Discrete Stochastic Processes

Lecture 14 (Revised)

Finite-State and Countable-State Markov Chains

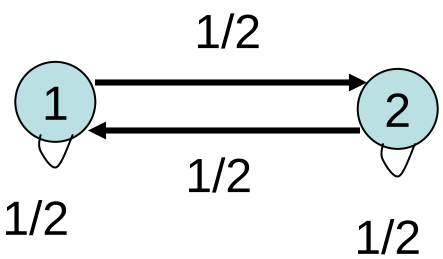
Convergence Rate Bounds for Finite Markov Chains

Branching Processes

Convergence Rate for Markov Chains

The following example illustrates the difference in rates at which two similar Markov chains converge to the same steady state probability distribution $\vec{\pi}$.

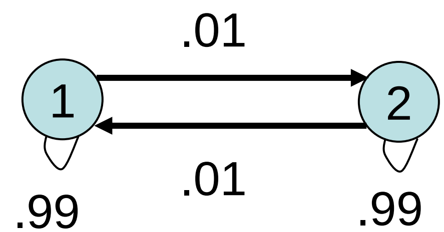
$$\vec{p}(t) = (p_1(t),\ p_2(t))$$



$$\vec{p}(0) = (1,\ 0)$$

$$\vec{p}(1) = (1/2,\ 1/2)$$

$$\vec{p}(2) = (1/2,\ 1/2)$$



$$\vec{p}(0) = (1,\ 0)$$

$$\vec{p}(1) = (.99,\ .01)$$

$$\vec{p}(1) = (.98,\ .02)$$

If we measure the distance between $\vec{p}(t)$ and $\vec{\pi}$ using the l_1 norm, i.e.,

$$\|\vec{p}(t) - \vec{\pi}\|_1 = \sum_{k=1}^{J} |p_k(t) - \pi_k|$$

then the following simple bound gives some feeling for the convergence rate:

Theorem Let $\vec{p}(t)$ represent the probability distribution at time t for any finite Markov chain with n states. Let

$$r_2 = 1 - \sum_{j=1}^{n} \min_i (p_{ij})$$

Then for all times t $\geq$ 1,

$$\| \vec{p}(t) - \vec{\pi} \|_1 \leq \| \vec{p}(0) - \vec{\pi} \|_1 (r_2)^t,$$

We begin with a few standard facts from linear algebra.

Definition 1: Let $\|\ \|$ be any norm on R^n. Let $\vec{x}$ denote a row vector in R^n and **A** be a matrix in R^{nxn}. Then the quantity $|||\ |||: R^{nxn} \to R$**,** given by

$$|||\mathbf{A}||| = \max_{\substack{\vec{x}\in R^n \\ \|\vec{x}\|=1}} \|\vec{x}A\| = \max_{\substack{\vec{x}\in R^n \\ \vec{x} \neq 0}} \frac{\|\vec{x}A\|}{\|\vec{x}\|}, \qquad (1)$$

is a norm on R^{nxn}. It is called an ***induced norm,*** since it arises from the norm $\|\ \|$ on R^n. It gives the ***maximum gain*** of **A,** since for any $\vec{x} \in R^n$, $\|\mathbf{A}\vec{x}\| \le |||\mathbf{A}|||\ \|\vec{x}\|$.

Lemma: Let $\|\ \|_1$ be the l_1 norm of any vector in R^n, i.e.,

$$\|\vec{x}\|_1 = \sum_{k=1}^{n} |x_k|.$$

Then the corresponding induced norm on R^{nxn} is given by

$$|||\mathbf{A}|||_1 = \max_i \sum_{j=1}^{n} |a_{ij}| \qquad (2)$$

Let **P** be the transition matrix for a Markov chain with n states, i.e., $\mathbf{P} \in R^{nxn}$**,**

$p_{ij} \geq 0$, $1 \leq i,j \leq n$**,** and $\sum_{j=1}^{n} p_{ij} = 1$, $1 \leq i \leq n$. Let $\vec{p}(0)$ be an initial distribution of probability for the chain. Then at any time $t \geq 1$, the probability distribution $\vec{p}(t)$ satisfies

$$\vec{p}(t) = \vec{p}(t-1)\,\mathbf{P}.$$

Let $\vec{\pi} = (\pi_1, \pi_2, \cdots, \pi_n)$ be a steady state probability distribution for the chain, i.e.,

$$\vec{\pi} = \vec{\pi}\,\mathbf{P}.$$

Theorem For any finite Markov chain and any $t \geq 1$,

$$\| \vec{p}(t) - \vec{\pi} \|_1 \leq [1 - \sum_{j=1}^{n} \min_i p_{ij}]^t \; \| \vec{p}(0) - \vec{\pi} \|_1, \; t \geq 1. \qquad (3)$$

