Markov Chains

One of the ways of generating a sequence of dependent random variables is to think of a system evolving in time. We have time points that are discrete say $T = 0, 1, \cdots, N, \cdots$. The state of the system is described by a point x in the state space $\mathcal{X}$ of the system. The state space $\mathcal{X}$ comes with a natural σ-field of subsets $\mathcal{F}$. At time 0 the system is in a random state and its distribution is specified by a probability distribution μ_0 on $(\mathcal{X}, \mathcal{F})$. At successive times $T = 1, 2, \cdots$, the system changes its state and given the past history $(x_0, \cdots, x_k)$ of the states of the system at times $T = 0, \cdots, k-1$ the probability that system finds itself at time k in a subset $A \in \mathcal{F}$ is given by $\pi_k(x_0, \cdots, x_{k-1}\,; A\,)$. For each $(x_0, \cdots, x_{k-1})$, π_k defines a probability measure on $(\mathcal{X}, \mathcal{F})$ and for each $A \in \mathcal{F}$, $\pi_k(x_0, \cdots, x_{k-1}\,; A\,)$ is assumed to be a measurable function of $(x_0, \cdots, x_{k-1})$, on the space $(\mathcal{X}^k, \mathcal{F}^k)$ which is the product of k copies of the space $(\mathcal{X}, \mathcal{F})$ with itself. We can inductively define measures μ_k on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1})$ that describe the probability distribution of the entire history $(x_0, \cdots, x_k)$ of the system through time k. To go from μ_{k-1} to μ_k we think of $(\mathcal{X}^{k+1},\ \mathcal{F}^{k+1})$ as the product of $(\mathcal{X}^k,\ \mathcal{F}^k)$ with $(\mathcal{X}, \mathcal{F})$ and construct on $(\mathcal{X}^{k+1},\ \mathcal{F}^{k+1}\,)$ a probability measure with marginal μ_{k-1} on $(\mathcal{X}^k,\ \mathcal{F}^k)$ and conditionals $\pi_k(x_0, \cdots, x_{k-1}; \cdot)$ on the fibers $(x_1, \cdots, x_{k-1}) \times \mathcal{X}$. This will define μ_k and the induction can proceed. We may stop at some finite terminal time N or go on indefinitely. If we do go on indefinitely, we will have a consitent family of finite dimensional distributions $\{\mu_k\}$ on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1})$ and we may try to use Kolmogorov's theorem to construct a probability measure P on the space $(\mathcal{X}^\infty, \mathcal{F}^\infty)$ of sequences $\{x_j : j \geq 0\}$ representing the total evolution of the system for all times. However Kolmogorov's theorem requires some assumptions on $(\mathcal{X}, \mathcal{F})$ that are satisfied if $\mathcal{X}$ is a complete separable metric space and $\mathcal{F}$ are the Borel sets. However, in the present context, there is a result known as Tulcea's theorem that proves the existence of a P on $(\mathcal{X}^\infty, \mathcal{F}^\infty)$ for any choice of $(\mathcal{X}, \mathcal{F})$, exploiting the fact that the consistent family of finite dimensional distributions μ_k arise from well defined successive regular conditional probability distributions.

An important subclass is generated when the transition probability depends on the past history only through the current state. In other words

$$\pi_k(x_0, \cdots, x_{k-1}; \cdot) = \pi_{k-1,k}\,(x_{k-1}; \cdot).$$

In such a case the process is called a Markov Process with transition probabilities $\pi_{k-1,k}\,(\cdot,\cdot)$. An even smaller subclass arises when we demand that $\pi_{k-1,k}\,(\cdot,\cdot)$ be the same for different values of k. A single transition probability $\pi(x, A)$ and the initial distribution μ_0 determine the entire process i.e. the measure P on $(\mathcal{X}^\infty, \mathcal{F}^\infty)$. Such processes are called time-homogeneous Markov Proceses or Markov Processes with stationary transition probabilities.

