Barebones Background for Markov Chains

by

Howard G. Tucker

Part I. Markov Dependence.

0. Introductory Remarks. This collection which I refer to as "Barebones Background for Markov Chains" is really a set of notes for lectures I gave during the spring quarters of 2001 and 2003 on Markov chains, leading up to Markov Chain Monte Carlo. The prerequisite needed for this is a knowledge of some basic probability theory and some basic analysis. This is really not a basic course in Markov chains. My main purpose here was to begin with a minimum definition of a Markov chain with a countable state space and to proceed in as direct a manner as possible from this definition to the strong law of large numbers for Markov chains, i.e., the ergodic theorem. This prepared the way for me to introduce in a mathematically rigorous way the Hastings-Metropolis algorithm for Markov Chain Monte Carlo, concluding with the optimality theorem by Billera and Diaconis. The only kind of Markov chain that I needed to deal with was an irreducible chain with a countable state space. Thus these notes should not be construed as a basic course in Markov chains but simply a course in the background mathematics needed for a rigorous justification of the Hastings-Metropolis algorithm. These are called notes for the simple reason that, other than statements of definitions and theorems and proofs of theorems, there is no expository prose included other than these introductory remarks.

1. Definition. An $n \times n$ matrix of real numbers, $\mathbf{P} = (p_{ij})$, where $2 \leq n \leq \infty$, is called a stochastic matrix if (i) $p_{ij} \geq 0$ for all $i \geq 1$, $j \geq 1$, (ii) $\sum_{j=1}^{n} p_{ij} = 1$ for all i, and (iii) for every j there exists an i such that $p_{ij} > 0$. (n is called the **order** of the matrix.)

2. Proposition. If $\mathbf{P} = (p_{ij})$ and if $\mathbf{Q} =(q_{ij})$ are stochastic matrices of the same order, then $\mathbf{PQ}$ is a stochastic matrix of the same order.

Proof: Let us denote $\mathbf{PQ} = (r_{ij})$, where

$$r_{ij} = \sum_{k\geq 1} p_{ik} q_{kj}.$$

Clearly, $r_{ij} \geq 0$. Also, for every $i \geq 1$,

$$\sum_{j\geq 1} r_{ij} = \sum_{j\geq 1} \sum_{k\geq 1} p_{ik} q_{kj} = \sum_{k\geq 1} p_{ik} \sum_{j\geq 1} q_{kj} = \sum_{k\geq 1} p_{ik} = 1.$$

Finally, for every $j \geq 1$, there exists a value of $k \geq 1$ such that $q_{kj} > 0$, and for this k, there exists an $i \geq 1$ such that $p_{ik} > 0$. Thus, for every $j \geq 1$ there exists an $i \geq 1$ such that $r_{ij} \geq p_{ik} q_{kj} > 0$. Q.E.D.

3. Corollary to Proposition 2. If $\mathbf{P} = (p_{ij})$ is a stochastic matrix, then $\mathbf{P}^n = (p_{ij}^n) = \prod_{j=1}^n \mathbf{P}$ is a stochastic matrix.

Proof: Apply Proposition 2.

4. Definition of Markov Chain. A Markov Chain is an infinite sequence, $X_0, X_1, X_2, \cdots$, of random variables, each taking values in a countable set $\mathbf{X}$ (which we shall sometimes refer to as the state space), and such that:

(i) $P([X_0 = j]) > 0$ for all $j \in \mathbf{X}$,

(ii) $P([X_{n+1} = j_{n+1}] | \bigcap_{k=0}^n [X_k = j_k]) = P([X_{n+1} = j_{n+1}] | [X_n = j_n]) = P([X_1 = j_{n+1}] | [X_0 = j_n])$ for $n = 0, 1, 2, \cdots$ and for all $j_i \in \mathbf{X}$ for which $P(\bigcap_{k=0}^n [X_k = j_k]) > 0$, and

(iii) for every $j \in \mathbf{X}$ there exists an $i \in \mathbf{X}$ such that $P([X_1 = j] | [X_0 = i]) > 0$.

5. Proposition. If $\{X_n\}$ is a Markov chain, then for every positive integer n and every state $i \in \mathbf{X}$,

$$P([X_n = i]) > 0.$$

Proof: We first prove the theorem for $n = 1$. Let $j \in \mathbf{X}$ be arbitrary. Then by (iii) in the definition of Markov chain, there exists an $i \in \mathbf{X}$ such that $P([X_1 = j] | [X_0 = i]) > 0$. Then, since also $P([X_0 = i]) > 0$, we have

$$\begin{aligned} P([X_1 = j]) &= \textstyle\sum_{k \in \mathbf{X}} P([X_1 = j] | [X_0 = k]) P([X_0 = k]) \\ &\geq P([X_1 = j] | [X_0 = i]) P([X_0 = i]) > 0. \end{aligned}$$

Now let n be any positive integer for which the theorem is true. Then $P([X_n = k]) > 0$ for all $k \in \mathbf{X}$, and, as before, there exists an $i \in \mathbf{X}$ such that $P([X_1 = j] | [X_0 = i]) > 0$. By requirement (ii) in the definition of a Markov chain, $P([X_1 = j] | [X_0 = i]) = P([X_{n+1} = j] | [X_n = i]) > 0$. Thus

$$\begin{aligned} P([X_{n+1} = j]) &= \textstyle\sum_{k \in \mathbf{X}} P([X_{n+1} = j] | [X_n = k]) P([X_n = k]) \\ &\geq P([X_{n+1} = j] | [X_n = i]) P([X_n = i]) > 0. \quad \text{Q.E.D.} \end{aligned}$$

6. Theorem for Existence of a Markov Chain. Given a countable state space $\mathbf{X}$, a probability measure π over $\mathbf{X}$ such that $\pi(i) > 0$ for all $i \in \mathbf{X}$

and a stochastic matrix $\mathbf{P} = (p_{ij})$ indexed by $\mathbf{X}$, there exist a probability space $(\Omega, \mathcal{A}, P)$ and $\mathbf{X}$-valued random variables $X_0, X_1, X_2, \cdots$ defined over Ω which form a Markov chain for which (i) $P([X_0 = i]) = \pi(i)$ for all $i \in \mathbf{X}$ and (ii) $P([X_{n+1} = j]|[X_n = i]) = p_{ij}$ for all $n \geq 0$ for all $j \in \mathbf{X}$ and all $i \in \mathbf{X}$.

Proof: Let us order the elements of X in the order type $1, 2, \cdots, n$ or $1, 2, \cdots, n, \cdots$, depending on whether $\mathbf{X}$ is finite or denumerable, so that for all intents and purposes the elements of $\mathbf{X}$ are consecutive positive integers. Then denote $\Omega = \times_{r=0}^{\infty}\mathbf{X}$. For every $j_0 \in \mathbf{X}$, define

$$A(0, j_0) = \{(x_0, x_1, x_2, \cdots) \in \Omega : x_0 = j_0\}.$$

For arbitrary j_0, j_1 in $\mathbf{X}$, define

$$\begin{aligned} A(0, j_0; 1, j_1) &= \{(x_0, x_1, x_2, \cdots) \in \Omega : x_0 = j_0, x_1 = j_1\} \\ &= \textstyle\bigcap_{k=0}^{1}\{(x_0, x_1, x_2, \cdots) \in \Omega : x_k = j_k\}, \end{aligned}$$

and, in general, for arbitrary $j_0, j_1, \cdots, j_n$ in $\mathbf{X}$, define

$$\begin{aligned} A(0, j_0; \ldots; n, j_n) &= \{(x_0, x_1, x_2, \cdots) \in \Omega : x_0 = j_0, \cdots, x_n = j_n\} \\ &= \textstyle\bigcap_{k=0}^{n}\{(x_0, x_1, x_2, \cdots) \in \Omega : x_k = j_k\} \end{aligned}$$

Over all subsets of the kind just defined, define a function P by $P(A(0, j_0)) = \pi(j_0)$, $P(A(0, j_0, 1, j_1)) = \pi(j_0)p_{j_0 j_1}$, and, in general, $P(A(0, j_0; \ldots; n, j_n)) = \pi(j_0)\prod_{k=1}^{n} p_{j_{k-1}j_k}$. Without ambiguity we may define, for integers $0 \leq k_0 < \cdots < k_r$ and r, arbitrary elements of $\mathbf{X}$, call them $j_{k_0}, \cdots, j_{k_r}$, the set $A(k_0, j_{k_0}; \cdots; k_r, j_{k_r})$ by taking the disjoint union of all sets of the form $A(0, j_0; \ldots; n, j_n)$, for any $n \geq k_r$, over all sets of the form

$$A(0, m_0; 1, m_1; 2, m_2; \cdots, n, m_n),$$

where $m_{k_0} = j_{k_0}, m_{k_1} = j_{k_1}, \cdots, m_{k_r} = j_{k_r}$, then $P(A(k_0, j_{k_0}, \cdots, k_r, j_{k_r}))$ is defined as the sum of $P(A(0, j_0, \ldots, n, j_n))$ of these sets. The manner in which we defined $P(A(k_0, j; \cdots; k_r, j_{k_r}))$ indicates that we can apply the Kolmogorov-Daniell theorem to extend uniquely the function P from the algebra of subsets generated by all sets of the form $A(k_0, j_{k_0}; \cdots; k_r, j_{k_r})$ to the sigma-algebra generated by this algebra. Then define for $k \geq 0$ the function X_k by $X_k(j_0, j_1, j_2, \cdots) = j_k$. We now verify that $\{X_0, X_1, \cdots\}$ is the Markov chain as claimed. Since $[X_0 = j_0] = A(0, j_0)$, it follows that $P([X_0 = j_0]) = \pi(j_0) > 0$. Also, since

$$A(0, j_0, \ldots, n, j_n) = \textstyle\bigcap_{k=0}^{n}[X_k = j_k]$$

for $n \geq 0$, it follows that

$$P\left(\bigcap_{k=0}^{n}[X_k = j_k]\right) = \pi(j_0)\prod_{k=1}^{n} p_{j_{k-1}j_k}.$$

Now suppose that $P(\bigcap_{k=0}^{n}[X_k = j_k]) > 0$. Then by the above

$$\begin{aligned} P([X_{n+1} = j_{n+1}] | \bigcap_{k=0}^{n}[X_k = j_k]) &= \frac{P(\bigcap_{k=0}^{n+1}[X_k = j_k])}{P(\bigcap_{k=0}^{n}[X_k = j_k])} \\ &= \frac{\pi(j_0)\prod_{k=1}^{n+1} p_{j_{k-1}j_k}}{\pi(j_0)\prod_{k=1}^{n} p_{j_{k-1}j_k}} \\ &= p_{j_n j_{n+1}}. \end{aligned}$$

Also

$$\begin{aligned} P([X_{n+1} = j_{n+1}] | [X_n = j_n]) &= \frac{P([X_{n+1} = j_{n+1}][X_n = j_n])}{P([X_n = j_n])} \\ &= \frac{\sum P([X_{n+1} = j_{n+1}][X_n = j_n]\bigcap_{k=0}^{n-1}[X_k = j_k])}{\sum P(\bigcap_{k=0}^{n-1}[X_k = j_k])} \\ &= p_{j_n j_{n+1}}, \end{aligned}$$

where the two sums are taken over all $n-$tuples of states, $j_0, j_1, \cdots, j_{n-1}$. From this we also obtain

$$P([X_{n+1} = j_{n+1}] | [X_n = j_n]) = P([X_1 = j_{n+1}] | [X_0 = j_n]),$$

and thus $\{X_0, X_1, \cdots\}$ satisfies the definition of a Markov chain with state space $\mathbf{X}$, stochastic matrix P and initial distribution $\pi(\cdot)$.

7. Corollary. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, if $n \geq 1$, and if $i_0, i_1, \cdots, i_n$ is any sequence of states in $\mathbf{X}$, then

$$P(\bigcap_{j=0}^{n}[X_j = i_j]) = \pi(i_0)\prod_{j=1}^{n} p_{i_{j-1}i_j}.$$

Proof: This follows immediately from the proof of theorem 6.

8. Notation. For i, j in $\mathbf{X}$, we denote $p_{ij}^1 = p_{ij}$; then

$$p_{ij}^2 = \sum_{k \in \mathbf{X}} p_{ik} p_{kj},$$

and in general

$$p_{ij}^{n+1} = \sum_{k \in \mathbf{X}} p_{ik}^n p_{kj}.$$

9. Proposition. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if $n \geq 1$, then for every positive n and k,

$$P\left([X_{n+k} = i_1]|[X_n = i_0)\right] = p^k_{i_0 i_1}.$$

Proof: By requirement (ii) in the definition of a Markov chain,

$$P([X_{n+k} = i_1]|[X_n = i_0]) = P([X_k = i_1]|[X_0 = i_0]).$$

But

$$P([X_k = i_1]|[X_0 = i_0]) = \sum P([X_k = i_1] \bigcap_{r=1}^{k-1} [X_r = u_r][X_0 = i_0])/P([X_0 = i_0]),$$

where the sum is taken over all $u_1, \cdots u_{k-1}$ in $\mathbf{X}$. Now apply Corollary 7, the definiton of a stochastic matrix and Notation 8 above to obtain the conclusion.

10. Theorem. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if $0 \leq k_0 < k_1 < \cdots < k_n$ are integers, then

$$P(\textstyle\bigcap_{j=0}^n [X_{k_j} = i_j]) = P([X_{k_0} = i_0]) \prod_{j=1}^n p^{k_j - k_{j-1}}_{i_{j-1} i_j}.$$

Proof: We prove this for $n = 1$. We first observe that

$$P([X_{k_0} = i_0][X_{k_1} = i_1])$$

is equal to

$$\sum P(\bigcap_{q=0}^{k_0-1} [X_q = u_q][X_{k_0} = i_0] \bigcap_{r=k_0+1}^{k_1-1} [X_r = v_r][X_{k_1} = i_1]),$$

where the sum is taken over all $u_q \in \mathbf{X}$, $0 \leq q \leq k_0 - 1$ and all $v_r \in \mathbf{X}$, $k_0 + 1 \leq r \leq k_1 - 1$. Applying the multiplication rule, Corollary 7, and the definition of Markov chain, and Notation 8, we obtain that the above sum is equal to the product of two sums of products, namely,

$$\sum \pi(u_0)(\prod_{m=1}^{k_0-1} p_{u_{m-1} u_m}) p_{u_{k_0-1} i_0},$$

where the sum is taken over all states $u_0, u_1, \cdots, u_{k_0-1}$ in $\mathbf{X}$, and

$$\sum p_{i_0 v_{k_0+1}} (\prod_{m=k_0+1}^{k_1-2} p_{v_m v_{m+1}}) p_{v_{k_1-1} i_1},$$

where the sum is taken over all states $v_{k_0+1}, \cdots, v_{k_1-1}$ in $\mathbf{X}$. The first sum is equal to $P([X_{k_0} = i_0])$, and the second sum is equal to $p_{i_0 i_1}^{k_1-k_0}$. The proof for general n involves the product of $n+1$ sums of products. Q.E.D.

