Chapter 8

Markov Chains

8.1 Non-Linear Stochastic Recursions

We have just finished studying the class of "state space models", in which the dynamics take the form

$$X_{n+1} = F_n X_n + W_n + u_n, \tag{8.1}$$

where $(W_n : n \geq 0)$ is independent, and the sequences $(F_n : n \geq 0)$ and $(u_n : n \geq 0)$ are deterministic. These models have affine/linear dynamics.

To describe non-linear phenomena, we can instead consider S-valued sequences $X = (X_n : n \geq 0)$ that arise as solutions to stochastic recursions of the form

$$X_{n+1} = f_{n+1}(X_n, Z_{n+1}), \tag{8.2}$$

where $(Z_n : n \geq 1)$ is a sequence of independent $\tilde{S}$-valued random variables, and $f_{n+1} : S \times \tilde{S} \to S$ is a deterministic map for $n \geq 0$. This clearly generalizes (8.1) .

8.2 The Markov Property

A sequence $X = (X_n : n \geq 0)$ is said to be *Markov* if for $n \geq 0$, $B \subseteq S$.

$$P(X_{n+1} \in B | X_0, \cdots, X_n) = P(X_{n+1} \in B | X_n). \tag{8.3}$$

When S is discrete, the *one-step transition kernel* $P_n = (P_n(x,y) : x, y \in S)$ can be viewed as a matrix in which the *(x,y)'th* entry is $P(X_n = y | X_{n-1} = x)$. The Markov sequence $X = (X_n : n \geq 0)$ (or equivalently, *Markov chain* X) is said to have *stationary transition probabilities* (or, equivalently, to be *time-homogeneous*) if $P_n = P$ for $n \geq 1$.

Remark A matrix $A = A(x,y) : x, y \in S)$ is said to be *stochastic* if $A(x,y) \geq 0$ for $x, y \in S$, and $\sum_y A(x,y) = 1$ for $x \in S$. Note that the P_n's defined above must necessarily be stochastic.

Exercise 8.1:
Prove that a sequence $X = (X_n : n \geq 0)$ is Markov if and only if for each $A \subseteq S^{n+1}, B \subseteq S^\infty, P((X_0, \cdots, X_n) \in A, (X_n, X_{n+1}, \cdots) \in B | X_n)$
$= P((X_0, \cdots, X_n) \in A | X_n) P((X_n, X_{n+1}, \cdots) \in B | X_n)$.
In other words, a sequence is Markov if and only if the past and future are conditionally independent, given the present state.

Suppose that X_0 is independent of the Z_n's in (8.2) . Then,

$$\begin{aligned} P(X_{n+1} \in B | X_j : 0 \leq j \leq n) &= P(X_{n+1} \in B | X_n) \\ &= P(f_{n+1}(X_n, Z_{n+1}) \in B | X_n), \end{aligned}$$

and hence the solution X to (8.2) must be Markov, with $P_n(x, dy) = P(f_{n+1}(x, Z_{n+1}) \in dy)$ for $x, y \in S$.

Exercise 8.2:
Suppose that $S = \mathbb{R}$, and that $X = (X_n : n \geq 0)$ is Markov. Prove that X can be represented as the solution to the stochastic recursion $X_{n+1} = f_{n+1}(X_n, U_{n+1})$, where $U = (U_n : n \geq 1)$ is a sequence of iid uniform [0,1] rv's, and $f_{n+1}(x, y) = F_{n+1}^{-1}(x, y)$. Here, $F_{n+1}(x, y) = P(X_{n+1} \leq y | X_n = x)$ and $F_{n+1}^{-1}(x, y) = \sup\{\ z : F_{n+1}(x, z) \leq y\}$. (This problem establishes that Markov chains and solutions to stochastic recursions are one and the same.)

Exercise 8.3: (continuation of Problem 8.2)
Prove that if $S = \mathbb{R}$, then $X = (X_n : n \geq 0)$ is a Markov chain with stationary transition probabilities if and only if it can be represented in the form $X_{n+1} = f(X_n, U_{n+1})$, where $f : \mathbb{R} \times [0, 1] \to \mathbb{R}$ and $(U_n : n \geq 1)$ is a sequence of iid uniform [0,1] rv's.

8.3 Examples of Markov Chains

Example 8.1: (simple random walk). A Markov chain $X = (X_n : n \geq 0)$ is said to be a *simple random walk* on $\mathbb{Z}$ if

$$P(i, i+1) = p, i \in \mathbb{Z}\ P(i, i-1) = q, i \in \mathbb{Z}$$

with $p + q = 1 (0 < p < 1)$. Graphically, we have:

Note that $X_n = X_0 + \beta_1 + \cdots + \beta_n$, where $P(\beta_i = +1) = p = 1 - P(\beta_i = -1)$, and the β_i's are iid.

Example 8.2: (a queuing example)
Let Z_n be the number of packets to arrive to a buffer in slot n. Suppose that $(Z_n : n \geq 1)$ is iid with

$$P(Z_n = 2) = p,\ P(Z_n = 1) = r,\ P(Z_n = 0) = q,$$

where p + q + r = 1. Assume that the server serves one packet per slot of time, assuming there is a packet to be served. Here, the transition graph for X_n = number of packets in the system at time n is

Example 8.3: (birth-death chains)
Here, the transition graph takes the form

Here, X_n can be interpreted as the number of members of a population at time n. (Increasing by 1 = birth, decreasing by 1 = death).

Example 8.4: (storage in reservoir).
Suppose that S_n = storage in reservoir at beginning of period n. Then, $S_{n+1} = S_n + Z_{n+1} - O_{n+1}$, where Z_{n+1} is the inflow in period n + 1, and O_{n+1} is the outflow in period n + 1. Assume that the outflow follows the power law $O_{n+1} = aS^b_{n+1}$ for $a, b > 0$. Hence, $S_{n+1} + aS^b_{n+1} = S_n + Z_{n+1}$ for $n \geq 0$. This defines a Markov chain on $S = [0, \infty)$, provided that $(Z_n : n \geq 1)$ is iid.

Example 8.5: (discretized stochastic differential equations)
Given a time increment h, consider the sequence $(X_n : n \geq 0)$ defined via the recursion

$$X_{n+1} - X_n = \mu(X_n)h + \sigma(X_n)[B((n+1)h) - B(nh)],$$

where $B = (B(t) : t \geq 0)$ is standard Brownian motion. Because B has stationary and independent increments, the Z_n's are iid, where $Z_{n+1} = B((n+1)h) - B(nh)$. Hence, $(X_n : n \geq 0)$ defines a Markov chain on $S = I\!R$. Such stochastic recursions arise when stochastic differential equations are discretized.

Remark Note that these Markov chains can all be easily simulated, provided that efficient algorithms for generating iid copies of Z_1 exist (and given an appropriate initial condition for X).

8.4 Computing the Distribution of the Markov Chain at Time n

For a typical discrete state space stochastic sequence $X = (X_m : m \geq 0)$, the computation of $P(X_n = y)$ takes the form

$$P(X_n = y) = \sum_{x_0, x_1, \cdots, x_{n-1}} P(X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}, X_n = y), \tag{8.4}$$

and hence (exact) computation of $P(X_n = y)$ requires summing over all $|S|^n$ paths of length n. One of the key features of Markov chain theory is that such computations simplify enormously in the presence of Markov structure.

In particular, suppose that the Markov chain $X = (X_n : n \geq 0)$ has stationary transition probabilities and discrete state space. Then,

$$P(X_n = y) = \sum_{x_0, x_1, \cdots, x_{n-1}} \mu(x_0) P(x_0, x_1) \cdots P(x_{n-1}, y), \tag{8.5}$$

where $\mu = (\mu(x) : x \in S)$ is the initial distribution of X. Note that if $\mu_n = (\mu_n(y) : y \in S)$ (with $\mu_n(y) = P(X_n = y)$) is written as a row vector, the above representation establishes that $\mu_n(y) = (\mu P^n)(y)$, i.e. $\mu_n = \mu P^n$.

Because μ_n can be computed recursively via vector-matrix multiplications, μ_n can be computed in $O(n|S|^2)$ operations. The fact that μ_n can be computed through linear algebraic methods dramatically reduces the amount of computation needed in the Markov setting.

Convention When X is a discrete state space Markov chain, we will always encode all the probability distributions on S as row vectors.
When $X = (X_n : n \geq 0)$ has a continuous state space, the analog to (8.5) is

$$P(X_n \in B) = \int_S \mu(dx_0) \int_s P(x_0, dx_1) \cdots \int_S P(x_{n-1}, B)$$

8.5 Computing the Conditional Expectation of $f(X_n)$

Again, we suppose that the Markov chain has stationary transition probabilities, so that the transition kernels / matrices satisfy $P_n = P$ for $n \geq 1$. Our goal here is to compute $r_n(x) = E_x f(X_n)$ for a given "reward" function f. (e.g. f would be an indicator function if we wish to compute a probability involving X_n)

Convention When X is a discrete state space Markov chain, we will always encode all functions $f : S \to \mathbb{R}$ as column vectors.
To ensure that $r_n(x)$ is well-defined, we suppose that f is non-negative. If S is discrete, we may then compute $r_n(x)$ as

$$r_n(x) = \sum_{x_1,\cdots,x_n} P(x,x_1)P(x_1,x_2)\cdots P(x_{n-1},x_n)f(x_n) = (P^n f)(x), \tag{8.6}$$

so that $r_n = P^n f$.

Note that r_n can be recursively computed via the recursion $r_n = Pr_{n-1}$ subject to $r_0 = f$.
When $X = (X_n : n \geq 0)$ has a continuous state space, the analog to to (8.6) is

$$r_n(x) = \int_S P(x,dx_1) \int_s P(x_1,dx_2) \cdots \int_S P(x_{n-1},dx_n)f(x_n)$$

8.6 First Transition Analysis

An enormous variety of expectations can be computed as solutions to systems of linear equalities (in discrete state space) or as solutions to linear integral equations (in continuous state space). To illustrate this principle, suppose that we wish to compute $u^*(x) = E_x \sum_{j=0}^{T-1} f(X_j)$, where $T = \inf\ \{n \geq 0 : X_n \in C^C\}$ is the so-called "first hitting time" of C^C. If one views $f(x)$ as the "reward" obtained by spending one unit of time in x, then $\sum_{j=0}^{T-1} f(X_j)$ is the total reward cumulated up to the hitting time T, and $u^*(x)$ is the expectation of this rv, conditional on $X_0 = x$. Such computations are of relevance in computing the total cash flow generated by an insurance customer up to the random time at which the customer terminates his or her policy (as just one example).
Assume that $X = (X_n : n \geq 0)$ has stationary transition probabilities and discrete state space. As usual, to guarantee that the expectation is well-defined, suppose f is non-negative. The principle of "first transition analysis" asserts that we should condition on the state X_1 at the first transition:

$$\begin{aligned}
u^*(x) &= E_x \sum_{j=0}^{T-1} f(X_j) \\
&= \sum_{y\in S} E_x[\sum_{j=0}^{T-1} f(X_j)|X_i = y]P(x,y) \\
&= \sum_{y\in S} E_x[f(x) + \sum_{j=1}^{T-1} f(X_j)|X_i = y]P(x,y) \\
&= f(x) + \sum_{y\in C} P(x,y)E_x(\sum_{j=1}^{T-1} f(X_j)|X_i = y) \\
&= f(x) + \sum_{y\in C} P(x,y)u^*(y),
\end{aligned}$$

In other words, u^* should satisfy the linear system

$$u = f + Bu, \tag{8.7}$$

where $B = (P(x, y) : x, y \in C)$. Note that B is the principal submatrix of P corresponding to "C to C" transitions.
In an ideal world, u^* would be the unique solution to the linear system (8.7) . Alas, things are more complicated.

Example 8.6: Let $X = (X_n : n \geq 0)$ be our queuing model, with transition diagram

Suppose $C^c = \{0\}$, so that T is the first time the router is idle. First transition analysis shows that $u^*(x) = E_x T$ should satisfy

$$1 + pu(x + 1) + qu(x - 1) = u(x) \tag{8.8}$$

for $x \geq 1$, subject to $u(0) = 0$. Intuitively, we expect that $u^*(x) = \infty$ whenever $p > \frac{1}{2}$. When $q > \frac{1}{2}$, we expect u(x) to be finite valued. (For $p = \frac{1}{2}$, it turns out that $u^*(x) = \infty$.) Note that $(u^*(x) : x \geq 0)$ satisfies the second-order linear difference equation (8.8) . As for second-order linear differential equations, such equations require two boundary conditions to uniquely determine the solution. But we have only a single boundary condition, namely $u(0) = 0$. So, (8.8) does not uniquely determine $(u^*(x) : x \geq 0)$.

So, how do we determine the probabilistically meaningful solution when solving the linear system (8.7) ?

Theorem 8.1. *Suppose $f \geq 0$. The function $u^* = (u^*(x) : x \in C)$ has the following properties*

- *i.)* $u^* = \sum_{n=0}^{\infty} B^n f$
- *ii.)* u^* *is the minimal non-negative solution to* $u = f + Bu$.

Proof. Note that for $x \in C$,

$$\begin{aligned} u^*(x) &= E_x \sum_{n=0}^{T-1} f(X_n) \\ &= E_x \sum_{n=0}^{\infty} f(X_n) I(T > n) \\ &= \sum_{n=0}^{\infty} E_x f(X_n) I(T > n) (\text{use Fubini and the fact that} f \geq 0) \end{aligned}$$

But,

$$\begin{aligned} E_x f(X_n) &= \sum_{x_i \in C, x_1, \cdots, x_n} P(x, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n) f(x_n) \\ &= (B^n f)(x). \end{aligned}$$

Hence, $u^* = \sum_{n=0}^{\infty} B^n f$.
As for *ii.)*, observe that any solution u to $u = f + Bu$ satisfies $u = f + Bf + \cdots + B^n f + B^{n+1} u$.
If u is non-negative, $B^{n+1} u \geq 0$, so $u \geq f + Bf + \cdots + B^n f$.
But n is arbitrary, so we can let $n \to \infty$, yielding the inequality $u \geq u^*$; this proves *ii.)*. □

Exercise 8.4:
Use Theorem 8.1 to compute u^* for Example 8.6.