Chapman-Kolmogorov Equations:. If we have the transition probabilities $\pi_{k,k+1}$ of transition from time k to $k+1$ of a Markov Chain it is possible to obtain directly the transition probabilities from time k to $k+\ell$ for any $l \geq 2$. We do it by induction on ℓ. Define

$$\pi_{k,k+\ell+1}(x, A) = \int_{\mathcal{X}} \pi_{k,k+\ell}(x, dy)\, \pi_{k+\ell,k+\ell+1}(y, A) \tag{3.1}$$

or equivalently, in a more direct fashion

$$\pi_{k,k+\ell+1}(x, A) = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \pi_{k,k+1}(x, dy_{k+1}) \cdots \pi_{k+\ell,k+\ell+1}(y_{k+\ell}, A)$$

Theorem 3.1. The transition probabilities $\pi_{k,m}(\cdot,\cdot)$ satisfy the relations

$$\pi_{k,n}(x, A) = \int_{\mathcal{X}} \pi_{k,m}(x, dy)\, \pi_{m,n}(y, A) \tag{3.2}$$

for any $k < m < n$ and for the Markov Process defined by the one step transition probabilities $\pi_{k,k+1}(\cdot,\cdot)$, for any $n > m$

$$P[x_n \in A|\Sigma_m] = \pi_{m,n}(x_m, A) \quad \text{a.e.}$$

where Σ_m is the σ-field of past history upto time m generated by the coordinates $x_0, x_1, \cdots, x_m$.

Proof: The identity is basically algebra. The multiple integral can be carried out by iteration in any order and after enough variables are integrated we get our identity. To prove that the conditional probabilities are given by the right formula we need to establish

$$P[\{x_n \in A\} \cap B] = \int_B \pi_{m,n}(x_m, A)\, dP$$

for all $B \in \Sigma_m$ and $A \in \mathcal{F}$. We write

$$
\begin{aligned}
&P[\{x_n \in A\} \cap B] = \int_{\{x_n \in A\} \cap B} dP \\
&= \int \cdots \int_{\{x_n \in A\} \cap B} d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m) \\
&\qquad\qquad \pi_{m,m+1}(x_m, dx_{m-1}) \cdots \pi_{n-1,n}(x_{n-1}, dx_n) \\
&= \int \cdots \int_B d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m) \\
&\qquad\qquad \pi_{m,m+1}(x_m, dx_{m-1}) \cdots \pi_{n-1,n}(x_{n-1}, A) \\
&= \int \cdots \int_B d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m)\, \pi_{m,n}(x_m, A) \\
&= \int_B \pi_{m,n}(x_m, A)\, dP
\end{aligned}
$$

and we are done.

Remark: If the chain has stationary transition probabilities then the transition probabilities $\pi_{m,n}(x, dy)$ from time m to time n depend only on the difference $k = n - m$ and are given by what are usually called the k step transition probabilities. They are defined inductively by

$$\pi^{(k+1)}(x, A) = \int_{\mathcal{X}} \pi^{(k)}(x, dy)\, \pi(y, A)$$

and satisfy the Chapman-Kolmogorov equations

$$\pi^{(k+\ell)}(x, A) = \int_{\mathcal{X}} \pi^{(k)}(x, dy) \pi^{(\ell)}(x, A) = \int_{\S} \pi^{(\ell)}(x, dy) \pi^{(k)}(y, A)$$

Let us look at some examples.

Suppose we have an urn containg a certain number of balls (nonzero) some red and others green. A ball is drawn at random and its color is noted. Then it is returned to the urn along with an extra ball of the same color. Then a new ball is drawn at random and the process continues ad infinitum. The current state of the system can be characterized by two integers r, g such that $r + g \geq 1$. The initial state if the system is some r_0, g_0 with $r_0 + g_0 \geq 1$. The system can go from (r, g)

to either $(r+1,g)$ with probability $\frac{r}{r+g}$ or to $(r,g+1)$ with probability $\frac{g}{r+g}$. This is clearly an example of a Markov Chain with stationary transition probabilities.

