Module 3

Long-run behavior of Markov chain

Topics: Classification of states, Long-run fractions, Stationary and limiting distributions

Classification of states

Recurrence vs. transience: We have studied long-run behavior of absorbing Markov chains. We classified states of absorbing Markov chains into absorbing and non-absorbing states. In the case of finite-state absorbing Markov chains, absorption is guaranteed, i.e., the chain will eventually end up in one of the absorbing states. We calculated probability of absorption into different absorbing states. For the infinite-state case, it is possible that the chain does not end up in any of the absorbing states and diverges. In two such chains, we were able to calculate probabilities of absorption and divergence.

For non-absorbing Markov chains, classification of states into absorbing and non-absorbing states does not work. There we classify the states into recurrent and transient states. As the names suggest, recurrent states are visited by the Markov chain again and again, while the transient states are visited less, as in the chain stops visiting the transient states after some time. Transient states have similarity with non-absorbing states, which too are not visited by the chain after absorption takes place. Recurrent states have similarity with absorbing states, but there's an important difference. Markov chain cannot leave an absorbing state, which is allowed for recurrent states. Formal definitions are given below.

Let $T_{ij} \coloneqq \min\{n \geq 1: X_n = j | X_0 = i\}$ denote time of first visit to state j from state i for all $i, j \in \Omega$. It's a random variable because the Markov chain can take different paths leading to different values for T_{ij}. The '$n \geq 1$' condition in the definition of T_{ij} may seem redundant, but it becomes important when $i = j$. T_{jj} without the condition '$n \geq 1$' is zero and it's not a random variable, but with the condition T_{jj} denotes the time till first return. The definition of recurrent and transient states is based on the time till first return.

In the Markov chain to the right, let us consider $T_{2,2}$. If the chain takes the path $2 \to 2$, then $T_{2,2} = 1$, for path $2 \to 3 \to 2$, $T_{2,2} = 2$, for $2 \to 1$, $T_{2,2} = \infty$, and so on. Since $T_{2,2}$ can be ∞ with positive probability, then the chain may never return to state 2. So, it is a transient state. Note that $T_{2,2}$ is an improper random variable. If it were a proper random variable, then return would have been guaranteed and state 2 would have been recurrent.

p_{ij}	1	2	3	4	5
1	1	0	0	0	0
2	0.1	0.5	0.4	0	0
3	0	0.2	0.6	0.2	0
4	0	0	0	0.5	0.5
5	0	0	0	0.3	0.7

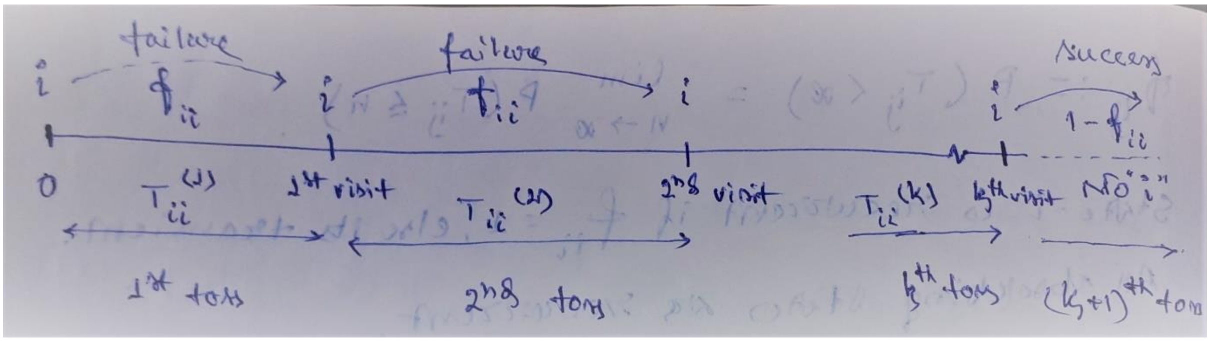
Based on the above illustration, state-j is recurrent if $P(T_{jj} < \infty) = 1$, else it's transient. Note that $P(T_{jj} < \infty) = \lim_{t\to\infty} P(T_{jj} \leq t) = \sum_{t=1}^{\infty} P(T_{jj} = t)$. Let $f_{ij} = P(T_{ij} < \infty)$ for all $i, j \in \Omega$. Then j is recurrent if $f_{jj} = 1$, else it's transient.

