Stat205B: Probability Theory (Spring 2003) **Lecture: 5**

Asymptotic Behaviour of Markov Chains (continued)

Lecturer: Jim Pitman *Scribe: Vinod Prabhakaran* `vinodmp@eecs.berkeley.edu`

Theorem 5.1. If P is

- irreducible
- positive recurrent (i.e., $\mathbb{E}_i T_i < \infty$ for some or all states i. See [1, p. 307, (4.7)])
- aperiodic (i.e., g.c.d$\{n : P^n(i,i) > 0\}$=1)

then

$$\lim_{n\to\infty} P^n(i,j) = \pi_j \text{ for all } i, j$$

where π_j is the unique stationary distribution of P.

Proof. The proof will be done in the next lecture. □

Recall the following fact from last lecture:

Fact. If P is irreducible and recurrent, and i is some fixed state, we defined

$$\mu_i(j) := \mathbb{E}_i\left(\sum_{n=0}^{T_i} 1(X_n = j)\right)$$

This is the expected number of j's in an i-block. We showed that

$$\mu_i P = \mu_i$$

$$\text{i.e., } \sum_j \mu_i(j)P(j,k) = \mu_i(k)$$

Next we show that such invariant measures have some uniqueness:

Theorem 5.2. If μ and ν are two invariant measures for the above setup (i.e., $\mu > 0$, $\nu > 0$, $\mu P = \mu$, $\nu P = \nu$, and P irreducible and recurrent), then $\mu = c\nu$, for some $c \geq 0$.

Proof. This proof is based on Exercise 4.6 in [1, p. 306][1]. It uses a *duality* idea which roughly says "P acting on left versus right is like *reversing time*". (The argument will be given just in the positive recurrent case, but can be generalized to cover the null-recurrent case too.)

We are considering the equation $\mu = \mu P$, where μ is a row vector. As we will see presently, it easier to deal with equations of the form $h = Ph$, where h is a column vector. Since the row sums of P are one, $h \equiv$ *constant* solves $h = Ph$. In fact, for an irreducible and recurrent P, this is the only possibility as shown below:

[1] See [1, p. 305, (4.4)] for another proof.

Claim. For an irreducible and recurrent Markov chain, every non-negative harmonic function is a *constant*.

Proof of claim: (See also [1, p. 299, Exercise 3.9]) Consider (X_n), a Markov chain with transition probability matrix P. It is easy to check that $h(X_n)$ is a non-negative martingale. Non-negativity implies that this martingale converges (by Martingale convergence theorem [1, p. 236, (2.11)]). But if $\exists\, i, j$ such that $h(i) \neq h(j)$, then $h(X_n) = h(i)$ i.o. and $h(X_n) = h(j)$ i.o.. This is because the chain is irreducible and recurrent and therefore hits all the states infinitely often. But since $h(X_n)$ converges, this implies that $h(i) = h(j)$.

Now suppose μ is an invariant probability measure and consider the chain $(X_0, X_1, \cdots, X_n)$ under $\mathbb{P}_\mu$. If we define

$$\hat{P}(x, y) = \frac{\mu(y)P(y, x)}{\mu(x)}$$

then the reversal $(X_n, X_n - 1, \cdots, X_0)$ is a Markov chain with a homogeneous transition probability matrix $\hat{P}$. Note that $\hat{P}$ is a valid transition probabilty matrix since

$$\sum_y \hat{P}(x, y) = \frac{\sum_y \mu(y)P(y, x)}{\mu(x)} = 1.$$

Also $\hat{P}(x, y) \geq 0$. Note that $\hat{P}$ is well-defined since P irreducible $\Rightarrow \mu(x) > 0$ as shown below:
P irreducible $\Rightarrow \forall x, y\ \exists n :\ P^n(y, x) > 0$. But, $\mu = \mu P \Rightarrow \mu = \mu P^n \Rightarrow \mu(x) = \sum_y \mu(y)P^n(y, x) \geq \mu(y)P^n(y, x)$. So either $\mu \equiv 0$ or $\mu(x) > 0,\ \forall x$. But μ is assumed to be a probability measure. So it cannot be identically 0.