11. Theorem. If $\{X_n\}$ is a Markov Chain determined by the state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, if $0 \leq k_0 < k_1 < \cdots < k_m < r_0 < r_1 < \cdots < r_n$, and if

$$P(\bigcap_{j=0}^{m} [X_{k_j} = u_j]) > 0,$$

where $\{u_0, \cdots, u_m\} \subset \mathbf{X}$, and if $\{v_0, \cdots, v_n\} \subset \mathbf{X}$, then

$$P(\bigcap_{j=0}^{n} [X_{r_j} = v_j] | \bigcap_{j=0}^{m} [X_{k_j} = u_j]) = P(\bigcap_{j=0}^{n} [X_{r_j} = v_j] | [X_{k_m} = u_m]).$$

Proof: We first write

$$P(\bigcap_{j=0}^{n} [X_{r_j} = v_j] | \bigcap_{j=0}^{m} [X_{k_j} = u_j]) = \frac{P(\bigcap_{j=0}^{n} [X_{r_j} = v_j] \bigcap_{j=0}^{m} [X_{k_j} = u_j])}{P(\bigcap_{j=0}^{m} [X_{k_j} = u_j])}.$$

Then, applying theorem 10 and doing appropriate cancellation, we obtain

$$P(\bigcap_{j=1}^{n} [X_{r_j} = v_j] | \bigcap_{j=0}^{m} [X_{k_j} = u_j]) = p_{u_m v_0}^{r_1-k_m} \prod_{t=1}^{n} p_{v_{t-1} v_t}^{r_t - r_{t-1}}.$$

Note that theorem 10 also yields

$$P(\bigcap_{j=1}^{n} [X_{k_j} = i_j] | [X_{k_0} = i_0]) = \prod_{j=1}^{n} p_{i_{j-1} i_j}^{k_j - k_{j-1}},$$

which, when applied to the above, gives us

$$P(\bigcap_{j=1}^{n} [X_{r_j} = v_j] | \bigcap_{j=0}^{m} [X_{k_j} = u_j]) = P(\bigcap_{j=0}^{n} [X_{r_j} = v_j] | [X_{k_m} = u_m]).$$

12. Theorem. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, if $S \subset \times_{k=0}^{n} \mathbf{X}$ for $n \geq 1$, and if $P([X_{n+1} = i][(X_0, X_1, \cdots, X_n) \in S]) > 0$, then

$$P([X_{n+2} = j] | [X_{n+1} = i][(X_0, X_1, \cdots, X_n) \in S]) = P([X_{n+2} = j] | [X_{n+1} = i])$$

for all j in $\mathbf{X}$.

Proof: The left side of the above may be expressed as

$$\frac{P([X_{n+2}=j][X_{n+1}=i][(X_0,X_1,\cdots,X_n)\in S])}{P([X_{n+1}=i][(X_0,X_1,\cdots,X_n)\in S])}.$$

The numerator is equal to

$$\sum P([X_{n+2}=j][X_{n+1}=i]\bigcap_{k=0}^{n}[X_k=u_k]),$$

where the sum is taken over all $(n+1)$-tuples $(u_0,\cdots,u_n)\in S$. Applying Corollary 7, this becomes

$$\sum\left(\pi(u_0)\prod_{k=1}^{n}p_{u_{k-1}u_k}\right)p_{u_ni}p_{ij},$$

again where the sum is taken over all $(n+1)$-tuples $(u_0,\cdots,u_n)\in S$. We may factor out the term p_{ij}. What remains in the numerator by applying Corollary 7 again is $P([X_{n+1}=i][(X_0,X_1,\cdots,X_n)\in S])$, which cancels out with the denominator, yielding the theorem.

13. Theorem. (Chapman-Kolmogorov equations) If $\{X_n\}$ is a Markov Chain, then

$$P([X_{m+n}=j]|[X_0=i])=\sum_{k\in\mathbf{X}}P([X_{m+n}=j]|[X_m=k])P([X_m=k]|[X_0=i]).$$

Proof: By theorem 11,

$$\begin{aligned}P([X_{m+n}=j]|[X_0=i]) &= \frac{\sum_{u\in\mathbf{X}}P([X_{m+n}=j][X_m=u][X_0=i])}{P([X_0=i])}\\ &= \frac{\sum_{u\in\mathbf{X}}P([X_{m+n}=j]|[X_m=u])P([X_m=u]|[X_0=i])P([X_0=i])}{P([X_0=i])},\end{aligned}$$

which yields the conclusion.

Exercises for Part I

1 Prove: If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i):i\in\mathbf{X}\}$ and stochastic matrix $\mathbf{P}=(p_{ij})$, and if $Y_n=X_{m+n}$ for some m and $n=0,1,2,\cdots$, then $\{Y_n\}$ is a Markov chain.

2. An equivalent definition of Markov chain appears to be: "The future and the past are conditionally independent, given the present." So check a

particular case of this. Prove that if $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then

$$P([X_{n-2} = i][X_{n-1} = j][X_{n+1} = k][X_{n+2} = l]|[X_n = r])$$

is equal to

$$P([X_{n-2} = i][X_{n-1} = j]|[X_n = r])P([X_{n+1} = k][X_{n+2} = l]|[X_n = r])$$

for $n \geq 2$ and for arbitrary elements i, j, k, l, r in $\mathbf{X}$.

3. Sometimes that part of the definition of Markov chain that states

$$P([X_{n+1} = j_{n+1}]|\bigcap_{k=0}^{n}[X_k = j_k]) = P([X_{n+1} = j_{n+1}]|[X_n = j_n])$$

for all states $j_0, j_1, \cdots, j_{n+1}$ in $\mathbf{X}$ which satisfy $P(\bigcap_{k=0}^{n}[X_k = j_k]) > 0$ is replaced by the requirement that $P([X_{n+1} = j_{n+1}]|\bigcap_{k=0}^{n}[X_k = j_k])$ does not depend on $j_0, j_1, \cdots, j_{n-1}$ for all states $j_0, j_1, \cdots, j_{n+1}$ in $\mathbf{X}$ which satisfy $P(\bigcap_{k=0}^{n}[X_k = j_k]) > 0$. Prove that these two statements are equivalent.

4. Consider the set of all 2×2 tables of nonnegative integers for which the numbers in the first row adds up to 6, the numbers in the first column adds up to 5, and the sum of all four numbers is 16. An example of one such table is

$$s_2 = \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix}$$

(i) List all six tables that satisfy the requirements.

(ii) These six tables constitute a state space for a Markov chain which we are about to construct. Let s_i denote the table with the number i in first row and the first column. A table like this one arises from time to time in statistical practice with a probability measure defined over it by

$$\pi(s_i) = \frac{\binom{5}{i}\binom{11}{6-i}}{\binom{16}{6}} \text{ for } 0 \leq i \leq 5.$$

The mechanism for generating a Markov chain here is to select a first state at random according to the distribution $\pi(\cdot)$. Whatever state is obtained, then the next state is selected according to the following procedure. First, select one of the following tables,

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

with probability $\frac{1}{2}, \frac{1}{2}$. Whichever table is obtained, add it termwise to the original table for your next state **unless** adding it termwise puts a negative number in one of the entries, in which case that next state is the same as the present state. So display the 6×6 stochastic matrix for this Markov chain. (For example, $p_{s_0,s_0} = \frac{1}{2}$, $p_{s_3 s_2} = \frac{1}{2}$, $p_{s_2 s_4} = 0$, etc.)

(iii) Compute the density of X_0 from the formula given above, then compute the density of X_1, and finally compute the joint density of X_0, X_1.

(iv) If P denotes the stochastic matrix from part (ii), display these matrices: P^{10}, P^{100} and P^{1000}.

Part II. Aperiodicity, Irreducibility and the First Limit Theorem.

1. Definition. If a and b are positive integers, then $a|b$ means there exists a positive integer n such that $b = na$. We shall say in this case that "a divides b".

2. Proposition. If a and b are positive integers, and if $a|b$, then $a \leq b$.

3. Definition. If $a|b$ and if $a|c$, then a is called a common divisor of b and c.

4. Definition. If a is a common divisor of b and c, then a is called a greatest common divisor of b and c if for every common divisor d of b and of c, then $a \geq d$. In this case we write $a = (b, c)$. Similarly, a greatest common divisor of more than two positive integers, $a_1, \cdots, a_r$, is a number d that divides each of $a_1, \cdots, a_r$ and is not smaller than any such common divisor; we denote this greatest common divisor by $(a_1, \cdots, a_r)$.

5. Remark. If a and b are positive integers, then (a, b) always exists and is unique. The same holds for $a_1, \cdots, a_r$. This follows from Proposition 2.

6. Theorem (Division Algorithm). If $0 < a < b$ are integers, then there exist unique positive integers c and d such that $0 \leq d < a$ and $b = ca + d$.

Proof: Since $0 < a < b$, and since $\{ka, k = 1, 2, \cdots\}$ is an unbounded set, we let c denote the largest value of k for which $ka \leq b$. Then take $d = b - ca$.

7. Euclidean Algorithm. If $0 < a < b$ are integers, then (a, b) is obtained in the following way. First use the division algorithm to write $b = q_0 a + r_0$, where $q_0 \geq 1$ and $0 \leq r_0 < a$. If $r_0 = 0$, then it is clear that $a = (a, b)$. But if $r_0 > 0$, and since $r_0 < a$, then by the division algorithm we may write $a = q_1 r_0 + r_1$ where $q_1 \geq 1$ and $0 \leq r_1 < r_0$. If $r_1 = 0$, then $r_0|a$, so then by the previous equation $b = q_0 a + r_0$, it follows that $r_0|b$. Thus

$r_0 \leq (a,b)$. But note that by $b = q_0 a + r_0$, it follows that $(a,b)|r_0$. Therefore by Proposition 2 above, $(a,b) \leq r_0$, so the two inequalities imply $(a,b) = r_0$. But if $r_1 > 0$, then as before we may write $r_0 = q_1 r_1 + r_2$ where $0 \leq r_2 < r_1$. Again, if $r_2 = 0$, then in the very same way, one can show that $r_1|a$ and $r_1|b$ so that $r_1 \leq (a,b)$, while at the same time showing that $(a,b)|r_1$. Thus we have, as before, $(a,b) = r_1$.Proceed until the last nonzero remainder is obtained.

8. Theorem. If $0 < a < b$ are integers, then there exist integers u and v such that $(a,b) = ua + vb$.

Proof: In the Euclidean algorithm above, let us take the the case where $(a,b) = r_1$. Then since $r_1 = a - q_1 r_0$ and since $r_0 = b - q_0 a$, we have $r_1 = a - q_1(b - q_0 a) = (1 + q_0 q_1)a - q_1 b$, which proves the theorem in the case when $(a,b) = r_1$ in the application of the Euclidean algorithm. From here on, the general proof is just a matter of notation.

9. Theorem. If a and b are positive integers, if $d|a$ and if $d|b$, then $d|(a,b)$.

Proof: This follows directly from Theorem 8.

10. Theorem. If $0 < k_1 < k_2 < \cdots < k_r$ are integers, where $r \geq 2$, then

$$(k_1, k_2, \cdots, k_r) = ((k_1, k_2, \cdots, k_{r-1}), k_r),$$

and there exist integers $t_1, \cdots, t_r$ such that

$$(k_1, \cdots, k_r) = \textstyle\sum_{j=1}^{r} t_j k_j.$$

Proof: We shall prove this by induction on n for the integers $0 < b < a_0 < a_1 < \cdots < a_n$. We first prove it is true for $n = 1$ by proving $(b, a_0, a_1) = ((b, a_0), a_1)$ and then by applying theorem. Let $d = (b, a_0, a_1)$. Then $d|b$ and $d|a_0$ and $d|a_1$, so by theorem 8, $d|(b, a_0)$; but also $d|a_1$, so again by theorem 9, $d|((b, a_0), a_1)$, from which follows from Proposition 2 above that $(b, a_0, a_1) \leq ((b, a_0), a_1)$. On the other hand, if $d = ((b, a_0), a_1)$, then $d|(b, a_0)$

and $d|a_1$, which in turn implies $d|b$ and $d|a_0$ and $d|a_1$ so $d \leq (b, a_0, a_1)$. Thus the first conclusion of the theorem,

$$(b, a_0, a_1) = ((b, a_0), a_1),$$

is true. Applying theorem 8 twice we obtain $(b, a_0, a_1) = k(b, a_0) + ma_1 = krb + ksa_0 + ma_1$ for some integers k, m, r, s. Now let n be any positive integer for which the two conclusions of the theorem are true, i.e.,

$(b, a_0, a_1, \cdots, a_n) = ((b, a_0, a_1, \cdots, a_{n-1}), a_n)$ and

$$(b, a_0, a_1, \cdots, a_n) = jb + \sum_{k=0}^{n} j_k a_k$$

for some integers $j, j_1, \cdots, j_n$. Then by the same proof as in the case of $n = 1$ we have

$$(b, a_0, a_1, \cdots, a_n, a_{n+1}) = ((b, a_0, a_1, \cdots, a_n), a_{n+1}).$$

Thus, for some integers u, v,

$$\begin{aligned}(b, a_0, a_1, \cdots, a_n, a_{n+1}) &= u(b, a_0, a_1, \cdots, a_n) + va_{n+1} \\ &= ujb + \sum_{k=0}^{n} uj_k a_k + va_{n+1}\end{aligned}$$

which proves the theorem for $n + 1$. Q.E.D.

11. Theorem. If $a_1, \cdots, a_r$ are positive integers satisfying $(a_1, \cdots, a_r) = 1$, then there exists a positive integer N such that for every positive integer m, there exist integers $d_1, \cdots, d_r$ such that

$$N + m = \sum_{j=1}^{r} d_j a_j.$$

Proof: By theorem 10 there exist integers $c_1, \cdots, c_n$ such that

$$1 = c_1 a_1 + \cdots + c_n a_n.$$

Without loss of generality we may assume that $c_1 > 0$. Now let us define

$$N = a_1 a_2 |c_2| + a_1 a_3 |c_3| + \cdots a_1 a_n |c_n|.$$

Claim: It is sufficient to prove the theorem for $0 \leq m \leq a_1 - 1$. Indeed, suppose the theorem is true for all values of m that satisfy $0 \leq m \leq a_1 - 1$. Now let $m \geq a_1$. Hence by the division algorithm,

$$N + m = N + ca_1 + r,$$

where $c \geq 1$ and $0 \leq r \leq a_1 - 1$. By the hypothesis of the claim, there exist integers $d_1, \cdots, d_n$ such that $N + r = \sum_{i=1}^{n} d_i a_i$. Thus

$$N + m = (c + d_1)a_1 + \sum_{i=2}^{n} d_i a_i,$$

which proves the claim. So now we have to prove it only for $0 \leq m \leq a_1 - 1$. In this case

$$N + m = a_1 a_2 |c_2| + a_1 a_3 |c_3| + \cdots a_1 a_n |c_n| + m$$

As noted above, $1 = c_1a_1 + \cdots + c_na_n$, so

$$\begin{aligned} N + m &= a_1a_2|c_2| + a_1a_3|c_3| + \cdots a_1a_n|c_n| + m\textstyle\sum_{i=1}^n c_ia_i \\ &= mc_1a_1 + (a_1|c_2| + mc_2)a_2 + \cdots + (a_1|c_n| + mc_n)a_n, \end{aligned}$$

which proves the theorem.

12. Definition. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then state j is said to be accessible from state i if there exists a nonnegative integer n such that $p_{ij}^n > 0$. (Note: for all $i \in \mathbf{X}$, i is accessible from i since $p_{ii}^0 = 1$.)

13. Proposition. If state j is accessible from state i, and if state k is accessible from state j, then state k is accessible from state i.

Proof: By hypothesis, there exist positive integers m and n such that $p_{ij}^m > 0$ and $p_{jk}^n > 0$. Thus by the Chapman-Kolmogorov equations,

$$\begin{aligned} p_{ik}^{m+n} &= \textstyle\sum_{t\in\mathbf{X}} p_{it}^m p_{tk}^n \\ &\geq p_{ij}^m p_{jk}^n > 0, \end{aligned}$$

which proves the proposition.

14. Definition. If a state j is accessible from state i, and if state i is accessible from state j, then states i and j are said to communicate, and this is expressed by: state i communicates with state j..

15. Proposition. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then "communicates with" is an equivalence relation and therefore partitions $\mathbf{X}$ in a unique manner into disjoint equivalence classes.

Proof: We must prove that "communicates with" is reflexive, symmetric and transitive. Refelxivity follows because $p_{ii}^0 = 1$ for all $i \in \mathbf{X}$. Transitivity follows from Proposition 13, and symmetry follows from the definition of "communicates with".

16. Definition. A Markov chain is said to be irreducible if there is only one equivalence class, i.e., if every two elements in the state space communicate with each other.