The above theory easily extends to continuous state space, with the obvious replacement of sums by integrals. To be specific, suppose that $g : C \to I\!R$ and define $B^n g$ via $B^0 g = g$ and $(B^{n+1})g(x) = \int_c P(x, dy)(B^n g)(y)$ for $x \in C$. If f is non-negative, then $u^*(x) = E_x \sum_0^{T-1} f(X_n)$ satisfies $u^* = \sum_{n=0}^{\infty} B^n f$ and is the minimal non-negative solution of the linear integral equation $u(x) = f(x) + \int_C P(x, dy)u(y)$ for $x \in C$.

Exercise 8.5:
In many economics and control settings, one wishes to compute an expected infinite horizon discounted reward of the form $u^+(x) = E_x \sum_{n=0}^{\infty} e^{-\alpha n} f(X_n)(\alpha > 0)$.
Suppose that $\|f\| \triangleq \sup \ \{|f(x)| : x \in S\} < \infty$. Prove that u^* is the unique bounded solution of the linear system $u = f + e^{-\alpha} P u$.

Exercise 8.6:
Suppose $T = \inf\{n \geq 0 : X_n \in C^c\}$, and let $u^*(x) = E_x \sum_{n=0}^{T-1} e^{-\alpha n} f(X_n)$ for $x \in C$.

1. If $f \geq 0$, prove that u^* is the minimal non-negative solution of $u = f + e^{-\alpha} B u$
2. If $\|f\| < \infty$ prove that u^* is the unique bounded solution of $u = f + e^{-\alpha} B u$
3. If $|C| < \infty$, prove that $u^* = (I - e^{-\alpha} B)^{-1} f$

For extensions of the ideas developed in this lecture to Markov chains with non-stationary transition probabilities, please read the handout on Markov chains that is posted on the class website.

8.7 More on "First Transition Analysis"

Consider the expectation

$$u^*(x) = \mathrm{E}_x \left[\sum_{j=0}^{T-1} f(X_j) \right],$$

where $T = \inf\{n \geq 0 : X_n \in C^C\}$ is the "first hitting time" of C^C and $f : C \to I\!R$ is a given reward function. When F is non-negative, we have already shown that u^* satisfies

$$u = f + Bu,$$

where $B = (P(x, y) : x, y \in C)$ is the restriction of P to C. In fact,

$$u^* = \sum_{n=0}^{\infty} B^n f. \tag{8.9}$$

Note that the infinite sum in (8.9) may diverge, so that $u^*(x) = \infty$. It is clearly of interest to know when u^* is finite-valued. In fact, we shall now show how to obtain computable bounds on u^*. This involves the use of "weighted vector/matrix norms". For $g : C \to I\!R$, define the norm of the vector/function g to be

$$\|g\| = \sup_{x \in C} \frac{|g(x)|}{w(x)};$$

where $w(\cdot)$ is a positive "weighting function". Now, define the norm of the matrix/kernel K to be

$$\|K\| = \sup_{\|g\| = 1} \|Kg\|.$$

These norms have the following properties (see the handout on the class webpage for additional details):

i.) $\|g_1 + \alpha g_2\| \le \|g_1\| + |\alpha|\|g_2\|$

ii.) $\|K_1 + \alpha K_2\| \le \|K_1\| + |\alpha|\|K_2\|$

iii.) $\|K_1 K_2\| \le \|K_1\|\|K_2\|$

It follows that

$$\begin{aligned} \|u^*\| &\le \sum_{n=0}^{\infty} \|B\|^n \|f\| \\ &= (1 - \|B\|)^{-1}\|f\| \end{aligned} \tag{8.10}$$

if $\|B\| < 1$. Now, $\|B\| < 1$ if and only if

$$\sup_x \int_C P(x, dy)\frac{w(y)}{w(x)} < 1$$

i.e. $\|B\| < 1$ if and only if there exists $r < 1$ such that

$$\int_C P(x, dy) w(y) \le r w(x) \tag{8.11}$$

for $x \in C$ (and r can be taken to be $\|B\|$). Relation (8.11) is equivalent to demanding that

$$\mathrm{E}_x \left[w(X_1) I(X_1 \in C)\right] \le r w(x) \tag{8.12}$$

for $x \in C$. If $w : s \to I\!R$ is non-negative, a sufficient condition for (8.12) is

$$\mathrm{E}_x \left[w(X_1)\right] \le r w(x) \tag{8.13}$$

for $x \in C$. In the presence of (8.13), (8.10) permits us to conclude that

$$\sup_{x \in C} \frac{|u^*(x)|}{w(x)} \le (1 - r)^{-1} \sup_{y \in C} \frac{|f(y)|}{w(y)}. \tag{8.14}$$

When f is positive, one possible choice for w is f, in which case (8.14) becomes

$$|u^*(x)| \le (1 - r)^{-1} |f(x)|$$

for $x \in C$. We summarize our discussion with the following theorem.

Theorem 8.2. *Suppose that f is positive and there exists an $r < 1$ such that*

$$\mathrm{E}_x \left[f(X_1)\right] \le r f(x)$$

for $x \in C$. Then,

$$|u^*(x)| \le (1 - r)^{-1} f(x)$$

for $x \in C$.

Example 8.7: Let $X = (X_n : n \ge 0)$ be the queueing example introduced in Lecture 8 with $p + q = 1$ and $p < 1/2$. Put $T = \inf\{n \ge 0 : X_n = 0\}$ and $f(x) = \exp(\gamma x)$. Then for $x \ge 1$,

$$\begin{aligned} \mathrm{E}_x \left[f(X_1)\right] &= p e^{\gamma(x+1)} + q e^{\gamma x - 1} \\ &= e^{\gamma x}(p e^{\gamma} + q e^{-\gamma}) \end{aligned}$$

So long as $e^{\gamma} < q/p$, $p e^{\gamma} + q e^{-\gamma} < 1$ and hence

$$\mathrm{E}_x \left[f(X_1)\right] \le r f(x)$$

for $r = p e^{\gamma} + q e^{-\gamma} < 1$ and $x \ge 1$. We conclude that if $e^{\gamma} < q/p$, then

$$\mathrm{E}_x \left[\sum_{j=0}^{T-1} e^{\gamma X_j}\right] \le (1 - p e^{\gamma} - q e^{-\gamma})^{-1} e^{\gamma x}$$

for $x \ge 1$.

8.8 Further Examples of First Transition Analysis

Consider a birth-death Markov chain on $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$, with $P(x, x+1) = p_x$, $P(x, x-1) = q_x$, and $P(x, x) = r_x$. In many applications, it is of interest to compute the "exit probabilities"

$$\mathrm{P}_x \{(T_b < T_a)\} \tag{8.15}$$

where $T_b = \inf\{n \geq 0 : X_n = b\}$ and $T_a = \inf\{n \geq 0 : X_n = a\}$ (with $a < x < b$). Hence (8.15) represents the probability that the process exits through the right boundary at b prior to exiting through the left boundary at a. One important example of such a calculation arises in the "gambler's ruin" chain. Here, X_n represents the fortune of a gambler at time n. The model leads to $p_x = p$ and $q_x = q$ for $x \geq 1$, where p is the probability of winning a one dollar bet. The left boundary $a = 0$ (where the gambler is "ruined"), whereas the right boundary b represents the level at which the gambler has decided to exit the game and leave with his/her winnings. Put $u^*(x) = \mathrm{P}_x \{T_b < T_a\}$. First transition analysis leads to the linear system

$$u(x) = p_x u(x+1) + r_x u(x) + q_x u(x-1) \tag{8.16}$$

for $a < x < b$, subject to $u(a) = 0$, $u(b) = 1$. It is easily seen that the solution to (8.16) is given by

$$\mathrm{P}_x \{T_b < T_a\} = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y},$$

where

$$\gamma_y = \frac{q_1 q_2 \cdots q_y}{p_1 p_2 \cdots p_y}.$$

This is a special case of what is known as an "exit distribution" calculation (sometimes known as an "absorption probability" calculation). Let $T = \inf\{n \geq 0 : X_n \in C^C\}$ be the "exit time" from C to C^C, and put

$$u^*(x) = \mathrm{P}_x \{X_T \in A\}$$

for $x \in C$. Then, $u^* = (u^*(x) : x \in C)$ satisfies

$$u^*(x) = \mathrm{P}_x \{X_1 \in A\} + \int_C P(x, dy) u(y)$$

and is the minimal non-negative such solution. The above birth-death calculation is the special case in which $C^C = \{a, b\}$.

8.9 The Concepts of Steady-State and Equilibrium for a Stochastic System

For a deterministic dynamical system $(x(t) : t \geq 0)$, a steady-state is said to exist if:

> There exists an x^* such that for each initial condition $x(0) = x$, $x(t) \to x^*$ as $t \to \infty$.

This notion of steady-state is very demanding in the stochastic setting. It is rarely the case that $X(t) \to X^*$ a.s. as $t \to \infty$. (Think of a two-state Markov chain as a counter-example.) A better generalization to the stochastic environment is to demand that the distribution of $X(t)$ converges. Here is one possible generalization to the stochastic environment:

> There exists $X(\infty)$ such that for each initial distribution for $X(0)$, the distribution of $X(t)$ converges to that of $X(\infty)$.

In what sense shall we demand that the distribution of $X(t)$ converges to that of $X(\infty)$?

Definition 8.1: A sequence $(Y_n : n \geq 0)$ of S-valued random variables is said to *converge in total variation* to X_∞ if

$$\sup_{A \subset S} |\,\mathrm{P}\{Y_n \in A\} - \mathrm{P}\{Y_\infty \in A\}\,| \to 0$$

as $n \to \infty$ (written $Y_n \xrightarrow{tv} Y_\infty$ as $n \to \infty$).

Remark: Total variation convergence implies weak convergence. If S is discrete, weak convergence and total variation convergence are identical concepts.

Definition 8.2: An S-valued time-homogeneous Markov chain $X = (X_n : n \geq 0)$ is said to *converge to a steady-state* if there exists an S-valued random variable X_∞ such that for each initial distribution μ,

$$X_n \xrightarrow{tv} X_\infty$$

as $n \to \infty$.

In many settings, one can interpret the steady-state X_∞ as describing the "time-average" behavior of the sequence $(X_n : n \geq 0)$:

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\{X_j \in A\} \to \mathrm{P}\{X_\infty \in A\} \ \text{ a.s.}$$

as $n \to \infty$. In other words, the steady-state probability $\mathrm{P}\{X_\infty \in A\}$ can be interpreted as the long-run fraction of time that $X = (X_n : n \geq 0)$ spends in the subset A.

Definition 8.3: An S-valued time-homogeneous Markov chain $X = (X_n : n \geq 0)$ is said to be *ergodic* if there exists an S-valued random variable X_∞ so that for each initial distribution μ,

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\{X_j \in A\} \to \mathrm{P}\{X_\infty \in A\} \text{ a.s.}$$

as $n \to \infty$.

In the Markov chain setting, we shall see that Definition 8.2 typically implies Definition 8.3 (but not necessarily vice versa). So, ergodicity is the weaker concept. We turn next to the concept of an equilibrium, When a deterministic dynamical system $(x(t) : t \geq 0)$ is governed by a differential equation of the form

$$\dot{x}(t) = g(x(t)),$$

the limiting value x^* will typically satisfy $g(x^*) = 0$. In that case, if we set $x(0) = x^*$, then $x(t) \equiv x^*$ for $t \geq 0$. In other words, x^* is an equilibruim for the dynamical system. In the stochastic setting, an equilibrium distribution is one for which initialization with distribution $\pi(\cdot)$ at $t = 0$ means that $X(t)$ has distribution $\pi(\cdot)$ for each $t \geq 0$. In the Markov chain setting, this means that if $\mathrm{P}\{X_0 \in \cdot\} = \pi(\cdot)$, then

$$\mathrm{P}\{X_1 \in \cdot\} = \pi(\cdot).$$

But conditioning on X_0 shows that such an equilibrium distribution π must satisfy

$$\int_S \pi(dx)\,\mathrm{P}\{x, \cdot\} = \pi(\cdot) \tag{8.17}$$

The integral equation (8.17) (in discrete state space, the linear system (8.17)) is the stochastic analog to $g(x^*) = 0$.

Definition 8.4: Let $X = (X_n : n \geq 0)$ be an S-valued time-homogeneous Markov chain. The distribution π is said to be a *stationary distribution* (often referred to as a steady-state distribution) of X if

$$\int_S \pi(dx)\,\mathrm{P}\{x, A\} = \pi(A)$$

for each (measurable) $A \subset S$.

Why is such a distribution π called a stationary distribution of X? The next exercise provides the answer.

Exercise 8.7: Prove that if π is a stationary distribution of X and if $\mathrm{P}\{X_0 \in \cdot\} = \pi(\cdot)$, then $X = (X_n : n \geq 0)$ is a stationary sequence. In other words, prove that for each $n \geq 0$,

$$(X_n, X_{n+1}, \ldots) \stackrel{\mathcal{D}}{=} (X_0, X_1, \ldots).$$

8.10 Stationary Distributions for Finite State Markov Chains

We assume here that $|S| < \infty$.

Theorem 8.3. *Let $X = (X_n : n \geq 0)$ be a finite-state Markov chain. Then, X possesses at least one stationary distribution π. In other words, there exists at least one probability mass function $\pi = (\pi(x) : x \in S)$ for which*

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y) \tag{8.18}$$

for $y \in S$.

Remark: The linear system (8.18) can be re-written in vector-matrix form as $\pi = \pi P$. (Recall that all probability distributions are always encoded as row vectors.)