Consider a queue for service in a store. Suppose at each of the times $1,2,\cdots$, a random number of new customers arrive and and join the queue. If the queue is non empty at some time, then exactly one customer will be served and will leave the queue at the next time point. The distribution of the number of new arrivals is specified by $\{p_j : j \geq 0\}$ where p_j is the probability that exactly j new customers arrive at a given time. The number of new arrivals at distinct times are assumed to be independent. The queue length is a Markov Chain on the state space $\mathcal{X} = \{0,1,\cdots,\}$ of nonegative integers. The transition probabilities $\pi(i,j)$ are given by $\pi(0,j) = p_j$ because there is no service and nobody in the queue to begin with and all the new arrivals join the queue. On the other hand $\pi(i,j) = p_{j-i+1}$ if $j+1 \geq i \geq 1$ because one person leaves the queue after being served.

Consider a reservoir into which water flows. The amount of additional water flowing into the reservoir on any given day is random, and has a distribution α on $[0,\infty)$. The demand is also random for any given day, with a probability distribution β on $[0,\infty)$. We may also assume that the inflows and demands on successive days are random variables ξ_n and η_n, that have α and β for their common distributions and are all mutually independent. We may wish to assume a percentage loss due to evaporation. In any case the storage level at successive days have a recurrence relation

$$S_{n+1} = [(1-p)S_n + \xi_n - \eta_n]^+$$

p is the loss and we have put the condition that the outflow is the demand unless the stored amount is less than the demand in which case the outflow is the available quantity. The current amount in storage is a Markov Process with Stationary transition probabilities.

Let $X_1,\cdots,X_n,\cdots$ be a sequence of independent random variables with a common distribution α. Let $S_n = Y + X_1 + \cdot + X_n$ for $n \geq 1$ with $S_0 = Y$ where Y is a random variable independent of $X_1,\ldots,X_n,\ldots$ with distribution μ. Then S_n is a Markov chain on R with one step transition probaility $\pi(x, A) = \alpha(A - x)$ and initial distribution μ. The n step transition probability is $\alpha^n(A - x)$ where α^n is the n-fold convolution of α.

These are examples that can described by models of the type

$$x_n = f(x_{n-1}, \xi_n)$$

where x_n is the current state and ξ_n is some random external disturbance. ξ_n are assumed to be independent and identically distributed. They could have two components like inflow and demand. The new state is a deterministic function of the old state and the noise.

Exercise 3.10. Verify that our previous two examples can be cast in the above form. In fact there is no loss of generality in assuming that ξ_j are mutually independent random variables having as common distribution the uniform distribution on the interval $[0, 1]$.

Given a Markov Chain with stationary transition probabilities $\pi(x, dy)$ on a state space $(\mathcal{X}, \mathcal{F})$, the behavior of $\pi^{(n)}(x, dy)$ for large n is an important and natural question. In the best situation of independent random variables $\pi^{(n)}(x, A) = \mu(A)$ are independent of x as well as n. Hopefully after a long time the Chain will "forget" its origins and $\pi^{(n)}(x, \cdot) \to \mu(\cdot)$, in some suitable sense, for some μ that does not depend on x . If that happens, then from the relation

$$\pi^{(n+1)}(x, A) = \int \pi^{(n)}(x, dy)\pi(y, A),$$

we conclude

$$\mu(A) = \int \pi(y, A)\, d\mu(y) \quad \text{for all} \quad A \in \mathcal{F}$$

Measures that satisfy the above property abbreviated as $\mu\pi = \mu$ are called invariant measures for the Markov Chain. If we start with the initial distribution μ which is inavariant then the probability measure P has μ as marginal at every time. In fact P is stationary i.e., invariant with respect to time translation, and can be extended to a stationary process where time runs from $-\infty$ to $+\infty$.

