

Stat 471: Lecture 20

Markov Chains III.

Moo K. Chung
mchung@stat.wisc.edu

October 24, 2003

1. Suppose we have state space S as the sample space and let $\pi = \{\pi_j : j \in S\}$ be a probability distribution defined on S , i.e. $\sum_{j \in S} \pi_j = 1$. Then π is a stationary distribution for the Markov chain with transition probability $\mathbf{P}$ if $\pi' = \pi \mathbf{P}$. Note that

$$\pi' = \pi' \mathbf{P} = \pi' \mathbf{P}^2 = \dots = \pi' \mathbf{P}^n.$$

This probability π is also called *invariant probability measure*. Read S.I. Resnick's *Adventures in Stochastic Processes* for theoretical details on Markov Chains and a stationary distribution. A couple of interesting properties on π_j . $\frac{1}{\pi_j}$ is the expected number of states a Markov chain should take to return to the original state j (the mean recurrence time). In our crazy rat example, the rat will return to position 2 in average 3 steps if it was at the position 2 initially.

2. If we let $X_0 \sim \pi$, i.e. $P(X_0 = j) = \pi_j$. From the law of total probability $P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i)$ if $\{B_i\}$ partition S . So $P(X_1 = j) = \sum_{i \in S} \pi_i P_{ij}$. Hence $X_1 \sim \pi' \mathbf{P} = \pi'$. Similarly $X_i \sim \pi$ for all i .
3. Actually you don't need to have $X_0 \sim \pi$ to have a stable chain. If $X_0 \sim \mu$ for any probability distribution, $\mu' \mathbf{P}^n \rightarrow \pi'$ as $n \rightarrow \infty$. For proof, see G. Winkler's *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods* Theorem 4.3.1.

```
>> mu' %generate from rand
    0.534161    0.465839
    0.986845    0.013155
    0.969502    0.030498
>> mu' * P
    0.73292    0.26708
    0.50658    0.49342
    0.51525    0.48475
>> mu' * P^10
    0.66654    0.33346
```

```
    0.66696    0.33304
>> mu' * P^100
    0.66667    0.33333
    0.66667    0.33333
    0.66667    0.33333
```

4. Ergodic Theorem. Suppose a Markov chain X_i with kernel $\mathbf{P}$ on a finite sample space S has invariant distribution π then for any distribution μ and function g ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n g(X_i) = \mathbb{E}_{\pi} g = \sum_{i \in S} g(i) \pi_i.$$

After a large number of run, say m , the above convergence can be written

$$\mathbb{E}_{\pi} g \doteq \frac{1}{n - m} \sum_{i=m+1}^n g(X_i).$$

The number of samples m that are discarded for the above estimation is called the *burn in*. This is the basis of MCMC. Given probability distribution π , we estimate $\mathbb{E}_{\pi} g$ by constructing a Markov chain X_i and computing the ergodic limit.

5. The difference between MCMC and a simple Monte-Carlo method is if X_i are iid. Generalized by Hastings (1970), the Metropolis-Hastings algorithm is the only known method of MCMC. All other MCMC-looking procedures are actually pseudo-MCMC. The number n can be determined by running several Markov chains in parallel, each with a different initial value. If the estimated $\mathbb{E}_{\pi} g$ has large variation, increase the number of chains.