Proof of Theorem

Since $\vec{p}(t) = \vec{\pi} + (\vec{p}(t) - \vec{\pi})$, and $\vec{\text{p}}(t)$ and $\vec{\pi}$ are both probability distributions, the difference $\vec{x}(\text{t}) = (\vec{p}(t) - \vec{\pi})$ between $\vec{p}(t)$ and $\vec{\pi}$ satisfies

$$\sum_{k=1}^{n} x_k(\text{t}) = \sum_{k=1}^{n} (p_k(t) - \pi_k) = 0.$$

And, since

$$\vec{x}(\text{t+1}) = (\vec{p}(t+1) - \vec{\pi}) = (\vec{p}(t) - \vec{\pi})\,\mathbf{P} = \vec{x}(t)\,\mathbf{P}, \qquad (4)$$

$\vec{x}(t)$ evolves in the subspace orthogonal to the column vector $\mathbf{1} = \begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}$.

Therefore, subtracting a constant times the column vector **1** from any column of **P** does not alter the evolution equation (4). Extending this to all the columns, we see that if [**1]** $\in R^{nxn}$ is the matrix consisting entirely on 1's, and **D** $\in R^{nxn}$ is diagonal, then

$$\vec{x}(\text{t+1}) = (\vec{p}(t+1) - \vec{\pi}) = \vec{x}(t)\,\mathbf{P} = \vec{x}(t)\,\mathbf{(P - [1]D)}, \ \forall t \geq 0.$$

In other words, subtracting a constant vector $d_{jj}\mathbf{1}$ from the jth column of **P** for each column does not alter the product $\vec{x}(t)\mathbf{P}$, because the entries of $\vec{x}(t)$ sum to 0.

Since the choice of the diagonal entries in the diagonal matrix **D** has no effect on the product $\vec{x}(t)$ **(P – [1]D),** and since

$$\|\vec{x}(t+1)\|_1 \le |||(P - [1]D)|||_1 \; \|\vec{x}(t)\|_1 = [\max_i \sum_{j=1}^{n} |\{P - [1]D\}_{ij}|] \|\vec{x}(t)\|_1 \quad \forall t \ge 1, \tag{5}$$

we can choose the diagonal elements of D in a way that attempts to make the bound above as tight as possible, i.e., we can attempt to minimize

$$|||(P - [1]D)|||_1 = [\max_i \sum_{j=1}^{n} |\{P - [1]D\}_{ij}|] \tag{6}$$

by a good choice for the diagonal elements of **D.**

A good (but not necessarily optimal) approach is to shrink the entries in each column of **(P-[1]D)** by choosing the entries d_{jj} for each j to reduce all the values in the j-th column of **P** so that the corresponding column of (**P – [1]D)** has the smallest element(s) of the j-th column shrunk to zero, with the remaining elements in that column remaining positive.

Put another way, for any diagonal $\mathbf{D} \in R^{nxn}$**,** all elements in the j-th column of (**P – [1]D)** are nonnegative, decreasing functions of d_{jj} for $0 \le \mathrm{d}_{jj} \le \min_{1 \le i \le n} (p_{ij})$, so choosing $\mathbf{D}^*$ as the diagonal matrix with $d^*_{jj} = \min_{1 \le i \le n}(p_{ij})$ reduces each element in the jth column as much as possible without having any element $(p_{ij} - d_{jj})$ become negative (and therefore contributing a positive term to $|(p_{ij} - d_{jj})|$ that would grow with further increases in d_{jj}). Though this approach may not always be optimal, it does succeed in reducing the values in each row sum

$$\sum_{j=1}^{n} |\{\mathrm{P - [1]D}\}_{\mathrm{ij}}| \;=\; \sum_{j=1}^{n} |\{\mathrm{P - [1]D^*}\}_{\mathrm{ij}}|$$

to

$$1 - \sum_{\mathrm{j=1}}^{\mathrm{n}} \min_{i}(p_{ij}),$$

This choice of $\mathbf{D}^*$ thereby brings the induced norm down to the value

$$|||(\mathrm{P} - [1]\mathrm{D}^*)|||_1 = [\max_i \sum_{j=1}^{n} |\{\mathrm{P} - [1]\mathrm{D}^*\}_{ij}|] = 1 - \sum_{j=1}^{n} \min_i (p_{ij}) .$$