Proof. Let $K_n = n^{-1} \sum_{j=0}^{n-1} p^j$. Then $(K_n : n \geq 0)$ is a sequence of $|S| \times |S|$ stochastic matrices with all entries bounded between 0 and 1. It follows that one can extract a subsequence n' for which

$$K_{n'} \to K_\infty$$

as $n' \to \infty$, where K_∞ is a stochastic matrix. Note that

$$\begin{aligned} K_{n'} p &= \frac{1}{n'} \sum_{j=1}^{n'} P^j \\ &= K_{n'} - \frac{1}{n'} (\mathrm{I}\{-\}\, P^{n'}) \\ &\to K_\infty \end{aligned}$$

as $n' \to \infty$. But $K_{n'} \to K_\infty p$ as $n' \to \infty$. So,

$$K_\infty = K_\infty p.$$

Each row of K_∞ is then a stationary distribution of X. □

Example 8.8: Let $X = (X_n : n \geq 0)$ be a two-state Markov chain with transition graph

with $0 < a, b < 1$. Here,

$$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b} \right).$$

Example 8.9: Let $X = (X_n : n \geq 0)$ be a Markov chain that describes the movement of molecules across a permeable barrier. Suppose m is the total number of molecules, with X_n representing the number of molecules on one side of the barrier (so that $m - X_n$ are in the other side). At each time n, we randomly choose one of the m molecules and move that molecule to the other side of the barrier. This leads to a Markov chain with birth-death structure given by

$$\begin{aligned} \mathrm{P}\{x, x+1\} &= (m - x)/m \\ \mathrm{P}\{x, x-1\} &= x/m \end{aligned}$$

for $0 \leq x \leq m$. The stationary distribution here is

$$\pi(x) = \begin{pmatrix} m \\ x \end{pmatrix} 2^{-m}$$

for $0 \leq x \leq m$. In other words, X_∞ is binomial$(m, 1/2)$ Note that the central limit theorem asserts that

$$X_\infty \overset{\mathcal{D}}{\approx} \frac{m}{2} + \frac{\sqrt{m}}{2} N(0,1)$$

for m large. The stochastic fluctuations of order $\sqrt{m}$ are tiny as compared to the mean (of order m) when m is large. Furthermore, the fluctuations are approximately Gaussian. So, this model (due to the physicist Ehrenfest) predicts behavior consistent with statistical thermodynamics.

Example 8.10: Returning to the two state example, suppose that the transition graph is

In this case, both states are "absorbing states". Here, the steady-state behavior clearly depends on the initial state chosen. This is reflected in the fact that this Markov chain has multiple stationary distributions. In particular, any stochastic vector (p, q), (with $p + q = 1$) is a stationary distribution for this example. It is of obvious interest to know when the stationary distribution is unique.

Definition 8.5: A Markov chain $X = (X_n : n \geq 0)$ in discrete state space S (and having stationary transition probabilities) is said to be *irreducible* if for each $x \in S$ and $y \in S$, there exists $n = n(x, y)$ such that $P^n(x, y) > 0$.

Remark: A discrete state space Markov chain is "irreducible" if the directed transition graph corresponding to the chain is connected (i.e. there exists a directed path from each x to each y). We state without proof the following intuitively reasonable result.

Theorem 8.4. *Let $X = (X_n : n \geq 0)$ be a discrete state space Markov chain (with either finitely or infinitely many states) with stationary transition probabilities. If X is irreducible, then X has at most one stationary distribution.*

Remark: If X is irreducible with finite state space, Theorems 8.3 and 8.4 guarantee existence of a unique stationary distribution. If X is irreducible with infinite state space, the chain either has a unique stationary distribution or none at all.

8.11 New Phenomena in the Presence of Infinite State Space

We have shown (see Theorem 9.2) that every finite state Markov chain possesses at least one stationary distribution. In the presence of infinite state space, there need not exist *any* stationary distribution. This (perhaps) is not surprising in light of the fact that existence of a stationary distribution is closely connected to the concept of stochastic stability; instability seems intuitively easier to generate in infinite state systems.

Example 8.11: Let $X = (X_n : n \geq 0)$ be the Markov chain on $\mathbb{Z}^+$ with transition graph

Note that the stationarity equations translate to $\pi(0) = 0$ and

$$\pi(x+1) \;=\; \pi(x)$$

for $x \geq 0$. Hence, $\pi(x) = 0$ for $x \geq 0$ is the only solution of $\pi = \pi P$. So there is no stationary distribution for this chain.

On the other hand, this chain evolves so that $X_n = X_0 + n$. This model is describing an unstable system for which $X_n \to \infty$ *a.s.* as $n \to \infty$.

This shows that infinite state Markov chains need not possess stationary distributions. As seen in Example 10.1, infinite chains can exhibit a wider range of behaviors than is the case for finite state systems. As one might expect, this is closely related to the fact that infinite state matrices can exhibit more complex behaviors than do finite matrices.

Example 8.12: An infinite stochastic matrix $P = (P(x, y) : x, y \in S)$ can satisfy $P^n \to 0$ as $n \to \infty$ (in the sense that $P^n(x, y) \to 0$ as $n \to \infty$, for each $x, y \in S$). In fact, the transition matrix of Example 10.1 has this characteristic. This occurs despite the fact that $P^n e = e$ for each $n \geq 1$, where $e = (1, 1, \ldots)^T$. (Here is another example where interchanging limits and infinite sums fails to hold!) Note that the subsequence argument followed in the proof of Theorem 9.2 fails for this example.

Example 8.13: For any finite matrix A, an eigenvalue λ must have both a corresponding row and column eigenvector. But this can fail for infinite matrices. Consider, for example, the transition matrix P of the chain introduced in Example 10.1. Clearly $Pe = e$ so e is a column eigenvector associated with eigenvalue 1. Our discussion of Example 10.1 actually proved that no row eigenvector corresponding to eigenvalue 1 exists.

The point here is that infinite state Markov chains need special care...like a wee baby.

8.12 Regenerative Structure of Markov Chains

A key idea in the study of stochastic stability for Markov chains is the concept of "regeneration." Consider the queueing chain introduced earlier in which 0, 1, or 2 packets arrive per slot of time with probabilities q, r, and p respectively. A simulation of this model might produce the realization below:

Note that every time the chain returns to state 0 (at times T_0, T_1, ...), the system "regenerates." In other words, the chain starts probabilistically afresh each time it returns to state 0. More precisely, the simulation algorithm's output subsequent to T_i is algorithmically identical to that subsequent to T_0. This ensures that

$$(X_{T_i}, X_{T_{i+1}}, \ldots) \stackrel{\mathcal{D}}{=} (X_{T_0}, X_{T_{0+1}}, \ldots)$$

for each $i \geq 1$. Furthermore, the (uniform) random numbers used to drive the simulation subsequent to T_i are independent of those used prior to T_i so that

$$(X_0, X_1, \ldots, X_{T_i - 1}) \text{ is independent of } (X_{T_i}, X_{T_i+1}, \cdots)$$

for $i \geq 1$. These two observations together imply that the "0-cycles" defined by

$$(X_{T_{i-1}}, \ldots, X_{T_i - 1})$$

are iid. Hence, the simulation of the Markov chain $X = (X_n : n \geq 0)$which describes a highly correlated and dependent process can be split into iid cycles known as *regenerative cycles.* This beautiful idea allows one to apply theorems for iid sequences (like the law of large numbers and central limit theorem) to Markov chains.

8.13 Transience versus Recurrence - the final showdown

Given a Markov chain $X = (X_n : n \geq 0)$having stationary transition probabilities and living in on discrete state space S, the regenerative cycle approach can be applied to any fixed state $z \in S$. The state x is called the "return state" or "regenerative state." In order to guarantee that the entire sample path of the chain can be decomposed into an infinite sequence of iid z-cycles, we need to know that the Markov chain returns to z infinitely often.

Definition 8.6: The state $z \in S$ is said to be *recurrent* if $\mathrm{P}_z\{X \text{ visits } z \text{ infinitely often}\} = 1$. Otherwise, the state $z \in S$ is said to be *transient*.

Remark: Let $N(z) \triangleq \sum_{n=1}^{\infty} I(X_n = z)$ be the total number of visits to z. Then z is recurrent if and only if $\mathrm{P}_z\{N(z) = \infty\} = 1$; otherwise, z is transient.
The following result is intuitively obvious.

Proposition 8.1: The state $z \in S$ is recurrent if and only if $\mathrm{P}_z\{\tau(z) < \infty\} = 1$, where $\tau(z) = \inf\{n \geq 1 : X_n = z\}$ is the "first return time to z." The state $z \in S$ is transient if and only if $\mathrm{P}_z\{\tau(z) < \infty\} < 1$.

One way of computing $u^*(x) = \mathrm{P}_x\{\tau(z) < \infty\}$ is via "first transition analysis."

Proposition 8.2: The function $u^* = (u^*(x) : x \neq z)$ is the minimal non-negative solution of

$$u(x) = P(x,z) + \sum_{y \neq z} P(x,y)u(y), \quad x \neq z.$$

Example 8.14: We introduce here an important model known as a "branching chain," This Markov chain model is a basic model that is widely used in different biology subdisciplines. Think of X_n as representing, for example, the number of females in a population at generation n. Each of the X_n females independently reproduces; the ith female has ξ_{ni} female children. So,

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{ni}.$$

We are interested in computing $u^*(x) = \mathrm{P}_x\{\tau(0) < \infty\}$ (i.e. the probability of "extinction"). Here u^* satisfies

$$u(x) = \mathrm{P}\{\xi_{n1} = 0\}^x + \mathrm{E}\left[u(\sum_{i=1}^{x} \xi_{ni}) I(\sum_{i=1}^{x} \xi_{ni} \geq 1)\right] \tag{8.19}$$

for $x \geq 1$. It seems intuitively reasonable to suppose that $u^*(x) = \gamma^x$ for some $\gamma \in [0,1)$, since extinction of the x family lines present at time 0 should occur independently and with the same distribution. Noting that $\gamma^0 = 1$, we can re-write (1) as requiring that

$$\gamma^x = \mathrm{E}\left[\gamma^{\sum_{i=1}^{x} \xi_{ni}}\right]. \tag{8.20}$$

Since the ξ_{ni}'s are iid, (2) forces us to choose γ so that

$$\gamma^x = (\mathrm{E}\left[\gamma^{\xi_{11}}\right]^x. \tag{8.21}$$

Suppose that $\mathrm{P}\{\xi_{ni} = 1\} < 1$ (so that we are not dealing with the trivial situation in which each member of the nth population is identically replaced by a single member in the $(n+1)$st, so that $X_n = X_0$ for $n \geq 1$). Then simple real variable arguments show that there exists a root $\gamma \in [0,1)$ precisely when $\mathrm{E}[\xi]_{n1} > 1$ (i.e. the number of "replacement females" is, on average, greater than one). If, on the other hand, $\mathrm{E}[\xi_{n1}] \leq 1$ (i.e. the number of "replacement females" is, on average, less than or equal to one), there is no $\gamma \in [0,1)$ satisfying (3), and the minimal non-negative solution to (1) is $u(x) = 1$ for $x \geq 1$ (i.e. extinction is certain!).

This model also plays a key role in neutron scattering calculations in computing the probability that a "chain reaction" will terminate.
For irreducible Markov chains one need not check recurrence/transience individually for each state.

Proposition 8.3: Let $X = (X_n : n \geq 0)$be an irreducible Markov chain on discrete state space S. Then either all states are recurrent or all states are transient.

Proposition 8.3 establishes that recurrence and transience are "class properties." If one state is recurrent (transient), then all states are recurrent(transient). See, for example, p. 17-25 of *Introduction to Stochastic Processes* by P.G. Hoel, S.C. Emerson, and C.J. Stone, Houghton-Mifflin Company (1972) for details.
One approach to verifying recurrence/transience is to appeal to Proposition 8.2. We now describe another means of dealing with this issue.
We first note that

$$\mathrm{P}_y\{N(y)=k\} = \mathrm{P}_y\{\tau(y)<\infty\}^k\, \mathrm{P}_y\{\tau(y)=\infty\}$$

This implies that $N(y)$ is a geometric random variable (conditional on $X_0 = y$) with mean

$$\mathrm{E}_y[N(y)] = \frac{\mathrm{P}_y\{\tau(y)<\infty\}}{\mathrm{P}_y\{\tau(y)=\infty\}}. \tag{8.22}$$

More generally, if $x \neq y$, $k \geq 1$

$$\mathrm{P}_x\{N(y)=k\} = \mathrm{P}_x\{\tau(y)<\infty\}\,\mathrm{P}_y\{\tau(y)<\infty\}^{k-1}\,\mathrm{P}_y\{\tau(y)=\infty\} \tag{8.23}$$

whereas $\mathrm{P}_x\{N(y)=0\} = \mathrm{P}_x\{\tau(y)=\infty\}$. Relations (4) and (5) together imply that $\mathrm{E}_x[N(y)] < \infty$ if and only if $\mathrm{P}_y\{\tau(y)<\infty\} < 1$ (i.e. y is transient). But

$$\begin{aligned}
\mathrm{E}_x[N(y)] &= \mathrm{E}_x\left[\sum_{n=0}^{\infty} I(X_n = y)\right] \\
&= \sum_{n=0}^{\infty} \mathrm{E}_x[\mathrm{I}\{X_n = y\}] \quad \text{by Fubini's Theorem} \\
&= \sum_{n=0}^{\infty} \mathrm{P}_x\{X_n = y\} \\
&= \sum_{n=0}^{\infty} P^n(x,y)
\end{aligned}$$

The following theorem summarizes the above discussion.

Theorem 8.5. *Let $X = (X_n : n \geq 0)$be an irreducible Markov chain on discrete state space S. Then*

(i) all states are transient if and only if

$$\sum_{n=0}^{\infty} \mathrm{I}\{X_n = y\} < \infty \quad a.s.$$

for some state y if and only if

$$\sum_{n=0}^{\infty} P^n(x,y) < \infty$$

for some $(x,y) \in S \times S$.

(ii) all states are recurrent if and only if

$$\sum_{n=0}^{\infty} \mathrm{I}\{X_n = y\} = \infty \quad a.s.$$

for some state y if and only if

$$\sum_{n=0}^{\infty} P^n(x,y) = \infty$$

for some $(x,y) \in S \times S$.

We now illustrate this transience/recurrence criterion by applying it to a simple symmetric random walk.