17. Definition. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then for each $i \in \mathbf{X}$, the period of i, denoted by d_i, is defined by $d_i = \gcd\{n \geq 1 : p_{ii}^n > 0\}$.

18. Definition. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then the state $i \in \mathbf{X}$ is called aperiodic if $d_i = 1$.

19. Definition. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, then it is said to satisfy the uniform irreducibility condition if there exists a positive integer, N, such that, for all $n \geq N$ and all i, j in $\mathbf{X}$, $p_{ij}^n > 0$.

20. Theorem: If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if two distinct states i and j communicate with each other, then the periods of the two states are equal.

Proof: Let n and m be positive integers that satisfy: $p_{ij}^n > 0$ and $p_{ji}^m > 0$. Let $k > 0$ satisfy $p_{ii}^k > 0$. Then, by the Chapman-Kolmogorov equations, it follows that

$$\begin{aligned} p_{jj}^{m+n+k} &= \textstyle\sum_{u,v \in \mathbf{X}} p_{ju}^m p_{uv}^k p_{vj}^n \\ &\geq p_{ji}^m p_{ii}^k p_{ij}^n = c p_{ii}^k, \end{aligned}$$

where $c = p_{ji}^m p_{ij}^n > 0$ does not depend on k. Since $p_{jj}^{m+n} > 0$, it follows that $m + n = k_1 d(j)$ for some positive integer k_1 and $m + n + k = k_2 d(j)$. For our particular value of k such that $p_{ii}^k > 0$, we have

$$k = m + n + k - (m + n) = k_2 d(j) - k_1 d(j) = (k_2 - k_1) d(j).$$

Thus $d(j)$ divides all values of k for which $p_{ii}^k > 0$, which implies that $d(j)|d(i)$. By the arbitrariness of i and j, one can also prove that $d(i)|d(j)$. Thus by Proposition 2, $d(i) = d(j)$.

21. Corollary. In an irreducible Markov chain, if one state is aperiodic, then all states are aperiodic. (In which case we shall say that the chain is aperiodic.)

Proof: This follows immediately from Theorem 20 and the definition of irreducibility.

22. Theorem. If $\{X_n\}$ is a Markov Chain determined by initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$ and with a **finite** state space $\mathbf{X}$, then it satisfies the **uniform irreducibility condition** if and only if it is irreducible and aperiodic.

Proof: Assume that the Markov chain satisfies the uniform irreducibility condition. Then there exists a positive integer m such that $p_{ij}^m > 0$ for all

ordered pairs of states i, j. For any state j, we then have, by the Chapman-Kolmogorov equations,

$$p_{jj}^{m+1} = \sum_{k \in \mathbf{X}} p_{jk}^m p_{kj}.$$

By the definition of Markov chain, there exists a state k_0, such that $p_{k_0 j} > 0$. From this last equation, we observe that $p_{jj}^{m+1} = \sum_{k \in \mathbf{X}} p_{jk}^m p_{kj} \geq p_{jk_0}^m p_{k_0 j} > 0$. Since $(m, m+1) = 1$, it follows that the chain is aperiodic. Since $p_{ij}^m > 0$ for all ordered pairs of states i, j, it easily follows that the Markov chain is irreducible. Conversely, suppose that the chain is irreducible and is aperiodic. Let $i \in \mathbf{X}$, and let $D_i = \{n : p_{ii}^n > 0\}$. Since

$$p_{ii}^{m+n} = \sum_{k \in \mathbf{X}} p_{ik}^m p_{ki}^n \geq p_{ii}^m p_{ii}^n,$$

it follows that D_i is closed under addition. By the aperiodicity hypothesis, there exist a finite number of distinct elements, $a_1, \cdots, a_r$, in D_i such that $(a_1, \cdots, a_r) = 1$. By theorem 11, it follows that there exists a positive integer M_{ii} such that $p_{ii}^n > 0$ for all $n > M_{ii}$. Also, for any pair of states $j \neq i$, it follows by irreducibility that there exists a positive integer k such that $p_{ij}^k > 0$. If we define $M_{ij} = M_{ii} + k$, it follows that for every $n \geq M_{ij}$, $p_{ij}^n > 0$. Since by hypothesis $\mathbf{X}$ is finite, we may take $N = \max\{M_{ij} : i \in \mathbf{X}, j \in \mathbf{X}\} < \infty$, and from this it follows that for every $n \geq N$, $p_{ij}^n > 0$ for all $i \in \mathbf{X}, j \in \mathbf{X}$. Q.E.D.

23. First Big Limit Theorem. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if it satisfies the following three conditions:

(I) it has an aperiodic state,

(ii) it has a finite state space, and

(iii) it is irreducible,

$$\lim_{n \to \infty} \mathbf{P}^n \text{ exists and } = \Pi,$$

where all rows of Π are the same.

Proof: By theorem 22, $\mathbf{P}$ satisfies the uniform irreducibility condition which is: there exists a positive integer m such that $p_{ij}^n > 0$ for all $n \geq m$ and for all i, j in $\mathbf{X}$. We first prove the theorem in Case 1: $m = 1$. Thus $p_{ij} > 0$ for all pairs of states i, j in $\mathbf{X}$. Let $\epsilon = \min\{p_{ij} : \text{all } i, j \text{ in } \mathbf{X}\}$. Since $\mathbf{P}$ has a finite number of rows and columns, it follows that $0 < \epsilon \leq p_{ij}$ for all i, j in $\mathbf{X}$. Next denote

$$m_j(n) \quad = \min\{p_{ij}^n : i \in \mathbf{X}\}$$

and

$$M_j(n) \quad = \max\{p_{ij}^n : i \in \mathbf{X}\}$$

for all $j \in \mathbf{X}$ and for all $n \geq 1$. Applying the Chapman-Kolmogorov equations, we obtain, for $n \geq 2$,

$$p_{ij}^n = \sum_{k\in\mathbf{X}} p_{ik} p_{kj}^{n-1} \geq \sum_{k\in\mathbf{X}} p_{ik} m_j(n-1) = m_j(n-1)$$

and

$$p_{ij}^n = \sum_{k\in\mathbf{X}} p_{ik} p_{kj}^{n-1} \leq \sum_{k\in\mathbf{X}} p_{ik} M_j(n-1) = M_j(n-1)$$

for all $i, j \in \mathbf{X}$ and for all $n \geq 2$. Taking the $\min_{i\in\mathbf{X}}$ of both sides of the first inequality, we get

$$m_j(n) \geq m_j(n-1),$$

and, taking the $\max_{i\in\mathbf{X}}$ of the second inequality, we obtain

$$M_j(n) \leq M_j(n-1)$$

for all $n \geq 2$ and all $j \in \mathbf{X}$. Thus for each $j \in \mathbf{X}$, m_j and M_j satisfy

$$m_j = \lim_{n\to\infty} m_j(n) \leq \lim_{n\to\infty} M_j(n) = M_j.$$

Note that for every j and for all ordered pairs of states $i, j \in \mathbf{X}$,

$$1 = \sum_{j\in\mathbf{X}} p_{ij}^n \geq \sum_{j\in\mathbf{X}} m_j(n-1)$$

and

$$1 = \sum_{j\in\mathbf{X}} p_{ij}^n \leq \sum_{j\in\mathbf{X}} M_j(n-1).$$

Hence, if we can prove that $m_j = M_j =$(some)π_j for all $j \in \mathbf{X}$, then $\sum_{j\in\mathbf{X}} \pi_j = 1$, and all of the rows are the same. So it is sufficient to prove $m_j = M_j$ for all $j \in \mathbf{X}$. Let $i_0 \in \mathbf{X}$ be such that

$$m_j(n) = \min_{i\in\mathbf{X}} p_{ij}^n = p_{i_0 j}^n,$$

and let $i_1 \in \mathbf{X}$ be such that

$$M_j(n-1) = \max_{i\in\mathbf{X}} p_{ij}^{n-1} = p_{i_1 j}^{n-1}.$$

Then

$$\begin{aligned} m_j(n) \quad &= p_{i_0 j}^n = \sum_{k\in\mathbf{X}} p_{i_0 k} p_{kj}^{n-1} \\ &= \sum_{k\neq i_1} p_{i_0 k} p_{kj}^{n-1} + p_{i_0 i_1} p_{i_1 j}^{n-1} \\ &= \epsilon p_{i_1 j}^{n-1} + (p_{i_0 i_1} - \epsilon) p_{i_1 j}^{n-1} + \sum_{k\neq i_1} p_{i_0 k} p_{kj}^{n-1} \\ &\geq \epsilon M_j(n-1) + \{p_{i_0 i_1} - \epsilon + \sum_{k\neq i_1} p_{i_0 k}\} m_j(n-1) \end{aligned}$$

or

$$m_j(n) \geq \epsilon M_j(n-1) + (1-\epsilon)m_j(n-1).$$

Taking the limit of both sides as $n \to \infty$, we obtain $m_j \geq \epsilon M_j + (1-\epsilon)m_j$ or $m_j \geq M_j$. Since we established above that $m_j \leq M_j$, we may conclude that $m_j = M_j = \pi_j$ for all $j \in \mathbf{X}$. Hence $\sum_{j\in\mathbf{X}} \pi_j = 1$, and all rows of the limit matrix are identical. This proves the theorem for $m = 1$. We now prove the theorem for $m > 1$. Since $\mathbf{P}^m$ is a stochastic matrix with all positive entries, we may replace $\mathbf{P}$ in the proof of the case above where $m = 1$ by $\mathbf{P}^m$ and thereby conclude that

$$\mathbf{P}^{mn} \to (\text{some})\Pi \text{ as } n \to \infty,$$

where Π is a stochastic matrix of which all rows are identical. Then for $1 \leq k \leq m-1$,

$$\lim_{n\to\infty} \mathbf{P}^{mn+k} = \mathbf{P}^k\Pi.$$

Since all rows of Π are the same , and since $\mathbf{P}^k$ is a stochastic matrix, then $\mathbf{P}^k\Pi = \Pi$. So $\mathbf{P}^r \to \Pi$ as $r \to \infty$.

24. Corollary. If $\{X_n\}$ is a Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if it satisfies the following three conditions:

(I) it has an aperiodic state,

(ii) it has a finite state space, and

(iii) it is irreducible,

if Π is as in theorem 23 above, then

$$\Pi\mathbf{P} = \mathbf{P}\Pi = \Pi.$$

Proof: First observe that $\mathbf{P}^{n+1} = \mathbf{P}^n\mathbf{P} \to \mathbf{\Pi P}$ as $n \to \infty$, but also $\mathbf{P}^{n+1} \to \Pi$, so $\Pi\mathbf{P} = \Pi$. Since $\mathbf{P}$ is a stochastic matrix, $\mathbf{P}\Pi = \Pi$, which proves the corollary.

25. Corollary. Under the conditions of the theorem and corollary above, if the initial distribution of X_0 is the same as that of the common row in Π, then the random variables $\{X_n\}$ are identically distributed, and the sequence is stationary.

Proof: This follows from the conclusion in Corollary 24 that states $\mathbf{P}\Pi = \Pi$.

Part III. Recurrence

1. Notation. If $\{X_n\}$ is a Markov chain with state space $\mathbf{X}$, then for $i \in \mathbf{X}$, we shall denote $T_i(0) = 0$, and for $k \geq 1$, we shall denote

$$T_i(k) = \min\{m : m > T_i(k-1), X_m = i\},$$

provided that this set, $\{m : m > T_i(k-1), X_m = i\}$, is not empty; otherwise, $T_i(k) = \infty$.

2. Proposition. If $\{X_n\}$ is a Markov chain, then state j is accessible from state i if and only if

$$P([T_j(1) < \infty] | [X_0 = i]) > 0.$$

Proof: We first note that

$$[T_j(1) < \infty] = \bigcup_{n \geq 1} \left([X_n = j] \bigcap_{r=1}^{n-1} [X_r \neq j] \right).$$

First suppose that state j is accessible from state i. Then by the definition of accessible, there exists a positive integer n such that $p_{ij}^n > 0$. Let $n_0 = \min\{n : p_{ij}^n > 0\}$. We now prove the

Claim: $P([X_{n_0} = j] \bigcap_{r=1}^{n_0-1} [X_r \neq j] | [X_0 = i]) = p_{ij}^{n_0}$. *Indeed,* we know that

$$p_{ij}^{n_0} = \sum P([X_{n_0} = j] \bigcap_{r=1}^{n_0} [X_r = k_r] | [X_0 = i])$$

where the sum is taken over all $k_r \in \mathbf{X}$, $1 \leq r \leq n_0 - 1$. However, if, in any of the summands, there is a $t \in \{1, \cdots, n_0 - 1\}$ such that $k_t = j$, then, by our definition of n_0, that term becomes

$$\begin{aligned} 0 &\leq P([X_{n_0} = j] \bigcap_{r=1, r \neq t}^{n_0} [X_r = k_r][X_t = j] | [X_0 = i]) \\ &\leq P([X_t = j] | [X_0 = i]) = p_{ij}^t = 0, \end{aligned}$$

which proves the claim. Thus by the identity that we first noted above, state j is accessible from state i if and only if $P([T_j(1) < \infty] | [X_0 = i]) > 0$.

3. Definition. If $\{X_n\}$ is a Markov chain, state i is said to be **recurrent** if

$$P([T_i(1) < \infty] | [X_0 = i]) = 1.$$

4. Proposition. If $\{X_n\}$ is a Markov chain, then state i is **recurrent** if and only if

$$P(\bigcup_{n=1}^{\infty} [X_n = i] | [X_0 = i]) = 1.$$

Proof: In the proof of Proposition 2, we noted that

$$[T_i(1) < \infty] = \bigcup_{n\geq 1} \left([X_n = i] \bigcap_{r=1}^{n-1} [X_r \neq i]\right).$$

But it is also easily shown that

$$\bigcup_{n=1}^{\infty} [X_n = i] = \bigcup_{n\geq 1} \left([X_n = i] \bigcap_{r=1}^{n-1} [X_r \neq i]\right).$$

and hence

$$[T_i(1) < \infty] = \bigcup_{n=1}^{\infty} [X_n = i],$$

from which the proposition follows.

5. Notation. For every pair of elements i, j in the state space $\mathbf{X}$ of a Markov chain $\{X_n\}$, we shall denote

$$f_{ij}^n = P\left(\left(\bigcap_{r=1}^{n-1} [X_r \neq j]\right) \cap [X_n = j] | [X_0 = i]\right)$$

and

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n = P\left(\bigcup_{n=1}^{\infty} [X_n = j] | [X_0 = i]\right),$$

with $f_{ii}^0 = 0$ for all $i \in \mathbf{X}$. Here $f_{ij}^0 = 0$ for all pairs i, j of states where $i \neq j$; also, $p_{ii}^0 = 1$ for all $i \in \mathbf{X}$, and $p_{ij}^0 = 0$ for $i \neq j$.

6. Remark. If $\{X_n\}$ is a Markov chain, then

$$f_{ij}^n = P([T_j(1) = n] | [X_0 = i]).$$

7. Definition of Generating Functions. In the notation given above, we define the generating functions

$$F_{ij}(s) = \sum_{n=0}^{\infty} f_{ij}^n s^n$$

and

$$P_{ij}(s) = \sum_{n=0}^{\infty} p_{ij}^n s^n,$$

both of which converge for $|s| < 1$.