Let us calculate $f_{2,2}$. $P(T_{2,2}=1) = P(\text{path } 2 \to 2) = 0.5$, $P(T_{2,2}=2) = P(\text{path } 2 \to 3 \to 2) = 0.4 \times 0.2$, $P(T_{2,2}=3) = P(\text{path } 2 \to 3 \to 3 \to 2) = 0.4 \times 0.6 \times 0.2$, $P(T_{2,2}=4) = P(\text{path } 2 \to 3 \to 3 \to 3 \to 2) = 0.4 \times 0.6^2 \times 0.2$, and so on. For any $t \geq 2$, $P(T_{2,2}=t) = 0.4 \times 0.6^{t-2} \times 0.2$. Then $f_{2,2} = \sum_{t=1}^{\infty} P(T_{2,2}=t) = 0.5 + \sum_{t=2}^{\infty} 0.4 \times 0.6^{t-2} \times 0.2 = 0.5 + 0.08 \times (1 + 0.6 + 0.6^2 + \cdots) = 0.5 + 0.08/(1-0.6) = 0.7$. So, state-2 is transient. In a similar manner, one can verify that state-3 is transient, and the rest are recurrent.

For an absorbing state, in any kind of Markov chain, $P(T_{jj}=1) = P(\text{path } j \to j) = 1$. Thus, absorbing states are recurrent. For any non-absorbing state i in an absorbing Markov chain, $\exists n < \infty$ s.t. $p_{ij}^{(n)} > 0$ for some absorbing state j. Then $f_{ii} < 1$, because there is a finite length path of 'no return' with positive probability. So, non-absorbing states in an absorbing Markov chain are transient, irrespective of absorption is guaranteed or not.

Alternate characterization: A recurrent state is visited by a Markov chain again and again, and a transient state may never be visited by the chain after some time. In light of this, if we consider the expected number of returns to a state, it would appear that the expectation is infinite for recurrent states and finite for transient states. Let $N_{ij} \coloneqq \sum_{n=1}^{\infty} \mathbb{1}(X_n = j | X_0 = i)$ denote the number of visits to state j from state i for all $i, j \in \Omega$. N_{ij} is random variable and $E[N_{ij}] = E[\sum_{n=1}^{\infty} \mathbb{1}(X_n = j | X_0 = i)] = \sum_{n=1}^{\infty} E[\mathbb{1}(X_n = j | X_0 = i)] = \sum_{n=1}^{\infty} p_{ij}^{(n)}$. Our guess is that $E[N_{ii}] = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ if i is recurrent, else $E[N_{ii}] < \infty$. This is true, as shown below. This equivalence provides an alternate characterization of recurrence vs. transience.

Let us proof the equivalence. Consider that a coin is tossed every time the Markov chain is in state i, including at the start. Consider return to state i (at some point in time) as failure and no return as success. Then probability of success is $P(T_{ii} = \infty) = 1 - f_{ii}$ and probability of failure is $P(T_{ii} < \infty) = f_{ii}$. This is depicted in the following diagram.



Clearly, the number of returns to state i is one less than the number of tosses, i.e., $N_{ii} = N - 1$, where N denotes the number of tosses. Due to the Markov property, these tosses can be regarded as independent and identical Bernoulli trials with success probability $1 - f_{ii}$. Then N is the number of tosses required to get the first success, i.e., $N \sim Geo(1 - f_{ii})$.

$$\Rightarrow E[N_{ii}] = E[N] - 1 = \begin{cases} \dfrac{1}{1-f_{ii}} - 1 = \dfrac{f_{ii}}{1-f_{ii}} < \infty & \text{if } f_{ii} < 1 \\ \infty - 1 = \infty & \text{if } f_{ii} = 1 \end{cases}$$

If i is recurrent, then $f_{ii} = 1$ and $E[N_{ii}] = \infty$. If i is transient, then $f_{ii} < 1$ and $E[N_{ii}] < \infty$, as claimed. It can be restated as: *i is recurrent if and only if $E[N_{ii}] = \infty$*.