Example 8.15: Consider the Markov chain $X = (X_n : n \geq 0)$on $\mathbb{Z}$ having transition probabilities $\mathrm{P}\{x, x+1\} = 1/2 = \mathrm{P}\{x, x-1\}$. We will show that 0 is recurrent.
Note that $P^m(0,0) = 0$ for m odd. On the other hand,

$$P^{2m}(0,0) = \binom{2m}{m} 2^{-2m}. \tag{8.24}$$

Stirling's formula states that

$$m! \sim \sqrt{2\pi m}\left(\frac{m}{e}\right)^m$$

as $m \to \infty$. Substituting this asymptotic relationship in (6) shows that

$$P^{2m}(0,0) \sim \frac{c}{\sqrt{m}}$$

as $m \to \infty$, where $0 < c < \infty$. Since $\sum_{m=0}^{\infty} m^{-1/2} = \infty$, it follows that a simple symmetric random walk on $\mathbb{Z}$ is recurrent.

Example 8.16: We now consider a simple symmetric random walk on $\mathbb{Z}^d$ for $d \geq 2$. To be precise, let

$$\vec{X}_n = (X_n^1, X_n^2, \ldots, X_n^d)$$

where each of the coordinate processes $(X_n^i : n \geq 0)$ is an independent simple symmetric random walk on $\mathbb{Z}$. Note that $(\vec{X}_{2n} : n \geq 0)$ is an irreducible Markov chain on $X = \{(k_1, k_2, \ldots, k_d) : k_i \text{ is even for } 1 \leq i \leq d\}$. We now test the recurrence/transience of state $\vec{0}$.
Note that

$$\begin{aligned} P^{2n}(\vec{0},\vec{0}) &= \prod_{i=1}^{d} \mathrm{P}\left\{X_{2n}^i = 0 | X_0^i = 0\right\} \\ &= \prod_{i=1}^{d} \binom{2n}{n} 2^{-2n} \\ &\sim \prod_{i=1}^{d} \left(\frac{c}{\sqrt{n}}\right) \\ &= \frac{c^d}{n^{d/2}} \end{aligned}$$

Since $\sum_{n=1}^{\infty} n^{-1} = \infty$, it follows that if $d = 2$, then $(X_{2n} : n \geq 0)$ is recurrent on S. On the other hand, for $d \geq 3$, $\sum_{n=1}^{\infty} n^{-d/2} < \infty$ so $(X_{2n} : n \geq 0)$ is then transient on S. In other words, a simple symmetric random walk is recurrent in dimensions 1 and 2 and transient in dimensions greater than 2. This is one of the most celebrated results of twentieth century probability! It implies that if you are lost at a probability conference contained on one floor of a building, simply begin a random walk and you will eventually find the room you seek. In fact, if you walk forever, you will find it infinitely many times.
Remark: Suppose we color each site that is visited by $(X_{2n} : n \geq 0)$ blue. Any site that is never visited is colored red. For one or two dimensional simple symmetric random walk, all sites are colored blue. For a random walk in dimension greater than two, some sites in S are colored blue and some are colored red. But if one looks at the "two-dimensional projections" (e.g. the set of sites (k_1, k_2) such that there exists at least one site of the form $(k_1, k_2, l_3, \ldots, l_d)$ with $l_i \in 2\mathbb{Z}$ that is blue), all the two-dimensional projections are blue. Thus, the additional "degree of freedom" represented by the third dimension renders the system transient.

8.14 Laws of Large Numbers for Recurrent Markov Chains

We now assume that $X = (X_n : n \geq 0)$is a recurrent irreducible Markov chain on discrete state space S. In this setting, we can choose any state $z \in S$ as our "return state" or "regeneration state." Let T_n be the nth time that X returns to z, and note that the recurrence implies that $T_n < \infty$ *a.s.* for $n \geq 0$. Furthermore, for $n \geq 1$,

$$(X_{T_{n-1}}, \ldots, X_{T_n - 1}) \tag{8.25}$$

are iid z-cycles. We wish to use this regenerative structure to study

$$n^{-1} \sum_{j=0}^{n-1} I(X_j = y), \tag{8.26}$$

the proportion of time that X spends in y over its first n steps.
Let $N(n)$ be the number of cycles completed in the first n steps, so that

$$N(n) = \max\{m : T_n \leq m\}.$$

Because we expect that averaging over $[0, n-1]$ is close to averaging over $[0, T_{N(n)})$,

$$n^{-1} \sum_{j=0}^{n-1} I(X_j = y) \approx T_{N(n)}^{-1} \sum_{j=0}^{T_{N(n)}-1} I(X_j = y).$$

To exploit the iid behavior of the z-cycles, we write

$$\sum_{j=0}^{T_{N(n)}-1} I(X_j = y) = \sum_{j=0}^{N(n)} Y_j,$$

$$T_{N(n)} = \sum_{j=0}^{N(n)} \tau_j,$$

where

$$Y_j = \sum_{i=T_{j-1}}^{T_j - 1} I(X_j = y)$$

$$\tau_j = T_j - T_{j-1}$$

Because of the iid structure of the z-cycles $(Y_j : j \geq 1)$ is iid and $(\tau_j : j \geq 1)$ is iid. Hence, the strong law of large numbers for iid sequences implies that

$$n^{-1} \sum_{j=0}^{n} Y_j \to \mathrm{E}\left[Y\right]_1 \quad a.s.,$$

$$n^{-1} \sum_{j=0}^{n} \tau_j \to \mathrm{E}\left[\tau\right]_1 \quad a.s.$$

as $n \to \infty$. So, this suggests that

$$\begin{aligned} n^{-1} \sum_{j=0}^{n-1} I(X_j = y) \quad &\approx \quad \sum_{j=0}^{N} (n) Y_j \ / \ \sum_{j=0}^{N} (n) \tau_j \\ &= \quad \frac{\sum_{j=0}^{N} (n) Y_j \ / \ N(n)}{\sum_{j=0}^{N} (n) \tau_j \ / \ N(n)} \\ &\to \quad \frac{\mathrm{E}\left[Y\right]_1}{\mathrm{E}\left[\tau\right]_1} \quad a.s. \end{aligned}$$

as $n \to \infty$. The following is a rigorous statement of what we have just heuristically derived.

Theorem 8.6. *Let $X = (X_n : n \geq 0)$ be an irreducible recurrent Markov chain on discrete state space S. Then, conditional on $X_0 = x$,*

$$\frac{1}{n}\sum_{j=0}^{n-1} I(X_j = y) \;\rightarrow\; \frac{\mathrm{E}_z\left[\sum_{i=0}^{\tau(z)-1} I(X_j = y)\right]}{\mathrm{E}_z\left[\tau(z)\right]} \quad a.s.$$

as $n \to \infty$.

The Bounded Convergence Theorem yields the following corollary.

Corallary 8.1: Let $X = (X_n : n \geq 0)$be an irreducible recurrent Markov chain on discrete state space S. Then

$$\frac{1}{n}\sum_{j=0}^{n-1} P^j(x, y) \;\rightarrow\; \frac{\mathrm{E}_z\left[\sum_{i=0}^{\tau(z)-1} I(X_j = y)\right]}{\mathrm{E}_z\left[\tau(z)\right]}$$

as $n \to \infty$.

Remark: In applying the strong law of large numbers for iid sequence $(Z_n : n \geq 0)$, we have used the following extension: If the Z_n's are non-negative, then

$$n^{-1}\sum_{i=1}^{n} Z_i \;\rightarrow\; \mathrm{E}\left[Z\right]_1 \quad a.s. \tag{8.27}$$

as $n \to \infty$, regardless of whether $\mathrm{E}\left[Z\right]_1$ is finite or not. (Of course, if $\mathrm{E}\left[Z\right]_1 = \infty$, then (9) asserts that $n^{-1}\sum_{i=1}^{n} Z_i \;\rightarrow\; \infty \quad a.s.$ as $n \to \infty$.)

Remark: Note that the ratio limit specified by Theorem 8.6, namely

$$\frac{\mathrm{E}_z\left[\sum_{i=0}^{\tau(z)-1} \mathrm{I}\left\{X_j = y\right\}\right]}{\mathrm{E}_z\left[tau(z)\right]}$$

is always well-defined, in the sense that the ratio ∞/∞ can never appear. To see this, note that

$$\mathrm{P}_z\left\{\sum_{j=0}^{\tau(z)-1} I(X_j = y) = k\right\} \;=\; \mathrm{P}_z\left\{\tau(y) < \tau(z)\right\}\mathrm{P}_y\left\{\tau(y) < \tau(z)\right\}^{k-1}\mathrm{P}_y\left\{\tau(z) < \tau(y)\right\}$$

for $k \geq 1$. It follows that

$$\mathrm{E}_z\left[\sum_{j=0}^{\tau(z)-1} \mathrm{I}\left\{X_j = y\right\}\right] < \infty$$

Put

$$\nu = \mathrm{E}_z\left[\sum_{i=0}^{\tau(z)-1} \mathrm{I}\left\{X_i = y\right\}\right] \;/\; \mathrm{E}_z\left[\tau(z)\right].$$

Note that Corallary 8.1 guarantees that $\nu = (\nu(y) : y \in S)$ is independence of the choice of return state z (since the left-hand side of the limit of Corallary 8.1 is defined independently of z). Obviously, if $\mathrm{E}_z\left[\tau(z)\right] = \infty$, then $\nu = 0$. However, if $\mathrm{E}_z\left[\tau(z)\right] < \infty$, then ν is a probability distribution on S, and Theorem 8.6 proves that

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\left\{X_j = y\right\} \;\rightarrow\; \nu(y) \quad a.s.$$

as $n \to \infty$. In other words, if $\mathrm{E}_z\left[\tau(z)\right] < \infty$ for some $z \in S$ (and hence for all $z \in S$), X is ergodic (see Definition 8.3).

Note that if there exists $z \in S$ such that $\mathrm{E}_z[\tau(z)] < \infty$, then

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\{X_j = y\} \;\to\; \nu(y) \quad a.s.$$

as $n \to \infty$, where all the $\nu(y)$'s are necessarily positive. (Remember that the current discussion presumes irreducibility!). On the other hand, if $\mathrm{E}_z[\tau(z)] = \infty$, then

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\{X_j = y\} \;\to\; 0 \quad a.s.$$

as $n \to \infty$. This motivates the following definition.

Definition 8.7: A recurrent state $z \in S$ is said to be *positive recurrent* if $\mathrm{E}_z[\tau(z)] < \infty$. Otherwise, a recurrent state $z \in S$ is said to be *null recurrent.*

Our above discussion can be summarized with the following theorem.

Theorem 8.7. *Let $X = (X_n : n \geq 0)$be an irreducible Markov chain on discrete state space S.*

(i) Either all states are positive recurrent or no states are positive recurrent.

(ii) Either all states are null recurrent or no states are null recurrent.

(iii) The chain is positive recurrent if and only if there exists a probability distribution ν for which

$$\frac{1}{n}\sum_{j=0}^{n-1} I(X_j = y) \;\to\; \nu(y) \quad a.s.$$

as $n \to \infty$ or

$$\frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) \;\to\; \nu(y)$$

as $n \to \infty$. The probability distribution ν can be defined as

$$\nu(y) = \frac{\mathrm{E}_z\left[\sum_{j=0}^{\tau(z)-1} I(X_j = y)\right]}{\mathrm{E}_z[\tau(z)]}$$

for any $z \in S$.

(iv) The chain is null recurrent if and only if

$$\begin{cases} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{I}\{X_j = y\} \;\to\; 0 \quad a.s. \quad \text{as } n \to \infty \\ \sum_{j=0}^{\infty} \mathrm{I}\{X_j = y\} \;=\; \infty \quad a.s. \end{cases}$$

or

$$\begin{cases} \frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) \;\to\; 0 \quad \text{as } n \to \infty \\ \sum_{j=0}^{\infty} P^j(x,y) \;=\; \infty \end{cases}$$

8.15 Proof of Theorem 8.7

Theorem 8.7 is fundamental to the analysis of Markov Chains. In the view of its importance, and the fact that its proof illustrates the use of "sample path arguments", we now offer a more complete proof.
We start by proving the strong law for renewal counting processes.

Definition 8.8: Suppose $(\tau_n : n \geq 1)$ is a sequence of iid positive rv's. For

$$N(n) = \max\{m \geq 0 : \sum_{i=0}^{m} \tau_i \leq n\}$$

Then, $N = (N(n) : n \geq 0)$ is called a *renewal counting process.*

Remark: Suppose the m^{th} event (e.g. an arrival to a queue) occurs at time $\sum_{i=0}^{m} \tau_i$. Then, $\mu(n)$ counts the number of events (e.g. arrivals) to occur by time n.

Theorem 8.8. *Let $N = (N(n) : n \geq 0)$ be a renewal counting process. Then,*

$$\frac{N(n)}{n} \to \frac{1}{\mathrm{E}[\tau]_1}$$

a.s. as $n \to \infty$.