Continuing to suppose that μ is a probability measure we can easily verify that the reversed chain is irreducible and recurrent. If ν is anotherinvariant measure, let $h(y) = \nu(y)/\mu(y)$. Then $\nu = \nu P$ is the same as $h = \hat{P}h$. Since h has to be constant by earlier discussion, μ is unique up to constant multiples.

□

An example on finding invariant measure:

Example. Consider the following population model: X_n=number of individuals at time n, $X_n \in \{0, 1, \cdots\}$. Between time n and $n + 1$, each individual dies with probability p independent of others. There is also an immigration of Y_{n+1} individuals independent of $X_1, X_2, \cdots, X_n$ according to Poission(λ). i.e.,

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i + Y_{n+1} \tag{1}$$

where $Z_1,\ Z_2,\ \cdots$ are independent Bernoulli(p) and $\mathbb{P}(Y_{n+1} = k) = e^{-\lambda}\lambda^k/k!,\ k = 0, 1, \cdots$.

Notice that (X_n) is a Markov chain with homogeneous probability transition matrix

$$P(j, k) = \mathbb{P}(X_{n+1} = k | X_n = j)$$

and
$P(j, .)$ =Binomial(j, p)$*$Poisson(λ). There is no simpler formula for $P(j, k)$ than a convolution, summing over possible values of the binomial variable. So the system of equations $\mu P = \mu$ for a stationary vector μ is easily handled. To understand the long-run behaviour of this chain, first recall a key-fact:

$$\text{Poisson}(\lambda) * \text{ Poisson}(\mu) = \text{ Poisson}(\lambda + \mu).$$

Also, if in (1), we assume that $X_n \sim$ Poisson(λ), then $\sum_{i=1}^{X_n} Z_i \sim$ Poisson(λp), by the *thinning property* of the Poisson distribution. So if we try $X_0 \sim$ Poisson(ν), we get $X_1 \sim$ Poisson($\nu p + \lambda$). In order to have $X_0 \stackrel{d}{=} X_1$, we just need $\nu = \lambda/(1 - p)$. We learn that Poisson($\lambda/(1 - p)$) is a stationary distribution. The chain is obviously irreducible since $P(j, k) > 0\ \forall j, k$. Using the fact below we can conclude that we have found a unique stationary probability measure for this Markov chain.

Fact. If a Markov chain P is irreducible and it has a stationary probability measure then it is recurrent. Then by previous discussion the stationary measure is unique.

Proof. Recall y is recurrent if $\sum_n P^n(y,y) = \infty$. Recall also that

$$\sum_n P^n(x,y) = \mathbb{P}_y(T_y = \infty) \sum_n P^n(y,y).$$

Let $N_y := \sum_{n=0}^{\infty} 1_{(X_n=y)}$ be the number of hits on y. Then

$$\begin{aligned}
\mathbb{E}_\mu(N_y) &= \mathbb{E}_u\left(\sum_{n=0}^{\infty} 1_{(X_n=y)}\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_\mu(X_n = y) \\
&= \infty \quad \text{ if } \mu(y) > 0
\end{aligned}$$

But

$$\mathbb{E}_\mu(N_y) = \sum_x \mu(x)\mathbb{E}_x(N_y) \leq \mathbb{E}_y(N_y) = \sum_n P^n(y,y)$$

where the inequality follows from the fact that μ is a probability measure. □

If π is the unique invariant probabaility measure, then

$$\pi(x) = 1/\mathbb{E}_x(T_x)$$

This follows immediately from

$$\mathbb{E}_x(T_x) = \sum_y \mathbb{E}_x(\text{number of } y\text{'s before } T_x) = \sum_y \mu_x(y)$$

But $\mu_x(y)$ is the invariant measure with $\mu_x(x) = 1$. If π is the invariant probability measure

$$\mu_x(y) = \pi(y)/\pi(x) \;\Rightarrow\; \mathbb{E}_x T_x = \sum_y \pi(y)/\pi(x) = 1/\pi(x).$$

References

[1] R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.