The alternate characterization of recurrence vs. transience provides another way of checking if a state is recurrent or transient. Earlier, we obtained mass function of T_{ii} to check if $f_{ii} = 1$ or not, and accordingly declared i to be recurrent or transient. However, this works for Markov chains with a few states and simpler structures. Consider the chain to the right. It's difficult to obtain f_{ii} here.

p_{ij}	1	2	3	4	5
1	0.8	0	0.1	0	0.1
2	0.1	0.7	0.2	0	0
3	0	0.1	0.6	0.2	0.1
4	0.3	0	0.2	0.5	0
5	0	0.3	0	0.3	0.4

We can use a computer to numerically check whether $E[N_{ii}] = \sum_{n=1}^{\infty} p_{ii}^{(n)}$ diverges to infinity or converges to some finite quantity in the above Markov chain, and accordingly declared i to be recurrent or transient. However, convergence may be slow in some cases leading to doubt whether a state is recurrent or transient. Also, we must check for every state separately, which can be time consuming or even impossible. We need simpler tools to identify nature of states. Then idea of communication and its implications provide us such tools.

Communication and nature of states: For $i \neq j$, we say that state i leads to state j, denoted by $i \rightarrow j$, if there exists $n < \infty$ s.t. $p_{ij}^{(n)} > 0$, i.e., j can be reached from i in finite time with positive probability. If $i \rightarrow j$ and $j \rightarrow i$, then we say that i and j communicate and denote it by $i \leftrightarrow j$. Earlier, we used '→' informally to denote paths taken by Markov chain. Now, we are using it formally to represent communication among states.

It's not difficult to identify which states communicate (with one another) in a Markov chain. For smaller chains, it can be done by visual inspection. For example, in the first chain $2 \leftrightarrow 3$ and $4 \leftrightarrow 5$; also, $2,3 \rightarrow 1,4,5$. In the second chain, all pairs of states communicate. For larger chains, we can obtain $p_{ij}^{(n)}$ numerically for $n = 1,2,\dots,|\Omega|$, and if any of these are positive, then $i \rightarrow j$. We don't need to go beyond $n = |\Omega|$, because paths of greater length involve repetition of states. Consider a path from i to j is of length $m > |\Omega|$. The path must be of the form $i \cdots (n_1 \text{ steps}) \cdots k \cdots (m - n_1 - n_2 \text{ steps}) \cdots k \cdots (n_2 \text{ steps}) \cdots j$, where $n_1 + n_2 \leq |\Omega|$ and $k \in \Omega$. If this path has positive probability, then the $i \cdots (n_1 \text{ steps}) \cdots k \cdots (n_2 \text{ steps}) \cdots j$ also has positive probability. Therefore, if $p_{ij}^{(m)} > 0$ for some $m > |\Omega|$, then there exists $n \leq |\Omega|$ s.t. $p_{ij}^{(n)} > 0$. This ensures that only finite number of matrix multiplications are required. For infinite-state Markov chains, transition probabilities have patterns, which can be used to tell if two states communicate or not, in a case-by-case basis.

Let $i \leftrightarrow j$ and i be recurrent. Since $i \leftrightarrow j$, $\exists m, n < \infty$ s.t. $p_{ij}^{(m)}, p_{ji}^{(n)} > 0$. Then

$$E[N_{jj}] = \sum_{k=1}^{\infty} p_{jj}^{(k)} \geq \sum_{k=1}^{\infty} p_{jj}^{(m+n+k)} \geq \sum_{k=1}^{\infty} p_{ji}^{(n)} p_{ii}^{(k)} p_{ij}^{(m)} = p_{ij}^{(m)} p_{ji}^{(n)} \sum_{k=1}^{\infty} p_{ii}^{(k)} = p_{ij}^{(m)} p_{ji}^{(n)} E[N_{ii}]$$

Since i is recurrent, $E[N_{ii}] = \infty$. Then $E[N_{jj}] \geq p_{ij}^{(m)} p_{ji}^{(n)} E[N_{ii}] = \infty$, i.e., j is recurrent. On the other hand, if $i \leftrightarrow j$ and i is transient, then j must be transient. Instead, if j is recurrent, then it violates the just established result. Hence, *if $i \leftrightarrow j$, then their nature, i.e., recurrence vs. transience, must be the same*.