Proof. Put $T_m = \tau_0 + \tau_1 + \ldots + \tau_m$. Then,

$$T_{N(n)} \leq n \leq T_{N(n)+1} \tag{8.28}$$

Because the τ_n's are iid,

$$\frac{T_m}{m} \to \mathrm{E}[\tau]_1 \qquad a.s.$$

as $m \to \infty$. It follows that

$$\frac{T_{N(n)}}{N(n)} \to \mathrm{E}[\tau]_1 \qquad a.s.$$

as $n \to \infty$.
(To prove this, let $A = \{\omega : \frac{T_m(\omega)}{m} \to \mathrm{E}[\tau]_1 \text{ as } n \to \infty\}$, and note that the strong law for the τ_n's ensures that $\mathrm{P}\{A\} = 1$. Since $N(n, \omega) \to \infty$ as $n \to \infty$, we see that for $\omega \in A$, $\frac{T_{N(n,\omega)}(\omega)}{N(n,\omega)} \to \mathrm{E}[\tau]_1$. So, $B \supseteq A$, where $B = \{\omega : \frac{T_{N(n,\omega)}(\omega)}{N(n,\omega)} \to \mathrm{E}[\tau_1] \text{ as } n \to \infty\}$. Because $\mathrm{P}\{A\} = 1$,evidently $\mathrm{P}\{B\} = 1$. This is what is called a "sample path argument".)
Now, (1) imples that

$$\frac{T_{N(n)}}{N(n)} \leq \frac{n}{N(n)} \leq \frac{T_{N(n)+1}}{N(n)+1} \cdot \frac{N(n)+1}{N(n)}$$

where

$$\frac{T_{N(n)}}{N(n)} \to \mathrm{E}[\tau]_1 \qquad a.s.$$

$$\frac{T_{N(n)+1}}{N(n)+1} \to \mathrm{E}[\tau]_1 \qquad a.s.$$

$$\frac{N(n)+1}{N(n)} \to 1$$

So, $\frac{n}{N(n)} \to \mathrm{E}[\tau]_1$ a.s. as $n \to \infty$, proving the result. □

Proof of Theorem 8.6: Observe that

$$\frac{1}{n}\sum_{i=0}^{N(n)} Y_i \le \frac{1}{n}\sum_{i=0}^{n-1} I(X_i = y) \le \frac{1}{n}\sum_{i=0}^{N(n)+1} Y_i$$

But

$$\frac{1}{n}\sum_{i=0}^{N(n)} Y_i = \frac{N(n)}{n}\frac{\sum_{i=0}^{N(n)} Y_i}{N(n)}$$

As in the proof of Theorem 8.8, we find that

$$\frac{\sum_{i=0}^{N(n)} Y_i}{N(n)} \to \mathrm{E}\left[Y\right]_1 \qquad a.s.$$

as $n \to \infty$. So $\mathrm{P}\{A\} = 1$ where $A = \{\omega : \frac{\sum_{i=0}^{N(n,\omega)} Y_i(\omega)}{N(n,\omega)} \to \mathrm{E}[Y_1] \text{ as } n \to \infty\}$. Theorem 8.8 proves that $\mathrm{P}\{B\} = 1$, where $B = \{\omega : \frac{N(n,\omega)}{n} \to \frac{1}{\mathrm{E}[\tau]_1} \text{ as } n \to \infty\}$.Observe that $A \cap B \subseteq C$, where $C = \{\omega : \frac{\sum_{i=0}^{N(n,\omega)} Y_i(\omega)}{n} \to \frac{\mathrm{E}[Y_1]}{\mathrm{E}[\tau_1]} \text{ as } n \to \infty\}$.
Since $\mathrm{P}\{A \cap B\} = 1$, $\mathrm{P}\{C\} = 1$. So,

$$\frac{1}{n}\sum_{i=0}^{N(n)} Y_i \to \frac{\mathrm{E}\left[Y_1\right]}{\mathrm{E}\left[\tau_1\right]} \qquad a.s.$$

as $n \to \infty$. Similarly,

$$\frac{1}{n}\sum_{i=0}^{N(n)+1} Y_i \to \frac{\mathrm{E}\left[Y_1\right]}{\mathrm{E}\left[\tau_1\right]} \qquad a.s.$$

as $n \to \infty$. Consequently,

$$\frac{1}{n}\sum_{i=0}^{n-1} \mathrm{I}\left\{X_i = y\right\} \to \frac{\mathrm{E}\left[Y_1\right]}{\mathrm{E}\left[\tau_1\right]} \qquad a.s.$$

as $n \to \infty$, proving the result. ■

Remark: Sample path arguments permit one to reduce probabilistic arguments to "real variables arguments" by applying these ideas to each individual ω (i.e a "sample path").

8.16 Positive Recurrent Markov Chains

Let $X = (X_n : n \ge 0)$ be an irreducible positive recurrent Markov chain on discrete state space S. Then,

$$\frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) \to \nu(y) \qquad a.s.$$

as $n \to \infty$, where

$$\nu(y) = \frac{\mathrm{E}_z\left[\sum_{j=0}^{\tau(z)-1} I(X_j = y)\right]}{\mathrm{E}_z\left[\tau(z)\right]}$$

But

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} P^j(x,y) &= \frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) + \frac{1}{n}(P^n(x,y) - P^0(x,y)) \\ &\to \nu(y)\end{aligned} \tag{8.29}$$

as $n \to \infty$. On the other hand,

$$\frac{1}{n}\sum_{j=0}^{n} P^j(x,y) = \sum_z \frac{1}{n}\sum_{j=0}^{n-1} P^j(x,z)P(z,y) \tag{8.30}$$

The right-hand side should converge to

$$\sum_z \nu(z)P(z,y)$$

as $n \to \infty$. Relations (2) and (3) suggest that ν must coincide with the stationary distribution π.

Exercise 8.8: Prove that if $X = (X_n : n \geq 0)$ is an irreducible positive recurrent Markov chain, then

$$\sum_z \frac{1}{n}\sum_{j=0}^{n-1} P^j(x,z)P(z,y) \to \sum_z \nu(z)P(z,y)$$

as $n \to \infty$. (This interchange of limit and sum is easy if $|S| < \infty$, but requires a bit of care when $|S| = \infty$.)
Suppose that $X = (X_n : n \geq 0)$ is an irreducible Markov chain for which a solution π to

$$\pi = \pi P$$

$$\text{s/t } \pi(x) \geq 0,\ x \in S,\ \sum_x \pi(x) = 1$$

exists (i.e. a stationary distribution exists). Note that

$$\pi(y) = \sum_x \pi(x)\frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) \tag{8.31}$$

If the chain is transient or null recurrent, then

$$\frac{1}{n}\sum_{j=0}^{n-1} P^j(x,y) \to 0$$

as $n \to \infty$. It follows from (4) that if the chain is transient or null recurrent, then $\pi(y) = 0$ for $y \in S$. Since π was assumed to be stationary distribution, this yields a contradiction. So, X must be positive recurrent.

We summarize this discussion with the following theorem.

Theorem 8.9. *Let $X = (X_n : n \geq 0)$ be an irreducible Markov chain on discrete state space S. Then, X is positive recurrent if and only if there exists a solution to*

$$\pi = \pi P$$

$$s/t\ \pi(x) \geq 0,\ x \in S,\ \sum_x \pi(x) = 1$$

Remark: This shows that an irreducible Markov chain is ergodic if and only if a stationary distribution π exists.
Remark: Note that $\pi = \pi P$ can be interpreted probabilistically as:
(equilibrium probability of being in state y) $= \sum_x$ (equilibrium probability of being in state x at previous transition) $\cdot P(x,y)$
These are called the "global balance equations".

Remark: Theorems 8.6 and 8.9 prove that if a stationary distribution π exists (for an irreducible Markov chain), then

$$\pi(z) = \frac{\mathrm{E}_z \left[\sum_{j=0}^{\tau(z)-1} \mathrm{I}\{X_j = z\} \right]}{\mathrm{E}_z [\tau(z)]}$$

But the number of visits to z over a z-cycle is exactly one. So,

$$\pi(z) = \frac{1}{\mathrm{E}_z [\tau(z)]}$$

or, equivalently,

$$\mathrm{E}_z [\tau(z)] = \frac{1}{\pi(z)}$$

For example, consider the Ehrenfest model. This proves that $\mathrm{E}_0 [\tau(0)] = 2^m$, where m is the number of molecules. In other words (given that m can easily be of the order of 10^{23} or more), the return time to 0 (all molecules on one side!) is exceptionally long!
In many applications, we are interested in developing approximations for cumulative rewards:

$$R_n = \sum_{j=0}^{n-1} f(X_j)$$

where $f(x)$ is the reward associated with spending one unit of time in x.

Theorem 8.10. *Suppose that $X = (X_n : n \geq 0)$ is an irreducible positive recurrent Markov chain in discrete state space S. If f is non-negetive, then*

$$\frac{1}{n} \sum_{j=0}^{n-1} f(X_j) \to \sum_y \pi(y) f(y) \qquad a.s.$$

as $n \to \infty$.

Exercise 8.9: Prove Theorem 8.10. Theorem 8.10 supports the approximation

$$R_n \approx n\pi f \tag{8.32}$$

where n is large.

Exercise 8.10: Suppose that the cumulative reward takes the form

$$R_n = \sum_{j=0}^{n-1} f(X_j, X_{j+1})$$

for some $f : S^2 \to [0, \infty)$. What should $\frac{R_n}{n}$ converge to as $n \to \infty$?

Exercise 8.11: Suppose that $X = (X_n : n \geq 0)$ satisfies a stochastic recursion of the form $X_{n+1} = g(X_n, Z_{n+1})$. Assume that the cumulative reward takes the form

$$R_n = \sum_{j=0}^{n-1} f(X_j, Z_{j+1})$$

for some non-negetive f. What should $\frac{R_n}{n}$ converge to as $n \to \infty$?

8.17 The Central Limit Theorem for Markov Chains

To improve upon the approximation (5), we appeal to the central limit theorem. The iid regenerative cycles can again be used to advantage here. For a given reward function f, put

$$Z_i = \sum_{j=T_{i-1}}^{T_i - 1} f_c(X_j)$$

where $f_c(x) = f(x) - \pi f$.

Theorem 8.11. *Let $X = (X_n : n \geq 0)$ be an irreducible positive recurrent Markov chain on discrete state space S. Suppose that $\sum_X \pi(x)|f(x)| < \infty$ and*

$$\mathrm{E}_z \left[\left(\sum_{j=0}^{\tau(z)-1} |f_c(X_j)| \right)^2 \right] < \infty$$

Then,

$$\frac{\sum_{j=0}^{n-1} f(X_j) - n\pi f}{n^{1/2}} \Rightarrow \sigma N(0,1)$$

as $n \to \infty$, where $\sigma^2 = \frac{\mathrm{E}[Z_1]^2}{\mathrm{E}[\tau_1]}$

Exercise 8.12: Suppose that $X = (X_n : n \geq 0)$ is an irreducible finite state Markov chain.
a) Argue that for each $x \in S$, $\mathrm{P}_x\{\tau(z) \leq m\} > 0$, where $m = |S|$.
b) Put $\delta = \min\{\mathrm{P}_x\{\tau(z) \leq m\} : x \in S\}$. Prove that $\mathrm{P}_x\{\tau(z) > nm\} \leq (1-\delta)^n$.
c) Prove that $\mathrm{E}_z\left[\left(\sum_{j=0}^{\tau(z)-1} |f_c(X_j)|\right)^2\right] < \infty$ (and hence the central limit theorem always holds for finite state irreducible Markov chains).
d) Recall that $\mathrm{E}_z[\tau(z)] = \frac{1}{\pi(z)}$. Show how to compute

$$\mathrm{E}_z \left[\left(\sum_{j=0}^{\tau(z)-1} f_c(X_j) \right)^2 \right] \tag{8.33}$$

(and hence σ^2) by using "first transition analysis" (Hint: Use "first transition analysis" to express (6) in terms of $\mathrm{E}_x\left[\left(\sum_{j=0}^{\tau(z)-1} f_c(X_j)\right)^k\right]$ for $k = 1, 2$.)
Put $R_n = \sum_{j=0}^{n-1} f(X_j)$. Theorem 8.11 supports the approximation

$$R_n \stackrel{\mathcal{D}}{\approx} N(n\pi f, n\sigma^2)$$

for n large.

8.18 Monte Carlo Computation of Steady-State Quantities

If the state space S of $X = (X_n : n \geq 0)$ is either large or infinite, numerical computation of the stationary distribution π via solution of

$$\pi = \pi P$$

subject to $\pi(x) \geq 0$, $x \in S$, and $\sum_X \pi(x) = 1$, becomes impractical. In this setting, one widely used alternative is to compute steady-state expectations via simulation.

To compute $\alpha = \pi f$, simulate $X = (X_j : j \geq 0)$ over a long time interval, say to time T_n with n large (where T_n is the n^{th} time at which X returns to the regeneration state $z \in S$). Note that σ^2 can be estimated via

$$\hat{\sigma}_n^2 = \frac{1}{T_n} \sum_{i=1}^{n} \Big(\sum_{j=T_{i-1}}^{T_i - 1} (f(X_j) - \alpha_n) \Big)^2$$

where

$$\alpha_n = \frac{1}{T_n} \sum_{i=0}^{T_n - 1} f(X_i)$$

This leads to the following algorithm for producing approximate $100(1-\delta)\%$ confidence intervals for α.
Algorithm 11.1: (To compute approximate $100(1-\delta)\%$ confidence intervals for $\alpha = \pi f$.)
1. Select n large.
2. Simulate X, conditional on $X_0 = z$, to time T_n (where T_n is the n^{th} time X visits z).
3. Compute

$$\alpha_n = \frac{1}{T_n} \sum_{i=0}^{T_n - 1} f(X_i)$$

$$\hat{\sigma}_n^2 = \frac{1}{T_n} \sum_{i=1}^{n} \Big(\sum_{j=T_{i-1}}^{T_i - 1} (f(X_j) - \alpha_n) \Big)^2$$

4. Select z so that $\mathrm{P}\left\{-z \leq N(0,1) \leq z\right\} = 1 - \delta$. Then,

$$[\alpha_n - z\sqrt{\frac{\hat{\sigma}_n^2}{T_n}}, \alpha_n + z\sqrt{\frac{\hat{\sigma}_n^2}{T_n}}]$$

is an approximate $100(1-\delta)\%$ confidence intervals for $\alpha = \pi f$.

8.19 Time-Reversed Markov Chains

Suppose that $X = (X_n : n \geq 0)$ is an irreducible positive recurrent Markov chain. We will study the Markov structure of the stochastic process obtained by "running time backwards". To define this time-reversed process, it is convenient to work with a stationary version of X that was initialized at time $n = -\infty$.
To be more precise, let $(X_n^* = -\infty < n < \infty)$ be a stationary version of X with joint distributions defined by

$$\mathrm{P}\left\{X_{k_1}^* = x_1, \ldots, X_{k_m}^* = x_m\right\} = \pi(x_1) P^{k_2 - k_1}(x_1, x_2) \ldots P^{k_n - k_{n-1}}(x_{k-1}, x_k)$$

for $k_1 < k_2 < \ldots < k_m$. This defines a legitimate distribution for the doubly-indexed process $(X_n^* = -\infty < n < \infty)$. Intuitively, one can envision $(X_n^* = -\infty < n < \infty)$ as the process X started at time $-\infty$, so that at finite times, it exhibits equilibrium behaviour.