8. Proposition. For every state $i \in \mathbf{X}$, and for every $n \geq 1$,

$$p_{ii}^n = \sum_{k=\,0}^{n} f_{ii}^k p_{ii}^{n-k},$$

and

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)} \text{ for all } s \in (0, 1).$$

Proof: We first observe that

$$[X_n = i] = \bigcup_{k=1}^{n} [T_i(1) = k][X_n = i],$$

so that

$$p_{ii}^n = P([X_n = i]|[X_0 = i]) = P\left(\bigcup_{k=1}^{n} [T_i(1) = k][X_n = i]|[X_0 = i]\right),$$

and since the union of intersections on the right side is a disjoint union, then, letting

$$W = P\left([X_k = i] \bigcap_{j=0}^{k-1} [X_j \neq i]|[X_0 = i]\right)$$

the above is

$$\begin{aligned}
&= \sum_{k=1}^{n} P\left(\bigcap_{j=0}^{k-1} [X_j \neq i][X_k = i][X_n = i]|[X_0 = i]\right) \\
&= \sum_{k=1}^{n} P\left([X_n = i]|[X_k = i] \bigcap_{j=0}^{k-1} [X_j \neq i][X_0 = i]\right) W \\
&= \sum_{k=1}^{n} P\left([X_n = i]|[X_k = i]\right) P\left([X_k = i] \bigcap_{j=0}^{k-1} [X_j \neq i]|[X_0 = i]\right) \\
&= \sum_{k=1}^{n} f_{ii}^k p_{ii}^{n-k}.
\end{aligned}$$

Now, since we agreed that $p_{ii}^0 = 1$, we have

$$P_{ii}(s) = \sum_{n=0}^{\infty} p_{ii}^n s^n = 1 + \sum_{n=1}^{\infty} p_{ii}^n s^n,$$

and this, plus the fact that $f_{ii}^0 = 0$, implies

$$\begin{aligned}
P_{ii}(s) - 1 &= \sum_{n=1}^{\infty} \left(\sum_{k=0}^{n} f_{ii}^k p_{ii}^{n-k}\right) s^n \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} f_{ii}^k p_{ii}^{n-k} s^n \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} (f_{ii}^k s^k)(p_{ii}^{n-k} s^{n-k}) \\
&= \left(\sum_{n=0}^{\infty} f_{ii}^k s^k\right) \left(\sum_{n=0}^{\infty} p_{ii}^n s^n\right) \\
&= F_{ii}(s) P_{ii}(s)
\end{aligned}$$

which gives the conclusion.

9. Proposition. If i and j are distinct states, then

$$p_{ij}^n = \sum_{k=0}^{n} f_{ij}^k p_{jj}^{n-j}$$

and

$$P_{ij}(s) = F_{ij}(s) P_{jj}(s)$$

for all $s \in (-1, 1)$.

Proof: This proof is the same as the proof of theorem 8, except in this case, one must use the agreement that $f_{ij}^0 = p_{ij}^0 = 0$.

10. Proposition. The following three statements for a Markov chain are equivalent: for every state $i \in \mathbf{X}$,

(i) state i is recurrent,

(ii) $f_{ii} = 1$, and

(iii) $\sum_{n=0}^{\infty} p_{ii}^n = \infty$.

Proof: Since

$$[T_i(1) < \infty] = \bigcup_{n \geq 1} \left([X_n = i] \bigcap_{r=1}^{n-1} [X_r \neq i] \right),$$

it follows that

$$P([T_i(1) < \infty] | [X_0 = i]) = P(\bigcup_{n \geq 1} \left([X_n = i] \bigcap_{r=1}^{n-1} [X_r \neq i] \right) | [X_0 = i]) = f_{ii},$$

which proves that (i) holds if and only if (ii) is true. We now prove that (ii) is equivalent to (iii). We first prove

$$\textit{Claim:} \ \lim_{s \uparrow 1} \sum_{n=1}^{\infty} f_{ii}^n s^n = \sum_{n=1}^{\infty} f_{ii}^n \text{ and } \lim_{s \uparrow 1} \sum_{n=1}^{\infty} p_{ii}^n s^n = \sum_{n=1}^{\infty} p_{ii}^n.$$

Let $0 < s_k \uparrow 1$ as $k \to \infty$. If one applies the Lebesgue monotone convergence theorem to measures $\{f_{ii}^n, n = 1, 2, \cdots\}$ and $\{p_{ii}^n, n = 1, 2, \cdots\}$ over the space $\{1, 2, \cdots\}$ with respect to the functions $f_k(n) = s_k^n$, $k = 1, 2, \cdots$, the claim is established. So, by the above claim and proposition 8, $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^n = 1$ if and only if

$$\lim_{s \uparrow 1} P_{ii}(s) = \lim_{s \uparrow 1} \frac{1}{1 - F_{ii}(s)} = \sum_{n=1}^{\infty} p_{ii}^n = \infty,$$

which concludes the proof.

11.Corollary. If $\{X_n\}$ is an irreducible Markov Chain determined by state space $\mathbf{X}$, initial distribution $\{\pi(i) : i \in \mathbf{X}\}$ and stochastic matrix $\mathbf{P} = (p_{ij})$, and if any one state is recurrent, then all states are recurrent. (In which case we may refer to the chain as a recurrent chain.)

Proof: Let i and j be any two states in $\mathbf{X}$, and suppose that state i is recurrent. Since the chain is irreducible, then these two states communicate

with each other, so there exist positive integers m and n such that $p_{ij}^m > 0$ and $p_{ji}^n > 0$. Since $\mathbf{P}^{m+k+n} = \mathbf{P}^m\mathbf{P}^k\mathbf{P}^n$, it follows that

$$\begin{aligned} p_{jj}^{m+k+n} &= \textstyle\sum_{u\in\mathbf{X},v\in\mathbf{X}} p_{ju}^n p_{uv}^k p_{vj}^m \\ &\geq p_{ji}^n p_{ii}^k p_{ij}^m \\ &= c p_{ii}^k, \end{aligned}$$

where $c = p_{ji}^n p_{ij}^m > 0$. Since state i is recurrent, then by theorem 10, $\sum_{k=1}^\infty p_{ii}^k = \infty$, which combined with the above display implies

$$\textstyle\sum_{k=1}^\infty p_{jj}^{m+k+n} = \infty.$$

Applying theorem 10 to this last equation implies that state j is recurrent.

12. Proposition. If $\{X_n\}$ is a recurrent Markov chain, then the random variables

$$\{T_i(k) - T_i(k-1), k \geq 1\}$$

are conditionally independent and identically distributed with respect to the probability measure $P(\cdot|[X_0 = i]$.

Proof: For ease of notation, we shall do the proof only for $T_i(1) - T_i(0) = T_i(1)$ and $T_i(2) - T_i(1)$; the proof given suggests the proof in the general case. Letting u, v be any positive integers, and applying theorem 7 in Part I, we have

$$\begin{aligned} &P([T_i(1) = u][T_i(2) - T_i(1) = v]|[X_0 = i]) \\ &= P\left(\textstyle\bigcap_{k=1}^{u-1}[X_k \neq i][X_u = i]\bigcap_{m=u+1}^{u+v-1}[X_m \neq i][X_{u+v} = i]|[X_0 = i]\right) \\ &= \textstyle\sum P\left(\bigcap_{k=1}^{u-1}[X_k = j_k][X_u = i]\bigcap_{m=u+1}^{u+v-1}[X_m = r_m][X_{u+v} = i]|[X_0 = i]\right) \\ &= \textstyle\sum p_{ij_1}p_{j_1j_2}\cdots p_{j_{u-1}i}p_{ir_{u+1}}p_{r_{u+1}r_{u+2}}\cdots p_{r_{u+v-1}i}, \end{aligned}$$

where the above sums are taken over all elements $j_i, \cdots, j_{u-1}, r_{u+1}, \cdots, r_{u+v-1}$ in $\mathbf{X}$ that are not equal to i. It is easily recognized that this last sum is the product of two sums, namely,

$$\begin{aligned} &= \left(\textstyle\sum p_{ij_1}p_{j_1j_2}\cdots p_{j_{u-1}i}\right)\left(\sum p_{ir_{u+1}}p_{r_{u+1}r_{u+2}}\cdots p_{r_{u+v-1}i}\right) \\ &= P([T_i(1) = u]|[X_0 = i])P([T_i(1) = v]|[X_0 = i]). \end{aligned}$$

Now sum both sides of this equation with respect to u, and we obtain

$$P([T_i(2) - T_i(1) = v]|[X_0 = i]) = P([T_i(1) = v]|[X_0 = i]),$$

thus proving the two random variables to be conditionally identically distributed. Substituting this last quantity into the earlier one, we get

$$\begin{aligned}&P([T_i(1)=u][T_i(2)-T_i(1)=v]|[X_0=i])\\&=P([T_i(1)=u]|[X_0=i])P([T_i(2)-T_i(1)=v]|[X_0=i]),\end{aligned}$$

which establishes independence. The equation that established conditional identically distributed also implies $P([T_i(2)-T_i(1)<\infty]|[X_0=i])=1$ and therefore $P([T_i(2)<\infty]|[X_0=i])=1$. Q.E.D.

13. Theorem. If $\{X_n, n\geq 0\}$ is an irreducible, recurrent Markov chain determined by $(\pi, \mathbf{X}, \mathbf{P})$, if $f:\mathbf{X}\to\mathbf{R}^1$ is any function, and if, for $i\in\mathbf{X}$, random variables $\{Z_n, n\geq 1\}$ are defined by

$$Z_n=\textstyle\sum_{k=T_i(n-1)+1}^{T_i(n)} f(X_k),$$

then $\{Z_n\}$ are independent and identically distributed with respect to $P(\cdot|[X_0=i])$.

Proof: This is proved in much the same way as Theorem 12. Again we prove it for Z_1 and Z_2; the general proof has only more notation. Let a and b be any real numbers. Then

$$\begin{aligned}&P([Z_1\leq a][Z_2\leq b]|[X_0=i])\\&=\sum_{r\geq 1,s\geq 1}\sum P([X_{r+s}=i]\bigcap_{q=r+1}^{r+s-1}[X_q=i_q][X_r=i]\bigcap_{t=1}^{r-1}[X_t=i_t]|[X_0=i]),\end{aligned}$$

where the inner sum is taken over all $(r+s-2)$-tuplets,

$$(i_1,\cdots,i_{r-1},i_{r+1},\cdots,i_{r+s-1}),$$

of elements in $\mathbf{X}\backslash\{i\}$ that satisfy

$$\begin{aligned}&f(i_1)+\cdots+f(i_{r-1})+f(i) &&\leq a \text{ and}\\&f(i_{r+1})+\cdots+f(i_{r+s-1})+f(i) &&\leq b.\end{aligned}$$

As in the proof of theorem 12, the above double sum becomes

$$\begin{aligned}&=\sum_{r\geq 1,s\geq 1}\sum p_{ii_1}p_{i_1i_2}\cdots p_{i_{r-1}i}p_{ii_{r+1}}\cdots p_{i_{r+s-1}i}\\&=\left(\sum p_{ii_1}\cdots p_{i_{r-1}i}\right)\left(\sum p_{ii_{r+1}}\cdots p_{i_{r+s-1}i}\right)\\&=P([Z_1\leq a]|[X_0=i])P([Z_1\leq b]|[X_0=i]).\end{aligned}$$

Now let $a \uparrow \infty$ on both sides, from which we get

$$P([Z_2 \leq b]|[X_0 = i]) = P([Z_1 \leq b]|[X_0 = i]),$$

which establishes independence and identical distributions with respect to $P(\cdot|[X_0 = i])$.

14. Corollary. If $\{X_n\}$ is an irreducible, recurrent Markov chain, then for every $i \in \mathbf{X}$,

$$P([X_n = i \text{ infinitely often}]|[X_0 = i]) = 1.$$

Proof: Denote $A = [X_n = i \text{ infinitely often}]$; then clearly

$$A = \bigcap_{k=1}^{\infty}[T_i(k) < \infty] = \bigcap_{k=1}^{\infty}[T_i(k) - T_i(k-1) < \infty].$$

By theorem 12, $\{T_i(k) - T_i(k-1), k \geq 1\}$ are independent and identically distributed with respect to $P(\cdot|[X_0 = i)$. By the hypothesis of recurrence, $P([T_i(k) - T_i(k-1) < \infty]|[X_0 = i]) = 1$ for $k = 1$ and hence for all $k \geq 1$, which implies $P(A|[X_0 = i]) = 1$. Q.E.D.

Exercises for Part III.

1. Refer to problem 2 in the exercises for Part II. Suppose that the outcome of 16 plays of the game you observed this table:

	B	B^c
A	4	2
A^c	1	9

In much more complicated cases than this, we feel that in as many as 4 trials out of 16 of having both A and B occur is too large relative to the other cell counts for the hypothesis of independence. So what we might wish to do is compute the conditional probability $P([X_{11} \geq 4]|[X_{1\cdot} = 6][X_{\cdot 1} = 5])$ to see if this "is possible". From your answer last time, this conditional probability is 0.0357. You might simulate this game until you obtain 10,000 games in which the first row sums are equal to 6 and the first column sums are equal to 4. As these occur in your program, find the relative frequency in which the event $[X_{11} \geq 4]$ occurs. To do so, you might first take $\alpha = \frac{6}{16}$ and $\beta = \frac{5}{16}$. Then you might obtain empirical evidence that this conditional probability does not depend on α and β by taking $\alpha = .5$ and $\beta = .5$.

2. One can also create a Markov chain that we shall use later on. The state space is

$$\mathbf{X} = \{s(i) : 0 \leq i \leq 5\},$$

where

$$s(i) = \begin{pmatrix} & B & B^c \\ A & i & 6-i \\ A^c & 5-i & 5+i \end{pmatrix}.$$

The mechanism for this Markov chain is as follows. Select the state $s(i)$ with probability

$$\pi(s(i)) = P([X_{11} = i]|[X_{1\cdot} = 6][X_{\cdot 1} = 5]), 0 \leq i \leq 5.$$

If $1 \leq i \leq 4$, then take $p_{s(i)s(i+1)} = \frac{1}{2}$, $p_{s(i)s(i-1)} = \frac{1}{2}$, but take $p_{s(0)s(0)} = \frac{1}{2}$, $p_{s(0)s(1)} = \frac{1}{2}$, $p_{s(5)s(4)} = \frac{1}{2}$ and $p_{s(5)s(5)} = \frac{1}{2}$, with all the remaining entries in the stochastic matrix taken as zeros.

(i) Compute $\pi(s(i))$ for all six states in the state space.

(ii) Display the stochastic matrix associated with this Markov chain.

(iii) Find $P([X_k = s(i)])$ for $0 \leq i \leq 5$ and for $k = 1$ and 2.

Part IV Invariant Measures

1. Definition. An infinite sequence $\{X_n : n \geq 0\}$ of random variables is said to be (strictly) stationary if for all positive integers n and k, the joint distribution of $\{X_0, X_1, \cdots, X_n\}$ is the same as that of $\{X_k, X_1, \cdots, X_{k+n}\}$.

2. Definition. If $\{X_n\}$ is a Markov chain with stochastic matrix $\mathbf{P} = (p_{ij})$ and state spce $\mathbf{X}$, and if

$$\pi = (\pi_1, \pi_2, \cdots)^t$$

is a probability distribution over $\mathbf{X}$, then π is called a stationary distribution for $\{X_n\}$ if $\pi_i > 0$ for all $i \in \mathbf{X}$ and $\pi = \pi\mathbf{P}$, i.e., $\pi_k = \sum_{i \in \mathbf{X}} \pi_i p_{ik}$ for all $k \in \mathbf{X}$. (π as well as other measures over $\mathbf{X}$ is a vertical vector whose coordinates are ordered in the same manner as the rows and columns of $\mathbf{X}$.)

3. Proposition. If $\{X_n\}$ is a Markov chain with stochastic matrix $\mathbf{P} = (p_{ij})$ and state space $\mathbf{X}$, and if the initial distribution $\pi = (\pi_1, \pi_2, \cdots)$

is a stationary distribution for $\{X_n\}$, then the Markov chain is strictly stationary.