Stopping Times and Renewal Times

One of the important notions in the analysis of Markov Chains is the idea of stopping times and renewal times. A function

$$\tau(\omega) : \Omega \to \{n : n \geq 0\}$$

is a random variable defined on the set $\Omega = \mathcal{X}^\infty$ such that for every $n \geq 0$ the set $\{\omega : \tau(\omega) = n\}$ (or equivalently for each $n \geq 0$ the set $\{\omega : \tau(\omega) \leq n\}$) is measurable with respect to the σ-field $\mathcal{F}_n$ generated by $X_j : 0 \leq j \leq n$. It is not necessary that $\tau(\omega) < \infty$ for every ω. Such random variable τ are called stopping times. Examples of stopping times are, constant times $n \geq 0$, the first visit to a state x, or the second visit to a state x. The important thing is that in order to decide if $\tau \leq n$ i.e. to know if what ever is supposed to happen did happen before time n the chain need be observed only up to time n. Examples of τ that are not stopping times are easy to find. The last time a site is visited is not a stopping time nor is is the first time such that at the next time one is in a state x. An important fact is that the Markov property extends to stopping times. Just as we have σ-fields $\mathcal{F}_n$ associated with constant times, we do have a σ field $\mathcal{F}_\tau$ associated to any stopping time. This is the information we have when we observe the chain upto time τ. Formally

$$\mathcal{F}_\tau = \Big\{A : A \in \mathcal{F}^\infty \quad \text{and} \quad A \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \text{for each} \quad n\Big\}$$

One can check from the definition that τ is $\mathcal{F}_\tau$ measurable and so is X_τ on the set $\tau < \infty$. If τ is the time of first visit to y then τ is a stopping time and the event that the chain visits a state z before visiting y is $\mathcal{F}_\tau$ measurable.

Lemma 3.2. At any stopping time τ the Markov Property holds in the sense that the conditional distribution of $X_{\tau+1}, \cdots, X_{\tau+n}, \cdots$ conditioned on $\mathcal{F}_\tau$ is the same as the original chain starting from the state $x = X_\tau$ on the set $\tau < \infty$. In other words

$$\begin{aligned}
&P_x\{X_{\tau+1} \in A_1, \cdots, X_{\tau+n} \in A_n | \mathcal{F}_\tau\} \\
&\quad = \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n)
\end{aligned}$$

a.e. on $\{\tau < \infty\}$.

Proof: Let $A \in \mathcal{F}_\tau$ be given with $A \subset \{\tau < \infty\}$. Then

$$\begin{aligned} &P_x\{A \cap \{X_{\tau+1} \in A_1, \cdots, X_{\tau+n} \in A_n\}\} \\ &\quad = \sum_k P_x\{A \cap \{\tau = k\} \cap \{X_{k+1} \in A_1, \cdots, X_{k+n} \in A_n\}\} \\ &\quad = \sum_k \int_{A\cap\{\tau=k\}} \int_{A_1} \cdots \int_{A_n} \pi(X_k, dx_{k+1}) \cdots \pi(x_{k+n-1}, dx_{k+n})\, dP_x \\ &\quad = \int_A \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n)\, dP_x \end{aligned}$$

We have used the fact that if $A \in \mathcal{F}_\tau$ then $A \cap \{\tau = k\} \in \mathcal{F}_k$ for every $k \geq 0$.

Remark: If $X_\tau = y$ a.e. with respect to P_x on the set $\tau < \infty$, then at time τ, when it is finite, the process starts afresh with no memory of the past and will have conditionally the same probabilities in the future as P_y. At such times the process renews itself and these times are called renewal times.

Countable State Space

From the point of view of analysis a particularly simple situation is when the state space $\mathcal{X}$ is a countable set. It can be taken as the integers $\{x : x \geq 1\}$. Many applications fall in this category and an understanding of what happens in this situation will tell us what to expect in general.