The above result greatly simplifies the task of identifying nature of states. If we can identify nature of one state, say i, by calculating f_{ii} or $E[N_{ii}]$, then we know the nature of every other state that communicate with i. For the first chain, it's sufficient to identify nature of state $2/3$ and $4/5$; 1, being an absorbing state, is recurrent. For the second chain it's sufficient to identify nature of any one of the states. We can do even better by considering decomposition of states of a Markov chain into communicating blocks.

Decomposition of states: A non-null subset of states $C \subseteq \Omega$ is called a communicating block or an equivalence class if $i \leftrightarrow j$ for all $i, j \in C$ and $i \nrightarrow j$ for all $i \in C, j \notin C$. So, all states in a communicating block leads to one another, but does not lead to any state outside. Note that a state outside a communicating block may lead to a state inside, but the other way is not true. In the first chain, $\{4,5\}$ is a communicating block, but $\{2,3\}$ or $\{3,4\}$ or $\{4\}$ are not. $\{1\}$ is a communicating block, as the 1st condition is not applicable and the 2nd condition is satisfied. With this logic, all absorbing states, individually, are communicating blocks. In the second chain, the whole Ω is a communicating block.

We can decompose states of a Markov chains into some number of communicating blocks and the set of remaining states, i.e., $\Omega = C_1 \cup C_2 \cup \cdots \cup C_k \cup R$, where $C_1, C_2, \dots, C_k$ for $k \geq 0$ are communicating blocks and $R = \Omega \setminus (C_1 \cup C_2 \cup \cdots \cup C_k)$ is the set of remaining states. This decomposition for the first chain is: $C_1 = \{1\}$, $C_2 = \{4,5\}$, $R = \{2,3\}$. For the second chain, the decomposition is: $C_1 = \Omega$, $R = \emptyset$. A Markov chain without a communicating block is: $p_{i,i} = p \in (0,1)$, $p_{i,i+1} = 1 - p$ for $i = 1,2,3, \dots$. Here, $R = \Omega$. Decomposition plays an important role in the study of long-run behavior of Markov chains.

There is only one way of decomposing states of a Markov chain into communicating blocks. Try to check this *yourself*. We can identify the unique decomposition by listing down the communications among states. Use of graph theoretic concepts can improve the efficiency of this procedure. This approach does not work for infinite-state Markov chains. There we have to rely on the pattern in transition probabilities.

Decomposition and the nature of states: We say that i is recurrent if $f_{ii} = P(T_{ii} < \infty) = 1$ or $E[N_{ii}] = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, otherwise i is transient. Alternately, if $i \leftrightarrow j$ and j is recurrent (or transient), then i is recurrent (or transient). We don't need to do any of these if i is in a finite communicating block or in the set of remaining states. In the first case, i is recurrent, and in the second case, i is transient. If i is in an infinite communicating block, then we cannot comment on the nature of i; it can go either way.

States not in any communicating block are transient

Consider $i \in R = \Omega \setminus (C_1 \cup C_2 \cup \cdots \cup C_k)$, where $C_1, C_2, \dots, C_k$ are the communicating blocks of Ω. Let $R(i) = \left\{j : p_{ij}^{(n)} > 0, n \geq 1\right\}$ denote the set of states that can be reached from i. If $i \notin R(i)$, then it is impossible to return to i, and thus, i is transient as claimed. If $i \in R(i)$, then there is at least one more member in $R(i)$; otherwise, i is an absorbing state and $\{i\}$ is a communicating block, which is impossible. If $j\ (\neq i)$ is in $R(i)$ and $j \nrightarrow i$, then there is a path of no-return to i (via j), which makes i transient as claimed. Finally, if all $j \in R(i) \setminus \{i\}$ lead to i, then $R(i)$ is a communicating block, which is impossible. This establishes the claim.

The last part of the above argument relies on the following: Let $R(i)$ denote the set of states that can be reached from i. If all $j \in R(i) \setminus \{i\}$ lead to i, then $R(i)$ is a communicating block. Verify validity of this assertion *yourself*.