Proof: We observe that

$$\begin{aligned} P([X_1 = i]) &= \textstyle\sum_{j\in\mathbf{X}} P([X_1 = i]|[X_0 = j])P([X_0 = j]) \\ &= \textstyle\sum_{j\in\mathbf{X}} \pi_j p_{ji} = \pi_i \\ &= P([X_0 = i]). \end{aligned}$$

Continuing inductively, one can prove $P([X_n = i]) = \pi_i$ for all $n \geq 1$ and all $i \in \mathbf{X}$. From this we obtain for all positive integers m and n,

$$\begin{aligned} P\left(\bigcap_{k=m}^{m+n} [X_k = i_k]\right) &= P([X_m = i_m]) \textstyle\prod_{t=m+1}^{m+n} P([X_t = i_t]| \bigcap_{q=m}^{t-1} [X_q = i_q]) \\ &= \pi_{i_m} \textstyle\prod_{t=m+1}^{n} p_{i_{t-1} i_t} \\ &= P([X_0 = i_m]) \textstyle\prod_{t=m+1}^{m+n} P([X_{t-m} = i_t]| \bigcap_{q=m}^{t-1} [X_{q-m} = i_q]) \\ &= P\left(\textstyle\bigcap_{k=m}^{m+n} [X_{k-m} = i_k]\right). \qquad \text{Q.E.D.} \end{aligned}$$

4. Definition. Let $\mathbf{P} = (p_{ij})$ be a stochastic matrix with state space $\mathbf{X}$, and let

$$\nu^t = (\nu_1, \nu_2, \cdots)$$

be a measure defined over the state space with $0 \leq \nu_i \leq \infty$ for all $i \in \mathbf{X}$. (In other words, ν is a nonnegative function that assigns to each $i \in \mathbf{X}$ a number ν_i.) Then ν is called an invariant measure over $\mathbf{X}$ if $\nu^t = \nu^t \mathbf{P}$, i.e., $\nu_k = \sum_{i\in\mathbf{X}} \nu_i p_{ik}$ for all $k \in \mathbf{X}$. (Like μ, the measure ν is regarded as a vertical vector indexed by the rows of $\mathbf{X}$.)

5. Proposition. If ν is an invariant measure for an irreducible, recurrent Markov chain, and if $\nu_i < \infty$ for at least one $i \in \mathbf{X}$, then $\nu_j < \infty$ for all $j \in \mathbf{X}$.

Proof: Let $j \in \mathbf{X}, j \neq i$. By hypothesis, the chain is irreducible, so there exists a positive integer m such that $p_{ji}^m > 0$. Since by hypothesis ν is an invariant measure for the chain, it follows that

$$\nu^t = \nu^t \mathbf{P} = \cdots = \nu^t \mathbf{P}^m,$$

from which there follows

$$\nu_i = \textstyle\sum_{k\in\mathbf{X}} \nu_k p_{ki}^m \geq \nu_j p_{ji}^m.$$

From the finiteness of ν_i and from $p_{ji}^m > 0$, it follows that $\nu_j < \infty$.

6. Theorem. If $\{X_n, n \geq 0\}$ is a recurrent and irreducible Markov chain, if $i \in \mathbf{X}$, and if ν is a measure over $\mathbf{X}$ defined for every $j \in \mathbf{X}$ by

$$\nu_j = E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]} | [X_0 = i])\right),$$

then

(1) $\nu_i = 1$

(2) ν is a finite invariant measure, and

(3) $\nu_j = \sum_{n=0}^{\infty} P([X_n = j][T_i(1) > n] | [X_0 = i]) < \infty$ for all $j \in \mathbf{X}$..

Proof: Note that in this theorem and its proof, the measure ν is a function of i. We first prove that (3) is true. By the monotone convergence theorem,

$$\begin{aligned}
E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]} | [X_0 = i])\right) &= \sum_{k=1}^{\infty} E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j][T_i(1)=k]} | [X_0 = i])\right) \\
&= \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} P\left([X_n = j][T_i(1) = k] | [X_0 = i])\right) \\
&= \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P([X_n = j][T_i(1) = k] | [X_0 = i]) \\
&= \sum_{n=0}^{\infty} P(\bigcup_{k=n+1}^{\infty} [X_n = j][T_i(1) = k] | [X_0 = i]) \\
&= \sum_{n=0}^{\infty} P([X_n = j][T_i(1) > n] | [X_0 = i]),
\end{aligned}$$

which proves (iii). We now wish to verify (1): $\nu_i = 1$. By our definition of ν_j given above,

$$\begin{aligned}
\nu_i &= E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=i]} | [X_0 = i])\right) \\
&= \frac{1}{P([X_0=i])} E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=i][X_0=i]}\right) \\
&= \frac{1}{P([X_0=i])} E\left(\sum_{m=1}^{\infty} \left(I_{\bigcap_{t=1}^{m-1} [X_t \neq i][X_m=i]} \sum_{n=0}^{m-1} I_{[X_n=i][X_0=i]}\right)\right) \\
&= \frac{1}{P([X_0=i])} E(I_{[X_0=i]}) = 1,
\end{aligned}$$

(note: this follows because each summand after $m = 1$ is the indicator of the empty set) which proves (1). We now wish to prove (2): $\nu_j = \sum_{k \in \mathbf{X}} \nu_k p_{kj}$ for all $j \in \mathbf{X}$. We first prove this to be true for $j \neq i$. By the definition of ν_j, and since $j \neq i$, we have

$$\nu_j = E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]} | [X_0 = i])\right) = E\left(\sum_{n=0}^{T_i(1)} I_{[X_n=j]} | [X_0 = i]\right).$$

Thus, for $j \neq i$, we may write

$$\begin{aligned}
\nu_j &= E(\textstyle\sum_{k=1}^{\infty} I_{[T_i(1)=k]} \sum_{n=0}^{T_i(1)} I_{[X_n=j]} | [X_0 = i]) \\
&= E(\textstyle\sum_{k=1}^{\infty} \sum_{n=0}^{k} I_{[T_i(1)=k][X_n=j]} | [X_0 = i]) \\
&= E(\textstyle\sum_{n=1}^{\infty} \sum_{k=n}^{\infty} I_{[T_i(1)=k][X_n=j]} | [X_0 = i]) \\
&= P([X_1 = j][T_i(1) \geq 1] | [X_0 = i]) + \textstyle\sum_{n=2}^{\infty} P([X_n = j][T_i(1) \geq n] | [X_0 = i]).
\end{aligned}$$

By hypothesis, this is a recurrent chain, so $P([T_i(1) \geq 1] | [X_0 = i]) = 1$, and thus the first term on the bottom line above becomes

$$P([X_1 = j][T_i(1) \geq 1] | [X_0 = i]) = P([X_1 = j] | [X_0 = i]) = p_{ij}.$$

So, since $j \neq i$, then $[X_n = j][T_i(1) \geq n] = [X_n = j][T_i(1) \geq n+1]$, and

$$\begin{aligned}
\nu_j &= p_{ij} + \textstyle\sum_{n=2}^{\infty} P([X_n = j][T_i(1) \geq n+1] | [X_0 = i]) \\
&= p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} \sum_{n=2}^{\infty} P([X_n = j][T_i(1) \geq n+1][X_{n-1} = k] | [X_0 = i]) \\
&= p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} \sum_{n=2}^{\infty} P([X_n = j] | [T_i(1) \geq n+1][X_{n-1} = k][X_0 = i]) Q(k, n)
\end{aligned}$$

where $Q(k, n) = P([T_i(1) \geq n+1][X_{n-1} = k] | [X_0 = i])$. But for $n \geq 2$ and $k \neq i$,

$$[T_i(1) \geq n+1][X_{n-1} = k][X_0 = i] = [X_1 \neq i] \cdots [X_{n-2} \neq i][X_{n-1} = k][X_0 = i].$$

Since $\nu_i = 1$, we have from the above that

$$\begin{aligned}
\nu_j &= \nu_i p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} \sum_{n=2}^{\infty} P([X_n = j] | \bigcap_{q=1}^{n-2} [X_q \neq i][X_{n-1} = k][X_0 = i]) Q(k, n) \\
&= \nu_i p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} \sum_{n=1}^{\infty} p_{kj} P([T_i(1) \geq n+1][X_n = k] | [X_0 = i]) \\
&= \nu_i p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} p_{kj} E\left(\sum_{m=1}^{\infty} I_{[T_i(1) > m][X_m = k]} | [X_0 = i]\right).
\end{aligned}$$

The inner sum vanishes for $m \geq T_i(1)$; also since $k \neq i$, we may start the inner sum at $m = 0$. Thus

$$\begin{aligned}
\nu_j &= \nu_i p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} p_{kj} E\left(\sum_{m=0}^{T_i(1)-1} I_{[X_m = k]} | [X_0 = i]\right) \\
&= \nu_i p_{ij} + \textstyle\sum_{k \in \mathbf{X}, k \neq i} p_{kj} \nu_k = \sum_{k \in \mathbf{X}} \nu_k p_{kj},
\end{aligned}$$

which establishes (2) for $j \neq i$. Last we wish to prove (2) for $j = i$. But first we establish a claim.

Claim. If $j \neq i$, if $n \geq 1$ and if $P([X_n = j][T_i(1) > n][X_0 = i]) > 0$, then

$$p_{ji} = P([X_{n+1} = i] | [X_n = j][T_i(1) > n][X_0 = i]).$$

Proof: By the technical theorems we obtained in Part I, and since $j \neq i$,

$$\begin{aligned}
&P([X_{n+1} = i]|[X_n = j][T_i(1) > n][X_0 = i]) \\
&= P([X_{n+1} = i]|[X_n = j](\textstyle\bigcap_{q=1}^{n}[X_q \neq i])[X_0 = i]) \\
&= P([X_{n+1} = i]|[X_n = j](\textstyle\bigcap_{q=1}^{n-1}[X_q \neq i])[X_0 = i]) \\
&= P([X_{n+1} = i]|[X_n = j]) = p_{ji},
\end{aligned}$$

which proves the claim. Continuing our proof of (2) for $j = i$, since in (1) we established that $\nu_i = 1$, it is therefore sufficient to prove $\sum_{k\in\mathbf{X}} \nu_k p_{ki} = 1$. Applying the above claim, the hypothesis of recurrence and (iii), we have, letting $W = P([X_{n+1} = i]|[X_n = k][T_i(1) > n][X_0 = i])$,

$$\begin{aligned}
&\textstyle\sum_{k\in\mathbf{X}} \nu_k p_{ki} \\
&= \textstyle\sum_{k\in\mathbf{X}} \sum_{n=0}^{\infty} P([X_n = k][T_i(1) > n]|[X_0 = i])p_{ki} \\
&= \textstyle\sum_{k\in\mathbf{X},k\neq i} \sum_{n=0}^{\infty} P([X_n = k][T_i(1) > n]|[X_0 = i])p_{ki} \\
&= \textstyle\sum_{k\in\mathbf{X},k\neq i} \sum_{n=0}^{\infty} P([X_n = k][T_i(1) > n]|[X_0 = i])P([X_{n+1} = i]|[X_n = k]) \\
&= \textstyle\sum_{k\in\mathbf{X},k\neq i} \sum_{n=0}^{\infty} P([X_n = k][T_i(1) > n]|[X_0 = i])W \\
&= \textstyle\sum_{n=0}^{\infty} \sum_{k\in\mathbf{X},k\neq i} P([X_n = k][T_i(1) > n][X_{n+1} = i]|[X_0 = i]) \\
&= \textstyle\sum_{n=0}^{\infty} P([X_{n+1} = i][T_i(1) = n+1]|[X_0 = i]) = 1.
\end{aligned}$$

Finally, the measure ν is finite by (1) and Proposition 5. Q.E.D.

7. Definition. A state i of a Markov chain $\{X_n : n \geq 0\}$ is said to be **positive recurrent** if $E(T_i(1)|[X_0 = i]) < \infty$.

8. Theorem. If $\{X_n\}$ is a recurrent, irreducible Markov chain, if state $i \in \mathbf{X}$ is a positive recurrent state, then the measure $\{\pi_j\}$, defined by

$$\pi_j = \frac{\nu_j}{E(T_i(1)|[X_0 = i])} = \frac{E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0 = i])\right)}{E(T_i(1)|[X_0 = i])}$$

is a stationary distribution.

Proof: By theorem 6*, we need only prove that* $\sum_{j\in\mathbf{X}} \pi_j = 1$. We shall make use of this well-known fact: if U is a nonnegative integer valued random variable with finite expectation, then

$$E(U) = \textstyle\sum_{u\geq 0} P([U \geq u]).$$

Making use of the Lebesgue monotone convergence theorem, we obtain

$$\begin{aligned}
\textstyle\sum_{j\in\mathbf{X}} \nu_j &= \textstyle\sum_{j\in\mathbf{X}} E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i]\right) \\
&= E\left(\textstyle\sum_{j\in\mathbf{X}} \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} I_{[X_n=j][T_i(1)=k]}|[X_0=i])\right) \\
&= E\left(\textstyle\sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \sum_{j\in\mathbf{X}} I_{[X_n=j][T_i(1)=k]}|[X_0=i])\right) \\
&= E\left(\textstyle\sum_{n=0}^{\infty} \sum_{k=n}^{\infty} I_{[T_i(1)=k]}|[X_0=i])\right) \\
&= \textstyle\sum_{n=0}^{\infty} P([T_i(1)\geq n]|[X_0=i]) = E(T_i(1)|[X_0=i]),
\end{aligned}$$

from which the conclusion follows.

9. Lemma. Let $\mathbf{P}$ be a stochastic matrix, and let ${}^r\mathbf{P}$ denote $\mathbf{P}$ with the column for state $r \in \mathbf{X}$ replaced by zeros. Let us denote $({}^r\mathbf{P})^n = ({}^r q_{ij})$. If $j \neq r$, then

$${}^r q_{ij} = P([X_n = j][T_r(1) > n]|[X_0 = i]).$$

Proof: First note that $\mathbf{P}^n = (p_{ij}^n)$, where

$$p_{ij}^n = \textstyle\sum_{k_1\in\mathbf{X}} \cdots \sum_{k_{n-1}\in\mathbf{X}} p_{ik_1} p_{k_1k_2} \cdots p_{k_{n-1}j}$$

Since the rth column of ${}^r\mathbf{P}$ contains all zeros, then

$$\begin{aligned}
{}^r q_{ij} &= \textstyle\sum_{k_1\neq r} \cdots \sum_{k_{n-1}\neq \mathbf{r}} p_{ik_1} p_{k_1k_2} \cdots p_{k_{n-1}j} \\
&= P([X_1 \neq r] \cdots [X_{n-1} \neq r][X_n = j]|[X_0 = i]) \\
&= P([X_n = j][T_r(1) > n]|[X_0 = i]). \qquad \text{Q.E.D.}
\end{aligned}$$

10. Theorem. Let $\{X_n\}$ be a recurrent and irreducible Markov chain with stochastic matrix $\mathbf{P}$. If μ and λ are invariant measures over $\mathbf{X}$, and if each is neither zero nor infinite at some state, then there exists $c > 0$ such that $0 < \mu_i = c\ \nu_i < \infty$ for all $i \in \mathbf{X}$.

Proof: Let $i \in \mathbf{X}$ be an arbitrary fixed state. By theorem 6, if ν is defined by

$$\nu_j = E\left(\textstyle\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i]\right)$$

for all $j \in \mathbf{X}$, then ν is an invariant measure and $\nu_i = 1$. Now let μ be a measure that satisfies the hypothesis.

Claim 1: $\mu_i > 0$ for all $i \in \mathbf{X}$. Indeed, by hypothesis there exists a state $j \in \mathbf{X}$ such that $\mu_j > 0$. Then

$$\mu^t = \mu^t\mathbf{P} = \cdots = \mu^t\mathbf{P}^m$$

for all positive integers m. By the hypothesis of irreducibility, there exists a positive integer m such that $p_{ji}^m > 0$. Thus

$$\mu_i = \sum_{k\in\mathbf{X}} \mu_k p_{ki}^m \geq \mu_j p_{ji}^m > 0,$$

which proves the claim. This claim and proposition 5 imply $0 < \mu_k < \infty$ for all $k \in \mathbf{X}$. Now fix the state i that defines the measure ν. By theorem 6, $\nu_i = 1$. Since the product of any positive constant with $\boldsymbol{\mu}$ is still an invariant measure, we may assume without loss of generality that $\mu_i = 1$.