The one step transition probability is a matrix $\pi(x, y)$ with nonnegative entries such that $\sum_y \pi(x, y) = 1$ for each x. Such matrices are called stochastic matrices. The n step transition matrix is just the n-th power of the matrix defined inductively by

$$\pi^{(n+1)}(x, y) = \sum_z \pi^{(n)}(x, z)\pi(z, y)$$

To be consistent one defines $\pi^{(0)}(x, y) = \delta_{x,y}$ which is 1 if $x = y$ and 0 otherwise. The problem is to analyse the behaviour for large n of $\pi^{(n)}(x, y)$. A state x is said to communicate with a state y if $\pi^{(n)}(x, y) >$

0 for some $n \geq 1$. We will assume for simplicity that every state communicates with every other state. Such Markov Chains are called **irreducible**. Let us first limit ourselves to the study of irreducible chains. Given an irreducible Markov chain with transition probabilities $\pi(x, y)$ we define $f_n(x)$ as the probability of returning to x for the first time at the n-th step assuming that the chain starts from the state x.. Using the convention that P_x refers to the measure on sequences for the chain starting from x and $\{X_j\}$ are the successive positions of the chain

$$\begin{aligned} f_n(x) &= P_x\Big\{X_j \neq x \text{ for } 1 \leq j \leq n-1 \text{ and } X_n = x\Big\} \\ &= \sum_{\substack{y_1 \neq x \\ \cdots \\ y_{n-1} \neq x}} \pi(x, y_1)\, \pi(y_1, y_2) \cdots \pi(y_{n-1}, x) \end{aligned}$$

Since $f_n(x)$ are probailities of disjoint events $\sum_n f_n(x) \leq 1$. The state x is called **transient** if $\sum_n f_n(x) < 1$ and **recurrent** if $\sum_n f_n(x) = 1$. The recurrent case is divided into two situations. If we denote by $\tau_x = \inf\{n \geq 1 : X_n = x\}$, the time of first visit to x, then recurrence is $P_x\{\tau_x < \infty\} = 1$. A recurrent state x is called **positive recurrent** if

$$E^{P_x}\{\tau_x\} = \sum_{n \geq 1} n\, f_n(x) < \infty$$

and **null recurrent** if

$$E^{P_x}\{\tau_x\} = \sum_{n \geq 1} n\, f_n(x) = \infty$$

Lemma 3.3. If for a (not necessarily irreducible) chain starting from x, the probability of ever visiting y is positive then so is the probability of visiting y before returning to x.

Proof: Assume that for the chain starting from x the probability of visiting y before returning to x is zero. But when it returns to x it starts afresh and so will not visit y until it returns again. This reasoning can be repeated and so the chain will have to visit x infinitely often before visiting y. But this will use up all the time and so it cannot visit y at all.

Lemma 3.4. For an irreducible chain all states x are of the same type.

Proof: Let x be recurrent and y be given. Since the chain is irreducible, for some k, $\pi^{(k)}(x,y) > 0$. By the previous lemma, for the chain starting from x, there is a positive probability of visiting y before returning to x. After each successive return to x, the chain starts afresh and there is a fixed positive probability of visiting y before the next return to x. Since there are infinitely many returns to x, y will be visited infinitely many times as well. Or y is also a recurrent state.

We now prove that if x is positive recurrent then so is y. We saw already that the probability $p = P_x\{\tau_y < \tau_x\}$ of visiting y before returning to x is positive. Clearly

$$E^{P_x}\{\tau_x\} \geq P_x\{\tau_y < \tau_x\}\, E^{P_y}\{\tau_x\}$$

and therefore

$$E^{P_y}\{\tau_x\} \leq \frac{1}{p} E^{P_x}\{\tau_x\} < \infty$$

On the other hand we can write

$$\begin{aligned}
E^{P_x}\{\tau_y\} &\leq \int_{\tau_y<\tau_x} \tau_x\, dP_x + \int_{\tau_x<\tau_y} \tau_y\, dP_x \\
&= \int_{\tau_y<\tau_x} \tau_x\, dP_x + \int_{\tau_x<\tau_y} \{\tau_x + E^{P_x}\{\tau_y\}\}\, dP_x \\
&= \int_{\tau_y<\tau_x} \tau_x\, dP_x + \int_{\tau_x<\tau_y} \tau_x\, dP_x + (1-p)\, E^{P_x}\{\tau_y\} \\
&= \int \tau_x\, dP_x + (1-p)\, E^{P_x}\{\tau_y\}
\end{aligned}$$