States in finite communicating blocks are recurrent

Consider a finite communicating block $C \subseteq \Omega$. If the Markov chain reaches a state in C, it can never leave. Then $\sum_{j \in C} p_{ij} = 1$ for all $i \in C$, which is the only requirement of a square matrix of non-negative entries to be the transition probability matrix of a Markov chain. So, C itself can be regarded as a Markov chain. In this chain, all pairs of states communicate. Then all states are either recurrent or transient. The claim is that the former is true.

Let $i \in C$ denote the state through which the Markov chain Ω enters C. Then $X_0 = i$ for the Markov chain C. Let j denote an arbitrary state in C. Then the claim is that j is recurrent. Assuming otherwise, let j be transient. It can be proved that if j is transient, then $E[N_{ij}] = \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all i. We have shown this for $i = j$. A proof for the general case appears in Appendix A. Since the series $\sum_{n=1}^{\infty} p_{ij}^{(n)}$ is convergent, its limiting term must be zero, i.e., $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. Then $\sum_{j \in C} \lim_{n\to\infty} p_{ij}^{(n)} = 0$.

Now, row sums of every n-step transition probability matrix is 1, i.e., $\sum_{j \in C} p_{ij}^{(n)} = 1$ for all $n \geq 1$. Taking limit as $n \to \infty$, we get $\lim_{n\to\infty} \sum_{j \in C} p_{ij}^{(n)} = 1$. Since C is finite, we can alter the order of limit and sum, and obtain $\sum_{j \in C} \lim_{n\to\infty} p_{ij}^{(n)} = 1$. This is in contradiction with the above observation that $\sum_{j \in C} \lim_{n\to\infty} p_{ij}^{(n)} = 0$. Therefore, our assumption that j is transient must be incorrect. Hence, j (and the whole C) is recurrent.

If C is infinite, then the change in the order of limit and sum is not permitted (even with the convergence theorems), and we do not reach the contradiction that $0 = 1$. In fact, infinite communicating blocks can be of either type, i.e., all states are either recurrent or transient. We provide examples of both kinds next.

Two examples of infinite communicating block

Let us consider random walk on integers. Here, $\Omega = \mathbb{Z}$ and $p_{i,i+1} = p \in (0,1)$ and $p_{i,i-1} = q = 1 - p$ for all $i \in \Omega$. For any pair of states $i < j$, $p_{ij}^{(j-i)} = p^{j-i} > 0$, $p_{ji}^{(j-i)} = q^{j-i} > 0$, and $j - i < \infty$. Thus, $i \leftrightarrow j$ for all $i < j$. So, the whole Ω is a communicating block and it's infinite. We show that all states are recurrent if $p = 1/2$, else all states are transient.

Since all pairs of states communicate, showing that a state is recurrent/transient is sufficient to show that the whole block is recurrent/transient. We show that $E[N_{00}] = \infty$ if $p = 1/2$, else it's finite. $E[N_{00}] = \sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} p_{00}^{(2n-1)} + \sum_{n=1}^{\infty} p_{00}^{(2n)}$; the odd and even cases are written separately. Starting with state 0, the random walk cannot visit state 0 in odd number of steps, i.e., $p_{00}^{(2n-1)} = 0$ for all n. For even number of steps, it is possible if and only if there are equal number of right and left movements. So, $p_{00}^{(2n)} = \binom{2n}{n} p^n q^n$ for all n. Then

$$
\begin{aligned}
E[N_{00}] = \sum_{n=1}^{\infty} \binom{2n}{n} p^n q^n &= \sum_{n=1}^{\infty} \frac{2n(2n-1)(2n-2)\cdots 21}{n!\,n!} p^n q^n \\
&= \sum_{n=1}^{\infty} 2^n \frac{(2n-1)(2n-3)\cdots 31}{n!} p^n q^n \\
&= \sum_{n=1}^{\infty} 2^n 2^n \left\{ \frac{2n-1}{2n} \cdot \frac{2n-3}{2(n-1)} \cdot \cdots \cdot \frac{3}{2\times 2} \cdot \frac{1}{2\times 1} \right\} p^n q^n \\
&= \sum_{n=1}^{\infty} (4pq)^n \left(\prod_{k=1}^{n} \frac{2k-1}{2k} \right)
\end{aligned}
$$