Definition 8.9: The *time-reversal* of $(X_n^* = -\infty < n < \infty)$ is the sequence $(Y_n = -\infty < n < \infty)$, where $Y_n = X_{-n}^*$.

Exercise 8.13: Prove that $Y = (Y_n : -\infty < n < \infty)$ is a Markov chain, in the sense that

$$\mathrm{P}\left\{Y_{n+1} = y | Y_j : j \leq n\right\} = \mathrm{P}\left\{Y_{n+1} = y | Y_n\right\}$$

Exercise 8.13 asserts that the time-reversal is a Markov chain. We now compute its transition matrix.

Proposition 8.4: Let $X^* = (X_n^* : -\infty < n < \infty)$ be the stationary version of an irreducible positive recurrent Markov chain on discrete state space S. The time-reversal Y of X^* is a Markov chain with transition matrix $\tilde{P} = (\tilde{P}(x,y) : x, y \in S)$, where

$$\tilde{P}(x,y) = \frac{\pi(y)P(y,x)}{\pi(x)}$$

Proof. Problem 11.6 establishes the Markov property. To compute $\tilde{P}$, note that

$$\begin{aligned}\tilde{P}(x,y) &= \mathrm{P}\left\{Y_{n+1} = y | Y_n = x\right\} \\ &= \frac{\mathrm{P}\left\{Y_{n+1} = y, Y_n = x\right\}}{\mathrm{P}\left\{Y_n = x\right\}} \\ &= \frac{\mathrm{P}\left\{X^*_{-n-1} = y, X^*_{-n} = x\right\}}{\mathrm{P}\left\{X^*_{-n} = x\right\}} \\ &= \frac{\pi(y)P(y,x)}{\pi(x)}\end{aligned}$$

□

8.20 Birth-Death Markov Chains

An important class of examples that arises in many applications is that of a “birth-death chain”. Such a Markov chain has state space $S = \mathbb{Z}^+$, with

$$P(x, x+1) = p_x$$

$$P(x, x) = r_x$$

$$P(x, x-1) = q_x$$

for $x \geq 1$, with $p_x + r_x + q_x = 1$, and $P(0,1) = p_0$ with $P(0,0) = r_0$, $(p_0 + r_0 = 1)$.
The global balance equations take the form

$$\pi(x) = p_{x-1}\pi(x-1) + r_x\pi(x) + q_{x+1}\pi(x+1)$$

for $x \geq 1$, and

$$\pi(0) = r_0\pi(0) + q_1\pi(1)$$

One can explicitly solve these equations, yielding

$$\pi(x) = \pi(0)\frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x}$$

for $x \geq 1$. In order that the $\pi(x)$'s sum to one, we must have

$$1 = \pi(0)(1 + \sum_{x=1}^{\infty} \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x}) \,. \tag{8.34}$$

If

$$\sum_{x=1}^{\infty} \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} = \infty \tag{8.35}$$

there is no value for $\pi(0)$ that solves (7). Hence, no stationary distribution exists if (8) holds. In other words, $X = (X_n : n \geq 0)$ is either transient or null recurrent when (8) occurs.

On the other hand, if

$$\sum_{x=1}^{\infty} \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} < \infty$$

then a stationary distribution π exists and

$$\pi(x) = \pi(0) \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x}$$

with

$$\pi(0) = \frac{1}{1 + \sum_{y=1}^{\infty} \frac{p_0 p_1 \cdots p_{y-1}}{q_1 q_2 \cdots q_y}}$$

Example 8.17: Recall our queueing chain on $\mathbb{Z}^+$. Here, $p_x = p$, $r_x = r$, and $q_x = q$ for $x \geq 1$. In this case, the queue is transient or null recurrent when $p \geq \frac{1}{2}$, and positive recurrent when $p < \frac{1}{2}$. For $p < \frac{1}{2}$, the stationary distribution is

$$\pi(x) = \left(\frac{p}{q}\right)^x \left(1 - \left(\frac{p}{q}\right)\right)$$

In particular, the steady-state probability of k or more packets in the queue is $(\frac{p}{q})^k$.

8.21 Detailed Balance and Reversibility

Note that when a birth-death chain is positive recurrent, the stationary distribution π satisfies the recursion

$$\pi(x+1) = \frac{p_x}{q_{x+1}} \pi(x)$$

so that

$$q_{x+1} \pi(x+1) = \pi(x) p_x$$

for $x \geq 0$. This is equivalent to the statement that

$$\pi(x+1) P(x+1, x) = \pi(x) P(x, x+1) \tag{8.36}$$

for $x \geq 0$. But $P(x, y) = P(y, x) = 0$ if $|y - x| \geq 2$. Hence, (9) implies that

$$\pi(x) P(x, y) = \pi(y) P(y, x) \tag{8.37}$$

for all $x, y \in S$. Recall that the time-reversal Y has transition matrix $\tilde{P}$, where

$$\tilde{P}(x, y) = \frac{\pi(y) P(y, x)}{\pi(x)}$$

Relation (9) asserts that $\tilde{P} = P$.

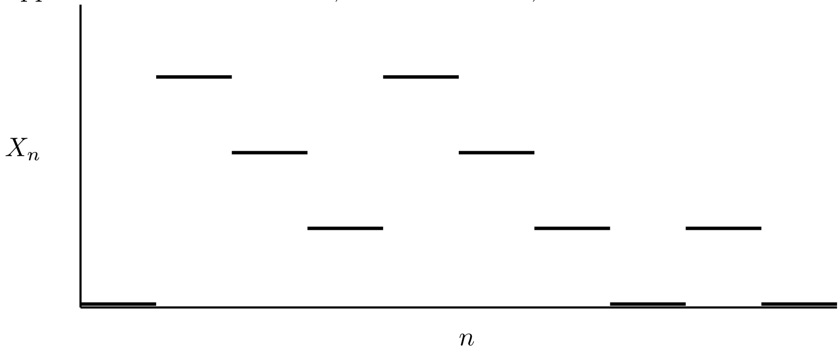
Definition 8.10: Let $X = (X_n : n \geq 0)$ be an irreducible positive recurrent Markov chain on discrete state space S. The Markov chain X is said to be *reversible* if $\tilde{P} = P$ or, equivalently, if

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

for all $x, y \in S$.

A reversible Markov chain has the property that, in equilibrium, the reversed process is statistically indistinguishable (i.e. has the same distribution) from the forward process. Note that all the positive recurrent birth-death chains are reversible.

Example 8.18: Suppose that a Markov chain, when simulated, has realizations of the form



Note that $X = (X_n : n \geq 0)$ can increase by jumps of size 1,2, or 3, but can decrease only with jumps of size 1. The time-reversal increases only by jumps of size 1, but can decrease by jumps of size 1, 2, or 3. It immediately follows that this Markov chain cannot be reversible.

Most Markov chain models are not reversible. In the next section, we will briefly discuss some of the nice properties that reversible Markov chains possess.
Note that when X is reversible, its stationary distribution satisfies

$$\pi(x)P(x,y) = \pi(y)P(y,x) \tag{8.38}$$

for $x, y \in S$. This can be interpreted probabilistically, as
"rate at which X jumps from x to y" = "rate at which X jumps from y to x".
The equations (11) are often called the "detailed balance equations".

Exercise 8.14: a) Suppose that one can find a probability distribution π that solves the detailed balance equations. Then, π solves the global equations (and hence is a stationary distribution).
b) Show, by example, that a Markov chain's stationary distribution need not satisfy the detailed balance equations.

8.22 Reversible Markov Chains

Let $X = (X_n : n \geq 0)$ be an irreducible reversible finite state Markov chain. Then, the stationary distribution $\pi = (\pi(x) : x \in S)$ satisfies

$$\pi(x)P(x,y) = \pi(y)P(y,x) \tag{8.39}$$

This seems closely related to the concept of symmetric matrices.
To be more precise, let $\tilde{D} = diag(\sqrt{\pi(x)} : x \in S)$ be a diagonal matrix with the $\sqrt{\pi(x)}$'s on its diagonal. Then, (12) asserts that

$$\sqrt{\frac{\pi(x)}{\pi(y)}}P(x,y) = \sqrt{\frac{\pi(y)}{\pi(x)}}P(y,x)$$

so that

$$B = \tilde{D}P\tilde{D}^{-1}$$

is symmetric. Consequently, P is similar to a symmetric matrix. But symmetric matrices are diagonalizable and have real eigenvalues. It follows that P is diagonalizable with real eigenvalues. We summarize this discussion with the following result.

Proposition 8.5: Let P be the transition matrix of a finite irreducible reversible Markov chain. Then, P has real eigenvalues and is diagonalizable.

As a result, P^n has a simple "spectral representation". In particular,

$$P^n = RD^nR^{-1}$$

where D is a diagonal matrix with the eigenvalues of P on its diagonal. It follows that $P^n(x, y)$ can be written as a linear combination of $\pi(y)$ and $|S| - 1$ decaying geometric terms.

8.23 Bayesian Statistics

Earlier in the course, we discussed the problem of estimating model parameters from observed data via the method of maximum likelihood. In the presence of additional "subjective" information on the parameters, an alternative is available.
Suppose that we are interested in estimating the statistical parameter $\theta \in \mathbb{R}^d$. We assume that the statistician has a "prior" $p(\cdot)$ on θ. The probability density function $p(\cdot)$ represents the statistician's subjective likelihood on θ in the absence of observed data. Once one observes data, the statistician re-computes her likelihood (now called the "posterior") to reflect the information about θ that is present in the data.
Mathematically, suppose that $L(\vec{X}|\theta)$ is the likelihood of the observed data $\vec{X}$ if the underlying parameter equals θ. Then, the posterior on θ is given by

$$p(\theta|\vec{X}) = \frac{L(\vec{X}|\theta)p(\theta)}{\int_{\boldsymbol{R}^d} L(\vec{X}|\theta')p(\theta')d\theta'} \tag{8.40}$$

Formula (8.40) is known as Bayes' formula.

Example 8.19: Suppose that you wish to estimate $p = P(A)$ for observed data. If one observes n independent trials, the data set consists of $I_1, I_2, \ldots, I_n$, where the I_j's are iid with $P(I_j = 1|\theta) = \theta = 1 - P(I_j = 0|\theta)$ (here, θ is the unknown Bernoulli parameter). Here, $\vec{X} = (I_1, \ldots, I_n)$ and

$$L_n(\vec{X}|\theta) = \binom{n}{S_n} \theta^{S_n}(1-\theta)^{n-S_n}$$

where $S_n = I_1 + \ldots + I_n$. The principle of maximum likelihood asserts that one should estimate θ via

$$\hat{\theta}_n = \frac{1}{n}S_n$$

So, if $I_1 = I_2 = \ldots = I_n = 0$, the maximum likelihood estimator asserts that the probability estimate for $P(A)$ is zero.

But this often seems unreasonable from a practical standpoint. For example, would you estimate the reliability of a component at 100% based on five successful experiments on the device? An alternative is to analyze the data using a Bayesian approach. Suppose that our prior on θ is a uniform distribution on $[0, 1]$. Then,

$$p(\theta|\vec{X}) = \frac{\theta^{S_n}(1-\theta)^{n-S_n}}{\int_0^1 \theta'^{S_n}(1-\theta')^{n-S_n}d\theta'}$$

This posterior is a distribution on θ that has posterior mean

$$(S_n + 1)/(n + 2) \tag{8.41}$$

Hence, even if $S_n = 0$ (as would occur if $I_1 = I_2 = \ldots = I_n = 0$), the posterior mean on θ is $1/(n + 2)$, which indicates a positive probability of component failure (even if all five experiments were successful). As a consequence, one might choose to take a Bayesian point of view in analyzing some data sets.
In Example 8.19, computing the posterior mean can be done in closed form. In general, Bayesian methods may require numerical integration to deal with the normalization factor that appears in (1).

8.24 Markov Chain Monte Carlo

Note that computing functionals of the posterior distribution for the parameter θ can, in principle, be accomplished by sampling from the density

$$\frac{L(\vec{X}|\theta)p(\theta)}{\int_{\mathbb{R}^d} L(\vec{X}|\theta')p(\theta')d\theta'}$$

Since the density above is typically known, in closed form (modulo the normalization constant), this sampling problem is a special case of sampling from a distribution of the form

$$\pi(B) = \frac{\int_B \nu(y)dy}{\int_S \nu(z)dz}. \tag{8.42}$$

where $\nu(\cdot)$ is easily computable. In addition to the key role that such computations play in Bayesian statistics, such computational problems arise in many other applied settings.

Example 8.20: Physicists are often interested in studying models in which the physical state $\vec{x}$ (i.e. the so-called "configuration") is chosen according to a distribution for which the mass function is in proportion to

$$\exp(-U(\vec{x})/kT) \tag{8.43}$$

The function $U(\cdot)$ is called the *potential energy* function, T is known as the *temperature*, and k is *Boltzmann's constant*. The mass function of such a *Boltzmann's distribution* (also known as a *Gibb's distribution*) is then given by

$$\frac{\exp(-U(\vec{x})/kT)}{\sum_{\vec{y}\in\mathcal{C}} \exp(-U(\vec{y})/kT)}$$

where $\mathcal{C}$ is the set of physically realizable configurations. The normalization constant

$$z(T) \triangleq \sum_{\vec{y}\in\mathcal{C}} \exp(-U(\vec{y})/kT)$$

is called the *partition function*.

Example 8.21: A special case of the above Boltzmann distribution is that associated with the *Ising model*. The Ising model is used to model the behavior of a magnetic material. The set of configurations $\mathcal{C} = \{(y_{ij} : y_{ij} \in \{0,1\}, 1 \le i,j \le N)\}$ models the set of sites associated with the atomic spins on N^2 lattice points. In particular, $y_{ij} = 0$ (1) if the spin at lattice point (i,j) is 0 (1). The potential energy of configuration $\vec{y}$ is

$$U(\vec{y}) = -J \sum_{\sigma\sim\sigma'} y_\sigma y_{\sigma'} + \sum_\sigma h_\sigma y_\sigma$$

where $\sigma \sim \sigma'$ means that σ and σ' are neighboring pairs on the lattice, J is the *interaction strength*, and h_σ is the strength of the external magnetic field at site σ.