by the renewal property at the stopping time τ_x. Therefore

$$E^{P_x}\{\tau_y\} \leq \frac{1}{p} E^{P_x}\{\tau_x\}$$

We also have

$$E^{P_y}\{\tau_y\} \leq E^{P_y}\{\tau_x\} + E^{P_x}\{\tau_y\} \leq \frac{2}{p} E^{P_x}\{\tau_x\}$$

proving that y is positive recurrent.

Transient Case: We have the following theorem regarding transience.

Theorem 3.5. An irreducible chain is transient if and only if

$$G(x,y) = \sum_{n=0}^{\infty} \pi^{(n)}(x,y) < \infty \quad \text{for all} \quad x,y$$

Moreover for any two states x and y,

$$G(x,y) = f(x,y)G(y,y)$$

and

$$G(x,x) = \frac{1}{1 - f(x,x)}$$

where $f(x,y) = P_x\{\tau_y < \infty\}$.

Proof: Each time the chain returns to x there is a probability $1-f(x,x)$ of never returning. The number of returns has then the geometric distribution

$$P_x\{ \text{ exactly } \ n \ \text{ returns to } \ x\} = (1 - f(x,x))f(x,x)^n$$

and the expected number of returns is given by

$$\sum_{k=1}^{\infty} \pi^{(k)}(x,x) = \frac{f(x,x)}{1 - f(x,x)}$$

The left hand side comes from the calculation

$$E^{P_x} \sum_{k=1}^{\infty} \chi_{\{x\}}(X_k) = \sum_{k=1}^{\infty} \pi^{(k)}(x,x)$$

and the right hand side from the calculation of the mean of a Geometric distribution. Since we count the visit at time 0 as a visit to x we add 1 to both sides to get our formula. If we want to calculate the expected number of visits to y when we start from x, first we have to get to y and the probability of that is $f(x,y)$. Then by the renewal property it is exactly the same as the expected number of visits to y starting from y, including the visit at time 0 and that equals $G(y,y)$.

Before we study the recurrent behavior we need the notion of periodicity. For each state x let us define $D_x = \{n : \pi^{(n)}(x,x) > 0\}$ to be the set of times at which a return to x is possible if one starts from x. We define d_x to be the greatest common divisor of D_x.

Lemma 3.6. For any irreducible chain $d_x = d$ for all $x \in \mathcal{X}$ and for each x, D_x contains all sufficiently large multiples of d.

Proof: Let us define

$$D_{x,y} = \{n : \pi^{(n)}(x,y) > 0\}$$

so that $D_x = D_{x,x}$. By the Chapman-Kolmogorov equations

$$\pi^{(m+n)}(x,y) \geq \pi^{(m)}(x,z)\pi^{(n)}(z,y)$$

for every z, so that if $m \in D_{x,z}$ and $n \in D_{z,y}$, then $m + n \in D_{x,y}$. In particular if $m, n \in D_x$ it follows that $m + n \in D_x$. Since any pair of states communicate with each other, given $x, y \in \mathcal{X}$, there are positive integers n_1 and n_2 such that $n_1 \in D_{x,y}$ and $n_2 \in D_{y,x}$. This implies that with the choice of $\ell = n_1 + n_2$, $n + \ell \in D_x$ whenever $n \in D_y$; similarly $n + \ell \in D_y$ whenever $n \in D_x$. Since ℓ itself belongs to both D_x and D_y both d_x and d_y divide ℓ. Suppose $n \in D_x$. Then $n + \ell \in D_y$ and therefore d_y divides $n + \ell$. Since d_y divides ℓ, d_y must divide n. Since this is true for every $n \in D_x$ and d_x is the greatest common divisor of D_x, d_y must divide d_x. Similarly d_x must divide d_y. Hence $d_x = d_y$. We now complete the proof of the lemma. Let d be the greatest common divisor of D_x. Then it is the greatest common divisor of a finite subset $n_1, n_2, \cdots, n_q$ of D_x and there will exist integers $a_1, a_2, \cdots, a_q$ such that