If $p = 1/2$, then $4pq = 1$ and we have

$$
\begin{aligned}
E[N_{00}] = \sum_{n=1}^{\infty} \prod_{k=1}^{n} \frac{2k-1}{2k} &= \prod_{k=1}^{1} \frac{2k-1}{2k} + \prod_{k=1}^{2} \frac{2k-1}{2k} + \prod_{k=1}^{3} \frac{2k-1}{2k} + \cdots \\
&= \frac{1}{2} + \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} + \cdots \geq \frac{1}{2} + \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \cdots = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots
\end{aligned}
$$

The above infinite series diverges. Hence, $E[N_{00}] = \infty$, as expected.

$4pq = 4p(1-p)$ is a strictly concave function with maxima $p = 1/2$ and maximum value 1. Therefore, if $p \neq 1/2$, $4pq < 1$. Also, $(2k-1)/2k < 1$ for all $k \geq 1$. Then

$$
E[N_{00}] = \sum_{n=1}^{\infty} (4pq)^n \left(\prod_{k=1}^{n} \frac{2k-1}{2k} \right) < \sum_{n=1}^{\infty} (4pq)^n = \frac{4pq}{1-4pq} < \infty, \text{as expected}
$$

In this case, with time, the Markov chain may drift towards $+\infty$ if $p > q$ or towards $-\infty$ if $p < q$, and never return to 0 again.

Long-run fractions

Stationary and limiting distributions

Appendix A

For a general Markov chain, we have shown that $E[N_{jj}] = f_{jj}/(1-f_{jj})$ by introducing $N \sim Geo(1-f_{jj})$ such that $N_{jj} = N - 1$. If j is transient, then $f_{jj} < 1$ and $E[N_{jj}] < \infty$. There $X_0 = j$. Here, with an arbitrary starting state i, we obtain $E[N_{ij}]$.

N_{ij} is the number of visits to state j from state i. In order to visit j from i for the first time, T_{ij} must be finite. Once the first visit to j has happened, then the subsequent visits are essentially returns to j. Therefore, the event $\{N_{ij} \geq k\}$ for $k \geq 1$ is same as first visit to j from i in finite time (i.e., $T_{ij} < \infty$) and then at least $k-1$ returns to j (i.e., $N_{jj} \geq k-1$).

$$\begin{aligned} P(N_{ij} \geq k) &= P(T_{ij} < \infty, N_{jj} \geq k-1) \\ &= P(T_{ij} < \infty) P(N_{jj} \geq k-1), \text{due to Markov property} \\ &= f_{ij} P(N-1 \geq k-1) = f_{ij} P(N > k-1) = f_{ij} f_{jj}^{k-1} \text{ for } k \geq 1 \end{aligned}$$

We have used the connection between N_{jj} and $N \sim Geo(1-f_{jj})$ in the above derivation. For $k = 0$, $P(N_{ij} \geq k) = 1$. Now, we obtain mass function as follows:

$$P(N_{ij} = k) = P(N_{ij} \geq k) - P(N_{ij} \geq k+1) = \begin{cases} f_{ij} f_{jj}^{k-1}(1-f_{jj}) \text{ for } k \geq 1 \\ 1 - f_{ij} \text{ for } k = 0 \end{cases}$$

Note that if $f_{ij} = 0$, then $P(N_{ij} = 0) = 1$ and $E[N_{ij}] = 0$. If $f_{ij} > 0$, then

$$E[N_{ij}] = \sum_{k=1}^{\infty} k f_{ij} f_{jj}^{k-1}(1-f_{jj}) = f_{ij} E[N] = \begin{cases} \dfrac{f_{ij}}{1-f_{jj}} < \infty \text{ if } f_{jj} < 1 \\ f_{ij} \times \infty = \infty \text{ if } f_{jj} = 1 \end{cases}$$

If j is transient, then $f_{jj} < 1$ and $E[N_{ij}] < \infty$, as desired. It shall be noted that if $f_{ij} = 0$, then $E[N_{ij}] = 0$ irrespective of the nature of j. The above expressions for the mass function and expectation of N_{ij} are valid for arbitrary i, j in any Markov chain.