Example 8.22: In molecular structure simulations, the configuration space $\mathcal{C}$ describes the position of k macromolecules. Hence, $\mathcal{C} = \{(\vec{x}_1, \ldots, \vec{x}_k) : \vec{x}_i \in \mathbb{R}^3\}$. A commonly used potential energy function for such simulations is

$$U(\vec{x}) = \sum_{i,j} \Phi(\|\vec{x}_i - \vec{x}_j\|)$$

where

$$\Phi(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

the so-called *Leonard-Jones pair potential*.

Example 8.23: Here, $\mathcal{C}$ is the set of connected graphs on the integer lattice with N (undirected) edges and without circuits. The potential energy $U(\cdot)$ is constant across all such configurations. Because the graph contains no loops, such a model puts uniform weight on all simple symmetric random walks of duration N which never self-intersect (i.e. a so-called "self-avoiding random walk"). Such models are used in studying the growth of chain polymers.

To deal with the above class of computations, it is common to use Markov chain Monte Carlo (MCMC). The idea behind MCMC is to simulate a Markov chain $X = (X_n : n \geq 0)$ having $\pi(\cdot)$ (see (3)) as its unique stationary distribution. Note that it is common in many such applications to need to simulate a Markov chain that is evolving in a high dimensional (or continuous) state space.

8.25 The Metropolis Algorithm

We wish to construct an easily simulatable Markov chain $X = (X_n : n \geq 0)$ having stationary distribution $\pi(\cdot)$ given by (3). To construct X, let $q : S \times S \to [0, \infty)$ be a transition density on S, so that

$$1 = \int_S q(x, y)dy$$

We assume that we have an algorithm available for making transitions according to the transition kernel

$$Q(x, B) = \int_B q(x, y)dy$$

The Metropolis algorithm suggests simulating the Markov chain $X = (X_n : n \geq 0)$ having transition kernel

$$P(x, dy) = p(x, y)dy$$

for $x \neq y$, where

$$p(x, y) = q(x, y) \min\left(1, \frac{\nu(y)q(y, x)}{\nu(x)q(x, y)}\right)$$

Note that for $x \neq y$,

$$\begin{aligned}\nu(x)p(x, y) &= \min(\nu(y)q(y, x), \nu(y)q(y, x)) \\ &= \nu(y)q(y, x)\end{aligned}$$

so that

$$\pi(y)dy = \frac{\nu(y)dy}{\int_S \nu(z)dz}$$

satisfies the "detailed balance conditions." Our earlier discussion of detailed balance establishes that π is therefore a stationary distribution for X.
It remains only to describe how X can be simulated. To generate X_{n+1} from X_n:

1. Generate $Y \in S$ from $Q(X_n, \cdot)$.
2. Generate a uniform r.v. U on $[0, 1]$.
3. If $U \leq \min(1, \frac{\nu(y)q(y,x)}{\nu(x)q(x,y)})$, $X_{n+1} = Y$. Else, $X_{n+1} = X_n$.

This algorithm leads to a Markov chain $X = (X_n : n \geq 0)$ that is reversible. There are more general versions of the Metropolis algorithm that lead to Markov chains that are not reversible.

8.26 Convergence to Stationarity

Because of initialization effects that may be present in the above Metropolis algorithm, it is of interest to know how many steps the Markov chain X must take before it is in (approximate) stationarity. This reduces, mathematically, to the question of convergence of $P_x(X_n \in \cdot)$ to $\pi(\cdot)$ and its associated rate of convergence. This problem also arises in the context of computing the n-step transition probabilities for a Markov chain. If n is large or the state space S is large or continuous, computing $P_x(X_n \in \cdot)$ may be expensive. In this case, one may wish to approximate $P_x(X_n \in \cdot)$ to avoid the need to do numerical computation. Again, this relates to the issue of convergence of $P_x(X_n \in \cdot)$ to $\pi(\cdot)$, for in that case $\pi(\cdot)$ may be used as an approximation for the n-step transition probabilities.
Let us study this problem in the context of a positive recurrent irreducible discrete-time Markov chain on discrete state space S.

Example 8.24: Let $X = (X_n : n \geq 0)$ be a Markov chain with transition graph

Note that $P^{2n}(1,1) = 1$, whereas $P^{2n+1}(1,1) = 1$. Here, different approximations are needed for P^k, depending on whether k is even or odd.

Example 8.25: Consider the birth-death chain

Note that if X_0 is even, then X_{2n} is even, whereas X_{2n+1} is odd. Again, different approximations are needed for P^k, depending on whether or not k is even or odd.

Definition 8.11: Let $X = (X_n : n \geq 0)$ be an irreducible Markov chain with transition matrix P. The state $x \in S$ is said to be *periodic of period* p if

$$\gcd\{n \geq 1 : P^n(x,x) > 0\} = p$$

(where gcd = "greatest common divisor"). If the period is 1, the state x is said to be *aperiodic*.

Proposition 8.6: Suppose X is irreducible. Then $x \in S$ is periodic of period p if and only if y is periodic of period p for all $y \in S$.

An irreducible Markov chain with transition matrix P is periodic of period p if and only if S can be partitioned into $P_1, P_2, \ldots, P_p$ for which P has the block structure

$$P = \begin{pmatrix} 0 & P_{12} & 0 & \cdots & 0 \\ 0 & & P_{23} & & \\ & & & \ddots & \\ & & & & P_{p-1,p} \\ P_{p,1} & 0 & & & 0 \end{pmatrix}$$

Note that if $P(x,x) > 0$ for some $x \in S$, X must be aperiodic. (The converse is false, however.)

Theorem 8.12. *Suppose X is irreducible, aperiodic, and positive recurrent, living discrete state space S. Then,*

$$X_n \xrightarrow{tv} X_\infty$$

as $n \to \infty$. i.e.

$$\sup_B |P_x(X_n \in B) - \pi(B)| \to 0$$

as $n \to \infty$.

See p.73-77 of P.E. Hoel, S.C. Port, and C.J. Stone, Introduction to Stochastic Processes, Houghton Mifflin, Boston (1972).
Remark. Note that convergence to stationarity is closely related to the notion of "mixing" (the asymptotic independence between X_0 and X_n when n is large). Observe that

$$\begin{aligned} P_x(X_0 \in A, X_n \in B) &= E_x I(X_0 \in A) P(X_n \in B | X_0) \\ &\to E_x I(X_0 \in A)\pi(B) \\ &= \lim_{n\to\infty} P_x(X_0 \in A) P_x(X_n \in B) \end{aligned}$$

when Theorem 8.12 pertains.

8.27 Coupling

We now prove a "convergence to stationarity" result using a probabilistic method known as "coupling." This approach will establish convergence to stationarity for an important class of Markov chains (including some continuous state space examples) and will even provide a computable bound on the rate of convergence. Throughout this section, we assume:

A1. There exists $m \geq 1$, a probability distribution $\varphi(\cdot)$ on S, and $\lambda > 0$ for which

$$P_x(X_m \in \cdot) \geq \lambda \varphi(\cdot)$$

Remark. Suppose that X is an aperiodic irreducible finite state Markov chain. Then, there exists $m \geq 1$ such that P^m is a strictly positive matrix. Hence, there exists $z \in S$ such that

$$\min_{x \in S} P^m(x, z) > 0$$

Hence,

$$P^m(x, y) \geq \lambda \varphi(y) \tag{8.44}$$

where

$$\varphi(y) = \delta_{yz}$$

$$\lambda = \min_{x \in S} P^m(x, z)$$

In particular, if P has a column of its transition matrix which is strictly positive, the condition (5) holds with $m = 1$.
Remark. Suppose that $X = (X_n : n \geq 0)$ is a Markov chain taking values in state space $S \subseteq \mathbb{R}^d$ for which

$$P(x, B) = \int_B p(x, y) dy$$

for some transition density p. If

$$\varphi(B) = \int_B \phi(y) dy$$

for some density $\phi(y)$, condition A1 for $m = 1$ requires that for $x \in S$,

$$\int_B p(x, y) dy \geq \int_B \lambda \phi(y) dy$$

for all (measurable) $B \subseteq S$. Condition (6) holds if

$$p(x, y) \geq \lambda \phi(y)$$

for $x, y \in S$, so that

$$\lambda\phi(y) = \inf_{x\in S} p(x,y)$$
$$\text{i.e. } \phi(y) = \inf_{x\in S} p(x,y) \,/ \int_S \inf_{x\in S} p(x,z)dz$$
$$\lambda = \int_S \inf_{x\in S} p(x,z)dz$$

Hence, (6) holds provided that

$$\int_S \inf_{x\in S} p(x,z)dz > 0$$

Example 8.26: Consider the storage sequence $S = (S_n : n \geq 0)$ satisfying the recursion

$$S_{n+1} + aS_{n+1}^b = S_n + Z_{n+1}$$

where $(Z_n : n \geq 1)$ is iid with common density $f_z(\cdot)$ which is continuous and positive on $[0, c]$. Then, S is a Markov chain on state space $[0, (c/a)^{1/b}]$. Note that

$$\begin{aligned} P_x(S_1 \geq y) &= P_x(S_1 + aS_1^b \leq y + ay^b) \\ &= P(x + Z_1 \leq y + ay^b) \\ &= P(Z_1 \leq y + ay^b - x) \end{aligned}$$

so that

$$\begin{aligned} p(x,y) &= \frac{d}{dy} P_x(S_1 \leq y) \\ &= f_z(y + ay^b - x)(1 + aby^{b-1}) \end{aligned}$$

Hence, if $(c/a)^{1/b} < c$, it follows that for $(c/a)^{1/b} \leq y + ay^b \leq c$,

$$\inf\{p(x,y) : 0 \leq x \leq (c/a)^{1/b}\} > 0$$

verifying (7).

A number of the MCMC algorithms described earlier can, under suitable conditions, be viewed as Markov chains satisfying A1.

Our main theorem here is:

Theorem 8.13. *Suppose that $X = (X_n : n \geq 0)$ is a Markov chain satisfying A1 Then, X has a unique stationary distribution $\pi(\cdot)$ such that:*

i.) *for each non-negative f,*

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \to \int_X f(y)\pi(dy) \;\; a.s.$$

as $n \to \infty$;

ii.) $\sup_{B\subseteq S} |P_x(X_n \in B) - \pi(B)| \leq (1-\lambda)^{\lfloor n/m \rfloor}$

Remark. The second part of Theorem 8.13 offers a computable bound on the rate of convergence to stationarity. It further asserts that $X_n \xrightarrow{tv} X_\infty$ as $n \to \infty$, where $P(X_\infty \in \cdot) = \pi(\cdot)$.
The proof of Theorem 8.13 proceeds via a series of steps.

Step 1. Show that X is suitably regenerative. Under condition A1, set

$$Q(x, dy) = (P_x(X_m \in dy) - \lambda\varphi(dy))/(1-\lambda)$$

Note that $Q(x, dy) \geq 0$ and $\int_S Q(x, dy) = 1$, so $(Q(x, dy) : x, y \in S)$ is a transition kernel. Furthermore,

$$P_x(X_m \in dy) = \lambda\varphi(dy) + (1-\lambda)Q(x, dy) \tag{8.45}$$

The equality (8) expresses $P_x(X_m \in \cdot)$ as a mixture of the distributions $\varphi(\cdot)$ and $Q(x, \cdot)$. Such a mixture can be viewed probabilistically as:

If $X_r = x$, flip a coin having probability λ of heads and $1-\lambda$ of tails. If the coin comes up heads, distribute X_{r+m} according to distribution φ. Otherwise, distribute X_{r+m} according to $Q(x, \cdot)$. Once X_{r+m} has been generated, generate $X_{r+1}, \ldots, X_{r+m-1}$ from the conditional distribution $P((X_{r+1}, \ldots, X_{r+m-1}) \in \cdot | X_r, X_{r+m})$.

Every time η at which a head occurs leads to a suitable regeneration. In particular, $X_{\eta+m}$ has distribution φ (thereby initiating a fresh regenerative cycle) and $X_{\eta+m}$ is independent of X_η (leading to suitable independence between cycles). If we flip a coin once every m steps, it follows that τ, the first time at which X has distribution φ, has distribution $P_x(\tau > km) = (1-\lambda)^k$. Hence, $E_x\tau = m/\lambda$. Thus, X exhibits regenerative behavior with finite expected cycle length. An argument like that of Theorem 10.2 proves part a.) of Theorem 8.13.

Remark. Note that if $m = 1$, $X_{\tau-1}$ and X_τ are independent, so the above coin flip idea leads to iid cycles (precisely the situation analyzed in Lectures 10 and 11). If $m > 1$, $X_{\tau-m}$ and X_τ are independent, but $X_{\tau-m+1}, \ldots, X_{\tau-1}$ may be correlated with X_τ. In this case, the cycles are identically distributed but not independent. Nevertheless, they exhibit enough independence to establish part a.) of Theorem 8.13.

Remark. Why do we use the coin flip idea to construct regenerations here? We wish to apply these ideas to continuous state Markov chains (like the storage model or Metropolis algorithm). Such chains typically do not return to any given point infinitely often, so the regenerative approach that worked on discrete state space (using returns to a fixed state) fails here.

Step 2. Establish the "coupling inequality."