$$a_1 n_1 + a_2 n_2 + \cdots + a_q n_q = d$$

Some of the a's will be positive and others negative. Seperating them out, and remembering that all the n_i are divisible by d, we find two integers $md, (m+1)d$ such that they both belong to D_x. If now $n = kd$ with $k > m^2$ we can write $k = \ell m + r$ with a large $\ell \geq m$ and the remainder r is less than m.

$$kd = (\ell m + r)d = \ell md + r(m+1)d - rmd = (\ell - r)md + r(m+1)d \in D_x$$

since $\ell \geq m > r$.

Remark: For an irreducible chain the common value d is called the period of the chain and an irreducible chain with period $d = 1$ is called aperiodic.

The simplest example of a periodic chain is one with 2 states and the chain shuttles back and forth between the two. $\pi(x, y) = 1$ if $x \neq y$ and 0 if $x = y$. A simple calculation yields $\pi^{(n)}(x, x) = 1$ if n is even and 0 otherwise. There is oscillatory behavior in n that persists. The main theorem for irreducible, aperiodic, recurrent chains is the following.

Theorem 3.7. Let $\pi(x, y)$ be the one step transition probability for a recurrent aperiodic Markov chain and let $\pi^{(n)}(x, y)$ be the n-step transition probabilities. If the chain is null recurrent then

$$\lim_{n\to\infty} \pi^{(n)}(x, y) = 0 \quad \text{for all} \quad x, y$$

If the chain is positive recuurrent then of course $E^{P_x}\{\tau_x\} = m(x) < \infty$ for all x, and in that case

$$\lim_{n\to\infty} \pi^{(n)}(x, y) = q(y) = \frac{1}{m(y)}$$

exist for all x and y is independent of the starting point x and $\sum_y q(y) = 1$.

The proof is based on

Theorem 3.8. **(Renewal Theorem.)** Let $\{f_n : n \geq 1\}$ be a sequence of nonnegative numbers such that

$$\sum_n f_n = 1, \qquad \sum_n n\, f_n = m \leq \infty$$

and the greatest common divisor of $\{n : f_n > 0\}$ is 1. Suppose that $\{p_n : n \geq 0\}$ are defined by $p_0 = 1$ and recursively

$$p_n = \sum_{j=1}^{n} f_j\, p_{n-j} \tag{3.3}$$

Then

$$\lim_{n\to\infty} p_n = \frac{1}{m}$$

where if $m = \infty$ the right hand side is taken as 0.

The proof is based on several steps.

Step 1. We have inductively $p_n \leq 1$. Let $a = \limsup_{n\to\infty} p_n$. We can choose a subsequence n_k such that $p_{n_k} \to a$. We can assume without loss of generality that $p_{n_k+j} \to q_j$ as $k \to \infty$ for all positive and negative integers j as well. Of course the limit q_0 for $j = 0$ is a. In relation 3.3 we can pass to the limit along the subsequence and use the dominated convergence theorem to obtain

$$q_n = \sum_{j=1}^{\infty} f_j \, q_{n-j} \tag{3.4}$$

valid for $-\infty < n < \infty$. In particular

$$q_0 = \sum_{j=1}^{\infty} f_j \, q_{-j} \tag{3.5}$$

Step 2: Because $a = \limsup p_n$ we can conclude that $q_j \leq a$ for all j. If we denote by $S = \{n : f_n > 0\}$ then $q_{-k} = a$ for $k \in S$. We can then deduce from equation 3.4 that $q_{-k} = a$ for $k = k_1 + k_2$ with $k_1, k_2 \in S$. By repeating the same reasoning $q_{-k} = a$ for $k = k_1 + k_2 + \cdots + k_\ell$. By lemma 3.6 because the greatest common factor of the integers in S is 1, there is a k_0 such that for $k \geq k_0$,we have $q_{-k} = a$. We now apply the relation 3.4 again to conclude that $q_j = a$ for all positve as well as negative j.