Claim 2. $\mu_j \geq \nu_j$ for all $j \in \mathbf{X}$. Indeed, let ${}^i\mathbf{P} = ({}^ip_{rs})$ denote $\mathbf{P}$ with the column determmined by state i replaced by zeros. We first verify that

$$\mu_j = \delta_{ij} + \sum_{k\in\mathbf{X}} \mu_k \ {}^ip_{kj};$$

indeed, if $j = i$, then $\delta_{ii} = 1$ and ${}^ip_{ki} = 0$, (so in this case $\mu_i = 1$),and if $j \neq i$, then $\delta_{ij} = 0$ and ${}^ip_{kj} = p_{kj}$ (and in this case, since μ is an invariant measure, $\mu_j = \sum_{k\in\mathbf{X}} \mu_k p_{kj}$). Next recall by theorem 6 that

$$\nu_j = \sum_{n=0}^{\infty} P([X_n = j][T_i(1) > n]|[X_0 = i]).$$

Then by lemma 9, ${}^ip_{ij}^n = P([X_n = j][T_i(1) > n]|[X_0 = i])$. So $\nu_i = 1$, and, for $j \neq i$, $\nu_j = \sum_{n=0}^{\infty} {}^ip_{ij}^n$. So denote $\boldsymbol{\delta}_i$ as the vertical vector with all zeros except for a 1 in the ith place. Then we may write

$$\begin{aligned} \mu^t &= (\mu_1, \mu_2, \cdots) \\ &= (\delta_{i1} + \sum_{k\in\mathbf{X}} \mu_k \ {}^ip_{k1}, \delta_{i2} + \sum_{k\in\mathbf{X}} \mu_k \ {}^ip_{k2}, \cdots) \\ &= \boldsymbol{\delta}_i^t + \mu^t \ {}^i\mathbf{P}. \end{aligned}$$

From this last identity and the invariant measure hypothesis,

$$\begin{aligned} \boldsymbol{\mu}^t &= \boldsymbol{\delta}_i^t + \mu^t \ {}^i\mathbf{P} \\ &= \boldsymbol{\delta}_i^t + (\boldsymbol{\delta}_i^t + \mu^t \ {}^i\mathbf{P}) \ {}^i\mathbf{P} \\ &= \boldsymbol{\delta}_i^t + \boldsymbol{\delta}_i^t \ {}^i\mathbf{P} + \mu^t(\ {}^i\mathbf{P})^2, \end{aligned}$$

and continuing this substitution of $\boldsymbol{\mu}^t = \boldsymbol{\delta}_i^t + \mu^t \ {}^i\mathbf{P}$, we get at stage N ,

$$\boldsymbol{\mu}^t = \sum_{n=0}^{N} (\boldsymbol{\delta}_i^t)(\ {}^i\mathbf{P})^n + \boldsymbol{\mu}^t(\ {}^i\mathbf{P})^{n+1} \geq \sum_{n=0}^{N} (\boldsymbol{\delta}_i^t)(\ {}^i\mathbf{P})^n,$$

where this last inequality means in a coordinate wise sense. So let $N \to \infty$, and we get

$$\mu^t \geq \sum_{n=0}^{\infty} (\boldsymbol{\delta}_i^t)(\ {}^i\mathbf{P})^n \text{ coordinatewise.}$$

Since we established above that $\nu_i = 1$ and $\nu_j = \sum_{n=0}^{\infty} p_{ij}^n$ for $j \neq i$, we obtain from what has been established up to this point that $\mu_j \geq \nu_j$ for all $j \neq i$, and in addition $\mu_i = \nu_i = 1$. This verifies claim (2). But from claim (2), $\mu - \nu$ is nonnegative and is also an invariant measure. If, for any $j \in \mathbf{X}$, $\mu_j - \nu_j > 0$, then by claim (i), $\mu_k - \nu_k > 0$ for all states k. But this contradicts the fact that $\mu_i - \nu_i = 0$. Hence the two invariant measures are identical. Q.E.D.

11. Theorem. If $\{X_n\}$ is a Markov chain with state space $\mathbf{X}$, stochastic matrix $\mathbf{P} = (p_{ij})$ and initial distribution π and if the following three conditions are satisfied,

(i) the Markov chain is irreducible,

(ii) the Markov chain is aperiodic, and

(iii) $\mathbf{X}$ is finite,

the Markov chain is positive recurrent.

Proof: The first big limit theorem in Part II and its first corollary state that the hypotheses here are sufficient for there to exist an invariant probability measure, $\boldsymbol{\pi}$, for the chain such that $\mathbf{P}^n \to \Pi$ as $n \to \infty$, where each row of Π is $\boldsymbol{\pi}^t$. We first claim that the chain is recurrent. Indeed, by Proposition 10 in Part III, state $i \in \mathbf{X}$ is recurrent if and only if $f_{ii} = 1$ if and only if $\sum_{n=1}^{\infty} p_{ii}^n = \infty$. Suppose to the contrary that there exists a state $i \in \mathbf{X}$ which is not recurrent. Then $\sum_{n=1}^{\infty} p_{ii}^n < \infty$. This implies that $p_{ii}^n \to 0$ as $n \to \infty$. But since $\mathbf{P}^n \to \Pi$ as $n \to \infty$ as established above, this means that $p_{ii}^n \to \pi_i$. But since π is a probability measure, it follows by theorem 10 above that $\pi_k > 0$ for all $k \in \mathbf{X}$, supplying us with a contradiction. Thus every state in the chain is recurrent. We now wish to prove that the chain is **positive** recurrent, in other words, to prove that if $\{X_n : n \geq 0\}$ is a Markov chain with initial distribution π and with transition matrix $\mathbf{P} = (p_{ij})$, then $E(T_i(1)|[X_0 = i]) < \infty$ for any state i. So take any state i. Since it is recurrent and if the measure ν is defined by

$$\nu_j = E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0 = i])\right)$$

for every $j \in \mathbf{X}$, then ν is a finite invariant measure, $\nu_i = 1$, so that ν_j is finite and positive for all states j. Further, by the uniqueness theorem above, there exists a finite $c > 0$ such that $\nu_j = c\pi_j$ for all $j \in \mathbf{X}$. By hypothesis,

the state space is finite, so $\sum_{j\in\mathbf{X}} \nu_j < \infty$. Thus

$$\begin{aligned}\infty &> \textstyle\sum_{j\in\mathbf{X}} E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i]\right)\\ &= E\left(\textstyle\sum_{n=0}^{T_i(1)-1}\sum_{j\in\mathbf{X}} I_{[X_n=j]}|[X_0=i]\right)\\ &= E\left(\textstyle\sum_{n=0}^{T_i(1)-1} 1|[X_0=i]\right) = E(T_i(1)|[X_0=i]),\end{aligned}$$

which proves the theorem.

12. Theorem. Suppose a Markov chain $\{X_n\}$ is irreducible and positive recurrent. For all $j \in \mathbf{X}$ define

$$\nu_j = E\left(\textstyle\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i]\right).$$

Then π_j defined by

$$\pi_j = \frac{\nu_j}{E(T_i(1)|[X_0=i])} = \frac{E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i]\right)}{E(T_i(1)|[X_0=i])}$$

is the unique stationary distribution for the chain.

Proof: Theorems 8 and 10.

Exercises for Part IV

1. Refer to problem 2 in exercises for Part III. There we discovered that the distribution of X_0 was not the same as the distribution for X_1, and so the Markov chain was not stationary. In certain applications to be met later we shall need to change the stochastic matrix so that the original $\pi(\cdot)$ that we took is an invariant measure. Here is the description of one such revision. Suppose at time 0 the event $[X_0 = s(i)]$ occurs. So sample from the six states according to the distribution $p_{s(i)s(0)}, p_{s(i)s(1)}, \cdots, p_{s(i)s(5)}$. If the outcome is $s(i)$, then decide that $[X_1 = s(i)]$ has occurred. But if the outcome is $s(j)$ for some $j \neq i$, then a decision must be made whether to decide that $[X_1 = s(j)]$ has occurred or $[X_1 = s(i)]$ has occurred, and it is made as follows. Take a random number in $[0, 1)$, call it U, and define

$$\alpha(s(i), s(j)) = \min\left\{1, \frac{\pi(s_j)p_{s(j)s(i)}}{\pi(s_i)p_{s(i)s(j)}}\right\}.$$

If $[U < a(s(i), s(j))]$ occurs, then decide that $[X_1 = s(j)]$ has occurred; otherwise, decide that $[X_0 = s(i)]$ has occurred. Repeat all of the above for $X_2, X_3, \cdots$.

(i) Display the stochastic matrix for this altered Markov chain.

(ii) Compute $P([X_0 = s(i)])$ for $0 \leq i \leq 5$, and determine whether $P([X_0 = s(i)]) = P([X_1 = s(i)])$ for $0 \leq i \leq 5$.

(iii) Verify that this Markov chain is irreducible, is aperiodic, has a finite state space and is positive recurrent.

Part V. The Ergodic Theorem.

1. Lemma. If $\{X_n\}$ is a sequence of nonnegative random variables, and if $\sum_{n=0}^{\infty} X_n$ converges almost surely, then for every $\epsilon > 0$,

$$P([X_n \geq \epsilon \text{ infinitely often}]) = 0.$$

Proof: Suppose the conclusion is not true. Then there exists an $\epsilon > 0$ such that

$$P([X_n \geq \epsilon \text{ infinitely often}]) > 0.$$

Let $\omega \in [X_n \geq \epsilon$ infinitely often$]$; then $\sum_{n=0}^{\infty} X_n(\omega) = \infty$ over this set of positive probability, which contradicts the hypothesis. Q.E.D.

2. Lemma. If $\{x_n\}$ is a sequence of real numbers, if $x_n \to \infty$ as $n \to \infty$, and if

$$r_n = \max\{k : x_k \leq n\},$$

then $r_n \to \infty$ as $n \to \infty$.

Proof: First note that $\{r_n\}$ is nondecreasing since

$$\{k : x_k \leq n\} \subset \{k : x_k \leq n + 1\}.$$

Thus it is sufficient to prove that $\{r_n\}$ is unbounded. If, to the contrary, $\{r_n\}$ is bounded, then there is a minimum value of n, call it n_0, and a maximum value of k, call it k_0, such that $r_n = r_{n_0}$ for all $n > n_0$ and such that $x_k \leq x_{n_0}$ for all $k > k_0$. This contradicts the hypothesis that $x_n \to \infty$ as $n \to \infty$.

3. Theorem (Strong Law of Large Numbers). If $\{X_n\}$ is a sequence of independent, identically distributed random variables, and if $\overline{X}_n = (X_1 + \cdots + X_n)/n$, then $P([\overline{X}_n \text{ converges}]) = 1$ if and only if $E|X_1| < \infty$, in which case $P([\overline{X}_n \to E(X_1) \text{ as } n \to \infty]) = 1$.

Theorem 3, due to A. N. Kolmogorov, is a classical theorem in measure-theoretic probability.

4. Lemma. If $\{X_n\}$ is a Markov chain that is irreducible and positive recurrent with unique stationary distribution $\{\pi_i, i \in \mathbf{X}\}$, and if $f : \mathbf{X} \to \mathbf{R}^1$ is a function that satisfies $\sum_{i\in\mathbf{X}} |f(i)|\pi_i < \infty$, then

$$\sum_{i\in\mathbf{X}} f(i)\pi_i = \frac{E\left(\sum_{n=1}^{T_i(1)} f(X_n)|[X_0 = i]\right)}{E(T_i(1)|[X_0 = i])}$$

for all $i \in \mathbf{X}$.

Proof: By Theorem 12 in Part IV,

$$\begin{aligned}
\sum_{j\in\mathbf{X}} f(j)\pi_j &= \sum_{j\in\mathbf{X}} f(j)\frac{E\left(\sum_{n=0}^{T_i(1)-1} I_{[X_n=j]}|[X_0=i])\right)}{E(T_i(1)|[X_0=i])} \\
&= \frac{E\left(\sum_{j\in\mathbf{X}}\sum_{n=0}^{T_i(1)-1} f(j)I_{[X_n=j]}|[X_0=i])\right)}{E(T_i(1)|[X_0=i])} \\
&= \frac{E\left(\sum_{k=1}^{\infty} I_{[T_i(1)=k]}\sum_{j\in\mathbf{X}}\sum_{n=0}^{k-1} f(j)I_{[X_n=j]}|[X_0=i])\right)}{E(T_i(1)|[X_0=i])} \\
&= \frac{E\left(\sum_{k=1}^{\infty} I_{[T_i(1)=k]}\sum_{n=0}^{k-1} f(X_n)|[X_0=i])\right)}{E(T_i(1)|[X_0=i])}.
\end{aligned}$$

At this point we should note that in this last sum, $f(X_0) = f(i)$ since everything takes place over the event $[X_0 = i]$, and when we take $n = k$, since we are then over the event $[T_i(1) = k]$, then also $f(X_n) = f(i)$, so we can start the innermost sum at $n = 1$ and end it at k. Thus

$$\begin{aligned}
\sum_{j\in\mathbf{X}} f(j)\pi_j &= \frac{E\left(\sum_{k=1}^{\infty} I_{[T_i(1)=k]}\sum_{n=1}^{k} f(X_n)|[X_0=i])\right)}{E(T_i(1)|[X_0=i])} \\
&= \frac{E\left(\sum_{n=1}^{T_i(1)} f(X_n)|[X_0=i])\right)}{E(T_i(1)|[X_0=i])},
\end{aligned}$$

which proves the lemma.

5. Theorem (Ergodic Theorem) If $\{X_n\}$ is a Markov chain that is irreducible and positive recurrent with unique stationary distribution $\{\pi_i, i \in \mathbf{X}\}$, and if $f : \mathbf{X} \to \mathbf{R}^1$ is any bounded function over $\mathbf{X}$, then

$$P\left(\left[\lim \frac{1}{N}\sum_{n=0}^{N} f(X_n) \text{ exists and } = \sum_{i\in\mathbf{X}} f(i)\pi_i\right] |[X_0 = i]\right) = 1$$

for every $i \in \mathbf{X}$, and

$$P\left(\left[\lim \frac{1}{N} \sum_{n=0}^{N} f(X_n) \text{ exists and } = \sum_{i \in \mathbf{X}} f(i)\pi_i\right]\right) = 1.$$

Proof: Let i be any state in $\mathbf{X}$. We first prove the theorem when f is nonnegative and bounded. For every positive integer N, let us define

$$B_i(N) = \max\{k \geq 0 : T_i(k) \leq N\},$$

i.e., $B_i(N)$ is the number of return visits to state i among $X_1, \cdots, X_N$. Since we have established that $\{T_i(k) - T_i(k-1), k \geq 1\}$ are independent, identically distributed and positive with respect to $P(\cdot|[X_0 = i])$, it follows that $T_i(k) \to \infty$ with probability 1 as $k \to \infty$. Next define

$$\eta_k = \sum_{n=T_i(k)+1}^{T_i(k+1)} f(X_n).$$

By a Proposition 12 in Part III, $\{\eta_k, k \geq 1\}$ are independent and identically distributed with respect to $P(\cdot|[X_0 = i])$. We now prove that in addition, their common expectation, $E(\eta_1|[X_0 = i])$, is finite. Indeed, since the chain is positive recurrent by hypothesis, we have

$$E(\eta_1|[X_0 = i]) \leq (\text{upper bound of } f)E(T_i(1)|[X_0 = i]) < \infty.$$

So now we may apply the classical strong law of large numbers of Kolmogorov to obtain

$$\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \sum_{n=T_i(k)+1}^{T_i(k+1)} f(X_n) = E\left(\sum_{n=1}^{T_i(1)} f(X_n)|[X_0 = i]\right) \text{ a.e.}[P(\cdot|[X_0 = i])].$$

Notice that the random variable $B_i(N)$ defined above is the largest value of k such that $T_i(k) \leq N$, then $T_i(B_i(N)) \leq N$, and for any random variable strictly greater than $B_i(N)$, then T_i evaluated at that random variable is $> N$. This means that

$$T_i(B_i(N)) \leq N \leq T_i(B_i(N) + 1).$$

Thus for $f \geq 0$ and f bounded,

$$\sum_{n=0}^{T_i(B_i(N))} f(X_n) \leq \sum_{n=0}^{N} f(X_n) \leq \sum_{n=0}^{T_i(B_i(N)+1)} f(X_n).$$

Let us look at the sum at the left of this last string of inequalities; this term can be written as

$$\sum_{n=0}^{T_i(B_i(N))} f(X_n) = \sum_{n=0}^{T_i(1)} f(X_n) + \sum_{k=1}^{B_i(N-1)} \eta_k = J + \sum_{k=1}^{B_i(N-1)} \eta_k,$$

where

$$J = \sum_{n=0}^{T_i(1)} f(X_n).$$

Similarly, the sum on the right side of the last string of inequalities becomes

$$\sum_{n=0}^{T_i(B_i(N)+1)} f(X_n) = J + \sum_{k=1}^{B_i(N)} \eta_k.$$

Now J is a fixed random variable, and so $\frac{1}{N}J \to 0$ a.e.$[P(\cdot|[X_0 = i])]$. We should take note of the following:

$$\frac{\sum_{k=1}^{B_i(N)} \eta_k}{N} = \frac{\sum_{k=1}^{B_i(N)} \eta_k}{B_i(N)} \frac{B_i(N)}{N}.$$

We next prove a claim.