Let $X = (X_n : n \geq 0)$ be a version of the Markov chain conditioned on $X_0 = x \in S$. Suppose that $X^* = (X_n^* : n \geq 0)$ is a stationary version of the Markov chain (that is initiated with the distribution π). Assume we can simulate X and X^* simultaneously in such a way that there exists a finite time T (called the "coupling time") at which $X_T = X_T^*$. Once the two chains "couple," simulate them together according to the transition dynamics of the chain, so that $X_n = X_n^*$ for $n \geq T$. Assuming such a coupling time T exists, note that

$$\begin{aligned} P_x(X_n \in B) - \pi(B) &= P_x(X_n \in B) - P(X_n^* \in B) \\ &= P_x(X_n \in B, T \leq n) + P_x(X_n \in B, T > n) \\ &\quad - P(X_n^* \in B, T \leq n) - P(X_n^* \in B, T > n) \end{aligned}$$

But $X_n = X_n^*$ on $\{T \leq n\}$. So, $P_x(X_n \in B, T \leq n) = P(X_n^* \in B, T \leq n)$. Hence,

$$\begin{aligned} P_x(X_n \in B) - \pi(B) &= P_x(X_n \in B, T > n) - P(X_n^* \in B, T > n) \\ &\leq P_x(T > n) \end{aligned}$$

Similarly, $\pi(B) - P_x(X_n \in B) \leq P_x(T > n)$. It follows that

$$\sup_{B \subseteq S} |P_x(X_n \in B) - \pi(B)| \leq P_x(T > n) \tag{8.46}$$

The inequality (9) is known as the "coupling inequality."

Step 3. Establish a suitable "coupling."

Note that when the state space is continuous, it is unlikely that independent simulation of X and X^* will lead to a finite coupling time T at which the two chains "meet." To get X and X^* to couple requires that we somehow correlate the simulations of X and X^*.

Recall our coin flip idea. At times 0, m, $2m$, $3m$, ..., the idea is to flip a coin having probability of heads equal to λ. Independent simulation of X and X^* would involve flipping two coins at each such time. To create a finite coupling time T, we instead flip just one coin having probability of heads equal to λ. If the coin flip at time rm is heads, we generate Y from the distribution φ and put $X_{(r+1)m} = X^*_{(r+1)m} = Y$. If the coin flip at time rm is tails, we independently simulate $X_{(r+1)m}$ and $X^*_{(r+1)m}$ from $Q(X_{rm}, \cdot)$ and $Q(X^*_{rm}, \cdot)$, respectively. As before, when $m > 1$, we "condition in" $(X_{rm+1}, \ldots, X_{(r+1)m-1})$ and $(X^*_{rm+1}, \ldots, X^*_{(r+1)m-1})$ independently (conditional on $(X_{rm}, X_{(r+1)m})$ and $(X^*_{rm}, X^*_{(r+1)m})$, respectively). With this approach to jointly simulating X and X^*, we find that

$$P_x(T > nm) = (1 - \lambda)^n,$$

leading to part ii) of Theorem 8.13 (upon use of the coupling inequality).

Remark. As noted before, part ii.) of Theorem 8.13 provides a computable bound on the rate of convergence to stationarity A1. For reversible Markov chains, other bounds can be computed. In particular, one can apply Poincare's inequality and Cheeger's inequality to the study of rates of convergence for reversible chains. This leverages off the fact that eigenvalues for reversible chains can be elegantly characterized via a minimax argument. For details, see p.256-268 of J.S. Liu, Monte Carlo Strategies in Scientific Computing, Springer-Verlag (2003).

8.28 Recurrence of Markov Chains on General State Space

We have previously developed a fairly complete steady-state theory for discrete state space Markov chains. In particular, we established that when $X = (X_n : n \geq 0)$ is an irreducible Markov chain on discrete state space S, then X has a steady-state (in the sense of the law of large numbers) precisely when X has a stationary distribution. One problem with this theory is that we cannot establish stability without essentially solving the linear system $\pi = \pi P$.

In the last lecture, we showed that if a Markov chain satisfies condition A1, then is has a stationary distribution, satisfies a law of large numbers, and exhibits geometrically fast convergence to stationarity. In practice, condition A1 tends to be valid only for Markov chains taking values in a compact state space.

We now wish to develop recurrence theory that permits us to establish stability even for models for which the stationary equations cannot be explicitly solved. Ideally, this theory will also permit us to verify stability for Markov chains evolving in non-compact state spaces.

We start with the following generalization of condition A1:

A2 There exists $A \subseteq S$, $\lambda > 0$, $m \geq 1$, and a distribution φ on S such that

(a) $P_x(T_A < \infty) = 1$, $x \in S$, where $T_A = \inf\{n \geq 0 : X_n \in A\}$

(b) $P_x(X_m \in \cdot) \geq \lambda\, \varphi(\cdot)$, $x \in A$

Remark. If $A = S$, A2 is just condition A1. So A1 generalizes to A2.

Remark. If a Markov chain $X = (X_n : n \geq 0)$ satisfies A2, then the Markov chain is said to be *recurrent in the sense of Harris.*

Remark. If S is discrete, then X satisfies A2 if and only if there exists $z \in S$ such that $P_x(\tau(z) < \infty) = 1$ for $x \in S$, where $\tau(z) = \inf\{n \geq 1 : X_n = z\}$.

Note that condition i.) guarantees that X visits A infinitely often a.s. Using the same coin flip idea as for A1, condition ii.) ensures that every time X visits A (not having visited A in the previous m time steps), there is a probability of λ of a successful coin toss (i.e. a "heads"). It follow that there will be infinitely many times $T_1, T_2, \cdots$ at which X has distribution φ. Each such random time T_i initiates a cycle having an identical distribution. Furthermore, X_{T_i-m} is independent of X_{T_i}. Hence, if $m = 1$, the resulting cycles are iid, where if $m > 1$, the cycles are correlated. However, then degree of correlation is modest, and the following theorem is easily established.

Theorem 8.14. *Let $X = (X_n : n \geq 0)$ satisfy A2. Put $\tau_i = T_i - T_{i-1}$. If $\mathrm{E}[\tau]_1 < \infty$, then X possesses a unique stationary distribution π and for each non-negative f,*

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_1) \longrightarrow \int_S \pi(dx) f(x) \;\; a.s$$

as $n \to \infty$.

Remark. A sufficient condition guaranteeing $\mathrm{E}[\tau]_1 < \infty$ is to require that

$$\sup_{x\in A} \mathrm{E}_x[\tau_A] < \infty, \tag{8.47}$$

where $\tau_A = \inf\{n \geq 1 : X_n \in A\}$.
To illustrate what is involved in verifying A2 ii.), suppose that

$$P_x(X_m \in B) = \int_B p_m(x,y)\xi(dy) \tag{8.48}$$

for some distribution $\chi(\cdot)$ and transition density p_m. Suppose that

$$\varphi(B) = \int_B \phi(y)\xi(dy)$$

for some density $\phi(\cdot)$. (The Radon-Nikodym theorem actually guarantees that φ must take this form in the presence of A2 ii.).) If $p_m(\cdot, y)$ is continuous and positive, with A compact, then we can take

$$\begin{aligned} \phi(y) &= \inf_{x\in A} p_m(x,y), \\ \lambda &= \int_S \inf_{x\in A} p_m(x,z)\xi(dz). \end{aligned}$$

Hence in some generality, it follows that any compact set A satisfies A2 ii.). (Caution: The above analysis assumes that the m-step transition probabilities $P_x(X_m \in \cdot)$ can be represented as in (8.48), with a transition density that is positive and continuous. This must be verified separately for each example. If this fails to be true, one must verify A2 ii.) from "first principles".)

8.29 Stochastic Lyapunov Functions

It remains to provide a technique for verifying A2 i.) and condition (8.47).

Proposition 8.7: Let $w : S \to [1, \infty)$ for which there exists $r \geq 1$ such that

$$\mathrm{E}_x[w(X_1)] \leq r w(x)$$

for $x \in A$. Then,

$$\mathrm{E}_x[\tau_A] \leq (1-r)^{-1} w(x)$$

for $x \in A^c$.

Proof. Via use of operator/function norms with weight function $w(\cdot)$, we conclude that

$$\mathrm{E}_x\left[\sum_{j=0}^{T_A-1} w(X_j)\right] \leq (1-r)^{-1}w(x)$$

for $x \in A^c$. But since $w(x) \geq 1$ for $x \in S$,

$$\tau_A \leq \sum_{j=0}^{\tau_A-1} w(X_j),$$

yielding the result. □

Using "first transition" analysis, we see that for $x \in A$,

$$\mathrm{E}_x[\tau_A] = 1 + \int_{A^c} P(x,dy)\,\mathrm{E}_y[\tau_A].$$

In view of proposition 13.1, we conclude that

$$\sup_{x\in A}\mathrm{E}_x[\tau_A] \leq 1 + (1-r)^{-1}\sup_{x\in A}\mathrm{E}_x[w(X_1)].$$

Corallary 8.2: Suppose that there exists $A \subseteq S$, $\lambda > 0$, $m \geq 1$, a distribution φ, $r < 1$, and $w : S \to [1,\infty)$ such that:

[i.)]

1. $\mathrm{E}_x[w(X_1)] \leq rw(x)$, $x \in A^c$
2. $\sup_{x\in A}\mathrm{E}_x[w(X_1)] < \infty$
3. $\mathrm{P}_x\{X_m \in \cdot\} \geq \lambda\varphi(\cdot)$, $x \in A$.

Then, X possesses a unique stationary distribution π and for each non-negative f,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \longrightarrow \int_S \pi(dx)f(x) \quad \text{a.s}$$

as $n \to \infty$.
Remark. The function $w(\cdot)$ is called a *stochastic Lyapunov function.*

Example 8.27: Suppose that $(S_n : n \geq 0)$ is a sequence of rv's describing reservoir storage in the presence of a linear release rule, so that

$$S_{n+1} = S_n + Z_{n+1} - aS_{n+1}$$

for $a > 0$. Assume that $(Z_n : n \geq 1)$ is a sequence of iid non-negative rv's for which $\mathrm{E}[Z_1] < \infty$. Put $w(x) = 1 + x$ for $x \geq 0$. Then,

$$\begin{aligned}
\mathrm{E}_x[w(S_1)] &= 1 + \mathrm{E}_x[S_1] \\
&= 1 + \frac{\mathrm{E}[x+Z_1]}{1+a} \\
&\leq \left(1+\frac{a}{2}\right)^{-1} w(x) + 1 + \frac{\mathrm{E}[Z_1]}{(1+a)} - ax(1+a)^{-1}(2+a)^{-1} \\
&\leq \left(1+\frac{a}{2}\right)^{-1} w(x)
\end{aligned}$$

for $x \geq (1+a)(2+a)/a + \mathrm{E}[Z]_1(2+a)/a$. Put $A = [0, (1+a)(2+a)/a + \mathrm{E}[Z_1](2+a)/a]$. Condition ii.) of Corallary 8.2 is easily verified here. Hence, it remains only to show that condition iii.) is satisfied. But this is straightforward to carry out if we assume that Z_1 has a continuous positive density on $[0,\infty)$; see Section 1.

A weaker Lyapunov condition is offered by our next result

Theorem 8.15. *Suppose that there exists $A \subseteq S$, $\lambda > 0$, $m \geq 1$, a distribution φ, $\varepsilon > 0$, and $g : S \to [0, \infty)$ such that:*

1. $\mathrm{E}_x [g(X_1)] \leq g(x) - \varepsilon$, $x \in A^c$
2. $\sup_{x \in A} \mathrm{E}_x [g(X_1)] < \infty$
3. $\mathrm{P}_x \{X_m \in \cdot\} \geq \lambda \varphi(\cdot)$, $x \in A$.

Then, X possesses a unique stationary distribution π and for each non-negative f,

$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \longrightarrow \int_S f(x) \pi(dx) \quad a.s.$$

as $n \to \infty$.

Proof. Let $B = (B(x, dy) : x, y \in A^c)$ be the restriction of $P = (P(x, dy) : x, y \in S)$ to A^c, so that $B(x, dy) = P(x, dy)$ for $x, y \in A^c$. Put $e(x) = 1$ for $x \in A^c$. Note that condition i.) implies that

$$\begin{aligned} \int_{A^c} B(x, dy) g(y) &= \int_{A^c} P(x, dy) g(y) \\ &\leq \int_S P(x, dy) g(y) \\ &= \mathrm{E}_x [s(X_1)] \\ &\leq g(x) - \varepsilon e(x), \end{aligned}$$

so that

$$Bg \leq g - \varepsilon e.$$

Hence,

$$\varepsilon e \leq g - Bg.$$

Since B is a non-negative operator, it follows that

$$\varepsilon B^j e \leq B^j g - B^{j+1} g.$$

Summing over $j \in \{0, 1, \cdots, n\}$, we get

$$\varepsilon \sum_{j=0}^{n} B^j e \leq g - B^{n+1} g \leq g.$$

Sending $n \to \infty$, we conclude that

$$\varepsilon \sum_{j=0}^{n} B^j e \leq g.$$

But

$$(B^j e)(x) = P_x(\tau_A > j),$$

so

$$\sum_{i=0}^{\infty} (B^j e)(x) = \sum_{j=0}^{\infty} P_x(\tau_A > j) = \mathrm{E}_x [\tau_A],$$

yielding the inequality

$$\mathrm{E}_x [\tau_A] \leq g(x)/\varepsilon$$

for $x \in A^c$. Consequently, $\mathrm{P}_x \{\tau_A < \infty\} = 1$ for $x \in A^c$ (and hence for all $x \in S$). Furthermore,

$$\sup_{x \in A} \mathrm{E}_x [\tau_A] \leq 1 + \sup_{x \in A} \mathrm{E}_x [g(X_1)].$$

The conclusion of the theorem therefore follows as for Corallary 8.2. □

Remark. The function g appearing in Theorem 8.15 is also known as a stochastic Lyapunov function. Note that any weight function $w(\cdot)$ satisfying proposition 13.1 automatically satisfies condition i.) of Theorem 8.15 (put $\varepsilon = 1 - r$) but not vice versa. So, the Lyapunov function of Theorem 13.2 legitimately generalizes the Lyapunov function of Corallary 8.2.

Remark. A nice physical way to think about g is to view $g(x)$ as representing the "potential energy" associated with x. Condition i.) asserts that, in expectation, the potential energy has a tendency to decrease by ε on A^c. Hence, since the system wishes to move to points of lower potential energy, it follows that the system should eventually enter A.

Remark. A high level of ingenuity may be required to find a suitable Lyapunov function (satisfying i.) and ii.) of Theorem 8.15). One approach to finding a suitable g is to try some candidate functions. When $S \subseteq \mathbb{R}^d$, typical candidates to try are:

1. $\|x\|^p$
2. $\exp(a\|x\|^p)$
3. $(\log(1 + \|x\|))^p$.