Step 3: If we add up equation 3.3 for $n = 1, \cdots, N$ we get

$$\begin{aligned} p_1 + p_2 + \cdots + p_N &= (f_1 + f_2 + \cdots + f_N) + (f_1 + f_2 + \cdots + f_{N-1})p_1 \\ &+ \cdots + (f_1 + f_2 + \cdots + f_{N-k})p_k + \cdots + f_1 p_{N-1} \end{aligned}$$

If we denote by $T_j = \sum_{i=j}^{\infty} f_i$, we have $T_1 = 1$ and $\sum_{j=0}^{\infty} T_j = m$. We can now rewrite

$$\sum_{j=1}^{N} T_j p_{N-j+1} = \sum_{j=1}^{N} f_j$$

Step 4: Because $p_{N-j} \to a$ for every j along the subsequence $N = n_k$, if $\sum_j T_j = m < \infty$, we can deduce from the dominated convergence theorem that $m\,a = 1$ and we conclude that

$$\limsup_{n\to\infty} p_n = \frac{1}{m}$$

If $\sum_j T_j = \infty$, by Fatou's Lemma $a = 0$. Exactly the same argument applies to liminf and we conclude that

$$\liminf_{n\to\infty} p_n = \frac{1}{m}$$

This concludes the proof of the theorem.

We now turn to the proof of our main theorem.

If we take a fixed $x \in \mathcal{X}$ and consider $f_n = P_x\{\tau_x = n\}$, then f_n and $p_n = \pi^{(n)}(x,x)$ are related by (1) and $m = E^{P_x}\{\tau_x\}$. In order to apply the renewal theorem we need to establish that the greatest common divisor of $S = \{n : f_n > 0\}$ is 1. In general if $f_n > 0$ so is p_n. So the greatest common divisor of S is always larger than that of $\{n : p_n > 0\}$. That does not help us because the greatest common divisor of $\{n : p_n > 0\}$ is 1. On the other hand if $f_n = 0$ unless $n = k\,d$ for some k, the relation 3.3 can be used inductively to conclude that the same is true of p_n. Hence both sets have the same greatest common divisor. We can now conclude that

$$\lim_{n\to\infty} \pi^{(n)}(x,x) = q(x) = \frac{1}{m(x)}$$

On the other hand if $f_n(x,y) = P_x\{\tau_y = n\}$, then

$$\pi^{(n)}(x,y) = \sum_{k=1}^{n} f_k(x,y)\,\pi^{(n-k)}(y,y)$$

and recurrence implies $\sum_{k+1}^{\infty} f_k(x,y) = 1$ for all x and y. Therefore

$$\lim_{n\to\infty} \pi^{(n)}(x,y) = q(y) = \frac{1}{m(y)}$$

and is independent of x, the starting point. In order to complete the proof we have to establish that

$$Q = \sum_y q(y) = 1$$

It is clear by Fatou's lemma that

$$\sum_y q(y) = Q \leq 1$$

By letting $n \to \infty$ in the Chapman-Kolmogorov equation

$$\pi^{(n+1)}(x, y) = \sum_z \pi^n(x, z)\pi(z, y)$$

and using Fatou's lemma we get

$$q(y) \geq \sum_z \pi(z, y)q(z)$$

Summing with repect to y we obtain

$$Q \geq \sum_{z,y} \pi(z, y)q(z) = Q$$

and equality holds in this relation. Therefore

$$q(y) = \sum_z \pi(z, y)q(z)$$

for every y or $q(\cdot)$ is an invariant measure. By iteration

$$q(y) = \sum_z \pi^n(z, y)q(z)$$

and if we let $n \to \infty$ again an application of the bounded convergence theorem yields

$$q(y) = Q\, q(y)$$

implying $Q = 1$ and we are done.