\pi(x_0) > 0$ for some x_0 then for any y there's r such that $P^r(x,y) > 0$ and thus $p(y) = \sum_x \pi(x)P^r(x,y) \geq \pi(x_0)P^r(x_0,y) > 0$. □

What happens if the chain is not irreducible?

Definition. *$x \to y$ if there is a directed path from x to y, i.e. $\exists t \geq 0 P^t(x, y) > 0$.*

Definition. *$x \leftrightarrow y$ if $x \to y$ and $y \to x$.*

Exercise. *prove that $\leftrightarrow$ is an equivalent relation.*

The chain is irreducible iff there is only one equivalent class. Otherwise, we get a Directed Acyclic Graph (DAG) of equivalent classes, where $C \to D$ if there are $x \in C$ and $y \in D$ such that $P(x, y) > 0$. Alternative definition with $x \to y$ yields the corresponding transitive closure, i.e. a partial order. This graph is called the communication graph (relation, classes, etc.).

Exercise. *prove it's a DAG.*

No matter where the chain is started, it will eventually find itself in one of the leaves of this DAG. Then, it will stay on this leaf forever, and will be an irreducible chain there. If there's only 1 leaf then the stationary distribution there is the unique stationary distribution of the whole chain. If there's more then one leaf, then there's more then one stationary distribution, one for each leaf (and their convex combinations).

If we start the chain at some vertex, x, what is the probability it'll end at a specific leaf? Denote this probability $f(x)$. Note that this is a function of the state, not the communication class.

Theorem. *f is an harmonic function.*

Proof. straightforward. □

So we see that harmonic functions are not just a technical tool, they represent something significant.

Exercise. *show that any harmonic function is a linear combination of functions of this form.*