Putting the whole story together, we see that

$$\| \vec{x}(t+1)\|_1 = \|(\vec{p}(t+1) - \vec{\pi})\| = \|(\vec{p}(t) - \vec{\pi})(P - [1]D^*)\| \leq$$

$$|||(P - [1]D^*)|||_1 \| (\vec{p}(t) - \vec{\pi}) \|_1 \leq [1 - \sum_{j=1}^{n} \min_i (p_{ij})] \| (\vec{p}(t) - \vec{\pi}) \|_1 = [1 - \sum_{j=1}^{n} \min_i (p_{ij})] \| \vec{x}(t)\|_1,$$

which implies that the distance between $\vec{p}(t)$ and the equilibrium distribution $\vec{\pi}$ decays with time at least as fast as the upper bound

$$\| \vec{x}(t)\|_1 \leq [1 - \sum_{j=1}^{n} \min_i (p_{ij})]^t \| \vec{x}(0)\|_1 ,$$

as claimed.

∎

Relation to Eigenvalues

Any finite Markov chain has an eigenvalue $\lambda_1 = 1$ with a corresponding left eigenvector $\vec{\pi}$. The difference in convergence rate arises from the other eigenvalues $\lambda_k,\ k \geq 2,$ which have magnitudes $|\lambda_k| \leq 1.$

If we order them so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_J|,$ we see that $|\lambda_2|$ determines the slowest decaying behavior.

<u>COROLLARY</u>

For any finite Markov chain,

$$|\lambda_2| \leq r_2 = 1 - \sum_{j=1}^{J} \min_i (p_{ij})$$

Branching Processes

X_n = number of elements in n^{th} generation.

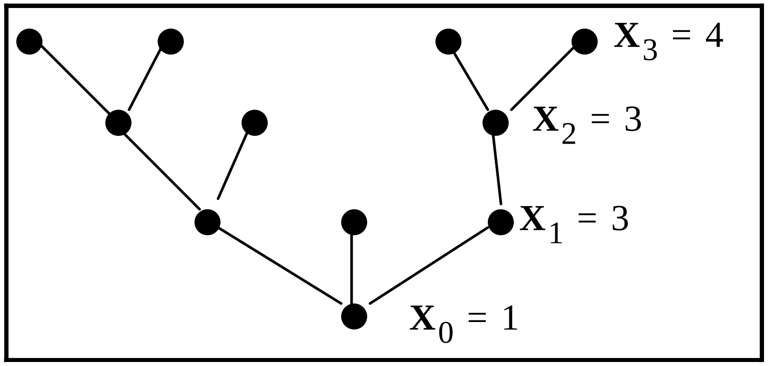
Each element k in generation n produces random number Y_{kn} of offspring. $\{Y_{kn}\}$ is set of IID rv's with pmf $p_j = P(Y_{kn} = j)$ (independent of n and k)

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{kn}$$

$Y_{12} = 2, Y_{22} = 0, Y_{32} = 2 \Rightarrow X_3 = 4$

$Y_{11} = 2, Y_{21} = 0, Y_{31} = 1 \Rightarrow X_2 = 3$

$Y_{10} = 3 \Rightarrow X_1 = 3$



$\{X_n;\ n \geq 0\}$ is a Markov chain with

$$P_{ij} = P\left(Y_{1n} + Y_{2n} + \ldots Y_{in} = j \,\middle|\, X_{n-1} = i\right) \quad \text{(independent of } n\text{)}$$

Expectations are easy

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{k,n}$$

$$E\{X_{n+1}\} = E\{X_n\}E\{Y\} = E\{X_{n-1}\}[E\{Y\}]^2 = X_0[E\{Y\}]^{n+1}$$

So, if $E\{Y\} > 1$, expected population explodes without limit and if $E\{Y\} < 1$, expected population goes to zero. That **does not guarantee** the population does not die out if $E\{Y\} > 1$.