Claim: $\frac{B_i(N)}{N} \to \frac{1}{E(T_i(1)|[X_0=i])}$ as $N \to \infty$ a.e. $[P(\cdot|[X_0 = i])]$. To prove this claim, let us recall that $T_i(n) = \sum_{k=0}^{n} \alpha_k$, where $\alpha_1, \cdots, \alpha_n$ are independent and identically distributed with common distribution $P([T_i(1) = \cdot]|[X_0 = i])$. Also, by hypothesis, $E(\alpha_1|[X_0 = i]) < \infty$. By the strong law of large numbers given above,

$$P\left([\frac{T_i(n)}{n} \to E(T_i(1)|[X_0 = i])|[X_0 = i]\right) = 1.$$

We then combine this with the fact that

$$\frac{T_i(B_i(N))}{B_i(N)} \leq \frac{N}{B_i(N)} \leq \frac{T_i(B_i(N)+1)}{B_i(N)+1} \frac{B_i(N)+1}{B_i(N)}$$

and the fact that $B_i(N) \uparrow \infty$ with $P(\cdot|[X_0 = i])$-probability 1 as $n \to \infty$ to obtain

$$\frac{T_i(B_i(N))}{B_i(N)} \to E(T_i(1)|[X_0 = i]) \text{ a.e. } [P(\cdot|[X_0 = i])],$$

which proves the claim. Recalling that we are assuming for the time being that $f \geq 0$, by the above claim, and by lemma 4, we have

$$\begin{aligned} \frac{1}{N}\sum_{n=0}^{N} f(X_n) &\geq \frac{1}{N}\sum_{n=0}^{T_i(B_i(N))} f(X_n) \\ &= \frac{B_i(N)}{N}\frac{1}{B_i(N)}\sum_{n=0}^{T_i(B_i(N))} f(X_n) \\ &\to \frac{1}{E(T_i(1)|[X_0=i])} E\left(\sum_{n=1}^{T_i(1)} f(X_n)\right), \end{aligned}$$

or

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(X_n) \geq \sum_{j\in\mathbf{X}} f(j)\pi_j. \text{ a.e. } [P(\cdot|[X_0 = i])].$$

Similarly, we can also make use of the inequality

$$\sum_{n=0}^{N} f(X_n) \leq \sum_{n=0}^{T_i(B_i(N)+1)} f(X_n)$$

to obtain

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(X_n) \leq \sum_{j\in\mathbf{X}} f(j)\pi_j \text{ a.e. } [P(\cdot|[X_0 = i])].$$

These last two inequalities verify that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(X_n) = \sum_{j\in\mathbf{X}} f(j)\pi_j \text{ a.e. } [P(\cdot|[X_0 = i])].$$

Now let

$$A = [\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(X_n) \text{ exists and} = \sum_{j\in\mathbf{X}} f(j)\pi_j].$$

Then the absolute probability of A

$$P(A) = \sum_{i\in\mathbf{X}} P(A|[X_0 = i])P([X_0 = i]) = \sum_{i\in\mathbf{X}} 1P([X_0 = i]) = 1.$$

For general bounded function f, we make use of $f = f^+ - f^-$. The above proof holds for each of f^+ and $f^{-\cdot}$. Hence it holds for f. Q.E.D.

Problems for Part V

1. Verify the following equation,

$$\sum_{j\in\mathbf{X}} \sum_{n=0}^{k-1} f(j) I_{[X_n=j]} = \sum_{n=0}^{k-1} f(X_n),$$

which is used without proof in the proof of Lemma 4.

2. Verify the following sentence that appeared in the proof of theorem 5: Since we have established that $\{T_i(k) - T_i(k-1), k \geq 1\}$ are independent, identically distributed and positive with respect to $P(\cdot|[X_0 = i])$, it follows that $T_i(k) \to \infty$ with probability 1 as $k \to \infty$. as well as with $P(\cdot|[X_0 = i])$-probability 1.

3. Prove that since $B_i(N) \uparrow \infty$ with $P(\cdot|[X_0 = i])$-probability 1 as $n \to \infty$, then

$$\frac{T_i(B_i(N))}{B_i(N)} \to E(T_i(1)|[X_0 = i]) \text{ a.e. } [P(\cdot|[X_0 = i])] \text{ as } n \to \infty.$$

Part VI. Markov Chain Monte Carlo.

1. Proposition. If $\{X_n\}$ is a Markov chain with initial distribution π and state space $\mathbf{X}$ and with stochastic matrix $\mathbf{P}$, then the Markov chain is stationary if and only if $P([X_0 = j]) = P([X_1 = j])$ for all $j \in \mathbf{X}$.

Proof: Observe that for every positive integer n and every positive integer k and all $k+1$ tuplets of states, $i_0, \cdots, i_k$,

$$P\left(\bigcap_{r=0}^{k}[X_r = i_r]\right) = P\left([X_0 = i_0]\right) \prod_{r=1}^{k} p_{i_{r-1}i_r}$$

and

$$P\left(\bigcap_{r=n}^{n+k}[X_r = i_r]\right) = P\left([X_n = i_0]\right) \prod_{r=1}^{k} p_{i_{r-1}i_r},$$

from which the conclusion follows.

2. Definition. A Markov chain $\{X_n\}$ is said to be reversible if

$$P([X_n = i][X_{n+1} = j]) = P([X_n = j][X_{n+1} = i])$$

for all pairs of states i, j and every nonnegative integer n.

3. Proposition. If a Markov chain $\{X_n\}$ is reversible, then it is strictly stationary.

Proof: Sum both sides of the display equation in definition 2 for all $j \in \mathbf{X}$, and then apply proposition 1.

4. Theorem. If $\{X_n\}$ is an irreducible and aperiodic Markov chain with initial distribution π, state space $\mathbf{X}$ and stochastic matrix $K = (K(i,j))$ that satisfies $K(i,j) = 0$ if and only if $K(j,i) = 0$, and if the stochastic matrix $MK = (MK(i,j))$ indexed by $\mathbf{X}$ is defined by

$$MK(i,j) = \begin{cases} K(i,j)\alpha(i,j) \text{ if } i \neq j \\ 1 - \sum_{k \in \mathbf{X} \setminus \{i\}} K(i,k)\alpha(i,k) \text{ if } i = j, \end{cases}$$

where

$$\alpha(i,j) = \min\left\{1, \frac{\pi(j)K(j,i)}{\pi(i)K(i,j)}\right\},$$

then the Markov chain determined by π, MK and $\mathbf{X}$ is a reversible (and thus stationary), irreducible, aperiodic Markov chain.

Proof: It is sufficient to prove: for every pair of distinct states i and j,

$$\pi(i)MK(i,j) = \pi(j)MK(j,i).$$

Indeed, if $\pi(j)K(j,i) = \pi(i)K(i,j)$ for all pairs i,j, there is no change needed to obtain the conclusion. But if there exists states $i \neq j$ such that $\pi(j)K(j,i) < \pi(i)K(i,j)$, then $\alpha(i,j) = \frac{\pi(j)K(j,i)}{\pi(i)K(i,j)}$ and $\alpha(j,i) = 1$, so that

$$\pi(i)MK(i,j) = \pi(i)K(i,j)\frac{\pi(j)K(j,i)}{\pi(i)K(i,j)} = \pi(j)K(j,i) = \pi(j)MK(j,i),$$

which establishes reversibility. Since $0 < \alpha(i,j) \leq 1$, then $K(i,j) > 0$ if and only if $MK(i,j) > 0$, so irreducibility and aperiodicity are preserved. Q.E.D.

5. Proposition. In Theorem 4, if the original Markov chain is not aperiodic and not reversible, then the conclusion of the theorem continues to be true.

Proof: Since the chain is not reversible, there exists a pair of states i and j such that

$$\pi(j)K(j,i) < \pi(i)K(i,j).$$

Thus if $\{Y_n : n \geq 0\}$ denotes the Markov chain determined by $\{\pi, MK\}$, it follows that $0 < \alpha_{ij} < 1$ and

$$P([X_1 = i]|[X_0 = i]) \geq 1 - \alpha_{ij} > 0,$$

from which aperiodicity follows. Reversibility of $\{Y_n\}$ follows from theorem 4.

6. Proposition. Suppose $\{X_n\}$ is a Markov chain determined by $\{\pi, \mathbf{P}\}$ that is irreducible and is reversible (and therefore stationary) but is not aperiodic. If $\{Y_n\}$ is a Markov chain determined by (the same) π but with stochastic matrix $\mathbf{Q} = (q_{ij}\}$, where

$$q_{ij} = \begin{cases} 0.9p_{ij} \text{ if } i \neq j \\ 1 - 0.9\sum\{p_{ik} : k \neq i\} \text{ if } i = j, \end{cases}$$

then $\{Y_n\}$ is a Markov chain that is irreducible, reversible and aperiodic.

Proof: Since $\{X_n\}$ is reversible, we know that

$$\pi(i)p_{ij} = \pi(j)p_{ji}.$$

Applying this, we get

$$\pi(i)q_{ij} = 0.9\pi(i)p_{ij} = 0.9\pi(j)p_{ji} = \pi(j)q_{ji},$$

which means that $\{Y_n\}$ is reversible and therefore stationary. Since

$$q_{ii} = (1 - 0.9\sum\{p_{ik} : k \neq i\}) : k \neq i\} > 0,$$

it follows that $\{Y_n\}$ is aperiodic. Finally, $\{Y_n\}$ is irreducible because $\{X_n\}$ is irreducible and $p_{ij} > 0$ if and only if $q_{ij} > 0$.

7. Markov Chain Monte Carlo Algorithm. Everything developed in these notes so far constitutes the prerequisites for this section. Suppose there is given a finite state space, i.e., a set, $\mathbf{X}$ of elements with a probability measure P defined over it. Suppose there is a function φ:$\mathbf{X} \to \mathbf{R}^1$ whose value, $\varphi(x)$, is known or can be determined for every $x \in \mathbf{X}$. An observation X is taken at random according to the probability measure P, and a state x is observed. One wishes to test the null hypothesis that P is actually some known probability π, whose value $\pi(y)$ is known for each $y \in \mathbf{X}$ at least up to a multiplicative constant, i.e., for any pair u, v in $\mathbf{X}$, the ratio $\pi(u)/\pi(v)$ is known or can be determined. The problem is to obtain an "accurate estimate" of $\pi(\{z \in \mathbf{X} : \varphi(z) \leq \varphi(x)\})$ for some particular $x \in \mathbf{X}$. The problem is made difficult in that one does not know π completely and thus cannot simulate repeated observations on X according to the probability measure π in order to estimate $\pi(\{z \in \mathbf{X} : \varphi(z) \leq \varphi(x)\})$ by means of the strong law of large numbers.

A solution to the problem is possible under the following condition. Suppose you can come up with a stochastic matrix $K(\cdot, \cdot)$ that is indexed by $\mathbf{X}$ and has a small number of positive entries in each row such that the Markov chain determined by $\{\pi, K\}$ is reversible (or at least stationary), irreducible and aperiodic. In this case, one starts with the state x. Since the stochastic matrix we have at our disposal contains only a relatively small number of positive entries in the row determined by x, one can simulate the probability distribution $K(x, \cdot)$ and take an observation on $\mathbf{X}$ according to this distribution.. The outcome of this will be called X_1. Then, whatever the outcome

is, say, $X_1 = x_1$, one may simulate an observation on $\mathbf{X}$ according to tne probability distribution $K(x_1, \cdot)$, and call the outcome X_2. In this way one can generate an observable Markov chain $\{X_n\}$ that is reversible, irreducible and aperiodic. After a very large number n of observations, our simulated observation on X will be, if our hypothesis is correct, none other than the distribution π; this follows from our first big limit theorem. So at some time after taking a large number N of observations on the Markov chain one may begin start observing $I_{[\varphi(X_n)\leq\varphi(x)]}$ for $n \geq N$. Keeping track of the number of 1's for the next M simulations yields the quantity

$$S_M = \sum_{n=N}^{N+M} I_{[\varphi(X_n)\leq\varphi(x)]}.$$

By the second big limit theorem, known as the ergodic theorem, the ratio $\frac{1}{M}S_M$ converges with probability one as $M \to \infty$ to $\pi(\{z \in \mathbf{X} : \varphi(z) \leq \varphi(x)\})$.

Next suppose that the only stochastic matrix that one can come up with that is indexed by $\mathbf{X}$ and that has a relatively small number of positive entries in each row is such that the Markov chain determined by $\{\pi, K\}$ is irreducible and reversible (and therefore stationary) but not aperiodic. In this case, for every pair states $i \neq j$, define a new Markov chain $\{\pi, AK\}$ by $AK(i,j) = 0.9K(i,j)$, and for $j = i$, let

$$AK(i,i) = 1 - \sum\{AK(i,j) : j \neq i\}.$$

In this case, this Markov chain is irreducible, reversible and aperiodic by proposition 6, and one can then proceed to estimate $\pi(\{z \in \mathbf{X} : \varphi(z) \leq \varphi(x)\})$ as above.

Next suppose that the only stochastic matrix K that one can come up with that is indexed by $\mathbf{X}$ and that has a relatively small number of positive entries in each row is such that the Markov chain determined by $\{\pi, K\}$ is irreducible but is not stationary. In this case, we would like to alter K to obtain a stochastic matrix denoted by MK such that the Markov chain determined by $\{\pi, MK\}$ is stationary, irreducible and aperiodic. In this case, define

$$MK(i,j) = \begin{cases} K(i,j)\alpha(i,j) \text{ if } i \neq j \\ 1 - \sum\{K(i,k)\alpha(i,k) : k \neq i\} \text{ if } j = i, \end{cases}$$

where

$$\alpha(i,j) = \min\left\{1, \frac{\pi(j)K(j,i)}{\pi(i)K(i,j)}\right\}.$$

Then by theorem 4, the Markov chain determined by $\{\pi, MK\}$ is reversible and therefore stationary, it is still irreducible, and by proposition 7, it is aperiodic. The procedure in this case and in a previous case is an application of what is referred to in the literature as the Hastings-Metropolis procedure.

In the case when the Hastings-Metropolis procedure is applied, the stochastic matrix MK enjoys a certain opimality property, which we now develop.

