Markov Chains Mixing Times

Ori Gurel-Gurevich Ariel Yadin

December 27, 2007

1 convergence to stationary distribution

Assume the chain is irreducible, so we have a stationary distribution, π. Does $\mu_0 P^t = \mu_t \to \pi$ for any μ_0?

What exactly do we mean by $\to$ here? There are several notions of convergence, the simplest is pointwise, i.e. $\mu_t(x) \to \pi(x)$ for any $x \in \Omega$. In other words, convergence as a vector in $\mathbb{R}^\Omega$.

However, sometimes we want to be more specific about how *close* are we to stationarity. So we want to define what is the distance between distributions. We can consider known norms on $\mathbb{R}^n$.

L_∞: $d_\infty(\mu,\nu) = \max_{x\in\Omega} |\mu(x) - \nu(x)|$

L_2: $d_2(\mu,\nu) = \sqrt{\sum_{x\in\Omega} (\mu(x) - \nu(x))^2}$

L_1: $d_1(\mu,\nu) = \sum_{x\in\Omega} |\mu(x) - \nu(x)|$

and more

They are all equivalent w.r.t. convergence, but may be different when considering mixing times.

Obvious inequalities:

$$d_\infty \leq d_1 \leq \sqrt{n} d_2 \leq n d_\infty$$

One may consider more "exotic" distances:

$$d(\mu,\nu) = \max_{x\in\Omega} |\log \mu(x) - \log \nu(x)|$$

What do they all mean? We will work mainly with the L_1 distance, which has a very natural probabilistic interpretation.

Suppose that we have μ and ν and we know $d_1(\mu, \nu)$. Given an event A, if we know $\mu(A)$ what does this tells us about $\nu(A)$?

Exercise. $d_1(\mu, \nu) = 2 \max_{A \subset \Omega} |\mu(A) - \nu(A)|$

L_1 is also called *total variation* norm.

Periodic example for non-convergence.

SRW on a bipartite graph.

These examples have a period - all the times for which $P^t(x,x) > 0$ are divisible by some $m > 1$.

Definition. *A Markov chain is* aperiodic *if* $\forall x \in \Omega \gcd\{t | P^t(x,x) > 0\} = 1$. *It is called periodic otherwise.*

Like irreducibility, this is a property of the underlying graph, not of the specific probabilities.

Exercise. *Show that a SRW on a non-bipartite graph is aperiodic.*

Lemma. *If P is irreducible and aperiodic, exists r such that for all $x, y \in \Omega$, $P^r(x,y) > 0$.*

Proof. First we prove that for all $x \in Omega$ we have $P^r(x,x) > 0$ for all large enough r. Since the gcd is 1, we have a finite number of loops with lengths with gcd $= 1$. If these lengths are n_i then there are integers a_i such that $\sum a_i n_i = 1$. Let $A = -\min\{a_i\}$ then any number greater then $A \prod n_i$ can be expressed as a linear combination of the n_i's with natural coefficients.

Now the same is true for also for paths from x to y. Since Ω is finite, the proof is concluded. □

Exercise. *Prove that if P is irreducible then* $\gcd\{t | P^t(x,x) > 0\}$ *is the same for all choices of $x \in \Omega$. Conclude that if* $\gcd\{t | \exists y \in \Omega P^t(y,y) > 0\} = 1$ *then the chain is aperiodic.*

Theorem. *If a chain is irreducible and aperiodic then* $\mu P^t \to \pi$*. More specifically, there exists* $\alpha < 1$ *such that* $|\mu P^t - \pi| \leq \alpha^t$*, for any distribution* μ *on* Ω*.*

Proof. Assume $P(x, y) > 0$ for all $x, y \in \Omega$. Since Ω is finite, there is $\beta > 0$ such that $P(x, y) > \beta\pi(y)$. This implies that for any distribution μ, one can decompose $\mu P = \beta\pi + \alpha\nu$, for $alpha = 1 - \beta$ and ν some probability measure depending on μ.

Fix μ_0 and let $mu_t = \mu_0 P^t$. Then $\mu_1 = \beta\pi + \alpha\nu_1$. Decomposing $\nu_1 P$ we get $\nu_1 P = \beta\pi + \alpha\nu_2$. Therefore,

$$\mu_2 = \beta\pi + \alpha(\beta\pi + \alpha\nu_2) = (1 - \alpha^2)\pi + \alpha^2\nu_2$$

By induction we see that

$$\mu_t = (1 - \alpha^t)\pi + \alpha^t\nu_t$$

Thus,

$$\|\pi - \mu_t\|_1 = \|\alpha^t(\pi - \nu_t)\| \leq 2\alpha^t$$

□

Do we really need irreducibility? Recalling our discussion of the communication graph and the stationary distributions on the irreducible components, it is easy to see that the theorem holds with π replaced by some stationary distribution, dependent on μ.

Exercise. *prove that.*