We seek the more fundamental quantity,

$$\lim_{n\to\infty} F_{10}(n) = P\{\text{population descending from 1 founder ever dies out.}\}$$

Note that $P_{00} = 1$ (if population dies out, it ain't coming back. "Death is forever.")

$$F_{ij}(n) = P_{ij} + \sum_{k \neq j} P_{ik} F_{kj}(n-1);\ n > 1;\ F_{ij}(1) = P_{ij}$$

We want to find $F_{i0}(n)$, i.e., the probability that the population dies out by time *n*, given $X_0 = i$.
$F_{i0}(n)$ is the probability all offspring of all *i* "original founders" die out by time *n*.
Assume $p_0 = P\{Y_{kn} = 0\} > 0$

$$F_{k0}(n) = \left[F_{10}(n)\right]^k;\ P_{1k} = p_k = \text{ prob of } k \text{ offspring}$$

Difference Equation for $F_{10}(n)$:

$$F_{10}(n) = P_{10} + \sum_{k \neq 0} P_{1k} F_{k0}(n-1) = p_0 + \sum_{k \geq 1} p_k \left[F_{10}(n-1)\right]^k = \sum_{k=0}^{\infty} p_k \left[F_{10}(n-1)\right]^k = g(F_{10}(n-1)),$$

where g(•) is (***interestingly!***) the z transform of the probability distribution p_k:

$$g(z) = \sum_{k=0}^{\infty} p_k z^k$$

Let $g(z) = \sum_{k=0}^{\infty} p_k z^k$ (z transform of $\{p_k\}$)

$g(0) = p_0$

$g(1) = 1$

$g'(z) = \sum_{k=1}^{\infty} k p_k z^{k-1} > 0$

$g'(1) = \sum_{k=1}^{\infty} k p_k = E\{Y\}$

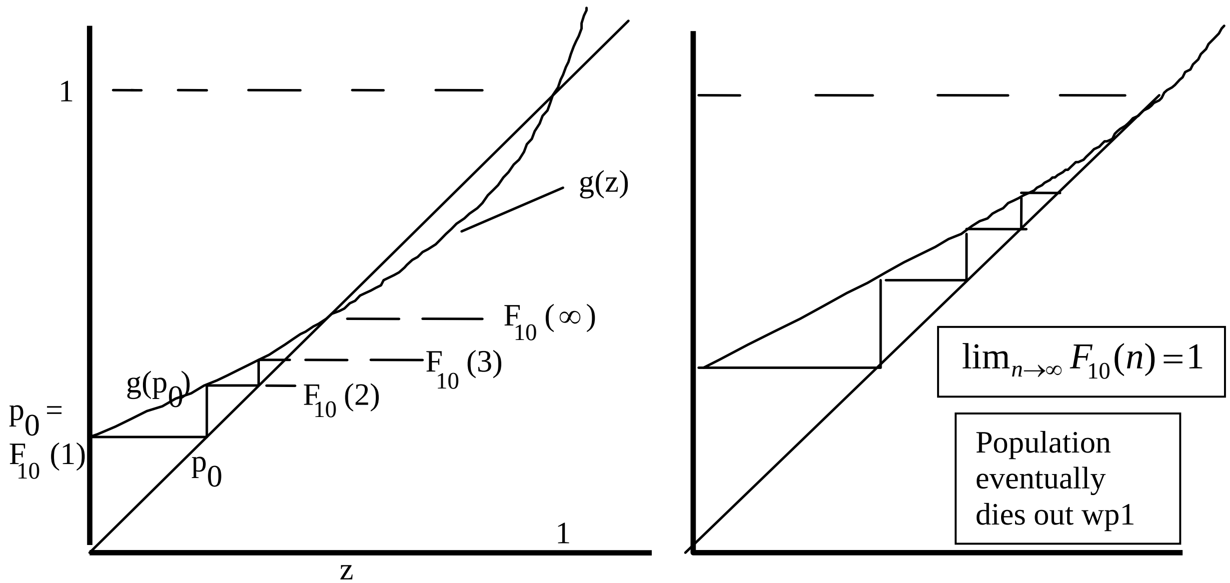
$g''(z) = \sum_{k=2}^{\infty} k(k-1) p_k z^{k-2} > 0$

Z real, $0 \le z \le 1$

Iterative Equation

$$F_{10}(n) = g\left(F_{10}(n-1)\right); \quad F_{10}(1) = p_0$$

Assume $p_0 = P\{Y_{kn} = 0\} > 0$



Lower crossing point is $\lim_{n\to\infty} F_{10}(n)$.

If $E[Y] \leq 1$, then $F_{10}(\infty) = 1$, and population dies out with probability 1 (but if $E[Y] = 1$, $E[X_n] = E[X_o]$ for all n).

Note that state 0 is positive recurrent and all other states are transient (state 0 is accessible from all of them). This suggests that for any value of E[Y], the process either dies out or else grows without bound.