Let $\mathbf{X}$ denote a finite state space, and let π denote a probability measure over $\mathbf{X}$ such that $\pi(x) > 0$ for all $x \in \mathbf{X}$. Let $S(\mathbf{X})$ denote the set of all stochastic matrices indexed by $\mathbf{X}$ such that for each $K \in S(\mathbf{X})$, the Markov chain determined by $\{\pi, K\}$, is irreducible and aperiodic. Let $R(\pi) \subset S(\mathbf{X})$ be defined as the set of all $K \in S(\mathbf{X})$ such that the Markov chain determined by $\{\pi, K\}$ is reversible. Finally over $S(\mathbf{X})$ we create a metric $d(K, K')$ defined by

$$\begin{aligned} d(K, K') &= \textstyle\sum\{\pi(x)|K(x,y) - K'(x,y)| : x \in \mathbf{X}, y \in \mathbf{X}, x \neq y\} \\ &= \textstyle\sum_{x \in \mathbf{X}} \pi(x) \sum_{\{y: y \neq x\}} |K(x,y) - K'(x,y)|. \end{aligned}$$

It should be noted that $d(\cdot, \cdot)$ satisfies the properties of a metric over $S(\mathbf{X})$.

Next we define the mapping $M : S(\mathbf{X}) \to R(\mathbf{X})$ by

$$\begin{aligned} MK(x,y) &= K(x,y) \wedge \tfrac{\pi(y)}{\pi(x)} K(y,x) \text{ if } x \neq y \\ &= 1 - \textstyle\sum_{\{z: z \neq x\}} MK(x,z) \text{ if } x = y. \end{aligned}$$

By theorem 4 above, the Markov chain determined by $\{\pi, MK\}$ is reversible, irreducible and aperiodic. The mapping $M : K \to MK$ is called the Metropolis map.

8. Theorem. (Billera and Diaconis) For every $K \in S(\mathbf{X}) \backslash R(\pi)$, MK is the unique closest element of $R(\pi)$ to K in the metric d that is coordinatewise not larger than K off the main diagonal.

Proof: Let x and y denote any two states in $\mathbf{X}$, where $x \neq y$. Let us define, for these two fixed states,

$$\begin{aligned} H_{xy} &= \{K \in S(\mathbf{X}) : \pi(x)K(x,y) = \pi(y)K(y,x)\}, \\ H^+_{xy} &= \{K \in S(\mathbf{X}) : \pi(x)K(x,y) > \pi(y)K(y,x)\}, \text{ and} \\ H^-_{xy} &= \{K \in S(\mathbf{X}) : \pi(x)K(x,y) < \pi(y)K(y,x)\}. \end{aligned}$$

Thus, for every $x \neq y$, $S(\mathbf{X})$ is the union of these three disjoint sets:

$$S(\mathbf{X}) = H_{xy} \cup H^+_{xy} \cup H^-_{xy}.$$

We shall break up the proof into a sequence of claims and subclaims.

Claim 1: If $K \in S(\mathbf{X})$, then $K \in H^+_{xy}$ if and only if $K \in H^-_{yx}$. This claim follows directly from the definitions.

Now let $N \in R(\mathbf{X})$ be arbitrary. Then by the definition of $R(\mathbf{X})$,

$$\pi(x)N(x,y) = \pi(y)N(y,x)$$

for all x, y in $\mathbf{X}$.

Claim 2: For all $N \in R(\pi)$ and all $K \in S(\mathbf{X})\backslash R(\pi)$,

$$d(K,N) \geq \sum\{\pi(x)|K(x,y) - N(x,y)| + \pi(y)|K(y,x) - N(y,x)|\},$$

where the sum is taken over only those pairs of elements x, y in $\mathbf{X}$ that satisfy: $x \neq y$ and $K \in H^+_{xy}$. To prove this we break up $d(K,N)$ into a sum of three sums:

$$\begin{aligned} d(K,N) &= \textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H^+_{xy}\}} |K(x,y) - N(x,y)| \\ &+ \textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H_{xy}\}} |K(x,y) - N(x,y)| \\ &+ \textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H^-_{xy}\}} |K(x,y) - N(x,y)| \\ &\geq \textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H^+_{xy}\}} |K(x,y) - N(x,y)| \\ &+ \textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H^-_{xy}\}} |K(x,y) - N(x,y)|. \end{aligned}$$

By claim 1, this last sum above is

$$\textstyle\sum_{x\in\mathbf{X}} \pi(x) \sum_{\{y:y\neq x, K\in H^+_{yx}\}} |K(x,y) - N(x,y)|,$$

and now replace in it each x by a y and each y by an x to obtain

$$\textstyle\sum_{y\in\mathbf{X}} \pi(y) \sum_{\{x:x\neq y, K\in H^+_{xy}\}} |K(y,x) - N(y,x)|,$$

which proves the claim.

Claim 3. $d(K, MK) \leq d(K,N)$. To prove this, we first note that since $K(x,y) \geq MK(x,y)$ for $x \neq y$, we obtain.

$$\begin{aligned} d(K,MK) &= \textstyle\sum_{x\neq y}\{\pi(x)|K(x,y) - MK(x,y)|\} \\ &= \textstyle\sum_{x\neq y}\{\pi(x)K(x,y) - \pi(x)MK(x,y)\} \\ &= \textstyle\sum_{x\neq y}\{\pi(x)K(x,y) - \pi(x)\{K(x,y) \wedge \frac{\pi(y)}{\pi(x)}K(y,x)\}\}. \end{aligned}$$

Then, by the definitions of H^+_{xy}, H_{xy} and H^-_{xy}, (and for this fixed K), we continue the above to obtain

$$\begin{aligned} d(K,MK) &= \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}}\{\pi(x)K(x,y) - \pi(y)K(y,x)\} \\ &+ \textstyle\sum_{\{x\neq y: K\in H_{xy}\}}\{\pi(x)K(x,y) - \pi(x)K(x,y)\} \\ &+ \textstyle\sum_{\{x\neq y: K\in H^-_{xy}\}}\{\pi(x)K(x,y) - \pi(x)K(x,y)\} \\ &= \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}}\{\pi(x)K(x,y) - \pi(y)K(y,x)\}\}. \end{aligned}$$

Since N satisfies $\pi(x)N(x,y) = \pi(y)N(y,x)$, the the last sum above is equal to

$$\sum_{x\neq y, K\in H^+_{xy}} \{\pi(x)(K(x,y) - N(x,y)) + \pi(y)(N(y,x) - K(y,x))\},$$

which is

$$\leq \sum_{x\neq y, K\in H^+_{xy}} \{\pi(x)|K(x,y) - N(x,y)| + \pi(y)|N(y,x) - K(y,x)|\}$$

By claim 2, we obtain $d(K, MK) \leq d(K, N)$, which proves claim 3.

By claim 3, MK is **a closest** element of $R(\pi)$ to K with respect to the metric d that is majorized off the main diagonal by K. We now wish to prove that MK is **the closest** element of $R(\pi)$ to K that is majorized off the main diagonal by K.

Claim 4. As before, $K \in S(\mathbf{X})\backslash R(\pi)$ is fixed. Let $N \in R(\pi)$ be such that $N(x,y) \leq K(x,y)$ for all $x \neq y$ in $\mathbf{X}$. Then there is at least one pair (x,y), $x \neq y$, such that $N(x,y) < K(x,y)$ and thus for at least this (x,y),

$$K \in H^+_{xy} \cup H^-_{xy}.$$

To show this, if, to the contrary, every pair (x,y) satisfies $N(x,y) = K(x,y)$, then K would be in $R(\pi)$, which contradicts the definition of K. Since there is at least one such pair, (x,y), for which $N(x,y) < K(x,y)$, for this pair, K cannot be in H_{xy}; consequently, $K \in H^+_{xy} \cup H^-_{xy}$.

As before, $K \in S(\mathbf{X})\backslash R(\pi)$, but let $N \in R(\pi)$ be such that $N(x,y) \leq K(x,y)$ for all $x \neq y$ in $\mathbf{X}$ with $N(x,y) < K(x,y)$ for at least one such pair. Let us define, for all pairs $x \neq y$ in $\mathbf{X}$, the function

$$\epsilon(x,y) = N(x,y) - K(x,y).$$

Note that, for all pairs if distinct elements x, y in $\mathbf{X}$, $\epsilon(x,y) \leq 0$, with $\epsilon(x,y) < 0$ for at least one such pair. Since $N(x,y) = K(x,y) + \epsilon(x,y)$ and $N \in R(\pi)$, we have

$$\begin{aligned} N(y,x) &= \tfrac{\pi(x)}{\pi(y)} N(x,y) \\ &= \tfrac{\pi(x)}{\pi(y)} (K(x,y) + \epsilon(x,y)). \end{aligned}$$

By claim 2 and this last expression, $d(K,N) \geq$

$$\begin{aligned} &\geq \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} \{\pi(x)|K(x,y) - N(x,y)| + \pi(y)|K(y,x) - N(y,x)|\} \\ &= \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} \{\pi(x)|\epsilon(x,y)| + \pi(y)|K(y,x) - \tfrac{\pi(x)}{\pi(y)}(K(x,y) + \epsilon(x,y))|\} \\ &= \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} \pi(x)|\epsilon(x,y)| \\ &+ \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} |\pi(y)K(y,x) - \pi(x)K(x,y) - \pi(x)\epsilon(x,y)|. \end{aligned}$$

Now for those $x \neq y$ for which $K \in H^+_{xy}$, the expression $\pi(y)K(y,x) - \pi(x)K(x,y)$ is negative, while $-\pi(x)\epsilon(x,y)$ is nonnegative, with at least one term strictly positive by claim 4. These two observations imply

$$\begin{aligned} d(K,N) &> \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} \pi(x)|\epsilon(x,y)| \\ &+ \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} |\pi(y)K(y,x) - \pi(x)K(x,y)| \\ &> \textstyle\sum_{\{x\neq y: K\in H^+_{xy}\}} (\pi(x)K(x,y) - \pi(y)K(y,x)). \end{aligned}$$

But recall that in our proof of claim 3 we proved that

$$d(K, MK) = \sum_{x\neq y, K\in H^+_{xy}} (\pi(x)K(x,y) - \pi(y)K(y,x)).$$

Thus $d(K,N) > d(K,MK)$, which proves the theorem.

Exercises for Part VI

These exercises are designed to demonstrate in a very simple manner an application of the Markov Chain Monte Carlo algorithm. Suppose we can take 10 independent observations a random vector $\begin{pmatrix} U \\ V \end{pmatrix}$, whose range is

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Suppose we desire to test the null hypothesis that U and V are independent, that is, for some unknown positive probabilities p_1, p_2, q_1, q_2, q_3, the joint density of U, V satisfies

$$P([U=i][V=j]) = p_i q_j, i = 1,2, j = 1,2,3.$$

So we take 10 independent observations on $\begin{pmatrix} U \\ V \end{pmatrix}$, and obtain results that can be arranged in a table with two rows and three columns, namely,

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix},$$

where X_{ij} = the number of observations which take U-values in row i and V-values in column j. Thus

$$\sum_{i=1}^{2}\sum_{j=1}^{3} X_{ij} = \sum_{j=1}^{3}\sum_{i=1}^{2} X_{ij} = 10.$$

1. Verify that the total number of tables produced in this way is 6^{10}.

2. Now suppose that when we take these ten observations, we obtain the following table:

$$\begin{bmatrix} 2 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

The row totals are 5 and 5, and the column totals are $3, 3, 4$ respectively.

3. Let us denote the row totals for any of the 6^{10} outcomes by $X_{1\cdot}, X_{2\cdot}$ and the column totals by $X_{\cdot 1}, X_{\cdot 2,} X_{\cdot 3}$. Compute

$$P([X_{11} = i][X_{12} = j]|[X_{1\cdot} = 5][X_{\cdot 1} = 3][X_{\cdot 2} = 3])$$

for all (i, j)'s that do not violate the conditioning event.

4. Note that in problem 3, this conditional joint distribution of X_{11} and X_{12} does not depend on the unknown parameters p_1, q_1, q_2. So now find the conditional density of χ^2 given $[X_{1\cdot} = 5][X_{\cdot 1} = 3][X_{\cdot 2} = 3]$, where

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(X_{ij} - 10X_{i\cdot}X_{\cdot j})^2}{10X_{i\cdot}X_{\cdot j}}.$$

5. For the particular table of results obtained in problem 2, compute the value, χ_0^2, of the χ^2-statistic defined in problem 4.

6. Assuming that it is really true that the selections of the row and column were independent of each other with unknown probabilities p_1, p_2 for the rows and q_1, q_2, q_3 for the columns, i.e., that U and V are independent, use your results in problem 3 to compute

$$P\left([\chi^2 \geq \chi_0^2]|[X_{1\cdot} = 5][X_{\cdot 1} = 3][X_{\cdot 2} = 3]\right).$$

If this is too small, we would reject the null hypothesis that the selections of the rows and the columns are independent. In this problem, this computation could be easily carried out. However, for more rows and/or columns and for larger sample sizes, this is an impossible task. But this probability can be simulated, and this is by means of of the Hastings-Metropolis algorithm, which you are now asked to carry out in a sequence of steps outlined in problems 7, 8, 9 and 10.

7. First note that the conditional density computed in problem 3 has 13 points in its domain. Thus we have a state space of 13 states. In brief, we would like to have at our disposal a Markov chain with this 13 state state space and a transition matrix for which this conditional density is the

invariant measure. Having these, we could then simulate the Markov chain, and for a large number of observations we could find the relative frequency with which the event $[\chi^2 \geq \chi_0^2]$ occurs. By the ergodic theorem for Markov chains, this relative frequency converges to $P([\chi^2 \geq \chi_0^2])$ with probability 1. So we need a stochastic matrix, which we construct and simulate as follows. Whichever state the Markov chain is at, select two columns in the table at random without replacement. These two columns and the two rows determine a 2×2 table or matrix. Then select one of the two matrices

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

at random with probabilities $\frac{1}{2}, \frac{1}{2}$. If you can add whatever you get to the two columns selected at random without replacement and not obtain any negative integers, then do so, and call the resulting 2×3 table the new state of the chain. On the other hand, if you do get a -1, then the next state is the same as the present one. Thus a Markov chain is created whose state space is the set of all 2×3 tables of nonnegative integers whose row sums are 5 and 5 and whose column sums are $3, 3$ and 4. Display the transition matrix, $\mathbf{P} = (p(i,j))$, indexing the rows and columns in the same manner as the states are indexed for the invariant probability measure, i.e., the conditional density computed in problem 3.

8. With respect to the Markov chain created in problem 7 which could be simulated, verify that this chain is irreducible and aperiodic.

9. If you let μ denote the desired invariant probability measure (as a row vector), and if the rows of $\mathbf{P}$ are indexed in the same order as the coordinates of μ, you should be able to verify that the equation $\mu = \mu\mathbf{P}$ (matrix multiplication) is not satisfied. Thus the stochastic matrix $\mathbf{P}$ is not the one we need for a stationary Markov chain.

10. We now know how to alter $\mathbf{P}$ to get a new stochastic matrix $\mathbf{Q} = (q_{ij})$ which together with the initial distribution,

$$\mu = (\mu(1), \cdots, \mu(13)),$$

will determine a stationary Markov chain. So define

$$q(i,j) = \begin{cases} p(i,j)\alpha(i,j) \text{ if } i \neq j \\ 1 - \sum\{p(i,k)\alpha(i,k) : k \neq i\} \text{ if } j = i, \end{cases}$$

where

$$\alpha(i,j) = \min\left\{1, \frac{\pi(j)p(j,i)}{\pi(i)p(i,j)}\right\}$$

for all those ordered pairs of states (i, j) for which $p(i, j) > 0$. Now simulate this chain, $(\mu, \mathbf{Q})$, say, $25,000$ times, starting from any state you choose, and then at each simulation for the next $25,000$ observations compute the relative frequency with which the event $[\chi^2 \geq \chi_0^2]$ occurs. This relative frequency should be very close to the conditional probability that was obtained in problem 3. (Epilogue: Why did we simulate the chain $25,00$ times before we started simulating $P([\chi^2 \geq \chi_0^2])$? For some reason, one wants to start the chain in which an estimate of $P([\chi^2 \geq \chi_0^2])$ is computed "according to the invariant measure μ.").